Phenome-wide Mendelian randomisation analysis identifies causal factors for age-related macular degeneration.

Thomas H Julian¹,², Johnathan Cooper-Knock³, Stuart MacGregor⁴, Hui Guo⁵ Tariq Aslam⁶, Eleanor Sanderson⁷, Graeme Black*¹,⁹, Panagiotis I Sergouniotis*¹,²,⁸,⁹,

1. Division of Evolution, Infection and Genomics, School of Biological Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, UK.
2. Manchester Royal Eye Hospital, Manchester University NHS Foundation Trust, Manchester, UK.
3. Department of Neuroscience, Sheffield Institute for Translational Neuroscience (SITraN), University of Sheffield, Sheffield, UK.
4. Statistical Genetics, QIMR Berghofer Medical Research Institute, Brisbane, Australia.
5. Centre for Biostatistics, Division of Population Health, Health Services Research & Primary Care, School of Health Sciences, Faculty of Biology, Medicine and Health, The University of Manchester, Manchester, UK.
6. Division of Pharmacy and Optometry, Faculty of Biology, Medicine and Health, School of Health Sciences, The University of Manchester, Manchester, UK.
7. MRC Integrative Epidemiology Unit, University of Bristol, Bristol, United Kingdom.
8. European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Cambridge, UK.
9. Manchester Centre for Genomic Medicine, Saint Mary’s Hospital, Manchester University NHS Foundation Trust, Manchester, UK.

*Corresponding authors: Panagiotis Sergouniotis & Graeme Black

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
ABSTRACT:

Introduction:
Age-related macular degeneration (AMD) is the leading cause of blindness in the industrialised world and is projected to affect 288 million people worldwide by 2040. Aiming to identify causal factors for this common disorder, we designed and applied a phenome-wide Mendelian randomisation (MR) approach to identify therapeutic targets and avenues for future research.

Methods:
We evaluated the effect of 4,591 exposure traits on early AMD using univariable MR. Statistically significant results were explored further using validation in an advanced AMD cohort, MR Bayesian model averaging (MR BMA) and multivariable MR.

Results:
44 traits were found to be putatively causal for early AMD in univariable analysis. MR BMA of lipid traits suggested a causal role for serum sphingomyelin (marginal inclusion probability=0.76, model averaged causal effect=0.29). Univariable MR analysis supported roles for complement and immune cell traits. Serum proteins were found to have significant relationships with AMD including S100-A5 (Odds ratio (OR)= 1.07, p = 6.80E-06), cathepsin F (OR=1.10, p = 7.16E-05) and serine palmitoyltransferase 2 (OR= 0.86, p value=1.00E-03).

Conclusions:
The results of this study support several putative causal factors in AMD and highlight avenues for future clinical research.
INTRODUCTION:

Age-related macular degeneration (AMD) is a progressive multifactorial disorder that affects individuals who are ≥50 years old and is caused by the complex interplay of multiple genetic and environmental risk factors. Genome-wide association studies (GWAS) have identified AMD-implicated variants in at least 69 loci including important risk alleles in the 1q32 and 10q26 genomic regions (corresponding to the CFH - complement factor H – and ARMS2/HTRA1 locus respectively) (Fritsche et al., 2016; Winkler et al., 2020). Other key risk factors include age, smoking, alcohol consumption and low dietary intake of antioxidants (carotenoids, zinc) (Chakravarthy et al., 2010).

AMD can be categorised according to severity (early, intermediate, or advanced) or based on the presence of neovascularization (neovascular or non-neovascular). Advanced AMD results in loss of central vision, often leading to severe visual impairment (Fleckenstein et al., 2021). Notably, AMD is a major cause of blindness in the elderly population and represents a substantial global burden that is expected to continue to grow into the future as an ageing population expands worldwide.

Mendelian randomization (MR) is a statistical approach that uses genetic variation to look for causal relationships between exposures (such as body mass index) and medically-relevant outcomes (such as risk of a specific disease) (Julian et al., 2021). MR is increasingly being used as it can overcome a major limitation of observational studies: unmeasured confounding (Sanderson et al., 2022). To minimise issues with certain types of confounding and to support causal inference statements, MR uses genetic variation as an “instrument” (i.e. as a variable that is associated with the exposure, independent of the confounders, and is not associated with the outcome, other than through the exposure). Its principles are based on Mendel’s laws of segregation and independent assortment, which state that offspring inherit alleles randomly from their parents and randomly with respect to other locations in the genome (Julian et al., 2021).

As a result, genetic variants that are related to an exposure of interest can be used to proxy the part of the exposure variable that is independent of possible confounding influences (e.g., from the environment and other traits). It is noted that analogies have been drawn between MR and randomised controlled trials with...
these two approaches considered proximal in terms of hierarchy of evidence. (Julian et al., 2021) To date, the use of MR approaches in the context of AMD has been limited, although these methods have been successfully implemented to explore the relationship between AMD and a small number of traits including lipids, thyroid function, CRP and complement factors. (Cipriani et al., 2021; Han et al., 2021, 2020b; Li et al., 2022; Zuber et al., 2020)

In this study, we have developed a systematic, broad (‘phenome-wide’) MR analysis methodology which serves to investigate the relationship between early AMD and several thousand exposure variables. Using this approach, we report a set of traits that are robustly associated with genetic liability to AMD.

