Association of Age-Related Cataract Risk with High Polygenetic Risk Scores Involved in Galactose-Related Metabolism and Dietary Interactions

Donghyun Jee^a Suna Kang^b Sunmin Park^b

^aDivision of Vitreous and Retina, Department of Ophthalmology, St. Vincent's Hospital, College of Medicine, The Catholic University of Korea, Suwon, South Korea; ^bFood and Nutrition, Obesity/Diabetes Research Center, Hoseo University, Asan, South Korea

Keywords

Age-related cataract · Polygenetic risk score · Galactose · Lactose · Milk · Dairy product

Abstract

Introduction: Cataracts are associated with the accumulation of galactose and galactitol in the lens. We determined the polygenetic risk scores for the best model (PRSBM) associated with age-related cataract (ARC) risk and their interaction with diets and lifestyles in 40,262 Korean adults aged over 50 years belonging to a hospital-based city cohort.

Methods: The genetic variants for ARC risk were selected in lactose and galactose metabolism-related genes with multivariate logistic regression using PLINK 1.9 version. PRSBM from the selected genetic variants were estimated by generalized multifactor dimensionality reduction (GMDR) after adjusting for covariates. The interactions between the PRSBM and each lifestyle factor were determined to modulate ARC risk. **Results:** The genetic variants for ARC risk related to lactose and galactose metabolism were *SLC2A1*_rs3729548, *ST3GAL3*_rs3791047, *LCT*_rs2304371, *GALNT5*_rs6728956, *ST6GAL1*_rs2268536, *GALNT17*_rs17058752, *CSGALNACT1*_rs1994788, *GALNTL4*_rs10831608, *B4GALT6*_rs1667288, and *A4GALT*_rs9623659. In GMDR, the best model included all ten genetic variants. The highest odds ratio for a single SNP

in the PRSBM was 1.26. However, subjects with a high-PRSBM had a higher ARC risk by 2.1-fold than a low-PRSBM after adjusting for covariates. Carbohydrate, dairy products, kimchi, and alcohol intake interacted with PRSBM for ARC risk, where participants with high-PRSBM had a much higher ARC risk than those with low-PRSBM when consuming diets with high carbohydrate and low dairy product and kimchi intake. However, only with low alcohol intake, the participants with high-PRSBM had a higher ARC risk than those with low-PRSBM. **Conclusion:** Adults aged >50 years having high-PRSBM may modulate dietary habits to reduce ARC risk.

© 2021 The Author(s)

Published by S. Karger AG, Basel

Introduction

A cataract is defined as eye lens opacity causing light not to focus on the retina, which is the most common cause for preventable low vision. Cataracts may be congenital or developed at an advanced age, considered an age-related cataract (ARC). ARC commonly occurs in people over 50 years old [1]. The potential risk factors of cataracts are classified as genetic and environmental factors. Environmental factors include age, smoking, diabetes, eye injury, chronic use of steroids, heavy alcohol drinking, and ultraviolet exposure. ARC is induced by in-

creased osmotic pressure, oxidative stress, and inflammation [2].

When monosaccharides, including glucose and galactose, are not properly metabolized in the lens, they can cause a cataract. Galactose can be produced in the body, and it is metabolized through the DeLey-Doudoroff pathway [3]. If the genes in the pathway are mutated, the expressions and functions of the mutated genes are changed, and galactose metabolism is impaired. Glucose and galactose are converted into sorbitol and galactitol by aldose reductase, and the concentrations of sorbitol and galactitol increase in the lens [3]. As a consequence of the process, oxidative stress is elevated, and sugar alcohol induces osmotic damage in the lens [4], causing cellular damage, leading to the development of cataracts. Additionally, galactokinase deficiency is an autosomal recessive genetic disease that causes hypergalactosemia, which induces cataracts [5]. Galactokinase (*GALK*) missense mutations (Ala198Val and Pro28Thr) and UDP-galactose 4-epimerase-A (Val94Met) can lead to hypergalactosemia [6, 7]. Persons with the *GALK*_Aal198Val variant have a higher risk of congenital cataract formation [7]. Mutation of some genes, including *FOXE3*, *HSF4*, *MAF*, *PITX3*, and *CRYAB*, are reported to be associated with cataract risk [8–10].

Dairy products, mainly milk, have a reputation for causing cataracts. However, there is no evidence that milk and dairy product consumption itself increases cataract risk. This belief may come from the association of *GALK* mutations and cataract risk [10]. Galactose mostly comes from the lactose in milk and dairy products by lactase, an inducible gene, in the intestines [11, 12]. However, galactose is also endogenously produced in the body, and it needs to be metabolized [3]. *GALK* is necessary to metabolize galactose, and persons with attenuated galactose utilization may be at increased risk of developing ARC. Continuous dairy product consumption by infants produces lactose- and galactose-metabolizing genes to metabolize lactose and galactose [13]. However, discontinuation of dairy products prevents the symptoms of lactose intolerance and impairs galactose metabolism in the body. Dairy product intake may protect against the impairment of galactose metabolism to protect against ARC. Since lactic acid bacteria also participate in lactose digestion [14, 15], bacteria from fermented foods such as kimchi may utilize lactose and galactose in the intestines. Therefore, lactose and galactose metabolism in the intestines may interact with modulating ARC risk [12].

Galactose is converted into glucose by several enzymes, including *GALK*, galactose-1-phosphate uridylyltransferase

(*GALT*), aldose transferase, galactose dehydrogenase (*GALDH*), and UDP-glucose/galactose pyrophosphorylase (*UGP*), and other genes in the body [16, 17]. Lactose intolerance due to an inborn error of metabolism with nonfunctional *GALT* is known to result in cataracts at a young age. However, genetic variants of *GALT* and other enzymes related to lactose and galactose metabolism may be associated with an increased risk of ARC due to disruption of normal galactose metabolism. The genetic variants may interact with each other, and they may have environmental interactions, including food intake (vegetables, kimchi, fruits, and dairy products), physical activity, smoking, alcohol drinking, and steroid usage, which may increase the risk of ARC. Therefore, we hypothesized that (1) the genetic variants related to lactose and galactose metabolism are associated with ARC risk and (2) the polygenic risk score generated from the ARC-related genetic variants with gene-gene interactions interacts with dairy products and other foods and nutrients. These hypotheses were examined in 40,262 Korean adults >50 years old in a hospital-based city cohort.

Methods

Participants

Middle-aged and elderly Korean adults volunteered to participate in this hospital-based city cohort study. During 2004–2013, 40,262 adults aged >50 years participated in the city hospital cohorts under the Korean Genome and Epidemiology Study (KoGES) established by the Korean Center for Disease and Control. The present study, including KoGES procedures, followed the Helsinki Declaration. The Institutional Review Board on KoGES approved this study (KBP-2015-055). All subjects that participated in KoGES provided written informed consent. The Institutional Review Board of Hoseo University approved the present study (1041231-150811-HR-034-01).

Definition of ARC and Metabolic Syndrome

The presence of ARC was defined by using a questionnaire whether participants had a diagnosis of cataracts by a physician after 50 years old, and those who answered yes were considered to have ARC. Since ARC primarily develops after age 50, only participants over 50 were included in this study. Among the participants, 1,972 participants were previously diagnosed with cataracts after age 50 and 38,290 without a cataract diagnosis.

