Meta-analysis of 542,934 subjects of European ancestry identifies new genes and mechanisms predisposing to refractive error and myopia

Supplementary Notes

Table of Contents

1 UK Biobank Eye and Vision Consortium Membership 3
2 The CREAM Consortium 6
3 23andMe 9
4 Cohort Description 10
4.1 UK Biobank 10
4.1.1 Phenotyping 10
4.1.2 Genotyping 14
4.1.3 Association analyses 15
4.2 23andMe 16
4.2.1 Phenotyping 16
4.2.2 Genotyping 17
4.2.3 Association analyses 18
4.3 GERA 18
4.3.1 Phenotyping 18
4.3.2 Genotyping 18
4.3.3 Association analyses 19
4.4 Cream Consortium 19
4.4.1 Phenotyping 19
4.4.2 Genotyping 19
4.4.3 Association analyses 19
4.5 EPIC 20
4.5.1 Phenotyping 20
4.5.2 Genotyping and imputation 20
4.5.3 Association analysis 21
5 STATISTICAL ANALYSES 21
5.1 Meta-analyses 21
5.2 Conditional analyses 21
5.3 Multiple testing correction 21
5.4 Genomic inflation 21
5.5 LDscore regression-based methods 22
5.5.1 Polygenicity vs. inflation 22
5.5.2 Calculation of genetic correlation 22
5.6 Associated SNPs and gene annotations 22
5.6.1 OMIM 22
5.6.2 The GWAS Catalog. 22

5.7 Graphical illustration of association 22
5.8 Mendelian randomization 22
5.9 Gene expression, GTEx and other transcription data 22
5.10 LD score regression applied to specifically expressed genes (LDSC-SEG) 23
5.11 SMR 23
5.11.1 Test description 23
5.11.2 Datasets for the SMR analyses: eQTL, *cis*-mQTL 23
5.12 Gene-set enrichment 24
5.12.1 GSEA definitions 24
5.13 Analyses of signals of natural selection 24
5.14 Estimation of effect size distributions for spherical equivalent 24
6 FULL ACKNOWLEDGEMENTS 25
6.1 UK Biobank 25
6.2 23andMe 25
6.3 GERA 25
6.4 Consortium for Refractive Error and Myopia (CREAM) 25
6.5 EPIC-Norfolk 25
6.6 King's College London authors 26
6.7 UK Biobank Eye Consortium members 26
6.8 University College London authors 26

1 UK Biobank Eye and Vision Consortium Membership

- Prof **Tariq ASLAM** - Manchester University, Manchester, United Kingdom
- Prof **Sarah BARMAN** - Kingston University, London, United Kingdom
- Prof **Jenny BARRETT** - University of Leeds, Yorkshire, United Kingdom
- Prof **Paul BISHOP** - Manchester University, Manchester, United Kingdom
- Mr **Peter BLOWS** - NIHR Biomedical Research Centre, Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology, London, United Kingdom
- Dr **Catey BUNCE** - King's College London, London, United Kingdom
- Dr **Roxana CARARE** - University of Southampton, Southampton, United Kingdom
- Prof **Usha CHAKRAVARTHY** - Queens University Belfast, Belfast, Ireland
- Miss **Michelle CHAN** - NIHR Biomedical Research Centre, Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology, London, United Kingdom
- Dr **Sharon CHUA** - NIHR Biomedical Research Centre, Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology, London, United Kingdom
- Prof **David CRABB** – City, University of London, London, United Kingdom
- Mrs **Philippa CUMBERLAND** - UCL Great Ormond Street Institute of Child Health, London, United Kingdom
- Dr **Alexander DAY** - NIHR Biomedical Research Centre, Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology, London, United Kingdom
- Miss **Parul DESAI** - NIHR Biomedical Research Centre, Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology, London, United Kingdom
- Prof **Bal DHILLON** - University of Edinburgh, Scotland, United Kingdom
- Prof Andrew DICK - University of Bristol, Bristol, United Kingdom
- Dr **Cathy EGAN** - NIHR Biomedical Research Centre, Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology, London, United Kingdom
- Prof **Sarah ENNIS** - University of Southampton, Southampton, United Kingdom
- Prof **Paul FOSTER** - NIHR Biomedical Research Centre, Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology, London, United Kingdom
- Dr **Marcus FRUTTIGER** - NIHR Biomedical Research Centre, Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology, London, United Kingdom
- Dr **John GALLACHER** - University of Oxford, Oxford, United Kingdom
- Prof **David (Ted) GARWAY-HEATH** - NIHR Biomedical Research Centre, Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology, London, United Kingdom
- Dr **Jane GIBSON** - University of Southampton, Southampton, United Kingdom
- Mr **Dan GORE** - NIHR Biomedical Research Centre, Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology, London, United Kingdom
- Prof **Jeremy GUGGENHEIM** - Cardiff University, Wales, United Kingdom
- Prof **Chris HAMMOND** - King's College London, London, United Kingdom
- Prof **Alison HARDCASTLE** - NIHR Biomedical Research Centre, Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology, London, United Kingdom
- Prof **Simon HARDING** - University of Liverpool, London, United Kingdom

- Dr **Ruth HOGG** - Queens University Belfast, Belfast, Ireland
- Dr **Pirro HYS**I - King's College London, London, United Kingdom
- Mr **Pearse A KEANE** - NIHR Biomedical Research Centre, Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology, London, United Kingdom
- Prof **Sir Peng Tee KHAW** - NIHR Biomedical Research Centre, Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology, London, United Kingdom
- Mr **Anthony KHAWAJA** - NIHR Biomedical Research Centre, Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology, London, United Kingdom
- Mr **Gerassimos LASCARATOS** - NIHR Biomedical Research Centre, Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology, London, United Kingdom
- Prof **Andrew LOTERY**- University of Southampton, Southampton, United Kingdom
- Prof Phil LUTHERT - NIHR Biomedical Research Centre, Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology, London, United Kingdom
- Dr **Tom MACGILLIVRAY** - University of Edinburgh, Scotland, United Kingdom
- Dr **Sarah MACKIE** - University of Leeds, Yorkshire, United Kingdom
- Prof **Keith MARTIN** - University of Cambridge, Cambridge, United Kingdom
- Ms **Michelle MCGAUGHEY** - Queen's University Belfast, Belfast, Ireland
- Dr **Bernadette MCGUINNESS** - Queen's University Belfast, Belfast, Ireland
- Dr **Gareth MCKAY** - Queen's University Belfast, Belfast, Ireland
- Mr **Martin MCKIBBIN** - Leeds Teaching Hospitals NHS Trust, Yorkshire, United Kingdom
- Dr **Danny MITRY** – NIHR Biomedical Research Centre, Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology, London, United Kingdom & Royal Free Hospital, London, United Kingdom
- Prof **Tony MOORE** - NIHR Biomedical Research Centre, Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology, London, United Kingdom
- Prof **James MORGAN** - Cardiff University, Wales, United Kingdom
- Ms **Zaynah MUTHY** – NIHR Biomedical Research Centre, Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology, London, United Kingdom
- Mr **Eoin O'SULLIVAN** - King's College Hospital NHS Foundation Trust, London, United Kingdom
- Dr **Chris OWEN** - St George's, University of London, London, United Kingdom
- Mr Praveen PATEL - NIHR Biomedical Research Centre, Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology, London, United Kingdom
- Mr **Euan PATERSON** - Queens University Belfast, Belfast, Ireland
- Dr **Tunde PETO** - Queen's University Belfast, Belfast, Ireland
- Dr **Axel PETZOLD** - UCL Institute of Neurology, London, United Kingdom
- Prof **Jugnoo RAHI** - UCL Great Ormond Street Institute of Child Health, London, United Kingdom
- Dr **Alicja RUDNICKA** - St George's, University of London, London, United Kingdom
- Mr **Jay SELF** - University of Southampton, Southampton, United Kingdom
- Prof **Sobha SIVAPRASAD** - NIHR Biomedical Research Centre, Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology, London, United Kingdom
- Mr **David STEEL** - Newcastle University, Newcastle, United Kingdom
- Mrs **Irene STRATTON** - Gloucestershire Hospitals NHS Foundation Trust

- Mr **Nicholas STROUTHIDIS** - NIHR Biomedical Research Centre, Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology, London, United Kingdom
- Prof **Cathie SUDLOW** - University of Edinburgh, Scotland, United Kingdom
- Dr **Caroline THAUNG** - NIHR Biomedical Research Centre, Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology, London, United Kingdom
- Miss **Dhanes THOMAS** - NIHR Biomedical Research Centre, Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology, London, United Kingdom
- Prof **Emanuele TRUCCO** - University of Dundee, Scotland, United Kingdom
- Prof **Adnan TUFAIL** - NIHR Biomedical Research Centre, Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology, London, United Kingdom
- Dr **Veronique VITART** - University of Edinburgh, Scotland, United Kingdom
- Prof **Stephen VERNON** – Nottingham University Hospitals NHS Trust, Nottingham, United Kingdom
- Mr **Ananth VISWANATHAN** - NIHR Biomedical Research Centre, Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology, London, United Kingdom
- Dr **Cathy WILLIAMS** - University of Bristol, Bristol, United Kingdom
- Dr **Katie WILLIAMS** - King's College London, London, United Kingdom
- Prof **Jayne WOODSIDE** - Queen's University Belfast, Belfast, Ireland
- Dr **Max YATES** - University of East Anglia, Norwich, United Kingdom
- Ms **Jennifer YIP** - University of Cambridge, Cambridge, United Kingdom
- Dr **Yalin ZHENG** - University of Liverpool, London, United Kingdom