METHODS:

Data sources

Outcome data

Two AMD phenotypes were used as outcome measures in this study. The first was early AMD. The GWAS summary statistics for this phenotype were taken from a meta-analysis by Winkler et al (Winkler et al., 2020). This study focused on populations of European ancestries and used data from the ARIC, AugUR, CHS, GHS, IAMDGC, KORA S4, LIFE-Adult NICOLA, UKBB and WHI studies (14,034 early AMD cases and 91,214 controls overall). A full description of how these studies classified participants as “early AMD” can be found in Winkler et al (Winkler et al., 2020). The second phenotype that we studied was advanced AMD. For this trait, we drew on GWAS summary statistics from a multiple trait analysis of GWAS (MTAG) study by Han et al. (Han et al., 2020a). This meta-analysis also focused on individuals of European ancestries and derived data from the IAMDGC 2013 (Fritsche et al., 2013) (17,181 cases and 60,074 controls) and 2016 (Fritsche et al., 2016) (16,144 advanced AMD cases and 17,832 controls) studies as well as the GER A study (Kvale et al., 2015) (4,017 cases and 14,984 controls). The summary statistics produced in this GWAS are primarily reflective of advanced AMD, but the GER A cohort included both advanced and intermediate AMD cases. Notably, the MTAG approach is able to leverage the high genetic correlation between the input phenotypes to detect genetic associations for advanced AMD.
Exposure data

In this study we performed a phenome-wide screen to make causal inferences on the role of an extensive range of traits in early and advanced AMD. To achieve this, we used both published and unpublished GWAS data from the IEU open GWAS database; these were accessible via the TwoSampleMR programme in R(Hemani et al., 2018). We included all European GWAS within the database but removed imaging phenotypes and eQTL data. In the early AMD analysis, we also excluded studies from the “ukb” and “met-d” batches as these are studies entirely derived from UK Biobank data and therefore had extensive population overlap with the early AMD GWAS (Sudlow et al., 2015). In the advanced AMD analysis, “ukb” and “met-d” batches were included. The early AMD analysis was conducted on 30/12/2021 and a total of 10,979 traits were considered for analysis. The advanced AMD analysis was conducted on 08/01/2022. On 26/01/2022 we added newly published ‘finn-b’ (n=2803) traits to our analysis in place of outdated finn-a traits (n=1489). As it is not possible to manually inspect the degree of population overlap for all traits prior to conducting the analysis, the degree of overlap for all significant traits was inspected after the analysis.

Instrument selection

In this study, we used a statistically driven approach to instrumental variable selection. In general, an arbitrary p-value threshold is set for the identification of appropriate single-nucleotide polymorphisms (SNPs) that are used as instrumental variables (referred to thereafter as instruments). A conventional value for selection of instruments is p=5E-08, but this can, in some cases, be problematic. For example, when the number of instruments exceeding this threshold is small, the analysis can be underpowered or, in certain cases of unbiased screens, the results can be inflated.(Boddy et al., 2022) With this in mind, we have set the p-value for instrument selection for each trait to the level where >5 instruments are available for each analysis. More specifically, for each trait, the analysis would first be conducted with a p-value for inclusion of 5E-8 and would sequentially increase by a factor of 10 each time until >5 eligible instruments are identified to a predefined maximum of p=5E-05. Final range of p values for inclusion was 5E-06 – 5E-08.

Proxies
Where an exposure instrument was not present in the outcome dataset, we sought to identify a suitable proxy (Hartwig et al., 2016). In the early AMD analysis, we were able to achieve this by using the TwoSampleMR software with a linkage disequilibrium $R^2$ value of 0.9 or above for identification of SNPs (Purcell et al., 2007). For the advanced AMD phenotype we used data which was not derived from the TwoSampleMR database and therefore the ensembl server was used to derive proxies (Cunningham et al., 2022; Hemani et al., 2018).

**Clumping**

SNPs were clumped using a linkage disequilibrium $R^2$ value of 0.001 and a genetic distance cut-off of 10,000 kilo-bases.

**Harmonisation**

The effects of instruments on outcomes and exposures were harmonised in order to ensure that the beta values (i.e. the regression analysis estimates of effect size) were signed with respect to the same alleles (Hartwig et al., 2016). For palindromic alleles (i.e., alleles that are the same on the forward as on the reverse strand), those with minor allele frequency (MAF) > 0.42 were omitted from the analysis in order to reduce the risk of errors due to strand issues.

**Removal of pleiotropic genetic variants and outliers**

In addition to using a range of robust measures of scale (which quantify the statistical dispersion in a sample of numerical data while resisting outliers) and quality control measures (as detailed below), we endeavoured to remove pleiotropic instruments and outliers from the analysis. We operated this using a statistical approach in which instruments which were more significant for the outcome than the exposure were removed (Hemani et al., 2017). Radial MR, a simulation-based approach that detects outlying instruments was also utilised (Bowden et al., 2018).

**Causal inference**
MR estimation was primarily performed using a multiplicative random effects (MRE) inverse variance weighted (IVW) analysis. MRE IVW was selected over a fixed effects (FE) approach as it allows inclusion of heterogeneous instruments, certain to occur within the breadth of this screen (Burgess et al., 2019). A range of robust measures of scale were utilised to ensure the accuracy of the results and to cover for a range of violations of the MR assumptions (Burgess et al., 2019). The robust measures selected were the weighted median (Bowden et al., 2016a), Egger (Burgess and Thompson, 2017a), weighted mode (Hartwig et al., 2017) and radial MR with modified second order weights (Bowden et al., 2018).

**Further quality control**

Instrument strength was determined using the F statistic (which tests the association between the instruments and the exposure) (Burgess et al., 2011). F statistics were calculated against the final included set of instruments. A mean F statistic >10 was considered sufficiently strong.

The Cochran’s Q test was performed for each analysis. Cochran’s Q is a measure of heterogeneity among causal estimates and serves as an indicator of the presence of horizontal pleiotropy (which occurs when an instrument exhibits effects on the outcome through pathways other than the exposure) (Bowden and Holmes, 2019). It is important to note that a heterogeneous instrument is not necessarily invalid, but rather calls for a primary assessment with an MRE IVW rather than an FE approach; this has been conducted as standard throughout our analysis.