Metabolic syndrome comprised 3 or more of the following categories: (1) waist circumference ≥ 90 cm for men and ≥ 85 cm for women (abdominal obesity); (2) average systolic blood pressure ≥ 130 mm Hg or diastolic blood pressure ≥ 85 mm Hg (hypertension); (3) < 40 mg/dL for men and < 50 mg/dL for women (hypo-HDL-cholesterolemia); (4) ≥ 150 mg/dL (hypertriglyceremia); and (5) ≥ 110 mg/dL (hyperglycemia) [18]. Participants who were taking medication for dyslipidemia, hyperglycemia, and hypertension were included in the metabolic syndrome group.

Clinical Parameters Measurements

Trained researchers interviewed the participants to collect health interview including age, education and income levels, the status of smoking and alcohol intake, and physical activity [19, 20]. We divided education levels into three groups: less than graduation of high school, graduation of high school, and graduation of college, or more. We categorized household income (USD/month) into four groups: very low (<USD 1,000), low (USD 1,000–2,000), intermediate (USD 2,000–4,000), and high (>USD 4,000) [21]. Current and no smoking status was classified with smoking 100 cigarettes in the lifetime while past smoking status was defined as quitting smoking for at least the last 6 months [21]. Alcohol consumption status was categorized into four groups according to average daily alcohol consumption: non-drinker (0 g), moderate drinker (0–20 g), and heavy drinker (>20 g). The consumption of dairy products including cheese, yogurt, and milk was classified into <0.75 serving/day (low dairy products) and ≥0.75 serving/day (high dairy products).

A standardized procedure was adopted to measure waist circumference, height, and body weight [22]. We have calculated BMI as the weight in kilograms per height in meter squared (kg/m²). Blood sampling was performed after an overnight fast. Biochemical measurements were conducted using plasma and serum samples [22]. An Automatic Analyzer (Hitachi 7600; Hitachi, Tokyo, Japan) was used for the determination of concentrations of serum glucose levels and hemoglobin A1c in serum. We measured blood pressure on the right arm at the same height as the subject's heart in a sitting position.

Genotyping

The Center for Genome Science in the Korea National Institute of Health provided the genotype data. Genomic DNA was extracted from whole blood, and genotypes determined on Korean Chip (Affymetrix, Santa Clara, CA, USA) made available to the scientists. Korean Chip was implemented for studying Korean genetic variants and included the critical SNPs related to metabolic diseases (diabetes, hypertension, dyslipidemia, osteoporosis, kidney disease, and autoimmune diseases) and cancers [23]. The genotyping accuracy was examined by Bayesian Robust Linear Modeling using the Mahalanobis Distance (BRLMM) Genotyping Algorithm. DNA samples included genotyping accuracies (≥98%), missing genotype call rates (<4%), low heterozygosity (<30%), or no gender biases. We included genetic variants that satisfy the inclusion criteria of HWE ($p > 0.05$) [24].

Association of ARC with the Genetic Variants Related to Lactose and Galactose Metabolism

Subjects were categorized into having ARC (case; $n = 1,971$) and noncataract (control; $n = 38,290$). The genes related to lactose and galactose metabolism were selected from the Kyoto Encyclopedia of Genes and Genomes database (<https://www.kegg.jp/kegg/kegg2.html>) using keywords of lactose and galactose metabolism including the DeLey-Doudoroff pathway. The locations of the related genes were used to explore the genetic variants associated with ARC. The association of ARC with the genetic variants related to lactose- and galactose-related metabolism was conducted with covariates including age, sex, residence area, education, income, BMI, and serum glucose concentrations using PLINK 1.9 version, a free, open-source whole genome association analysis toolset (<https://zzz.bwh.harvard.edu/plink/>). Covariates were used to eliminate the confounding effects to find the genetic variants

associated with ARC. The genetic variants associated with ARC were selected at $p < 0.05$. Linkage disequilibrium (LD) analyses were performed in the selected genetic variants in the same chromosome using Haploview 4.2 in PLINK, and the 10 potential genetic variants for the best model in the same chromosome did not show a strong correlation ($r^2 < 0.4$) in LD.

Identification of the Best Model for Genetic Variant-Genetic Variant Interaction by Generalized Multifactor Dimensionality Reduction

The genetic variants selected were used to find the best genetic variant-genetic variant interaction model using the GMDR method [19]. The best genetic variant-genetic variant interaction model was searched in a sign rank test of trained balanced accuracy (TRBA) and testing balanced accuracy (TEBA) with or without adjusting for covariates using a GMDR program and a p value threshold of 0.05 [20]. The covariates included age, gender, residence area, education, income level, and BMI. Ten-fold cross-validation was also checked by cross-validation consistency (CVC) since the sample size was greater than 1,000 [20]. Ten out of 10 in CVC met the perfect cross-validation. Therefore, the best model was selected from the criteria of the sign test result between TRBA and TEBA and CVC value.

The risk allele of each SNP in the selected best model was counted as 1. If persons had AA, AG, GG of one SNP, and the A allele was the risk allele, the genetic score for the SNP was 2, 1, and 0. Polygenic risk scores of the best model (PRSBM) were computed by the summation of risk allele scores in each SNP in the PRSBM. According to the PRSBM, the participants were divided into low-PRSBM, medium-PRSBM, and high-PRSBM, corresponding to the low-risk, medium-risk, and high-risk groups of genetic impacts. A high score of the PRSBM indicated a higher number of risk alleles in the SNPs in the best gene-gene interaction model.

Assessment of Food and Nutrient Intake by Using the Semiquantitative Food Frequency Questionnaire

Dietary intake was estimated using an SQFFQ developed and validated in the KoGES [25]. This questionnaire requested information regarding the participants' consumption of food items. The SQFFQ included 106 food items, and they were used to calculate the intake of 23 nutrients from the Korean food composition table [25]. Average daily nutrient intake was calculated from the food intake measured by SQFFQ using the Computer-Aided Nutritional Analysis Program (CAN Pro) 3.0, a nutrient database developed by the Korean Nutrition Society. Intakes of carbohydrate, protein, and fat were calculated as energy percent, and they were categorized into 2 groups by 65, 13, and 15 energy percent, respectively, and they corresponded to about 25, 75, and 70 percentiles of Korean adults represented in KoGES and KNHANES [26].

Statistical Analyses

Statistical analysis was performed using PLINK version 2.0 (<http://pngu.mgh.harvard.edu/~purcell/plink>) for GWAS and SAS (version 9.3; SAS Institute, Cary, NC, USA) for other statistical analyses. Proc univariate was used to check the normality of the variables, and they all had a normal distribution. Statistical differences of continuous variables between ARC and control groups were compared by analysis of covariance after adjusting covariates. Those of the categorical variables were determined by the χ^2 test. The association of general characteristics with ARC incidence was

assessed through logistic regression analysis after adjusting covariates including residence area, sex, age, levels of income and education, BMI, habits of smoking and alcohol, the amount of daily energy intake, dairy product consumption, physical activity, and medication for dermatitis and arthritis [27].

For the descriptive statistics, the numbers and percentages of categorical variables such as gender, education, income, exercise, and smoking status and alcohol and coffee intake were calculated according to the three groups of PRSBM (low-, middle-, and high-PRSBM groups). Frequency distributions of the classification variables were analyzed using the χ^2 test. Means and standard deviations were calculated for the continuous variables according to the PRSBM categories. The significant differences were determined by one-way analysis of variance with or without the adjustments for covariates. Multiple comparisons among the groups were determined with the Tukey test.

