Journal of the American Heart Association

ORIGINAL RESEARCH

Association of Common and Rare Genetic Variation in the 3-Hydroxy-3-Methylglutaryl Coenzyme A Reductase Gene and Cataract Risk

Jonas Ghouse, MD, PhD; Gustav Ahlberg, MSc, PhD; Anne Guldhammer Skov, MD; Henning Bundgaard, MD, DMSc*; Morten S. Olesen, MSc, PhD*

BACKGROUND: Results from animal models and observational studies have raised concerns regarding the potential cataractogenic effects of statin treatment. We investigated whether common and rare genetic variants in *HMGCR* are associated with cataract risk, to gauge the likely long-term effects of statin treatment on lenticular opacities.

METHODS AND RESULTS: We used genotyping data and exome sequencing data of unrelated European individuals in the UK Biobank to test the association between genetically proxied inhibition of *HMGCR* and cataract risk. First, we constructed an *HMGCR* genetic score consisting of 5 common variants weighted by their association with low-density lipoprotein cholesterol. Second, we analyzed exome sequencing data to identify carriers of predicted loss-of-function mutations in *HMGCR*. Common and rare variants in aggregate were then tested for association with cataract and cataract surgery. In an analysis of >402 000 individuals, a 38.7 mg/dL (1 mmol/L) reduction in low-density lipoprotein C by the *HMGCR* genetic score was associated with higher risk for cataract (odds ratio, 1.14 [95% CI, 1.00–1.39], P=0.045) and cataract surgery (odds ratio, 1.25 [95% CI, 1.06–1.48], P=0.009). Among 169 172 individuals with *HMGCR* sequencing data, we identified 32 participants (0.02%), who carried a rare *HMGCR* predicted loss-of-function variant. Compared with noncarriers, heterozygous carriers of *HMGCR* predicted loss-of-function had a higher risk of developing cataract (odds ratio, 4.54 [95% CI, 1.96–10.53], P=0.001) and cataract surgery (odds ratio, 5.27 [95% CI, 2.27–12.25], $P=5.37\times10^{-4}$). In exploratory analyses, we found no significant association between genetically proxied inhibition of *PCSK9, NPC1L1*, or circulating low-density lipoprotein cholesterol levels (P>0.05 for all) and cataract risk.

CONCLUSIONS: We found that genetically proxied inhibition of the *HMGCR* gene mimicking long-term statin treatment associated with higher risk of cataract. Clinical trials with longer follow-up are needed to confirm these findings.

Key Words: ADR ■ cataract ■ HMG-CoA reductase ■ HMGCR ■ loss-of-function ■ statins

Cataract is the leading cause of blindness worldwide and >20 million people in the United States are reported to have cataracts.[1] With a growing elderly population, the incidence of cataracts is likely to rise. Therefore, identification of risk factors for developing lens opacities must be prioritized from a public health standpoint.

Statins, which inhibit the 3-hydroxy-3-methylglutaryl coenzyme A (HMG-CoA) reductase, are low-density lipoprotein (LDL) cholesterol–lowering drugs commonly prescribed for the prevention and management of cardiovascular disease.[2] Although the overall safety profile of the marketed statins was shown to be favorable in humans, cataractogenic safety concerns were

Correspondence to: Jonas Ghouse, MD, Laboratory for Molecular Cardiology, Department of Cardiology, Copenhagen University Hospital, Rigshospitalet, Henrik Harpestrengsvej 4C, 2100 Copenhagen, Denmark. Email: jonasghouse@gmail.com

*H. Bundgaard and M. S. Olesen contributed equally.

Supplemental Material for this article is available at https://www.ahajournals.org/doi/suppl/10.1161/JAHA.122.025361

For Sources of Funding and Disclosures, see page 7.

JAHA is available at: www.ahajournals.org/journal/jaha

CLINICAL PERSPECTIVE

What Is New?

- This study suggests that genetically proxied inhibition of HMG-CoA reductase (target of statins) is associated with increased risk of cataract.

What Are the Clinical Implications?

- Since our associations mimic long-term statin-treatment, primarily younger patients with familial hypercholesterolemia, who are subject to almost life-long statin therapy, may warrant closer clinical follow-up.

Nonstandard Abbreviations and Acronyms

HMG-CoA reductase	3-hydroxy-3-methylglutaryl coenzyme A
pLoF	predicted loss-of-function
RCT	randomized control trial
UKB	UK Biobank

mainly spurred by animal studies that showed statin induced subcapsular lens opacities when administered at excessive doses.[3] This association has been subject to extensive evaluation in humans with conflicting evidence from clinical and observational studies on cataract risk. While the Heart Outcomes Prevention Evaluation-3 trial[4] reported a significant increase in cataract surgery in participants randomized to rosuvastatin, other clinical trials did not find an increased risk with statin use.[5–7] Conflicting results have also been reported in observational studies ranging from deleterious, neutral, to protective effects of statin treatment.[8–11] Although informative, the majority of studies that reported on effects of statins in the human lens were limited by, for example, short treatment duration, lack of standardized ophthalmic assessment, or inherent bias associated with observational study designs.

Effects of naturally occurring variation in genes encoding pharmacological targets can be used as proxies of potential long-term risk of adverse drug reactions. Since genetic variants are inherited randomly and fixed at the time of conception, analyses using effects of genetic variants as proxies for intervention targets are in principle independent of confounding and cannot be influenced by reverse causation. In this study, we examined the effect of common and rare genetic variants in the *HMGCR* gene, serving as models for lifelong inhibition of HMG-CoA reductase, on cataract risk using genotyping and exome sequencing data from the UK Biobank (UKB).

METHODS

UKB has received ethical approval from the UK national health service's National Research Ethics Service (ref 11/NW/0382). Individual-level data used to derive these results can be obtained with an approved application to the UK Biobank study. The UKB has been approved by the Northwest Multicenter Research Ethics Committee, UK (Ref: 16/NW/0274). Written informed consent has been obtained from all study participants.

Study Population

The UKB is a large and prospective study of ≈500 000 participants aged 40 to 69 years, recruited between 2006 and 2010.[12] All analyses were conducted under the application number 43247.

Selection of Common Genetic Variants

To proxy HMG-CoA reductase inhibition, we used a previously published 5 single-nucleotide polymorphism *HMGCR* genetic score (Table S1 and S2).[13] In brief, the genetic score was constructed by combining all variants within 100 kb on either side of the gene, which were associated with LDL-C levels at the genome-wide level ($P<5\times10^{-8}$) as reported by the Global Lipids Genetics Consortium[14] and that were in low linkage disequilibrium ($r^2<0.30$) with all other variants included in the score.[15] For each variant, we defined the exposure allele as the allele associated with lower LDL cholesterol levels. For each individual, we calculated a weighted *HMGCR* score by adding the number of LDL cholesterol–lowering alleles weighted by the effect of each variant on LDL cholesterol levels measured in milligrams per deciliter. Details on sample quality control and ancestry definition is provided in Data S1.

Selection of Rare Genetic Variants

To provide complementary evidence to support a causal role of HMG-CoA reductase inhibition on cataract risk, we also extended our analyses to rare variation in *HMGCR*. Cases were defined as persons carrying a rare *HMGCR* predicted loss-of-function variant and controls were ascertained as persons not carrying an *HMGCR* predicted loss-of- function(pLoF) variant (ie, noncarriers). To identify carriers of rare *HMGCR* pLoF variants, we used exome sequenced data for *HMGCR* from unrelated individuals of European ancestry (≈170 000 individuals) provided by the UKB.

Relatedness and ancestry definitions are provided in Data S1 and Figure S1. Detailed information on exome sequencing methodology, alignment, variant calling, and annotation has been described previously.[16] We excluded variants with a genotype quality <20, genotype depth <10, missing genotypes >0.1 and minor allele frequency >0.01. Variants were annotated using SnpEff.[17] Variants in *HMGCR* were identified via positional intersection with Ensembl transcript (ENSG00000113161). We defined pLoF variants as variants leading to loss of a start codon, or loss of a stop codon, or to a premature stop codon; open reading frame shifting indels leading to the formation of a premature stop codon; and variants disrupting splice acceptor or donor sites in the canonical isoform of the gene (Ensembl annotation). We prioritized Sequence Ontology standardized terms (URL: http://www.sequenceontology.org/) with putative high impact. Furthermore, we evaluated splice site region variants with dbscSNV.[18] AdaBoost and Random Forest scores >0.8 were set as splice-altering effects and classified as pLoF variants (Table S3). We also annotated rare missense variants (minor allele frequency <0.01) using the following five *in silico* prediction tools: SIFT, PolyPhen2 HDIV, PolyPhen2 HVAR, LRT, and MutationTaster, as performed previously.[19] A missense variant was considered damaging if each of the 5 prediction tools predicted it to be deleterious (Table S4).

Outcome Definitions

Cataract outcomes were defined using a combination of self-reported data (confirmed by a health care professional) and/or hospital admission diagnosis codes. We defined cataract using the *International Statistical Classification of Diseases* codes for "Senile cataract" (*ICD-9* 366.1; *ICD-10* H25) and "Other cataract" (*ICD-9* 366.0, 366.8 and 366.9; *ICD-10* H26). Although lens opacities are objective changes, they may not associate with visual impairment or functional consequences. Therefore, we also used cataract surgery as a hard end point. Cataract surgery was defined as surgery, using Office of Population Censuses and Surveys Classification of Interventions and Procedures (OPCS) 3 codes for "Discission of cataract and capsulectomy" (170), "Extracapsular extraction of cataract" (173), "Intracapsular extraction of cataract" (174), and OPCS4 codes for "Phacoemulsification of lens" (C71.2) and "Insertion of prosthetic replacement for lens" (C75.1). Additional details on UKB data fields that were used to define study outcomes are provided in Table S5.

