BMC Medical Genomics

RESEARCH

Open Access

The inverse relationship between copy number variation burden and age of onset of metabolic syndrome in Korean middle-aged adults

Seong-Hee Ko[1†], Minyeong Kim[1†] and Dayeon Shin[2*]

Abstract

Background Metabolic syndrome (MetS) shows wide inter-individual variation in age at onset, yet the contribution of structural genomic variation is underexplored. We posited that copy number variation (CNV) burden—an aggregate of autosomal deletions and duplications—anticipates the timing of MetS. Our objective was to test whether higher CNV burden is associated with earlier onset of MetS and to nominate CNV regions enriched for metabolic biology, clarifying genetic architecture and guiding prevention. We leveraged a longitudinal Korean cohort to address this gap and provide evidence.

Methods A total of 546 adults from the Korean Genome and Epidemiology Study (KoGES) cohort, with baseline genotyping and up to 20 years of follow-up, were analyzed. CNV burden was quantified as the total number of autosomal CNV segments identified after quality control. For each genomic segment, log2-transformed copy number ratios outside the diploid range were classified as CNV-present (value = 1) and those within the diploid interval were set to zero. Individual CNV burden was calculated by summing all non-zero segments and log-transformed to reduce skewness prior to analysis. MetS was defined according to NCEP ATP III and Korean waist circumference reference values. Onset age of MetS was calculated by adding the elapsed follow-up time (years) until the first diagnosis of MetS to each participant's age at baseline. Participants who did not develop MetS were excluded from the onset-age calculation. We used multivariable logistic regression, multivariable linear regression, Kaplan–Meier with log-rank tests, and Cox proportional hazards models.

Results A total of 403 CNVs showed nominal associations with MetS based on multivariable logistic regression adjusted for demographic, lifestyle, and clinical covariates ($P < 0.05$). The strongest signals were located in the amylase gene clusters (*AMY1A/B/C*, *AMY2A/B*) and *KDM4C*, although no CNV survived multiple-testing correction. Higher CNV burden was associated with earlier MetS onset in men ($\beta = -21.42$, $P = 0.0249$) and women ($\beta = -14.09$, $P = 0.0451$). In sex-stratified Cox proportional hazards models, CNV burden was consistently associated with an increased risk

[†]Seong-Hee Ko and Minyeong Kim contributed equally to this work as co-first authors..

*Correspondence:
Dayeon Shin
dayeonshin@ewha.ac.kr

Full list of author information is available at the end of the article

© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

of incident MetS in men across all adjustment models, with hazard ratios ranging from 1.18 to 1.21. In women, the association showed a similar positive direction, with significance observed in partially adjusted models, although attenuated after full adjustment.

Conclusions Greater CNV burden consistently was associated with earlier MetS onset in a population-based Korean cohort, positioning structural variation as a contributor to cardiometabolic trajectories. The prominence of *AMY1/AMY2* and *KDM4C* suggests convergent pathways linking starch metabolism and epigenetic regulation to disease timing. CNV burden may serve as a biomarker to prioritize early clinical surveillance and tailored prevention, warranting mechanistic and external validation studies.

Keywords Copy number variations, Metabolic syndrome, Onset ages, Genome-wide association study, Korean Genome and Epidemiology Study (KoGES)

Introduction

Metabolic syndrome (MetS) is defined by the co-occurrence of abdominal obesity, elevated blood pressure, impaired fasting glucose, high triglycerides, and low HDL cholesterol, and is associated with increased risks of cardiovascular disease and type 2 diabetes [1–3]. In Korea, the prevalence of MetS rises sharply in older adults and exceeds that of many Western populations, which may be attributed to rapid dietary westernization, high refined-carbohydrate intake, increasing rates of central obesity, and possible East Asian genetic susceptibility [4, 5]. Importantly, recent epidemiologic studies have shown that MetS developing at a younger age is associated with a markedly higher risk of later cardiovascular complications compared with MetS that develops later in life [6–8]. Therefore, understanding factors that contribute not only to MetS risk but also to its age of onset has become a critical research priority.

Genetic studies of MetS have primarily focused on single nucleotide polymorphisms (SNPs) and have identified numerous risk alleles across metabolic pathways [9–11]. However, SNP-based findings explain only a modest fraction of the variability in metabolic traits and do not sufficiently account for inter-individual differences in the timing of MetS onset. Because MetS arises from the cumulative influence of many genetic variants with small effects interacting with lifestyle and environmental exposures, it is considered a polygenic and heterogeneous metabolic condition [12, 13].

Copy number variations (CNVs), defined as DNA deletions or duplications typically ranging from kilobases to megabases, represent a major class of structural variation that can alter gene dosage, disrupt coding or regulatory regions, and influence gene expression more substantially than single nucleotide changes [14, 15]. CNVs have previously been associated with type 2 diabetes [16], dyslipidemia [17], obesity-related traits, and BMI variation, including well-characterized CNVs at the *AMY1* and *AMY2* amylase loci that affect carbohydrate metabolism and adiposity [18–22]. In addition, CNVs involving epigenetic regulators such as KDM4C have been linked to pathways relevant to glucose, lipid, and amino acid metabolism [23–28]. Despite these biologically plausible links, prior studies have focused largely on the presence or absence of disease and have not examined whether CNVs contribute to differences in the age at which MetS develops.

To our knowledge, no population-based study has evaluated whether genome-wide CNV burden, an aggregate measure of the total structural variation carried by an individual, is associated with the timing of MetS onset. This represents a critical gap in the literature because CNV burden may capture genomic instability or cumulative metabolic susceptibility not reflected in SNP-based approaches.

The present study addresses this gap using genome-wide CNV profiles from Korean adults followed for up to 20 years. We first identify CNVs associated with MetS and then investigate whether individual CNV burden is associated with variation in the age at MetS onset. By integrating structural genomic variation with long-term epidemiologic follow-up, this study provides new insights into the genomic architecture underlying the development and timing of MetS.

Materials and methods

Study population

This study utilized data from the Ansan–Anseong cohort of the Korean Genome and Epidemiology Study (KoGES), a population-based prospective cohort of Korean adults aged 40–69 years. The Ansan (urban) and Anseong (rural) sites constitute one of the core KoGES Projects. CNV data were obtained from 1,000 participants in this cohort whose genomic data were curated as part of the Korea Association Resource (KARE) consortium, which is a researcher network established to analyze KoGES genomic resources. The KoGES-Ansan and Anseong studies enrolled 10,030 participants aged 40–69 years reside in urban (Ansan) and rural (Anseong) cities. At the baseline interview, all of the participants offered informed written consent and were examined regarding demographic data, biochemical data, anthropometric

measurement, medical history, including the analysis of dietary intake at medical institutions [29]. Among the total cohort of 10,030 participants, DNA samples from 1,000 participants (499 men and 501 women) from the Ansan and Anseong cohort populations were selected and genotyped at baseline. Of these 1,000 genotyped participants, we excluded those without MetS components data (n = 51), without covariates information (n = 69), and those with MetS at baseline or without data on onset age (n = 334). In summary, 546 participants (287 men and 259 women) were contained in the analysis of CNV-based association (Fig. 1). The study protocol was examined and approved by the Institutional Review Board (IRB) of Ewha Womans University (IRB No. ewha-202603-0023-01).

Demographic characteristics and anthropometric measurements

Demographic information was analyzed using an interviewer-administered questionnaire collected at baseline. Alcohol consumption was divided into non-drinkers, former drinkers, and current drinkers. Smoking status was divided as non-smokers if they answered "I have never smoked" to the question "Do you currently smoke?", former smokers if they answered "I have smoked before but do not currently smoke", and current smokers if they answered "I smoke occasionally" or "I continue to smoke habitually". Physical activity was measured by considering the length of time of the participants' activity across diverse intensity levels. The metabolic equivalent of task-hour/week (MET-h/week) was calculated by assigning specific weights to each activity intensity: 1.5 for sedentary, 3 for light, 5 for moderate, and 7 for vigorous activity. Socioeconomic characteristics were assessed using two indicators: educational attainment and household income. Educational status was categorized into three groups based on the highest level of completed schooling: (1) high school or below (2), college graduate, and (3) graduate-level education or higher. Household income was evaluated using participants' self-reported average monthly earnings and classified into four brackets: less than 2,000,000 won; 2,000,000–4,000,000 won; 4,000,000–6,000,000 won; and 6,000,000 won or more [30]. Energy intake (kcal/day) was estimated using a 103-item semiquantitative food frequency questionnaire (SQ-FFQ) offered by the KoGES dietary assessment [31].

