Genome-wide association analysis of 95,549 individuals identifies novel loci and genes influencing optic disc morphology

Supplementary

Table of Contents:

Supplementary methods

Web resources

BOLT-LMM software: https://data.broadinstitute.org/alkesgroup/BOLT-LMM/

GCTA software: http://cnsgenomics.com/software/gcta/

GWAS Catalog: https://www.ebi.ac.uk/gwas/

Haplotype Reference Consortium: http://www.haplotype-reference-consortium.org/

International Glaucoma Genetic Consortium dataset: https://goo.gl/73qHqk

LOCUSZOOM: http://locuszoom.sph.umich.edu/

LDlink: https://analysistools.nci.nih.gov/LDlink/

LD score regression software: https://github.com/bulik/ldsc

METAL software: http://csg.sph.umich.edu/abecasis/Metal/

PLINK software: http://www.cog-genomics.org/plink2

R: https://cran.r-project.org/

UK Biobank: http://www.ukbiobank.ac.uk/

Rank-based inverse-normal transformation

Rank-based inverse-normal transformation is widely used in quantitative genetics to generate normal distribution from any distribution, unless heavy ties are present in quantitative traits.(1, 2) The transformed phenotype data has a mean value of zero, and a standard deviation (SD) of one. For the GWAS results from rank-based inverse-normal transformed trait, we need to scale the SNP effect sizes by multiplying by the SD of the raw phenotype data. In the assessment of vertical disc diameter in UK Biobank, we cropped and enlarged the fundus retinal eye images to facilitate grading using a custom Java program. Thus, we do not have the actual SD value for vertical disc diameter. In IGGC dataset, the disc size was measured by disc area.(3) In the traditional inverse-variance meta-analysis of UK Biobank and IGGC disc size GWAS summary statistics, we scaled the UK Biobank rank inverse-normal transformed disc diameter by the SD (= 0.4) of IGGC disc area to make sure they are on the same scale. In the sensitivity analysis, we also used multiple trait analysis of GWAS (MTAG), a framework to generalize the standard inverse-variance meta-analysis method that can joint analyse the same trait with different measures or even different traits with a high genetic

correlation,(4) to joint analyse UKBB and IGGC disc size summary statistics, the results are essentially identical to traditional inverse variance meta-analysis.

MTAG method

The multiple trait analysis of GWAS (MTAG) is a generalized inverse-variance meta-analysis method to perform joint analysis of GWAS summary results from related traits to improve statistical power to identify new genes.(4) In MTAG, GWAS summary results from related traits are used to construct the variance–covariance matrix of their SNP effects and estimation error; MTAG improves the power and accuracy of effect estimates by incorporating information from other genetic correlated traits. In our analysis, the disc diameter data in UK Biobank and the disc area data in IGGC are the same trait (disc size) of different measurements. The genetic correlation between UK Biobank dataset and IGGC dataset is very high (genetic correlation coefficient = 0.83, $P = 1.31\times10^{-76}$). In our main analysis, we performed the traditional inverse-variance meta-analysis in METAL. In our sensitivity analysis, we conducted MTAG to joint analysis UK Biobank and IGGC disc size data (default parameters, NO --perfect_gencov and NO --equal_h2). We showed their results are essentially identical.

Definition of independent genome-wide significant loci

To identify independent genome-wide significant loci, we conducted stepwise model selection procedures in GCTA-COJO software (version 1.26)(5). GCTA-COJO uses GWAS summary results and estimates LD from a reference sample (randomly selected 5,000 UKBB white British ancestry individuals, considering SNPs within a two megabase window) for the conditional and joint association analysis. Although the joint analysis can uncover SNPs with $P<5\times10^{-8}$ in the joint test and $P<5\times10^{-8}$ in the standard (unconditional) test, here we only report SNPs with both unconditional P values and joint P values less than 5×10^{-8}.