2 The CREAM Consortium

Joan E. Bailey-Wilson[1], Paul Nigel Baird[2], Amutha Barathi Veluchamy[3-5], Ginevra Biino[6], Kathryn P. Burdon[7], Harry Campbell[8], Li Jia Chen[9], Ching-Yu Cheng[10-12], Emily Y. Chew[13], Jamie E. Craig[14], Phillippa M. Cumberland[15], Margaret M. Deangelis[16], Cécile Delcourt[17], Xiaohu Ding[18], Cornelia M. van Duijn[19], David M. Evans[20-22], Qiao Fan[23], Maurizio Fossarello[24], Paul J. Foster[25], Puya Gharahkhani[26], Adriana I. Iglesias[19,27,28], Jeremy A. Guggenheim[29], Xiaobo Guo[18,30], Annechien E.G. Haarman[19,28], Toomas Haller[31], Christopher J. Hammond[32], Xikun Han[26], Caroline Hayward[33], Mingguang He[2,18], Alex W. Hewitt[2,7,34], Quan Hoang[3,35] , Pirro G. Hysi[32], Robert P. Igo Jr.[36], Sudha K. Iyengar[36-38], , Jost B. Jonas[39,40], Mika Kähönen[41,42], Jaakko Kaprio[43,44], , Anthony P. Khawaja[25,45], Caroline C. W. Klaver[19,28,46], Barbara E. Klein[47], Ronald Klein[47], Jonathan H. Lass[36,37], Kris Lee[47], Terho Lehtimäki[48,49], Deyana Lewis[1], Qing Li[50], Shi-Ming Li[40], Leo-Pekka Lyytikäinen[48,49], Stuart MacGregor[26], David A. Mackey[2,7,34], Nicholas G. Martin[51], Akira Meguro[52], Andres Metspalu[31], Candace Middlebrooks[1], Masahiro Miyake[53], Nobuhisa Mizuki[52], Anthony Musolf[1], Stefan Nickels[54], Konrad Oexle[55], Chi Pui Pang[9], Olavi Pärssinen[56,57], Andrew D. Paterson[58], Norbert Pfeiffer[54], Ozren Polasek[59,60], Jugnoo S. Rahi[15,25,61], Olli Raitakari[62,63], Igor Rudan[8], Srujana Sahebjada[2], Seang-Mei Saw[64,65], Dwight Stambolian[66], Claire L. Simpson[1,67], E-Shyong Tai[65], Milly S. Tedja[19,28], J. Willem L. Tideman[19,28], Akitaka Tsujikawa[53], Virginie J.M. Verhoeven[19,27,28], Veronique Vitart[33], Ningli Wang[40], Juho Wedenoja[43,68], Wen Bin Wei[69], Cathy Williams[22], Katie M. Williams[32], James F. Wilson[8,33], Robert Wojciechowski[1,70,71], Ya Xing Wang[40], Kenji Yamashiro[72], Jason C. S. Yam[9], Maurice K.H. Yap[73], Seyhan Yazar[34], Shea Ping Yip[74], Terri L. Young[47], Xiangtian Zhou[75]

Affiliations

1. Computational and Statistical Genomics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland, USA.
2. Centre for Eye Research Australia, Ophthalmology, Department of Surgery, University of Melbourne, Royal Victorian Eye and Ear Hospital, Melbourne, Australia.
3. Singapore Eye Research Institute, Singapore National Eye Centre, Singapore.
4. Duke-NUS Medical School, Singapore, Singapore.
5. Department of Ophthalmology, National University Health Systems, National University of Singapore, Singapore.
6. Institute of Molecular Genetics, National Research Council of Italy, Pavia, Italy.
7. Department of Ophthalmology, Menzies Institute of Medical Research, University of Tasmania, Hobart, Australia.
8. Centre for Global Health Research, Usher Institute for Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK.
9. Department of Ophthalmology and Visual Sciences, The Chinese University of Hong Kong, Hong Kong Eye Hospital, Kowloon, Hong Kong.
10. Department of Ophthalmology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore.
11. Ocular Epidemiology Research Group, Singapore Eye Research Institute, Singapore National Eye Centre, Singapore.
12. Ophthalmology & Visual Sciences Academic Clinical Program (Eye ACP), Duke-NUS Medical School, Singapore.
13. Division of Epidemiology and Clinical Applications, National Eye Institute/National Institutes of Health, Bethesda, USA.
14. Department of Ophthalmology, Flinders University, Adelaide, Australia.

15. Great Ormond Street Institute of Child Health, University College London, London, UK.
16. Department of Ophthalmology and Visual Sciences, John Moran Eye Center, University of Utah, Salt Lake City, Utah, USA.
17. Université de Bordeaux, Inserm, Bordeaux Population Health Research Center, team LEHA, UMR 1219, F-33000 Bordeaux, France.
18. State Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Center, Sun Yat-sen University, Guangzhou, China.
19. Department of Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands.
20. Translational Research Institute, University of Queensland Diamantina Institute, Brisbane, Queensland, Australia.
21. MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK.
22. Department of Population Health Sciences, Bristol Medical School, Bristol, UK.
23. Centre for Quantitative Medicine, DUKE-National University of Singapore, Singapore.
24. University Hospital 'San Giovanni di Dio', Cagliari, Italy.
25. NIHR Biomedical Research Centre, Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology, London, UK.
26. Statistical Genetics, QIMR Berghofer Medical Research Institute, Brisbane, Australia.
27. Department of Clinical Genetics, Erasmus Medical Center, Rotterdam, The Netherlands.
28. Department of Ophthalmology, Erasmus Medical Center, Rotterdam, The Netherlands.
29. School of Optometry & Vision Sciences, Cardiff University, Cardiff, UK.
30. Department of Statistical Science, School of Mathematics, Sun Yat-Sen University, Guangzhou, China.
31. Institute of Genomics, University of Tartu, Tartu, Estonia.
32. Section of Academic Ophthalmology, School of Life Course Sciences, King's College London, London, UK.
33. MRC Human Genetics Unit, MRC Institute of Genetics & Molecular Medicine, University of Edinburgh, Edinburgh, UK.
34. Centre for Ophthalmology and Visual Science, Lions Eye Institute, University of Western Australia, Perth, Australia.
35. Department of Ophthalmology, Columbia University, New York, USA.
36. Department of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, Ohio, USA.
37. Department of Ophthalmology and Visual Sciences, Case Western Reserve University and University Hospitals Eye Institute, Cleveland, Ohio, USA.
38. Department of Genetics, Case Western Reserve University, Cleveland, Ohio, USA.
39. Department of Ophthalmology, Medical Faculty Mannheim of the Ruprecht-Karls-University of Heidelberg, Mannheim, Germany.
40. Beijing Tongren Eye Center, Beijing Tongren Hospital, Beijing Institute of Ophthalmology, Beijing Key Laboratory of Ophthalmology and Visual Sciences, Capital Medical University, Beijing, China.
41. Department of Clinical Physiology, Tampere University Hospital and School of Medicine, University of Tampere, Tampere, Finland.
42. Finnish Cardiovascular Research Center, Faculty of Medicine and Life Sciences, University of Tampere, Tampere, Finland.
43. Department of Public Health, University of Helsinki, Helsinki, Finland.
44. Institute for Molecular Medicine Finland FIMM, HiLIFE Unit, University of Helsinki, Helsinki, Finland.
45. Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK.
46. Department of Ophthalmology, Radboud University Medical Center, Nijmegen, The Netherlands.
47. Department of Ophthalmology and Visual Sciences, University of Wisconsin–Madison, Madison, Wisconsin, USA.

48. Department of Clinical Chemistry, Finnish Cardiovascular Research Center-Tampere, Faculty of Medicine and Life Sciences, University of Tampere.
49. Department of Clinical Chemistry, Fimlab Laboratories, University of Tampere, Tampere, Finland.
50. National Human Genome Research Institute, National Institutes of Health, Baltimore, USA.
51. Genetic Epidemiology, QIMR Berghofer Medical Research Institute, Brisbane, Australia.
52. Department of Ophthalmology, Yokohama City University School of Medicine, Yokohama, Kanagawa, Japan.
53. Department of Ophthalmology and Visual Sciences, Kyoto University Graduate School of Medicine, Kyoto, Japan.
54. Department of Ophthalmology, University Medical Center of the Johannes Gutenberg University Mainz, Mainz, Germany.
55. Institute of Neurogenomics, Helmholtz Zentrum München, German Research Centre for Environmental Health, Neuherberg, Germany.
56. Department of Ophthalmology, Central Hospital of Central Finland, Jyväskylä, Finland.
57. Gerontology Research Center, Faculty of Sport and Health Sciences, University of Jyväskylä, Jyväskylä, Finland.
58. Program in Genetics and Genome Biology, Hospital for Sick Children and University of Toronto, Toronto, Ontario, Canada.
59. Gen-info Ltd, Zagreb, Croatia. .
60. University of Split School of Medicine, Soltanska 2, Split, Croatia.
61. Ulverscroft Vision Research Group, University College London, London, UK.
62. Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland.
63. Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland.
64. Myopia Research Group, Singapore Eye Research Institute, Singapore National Eye Centre, Singapore.
65. Saw Swee Hock School of Public Health, National University Health Systems, National University of Singapore, Singapore.
66. Department of Ophthalmology, University of Pennsylvania, Philadelphia, Pennsylvania, USA.
67. Department of Genetics, Genomics and Informatics, University of Tennessee Health Sciences Center, Memphis, Tenessee.
68. Department of Ophthalmology, University of Helsinki and Helsinki University Hospital, Helsinki, Finland.
69. Beijing Tongren Eye Center, Beijing Key Laboratory of Intraocular Tumor Diagnosis and Treatment, Beijing Ophthalmology & Visual Sciences Key Lab, Beijing Tongren Hospital, Capital Medical University, Beijing, China.
70. Department of Epidemiology and Medicine, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA.
71. Wilmer Eye Institute, Johns Hopkins Medical Institutions, Baltimore, Maryland, USA.
72. Department of Ophthalmology, Otsu Red Cross Hospital, Nagara, Japan. .
73. Centre for Myopia Research, School of Optometry, The Hong Kong Polytechnic University, Hong Kong, Hong Kong.
74. Department of Health Technology and Informatics, The Hong Kong Polytechnic University, Hong Kong, Hong Kong.
75. School of Ophthalmology and Optometry, Eye Hospital, Wenzhou Medical University, China.