Participants' height (cm) and weight (kg) were recorded by measuring while barefoot and wearing light clothing. Body mass index (BMI) was measured by dividing an participant's weight by the square of height (m^2). Waist circumference (cm) was measured horizontally at the midpoint between the lowest rib cage and iliac crest. Participants took a rest for 10 min in a sitting position and then their blood pressure (mmHg) was measured using a mercury sphygrometer (Baumanometer; WA Baum, Copiague, NY, USA). Blood samples of each participant were gathered in the morning after participants fasted for 12 h. The various biomarker such as fasting plasma

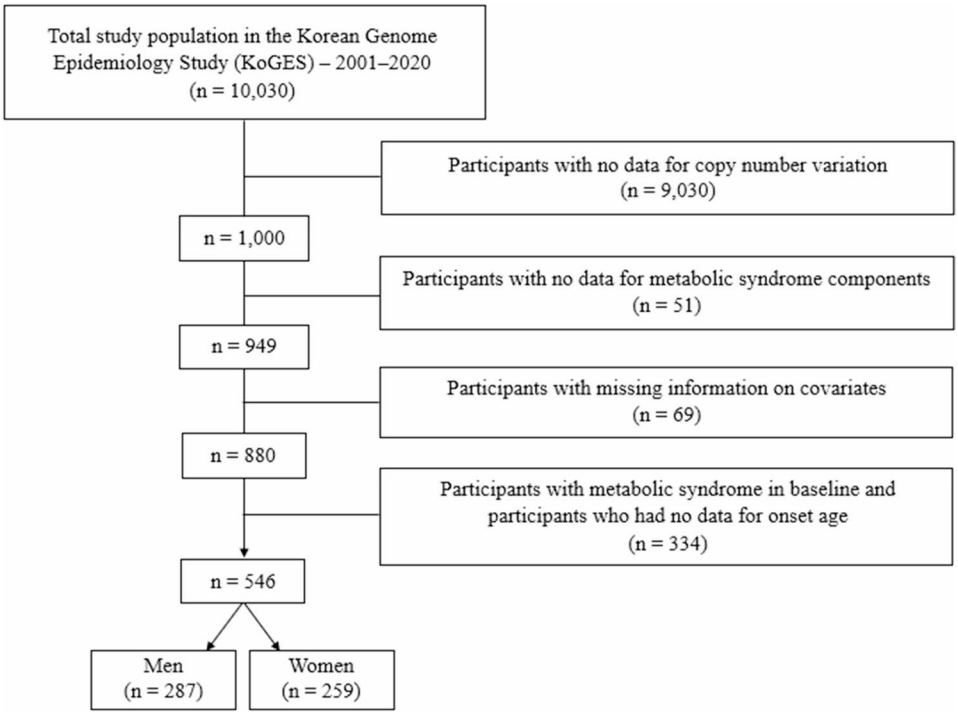


Fig. 1 Flowchart illustrating the selection of participants included in the study

glucose, triacylglycerol, and HDL cholesterol were measured using a chemical analyzer (Hitachi 7600; Hitachi, Tokyo, Japan, until August 2002, ADVIA 1650; Siemens, Tarrytown, NY, USA, from September 2002).

Definition of MetS

The guidelines of the National Cholesterol Education Program (NCEP) Adult Treatment Panel III (ATP III) were used as the definition of MetS in this study [32]. MetS was defined as satisfying three or more of the following five characteristics, and abdominal obesity in Koreans was defined based on the Korean Society for the Study of Obesity's treatment guidelines [33]:

(1) Systolic blood pressure (SBP) ≥ 130 mmHg, diastolic blood pressure (DBP) ≥ 85 mmHg, use of antihypertensive medication, history of hypertension diagnosis, or treatment for hypertension; (2) Serum triglyceride level ≥ 150 mg/dL; (3) Fasting blood glucose (FBG) ≥ 100 mg/dL, use of insulin injection, use of oral hypoglycemic agents, or history of diabetes diagnosis; (4) Serum HDL cholesterol level < 40 mg/dL in men and < 50 mg/dL in women; and (5) Abdominal obesity: waist circumference ≥ 90 cm in men and ≥ 85 cm in women [33]. Non-MetS was defined as not meeting the criteria for MetS; that is, individuals with zero, one, or two of the five components.

The onset age of MetS

In this study, the onset age of MetS was calculated by adding the elapsed follow-up time (years) until the first diagnosis of MetS to each participant's age at baseline [34]. Participants who did not develop MetS were excluded from the onset-age calculation. Based on this definition, the mean onset age among individuals who developed MetS was 60.7 ± 9.1 years. When stratified by sex, the mean onset age was 59.1 ± 9.1 years in men and 62.2 ± 8.9 years in women. The onset age variable was used as the outcome measure in multivariable linear regression analyses.

Quality control of SNP genotype

Genotype data generated from the Affymetrix Genome-Wide Human SNP Array 5.0 were subjected to a standardized multi-step quality control (QC) procedure using PLINK version 1.9. At the individual level, participants with a genotype call rate below 95% were removed. Sex concordance was evaluated using the PLINK --check-sex function, and samples exhibiting inconsistencies between reported sex and genetically inferred sex were excluded.

For SNP-level QC, variants with a missing call rate > 5% (--geno 0.05) were removed. SNPs exhibiting significant deviation from Hardy–Weinberg equilibrium ($P < 1 \times 10^{-6}$) or with a minor allele frequency (MAF) < 5% were also filtered out. To derive a linkage disequilibrium (LD)-independent set of markers for assessment of population structure, LD pruning was performed using a 50 kb sliding window, shifting by 5 SNPs at each step, and applying an LD threshold of $r^2 < 0.2$. The pruned SNP set was subsequently used to compute principal components (PCs) via the PLINK --pca algorithm [35, 36].

The first two principal components (PC1 and PC2), which captured the majority of ancestry-related variation within the cohort, were extracted and incorporated as covariates in all downstream association analyses to account for potential confounding from subtle population stratification.

Genotyping and CNV determination

The CNV data used in this study were 1,000 samples offered by the National Biobank of Korea, Korea Disease Control and Prevention Agency. CNVs were detected using a 720 K comparative genomic hybridization (CGH) array (Roche NimbleGen, Madison, WI, USA) provided by the National Biobank of Korea [37]. Raw data acquisition used data in CEL file format obtained from CGH array chip 720 K, and each file contains signal intensity information for each sample [38]. Signal intensity extraction was performed from the CEL files using PennCNV (version 1.0.3), and CNV calling was performed based on this. Signal intensity extraction was performed using PennCNV (version 1.0.3), and CNV calling was performed based on this software [39]. CNV determination was also performed using PennCNV, and the results were saved in text file format. Afterwards, GC adjustment was applied to minimize data distortion and increase the accuracy of analysis. GC adjustment for signal intensity was performed using the GC-modeling approach implemented in PennCNV, which reduces genomic 'wave' artifacts caused by local GC content [40]. The preprocessing steps such as raw data acquisition, signal intensity extraction, CNV calling, and GC ratio correction were performed in advance by the distributor, the National Biobank of Korea of Korea Disease Control and Prevention Agency.

The CNV text files provided by KoGES included a segment identifier (SEQ_ID), which groups probes belonging to the same contiguous genomic interval. The use of SEQ_ID as a segment-level identifier is consistent with previous bioinformatics workflows, where SEQ_ID denotes chromosomal intervals that link probe-level intensity data to genomic segments [41].

For CNV calling, probes with the same SEQ_ID were grouped into one CNV and then, merged into unique CNV segments. Outlier detection was performed by excluding segments that fall outside the quartile (Q) 1 – 1.5×interquartile range (IQR) and Q3 + 1.5×IQR ranges according to the IQR method, and calculating the average of the log2-transformed corrected signal intensity values for the remaining CNV segments. For QC, CNV

segments with less than 5 probes or a positive to negative probe ratio of less than 90% were removed, and this ratio was determined as the ratio of the number of positive and negative probes among the total probes. In addition, segments with a CNV value of 0 in all samples were considered meaningless regions and then removed, and CNVs located on sex chromosomes were also excluded. CNV coordinate information was provided based on hg18 (Human Genome Assembly 18; GRCh36) and converted to hg19 (GRCh37) using the UCSC LiftOver tool and the hg18ToHg19 chain file to apply the latest analysis standard [42]. A total of 15,766 CNVs were finally included in the analysis. A detailed overview of the CNV preprocessing and quality-control pipeline is presented in Fig. 2.

For each individual, CNV burden was calculated through a systematic multi-step process. As described in the previous section, autosomal CNV segments were first filtered by QC criteria and segments located on sex chromosomes were excluded. In addition, segments with log2 transformed copy number ratios within the diploid range, defined as log2 (1.32/2) to log2 (2.64/2) [43], were considered normal and set to zero. Following preprocessing, we quantified CNV burden at the individual level. Each genomic segment was first classified as either normal or CNV-present based on its log2 ratio. Segments with log2 ratios within the diploid range (i.e., values expected for normal copy number) were set to zero. Any segment with a log2 ratio outside this interval—indicating either a deletion or duplication—was assigned a value of one. Thus, CNV burden for each participant was calculated by summing the number of segments labeled as one. For example, if Participant A had 150 autosomal segments showing non-zero values (i.e., CNVs) and Participant B had 200 such segments, their CNV burdens were 150 and 200, respectively. This approach provides a comprehensive measure of an individual's overall structural genomic variation load without weighting by CNV type, size, or associated with functional impact [44]. The distribution of CNV burden exhibited substantial left-skewness; therefore, log-transformation was applied prior to statistical modeling to reduce skewness and improve normality (Supplementary Figure S1). To