Statistical Analysis

For both common and rare variant association analysis, serving as a positive control, we first tested the association with lipids (LDL-C and total cholesterol) using linear regression. We then estimated the odds ratios (OR) with 95% CI for cataract and cataract surgery using logistic regression (with Firth-bias correction for rare variant analysis).[20] For the common variant analysis, ORs were scaled to per 38.7 mg/dL (1 mmol/L) lower LDL-C. All models were adjusted for age at enrollment, sex, assessment center, and genetic ancestry (as quantified by the first 5 principal components).

Sensitivity Analysis

To complement the rare variant association signal from logistic regression, we conducted Fisher's exact test. OR and 95% CIs were obtained as conditional maximum likelihood estimates from Fisher's noncentral hypergeometric distribution. To provide evidence that our *HMGCR* genetic score affects gene expression, we filtered variants that did not associate with mRNA expression levels of *HMGCR* in any of the 49 tissues available in The Genotype-Tissue Expression project The Genotype-Tissue Expression project v8 release (Table S6). To investigate whether the association between *HMGCR* and cataract is specific to the mevalonate synthesis pathway by means of HMG-CoA reductase inhibition rather than general LDL-C lowering, we conducted additional sensitivity analyses. First, we tested whether genetically proxied inhibition of *PCSK9* (target of PCSK9- (proprotein convertase subtilisin/kexin type 9) inhibitors) and *NPC1L1* (target of ezetimibe) yielded similar results, by using both pLoF variants and gene-specific genetic scores with known effects on LDL-C. For selection of pLoF variants, we applied the same approach as in our main analysis (Table S7 and S8). For selection of common variants to be included in the genetic scores, we considered a 7-single-nucleotide polymorphism *PCSK9* (Table S9 and S10) and a 5-single-nucleotide polymorphism *NPC1L1* genetic score (Table S11 and 12) previously published by Ference et al[15] Pharmacologic inhibition of these 2 drug targets lead to an additional 20% to 60% reduction in LDL-C compared with statin therapy alone.[21,22] If the risk-increasing effect of *HMGCR* variants is via LDL-C reduction, one would expect similar-sized association with variants in *NPC1L1* and *PCSK9*. Second, beyond gene-specific associations, we also assessed whether generally lower LDL-C associated with cataract risk. For this we used a previously published genetic instrument for low LDL-C, consisting of 58 independent genetic variants (r^2<0.01) that associated with LDL-C levels at the genome-wide level ($P<5\times10^{-8}$, Table S13).[23] To test whether variants that reside in the last 5% of the resulting protein and are predicted to escape nonsense mediated decay[24] may have influenced our results, we reran the rare variant analysis without such variants. To evaluate nonlinear

effects of age on cataract risk, we modeled age as a cubic spline, with knots introduced at the 25th, 50th, and 75th percentile. Lastly, to evaluate the influence of horizontal pleiotropy on our analyses (ie, that genetic variants influence multiple traits via independent biological pathways), we conducted additional sensitivity analyses in which we adjusted our main models for common cataract risk factors (eg, type 2 diabetes, uveitis, steroid treatment, and smoking).

RESULTS

Association Between Common Genetic Variants and Cataract Risk

Using genotyped data on up to 402 750 unrelated individuals of European ancestry, we found an expected strong association between the *HMGCR* genetic score and circulating LDL-C levels (–1.32 mg/dL per SD increase; 95% CI –1.42, –1.21; $P=6.11\times10^{-131}$; Figure 1). For each 38.7 mg/dL (1 mmol/L) reduction in LDL-C, we found that genetically proxied HMG-CoA reductase inhibition was significantly associated with both cataract (OR, 1.14 [95% CI, 1.00–1.29], $P=0.045$; Figure 1) and cataract surgery (OR, 1.19 [95% CI, 1.04–1.36]; $P=0.009$; Figure 1). These results did not materially change when we restricted our genetic instruments to include variants that affected *HMGCR* mRNA expression only (Table S6). A breakdown on the relationship between cataract and cataract surgery diagnoses is provided in Table S14.

Association Between Rare Genetic Variants and Cataract Risk

Beyond common genetic variants, we also investigated whether a similar pattern of association could be reproduced using exome sequenced data to analyze rare *HMGCR* pLoF variants. In total, we identified 32 individuals with 1 of 17 unique *HMGCR* pLoF variants (Table S3), equating to a carrier frequency of 0.02%. The most frequent variant, p. Leu521Phe (37.5%; 12/32 carriers), is a nonsynonymous splice region variant. As expected, we found that *HMGCR* pLoF carriers had significantly lower LDL-C (–13.1 mg/dL; 95% CI –25.1, –1.2; $P=0.034$) and total cholesterol (–17.4 mg/dL; 95%

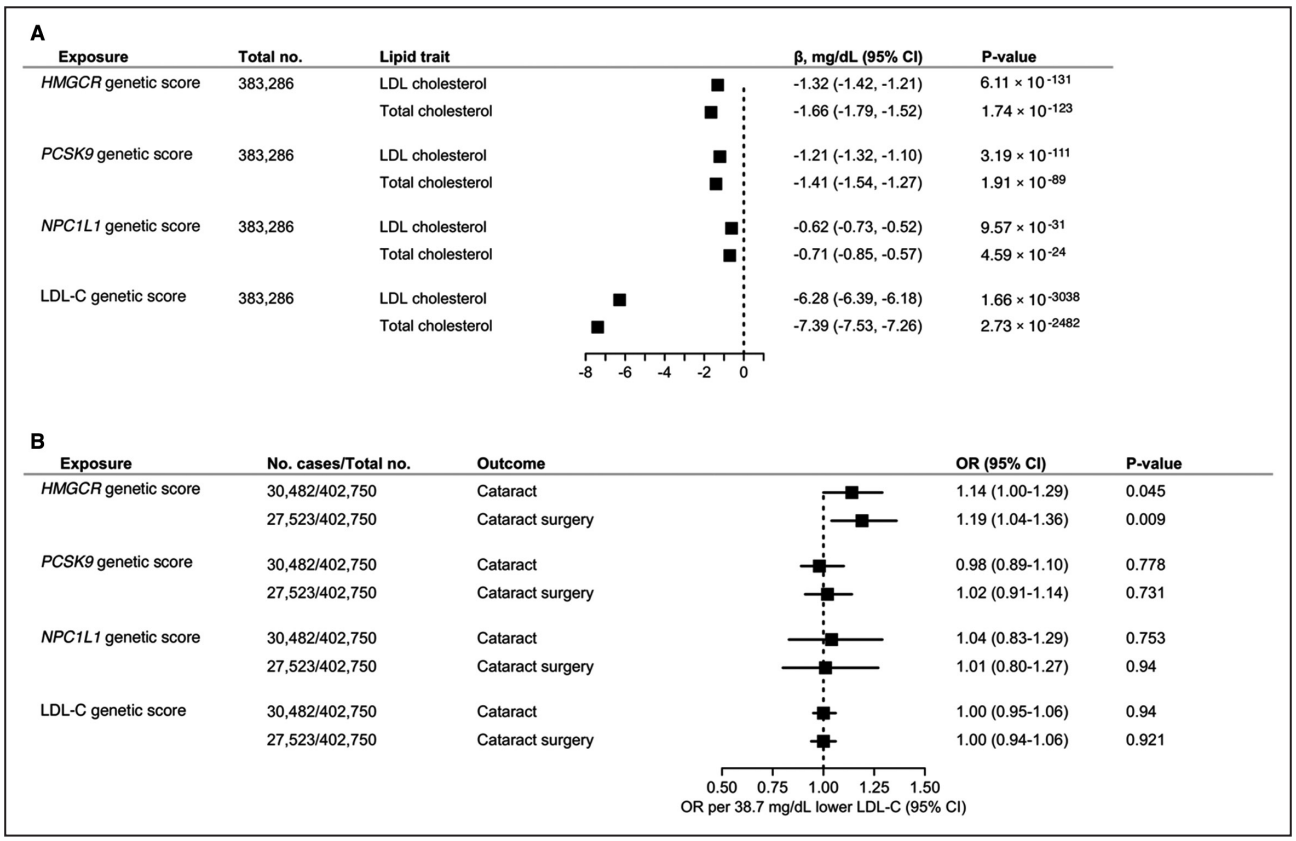


Figure 1. Association between *HMGCR* genetic scores, lipids, and cataract risk.
Shown are the associations between *HMGCR* genetic score and lipids (**A**) and cataract risk (**B**). Results from the sensitivity analysis, showing the association between *NPC1L1, PCSK9*, and LDL-C genetic scores, lipids, and cataract risk are displayed for comparison. ORs and 95% CIs were calculated using a logistic regression model adjusted for age at enrollment, sex, assessment center, and 5 first principal components. ORs were scaled to per 38.7 mg/dL (1 mmol/L) lower LDL-C. LDL-C indicates low-density lipoprotein cholesterol; and OR, odds ratios.

