

Automated AI labeling of optic nerve head enables insights into cross-ancestry glaucoma risk and genetic discovery in >280,000 images from UKB and CLSA

Xikun Han,^{1,2,*} Kaiah Steven,³ Ayub Qassim,⁴ Henry N. Marshall,⁴ Cameron Bean,³ Michael Tremeer,³ Jiyuan An,⁵ Owen M. Siggs,⁴ Puya Gharahkhani,¹ Jamie E. Craig,⁴ Alex W. Hewitt,^{6,7} Maciej Trzaskowski,³ and Stuart MacGregor¹

Summary

Cupping of the optic nerve head, a highly heritable trait, is a hallmark of glaucomatous optic neuropathy. Two key parameters are vertical cup-to-disc ratio (VCDR) and vertical disc diameter (VDD). However, manual assessment often suffers from poor accuracy and is time intensive. Here, we show convolutional neural network models can accurately estimate VCDR and VDD for 282,100 images from both UK Biobank and an independent study (Canadian Longitudinal Study on Aging), enabling cross-ancestry epidemiological studies and new genetic discovery for these optic nerve head parameters. Using the AI approach, we perform a systematic comparison of the distribution of VCDR and VDD and compare these with intraocular pressure and glaucoma diagnoses across various genetically determined ancestries, which provides an explanation for the high rates of normal tension glaucoma in East Asia. We then used the large number of AI gradings to conduct a more powerful genome-wide association study (GWAS) of optic nerve head parameters. Using the AI-based gradings increased estimates of heritability by ~50% for VCDR and VDD. Our GWAS identified more than 200 loci associated with both VCDR and VDD (double the number of loci from previous studies) and uncovered dozens of biological pathways; many of the loci we discovered also confer risk for glaucoma.

Introduction

The optic nerve head is the exit point of retinal ganglion cell axons from the eye to the brain.¹ It is commonly assessed during ophthalmic examinations using funduscopy or optical imaging technology for multiple ocular diseases, such as glaucoma, which is the leading cause of irreversible blindness globally and is characterized by characteristic cupping of the optic disc as a result of retinal ganglion cell apoptosis.^{2,3} Enlarged vertical cup-to-disc ratio (VCDR) is considered a hallmark of glaucomatous optic neuropathy and is often used to define glaucoma in general population-based prevalence surveys.⁴ However, VCDR alone is not adequate to assess glaucomatous damage in part because of the variation of optic disc size. For instance, a vertical cup-to-disc ratio of 0.5 in a small optic disc could be pathologic, whereas a vertical cup-to-disc ratio of 0.8 in a large disc size may represent physiologic cupping. Adjusting for optic disc size is hence important to maximizing the clinical utility of VCDR in diagnosing glaucoma.

Family studies have shown that optic disc morphology traits are highly heritable with an estimated heritability of 0.48 and 0.57 for VCDR and optic disc diameter, respectively.⁵ Large-scale genome-wide association studies (GWASs) for optic disc morphology have begun to shed light on the development and pathogenesis of glaucoma

and other optic nerve diseases.^{6–8} However, both large sample sizes and accurate phenotyping are critical in GWASs and further progress will be difficult under the existing manual phenotype paradigm. Manual assessment of optic disc photographs is time intensive and often suffers from poor inter-observer concordance, even when performed by trained specialists, and an alternative approach is required.^{9,10} Clinical estimates of VCDR are more difficult from monoscopic photographs compared with stereoscopic viewing of the optic nerve head, which can be achieved during slit-lamp biomicroscopy or from stereoscopic photographs.

Recent advances in artificial intelligence (AI) algorithms have shown exciting promise in healthcare,¹¹ including the automated diagnosis of eye diseases.^{12,13} With the high performance of AI technology, the U.S. Food and Drug Administration approved the first medical device to use AI technology to detect diabetic retinopathy in 2018.^{14,15} The probabilistic nature and non-linear capabilities, as well as analytical capabilities to deal with single and multimodal high-dimensional data, has seen the application of AI experience lower resistance to adoption in the medical field when applied to computer vision applications. Two fundamental properties have facilitated AI application to medical diagnostics. First, the problem space (medical imaging) is, relative to other medical domains,

¹Statistical Genetics Lab, QIMR Berghofer Medical Research Institute, Brisbane, Herston, QLD 4006, Australia; ²School of Medicine, University of Queensland, Brisbane, Herston, QLD 4006, Australia; ³Max Kelsen, Brisbane, Spring Hill, QLD 4000, Australia; ⁴Department of Ophthalmology, Flinders University, Flinders Medical Centre, Bedford Park, SA 5042, Australia; ⁵Queensland University of Technology, Brisbane, Brisbane City, QLD 4000, Australia; ⁶Menzies Institute for Medical Research, University of Tasmania, Hobart, TAS 7000, Australia; ⁷Centre for Eye Research Australia, University of Melbourne, East Melbourne, VIC 3002, Australia

*Correspondence: xikun.han@qimrberghofer.edu.au

<https://doi.org/10.1016/j.ajhg.2021.05.005>

© 2021 American Society of Human Genetics.



well studied and very well understood. Second, an observation of the output can be quickly validated by a clinical practitioner, who by having access to additional clinical or historical data about that patient, may suggest alternative diagnosis. A motivating factor driving utilization of AI on data such as fundus images is the large volume of images available for algorithms to be trained on. Furthermore, standardized imaging techniques can drastically reduce the dataset heterogeneity. This is highlighted by the collection of images as part of the UK Biobank (UKB) and Canadian Longitudinal Study on Aging (CLSA) biobank completed over a decade. Automated diagnosis from retinal fundus imaging has been approached through a number of different algorithms, ranging from multi-stage “classical” learning algorithms to end-to-end deep learning models.^{16–18} The fast and automatic phenotyping with AI algorithms would enable large-scale genetic studies.^{19,20}

In this study, a convolutional neural network (CNN) model was utilized in a transfer learning approach, training on clinical assessments of the optic nerve head in ~70,000 photographs (labeled training data) of UKB participants. Automatic labeling by the CNN model dramatically boosts the effective sample size ($n = 282,100$ total images graded), presenting an opportunity to greatly expand the utility of the GWAS approach for VCDR and optic disc diameter. We also apply the AI labels systematically across the multiple different ancestries in UKB and CLSA and investigate how VCDR and other glaucoma risk factors, such as intraocular pressure (IOP), relate to glaucoma risk in different ancestries.

Subjects and methods

Study design and overview

The overall study design is summarized in [Figure 1](#). We use transfer learning to train three CNN models for image gradability, VCDR, and vertical disc diameter (VDD) values from ~70,000 UKB fundus images graded by clinicians. These models were then applied to all UKB fundus images (85,736 participants and 175,770 images in total) and another independent cohort—the Canadian Longitudinal Study on Aging (CLSA, 29,635 participants and 106,330 images in total). We performed AI-based GWASs for VCDR and VDD and replicated genetic discoveries in clinician-graded fundus images from the International Glaucoma Genetics Consortium (IGGC) and in IGGC glaucoma case-control studies.²¹ The large-scale biobank data for both VCDR and IOP also allow us to systematically compare the glaucoma risk and optic nerve head parameters across different ancestries.

The UKB and the CLSA have been reviewed and approved by local research ethics boards. All participants provided informed written consent, and study procedures were performed in accordance with the World Medical Association Declaration of Helsinki ethical principles for medical research.

Study populations

UK Biobank

The UK Biobank (UKB) is a population-based cohort study with deep genetic and phenotypic data from ~500,000 participants

aged between 40 to 69 years at the time of recruitment (2006–2010) and living in the United Kingdom.²² Retinal fundus images, obtained with the Topcon 3D OCT 1000 Mk2 instrument, were available for both left and right eyes from two assessment visits, covering ~85,000 participants (~68,000 participants in the baseline visit and ~19,000 participants in the first repeat assessment visit [2012–2013], with approximately 2,000 overlap participants in the baseline and repeat visits). In our previous study, VCDR and VDD were graded by two clinicians via a custom Java program.²³ Detailed image processing and quality control methods were described previously.²³ Briefly, given the time-consuming nature of manual grading, we only graded the left eye images (if the left eye images were ungradable, the right eye images were used instead) and one visit (if the second visit measurements were unavailable, the first visit measurements were used instead) of white British ancestry participants. A total of 67,040 participants with both VCDR and VDD measurements were included in our previous GWAS. In this study, we used a CNN model to grade left and right eye images from two visits for all participants, irrespective of ancestry, with a total of 175,770 images.