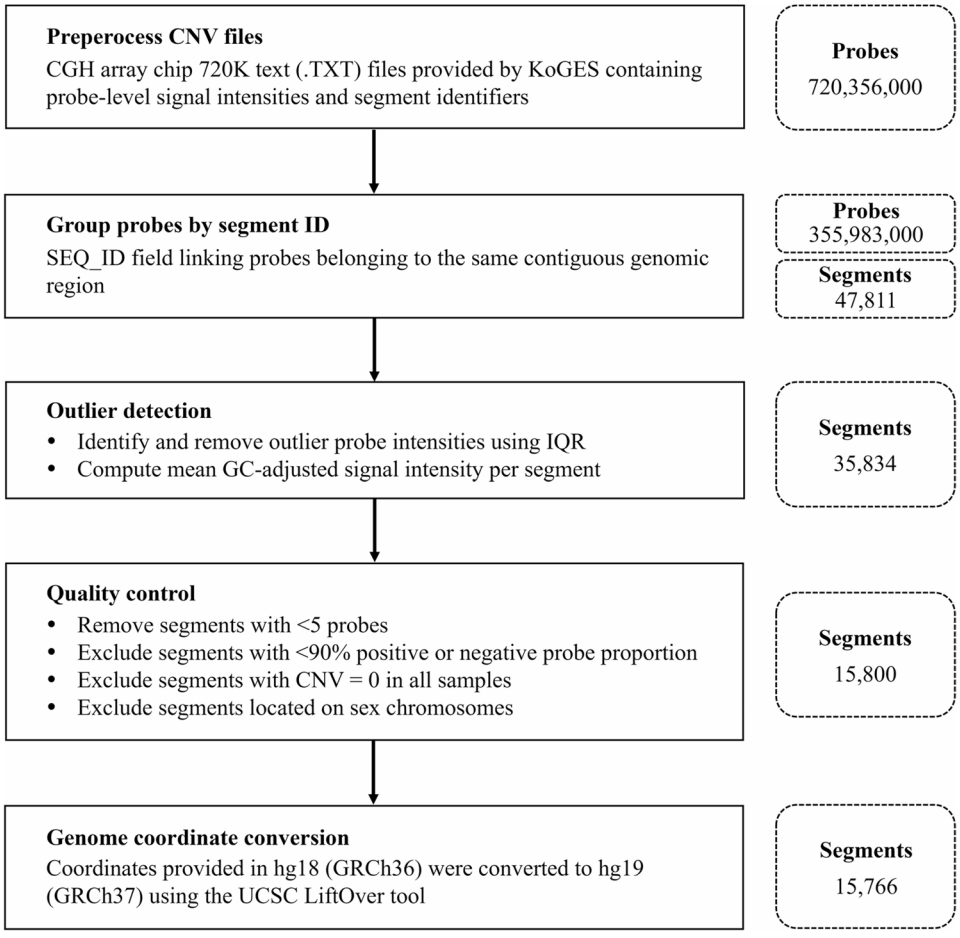


Fig. 2 Segment-level CNV preprocessing and quality-control workflow from 720 K CGH array data. The diagram summarizes the sequential filtering steps applied to probe-level signals and the resulting number of retained segments at each stage of CNV extraction. CGH, comparative genomic hybridization; CNV, copy number variation; IQR, interquartile range; KoGES, Korean Genome and Epidemiology Study; SEQ_ID, segment identifier

construct CNV burden groups, sex-specific tertiles were defined using quantile-based cut points, ensuring equal sample sizes in each tertile category (n = 182). CNV burden showed a right-skewed distribution without natural cut-points; therefore, tertile-based grouping was applied to ensure balanced sample sizes and facilitate interpretable descriptive comparisons, including Kaplan–Meier curves. Quantile-based grouping of genetic burden is widely used in genomic epidemiology, with studies commonly dividing burden indices into categories such as tertiles or quartiles, particularly when these indices exhibit skewed distributions. We emphasize that tertile categories were used only for visualization and secondary descriptive analyses, whereas all primary inferential models (linear regression and Cox models) treated CNV burden as a continuous variable.

Data analyses

The demographic characteristics of the individuals were classified and compared based on the presence or absence of MetS and CNV burden tertiles, respectively. For comparison of continuous variables, Student's t-test was used to analyze mean differences between two groups (MetS vs. non-MetS), while general linear model (GLM) was employed to test mean differences among the three CNV burden groups. Categorical variables were presented as frequency (%), and the difference among groups was tested using the Chi-square test. To examine the association between CNV burden and incidence of MetS during follow-up, multivariable logistic regression analyses were performed. Because participants with MetS at baseline were excluded, the binary MetS outcome used in logistic regression represents incidence of MetS during follow-up, not baseline prevalence. Incidence of MetS was defined as the first occurrence of MetS at any follow-up examination from baseline to the last available follow-up visit. Participants who did not develop MetS during follow-up were classified as non-MetS. Covariates included baseline age (continuous, years), sex (men, women), BMI (continuous, kg/m^2), alcohol intake (non-drinker, former drinker, current drinker), smoking status (non-smoker, former smoker, current smoker), physical activity (continuous, MET-h/week), energy intake (continuous, kcal/day), educational status (high school graduate or less, college graduate, graduate school or more), household income (< 2,000,000, 2,000,000 ~ 4,000,000, 4,000,000 ~ 6,000,000, ≥ 6,000,000), residential area (urban, rural), and the first two genetic principal components (PC1 and PC2). Multiple testing was addressed using the Benjamini–Hochberg false discovery rate (FDR) method with a 5% threshold [45]. All covariates were selected a priori based on established epidemiologic evidence regarding factors that influence MetS risk. No data-driven variable selection was performed; covariates were included to minimize confounding from demographic, lifestyle, and metabolic factors known to affect MetS development. To evaluate whether the association between CNV burden and the age at MetS onset followed a nonlinear pattern, we additionally fitted sex-stratified multivariable regression models that included both the linear and quadratic terms of onset age. Model fit was compared using the statistical significance of the quadratic term as well as Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC). Kaplan–Meier survival curves and log-rank tests were additionally conducted to compare the cumulative incidence of MetS across sex-specific CNV burden tertiles (low, medium, high) [46, 47]. Lastly, to assess whether gene-disrupting CNVs exerted stronger effects on MetS onset, we conducted a subgroup analysis restricted to deletion-type CNVs (heterozygous loss and deep deletion) that overlapped functional gene regions. Genomic intervals were intersected with RefGene-annotated gene bodies, exons, and UTRs to identify autosomal deletion segments affecting functional genomic elements. For each individual, a functional deletion burden was calculated as the total count of deletion-type CNVs overlapping these regions, and sex-stratified multivariable linear regression models were fitted adjusting for the same covariates described above. Time-to-event analyses were conducted using sex-stratified Cox proportional hazards models. Survival time was defined as the interval from baseline to the first occurrence of MetS, and participants who did not develop MetS were censored at their last available follow-up visit. CNV burden was quantified as the total number of non-diploid CNV events (DeepDel, HetLoss, Gain, or Amp) per participant. To examine the robustness of the association between CNV burden and incident MetS, four sequentially adjusted models were fitted. Model 1 adjusted for baseline age and population structure (PC1, PC2). Model 2 additionally controlled for socioeconomic and environmental variables (residential area, educational status, household income). Model 3 further included lifestyle factors (smoking status, drinking status, physical activity, and energy intake). Model 4 additionally adjusted for baseline BMI to assess potential mediation through adiposity. All covariates were selected a priori based on established associations with MetS and to ensure consistency with the analyses performed for onset age. All analyses were conducted using Python (version 3.12.7), and a two-sided $P < 0.05$ was considered statistically significant.

Results

Characteristics of the study participants based on incidence of MetS during follow-up

Table 1 compares general characteristics between individuals with and without MetS, stratified by men and women. A total of 287 men and 259 women were

Table 1 General characteristics of the study participants based on incidence of metabolic syndrome during follow-up

Variables	Men (*n* = 287)		*P*-value	Women (*n* = 259)		*P*-value
	MetS (*n* = 122)	Non-MetS (*n* = 165)		MetS (*n* = 128)	Non-MetS (*n* = 131)	
CNV (n, %)			0.3157			0.6479
Low CNV	38 (31.1)	50 (30.3)		43 (33.6)	51 (38.9)	
Medium CNV	34 (27.9)	59 (35.8)		45 (35.2)	44 (33.6)	
High CNV	50 (41.0)	56 (33.9)		40 (31.2)	36 (27.5)	
Age (years)	51.2 ± 8.9	51.5 ± 9.6	0.7755	54.8 ± 8.8	51.1 ± 9.2	0.0012
BMI (kg/m^2)	23.8 ± 2.4	23.8 ± 2.4	< 0.0001	24.3 ± 2.9	23.4 ± 2.9	0.0151
Waist circumference (cm)	82.3 ± 5.6	78.3 ± 6.03	< 0.0001	80.0 ± 7.5	75.7 ± 7.6	< 0.0001
Systolic blood pressure (mmHg)	124.5 ± 17.1	117.7 ± 18.7	0.0014	122.4 ± 17.7	111.6 ± 17.4	< 0.0001
Diastolic blood pressure (mmHg)	83.2 ± 10.9	79.5 ± 12.03	0.0071	80.0 ± 10.3	73.6 ± 10.6	< 0.0001
Triglyceride level (mg/dL)	177.2 ± 110.0	127.3 ± 64.1	< 0.0001	120.5 ± 44.0	103.6 ± 39.0	0.0012
Glucose level (mg/dL)	88.3 ± 22.4	84.3 ± 14.3	0.0845	82.0 ± 10.6	80.2 ± 10.4	0.1719
HDL cholesterol level (mg/dL)	44.4 ± 7.9	48.29 ± 9.91	0.0003	48.5 ± 9.6	51.9 ± 9.3	0.0041
Nutritional intake						
Total calories (kcal/day)	2182.9 ± 851.6	2005.0 ± 601.3	0.0500	1837.7 ± 704.1	1891.5 ± 774.5	0.9354
Residential area (n, %)			0.8213			0.0636
Anseong	55 (45.1)	71 (43.0)		82 (64.1)	68 (51.9)	
Ansan	67 (54.9)	94 (57.0)		46 (35.9)	63 (48.1)	
Smoking status (n, %)			0.6361			0.9644
Non-smoker	23 (18.9)	31 (18.8)		121 (94.5)	123 (93.9)	
Former smoker	36 (29.5)	57 (34.5)		2 (1.6)	2 (1.5)	
Current smoker	63 (51.6)	77 (46.7)		5 (3.9)	6 (4.6)	
Alcohol intake (n, %)			0.4682			0.8811
Non-drinker	25 (20.5)	31 (18.8)		89 (69.5)	93 (71.0)	
Former drinker	10 (8.2)	21 (12.7)		3 (2.3)	2 (1.5)	
Current drinker	87 (71.3)	113 (68.5)		36 (28.1)	36 (27.5)	
Physical activity						
MET (hours/week)	190.4 ± 101.1	171.4 ± 106.7	0.1253	181.5 ± 108.3	168.0 ± 94.2	0.2853
Household income, won/month (n, %)			0.1612			0.2680
< 2,000,000	65 (53.3)	100 (60.6)		97 (75.8)	87 (66.4)	
2,000,000 ~ 4,000,000	46 (37.7)	58 (35.2)		25 (19.5)	32 (24.4)	
4,000,000 ~ 6,000,000	6 (4.9)	6 (3.6)		4 (3.1)	10 (7.6)	
≥ 6,000,000	5 (4.1)	1 (0.6)		2 (1.6)	2 (1.5)	
Education level (n, %)			0.0363			0.2251
High school graduate or less	95 (77.9)	137 (83.0)		121 (94.5)	119 (90.8)	
College graduate	27 (22.1)	23 (13.9)		6 (4.7)	12 (9.2)	
Graduate school or more	0 (0.0)	5 (3.0)		1 (0.8)	0 (0.0)	