CI −32.9, −1.5; *P*=0.032), compared with noncarriers. Similar to the main results, we found a significant association between *HMGCR* pLoF carrier status and cataract risk. Eight of 32 (25%) *HMGCR* pLoF carriers developed cataract, compared with 12 928 of 169 140 noncarriers (7.6%), equating to an OR of 4.54 (95% CI, 1.96–10.53, *P*=0.001; Figure 2). Moreover, 8 of 32 (25%) carriers underwent cataract surgery, compared with 11 366/169 140 noncarriers (6.7%), resulting in an OR of 5.27 (95% CI, 2.27–12.25, $P=5.37\times10^{-4}$; Figure 2). Comparable effect estimates were obtained when using Fisher's exact test (Table S15) or when adjusting for age as a cubic spline (Table S16). We also found similar results in analyses excluding the p. Leu521Phe variant, indicating that the reported associations are not driven by this variant in isolation (Table S17) or when excluding 2 variants located within 5% of the protein coding sequence (p. Met867fs and p. Thr887fs; Table S18). In a complementary analysis, we also investigated the aggregate effects of rare deleterious variants on cataract risk, including both rare pLoF *and* missense variants predicted to be deleterious. Here, we found concordant direction of effects, however with attenuated effect estimates (cataract: OR, 1.34 [95% CI, 0.97–1.84]; *P*=0.078 and cataract surgery: OR, 1.51 [95% CI, 1.09–2.08]; *P*=0.012).

Exploratory Analyses

To evaluate whether our results were biased through horizontal pleiotropy, we adjusted the main models for risk factors that have previously been linked with cataracts. We found that the results were largely consistent (Table S19). To explore whether the association was specific to the mevalonate synthesis pathway, rather than LDL-C lowering, we conducted additional sensitivity analyses, where we evaluated the effect of genetically proxied inhibition of *NPC1L1* and *PCSK9* on cataract risk. Using exome sequencing data, we found no significant association with *PCSK9* (cataract: OR, 1.09 [95% CI, 0.74–1.60], *P*=0.679; cataract surgery: OR, 1.22 [95% CI, 0.82–1.80], *P*=0.343; Figure 2) or *NPC1L1* (cataract: OR, 1.13 [95% CI, 0.71–1.79], *P*=0.621; cataract surgery: OR, 1.16 [95% CI, 0.71–1.89], *P*=0.551; Figure 2) pLoF carrier status and cataract risk. Nor did we observe any significant relationship between genetically proxied inhibition of *PCSK9* (OR, 0.98 per 38.7 mg/dL lower LDL-C [95% CI, 0.89–1.10], *P*=0.778; cataract surgery: OR, 1.02 [95% CI, 0.91–1.14], *P*=0.731; Figure 1) *NPC1L1* (Cataract: OR 1.04 per 38.7 mg/dL lower LDL-C, 95% CI 0.83–1.29], *P*=0.753; cataract surgery: OR, 1.01 [95% CI, 0.80–1.27], *P*=0.940; Figure 1) nor with genetically proxied lower LDL-C levels (cataract: OR, 1.00 per 38.7 mg/

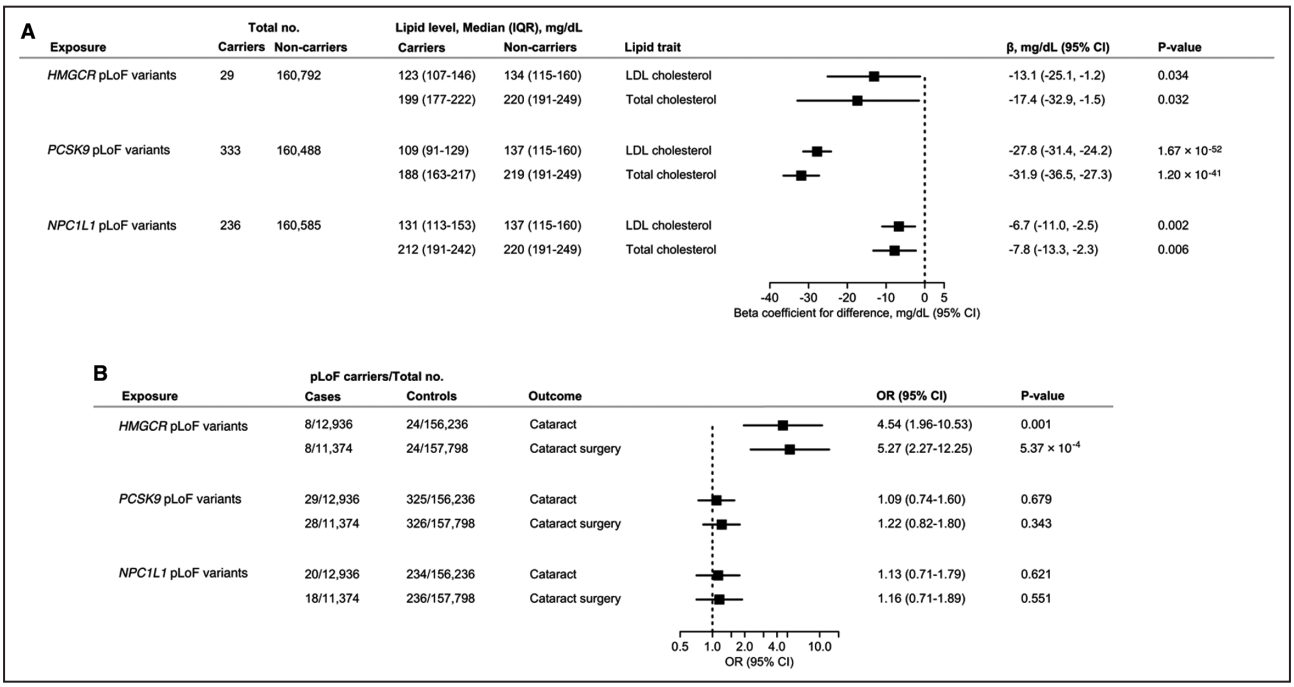


Figure 2. Association between *HMGCR* pLoF carrier status, lipids, and cataract risk.
Shown are the associations between *HMGCR* pLoF carrier status and lipids (**A**) and cataract risk (**B**). Results from the sensitivity analysis, showing the association between *NPC1L1* and *PCSK9* pLoF, lipids, and cataract risk are displayed for comparison. ORs and 95% CIs were calculated using a Firth bias-corrected logistic regression model adjusted for age at enrollment, sex, assessment center, and 5 first principal components. IQR indicates interquartile range; LDL, low-density lipoprotein; ORs, odds ratios; PCSK9, proprotein convertase subtilisin/kexin type 9; and pLoF, predicted loss of function.

dL lower LDL-C [95% CI, 0.95–1.06], *P*=0.940; cataract surgery: OR, 1.00 [95% CI, 0.94–1.06], *P*=0.921; Figure 1), when using array data to mimic pharmacological inhibition (Table S20) or generally lower LDL-C levels.

DISCUSSION

By leveraging both large-scale genotyping and exome sequencing data on >400 000 individuals, we used genetic variants in *HMGCR* to gain insight into the expected effects of long-term therapeutic inhibition of HMG-CoA reductase on cataract risk. We found that lifelong genetic inhibition of *HMGCR* variants was associated with higher risks of both cataract and cataract surgery.

Adverse drug reactions are often detected during early clinical evaluation and more systematically during randomized control trials (RCTs). However, many RCTs are of relatively short duration and may not detect adverse drug reactions that are either rare or only materialize during longer-term treatment. RCTs that have evaluated the potential cataractogenic risk of statins have indeed either been shorter-term studies (ranging from 48 weeks[25] to 18 months[7] of follow-up) or lacked standardized ophthalmic assessment that is needed to detect lenticular opacities.[4,6] Also, patients with poor visual acuity or with a history of cataract before study enrollment were excluded from some RCTs, precluding the possibility of studying cataract progression rather than incident events.[7,25] Our results permit several conclusions. First, since germline genetic variants are fixed at the time of conception, such variants can provide complementary evidence to shorter-term RCTs on the long-term consequences HMG-CoA reductase inhibition. Second, we provide evidence for a dose-dependent relationship between the degree of genetic inhibition and higher risk, as exemplified by the effect estimate gradient (ie, higher relative risk with more "functional" genetic variants). Third, we found stronger associations with cataract surgery compared with a more universal cataract diagnosis, which could reflect higher specificity or a more severe phenotype.