In the UKB, ~488,000 participants were genotyped for 805,426 variants on Axiom arrays (Affymetrix Santa Clara, USA). The genetic data, quality control procedures, and imputation methods have been described previously.²² Briefly, ~96 million variants were imputed with the Haplotype Reference Consortium (HRC) and UK10K haplotype resources,^{22,24,25} and 487,409 individuals passed genotyping quality control. Of them, 438,870 individuals were genetically similar to those of white British ancestry.^{22,26} For the GWAS in UKB, we retained SNPs with minor allele frequency (MAF) > 0.01 and imputation quality score > 0.8. To verify self-reported diverse ancestry information (data field 21000 in UKB), we used a K-means clustering method based on genetic principal components (PCs). The genetic clusters were compared with self-reported ancestry. Participants within the same self-reported ancestry groups were largely in the same genetic clusters (e.g., African [$N = 9,791$], East Asian [$N = 2,594$], and South Asian [$N = 9,941$], detailed in [Figure S1](#)), and on average, ~20% of them have fundus retinal images.

The Canadian Longitudinal Study on Aging

The Canadian Longitudinal Study on Aging (CLSA) is a national, longitudinal cohort study of 51,338 participants, aged 45 to 85 years at enrollment, from ten Canadian provinces.^{27,28} Recruitment and baseline data collection were completed in 2015, and participants were followed-up on every 3 years and had an initial follow-up visit completed in 2018. In this study, the nerve head photographs are available for 30,097 participants from the CLSA “comprehensive cohort” (for both left and right eyes and the baseline and first follow-up visit). Retinal fundus imaging was performed with a Topcon (TRC-NW8) non-mydratric retinal camera, and images were saved in jpg format. A random sample of 1,000 images was graded by a clinician for both VCDR and VDD via a custom Java program. The latest genome-wide genotype data (August 2019 release) are available for 19,669 participants of the comprehensive cohort, comprising 794,409 genetic variants genotyped on the Affymetrix Axiom array and ~40 million genetic variants imputed via the Haplotype Reference Consortium.²⁴ Variant- and sample-based quality control procedures were consistent with standards of the UKB,²² and detailed steps are presented in the CLSA support document (see [web resources](#)). For the GWAS analysis, we included 18,304 participants of European ancestry on the basis of the K-means cluster method on genetic principal components, and the largest cluster also contains

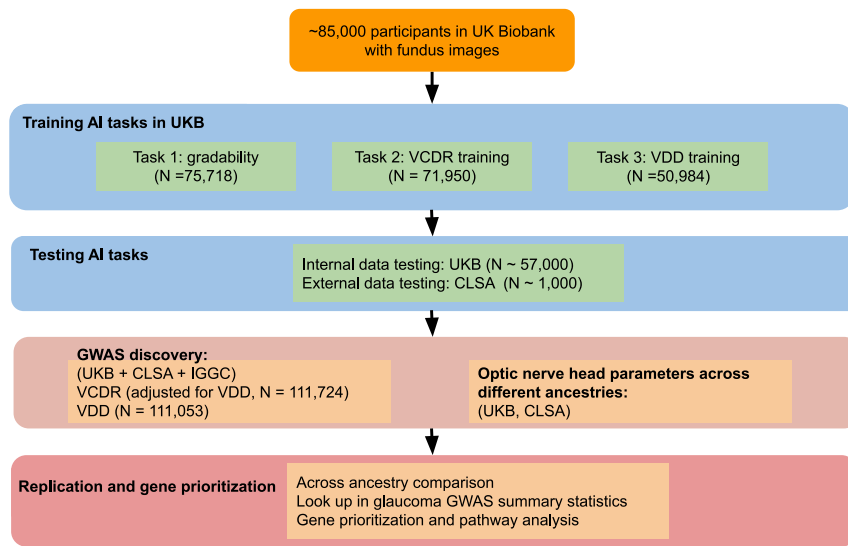


Figure 1. Flowchart of AI framework and datasets

In UK Biobank (UKB), the fundus retinal eye images were available for ~85,000 participants (~68,000 participants in the baseline visit and ~19,000 participants in the first repeat assessment visit). In our previous study, vertical cup-to-disc ratio (VCDR) and vertical disc diameter (VDD) were graded by two clinicians in ~70,000 photographs via a custom Java program. These clinical assessments were used as training data for three convolutional neural network (CNN) models for image gradability, VCDR, and VDD values. The learned models were then applied to all UKB fundus images (85,736 participants and 175,770 images in total) and another independent cohort—the Canadian Longitudinal Study on Aging (CLSA; 29,635 participants and 106,330 images in total). We further used the AI labels to systematically evaluate optic nerve head parameters across the multiple different ancestries in UKB and CLSA and to perform AI-based GWASs for VCDR and VDD.

the majority of individuals that self-report European ancestry. SNPs with MAF > 0.01 and imputation quality score > 0.8 were retained in association analysis. From the K-means clustering method, the sample sizes for African, South Asian, and East Asian are 135, 219, and 217, respectively (PC plot shown in Figure S2).

The International Glaucoma Genetic Consortium

The International Glaucoma Genetic Consortium (IGGC) is an international consortia established to identify glaucoma genetic risk variants through large-scale meta-analysis. The phenotype and genotype data of VCDR and optic disc area for IGGC have been previously described elsewhere.^{7,29} It should be noted the optic disc area is not in the same scale as VDD from the AI gradings. When comparing and meta-analyzing the VDD and disc area data, we applied a rank-based inverse normal transformation to AI gradings and rendered them back to disc area scale, as detailed in our previous study.⁸ Publicly available summary statistics were downloaded for individuals of European descent ($N_{\text{VCDR}} = 25,180$, $N_{\text{disc}} = 24,509$, from the latest HRC imputation), as well as Asian descent ($N_{\text{VCDR}} = 8,373$, $N_{\text{disc}} = 7,307$).^{7,29}

Glaucoma GWAS dataset

The glaucoma datasets were described in our previous study, including 34,179 primary open-angle glaucoma cases and 349,321 controls from a large-scale multi-ethnic meta-analysis.²¹ The detailed information of phenotype definition and genetic association analyses were presented in detail previously.²¹ We used the GWAS summary statistics to look up each of the VCDR loci (adjusted for VDD) and assess their effects on glaucoma.

AI algorithm on retinal images

We used three separate CNN models to make inferences about image gradability, VCDR, and VDD values of retinal fundus images in UKB. The image gradability (gradable or ungradable) was defined as a binary classification based on VCDR values (which is a slightly more complex measurement than VDD), while the latter two tasks were modeled as regression problems. Images with a higher likelihood of gradability (i.e., designated softmax probability more than 0.5) were assigned as gradable. While a variety of CNN model architectures were tested, the final architecture used for all CNN

models was ResNet-34.³⁰ Pre-trained weights, initially trained on ImageNet³¹ classification tasks, were utilized for each model as a form of transfer learning. Untrained layers specific to each model were additionally added, forming a custom regression (Relu) and classification (softmax) heads for each respective task. All fundus images were cropped to a pixel ratio of 1,080:800 before training or validation. We used the highest native resolution for the UKB training images because we found that using lower resolution negatively impacted inference metrics. The total dataset sizes used for the VCDR, VDD, and gradability tasks were 71,950, 50,984, and 75,718, respectively. Each dataset was randomly split into 80% training and 20% validation. The model performance was validated by sample hold out, and final testing was performed on images from the CLSA dataset. Model requirements for regression tasks were defined achieving a validation loss equal or lower than human inter-rater loss. The gradability task criteria were defined as accuracy above 95%. Both regression tasks utilized mean square error loss function, while the classification model optimized over the binary cross entropy loss function. Training of all models was completed via the FastAI framework,³² while the in-built data augmentations functionality was utilized to improve accuracy and generalizability. The specifics of which augmentations were used can be found in Table S1. It should be noted that the regression task for VDD was dependent on image scale, and as such, augmentations that introduced scaling were omitted. Training was carried out in two stages: the first involved freezing the pre-trained weights and only training the task head and in the second, the “fine tuning” stage, all model weights were unfrozen. Each stage was trained with cyclical training rate, as described elsewhere,³³ and performed until the validation loss reached a plateau.

