Research Article

Check for updates

Nutriomics and artificial intelligence nutrition obesity cohort (NAINOC): a design paper for a prospective cohort for nutrition and obesity research

*Correspondence:

Sungha Park
Division of Cardiology, Department of Internal Medicine, Severance Hospital, Yonsei University College of Medicine, 50-1 Yonsei-ro, Seodaemun-gu, Seoul 03722, Republic of Korea.
Email: shpark0530@yuhs.ac

Jae-Ho Park
Food Functionality Research Division, Korea Food Research Institute, 245 Nongsaengmyeong-ro, Iseo-myeon, Wanju-gun 55365, Republic of Korea.
Email: jaehopark@kfri.re.kr

ORCID iDs
Minyoung Lee
https://orcid.org/0000-0002-9333-7512
Sungha Park
https://orcid.org/0000-0001-5362-478X
Soo-Hyun Park
https://orcid.org/0000-0003-0968-558X

Minyoung Lee [1,2], **Sungha Park** [3,4*], **Soo-Hyun Park** [5], **Ho-Young Park** [5], **Yu Ra Lee** [5], **Min-Sun Kim** [6], **Miso Nam** [6], **Jangho Lee** [5], **Hyein Seo** [7], **Yong-ho Lee** [1,2], **Chan Joo Lee** [4], **Jae-Ho Park** [5*], **Hye Hyun Yoo** [8], **Hyun-Jin Kim** [9], **Kyong-Oh Shin** [10], **Yoshikazu Uchida** [10], and **Kyungho Park** [10]

[1]Division of Endocrinology and Metabolism, Department of Internal Medicine, Yonsei University College of Medicine, Seoul, Republic of Korea
[2]Institute of Endocrine Research, Yonsei University College of Medicine, Seoul, Republic of Korea
[3]Integrative Research Center for Cerebrovascular and Cardiovascular diseases, Yonsei University College of Medicine, Seoul, Republic of Korea
[4]Division of Cardiology, Department of Internal Medicine, Severance Hospital, Yonsei University College of Medicine, Seoul, Republic of Korea
[5]Food Functionality Research Division, Korea Food Research Institute, Wanju, Republic of Korea
[6]Food Industry Research Division, Korea Food Research Institute, Wanju, Republic of Korea
[7]Intelligence Policy Team, Korea Food Research Institute, Wanju, Republic of Korea
[8]Pharmacomicrobiomics Research Center, College of Pharmacy, Hanyang University, Ansan, Republic of Korea
[9]Division of Applied Life Science (BK21 Plus), Department of Food Science & Technology, and Institute of Agriculture and Life Science, Gyeongsang National University, Jinju, Republic of Korea
[10]Department of Food Science and Nutrition, and Convergence Program of Material Science for Medicine and Pharmaceutics, Hallym University, Chuncheon, Republic of Korea

ABSTRACT

Background: The increase in obesity is becoming a world-wide health issue. However, no prospective cohorts in East Asia have thoroughly explored comprehensive nutritional and multiomic data in individuals with obesity. This study is designed to establish an obesity cohort that constitutes clinical characteristics, nutritional status, laboratory profiles, metabolic complication studies, and multiomic profiles with the goal of artificial intelligence platform-based nutriomic analysis.

Methods: This study aims to enroll at least 400 obese adults (aged ≥ 19 years; body mass index ≥ 25 kg/m²) and 100 non-obese adults as controls. Obese participants have to have at least one of the following chronic metabolic diseases: hypertension, type 2 diabetes mellitus, cardiovascular disease, and metabolic syndrome. Participants will undergo assessment for demographic data, clinical, lifestyle, and dietary assessments, laboratory examination, coronary calcium/visceral fat scan, liver fibroscan, carotid ultrasound, and continuous glucose monitoring. Metabolite analysis will be conducted for blood/stool/urine/saliva samples. Deoxyribonucleic acid methylation analysis, peptidomic analysis, and lipidomic analysis will be performed on blood samples. Obese individuals will have annual study visits for collection of clinical measures and multiomics data over a 5-year period. Control individuals will have a baseline hospital visit with annual telephone follow-up for clinical event monitoring.

Ho-Young Park [ID]
https://orcid.org/0000-0002-9966-9059
Yu Ra Lee [ID]
https://orcid.org/0000-0002-1258-5363
Min-Sun Kim [ID]
https://orcid.org/0009-0002-9681-6006
Miso Nam [ID]
https://orcid.org/0000-0001-8202-9541
Jangho Lee [ID]
https://orcid.org/0000-0002-5604-821X
Hyein Seo [ID]
https://orcid.org/0000-0002-5722-7957
Yong-ho Lee [ID]
https://orcid.org/0000-0002-6219-4942
Chan Joo Lee [ID]
https://orcid.org/0000-0002-8756-409X
Jae-Ho Park [ID]
https://orcid.org/0000-0003-4428-436X
Hye Hyun Yoo [ID]
https://orcid.org/0000-0001-8282-852X
Hyun-Jin Kim [ID]
https://orcid.org/0000-0002-7284-3547
Kyong-Oh Shin [ID]
https://orcid.org/0000-0001-8568-8905
Yoshikazu Uchida [ID]
https://orcid.org/0000-0002-4980-3416
Kyungho Park [ID]
https://orcid.org/0000-0002-1552-9914

Abbreviations

A.U., arbitrary unit; ACN, acetonitrile; AI, artificial intelligence; BMI, body mass index; BP, blood pressure; CAC, coronary artery calcium; CAD, collision gas; CT, computed tomography; CUR, curtain gas; CVD, cardiovascular disease; D2O, deuterium oxide; DM, diabetes mellitus; DMR, differentially methylated region; DNA, deoxyribonucleic acid; FA, formic acid; FFQ, food frequency questionnaire; GS1, Gas 1; GS2, Gas 2; HbA1c, glycated hemoglobin A1c; HOMA-IR, homeostatic model assessment of insulin resistance; HOMA-β, homeostasis model assessment of β-cell function; IMT, intima-media thickness; LC, liquid chromatography; MeOH, methanol; MRM, multiple reaction monitoring; MS, mass spectrometer; MS/MS, tandem mass spectrometry; NAINOC, Nutriomics and Artificial Intelligence Nutrition Obesity Cohort; NMR, nuclear magnetic resonace; QC, quality control; SFA, subcutaneous fat area; SNP, single nucleotide polymorphism; TEM, temperature; VFA, visceral fat area; WHOQOL-BREF, World Health Organization Quality of Life Brief Version.

Conclusions: The strength of this cohort will be as follows. First, the cohort will enable the integration of nutritional intake data with other multiomics data for a comprehensive analysis. Second, inclusion of both obese individuals with various metabolic traits and non-obese individuals as controls is advantageous for studying a wide range of obesity phenotypes in comparison with non-obese conditions. Third, diverse modalities to assess metabolic and complication status will facilitate multifaceted analysis. Lastly, beyond the typical blood and stool samples in multiomic studies, the inclusion of urine, saliva, and skin samples will further refine obesity characterization.

Keywords: Nutriomics; Multiomics; Nutrition; Obesity; Cohort

BACKGROUND

Obesity is a chronic disease marked by excess adiposity and rising at epidemic proportions, leading to heightened comorbidity and mortality risks [1]. According to the WHO, from 1975 to 2016, the prevalence of obesity has increased by 3 folds and in 2017, the prevalence of obesity in the US was reported to be 42% and by 2030, 78% of US adults are projected to be either overweight or obese [1-3]. The increase in obesity is becoming a world-wide health issue of epidemic proportion. For example, in South Korea, the prevalence of obesity, defined as body mass index (BMI) of more than 25 kg/m², has increased from 29.7% in 2009 to 38.4% in 2021 [4]. Obesity is associated with various health problems such as type 2 diabetes mellitus (DM), dyslipidemia, hypertension, cerebrovascular disease, cancer, respiratory disorders, obstructive sleep apnea, and degenerative arthritis [1,5,6]. Due to higher risk of comorbidities, obesity is associated with shortening of life expectancy by 2–4 years in those with BMI of 30–35 kg/m² and shortened life expectancy of 8–10 years in those with BMI of 40–45 kg/m² [5]. The major cause of mortality in subjects with obesity is cardiovascular disease (CVD), which has been shown to increase progressively with increasing BMI with no significant racial difference [5,6].

