

Title Page

Title:

Sex differences in the risk of cataract associated with type 2 diabetes: a Mendelian randomization study

Authors and affiliations:

Haoyang Zhang^{a,b}, Xuehao Xiu^b, Yuedong Yang^{a,*}, Yuanhao Yang^{c,d,*}, Huiying Zhao^{b,*}

^aSchool of Data and Computer Science, Sun Yat-sen University, Guangzhou, China

^bDepartment of Medical Research Center, Sun Yat-sen Memorial Hospital; Guangdong Provincial Key Laboratory of Malignant Tumor Epigenetics and Gene Regulation; Guangzhou, China

^cInstitute for Molecular Bioscience, The University of Queensland, Brisbane, QLD, Australia

^dMater Research, Translational Research Institute, Brisbane, QLD, Australia

Contact information of corresponding authors:

Huiying Zhao: zhaohy8@mail.sysu.edu.cn

Yuanhao Yang: yuanhao.yang@mater.uq.edu.au

Yuedong Yang: yangyd25@mail.sysu.edu.cn

Abstract

Type 2 diabetes (T2D) is a recognized risk factor for developing cataract. However, it is unclear if the shared genetic variance and potential genetic causal relationship between T2D and cataract are different for males and females. We evaluated sex-specific genetic correlation (r_g) and putative genetic causality between the two diseases by using linkage disequilibrium score regression (LDSC) and six Mendelian randomization (MR) approaches after leveraging large-scale population-based genome-wide association studies (GWAS) summary of T2D and cataract. Application of LDSC found a significant genetic correlation between T2D and cataract in East Asian males ($r_g=0.68$, 95% confident interval [CI]=0.17 to 1, p-value= 8.60×10^{-3}) but a non-significant genetic correlation in East Asian females ($r_g=0.25$, CI= -0.02 to 0.52, p-value= 8.38×10^{-2}). MR analyses indicated a consistently stronger (paired t-test $|t|=5.87$, p-value= 2.04×10^{-3}) causal effect of T2D on cataract in East Asian males (liability OR=1.20 to 1.41, p-value= 5.86×10^{-27} to 6.60×10^{-6}) than in females (liability OR=1.12 to 1.21, p-value= 2.02×10^{-14} to 1.82×10^{-2}). In Europeans, the LDSC analysis suggested a close significant genetic correlation between the two diseases in males ($r_g=0.20$, 95% confident interval [CI]=0.08 to 0.32, p-value= 7.00×10^{-4}) and females ($r_g=0.17$, CI= 0.05 to 0.29, p-value= 4.90×10^{-3}); but the MR analyses provided weak evidences on a causal relationship between the two diseases in both sexes. These results presented the first evidence on sex difference of the casual relationship between cataract and T2D in East Asians, and supported a potential genetic heterogeneity of the shared genetics underlying T2D and cataract between East Asians and Europeans in both sexes.

Keywords Sex difference; Mendelian randomization; Type 2 diabetes; Cataract

Introduction

Type 2 diabetes (T2D) is one of the most prevalent chronic diseases(1) and a recognized risk factor for cataract development(2). Previous studies have revealed both phenotypic(2, 3) and genotypic(4, 5) associations between the two diseases. Although a higher risk of cataract in female than in male was well-established by multiple phenotypic studies(6-8), the sex difference of cataract risk in T2D patients was still inconclusive and being studied. Olafsdottir et al.(9) observed the risk of nuclear cataract in T2D females was higher than that of T2D males from an Icelandic cohort ($N=531$). Another Indian T2D cohort study ($N=779$) conducted by Srinivasan et al. also identified an increased risk of posterior subcapsular cataract or nuclear cataract in females compared to males. However, a recent study(10) in Australia, Fremantle Diabetes Study Phase II (FDS2) ($N=6316$), suggested sex was not a risk factor for incidence of intraocular lens implantation. Similarly, another Taiwan cohort study of 578 T2D participants also found that the sex difference of cataract risk did not exist in the multivariate analysis (controlling covariates like Body mass index [BMI], Systolic blood pressure [SBP], and Hemoglobin A1c[HbA1c])(11). These results suggested that potential confounding factors and limit sample size may influence the estimation of sex difference in cataract risk. Thus, it is worthwhile to investigate such a sex-differential effect on the basis of genetic correlation and causal relationships by leveraging large population-based genetics studies.

With the development of genome-wide association studies (GWAS) during past decades, multiple methods were exploited to investigate the shared genetic architecture and genetic casual relationships between traits. For example, by examining the linkage disequilibrium (LD) score from single nucleotide polymorphisms (SNPs) across two traits, linkage disequilibrium score regression (LDSC)(12) was designed to quantify the genetic correlation feasibly by measuring the correlation of LD scores from two traits. For traits with significant genetic correlation,

Mendelian randomization (MR)(13) is a popular tool for searching potential causal effects from a trait (i.e., exposure) to another (i.e., outcome) by using independent SNPs strongly associated with the exposure trait as instruments. MR exploits the concept of natural random allocation of genetic variant to mimic the randomization procedure in randomized controlled trials (RCTs). The genetic causality inferred by MR is less-biased because the allocation of genetic variant is generally not correlated with both exposure and outcome(13). A special application of MR is summary-data-based Mendelian randomization (SMR)(14) that treats gene expression in expression quantitative trait loci (eQTL) summary data as “exposure” and a focal trait as “outcome”. Identifying potential functional genes related with a genetic correlation, SMR may provide new insights for the understanding of shared genetic etiology (15). Application of MR analysis indicated a potential causal effect of T2D on cataract in East Asians(15), but it is far from clear if a sex difference exists in the causal relationship between T2D and cataract.

In this study, we utilized sex-stratified GWAS summary statistics of T2D and cataract from multiple sources (i.e., Asian Genetic Epidemiology Network [AGEN](16), BioBank Japan [BBJ](17), and UK Biobank (UKB)(18)) to investigate the genetic relationship between T2D and cataract in both East Asians and Europeans. We performed sex-stratified LDSC and MR analysis to measure the genetic correlation and genetic causality between T2D and cataract, and tested if the genetic correlation and genetic causality were different between males and females. Then, we implemented SMR analysis (14) to identify potential functional genes underlining the putative sex-differential causality. The flowchart of our study was shown in Fig. 1.

Methods and materials

Datasets

We downloaded the sex-stratified hg19-based GWAS summary statistics of T2D ($N_{\text{case}}=40,250$ [male=25,705, female=14,545], $N_{\text{control}}=170,615$ [male=82,774, female=87,841])(19) and cataract ($N_{\text{case}}=24,622$ [male=11,641, female=12,981], $N_{\text{control}}=187,831$ [male=97,706, female=90,125])(19) from the BBJ (<http://jenger.riken.jp/en/>), namely BBJ-T2D GWAS and BBJ-cataract GWAS. BBJ(17) was a large database including phenotype and genotype data from around 200,000 Japanese individuals (proportion of male=53.10%; average baseline age at 62.70 for men and 61.50 for women). In BBJ, both T2D and cataract were diagnosed by physicians; the sex-stratified GWAS of two diseases were performed using a linear mixed model via SAIGE (20) adjusted by age and top five genetic principal components.