Optic nerve head parameters, intraocular pressure, and glaucoma risk across different ancestries

Previous studies have reported differences in VCDR and VDD values across different ancestry groups.^{34,35} Taking advantage of the diverse ancestries available in UKB and CLSA, we compared our AI-derived VCDR and VDD values, as well as IOP

(corneal-compensated²⁶) values across different ancestry groups. We used the K-means clustering method to define ancestry groups on the basis of genetic data (detailed above). We used boxplots to show the differences of optic nerve head measurements across different ancestry groups (e.g., median value and upper and lower quartiles). The mean values of VCDR across different ancestries were estimated after adjusting for age, sex, and VDD. The 97.5th percentile of optic nerve head measurements and its 95% confidence interval (2.5% to 97.5% quantiles) were also calculated on the basis of 1,000 bootstrapped samples, on account of the substantially smaller sample size for individuals of African, East Asian, and South Asian ancestry. We then investigated how VCDR and IOP relate to glaucoma risk in different ancestries. The definition of glaucoma cases and controls was detailed in our previous study.²³ Briefly, in UKB, glaucoma cases were ascertained from International Classification of Diseases diagnosis, record-linkage data from general practitioners, and self-reported previous diagnosis. In the CLSA, participants were interviewed in-person with the question “has a doctor ever told you that you have glaucoma.” We used logistic regression models to evaluate the association between genetically defined ancestry groups and glaucoma risk. In each different model, we adjusted different covariates to evaluate the association of ethnic groups and glaucoma risk. In the base model, only sex and age were adjusted for; the other models also include either IOP, VCDR, or both (IOP and VCDR).

Genome-wide association analysis and meta-analysis

For both UKB and CLSA, we carried out the VCDR and VDD GWAS association tests by using a linear mixed model (with BOLT-LMM v.2.3)³⁶ to account for cryptic relatedness and population stratification, adjusting for sex and age. We also included the first ten PCs in the model to speed up the convergence of computations.³⁷ The average values of measurements from left and right eyes and multiple visits (if available) were used and were first transformed via a rank-based inverse-normal method before association tests.³⁸ To account for optic disc size covariation, we adjusted VCDR grading for VDD in GWAS analyses.^{8,39} The VDD-adjusted VCDR and VDD GWASs results from UKB and CLSA were then meta-analyzed with those from the IGGC on the basis of the inverse variance-weighted method (METAL software 2011-03-25 release).⁴⁰ We also conducted association tests for VCDR and VDD in African and South Asian populations in UKB. Because of the relatively small size of each of these populations (Table S2, less than the recommended sample size of 5,000 in BOLT-LMM), PLINK was used instead, after removing related individuals.⁴¹

SNP-based heritability was calculated by LD score regression (LDSC) from GWAS summary statistics.^{42,43} We used bivariate LD score regression to estimate the genetic correlation between pairs of traits in European ancestry.⁴² We selected independent SNPs on the basis of the PLINK clumping method with p value $< 5 \times 10^{-8}$, $r^2 < 0.01$, and a window of 1 Mb from the index variant.⁴¹ To define “novel loci” from the AI-based GWAS, we checked previous UKB VCDR and VDD GWASs based on clinician gradings;^{8,23} we also looked up the proxy SNPs ($r^2 > 0.8$) of top loci and their nearest genes in the GWAS Catalog.⁴⁴

Cross-population genetic effects on optic nerve head parameters

We evaluated the effects of genetics variants on VCDR and VDD across different populations on the basis of the following methods. First, we compared and replicated the AI-based top loci

from European ancestry with the GWAS from African and South Asian samples. The effect sizes and standard errors of top loci were shown in a scatterplot for different ancestries. Second, we calculated the trans-ethnic genetic effect correlation for VCDR and VDD by using the “Popcorn” package.⁴⁵ Specifically, the GWAS summary statistics for VCDR and VDD from European ancestry were compared with that in Asian and African ancestry.

Transcriptome-wide association study and pathway analysis

To prioritize potential causal genes, we performed transcriptome-wide association study (TWAS) analysis in FUSION by using GWAS summary statistics and retina gene expression data.⁴⁶ In FUSION, we used reference data with both gene expression and genetic variants (SNPs) to train predictive models, which we used to impute the expression-trait association directly from large-scale GWAS summary statistics.⁴⁶ The weights of retina gene expression were obtained from 406 individuals from Eye Genotype Expression (EyeGEx) database.^{46,47} We also used the EyeGEx database to perform a summary data-based Mendelian randomization (SMR) to investigate the association of gene expression levels (exposure) and phenotype (outcome).⁴⁸ We used the heterogeneity in dependent instruments (HEIDI) tests to evaluate the null hypothesis that a single causal variant affects both gene expression and outcome, and the significance threshold was set at 0.05 ($P_{\text{HEIDI}} \geq 0.05$, not reject the null hypothesis).⁴⁸ Pathway analyses were conducted in MAGMA as implemented in FUMA (v.1.3.6).^{49,50} All other analyses were performed with R software.⁵¹

Results

Study data and performance of the trained AI model

In the UKB, 85,736 participants had at least one fundus retinal image and there was a total of 175,770 images available (Table 1). The mean age at baseline was 57.0 (SD = 8.1) years, and 54% were women. In the CLSA cohort, 29,635 participants with 106,330 images were included in analysis; of these participants, 50% were women and the mean age at recruitment was 62.6 (SD = 10.0) years.

We first trained a convolution neural network to assess whether each image was gradable in the UKB training sample. We found that most participants (>95%) had gradable images in the UKB and the CLSA cohort (Figure S3). We then predicted the measurements of both VCDR and VDD and compared the AI-based measures with clinician gradings. The AI-based VCDR and VDD measurements exhibited a higher concordance to clinician gradings compared with previous gradings by two clinicians.^{8,23,52,53} For instance, the Pearson’s correlation coefficient of the VCDR measurements in the UKB samples was 0.81 (95% confidence interval [CI] = 0.80–0.81), and it was 0.84 (95% CI = 0.82–0.86) for an independent Canadian dataset (CLSA) (Figure S4). We therefore speculated that with the improved accuracy of VCDR and VDD measurements and the larger number of images graded, the optic nerve head assessment would increase the power for genetic discovery.

Table 1. Characteristics of retinal fundus images from the UK Biobank and Canadian Longitudinal Study on Aging participants

Variable		UKB	CLSA
Number of images		175,770	106,330
Number of participants		85,736	29,635
% with at least one gradable image		95%	99%
Sex	women (%)	44,017 (54%)	14,941 (51%)
Age at recruitment	mean (SD), years	57 ± 8	63 ± 10
Vertical cup-disc-ratio	unit in 0–1	0.37 ± 0.14	0.35 ± 0.15
Vertical disc diameter	unit in pixel count	129.0 ± 10.5	121.4 ± 10.6

Abbreviations are as follows: CLSA, Canadian Longitudinal Study on Aging cohort; SD, standard deviation; UKB, UK Biobank.

Optic nerve head parameters and intraocular pressure across different ancestries

We compared AI model-derived VCDR and VDD measurements across different genetically defined ancestry groups. VDD was similar across three ancestral groups (Europeans, East Asians, and South Asians) and larger in Africans (Figures 2B and 2E). On average, after adjusting for age, sex, and VDD, VCDR was markedly higher in Asians and Africans than it was in Europeans (similar results in UKB [Figure 2A] and in CLSA [Figure 2D]). A different ancestry-based trend was also observed for IOP; relative to Europeans, South Asians had similar IOP, East Asians had lower IOP, and Africans had higher IOP (Figures 2C and 2F).