3 23andMe

The following members of the 23andMe Research Team contributed to this study:

Michelle Agee

Babak Alipanahi

Adam Auton

Robert K. Bell

Katarzyna Bryc

Sarah L. Elson

Pierre Fontanillas

Nicholas A. Furlotte

David A. Hinds

Karen E. Huber

Aaron Kleinman

Nadia K. Litterman

Jennifer C. McCreight

Matthew H. McIntyre

Joanna L. Mountain

Elizabeth S. Noblin

Carrie A.M. Northover

Steven J. Pitts

J. Fah Sathirapongsasuti

Olga V. Sazonova

Janie F. Shelton

Suyash Shringarpure

Chao Tian

Joyce Y. Tung

Vladimir Vacic

Catherine H. Wilson.

4 Cohort Description

4.1 UK Biobank

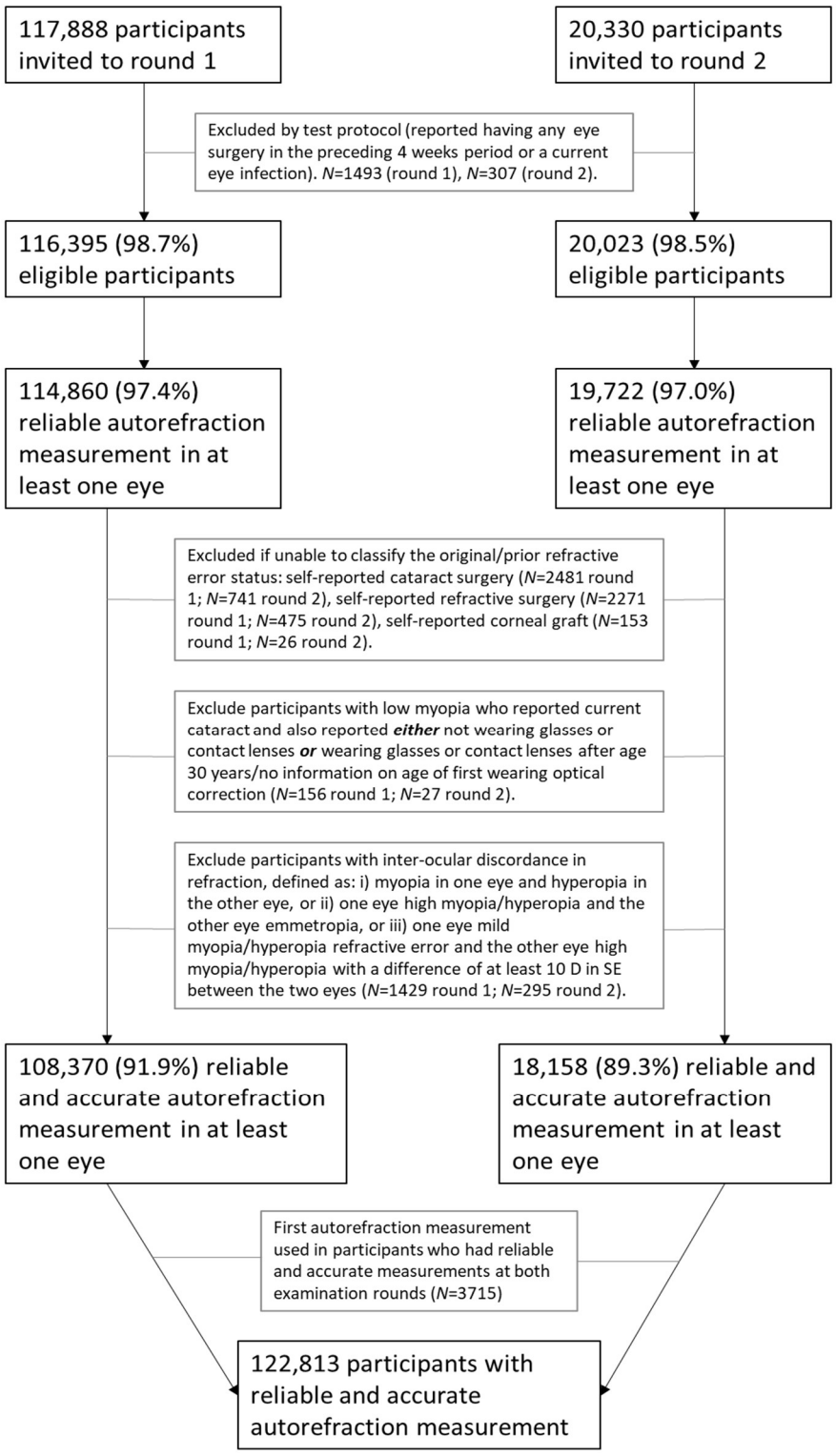
The UK Biobank is a very large multisite cohort study established by the Medical Research Council, Department of Health, Wellcome Trust medical charity, Scottish Government and Northwest Regional Development Agency. A baseline questionnaire, measurements, and biological samples were undertaken in 22 assessment centers across the UK between 2006 and 2010.

4.1.1 Phenotyping

4.1.1.1 Spherical equivalent (UK Biobank – 1)

Ophthalmic assessment was not part of the original baseline assessment and was introduced as an enhancement in 2009 for 6 assessment centers which are spread across the UK (Liverpool and Sheffield in North England, Birmingham in the Midlands, Swansea in Wales, and Croydon and Hounslow in Greater London). Participants completed a touch-screen self-administered questionnaire. The response options for ethnicity included White (English/Irish or other white background), Asian or British Asian (Indian/Pakistani/Bangladeshi or other Asian background), Black or Black British (Caribbean, African, or other black background), Chinese, mixed (White and Black Caribbean or African, White and Asian, or other mixed background), or other, non-defined, ethnic group.

Refractive error (RE) was measured by non-cycloplegic autorefraction in both eyes using the Tomey RC 5000 Auto Refkeratometer (Tomey Corp., Nagoya, Japan). The right eye was measured first and up to 10 measurements were taken per eye. The most representative result was automatically recorded. To ensure reliable and accurate RE data, we excluded participants based on previously published criteria[1] (Supplementary Note Figure 1). Spherical equivalent was calculated as spherical refractive error (UK Biobank codes 5084 and 5085) plus half the cylindrical error (UK Biobank 5086 and 5087) for each eye. If reliable data were only available for one eye, the RE of that eye was considered as the participant's RE. If reliable data were available for both eyes, we calculated the mean of right and left RE as the participant's RE.



Supplementary Note Figure 1. Flowchart explaining the exclusion of UK Biobank participants, as described elsewhere[1], for whom spherical equivalent measurements were available, based on existing clinical data.

A summary of the basic demographic characteristics of the subset of the UK Biobank for which the spherical equivalent was directly measured and which was used for our analysis is given below (Supplementary Note Table 1).

		Participants	**Spherical Equivalent (diopters)**	**Age (years)**
		N (%)	Mean ± SD	Mean ± SD
All		102,117	-0.28D (2.74)	57.26 (7.86)
Sex	**Male**	47,774	-0.27D (2.65)	57.59 (7.92)
	Female	54,343	-0.29D (2.82)	56.97(7.79)

Supplementary Note Table 1**.** Characteristics of UK Biobank participants for whom spherical equivalent measurements were available and included in the quantitative spherical equivalent analysis.
Abbreviations: N, number; SD, standard deviation; Age, age at first documented spherical equivalent assessment

4.1.1.2 Inferring myopia case-control status from self-reported age of spectacle wear (UKB2 sample)

Direct measurements of RE were only available for just 22.7% of the entire UK Biobank sample. However refractive error is strongly correlated with several clinical parameters and demographic factors[2], which can be used to predict refractive error and myopia. Some of that indirect information was present for significant numbers among the UK Biobank participants.