Although numerous cohort studies have demonstrated the association of obesity with many chronic diseases, there are currently no prospective cohort studies in East Asia that explored comprehensive nutritional and multiomic data in individuals with obesity. Multiomic approaches have gained significant attention for elucidating the mechanisms of chronic diseases and advancing treatment and clinical applications [7,8]. Dietary factors have been recognized as crucial in the development of chronic diseases, and their integration into multiomic analyses has become increasingly important since environmental contributions were first suggested in 1967 [9]. Recently, nutriomic approaches that combine multiomics and nutritional evaluation have revealed novel features of cardiometabolic disease in middle-aged Europeans [10]. Based on these backgrounds, the objective of this study is to establish an obesity cohort that constitutes clinical and nutritional characteristics, blood chemistry profiles, metabolic complication studies, and multiomic profiles with the goal of nutriomic analysis [11]. We plan to obtain genomic and epigenomic data, microbiome metagenomic data, and metabolomic data from blood, feces, urine, saliva, and skin in obese and non-obese subjects. Learning-based data reduction algorithms, a subset of artificial intelligence (AI), will be used to identify genomic traits that contribute to features and comorbidities of obesity. This cohort study will afford the classification of obesity phenotypes through an AI platform, facilitating personalized dietary strategies for the treatment and prevention of obesity and frequently comorbid chronic metabolic diseases.

Authors' contributions
Conceptualization: Lee M, Park SH[1], Park JH; Data curation: Lee M, Park SH[1], Park SH[2], Park JH; Formal analysis: Lee M, Lee J; Funding acquisition: Park SH[1]; Investigation: Lee M, Park SH[1], Park SH[2], Park HY, Lee YR, Kim MS, Nam M, Lee YH, Lee CJ, Park JH, Yoo HH, Kim HJ, Shin KO; Methodology: Lee M, Park SH[1], Park SH[2], Park HY, Lee YR, Kim MS, Nam M, Lee J, Seo H, Lee YH, Lee CJ, Park JH, Yoo HH, Kim HJ, Shin KO, Uchida Y, Park K; Project administration: Park SH[1], Park JH; Resources: Park SH[1], Park JH; Software: Lee YR, Kim MS, Nam M, Lee J, Seo H, Kim HJ, Shin KO, Park K;

METHODS

This study was approved by the Institutional Review Board of Severance Hospital, Seoul, South Korea (4-2022-0645). This study will be performed as a collaborative study between the Korea Food Research Institute and Yonsei University College of Medicine, Severance hospital. This study is originally designed to follow each participant for five years after enrollment; however, extended follow-up may be considered depending on factors such as participant retention, the sustainability of the study, and the significance of the findings.

The study aims to enroll at least 400 obese adults aged ≥ 19 years with BMI ≥ 25 kg/m^2 who have at least one or more chronic diseases such as hypertension, type 2 DM, and CVD. The obesity cohort is to be followed up every year for 5 years. One hundred control subjects are to be enrolled over the five-year study period for comparison with the obese subjects. Over the entire study period, the cohort will comprise 500 participants with complete dietary, clinical, and multi-omics data collected at multiple time points. The sample size was informed by prior multi-omic cohort studies and literature on sample size estimation for epigenomics and metabolomics [12-15].

Inclusion criteria for individuals with obesity
1. Adults aged ≥ 19 years and with BMI ≥ 25 kg/m^2

2. The subject has to have at least one of the following chronic metabolic diseases: hypertension, type 2 DM, CVD, and metabolic syndrome.

Obesity is categorized into class I (25.0–29.9), class II (30.0–34.9), and class III (≥ 35.0) according to BMI and the WHO guidelines for the Asia-Pacific region [16]. Type 2 DM is defined by either the diagnostic criteria of Korean Diabetes Association guidelines [16], a previous diagnosis by a physician, or the use of glucose-lowering drugs. Prediabetes is defined by fasting glucose levels of 100–125 mg/dL, 2-hour glucose of 140–199 mg/dL after a 75 g glucose load or meal, or glycated hemoglobin A1c (HbA1c) levels of 5.7–6.4% [16]. Definition of metabolic syndrome is based on previously established criteria [17].

Exclusion criteria for the obese group
1. Life expectancy of less than 6 months due to severe non-CVD
2. Pregnancy, suspected pregnancy, or lactation
3. Within 3 months of organ transplantation
4. Currently being treated for acute rejection of organ transplantation
5. Within 6 months of discharge after treatment for acute coronary syndrome (myocardial infarction or unstable angina)
6. Within 6 months of discharge after hospitalization for acute ischemic stroke
7. Type 1 DM
8. Currently receiving steroids or hormone replacement therapy
9. Inability to use a smartphone

Inclusion criteria for individuals without obesity as control group
Adult subjects aged ≥ 19 years with BMI between 18.5–24.9 kg/m^2. The patient may have hypertension or dyslipidemia but should not have other chronic metabolic diseases such as type 2 DM and CVD. Based on the 2019–2021 Korea National Health and Nutrition Examination Survey, conducted among adults aged ≥ 19 years with a mean BMI of 23 kg/m^2 representing

Supervision: Park SH[1]; Validation: Lee M, Park SH[1], Park SH[2], Park HY, Nam M, Seo H, Park JH, Kim HJ, Park K; Visualization: Lee M; Writing - original draft: Lee M, Park SH[1], Park SH[2], Park HY, Kim MS, Lee J, Seo H, Park JH, Kim HJ, Park K; Writing - review & editing: Lee YR, Nam M, Lee YH, Lee CJ, Yoo HH, Shin KO, Uchida Y.

Park SH[1], Sung-Ha Park; Park SH[2], Soo-Hyun Park.

a predominantly non-obese population, the prevalence of hypertension alone, dyslipidemia alone, and their comorbidity was 8.7%, 24.6%, and 15.0%, respectively [18]. Therefore, excluding individuals with hypertension and dyslipidemia would substantially limit the pool of eligible participants for the non-obese control group. Accordingly, hypertension and dyslipidemia were permitted as comorbid conditions in the control group. Nevertheless, the presence of hypertension and dyslipidemia in the control group may influence the risk of developing CVD and other chronic metabolic diseases within this group [19,20]. Subgroup analyses restricted to control group participants without hypertension and dyslipidemia may be considered to account for the potential confounding effects of these conditions.

Exclusion criteria for the control group
1. Life expectancy of less than 6 months due to severe non-CVD
2. Pregnancy, suspected pregnancy, or lactation
3. Current use of physician-prescribed medications for chronic diseases other than hypertension or dyslipidemia
4. Diagnosis of malignant tumor within the past 5 years
5. Inability to use a smartphone

Cohort protocol
The research protocol for each year is summarized in **Table 1**.

1. Baseline demographic data: age, gender, medical history, medication history, gynecological history, and family history

2. Blood pressure (BP) measurement: after 5 minutes of rest in a sitting position, office BP will be taken in the research examination room by a trained nurse using a validated automated device (HEM 7080-IC; Omron, Kyoto, Japan). After positioning the subjects

Table 1. Clinical research timeline

Variables	Year 1[a]	Year 2	Year 3	Year 4	Year 5
Baseline demographic data collection	•				
Vital signs measurement, physical examination	•	•	•	•	•
Body composition analysis	•	•	•	•	•
Lifestyle assessment (smoking, alcohol consumption, sleep, and physical activity)	•				
Quality of life assessment (WHOQOL-BREF)	•				
Nutritional assessment (FFQ)[b]	•	•	•	•	•
14-day dietary intake diary assessment	•	•	•	•	•
14-day dietary intake photography collection	•	•	•	•	•
Laboratory test	•	•	•	•	•
Human-derived sample collection	•	•	•	•	•
Coronary calcium scan, abdominal CT[c]		•			
Liver fibroscan	•				•
Carotid ultrasound[d]		•			
14-day continuous glucose monitoring[e]		•		•	
Cardiovascular event adjudication		•	•	•	•

WHOQOL-BREF, world health organization quality of life - brief version; FFQ, food frequency questionnaire; CT, computed tomography.
[a]Participants in the obese group are scheduled to visit the hospital once a year from the time of enrollment until the fifth year. For participants in the control group, only the first-year hospital visit is required, followed by annual telephone follow-up for clinical event monitoring.
[b]All participants will complete the FFQ developed for targeting metabolic diseases (Met-FFQ). During the first year, some voluntary participants will also complete the FFQ developed as part of the Korea Genome and Epidemiology Study.
[c]Although the assessment is primarily scheduled to be performed within the calendar year of study enrollment, it may be conducted in the following year if there are scheduling difficulties or limited scanner availability, provided that the scan is not substantially delayed beyond the one-year mark.
[d]This test is limited to obese participants and will be conducted only during the second year.
[e]Among obese participants, the test will be conducted for those with diabetes during the second year and for those with diabetes or prediabetes during the fourth year.

and setting the device, the average of three BP readings taken at 1-minute intervals after 5 minutes of rest will be used in this study using the nocturnal automatic BP measurement mode [21,22].

3. Body composition using Inbody 770 (InBody Co., Seoul, Korea).

4. Questionnaire for lifestyle assessment (smoking, alcohol consumption, sleep, and physical activity): information on lifestyle factors, including smoking, alcohol consumption, sleep, and physical activity, will be collected through standardized questionnaires (**Table 2**).