The precise mechanism that leads to cataract in humans is not clear. In animals, no relationship between decrease in cholesterol levels and lenticular opacities was found, but a direct relationship between plasma statin levels and cataract incidence was established.[3] To investigate whether the reported association was specific to the HMG-CoA pathway rather than reflecting a general association of low circulating cholesterol levels, we conducted exploratory analyses of other LDL-C-lowering pathways. We found no association between genetically proxied inhibition of *NPC1L1*, *PCSK9*, and cataract risk, nor did we observe any association with genetically lower LDL-C levels. This indicates that low circulating cholesterol may not be driving the observed association between HMG-CoA reductase inhibition and cataract risk. Instead the association is likely related to the intrinsic role of HMG-CoA reductase in lens sterol synthesis, which is important for membrane formation and transparency.[26]

Strengths of this study include a large population size, with sufficient statistical power to detect associations between genetically proxied inhibition of *HMGCR* and cataract risk; a 2-stage complementary design, including sequencing and genotyping data, which enabled dual assessment of the effect of common *and* rare genetic variants on cataract risk; a parallel investigation of other LDL-lowering pathways (*PCSK9* and *NPC1L1*) on cataract risk; the Mendelian randomization design, in which biases from reverse causation and confounding are reduced; and ability to evaluate both cataract and cataract surgery, where the latter serves as a "harder" end point. Our study needs to be interpreted within the context of its limitations. First, rather than studying the effect of genetic variants on enzymatic activity, we weighted our genetic score by their effect on LDL-C, which is an indirect measure of their function. Second, our results were based on data derived from individuals of European ancestry, limiting the generalizability to non-European populations. Third, the effect estimates that we report reflect life-long exposure to HMG-CoA inhibition, and do not necessarily translate into clinically meaningful risk in high-risk adults with comparably shorter treatment duration. Fourth, although the Genotype-Tissue Expression project v8 release contains mRNA expression data from 49 different tissues, we did not have data derived from the human lens, limiting the ability to explore site-specific functional effects. Fifth, as noted by Backman et al, ≈4% of the coding variants that were identified in the UKB exome sequencing data were highlighted as having "low quality."[19] However, sequencing errors are unlikely to be related to disease status and would likely bias towards the null.

In conclusion, this study found that both common and rare *HMGCR* variants were associated with an increased risk for cataract. The beneficial effects of statins are unequivocal and the reported link with lens opacities should not prevent statin initiation in high-risk adults, but should be disclosed to the patient, especially when indicated for primary prevention. Moreover, younger patients with familial hypercholesterolemia, who are subject to almost life-long statin therapy often initiated in childhood or early adolescence, may warrant closer clinical follow-up.

ARTICLE INFORMATION

Received February 28, 2022; accepted April 14, 2022.

Affiliations

Laboratory for Molecular Cardiology, Department of Cardiology, Copenhagen University Hospital, Rigshospitalet, Copenhagen, Denmark (J.G., G.A., M.S.O.);

Laboratory for Molecular Cardiology, Department of Biomedical Sciences (J.G., G.A., M.S.O.) and Department of Cardiology, Copenhagen University Hospital, Rigshospitalet (H.B.), University of Copenhagen, Denmark; and Department of Ophthalmology, Copenhagen University Hospital, Rigshospitalet-Glostrup, University of Copenhagen, Denmark (A.G.S.).

Acknowledgments

We thank the participants of the UK Biobank study and the Global Lipids Consortium who made their summary statistics publicly available for this study. UK Biobank data were analyzed under application 43247. JG and GA analyzed the data, drew the figures, and wrote the first draft of the manuscript. All authors contributed to the interpretation of the results and critical revision of the manuscript for important intellectual content and approved the final version of the manuscript. JG is guarantor.

Sources of Funding

This work was funded by BRIDGE - Translational Excellence Programme (#NNF18SA0034956 and #NNF20SA0064340), The John and Birthe Meyer Foundation, The Innovation Fund Denmark (PM Heart), NordForsk, and The Hallas-Møller Emerging Investigator. No conflicts of interest relevant to this article were reported. The funding sources had no role in the design or conduct of the study; collection, management, analysis, and interpretation of the data; or preparation, review, or approval of the manuscript.

Disclosures

HB receives lecture fees from Bristol-Myers Squibb, and Merck Sharp and Dohme. The remaining authors have no disclosures to report.

Supplemental Material

Data S1
Tables S1–S20
Figure S1
References[23, 27]

REFERENCES

1. Ackland P, Resnikoff S, Bourne R. World blindness and visual impairment: despite many successes, the problem is growing. *Community Eye Health*. 2017;30:71–73.
2. Unit ES. Efficacy and safety of cholesterol-lowering treatment. prospective meta-analysis of data from 90 056 participants in 14 randomised trials of statins. *Lancet*. 2005;366:1267–1278. doi: 10.1016/S0140-6736(05)67394-1
3. Gerson RJ, Macdonald JS, Alberts AW, Kornbrust DJ, Majka JA, Stubbs RJ, Bokelman DL. Animal safety and toxicology of simvastatin and related hydroxy-methylglutaryl-coenzyme a reductase inhibitors. *Am J Med*. 1989;87:S28–S38. doi: 10.1016/S0002-9343(89)80596-0
4. Yusuf S, Bosch J, Dagenais G, Zhu J, Xavier D, Liu L, Pais P, López-Jaramillo P, Leiter LA, Dans A, et al. Cholesterol lowering in intermediate-risk persons without cardiovascular disease. *N Engl J Med*. 2016;374:2021–2031. doi: 10.1056/NEJMoa1600176
5. Laties AM, Keates EU, Taylor HR, Chremos AN, Shear CL, Lippa EA, Gould AL, Hurley DP. The human lens after 48 weeks of treatment with lovastatin. *N Engl J Med*. 1990;323:683–684. doi: 10.1056/NEJM199009063231015
6. Pedersen TR, Berg K, Cook TJ, Faergeman O, Haghfelt T, Kjekshus J, Miettinen T, Musliner TA, Olsson AG, Pyörälä K, et al. Safety and tolerability of cholesterol lowering with simvastatin during 5 years in the Scandinavian Simvastatin Survival Study. *Arch Intern Med*. 1996;156:2085–2092.
7. Harris ML, Bron AJ, Brown NA, Keech AC, Wallendszus KR, Armitage JM, MacMahon S, Snibson G, Collins R. Absence of effect of simvastatin on the progression of lens opacities in a randomised placebo controlled study. Oxford Cholesterol Study Group. *Br J Ophthalmol*. 1995;79:996–1002. doi: 10.1136/bjo.79.11.996
8. Leuschen J, Mortensen EM, Frei CR, Mansi EA, Panday V, Mansi I. Association of statin use with cataracts: a propensity score-matched analysis. *JAMA Ophthalmol*. 2013;131:1427–1434. doi: 10.1001/jamaophthalmol.2013.4575
9. Wise SJ, Nathoo NA, Etminan M, Mikelberg FS, Mancini GBJ. Statin use and risk for cataract: a nested case-control study of 2 populations in Canada and the United States. *Can J Cardiol*. 2014;30:1613–1619. doi: 10.1016/j.cjca.2014.08.020
10. Tan JSL, Mitchell P, Rochtchina E, Wang JJ. Statin use and the long-term risk of incident cataract: the Blue Mountains Eye Study. *Am J Ophthalmol*. 2007;143:687–689. doi: 10.1016/j.ajo.2006.11.027
11. Hippisley-Cox J, Coupland C. Unintended effects of statins in men and women in England and Wales: population based cohort study using the QResearch database. *BMJ*. 2010;340:c2197. doi: 10.1136/bmj.c2197
12. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, Downey P, Elliott P, Green J, Landray M, 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:e1001779. doi: 10.1371/journal.pmed.1001779
13. Ference BA, Robinson JG, Brook RD, Catapano AL, Chapman MJ, Neff DR, Voros S, Giugliano RP, Davey Smith G, Fazio S, et al. Variation in PCSK9 and HMGCR and risk of cardiovascular disease and diabetes. *N Engl J Med*. 2016;375:2144–2153. doi: 10.1056/NEJMoa1604304
14. Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, Ganna A, Chen J, Buchkovich ML, Mora S, et al. Discovery and refinement of loci associated with lipid levels. *Nat Genet*. 2013;45:1274–1283. doi: 10.1038/ng.2797
15. Ference BA, Ray KK, Catapano AL, Ference TB, Burgess S, Neff DR, Oliver-Williams C, Wood AM, Butterworth AS, Di Angelantonio E, et al. Mendelian randomization study of ACLY and cardiovascular disease. *N Engl J Med*. 2019;380:1033–1042. doi: 10.1056/NEJMoa1806747
16. Szustakowski JD, Balasubramanian S, Kvikstad E, Khalid S, Bronson PG, Sasson A, Wong E, Liu D, Wade Davis J, Haefliger C, et al. Advancing human genetics research and drug discovery through exome sequencing of the UK Biobank. *Nat Genet*. 2021;53:942–948. doi: 10.1038/s41588-021-00885-0
17. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin)*. 2012;6:80–92. doi: 10.4161/fly.19695
18. Jian X, Boerwinkle E, Liu X. In silico prediction of splice-altering single nucleotide variants in the human genome. *Nucleic Acids Res*. 2014;42:13534–13544. doi: 10.1093/nar/gku1206
19. Backman JD, Li AH, Marcketta A, Sun D, Mbatchou J, Kessler MD, Benner C, Liu D, Locke AE, Balasubramanian S, et al. Exome sequencing and analysis of 454,787 UK Biobank participants. *Nature*. 2021;599:628–634. doi: 10.1038/s41586-021-04103-z
20. Clayton D, Hills M. *Statistical Models in Epidemiology*. Oxford University Press; 2013.
21. Sabatine MS, Giugliano RP, Wiviott SD, Raal FJ, Blom DJ, Robinson J, Ballantyne CM, Somaratne R, Legg J, Wasserman SM, et al. Efficacy and safety of evolocumab in reducing lipids and cardiovascular events. *N Engl J Med*. 2015;372:1500–1509. doi: 10.1056/NEJMoa1500858
22. Cannon CP, Blazing MA, Giugliano RP, McCagg A, White JA, Theroux P, Darius H, Lewis BS, Ophuis TO, Jukema JW, et al. Ezetimibe added to statin therapy after acute coronary syndromes. *N Engl J Med*. 2015;372:2387–2397. doi: 10.1056/NEJMoa1410489
23. Lotta LA, Stewart ID, Sharp SJ, Day FR, Burgess S, Luan J, Bowker N, Cai L, Li C, Wittemans LBL, et al. Association of genetically enhanced lipoprotein lipase-mediated lipolysis and low-density lipoprotein cholesterol-lowering alleles with risk of coronary disease and type 2 diabetes. *JAMA Cardiol*. 2018;3:957–966. doi: 10.1001/jamacardio.2018.2866
24. Van Hout CV, Tachmazidou I, Backman JD, Hoffman JD, Liu D, Pandey AK, Gonzaga-Jauregui C, Khalid S, Ye B, Banerjee N, et al. Exome sequencing and characterization of 49,960 individuals in the UK Biobank. *Nature*. 2020;586:749–756. doi: 10.1038/s41586-020-2853-0
25. Laties AM, Shear CL, Lippa EA, Gould AL, Taylor HR, Hurley DP, Stephenson WP, Keates EU, Tupy-Visich MA, Chremos AN. Expanded clinical evaluation of lovastatin (EXCEL) study results II. Assessment of the human lens after 48 weeks of treatment with lovastatin. *Am J Cardiol*. 1991;67:447–453. doi: 10.1016/0002-9149(91)90002-3
26. Cenedella RJ. Cholesterol and cataracts. *Surv Ophthalmol*. 1996;40:320–337. doi: 10.1016/s0039-6257(96)82007-8
27. Bellenguez C, Strange A, Freeman C, Donnelly P, Spencer CCA. A robust clustering algorithm for identifying problematic samples in genome-wide association studies. *Bioinformatics*. 2012;28:134–135. doi: 10.1093/bioinformatics/btr599