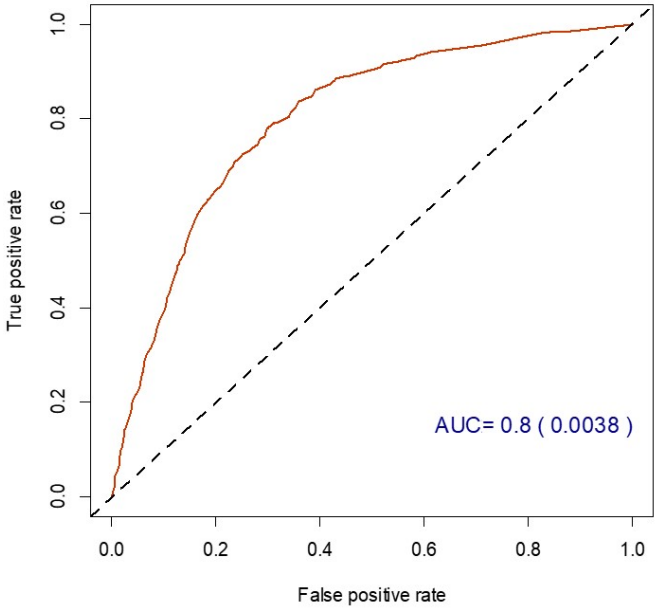
For example, age when the first lens correction is prescribed is strongly correlated to spherical equivalent both at a genetic[3] and phenotypic level[4]. It has used before as a proxy for RE previously[5] with reasonably low levels of genetic effect heterogeneity[6].A total of 87% of the participants, responded to the question on whether their vision needed correction when they were asked "Do you wear glasses or contact lenses to correct your vision?" (potential answers: Yes/No/Prefer not to answer) in a touch-screen self-administered questionnaire. If a participant answered "Yes" to this question, they were further asked "What age did you first start to wear glasses or contact lenses?", to which 67% of the participants responded. In addition, spherical equivalent and myopia affection status is highly correlated with age, sex, birth year[4,7] , all of which were available for UK Biobank participants.

We therefore aimed at harnessing the demographic and clinical information to obtain an estimate about the individual's likely myopia status. This general approach has been used successfully before[6], and to better classify the non-refracted subjects into myopia cases and non-myopia controls.

We proceeded in three steps: 1) training a Support Vector Machine model in 80% randomly selected UK Biobank participants of European descent for whom direct spherical equivalent and refractive error

status were available, 2) validate the prediction in the remaining 20% of UK Biobank participants of European descent for whom direct spherical equivalent and refractive error status were available, 3) make the SVM predictions in the remaining individuals with no direct spherical error measurements available using the model developed for the training data.

We initially fine-tuned the prediction model, in order to optimize the γ (gamma) and "cost" parameters in the 80% training data sample. The performance of the model was subsequently tested in the remaining 20% of the UK Biobank participants with spherical equivalent measurements. Receiver Operating Characteristic (ROC) curves were drawn and the area under the curve (AUC) calculated (Supplementary Note Figure 2). For the optimal SVM model, an Area Under the Curve (AUC) of 0.8 was obtained, providing strong evidence that myopia status could be inferred from 'age of first spectacle wear' with sufficient reliability to serve as a proxy for myopia status in a GWAS analysis. When we compared predicted vs observed myopia case-control status in the 20% validation subset of the UK Biobank for which spherical equivalent were available (cases defined as < -0.75 Diopters and controls > -0.5Diopters), the inferred myopia status was a strong predictor for actual (observed) myopia case-control status (OR=16.79, 95%CI 15.99-17.62).



Supplementary Note Figure 2. ROC curve and AUCs for the SVM model trained in 80% of fully phenotyped (i.e. spherical equivalent available) UK Biobank participants. For both ROC and AUC, the remaining 20% (i.e. not part of the initial training set) of the UK Biobank were used. Cases defined as < -0.75 Diopters and controls > -0.5Diopters. The standard error for the AUC is shown within brackets.

The optimal SVM model was applied to the UK Biobank sample with known 'age of first spectacle wear' but for whom no spherical equivalent measurements were available. That is, the GWAS for SVM-inferred myopia case-control status did not include participants with known (measured) refractive error. This inferred phenotypic status was used for a GWAS analysis, for which we followed the previously

described workflow (sections 4.1.3). In further support of SVM-inferred myopia case-control status as a valid proxy measure, the genetic effect sizes for variants assessed in the SVM-inferred phenotype were found to be highly correlated to those from the GWAS analyses for autorefractor-measured refractive error (Supplementary Note Table 2, Supplementary File 1**Error! Reference source not found.**).

	UKB1[1]	GERA[1]	23andMe[2,3]	UKB2[2,4]
Beta - UKB1	1.000	0.926	0.953	0.935
Beta - GERA	0.926	1.000	0.927	0.901
23andMe - log(OR)	0.953	0.927	1.000	0.941
UKB-2 log(OR)	0.935	0.901	0.941	1.000

Supplementary Note Table 2. Correlation of effects sizes between the UKB-2 subset (for which spherical equivalent was not available and myopia status was imputed using the SVM) with other cohorts in which refraction was directly measured or self-reported. "Beta" linear regression coefficients; log(OR) the logarithm of the logistic regression Odds Ratios; UKB-1 denotes the first subset of the UK Biobank participants (spherical equivalent available), UKB -2 the second subset (myopia case-control status inferred). For a description of the 23andMe and GERA cohorts, please refer to subsequent sections for further cohort descriptions.

The basic characteristics of the participants that were selected for the second UK Biobank subset (UKB-2) of our meta-GWAS are shown in the Supplementary Note Table 3.

		inferred cases/controls	**Age first spectacle wear (years)**	**Age (years)**
		N cases (%)	Mean ± SD	Mean ± SD
All		108,956/70,941 (60.56)	26.1 ± 13.58	56.61 ± 7.79
Sex	**Male**	45,994/30,388 (60.25)	26.38 ± 13.48	56.78 ± 7.88
	Female	62,962 / 40,553 (60.82)	25.74 ± 13.65	56.48 ± 7.73

Supplementary Note Table 3**.** Characteristics of UK Biobank participants included in the inferred myopia qualitative analysis (UKB-2). Abbreviations: N, number; SD, standard deviation; Age, age at first documented spherical equivalent assessment.

4.1.2 Genotyping

DNA extraction, genotyping and imputation of UK Biobank participants has been reported elsewhere[8]. DNA extraction begun on buffy coat samples. DNA was extracted from 850 µl buffy coat (recovered from 9 ml of whole blood) on customized TECAN Freedom EVO® 200 platform[9]. The samples were then processed in the approximate order received to produce genotype data. Genotyping was done using two arrays. The first array was the Affymetrix Axiom® platform with a custom-designed array described in the UK Biobank Axiom® Array Content Summary[10]. Processing was done using a LIMS system to track instrumentation, Axiom consumables arrays and reagents and operators. The process is described

elsewhere[11]. Details on genotyping procedure and quality control can be found elsewhere[12]. The second array is what was the UK BiLEVE, described elsewhere[13].

Phasing on the autosomes was carried out using a modified version of the SHAPEIT2[14] program modified to allow for very large sample sizes. This new method (which we refer to as SHAPEIT3) modifies the SHAPEIT2 surrogate family approach to remove a quadratic complexity component of the algorithm[15]. In small sample sizes of a few thousand samples, this part of the algorithm, which involves calculating Hamming distances between current haplotypes estimates, contributes only a relatively small part to the computational cost. As sample sizes increase over 10,000 samples then this component becomes significant. The new algorithm uses a divisive clustering algorithm to identify clusters of haplotypes, and then calculates Hamming distances only between pairs of haplotypes within each cluster. Only haplotypes within each cluster are used as candidates for the surrogate family copying states in the HMM model.

A total of 806,466 directly genotyped DNA sequence variants were available after variant quality control. The UK Biobank team then performed imputation from a combined Haplotype Reference Consortium (HRC) and UK10K reference panel; phasing was performed using SHAPEIT3 and imputation was carried out via the IMPUTE4 program[16]. Only HRC-imputed variants were used for the purpose our analyses of the UK Biobank participants. The variant-level quality control exclusion metrics applied to imputed data for GWAS included the following: call rate < 95%, Hardy–Weinberg equilibrium $P < 1 \times 10^{-6}$, posterior call probability < 0.9, imputation quality < 0.4, and MAF < 0.005. The Y chromosome and mitochondrial genetic data were excluded from this analysis. In total, 10,263,360 imputed DNA sequence variants were included in our analysis.

For sample quality control, we removed individuals of non-European ancestry and participants with relatedness corresponding to third-degree relatives or closer, and an additional 480 samples with an excess of missing genotype calls or more heterozygosity than expected were excluded. In total, genotypes were available for 102,117 participants of European ancestry with spherical equivalent data**.**

4.1.3 Association analyses

The basic model tested was the average of spherical equivalent measured in the left and right eye as an outcome of a regression model whose predictor is the allele dosage at a given polymorphic locus, adjusted for the effect of relevant covariables (see table below). The empirical association between spherical equivalent and other covariables is shown in Supplementary Note Table 4.

Effect size estimates (Beta) and P-values are from the multivariate regression model. Since demographic factors and principal components had a small yet real effect over Spherical Equivalent, the above variables were included in the model.
Therefore, models of mixed linear regressions, as described before[17], where the spherical equivalent was the outcome, the allele dosage the predictor, adjusted for age, sex and the first 10 principal components. Since there was, there is evidence of cryptic relatedness among the UK Biobank participants, a linear mixed model that controls for population structure was used as implemented in the Bolt-LMM software [18].