Continuous variables are expressed as mean ± standard deviation, while categorical variables are expressed as frequencies (%). The p-values were based on Student's t-test for continuous variables and the chi-square test for categorical variables. CNV were divided into three groups (Low, Medium, High) based on total CNV counts across all loci on autosomal chromosomes 1 through 22, using sex-specific tertile-based grouping

BMI Body mass index, *CNV* Copy number variation, *HDL* High-density lipoprotein, *MET* Metabolic equivalent of task, *MetS* Metabolic syndrome

included in the baseline assessment. CNV burden was categorized into sex-specific tertile groups (low, medium, and high). Across both sexes, the distribution of CNV tertile groups did not differ significantly between participants with and without MetS.

In men, BMI, waist circumference, systolic blood pressure, diastolic blood pressure, triglycerides, and HDL cholesterol levels displayed significant differences between MetS and non-MetS individuals (all $P < 0.05$), with those in the MetS group exhibiting higher adiposity, higher blood pressure, elevated triglyceride levels, and lower HDL cholesterol levels. Blood glucose did not differ significantly between the two groups. Total calorie intake was marginally higher in the MetS group ($P = 0.0500$). Residential area, smoking status, and alcohol intake did not show statistically meaningful differences. Furthermore, physical activity levels measured by MET-hours/week were comparable between the groups ($P = 0.1478$). Household income also did not differ between MetS and non-MetS men ($P = 0.1612$). Education level showed a statistically significant difference between MetS and non-MetS men ($P = 0.0363$). However, both groups were

predominantly composed of individuals with a high school education or less, and no clear trend in educational attainment was observed.

In women, age was significantly higher among individuals with MetS compared to those without ($P = 0.0012$). Similarly to men, women with MetS exhibited significantly higher BMI, waist circumference, systolic and diastolic blood pressure, and triglyceride levels, along with lower HDL cholesterol levels (all $P < 0.05$). Blood glucose levels did not differ between groups. Total calorie intake and physical activity were comparable between MetS and non-MetS women. Household income and education level also showed no statistically significant group differences.

Overall, anthropometric measures, blood pressure, and lipid metabolism markers were consistently associated with MetS across both sexes, whereas CNV tertile grouping, lifestyle factors, physical activity, and socioeconomic indicators exhibited limited discriminatory value in differentiating MetS status in this population.

Characteristics of the study participants based on CNV burden groups

Table 2 summarizes the characteristics of participants according to CNV burden tertiles (low, medium, and high), stratified by sex. In men, age differed significantly across CNV burden groups ($P = 0.0096$), with the medium CNV group showing the youngest mean age compared to the low and high CNV groups. BMI also varied significantly among groups ($P = 0.0069$). Values were lowest in the low CNV group and higher in both the medium and high CNV groups. Systolic ($P = 0.0407$) and diastolic blood pressure ($P = 0.0303$) differed significantly across CNV groups, with the high CNV group showing the highest levels. For metabolic components such as triglycerides, fasting glucose, and HDL-cholesterol, no significant differences were observed among CNV burden groups. Total calorie intake also did not differ significantly ($P = 0.5849$). However, residential area showed a highly significant difference among CNV burden groups ($P < 0.0001$), with the high CNV group having a greater proportion of Ansan residents compared to the low CNV group. Lifestyle factors including smoking status ($P = 0.3323$) and alcohol intake ($P = 0.0875$) did not differ significantly across groups. Physical activity differed significantly among men ($P = 0.0135$). The medium CNV group had the lowest MET levels, while the low CNV group reported the highest activity level. Household income also showed significant differences ($P = 0.0205$), with the lowest-income category being most prevalent in the low CNV group and least prevalent in the medium CNV group. In contrast, education level did not significantly differ across CNV burden categories in men ($P = 0.1337$), although the proportion of individuals with college education tended to increase from the low to high CNV groups.

In women, age, BMI, waist circumference, blood pressure, triglycerides, glucose, HDL-cholesterol, and total calorie intake did not differ significantly among CNV burden tertiles (all $P > 0.05$). Residential area showed a marked difference by CNV burden ($P < 0.0001$). The low CNV group had the highest proportion of residents from Anseong, whereas the high CNV group had the lowest. Smoking status ($P = 0.3739$) and alcohol intake ($P = 0.2261$) did not differ significantly across CNV groups, though current drinking tended to be more common in the high CNV group. Physical activity also did not significantly differ ($P = 0.2873$), although MET tended to be higher in the low CNV group compared to the medium and high CNV groups. Household income showed a significant difference across CNV tertiles ($P = 0.0277$): the proportion of participants with income below 2,000,000 KRW/month was highest in the low CNV group and lowest in the high CNV group. In contrast, education level did not significantly differ across CNV tertiles in women ($P = 0.2071$), although the proportion of high school graduate or less appeared lowest in the high CNV group.

CNV associated with MetS

Multivariable logistic regression identified 403 CNVs that showed nominal associations with MetS based on unadjusted *P*-values ($P < 0.05$), after adjusting for baseline age, sex, BMI, smoking status, drinking status, household income, educational level, physical activity, total energy intake, residential area, and two principal component scores. Among these, the five CNVs with the smallest *P*-values are presented in Table 3. The candidate CNVs that met the nominal significance threshold ($P < 0.05$), along with their corresponding effect estimates, confidence intervals, and FDR-adjusted *P*-values, are provided in Supplementary Table S1.

The top candidate variants were distributed across multiple chromosomes, including chromosomes 1, 2, 9, and 14. Several CNV regions overlapped with genes related to metabolic functions, including *AMY1A*, *AMY1B*, *AMY1C*, *AMY2A*, *AMY2B*, *ACTG1P4*, and *KDM4C*. Notably, the CNV nsv547016, which spans the *AMY1* gene cluster on chromosome 1, exhibited an increased ORs for MetS (OR = 3.98; 95% CI: 1.90–8.34) and the lowest nominal p-value (2.55×10^{-4}). Another candidate CNV within the broader amylase locus (nsv517215) also showed a positive association (OR = 3.91; 95% CI: 1.75–8.74). Both segments encompass genes from the *AMY1*/*AMY2* clusters, reinforcing the biological relevance of variation in locus. Four of the five CNVs demonstrated nominal *P*-values below 0.001.

Table 2 General characteristics of the study participants based on the presence of copy number variation burden