The MR-Egger intercept test was used to detect horizontal pleiotropy. When this occurs, the Egger regression is robust to horizontal pleiotropy under the assumption that that pleiotropy is uncorrelated with the association between the SNP and the exposure (Burgess and Thompson, 2017b). In this study, assuming there is not demonstrable horizontal pleiotropy we do not utilise the Egger regression to determine causal effect given its low power with a small number of SNPs (Bowden et al., 2015).
The $I^2$ statistic was calculated as a measure of heterogeneity between variant specific causal estimates. A $I^2 < 0.9$ indicates that Egger is more likely to be biased towards the null through violation of the ’NO Measurement Error’ (NOME) assumption (Bowden et al., 2016b).

Leave-one-out (LOO) cross-validation was performed for every analysis to determine if any particular SNP was driving the significance of the causal estimates.

**Management of duplicate traits/GWAS**

As the GWAS database used contains multiple different GWAS studies for the certain traits, some exposures were analysed on multiple occasions. Where this occurred, we considered the largest sample size study to be the primary analysis. Where there were duplicate studies in the same population, we used the study with the largest F statistic.

**Identification of significant results**

Before considering an MR result to be significant, the results of a range of causal inference and quality control tests should be considered. Notably, it is not necessary for a study to find significance in all measures to determine a true causal relationship. MR is a low power study type and as such an overly conservative approach to multiple testing can be excessive (Burgess et al., 2019). However, in the context of the present study we have chosen to consider the results of our early AMD phenome-wide screen as significant only if they remained: significant after false discovery rate correction (FDR) in the MRE IVW; nominally significant in weighted mode and weighted median; and nominally significant throughout the leave-one-out analysis (IVW-MRE) (Benjamini and Hochberg, 1995). This conservative approach was selected as a large number of phenotypes was considered and because we wanted to focus on high confidence signals.

Where we identified causal traits for early AMD, we have additionally explored their relationship with advanced AMD so as to ascertain whether these traits are also implicated in more severe disease. These two AMD classifications are phenotypically distinct but are generally part of the same disease spectrum.
For this reason, where traits failed to replicate as causal factors in the advanced AMD dataset this would not necessarily imply that they are not causal for early AMD. However, significance in both AMD traits provides support for the detected causal links and evidence that a factor plays a role across the spectrum of disease.

Multivariable mendelian randomisation

Multivariable MR was performed in circumstances where we aimed to estimate the effect of >1 closely related or potentially confounding exposure traits. (Sanderson et al., 2019) P-values for the inclusion of instruments for the exposures of interest were optimised to obtain sufficiently high (>10) conditional F-statistics for reliable analysis (Sanderson et al., 2021). With this in mind, selection for exposures began at p>5E-08 and where a trait’s instruments had a conditional F-statistic <10 the p-value for selection was reduced in an automated manner by factor of 10 until an F-statistic >10 was obtained in the final analysis. We utilised the same clumping procedure as in our univariable MR analysis. Adjusted Cochran’s Q statistics were calculated, with a p-value of <0.05 indicating significant heterogeneity. Where the Cochran’s Q statistic indicated heterogeneity, a Q-statistic minimisation procedure was used to evaluate the causal relationship, with testing both assuming high levels (0.9) and low levels (0.1) of phenotypic correlation (Sanderson et al., 2021).

Two-sample multivariable Mendelian randomisation approach based on Bayesian model averaging (MR-BMA)

Multivariable MR can be used to obtain effect estimates for a few (potentially related) traits, however it cannot be directly applied when there are many traits to be considered. MR-BMA is a Bayesian approach developed by Zuber et al (Zuber et al., 2020). Their approach searches over sets of potential risk factors to determine which are most likely to be causal risk factors.

Notably, Zuber et al. have previously performed an excellent analysis which considered the role of lipids against an older AMD GWAS. This study served as the proof of the MR-BMA method and also
demonstrated that several lipid traits have causal roles in AMD. However, this analysis has two potential limitations. First, it downweighed fatty acid traits through limiting composite traits for SNP identification to HDL, LDL and triglycerides. Second, numerous lipids with a potential role in AMD were not included in the analysis. For these reasons, we chose to conduct a more inclusive analysis with a slightly altered approach.

The following study design modifications were made compared to Zuber et al. (Zuber et al., 2020):

1. Fatty acids were included as a composite trait (utilising GWAS data for serum fatty acids derived from Nightingale Health 2020 listed in TwoSampleMR (Hemani et al., 2018)).
2. All lipid and fatty acid measures in the Kettunen et al. GWAS were considered as potential causal traits (n=102 traits) (Kettunen et al., 2016);
3. We utilised a more recent, early AMD GWAS (Winkler et al., 2020);
4. In general, multivariable MR (of any sort) cannot produce reliable results where the included traits are \( \geq 0.99 \) correlated with respect to the included instruments. For this reason, where two traits were highly correlated, one was removed at random rather than manually selecting traits in a manner which risks selection bias.

Additionally, we performed MR-BMA for immune cell and complement phenotypes. In this analysis instruments were obtained at genome wide significance (5E-08) for every included exposure in the model given that composite traits were not available and/or applicable. For the immune cell MR-BMA, all immune traits in the Orru et al. GWAS present in TwoSampleMR were used as exposures (Orrù et al., 2020). In the complement analysis, we used all complement traits available in both Sun et al. and Suhre et al. with the exception of complement subfractions (Suhre et al., 2017; Sun et al., 2018).

For our MR-BMA analysis, the prior probability was set to 0.1 and the prior variance was set to 0.25. We operated our stochastic search with 10,000 iterations and calculated empirical p-values with 100,000 permutations.

