scientific reports

Check for updates

OPEN

Multipartite network analysis to identify environmental and genetic associations of metabolic syndrome in the Korean population

Ji-Eun Shin[1,5], Nari Shin[2,5], Taesung Park[3] & Mira Park[4]✉

Network analysis has become a crucial tool in genetic research, enabling the exploration of associations between genes and diseases. Its utility extends beyond genetics to include the assessment of environmental factors. Unipartite network analysis is commonly used in genomics to visualize initial insights and relationships among variables. Syndromic diseases, such as metabolic syndrome, are characterized by the simultaneous occurrence of various signs, symptoms, and clinicopathological features. Metabolic syndrome encompasses hypertension, diabetes, obesity, and dyslipidemia, and both genetic and environmental factors contribute to its development. Given that relevant data often consist of distinct sets of variables, a more intuitive visualization method is needed. This study applied multipartite network analysis as an effective method to understand the associations among genetic, environmental, and disease components in syndromic diseases. We considered three distinct variable sets: genetic factors, environmental factors, and disease components. The process involved projecting a tripartite network onto a two-mode bipartite network and then simplifying it into a one-mode network. This approach facilitated the visualization of relationships among factors across different sets and within individual sets. To transition from multipartite to unipartite networks, we suggest both sequential and concurrent projection methods. Data from the Korean Association Resource (KARE) project were utilized, including 352,228 SNPs from 8840 individuals, alongside information on environmental factors such as lifestyle, dietary, and socioeconomic factors. The single-SNP analysis step filtered SNPs, supplemented by reference SNPs reported in a genome-wide association study catalog. The resulting network patterns differed significantly by sex: demographic factors and fat intake were crucial for women, while alcohol consumption was central for men. Indirect relationships were identified through projected bipartite networks, revealing that SNPs such as rs4244457, rs2156552, and rs10899345 had lifestyle interactions on metabolic components. Our approach offers several advantages: it simplifies the visualization of complex relationships among different datasets, identifies environmental interactions, and provides insights into SNP clusters sharing common environmental factors and metabolic components. This framework provides a comprehensive approach to elucidate the mechanisms underlying complex diseases like metabolic syndrome.

Keywords Environment, Genome-wide association study, Metabolic Syndrome, Multipartite network, Projection, Tripartite network

Metabolic syndrome (MetS) is defined as a cluster of metabolic abnormalities conditions, including abdominal obesity, hypertension, diabetes, and dyslipidemia[1,2]. The prevalence of MetS has been reported as 20–25% worldwide[3,4]. MetS is known to be associated with an increased risk of type 2 diabetes mellitus, cardiovascular

[1]Department of Biomedical Informatics, Konyang University, Daejeon, Republic of Korea. [2]Department of Statistics, Korea University, Seoul, Republic of Korea. [3]Department of Statistics, Seoul National University, Seoul, Republic of Korea. [4]Department of Preventive Medicine, Eulji University, Daejeon, Republic of Korea. [5]These authors contributed equally: Ji-Eun Shin and Nari Shin. ✉email: mira@eulji.ac.kr

disease, and premature mortality. The individual components of MetS are also known to be important risk factors for cardiovascular diseases[5–7].

Various attempts have been made to discover the genetic risk factors of MetS. The heritability of MetS has been estimated at over 30%[8–10]. Many genes and variants associated with MetS have been identified through genome-wide association studies (GWAS)[11–14]. Many studies have also sought to find genetic variants associated with each MetS component by population. For example, variants near insulin receptor substrate 1 were found to be associated with various traits of MetS, such as insulin resistance, HDL cholesterol, and triglycerides in a French population[15]. *GCKR* has been reported to be associated with fasting glucose and insulin levels in individuals of European ancestry[16,17]. *UGT1A1* has been reported to impact MetS in both men and women in a Mediterranean population[18]. In the Korean population, *CCDC63, LPL, MYL2*, and *APOA5* were found to be associated with MetS[14,19]. The number of variants associated with MetS continues to increase[1,2].

Meanwhile, studies on pleiotropic single-nucleotide polymorphisms (SNPs) for MetS-related traits have also been conducted. Kraja et al.[20] reported that the same loci were associated with more than one MetS-related trait. Based on pleiotropic associations, their research revealed relationships between SNPs, lipids, inflammation, and obesity[20]. More recently, pleiotropic SNPs and genes related to type 2 diabetes and obesity have been identified by applying genetic analyses incorporating pleiotropy and annotations using GWAS datasets[21]. A study found that *IGF2BP2* and *TNFRSF13B* predisposed individuals to MetS from a pleiotropic standpoint[22]. These results suggest that examining pleiotropy among metabolic traits is essential.

Since MetS is a multifactorial disease, environmental factors influence MetS. Numerous studies have investigated the influence of both genetic and environmental factors on the development of MetS. Specifically, environmental factors including dietary patterns, physical activity levels, and smoking status have been extensively explored[23–29]. For instance, a sedentary lifestyle and consumption of energy-dense diets have been linked to patterns in the clustering of different MetS traits MetS[30]. Moreover, research indicates that weight loss and increased physical activity are prioritized over pharmacological interventions in managing MetS[31]. Similarly, risk factors related to overnutrition and sedentary behavior have been identified as significant contributors to MetS, alongside a genetic predisposition[27]. Furthermore, a study highlighted the substantial role of various environmental factors, including diet, physical inactivity, stress, education levels, exposure to pollutants, and addictive behaviors, in the development of obesity-related MetS[28]. Recent investigations in Korea have explored the associations between environmental factors—such as sleep duration, sedentary behavior, alcohol consumption, smoking habits, and dietary patterns—and the risk of developing MetS. These studies have reinforced the observation that individuals with unhealthy lifestyle habits are more prone to developing MetS[32].

Several studies have emphasized the importance of simultaneously considering both environmental and genetic factors[8,30,31,33–37]. For instance, a multivariate genetic analysis was conducted on nine endophenotypes associated with MetS, utilizing twin data to identify common genetic and environmental factors[37]. Additionally, Prone-Olazabal et al.[36] provided an updated perspective on the genetics of MetS as a cohesive entity, examining SNPs and gene-diet interactions concerning cardiometabolic markers. In light of the understanding that genetic interactions intersect with an individual's environment, the distinction between genetic disorders and traits from environmental influences remains challenging[35].

Network analysis has recently been used for genetic data to investigate disease-gene associations[38–40]. A network is a collection of nodes and edges connecting the nodes. It can be used to visualize biological processes by taking biological entities such as genes, proteins, and diseases as nodes and representing the relationships between the entities by edges[41,42]. One-mode unipartite network analysis for each variable set or the whole variable set is widely used for genomic data.

Since it is not easy to investigate complex relationships through statistical models, we consider a more intuitive representation via a smart visualization method. Among several visualization methods, multipartite network analysis has the advantage of enabling researchers to easily grasp the relationships among genes, environments, and diseases. A multipartite network, often referred to as a k-partite graphs, can be seen as a complicated form of a network. The distinctive characteristic of a k-partite network is that the nodes can be divided into k disjoint sets. The edges do not connect nodes in the same set; instead, they only link nodes in different sets[43].

We applied tripartite network analysis for the case of $k = 3$, considering that there are three different variable sets relevant to MetS—namely, MetS components, environmental factors, and genetic factors. We considered dichotomous variables for the diagnosis of metabolic syndrome as MetS components, demographic variables, and dietary habits as environmental factors, and selected SNPs from GWAS data as genetic factors. To represent the relationship between two sets of variables, we used projections with weights[38]. A tripartite network was projected onto a two-mode bipartite network, and the projected bipartite network was projected again onto a one-mode network with the least loss of information. Through this procedure, we could visualize not only the relationship among factors in the different sets but also the compressed relationship among factors within the sets.

Materials and methods

Data

We used data from the Korean Association Resource (KARE) project (http://biobank.nih.go.kr). This project, a part of the Korean Genome Epidemiological Study (KoGES), started in 2007 and is still in progress. The data comprise two community-based cohorts from a rural area (Ansung) and an urban area (Ansan). The cohorts consist of community dwellers and participants recruited from the national health examinee registry. For baseline recruitment, eligible participants were asked to volunteer. Participants completed consent forms and then underwent surveys and examinations to assess their current health status and lifestyle habits. Anthropometric and clinical measurements such as weight, height, waist circumference, and blood pressure were measured. Human biological materials, including blood, urine, and DNA, were collected for analysis. The data include information

on genetic variants and environmental factors affecting chronic diseases such as type 2 diabetes, hypertension, obesity, MetS, osteoporosis, cardiovascular disease, and cancer in Koreans[44]. All participants provided their written informed consent to participate in this study. All methods were carried out following relevant guidelines and regulations (Declaration of Helsinki). This study was approved by the Institutional Review Board (IRB) of Eulji University (EU21-003-01).

Among the participants, 10,030 samples from individuals aged between 40 and 69 were genotyped using an Affymetrix Genome-Wide Human SNP Array 5.0. Quality control for the samples and genotypes was performed as previously described by Cho et al.[44]. SNPs with minor allele frequencies (< 0.01), low genotype calling rates (< 95%), and violation of Hardy–Weinberg equilibrium (p-values < 1E − 06) were removed. Participants whose sex/gender did not match or had a low calling rate (< 95%) were excluded. After quality control, 352,228 SNPs in 8840 individuals remained.