We downloaded another sex-stratified T2D GWAS summary statistics from AGEN (<https://blog.nus.edu.sg/agen/summary-statistics/t2d-2020/>), namely AGEN-T2D GWAS. The AGEN-T2D GWAS (16) was generated using a fixed-effect inverse-variance-weighted meta-analysis via METAL(21) from 22 East Asian population-based cohorts ($N_{\text{case}}=117,339$ [male=28,027, female=89,312], $N_{\text{control}}=162,425$ [male=27,370, female=135,055]). The East Asian samples of AGEN-T2D GWAS was approximately independent to that of BBJ-T2D GWAS. Thus, we employed AGEN-T2D and BBJ-cataract to investigate the genetic correlation, as well as the causal relationship between two diseases in East Asians. Parallel analysis between BBJ-T2D and BBJ-cataract was also performed to investigate the robustness of the findings.

Genome-wide association study on T2D and cataract of Europeans from UK Biobank

We generated the sex-stratified GWAS summary statistics of T2D ($N_{\text{case}}=29,267$ [male=17,688, female=11,579], $N_{\text{control}}=473,237$ [male=211,434, female=261,803]) and cataract ($N_{\text{case}}=33,842$ [male=14,758, female=19,084], $N_{\text{control}}=468,662$ [male=214,364, female=254,298]) using the European cohorts in UK Biobank (18), namely

UKB-T2D GWAS and UKB-cataract GWAS. UK Biobank is a prospective study with over 500,000 participants across different population (proportion of male=45.60%; average baseline age at 56.74 for men and 56.35 for women). We determined the case of T2D and cataract patients by ICD-10 code (UKB field 41270). An individual with ICD-10 code of E11 (non-insulin-dependent diabetes mellitus) was collected as a T2D patient while an individual with ICD-10 codes of H25 (senile cataract) or H26 (other cataract) was collected as a cataract patient. The genetic location of the genotypic data was based on hg19. The sex-stratified GWAS was performed by BOLT-LMM (22) adjusting for age, sex, and cryptic relatedness.

Estimation of heritability and genetic correlation

We performed sex-stratified LDSC(12, 23) to estimate the single-trait liability-scale heritability (h^2) of T2D and cataract, and the cross-trait genetic correlation (r_g) between: 1. AGEN-T2D and BBJ-cataract, 2. BBJ-T2D and BBJ-cataract, and 3. UKB-T2D and UKB-cataract.

We used the genotype data of 481 East Asians and 489 Europeans in the 1000 Genomes as reference genome. The reference genome and the pre-calculated LD scores were downloaded from the LDSC website (<https://alkesgroup.broadinstitute.org/LDSCORE/>). We conducted quality control for SNPs inputting into LDSC analysis by: 1. filtering SNPs using the HapMap3(24); 2. removing the SNPs if they were strand-ambiguous; 3. removing SNPs whose minor allele frequency (MAF)<0.01; 4. removing SNPs located in the major histocompatibility complex (MHC) region (chromosome 6: 28,477,797–33,448,354)(25).

Single-trait LDSC was performed to estimate the sex-stratified liability-scale h^2 of T2D and cataract. For East Asians, we set the population prevalence of T2D (male: 5.90%(26), female: 7.00%(26)) and cataract (male:

40.60%(27), female: 42.33%(27)) according to cohort studies conducted by Wang et al.(26) and Tang et al.(27).

For Europeans, we set the population prevalence of T2D (male: 3.80%(28); female: 3.10%(28)) and cataract (male: 30.02%(29), female: 40.69%(29)) according to a health survey conducted by National Health Service (NHS) Health and Social Care Information Centre(28) and a cross-sectional study conducted by Reidy et al.(29).

Cross-trait LDSC was applied to measure the sex-stratified genetic correlation between T2D and cataract. We performed cross-trait LDSC twice, with and without a constrained intercept, since constraining the intercept may increase the accuracy of estimation when the impact from population stratification is limited. A significant r_g was determined with $p\text{-value} < 0.05/2 = 0.025$. The differences of r_g between two sexes were evaluated by Z test.

MR analyses for genetic causality

The sex-stratified causal relationship between T2D and cataract was evaluated by multiple MR approaches. Inverse-variance-weighted (IVW) method(30) is a basic MR method that integrates the GWAS effect ratios of instrumental SNPs between exposure and outcome in a fixed effect meta-analysis model. However, the horizontal pleiotropy(13) (i.e., a potential confounding that instrumental SNPs effect on outcome through non-causal pathways) may affect measurement of casual relationship between traits for IVW. Horizontal pleiotropy contains uncorrelated pleiotropy and correlated pleiotropy. Uncorrelated pleiotropy is a type of pleiotropy whose effect on outcome is independent to exposure, and correlated pleiotropy is another type of pleiotropy whose effect on outcome is related with exposure(13). To reduce the probability of false positive in MR analysis, except for IVW, five other MR methods(i.e., MR-Egger (31), generalized summary-data-based Mendelian randomization [GSMR] (32), weighed median (33), weighted mode (34), and the causal analysis using summary effect estimates [CAUSE] (35)) with different assumptions on horizontal pleiotropy were also employed(36).

Comparing to IVW, MR-Egger made an improvement to account for uncorrelated pleiotropy by adding an intercept in the IVW regression model. In GSMR analysis, SNPs were excluded if they involve in uncorrelated pleiotropy by a heterogeneity in dependent instrument (HEIDI) outlier test. The weighted median model measures the causal effect as the weighted median of GWAS estimates ratios to reduce the impact from pleiotropic SNPs. The weighted mode model loosens the assumption of weighted median model (i.e., pleiotropy is existed in less than half of instrumental SNPs), and measures the causal effect from the most frequent (mode) instrumental SNPs. For these five methods (IVW, MR-Egger, GSMR, weighted median, and weight mode), we selected independent instrumental SNPs with GWAS p -value $< 5 \times 10^{-8}$, and clumped with LD $r^2 < 0.05$ within 1000-kb window using PLINK 1.9(37) according to the same 1000 Genomes reference described above(38). CAUSE (35) is a more powerful method to identify the causality from both types of pleiotropy by pruning SNPs with a GWAS p -value threshold at 1×10^{-3} and a LD r^2 threshold at 0.1. It also conducted a model comparison between the causal model (i.e., instrumental SNPs act on exposure and outcome through both causal pathways and shared factors) and sharing model (i.e., instrumental SNPs act on exposure and outcome only through shared factors) based on the expected log pointwise posterior density (ELPD) test.

We used six MR analysis approaches to examine the genetic causality between T2D and cataract. The MR analyses were performed by R packages “cause”, “gsmr” and “TwoSampleMR”. Instrumental SNPs with MAF > 0.01 or located within the MHC region were excluded(22). A significant causal relationship was determined if the MR p -value reached Bonferroni-corrected level (i.e., less than $0.05/12 \approx 4.17 \times 10^{-3}$, adjusted by six bi-directional tests). The logit-scale MR estimates (i.e., β_{MR}) were further transformed to liability-scale according to the formula proposed by Byrne et al⁽³⁹⁾:

$$\beta_{\text{MR}}(\text{liability}) = \frac{Z_{K_{\text{exposure}}}K_{\text{outcome}}(1-K_{\text{outcome}})}{Z_{K_{\text{outcome}}}K_{\text{exposure}}(1-K_{\text{exposure}})}\beta_{\text{MR}}(\text{logit}),$$

where K_{exposure} and K_{outcome} represent the population prevalence of exposure and outcome while Z_{K_x} and Z_{K_y} represent the value of standard normal distribution at such prevalence. We assumed the population prevalence is the same as that used in estimating the liability heritability. The liability β_{MR} was then converted into odds ratio (OR). The sex difference of causal effect between two sexes estimated with liability β_{MR} was evaluated by paired t-test.