The association of PRSBM in the best model with ARC risk was evaluated using logistic regression analysis after adjusting for covariates to exclude the effect of confounding factors. ARC risk was assessed as the odds ratios (ORs) and 95% confidence intervals (CI) using low-PRSBM as a reference. The variables were divided into two groups with the cutoff points to analyze their associations with PRSBM for ARC risk in logistic regression analysis. The variables' cutoffs were mainly 70th percentiles or known values such as dietary recommended intake of energy intake. The logistic regression analysis was conducted considering two adjusted models. The first model included adjustment for residence area, sex, age, BMI, education, and income. The second model included adjustments for residence area, sex, age, BMI, education and income levels, smoking and drinking status, daily energy intake, dairy product consumption, physical activity, and medication for arthritis and dermatitis. The covariates were selected because they are known to influence ARC risk. Arthritis and dermatitis medicine intakes were used to adjust taking corticosteroids for treating dermatitis and arthritis since corticosteroid intake is involved in cataract etiology. According to the dietary intake, unadjusted and adjusted ORs and 95% CI for ARC risk compared to the low-PRSBM were calculated.

We divided the participants into two groups with high or low intake using the classification criteria mentioned above to determine the interaction between the PRSBM and each lifestyle factor, including intake of fat, carbohydrate, protein, kimchi, milk, dairy products, and alcohol. The interaction between PRSBM and lifestyles was examined using two-way analysis of variance with PRSBM and lifestyles as the main effects with their interaction terms, and the model also included covariates. p value <0.05 indicated statistical significance.

Results

Socioeconomic and Lifestyle Characteristics of the Participants

Of 40,262 participants in the hospital-based cohort, 1,972 (5.2%) had ARC. The adjusted mean of age was higher in the cataract group than the noncataract group, and men had a lower incidence of ARC than women. The

covariates to calculate adjusted means were as follows: residence area, surveyed year, BMI, gender, age, the status of smoking and alcohol, education levels, income, job, energy intake, and medication for arthritis and dermatitis. BMI was not significantly different between the two groups (Table 1). However, adjusted ORs indicated that age and BMI were positively associated with ARC incidence by 4.2- and 1.4-fold, respectively. Adjusted means of serum glucose levels and hemoglobin A1c levels were higher in the cataract group than the noncataract group. Adjusted ORs for serum glucose and HbA1c were also elevated by 1.4- and 1.6-fold in the cataract group, respectively (Table 1). ARC incidence was lower in the participants with higher education ($p < 0.001$), higher income ($p < 0.001$), moderate alcohol drinking ($p < 0.001$), and high coffee intake ($p < 0.01$) than those with lower education, lower income, heavy exercise, no alcohol, and coffee intake without any adjustment. However, adjusted ORs for cataract incidence did not show any association with education, income, exercise, and alcohol and coffee drinking after adjusting for residence area, survey year, BMI, gender, age, levels of education and income, the status of smoking and alcohol, job, and energy intake. The cataract incidence was lower with higher daily energy and kimchi intake by 0.90- and 0.82-fold, respectively, although the adjusted means of daily energy and kimchi intake were not significantly different between the cataract and noncataract groups (Table 1). High carbohydrate (≥ 65 energy %), fat (≥ 15 energy %), and protein (≥ 13 energy %) intakes were associated with cataract incidence by 1.37-, 0.79-, and 0.84-fold, respectively (Table 1). The cutoffs of carbohydrates, fat, and protein intake were based on the 75th percentiles of Korean intake [28].

Characteristics of 10 Selected Genetic Variants Related to Lactose and Galactose Metabolism

According to the candidate gene-based approach, ten SNPs were selected to be significantly associated with ARC risk among the genes related to lactose and galactose metabolism (online suppl. Fig. 1S; for all online suppl. material, see www.karger.com/doi/10.1159/000521548). The selected SNPs and their characteristics are shown in Table 2. The minor alleles of *SLC2A1*, *ST3GAL3*, *LCT*, and *GALNT5* had a negative association with ARC (ORs = 0.8646–0.9033) after adjusting for cataract risk factors, whereas those of *ST6GAL1*, *GALNT17*, *CSGALNACT1*, *GALNTL4*, *B4GALT6*, and *A4GALT* had a positive association (ORs = 1.111–1.256; Table 2) (online suppl. Table 1S). All SNPs met the criteria of MAF ($\geq 0.05\%$) and Hardy-Weinberg equilibrium (HWE; $p > 0.05$) criteria (Ta-

Table 1. Socioeconomic and lifestyle characteristics of the participants according to ARC incidence

	Non-ARC (n = 38,290)	ARC (n = 1,972)	Adjusted OR (95% CI)
Age, ⁴ years	57.8±5.5	61.8±5.4***, ¹	4.161 (3.722–4.652)***, ²
Gender (male)	14,001 (36.6)	805 (40.8)***, ³	1.081 (0.971–1.204)
BMI, ⁵ kg/m ²	19.8±2.8	20.1±2.9	1.112 (1.006–1.230)*
Fasting serum glucose, ⁶ mg/dL	96.5±20.2	100.4±26.5***	1.385 (1.250–1.535)***
HbA1c ⁷	5.79±0.7	5.91±0.9***	1.636 (1.388–1.928)***
Blood pressure ⁸	1,169 (4.17)	803 (6.56)***	1.454 (1.314–1.609)***
Education ⁹			
<High school college	6,835 (23.1)	480 (32.3)***	
High school	7,640 (25.9)	395 (26.6)	0.920 (0.815–1.039)
College more	15,076 (51.0)	612 (41.2)	
Income (per month)			
<USD 1,000/yr	4,862 (13.5)	433 (22.9)***	
USD 1,000–2,000	8,791 (24.4)	530 (28.7)	0.953 (0.854–1.062)
USD 2,000–4,000	14,849 (41.3)	641 (34.7)	
>USD 4,000	7,480 (20.8)	254 (13.8)	
Exercise ¹⁰			
No	16,593 (43.5)	808 (41.2)*	1.082 (0.979–1.196)
Yes	21,589 (41.2)	1,153 (58.8)	
Smoking ¹¹			
No	35,113 (91.7)	1,831 (92.9)	
Former smoking	1,994 (5.21)	90 (4.57)	1.069 (0.787–1.452)
Current smoking	1,168 (3.1)	50 (2.54)	
Alcohol intake ¹²	100.2±260.5	111.1±511.6	
No	23,123 (60.4)	1,290 (65.4)***	0.901 (0.783–1.036)
Mild drink (0–20 g)	7,065 (18.5)	316 (16.0)	
Moderate drink (≥20 g)	8,102 (21.2)	366 (18.6)	
Coffee intake ¹³	3.44±2.99	3.33±2.90*	
Low (<3 g/day)	17,905 (46.8)	997 (50.6)**	0.970 (0.878–1.071)
Medium (3–16 g/day)	6,539 (17.1)	332 (16.8)	
High (≥16 g/day)	13,846 (36.2)	643 (32.6)	
Energy intake, ¹⁴ kcal	1,713±531	1,691±516	0.902 (0.814–0.999)*
CHO percent intake ¹⁵	70.5±20.8	71.0±20.0**	1.368 (1.142–1.639)***
Fat percent intake ¹⁶	13.5±8.7	13.1±8.0**	0.786 (0.681–0.907)**
Protein percent intake ¹⁷	13.2±5.8	13.1±5.6*	0.840 (0.722–0.979)*
Dairy product intake (0.75 serving/day) ¹⁸	113±133	115±142	0.997 (0.895–1.111)
Ca intake, ¹⁹ mg/day	442±256	443±265	0.991 (0.875–1.121)
Kimchi, ²⁰ g/day	138±113	130±142	0.817 (0.732–0.913)***