To diagnose MetS, the National Cholesterol Education Program (NCEP) Adult Treatment Panel III (ATP III) criteria are widely used[45]. These include different criteria for the Asian population[1,46]. A person who has three or more of the five MetS components is diagnosed with MetS. The five metabolic syndrome components are hypertension (> 130/85 mmHg), abdominal obesity (a waist circumference of ≥ 90 cm in Asian-American men, and ≥ 80 cm in Asian-American women), elevated triglycerides (≥ 150 mg/dL), reduced plasma high-density lipoprotein cholesterol (HDL-C; < 40 mg/dL in men and < 50 mg/dL in women), and impaired glucose tolerance (> 100 mg/dL). We followed these criteria and obtained five dichotomous variables as MetS components.

We also considered 10 variables of demographic characteristics, lifestyle factors, and dietary habits as environmental factors. The demographic variables comprised age, education level, and monthly household income. As lifestyle factors, we analyzed alcohol consumption, smoking, and physical activity (metabolic equivalents of task). The participants were questioned by trained interviewers regarding their socio-demographic status (age, education, household income) and lifestyle (diet, smoking, alcohol consumption, physical activity). Education level was categorized into six groups, and monthly household income was classified into eight groups. In the analysis, low-frequency items were integrated and finally, the education and household income items were reduced to four items and three items, respectively.

Protein, carbohydrates, and fat intake, as well as total energy, were used as variables for dietary habits. For dietary assessment, a food-frequency questionnaire (FFQ) involving 103 semi-quantitative items was developed[47]. Information regarding the protocol of the FFQ has been described elsewhere[48]. The frequencies of food consumption were categorized into nine groups, ranging from "rarely" to "more than three times per day." Portion sizes for each food item could be selected from three options: "small", "medium", or "large". The duration of seasonal fruit intake was classified into four categories (3, 6, 9, and 12 months). To assess the overall intake of nutrients such as protein and carbohydrates, the consumption frequency of each food item was multiplied by its nutrient content using the CAN-Pro 2.0 nutrient database developed by the Korean Nutrition Society[49]. Subsequently, the amounts of macronutrients were converted into calories, and the percentages of total calorie intake from each macronutrient were calculated. More details on the KoGES cohort profile can be found in Ref.[50]. The data description is summarized in Table 1.

Methods

Foundations of multipartite network analysis

A multipartite network or a k-partite network consists of mutually exclusive sets of nodes. Edges can exist only between nodes belonging to different sets. A graph is called k-partite if it can be partitioned into k nonempty, vertex-disjoint, edgeless subgraphs[40]. A k-partite graph can be represented as $G = (V, E)$, where

Chip	Affymetrix genome-wide human SNP array 5.0
Number of SNPs	352,228
Number of samples	8840
Cohort	
Urban area	Ansan
Rural area	Ansung
Environmental factors	
Demographic	Age, education level, household income
Lifestyle	Alcohol consumption, smoking, physical activity
Dietary habits	Protein intake, carbohydrate intake, fat intake, total energy
MetS diagnosis criteria	
Abdominal obesity	WC ≥ 90 for men, WC ≥ 80 for women
Triglycerides	≥ 150 mg/dL
HDL-C	< 40 mg/dL in men, < 50 mg/dL in women
Hypertension	> 130/85 mmHg
Fasting glucose	> 100 mg/dL

Table 1. Summary of the data. *SNP* single-nucleotide polymorphism, *MetS* metabolic syndrome, *HDL-C* high-density lipoprotein cholesterol, *WC* waist circumference.

V and E represent vertices and edges satisfying $V = V_1 \cup V_2 \cup \cdots \cup V_k$ and $V_i \cap V_j = \emptyset$ for $i \neq j$ and $E = \{(u, v) : u \in V_i, v \in V_j, i \neq j\}$, respectively[39].

There are two types of multipartite networks: closed and open networks. While a closed network has no restriction on its structure, an open network does not allow a circular structure. The adjacent matrix for a k-partite graph is given by

$$A = \begin{bmatrix} 0 & A_{12} & A_{13} & \cdots & A_{1k} \\ A_{12}^T & 0 & A_{23} & \cdots & A_{2k} \\ A_{13}^T & A_{23}^T & 0 & \cdots & A_{3k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ A_{1k}^T & A_{2k}^T & A_{3k}^T & \cdots & 0 \end{bmatrix} \tag{1}$$

for a closed network, and

$$A^* = \begin{bmatrix} 0 & A_{12} & 0 & \cdots & 0 \\ A_{12}^T & 0 & A_{23} & \cdots & 0 \\ 0 & A_{23}^T & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & A_{3k}^T & \cdots & 0 \end{bmatrix} \tag{2}$$

for an open network, respectively. Here, A_{ij} is a rectangular matrix called an incidence matrix. The (m, n)-th element of A_{ij} is 1 if there is an edge between vertices m of part i and n of part j, and 0 otherwise. Networks can also be classified into directed and undirected networks. As the terms indicate, vertices in a directed network are connected by directed edges, while the nodes of an undirected network are interconnected.

To understand the structure of a multipartite network, various measures can be employed. Degree distribution, where the degree of a node represents the number of edges it connects to other nodes, provides insights into the network's structure[35,51]. Connectivity measures the minimum number of vertices required to separate remaining nodes, indicating strong or weak graph linkage[42,52,53]. Closeness centrality gauges a node's proximity to others by calculating the inverse of the average shortest distance to all nodes. Betweenness centrality quantifies a node's importance by assessing its role in shortest paths[54]. Nodes with high closeness or betweenness centrality act as significant hubs. Additionally, the clustering coefficient indicates the likelihood of neighboring nodes being connected[54].

When k = 2, the network is a bipartite network. From a bipartite network, a one-mode projection can be created to compress the network and reveal connections within one dataset[55]. This results in two one-mode projections for each dataset:$P_1 = A_{12}^T A_{12}$ and $P_2 = A_{12} A_{12}^T$, where A_{12} is a bi-adjacency matrix encoding the edges from the first dataset to the second dataset. Similarly, a k-partite network produces k different (k − 1)-mode projections by consolidating information across the remaining set. However, a multi-stage projection onto the $k - i\,(i > 1)$ mode from a k-partite network is not well established. Assigning weights, which can be simple, hyperbolic, or resource allocation-based, to edges can reduce information loss during the projection process[38,56,57].

We propose utilizing k-partite networks to elucidate the complex relationship among genes, environments, and disease components in syndromic diseases. There is potential for k-partite networks to be applied in various fields, but no research has yet used this method to integrate multiple aspects of genetics, environment, and disease. We provide a series of analysis processes and propose concurrent and sequential projections to offer various visualizations of hidden relationships.

Implementing multipartite network analysis in GWAS

We employed a multipartite network to identify the environmental and genetic associations for a syndromic disease. We considered three distinct datasets: genetic factors, environmental factors, and MetS components. As genetic factors, we used SNPs. Ten environmental factors— E0 (age), E1 (education), E2 (income), E3 (alcohol), E4 (smoking), E5 (physical activity), E6 (total energy), E7 (protein intake), E8 (fat intake), E9 (carbohydrate intake)—were considered, as in previous research[29]. Five components of MetS—MetS1 (abdominal obesity), MetS2 (triglycerides), MetS3 (HDL-C), MetS4 (hypertension), and MetS5 (fasting glucose)—were considered as variables in the disease dataset. All analyses were stratified by sex since it is known that there are sex differences in metabolic homeostasis[58,59].

The procedure for the analysis was as follows:

(a) Variable selection: To reduce the number of SNPs in GWAS data, we performed a single SNP analysis using logistic regression. Age and area were considered as covariates. To address the issue of multiple comparisons, we adjusted the p-values using the Bonferroni correction. Since the number of SNPs selected based on this criterion was not sufficient, we used a less stringent threshold of p < 1E − 05. By filtering with this threshold, 42 SNPs for women and 57 SNPs for men were selected.

(b) Addition of reference SNPs: To improve the validity of our study, we also included 131 referenced SNPs that were reported to affect each component of MetS in a GWAS catalog (https://www.ebi.ac.uk/gwas/home). We used the five components of MetS as search terms and targeted studies focusing on Asian populations. After reviewing the content of the selected papers, we retrieved a list of relevant SNPs. Using the same threshold of p < 1E − 05, significant SNPs were selected. Excluding overlapping SNPs, 168 SNPs for men and 160 SNPs for women were used in the analysis.

(c) Construction of an adjacency matrix A, as shown in Eq. (1): The incidence matrix A_{ij} was set based on Pearson correlation coefficients. The *(m,n)*-th elements of A_{ij} were set to 1 if the correlation coefficient between the m-th variable in dataset i and the n-th variable in dataset j is significant ($p < 0.001$), and 0 otherwise. Through this procedure, 67 nodes for women and 65 nodes for men in the genetic factor set remained. The environmental factor set and MetS component set still had 10 and 5 nodes, respectively.

(d) Building the tripartite network: A tripartite graph was drawn using the adjacency matrix. To represent the strength of the connections between nodes, correlation coefficients were used as the weights of edges. An undirected and closed network was created.

(e) Projection to a bipartite network: We constructed two-mode projections composed of two sets of variables using projection from a tripartite network. They were connected by an edge if they shared a common variable in a third dataset. For example, if an SNP and a MetS component were connected in the two-mode projection, they shared at least one environmental factor. We used the simple weighting method—that is, the strength of the connection between two nodes is proportional to the number of nodes that they shared in the original graph. In total, three two-mode projections were created.