SMR analysis to identify candidate functional genes underling the genetic relationship of T2D and cataract

If a sex-differential genetic relationship was suggested by MR analysis, we further explored potential functional genes associated with the sex-differential causal relationship via SMR(14). SMR can detect functional genes associated with a trait by treating gene expression as “exposure” and a trait as an “outcome” (e.g., T2D) with top cis-eQTL (GWAS p-value $<5\times 10^{-8}$) associated with each gene (19,250 probes, assessed from a blood-based cis-eQTL summary dataset(40), URL: <https://eqtlgen.org/cis-eqtls.html>). We applied SMR analysis on each single-trait GWAS, as well as the cross-trait GWAS of T2D and cataract generated with inverse-variance-weighted meta-analysis via MTAG(41).

To exclude associations introduced by linkage (e.g., a top SNP was in LD with causal SNPs which affecting a gene expression and the trait, respectively), we conducted HEIDI test on each SMR estimates. Significant functional genes were determined with SMR p-value $<0.05/19250\approx 2.60\times 10^{-6}$ and HEIDI p-value >0.05 from at least 10 SNPs.

Results

The genetic correlation between T2D and cataract in males and females

Single-trait LDSC indicated the liability-scale heritability of T2D and cataract in males (AGEN-T2D: $h^2=0.22$, $p\text{-value}=6.94\times 10^{-20}$; BBJ-cataract: $h^2=0.02$, $p\text{-value}=1.51\times 10^{-1}$) and in females (AGEN-T2D: $h^2=0.32$, $p\text{-value}=2.15\times 10^{-11}$; BBJ-cataract: $h^2=0.02$, $p\text{-value}=1.45\times 10^{-1}$) were approximately equal in East Asians. Using cross-trait LDSC, we identified a significant genetic correlation between AGEN-T2D and BBJ-cataract in males ($r_g=0.68$, 95% confident interval [CI]=0.17 to 1, $p\text{-value}=8.60\times 10^{-3}$) but not in females ($r_g=0.25$, CI= -0.02 to 0.52, $p\text{-value}=8.38\times 10^{-2}$). The genetic correlation in females were lower than that in males although the gap was non-significant ($|Z|=1.68$, $p\text{-value}=9.30\times 10^{-2}$). The analysis using BBJ-T2D GWAS and BBJ-cataract estimated a higher genetic correlation (males: $r_g=0.72$, 95%CI=0.23 to 1, $p\text{-value}=4.10\times 10^{-3}$; females: $r_g=0.45$, 95%CI=0.08 to 0.82, $p\text{-value}=1.57\times 10^{-2}$) than using AGEN-T2D and BBJ-cataract, suggesting a potential inflation introduced by partial sample overlap from two BBJ cohorts.

To estimate the sex-stratified genetic correlation between T2D and cataract in Europeans, we carried out GWAS analysis based on the European cohorts in UK Biobank. The Manhattan and quantile–quantile plots of the distribution of GWAS p-value were displayed in Fig. S1 and Fig. S2. The effect size of genome-wide significant SNPs with GWAS $p\text{-value}<5\times 10^{-8}$ were detailed in Table S1. Single-trait LDSC detected a mild inflation of our GWAS analysis (UKB-T2D male: $\lambda_{gc}=1.20$, LDSC intercept=1.03; UKB-T2D female: $\lambda_{gc}=1.25$, LDSC intercept=1.04; UKB-cataract male: $\lambda_{gc}=1.10$, LDSC intercept=1.01; UKB-cataract female: $\lambda_{gc}=1.09$, LDSC intercept=1.02). The liability-scale heritability of cataract in males ($h^2=0.15$, $p\text{-value}=7.03\times 10^{-15}$) is close to the heritability of cataract in females ($h^2=0.11$, $p\text{-value}=4.80\times 10^{-9}$) while that of T2D in males ($h^2=0.15$, $p\text{-value}=6.03\times 10^{-45}$) was lower than that in females ($h^2=0.29$, $p\text{-value}=5.16\times 10^{-52}$). The genetic correlation be-

tween the two diseases in males ($r_g=0.20$, $p\text{-value}=7.00\times 10^{-4}$) was approximately equal ($|Z|=0.33$, $p\text{-value}=0.74$) to that in females ($r_g=0.17$, $p\text{-value}=4.90\times 10^{-3}$). The detailed statistics of the LDSC analysis was listed in Table S2.

Sex differences in the putative causality of T2D on cataract

Based on the strong genetic correlation between T2D and cataract in East Asians, we implemented multiple MR methods to infer the sex-specific causal relationship between the two diseases (see Fig. 2 and Table. S3). By analyzing AGEN-T2D GWAS and BBJ-cataract GWAS, we identified robust evidences for a putative causal effect of T2D on cataract for East Asian males (liability OR=1.20 to 1.41, $p\text{-value}=5.86\times 10^{-27}$ to 6.60×10^{-6}), with all of six MR models surpassing the Bonferroni-corrected significance threshold ($p\text{-value}<4.17\times 10^{-3}$) (Fig. 2a). We also observed suggestive evidence for the putative causal effect of T2D on cataract in females (liability OR=1.12 to 1.21, $p\text{-value}=2.02\times 10^{-14}$ to 1.82×10^{-2}) with four of six MR models (except for MR-Egger and weighted mode model) reaching the Bonferroni-corrected significance threshold. Notably, the estimated MR odds ratios (ORs) for males was significantly stronger than that for females (paired t-test $|t|=5.87$, $p\text{-value}=2.04\times 10^{-3}$). The results of CAUSE suggested that comparing to healthy individuals, a male with T2D had approximately 1.20 times risks for cataract, and the risk would be reduced to 1.12 for females (Fig. 3a). The sex difference in causal effect from T2D to cataract was also supported by a parallel analysis between BBJ-T2D and BBJ-cataract (paired t-test $|t|=20.45$, $p\text{-value}=5.18\times 10^{-6}$) (Fig. 2b & Fig. 3b). In the reverse direction, there was no evidence for a causal effect of cataract to T2D in both sexes (male: liability OR = 1.01 to 1.24, $p\text{-value}=1.97\times 10^{-1}$ to 9.37×10^{-1} ; female: liability OR = 0.97 to 1.09, $p\text{-value}=5.20\times 10^{-1}$ to 9.58×10^{-1}) (Fig. 2c).

These MR results from T2D to cataract were less likely to be influenced by the horizontal pleiotropy (Table S4) because of the non-significant and close-to-zero MR-Egger intercept (-0.008 for male and 0.004 for female, both

p-value>0.05) and the better fitness of causal model than sharing model in CAUSE (ELPD p-value<0.05 in both sexes).

When the MR analysis was applied in Europeans, we found the causal effect of T2D on cataract was non-significant (male: liability OR=1.05 to 1.07, p-value= 3.01×10^{-2} to 6.33×10^{-1} ; female: liability OR=1.04 to 1.13, p-value= 8.97×10^{-3} to 2.62×10^{-1}) in both males and females, as well as in the reverse direction (Fig. 2c). As revealed by CAUSE, comparing to the sharing model, the better fitness of causal model in Europeans was less significant (ELPD p-value= 1.94×10^{-1} for males and 7.14×10^{-2} for females) than that in East Asians (ELPD p-value= 1.33×10^{-2} for males and 1.31×10^{-2} for females), suggesting a higher pleiotropy in the analysis of Europeans than in East Asians (Fig. 3 & Table S4).