**Presentation of effect sizes:**
We aim to present effect sizes to enable readers to appraise the impact of putative risk factors. In univariable MR, these are presented as OR per 1SD change in the exposure for continuous traits and beta values for binary exposure traits. We present this way because odds ratios are uninformative for binary exposure traits (Burgess and Labrecque, 2018). Similarly, beta values are not presented by all MR authorities where an exposure variable is binary in nature and it’s important to recognise that these are indicative only. Multivariable MR effect sizes are presented as beta estimates irrespective of the nature of the exposure variable, given that the role of Multivariable MR within this study is purely to identify confounders.

MR-BMA effect sizes are presented in the form of MACE. The MACE is a conservative estimate of the direct causal effect of an exposure on an outcome averaged across models. It should be noted that the primary function of MR-BMA is to highlight the probable causal trait amongst a number of candidate causal risk factors, and the authors of this study warn that the causal effects should be used to interpret direction of effect but not necessarily as an absolute guide to magnitude (Zuber et al., 2020).

Software

R version 4.1.0.

TwoSampleMR version 0.5.6.

RadialMR version 1.0.

MVMR version 0.3

MR BMA code was sourced from https://github.com/verena-zuber/demo_AMD.

RESULTS:

Overview

Focusing on early AMD, we applied univariable MR analysis to a broad range of traits. Following careful quality control, we sought to replicate significant results in an advanced AMD dataset and conducted further analyses using multivariable MR and MR-BMA.
Overall, 4,591 traits were eligible for analysis. Among these, 44 were found to be putatively causal for early AMD (Table 1 and Supplementary Data 1 & 2). Most of these causal traits were serum lipoprotein concentration and compositional measures (n=29). Other significant traits identified included immune cell phenotypes (n=5), serum proteins (n=6) and disease phenotypes (n=4).

<table>
<thead>
<tr>
<th>Trait Name</th>
<th>Odds Ratio</th>
<th>FDR Adjusted IVW P Value</th>
<th>MRE IVW Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rheumatoid arthritis</td>
<td>NA</td>
<td>1.51E-06</td>
<td>0.08</td>
</tr>
<tr>
<td>Unswitched memory B cell %B cell</td>
<td>1.11</td>
<td>5.40E-05</td>
<td>0.10</td>
</tr>
<tr>
<td>CD62L- Dendritic Cell %Dendritic Cell</td>
<td>0.95</td>
<td>3.32E-02</td>
<td>-0.05</td>
</tr>
<tr>
<td>Effector Memory CD8+ T cell Absolute Count</td>
<td>1.08</td>
<td>1.79E-04</td>
<td>0.07</td>
</tr>
<tr>
<td>CD25 on IgD+ CD38- naive B cell</td>
<td>0.93</td>
<td>1.25E-04</td>
<td>-0.07</td>
</tr>
<tr>
<td>CD80 on plasmacytoid Dendritic Cell</td>
<td>0.96</td>
<td>1.63E-02</td>
<td>-0.04</td>
</tr>
<tr>
<td>Total cholesterol in IDL</td>
<td>0.80</td>
<td>6.48E-08</td>
<td>-0.22</td>
</tr>
<tr>
<td>Free cholesterol in IDL</td>
<td>0.80</td>
<td>5.51E-08</td>
<td>-0.22</td>
</tr>
<tr>
<td>Total lipids in IDL</td>
<td>0.79</td>
<td>3.35E-09</td>
<td>-0.23</td>
</tr>
<tr>
<td>Concentration of IDL particles</td>
<td>0.80</td>
<td>9.83E-09</td>
<td>-0.23</td>
</tr>
<tr>
<td>Phospholipids in IDL</td>
<td>0.79</td>
<td>1.86E-09</td>
<td>-0.23</td>
</tr>
<tr>
<td>Triglycerides in IDL</td>
<td>0.84</td>
<td>2.40E-07</td>
<td>-0.18</td>
</tr>
<tr>
<td>Total cholesterol in large LDL</td>
<td>0.83</td>
<td>9.85E-08</td>
<td>-0.19</td>
</tr>
<tr>
<td>Cholesterol esters in large VLDL</td>
<td>0.83</td>
<td>1.41E-07</td>
<td>-0.19</td>
</tr>
<tr>
<td>Free cholesterol in large LDL</td>
<td>0.83</td>
<td>7.07E-07</td>
<td>-0.18</td>
</tr>
<tr>
<td>Total lipids in large LDL</td>
<td>0.83</td>
<td>1.04E-07</td>
<td>-0.19</td>
</tr>
<tr>
<td>Concentration of large LDL particles</td>
<td>0.83</td>
<td>2.40E-07</td>
<td>-0.19</td>
</tr>
<tr>
<td>Phospholipids in large LDL</td>
<td>0.83</td>
<td>1.33E-06</td>
<td>-0.18</td>
</tr>
<tr>
<td>Cholesterol esters in large VLDL</td>
<td>0.82</td>
<td>6.55E-03</td>
<td>-0.20</td>
</tr>
<tr>
<td>18:2, linoleic acid (LA)</td>
<td>0.80</td>
<td>2.93E-07</td>
<td>-0.23</td>
</tr>
<tr>
<td>Total cholesterol in LDL</td>
<td>0.82</td>
<td>4.46E-08</td>
<td>-0.19</td>
</tr>
<tr>
<td>Total cholesterol in medium LDL</td>
<td>0.82</td>
<td>3.64E-09</td>
<td>-0.20</td>
</tr>
<tr>
<td>Cholesterol esters in medium LDL</td>
<td>0.82</td>
<td>5.93E-09</td>
<td>-0.20</td>
</tr>
<tr>
<td>Lipoprotein Fraction</td>
<td>FDR Adjusted p-Value</td>
<td>Log Odds Ratio</td>
<td>Beta</td>
</tr>
<tr>
<td>----------------------</td>
<td>----------------------</td>
<td>----------------</td>
<td>------</td>
</tr>
<tr>
<td>Total lipids in medium LDL</td>
<td>0.81</td>
<td>3.35E-09</td>
<td>-0.21</td>
</tr>
<tr>
<td>Concentration of medium LDL particles</td>
<td>0.81</td>
<td>3.64E-09</td>
<td>-0.21</td>
</tr>
<tr>
<td>Phospholipids in medium LDL</td>
<td>0.81</td>
<td>1.86E-09</td>
<td>-0.21</td>
</tr>
<tr>
<td>Total cholesterol in small LDL</td>
<td>0.81</td>
<td>9.85E-08</td>
<td>-0.21</td>
</tr>
<tr>
<td>Total lipids in small LDL</td>
<td>0.81</td>
<td>2.65E-07</td>
<td>-0.21</td>
</tr>
<tr>
<td>Total cholesterol in small VLDL</td>
<td>0.85</td>
<td>2.54E-02</td>
<td>-0.16</td>
</tr>
<tr>
<td>Serum total cholesterol</td>
<td>0.77</td>
<td>3.67E-08</td>
<td>-0.26</td>
</tr>
<tr>
<td>Total phosphoglycerides</td>
<td>0.81</td>
<td>3.93E-03</td>
<td>-0.21</td>
</tr>
<tr>
<td>Triglycerides in very large HDL</td>
<td>0.87</td>
<td>4.37E-04</td>
<td>-0.14</td>
</tr>
<tr>
<td>Total lipids in very small VLDL</td>
<td>0.84</td>
<td>1.97E-03</td>
<td>-0.18</td>
</tr>
<tr>
<td>Concentration of very small VLDL particles</td>
<td>0.84</td>
<td>9.04E-04</td>
<td>-0.18</td>
</tr>
<tr>
<td>Phospholipids in very small VLDL</td>
<td>0.83</td>
<td>3.63E-03</td>
<td>-0.19</td>
</tr>
<tr>
<td>Interferon alpha-10</td>
<td>1.14</td>
<td>1.83E-02</td>
<td>0.13</td>
</tr>
<tr>
<td>Protein S100-A5</td>
<td>1.07</td>
<td>6.94E-04</td>
<td>0.07</td>
</tr>
<tr>
<td>Serine palmitoyltransferase 2</td>
<td>0.86</td>
<td>3.17E-02</td>
<td>-0.15</td>
</tr>
<tr>
<td>CD59 glycoprotein</td>
<td>1.10</td>
<td>1.77E-03</td>
<td>0.09</td>
</tr>
<tr>
<td>Complement factor H-related protein 5</td>
<td>1.09</td>
<td>9.30E-05</td>
<td>0.09</td>
</tr>
<tr>
<td>Cathepsin F</td>
<td>1.10</td>
<td>4.44E-03</td>
<td>0.10</td>
</tr>
<tr>
<td>Benign neoplasm: Skin, unspecified</td>
<td>NA</td>
<td>8.23E-03</td>
<td>0.07</td>
</tr>
<tr>
<td>Psychiatric diseases</td>
<td>NA</td>
<td>8.54E-05</td>
<td>-0.43</td>
</tr>
<tr>
<td>Mytonic disorders</td>
<td>NA</td>
<td>2.32E-02</td>
<td>0.00</td>
</tr>
</tbody>
</table>