(f) Projection to a unipartite network: Unlike one-level projection, a method for conducting a multi-stage projection onto $k - i\,(i > 1)$ from a k-partite network has not been well established. There can be various paths from a k-partite to a unipartite network. We proposed two types of projections for obtaining unipartite projections: sequential projection and concurrent projection.

A. Sequential projection: A $(k - i + 1)$-mode projection is compressed to a to $(k - i)$-mode projection by aggregating information over the remaining set for $i = 1, \cdots, k - 1$. That is, for the i-th stage, a $(k - i)$ -mode network is constructed by connecting two nodes within the same dataset if they share at least one node in a different dataset on the $(k - i + 1)$-mode projection. For example, if we have three different datasets, three two-mode bipartite networks are produced in the first stage, and three one-mode unipartite projections are obtained for each two-mode projection in the second stage. The final network varies depending on its route of derivation.

B. Concurrent projection: A k-partite network is compressed to a unipartite network at once. To draw a unipartite network of one set, the nodes of the other set are treated as if they belong to the same dataset. For example, if we have three different datasets, only one one-mode projection is obtained for each dataset.

To create a unipartite network via sequential projection, a two-mode projection is compressed in a similar way to (d). In this process, nodes within the same dataset are connected if they share at least one node in another dataset on the two-mode projection, resulting in three one-mode projections per two-mode projection. Concurrent projection compresses a tripartite network directly into a unipartite network without utilizing method (d). Here, nodes from the other set are considered part of the same dataset, yielding three one-mode projections in total.

(g) Construction of one-mode projections from (d). In the network, nodes in the same dataset were connected. For the above procedure (c)–(e), separate networks were established for men and women.

Single SNP association testing was performed using PLINK (http://pngu.mgh.harvard.edu/~purcell/plink/). To draw a multipartite network, the *igraph* package in R can be used. Since the *igraph* package does not provide projections of tripartite networks, we modified the algorithm to enable projections using simple weights.

Results

Descriptive statistics

Table 2 demonstrates the descriptive statistics for participants by sex. Although the average age of men and women was similar, there were significant differences in every environmental factor ($p < 0.001$). Each component of MetS showed a higher proportion in women than in men except for MetS5, and a particularly high value for MetS1 was found in women. The overall proportion of individuals with MetS was 32.8%. Therefore, we constructed a separate network for each sex.

Constructing tripartite networks

Figure 1 shows the tripartite network using the data from men. The nodes that had weak connections with other nodes ($p \geq 0.001$) were eliminated in the drawing process. Total of 80 nodes were used to draw the network. The Rs numbers of the SNPs used in the graph are listed in Supplementary Table 1. In the graph, MetS3 (HDL-C) seemed to have the most connections, followed by MetS2 (triglycerides). E3 (alcohol) was located at the center connecting metabolic components except for MetS1 (abdominal obesity). A group of SNPs, including S64–S67, S70–S73, S76, S78, S105–S107, and S110, showed connections to both MetS2 and MetS3. It is remarkable that MetS1 had no direct connection with most SNPs except for S20 and was mainly related to nutritional factors such as E6 (total energy), E7 (protein intake), E8 (fat intake), and E9 (carbohydrate intake).

The node with the largest degree (i.e., the node that was connected to the most nodes) was MetS3, with a degree of 42. MetS2 had the second-highest degree (34). Among the nodes in the environmental factor set, E3 (alcohol) showed the largest degree (16), and among the nodes in the SNP set, S129 (rs12903590) and S130 (rs4821116) showed the largest degree (7). S129 is mapped to the *ALDH1A2* gene and has been reported to be related to HDL-C levels[60,61]. S130 has been reported to be located in *UBE2L3* and related to hepatitis B virus infections and HDL-C levels[62,63]. The nodes with large degrees also showed high centrality. A node with high

	Male	Female
Number of samples	4182 (47.3%)	4658 (52.7%)
E0: Age (year)	51.78 ± 8.78	52.61 ± 9.01
E1: Education		
≤ Elementary school	834 (20.0%)	2074 (45.0%)
Middle-high school	2439 (58.6%)	2262 (49.1%)
≥ College	888 (21.4%)	271 (5.9%)
E2: Household income (monthly)		
< 1000 USD**	556 (13.4%)	1125 (24.8%)
1000–1999 USD	1209 (29.2%)	1498 (33.0%)
2000–2999 USD	1458 (35.2%)	1306 (28.8%)
≥ 3000 USD	918 (22.2%)	611 (13.5%)
E3: Alcohol (g/day)	18.96 ± 28.8	1.31 ± 5.8
E4: Smoking (pack-years)	19.43 ± 18.4	0.41 ± 2.9
E5: Physical activity (MET min/week)	9747.33 ± 6450.82	9203.32 ± 6228.25
E6: Total energy(kcal/day)	2029.71 ± 654.08	1887.96 ± 742.03
E7: Protein intake (g)	70.02 ± 28.44	63.60 ± 30.99
E8: Fat intake (g)	36.13 ± 20.77	29.34 ± 20.90
E9: Carbohydrate intake (g)	350.59 ± 106.93	338.13 ± 127.63
MetS1: Abdominal obesity	803 (19.2%)	2500 (53.7%)
MetS2: Triglyceride	2032 (48.6%)	1666 (35.8%)
MetS3: HDL-C	1549 (37.1%)	3267 (70.2%)
MetS4: Hypertension	1952 (46.7%)	1912 (41.1%)
MetS5: Fasting glucose	731 (18.6%)	508 (11.6%)

Table 2. Basic statistics for participants by sex. Values are presented as mean ± SD or number (%). *1 USD ≈ 1000 KRW.

closeness centrality tends to be in the center of the network, while many other nodes are connected. In contrast, a node with high betweenness centrality builds a bridge that connects a lateral node and a central node rather than being connected to many nodes. MetS3 showed the highest closeness centrality (0.0067) followed by MetS2, (0.0064), E3 (0.0064), S129 (0.0063), and S130 (0.0063). For betweenness centrality, MetS3 (1442.06), MetS2 (1256.88), and MetS4 (hypertension;492.39) showed high values. E3 (373.67) and E9 (370.14) also showed high betweenness centrality. Thus, these nodes played the role of hubs in the network for men.

The thickness of the edges denotes the strength of the connection between the nodes. Not only S129 and S130, but also S127 (rs17411126) and S138 (rs6805251) showed strong connection with E3 in Fig. 1. S127 is mapped to the *LPL* gene and is known to be related to the cholesterol ratio in the Korean population[64]. S138 is mapped to the *GSK3B* gene and has been reported to be associated with HDL-C[65]. Table 3 shows the top five edges based on the absolute value of the correlation coefficients and their p-values. All relationships for which the absolute value of correlations was greater than 0.10 are presented in Supplementary Table 2. The information obtained from the graph is confirmed.

Similarly, Fig. 2 represents the tripartite network for women. In this network, after eliminating the nodes with weak correlations ($p \geq 0.001$), 82 nodes were used to form the network. The same process as with the data from men was conducted to filter the significant nodes. Instead of E3 (alcohol), which played an important role in the network for men, socioeconomic variables such as E0 (age), E1 (education), E2 (income), and E8 (fat intake) were located at the center, connecting various MetS components in the network for women. Moreover, these showed strong connections. As in the network for men, E3 was linked to MetS3, and there were several SNPs (S104-S107) linking MetS2 and MetS3, playing the role of bridges. S120 (rs10899345) connected MetS4 to the environmental variables of E0, E1, E2, and E8, while S162 (rs2156552) linked these environment variables to E5 (physical activity). S120 has been identified in the *B3GNT6* gene, and the reported trait is waist circumference[66]. S162 is in *LOC105372112* and is known to be associated with HDL-C and LDL-C in various populations[67,68].

MetS3 had the largest degree (41). Among the environmental nodes, E2 showed the largest degree (8). Among the SNPs, the degrees of S162 and S120 were high (6 and 4, respectively). MetS3, E2, E0, E1, E8 were the five nodes with the highest closeness centrality (MetS3:0.0069, E2:0.0063, E0:0.0062, E1:0.0062, E8:0.0061) while MetS3, MetS2, MetS4, E3, E2 were the five nodes with the highest betweenness centrality (MetS3:2404.22, MetS2:863.22, MetS4:738.17, E3:532.00, E2:484.00). No SNP seemed to be important in terms of centrality. MetS4 and MetS1 showed the highest correlations with E0 (age). It is remarkable that they had high negative correlation coefficients with E1 (education) (Table 3).