Variable	Beta	SE	Z	P-value
age	0.05247	0.001122	46.773	<E-308
sex	-0.08013	0.017988	-4.455	8.42E-06
PC1	0.000297	0.000125	2.367	0.017935
PC2	0.001096	0.000272	4.032	5.54E-05
PC3	0.001369	0.000516	2.655	0.007935
PC4	0.005147	0.000799	6.439	1.21E-10
PC5	0.003194	0.001253	2.55	0.010786
PC6	0.004701	0.001314	3.577	0.000348
PC7	0.008073	0.001614	5.002	5.68E-07
PC8	0.000269	0.001605	0.168	0.866871
PC9	0.008165	0.002385	3.424	0.000618
PC10	-0.00039	0.001885	-0.207	0.836117

Online Note Table 4. The association between spherical equivalent, age, sex and the first 10 Principal Components.

For the second UK Biobank subset, for which no spherical equivalent information was available, the mixed linear model was built with the predicted myopia status as outcome and using the same covariates as for the previously described linear regression analysis on spherical equivalent (paragraph 4.1.3). Odds Ratios were obtained from the beta regression coefficient using the equation:

$$\ln(\mathrm{OR}) = \frac{\beta}{\mu(1-\mu)}$$

where μ is the fraction of the cases in the sample (μ=0.606). Although the case-control analysis was quite balanced, we opted to remove genotypes with MAF <0.01 and MAC< 400 recommended elsewhere[19] (which in our samples, most often would be correspond to MAF < 0.001).

4.2 23andMe

4.2.1 Phenotyping

The subjects were all volunteers from the 23andMe (Mountain View, CA, USA) personal genomics company. All participants included in the analyses provided informed consent and answered surveys online according to the approved 23andMe human subjects protocol, which was reviewed and approved by Ethical & Independent Review Services, a private institutional review board (http://www.eandireview.com). The participants were identified as myopia cases if they responded positively to any of the following questions:

1. "Have you ever been diagnosed by a doctor with nearsightedness (near objects are clear, far objects are blurry)?"
2. "Are you nearsighted (near objects are clear, far objects are blurry)?"
3. "What vision problems do you have? Please check all that apply." - Nearsightedness (near objects are clear, far objects are blurry.
4. "Prior to your LASIK eye surgery, what vision problems did you have? Please check all that apply." - Nearsightedness (near objects are clear, far objects are blurry.

Controls were defined as having said "No" or not checking nearsightedness to at least one of the questions above. Subjects who gave discordant answers were removed.

4.2.2 Genotyping

DNA extraction and genotyping were performed on saliva samples by CLIA-certified and CAP-accredited clinical laboratories of Laboratory Corporation of America. Samples were genotyped on one of four genotyping platforms. The V1 and V2 platforms were variants of the Illumina HumanHap550+ BeadChip, including about 25,000 custom SNPs selected by 23andMe, with a total of about 560,000 SNPs. The V3 platform was based on the Illumina OmniExpress+ BeadChip., with custom content to improve the overlap with our V2 array, with a total of about 950,000 SNPs. The V4 platform in current use is a fully custom array, including a lower redundancy subset of V2 and V3 SNPs, with additional coverage of lower-frequency coding variation, and about 570,000 SNPs. Samples that failed to reach 98.5% call rate were re-analyzed. For the GWAS only participants who have >97% European ancestry, as determined through an analysis of local ancestry, were included. For the purposes of ethnic categorization, an algorithm first partitioned phased genomic data into short windows of about 100 SNPs and used a support vector machine (SVM) to classify individual haplotypes into one of 31 reference populations. The SVM classifications then fed into a hidden Markov model (HMM) that accounts for switch errors and incorrect assignments and gives probabilities for each reference population in each window. The reference population data are derived from public datasets (the Human Genome Diversity Project, HapMap, and 1000 Genomes), as well as 23andMe customers who have reported having four grandparents from the same country. A maximal set of unrelated individuals was chosen for each analysis using a segmental identity-by-descent (IBD) estimation algorithm[20]. Individuals were defined as related if they shared more than 700 cM IBD, including regions where the two individuals share either one or both genomic segments identical-by-descent. This level of relatedness corresponds approximately to the minimal expected sharing between first cousins in an outbred population.

Participant genotype data were imputed against the September 2013 release of 1000 Genomes Phase1 reference haplotypes, phased with ShapeIt2[21]. We phased and imputed data for each genotyping platform separately. We phased using an internally developed phasing tool which implements the Beagle haplotype graph-based phasing algorithm[22].

SNPs with Hardy-Weinberg equilibrium $P<10^{-20}$, call rate < 95%, or with large allele frequency discrepancies compared to European 1000 Genomes reference data were excluded from imputation. Imputation was done against all-ethnicity 1000 Genomes haplotypes (excluding monomorphic and singleton sites) using Minimac[23]. For the X chromosome, separate haplotype graphs were built for the non-pseudoautosomal region and each pseudoautosomal region, and these regions were phased separately. Males and females were imputed together using Minimac2[23], as with the autosomes, treating males as homozygous pseudo-diploids for the non-pseudoautosomal region.

HLA allele dosages were imputed from SNP genotype data using HIBAG[24]. We imputed alleles for HLA-A, B, C, DPB1, DQA1, DQB1, and DRB1 loci at four-digit resolution. To test associations between HLA allele dosages and phenotypes, we performed logistic or linear regression using the same set of covariates used in the SNP-based GWAS for that phenotype. We performed separate association tests for each imputed allele.

4.2.3 Association analyses

Association test results were computed by linear regression assuming additive allelic effects. For tests imputed dosages rather than best-guess genotypes were used. Covariates for age, gender, the first ten principal components to account for residual population structure were also included into the model. Results for the X chromosome are computed similarly, with male genotypes coded as if they were homozygous diploid for the observed allele.

4.3 GERA

The Genetic Epidemiology Research in Adult Health and Aging (GERA) cohort is part of the Kaiser Permanente Research Program on Genes, Environment, and Health (RPGEH) and has been described in detail elsewhere[25,26]. The GERA cohort comprises 110,266 adult men and women who are consented participants in the RPGEH, an unselected cohort of adult participants who are members of Kaiser Permanente Northern California (KPNC), an integrated health care delivery system, with ongoing longitudinal records from vision examinations. For this analysis, 34,998 adults (25 years and older), who self-reported as non-Hispanic white, and who had at least one assessment of spherical equivalent obtained between 2008 and 2014 were included (Supplementary Note Table 5). All study procedures were approved by the Institutional Review Board of the Kaiser Foundation Research Institute.

		Participants	**Spherical Equivalent (diopters)**	**Age (years)**
		N (%)	Mean ± SD	Mean ± SD
All		34,998 (100)	-0.35 ± 2.56	66.54 ± 11.55
Sex	**Male**	14,431 (41.23)	-0.32 ± 2.46	68.84 ± 10.70
	Female	20,567 (58.77)	-0.38 ± 2.64	64.93 ± 11.84

Supplementary Note Table 5**.** Characteristics of GERA non-Hispanic white subjects included in the GWAS of spherical equivalent by sex. Abbreviations: N, number; SD, standard deviation; Age, age at first documented spherical equivalent assessment

4.3.1 Phenotyping

All participants underwent vision examinations, and most subjects had multiple measures for both eyes. Spherical equivalent was assessed as the sphere + cylinder/2. For this analysis, spherical equivalent was selected from the first documented assessment, and the mean of both eyes was used. As previously described[27], individuals with histories of cataract surgery (in either eye), refractive surgery, keratitis, or corneal diseases were excluded.

4.3.2 Genotyping

DNA samples from GERA individuals were extracted from Oragene kits (DNA Genotek Inc., Ottawa, ON, Canada) at KPNC and genotyped at the Genomics Core Facility of the University of California, San Francisco (UCSF). DNA samples were genotyped at over 665,000 single nucleotide polymorphisms (SNPs) on Affymetrix Axiom arrays (Affymetrix, Santa Clara, CA, USA)[28,29]. SNPs with initial genotyping call rate ≥97%, allele frequency difference ≤0.15 between males and females for autosomal markers, and genotype concordance rate >0.75 across duplicate samples were included[26]. Around 94% of samples and more than 98% of genetic markers assayed passed quality control (QC) procedures. In addition to those QC criteria,

SNPs with genotype call rates <90% were removed, as well as SNPs with a minor allele frequency (MAF) < 1%.
Following genotyping QC, we conducted statistical imputation of additional genetic variants. Following the pre-phasing of genotypes with Shape-IT v2.r72719[30], variants were imputed from the cosmopolitan 1000 Genomes Project reference panel (phase I integrated release; http://1000genomes.org) using IMPUTE2 v2.3.0.[31-33] As a QC metric, we used the info r^2 from IMPUTE2, which is an estimate of the correlation of the imputed genotype to the true genotype[34]. Variants with an imputation $r^2 < 0.3$ were excluded, and we restricted to SNPs that had a minor allele count (MAC) ≥ 20.

4.3.3 Association analyses

A linear regression of each individual's spherical equivalent was performed with the following covariates: age at first documented spherical equivalent assessment, sex, and genetic principal components. A linear regression of the residuals on each SNP was then performed using PLINK[35] v1.9 (www.cog-genomics.org/plink/1.9/) to assess genetic associations. Data from each SNP were modeled using additive dosages to account for the uncertainty of imputation[36]. Eigenstrat[37] v4.2 was used to calculate the PCs[25]. The top 10 ancestry PCs were included as covariates, as well as the percentage of Ashkenazi ancestry to adjust for genetic ancestry, as described previously[25].