Variables	Men (*n* = 287)			*P*-value	Women (*n* = 259)			*P*-value
	Low CNV (*n* = 88)	Medium CNV (*n* = 93)	High CNV (*n* = 106)		Low CNV (*n* = 94)	Medium CNV (*n* = 89)	High CNV (*n* = 76)	
Age (years)	53.7 ± 9.5	49.5 ± 8.6	51.21 ± 9.3	0.0096	53.8 ± 9.4	53.2 ± 9.1	51.6 ± 8.9	0.2890
BMI (kg/m^2)	22.4 ± 2.5	23.3 ± 2.5	23.4 ± 2.4	0.0069	23.3 ± 2.6	24.0 ± 2.9	24.3 ± 3.1	0.0523
MetS status (n, %)				0.3157				0.6479
MetS group	50 (56.8)	59 (63.4)	56 (52.8)		51 (54.3)	44 (49.4)	36 (47.4)	
Non-MetS group	38 (43.2)	34 (36.6)	50 (47.2)		43 (45.7)	45 (50.6)	40 (52.6)	
Waist circumference (cm)	80.0 ± 6.3	79.9 ± 5.8	80.1 ± 6.4	0.9725	77.6 ± 7.6	77.8 ± 8.3	78.3 ± 7.6	0.8364
Systolic blood pressure (mmHg)	119.2 ± 15.7	117.9 ± 17.2	124.1 ± 20.7	0.0407	115.2 ± 17.9	116.4 ± 17.2	119.6 ± 20.1	0.2815
Diastolic blood pressure (mmHg)	80.4 ± 9.5	79.1 ± 11.2	83.4 ± 13.4	0.0303	76.1 ± 10.5	75.9 ± 10.2	78.5 ± 12.1	0.2255
Triglyceride level (mg/dL)	147.7 ± 77.5	155.6 ± 117.3	143.0 ± 70.0	0.6133	113.5 ± 45.9	105.9 ± 39.3	117.1 ± 40.8	0.2162
Glucose level (mg/dL)	84.0 ± 15.6	87.8 ± 25.0	86.1 ± 12.1	0.3629	81.4 ± 10.1	80.5 ± 6.7	81.3 ± 14.2	0.8312
HDL cholesterol level (mg/dL)	46.8 ± 10.8	46.6 ± 9.1	46.5 ± 8.2	0.9693	49.0 ± 9.8	50.6 ± 9.1	51.0 ± 9.9	0.3501
Nutritional intake								
Total calories (kcal/day)	2071.9 ± 645.8	2028.2 ± 677.8	2133.9 ± 817.4	0.5849	1903.9 ± 817.0	1895.6 ± 813.0	1780.8 ± 518.0	0.4993
Residential area (n, %)				< 0.0001				< 0.0001
Anseong	60 (68.2)	34 (36.6)	32 (30.2)		74 (78.7)	48 (53.9)	28 (36.8)	
Ansan	28 (31.8)	59 (63.4)	74 (69.8)		20 (21.3)	41 (46.1)	48 (63.2)	
Smoking status (n, %)				0.3323				0.3739
Non-smoker	16 (18.2)	17 (18.3)	21 (19.8)		87 (92.6)	87 (97.8)	70 (92.1)	
Former smoker	22 (25.0)	36 (38.7)	35 (33.0)		1 (1.1)	1 (1.1)	2 (2.6)	
Current smoker	50 (56.8)	40 (43.0)	50 (47.2)		6 (6.4)	1 (1.1)	4 (5.3)	
Alcohol intake (n, %)				0.0875				0.2261
Non-drinker	23 (26.1)	12 (12.9)	21 (19.8)		73 (77.7)	62 (69.7)	47 (61.8)	
Former drinker	6 (6.8)	15 (16.1)	10 (9.4)		2 (2.1)	1 (1.1)	2 (2.6)	
Current drinker	59 (67.0)	66 (71.0)	75 (70.8)		19 (20.2)	26 (29.2)	27 (35.5)	
Physical activity								
MET (hours/week)	205.5 ± 113.8	161.6 ± 103.3	173.5 ± 94.0	0.0135	186.4 ± 110.8	173.2 ± 95.4	161.7 ± 95.6	0.2873
Household income, won/month (n, %)				0.0205				0.0277
< 2,000,000	61 (69.3)	44 (47.3)	60 (56.6)		77 (81.9)	62 (69.7)	45 (59.2)	
2,000,000 ~ 4,000,000	22 (25.0)	43 (46.2)	39 (36.8)		14 (14.9)	22 (24.7)	21 (27.6)	
4,000,000 ~ 6,000,000	5 (5.7)	2 (2.2)	5 (4.7)		3 (3.2)	3 (3.4)	8 (10.5)	
≥ 6,000,000	0 (0.0)	4 (4.3)	2 (1.9)		0 (0.0)	2 (2.2)	2 (2.6)	
Education level (n, %)				0.1337				0.2071
High school graduate or less	78 (88.6)	75 (80.6)	79 (74.5)		89 (94.7)	84 (94.4)	67 (88.2)	
College graduate	10 (11.4)	16 (17.2)	24 (22.6)		5 (5.3)	4 (4.5)	9 (11.8)	
Graduate school or more	0 (0.0)	2 (2.2)	3 (2.8)		0 (0.0)	1 (1.1)	0 (0.0)	

Continuous variables are expressed as mean ± standard deviation, while categorical variables are expressed as frequencies (%). The p-values were based on the general linear model for continuous variables and the chi-square test for categorical variables. CNV were divided into three groups (Low, Medium, High) based on total CNV counts across all loci on autosomal chromosomes 1 through 22, using sex-specific tertile-based grouping

BMI Body mass index, *CNV* Copy number variation, *HDL* High-density lipoprotein, *MET* Metabolic equivalent of task

To account for multiple testing, we applied the Benjamini–Hochberg FDR procedure at a 5% threshold. None of the associations remained statistically significant after FDR correction. Accordingly, these CNVs are interpreted as candidate MetS-associated CNVs, reflecting suggestive associations that warrant replication in future studies. Genome-wide CNV association analysis did not identify any CNV segment surpassing the Bonferroni-corrected significance threshold to $0.05/15{,}766$ ($= 3.17 \times 10^{-6}$). The Manhattan plot demonstrated uniformly distributed test statistics, and the QQ plot showed minimal deviation from expectation ($\lambda GC = 1.213$), indicating no substantial inflation. Manhattan and QQ plots are shown in Supplementary Figures S2–S3.

Table 3 Top 5 copy number variations associated with metabolic syndrome

CNV region	Size (bp)	CNV ID	Gene	OR (95% CI)	*P*-value	FDR
chr1:104144265–104,220,453	76,188	nsv547016	*AMY1A, AMY1B, AMY1C, AMY2A*	3.98 (1.90–8.34)	2.55×10^{-4}	0.88
chr14:82499126–82,503,353	4227	nsv821436	-	0.73 (0.61–0.87)	4.26×10^{-4}	0.88
chr2:164643458–164,645,094	2049	nsv514101	-	0.05 (0.01–0.28)	5.24×10^{-4}	0.88
chr1:104136714–104,317,018	180,304	nsv517215	*ACTG1P4, AMY1A, AMY1B, AMY1C, AMY2A, AMY2B*	3.91 (1.75–8.74)	9.15×10^{-4}	0.88
chr9:6729261–6,729,815	555	nsv820480	*KDM4C*	2.56 (1.46–4.49)	1.06×10^{-3}	0.88

Adjusted for baseline age, sex, BMI, smoking, drinking, income, education, physical activity, energy intake, residential area, PC1, and PC2. FDR (Benjamini-Hochberg) adjusted *P*-value was applied (5% level). Database of genomic variants (https://www.ncbi.nlm.nih.gov/dbvar). CNV region were based on Database for Genomic Variants (Build GRCh37: Feb. 2009, hg19)

PC Principal component, *CNV* Copy number variation, *FDR* False discovery rate

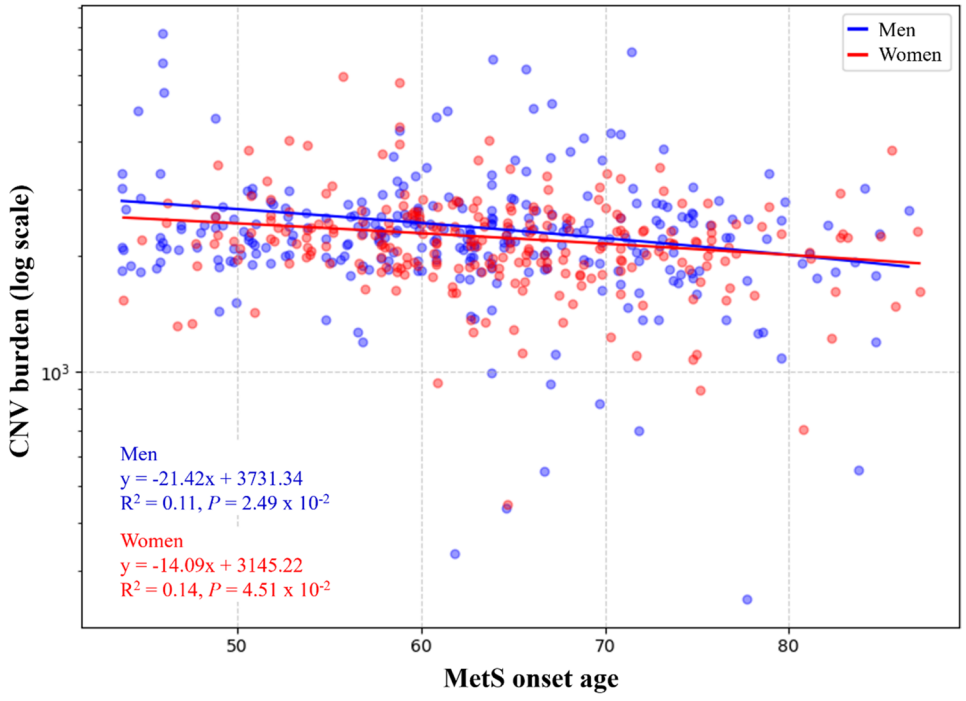


Fig. 3 Linear regression between copy number variations burden and onset age of metabolic syndrome. The x-axis (MetS onset age) indicates the age at which metabolic syndrome was first observed in each individual with MetS, while the y-axis (CNV burden) represents the total number of CNVs carried by th at individual. The CNV burden was log-transformed to normalize its skewed distribution and to facilitate more accurate modeling of its association with the age at MetS onset. The analysis was stratified by sex, and regression was adjusted for baseline age, BMI, smoking, drinking, income, education, physical activity, energy intake, residential area, PC1, and PC2. PC, principal component; BMI, body mass index; CNV, copy number variation; MetS, metabolic syndrome

Association between CNV burden and the onset age of MetS

As shown in Fig. 3, a multivariable linear regression analysis examines the relationship between CNV burden which was log-transformed to normalize the skewed distribution and more accurately model the association with age of MetS onset. The analysis was further stratified by sex after adjusting for baseline age, BMI, smoking status, drinking status, physical activity, energy intake, educational status, household income, residential area, PC1 and PC2. The statistical regression analysis showed that the slopes were significantly negative for both sexes (men: β = -21.42, $P=0.0249$; women: β = -14.09, $P=0.0451$). We further examined whether the association between CNV burden and MetS onset age followed a nonlinear (quadratic) pattern. In sex-stratified models including an additional squared term of onset age, the quadratic term was not statistically significant in either men ($P=0.71$) or women ($P=0.51$), and model fit indices (R^2, AIC, and BIC) were virtually unchanged compared with the linear model. These findings indicate that a linear specification adequately captures the relationship between CNV burden and the timing of MetS onset (Supplementary Figure S4). Subgroup analyses restricted to deletion-type CNVs overlapping functional gene regions yielded similar results, with higher gene-disrupting deletion burden associated with earlier MetS onset in both men (β =