Table 2. Questionnaire for lifestyle assessment (smoking, alcohol consumption, sleep, and physical activity)

No.	Question	Unit	Response options
Smoking			
1.	Have you smoked a total of 5 packs (100 cigarettes) or more in your lifetime?		No/Yes/Don't know
1-1.	Do you currently smoke?		No/Yes/Don't know
1-2.	At what age did you start smoking?	Years	
1-2-1.	Don't know age started		Don't know
1-3.	For how many years have you smoked in total?	Years	
1-3-1.	Don't know total smoking duration	Years	Don't know
1-3-2.	For how many months have you smoked in total?	Months	
1-4.	On average, how many cigarettes do/did you smoke per day?	Cigarettes (approx.)	
1-4-1.	Don't know number of cigarettes per day		Don't know
Alcohol consumption			
2.	Do you abstain from alcohol or have never consumed it (e.g., for religious reasons)?		Do not drink/Drink/Don't know
2-1.	Have you quit drinking?		No/Yes
2-1-1.	If yes, since when?	Years	
2-1-2.	If yes, since when?	Months	
2-1-3.	Don't know when quit		Don't know
2-2.	Frequency of Makgeolli (rice wine) consumption		Never/Once a month/2–3 times a month/Once a week/2–3 times a week/4–6 times a week/Once a day/Twice or more per day
2-2-1.	Average amount of Makgeolli consumed	Makgeolli cups (250 mL)	
2-3.	Frequency of Cheongju (refined rice wine) consumption		Never/Once a month/2–3 times a month/Once a week/2–3 times a week/4–6 times a week/Once a day/Twice or more per day
2-3-1.	Average amount of Cheongju consumed	Soju glasses (50 mL)	
2-4.	Frequency of wine consumption		Never/Once a month/2–3 times a month/Once a week/2–3 times a week/4–6 times a week/Once a day/Twice or more per day
2-4-1.	Average amount of wine consumed	Wine glasses (90 mL)	
2-5.	Frequency of soju consumption		Never/Once a month/2–3 times a month/Once a week/2–3 times a week/4–6 times a week/Once a day/Twice or more per day
2-5-1.	Average amount of soju consumed	Soju glasses (50 mL)	
2-6.	Frequency of beer consumption		Never/Once a month/2–3 times a month/Once a week/2–3 times a week/4–6 times a week/Once a day/Twice or more per day
2-6-1.	Average amount of beer consumed	Beer glasses (250 mL)	
2-7.	Frequency of liquor/spirits (Western alcohol) consumption		Never/Once a month/2–3 times a month/Once a week/2–3 times a week/4–6 times a week/Once a day/Twice or more per day
2-7-1.	Average amount of liquor consumed	Liquor glasses (30 mL)	
2-8.	Other types of alcohol (please specify)		
2-8-1.	Frequency of other alcohol consumption		Never/Once a month/2–3 times a month/Once a week/2–3 times a week/4–6 times a week/Once a day/Twice or more per day
2-8-2.	Average amount of other alcohol consumed	Other glass	
Sleep			
3.	Sleep and wake-up time		
3-1.	What time do you usually go to bed?	Hours	
3-1-1.	Don't know bedtime		Don't know
3-2.	What time do you usually go to bed?	Minutes	
3-3.	What time do you usually wake up?	Hours	
3-3-1.	Don't know wake-up time		Don't know

(continued to the next page)

Table 2. (Continued) Questionnaire for lifestyle assessment (smoking, alcohol consumption, sleep, and physical activity)

No.	Question	Unit	Response options
3-4.	What time do you usually wake up?	Minutes	
4.	Sleep duration		
4-1.	Average sleep duration per day	Hours	
4-1-1.	Don't know average sleep duration		Don't know
4-2.	Average sleep duration per day	Minutes	
5.	Do you currently have difficulty falling asleep?		No/Occasionally (1–2 days/week)/Often (3–4 days/week)/Almost daily (≥5 days/week)/Don't know
6.	Do you currently have difficulty falling back asleep after waking during the night?		No/Occasionally (1–2 days/week)/Often (3–4 days/week)/Almost daily (≥5 days/week)/Don't know
7.	Snoring		
7-1.	Do you snore?		No/Yes/Don't know
7-1-1.	How loud is your snoring?		Slightly louder than breathing/As loud as speaking/Louder than speaking/Very loud (audible from other rooms)
7-1-2.	How often do you snore?		Almost daily/3–4 times per week/1–2 times per week/1–2 times per month/Rarely or never
7-1-3.	Have you ever bothered others with your snoring?		No/Yes/Don't know
7-2.	Has anyone ever noticed that you stop breathing while sleeping?		Almost daily/3–4 times per week/1–2 times per week/1–2 times per month/Rarely or never
7-3.	How often do you feel tired after sleeping?		Almost daily/3–4 times per week/1–2 times per week/1–2 times per month/Rarely or never
7-4.	Do you feel tired while awake?		Almost daily/3–4 times per week/1–2 times per week/1–2 times per month/Rarely or never
7-5.	Have you ever fallen asleep while driving?		Almost daily/3–4 times per week/1–2 times per week/1–2 times per month/Rarely or never
Physical activity			
8.	Compared to usual, how was your level of physical activity during the past 7 days?		Similar to usual/More active than usual/Less active than usual
9.	During the past 7 days, on how many days did you engage in vigorous physical activities (e.g., carrying heavy objects, running, aerobics)?		Days per week/None
9-1.	Number of days of vigorous physical activity	Days (per week)	
9-2.	On such days, how many hours did you usually spend doing vigorous physical activity?	Hours (per day)	
9-2-1.	On such days, how many hours did you usually spend doing vigorous physical activity?		Don't know/Not sure
9-3.	On such days, how many minutes did you usually spend doing vigorous physical activity?	Minutes (per day)	
10.	During the past 7 days, on how many days did you engage in moderate physical activities (e.g., carrying light loads, biking at a regular pace, doubles tennis)? Do not include walking.		Days per week/None
10-1.	Number of days of moderate physical activity	Days (per week)	
10-2.	On such days, how many hours did you usually spend doing moderate physical activity?	Hours (per day)	
10-2-1.	On such days, how many hours did you usually spend doing moderate physical activity?		Don't know/Not sure
10-3.	On such days, how many minutes did you usually spend doing moderate physical activity?	Minutes (per day)	
11.	During the past 7 days, on how many days did you walk for at least 10 min at a time?		Days per week/None
11-1.	Number of days walked at least 10 min	Days (per week)	
11-2.	Total hours spent walking (≥ 10 min per episode) over the past 7 days	Hours (per week)	
11-2-1.	Total hours spent walking (≥ 10 min per episode) over the past 7 days		Don't know/Not sure
11-3.	Total minutes spent walking (≥ 10 min per episode) over the past 7 days	Minutes (per week)	
12.	During the past 7 days, how many hours did you usually spend sitting on a weekday?		Hours per day/Not sure
12-1.	Average sitting time on weekdays	Hours (per day)	
12-1-1.	Average sitting time on weekdays		Don't know/Not sure
12-2.	Average sitting time on weekdays	Minutes (per day)	

5. Questionnaire for quality of life: quality of life will be evaluated through World Health Organization Quality of Life Brief Version (WHOQOL-BREF), which covers four domains: physical health, psychological aspects, social relationships, and environment [23].

6. Questionnaire for dietary intakes: information on subjects' dietary intake will be collected using two methods. First, food frequency questionnaire (FFQ) specifically designed for metabolic diseases will be used to assess dietary intake over the course of one year [24]. This semi-quantitative FFQ contains 129 items in 12 categories reflecting the dietary characteristics of obese Korean adults [24]. Additionally, voluntary participants will complete another FFQ consisting of 103 items, only once. The 103-FFQ, developed as part of the Korea Genome and Epidemiology Study, encompasses a broader range of dietary categories [25]. Second, dietary intake will be collected using dietary record application for 14 days. Subjects will upload photos of the food they consume during that period via the application, and automatic recognition will collect information on the food items and their nutritional information. For subjects who have difficulty in accessing the application, paper-based food records, along with food images, will be collected instead.

7. Laboratory test: complete blood count, routine serum chemistry, HbA1c, gamma-glutamyl transferase, C-reactive protein, 25(OH) vitamin D, total vitamin D, free fatty acid, fibrinogen, adiponectin, leptin, and routine urinalysis will be measured yearly. Glucose and insulin levels for both fasting and postprandial states are tested only for patients attending the diabetes center.

8. Human-derived sample collection: blood (10 mL, EDTA tube including anticoagulant, BD, Belliver Industrial Estate, Plymouth, UK), serum (0.5 mL, cryotube aliquoted into 10 vials, AXYGEN, Corning, Shanghai, China), stool (20 g, 1 NGS sample kit, OMNIgeneGut, Genotek, Ottawa, Canada), urine (10 mL, 15 mL conical tube, SPL Life Sciences Co., Ltd, Pocheon, Korea), saliva (3 mL, 15 mL conical tube, SPL Life Sciences Co., Ltd), and skin (×3, D-Square tape, CuDerm Corporatin, Dallas, USA) samples will be collected and stored at −70°C for future analysis.