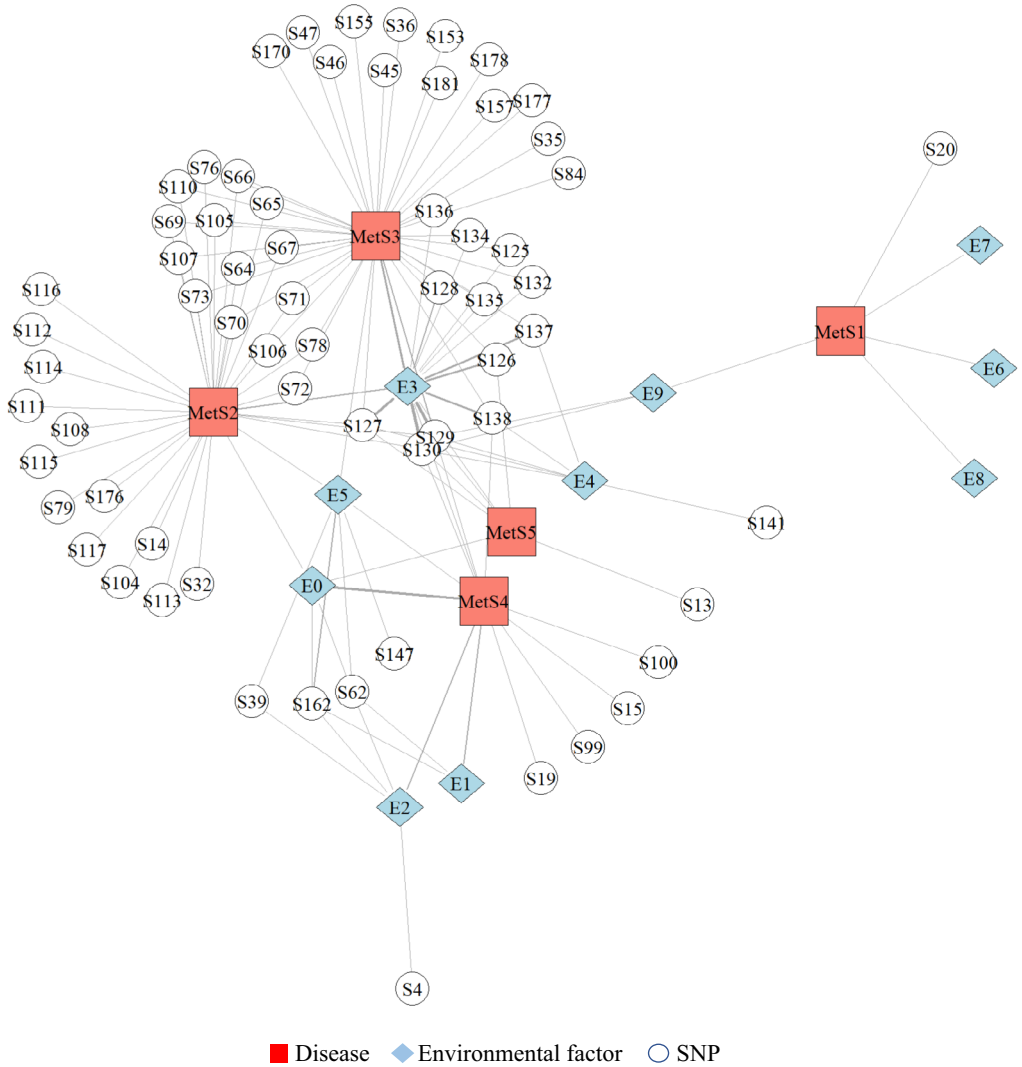

Fig. 1. Tripartite network of data from men. E0 (age), E1 (education), E2 (income), E3 (alcohol), E4 (smoking), E5 (physical activity), E6 (total energy), E7 (protein intake), E8 (fat intake), E9 (carbohydrate intake); MetS1 (abdominal obesity), MetS2 (triglycerides), MetS3 (HDL-C), MetS4 (hypertension), MetS5 (fasting glucose); S1–S190 (SNPs); Line thickness (degree of association).

Male		Female	
Edge	**Correlation coefficient***	**Edge**	**Correlation coefficient**
E3 ↔ S130	− 0.2780	MetS4 ↔ E0	0.4047
E3 ↔ S129	− 0.2758	MetS1 ↔ E0	0.3244
E0 ↔ MetS4	0.2405	MetS1 ↔ E1	− 0.2884
E3 ↔ S127	− 0.2305	MetS4 ↔ E1	− 0.2835
E3 ↔ S138	− 0.1629	MetS1 ↔ E2	− 0.2731

Table 3. Top five edges with high correlations. *Pearson correlation coefficient.

Constructing projected bipartite and unipartite networks

To elucidate the relationship between the nodes in two different sets, we projected the tripartite network into a two-mode bipartite network. Three two-mode projections were created for each sex. Among them, bipartite networks with the metabolic component set and SNP set for each sex are shown in Fig. 3. The projected bipartite network implies an indirect relationship between the nodes. For instance, in the network for men, MetS4 and S62 (rs4244457) are connected because they share E0, E1, E2, and E5 in the tripartite network. To reduce the loss of information, we applied simple weighting for the projection. The thickness of the edges was proportional

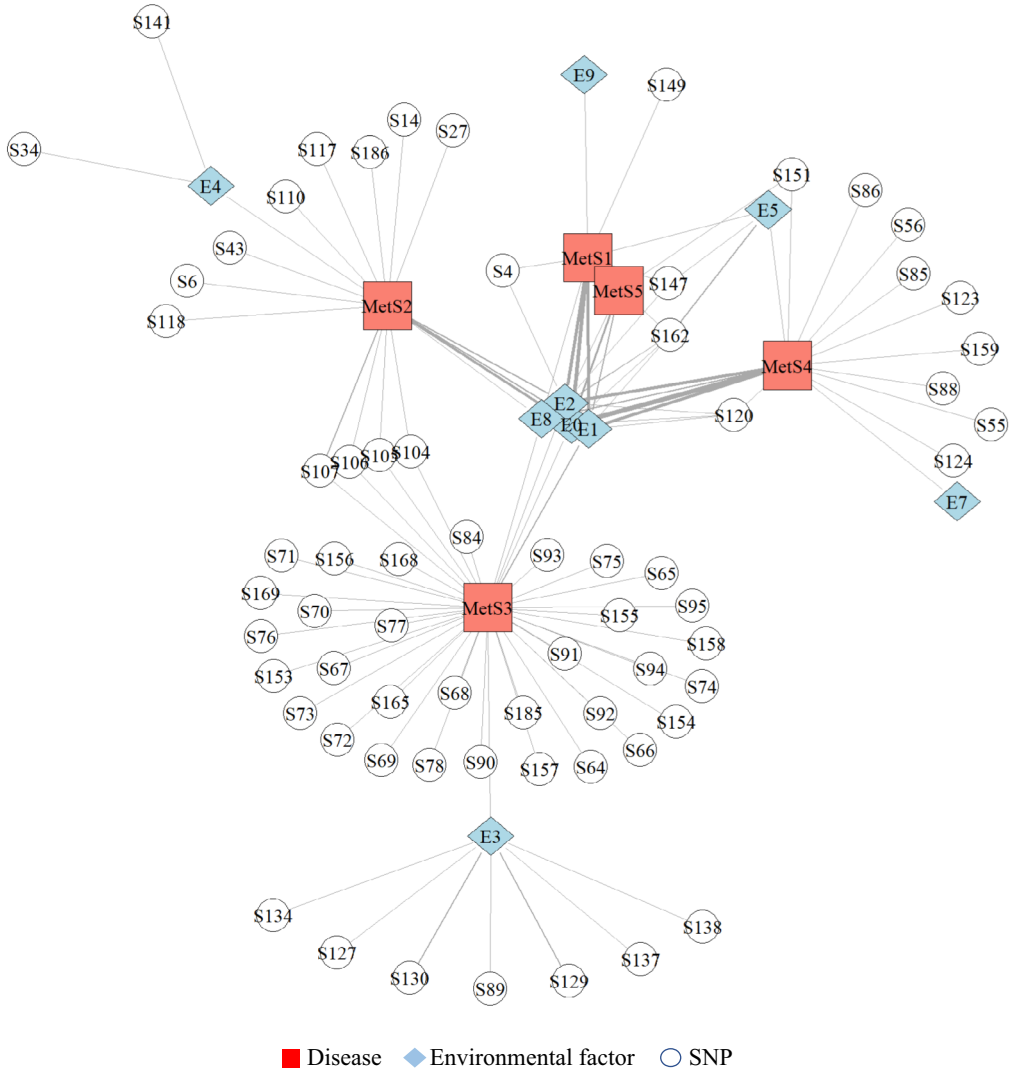

Fig. 2. Tripartite network of data from women. E0 (age), E1 (education), E2 (income), E3 (alcohol), E4 (smoking), E5 (physical activity), E6 (total energy), E7 (protein intake), E8 (fat intake), E9 (carbohydrate intake); MetS1 (abdominal obesity), MetS2 (triglycerides), MetS3 (HDL-C), MetS4 (hypertension), MetS5 (fasting glucose); S1–S190 (SNPs); Line thickness (degree of association).

to the number of environmental factors shared by the two nodes. We can interpret this as indicating a large indirect association between MetS4 and S62 through environmental factors, although there was no significant direct association, as shown in the tripartite network. S162 (rs2156552) and MetS4 also showed slightly stronger indirect relationships than other nodes. In the network for women, S120 (rs10899345) and S162 (rs2156552) showed strong connections with every component of metabolic syndrome. These SNPs showed high degrees in the tripartite network, but their direct correlations with metabolic syndrome components were low. However, the projected bipartite network indicated that they had strong indirect relationships with metabolic components, reflecting environmental factors. The bipartite networks of MetS components and environmental factors, as well as environmental factors and the SNP set, can be interpreted similarly (Supplementary Figs. 1, 2).

