nature portfolio

Corresponding author(s): Jian Yang, Jian Zeng, Chikashi Terao, Kun Zhou, Gang Chen

Last updated by author(s): Mar 6, 2026

# Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

## Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

| n/a | Confirmed |
|---|---|
| ☐ | ☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☒ | ☐ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☒ | ☐ The statistical test(s) used AND whether they are one- or two-sided<br>*Only common tests should be described solely by name; describe more complex techniques in the Methods section.* |
| ☐ | ☒ A description of all covariates tested |
| ☐ | ☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☐ | ☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| ☐ | ☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted<br>*Give P values as exact values whenever suitable.* |
| ☐ | ☒ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☒ | ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ☐ | ☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated |

*Our web collection on statistics for biologists contains articles on many of the points above.*

## Software and code

Policy information about availability of computer code

| | |
|---|---|
| Data collection | Data was obtained from the UK Biobank study, the 23andMe Inc cohort, the Genetic Epidemiology Research on Adult Health and Aging (GERA) cohort, the FinnGen cohort, the Tohoku Medical Megabank Organization's (ToMMo) community-based cohort study, a published GWAS study of high myopia by Meguro et al. 2020, Ophthalmology, the WeGene Inc cohort, the European Prospective Investigation into Cancer (EPIC)-Norfolk cohort, the Avon Longitudinal Study of Parents and Children (ALSPAC) cohort, the Million Veteran Program (MVP) cohort, the All of Us research program cohort, the 23mofang, Inc cohort, and the Estonian Biobank cohort. |

| | |
|---|---|
| Data analysis | GCTA v1.93.0 beta (https://yanglab.westlake.edu.cn/software/gcta/#Overview) was used for heritability estimation (GREML, GREML-LDMS), genome-wide association study (fastGWA-lr, fastGWA-mlm), Mendelian randomization analysis (GSMR), and conditional analysis (COJO, mtCOJO). LDSC v1.0.1 (https://github.com/bulik/ldsc), and GCTB v2.01 beta (https://cnsgenomics.com/software/gctb/#Overview) were also used for heritability estimation. MTAG v1.0.8 (https://github.com/JonJala/mtag), and METAL (version released on 2011-03-25, https://genome.sph.umich.edu/wiki/METAL) were used for GWAS meta-analysis. Popcorn v1.0 ( https://github.com/brielin/Popcorn) was used for estimating genetic correlation across ancestries. MAGMA v1.10 (https://cloufield.github.io/GWASTutorial/09_Gene_based_analysis/),GCTA-mBAT-combo v1.94.1 (https://yanglab.westlake.edu.cn/software/gcta/#mBAT-combo), SMR v1.3.1 (https://yanglab.westlake.edu.cn/software/smr/#Overview), FUSION-TWAS (https://github.com/gusevlab/fusion_twas), the R package 'COLOC' (https://chr1swallace.github.io/coloc/), PoPS v0.2 (https://github.com/FinucaneLab/pops) were used for gene prioritization analyses. 'CARMA' R package v1.0 (https://github.com/ZikunY/CARMA), PolyFun v 1.0.0 (https://github.com/omerwe/polyfun), and SuSiEx (https://github.com/getian107/SuSiEx) were used for fine-mapping analysis. DENTIST v1.2.0.0 (https://github.com/Yves-CHEN/DENTIST) was used to detect problematic variants in the quality control process for fine-mapping. GCTB and SBayesRC (https://github.com/zhilizheng/SBayesRC) were used to re-weight SNP effect sizes for polygenic risk prediction. PLINK v1.90 (https://zzz.bwh.harvard.edu/plink/) was used to calculate polygenic risk score. The R package 'pROC' v1.18.4 , 'survival' v3.5-7, 'survminer' v0.4.9 were used to evaluate the performance of polygenic risk score. |