9. Coronary calcium scan and abdominal fat computed tomography (CT): coronary artery calcium (CAC) scan will be performed with the latest generation of 256-slice CT scanner (Revolution, GE Healthcare, Waukesha, USA) for all study participants, which consists of standard scanning with 120 kV as described previously [26].

   Subcutaneous fat area (SFA) and visceral fat area (VFA) is quantified using an abdominal fat CT scan (Siemens somatom sensation 64, Siemens healthcare Gmbh, Erlangen, Germany). A single 3-mm-thick axial slice at the midpoint of the L4 vertebra is obtained with participants in the supine position. Fat measurement is carried out using the TeraRecon Aquarius workstation (Aquarius iNtuition, version 4.4.6; TeraRecon, Foster City, USA). SFA is defined as adipose tissue within an attenuation range of −190 to −30 Hounsfield units, and VFA as tissue between −150 and −50 Hounsfield units. VFA refers to fat within the abdominal cavity, bounded by the abdominal and oblique muscles and the posterior surface of the vertebral body. Fat located between the musculature and skin is classified as SFA. Measurements are expressed in cm^2.

10. Liver fibroscan: liver fibroscan will be conducted at the first and the fifth year according to the method described previously [27]. Fibroscan will be done using one of three

models; Fibro scan 502, 530, and 630 (Echosens, Paris, France). Hepatic steatosis will be quantified using controlled attenuation parameter [28]. The presence of hepatic steatosis is defined as being ≥ S1. Hepatic fibrosis will be determined using liver stiffness value (kPa) and classified into stages F0 through F4, with stages ≥ F3 defined as advanced fibrosis [29].

11. Carotid intima-media thickness (IMT): baseline carotid IMT and plaque measurement will be performed as described previously using a high-resolution real time 8-MHz linear scanner (Acuson Juniper, Siemens Inc., Berlin, Germany) [30,31]. Carotid plaques will be defined as focal thickening encroaching into the lumen by ≥ 50% or absolute thickness of ≥ 1.5 mm at the common carotid artery, carotid bulb, or the internal carotid artery.

12. Continuous glucose monitoring: device for continuous glucose monitoring (FreeStyle Libre®; Abbott Diabetes Care, Alameda, CA, USA) will be applied for 14 consecutive days at the second year for individuals with diabetes and at the fourth year for individuals with diabetes and prediabetes. The FreeStyle Libre Pro® sensor will be worn on the posterior upper arm in accordance with the manufacturer's guidelines and using aseptic technique. Glucose management indicator (%), coefficient of variation, time in range (70–180 mg/dL, %), time above range (> 180 or 250 mg/dL, %), time below range (< 70 or 54 mg/dL, %) will be analyzed [32,33].

13. Comorbidities and medication use: comorbidities and medication use will be re-evaluated at least once every year throughout the cohort follow-up period, starting from baseline enrollment. Hospitalizations and emergency room visits due to illness or accidents since the last visit, including the reason and timing of each event, will be checked every year.

Information on the presence or absence of the following diseases at baseline, as well as the age at diagnosis, will be obtained: stroke, transient ischemic attack, myocardial infarction, angina, heart failure, chronic kidney disease, hypertension, dyslipidemia, DM, thyroid disease, fatty liver disease, chronic hepatitis, liver cirrhosis, asthma or chronic obstructive pulmonary disease, osteoporosis, arthritis, autoimmune disease, and malignant tumor.

The list of medications for which usage information is collected is summarized in **Table 3**. For each medication, the product name, ingredient name, dosage per administration, and frequency of daily use will be recorded.

14. Cardiovascular event and chronic metabolic disease adjudication: clinical event adjudication for newly developed stroke, transient ischemic attack, myocardial infarction, angina, heart failure, chronic kidney disease, hypertension, dyslipidemia, and DM will be assessed based on medical records or in-person/telephone interview. The timing

Table 3. List of concomitant medications

Drug category	Drugs
Antihypertensive agents	Thiazide diuretics, thiazide-like diuretics, calcium channel blockers, angiotensin-converting enzyme inhibitors, angiotensin II receptor blockers, beta-blockers, mineralocorticoid receptor antagonists, alpha-blockers, nitrates, hydralazine, minoxidil
Antiplatelet/Anticoagulant agents	Antiplatelet agents, warfarin, direct oral anticoagulants
Lipid-lowering agents	Statins, ezetimibe
Gastrointestinal agents	Proton pump inhibitors
Antidiabetic agents	Metformin, sulfonylureas, meglitinides, thiazolidinediones, dipeptidyl peptidase-4 inhibitors, sodium-glucose cotransporter 2 inhibitors, alpha-glucosidase inhibitors, glucagon-like peptide-1 receptor agonists, insulins
Anti-obesity agents	Orlistat, naltrexone/bupropion, phentermine/topiramate, liraglutide, semaglutide

of diagnosis and related medication use, will be also identified. In medical records, the documentation of each diagnosis and the prescription history of related medications will be verified. Two investigators will review the events independently for agreement. This investigation will be conducted annually, a timeframe considered to be appropriate based on the prospective studies that captured meaningful changes in CVD incidence and associated metabolic risk factors in a yearly assessment [34-36].

15. Metabolic indices: insulin resistance is assessed using homeostatic model assessment of insulin resistance (HOMA-IR), calculated as fasting insulin [μIU/mL] × fasting glucose [mg/dL])/405 [37]. Insulin secretory pancreatic β-cell function is assessed using homeostasis model assessment of β-cell function (HOMA-β), calculated as (360 × fasting insulin [μIU/mL])/(fasting glucose [mg/dL] − 63) [37]. Insulin resistance is defined by a HOMA-IR value > 2.5 (mg/dL*μIU/mL) [38]. A HOMA-β value < 50% is considered to indicate reduced insulin secretory function [39].

16. Weight loss interventions: the current study is designed as a cohort-based observational study rather than an interventional trial, and no formal nutritional counseling or exercise programs are provided to either the obese or control group. However, participants with chronic metabolic conditions such as hypertension, type 2 DM, CVD, or metabolic syndrome may receive general recommendations or guidance regarding weight loss and lifestyle modification as part of their routine clinical care [40-42]. In addition, some individuals with obesity may pursue active weight reduction with the use of anti-obesity medications, such as orlistat, naltrexone/bupropion, phentermine/topiramate, liraglutide, and semaglutide [43]. Certain antidiabetic agents, such as sodium-glucose cotransporter 2 inhibitors and glucagon-like peptide-1 receptor agonists, are also known to promote weight loss [44]. This study will collect data on both anti-obesity medications and weight-reducing antidiabetic drugs (**Table 3**). The potential impact of these weight-modifying pharmacologic interventions will be considered in subsequent analyses.

Whole blood deoxyribonucleic acid (DNA) methylation analysis

Genomic DNA samples are extracted from 500 μL of whole blood samples, and 500 μg genomic DNA is bisulfite-converted using the EZ DNA Methylation Kit (Zymo Research, Irvine, USA) according to the manufacturer's instructions. The samples are analyzed for DNA methylation using the Infinium MethylationEPIC v2.0 BeadChip and Illumina iSan System (Illumina, Inc., San Diego, USA). Raw intensity data files (IDAT) obtained from methylation analysis are processed using the ChAMP package (ver. 2.29.1) in R 3.6.3 [45]. The champ.qc function is used to check the quality of low quality probes and samples. This includes removal of poor quality probes based on detection p-values, correction for dye bias, and evaluation of bisulfite conversion efficiency. Samples or probes that do not meet quality control (QC) thresholds are excluded from subsequent analyses. To correct for technical variations, data normalization is performed using the Beta Mixture Quantile dilation method [46] using champ.norm function in the ChAMP package. Batch effects are assessed and corrected using the champ.svd and champ.combat functions, respectively. Differential methylation analysis aims to identify differentially methylated positions and regions (DMRs) between defined sample groups. CpG sites are annotated using the manifest with mapping information for EPIC_v2 available at https://zwdzwd.github.io/InfiniumAnnotation [47]. This resource provides detailed information on the genomic context of each CpG site, including gene associations, regulatory regions, and proximity to CpG island. The champ. dmp function is used to identify individual CpG sites showing significant differences in

methylation levels, employing a statistical threshold of $P < 0.05$ after adjustment for multiple comparisons using the false discovery rate method. For DMR identification, the champ. dmr function is applied using the Bumphunter algorithm [48]. Regions are considered differentially methylated if they contain at least three consecutive CpG sites with an average methylation difference of > 10% between groups.