Figure 4 demonstrates the projected unipartite graph of data from men and women using concurrent projection. By the definition of a closed tripartite network, nodes in the same set were disconnected in the original tripartite network. However, through the projection, indirect relationships between the nodes in the same set could be discovered in the unipartite network. For men, MetS2 and MetS3 were strongly related through SNPs and environmental factors, and MetS1 was not related to other MetS components. In the data from women, the relationships involving MetS2 and MetS3 were weaker, but all the components were connected. For the environmental network, E0, E3, and E5 for men and E0, E1, E2, and E8 for women were strongly related through SNPs and MetS components. The unipartite network for SNPs showed several SNP clusters, each of which shared the same environments and MetS components.

To obtain unipartite networks using sequential projection, we re-compressed the projected bipartite network. For each dataset, two different unipartite networks were produced. The resulting structure of the unipartite network and the relationship between the nodes differed according to the order of the aggregating dataset. For

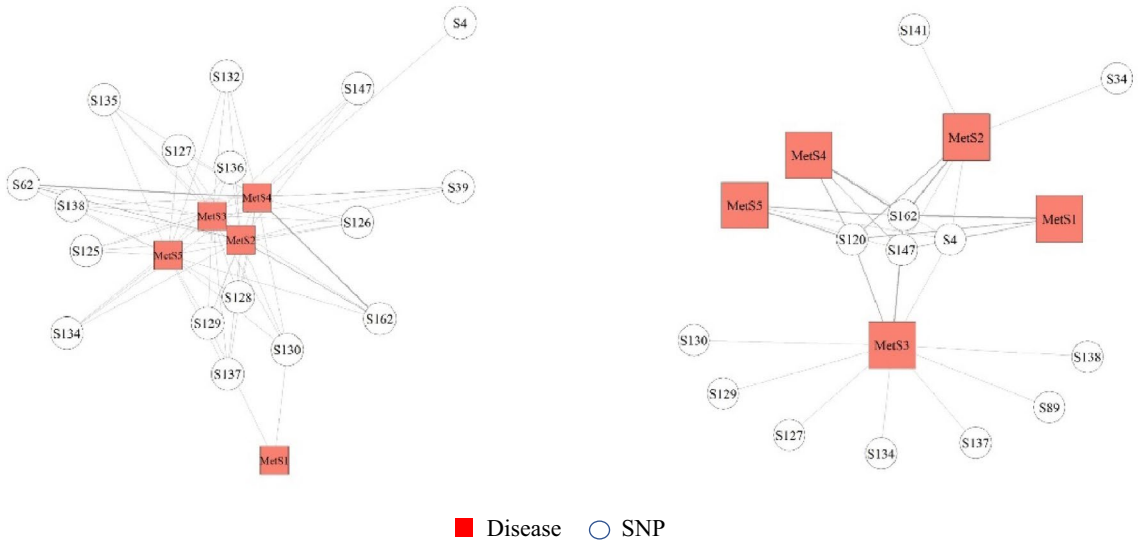

Fig. 3. Projected bipartite network of MetS components and the SNP set. MetS1 (abdominal obesity), MetS2 (triglycerides), MetS3 (HDL-C), MetS4 (hypertension), MetS5 (fasting glucose); S1–S190 (SNPs) ; Line thickness (degree of association). (**a**) Data from men (**b**) Data from women.

instance, although (b) and (d) in Fig. 5 both denote relationships between environmental factors, the graphs are completely different. This is because (b) was obtained by aggregating information over MetS component information from the environment-MetS components bipartite network (a), whereas (d) was obtained by aggregating information using the SNPs from the environment-SNP bipartite network (c). In the data from men, E3 and E4 were strongly connected by metabolic components, but nothing was connected to E4 via SNPs. The remaining unipartite networks obtained by sequential projection can be seen in Supplementary Figs. 3–7.

Discussion

To visualize the structure of numerous relationships among different variable sets—corresponding to genes, environments, and diseases—at once, a multipartite network was used. From a methodological perspective, the following novel points are proposed.

i. Utilizing multipartite networks to explore genetic and environmental influences on syndromic diseases: To understand genetic and environmental influences on syndromic diseases, we constructed independent variable sets and utilized multipartite networks to visualize the relationships between each set of variables at a glance.
ii. Identifying indirect relationships via projected bipartite networks: Using a projected bipartite network enabled the identification of indirect relationships between nodes that could not be discovered in a usual network. Connections in the lower mode network graphs derived through projection do not indicate direct associations, but rather indirect associations reflecting the factors in a hidden set. For example, if an SNP and a metabolic component are connected in a projected bipartite network, this does not imply a direct association between them, but rather that they share a hidden environmental factor.
iii. Proposing two different multi-stage projection methods: To elucidate the relationship between nodes in the same set, we suggested two different projection methods. Using the concurrent method allows us to represent associations between variables explained by variables from different groups. For instance, in a graph of diseases obtained from the projection into a one-mode unipartite network, diseases with significant indirect influences from both environment and genetics are strongly connected. In our data, the concurrent projection method was preferred due to its ease of interpretation. However, if the variable sets are nested, such as SNPs, genes, and pathways, sequential projection would be more meaningful. It is recommended to choose a projection method considering the relationship between sets.
iv. Applicability to small samples: Many studies did not conduct sex-stratified analyses due to sample size limitations or analytical complexities[69,70]. Moreover, the effects of individual variables can be weak in complex diseases. Multipartite networks serve as exploratory tools, capable of revealing not only strictly significant variables but also potential underlying associations. Therefore, they can offer advantages in small-scale studies.
v. Applicability to pleiotropy: We set each component of MetS as a node. However, if a node is defined as a disease, a pleiotropic effect can also be seen through a tripartite network graph.

From the perspective of MetS analysis results, the study's novel findings can be summarized as follows:

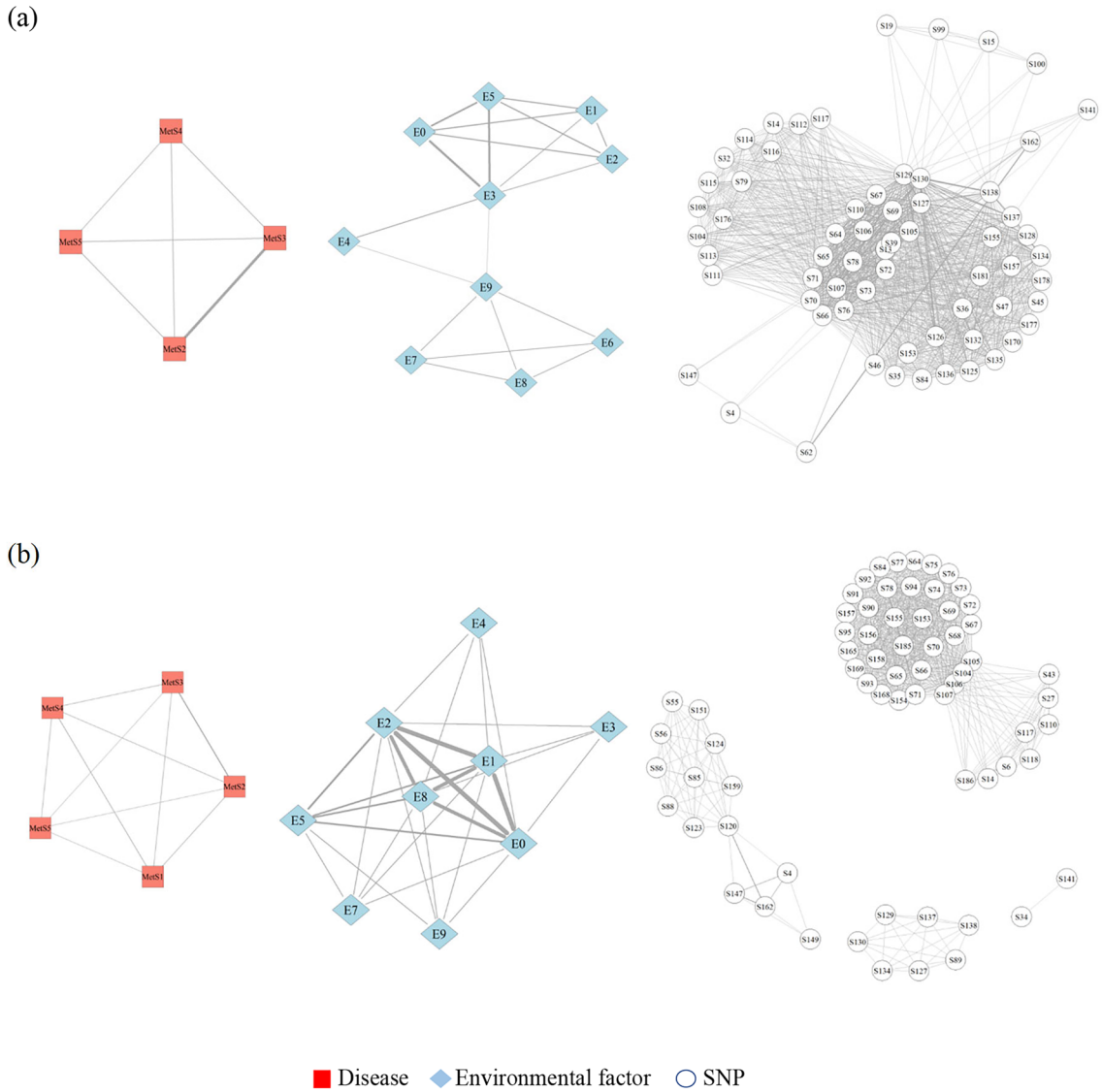

Fig. 4. Projected unipartite networks using concurrent projection. E0 (age), E1 (education), E2 (income), E3 (alcohol), E4 (smoking), E5 (physical activity), E6 (total energy), E7 (protein intake), E8 (fat intake), E9 (carbohydrate intake); MetS1 (abdominal obesity), MetS2 (triglycerides), MetS3 (HDL-C), MetS4 (hypertension), MetS5 (fasting glucose); S1–S190 (SNPs) ; Line thickness (degree of association). (**a**) Data from men. (**b**) Data from women.