To define novel disc size SNPs, we excluded known disc size-associated loci. To check against the previous literature, we first used HaploReg to identity all the proxy SNPs ($r^2>=0.8$) of the lead SNPs

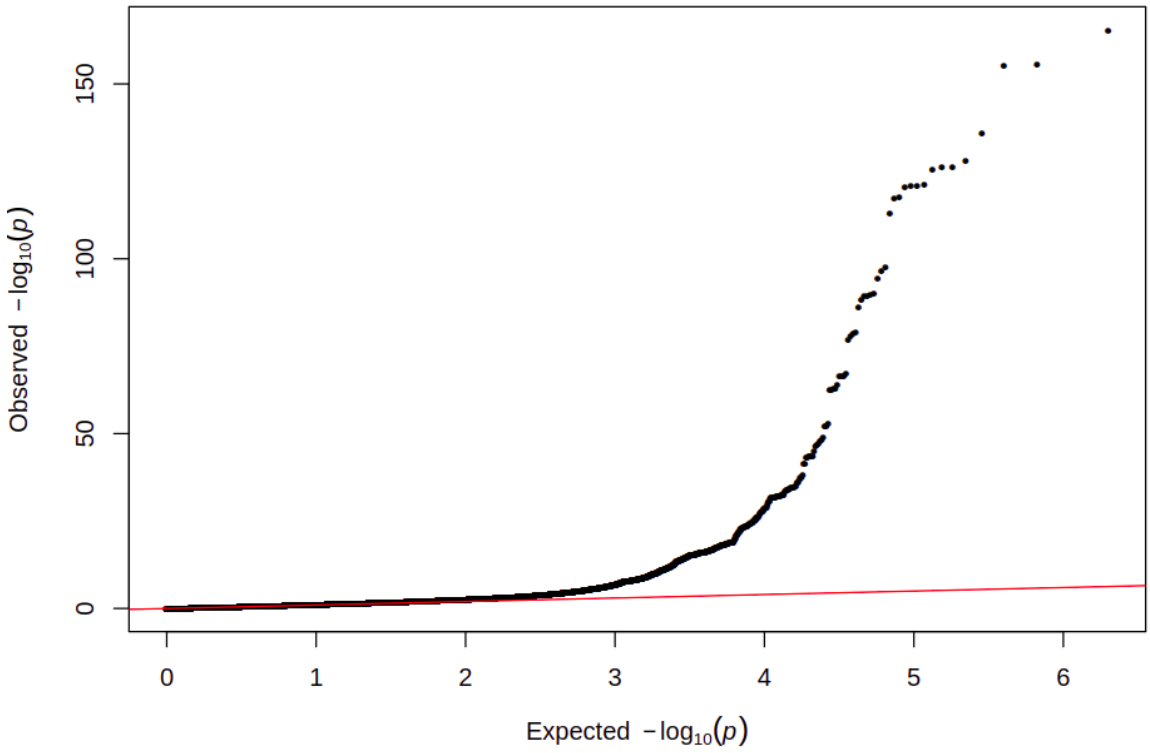
from GCTA-COJO.(6) The lead SNPs, their proxy SNPs, and the located genes were checked in LDlink(7), GWAS Catalog,(8) and PubMed.

Supplementary Figures

Figure S1. Locuszoom plot of UKBB disc diameter lead SNPs

Locuszoom figures are displayed in a separate file: FigureS1_locuszoom_DISC_UKBB.pdf.

Figure S2. Quantile-quantile plot for UK Biobank disc diameter GWAS



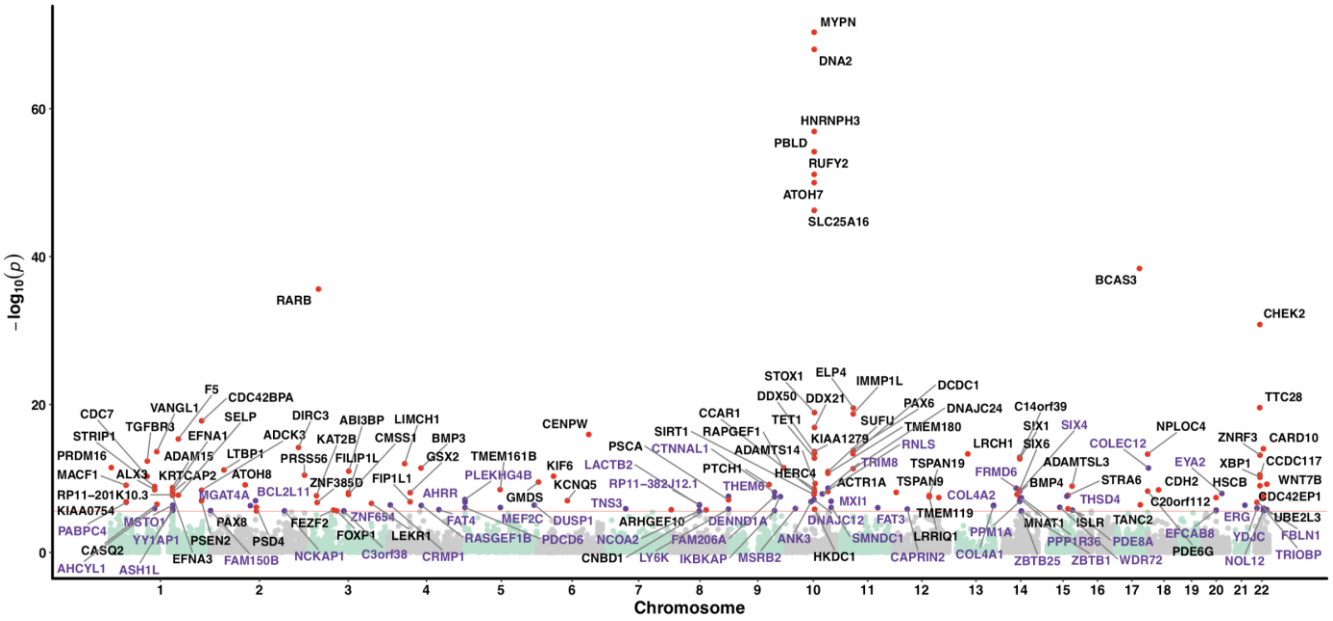
Linkage disequilibrium (LD) score regression intercept is used to assess the genomic inflation, and the intercept is 1.05 (standard error, SE=0.01) with lambda 1.20 and attenuation ratio ((Intercept -1) / ($\underline{\chi 2}$ -1)) 0.12 (SE=0.02). The quantile-quantile plot is based on 1×10^6 randomly selected SNPs.