**Table 1.**

This table contains data from analysis in an early AMD cohort. Only traits passing stringent quality control described in the methods are listed. FDR = false discovery rate. IVW = inverse variance weighted. MRE = multiplicative random effects. % = as a proportion of. NA = not applicable, exposure traits of binary nature do not produce accurate odds ratios and as such beta can be used to infer direction of effect but not necessarily magnitude.

Lipoprotein metabolism is linked to AMD risk
Univariable MR demonstrated significant causal relationships for 28 serum lipoprotein measures and 1 serum fatty acid concentration (18:2 linoleic acid) in early AMD (Table 1). These relationships were also strongly supported by the results of an advanced AMD analysis (Kettunen et al., 2016) (Supplementary data 3).

Serum metabolites are highly correlated traits, and the instruments for serum lipoprotein and fatty acid measures in our study were demonstrably correlated (Supplementary Figure 1). MR-BMA was used in order to discern which traits were driving the causal relationships. In our initial analysis, two genetic variants (rs11065987 [BRAP] and rs10455872 [LPAL2]) were found to be outliers in terms of Q statistic (Q statistic 13.84-15.71 and 6.79-10.46 respectively across models) (Figures 1A-D). These SNPs were therefore omitted and the analysis was re-run. In the subsequent analysis, no outlier SNPs were identified. The top 10 models ranked according to their posterior probability are presented in Supplementary Data 4 and the top 10 individual risk factors with respect to their marginal inclusion probability (MIP; defined as the sum of the posterior probabilities over all the models where the risk factor is present) are shown in Table 2. The MIPs of all traits included in our model are plotted in Figure 2. The top four traits with respect to their MIP are serum sphingomyelins (MIP=0.76, model averaged causal effects (MACE) =0.29), triglycerides in IDL (MIP=0.32, MACE= -0.16), free cholesterol (MIP=0.20, MACE=0.07) and phospholipids in very small VLDL (MIP=0.63, MACE = -0.31). It is important to note that whilst MR-BMA functions to select the likely causal risk factor amongst a set of candidate causal traits, it is not possible to achieve this with certainty. As such, a definitive statement of causality for individual lipid traits cannot be obtained within the MR causal inference framework.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Risk factor (Trait ID)</th>
<th>Marginal Inclusion</th>
<th>Average Effect</th>
<th>P value</th>
<th>FDR p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Sphingomyelins (met-c-935)</td>
<td>0.76</td>
<td>0.30</td>
<td>2.40E-04</td>
<td>5.76E-03</td>
</tr>
<tr>
<td>2</td>
<td>Phospholipids in very small VLDL (met-c-955)</td>
<td>0.63</td>
<td>-0.31</td>
<td>1.00E-05</td>
<td>7.20E-04</td>
</tr>
</tbody>
</table>
### Table 2:

The lead causal traits identified by MR BMA of lipids in early AMD ranked according to their marginal inclusion probability (MIP).