The values represent adjusted means±standard deviations or the number of the subjects (percentage of each group). BMI, body mass index. Statistical significance was set at * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$. ¹ By ANCOVA after adjusted covariates. ² Adjusted ORs by logistic regression analysis with adjustment of covariates such as the location of residency, examination year, BMI, gender, age, the status of smoking and alcohol, levels of education and income, job, energy intake, and medication for arthritis and dermatitis in logistic regression models. ³ By χ^2 test. In logistic regression analysis, each parameter needed to be categorized into the low and high groups with cutoff point that was as following: ⁴ <60 years old, ⁵ <25 kg/m² BMI, ⁶ <126 mL/dL fasting serum glucose or medication for diabetes, ⁷ <6.5% HbA1c or medication for diabetes, ⁸ <130/90 mm Hg for blood pressure, ⁹ <college, formal and current smoking, ¹⁰ <USD 2,000/month, ¹¹ nonsmoking, ¹² <20 g/day alcohol drinking, ¹³ <3 g/day coffee drinking, ¹⁴ <estimated energy intake, ¹⁵ <65 energy % carbohydrate intake, ¹⁶ <15 energy % fat, ¹⁷ <13 energy % protein, ¹⁸ <0.75 serving/day of dairy products adding milk, yogurt, and cheese, ¹⁹ <500 mg/day Ca intake, and ²⁰ <100 g/day kimchi intake.

ble 2). Each selected SNP exhibited a low OR value (0.8646–1.256), indicating the weak association between each SNP and ARC. Therefore, a stronger association of ARC with genetic variants related to lactose and galactose

metabolism needed to be found. We hypothesized that polygenetic risk scores of the genetic variant-genetic variant interactions might be better to show the impact of lactose- and galactose-related genes on ARC risk.

Table 2. The characteristics of the ten genetic variants of genes related to lactose and galactose metabolism in age-related cataract risk used for the generalized multifactor dimensionality reduction analysis

CHR ¹	SNP ²	Location	Mi ³	OR ⁴	L95 ⁵	U95 ⁶	p value for OR ⁷	Genes	Feature	MAF ⁸	HWE ⁹
1	rs3729548	43400352	T ¹⁰	0.9033	0.8201	0.9948	0.0389	SLC2A1	Intron	0.3911	0.5384
1	rs3791047	44236114	G ¹⁰	0.8791	0.7782	0.9931	0.0384	ST3GAL3	Intron	0.1878	0.4017
2	rs2304371	136561557	G ¹⁰	0.8646	0.7685	0.9727	0.0155	LCT	Reference	0.2068	0.7531
2	rs6728956	158165004	C ¹⁰	0.8949	0.8069	0.9924	0.0354	GALNT5	Intron	0.3029	0.8989
3	rs2268536	186756908	G ¹¹	1.111	1.006	1.228	0.0382	ST6GAL1	Intron	0.3054	0.2934
4	rs17058752	173613358	C ¹¹	1.121	1.006	1.248	0.0379	GALNT17	Intron	0.2576	0.5314
8	rs1994788	19578268	C ¹¹	1.189	1.037	1.363	0.0129	CSGALNACT1	Intron	0.1269	0.6282
11	rs10831608	11404814	T ¹¹	1.164	1.006	1.355	0.0497	GALNTL4	Intron	0.1027	0.5454
18	rs1667288	29235544	T ¹¹	1.124	1.005	1.257	0.0411	B4GALT6	Intron	0.2274	0.2696
22	rs9623659	43088971	T ¹¹	1.256	1.111	1.419	0.0003	A4GALT	Coding-synonymous	0.1662	0.8116

SLC2A1, solute carrier family 2 member 1; ST3GAL3, ST3 beta-galactoside alpha-2,3-sialyltransferase 3; LCT, lactase; GALNT5, polypeptide N-acetylgalactosaminyltransferase 5; ST6GAL1, ST6 beta-galactoside alpha-2,6-sialyltransferase 1; GALNT17, polypeptide N-acetylgalactosaminyltransferase 17; CSGALNACT1, chondroitin sulfate N-acetylgalactosaminyltransferase 1; GALNTL4, UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase-like 4; B4GALT6, beta-1,4-galactosyltransferase 6; A4GALT, alpha 1,4-galactosyltransferase. ¹ Chromosome. ² Single-nucleotide polymorphism. ³ Minor allele. ⁴ Odds ratio. ⁵ Lower and ⁶ upper ends of 95% confidence interval. ⁷ p value for OR after adjusting for age, gender, residence area, survey year, BMI, daily energy intake, and levels of education and income. ⁸ Minor allele frequency. ⁹ Hardy-Weinberg equilibrium. ¹⁰ Nonrisk allele. ¹¹ Risk allele.

The Best Model for the Association with ARC Assessed by Gene-Gene Interaction Using GMDR

Since each SNP selected for ARC was weakly associated with ARC, a gene-gene interaction study was conducted to better explain the ARC genetic effect by combining multiple genetic variants. The GMDR approach is used to identify the gene-gene interactions through multifactor dimensionality reduction [29]. GMDR provided multiple sets of the best SNP combinations with gene-gene interaction effects. Although SNPs involved in lactose metabolism were identified, 10 SNPs were considered core SNPs to explain ARC risk. In the process of generating the best model using GMDR, TRBA, TEBA, and CVC were used for validating the gene-gene interaction model [30, 31]. The best set of gene-gene interactions was selected by minimum prediction error to represent high balance accuracy and maximum CVC values [30, 31]. TRBA calculated the accuracy of the prediction model of gene-gene interaction with 80% of the participants after training the program [19]. TEBA represented the accuracy of the prediction model in the rest of the participants. A sign test was used to analyze the statistical significance of TEBA. CVC indicated that the number of genetic variants had cross-validation, and the maximum was 10 out of 10 genetic variants. GMDR was conducted with the risk alleles of the 10 SNPs adjusting for age, gender, and BMI or age, gender, survey year, residence area, BMI, education, and income. The best model for ARC with gene-gene interaction was explored with the selected 10 SNPs by GMDR in consideration of CVC (p value for the GMDR model was 0.001 in both models, and CVC was 10/10) (Table 3). Model 9 plus GALNTL4_rs10831608 showed the highest training accuracy (0.7962) and CVC (10/10). The best model included SLC2A1_rs3729548, ST3GAL3_rs3791047, LCT_rs2304371, GALNT5_rs6728956, ST6GAL1_rs2268536, GALNT17_rs17058752, CSGALNACT1_rs1994788, GALNTL4_rs10831608, B4GALT6_rs1667288, and A4GALT_rs9623659. Therefore, PRSBM, including 10 SNPs, was used for genetic variant-lifestyles interactions.

Adjusted ORs for ARC according to the PRSBM

The PRSBM was calculated by summing the number of risk alleles, and it was divided into three categories: low-PRSBM, 0–7; medium-PRSBM, 8–9; and high-PRSBM, ≥10. High-PRSBM indicated a genetically higher incidence of ARC. The ARC incidence was higher in the high-PRSBM group by 1.874- and 2.140-folds in models 1 and 2 compared to the low-PRSBM (Table 4). Models 1 and 2 included different adjusting covariates. Covariates of model 1 were age, gender, residency location, survey year, BMI, lev-