Serum metabolite analysis

For the analysis of serum metabolites, 50 μL of serum sample is added to 50 μL of acetonitrile (ACN) containing an internal standard (200 μg/mL) and incubated at −20°C for 20 minutes. After extraction, centrifugation is performed at 12,000 g at 4°C for 10 minutes, and 70 μL of supernatant is diluted and mixed with 70 μL of 1% trifluoroacetic acid in water. The resulting solution is transferred to an liquid chromatography (LC) vial, and 5 μL is injected for analysis. QC samples are prepared by pooling equal volumes of all individual samples to create a representative matrix. Retention time consistency and peak intensity variation in QC injections are monitored to ensure reliable data acquisition.

All tandem mass spectrometry (MS/MS) experiments are performed using an Agilent 6470 triple quadrupole mass spectrometer (MS) instrument (Agilent, Santa Clara, USA) coupled to an Agilent 1260 infinity LC system (Agilent). A gradient eluent (A, 0.1% formic acid [FA] in water; B, 0.1% FA in ACN) is used 0.3 mL/min with ACQUITY UPLC BEH C18 column (1.7 μm, 2.1 × 100 mm) (Manchester, UK) at 40°C. The gradient elution system is controlled as follows: 0 minutes, 5% B; 0–1 minutes, 5% B; 1–5.5 minutes, 5–25% B; 5.5–7.5 minutes, 25–27% B; 7.5–8.5 minutes, 27–30% B; 8.5–10.5 minutes, 30% B; 10.5–13.5 minutes, 30–40% B; 13.5–16.5 minutes, 40–95% B; 16.5–17 minutes, 95–5% B. The gradient is finally returned to the initial condition (5% B) and held there for 5 min before running the next sample.

Fecal sample collection and microbioal 16S rRNA sequencing analysis

Fecal samples from each participant are immediately mixed with OMNIgene-GUT tubes (OMR-200; DNA Genotek, Stittsville, Canada) and shaking for a minimum of 30 seconds. And then, 250 μL of samples is placed into a 2-mL tube containing 0.3 g sterile 0.1 mm zirconia beads (BioSpec, Bartlesville, USA) with 1.2 mL ASL lysis buffer from QIAamp DNA Stool Mini Kit (Qiagen, Hilden, Germany) and vortexed for 1 minute. The samples are subsequently heated at 95°C for 15 minutes and then subjected to two cycles of bead beating at a frequency of 30 Hz for 1 minute using the Qiagen TissueLyser II. After centrifugation, the supernatant (1.2 mL) is treated with an InhibitEX Tablet, and 350 μL of the resulting supernatant is used in the subsequent steps with a QIAcube system (Qiagen). Library preparation of the V3–V4 hypervariable region of the 16S rRNA gene is performed according to the 16S Metagenomic Sequencing Library Preparation Illumina protocol (Part #15044223 Rev. B; Illumina). The library pool containing equal molar quantities of each sample is sequenced using the MiSeq system (Illumina), according to the manufacturer's instructions. The QIIME software is used to assess microbial diversity, and QIIME Uclust is used to generate taxonomic composition.

Fecal metabolites analysis

The leftover fecal samples (200 μL) in the OMNIgene-GUT tubes are mixed with ACN (200 μL) and sonicated in ice bath for 10 minutes. Then, centrifugation is performed at 13,000 rpm at 4°C for 10 minutes. After centrifugation, the supernatant (50 μL), 50 μL of internal standard, and 400 μL of water are added. The solution is transferred to an LC vial, and 5 μL is injected for analysis. Also, QC samples are prepared by pooling equal volumes of

all individual fecal extracts. Retention time consistency and peak intensity variation in QC injection are monitored to ensure reliable data acquisition.

All MS/MS experiments are performed using a Waters Xevo triple quadrupole MS instrument (Waters, Manchester, UK) coupled to a Waters ACQUITY H-Class LC system (Waters). A gradient eluent (A, 0.1% FA in water; B, 0.1% FA in ACN) is used 0.35 mL/min with ACQUITY UPLC BEH C18 column (1.7 μm, 2.1 × 100 mm) at 40°C. The gradient elution system is controlled as follows: 0 minutes, 5% B; 0–17.5 minutes, 5–95% B; 17.5–19 minutes, 95–89% B; 19–19.5 minutes, 89–5% B. The gradient is finally returned to the initial condition (5% B) and held there for 2.5 minutes before running the next sample.

Urine metabolite analysis

For urine samples, 400 μL is combined with 230 μL of 0.2 M sodium phosphate buffer (pH 7.0) in deuterium oxide (D_2O). After centrifuging the sample at 13,000 rpm for 10 minutes at 4°C, the pH is adjusted to 7.0 ± 0.1. 540 μL of sample is combined with 60 μL of 6 mM 3-(trimethylsilyl)propionic 2,2,3,3-d4 acid in D_2O. The resulting 600 μL samples are transferred to 5 mm nuclear magnetic resonace (NMR) tubes.

Urine samples are analyzed using Bruker Avance III HD 800 MHz FT-NMR Spectrometer (Bruker BioSpin, Ettlingen, Germany) equipped with a 5-mm CPTIC cryogenic probe at 298K. ^{1}H NMR spectra using a NOESYPRESAT pulse sequence to suppress the residual water signal are collected into 64,000 data points with 64 transients, an acquisition time of 2 seconds, and a relaxation delay of 2 seconds. Phase and baseline correction of all acquired ^{1}H NMR spectra are performed using TopSpin 3.6 and AMIX software (Bruker BioSpin), respectively. Spectral intensities are normalized by the median fold-change method to correct for systematic variation. Peaks are retained only if their QC-to-blank ratio exceeds 3 and their coefficient of variation in QC samples is ≤ 30%. Equal aliquots of all study samples are then combined to generate a pooled QC sample. Identification of metabolites is conducted using 800 MHz NMR database in Chenomx NMR Suite Version 8.6 (Chenomx Inc., Edmonton, Canada), 2D total correlation spectroscopy, and spiking experiments. Metabolite quantification is performed using the Chenomx and the metabolites concentrations are normalized to creatinine levels.

Saliva metabolite analysis

For saliva collection, all participants are instructed to fast and avoid smoking as well as dental care practices such as tooth brushing or flossing, for at least 2 hours prior to sample collection. Before collecting the samples, participants rinse their mouths with water to remove any residual debris. They then drool directly into a 15 mL conical tube until at least 3 mL of saliva is collected. The saliva samples are centrifuged at 12,000 rpm for 5 minutes at 4°C, after which 1 mL of the supernatant is transferred to a new tube for further analysis.

To 100 μL of saliva, 400 μL of ACN is combined and vortexed for 2 minutes. The mixture is centrifuged at 13,000 rpm for 15 minutes at 4°C. The supernatants are then transferred to new tube and evaporated using a nitrogen stream. The samples are stored at −80°C until the LC-MS/MS analysis, and are diluted with 100 μL of water prior to analysis.

Salivary metabolites are analyzed using a Vanquish ultra-high performance liquid chromatography system (Thermo Fisher Scientific, Santa Clara, USA) with a Hypersil GOLD™ VANQUISH™ C18 column (2.1 mm × 150 mm, 1.9 μm; Thermo Fisher Scientific) coupled to an Orbitrap Exploris 120 MS (Orbitrap MS, Thermo Fisher Scientific). The chromatographic

separation analysis is conducted using a binary gradient separation with a flow rate of 0.3 mL/min at 45°C. The mobile phase consists of (A) 0.1% FA, and (B) 0.1% FA in methanol (MeOH). The acquisition methods are performed using data dependent acquisition with spray voltages of 3,500 V and −2,500 V for the positive and negative modes. The vaporizer temperature is 350°C and ion transfer tube temperature is 325°C with the auxiliary gas and sheath gas at 10 and 45 arbitrary units (A.U.), respectively. The pooled QC sample is generated using the same procedure as described for the urine sample analysis.

Peptidomic analysis

For blood peptide extraction, serum (100 μL) mixed with 100 μL of distilled water is loaded into the Oasis HLB 96-well μElution Plate, which is conditioned with 200 μL of MeOH and then equilibrated with 200 μL of distilled water. After washing with 200 μL of 5% MeOH, the peptides are eluted with 80 μL of 40% MeOH. The eluted peptide samples mixed with 10 μL of terfenadine as an internal standard are analyzed using ultra-performance liquid chromatography-quadrupole-time of flight MS (Xevo G2-S, Waters, Milford, USA). Samples are injected into an Acquity UPLC BEH C18 column (2.1 mm × 100 mm, 1.7 um; Waters). The mobile phase consists of water with 0.1% FA (A) and ACN with 0.1% FA (B) at a flow rate of 0.35 mL/min for 12 minutes. The eluted peptides are ionized by positive mode electrospray ionization and detected using Q-TOF MS using the following optimized MS conditions: 3 kV of capillary voltages, 40 V of sampling cone voltage, 800 L/h of desolvation flow rate, 400°C of desolvation temperature, and 100°C of ion source temperature. Leucine-enkephalin is used as a reference for the lock mass at a frequency of 10 seconds. A pooled QC sample—prepared by combining equal aliquots of randomly selected study samples—is run between each analysis batch. MS/MS spectra are obtained using collision energy ramps from 10 to 30 eV. Mass spectrometry data are collected from m/z 50 to 1,500 with a scan time of 0.2 seconds. The peptide sequence is analyzed by *de novo* peptide sequencing using BioLynx (Waters) and the peptide sequences are confirmed by BLAST sequence similarity search.