(i) Sex-based variations in network patterns on metabolic syndrome: Utilizing Korean GWAS data, we identified distinct patterns between men and women. A notable contrast is the central and hub role of alcohol in the network for men, whereas its significance was lower within the female network. While the impact of alcohol consumption on health issues such as hypertension and dyslipidemia has been acknowledged[71–73], the use of multipartite networks helped confirm its influence on MetS components, particularly in men. Furthermore, within the male network structure, HDL-C, triglyceride, and hypertension from the MetS component set; rs12903590 and rs4821116 from the SNP set; and carbohydrate intake and alcohol from the environmental set served as central and bridging nodes. In contrast, key nodes in the female data comprised age, education, income, and fat intake from the environmental set, which were strongly linked with MetS components, displaying a distinct pattern compared to men. Previous studies have underscored sex/gender differences in the risk and genetic effects of MetS[58,59,74]. Additionally, the effects of socioeconomic variables and dietary habits on MetS have been reported[75–77]. Certain SNPs, such as rs12903590 and rs4821116, have been associated with HDL-C cholesterol levels in the Asian population[78,79]. However, network graphs offer a clear depiction of their associations with pertinent SNPs.

(ii) Environmental interactions on MetS and genes: The projected bipartite network enabled the identification of indirect relationships between MetS components influenced by environmental factors and SNPs can be identified. The analysis indicated that in men, rs4244457 is associated with hypertension through age, education, and physical activity, while rs2156552 appears to be prominently linked to hypertension

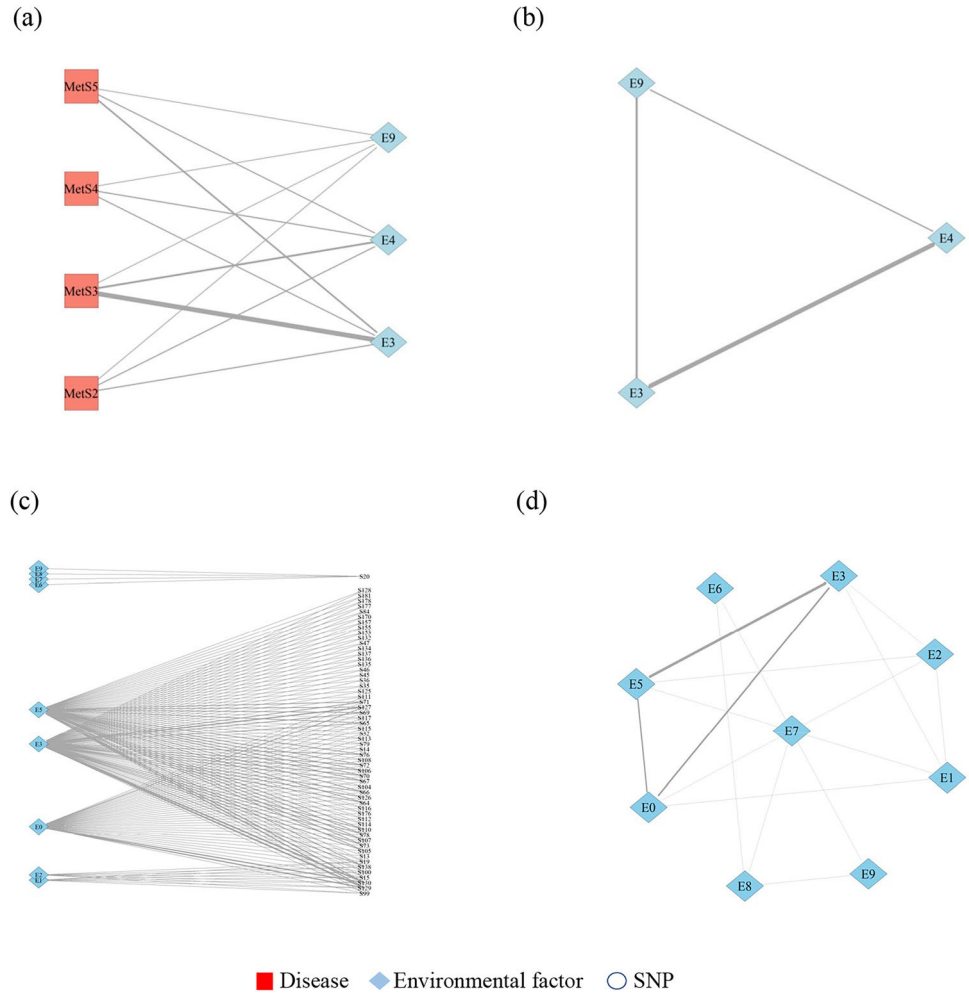

Fig. 5. Projected unipartite networks using sequential projection for the data from men. (**a**) Projected bipartite network of the metabolic component set and the environmental set. (**b**) Projected unipartite network of environmental factors using sequential projection to (**a**). (**c**) Projected bipartite network of the environmental set and the SNP set. (**d**) Projected unipartite network of the environmental set using sequential projection to (**c**). E0 (age), E1 (education), E2 (income), E3 (alcohol), E4 (smoking), E5 (physical activity), E6 (total energy), E7 (protein intake), E8 (fat intake), E9 (carbohydrate intake); MetS1 (abdominal obesity), MetS2 (triglycerides), MetS3 (HDL-C), MetS4 (hypertension), MetS5 (fasting glucose); S1–S190 (SNPs) ; Line thickness (degree of association).

through age, income level, and physical activity. In women, rs10899345 and rs2156552 are associated with all MetS components through age, education level, income, and fat intake. These findings could not be obtained through simple correlation analysis, underscoring the need for further analyses such as gene-environmental interaction analysis or mediation analysis. While there have been various prior studies on this topic[80–82], the specific SNPs with lifestyle interactions on MetS addressed are, to the best of the authors' knowledge, not covered in those studies.

Among the identified SNPs in this study, rs1290359, which showed a direct relationship with metabolic components, maps to the ALDH1A2 gene. ALDH1A2 is involved in converting retinol into retinoic acid (RA), a critical regulator of lung and cardiovascular development during human embryogenesis. Additionally, this gene is implicated in T-cell acute lymphoblastic leukemia and is considered a candidate tumor suppressor in prostate cancer[83,84]. ALDH1A2 may also promote a progressive phenotype in glioblastoma[85]. Furthermore, rs4821116 is located in the UBE2L3 gene, which has been associated with various autoimmune diseases, including rheumatoid arthritis, celiac disease, Crohn's disease, and systemic lupus erythematosus, through its role in ubiquitination of the NF-κB precursor[86–88]. The SNP identified via indirect relationships, rs2156552, maps to the ACAA2 gene. ACAA2 is a rate-limiting enzyme in mitochondria responsible for catalyzing the final step of the mitochondrial beta-oxidation pathway[89]. Dysfunction of this enzyme may contribute to several metabolic disorders and diseases. The ACAA2 expression has been proposed as a potential molecular marker for small-cell neuroendocrine cancers[89]. The ACAA2 locus also has been linked to blood lipid abnormalities, particularly in HDL and LDL

cholesterol levels[68]. Considering this information, future studies could explore potential associations with diseases related to these genes.

Although the data were obtained according to systematic and standardized epidemiological data quality control procedures, this study still has several limitations. First, bias is possible since the variables related to lifestyle and diet were obtained from self-reported survey forms. Second, we used SNP chip data, which could be impacted by bias according to the direct genotyping approach without imputation analysis.

A few noteworthy methodological points are as follows. First, in selecting the threshold of $p < 1E-05$, we aimed to balance between the rigorous control of false positives, as done with the Bonferroni correction, and the need to include a sufficient number of SNPs to catch meaningful signals for exploratory analysis. This threshold enables more SNPs in our graph while still maintaining a reasonable level of statistical significance. Various studies have used the same threshold in the analysis of GWAS data[81,90–94]. Second, linkage disequilibrium (LD) pruning was not performed in the variable selection stage. Unlike regression-based methods, LD pruning is not required in the variable selection stage, because representing SNPs in LD does not influence the results of network-based methods. Instead, we investigated the selected SNPs in a post hoc analysis. A list of the SNP pairs with high LD ($r^2 \geq 0.9$) is presented in Supplementary Table 2.

For the indirect relationships identified in this study, validation through mediation analysis or Mendelian randomization could be considered. These avenues could be pursued in future research endeavors.

Data availability

The Korea Association Resource (KARE) project data will be publicly distributed by the Distribution Desk of the Korea Biobank Network. Researchers who wish to receive epidemiological and genomic information data should apply through the 'Human Resources Distribution Desk (http://biobank.nih.go.kr).' After completing the application form and submitting the research plan and IRB approval (or waiver), it goes through deliberation by the Distribution Review Committee, which meets once a month. The researchers will directly receive the distributed resources after approval. For any inquiries, contact admin@koreabiobank.re.kr.