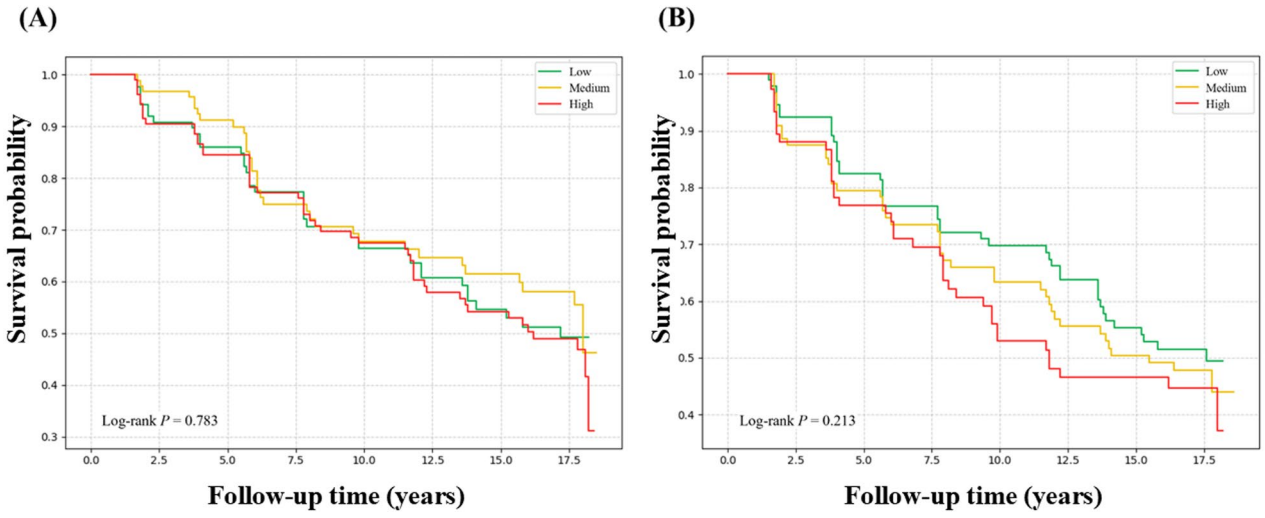


Fig. 4 Kaplan–Meier survival curves illustrating time to metabolic syndrome onset across copy number variations burden tertiles in men **A** and women **B**. A log-rank test was performed to evaluate whether the survival curves for the low and high CNV groups differed significantly. CNV were divided into three groups (low, medium, high) based on total CNV counts across all loci on autosomal chromosomes 1 through 22, using sex-specific tertile-based grouping. CNV, copy number variation; MetS, metabolic syndrome

Table 4 Adjusted HRs and 95% CIs of metabolic syndrome incidence based on copy number variation burden.

Men (*n* = 287)				**Women (*n* = 259)**			
Variable	**HR (95% CI)**		***P*-value**	**Variable**	**HR (95% CI)**		***P*-value**
	Non-MetS	**MetS**			**Non-MetS**	**MetS**	
Model 1	1 (Reference)	1.18 (1.01–1.37)	0.0378	Model 1	1 (Reference)	1.20 (0.97–1.48)	0.0981
Model 2	1 (Reference)	1.18 (1.01–1.39)	0.0399	Model 2	1 (Reference)	1.25 (1.00-1.55)	0.0470
Model 3	1 (Reference)	1.20 (1.01–1.42)	0.0397	Model 3	1 (Reference)	1.24 (1.00-1.54)	0.0550
Model 4	1 (Reference)	1.21 (1.01–1.45)	0.0366	Model 4	1 (Reference)	1.19 (0.95–1.49)	0.1300

Model 1: adjusted for baseline age, PC1, and PC2; Model 2: adjusted for baseline age, PC1, PC2, residential area, income, and education; Model 3: adjusted for baseline age, PC1, PC2, residential area, income, education, smoking, drinking, physical activity, and energy intake; Model 4: adjusted for baseline age, PC1, PC2, residential area, income, education, smoking, drinking, physical activity, energy intake, and BMI

PC Principal component, *BMI* Body mass index, *CI* Confidence interval, *CNV* Copy number variation, *HR* Hazard ratio

Significance was set at *P*-value < 0.05

−0.0087, $P = 0.044$) and women (β = −0.0122, $P = 0.010$) (Supplementary Table S2).

The Kaplan–Meier survival curves were generated to examine the cumulative incidence of MetS across tertiles of CNV burden separately in men and women (Fig. 4). Across both sexes, the three CNV burden groups (low, medium, high) exhibited broadly similar survival trajectories throughout the follow-up period. Although minor visual differences appeared—particularly a slightly earlier decline in survival probability in the high CNV burden group—these patterns were not statistically significant. In men, the log-rank test showed no significant difference in MetS-free survival among the three CNV burden groups ($P = 0.783$). Similarly, in women, survival curves did not diverge meaningfully, and the log-rank test indicated no significant group difference ($P = 0.213$).

Survival analysis using Cox proportional hazards models

To evaluate the prospective association between CNV burden and incident MetS, sex-stratified Cox proportional hazards models were fitted with sequential adjustment for demographic and lifestyle factors (Table 4).

Among men, higher CNV burden was consistently associated with an increased risk of developing MetS across all models. In the minimally adjusted model (Model 1: baseline age, PC1, PC2), CNV burden was associated with a higher hazard of incident MetS (HR = 1.18, 95% CI: 1.01–1.37; $P = 0.0378$). The association remained stable after additionally adjusting for residential area, income, and education (Model 2: HR = 1.18, 95% CI: 1.01–1.39; $P = 0.0399$) and after further controlling for lifestyle factors including smoking, drinking, physical activity, and energy intake (Model 3: HR = 1.20, 95% CI: 1.01–1.42; $P = 0.0397$). The association persisted when BMI was additionally included (Model 4: HR = 1.21, 95% CI: 1.01–1.45; $P = 0.0366$).

In women, the direction of association was similar, and statistical significance emerged in Model 2. CNV burden was associated with a higher risk of MetS incidence (HR = 1.25, 95% CI: 1.00–1.55; $P = 0.0470$). Although the fully adjusted model including BMI (Model 4) showed a weaker association (HR = 1.19, 95% CI: 0.95–1.49; $P = 0.1300$), the HR estimates consistently indicated a positive association between CNV burden and MetS incidence.

Overall, CNV burden demonstrated a robust and directionally consistent association with MetS incidence across multiple adjustment models, particularly in men.

Discussion

Our comprehensive analysis of 546 Korean adults aged ≥ 40 years demonstrates that CNV burden is significantly associated with the onset age of MetS. The primary objective of this study was not to predict MetS onset age but rather to investigate whether structural genomic variation contributes to inter-individual variation in disease timing. Using 20 years of follow-up data, we observed that individuals with a higher CNV burden developed MetS at an earlier age, suggesting that CNVs may influence susceptibility to earlier metabolic deterioration.

Traditionally, genetic studies on MetS and related metabolic traits have focused largely on SNP-based variation, assuming that most phenotypic diversity is attributable to single nucleotide differences [48, 49]. However, structural genetic variation, including CNVs, plays a substantial role in human evolution, gene expression, and disease phenotypes [15]. One of the most notable examples involves dietary adaptation: human populations with historically high starch intake exhibit higher *AMY1* copy numbers, reflecting positive selection on salivary amylase availability [18, 50–52]. This evolutionary evidence aligns with our finding that multiple MetS-associated CNV regions overlap with the *AMY1* and *AMY2* gene clusters.

CNVs are large-scale structural variants spanning ≥ 1 kb, including deletions, duplications, and more complex rearrangements [15, 53]. Accumulating evidence indicates that CNVs constitute a meaningful portion of genetic susceptibility to complex diseases [54]. Roughly 18% of gene expression variability is attributable to CNVs larger than 40 kb [55]. Furthermore, CNVs have been implicated in diverse disorders, including HIV susceptibility [56], autoimmune conditions [57], and asthma [58], in some cases conferring up to a 30% increase in disease risk [59]. Despite this evidence, there have been no prior investigations into whether CNV burden influences the timing of MetS onset, even though MetS is a classical polygenic disorder involving intricate metabolic regulation [9, 60].

To explore potential mechanisms underlying the CNV–MetS association, we examined phenotypic patterns across CNV burden tertiles. BMI increased progressively with CNV burden, consistent with previous findings that *AMY1*/*AMY2* CNVs, key determinants of carbohydrate metabolism, are associated with BMI, obesity, and metabolic traits. Additionally, Falchi et al. demonstrated that lower *AMY1* copy numbers predict obesity risk through decreased amylase expression and impaired starch metabolism [22], and subsequent studies have linked *AMY1A* CNV to glucose and lipid homeostasis, inflammatory markers, and gut microbiome composition [61]. Collectively, these findings support the role of *AMY1*/*AMY2* CNV variation as a metabolic modifier that may contribute to earlier MetS onset in individuals with high CNV burden.