SUPPLEMENTAL MATERIAL

Data S1. Sample quality control and ancestry definition for exome sequencing and genotyping data.

In the present analysis, samples that were related, outliers for heterozygosity, missingness or excess relativeness were removed. Samples were further excluded if they were not used in the central kinship inference. We used the ukb_gen_samples_to_remove() function from the R package *ukbtools*/v0.11 to choose a subset of individuals within which no pair had a kinship coefficient exceeding 0.0884, equivalent of up to third-degree relatives. For each related pair, this function removes whichever member has the highest number or relatives above the provided threshold, resulting in a maximal set. In addition, individuals with putative sex chromosome aneuploidy or with a mismatch between self-reported and genetically inferred sex were excluded. In order to identify a subset of individuals with European ancestry, we clustered individuals in UKB that had self-reported as "White" (UKB data field 21000; data codes: 1. 1001. 1002 and 1003). We then applied the Bayesian outlier detection algorithm implemented in the R package *aberrant* with principal components 1-2. 3-4 and 5-6, respectively.[26] The intersection of these clustered sets defined "individuals with European ancestry" (**Figure S1**).

Table S1. *HMGCR* variants included in *HMGCR* genetic score and their association with LDL-C in the Global Lipids Genetics Consortium.

SNP	Exposure Allele	Exposure Allele Frequency	Effect Size (mg/dl)	Standard error	*P*
rs12916	T	0.569	-2.3456	0.1216	7.79E-78
rs5909	G	0.898	-1.9744	0.2816	4.93E-13
rs2303152	G	0.880	-1.3536	0.2048	1.04E-09
rs10066707	G	0.583	-1.5904	0.1728	2.97E-19
rs2006760	C	0.814	-1.7056	0.2432	1.67E-13

For each variant, the exposure allele is the allele associated with lower LDL-C in Global Lipids Genetics Consortium.

Table S2. Linkage disequilibrium matrix for common variants included in the *HMGCR* genetic score.

	rs12916	rs5909	rs2303152	rs10066707	rs2006760
rs12916	1	0.16	0.18	0.27	0.18
rs5909	0.16	1	0.01	0.06	0.21
rs2303152	0.18	0.01	1	0.09	0.27
rs10066707	0.27	0.06	0.09	1	0.26
rs2006760	0.18	0.21	0.27	0.26	1

All measures of LD are in R^2 and were derived from the LDlink software using five European ancestry populations from phase 3 of the 1000 genomes project

Table S3. Loss-of-function variants in *HMGCR*.

Chromosome	Position	Reference allele	Alternate allele	Minor allele frequency (European)	Mutation	Variant type	Carriers
5	75342606	A	G	2.85E-06	p.Met1?	Start lost	1
5	75342613	C	G	2.85E-06	p.Ser3*	Stop gained	1
5	75342687	AC	A	2.86E-06	p.Ile29fs	Frameshift variant	1
5	75343892	C	T	5.78E-06	p.Arg69*	Stop gained	2
5	75347307	C	A	2.94E-06	p.Ser185*	Stop gained and splice region variant	1
5	75350312	C	T	8.56E-06	p.Arg240*	Stop gained	3
5	75350373	T	C	3.16E-06	p.Met260Thr	Splice region variant	1
5	75351159	G	GA	2.85E-06	p.Ser346fs	Frameshift variant	1
5	75351167	ACT	A	2.86E-06	p.Ser349fs	Frameshift variant	1
5	75354620	C	T	2.85E-06	p.Arg496*	Stop gained	1
5	75354697	G	T	3.57E-05	p.Leu521Phe	Splice region variant	12
5	75356414	CAG	C	2.89E-06	p.Asp653fs	Frameshift variant	1
5	75359168	A	G	2.86E-06	c.2158-2A>G	Splice acceptor variant and intron variant	1
5	75359427	CTG	C	2.85E-06	p.Cys777fs	Frameshift variant	1
5	75360045	C	T	2.85E-06	p.Arg840*	Stop gained	1
5	75360125	CAT	C	5.70E-06	p.Met867fs	Frameshift variant	2
5	75360331	AAGAC	A	2.85E-06	p.Thr887fs	Frameshift variant	1
Total							**32**

Table S4. Rare predicted deleterious missense variants in *HMGCR*.

Chromosome	Position	Reference allele	Alternate allele	Minor allele frequency (European)	Mutation	Carriers
5	75342646	C	G	8.56E-06	p.Ala14Gly	3
5	75342736	G	A	2.85E-06	p.Cys44Tyr	1
5	75343893	G	A	5.78E-06	p.Arg69Gln	2
5	75343911	A	G	2.88E-06	p.Tyr75Cys	1
5	75343937	C	T	5.94E-06	p.Arg84Cys	2
5	75344287	G	A	5.83E-06	p.Ser107Asn	2
5	75345605	G	A	2.86E-06	p.Asp133Asn	1
5	75345639	C	T	2.89E-06	p.Ala144Val	1
5	75345641	C	T	9.03E-05	p.Leu145Phe	28
5	75347228	C	T	2.58E-05	p.Arg159Cys	8
5	75347253	C	T	5.71E-06	p.Thr167Met	1
5	75347268	C	T	1.14E-05	p.Ala172Val	4
5	75347289	T	C	2.87E-06	p.Ile179Thr	1
5	75350052	G	C	1.43E-05	p.Gln189His	5
5	75350096	A	G	9.21E-04	p.Asn204Ser	290
5	75350132	T	G	2.86E-06	p.Val216Gly	1
5	75350264	C	T	1.71E-05	p.Arg224Trp	6
5	75350273	C	T	2.85E-06	p.Arg227Cys	1
5	75350282	C	T	8.56E-06	p.Arg230Cys	3
5	75350356	G	T	5.77E-06	p.Gln254His	2
5	75350365	G	T	2.98E-06	p.Lys257Asn	1

5	75350822	C	T	2.85E-06	p.Arg272Cys	1
5	75351153	G	A	2.85E-06	p.Glu343Lys	1
5	75354603	G	A	5.71E-06	p.Arg490His	2
5	75354617	C	T	2.85E-06	p.Arg495Cys	1
5	75354621	G	A	2.85E-06	p.Arg496Gln	1
5	75354630	T	A	2.85E-06	p.Leu499His	1
5	75355069	G	C	1.71E-05	p.Cys526Ser	4
5	75355096	C	G	2.85E-06	p.Pro535Arg	1
5	75355155	A	G	8.56E-06	p.Met555Val	3
5	75355185	A	G	2.85E-06	p.Ser565Gly	1
5	75355393	T	C	2.85E-06	p.Leu531Pro	1
5	75355410	C	T	2.85E-06	p.Arg590Cys	1
5	75355411	G	A	5.70E-06	p.Arg590His	2
5	75355425	C	T	2.85E-06	p.Arg595Cys	1
5	75355426	G	A	2.85E-06	p.Arg595His	1
5	75356378	C	T	5.76E-06	p.Ala639Val	2
5	75356383	C	T	2.87E-06	p.Arg641Cys	2
5	75358786	G	C	2.85E-06	p.Val706Leu	1
5	75358814	A	T	2.85E-06	p.Lys715Met	1
5	75359237	T	C	2.85E-06	p.Met742Thr	1
5	75359239	G	A	5.71E-06	p.Ala743Thr	1
5	75359297	T	C	2.85E-06	p.Ile762Thr	1
5	75359474	A	G	2.85E-06	p.Tyr792Cys	1
5	75359524	A	C	2.85E-06	p.Thr809Pro	1
5	75359525	C	A	2.85E-06	p.Thr809Asn	1
5	75360001	G	T	2.85E-06	p.Gly825Val	1

5	75360033	C	T	8.56E-06	p.Arg836Trp	3
5	75360043	C	T	2.85E-06	p.Ala839Val	1
5	75360063	G	A	1.14E-05	p.Val846Ile	4
5	75360085	T	C	2.85E-06	p.Leu853Pro	1
5	75360088	T	C	2.85E-06	p.Met854Thr	1
5	75360094	C	G	2.85E-06	p.Ala856Gly	1
5	75360126	A	G	2.85E-06	p.Met867Val	1
Total						**411**

Missense variants with a minor allele frequency <1 % and predicted to be deleterious by each of the five *in silico* prediction tools (i.e. SIFT, PolyPhen2 HDIV, PolyPhen2 HVAR, LRT and MutationTaster).