Table 3. GMDR results of multilocus interaction with genes related to ARC

GMDR models model	Adjusted for age, gender, BMI				Adjusted for age, gender, survey year, residence area, BMI, education, income			
	TRBA ¹	TEBA ²	<i>p</i> value ³	CVC ⁴	TRBA	TEBA	<i>p</i> value	CVC
Model 1 GALNT5_rs6728956	0.5215	0.4988	4 (0.828)	5/10	0.5216	0.4957	3 (0.945)	4/10
Model 1 plus ST6GAL1_rs2268536	0.5314	0.5041	8 (0.055)	6/10	0.5313	0.5041	8 (0.547)	6/10
Model 2 plus B4GALT6_rs1667288	0.5431	0.4957	5 (0.623)	2/10	0.5431	0.4966	5 (0.623)	2/10
Model 3 plus GALNT17_rs17058752	0.5614	0.5277	9 (0.011)	4/10	0.5613	0.5276	9 (0.011)	4/10
Model 4 plus SLC2A1_rs3729548	0.5863	0.5280	9 (0.011)	7/10	0.5862	0.5271	8 (0.055)	7/10
Model 5 plus ST3GAL3_rs3791047	0.6179	0.5310	8 (0.055)	7/10	0.6178	0.5295	8 (0.055)	7/10
Model 6 plus A4GALT_rs9623659	0.6606	0.5223	8 (0.055)	7/10	0.6605	0.5223	8 (0.055)	6/10
Model 7 plus LCT_rs2304371	0.7096	0.5212	7 (0.172)	10/10	0.7096	0.521	7 (0.172)	10/10
Model 8 plus CSGALNACT1_rs1994788	0.7580	0.5232	7 (0.172)	10/10	0.7581	0.5237	7 (0.172)	10/10
Model 9 plus GALNTL4_rs10831608	0.7962	0.5250	10 (0.001)	10/10	0.7962	0.5261	10 (0.001)	10/10

BMI, body mass index. ¹ Trained balanced accuracy; ² test balance accuracy; ³ sign rank (*p* value for the significance of the GMDR model by the sign test with adjusting for covariates designated in the table); ⁴ cross-validation consistency.

Table 4. Adjusted odds ratios for ARC according to the PRSBM for gene-gene interaction after covariate adjustments

	Model 1			Model 2	
	Low-PRSBM (<i>n</i> = 14,420)	Medium-PRSBM (<i>n</i> = 21,641)	High-PRSBM (<i>n</i> = 4,201)	Medium-PRSBM (<i>n</i> = 21,641)	High-PRSBM (<i>n</i> = 4,201)
ARC	1	1.411 (1.219~1.634)	1.874 (1.507~2.332)***	1.369 (1.135~1.650)	2.140 (1.622~2.822)***, 9
BMI ¹	1	0.973 (0.912~1.063)	0.984 (0.912~1.063)	0.956 (0.897~1.051)	0.946 (0.852~1.051)
Metabolic syndrome	1	1.027 (0.964~1.095)	0.931 (0.838~1.034)	1.028 (0.964~1.096)	0.926 (0.832~1.029)
Waist circumference ²	1	1.065 (0.994~1.141)	1.115 (0.997~1.247)	1.085 (0.986~1.195)	1.135 (0.971~1.327)
Fasting serum glucose ³	1	1.004 (0.935~1.077)	1.002 (0.893~1.124)	0.953 (0.811~1.119)	1.033 (0.797~1.339)
HbA1c ⁴	1	1.001 (0.907~1.104)	1.066 (0.911~1.247)	1.008 (0.921~1.114)	1.080 (0.921~1.268)
Hypertension ⁵	1	1.021 (0.918~1.136)	1.055 (0.890~1.250)	0.964 (0.835~1.114)	1.034 (0.829~1.300)
Plasma LDL cholesterol ⁶	1	1.132 (0.993~1.291)	1.040 (0.843~1.283)	1.050 (0.888~1.243)	0.997 (0.763~1.304)
Plasma HDL cholesterol ⁷	1	0.990 (0.942~1.040)	0.946 (0.873~1.026)	0.968 (0.904~1.037)	0.949 (0.848~1.062)
Plasma TG ⁸	1	1.015 (0.910~1.132)	0.891 (0.747~1.064)	1.024 (0.884~1.187)	0.806 (0.634~1.026)

Data were expressed as odds ratios and 95% confidence intervals. PRSBM from 10 SNPs were divided into three categories (0–7, 8–9, and >10) corresponding to the low-risk, medium-risk, and high-risk groups, respectively. Low-PRSBM was the reference for both model 1 and model 2. Model 1: adjusted for sex, age, location of residence, examination year, BMI, levels of income and education, and job. Model 2: adjusted for covariate of model 1 and job, hypertension, activity, energy, milk, percent intake of fat and carbohydrate, and medication for arthritis and dermatitis. The cutoff points of the parameters were: ¹ <25 kg/m² BMI, ² <90 and 85 cm in men and women, respectively, for waist circumferences, ³ <126 mg/dL fasting serum glucose concentrations plus diabetic drug intake, ⁴ <6.5% HbA1c plus diabetic drug intake, ⁵ <130/90 mm Hg in systolic blood pressure/diastolic blood pressure, ⁶ <160 mg/dL for plasma LDL, ⁷ >40 and 50 mg/dL in men and women, respectively, for plasma HDL, and ⁸ <150 mg/dL for plasma TG. ⁹ Statistical significance from low-PRSBM in logistic regression analysis was set at * *p* < 0.05, ** *p* < 0.01, and *** *p* < 0.001.

els of income and education, and job. Those of model 2 were age, gender, location of residency, examination year, the status of smoking and drinking, levels of income and education, job, energy intake, amount of physical activity, hypertension, dairy products, percent intake for carbohydrate and fat, and medication of arthritis and dermatitis. The

PRSBM groups did not significantly associate metabolic syndrome and its individual components including BMI, waist circumferences, fasting serum glucose concentrations, HbA1c, blood pressure, and plasma LDL, HDL, and triglyceride concentrations (Table 4). Therefore, the strong association of PRSBM was specific for ARC.

Table 5. Adjusted odds ratios for the incidence of age-related cataract by PRSBM for gene-gene interaction after covariate adjustments according to the patterns of lifestyles

	Low-PRSBM ² (n = 14,420)	Medium-PRSBM (n = 21,641)	High-PRSBM (n = 4,201)	Gene-nutrient interaction p value ¹¹
Low energy ³	1	1.211 (1.021~1.436) ¹	1.445 (1.121~1.863)*	0.573
High energy		1.091 (0.863~1.379)	1.934 (1.408~2.656)***	
Low CHO ⁴	1	1.082 (0.880~1.331)	1.210 (0.874~1.675)	
High CHO		1.248 (1.037~1.502)	1.986 (1.542~2.557)**	0.026
Low protein ⁵	1	1.047 (0.556~1.970)	1.258 (1.116~1.419)***	
High protein		1.090 (0.869~1.369)	1.633 (1.182~2.258)**	
Low fat ⁶	1	1.183 (0.692~2.021)	1.225 (1.097~1.368)***	0.472
Moderate fat		1.227 (0.924~1.630)	1.792 (1.204~2.669)**	
Low Ca ⁷	1	1.117 (0.951~1.312)	1.507 (1.193~1.903)**	
High Ca		1.304 (0.998~1.704)	1.926 (1.323~2.804)**	0.586
Low alcohol ⁸	1	1.147 (0.967~1.361)	1.803 (1.421~2.289)***	
High alcohol		1.232 (0.976~1.554)	1.305 (0.911~1.870)	
Low dairy products ⁹	1	1.118 (0.949~1.317)	1.556 (1.228~1.971)**	0.049
High dairy products		1.281 (0.993~1.653)	1.318 (1.024~1.648)*	
Low kimchi ¹⁰	1	1.254 (1.049~1.498)	1.622 (1.251~2.102)**	
High kimchi		1.049 (0.843~1.304)	1.639 (1.205~2.229)**	0.041

Reference was the low-PRSBM. Statistical significance by low-PRSBM in logistic regression analysis was set at * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$. ¹ Values were expressed as odds ratio and 95% confidence intervals. ² PRSBM was divided into three categories (0–7, 8–9, and >10) by three groups as the low, medium, and high groups of the best model of GMDR. The cutoff points of the parameters were defined as ³ <estimated energy intake, ⁴ <13% protein, ⁵ <15% fat, ⁶ <65% carbohydrate (CHO), ⁷ <20 g/day alcohol, ⁸ <0.75 serving/day of dairy products adding milk, yogurt, and cheese, ⁹ <500 mg/day Ca intake, and ¹⁰ <100 g/day kimchi. ¹¹ Multivariate regression models include the main effects, interaction terms of gene and main effects (energy and nutrient intake), and potential confounders including sex, age, BMI, the status of smoking and drinking, levels of income and education, job, physical activity, hypertension, energy, milk, percent intake for carbohydrate and fat, and medication for arthritis and dermatitis.