4.4 Cream Consortium

4.4.1 Phenotyping

All participants included in this analysis from CREAM were 25 years of age or older. RE was represented by measurements of refraction and spherical equivalent (SphE = spherical refractive error +1/2 cylinder refractive error) was the outcome variable for CREAM. Participants with conditions that might alter refraction, such as cataract surgery, laser refractive procedures, retinal detachment surgery, keratoconus, or ocular or systemic syndromes were excluded from the analyses. Recruitment and ascertainment strategies varied by study and were previously published elsewhere[6].

4.4.2 Genotyping

The genotyping process has been described elsewhere[6]. Samples were genotyped on different platforms, and study-specific QC measures of the genotyped variants were implemented before association analysis. Genotypes were imputed with the appropriate ancestry-matched reference panel for all cohorts from the 1000 Genomes Project (Phase I version 3, March 2012 release) with either minimac[23] or IMPUTE[16]. The metrics for preimputation QC varied among studies, but genotype call-rate thresholds were set at a high level (≥0.95). These metrics were similar to those described in a previous GWAS analyses[38]; detailed information for each cohort is described elsewhere[6].

4.4.3 Association analyses

To prevent overlap of samples, cohorts from the United Kingdom (1985BBC, ALSPAC-Mothers, EPIC-Norfolk, ORCADES and Twins UK) were excluded from the GWAS meta-analysis. Association analyses were performed following the workflow elsewhere[6]: All samples analyzed were of European descent, for each CREAM cohort, a single-marker analysis for the phenotype of Spherical equivalent (in diopters) was carried out with linear regression with adjustment for age, sex and up to the first five principal components. For all non-family-based cohorts, one of each pair of relatives was removed (after detection through either GCTA or identity by sequence (IBS)/identity by descent (IBD) analysis). In

family-based cohorts, a score test-based association was used to adjust for within-family relatedness. We used an additive SNP allelic-effect model.

4.5 EPIC

The European Prospective Investigation into Cancer (EPIC) study is a pan-European prospective cohort study designed to investigate the etiology of major chronic diseases[39]. EPIC-Norfolk , one of the UK arms of EPIC, recruited and examined 25,639 participants between 1993 and 1997 for the baseline examination[40]. Recruitment was via general practices in the city of Norwich and the surrounding small towns and rural areas, and methods have been described in detail previously[41]. Since virtually all residents in the UK are registered with a general practitioner through the National Health Service, general practice lists serve as population registers. Ophthalmic assessment formed part of the third health examination and this has been termed the EPIC-Norfolk Eye Study[42]. In total, 8,623 participants were seen for the Eye Study, between 2004 and 2011. The EPIC-Norfolk Eye Study was carried out following the principles of the Declaration of Helsinki and the Research Governance Framework for Health and Social Care. The study was approved by the Norfolk Local Research Ethics Committee (05/Q0101/191) and East Norfolk & Waveney NHS Research Governance Committee (2005EC07L). All participants gave written, informed consent.

4.5.1 Phenotyping

Refractive error was measured in both eyes using a Humphrey Auto-Refractor 500 (Humphrey Instruments, San Leandro, California, USA). Spherical equivalent was calculated as spherical refractive error plus half the cylindrical error for each eye.
Some basic demographic and clinical information about the samples used for the validation analyses is given below (Supplementary Note Table 6).

		Participants	**Spherical Equivalent (diopters)**	**Age (years)**
		N (%)	Mean ± SD	Mean ± SD
All		7,117 (100)	+0.16 ± 2.25	68.80 ± 8.18
Sex	**Male**	3,253 (45.71)	+0.15 ± 2.23	69.60 ± 8.20
	Female	3,864 (54.29)	+0.18 ± 2.27	68.13 ± 8.11

Supplementary Note Table 6**.** Characteristics of the participants in the EPIC-Norfolk cohort, included in the heritability and risk prediction analyses. Abbreviations: N, number; SD, standard deviation; Age, age at first documented spherical equivalent assessment

4.5.2 Genotyping and imputation

Genotypes obtained using the Affymetrix UK Biobank Axiom Array on 7,117 subjects contributed to the current study were excluded if they had low call rates, poor clustering, batch effects across genotyping plates and/or Hardy-Weinberg equilibrium $P < 10^{-7}$. Samples were excluded on grounds of poor genotyping across all SNPs, sex discordance , exces or low heterozygosity and unexplainable identity-by-descent values. Third-degree relatives or closer participants were also removed. Data were pre-phased using SHAPEIT[14] version 2 and imputed to the Phase 3 build of the 1000 Genomes project[43] (October 2014) using IMPUTE[16] version 2.3.2.

4.5.3 Association analysis

We examined the relationship between allele dosage and mean spherical equivalent using linear regression adjusted for age, sex and the first 5 principal components. Analyses were carried out using SNPTEST version 2.5.1.

5 STATISTICAL ANALYSES

5.1 Meta-analyses

For all meta-analyses we applied a Z-score method, weighted by the effective population sample size, as implemented in METAL[44]. No genomic control adjustment was applied during the meta-analysis.

5.2 Conditional analyses

The conditional and joint analysis on summary data (COJO) [45] as implemented in the GCTA program [46] was used to identify independent effects within associated loci as well as the calculation of the phenotypic variance explained[47] by all polymorphisms associated with the trait after the conditional analyses. Default parameters were used for the analysis. The LD estimates were derived from a randomly selected sample of 10,000 unrelated subjects the UK Biobank cohort.

5.3 Multiple testing correction

Two methods of correcting for multiple testing were used. The first was a classic Bonferroni correction, in which the threshold of significance (0.05) was divided by the number of tests (n):

$$\alpha = \frac{0.05}{n}$$

Given the large number of loci for which replication was needed, we additionally calculated the False Discovery Rates, using the Benjamini-Hochberg method[48].

5.4 Genomic inflation

To assess the potential inflation of association probabilities, genomic inflation factors[49] were calculated and Q-Q plots were drawn using the package 'gap' in R (https://cran.r-project.org/).

5.5 LDscore regression-based methods

5.5.1 Polygenicity vs. inflation

To distinguish between the effect of polygenicity and those arising from sample stratification or uncontrolled population admixture, we followed previously suggested approaches[50] to calculate the LD score regression intercepts using the program LD Score (https://github.com/bulik/ldsc).

5.5.2 Calculation of genetic correlation

Bivariate genetic correlations between refractive error and other complex traits whose summary statistics are publicly available were assessed following previously described methodologies[51], using the program LD Score (https://github.com/bulik/ldsc).

5.6 Associated SNPs and gene annotations

Polymorphisms associated at a GWAS level (P<5x10^{-08}) were clustered within an "associated genomic region", defined as a contiguous genomic region where GWAS-significant markers were within 1 million base pairs from each other, as suggested elsewhere[52]. Significant polymorphisms were annotated with the gene inside whose transcript-coding region they are located, or alternatively, if located between two genes, with the gene nearest to it. The associated genomic regions were collectively annotated with the gene overlapping, or nearest the most significantly associated variant within that region. In addition, the polymorphic sites were functionally annotated using SNPnexus[53].

5.6.1 OMIM

The Online Mendelian Inheritance In Man (OMIM) is a continuously curated catalog of human genes and phenotypic changes their polymorphic forms cause in humans[54]. This catalogue contains a still partial, but highly reliable list of gene-phenotype pairs and was used retrieve data that could inform about the functionality of specific genes with particular focus on phenotypic expressions of extremely penetrant mutations.

5.6.2 The GWAS Catalog.

Previous GWAS association of SNPs or genes with other phenotypic traits was conducted through queries of the GWAS Catalog[55]. Results were downloaded from the official site hosted at the European Bioinformatics Institute: https://www.ebi.ac.uk/gwas/downloads.

5.7 Graphical illustration of association

LocusZoom[56] was used to generate plot that visualize regional association and its genomic context. Data from the European participants in the 1000 Genome Project, November 2014 was used, and the graphs were generated using the online LocusZoomserver (http://locuszoom.org/).

5.8 Mendelian randomization

The R (https://cran.r-project.org) package MendelianRandomization v3.4.4 was used for Mendelian randomization analyses.

5.9 Gene expression, GTEx and other transcription data

We obtained data on tissue expression from several sources for genes that map within RE associated loci defined as described before (section **Error! Reference source not found.**). Information about the expression of the genes of interest in systemic (i.e. non-ocular) tissues was obtained from the GTEx Portal for GTEx release v7 (https://gtexportal.org/home/datasets). RNA sequencing data was obtained

for both fetal and adult corneal, trabecular meshwork and ciliary body, as described elsewhere[57], which we downloaded from the authors' supplementary information. In addition, we extracted data from the subset of subjects with presumed healthy adult retinas (AMD=1), described elsewhere[58] that obtained from the GTEx Portal (https://gtexportal.org/home/datasets).

Transcription data was processed using different platforms and were available in different units (Transcripts per Million bases, TPM, for the retina and GTEx tissues, and Fragments per Kilobase, FPKM for the other tissues). For purposes of comparing expression across different tissues for which different methodologies may have been used, expression levels for all tissues were rank-transformed. Hierarchical clustering was used to help visualize similarities and differences of patterns of transcript expression across different tissues ('hclust' package in R).