Table S5. Disease phenotype definitions

Disease phenotype	Definition including UK Biobank data-fields
Lipid traits	
LDL cholesterol (mmol/L)	30780
Total cholesterol (mmol/L)	30690
Binary outcomes	
Cataract	Self-reported history of cataract during verbal interview with trained nurse; **or** hospitalization due to ICD-10 code for cataract (H25, H26); **or** hospitalization due to ICD-9 code for cataract (366.0, 366.1, 366.8, 366.9).
Cataract surgery	Self-reported history of cataract surgery during verbal interview with trained nurse; **or** hospitalization due to OPCS-4 code for cataract surgery (C751, C712); **or** hospitalization due to OPCS-3 code for cataract surgery (170, 172, 173, 174).

Self-reported data was extracted using UK Biobank (UKB) data-field 20002; ICD-9 codes were extracted using UKB data-field 41271 and 41281; ICD-10 codes were extracted using UKB data-fields 41270 and 41280; OPCS-3 data were extracted using UKB data-field 41273 and 41283; OPCS-4 data were extracted using UKB data-field 41272 and 41282.

Table S6. Genetic variants included in the *HMGCR* score that associates with lower *HMGCR* mRNA expression and association results for genetic score using eQTL variants.

					HMGCR genetic score using eQTL variants	
SNP	Exposure allele	GTEx tissue	Normalized effect size	P-value	OR (95% CI) for cataract	OR (95% CI) for cataract surgery
rs12916	T	Skeletal muscle	-0.17	7.9E-09	1.21 (0.99-1.47)	1.30 (1.06-1.60)
rs10066707	G	Skeletal muscle	-0.12	3.0E-05		

Odds ratio (OR) are reported per 38.7 mg/dL (1 mmol/L) lower LDL-C.

Table S7. Loss-of-function variants in *PCSK9*

Chromosome	Position	Reference allele	Alternate allele	Mutation	Variant type	Carriers
1	55039838	A	G	p.Met1?	Start lost	2
1	55039902	T	TG	p.Leu23fs	Frameshift variant	2
1	55039925	G	GCGCA	p.Gln31fs	Frameshift variant	20
1	55039951	C	A	p.Tyr38Ter	Stop gained	1
1	55040006	G	T	p.Glu57Ter	Stop gained	3
1	55040024	AC	A	p.Phe64fs	Frameshift variant	3
1	55040045	G	A	c.207+1G>A	Splice donor variant	1
1	55043851	G	A	p.Trp72Ter	Stop gained	1
1	55043903	C	T	p.Gln90Ter	Stop gained	1
1	55044012	T	A	p.Met1?	Start lost	9
1	55046570	CT	C	p.Phe150fs	Frameshift variant	1
1	55046570	CTT	C	p.Phe150fs	Frameshift variant	1
1	55046646	G	A	p.Asp175Asn	Splice region variant	3
1	55052273	GA	G	c.524-2delA	Splice acceptor variant	1
1	55052374	ATG	A	p.Val208fs	Frameshift variant	1
1	55052412	G	T	c.657+1G>T	Splice donor variant	38
1	55052648	A	C	c.658-2A>C	Splice acceptor variant	1
1	55052661	T	A	p.Cys223Ter	Stop gained	1
1	55052704	G	GA	p.Asp238fs	Frameshift variant	2
1	55055992	G	A	c.800-1G>A	Splice acceptor variant	1
1	55056006	TC	T	p.Arg272fs	Frameshift variant	2
1	55056009	GA	G	p.Ser274fs	Frameshift variant	6
1	55057364	C	T	p.Gln344Ter	Stop gained	4
1	55057478	C	T	p.Gln382Ter	Stop gained	8

1	55057514	G	A	p.Gly394Ser	Splice region variant	33
1	55058036	G	T	p.Gly394Val	Splice region variant	9
1	55058113	GC	G	p.Asp422fs	Frameshift variant	11
1	55058139	G	A	p.Trp428Ter	Stop gained	1
1	55058597	A	AG	p.Ser485fs	Frameshift variant	1
1	55058649	T	C	c.1503+2T>C	Splice donor variant	1
1	55059491	A	AG	p.Lys506fs	Frameshift variant	1
1	55059573	C	T	p.Gln531Ter	Stop gained	1
1	55061368	G	GC	c.1682-3dup	Splice acceptor variant	19
1	55061374	G	C	c.1682-1G>C	Splice acceptor variant	3
1	55061437	C	T	p.Arg582Ter	Splice acceptor variant	2
1	55061473	GC	G	p.Ser470fs	Frameshift variant	2
1	55061517	CA	C	p.Lys609fs	Frameshift variant	1
1	55061557	G	A	c.1863+1G>A	Splice donor variant	91
1	55063542	C	A	p.Cys679Ter	Stop gained	61
1	55063570	C	T	p.Gln689Ter	Stop gained	3
1	55063582	T	G	p.Ter693Gly?	Stop lost	1
Total						**354**

Table S8. Loss-of-function variants in *NPC1L1*

Chromosome	Position	Reference allele	Alternate allele	Mutation	Variant type	Carriers
7	44513554	G	A	p.Arg1298*	stop_gained	2
7	44515801	AC	A	c.3796+1delG	stop_gained	2
7	44516085	G	A	p.Ala1211Val	splice_region_variant	6
7	44516110	CTT	C	p.Lys1202fs	frameshift_variant	1
7	44516865	G	T	p.Cys1119*	stop_gained	2
7	44516907	G	T	p.Tyr1105*	stop_gained	2
7	44516909	AC	A	p.Gln1104fs	frameshift_variant	1
7	44517206	C	T	c.3287+1G>A	splice_donor_variant	4
7	44517207	G	A	p.Thr1096Met	splice_region_variant	3
7	44517286	G	A	p.Arg1070*	stop_gained	4
7	44520822	T	C	c.3081-2A>G	splice_acceptor_variant	4
7	44521030	C	T	p.Trp1014*	stop_gained	1
7	44521053	G	A	p.Gln1007*	stop_gained	1
7	44522076	TG	T	p.Gln935fs	frameshift_variant	1
7	44522242	C	T	p.Asp880Asn	splice_region_variant	1
7	44531753	A	C	c.2637+2T>G	splice_donor_variant	71
7	44531754	C	T	c.2637+1G>A	splice_donor_variant	1
7	44531781	G	GT	p.Leu871fs	frameshift_variant	1
7	44532079	C	T	c.2547+1G>A	splice_donor_variant	1
7	44532091	G	A	p.Arg846*	stop_gained	7
7	44532173	C	A	p.Pro819fs	frameshift_variant	1
7	44532188	GAC	A	p.Val813fs	frameshift_variant	1
7	44532218	C	G	c.2410-1G>C	splice_acceptor_variant	3

7	44532219	T	C	c.2410-2A>G	splice_acceptor_variant	2
7	44533433	C	A	p.Glu803*	stop_gained	4
7	44533519	AG	A	p.Leu774fs	frameshift_variant	2
7	44533550	T	C	p.Ter725Trp?	stop_lost	1
7	44533808	G	A	p.Arg738*	stop_gained	3
7	44533826	G	A	p.Arg732*	stop_gained	3
7	44533851	CCT	C	p.Arg723fs	frameshift_variant	7
7	44533854	CTGTG	C	c.2167-5_2167-2delCACA	splice_acceptor_variant	1
7	44534450	G	C	p.Tyr721*	stop_gained	1
7	44534595	AC	A	p.Val673fs	frameshift_variant	1
7	44534630	C	A	c.1984-1G>T	splice_acceptor_variant	2
7	44535848	G	A	p.Arg659*	stop_gained	7
7	44535932	C	A	p.Glu631*	stop_gained	1
7	44536334	C	T	p.Trp592*	stop_gained	1
7	44536394	CA	C	p.Met572fs	frameshift_variant	1
7	44536841	C	T	c.1681+1G>A	splice_donor_variant	7
7	44536842	C	T	p.Gly561Arg	splice_region_variant	6
7	44538823	CA	C	p.Cys525fs	frameshift_variant	1
7	44538854	G	A	p.Gln515*	stop_gained	3
7	44538875	TCTGGTTGG	T	p.Ala505fs	frameshift_variant	2
7	44538973	A	AG	p.Leu475fs	frameshift_variant	5
7	44539112	G	GC	p.Pro429fs	frameshift_variant	1
7	44539181	G	A	p.Arg406*	stop_gained	54
7	44539509	AG	A	p.Ala296fs	frameshift_variant	1
7	44539658	G	A	p.Gln247*	stop_gained	1
7	44539675	GCAAC	G	p.Val240fs	frameshift_variant	5
7	44539896	CT	C	p.Gln167fs	frameshift_variant	1
7	44539898	G	A	p.Gln167*	stop_gained	5

7	44539918	TAGAA	T	p.Phe159fs	frameshift_variant	1
7	44540099	C	A	p.Glu100*	stop_gained	1
7	44540135	G	A	p.Gln88*	stop_gained	1
7	44540219	TGG	T	p.Ser59fs	frameshift_variant	1
Total						**254**

Table S9. *PCSK9* variants included in *PCSK9* genetic score and their association with LDL-C in the Global Lipids Genetics Consortium.