Interaction of PRSBM for the Best Model and Nutrient and Food Intakes in ARC Incidence

ARC incidence exhibited a significant interaction with the intake of carbohydrates, alcohol, dairy product (milk, yogurt, and cheese), and kimchi at the significance level of 0.05 (Table 5). Other dietary factors, including fruits, vegetables, and nuts, did not interact with PRSBM for ARC risk (data not shown). High carbohydrate intake (≥ 65 energy %), but not low carbohydrate intake (< 65 energy %), was associated with ARC by 1.986-folds. PRSBM was positively associated with ARC risk in both low and high intake of dairy products (milk, yogurt, and cheese) and kimchi (Table 5). Only a low alcohol intake showed a positive association of 1.803-energyfold with ARC (Table 5). In Figure 1a, ARC risk was much higher in the high-PRSBM group than the low-PRSBM in the low-carbohydrate category, but there was no significant association with PRSBM with ARC risk in the high-carbohydrate category. Interestingly, ARC risk was higher in subjects with the high-PRSBM than those with the low-PRSBM with low dairy product consumption

(Fig. 1b). In the high dairy product category, subjects with high-PRSBM had a higher ARC risk, but they had a lower risk in the low dairy product category. Kimchi intake showed a similar trend to dairy product intake. In low kimchi intake, ARC risk was higher in the high-PRSBM than the low-PRSBM (Fig. 1c). The trend for higher ARC risk in the low kimchi intake was similar in the high-kimchi intake, but the ARC risk in the high-PRSBM was lower in the high-kimchi category than the low-kimchi category. The participants with high-PRSBM had a high ARC risk only in low alcohol intake but not moderate alcohol intake (Fig. 1d). With moderate alcohol intake (< 20 g/day), PRSBM did not associate with ARC incidence.

Discussion

In this study, we established a PRSBM for cataracts using GMDR and evaluated whether they are related to the incidence of cataracts. We also evaluated the gene-nutri-

Fig. 1. The frequency distribution of ARC in the three groups of PRSBM according to nutrient and food intake. PRSBM was categorized into three groups (low-PRSBM, medium-PRSBM, and high-PRSBM). The PRSBM had interactions with carbohydrates, dairy products, kimchi, and alcohol intake. ARC incidence (%) in the three PRSBM groups according to the following categories:

a The low- and high-carbohydrate (CHO) intake by 65 energy % of carbohydrate. **b** The low- and high-dairy product intake by 150 g/day. **c** The low- and high-kimchi (Korean fermented cabbage) by 100 g/day. **d** The low- and moderate-alcohol intake by 20 g/day. Significant differences among the PRSBM groups at * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$.

ent interaction by analyzing the nutritional effect on the risk of ARC by PRSBM. We observed that the cataract risk was 2.1 times higher in the high-PRSBM than in the low-PRSBM after adjusting for covariates. The genetic variants in the PRSBM were selected from the genes related to galactose metabolism since the disturbance of galactose metabolism, especially *GALK*, is associated with ARC risk. The A198V mutation of *GALK* is prevalent in Asians (4.1% in Japanese and 2.8% in Koreans) but not Caucasians and Blacks [5]. Genetic variants in *GALK* had no association with ARC in the present study. ARC's genetic risk was modified by nutrients, including consuming carbohydrates and foods such as alcohol, dairy product, and kimchi. Specifically, ARC's genetic risk significantly increased in groups with high carbohydrates, low alcohol drinking, low dairy product consumption, and low kimchi consumption.

The origin of galactose is mostly from lactose derived from dairy products, mainly milk. Galactose is converted to glucose, primarily in the liver and brain [32]. Three enzymes that convert galactose to glucose-6-phosphate are encoded by the *GALK*, *GALT*, and phosphoglucomutase (*PGM*) genes. UDP-galactose and UDP-glucose interconversion require the enzyme encoded by the *GALE* gene UDP-glucose pyrophosphorylase 2 encoded by *UGP2*. Other genes related to galactose and lactose metabolism were also considered possible links to ARC, and 10 SNPs were selected for gene-gene interaction since no individual SNP had a significant impact on ARC risk. In GMDR, the best model included all 10 SNPs, and PRSBM had much higher adjusted ORs (2.14) for ARC risk. Interestingly, the PRSBM had interactions with dietary intakes that lower ARC incidence in Koreans.

Low dairy product consumption (<150 g/day) was found to increase the genetic risk of ARC, which increased in the participants with high-PRSBM in the present study. These results suggest that low dairy product intake might reduce the expression of the selected genes and that the galactose metabolism might be increased with high-PRSBM when galactose metabolism is needed. It remains controversial whether dairy product consumption is a risk factor for cataract formation [33]. An experimental animal study provided quantitative evidence that excessive milk intake aggravates ARC in rats due to oxidative damage caused by increased reactive oxygen species [34]. However, a recent longitudinal epidemiologic study reported no significant association between dairy product consumption and the risk of incident cataracts in 5,860 subjects [35]. The finding in the present study suggests that the cataractogenic action of lactose in dairy products, which is linked to the ARC genetic risk, could be exaggerated in low dairy product intake groups, possibly due to less expression of the related genes. Therefore, a low intake of dairy products may reduce the expression of genes encoding lactose and galactose enzymes, reducing the activity of those enzymes in people with high-PRSBM, resulting in reduced galactose metabolism and a concomitant increased ARC risk. Therefore, the continuous and proper amount of dairy product intake may normalize the gene expression related to lactose and galactose metabolism, and ARC might be lessened, especially in adults with high-PRSBM. However, this hypothesis has not been tested in human trials and needs confirmation.

ARC risk was highly positively associated with a high carbohydrate intake (≥ 65 energy %) by 1.37-fold in the present study. It was a consistent finding given that high carbohydrate consumption was regarded as a risk factor for cataracts. A cross-sectional epidemiologic study with 1,609 subjects reported that nuclear cataract was associated with total carbohydrate intake, although this study failed to show a dose-dependent relationship [36]. A recent meta-analysis confirmed that the risk of ARC was 1.18 and 1.15 times higher for the highest versus the lowest category of carbohydrate intake and glycemic index [37]. The effect of high carbohydrate intake on ARC risk in Koreans can explain this finding; an environmental risk factor could attenuate the impact of the genetic risk over the ARC risk. Furthermore, only the group with high carbohydrate intake showed a significant genetic risk of developing cataracts. Persons with high-PRSBM could reduce their risk of cataracts by decreasing carbohydrate intake (<65 energy %).