<table>
<thead>
<tr>
<th></th>
<th>Trait</th>
<th>MIP</th>
<th>p</th>
<th>p₀</th>
<th>p₁</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>Triglycerides in IDL (met-c-872)</td>
<td>0.32</td>
<td>-0.16</td>
<td>2.10E-04</td>
<td>5.76E-03</td>
</tr>
<tr>
<td>4</td>
<td>Free cholesterol (met-c-858)</td>
<td>0.20</td>
<td>0.07</td>
<td>2.83E-03</td>
<td>5.09E-02</td>
</tr>
<tr>
<td>5</td>
<td>Omega-3 fatty acids (met-c-855)</td>
<td>0.07</td>
<td>0.01</td>
<td>3.70E-02</td>
<td>2.34E-01</td>
</tr>
<tr>
<td>6</td>
<td>Free cholesterol in very large HDL (met-c-944)</td>
<td>0.07</td>
<td>0.01</td>
<td>2.74E-02</td>
<td>2.34E-01</td>
</tr>
<tr>
<td>7</td>
<td>Total lipids in very small VLDL (met-c-953)</td>
<td>0.06</td>
<td>-0.02</td>
<td>2.08E-02</td>
<td>2.34E-01</td>
</tr>
<tr>
<td>8</td>
<td>Cholesterol esters in medium VLDL (met-c-910)</td>
<td>0.05</td>
<td>0.01</td>
<td>3.66E-02</td>
<td>2.34E-01</td>
</tr>
<tr>
<td>9</td>
<td>Cholesterol esters in very large HDL (met-c-943)</td>
<td>0.05</td>
<td>0.01</td>
<td>4.87E-02</td>
<td>2.34E-01</td>
</tr>
<tr>
<td>10</td>
<td>Ratio of bisallylic groups to double bonds (met-c-844)</td>
<td>0.05</td>
<td>0.01</td>
<td>4.41E-02</td>
<td>2.34E-01</td>
</tr>
</tbody>
</table>
Figure 1A-D:

Plots generated for the top-ranking models in the first run of MR-BMA of lipid traits and early AMD. Plots 2A and 2C are for Cook's distance and 2B and 2D of Cochran's Q. Outlier instruments are annotated.
Figure 2: A graph detailing the results of MR-BMA. Lipids are ranked according to their MIP. The four likely true-causal traits are identified.

Serine palmitoyltransferase 2, an enzyme which catalyses the first committed step in sphingolipid biosynthesis, was robustly associated with early AMD. The relevant odds ratio highlighted that genetic liability to increasing serum enzyme levels is protective of AMD (Table 1, Supplementary Data 1, OR 0.86, \( p = 1.00 \times 10^{-3} \)). It is of interest that increasing serum levels of the associated enzyme serine palmitoyltransferase 1 also appeared to be protective of early AMD in most measures (Supplementary Data 2, OR 0.94, IVW \( p = 2.00 \times 10^{-3} \), FDR IVW \( p = 0.05 \), weighted median \( p = 0.05 \), significant throughout).
leave-one-out analysis). Serine palmitoyltransferase enzymes were not significantly related to risk of advanced AMD (Supplementary Data 3).

Reinforcing a causal role for complement

Increasing serum levels of the complement proteins CD59 glycoprotein (OR 1.10, IVW p= 2.04E-05) and Complement factor H-related protein 5 (OR 1.10, IVW p = 7.70E-07) were found to be associated with genetic liability to early AMD and passed multiple testing correction. Whilst Complement C4, Complement factor B and Complement factor I were also related to early AMD risk at nominal level of significance, they did not exceed the conservative criteria for causal inference imposed in this study with respect to robust measures (Supplementary Data 2). In subsequent advanced AMD analyses (Supplementary Data 3), CD59 glycoprotein remained related to disease risk (OR 1.07, IVW p= 0.04, weighted median p=9.00E-04, MR-Egger p=2.00E-03, weighted mode p =7.00E-03). Complement factor I reached a nominal level of significance in the IVW measure (OR 0.95, p=0.05) but did not remain significant in robust tests. No other complement measures nominally significant for early AMD were significantly related to advanced AMD.

Complement traits are correlated with one another, and as such MR BMA was attempted. In the initial analysis, the relationship between complement proteins was strongly suggested to be driven by complement factor H (Posterior probability=0.83, MIP=1.00, MIP=-0.68). However, rs2274700 (CFHR3, Cook's distance 26.26-55.24 & Cochran's Q 0.87-2.12 across models) and rs10824796 (MBL2, Cook's distance 0.13-9.93 & Cochran's Q 0.87-22.12 across models) were identified as outlier instruments and therefore warranted removal from subsequent analysis. rs2274700, a genetic variant known to be associated with AMD risk (Liao et al., 2016), represented the only instrument strongly associated with complement factor H in the MR BMA analysis and as such its removal precluded an informative high-throughput analysis of the complement cascade. Whilst it is clear that there is a causal relationship between complement and AMD, we cannot comment with precision about the specific complement-related molecule driving this.

Other immune traits have mixed protective and causal effects in AMD
Our screen identified other serum immune traits with potential causal roles in early AMD, with several of these related to dendritic cell populations. These traits were Unswitched memory B cell as a proportion of B cell (OR=1.11, p=4E-07), CD62L- Dendritic Cell as a proportion of Dendritic Cells (OR=0.95, p=1.22E-03), Effector Memory CD8+ T cell Absolute Count (OR=0.95, p=1.56E-06), CD80 on plasmacytoid Dendritic Cell (OR=0.96, p=4.20E-04) and interferon alpha 10 (OR=1.14, p=5.38E-04). Notably, numerous immune traits were also found to be significantly associated with advanced AMD (Supplementary Data 3).