Figure S3. Locuszoom plot of disc size meta-analysis (UK Biobank and IGGC datasets) lead SNPs.

Locuszoom figures are displayed in a separate file:

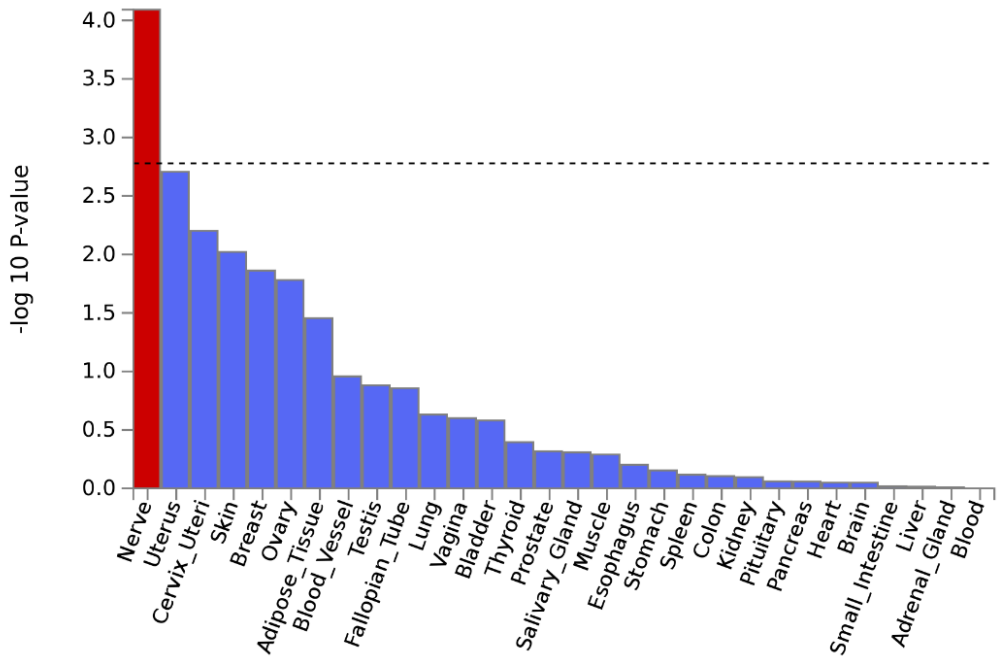
FigureS3_locuszoom_DISC_meta_UKBB_IGGC.pdf.