SNP	Exposure Allele	Exposure Allele Frequency	Effect Size (mg/dl)	Standard error	*P*
rs11206510	C	0.154	-2.6592	0.005	2.38E-53
rs2479409	A	0.668	-2.0544	0.0041	2.52E-50
rs2149041	C	0.839	-2.0352	0.0049	1.44E-35
rs2479394	A	0.715	-1.2352	0.0041	1.58E-19
rs10888897	T	0.395	-1.6224	0.0042	8.43E-31
rs7552841	C	0.635	-1.1776	0.0044	5.40E-15
rs562556	G	0.194	-2.048	0.0066	6.16E-21

For each variant, the exposure allele is the allele associated with lower LDL-C in Global Lipids Genetics Consortium.

Table S10. *NPC1L1* variants included in *NPC1L1* genetic score and their association with LDL-C in the Global Lipids Genetics Consortium.

SNP	Exposure Allele	Exposure Allele Frequency	Effect Size (mg/dl)	Standard error	*P*
rs217386	A	0.408	-1.1253	0.118	1.20E-19
rs2073547	A	0.806	-1.5035	0.152	1.92E-21
rs7791240	T	0.909	-1.3175	0.202	1.84E-10
rs10234070	C	0.904	-0.9145	0.183	1.52E-06
rs2300414	G	0.930	-1.0943	0.248	5.45E-06

For each variant, the exposure allele is the allele associated with lower LDL-C in Global Lipids Genetics Consortium.

Table S11. Linkage disequilibrium matrix for common variants included in the *PCSK9* genetic score

	rs11206510	rs2479409	rs2149041	rs2479394	rs10888897	rs7552841	rs562556
rs11206510	1	0.05	0.04	0.07	0.07	0.02	0.02
rs2479409	0.05	1	0.14	0.04	0.11	0.00	0.02
rs2149041	0.04	0.14	1	0.07	0.10	0.01	0.02
rs2479394	0.07	0.07	0.07	1	0.01	0.02	0.00
rs10888897	0.01	0.07	0.10	0.01	1	0.00	0.07
rs7552841	0.02	0.00	0.01	0.02	0.00	1	0.06
rs562556	0.02	0.02	0.02	0.00	0.07	0.06	1

All measures of LD are in R^2 and were derived from the LDlink software using five European ancestry populations from phase 3 of the 1000 genomes project

Table S12. Linkage disequilibrium matrix for common variants included in the *NPC1L1* genetic score

	rs217386	rs2073547	rs7791240	rs10234070	rs2300414
rs217386	1.00	0.13	0.06	0.03	0.04
rs2073547	0.13	1.00	0.27	0.17	0.07
rs7791240	0.06	0.27	1.00	0.02	0.31
rs10234070	0.03	0.17	0.02	1.00	0.02
rs2300414	0.04	0.07	0.31	0.02	1.00

All measures of LD are in R^2 and were derived from the LDlink software using five European ancestry populations from phase 3 of the 1000 genomes project

Table S13. LDL-C lowering variants included in LDL-C genetic score.

SNP	Chromosome	Position	Exposure allele	Effect size (mg/dL)	SE
rs2479409	1	55504650	A	-2.0544	0.131
rs629301	1	109818306	G	-5.3408	0.157
rs12027135	1	25775733	A	-0.96	0.122
rs2642442	1	220973563	C	-1.152	0.173
rs514230	1	234858597	A	-1.1648	0.173
rs2131925	1	63025942	G	-1.5648	0.125
rs12748152	1	27138393	C	-1.5968	0.211
rs267733	1	150958836	G	-1.0592	0.170
rs1367117	2	21263900	G	-3.7952	0.128
rs4299376	2	44072576	T	-2.5984	0.144
rs2710642	2	63149557	G	-0.7648	0.122
rs10490626	2	118835841	A	-1.6256	0.221
rs2030746	2	121309488	C	-0.6848	0.122
rs1250229	2	216304384	T	-0.7776	0.134
rs11563251	2	234679384	C	-1.104	0.198
rs7640978	3	32533010	T	-1.2544	0.221
rs17404153	3	132163200	T	-1.0752	0.173
rs6831256	4	3473139	A	-0.6016	0.122
rs12916	5	74656539	T	-2.3456	0.122
rs6882076	5	156390297	T	-1.4592	0.122
rs4530754	5	122855416	G	-0.88	0.115
rs3757354	6	16127407	T	-1.2224	0.141
rs1800562	6	26093141	A	-1.968	0.256
rs1564348	6	160578860	T	-1.5392	0.160
rs3177928	6	32412435	G	-1.4464	0.166
rs9488822	6	116312893	T	-0.9952	0.173

rs12670798	7	21607352	T	-1.1008	0.138
rs2072183	7	44579180	G	-1.2352	0.150
rs4722551	7	25991826	T	-1.2512	0.157
rs9987289	8	9183358	A	-2.2848	0.211
rs11136341	8	145043543	A	-1.4304	0.198
rs2081687	8	59388565	C	-0.9952	0.173
rs2954029	8	126490972	T	-1.8048	0.115
rs10102164	8	55421614	G	-1.0112	0.144
rs635634	9	136155000	C	-2.4704	0.176
rs3780181	9	2640759	G	-1.424	0.237
rs2255141	10	113933886	G	-0.9568	0.128
rs11220462	11	126243952	G	-1.888	0.189
rs174546	11	61569830	T	-1.6384	0.122
rs964184	11	116648917	C	-2.736	0.250
rs11065987	12	112072424	G	-0.8608	0.122
rs1169288	12	121416650	A	-1.2	0.128
rs4942486	13	32953388	C	-0.7776	0.118
rs8017377	14	24883887	G	-0.9696	0.122
rs3764261	16	56993324	A	-1.6896	0.134
rs2000999	16	72108093	G	-2.08	0.147
rs7206971	17	45425115	G	-0.9344	0.176
rs1801689	17	64210580	A	-3.2896	0.445
rs314253	17	7091650	C	-0.7744	0.122
rs6511720	19	11202306	T	-7.0688	0.195
rs4420638	19	45422946	A	-7.2032	0.246
rs10401969	19	19407718	C	-3.7888	0.230
rs6029526	20	39672618	T	-1.3952	0.166
rs2902940	20	39091487	G	-0.8768	0.131

rs364585	20	12962718	A	-0.7968	0.122
rs2328223	20	17845921	A	-0.9568	0.160
rs5763662	22	30378703	C	-2.4544	0.387
rs4253772	22	46627603	C	-1.0016	0.192

Variants selected for the LDL-C genetic score were obtained from Lotta *et al*, 2018[28] (PMID: 30326043)

Table S14. Confusion matrix on the relationship between cataract and cataract surgery for unrelated Europeans sampled from the UK Biobank

		Ever cataract surgery	
		No	**Yes**
Ever cataract	**No**	371051	1217
	Yes	4176	26306

Table S15. Association results between *HMGCR* loss-of-function variants and cataract risk using Fisher's exact test

	pLoF carriers/Total no.					
	Cases	Controls	Outcome	OR	95% CI	P-value
HMGCR pLoF	8/12,936	24/156,236	Cataract	4.03	1.56-9.27	0.002
	8/11,374	24/157,798	Cataract surgery	4.63	1.80-10.65	0.001

Table S16. Evaluation of non-linear effects of age using cubic spline regression.

Model	Outcome	OR (95% CI)	*P*
Genetic risk score analysis	Cataract	1.14 (1.00-1.29)	0.044
	Cataract surgery	1.19 (1.05-1.36)	0.009
pLoF variant analysis	Cataract	4.58 (2.00-10.47)	1.32E-03
	Cataract surgery	5.34 (2.33-12.22)	4.93E-04

Associations between genetically proxied inhibition of *HMGCR* and cataract risk was modeled using logistic regression, adjusted for age (modeled as a cubic spline), sex, assessment center and five principal components. Knots for the cubic spline were set at the 25^{th}, 50^{th} and the 75^{th} percentile for age.

Table S17. Rare variant association results without p.Leu521Phe

	pLoF carriers/Total no.					
	Cases	Controls	Outcome	OR	95% CI	P-value
HMGCR pLoF	7/12,935	13/156,725	Cataract	6.48	2.51-16.71	5.2E-4
	7/11,373	13/157,787	Cataract surgery	7.50	2.91-19.35	2.1E-4

Table S18. Rare variant association excluding variants residing within the last 5 % of the coding sequence.

	pLoF carriers/Total no.					
	Cases	Controls	Outcome	OR	95% CI	P-value
HMGCR pLoF	7/12,936	22/156,236	Cataract	4.02	1.68-9.61	5.00E-03
	7/11,374	13/157,798	Cataract surgery	4.67	1.95-11.18	2.00E-03

Table S19. Evaluation of horizontal pleiotropy, by evaluation of models adjusted for cataract risk factors.