Lipidomic analysis

The total lipid extraction is performed according to the Bligh–Dyer method with minor modification. Briefly, 1.2 mL of MeOH and chloroform (1:2, v/v) with 3 μL of 10N HCl are added to 50 μL of a human plasma sample or human stratum corneum tissues stripped using adhesive tape (D-Squame tape, CuDerm Corporatin, Dallas, USA), followed by the addition of 100 pmol of each lipid as an internal standard. After vortex mixing for 10 minutes, centrifugation at 15,000 × g for 5 minutes, the lower phase is transferred to a new tube and dried under nitrogen stream. The dried samples are dissolved in MeOH and ACN (1:1, v/v). Aliquots (5 μL) are injected into a reverse-phase column (ACQUITY UPLC CSH C18 column, length 100 mm, particle size 1.7 μm, inner diameter 2.1 mm) (Waters) using a binary gradient with mobile phase A (0.2% FA with 10 mM ammonium formate) and mobile phase B (MeOH:2-propanol in 0.2% FA with 10 mM ammonium formate). The effluent from the column is directly introduced into a hybrid quadrupole linear ion trap tandem MS (API5500 QTRAP, AB/SCIEX, Framingham, USA) equipped with an ESI in positive mode or negative mode and multiple reaction monitoring (MRM). Each of lipids are detected in the MRM mode by selecting the mass to charge ratio (m/z) of each species at the precursor ion at the product ion. The following conditions are used in positive-ion MRM: ion spray voltage, 5,500 V; temperature (TEM), 600°C; curtain gas (CUR), 30 A.U.; collision gas (CAD), "Medium"; ion nebulizer gas (Gas 1 [GS1]), 50 A.U.; and auxiliary gas (Gas 2 [GS2]), 60 A.U. For negative-ion MRM, the conditions are as follows: ion spray voltage, −4,500 V; TEM, 600°C; CUR, 30 A.U.; CAD, "Medium"; GS1, 50 A.U.; and GS2, 60 A.U. Nitrogen is used as the nebulizer, curtain,

and CAD. Equal volumes of each serum sample are pooled to generate a QC sample, which is injected after every ten study samples to monitor analytical stability and reproducibility. Isotopically labeled internal standards representing major lipid classes are spiked into all samples prior to extraction to correct for matrix effects and technical variability. Quantification is performed using external calibration curves prepared from authentic lipid standards at six concentration levels ($r^2 \geq 0.99$), with each point analyzed in duplicate. Lipid species are retained for downstream analysis if they meet the following criteria: QC-to-blank ratio ≥ 3, intra-QC coefficient of variation $\leq 15\%$, and consistent retention time within ± 0.15 minutes. Extraction and solvent blanks are included to exclude background contaminants. Lipid concentrations are reported as pmol/mg protein or nM, and data are processed using Analyst software (v1.7.1; Applied Biosystems, Foster City, USA) and MultiQuant (v3.0.3; SCIEX), with manual inspection of chromatographic peaks as needed.

Data collections and data cleaning

Participants in the obese group are scheduled to visit the hospital once a year from the time of enrollment until the fifth year. To minimize missed visits, research staff contact participants in accordance with the visit window, typically every 9 to 12 months, to coordinate and confirm upcoming appointments. For participants in the control group, a hospital visit is conducted only in the year of enrollment. Thereafter, annual follow-up is performed via telephone to monitor clinical events.

To encourage continued participation, several incentives are provided. When abnormal findings are detected through blood tests, functional assessments, or imaging studies, participants are offered referrals for further medical evaluation. Additionally, participants receive body composition analysis reports, enabling them to understand changes in physical and body composition parameters compared to the previous year. Furthermore, genetic reports based on single nucleotide polymorphism (SNP) analysis are distributed, providing personalized genetic risk for obesity and chronic metabolic diseases, thereby promoting awareness and engagement in personal health management.

Clinical measures and multi-omics data closest to the expected annual follow-up date, defined as 12 months from the previous year's measurement, are selected as that year's values. Missing values will be handled using the following approach.

Clinical variables
Clinical variables with a missing data rate below 5% are considered statistically inconsequential [49] and included in the analysis without imputation.

Multi-omics data
For metabolite analyses using serum, stool, urine, and saliva samples, as well as for DNA methylation, peptidomic, and lipidomic analyses, variables with more than 10% missingness across all samples are excluded first. Subsequently, participants with more than 10% missing values among the remaining variables are also excluded [50].

Genomic analysis with AI

DNA samples were extracted from whole blood using QIAamp DNA Blood Kits (Qiagen) according to the manufacturer's protocol. SNP genotyping is conducted using Affymetrix Axiom KORv1.1 array (Thermo Fisher Scientific). To identify a subset of obesity-related SNPs within the cohort, both univariate and multivariate feature selection methods are applied.

Initially, SNP array data from the K-chip are preprocessed using PLINK. SNPs with low minor allele frequency, high missing genotype rate, and low Hardy-Weinberg equilibrium values are filtered out to remove low-quality SNPs. Subsequently, regression-based association analysis is employed as a univariate feature selector. In this analysis, obesity-related phenotypes, such as BMI and waist circumference, are used as the independent targets, with sex and age included as covariates to remove their effects. Then, the reduced SNP data undergo multivariate feature selection. A learning-based feature selection algorithm identifies a subset of SNPs that are highly associated with obesity but not with each other. Additionally, the complexity of the selection process and the number of SNPs to be derived are important factors in this procedure.

For the classification model of diabetes based on DNA methylation status, we utilize DNA methylation array data from blood samples of obese patients and non-obese controls. During data preprocessing, the raw DNA methylation array data (.idat files) are processed using the PyMethylProcess (Version 0.1.3) Python package. This preprocessing includes normalization, removal of non-autosomal CpGs, imputation, and feature selection. After preprocessing, the dataset is divided into training (80%), validation (12.5%), and testing (7.5%) subsets. The segmented dataset is then used to build a classification model with the MethylNet (Version 0.1) Python package. First, latent space features of DNA methylation are extracted using a variational autoencoder, and a disease state prediction model is built using transfer learning. Finally, we employ SHAP to identify the CpGs important for disease state prediction and to interpret the predictions for each individual and class.

DISCUSSION

Metabolic disorders, including obesity, DM, metabolic syndrome, and CVD, are highly heterogeneous, displaying a wide range of characteristics and complexities [50]. Obesity is determined by the BMI, calculated by dividing a person's weight in kilograms by the square of their height in meters, and categorized according to the WHO's guidelines [51]. However, BMI fails to consider body composition, such as muscle and fat mass [50], and it also does not accurately reflect the distribution of fat, particularly excessive visceral fat [51]. Beyond the BMI itself, the accumulation of visceral/ectopic fat is considered a major contributor to cardiovascular and metabolic risk [51]. As a result, obesity categories based on BMI alone inherently show diverse risk levels for CVD, leading to the identification of phenotypes such as metabolically healthy obesity [51]. Therefore, it is crucial to classify the phenotypes of obesity and assess the cardiovascular risk for each phenotype to develop tailored strategies accordingly.

Recently, even beyond body composition and visceral adiposity, there are a wide variety of individual factors contributing to metabolic heterogeneity of obesity. Genetic and epigenetic diversity interacts differently with non-genetic factors such as health behavior and environment, leading to obesity through various etiologies [52]. In addition, alteration in the profiles of proteins belonging to metabolic pathways, oxidative stress, and inflammatory process was associated with adverse outcomes of obesity and its related diseases [53]. Individual characteristics of metabolites are thought to reflect changes in amino acid and nucleotide metabolism influenced by insulin resistance [54]. Apart from genetic background, disruption of the metabolome also predicted a higher risk of cardiovascular events than healthy metabolome [54]. Gut microbiome exhibits a large between-subjects diversity, and a disturbance in microbiota composition may induce weight gain, fat storage, and insulin

resistance [55]. An individual's genomic, proteomic, metabolomic, and microbial signatures may mutually influence each other and be correlated, but each omics profile has aspects that independently contribute to the development of obesity [50,54]. Thus, multiomics could be a suitable approach to comprehensively elucidate diverse physiological states of obesity [50].