Additional biological support stems from experimental evidence linking CNVs directly to metabolic phenotypes. Lacaria et al. showed that CNVs at chromosome 17p11.2 in mice produce MetS–like features, including increased adiposity, lower HDL, and reduced insulin sensitivity; these findings were corroborated in human SMS patients carrying deletion CNVs [62]. In our study, the recurrent identification of CNVs in *AMY1*/*AMY2* regions suggests a strong relationship with dietary carbohydrate utilization—a pathway known to affect metabolic risk. Similarly, the identification of CNV variation in *KDM4C* is notable, given its established role as a histone demethylase regulating metabolic gene expression via H3K9me3/H3K36me3 demethylation [23]. Through interactions with PPARγ, *KDM4C* modulates adipocyte differentiation [23, 24], influences cholesterol metabolism through ApoE regulation [25–27, 63], and cooperates with *ATF4* to regulate amino acid metabolic pathways [28]. These mechanisms provide an integrated framework through which *KDM4C* CNVs may modulate MetS development.

Sex-specific analyses further supported the contribution of CNV burden to variability in MetS onset. In the regression models, the inverse association between CNV burden and MetS onset age reached statistical significance in both men ($\beta = -21.42$, $P = 0.0249$) and women ($\beta = -14.09$, $P = 0.0451$). Although the effect sizes differed modestly between sexes, the consistent negative slopes suggest that higher CNV burden may accelerate the development of MetS across both groups. These findings indicate the potential biological relevance of structural genomic variation in shaping metabolic risk progression.

Our results are consistent with findings from early-onset obesity research showing that higher CNV burden is enriched in individuals with severe early metabolic phenotypes. Rare CNVs were significantly more common in severely obese children (BMI > + 3 SD) and were enriched in G protein–coupled receptors pathways affecting energy homeostasis [64]. Similarly, Finnish cohorts demonstrated CNV enrichment in early-onset obesity cases relative to controls [65]. These converging

lines of evidence highlight CNV burden as a meaningful genetic factor influencing early metabolic dysregulation.

Several limitations merit consideration. First, because the KoGES cohort begins at age 40, CNV-related effects contributing to very early-onset MetS could not be evaluated. Second, replication in larger or younger cohorts is required to confirm generalizability. Third, longitudinal CNV accumulation was not measured; thus, it is unclear whether CNV burden increases with age or precedes metabolic alterations.

Despite these limitations, our study has several strengths: it represents the first population-based investigation of CNV burden in relation to MetS onset timing; it incorporates 20-year longitudinal data from a well-characterized national cohort; and it provides mechanistic insights integrating carbohydrate metabolism, lipid regulation, and epigenetic modulation. In conclusion, our findings indicate that higher CNV burden is associated with earlier MetS onset in a Korean population. The identification of 403 MetS-related CNVs reaching nominal significance, including prominent loci at *AMY1A*/*AMY1B*/*AMY1C*/*AMY2A*/*AMY2B* and *KDM4C*, supports a role for structural genomic variation in shaping metabolic trajectories. These results highlight CNV burden as a potential biomarker for early metabolic risk stratification and underscore the need for further mechanistic and external validation studies.

Abbreviations

ApoE	Apolipoprotein E
ATP III	Adult Treatment Panel III
BMI	Body mass index
CGH	Comparative genomic hybridization
CI	Confidence interval
CNV	Copy number variation
CVD	Cardiovascular disease
DBP	Diastolic blood pressure
DM	Diabetes millitus
FBG	Fasting blood glucose
GWAS	Genome-wide association study
HDL	High density lipoprotein
HR	Hazard ratio
IRB	Institutional review board
KARE	Korea Association Resource
KCHIP	Korean Chip
KoGES	Korean Genome and Epidemiology Study
LDL	Low density lipoprotein
MET	Metabolic equivalent of task
MetS	Metabolic syndrome
NCEP	National Cholesterol Education Program
OR	Odds ratio
SBP	Systolic blood pressure
SNP	Single nucleotide polymorphism
SQ-FFQ	Semi-quantitative food frequency questionnaire

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12920-026-02353-2.

Supplementary Material 1.

Supplementary Material 2.

Acknowledgements

This study was conducted using bioresources from the National Biobank of Korea and the Korea Disease Control and Prevention Agency (NBK-2024-019).

Authors' contributions

Conceptualization: Shin D; investigation: Ko S-H, Kim M, Shin DS; supervision: Shin D; writing - original draft: Ko S-H, Kim M, Shin D; writing - review & editing: Ko S-H, Kim M, Shin D; funding acquisition: Shin D.

Funding

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. RS-2024-00340086).

Data availability

Data from the Korean Genome and Epidemiology Study are available through the procedure described at https://biobank.nih.go.kr/eng/.

Declarations

Ethics approval and consent to participate

The study protocol was reviewed and approved by the Institutional Review Board of Ewha University, Korea (IRB No. ewha-202603-0023-01). All KARE participants provided written informed consent prior to participation in the KoGES. All methods were carried out in accordance with relevant guidelines and regulations (Declaration of Helsinki).

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

[1]Department of Food and Nutrition, Inha University, Incheon 22212, Republic of Korea
[2]Department of Nutritional Science and Food Management, Ewha Womans University, 52, Ewhayeodae-gil, Seodaemun-gu, Seoul 03760, Republic of Korea

Received: 3 September 2025 / Accepted: 25 March 2026
Published online: 02 April 2026

References

1. Fahed G, Aoun L, Bou Zerdan M, Allam S, Bou Zerdan M, Bouferraa Y et al. Metabolic Syndrome: Updates on Pathophysiology and Management in 2021. Int J Mol Sci. 2022;23(2):786.
2. Saklayen MG. The Global Epidemic of the Metabolic Syndrome. Curr Hypertens Rep. 2018;20(2):12.
3. Pitliya A, Kakarlapudi Y, Vasudevan SS, Kancherla N, Kumar L, Cheruvu NP, et al. The Global Prevalence of Metabolic Syndrome in Connection with Sleep Duration: A Systematic Review and Meta-Analysis. Metab Syndr Relat Disord. 2024;22(6):411–21.
4. Kim H-J, Kang DR, Kim JY, Kim W, Jeong YW, Chun K-H, et al. Metabolic Syndrome Fact Sheet 2024: Executive Report. Cardiometab Syndr J. 2024;4(2):70–80.
5. Yoo JS, Choe EY, Kim YM, Kim SH, Won YJ. Predictive costs in medical care for Koreans with metabolic syndrome from 2009 to 2013 based on the National Health Insurance claims dataset. Korean J Intern Med. 2020;35(4):936–45.
6. Huang Z, Wang X, Ding X, Cai Z, Li W, Chen Z et al. Association of Age of Metabolic Syndrome Onset With Cardiovascular Diseases: The Kailuan Study. Front Endocrinol. 2022;13:857985.
7. Kraja AT, Borecki IB, North K, Tang W, Myers RH, Hopkins PN, et al. Longitudinal and age trends of metabolic syndrome and its risk factors: The Family Heart Study. Nutr Metabolism. 2006;3(1):41.