Figure S4: Manhattan plot of the gene-based test for disc size meta-analysis



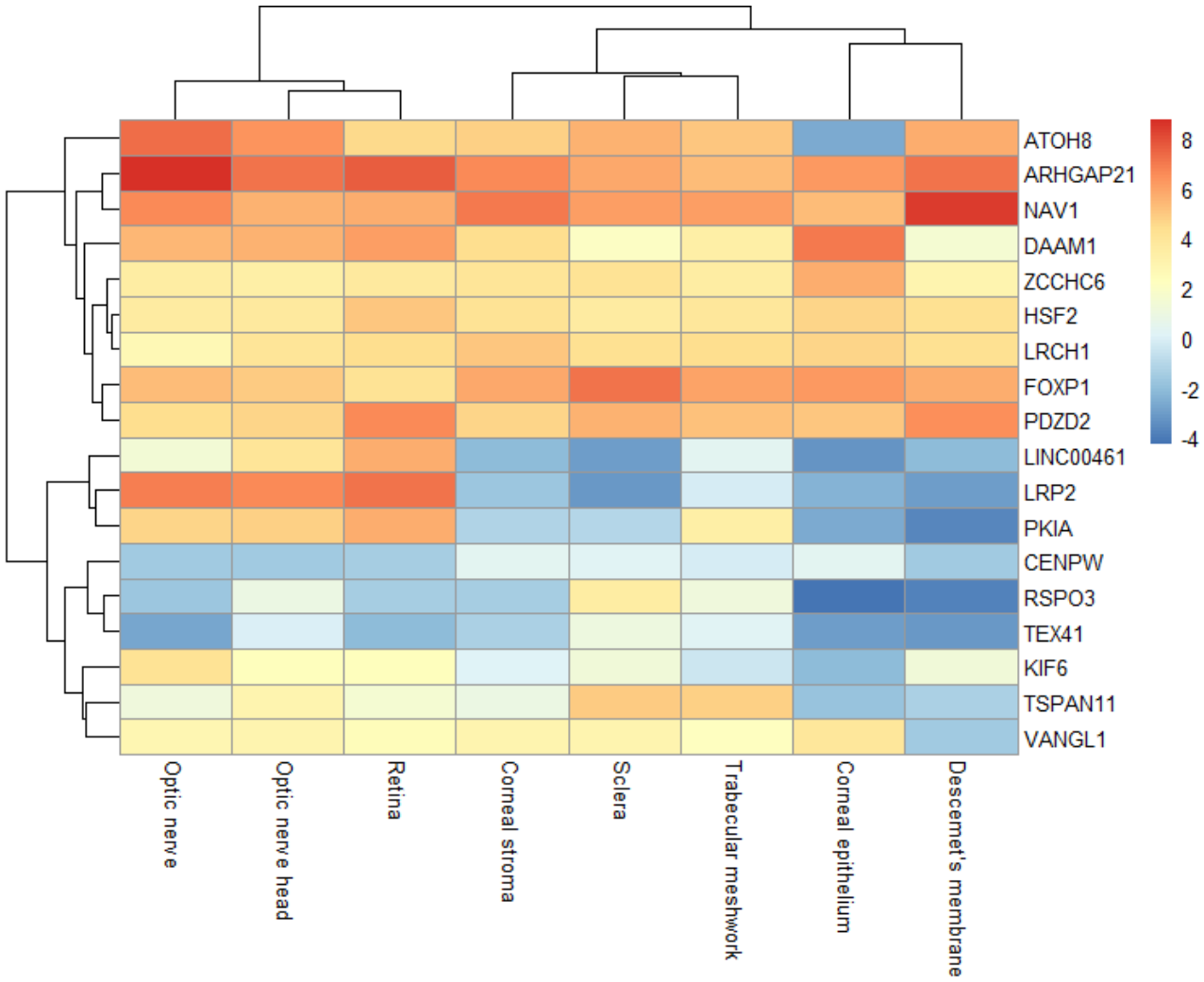
Genes with genome-wide significant SNPs are highlighted in red dots, with the gene names in black text. Genes with no genome-wide significant SNPs (novel genes from gene-based test) are highlighted in purple dots, with the gene names in purple text. The red line is the gene-based significant level (P = 0.05/18,619).