Candidate genes likely to be involved in the shared genetics between T2D on cataract in East Asian

To identify functional genes associated with the sex difference on the risk of cataract in T2D patients, we applied SMR to the cross-trait meta-analysis GWAS (i.e., T2D and cataract) as well as the single-trait GWAS. The cross-trait meta-analysis identified five genes (*TLE1*, *BRD3*, *AP3S2*, *KCNJ11*, and *KLHL42*) associated with the cross-trait of T2D and cataract in East Asian males (Table 1). Among them, two genes, *BRD3* ($\beta_{\text{SMR}}=-0.14 \pm 0.03$, SMR p-value= 9.28×10^{-7} , HEIDI p-value=0.20 from 14 SNPs) and *KLHL42* ($\beta_{\text{SMR}}=0.15 \pm 0.03$, SMR p-value= 1.56×10^{-6} , HEIDI p-value=0.06 from 20 SNPs) were not identified to be associated with individual traits by SMR analysis. These two genes may play important roles in causing cataract in males of East Asian with T2D. When analyzing on East Asian females, we only identified one gene, *KCNJ11* ($\beta_{\text{SMR}}=0.38 \pm 0.07$, SMR p-value= 1.15×10^{-8} , HEIDI p-value=0.11 from 16 SNPs) as associated with the cross-trait of T2D and cataract.

Interestingly, *KCNJ11* was also identified as a functional gene in both the single trait analysis of T2D and the cross-trait analysis of East Asian males.

Discussion

Previous epidemiology studies have paid a lot of efforts to explore the sex difference on pathways associated with co-occurrence of T2D and cataract by cross-sectional investigations and cohort studies (9, 11, 42). However, these studies are potentially limited due to their small sample size of patients co-occurring both diseases and laborious covariates collection process. Importantly, the knowledge of sex difference on the basis of shared genetics underlying two diseases was limited. In the study, by using the large population-based sex-stratified GWAS summary, we performed LDSC and MR analysis to investigate the sex-differential effect in genetic correlation and causal relationship between T2D and cataract. We found that T2D patients in East Asian males have probably higher risk in developing cataract than females, and such a sex-differential casual effect was not observed in Europeans. To our knowledge, this is the first attempt to explore the sex difference in potential genetic causality between T2D and cataract using population-based GWAS summary statistics.

Application of six MR methods consistently disclosed that T2D has higher putative causal effect on cataract in East Asian males than in females. Remarkably, the sex difference in the causal relationship between T2D and cataract were incomparable between East Asians and Europeans. The potential heterogeneity of the shared genetics between two diseases in East Asians and Europeans was investigated by LDSC analysis (Table S5). In results, we observed a significant genetic correlation of T2D between two populations for both sexes, with estimated r_g in the range of 0.67 to 0.71. However, LDSC provided weak evidence ($p\text{-value}>0.1$) for a

cross-population genetic correlation on cataract. These results suggested the cross-population heterogeneity of genetic correlations, as well as causal relationships, may be mostly due to the heterogeneity of cataract.

The previous epidemiology studies have widely accepted that the females have higher risk to suffer from cataract than males (6-8, 10, 43-46) while a number of studies suggested that sex was not a risk factor for cataract in people with T2D (10, 11, 47-49). A possible explanation to this difference is that some male-specific covariates may have a role in the development of cataract via intersecting with T2D. Moreover, our study further revealed a stronger putative casual effect of T2D on cataract in males than females in East Asians, suggesting the consequence of the interaction between male-specific covariates and T2D on cataract development may be not merely an equal risk of cataract. Further investigations are required to clarify such sex-specific factors in relationship between T2D and cataract.

The SMR analysis on the cross-trait of T2D and cataract revealed two male-specific functional genes, *KLHL42* and *BRD3*, to be relevant to the genetic causality of T2D on cataract in East Asian males. *KLHL42* encoded a protein in Kelch-like family (kelch like family member 42), and it was known to be a susceptible gene of T2D (50, 51). *BRD3* was reported to be associated with Brain-derived neurotrophic factor (BDNF) by a GAWs on Europeans (52). BDNF was revealed to be associated with increased risks of various diseases including T2D(52). The SMR analysis on East Asian males identified *KCNJ11* as associated with T2D while the SMR analysis on the cross-trait of T2D and cataract of East Asian females also revealed *KCNJ11* as a female-specific functional gene relevant to the genetic causality of T2D on cataract. *KCNJ11* is involved in the encoding of potassium channel and is reported to be associated with both T2D and congenital cataract in GWAS study of East Asians(53, 54). This

study offered us the evidence that *KCNJ11* may be a candidate gene associated with putative causal effect of T2D on cataract in East Asian males, but its role in females is still unclear.

Our study has two limitations. First, since there were no other published European sex-stratified T2D and cataract GWAS apart from that in UKB, we conducted analysis between two GWAS summary based on the sample cohort. Such an operation may introduce a sample-overlap bias in our European analysis. Secondly, we used a loosen p-value threshold, 1×10^{-5} to obtain at least 10 instruments in testing the causal effect of cataract on T2D in East Asians in MR analysis. This operation may violate the assumption of GSMR (GSMR recommended at least 10 independent SNPs with GWAS p-value $< 5 \times 10^{-8}$ associated with exposure to maintain the study power). Nevertheless, our MR analysis provided universally weak evidences to support the causal effect from cataract to T2D, which implicated the little influence of the loosen P-value threshold in this study.

In summary, we provided first and robust evidences for a sex-differential causality of T2D on cataract in East Asians, and identified several candidate genes associated with the putative causality. The results of this study provided theoretical fundament for early-stage prevention of cataract in East Asian T2D patients.

Declarations

Ethics approval

The BBJ project was approved by research ethics committees at the Institute of Medical Science, the University of Tokyo, the RIKEN Yokohama Institute, and the 12 cooperating hospitals. The UK Biobank project was approved by the North West Multi-center Research Eth-

ics Committee. The 22 sub cohorts used in AGEN-T2D GWAS were approved by relevant institutions.

Funding

The work was partly funded by the National Key R&D Program of China (2020YFB0204803; YD.Y, Sun Yat-sen University), the Natural Science Foundation of China (81801132, and 81971190; HY.Zhao, Sun Yat-sen Memorial Hospital; 61772566, 62041209, and U1611261; YD.Y, Sun Yat-sen University), Guangdong Key Field R&D Plan (2019B020228001 and 2018B010109006; YD.Y, Sun Yat-sen University), Introducing Innovative and Entrepreneurial Teams (2016ZT06D211, YD.Y, Sun Yat-sen University), Guangzhou S&T Research Plan (202007030010, YD.Y, Sun Yat-sen University), and Mater Foundation (YH.Y, Mater Research).

Conflict of Interest

All authors state they have no competing interests.

Availability of data and material

The European genotypic data and phenotypic data of T2D and cataract can be assessed from the UKB project (<https://www.ukbiobank.ac.uk/>) after application.

The East Asian GWAS summary statistics of T2D and cataract were available in the BBJ project (<http://jenger.riken.jp/en/result>) and AGEN database (<https://blog.nus.edu.sg/agen/summary-statistics/t2d-2020/>).

References

1. Charvat H, Goto A, Goto M, et al. Impact of population aging on trends in diabetes prevalence: A meta-regression analysis of 160,000 Japanese adults. *Journal of Diabetes Investigation*. 2015;6(5):533-42.