Received: 29 October 2023; Accepted: 26 August 2024
Published online: 31 August 2024

References

1. Alberti, K. G. *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* **120**(16), 1640–1645 (2009).
2. Wan, J. Y. *et al.* Genome-wide association analysis of metabolic syndrome quantitative traits in the GENNID multiethnic family study. *Diabetol. Metab. Syndr.* **13**(1), 59 (2021).
3. Ranasinghe, P. *et al.* Prevalence and trends of metabolic syndrome among adults in the Asia–Pacific Region: A systematic review. *BMC Public Health* **17**(1), 101 (2017).
4. Huh, J. H. *et al.* Metabolic syndrome fact sheet 2021: Executive report. *CardioMetabolic Syndrome J.* **1**(2), 125–134 (2021).
5. Mottillo, S. *et al.* The metabolic syndrome and cardiovascular risk: A systematic review and meta-analysis. *J. Am. Coll. Cardiol.* **56**(14), 1113–1132 (2010).
6. Kaur, J. A comprehensive review on metabolic syndrome. *Cardiol. Res. Pract.* **2014**, 943162 (2014).
7. Esposito, K. *et al.* Metabolic syndrome and risk of cancer: A systematic review and meta-analysis. *Diabetes Care* **35**(11), 2402–2411 (2012).
8. Chen, X. *et al.* Genetic and environmental influences on the correlations between traits of metabolic syndrome and CKD. *Clin. J. Am. Soc. Nephrol.* **14**(11), 1590–1596 (2019).
9. Zhu, Y. *et al.* Susceptibility loci for metabolic syndrome and metabolic components identified in Han Chinese: A multi-stage genome-wide association study. *J. Cell. Mol. Med.* **21**(6), 1106–1116 (2017).
10. Musani, S. K. *et al.* Heritability of the severity of the metabolic syndrome in whites and blacks in 3 large cohorts. *Circulat. Cardiovasc. Genet.* **10**(2), e001621 (2017).
11. Povel, C. M. *et al.* Genetic variants and the metabolic syndrome: A systematic review. *Obes. Rev.* **12**(11), 952–967 (2011).
12. Carty, C. L. *et al.* Analysis of metabolic syndrome components in >15,000 African Americans identifies pleiotropic variants: Results from the population architecture using genomics and epidemiology study. *Circ. Cardiovasc. Genet.* **7**(4), 505–513 (2014).
13. Tekola-Ayele, F. *et al.* Genome-wide association study identifies African-ancestry specific variants for metabolic syndrome. *Mol. Genet. Metab.* **116**(4), 305–313 (2015).
14. Oh, S.-W. *et al.* Genome-wide association study of metabolic syndrome in Korean populations. *PloS One* **15**(1), e0227357 (2020).
15. Rung, J. *et al.* Genetic variant near IRS1 is associated with type 2 diabetes, insulin resistance and hyperinsulinemia. *Nat. Genet.* **41**(10), 1110–1115 (2009).
16. Lanktree, M. B. & Hegele, R. A. Metabolic syndrome. In *Genomic and Precision Medicine* 283–299 (Elsevier, 2017).
17. Teslovich, T. M. *et al.* Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**(7307), 707–713 (2010).
18. Coltell, O. *et al.* Genome-wide association study (GWAS) on bilirubin concentrations in subjects with metabolic syndrome: Sex-specific gwas analysis and gene-diet interactions in a mediterranean population. *Nutrients* **11**(1), 90 (2019).
19. Lee, H.-S., Kim, Y. & Park, T. New common and rare variants influencing metabolic syndrome and its individual components in a Korean Population. *Sci. Rep.* **8**(1), 5701 (2018).
20. Kraja, A. T. *et al.* Pleiotropic genes for metabolic syndrome and inflammation. *Mol. Genet. Metab.* **112**(4), 317–338 (2014).
21. Zeng, Y. *et al.* GWA-based pleiotropic analysis identified potential SNPs and genes related to type 2 diabetes and obesity. *J. Hum. Genet.* **66**(3), 297–306 (2021).
22. Zhang, Y. *et al.* QTL-based association analyses reveal novel genes influencing pleiotropy of metabolic syndrome (MetS). *Obesity* **21**(10), 2099–2111 (2013).
23. Al-Qawasmeh, R. H. & Tayyem, R. F. Dietary and lifestyle risk factors and metabolic syndrome: Literature review. *Curr. Res. Nutr. Food Sci. J.* **6**(3), 594–608 (2018).
24. Takahara, M. & Shimomura, I. Metabolic syndrome and lifestyle modification. *Rev. Endocr. Metab. Disord.* **15**, 317–327 (2014).
25. de Lorgeril, M. Commentary on the clinical management of metabolic syndrome: Why a healthy lifestyle is important. *BMC Med.* **10**, 1–3 (2012).

26. Deng, Y.-Y. *et al.* Combined influence of eight lifestyle factors on metabolic syndrome incidence: A prospective cohort study from the MECH-HK Study. *Nutrients* **16**(4), 547 (2024).
27. Magueresse-Battistoni, L., Vidal, H. & Naville, D. Environmental pollutants and metabolic disorders: The multi-exposure scenario of life. *Front. Endocrinol.* **9**, 413568 (2018).
28. Ghosh, S. *et al.* Contribution of environmental, genetic and epigenetic factors to obesity-related metabolic syndrome. *Nucleus* **66**(2), 215–237 (2023).
29. Paik, J.K., *et al*. Dietary protein to carbohydrate ratio and incidence of metabolic syndrome in Korean adults based on a long-term prospective community-based cohort. *Nutrients*. **12**(11) (2020).
30. Bosy-Westphal, A. *et al.* Common familial influences on clustering of metabolic syndrome traits with central obesity and insulin resistance: The Kiel obesity prevention study. *Int. J. Obesity* **31**(5), 784–790 (2007).
31. Adamo, K.B. & F. Tesson. *Gene-environment interaction and the metabolic syndrome*. in *Novartis Foundation Symposium*. 2008. (John Wiley, 1999).
32. Park, Y. S. *et al.* Association between lifestyle factors and the risk of metabolic syndrome in the South Korea. *Sci. Rep.* **12**(1), 13356 (2022).
33. Ordovas, J. M. & Shen, J. Gene-environment interactions and susceptibility to metabolic syndrome and other chronic diseases. *J. Periodontol.* **79**(8S), 1508–1513 (2008).
34. Maistry, T. *et al.* Gene-environmental interaction and the metabolic syndrome in Asian Indians with insulin resistance. *Atherosclerosis* **275**, e183 (2018).
35. Darabos, C., Harmon, S. H. & Moore, J. H. Using the bipartite human phenotype network to reveal pleiotropy and epistasis beyond the gene. In *Biocomputing 2014* 188–199 (World Scientific, 2014).
36. Prone-Olazabal, D., Davies, I. & González-Galarza, F. F. Metabolic syndrome: An overview on its genetic associations and gene–diet interactions. *Metab. Syndrome Relat. Disord.* **21**(10), 545–560 (2023).
37. Benyamin, B. *et al.* Are there common genetic and environmental factors behind the endophenotypes associated with the metabolic syndrome?. *Diabetologia* **50**, 1880–1888 (2007).
38. Zhou, T. *et al.* Bipartite network projection and personal recommendation. *Phys. Rev. E* **76**(4), 046115 (2007).
39. Ferreri, L., M. Ivaldi, & M.D.L. Giacobini. Tripartite Networks: A first exploratory step towards the understanding of multipartite networks. in *NETSCI12 The International School and Conference on Network Science*. (2012).
40. Phillips, C. A. *et al.* On finding and enumerating maximal and maximum k-partite cliques in k-partite graphs. *Algorithms* **12**(1), 23 (2019).
41. Zhang, Y., E.D. Kolaczyk, & B.D. Spencer. Estimating network degree distributions under sampling: An inverse problem, with applications to monitoring social media networks. (2015).
42. Diestel, R. The basics. In *Graph Theory* 1–34 (Springer, 2017).
43. Koc, I., Yuksel, I. & Caetano-Anollés, G. Metabolite-centric reporter pathway and tripartite network analysis of *Arabidopsis* under cold stress. *Front. Bioeng. Biotechnol.* **6**, 121 (2018).
44. Cho, Y. S. *et al.* A large-scale genome-wide association study of Asian populations uncovers genetic factors influencing eight quantitative traits. *Nat. Genet.* **41**(5), 527–534 (2009).
45. Expert Panel on Detection, E. Executive summary of the third report of the National Cholesterol Education Program (NCEP) expert panel on detection, evaluation, and treatment of high blood cholesterol in adults (adult treatment panel III). *JAMA*. **285**(19), 2486–2497 (2001).
46. Grundy, S. M. *et al.* Diagnosis and management of the metabolic syndrome: An American Heart Association/National Heart, Lung, and Blood Institute scientific statement. *Circulation* **112**(17), 2735–2752 (2005).
47. Paik, J. K. *et al.* Dietary protein to carbohydrate ratio and incidence of metabolic syndrome in Korean adults based on a long-term prospective community-based cohort. *Nutrients* **12**(11), 3274 (2020).
48. Ahn, Y. *et al.* Validation and reproducibility of food frequency questionnaire for Korean genome epidemiologic study. *Eur. J. Clin. Nutr.* **61**(12), 1435–1441 (2007).
49. Society, K.N. *Computer aided nutritional analysis program for professionals*. The Korean Nutrition Society Seoul. (2011).
50. Kim, Y., Han, B.-G., K. Group. Cohort profile: The Korean genome and epidemiology study (KoGES) consortium. *Int. J. Epidemiol.* **46**(2), e20 (2017).
51. Zhang, Y., Kolaczyk, E. D. & Spencer, B. D. Estimating network degree distributions under sampling: An inverse problem, with applications to monitoring social media networks. *Ann. Appl. Stat.* **9**(1), 166–199 (2015).
52. Abdallah, M. & Hung, C.-N. Neighbor connectivity of the alternating group graph. *J. Interconnect. Netw.* **21**(03), 2150014 (2021).
53. Diestel, R. *Graph Theory 3rd ed.* Graduate texts in mathematics. **173**, 33 (2005).
54. Zhang, J. & Y. Luo. *Degree centrality, betweenness centrality, and closeness centrality in social network*. in *2017 2nd International Conference on Modelling, Simulation and Applied Mathematics (MSAM2017)*. (Atlantis Press, 2017).
55. Newman, M.E. *Networks—An Introduction*, 124–125. (Oxford University Press, 2010).
56. Neal, Z. The backbone of bipartite projections: Inferring relationships from co-authorship, co-sponsorship, co-attendance and other co-behaviors. *Social Netw.* **39**, 84–97 (2014).
57. Cann, T. J. B., Weaver, I. S. & Williams, H. T. P. Is it correct to project and detect? How weighting unipartite projections influences community detection. *Netw. Sci.* **8**(S1), S145–S163 (2020).
58. Sugiyama, M. G. & Agellon, L. B. Sex differences in lipid metabolism and metabolic disease risk. *Biochem. Cell Biol.* **90**(2), 124–141 (2012).
59. Mauvais-Jarvis, F. Sex differences in metabolic homeostasis, diabetes, and obesity. *Biol. Sex Differences* **6**(1), 14 (2015).
60. Kanai, M. *et al.* Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. *Nat. Genet.* **50**(3), 390–400 (2018).
61. Noordam, R. *et al.* Multi-ancestry sleep-by-SNP interaction analysis in 126,926 individuals reveals lipid loci stratified by sleep duration. *Nat. Commun.* **10**(1), 5121 (2019).
62. Li, C. *et al.* Variants identified by hepatocellular carcinoma and chronic hepatitis B virus infection susceptibility GWAS associated with survival in HBV-related hepatocellular carcinoma. *PLoS One* **9**(7), e101586 (2014).
63. Spracklen, C. N. *et al.* Association analyses of East Asian individuals and trans-ancestry analyses with European individuals reveal new loci associated with cholesterol and triglyceride levels. *Hum. Mol. Genet.* **26**(9), 1770–1784 (2017).
64. Lee, J. S., Cheong, H. S. & Shin, H. D. Prediction of cholesterol ratios within a Korean population. *R. Soc. Open Sci.* **5**(1), 171204 (2018).
65. Discovery and refinement of loci associated with lipid levels. *Nat. Genet*. **45**(11), 1274–1283 (2013).
66. Li, D. *et al.* Progressive effects of single-nucleotide polymorphisms on 16 phenotypic traits based on longitudinal data. *Genes Genom.* **42**(4), 393–403 (2020).
67. Edmondson, A. C. *et al.* Dense genotyping of candidate gene loci identifies variants associated with high-density lipoprotein cholesterol. *Circ. Cardiovasc. Genet.* **4**(2), 145–155 (2011).
68. Kathiresan, S. *et al.* Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. *Nat. Genet.* **40**(2), 189–197 (2008).
69. Shin, J.-A. *et al.* Metabolic syndrome as a predictor of type 2 diabetes, and its clinical interpretations and usefulness. *J. Diabetes Investig.* **4**(4), 334–343 (2013).