As with lipids, these traits are highly correlated and it is not possible to confidently address which immune cell trait is truly causal in a univariable analysis due to potential overlap in the genetic instruments for each trait. MR BMA using the Orru et al. GWAS for the above immune traits was not possible due to too few genetic variants strongly instrumenting across the exposure variables leading to a failure to construct a meaningful model (Orrù et al., 2020). For this reason, we are unable to determine which of these specific immune traits are driving the casual relationship between immune cell traits and AMD, though dendritic cell traits dominate the univariable analysis.

**Disease phenotypes are related to AMD risk:**

There are four disease phenotypes/groups which were found to be significant at a level which satisfy the criteria for causal inference defined in this study. These traits are rheumatoid arthritis (beta= 0.08, IVW p=9.86E-09), psychiatric diseases (beta=-0.43, IVW p=6.88E-07), ‘benign neoplasm: skin, unspecified’ (beta= 0.07, IVW p= 1.76E-04) and myotonic disorders (beta=-0.003, IVW p= 7.49E-04). None of these disorders were significant in advanced AMD. The very broad and non-specific definitions encapsulated in the GWAS of benign skin neoplasms, psychiatric disease and myotonic disorder make it challenging to apply further analyses within the scope of this study, but these may be interesting phenotypes for exploration in future studies.

In univariable MR, rheumatoid arthritis was identified to be causally related to AMD as detailed above. Multivariable MR with effector memory CD8+ T cell absolute count as a second exposure variable (utilising a Q statistic minimisation procedure) demonstrated that the causal effect of rheumatoid arthritis was either mediated by immune cells or underpinned by correlated pleiotropy (rheumatoid arthritis effect size: 0.03,
95% CI: -0.007-0.055; effector memory CD8+ T cell absolute count effect size 0.07, 95% CI: 0.001-0.19).

This finding persists in models assuming both very high (0.9) and low (0.1) phenotypic correlation.

With respect to the relationship between AMD and psychiatric disease, no reverse causation was identified (p=0.31). More specific psychiatric traits which could be analysed were not significantly related to genetic liability to early AMD (Supplementary Data 2). We also explored an extensive range of psychiatric disorders in the advanced AMD dataset and found no significant relationships (Supplementary Data 3).

Due to the broad nature of this trait, and negative results for all disorders explored, it is not possible to discern what the driving force behind the significant result for psychiatric disease. It appears probable that psychiatric disease is a false positive result given that it lacks the supporting evidence of biological plausibility, validation in an advanced AMD dataset, or supportive evidence through significance of other similar traits in the early AMD analysis.

Serum cathepsin F and S100 proteins have a causal role in AMD

In addition to the four previously mentioned serum proteins causally associated with early AMD (Table 1, Interferon alpha-10, CD59 glycoprotein, complement factor H-related protein and serine palmitoyltransferase 2) a further two serum proteins were identified to have causal effects on AMD. These are S100-A5 (OR 1.07, IVW p = 6.80E-06) and Cathepsin F (OR= 1.10, IVW p = 7.16E-05).

Serum cathepsin F (OR= 1.10, IVW p = 7.16E-05) was shown to be causally related to early AMD risk. However, no other serum cathepsin group proteins produced significant results in early AMD and no cathepsin group proteins were robustly significant in the advanced AMD analysis (although cathepsin S and G have significant IVW values of p=1.71E-05 and p=0.001 respectively, Supplementary data 3).

Serum protein S100-A5 was demonstrated to be causal for AMD (OR 1.07, IVW p = 6.80E-06). Additionally, protein S100-A13 was significantly related to early AMD risk in both IVW (OR=1.07, p=1.22E-05) and robust measures (weighted median p=3.31E-05, weighted mode p=9.00E-04, Egger p= 1.00E-03) but did not remain significant throughout the leave-one-out analysis, suggesting the relationship for this particular
protein could be driven by a small number of influential variants. Whilst protein S100-A2 was nominally significant in the IVW (p=0.04), it did not pass any robust measures. No S100 group proteins were significantly related to advanced AMD risk (Supplementary Data 3).

DISCUSSION:
In this study we used MR to advance our understanding of AMD pathogenesis. Our findings indicate protective effects for serum VLDL and IDL compositional and concentration traits in AMD. Causal effects for serum free cholesterol and sphingolipid metabolism were also highlighted. Notably, our results suggest that increasing serum levels of sphingomyelins are causally linked to AMD and we report protective relationships for specific enzymes implicated in sphingolipid biosynthesis. It is known that sphingolipids play key roles in retinal physiology, and there is emerging evidence that some sphingolipids may play a role in AMD pathogenesis. (Simon et al., 2021) Further, our findings are in keeping with those of a meta-analysis of metabolomic studies which described pathway enrichment analysis suggesting a role for sphingolipid metabolism in AMD (Hou et al., 2020). Whilst sphingomyelin itself has not previously been described as a causative factor for AMD, other substances involved in sphingolipid metabolism (including ceramide, which can serve as both a substrate for sphingomyelin synthesis or a product of sphingomyelin metabolism) have been suggested as causal factors through regulation of retinal cell death, inflammation and neovascularization (Simon et al., 2021). It is known that the balance of sphingolipids have crucial roles in determining cell fates, and therefore it is interesting to note the contrasting effects of serine palmitoyltransferase and sphingomyelin in this study (Taniguchi and Okazaki, 2020). Our study therefore adds weight to the assertion that sphingolipids play a key role in AMD pathophysiology. Further study of sphingolipid metabolism in AMD subjects is expected to provide important insights and to highlight potential treatment targets.