2. Tan AG, Kifley A, Tham YC, et al. Six-Year Incidence of and Risk Factors for Cataract Surgery in a Multi-ethnic Asian Population: The Singapore Epidemiology of Eye Diseases Study. *Ophthalmology*. 2018;125(12):1844-53. doi:10.1016/j.ophtha.2018.07.026
3. Foster PJ, Wong TY, Machin D, Johnson GJ, Seah SKL. Risk factors for nuclear, cortical and posterior subcapsular cataracts in the Chinese population of Singapore: the Tanjong Pagar Survey. *Br J Ophthalmol*. 2003;87(9):1112-20.
4. Lee SC, Wang Y, Ko GTC, Ma RCW, Chan JCN. Risk factors for cataract in Chinese patients with type 2 diabetes: evidence for the influence of the aldose reductase gene. *Clinical Genetics*. 2010;59(5):356-9.
5. Lin HJ, Huang YC, Lin JM, et al. Novel susceptibility genes associated with diabetic cataract in a Taiwanese population. *Ophthalmic genetics*. 2013;34(1-2):35-42. doi:10.3109/13816810.2012.736590
6. Lou L, Ye X, Xu P, et al. Association of Sex With the Global Burden of Cataract. *JAMA Ophthalmol*. 2018;136(2):116-21. doi:10.1001/jamaophthalmol.2017.5668
7. Hennis, Anselm. Risk Factors for Incident Cortical and Posterior Subcapsular Lens Opacities in the Barbados Eye Studies. *Archives of Ophthalmology*. 2004;122(4):525.
8. Varma R, Torres M. Prevalence of lens opacities in Latinos: the Los Angeles Latino Eye Study. *Ophthalmology*. 2004;111(8):1449-56. doi:10.1016/j.ophtha.2004.01.024
9. Olafsdottir E, Andersson DK, Stefánsson E. The prevalence of cataract in a population with and without type 2 diabetes mellitus. *Acta Ophthalmol*. 2012;90(4):334-40. doi:10.1111/j.1755-3768.2011.02326.x
10. Drinkwater JJ, Davis TME, Turner AW, Bruce DG, Davis WA. Incidence and Determinants of Intraocular Lens Implantation in Type 2 Diabetes: The Fremantle Diabetes Study Phase II. *Diabetes care*. 2019;42(2):288-96. doi:10.2337/dc18-1556
11. Tung TH, Liu JH, Lee FL, Chen SJ, Tsai CY, Chou P. Community-based study of cataracts among type 2 diabetics in Kinmen. *Eur J Epidemiol*. 2005;20(5):435-41. doi:10.1007/s10654-004-7537-9
12. Bulik-Sullivan BK, Loh P-R, Finucane HK, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics*. 2015;47(3):291-5. doi:10.1038/ng.3211
13. Davies NM, Holmes MV, Davey Smith G. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. *Bmj*. 2018;362:k601. doi:10.1136/bmj.k601
14. Zhu Z, Zhang F, Hu H, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet*. 2016;48(5):481-7. doi:10.1038/ng.3538
15. Zhang H, Xiu X, Xue A, Yang Y, Yang Y, Zhao H. The putative causal effect of type 2 diabetes in risk of cataract: a Mendelian randomization study in East Asian. 2021:2021.02.08.430342. doi:10.1101/2021.02.08.430342 %J bioRxiv
16. Spracklen CN, Horikoshi M, Kim YJ, et al. Identification of type 2 diabetes loci in 433,540 East Asian individuals. *Nature*. 2020;582(7811):240-5. doi:10.1038/s41586-020-2263-3
17. Nagai A, Hirata M, Kamatani Y, et al. Overview of the BioBank Japan Project: Study design and profile. *J Epidemiol*. 2017;27(3S):S2-S8. doi:10.1016/j.je.2016.12.005
18. Sudlow C, Gallacher J, Allen N, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med*. 2015;12(3):e1001779. doi:10.1371/journal.pmed.1001779
19. Ishigaki K, Akiyama M, Kanai M, Takahashi A, Kamatani Y. Large-scale genome-wide association study in a Japanese population identifies novel susceptibility loci across different diseases. *Nature Genetics*. 2020.
20. Zhou W, Nielsen JB, Fritsche LG, et al. Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nature genetics*. 2018;50(9):1335-41. doi:10.1038/s41588-018-0184-y
21. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans.

- Bioinformatics. 2010;26(17):2190-1. doi:10.1093/bioinformatics/btq340
22. Loh P-R, Tucker G, Bulik-Sullivan BK, et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nature Genetics*. 2015;47(3):284-90. doi:10.1038/ng.3190
 23. Bulik-Sullivan B, Finucane HK, Anttila V, et al. An atlas of genetic correlations across human diseases and traits. *Nature Genetics*. 2015;47(11):1236-41. doi:10.1038/ng.3406
 24. International HapMap C, Altshuler DM, Gibbs RA, et al. Integrating common and rare genetic variation in diverse human populations. *Nature*. 2010;467(7311):52-8. doi:10.1038/nature09298
 25. da Silva JS, Wowk PF, Poerner F, Santos PS, Bicalho Mda G. Absence of strong linkage disequilibrium between odorant receptor alleles and the major histocompatibility complex. *Human immunology*. 2013;74(12):1619-23. doi:10.1016/j.humimm.2013.08.005
 26. Wang Z, Wu Y, Wu J, et al. Trends in prevalence and incidence of type 2 diabetes among adults in Beijing, China, from 2008 to 2017. *Diabetic medicine : a journal of the British Diabetic Association*. 2020:e14487. doi:10.1111/dme.14487
 27. Tang Y, Wang X, Wang J, et al. Prevalence of Age-Related Cataract and Cataract Surgery in a Chinese Adult Population: The Taizhou Eye Study. *Investigative ophthalmology & visual science*. 2016;57(3):1193-200. doi:10.1167/iovs.15-18380
 28. The Information Centre. Health Survey for England 2004: health of ethnic minorities:. 2006.
 29. Reidy A, Minassian DC, Vafidis G, et al. Prevalence of serious eye disease and visual impairment in a north London population: population based, cross sectional study. *Bmj*. 1998;316(7145):1643-6. doi:10.1136/bmj.316.7145.1643
 30. Burgess S, Dudbridge F, Thompson SG. Combining information on multiple instrumental variables in Mendelian randomization: comparison of allele score and summarized data methods. *Stat Med*. 2016;35(11):1880-906. doi:10.1002/sim.6835
 31. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *International journal of epidemiology*. 2015;44(2):512-25.
 32. Zhu Z, Zheng Z, Zhang F, et al. Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nature Communications*. 2018;9(1):224. doi:10.1038/s41467-017-02317-2
 33. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. *Genet Epidemiol*. 2016;40(4):304-14. doi:10.1002/gepi.21965
 34. Hartwig FP, Davey Smith G, Bowden J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. *International journal of epidemiology*. 2017;46(6):1985-98. doi:10.1093/ije/dyx102
 35. Morrison J, Knoblach N, Marcus JH, Stephens M, He X. Mendelian randomization accounting for correlated and uncorrelated pleiotropic effects using genome-wide summary statistics. *Nature Genetics*. 2020;52(7):740-7. doi:10.1038/s41588-020-0631-4
 36. Burgess S, Davey Smith G, Davies NM, et al. Guidelines for performing Mendelian randomization investigations. *Wellcome Open Res*. 2020;4:186. doi:10.12688/wellcomeopenres.15555.2
 37. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81(3):559-75. doi:10.1086/519795
 38. Auton A, Abecasis GR, Altshuler DM, et al. A global reference for human genetic variation. *Nature*. 2015;526(7571):68-74. doi:10.1038/nature15393
 39. Byrne EM, Zhu Z, Qi T, et al. Conditional GWAS analysis to identify disorder-specific SNPs for psychiatric disorders. *Molecular Psychiatry*. 2020. doi:10.1038/s41380-020-0705-9
 40. Vösa U, Claringbould A, Westra H-J, et al. Unraveling the polygenic architecture of complex traits using