Figure S5: Tissue Expression analysis of GTEx v7 30 general tissue types



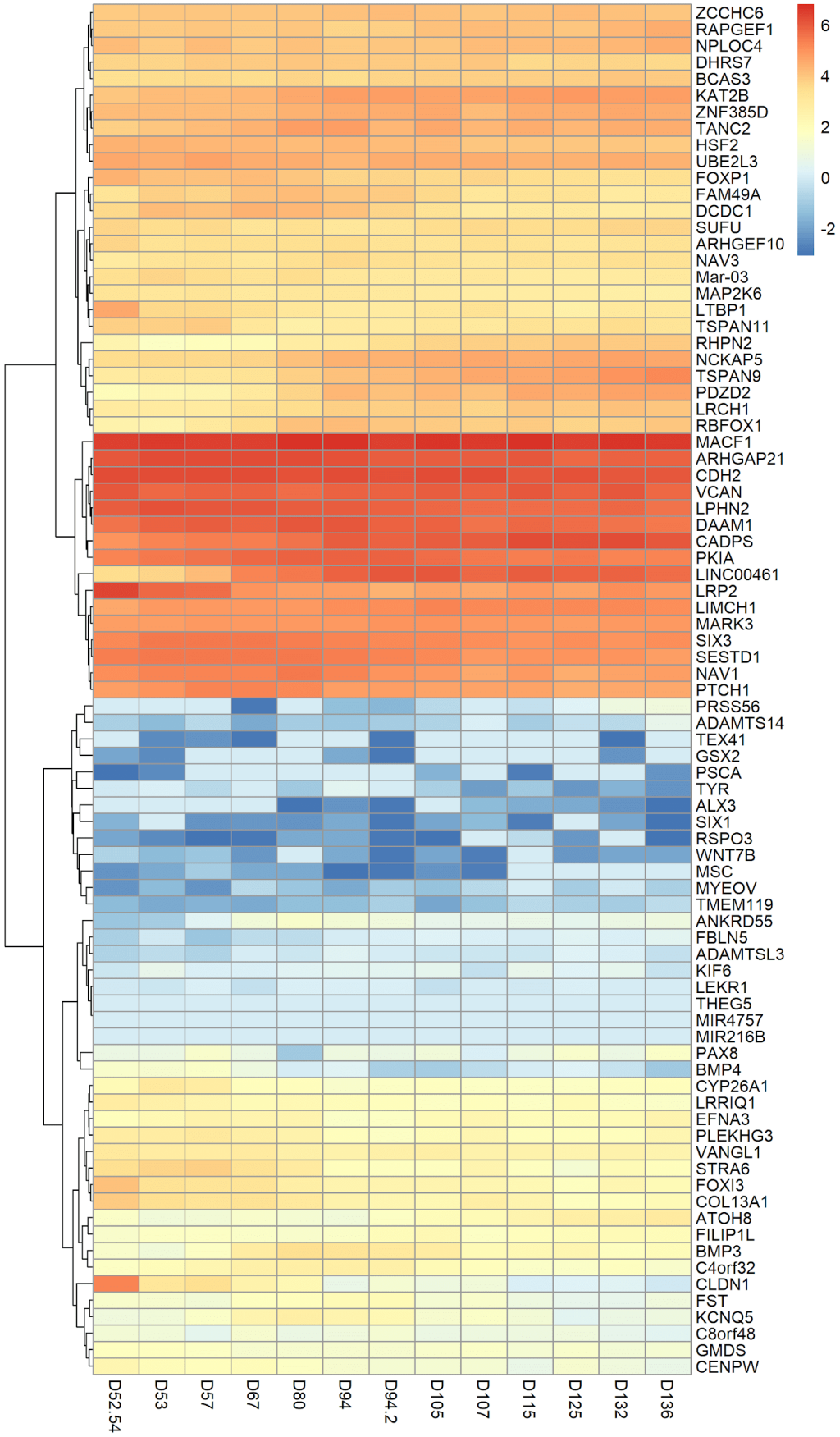
The expression of GTEx v7 30 general tissue types, where the x-axis represents the tissue types, the y-axis represents the negative of the log base 10 of p-values.

Figure S6. Normalised RNA-Seq expression in ocular tissue of the novel genes associated with vertical disc diameter that have been reproduced after Bonferroni correction.



The heat plot is a measure of normalised log counts per million of the 19 novel genes that have been replicated at Bonferroni correction in IGGC and EPIC samples. Genes and tissue types are clustered using complete-linkage hierarchical clustering method. Expression data for the 19th gene (*MIR216B*) was not available.