Alcohol drinking is a well-known risk factor for cataracts. Alcohol drinking also modifies the genetic risk factor of cataract development. However, alcohol intake did not play as a risk factor for ARC incidence in the present study. A meta-analysis involving a total of 119,706 participants demonstrated that relative risks in heavy drinkers, which were defined by >20 g/day of alcohol intake versus non-drinkers, were 1.25 for cataract surgery, 1.06 for cortical cataracts, and 1.26 for nuclear cataracts, relatively. These association between alcohol consumption and cataracts is not significant, consistent with the present study [38]. On the other hand, alcohol intake had an interaction with PRSBM in the present study. We found that the ARC risk by PRSBM was significant only in groups with low alcohol consumption, but more than 20 g/day alcohol intake was not associated with ARC's genetic risk. It suggested that moderate alcohol intake had a tendency to lower risk in the participants with high-PRSBM who elevated ARC incidence in low alcohol intake. The possible mechanism is that alcohol metabolism generates free radicals, aggregates lens proteins, and subsequently forms cataracts, and the genetic variants in the PRSBM may be involved in removing free radicals and preventing the aggregation of lens proteins by alcohol [39]. This strong environmental risk factor of alcohol consumption may attenuate the genetic risk factor's effect on the overall risk of cataract development.

On the other hand, physical exercise did not exhibit a significant association with ARC in the present study. It may be related to spending more time outside with greater exposure to UV light from the sun during exercise. A recent meta-analysis has demonstrated that increased physical activity is associated with reduced ARC risk by 10%, related to improving oxidative stress [40]. Although a positive relationship between ARC and physical activity has been reported [40], ambient UV light exposure increases ARC risk [41]. Therefore, UV light exposure needs to be avoided during outdoor physical activity to prevent ARC in older persons.

Interestingly, the consumption of kimchi, the Korean traditional probiotic vegetable food, significantly modified the genetic risk effect on ARC in this study. The group with low kimchi intake had a higher odds ratio for genetic risk of cataract of 1.254 and 1.622 for medium- and high-PRSBM, respectively, whereas the group with the intake of kimchi over 100 g/day showed that OR for genetic risk of cataract was 1.049 and 1.639 for medium- and high-PRSBM, respectively. Although ARC's genetic risk showed significantly different patterns, with the medium intake being adequate for decreasing ARC risk but not high intake of Kimchi consumption, we could not

draw any meaningful clinical conclusion due to the small effect size and inconsistent pattern.

This study has limitations as well as strengths. This study's strength includes a relatively large number of participants aged >50 years old, including 1,971 subjects having ARC and 38,290 participants having no cataracts, which could increase the robustness of the present study. Our study also has several limitations. First, the cataract incidence was assessed by using a questionnaire of cataract diagnosis by a physician, which may not be as accurate as a slit-lamp examination of the lens due to the recall bias. The second limitation is that the current study has a cross-sectional design, which makes inferring causality difficult. However, it is unlikely that the presence of ARC changes the eating patterns. The third limitation was no duplication study in another cohort since no cohort analyzing genetic variants included sufficient ARC patients to conduct a duplication study. The fourth limitation is that the SQFFQ method used in this study for nutrition evaluation may have some shortfalls. It might overestimate the nutrition intakes at individual levels since it relies on the participant's memory to recall past intake. However, SQFFQ is widely used in many other studies after developing and validating the KoGES [25, 42, 43].

In conclusion, ARC incidence might be positively associated about 2-fold with PRSBM involved in lactose and galactose metabolism and the nutrition-gene interaction for the ARC risk. The lactose- and galactose-related genetic impact on ARC risk interacted with carbohydrate intake, alcohol drinking, dairy products, and kimchi intake. Adults aged >50 years with high-PRSBM might benefit from consuming dairy products and kimchi daily and avoiding a high carbohydrate diet (≥ 65 energy %) associated with a lower incidence of ARC. Given that cataract is the most common cause for low vision, further study is warranted to assess the utility of nutrition-gene interactions to modify and decrease cataract risk.

References

- 1 Liu YC, Wilkins M, Kim T, Malyugin B, Mehta JS. Cataracts. *Lancet*. 2017 Aug 5; 390(10094):600–12.
- 2 Kasuya M, Itoi M, Kobayashi S, Sunaga H, Suzuki KT. Changes of glutathione and taurine concentrations in lenses of rat eyes induced by galactose-cataract formation or ageing. *Exp Eye Res*. 1992 Jan;54(1):49–53.
- 3 Bosch AM. Classical galactosaemia revisited. *J Inherit Metab Dis*. 2006;29(4):516–25.
- 4 Lorenzi M. The polyol pathway as a mechanism for diabetic retinopathy: attractive, elusive, and resilient. *Exp Diabetes Res*. 2007; 2007:61038.
- 5 Okano Y, Asada M, Fujimoto A, Ohtake A, Murayama K, Hsiao KJ, et al. A genetic factor for age-related cataract: identification and characterization of a novel galactokinase variant, “Osaka,” in Asians. *Am J Hum Genet*. 2001 Apr;68(4):1036–42.
- 6 Timson DJ, Lindert S. Comparison of dynamics of wildtype and V94M human UDP-galactose 4-epimerase-A computational perspective on severe epimerase-deficiency galactosemia. *Gene*. 2013 Sep 10;526(2):318–24.
- 7 Sneha P, Ebrahimi EA, Ghazala SA, Thirumal Kumar D, Siva R, Priya Doss CG, et al. Structural analysis of missense mutations in galactokinase 1 (GALK1) leading to galactosemia type-2. *J Cell Biochem*. 2018 Sep;119(9):7585–98.
- 8 Shiels A, Hejtmancik JF. Mutations and mechanisms in congenital and age-related cataracts. *Exp Eye Res*. 2017 Mar;156:95–102.
- 9 Anand D, Agrawal SA, Slavotinek A, Lachke SA. Mutation update of transcription factor genes FOXE3, HSF4, MAF, and PITX3 causing cataracts and other developmental ocular defects. *Hum Mutat*. 2018 Apr;39(4):471–94.

Statement of Ethics

The study was conducted according to the Declaration of Helsinki's guidelines and approved by the Institutional Review Board of the Korean National Institute of Health (1041231-190902-BR-099-01). We additionally obtained approval from the Institutional Review Board of Hoseo University (1041231-150811-HR-034-01). We obtained written informed consent from all participants involved in the study.

Conflict of Interest Statement

The authors declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding Sources

This research and the study design and methods were evaluated and approved for funding by the National Research Foundation (NRF-2019R1A2C1007203), South Korea.

Author Contributions

S.P. and D.J. participated in the experimental design, analyzed the data, and wrote the manuscript. S.K. designed the research and analyzed the data. S.P. and S.K. had primary responsibility for the final content. All authors listed as authors have substantially contributed to the work and seen and approved the submitted version.

Data Availability Statement

All data generated or analyzed during this study are included in this article. Further enquiries can be directed to the corresponding author.