Nutriomics is a compound word of nutrition and omics, and it is a field of multiomics that integratively analyzes the effects of food intake on the various biological data (genes, proteins, metabolites, and microbiome) [56]. The current "Nutriomics and Artificial Intelligence Nutrition Obesity Cohort (NAINOC)" study has several distinct strengths. First, the most notable strength is the comprehensive investigation of nutritional intake to analyze these data in relation to other multiomics data. Semi-quantitative FFQ used in the current study provides information on the intake frequency of 129 dish/food items, grouped according to similar recipes and main ingredients, over the past year [24]. The FFQ data can be used to estimate calorie intake and consumption of nutrients such as carbohydrate [57]. Furthermore, during the period of continuous glucose monitoring, data on dynamic changes in glucose corresponding to dietary intake will be obtained through the use of a dietary record application. This will allow to integrate real-time glucose responses to dietary intake with other multiomics data, which is unique to this study. Second, the inclusion of both obese individuals with various metabolic states and non-obese individuals as controls is another advantage in regard to classify obesity phenotypes. We included obese individuals with at least one chronic metabolic disorders other than obesity, which likely added to the metabolic diversity of the study population. Third, the use of multiple modalities to assess metabolic and complication status will facilitate multifaceted analysis. The present study not only assesses the presence or absence of metabolic disorders including type 2 DM and CVD, but also performs fibroscan to evaluate steatotic liver disease, as well as CAC scan and carotid IMT to assess atherosclerosis. Fourth, in addition to blood and stool samples frequently used in previous multiomic studies [50,54,58], samples of urine, saliva, and skin will further refine the individual characterization of obesity.

CONCLUSIONS

The nutriomic approach in this study was designed to classify the body's responses to food intake using various biomarkers in both obese and non-obese individuals, thereby enhancing the classification of obesity phenotypes and the development of individualized dietary treatment strategies.

ACKNOWLEDGEMENTS

REFERENCES

1. Elmaleh-Sachs A, Schwartz JL, Bramante CT, Nicklas JM, Gudzune KA, Jay M. Obesity management in adults: a review. JAMA. 2023;330:2000-15. PUBMED | CROSSREF

2. Abarca-Gómez L, Abdeen ZA, Hamid ZA, Abu-Rmeileh NM, Acosta-Cazares B, Acuin C, et al. Worldwide trends in body-mass index, underweight, overweight, and obesity from 1975 to 2016: a pooled analysis of 2416 population-based measurement studies in 128·9 million children, adolescents, and adults. Lancet. 2017;390:2627-42. PUBMED | CROSSREF

3. Koliaki C, Dalamaga M, Liatis S. Update on the obesity epidemic: after the sudden rise, is the upward trajectory beginning to flatten? Curr Obes Rep. 2023;12:514-27. PUBMED | CROSSREF

4. Jeong SM, Jung JH, Yang YS, Kim W, Cho IY, Lee YB, et al. 2023 obesity fact sheet: prevalence of obesity and abdominal obesity in adults, adolescents, and children in Korea from 2012 to 2021. J Obes Metab Syndr. 2024;33:27-35. PUBMED | CROSSREF

5. Whitlock G, Lewington S, Sherliker P, Clarke R, Emberson J, Halsey J, et al. Body-mass index and cause-specific mortality in 900 000 adults: collaborative analyses of 57 prospective studies. Lancet. 2009;373:1083-96. PUBMED | CROSSREF

6. Ni Mhurchu C, Rodgers A, Pan WH, Gu DF, Woodward M; Asia Pacific Cohort Studies Collaboration. Body mass index and cardiovascular disease in the Asia-Pacific Region: an overview of 33 cohorts involving 310 000 participants. Int J Epidemiol. 2004;33:751-8. PUBMED | CROSSREF

7. Li C, Stražar M, Mohamed AMT, Pacheco JA, Walker RL, Lebar T, et al. Gut microbiome and metabolome profiling in Framingham heart study reveals cholesterol-metabolizing bacteria. Cell. 2024;187:1834-1852.e19. PUBMED | CROSSREF

8. Prapiadou S, Živković L, Thorand B, George MJ, van der Laan SW, Malik R, et al. Proteogenomic data integration reveals CXCL10 as a potentially downstream causal mediator for IL-6 signaling on atherosclerosis. Circulation. 2024;149:669-83. PUBMED | CROSSREF

9. Mayer J, Thomas DW. Regulation of food intake and obesity. Science. 1967;156:328-37. PUBMED | CROSSREF

10. Fromentin S, Forslund SK, Chechi K, Aron-Wisnewsky J, Chakaroun R, Nielsen T, et al. Microbiome and metabolome features of the cardiometabolic disease spectrum. Nat Med. 2022;28:303-14. PUBMED | CROSSREF

11. Di Renzo L, Gualtieri P, Romano L, Marrone G, Noce A, Pujia A, et al. Role of personalized nutrition in chronic-degenerative diseases. Nutrients. 2019;11:1707. PUBMED | CROSSREF

12. Braun T, Feng R, Amir A, Levhar N, Shacham H, Mao R, et al. Diet-omics in the Study of Urban and Rural Crohn disease Evolution (SOURCE) cohort. Nat Commun. 2024;15:3764. PUBMED | CROSSREF

13. Zeevi D, Korem T, Zmora N, Israeli D, Rothschild D, Weinberger A, et al. Personalized nutrition by prediction of glycemic responses. Cell. 2015;163:1079-94. PUBMED | CROSSREF

14. Graw S, Henn R, Thompson JA, Koestler DC. pwrEWAS: a user-friendly tool for comprehensive power estimation for epigenome wide association studies (EWAS). BMC Bioinformatics. 2019;20:218. PUBMED | CROSSREF

15. Nyamundanda G, Gormley IC, Fan Y, Gallagher WM, Brennan L. MetSizeR: selecting the optimal sample size for metabolomic studies using an analysis based approach. BMC Bioinformatics. 2013;14:338. PUBMED | CROSSREF

16. Hur KY, Moon MK, Park JS, Kim SK, Lee SH, Yun JS, et al. 2021 clinical practice guidelines for diabetes mellitus of the Korean Diabetes Association. Diabetes Metab J. 2021;45:461-81. PUBMED | CROSSREF

17. Alberti KGMM, Eckel RH, Grundy SM, Zimmet PZ, Cleeman JI, Donato KA, et al. Harmonizing the metabolic syndrome: a joint interim statement of the International Diabetes Federation Task Force on Epidemiology and Prevention; National Heart, Lung, and Blood Institute; American Heart Association; World Heart Federation; International Atherosclerosis Society; and International Association for the Study of Obesity. Circulation. 2009;120:1640-5. PUBMED | CROSSREF

18. Kong JS, Kim MK. Association of healthy lifestyle factors with the risk of hypertension, dyslipidemia, and their comorbidity in Korea: results from the Korea National Health and Nutrition Examination Survey 2019-2021. Epidemiol Health. 2024;46:e2024049. PUBMED | CROSSREF

19. Tu Q, Hyun K, Lin S, Hafiz N, Manandi D, Li E, et al. Impacts of hypertension and diabetes on the incidence of cardiovascular diseases and all-cause mortality: findings from the China Health and Retirement Longitudinal Study cohort. J Hypertens. 2025;43:623-30. PUBMED | CROSSREF

20. Kobayashi G, Okada H, Hamaguchi M, Kurogi K, Murata H, Ito M, et al. Dyslipidemia and 10-year diabetes incidence in Japanese people: Population-based Panasonic cohort study 9. Front Endocrinol (Lausanne). 2022;13:957728. PUBMED | CROSSREF

21. Kario K, Hoshide S, Okawara Y, Tomitani N, Yamauchi K, Ohbayashi H, et al. Effect of canagliflozin on nocturnal home blood pressure in Japanese patients with type 2 diabetes mellitus: the SHIFT-J study. J Clin Hypertens (Greenwich). 2018;20:1527-35. PUBMED | CROSSREF

22. Kario K, Okada K, Kato M, Nishizawa M, Yoshida T, Asano T, et al. Twenty-four-hour blood pressure-lowering effect of a sodium-glucose cotransporter 2 inhibitor in patients with diabetes and uncontrolled

nocturnal hypertension: results from the randomized, placebo-controlled SACRA study. Circulation. 2019;139:2089-97. **PUBMED** | **CROSSREF**

23. Harper A, Power M, Grp W; The WHOQOL Group. Development of the World Health Organization WHOQOL-BREF quality of life assessment. Psychol Med. 1998;28:551-8. **PUBMED** | **CROSSREF**

24. Chung J, Ahn S, Joung H, Shin S. Development of a semi-quantitative food-frequency questionnaire for Korean adults with obesity. Nutrients. 2023;15:4848. **PUBMED** | **CROSSREF**

25. 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. **PUBMED** | **CROSSREF**