- blood eQTL metaanalysis. 2018:447367. doi:10.1101/447367 %J bioRxiv
41. Turley P, Walters RK, Maghizian O, et al. Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nat Genet.* 2018;50(2):229-37. doi:10.1038/s41588-017-0009-4
 42. Srinivasan S, Raman R, Swaminathan G, Ganesan S, Kulothungan V, Sharma T. Incidence, Progression, and Risk Factors for Cataract in Type 2 Diabetes. *Invest Ophthalmol Vis Sci.* 2017;58(13):5921-9. doi:10.1167/iovs.17-22264
 43. Zetterberg M, Celojevic D. Gender and cataract--the role of estrogen. *Current eye research.* 2015;40(2):176-90. doi:10.3109/02713683.2014.898774
 44. Klein BE, Klein R, Linton KL. Prevalence of age-related lens opacities in a population. The Beaver Dam Eye Study. *Ophthalmology.* 1992;99(4):546-52. doi:10.1016/s0161-6420(92)31934-7
 45. Leske MC, Connell AM, Wu SY, Hyman L, Schachat A. Prevalence of lens opacities in the Barbados Eye Study. *Archives of ophthalmology (Chicago, Ill. : 1960).* 1997;115(1):105-11. doi:10.1001/archophth.1997.01100150107018
 46. Mitchell P, Cumming RG, Attebo K, Panchapakesan J. Prevalence of cataract in Australia: the Blue Mountains eye study. *Ophthalmology.* 1997;104(4):581-8. doi:10.1016/s0161-6420(97)30266-8
 47. Klein BE, Klein R, Moss SE. Incidence of cataract surgery in the Wisconsin Epidemiologic Study of Diabetic Retinopathy. *Am J Ophthalmol.* 1995;119(3):295-300. doi:10.1016/s0002-9394(14)71170-5
 48. Becker C, Schneider C, Aballéa S, et al. Cataract in patients with diabetes mellitus-incidence rates in the UK and risk factors. *Eye (London, England).* 2018;32(6):1028-35. doi:10.1038/s41433-017-0003-1
 49. Adler AI, Stratton IM, Neil HA, et al. Association of systolic blood pressure with macrovascular and microvascular complications of type 2 diabetes (UKPDS 36): prospective observational study. *Bmj.* 2000;321(7258):412-9. doi:10.1136/bmj.321.7258.412
 50. Huopio H, Cederberg H, Vangipurapu J, et al. Association of risk variants for type 2 diabetes and hyperglycemia with gestational diabetes. *European journal of endocrinology.* 2013;169(3):291-7. doi:10.1530/eje-13-0286
 51. Viñuela A, Varshney A, van de Bunt M, et al. Genetic variant effects on gene expression in human pancreatic islets and their implications for T2D. *Nature Communications.* 2020;11(1):4912. doi:10.1038/s41467-020-18581-8
 52. Li S, Weinstein G, Zare H, et al. The genetics of circulating BDNF: towards understanding the role of BDNF in brain structure and function in middle and old ages. *Brain Commun.* 2020;2(2):fcaa176-fcaa. doi:10.1093/braincomms/fcaa176
 53. Suzuki K, Akiyama M, Ishigaki K, et al. Identification of 28 new susceptibility loci for type 2 diabetes in the Japanese population. *Nat Genet.* 2019;51(3):379-86. doi:10.1038/s41588-018-0332-4
 54. Cao B, Gong C, Wu D, et al. Genetic Analysis and Follow-Up of 25 Neonatal Diabetes Mellitus Patients in China. *J Diabetes Res.* 2016;2016:6314368. doi:10.1155/2016/6314368

Table 1. Candidate genes involving in the shared genetics between T2D and cataract in East Asians.

Sex	Trait	Probe and located gene			Top SNP								SMR test			
		Chr	BP	Gene symbol	SNP	BP	A1	A2	β_{GWAS}	P_{GWAS}	β_{eQTL}	P_{eQTL}	β_{SMR}	P_{SMR}	P_{HEIDI}	$N_{HEIDI\ SNP}$
Male	T2D (BBJ)	9	84251409	<i>TLE1</i>	rs7024675	84296357	C	G	0.10±0.01	2.69E-12	0.09±0.01	3.92E-24	1.14±0.20	8.64E-09	2.70E-01	20
	T2D (BBJ)	15	90405702	<i>AP3S2</i>	rs3759831	90414642	C	G	0.07±0.01	5.39E-08	0.52±0.01	0.00E+00	0.14±0.03	6.22E-08	2.12E-01	12
	T2D (AGEN)	2	27590286	<i>EIF2B4</i>	rs6547692	27734972	G	A	-0.08±0.01	1.18E-12	-0.07±0.01	1.48E-16	1.12±0.21	6.90E-08	6.05E-02	20
	T2D (AGEN)	2	27657891	<i>NRBP1</i>	rs4665972	27598097	T	C	-0.09±0.01	3.93E-11	0.18±0.01	7.55E-104	-0.50±0.08	2.43E-10	5.08E-02	20
	T2D (AGEN)	4	153000000	<i>MIR4453HG</i>	rs6832998	153470666	T	A	-0.07±0.01	2.63E-09	0.15±0.02	3.32E-15	-0.46±0.10	1.97E-06	3.31E-01	15
	T2D (AGEN)	11	17409142	<i>KCNJ11</i>	rs2074310	17421886	T	C	0.08±0.01	1.37E-10	0.16±0.01	4.37E-38	0.49±0.09	8.55E-09	1.25E-01	17
	T2D (AGEN)	15	90405702	<i>AP3S2</i>	rs3759831	90414642	C	G	0.07±0.01	5.64E-07	0.52±0.01	0.00E+00	0.14±0.03	5.92E-07	5.79E-02	17
	Cross-trait (BBJ-T2D and BBJ-cataract)	9	84251409	<i>TLE1</i>	rs3814535	84304985	A	G	0.07±0.01	2.56E-09	0.09±0.01	4.31E-24	0.81±0.16	2.42E-07	4.11E-01	20
	Cross-trait (BBJ-T2D and BBJ-cataract)	9	137000000	<i>BRD3</i>	rs467369	136905765	C	T	-0.05±0.01	7.28E-07	0.32±0.01	7.19E-274	-0.14±0.03	9.28E-07	2.04E-01	14
	Cross-trait (BBJ-T2D and BBJ-cataract)	15	90405702	<i>AP3S2</i>	rs12440156	90420673	G	A	0.06±0.01	4.46E-07	0.52±0.01	0.00E+00	0.11±0.02	4.73E-07	2.03E-01	11
	Cross-trait (AGEN-T2D and BBJ-cataract)	11	17409142	<i>KCNJ11</i>	rs2074310	17421886	T	C	0.05±0.01	2.46E-07	0.16±0.01	4.37E-38	0.32±0.07	1.62E-06	3.23E-01	16
	Cross-trait (AGEN-T2D and BBJ-cataract)	12	27944463	<i>KLHLA2</i>	rs12578595	27964996	T	C	-0.05±0.01	1.18E-06	-0.33±0.01	3.21E-208	0.15±0.03	1.56E-06	6.14E-02	20
	Cross-trait (AGEN-T2D and BBJ-cataract)	15	90405702	<i>AP3S2</i>	rs12440156	90420673	G	A	0.05±0.01	2.33E-06	0.52±0.01	0.00E+00	0.10±0.02	2.56E-06	5.55E-02	12
	Female	T2D (AGEN)	4	71828104	<i>MOB1B</i>	rs12649012	71748245	G	A	0.08±0.01	7.63E-10	0.15±0.02	9.91E-16	0.56±0.12	1.00E-06	8.57E-02
T2D (AGEN)		15	75654552	<i>MAN2C1</i>	rs11636031	75815758	C	T	-0.06±0.01	6.04E-08	0.92±0.01	0.00E+00	-0.07±0.01	5.80E-08	2.36E-01	20
T2D (AGEN)		15	75904617	<i>SNUPN</i>	rs8036758	75817562	A	G	-0.06±0.01	1.02E-07	0.41±0.01	0.00E+00	-0.15±0.03	1.23E-07	4.59E-01	20
Cross-trait (AGEN-T2D and BBJ-cataract)		11	17409142	<i>KCNJ11</i>	rs2074310	17421886	T	C	0.06±0.01	2.07E-10	0.16±0.01	4.37E-38	0.38±0.07	1.15E-08	1.13E-01	16