5.10 LD score regression applied to specifically expressed genes (LDSC-SEG)

Disease-relevant tissues and cell types were identified by analyzing gene expression data together with summary statistics from the meta-analysis of refractive error in all five cohorts, as described elsewhere[59]. Briefly, genes were ranked based on the t-statistic of their expression in each tisue and the 10% most expressed genes for each tissue were considered "specifically expressed genes". A stratified LD score regression was applied to the meta-analysis summary statistics to evaluate the contribution of the focal genome annotation to trait heritability.

5.11 SMR

SMR (Summary data–based Mendelian randomization) assesses the relationship between genetic variant, intermediate variables such as gene expression levels or methylation levels as mediating traits, to test causality on a specific phenotype[60].

5.11.1 Test description

The SMR package helps perform two tests. The first is an SMR test, which correlates GWAS effects with eQTL or methylation effects (or any other intermediate trait)[60]. This test suggests causation, although it is unable to fully differentiate between it and pleiotropy. The second test is that of Heterogeneity in Dependent Instrument (HEIDI). This test against the null hypothesis that changes in both eQTL (or other intermediary traits) and the phenotype of interest are caused by one single SNP, which is therefore considered as the candidate for the putative causal effect.

5.11.2 Datasets for the SMR analyses: eQTL, *cis*-mQTL

To perform the above-mentioned tests of causation/pleiotropy, we used three different datasets of association between genetic variants and intermediate traits. The first was the summary statistics of eQTL associations in the untransformed peripheral blood samples of 5,311 subjects[61]. There were two advantages in using these data: 1) this was the largest eQTL dataset available and 2) the use of a highly heterogeneous tissue such as peripheral blood would be more likely than any other single more homogeneous tissue to overcome any heterogeneity of eQTL effects with eye and retinal tissues that were unavailable at the time of the analysis and manuscript writing.

Assuming that tissues relevant to the development of refractive error are similar to the brain, we also used two datasets, one with eQTL effects and the other with results of a *cis*- methylation analysis (*cis*-mQTL), both in brain tissues[62].

5.12 Gene-set enrichment

To identify pathways or other gene sets that were over-represented among our results, we used a Gene-Set Enrichment Analysis (GSEA) as implemented in the Meta-Analysis Gene Set Enrichment of Variant (MAGENTA) software[63]. This program assigns scores to each gene based on the strength of association with refractive error, adjusting for potential confounders such as gene length and linkage disequilibrium. Enrichment for any gene set was assessed within genes above the cut-off of the highest 75th centile of significant gene scores. For the current study, the most recent versions of Gene Ontology (GO), Panther, KGG, Biocarta and MSigDB databases were used. We also carried out a similar enrichment analysis for the presence of transcription factor binding sites. A permutational procedure and false-discovery rates were used to calculate significance of enrichment and control for multiple testing.

5.12.1 GSEA definitions

For the enrichment analyses we used updated versions of the GSEA gene sets as described before[64]. We used the versions from September 2017 which were downloaded from:
http://software.broadinstitute.org/gsea/login.jsp

5.13 Analyses of signals of natural selection

Results of three statistical tests for natural selection were imported from the 1000 Genomes Selection Browser[65]. We downloaded and reported results from several tests such as iHS[66] and a cross-population comparison, XP-EHH, based on extended haplotype homozygosity test (average and maximum CEU, CEU vs YRI) [67] and the Tajima's D test. The absolute test scores and the rank scores (-log10 of the centile of the absolute test score across the genome) were reported.

5.14 Estimation of effect size distributions for spherical equivalent

We used a maximum-likelihood model to estimate the distribution of effect sizes, based on summary statistics of observations and linkage disequilibrium patterns to predict the likely number of SNPs that explain spherical equivalent heritability as well as explore the relationship between future sample sizes and the number of SNPs identified and variance or heritability explained as described elsewhere[68] and implemented in the GENESIS R package (https://github.com/yandorazhang/GENESIS).

6 FULL ACKNOWLEDGEMENTS

6.1 UK Biobank

UK Biobank was established by the Wellcome Trust medical charity, Medical Research Council, Department of Health, Scottish Government, and Northwest Regional Development Agency. It also had funding from the Welsh Assembly Government, British Heart Foundation, and Diabetes UK.

Development of the eye and vison dataset was led by Prof Sir Peng T Khaw and Prof Paul Foster, funded by a grant from The National Institute for Health Research (NIHR) to Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology for a biomedical research centre.

6.2 23andMe

23andMe would like to thank the research participants and employees of 23andMe for making this work possible.

6.3 GERA

GERA are grateful to the Kaiser Permanente Northern California members who have generously agreed to participate in the Kaiser Permanente Research Program on Genes, Environment, and Health. Support for participant enrollment, survey completion, and biospecimen collection for the RPGEH was provided by the Robert Wood Johnson Foundation, the Wayne and Gladys Valley Foundation, the Ellison Medical Foundation, and Kaiser Permanente Community Benefit Programs. Genotyping of the GERA cohort was funded by a grant from the National Institute on Aging, National Institute of Mental Health, and National Institute of Health Common Fund (RC2 AG036607). Data analyses were facilitated by National Eye Institute (NEI) grant R01 EY027004 (E.J.), National Institute of Diabetes and Digestive and Kidney Diseases grant R01 DK116738 (E.J.).

6.4 Consortium for Refractive Error and Myopia (CREAM)

CREAM investigators are gratefully thank all study participants, their relatives and the staff at the recruitment centers for their invaluable contributions. Funding for this particular GWAS mega-analysis was provided by the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Program (grant 648268), the Netherlands Organisation for Scientific Research (NWO, grant 91815655) and the National Eye Institute (grant R01EY020483). VJMV acknowledges funding from the Netherlands Organisation for Scientific Research (NWO, grant 91815655 & 91617076).

6.5 EPIC-Norfolk

EPIC-Norfolk infrastructure and core functions are supported by grants from the Medical Research Council (G1000143) and Cancer Research UK (C864/A14136). The clinic for the third health examination was funded by Research into Ageing (262). Genotyping was funded by the Medical Research Council (MC_PC_13048). We thank all staff from the MRC Epidemiology laboratory team for the preparation and quality control of DNA samples. Mr. Khawaja is supported by a Moorfields Eye Charity grant. Professor Foster has received additional support from the Richard Desmond Charitable Trust (via Fight for Sight) and the Department for Health through the award made by the National Institute for Health Research to

Moorfields Eye Hospital and the UCL Institute of Ophthalmology for a specialist Biomedical Research Centre for Ophthalmology.

6.6 King's College London authors

M.J. Simcoe is a recipient of a Fight for Sight PhD studentship. K. Patasova is a recipient of a Fight for Sight PhD studentship. P.G. Hysi the recipient of a FfS ECI fellowship. P.G. Hysi and C.J. Hammond acknowledge the TFC Frost Charitable Trust Support for the KCL Department of Ophthalmology. The statistical analyses were run in King's College London Rosalind HPC LINUX Clusters and cloud server. The UK Biobank data was accessed as part of the UK Biobank projects 669 and 17615.

6.7 UK Biobank Eye Consortium members

Data analyses were carried out using the RAVEN computing cluster, maintained by the ARCCA group. UK Biobank data was accessed as part of the UK Biobank Project 17615. Veronique Vitart is supported by core grant MC_UU_00007/10 from the UK Medical Research Council. The UK Biobank Eye and Vision Consortium has been supported by grants from NIHR (BRC3_026), Moorfields Eye Charity (ST 15 11 E), Fight for Sight (1507/1508), The Macular Society, The International Glaucoma Association (IGA, Ashford UK) and Alcon Research Institute.

6.8 University College London authors

Jugnoo Rahi is a National Institute for Health Research (NIHR) Senior Investigator, supported by the NIHR Biomedical Research Centres at Moorfields Eye Hospital/UCL Institute of Ophthalmology, and at the UCL Institute of Child Health/Great Ormond Street Hospital. Views expressed are those of the authors and not necessarily those of the NHS/NIHR. Phillippa Cumberland was funded by the Ulverscroft Foundation. Omar Mahroo is supported by Wellcome Trust grant 206619_Z_17_Z and the NIHR Biomedical Research Centre at Moorfields Eye Hospital and the UCL Institute of Ophthalmology. The UK Biobank data was accessed as part of the UK Biobank projects 669 and 17615.