Model	Outcome	OR (95% CI)	*P*
Genetic risk score analysis			
Main model	Cataract	1.14 (1.00-1.29)	0.045
	Cataract surgery	1.19 (1.04-1.36)	0.009
+ Type 2 diabetes	Cataract	1.14 (1.00-1.29)	0.044
	Cataract surgery	1.19 (1.05-1.36)	0.009
+ Uveitis	Cataract	1.14 (1.00-1.29)	0.047
	Cataract surgery	1.19 (1.04-1.36)	0.010
+ Smoking	Cataract	1.13 (1.00-1.28)	0.055
	Cataract surgery	1.19 (1.04-1.35)	0.011
+ Steroid treatment	Cataract	1.13 (1.00-1.29)	0.048
	Cataract surgery	1.19 (1.04-1.36)	0.010
pLoF variant analysis			
Main model	Cataract	4.54 (1.96-10.53)	1.40E-03
	Cataract surgery	5.27 (2.27-12.25)	5.37E-04
+ Type 2 diabetes	Cataract	4.53 (1.98-10.37)	1.40E-03
	Cataract surgery	5.26 (2.30-12.04)	5.50E-04
+ Uveitis	Cataract	4.58 (2.00-10.46)	1.33E-03
	Cataract surgery	5.31 (2.32-12.16)	5.10E-04
+ Smoking	Cataract	4.92 (2.13-11.32)	8.93E-04
	Cataract surgery	5.72 (2.48-13.20)	3.31E-04
+ Steroid treatment	Cataract	4.62 (2.02-10.55)	1.20E-03
	Cataract surgery	5.35 (2.34-12.25)	4.83E-04

The table displays the results from mediation analysis, in which the main model was adjusted for important risk factors for cataract. Below are the data fields used to infer covariates: T2D (130709); Uveitis (131158 and 131172); Steroid treatment (extracted from self-reported medication [20003], in which eye drops, oral, intravenous, inhaled routes of administration were considered); Smoking status (ever vs. never [20116]).

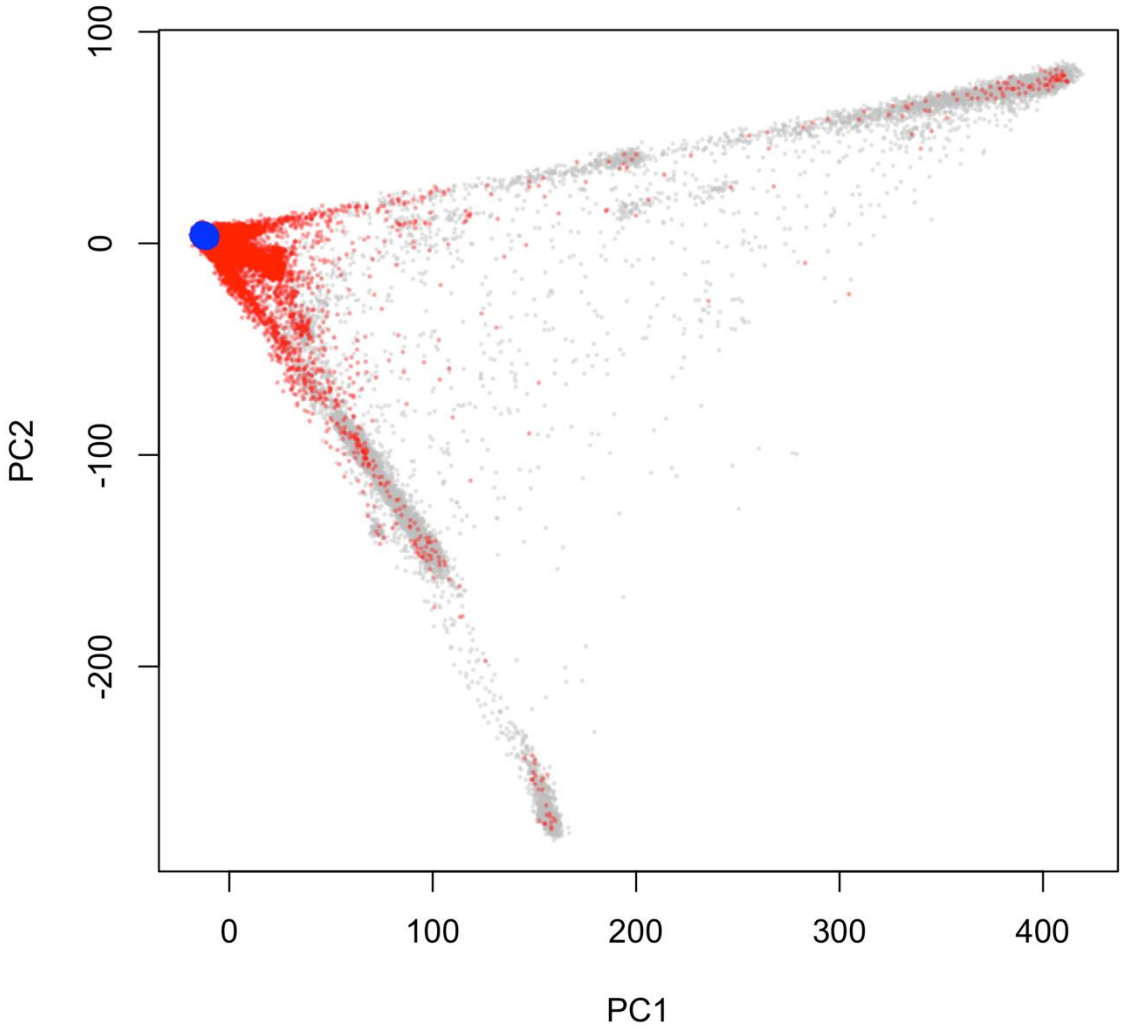
Table S20. LDL-C lowering alleles in *HMGCR, PCSK9* and *NPC1L1* and risk of cataract and cataract surgery.

SNP	Genomic coordinate, chromosome and position	Exposure allele[a]	Exposure Allele Frequency	Effect Size (SE) per allele in standardized LDL-C levels (mg/dl)[b]	OR of cataract (95% CI) per allele	P-value	OR of cataract surgery (95% CI) per allele	P-value
HMGCR variants								
rs12916	5:74656539	T	0.569	-2.345 (0.122)	1.01 (1.00-1.03)	0.138	1.02 (1.00-1.04)	0.027
rs5909	5:74656175	G	0.898	-1.974 (0.282)	1.01 (0.98-1.04)	0.687	1.02 (0.99-1.05)	0.288
rs2303152	5:74641707	G	0.880	-1.354 (0.205)	1.02 (0.99-1.05)	0.144	1.03 (1.00-1.06)	0.069
rs10066707	5:74560579	G	0.583	-1.590 (0.173)	1.02 (1.00-1.04)	0.049	1.02 (1.00-1.04)	0.022
rs2006760	5:74562029	C	0.814	-1.706 (0.243	1.02 (1.00-1.04)	0.069	1.02 (1.00-1.04)	0.071
PCSK9 variants								
rs11206510	1:55496039	C	0.154	-2.659 (0.005)	1.00 (0.98-1.03)	0.766	1.00 (0.98-1.03)	0.673
rs2479409	1:55504650	A	0.668	-2.054 (0.004)	1.00 (0.98-1.01)	0.698	1.00 (0.98-1.02)	0.863
rs2149041	1:55502137	C	0.839	-2.035 (0.005)	1.00 (0.98-1.02)	0.949	1.01 (0.98-1.03)	0.594
rs2479394	1:55486064	A	0.715	-1.235 (0.004)	1.00 (0.98-1.02)	0.765	1.01 (0.99-1.03)	0.432
rs10888897	1:55513061	T	0.395	-1.622 (0.004)	0.99 (0.98-1.01)	0.379	1.00 (0.98-1.01)	0.673
rs7552841	1:55518752	C	0.635	-1.178 (0.004)	1.01 (0.99-1.03)	0.306	1.01 (0.99-1.03)	0.486
rs562556	1:55524237	G	0.194	-2.048 (0.007)	0.99 (0.97-1.01)	0.317	1.00 (0.98-1.02)	0.885
NPC1L1 variants								
rs217386	7:44600695	A	0.408	-1.125 (0.118)	0.99 (0.98-1.01)	0.443	0.99 (0.97-1.01)	0.321
rs2073547	7:44582331	A	0.806	-1.126 (1.152)	1.00 (0.98-1.02)	0.806	1.00 (0.98-1.03)	0.766
rs7791240	7:44602589	T	0.909	-1.127 (0.202)	1.02 (0.99-1.05)	0.122	1.02 (0.99-1.06)	0.141

rs10234070	7:44537696	C	0.904	-1.128 (0.183)	0.99 (0.97-1.02)	0.540	0.99 (0.96-1.01)	0.278
rs2300414	7:44682938	G	0.930	-1.129 (0.248)	1.02 (0.99-1.06)	0.211	1.02 (0.98-1.05)	0.354

a) The exposure allele is the LDL-C lowering allele. b) Data from Ference et al (PMID 30865797)

Figure S1. Principal component plots.



Population substructure shown by principal components (PC). We plotted PC 1 and 2 from all participants with available exomes (n = 200,643). Dots in grey represents individuals who did not self-report as “White”. Dots in red are individuals who self-reported as “White” but was not clustered by aberrant. Dots in blue represent the set of individuals with European ancestry (n = 175,336).