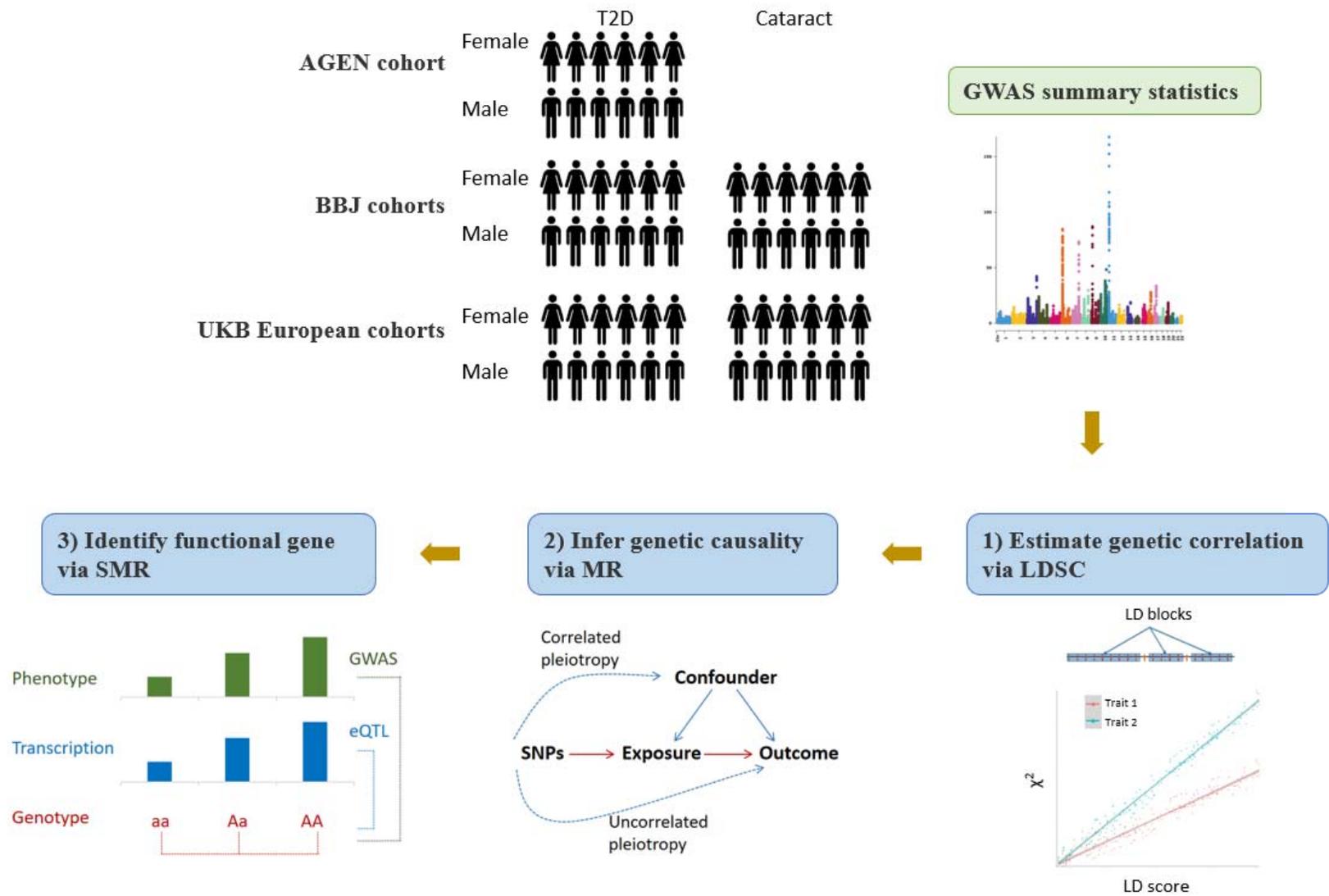
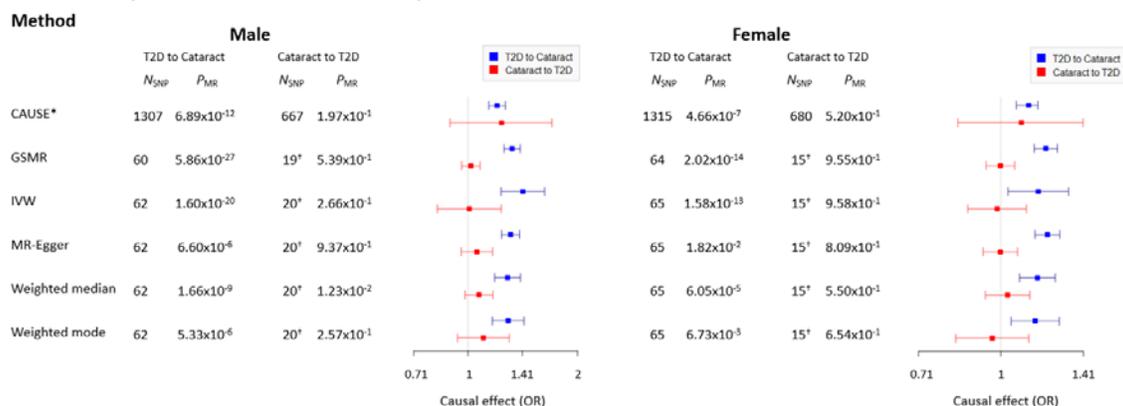


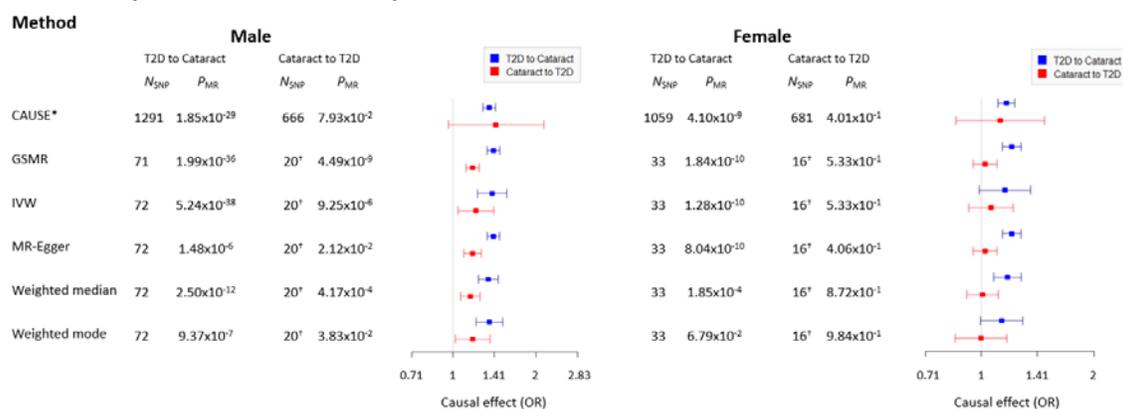
Figure 1. Outline of analysis performed in the study. Utilizing large-scale population-based GWAS summary statistics from multiple source, the study conducted a series of sex-stratified analysis to investigate the shared genetics underling T2D and cataract. It included three sections: 1. Measuring the sex-stratified genetic correlation between T2D and cataract with LDSC; 2. Inferring the sex-stratified causal relationship between T2D and cataract using six MR approaches; 3. Identifying candidate functional genes underlying the putative sex-differential causality between T2D and cataract with SMR.