26. Kim S, Lee CJ, Han K, Lee KH, Lee HJ, Park S. Histogram-derived modified thresholds for coronary artery calcium scoring with lower tube voltage. Sci Rep. 2021;11:17450. **PUBMED** | **CROSSREF**

27. Kim SU, Kim JK, Park YN, Han KH. Discordance between liver biopsy and Fibroscan® in assessing liver fibrosis in chronic hepatitis B: risk factors and influence of necroinflammation. PLoS One. 2012;7:e32233. **PUBMED** | **CROSSREF**

28. Kim JK, Lee KS, Choi JR, Chung HJ, Jung DH, Lee KA, et al. Usefulness of the controlled attenuation parameter for detecting liver steatosis in health checkup examinees. Gut Liver. 2015;9:405-10. **PUBMED** | **CROSSREF**

29. Jung CY, Ryu GW, Kim HW, Ahn SH, Kim SU, Kim BS. Advanced liver fibrosis measured by transient elastography predicts chronic kidney disease development in individuals with non-alcoholic fatty liver disease. Diabetologia. 2022;65:518-27. **PUBMED** | **CROSSREF**

30. Lee CJ, Rim TH, Kang HG, Yi JK, Lee G, Yu M, et al. Pivotal trial of a deep-learning-based retinal biomarker (Reti-CVD) in the prediction of cardiovascular disease: data from CMERC-HI. J Am Med Inform Assoc. 2023;31:130-8. **PUBMED** | **CROSSREF**

31. Lee CJ, Lee YH, Park SW, Kim KJ, Park S, Youn JC, et al. Association of serum proprotein convertase subtilisin/kexin type 9 with carotid intima media thickness in hypertensive subjects. Metabolism. 2013;62:845-50. **PUBMED** | **CROSSREF**

32. Bergenstal RM, Beck RW, Close KL, Grunberger G, Sacks DB, Kowalski A, et al. Glucose management indicator (GMI): a new term for estimating A1C from continuous glucose monitoring. Diabetes Care. 2018;41:2275-80. **PUBMED** | **CROSSREF**

33. Kim K, Choi SH, Jang HC, Park YS, Oh TJ. Glucose profiles assessed by intermittently scanned continuous glucose monitoring system during the perioperative period of metabolic surgery. Diabetes Metab J. 2022;46:713-21. **PUBMED** | **CROSSREF**

34. Lee EY, Han K, Kim DH, Park YM, Kwon HS, Yoon KH, et al. Exposure-weighted scoring for metabolic syndrome and the risk of myocardial infarction and stroke: a nationwide population-based study. Cardiovasc Diabetol. 2020;19:153. **PUBMED** | **CROSSREF**

35. Teo K, Chow CK, Vaz M, Rangarajan S, Yusuf S; PURE Investigators-Writing Group. The Prospective Urban Rural Epidemiology (PURE) study: examining the impact of societal influences on chronic noncommunicable diseases in low-, middle-, and high-income countries. Am Heart J. 2009;158:1-7.e1. **PUBMED** | **CROSSREF**

36. Yusuf S, Joseph P, Rangarajan S, Islam S, Mente A, Hystad P, et al. Modifiable risk factors, cardiovascular disease, and mortality in 155 722 individuals from 21 high-income, middle-income, and low-income countries (PURE): a prospective cohort study. Lancet. 2020;395:795-808. **PUBMED** | **CROSSREF**

37. Matthews DR, Hosker JP, Rudenski AS, Naylor BA, Treacher DF, Turner RC. Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man. Diabetologia. 1985;28:412-9. **PUBMED** | **CROSSREF**

38. Tsai SF, Yang CT, Liu WJ, Lee CL. Development and validation of an insulin resistance model for a population without diabetes mellitus and its clinical implication: a prospective cohort study. EClinicalMedicine. 2023;58:101934. **PUBMED** | **CROSSREF**

39. Wang C, Wang Y, Wu J, Liu S, Zhu Y, Lv S, et al. Current smoking dose-dependently associated with decreased β-cell function in chinese men without diabetes. J Diabetes Res. 2015;2015:841768. **PUBMED** | **CROSSREF**

40. Unger T, Borghi C, Charchar F, Khan NA, Poulter NR, Prabhakaran D, et al. 2020 International Society of Hypertension Global Hypertension Practice Guidelines. Hypertension. 2020;75:1334-57. **PUBMED** | **CROSSREF**

41. Moon JS, Kang S, Choi JH, Lee KA, Moon JH, Chon S, et al. 2023 clinical practice guidelines for diabetes management in Korea: full version recommendation of the Korean Diabetes Association. Diabetes Metab J. 2024;48:546-708. **PUBMED** | **CROSSREF**

42. Eckel RH, Jakicic JM, Ard JD, de Jesus JM, Houston Miller N, Hubbard VS, et al. 2013 AHA/ACC guideline on lifestyle management to reduce cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. Circulation. 2014;129:S76-99. **PUBMED** | **CROSSREF**

43. Jeon E, Lee KY, Kim KK. Approved anti-obesity medications in 2022 KSSO guidelines and the promise of phase 3 clinical trials: anti-obesity drugs in the sky and on the horizon. J Obes Metab Syndr. 2023;32:106-20. **PUBMED** | **CROSSREF**

44. Bae JH. SGLT2 inhibitors and GLP-1 receptor agonists in diabetic kidney disease: evolving evidence and clinical application. Diabetes Metab J. 2025;49:386-402. **PUBMED** | **CROSSREF**

45. Tian Y, Morris TJ, Webster AP, Yang Z, Beck S, Feber A, et al. ChAMP: updated methylation analysis pipeline for Illumina BeadChips. Bioinformatics. 2017;33:3982-4. **PUBMED** | **CROSSREF**

46. Teschendorff AE, Marabita F, Lechner M, Bartlett T, Tegner J, Gomez-Cabrero D, et al. A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. Bioinformatics. 2013;29:189-96. **PUBMED** | **CROSSREF**

47. Kaur D, Lee SM, Goldberg D, Spix NJ, Hinoue T, Li HT, et al. Comprehensive evaluation of the infinium human MethylationEPIC v2 BeadChip. Epigenetics Commun. 2023;3:6. **PUBMED** | **CROSSREF**

48. Jaffe AE, Murakami P, Lee H, Leek JT, Fallin MD, Feinberg AP, et al. Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies. Int J Epidemiol. 2012;41:200-9. **PUBMED** | **CROSSREF**

49. Dong Y, Peng CYJ. Principled missing data methods for researchers. Springerplus. 2013;2:222. **PUBMED** | **CROSSREF**

50. Watanabe K, Wilmanski T, Diener C, Earls JC, Zimmer A, Lincoln B, et al. Multiomic signatures of body mass index identify heterogeneous health phenotypes and responses to a lifestyle intervention. Nat Med. 2023;29:996-1008. **PUBMED** | **CROSSREF**

51. Neeland IJ, Poirier P, Després JP. Cardiovascular and metabolic heterogeneity of obesity: clinical challenges and implications for management. Circulation. 2018;137:1391-406. **PUBMED** | **CROSSREF**

52. Trang K, Grant SFA. Genetics and epigenetics in the obesity phenotyping scenario. Rev Endocr Metab Disord. 2023;24:775-93. **PUBMED** | **CROSSREF**

53. Rodriguez-Muñoz A, Motahari-Rad H, Martin-Chaves L, Benitez-Porres J, Rodriguez-Capitan J, Gonzalez-Jimenez A, et al. A systematic review of proteomics in obesity: unpacking the molecular puzzle. Curr Obes Rep. 2024;13:403-38. **PUBMED** | **CROSSREF**

54. Cirulli ET, Guo L, Leon Swisher C, Shah N, Huang L, Napier LA, et al. Profound perturbation of the metabolome in obesity is associated with health risk. Cell Metab. 2019;29:488-500.e2. **PUBMED** | **CROSSREF**

55. Puljiz Z, Kumric M, Vrdoljak J, Martinovic D, Ticinovic Kurir T, Krnic MO, et al. Obesity, gut microbiota, and metabolome: from pathophysiology to nutritional interventions. Nutrients. 2023;15:2236. **PUBMED** | **CROSSREF**

56. Anaya-Morua W. Omics analysis in nutrition science. Mex J Med Res ICSA. 2022;10:59-63. **CROSSREF**

57. Signorello LB, Munro HM, Buchowski MS, Schlundt DG, Cohen SS, Hargreaves MK, et al. Estimating nutrient intake from a food frequency questionnaire: incorporating the elements of race and geographic region. Am J Epidemiol. 2009;170:104-11. **PUBMED** | **CROSSREF**

58. Liu S, Li F, Cai Y, Ren L, Sun L, Gang X, et al. Unraveling the mystery: a Mendelian randomized exploration of gut microbiota and different types of obesity. Front Cell Infect Microbiol. 2024;14:1352109. **PUBMED** | **CROSSREF**