References:

1. Cumberland, P.M. *et al.* Frequency and Distribution of Refractive Error in Adult Life: Methodology and Findings of the UK Biobank Study. *PLoS One* **10**, e0139780 (2015).
2. Williams, K.M. *et al.* Early life factors for myopia in the British Twins Early Development Study. *British Journal of Ophthalmology*, bjophthalmol-2018-312439 (2018).
3. Wojciechowski, R. & Hysi, P.G. Focusing in on the complex genetics of myopia. *PLoS Genet* **9**, e1003442 (2013).
4. Williams, K.M. *et al.* Age of myopia onset in a British population-based twin cohort. *Ophthalmic Physiol Opt* **33**, 339-45 (2013).
5. Kiefer, A.K. *et al.* Genome-wide analysis points to roles for extracellular matrix remodeling, the visual cycle, and neuronal development in myopia. *PLoS Genet* **9**, e1003299 (2013).
6. Tedja, M.S. *et al.* Genome-wide association meta-analysis highlights light-induced signaling as a driver for refractive error. *Nat Genet* **50**, 834-848 (2018).
7. Williams, K.M. *et al.* Increasing Prevalence of Myopia in Europe and the Impact of Education. *Ophthalmology* **122**, 1489-97 (2015).
8. Bycroft, C. *et al.* Genome-wide genetic data on~ 500,000 UK Biobank participants. *bioRxiv*, 166298 (2017).
9. http://www.ukbiobank.ac.uk/wp-content/uploads/2014/04/DNA-Extraction-at-UK-Biobank-October-2014.pdf.
10. http://www.ukbiobank.ac.uk/wp-content/uploads/2014/04/UK-Biobank-Axiom-Array-Content-Summary-2014.pdf.
11. http://biobank.ctsu.ox.ac.uk/crystal/refer.cgi?id=155583.
12. https://biobank.ctsu.ox.ac.uk/crystal/docs/impute_ukb_v1.pdf.
13. Wain, L.V. *et al.* Novel insights into the genetics of smoking behaviour, lung function, and chronic obstructive pulmonary disease (UK BiLEVE): a genetic association study in UK Biobank. *Lancet Respir Med* **3**, 769-81 (2015).
14. Delaneau, O., Marchini, J. & Zagury, J.F. A linear complexity phasing method for thousands of genomes. *Nat Methods* **9**, 179-81 (2011).
15. O'Connell, J. *et al.* Haplotype estimation for biobank-scale data sets. *Nat Genet* **48**, 817-20 (2016).
16. Howie, B.N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* **5**, e1000529 (2009).
17. Loh, P.R. *et al.* Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat Genet* **47**, 284-90 (2015).
18. Loh, P.-R., Kichaev, G., Gazal, S., Schoech, A.P. & Price, A.L. Mixed-model association for biobank-scale datasets. *Nature genetics*, 1 (2018).
19. Ma, C., Blackwell, T., Boehnke, M., Scott, L.J. & Go, T.D.i. Recommended joint and meta-analysis strategies for case-control association testing of single low-count variants. *Genet Epidemiol* **37**, 539-50 (2013).
20. Henn, B.M. *et al.* Cryptic distant relatives are common in both isolated and cosmopolitan genetic samples. *PLoS One* **7**, e34267 (2012).
21. Delaneau, O., Zagury, J.F. & Marchini, J. Improved whole-chromosome phasing for disease and population genetic studies. *Nat Methods* **10**, 5-6 (2013).
22. Browning, S.R. & Browning, B.L. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am J Hum Genet* **81**, 1084-97 (2007).

23. Fuchsberger, C., Abecasis, G.R. & Hinds, D.A. minimac2: faster genotype imputation. *Bioinformatics* **31**, 782-4 (2015).
24. Zheng, X. *et al.* HIBAG--HLA genotype imputation with attribute bagging. *Pharmacogenomics J* **14**, 192-200 (2014).
25. Banda, Y. *et al.* Characterizing Race/Ethnicity and Genetic Ancestry for 100,000 Subjects in the Genetic Epidemiology Research on Adult Health and Aging (GERA) Cohort. *Genetics* **200**, 1285-95 (2015).
26. Kvale, M.N. *et al.* Genotyping Informatics and Quality Control for 100,000 Subjects in the Genetic Epidemiology Research on Adult Health and Aging (GERA) Cohort. *Genetics* **200**, 1051-60 (2015).
27. Shen, L. *et al.* The Association of Refractive Error with Glaucoma in a Multiethnic Population. *Ophthalmology* **123**, 92-101 (2016).
28. Hoffmann, T.J. *et al.* Next generation genome-wide association tool: design and coverage of a high-throughput European-optimized SNP array. *Genomics* **98**, 79-89 (2011).
29. Hoffmann, T.J. *et al.* Design and coverage of high throughput genotyping arrays optimized for individuals of East Asian, African American, and Latino race/ethnicity using imputation and a novel hybrid SNP selection algorithm. *Genomics* **98**, 422-30 (2011).
30. Delaneau, O., Marchini, J. & Zagury, J.F. A linear complexity phasing method for thousands of genomes. *Nature methods* **9**, 179-81 (2012).
31. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. & Abecasis, G.R. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nature genetics* **44**, 955-9 (2012).
32. Howie, B., Marchini, J. & Stephens, M. Genotype imputation with thousands of genomes. *G3* **1**, 457-70 (2011).
33. Howie, B.N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS genetics* **5**, e1000529 (2009).
34. Marchini, J. & Howie, B. Genotype imputation for genome-wide association studies. *Nature reviews. Genetics* **11**, 499-511 (2010).
35. Chang, C.C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* **4**, 7 (2015).
36. Huang, L., Wang, C. & Rosenberg, N.A. The relationship between imputation error and statistical power in genetic association studies in diverse populations. *American journal of human genetics* **85**, 692-8 (2009).
37. Price, A.L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nature genetics* **38**, 904-9 (2006).
38. Verhoeven, V.J. *et al.* Genome-wide meta-analyses of multiancestry cohorts identify multiple new susceptibility loci for refractive error and myopia. *Nat Genet* **45**, 314-8 (2013).
39. Riboli, E. & Kaaks, R. The EPIC Project: rationale and study design. European Prospective Investigation into Cancer and Nutrition. *Int J Epidemiol* **26 Suppl 1**, S6-14 (1997).
40. Day, N. *et al.* EPIC-Norfolk: study design and characteristics of the cohort. European Prospective Investigation of Cancer. *Br J Cancer* **80 Suppl 1**, 95-103 (1999).
41. Hayat, S.A. *et al.* Cohort profile: A prospective cohort study of objective physical and cognitive capability and visual health in an ageing population of men and women in Norfolk (EPIC-Norfolk 3). *Int J Epidemiol* **43**, 1063-72 (2014).
42. Khawaja, A.P. *et al.* The EPIC-Norfolk Eye Study: rationale, methods and a cross-sectional analysis of visual impairment in a population-based cohort. *BMJ Open* **3**(2013).

43. Delaneau, O., Marchini, J., Genomes Project, C. & Genomes Project, C. Integrating sequence and array data to create an improved 1000 Genomes Project haplotype reference panel. *Nat Commun* **5**, 3934 (2014).
44. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-1 (2010).
45. Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet* **44**, 369-75, S1-3 (2012).
46. Yang, J., Lee, S.H., Goddard, M.E. & Visscher, P.M. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* **88**, 76-82 (2011).
47. Yang, J. *et al.* Common SNPs explain a large proportion of the heritability for human height. *Nat Genet* **42**, 565-9 (2010).
48. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)*, 289-300 (1995).
49. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997-1004 (1999).
50. Bulik-Sullivan, B.K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* **47**, 291-5 (2015).
51. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat Genet* **47**, 1236-41 (2015).
52. Wood, A.R. *et al.* Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat Genet* **46**, 1173-86 (2014).
53. Dayem Ullah, A.Z., Lemoine, N.R. & Chelala, C. SNPnexus: a web server for functional annotation of novel and publicly known genetic variants (2012 update). *Nucleic Acids Res* **40**, W65-70 (2012).
54. McKusick, V.A. Mendelian Inheritance in Man and its online version, OMIM. *Am J Hum Genet* **80**, 588-604 (2007).
55. Buniello, A. *et al.* The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res* **47**, D1005-D1012 (2019).
56. Pruim, R.J. *et al.* LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* **26**, 2336-7 (2010).
57. Carnes, M.U., Allingham, R.R., Ashley-Koch, A. & Hauser, M.A. Transcriptome analysis of adult and fetal trabecular meshwork, cornea, and ciliary body tissues by RNA sequencing. *Exp Eye Res* **167**, 91-99 (2018).
58. Ratnapriya, R. *et al.* Retinal transcriptome and eQTL analyses identify genes associated with age-related macular degeneration. *Nat Genet* **51**, 606-610 (2019).
59. Finucane, H.K. *et al.* Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat Genet* **50**, 621-629 (2018).
60. Zhu, Z. *et al.* Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet* **48**, 481-7 (2016).
61. Westra, H.J. *et al.* Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet* **45**, 1238-1243 (2013).
62. Qi, T. *et al.* Identifying gene targets for brain-related traits using transcriptomic and methylomic data from blood. *Nat Commun* **9**, 2282 (2018).
63. Segre, A.V. *et al.* Common inherited variation in mitochondrial genes is not enriched for associations with type 2 diabetes or related glycemic traits. *PLoS Genet* **6**(2010).
64. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-50 (2005).

65. Pybus, M. *et al.* 1000 Genomes Selection Browser 1.0: a genome browser dedicated to signatures of natural selection in modern humans. *Nucleic Acids Res* **42**, D903-9 (2014).
66. Voight, B.F., Kudaravalli, S., Wen, X. & Pritchard, J.K. A map of recent positive selection in the human genome. *PLoS Biol* **4**, e72 (2006).
67. Sabeti, P.C. *et al.* Genome-wide detection and characterization of positive selection in human populations. *Nature* **449**, 913-8 (2007).
68. Zhang, Y., Qi, G., Park, J.H. & Chatterjee, N. Estimation of complex effect-size distributions using summary-level statistics from genome-wide association studies across 32 complex traits. *Nat Genet* **50**, 1318-1326 (2018).