Our MR findings support the well-documented assertion that complement plays a key role in AMD pathogenesis. (Clark and Bishop, 2018; Jha et al., 2007; Sivaprasad and Chong, 2006). Whilst MR-BMA was not feasible, univariable analysis supported roles for complement factor H related protein 5 and CD59 glycoprotein. A causal signal was obtained for Complement factor H-related protein 5, a finding in keeping
with recent studies highlighting the role of this and other factor-H-related proteins (FHR-1 to FHR4; all involved in the regulation of complement factor C3b turnover) in AMD (Cipriani et al., 2021). An intriguing observation was the apparently causal role of increasing serum levels of the complement-related CD59 glycoprotein (an inhibitor of membrane attack complex formation) in both early and advanced AMD. This finding appears to be at odds with previous studies that investigated the impact of high levels of CD59 in a model of AMD and found a potential therapeutic benefit (Cashman et al., 2011). The relationship between systemic and local complement activation is poorly understood and studies have produced conflicting results with respect to the roles of each in AMD (Clark and Bishop, 2018; Jha et al., 2007). It can therefore be speculated that the superficially inconsistent results between our study and previous work on CD59 reflects the complicated relationship between serum and ocular CD59 expression.

It is interesting to note that a connection between complement and lipid accumulation has been proposed in AMD. This includes evidence that dysregulation of the complement system results in lipid deposition both systemically and in the retina, and metabolomic evidence that decreased VLDL and increased HDL are associated with increased complement activation independent of AMD status (Armento et al., 2021). Given our results, the relationship between VLDL and complement is an area warranting further study.

Cathepsins are a very diverse family of lysosomal proteases which play an important part in cellular homeostasis by participating in antigen processing and degrade chemokines and proteases (Yadati et al., 2020). Cathepsins are susceptible to age-related alterations, and have been linked to modulation of pro-inflammatory signalling pathways. Additionally, cathepsins D and S have a direct role to play in the retina in the degradation of photoreceptor outer segments. It has also been shown that homozygosity for variant B cystatin C (an inhibitor of cysteine proteases) causes reduced secretion of mature cystatin C and is associated with increased susceptibility to neovascular AMD (Turk et al., 2012; Zurdel et al., 2002). For these reasons, it has been suggested that dysregulation of cathepsin activity may be a factor in AMD pathophysiology. Here, we show data that suggests serum cathepsin F, a ubiquitously expressed cysteine cathepsin, has a causal relationship with early AMD. The roles of specific cathepsins in AMD is an area which has not yet been mechanistically explored.
S100 proteins act as intracellular regulators and extracellular signalling proteins. Intracellularly, S100 proteins have a wide range of roles including regulating proliferation, differentiation, apoptosis, calcium homeostasis, metabolism and inflammation (Donato et al., 2013). Extracellular S100 proteins have important roles in immunity and inflammation. Of the S100 protein family, serum S100-A5 is the only protein to be robustly identified as a causal factor for AMD in our study. Whilst S100 proteins have received little study in AMD thus far, the binding of S100B to RAGE and the subsequent increase in VEGF have previously been shown to be linked to the development of AMD (Ma et al., 2007; Xia et al., 2017). S100-A5 is a protein which has received very little study and as such its biological functions are mostly uncharacterised, but it is known that it is involved in inflammation via the activation of RAGE (Wheeler and Harms, 2017). Whilst it is outside the scope of this study to assert a mechanistic hypothesis for this finding, it seems likely that the role of S100-A5 would be linked to the inflammatory processes.

For the most part, the present study has captured the known AMD risk factors. A notable omission however is smoking, an accepted AMD risk factor. In this study, cigarettes smoked per day was associated with early AMD at a nominal level of significance (IVW p=0.009) but was not significant with FDR or in robust measures. This finding is most probably reflective of the underpowered nature of MR. Given the hypothesis free nature of this study and the low power of MR, we have not commented on negative results as evidence of a null relationship between environmental exposures and AMD.

In summary, this study has identified previously undescribed causal factors for AMD in addition to reinforcing several previously suggested AMD risk factors. Future research will follow-up these causal factors from a mechanistic perspective in order to open up avenues for therapeutic intervention.

**Ethical approval**

This study utilised publicly available data and no additional ethical approval is required.

**AUTHOR CONTRIBUTIONS:**
THJ, JCK, PIS, GB, HG and ES were involved in study conception and design. THJ and JCK performed the statistical analysis. PIS, GB, GH, TA, SM and ES reviewed the methodology and statistical analysis. All authors aided in interpretation of the results. All authors contributed to manuscript drafting and editing. All authors approved the final manuscript. All authors meet the ICMJE authorship criteria.

ACKNOWLEDGEMENTS:
This project was funded by the NIHR and the Wellcome Institutional Strategic Support Fund for Causal Inference. SM acknowledges Program Grant (1150144) and Senior Research Fellowship funding from the Australian National Health and Medical Research Council (NHMRC).

COMPETING INTERESTS:
TA is involved on advisory boards and has received grants and speaker fees from Allergan, Novartis, Bayer, Roche, Bausch and Lomb, Heidelberg, Topcon and Canon.
The remaining authors have no declarations.

DATA AVAILABILITY:
All data associated with this study are provided within the supplementary data.
REFERENCES:


Han X, Ong J-S, Hewitt AW, Gharahkhani P, MacGregor S. 2021. The effects of eight serum lipid biomarkers on age-related macular degeneration risk: a Mendelian randomization study. *Int J*


Kvale MN, Hesselson S, Hoffmann TJ, Cao Y, Chan D, Connell S, Croen LA, Dispensa BP, Eshragh J,


Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**:559–575.


Sanderson E, Spiller W, Bowden J. 2021. Testing and correcting for weak and pleiotropic instruments in...