LDSC: linkage disequilibrium score regression; MR: Mendelian randomization. SMR: summary data-based Mendelian randomization.

a. East Asian (AGEN-T2D and BBJ-cataract)



b. East Asian (BBJ-T2D and BBJ-cataract)



c. European (UKB-T2D and UKB-cataract)

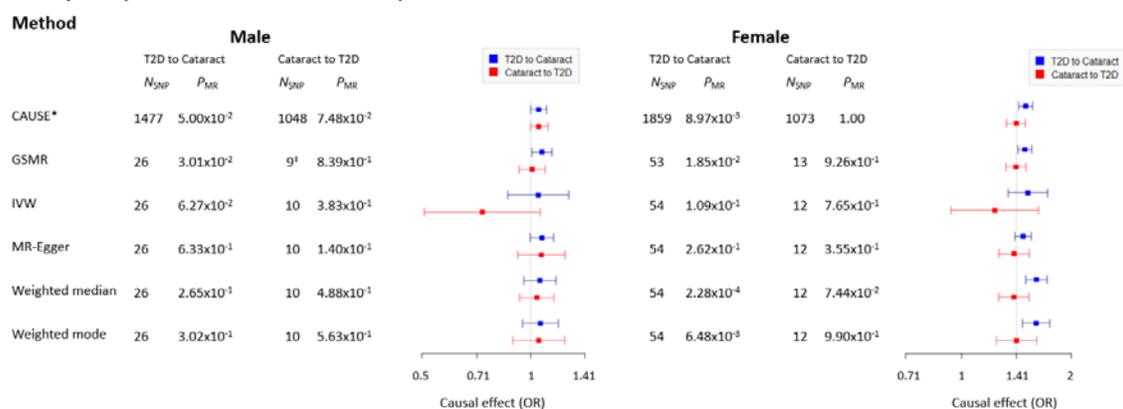


Figure 2. a. The sex-stratified bi-directional causal effect between T2D and cataract in East Asians estimated by six MR methods using AGEN-T2D and BBJ-cataract. b. The sex-stratified bi-directional causal effect between T2D and cataract in East Asians estimated by six MR methods using BBJ-T2D and BBJ-cataract. c. The sex-stratified bi-directional causal effect between T2D and cataract in European estimated by six MR methods

using UKB-T2D and BBJ-cataract. In the forest plot, the blue bar represents the MR estimates at direction of T2D to cataract (i.e., T2D as exposure and cataract as outcome), the red bar represents the MR estimates at direction of cataract to T2D (i.e., cataract as exposure and T2D as outcome).

CAUSE: causal analysis using summary effect estimates; GSMR: generalized summary-data-based Mendelian randomization; IVW: inverse variance weighted. *CAUSE recruited independent instrumental SNPs with GWAS p-value $<1 \times 10^{-3}$. †IVW, MR-Egger, GSMR, weighted median, and weighted mode recruited 'proxy' independent instrumental SNPs with GWAS p-value $<1 \times 10^{-5}$. ‡10 instrumental SNPs were provided for GSMR analysis, and 1 SNP was removed by HEIDI test.

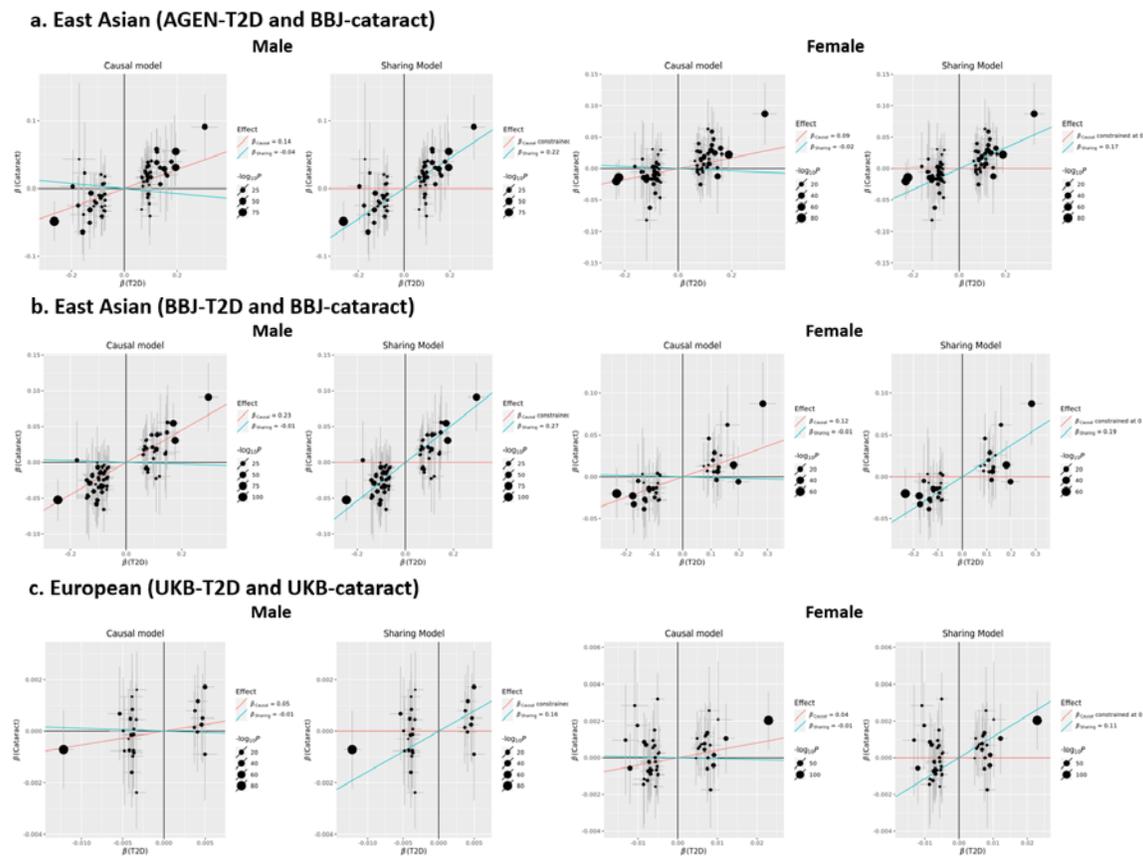


Figure 3. The sex-stratified CAUSE fitted causal model and sharing model using T2D as exposure and cataract as outcome. a. CAUSE models in East Asians using GWAS summary of AGEN-T2D and BBJ-cataract. b. CAUSE models in East Asians using GWAS summary of BBJ-T2D and BBJ-cataract. c. CAUSE models in Europeans using GWAS summary of UKB-T2D and UKB-cataract.

Points represent instrumental SNPs with GWAS p -value $< 5 \times 10^{-8}$ and their size represent the magnitude of GWAS p -value. Error bars stand for the 95% confidence intervals (CIs) of the SNP effect sizes based on the GWAS of T2D (x-axis) and cataract (y-axis). Blue line represents the estimated sharing effect between T2D and cataract. Red line represents the estimated causal effect of T2D on cataract, which would be constrained at zero in the sharing model.

CAUSE: causal analysis using summary effect estimates.