We then examined whether the systematically assessed VCDR, VDD, and IOP can explain the observed prevalence of glaucoma seen across different ancestries in the UK and Canada. Figure 3 shows the glaucoma risk of Africans, East Asians, and South Asians with European ancestry (the most common ancestry in UKB and CLSA datasets) as the baseline. Consistent with previous epidemiological studies, Africans have the highest glaucoma risk (Figure 3 base model, correcting for only age and sex OR = 2.5 relative to the reference of Europeans). As seen in Figure 2, Africans have higher VCDR and higher IOP than Europeans, and when these were corrected for, the glaucoma risk approached that of Europeans in both CLSA and UKB. East Asians had a similar base model risk to Europeans, although the contribution of IOP and VDR differs; on average, their IOP is lower and their VCDR is larger (Figure 2), and there was a pattern of glaucoma risk changing as either IOP alone or VCDR alone were adjusted for in the regression model. Adjusting for both IOP and VCDR, the risk of glaucoma in East Asians was not significantly different to Europeans, suggesting that the higher VCDR and lower IOP in this group relative to Europeans counteract each other, explaining the similar glaucoma incidences between these ancestries. Interestingly, in South Asians, IOP is similar to Europeans, but VCDR is higher (Figure 2). This means that South Asian base model risk does not change when IOP is included in the model, but when VCDR is included, the glaucoma risk decreases to become indistinguishable from the incidence in Euro-

peans. In summary, by examining individuals of varying ancestry living in the UK and Canada, we show that relative to European ancestry, African ancestry glaucoma incidence is driven by both elevated VCDR and IOP, East Asian ancestry glaucoma is driven by elevated VCDR but ameliorated by lower IOP, and finally, South Asian glaucoma is driven by elevated VCDR but not by changes in IOP (relative to that in Europeans).

AI-based phenotypes greatly increase SNP-based heritability and identify more loci

In the GWAS of VDD-adjusted VCDR, 145 and 19 statistically independent genome-wide significant SNPs were respectively identified in the UKB alone and CLSA alone (Figure S5). The results for the VCDR GWAS are essentially the same after further adjusting for both VDD and IOP (Figure S6). The analogous numbers of SNPs for VDD were 142 and 17 for UKB and CLSA, respectively. We found weak evidence of genomic inflation from linkage disequilibrium score regression (Table S3). From UKB, the AI-based GWAS of VDD-adjusted VCDR and VDD identified substantially more loci than our previous GWAS based on clinician gradings (76 for VDD-adjusted VCDR and 91 for VDD).^{8,23} Strikingly, the SNP-based heritability increased by ~50% for VCDR and VDD (Figure S7). For instance, the SNP-based heritability for VCDR was 0.22 from clinician gradings (only single measure), whereas the heritability increased to 0.35 from AI-based GWASs (average of multiple measures). The increased heritability indicated that AI-based phenotyping was substantially cleaner than clinician gradings, which may be a result of two aspects: (1) higher accuracy of AI-based gradings and (2) improved accuracy from multiple measures per individual. We further tested the hypothesis in UKB and CLSA by using only one measure per individual from AI-based gradings. The SNP-based heritability from a single measure (left or right eyes in the baseline or first follow-up visit) was ~0.3, which is roughly in the middle of heritability estimation from clinician gradings and AI-based multiple measures (Figure S7). These results indicate the higher accuracy of AI-based single measure per individual contributes to the increase of heritability estimation and averaging of multiple measures per individual can further increase the

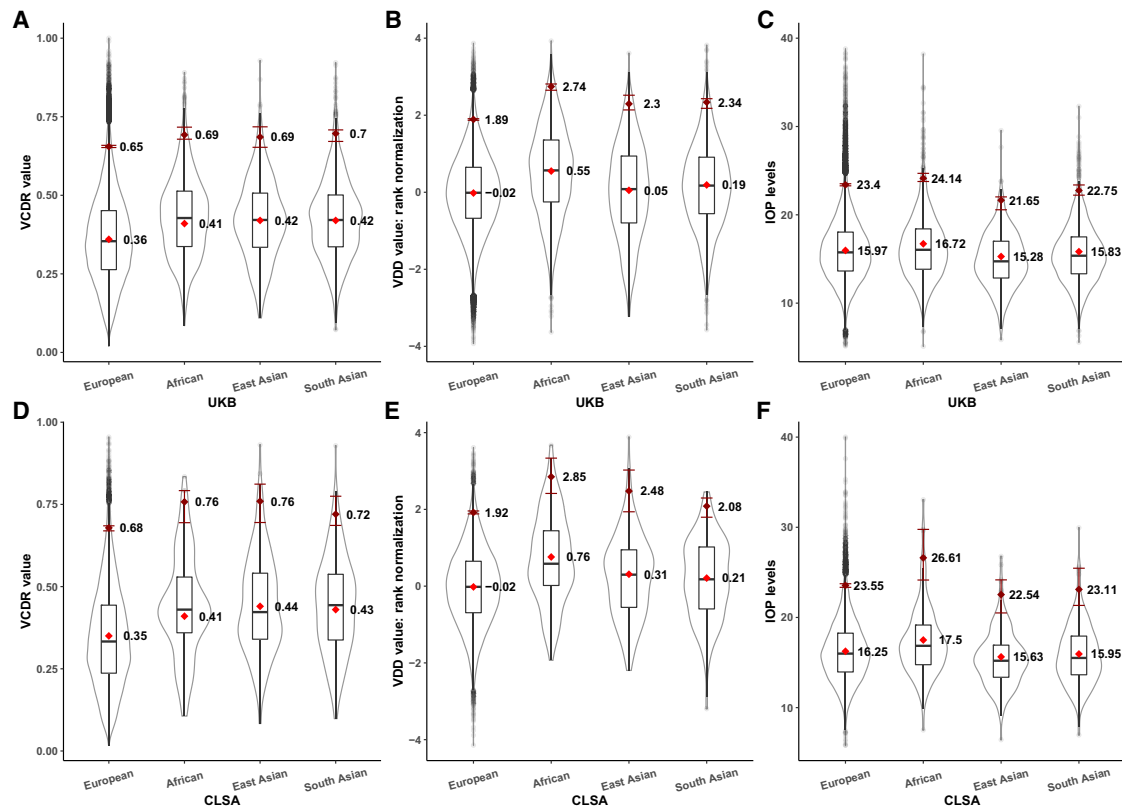


Figure 2. Optic nerve head measurements and intraocular pressure across different ancestry groups

(A) The boxplot for VCDR values from different ancestry groups in UK Biobank. The sample sizes of European, African, East Asian, and South Asian are 66,838, 2,220, 567, and 2,087, respectively. The box represents median value with first and third quartiles. The red diamond is the mean value of VCDR after accounting for age, sex, and VDD, where the mean value is annotated as text. The dark red diamond is the 97.5th percentile of VCDR value. The dark red error bar is the 95% confidence interval (2.5% to 97.5% quantiles) of the 97.5th percentile based on 1,000 bootstrapped samples, which is essential for CLSA data where the sample sizes for African, East Asian, and South Asian were substantially smaller (132, 213, and 213, respectively), and the sample size of white ancestry is 18,123.

(B) The boxplot for VDD values from different ancestry groups in UK Biobank. Because of the scale from fundus images, the VDD was rank normalized (mean = 0, SD = 1). The red diamond is the mean value of VDD after accounting for age and sex.

(C) The boxplot for IOP levels from different ancestry groups in the UK Biobank (truncated at 40 mm Hg, with 15 participants between 40–60 mm Hg).

(D–F) (D), (E), and (F) show the boxplots for VCDR, VDD, and IOP in the CLSA cohort, respectively.

heritability. Consistent with our previous study,²³ correcting for VDD in the VCDR GWAS also improved the relevance to glaucoma; there was a higher genetic correlation with glaucoma in the VDD-adjusted VCDR compared with the unadjusted VCDR GWAS (genetic correlation $r_g = 0.502$ versus 0.457 in UKB and 0.543 versus 0.481 in CLSA).