8. Tsay YC, Chen CH, Pan WH. Ages at Onset of 5 Cardiometabolic Diseases Adjusting for Nonsusceptibility: Implications for the Pathogenesis of Metabolic Syndrome. Am J Epidemiol. 2016;184(5):366–77.
9. van Walree ES, Jansen IE, Bell NY, Savage JE, de Leeuw C, Nieuwdorp M, et al. Disentangling Genetic Risks for Metabolic Syndrome. Diabetes. 2022;71(11):2447–57.
10. Chial H. Mendelian genetics: Patterns of inheritance and single-gene disorders. Nat Educ. 2008;1(1):63.
11. Prone-Olazabal D, Davies I, González-Galarza FF. Metabolic Syndrome: An Overview on Its Genetic Associations and Gene-Diet Interactions. Metab Syndr Relat Disord. 2023;21(10):545–60.
12. Lanktree M, Hegele RA. Copy number variation in metabolic phenotypes. Cytogenet Genome Res. 2008;123(1–4):169–75.
13. Yan YX, Li JJ, Xiao HB, Wang S, He Y, Wu LJ. Association analysis of copy number variations in type 2 diabetes-related susceptible genes in a Chinese population. Acta Diabetol. 2018;55(9):909–16.
14. Almal SH, Padh H. Implications of gene copy-number variation in health and diseases. J Hum Genet. 2012;57(1):6–13.
15. Hurles ME, Dermitzakis ET, Tyler-Smith C. The functional impact of structural variation in humans. Trends Genet. 2008;24(5):238–45.
16. Lee H-S, Moon S, Yun JH, Lee M, Hwang MY, Kim Y-J, et al. Genome-wide copy number variation study reveals KCNIP1 as a modulator of insulin secretion. Genomics. 2014;104(2):113–20.
17. Iacocca MA, Dron JS, Hegele RA. Progress in finding pathogenic DNA copy number variations in dyslipidemia. Curr Opin Lipidol. 2019;30(2):63–70.
18. Perry GH, Dominy NJ, Claw KG, Lee AS, Fiegler H, Redon R, et al. Diet and the evolution of human amylase gene copy number variation. Nat Genet. 2007;39(10):1256–60.
19. Carpenter D, Dhar S, Mitchell LM, Fu B, Tyson J, Shwan NA, et al. Obesity, starch digestion and amylase: association between copy number variants at human salivary (AMY1) and pancreatic (AMY2) amylase genes. Hum Mol Genet. 2015;24(12):3472–80.
20. Hasegawa T, Kakuta M, Yamaguchi R, Sato N, Mikami T, Murashita K, et al. Impact of salivary and pancreatic amylase gene copy numbers on diabetes, obesity, and functional profiles of microbiome in Northern Japanese population. Sci Rep. 2022;12(1):7628.
21. Zhan F, Chen J, Yan H, Wang S, Zhao M, Zhang S, et al. Association of Serum Amylase Activity and the Copy Number Variation of AMY1/2A/2B with Metabolic Syndrome in Chinese Adults. Diabetes Metab Syndr Obes. 2021;14:4705–14.
22. Falchi M, El-Sayed Moustafa JS, Takousis P, Pesce F, Bonnefond A, Andersson-Assarsson JC, et al. Low copy number of the salivary amylase gene predisposes to obesity. Nat Genet. 2014;46(5):492–7.
23. Lizcano F, Romero C, Vargas D. Regulation of adipogenesis by nuclear receptor PPARγ is modulated by the histone demethylase JMJD2C. Genet Mol Biol. 2011;34(1):19–24.
24. Sugii S, Olson P, Sears DD, Saberi M, Atkins AR, Barish GD, et al. PPARgamma activation in adipocytes is sufficient for systemic insulin sensitization. Proc Natl Acad Sci U S A. 2009;106(52):22504–9.
25. Zhu K, Zhang H, Luan Y, Hu B, Shen T, Ma B, et al. KDM4C promotes mouse hippocampal neural stem cell proliferation through modulating ApoE expression. Faseb j. 2024;38(5):e23511.
26. Ko SH, Kim HS. Menopause-Associated Lipid Metabolic Disorders and Foods Beneficial for Postmenopausal Women. Nutrients. 2020;12(1):202.
27. Tersigni C, Furqan Bari M, Cai S, Zhang W, Kandzija N, Buchan A, et al. Syncytiotrophoblast-derived extracellular vesicles carry apolipoprotein-E and affect lipid synthesis of liver cells in vitro. J Cell Mol Med. 2022;26(1):123–32.
28. Zhao E, Ding J, Xia Y, Liu M, Ye B, Choi JH, et al. KDM4C and ATF4 Cooperate in Transcriptional Control of Amino Acid Metabolism. Cell Rep. 2016;14(3):506–19.
29. Kim Y, Han B-G, group tK. Cohort Profile: The Korean Genome and Epidemiology Study (KoGES) Consortium. Int J Epidemiol. 2016;46(2):e20–e.
30. Kwon H, Kang M, Lim H. Influence of dietary cholesterol on metabolic syndrome risk in middle-aged Korean adults: using the Korean Genome and Epidemiology Study (KoGES). Lipids Health Dis. 2024;23(1):315.
31. Park SJ, Lyu J, Lee K, Lee HJ, Park HY. Nutrition survey methods and food composition database update of the Korean Genome and Epidemiology Study. Epidemiol Health. 2024;46:e2024042.
32. Alberti KGMM, Eckel RH, Grundy SM, Zimmet PZ, Cleeman JI, Donato KA, et al. Harmonizing Metabolic Syndrome Circulation. 2009;120(16):1640–5.
33. Haam JH, Kim BT, Kim EM, Kwon H, Kang JH, Park JH, et al. Diagnosis of Obesity: 2022 Update of Clinical Practice Guidelines for Obesity by the Korean Society for the Study of Obesity. J Obes Metab Syndr. 2023;32(2):121–9.
34. Driver JA, Djoussé L, Logroscino G, Gaziano JM, Kurth T. Incidence of cardiovascular disease and cancer in advanced age: prospective cohort study. BMJ. 2008;337:a2467.
35. Kim YJ, Go MJ, Hu C, Hong CB, Kim YK, Lee JY, et al. Large-scale genome-wide association studies in East Asians identify new genetic loci influencing metabolic traits. Nat Genet. 2011;43(10):990–5.
36. Cho YS, Go MJ, Kim YJ, Heo JY, Oh JH, Ban H-J, et al. A large-scale genome-wide association study of Asian populations uncovers genetic factors influencing eight quantitative traits. Nat Genet. 2009;41(5):527–34.
37. Przybytkowski E, Ferrario C, Basik M. The use of ultra-dense array CGH analysis for the discovery of micro-copy number alterations and gene fusions in the cancer genome. BMC Med Genom. 2011;4(1):16.
38. Gautier L, Cope L, Bolstad BM, Irizarry RA. affy—analysis of Affymetrix GeneChip data at the probe level. Bioinformatics. 2004;20(3):307–15.
39. Wang K, Li M, Hadley D, Liu R, Glessner J, Grant SF, et al. PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. Genome Res. 2007;17(11):1665–74.
40. Diskin SJ, Li M, Hou C, Yang S, Glessner J, Hakonarson H, et al. Adjustment of genomic waves in signal intensities from whole-genome SNP genotyping platforms. Nucleic Acids Res. 2008;36(19):e126.
41. Ryba T, Battaglia D, Pope BD, Hiratani I, Gilbert DM. Genome-scale analysis of replication timing: from bench to bioinformatics. Nat Protoc. 2011;6(6):870–95.
42. Genovese G, Rockweiler NB, Gorman BR, Bigdeli TB, Pato MT, Pato CN et al. BCFtools/liftover: an accurate and comprehensive tool to convert genetic variants across genome assemblies. Bioinformatics. 2024;40(2):btae038.
43. Mina M, Iyer A, Tavernari D, Raynaud F, Ciriello G. Discovering functional evolutionary dependencies in human cancers. Nat Genet. 2020;52(11):1198–207.
44. Willis J, Mukherjee S, Orlow I, Viale A, Offit K, Kurtz RC et al. Genome-wide analysis of the role of copy-number variation in pancreatic cancer risk. Front Genet 2014;5:29.
45. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J Roy Stat Soc: Ser B (Methodol). 1995;57(1):289–300.
46. Kaplan EL, Meier P. Nonparametric Estimation from Incomplete Observations. J Am Stat Assoc. 1958;53(282):457–81.
47. Mantel N. Evaluation of survival data and two new rank order statistics arising in its consideration. Cancer Chemother Rep. 1966;50(3):163–70.
48. Lee C, Cheon KS, Shin Y, Oh H, Jeong YM, Jang H et al. Development and Application of a Target Capture Sequencing SNP-Genotyping Platform in Rice. Genes (Basel). 2022;13(5):794.
49. Morcillo-Suarez C, Alegre J, Sangros R, Gazave E, de Cid R, Milne R, et al. SNP analysis to results (SNPator): a web-based environment oriented to statistical genomics analyses upon SNP data. Bioinformatics. 2008;24(14):1643–4.
50. Wu Y, Tao D, Wu X, Liu W, Cai Y. Diet of the earliest modern humans in East Asia. Front Plant Sci. 2022;13:989308.
51. Ströhle A, Hahn A. Diets of modern hunter-gatherers vary substantially in their carbohydrate content depending on ecoenvironments: results from an ethnographic analysis. Nutr Res. 2011;31(6):429–35.
52. Lebenthal E. Role of salivary amylase in gastric and intestinal digestion of starch. Dig Dis Sci. 1987;32(10):1155–7.
53. Freeman JL, Perry GH, Feuk L, Redon R, McCarroll SA, Altshuler DM, et al. Copy number variation: new insights in genome diversity. Genome Res. 2006;16(8):949–61.
54. Kun-Rodrigues C, Orme T, Carmona S, Hernandez DG, Ross OA, Eicher JD, et al. A comprehensive screening of copy number variability in dementia with Lewy bodies. Neurobiol Aging. 2019;75:e2231–10.
55. Stranger BE, Forrest MS, Dunning M, Ingle CE, Beazley C, Thorne N, et al. Relative impact of nucleotide and copy number variation on gene expression phenotypes. Science. 2007;315(5813):848–53.
56. Gonzalez E, Kulkarni H, Bolivar H, Mangano A, Sanchez R, Catano G, et al. The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. Science. 2005;307(5714):1434–40.
57. Fanciulli M, Norsworthy PJ, Petretto E, Dong R, Harper L, Kamesh L, et al. FCGR3B copy number variation is associated with susceptibility to systemic, but not organ-specific, autoimmunity. Nat Genet. 2007;39(6):721–3.
58. Brasch-Andersen C, Christiansen L, Tan Q, Haagerup A, Vestbo J, Kruse TA. Possible gene dosage effect of glutathione-S-transferases on atopic asthma:

using real-time PCR for quantification of GSTM1 and GSTT1 gene copy numbers. Hum Mutat. 2004;24(3):208–14.
59. Ionita-Laza I, Rogers AJ, Lange C, Raby BA, Lee C. Genetic association analysis of copy-number variation (CNV) in human disease pathogenesis. Genomics. 2009;93(1):22–6.
60. Rana S, Ali S, Wani HA, Mushtaq QD, Sharma S, Rehman MU. Metabolic syndrome and underlying genetic determinants-A systematic review. J Diabetes Metab Disord. 2022;21(1):1095–104.
61. Hariharan R, Mousa A, de Courten B. Influence of AMY1A copy number variations on obesity and other cardiometabolic risk factors: A review of the evidence. Obes Rev. 2021;22(6):e13205.
62. Lacaria M, Saha P, Potocki L, Bi W, Yan J, Girirajan S, et al. A duplication CNV that conveys traits reciprocal to metabolic syndrome and protects against diet-induced obesity in mice and men. PLoS Genet. 2012;8(5):e1002713.
63. Marais AD. Apolipoprotein E in lipoprotein metabolism, health and cardiovascular disease. Pathology. 2019;51(2):165–76.
64. Iakoubov L, Mossakowska M, Szwed M, Duan Z, Sesti F, Puzianowska-Kuznicka M. A common copy number variation (CNV) polymorphism in the CNTNAP4 gene: association with aging in females. PLoS ONE. 2013;8(11):e79790.
65. Pettersson M, Viljakainen H, Loid P, Mustila T, Pekkinen M, Armenio M, et al. Copy Number Variants Are Enriched in Individuals With Early-Onset Obesity and Highlight Novel Pathogenic Pathways. J Clin Endocrinol Metab. 2017;102(8):3029–39.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.