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

## Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The ancestry-stratified and cross-ancestry GWAS summary datasets and polygenic risk scores from this study are available for research use at https://yanglab.westlake.edu.cn/pub_data.html, excluding the 23andMe participants due to the 23andMe data transfer agreement. For bona fide researchers who meet the criteria for accessing the GWAS summary statistics from 23andMe (https://research.23andme.com/dataset-access/), we provide easy-to-use scripts for meta-analyzing our released summary statistics and the 23andMe summary statistics, and for generating the PRS given the meta-analyzed summary statistics at https://yanglab.westlake.edu.cn/pub_data.html. Individual-level data from the UK Biobank are publicly available to bona fide researchers (https://www.ukbiobank.ac.uk); GWAS summary statistics from the FinnGen cohort can be downloaded through the link https://storage.googleapis.com/finngen-public-data-r9/summary_stats/finngen_R9_H7_MYOPIA.gz; GWAS summary statistics from the MVP cohort can be downloaded at https://ftp.ncbi.nlm.nih.gov/dbgap/studies/phs002453/analyses/GIA/phs002453.MVP_R4.1000G_AGR.GIA.PheCodes_SenseOrgans_batch1.analysis-PI.MULTI.tar; GWAS summary statistics from the All of Us Research Program are accessible to researchers via the All of Us Researcher Workbench at https://www.researchallofus.org; Individual-level data from the ALSPAC cohort can be accessed upon request (https://www.bristol.ac.uk/alspac/researchers/access/); Individual-level data from the EPIC-Norfolk cohort can be accessed upon request (https://www.epic-norfolk.org.uk/for-researchers/data-sharing/data-requests/); GWAS summary statistics from the Meguro et al. study, the WeGene cohort, the GERA cohort, the 23mofang cohort, and the ToMMo cohort can be accessed upon request.

## Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

| | |
|---|---|
| Reporting on sex and gender | In the study, we included sex as a covariate for the association analysis. The sex was determined based on self-reporting or recorded by National Health Service (NHS). |
| Reporting on race, ethnicity, or other socially relevant groupings | For the genetic discovery, we focused on the European, East Asian and African ancestries due to the availability of the datasets. For the polygenic risk prediction, we evaluated the performance in the ancestries of European, East Asian, South Asian, and African. |
| Population characteristics | See detailed description in Section 1 of the Supplementary Note for the cohort information. |
| Recruitment | See detailed description in Section 1 of the Supplementary Note for the cohort information. |
| Ethics oversight | This study is approved by the Ethics Committee of Westlake University (approval number: 20200722YJ001) and the University of Queensland Human Research Ethics Committee 8 (approval number: 2011001173). |

Note that full information on the approval of the study protocol must also be provided in the manuscript.

# Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

# Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

| | |
|---|---|
| Sample size | For the European ancestry, we performed GWAS analyses using directly measured mean spherical equivalent (MSE) in 107,247 UK Biobank (UKB) participants, measured MSE in 34,998 Genetic Epidemiology Research on Adult Health and Aging (GERA) cohort, imputed MSE based on age of onset of spectacle wear (AOSW) from a polynomial model in 301,121 UKB samples, binary myopia status of 3,534 cases and 361,237 controls from the FinnGen cohort, binary myopia status of 106,086 cases and 85,757 controls from 23andMe, 65,574 cases and 333,242 controls from the Million Veteran Program (MVP), and 5,339 cases and 91,024 controls from the All of Us research program (AllofUs). For the East Asian ancestry, we conducted a GWAS based on ophthalmic examination data from the ToMMo, which comprised 14,786 myopes and 35,828 controls, and obtained summary statistics of a published GWAS study of high myopia (1,632 cases and 1,586 controls). We also performed a GWAS based on self-reported categorical myopia status in 5,952 samples from WeGene, and in 47,305 self-reported myopes and 8,209 non-myopes from 23mofang, Inc. For the AFR ancestry, we obtained GWAS summary statistics for myopia from the MVP cohort (20,300 cases and 84,944 controls) and the AllofUs cohort (924 cases and 38,569 controls). Additional individual-level data from multiple cohorts were employed for validation analyses, including directly measured MSE from the European Prospective Investigation into Cancer (EPIC)-Norfolk (n = 7,069), the Avon Longitudinal Study of Parents and Children (ALSPAC) (n = 1,493), non-EUR UKB participants (n = 10,088), as well as self-reported MSE from the Estonian Biobank (EBB) (n = 10,441). |
| Data exclusions | We removed data from participants whose other medical conditions or episodes could have altered refractive error related measurement(for example eye surgery, eye infections, etc.). |
| Replication | Reproducibility of results was assured by comparing association results and significance across different cohorts and ancestries. |
| Randomization | No randomization procedure was needed. |
| Blinding | No blinding was needed. |

# Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

## Materials & experimental systems

| n/a | Involved in the study |
|---|---|
| ☒ | ☐ Antibodies |
| ☒ | ☐ Eukaryotic cell lines |
| ☒ | ☐ Palaeontology and archaeology |
| ☒ | ☐ Animals and other organisms |
| ☒ | ☐ Clinical data |
| ☒ | ☐ Dual use research of concern |
| ☒ | ☐ Plants |

## Methods

| n/a | Involved in the study |
|---|---|
| ☒ | ☐ ChIP-seq |
| ☒ | ☐ Flow cytometry |
| ☒ | ☐ MRI-based neuroimaging |

## Plants

| | |
|---|---|
| Seed stocks | *Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.* |
| Novel plant genotypes | *Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.* |
| Authentication | *Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.* |