Validation AI-based GWASs

We then compared AI-based and clinician grading-based GWASs by using independent samples from the IGGC. The concordance of SNP effect sizes of top SNPs between the AI-based and clinician gradings was essentially one (Figures 4A and 4D), and nearly all previously published loci using clinician ratings were replicated. The estimated effect sizes at the top SNPs from AI-based GWASs were also highly concordant between UKB and CLSA (Figures 4B and 4E). When combining UKB and CLSA AI-based GWASs, we identified 193 and 188 loci for VDD-adjusted

VCDR and VDD, respectively, again exhibiting very high concordance with IGGC (Figures 4C and 4F). The high concordance and more loci support the better-powered GWAS from AI-based measurements.

New genetic discovery of optic nerve head measures, cross-ancestry comparison, and implications for glaucoma

To maximize power for locus discovery, we combined UKB, CLSA, and IGGC GWASs (European ancestry) and identified 230 and 231 independent genome-wide significant SNPs for VDD-adjusted VCDR and VDD, respectively (Figure 5). Of them, we found 111 and 107 novel loci for VDD-adjusted VCDR and VDD, respectively (Tables S4 and S5). The pleiotropy effects of VCDR and VDD lead SNPs on other traits were looked up in the GWAS Catalog (Table S6). We then compared the effect sizes of top VDD-adjusted VCDR and VDD loci across different ancestries (Asian and African GWASs); because of the much

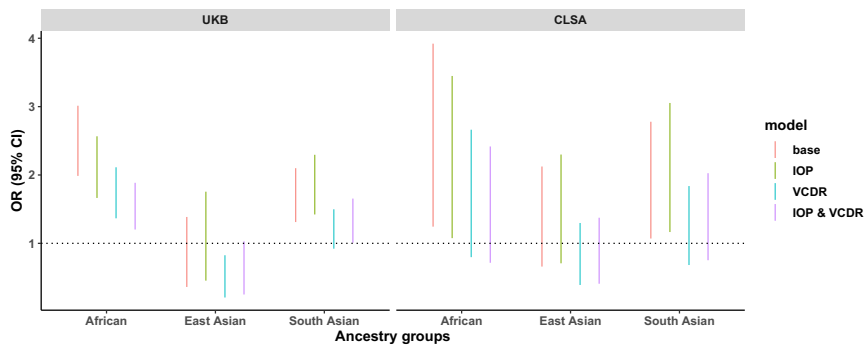


Figure 3. Glaucoma risk across different ancestry groups

The figure shows the risk of glaucoma in different ancestry groups. The horizontal line at OR = 1 is the reference for European ancestry. The y axis is the odds ratio (OR) and 95% confidence interval (CI) for three ethnic groups (African, South Asian, and East Asian). In each different model, we adjusted different covariates to evaluate the association of ethnic groups and glaucoma risk. In the base model, only sex and age were adjusted for; the other models also include either IOP, VCDR, or both (IOP and VCDR).

smaller available sample sizes, their confidence intervals were very large, however the clear linear trend indicated the loci identified from European ancestry also had an effect on Asian populations (Figures 6A and 6B; for VCDR and VDD the Pearson's correlation coefficient is 0.65 [p value 3.6×10^{-27}] and 0.62 [p value 9.3×10^{-23}], respectively). The sample size of African ancestry was much smaller than Asian ancestry ($N = 2,245$ versus 8,373 for VCDR) and showed a lower concordance (Figure S8). The genetic correlations across the genome were essentially one on the basis of the Popcorn approach for VCDR and VDD (Table S7). We also compared the effect sizes of VDD-adjusted VCDR top loci with their effect sizes on glaucoma (Figure 6C) and found a relatively high concordance (Pearson's correlation coefficient 0.71, $p = 4.1 \times 10^{-37}$). Of the 230 VCDR (adjusted for VDD) loci, 205 (89%) were in the same direction, 131 were associated with glaucoma at a nominal significance level ($p < 0.05$), and 68 were associated with glaucoma after Bonferroni correction ($p < 0.05/230 = 2.2 \times 10^{-4}$, the nearest gene names are highlighted in Figure 6C, e.g., *RBPMS*, *AFAP1*, *LMX1B*, *ABCA1*, *CAV1*, and *GAS7*).

Gene prioritization and pathway analysis

We performed TWAS analysis in FUSION on the basis of the VDD-adjusted VCDR and VDD GWAS summary statistics and retinal gene expression data. For VDD-adjusted VCDR, we identified 101 genes that were significant after Bonferroni correction for multiple testing, nine of which were not genome-wide significant in the per-SNP analysis (Figures S9A and S9B). For VDD, we identified 64 genes that were significant after Bonferroni correction for multiple testing, 13 of which were not genome-wide significant in the per-SNP analysis. From SMR analysis, we identified 29 and 24 genes for VDD-adjusted VCDR and VDD, respectively, that were significant after multiple testing. We also compared the genes identified from both FUSION and SMR, and 11 and eight genes overlap from the two methods for VDD-adjusted VCDR and VDD, respectively (Figures S9C and S9D). For instance, of the 11 genes that were associated with VDD-adjusted VCDR for the two approaches, six genes also passed the HEIDI tests (*P4HTM*, *SNX32*, *RASGRF*, *HAUS4*, *LRP11*, and *AC012613.2*), suggesting the effects on VCDR may be mediated via these gene expression in

retina tissue. The large increase in power resulting from the use of AI grading to improve accuracy and enable substantially larger datasets with multiple images per participant meant we were able to discover many new biological pathways influencing optic nerve head development and aging. Our pathway enrichment analysis uncovered 65 pathways for VCDR and 82 pathways for VDD after Bonferroni correction for multiple testing (Tables S8 and S9). For VCDR, as well as extracellular matrix pathways uncovered by our previous work, these new pathway analyses uncovered associations with telencephalon (forebrain) regionalization, embryo development, and anatomical tube development. There were several unexpected but statistically robust associations with kidney development (e.g., GO mesonephros development, $p_{\text{raw}} = 3.45 \times 10^{-8}$, $p = 0.00053$ after correction for multiple comparisons). The genes driving the kidney development pathway enrichment included *BMP2*, *BMP4*, *EYA1*, *FAT4*, *FOXC1*, *GLI3*, *PAX2*, *RARB*, *SIX1*, and *SALL1*. Several kidney pathways were also significant in the pathway enrichment analysis applied to our VDD GWAS.

Discussion

Our results show the promising application of AI algorithms in genetics studies. Large-scale biobanks such as UKB and CLSA have accumulated hundreds of thousands of optic nerve images containing important information for glaucomatous optic neuropathy. However, the time-intensive and moderate agreement of manual assessment has impeded the usage of retinal fundus images. We trained a deep learning model by using clinically estimated VCDR and VDD and found the trained model has a high accuracy. The large-scale biobank data for both VCDR and IOP allow us to systematically compare the glaucoma risk and optic nerve head parameters across different ancestries. Combining genetic and image data, we doubled the number of loci for both VCDR and VDD and did so with increased heritability.