- 10 Rubio-Gozalbo ME, Derks B, Das AM, Meyer U, Möslinger D, Couce ML, et al. Galactokinase deficiency: lessons from the GalNet registry. *Genet Med*. 2021 Jan;23(1):202–10.
- 11 Fassio F, Facioni MS, Guagnini F. Lactose maldigestion, malabsorption, and intolerance: a comprehensive review with a focus on current management and future perspectives. *Nutrients*. 2018 Nov 1;10(11):1599.
- 12 Misselwitz B, Butter M, Verbeke K, Fox MR. Update on lactose malabsorption and intolerance: pathogenesis, diagnosis and clinical management. *Gut*. 2019 Nov;68(11):2080–91.
- 13 Malik TF, Panuganti KK. Lactose intolerance. In: *StatPearls [Internet]*. Treasure Island, FL: StatPearls Publishing LLC; 2020.
- 14 Solomons NW. Fermentation, fermented foods and lactose intolerance. *Eur J Clin Nutr*. 2002 Dec;56(Suppl 4):S50–5.
- 15 Kaur M, Singh H, Jangra M, Kaur L, Jaswal P, Dureja C, et al. Lactic acid bacteria isolated from yak milk show probiotic potential. *Appl Microbiol Biotechnol*. 2017 Oct;101(20):7635–52.
- 16 Oh SL, Cheng LY, Zhou JZJ, Henke W, Hagen T. Galactose 1-phosphate accumulates to high levels in galactose-treated cells due to low GALT activity and absence of product inhibition of GALK. *J Inherit Metab Dis*. 2020; 43(3):529–39.
- 17 Rasmussen SA, Daenzer JMI, MacWilliams JA, Head ST, Williams MB, Geurts AM, et al. A galactose-1-phosphate uridylyltransferase-null rat model of classic galactosemia mimics relevant patient outcomes and reveals tissue-specific and longitudinal differences in galactose metabolism. *J Inherit Metab Dis*. 2019 Dec 17;43(3):518–28.
- 18 Park S, Ham JO, Lee BK. Effects of total vitamin A, vitamin C, and fruit intake on risk for metabolic syndrome in Korean women and men. *Nutrition*. 2015 Jan;31(1):111–8.
- 19 Hong KW, Kim SH, Zhang X, Park S. Interactions among the variants of insulin-related genes and nutrients increase the risk of type 2 diabetes. *Nutr Res*. 2018 Mar;51:82–92.
- 20 Uma Jyothi K, Reddy BM. Gene-gene and gene-environment interactions in the etiology of type 2 diabetes mellitus in the population of Hyderabad, India. *Meta Gene*. 2015 Sep;5:9–20.
- 21 Park S, Ahn J, Lee BK. Self-rated subjective health status is strongly associated with sociodemographic factors, lifestyle, nutrient intakes, and biochemical indices, but not smoking status: KNHANES 2007–2012. *J Korean Med Sci*. 2015 Sep;30(9):1279–87.
- 22 Kim Y, Han BG. Cohort profile: the Korean Genome and Epidemiology Study (KoGES) Consortium. *Int J Epidemiol*. 2017 Aug 1; 46(4):1350.
- 23 Lee SK, Kim MK. Relationship of sodium intake with obesity among Korean children and adolescents: Korea National Health and Nutrition Examination Survey. *Br J Nutr*. 2016 Mar 14;115(5):834–41.
- 24 Rabbie N, Speed TP. A genotype calling algorithm for affymetrix SNP arrays. *Bioinformatics*. 2006 Jan 1;22(1):7–12.
- 25 Ahn Y, Kwon E, Shim JE, Park MK, Joo Y, Kimm K, et al. Validation and reproducibility of food frequency questionnaire for Korean genome epidemiologic study. *Eur J Clin Nutr*. 2007 Dec;61(12):1435–41.
- 26 Park S, Ahn J, Lee BK. Very-low-fat diets may be associated with increased risk of metabolic syndrome in the adult population. *Clin Nutr*. 2016;35(5):1159–67.
- 27 Nam GE, Han K, Ha SG, Han BD, Kim DH, Kim YH, et al. Relationship between socioeconomic and lifestyle factors and cataracts in Koreans: the Korea National Health and Nutrition Examination Survey 2008–2011. *Eye*. 2015;29(7):913–20.
- 28 Ahn J, Kim NS, Lee BK, Park Kim S. Trends in the intake of fatty acids and their food source according to obese status among Korean adult population using KNHANES 2007–2017. *Food Nutr Bull*. 2020 Mar;41(1):77–88.
- 29 Xu H-M, Xu L-F, Hou T-T, Luo L-F, Chen G-B, Sun X-W, et al. GMDR: versatile software for detecting gene-gene and gene-environment interactions underlying complex traits. *Curr Genom*. 2016;17(5):396–402.
- 30 Zhu Z, Tong X, Zhu Z, Liang M, Cui W, Su K, et al. Development of GMDR-GPU for gene-gene interaction analysis and its application to WTCCC GWAS data for type 2 diabetes. *PLoS One*. 2013;8(4):e61943.
- 31 Daily JW, Liu M, Park S. High genetic risk scores of SLIT3, PLEKHA5 and PPP2R2C variants increased insulin resistance and interacted with coffee and caffeine consumption in middle-aged adults. *Nutr Metab Cardiovasc Dis*. 2019 Jan;29(1):79–89.
- 32 Roser M, Josic D, Kontou M, Mosetter K, Maurer P, Reutter W. Metabolism of galactose in the brain and liver of rats and its conversion into glutamate and other amino acids. *J Neural Transm*. 2009 Feb;116(2):131–9.
- 33 Mustafa OM, Daoud YJ. Is dietary milk intake associated with cataract extraction history in older adults? An analysis from the US population. *J Ophthalmol*. 2020;2020:2562875.
- 34 Chen Y, Jiang YX, Yi L, Miu AZ, Zhang SJ, Wu JH, et al. Excessive milk intake as a risk factor, probably associated with oxidative stress, in experimental naphthalene-initiated cataract in rats. *Ophthalmic Res*. 2012;47(2):87–97.
- 35 Camacho-Barcia L, Bullo M, Garcia-Gavilan JF, Martinez-Gonzalez MA, Corella D, Estruch R, et al. Dairy products intake and the risk of incident cataracts surgery in an elderly Mediterranean population: results from the PREDIMED study. *Eur J Nutr*. 2019 Mar; 58(2):619–27.
- 36 Chiu C-J, Robman L, McCarty CA, Mukesh BN, Hodge A, Taylor HR, et al. Dietary carbohydrate in relation to cortical and nuclear lens opacities in the Melbourne visual impairment project. *Invest Ophthalmol Vis Sci*. 2010; 51(6):2897–905.
- 37 Wu H, Zhang H, Li P, Gao T, Lin J, Yang J, et al. Association between dietary carbohydrate intake and dietary glycemic index and risk of age-related cataract: a meta-analysis. *Invest Ophthalmol Vis Sci*. 2014;55(6):3660–8.
- 38 Wang W, Zhang X. Alcohol intake and the risk of age-related cataracts: a meta-analysis of prospective cohort studies. *PLoS One*. 2014;9(9):e107820.
- 39 Balog Z, Sikić J, Vojniković B, Balog S. Senile cataract and the absorption activity of cytochrome C oxidase. *Coll Antropol*. 2001; 25(Suppl 1):33–6.
- 40 Jiang H, Wang L-N, Liu Y, Li M, Wu M, Yin Y, et al. Physical activity and risk of age-related cataract. *Int J Ophthalmol*. 2020;13(4):643–9.
- 41 Delcourt C, Cougnard-Grégoire A, Boniol M, Carrière I, Doré J-F, Delyfer M-N, et al. Lifetime exposure to ambient ultraviolet radiation and the risk for cataract extraction and age-related macular degeneration: the Alienor Study. *Invest Ophthalmol Vis Sci*. 2014; 55(11):7619–27.
- 42 Kim SH, Choi HN, Hwang JY, Chang N, Kim WY, Chung HW, et al. Development and evaluation of a food frequency questionnaire for Vietnamese female immigrants in Korea: the Korean Genome and Epidemiology Study (KoGES). *Nutr Res Pract*. 2011 Jun;5(3):260–5.
- 43 Kim D, Kim J. Dairy consumption is associated with a lower incidence of the metabolic syndrome in middle-aged and older Korean adults: the Korean Genome and Epidemiology Study (KoGES). *Br J Nutr*. 2017 Jan;117(1):148–60.