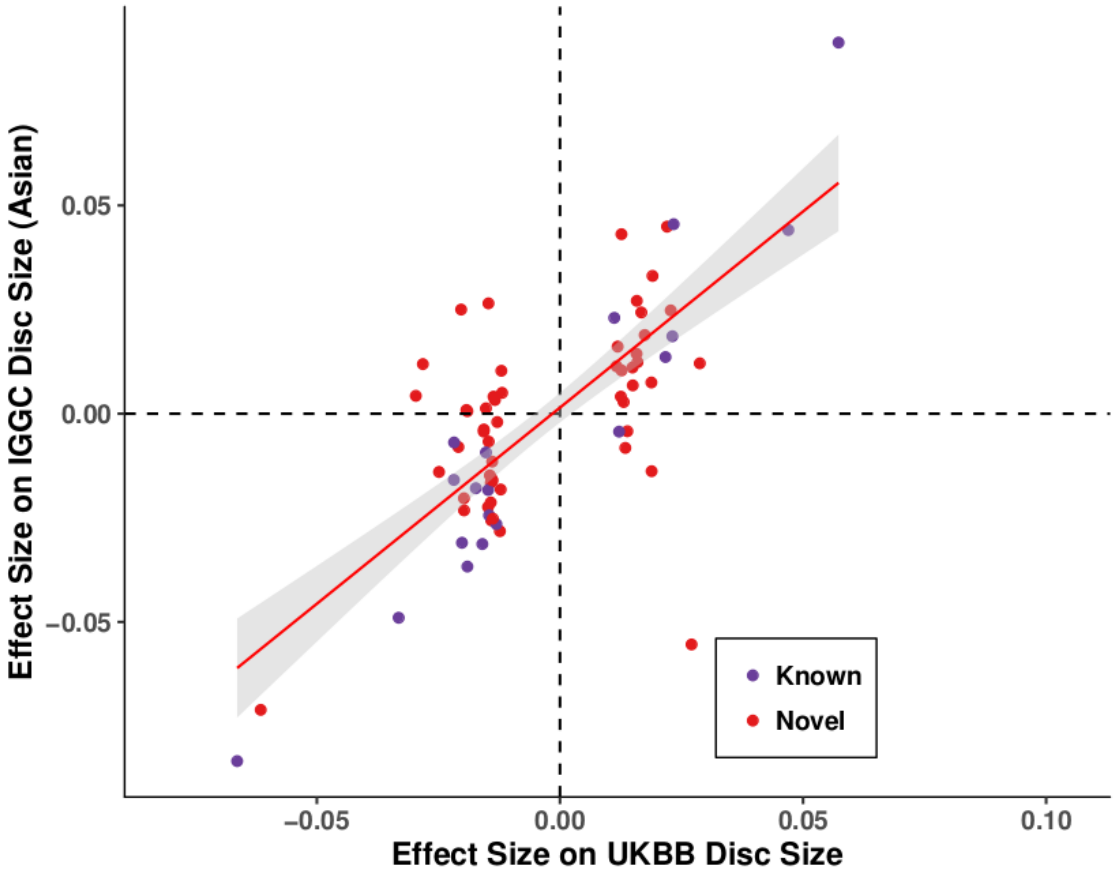
Figure S7. Normalised RNA expression of the novel genes associated with vertical disc diameter in foetal retinal tissue at various developmental ages



The heat plot is a measure of normalised log counts per million of the novel genes associated with vertical disc

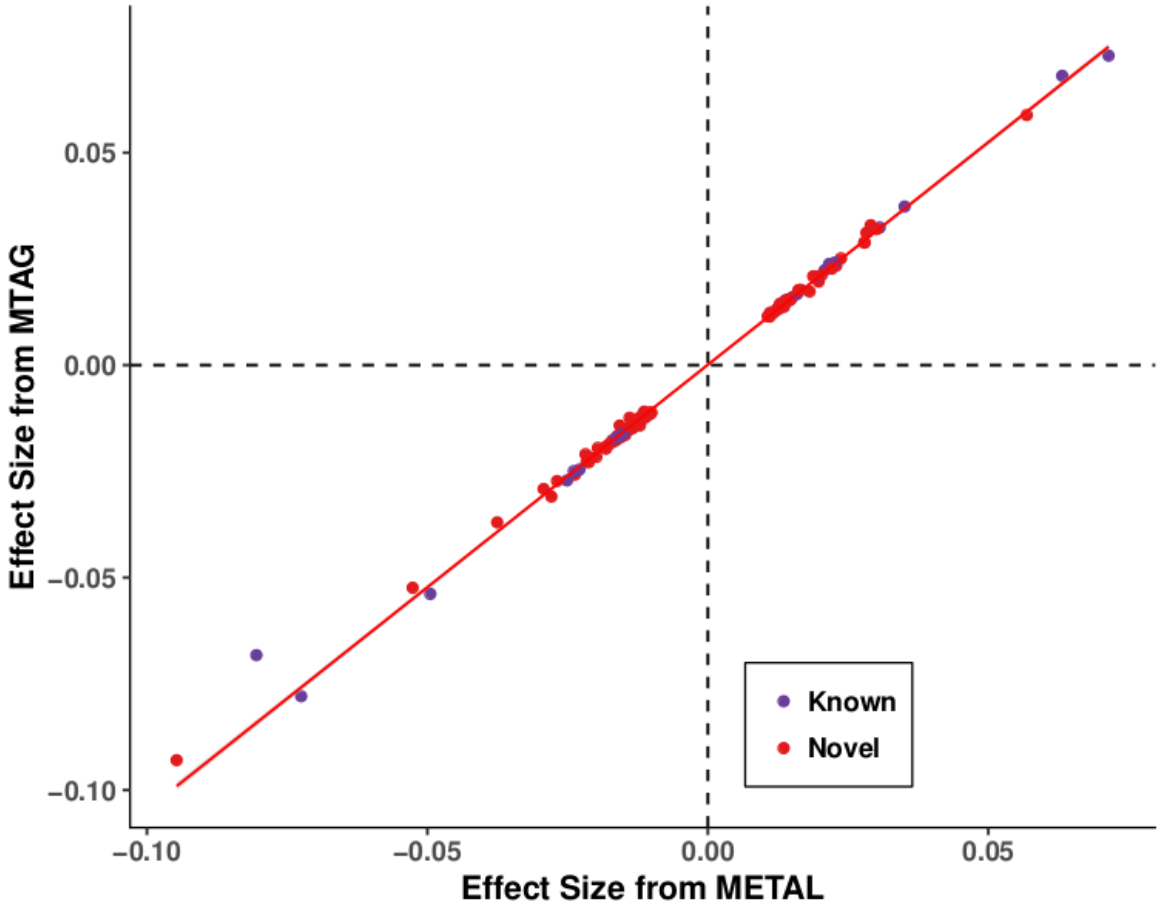
diameter. The x-axis (D52-136) represent whole fetal retinae tissue through various ages, in days. Genes are clustered using the complete-linkage hierarchical clustering method.(9)