The scope of available deep learning models for computer vision tasks is extensive and continuously developing. Various approaches to grade fundus images often utilize intricate data preprocessing methods^{54–56} as well as

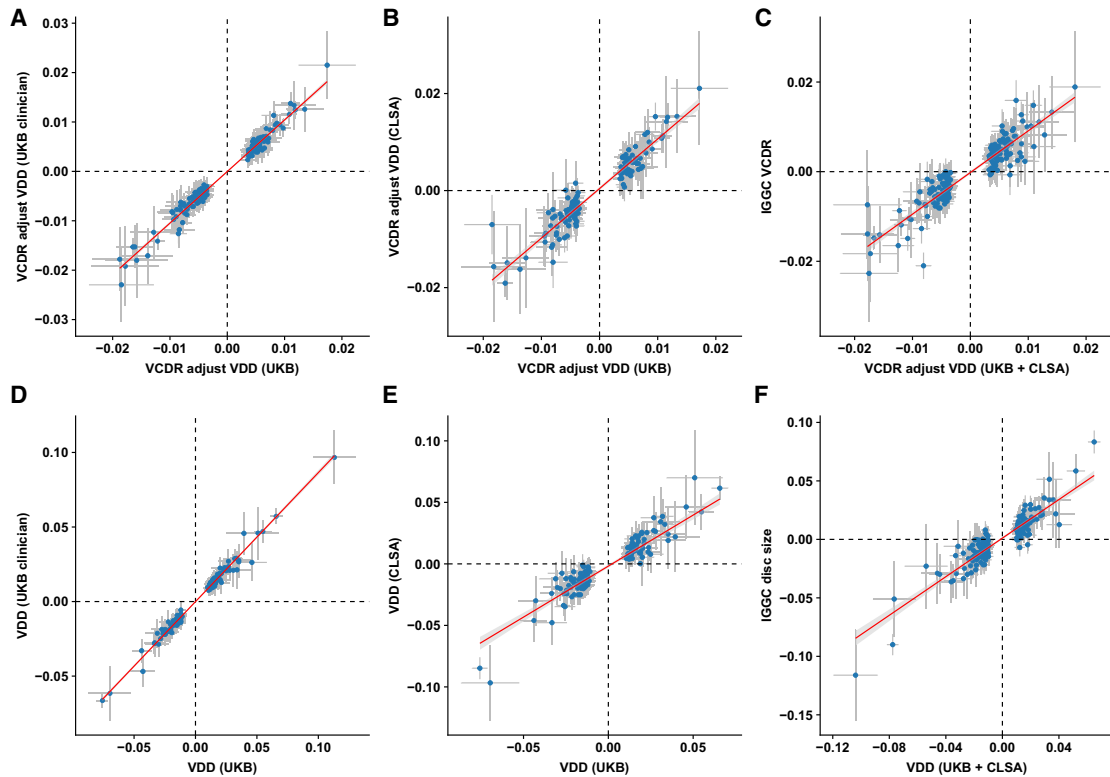


Figure 4. Validation AI-based GWAS

The figure shows the effect sizes for VDD-adjusted VCDR and VDD from different datasets. The vertical and horizontal error bars are the 95% confidence interval for SNP effect sizes. The red line is the best fit line and the 95% confidence interval region is in gray.

computationally heavy models and training methods.^{17,57}

In the instance of statistically powered, large-scale population study, fast inference and quick iterations are key, making heavy computational and design costs harder to justify. Here, we demonstrate that a relatively lightweight, pre-trained CNN model is capable of producing highly accurate estimations of VCDR and VDD as evinced by high correlation with clinical grading, improved genetic discovery, and further validations in independent samples. Although our model was trained on data from the white British ancestry group who form the bulk of UKB, we believe our approach is likely to perform acceptably in other ancestries. As part of the CNN modeling process, we augmented and perturbed the fundus images to increase robustness. Moreover, in the non-European ancestries we examined, the range of values seen for VCDR (or VDD) was broadly similar to that in Europeans. We were further reassured by the fact that in the models trained in UKB white British samples, our approach performed well in the independently collected CLSA data.

Our AI approach has dramatically accelerated the pace of genetic discoveries. In our previous work,^{8,23} we laboriously manually assessed a subset of UKB images. With the deep learning model trained on clinical measurements, we were able to predict on a new subject within a fraction of a second, making time and effort of image labeling trivial, even when applied to large-scale datasets (~1 h for ~0.3 million images). Sample size is one of the most important limiting factors for genetic discovery. Leveraging the

AI-based algorithm and large-scale data, we were able to conduct the most powerful GWAS of optic nerve head parameters to date. We doubled the number of genome-wide significant loci for both VCDR and VDD. Interestingly, the estimated SNP-based heritability also increased by ~50% for VCDR and VDD (Figure S7); the estimate for VCDR is not substantially lower than the heritability estimate from twin studies (~50%), although given more accurate (AI-based) phenotypes, the twin study-based heritability estimate may increase. The increased heritability is a result of more accurate measurements, which arise in part because of the higher accuracy of AI-based predictions and in part because of the AI approach's allowing time-efficient grading of multiple measures per individual.

Many of the newly identified VCDR genes are associated with other eye traits (e.g., glaucoma, IOP, exfoliation syndrome, myopia). For some loci associated with IOP, it is likely that they have an effect on VCDR as a secondary effect of the locus first acting on IOP. Loci including genes such as *ABCA1*, *CAVI*, *AFAP1*, and *LMX1B* were associated with VCDR for the first time; a likely explanation for this association is that the associated variant alters IOP and subsequently VCDR. Over 20 of the VCDR loci are also associated with refractive error and multiple aspects of eye physiology are most likely involved (axial length, corneal thickness, retinal ganglion cell function). We also found a significant genome-wide genetic correlation between VCDR (adjusted for VDD) and myopia ($rg = 0.3$, $p = 1 \times 10^{-14}$), as well as

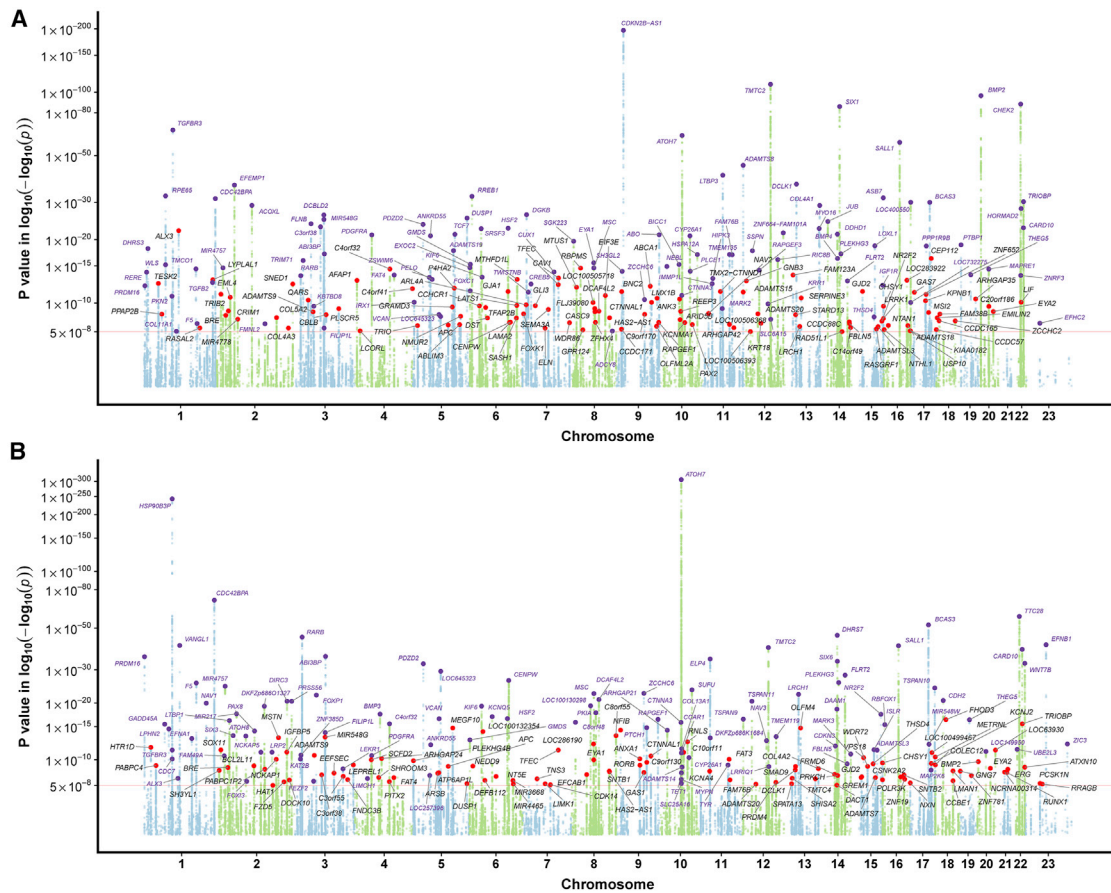


Figure 5. AI enables new genetic discovery for optic nerve head measures

(A and B) Manhattan plot in (A) shows p values for VDD-adjusted VCDR from the meta-analysis of UKB, CLSA, and IGGC (European ancestry). (B) shows p values for VDD from the meta-analysis of UKB, CLSA, and IGGC (European ancestry). The red horizontal line is the genome-wide significance level at $p = 5 \times 10^{-8}$. SNPs with p value less than 1×10^{-4} are not shown in the Manhattan plot. Previously unknown loci are highlighted with red dots, and the nearest gene names are in black text. Known SNPs are highlighted with purple dots, and the nearest gene names are in purple text.