70. Shang, X. *et al.* Dietary protein from different food sources, incident metabolic syndrome and changes in its components: An 11-year longitudinal study in healthy community-dwelling adults. *Clin. Nutr.* **36**(6), 1540–1548 (2017).
71. Freiberg, M. S. *et al.* Alcohol consumption and the prevalence of the metabolic syndrome in the US: A cross-sectional analysis of data from the Third National Health and Nutrition Examination Survey. *Diabetes care* **27**(12), 2954–2959 (2004).
72. Stranges, S. *et al.* Relationship of alcohol drinking pattern to risk of hypertension: A population-based study. *Hypertension* **44**(6), 813–819 (2004).
73. Magis, D., Jandrain, B. & Scheen, A. Alcohol, insulin sensitivity and diabetes. *Revue Medicale de Liege* **58**(7–8), 501–507 (2003).
74. Yi, Y. & An, J. Sex differences in risk factors for metabolic syndrome in the Korean Population. *Int. J. Environ. Res. Public Health.* **17**(24), 9513 (2020).
75. Julibert, A. *et al.* Dietary fat intake and metabolic syndrome in older adults. *Nutrients* **11**(8), 1901 (2019).
76. Silventoinen, K. *et al.* Educational inequalities in the metabolic syndrome and coronary heart disease among middle-aged men and women. *Int. J. Epidemiol.* **34**(2), 327–334 (2005).
77. Dallongeville, J. *et al.* Household income is associated with the risk of metabolic syndrome in a sex-specific manner. *Diabetes Care* **28**(2), 409–415 (2005).
78. Kanai, M. *et al.* Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. *Nat. Genet.* **50**, 390–400 (2018).
79. Spracklen, C. N. *et al.* Association analyses of East Asian individuals and trans-ancestry analyses with European individuals reveal new loci associated with cholesterol and triglyceride levels. *Hum. Mol. Genet.* **27**(6), 1122–1122 (2018).
80. Adamo, K.B. & F. Tesson. Gene–environment interaction and the metabolic syndrome. in *Genetic Effects on Environmental Vulnerability to Disease*, 103–121 (2008).
81. Jeon, S. *et al.* Structural equation modeling for hypertension and type 2 diabetes based on multiple SNPs and multiple phenotypes. *PLoS One* **14**(9), e0217189 (2019).
82. Lutz, S.M. & J.E. Hokanson. Mediation analysis in genome-wide association studies: current perspectives. *Open Access Bioinform.* 1–5 (2015).
83. Kim, H. *et al.* The retinoic acid synthesis gene ALDH1a2 is a candidate tumor suppressor in prostate cancer. *Cancer Res.* **65**(18), 8118–8124 (2005).
84. Liu, Y. *et al.* The genomic landscape of pediatric and young adult T-lineage acute lymphoblastic leukemia. *Nat. Genet.* **49**(8), 1211–1218 (2017).
85. Sanders, S. *et al.* The presence and potential role of ALDH1A2 in the glioblastoma microenvironment. *Cells* **10**(9), 2485 (2021).
86. Hu, Z. *et al.* New loci associated with chronic hepatitis B virus infection in Han Chinese. *Nat. Genet.* **45**(12), 1499–1503 (2013).
87. Zuo, X.-B. *et al.* Variants in TNFSF4, TNFAIP3, TNIP1, BLK, SLC15A4 and UBE2L3 interact to confer risk of systemic lupus erythematosus in Chinese population. *Rheumatol. Int.* **34**, 459–464 (2014).
88. Wang, S. *et al.* A functional haplotype of UBE2L3 confers risk for systemic lupus erythematosus. *Genes Immunity* **13**(5), 380–387 (2012).
89. Shen, M. *et al.* ACAA2 is a novel molecular indicator for cancers with neuroendocrine phenotype. *Br. J. Cancer* **129**(11), 1818–1828 (2023).
90. Valette, K. *et al.* Prioritization of candidate causal genes for asthma in susceptibility loci derived from UK Biobank. *Commun. Biol.* **4**(1), 700 (2021).
91. Lafarge, T. *et al.* Genome-wide association analysis for heat tolerance at flowering detected a large set of genes involved in adaptation to thermal and other stresses. *PLOS ONE* **12**(2), e0171254 (2017).
92. Chen, L. *et al.* Genome-wide assessment of genetic risk for systemic lupus erythematosus and disease severity. *Hum. Mol. Genet.* **29**(10), 1745–1756 (2020).
93. Chen, Z.-Q. *et al.* Leveraging breeding programs and genomic data in Norway spruce (*Picea abies* L. Karst) for GWAS analysis. *Genome Biol.* **22**, 1–30 (2021).
94. Yan, Q. *et al.* Genome-wide association study of brain amyloid deposition as measured by Pittsburgh Compound-B (PiB)-PET imaging. *Mol. Psychiatry* **26**(1), 309–321 (2021).

Acknowledgements

This study was conducted with bioresources from the National Biobank of Korea, the Korea Disease Control and Prevention Agency, Republic of Korea (NBK-2021-059).

Author contributions

Conceptualization: MP. Data curation: JS. Formal analysis: JS, NS. Funding acquisition: MP. Methodology: MP. Writing—original draft: MP, NS. Writing—review & editing: TP, MP.

Funding

This research was supported by a National Research Foundation of Korea (NRF) Grant funded by the Korean government (MSIT) (NRF-2021R1A2C1007788).

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-71217-5.

Correspondence and requests for materials should be addressed to M.P.