Figure S8. Comparison of the effect sizes for 91 genome-wide significant independent SNPs identified from UK Biobank disc size GWAS versus those in independent cohort IGGC Asian disc size GWAS.



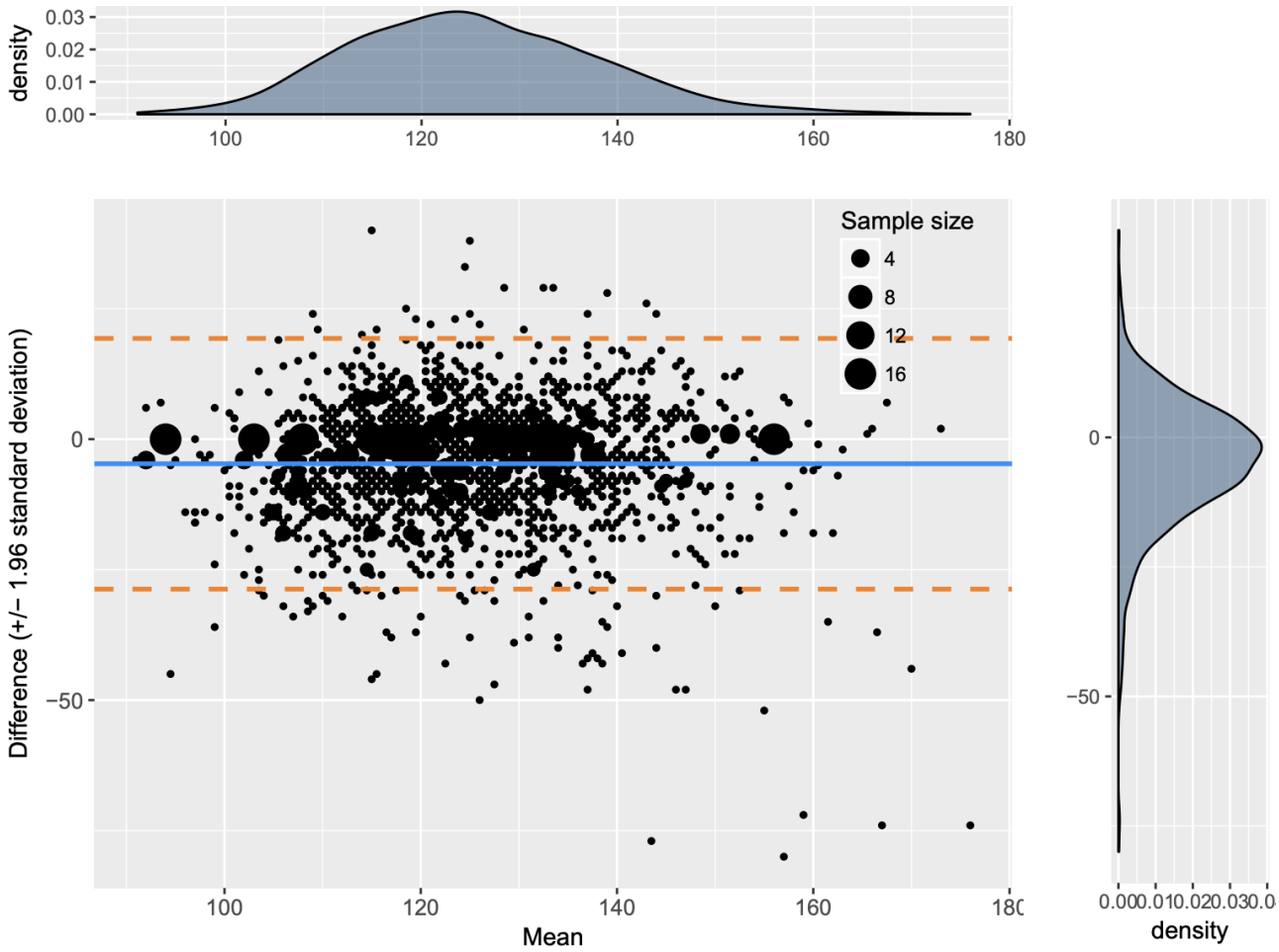
Pearson's correlation coefficient is 0.72 (P value=7.79×10^{-13}). The red line is the best fit line, with the 95% confidence interval region in grey. Novel disc size SNPs are highlighted in red and known SNPs in purple.