with well-studied traits that are associated with myopia, such as years of education.⁵⁸

In addition, several of the new VCDR genes provide possible links to retinal ganglion cell biology and they may constitute possible drug repositioning candidates. There are too many to discuss individually, but one SNP of interest is rs17855988; this missense variant in the elastin gene (*ELN*) has been associated with facial aging. Elastin in the sclera is most dense around the optic nerve head,⁵⁹ and *ELN* expression has been shown to be high in exfoliation glaucoma lens.⁶⁰ A subset of the VDD loci have been found to be associated at genome-wide significance levels in previous glaucoma GWASs. However, in the majority of cases, the association with glaucoma appears to be driven by the lead SNP's having a primary effect on VCDR (where the variance explained in VCDR for the peak SNP is larger than that for VDD: e.g., SNPs in or near *GMDS*, *CAV1*, *MYOF*, *SIX6*, *CHEK2*, *TMTC2*). Hence, the primary link between the disc parameters and glaucoma is via VCDR rather than via VDD. This is also shown in the lower genetic correlation between glaucoma and VDD ($rg = 0.01$) compared with glaucoma and VCDR ($rg = 0.5$).^{8,23} With the

high genetic correlation between VCDR and glaucoma, a multitrait analysis has recently shown that including VCDR can improve the power to identify glaucoma genes and to enable the development of polygenic risk score.²³ Future studies of glaucoma would benefit from incorporating these accurate AI-derived VCDR estimates.

Previous studies have looked at the differences between VDD across different ancestries.^{34,35} Our results were consistent with these: Africans had the largest disc size, followed by those of Asian ancestry. For VCDR, an early study (100 Black individuals and 100 white individuals) found that Black individuals had larger VCDR (mean values: Black = 0.35, white = 0.24).⁶¹ A subsequent larger study (1,534 Black individuals and 1,853 white individuals) reported larger VCDR in Black individuals (mean values: Black = 0.56, white = 0.49).⁶² A subsequent study in three different Asian ancestries showed that VCDR values were similar between the studied ancestries (mean VCDR 0.40, 0.42, and 0.40 in Malay, Chinese, and Indian, respectively).⁶³ It is striking that despite VCDR's theoretically being a simple parameter to assess, the mean VCDR varies widely across studies, possibly because of differences in

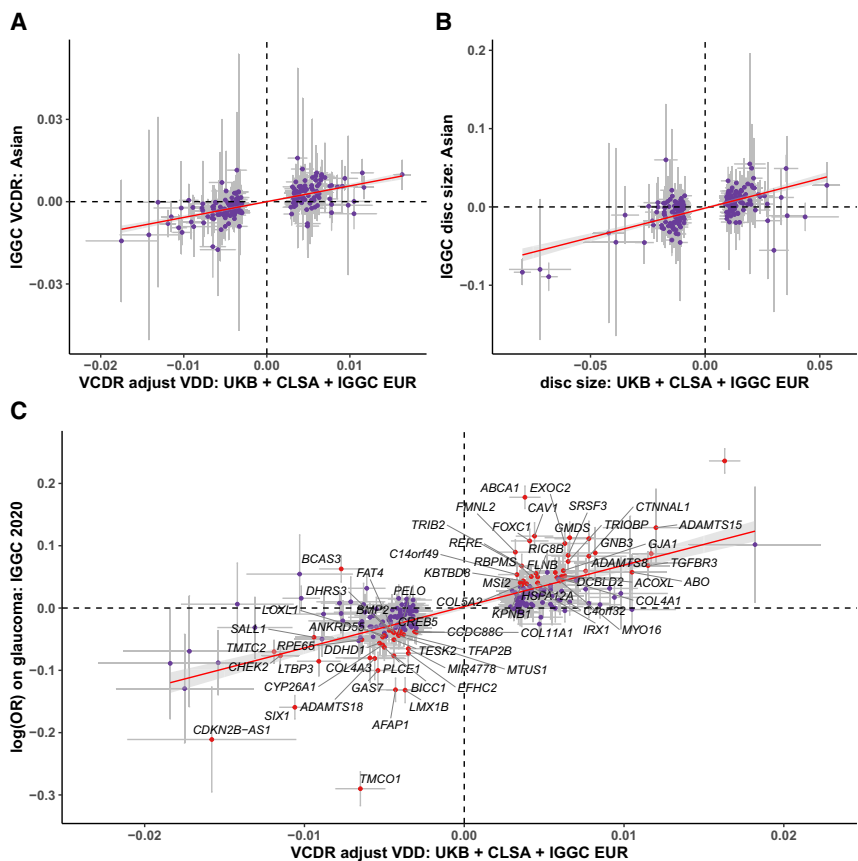


Figure 6. Comparison of the effect sizes for VCDR (adjusted for VDD) and VDD lead SNPs versus those observed in the Asian ancestry group and in independent glaucoma cohorts

(A–C) (A) and (B) show the effect sizes for lead VCDR (adjusted for VDD) and VDD loci (European versus Asian population). (C) shows the effect sizes for VCDR (adjusted for VDD) lead SNPs versus log odds ratio in meta-analysis of IGCC glaucoma GWAS.²¹ The 68 SNPs associated with glaucoma after Bonferroni correction ($p < 0.05/230 = 2.2 \times 10^{-4}$) are highlighted with red dots, and the nearest gene names are in black text.

where available, we have removed known cases of ACG in the broad glaucoma definition, but some ACG cases will remain.

A strength of our study is that we used a large number of clinically assessed images to train the deep learning model for VCDR and VDD; this allowed us to generate accurate predictions. Our study has shown that the AI-based measurements have a high accuracy. The AI-based optic nerve head assessment has also boosted the avail-

able sample size and dramatically expanded gene discovery for these key ocular phenotypes. We show that this deep learning model can also be used to assess future fundus images automatically and rapidly, especially in population-based studies with a large number of images. Moreover, the implementation of transfer-learning techniques shows that AI-aided labeling, with adequate sample size, has a potential to deliver equally successful genetic discoveries in other image-based biological phenotypes. Our study has several limitations. First, although our AI approach was able to grade a large proportion of images (particularly in the CLSA), a small proportion remained ungradable because of poor picture quality. Future studies could explore adversarial architectures to improve clinical ratings of VCDR and VDD. However, a set of high-quality truth labels would still be necessary for initial pre-training. Moreover, although we were able to use genetic data to clearly identify the major ancestries within UKB and CLSA (European, African, South Asian, and East Asian), there remained a group of uncategorized individuals with mixed ancestries that we did not include in our epidemiological or genetic analyses. Finally, we have shown the generalizability of the trained CNN model in UKB to the independent datasets in the CLSA (fundus images from the two studies were taken with different instruments from the Topcon), however, the robustness of the model to other different platforms/brands still needs further investigation.

measurement protocol, sex, age, and eye disease status. A further study⁴ looked at the 97.5th percentile of VCDR instead of the mean and reported broadly similar values in the Netherlands (0.73), Bangladesh (0.7), Mongolia (0.70), Singapore (0.7), and Tanzania (0.7). A major advantage of our study is that we use our AI-derived gradings in two population-based cohort studies to systematically assess VCDR differences across ancestries in a consistent study design. By leveraging large sample sizes, we are able to clearly show both Asian and African ancestry individuals have larger VCDR values than Europeans. Our primary results in Figure 2 correct VCDR for VDD, given previous studies' showing that correcting for VDD enhances the relevance to glaucoma.⁶⁴