Figure S9. Comparison of the effect sizes for 115 genome-wide significant independent SNPs identified from meta-analysis from METAL software versus those from MTAG method.



Pearson's correlation coefficient is 0.998 (P value=1.81×10^{-134}). The red line is the best fit line. Novel disc size SNPs are highlighted in red and known SNPs in purple.

Figure S10. Correlation between two examiner's disc diameter measurements



Bland-Altman plots for vertical disc diameter, where the x-axis represents the mean value of two measurements, the y-axis represents the difference between two measurements, the blue line is the mean value of difference, and the dashed orange lines are the 95% limits of agreement (95% confidence interval for the mean value of difference). The black dots are scaled by the number of samples. The right and top panels are the density plots for the difference of the measurements and the mean value of measurements, respectively.

Supplementary Tables

Supplementary Tables are displayed in a separate file: Supplementary_Tables_DD.xlsx.

References

1. Aulchenko,Y.S., Ripke,S., Isaacs,A. and van Duijn,C.M. (2007) GenABEL: an R library for genome-wide association analysis. *Bioinformatics*, **23**, 1294–1296.

2. Wu,Y., Zheng,Z., Visscher,P.M. and Yang,J. (2017) Quantifying the mapping precision of genome-wide association studies using whole-genome sequencing data. *Genome Biol.*, **18**, 86.

3. Bartling,H., Wanger,P. and Martin,L. (2008) Measurement of optic disc parameters on digital fundus photographs: algorithm development and evaluation. *Acta Ophthalmol.*, **86**, 837–841.

4. Turley,P., Walters,R.K., Maghzian,O., Okbay,A., Lee,J.J., Fontana,M.A., Nguyen-Viet,T.A., Wedow,R., Zacher,M., Furlotte,N.A., *et al.* (2018) Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nat. Genet.*, **50**, 229–237.

5. Yang,J., Ferreira,T., Morris,A.P., Medland,S.E., Genetic Investigation of ANthropometric Traits (GIANT) Consortium, DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium, Madden,P.A.F., Heath,A.C., Martin,N.G., Montgomery,G.W., *et al.* (2012) Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.*, **44**, 369–75, S1–3.

6. Ward,L.D. and Kellis,M. (2012) HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.*, **40**, D930–4.

7. Machiela,M.J. and Chanock,S.J. (2015) LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics*, **31**, 3555–3557.

8. MacArthur,J., Bowler,E., Cerezo,M., Gil,L., Hall,P., Hastings,E., Junkins,H., McMahon,A., Milano,A., Morales,J., *et al.* (2017) The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res.*, **45**, D896–D901.

9. Hoshino,A., Ratnapriya,R., Brooks,M.J., Chaitankar,V., Wilken,M.S., Zhang,C., Starostik,M.R., Gieser,L., La Torre,A., Nishio,M., *et al.* (2017) Molecular Anatomy of the Developing Human Retina. *Dev. Cell*, **43**, 763–779.e4.