The raised VCDR in Asian and African ancestry individuals living in the UK and Canada is in keeping with elevated glaucoma rates in these ancestries.⁶⁵ When combined with data on IOP, a combination of VCDR and IOP explains the vast majority of the variation between glaucoma rates in Europeans relative to Africans, South Asians, and East Asians. Although crucially, our data show (Figure 3) that the relative contributions of VCDR and IOP are clearly different between all four major population groups that we consider. For individuals of European, South Asian, or African ancestry, the vast majority of broadly defined glaucoma cases are open angle glaucoma (OAG). In East Asia, angle closure glaucoma (ACG) is common, and a limitation of our analysis is that we cannot distinguish between ACG and OAG in all cases—

To conclude, we showed that AI-based optic nerve head assessment has a high accuracy, and this greatly improves

our power to discover new genes. These findings provide insights into the pathogenesis of glaucomatous optic neuropathy. We also use the systematic assessment of VCDR across different ancestries to help explain how the pattern of IOP and VCDR measures underpin observed glaucoma risk; such findings in mixed ancestry groups living in the UK and Canada help explain the differing characteristics of glaucoma across ancestries. For example, relative to Europeans, individuals with East Asian ancestry are more likely to have lower IOP and increased VCDR. Given these East Asians are genetically similar to East Asians in countries such as China and Japan, this provides support for the assertion that normal tension forms of glaucoma predominate in East Asia because of genetic predisposition for high VCDR despite low IOP.

Data and code availability

UK Biobank data are available through the UK Biobank Access Management System (<https://www.ukbiobank.ac.uk/>). We will return the derived vertical cup-to-disc ratio and vertical disc diameter values following the UK Biobank policy, and in due course, they will be available through the UK Biobank Access Management System. Data are available from the Canadian Longitudinal Study on Aging (<https://www.clsa-elcv.ca>) for researchers who meet the criteria for access to de-identified CLSA data. Relevant code can be found at the following GitHub repository: <https://github.com/MaxKelsen/Glaucoma-Phenotype-ML-estimation>. The GWAS summary statistics from meta-analysis are available for research use at <https://xikunhan.github.io/site/publication/>.

Supplemental information

Supplemental information can be found online at <https://doi.org/10.1016/j.ajhg.2021.05.005>.

Acknowledgments

The opinions expressed in this manuscript are the authors' own and do not reflect the views of the Canadian Longitudinal Study on Aging. This work was conducted using the UK Biobank Resource (application number 25331). Full acknowledgments are provided in the [supplemental information](#).

Declaration of interests

K.S., C.B., M. Tremere, and M. Trzaskowski are employees of Max Kelsen. M. Tremere is the head of ML Delivery and M. Trzaskowski is the head of research at Max Kelsen.

Received: December 28, 2020

Accepted: May 10, 2021

Published: June 1, 2021

Web resources

The Canadian Longitudinal Study on Aging (CLSA), <https://www.clsa-elcv.ca/researchers/data-support-documentation>
UK Biobank, <https://www.ukbiobank.ac.uk/>

References

1. Meyer, C.H., Rodrigues, E.B., and Schmidt, J.C. (2003). Congenital optic nerve head pit associated with reduced retinal nerve fibre thickness at the papillomacular bundle. *Br. J. Ophthalmol.* 87, 1300–1301.
2. Kwon, Y.H., Fingert, J.H., Kuehn, M.H., and Alward, W.L. (2009). Primary open-angle glaucoma. *N. Engl. J. Med.* 360, 1113–1124.
3. Jonas, J.B., Aung, T., Bourne, R.R., Bron, A.M., Ritch, R., and Panda-Jonas, S. (2017). Glaucoma. *Lancet* 390, 2183–2193.
4. Foster, P.J., Buhmann, R., Quigley, H.A., and Johnson, G.J. (2002). The definition and classification of glaucoma in prevalence surveys. *Br. J. Ophthalmol.* 86, 238–242.
5. Klein, B.E.K., Klein, R., and Lee, K.E. (2004). Heritability of risk factors for primary open-angle glaucoma: the Beaver Dam Eye Study. *Invest. Ophthalmol. Vis. Sci.* 45, 59–62.
6. Springelkamp, H., Mishra, A., Hysi, P.G., Gharahkhani, P., Höhn, R., Khor, C.-C., Cooke Bailey, J.N., Luo, X., Ramdas, W.D., Vithana, E., et al.; NEIGHBORHOOD Consortium (2015). Meta-analysis of Genome-Wide Association Studies Identifies Novel Loci Associated With Optic Disc Morphology. *Genet. Epidemiol.* 39, 207–216.
7. Springelkamp, H., Iglesias, A.I., Mishra, A., Höhn, R., Wojciechowski, R., Khawaja, A.P., Nag, A., Wang, Y.X., Wang, J.J., Cuellar-Partida, G., et al.; NEIGHBORHOOD Consortium (2017). New insights into the genetics of primary open-angle glaucoma based on meta-analyses of intraocular pressure and optic disc characteristics. *Hum. Mol. Genet.* 26, 438–453.
8. Han, X., Qassim, A., An, J., Marshall, H., Zhou, T., Ong, J.-S., Hassall, M.M., Hysi, P.G., Foster, P.J., Khaw, P.T., et al. (2019). Genome-wide association analysis of 95 549 individuals identifies novel loci and genes influencing optic disc morphology. *Hum. Mol. Genet.* 28, 3680–3690.
9. Tielsch, J.M., Katz, J., Quigley, H.A., Miller, N.R., and Sommer, A. (1988). Intraobserver and interobserver agreement in measurement of optic disc characteristics. *Ophthalmology* 95, 350–356.
10. Varma, R., Steinmann, W.C., and Scott, I.U. (1992). Expert agreement in evaluating the optic disc for glaucoma. *Ophthalmology* 99, 215–221.
11. Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., Cui, C., Corrado, G., Thrun, S., and Dean, J. (2019). A guide to deep learning in healthcare. *Nat. Med.* 25, 24–29.
12. Gulshan, V., Peng, L., Coram, M., Stumpe, M.C., Wu, D., Narayanaswamy, A., Venugopalan, S., Widner, K., Madams, T., Cuadros, J., et al. (2016). Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *JAMA* 316, 2402–2410.
13. Yan, Q., Weeks, D.E., Xin, H., Swaroop, A., Chew, E.Y., Huang, H., Ding, Y., and Chen, W. (2020). Deep-learning-based Prediction of Late Age-Related Macular Degeneration Progression. *Nat. Mach. Intell.* 2, 141–150.
14. U.S. FDA (2018). FDA permits marketing of artificial intelligence-based device to detect certain diabetes-related eye problems. FDA. <https://www.fda.gov/news-events/press-announcements/fda-permits-marketing-artificial-intelligence-based-device-detect-certain-diabetes-related-eye>.
15. Abràmoff, M.D., Lavin, P.T., Birch, M., Shah, N., and Folk, J.C. (2018). Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. *NPJ Digit. Med.* 1, 39.

16. An, G., Omodaka, K., Hashimoto, K., Tsuda, S., Shiga, Y., Takada, N., Kikawa, T., Yokota, H., Akiba, M., and Nakazawa, T. (2019). Glaucoma Diagnosis with Machine Learning Based on Optical Coherence Tomography and Color Fundus Images. *J. Healthc. Eng.* *2019*, 4061313.
17. Sahlsten, J., Jaskari, J., Kivinen, J., Turunen, L., Jaanio, E., Hietala, K., and Kaski, K. (2019). Deep Learning Fundus Image Analysis for Diabetic Retinopathy and Macular Edema Grading. *Sci. Rep.* *9*, 10750.
18. Sengupta, S., Singh, A., Leopold, H.A., Gulati, T., and Lakshminarayanan, V. (2020). Ophthalmic diagnosis using deep learning with fundus images - A critical review. *Artif. Intell. Med.* *102*, 101758.
19. Ash, J.T., Darnell, G., Munro, D., and Engelhardt, B.E. (2021). Joint analysis of expression levels and histological images identifies genes associated with tissue morphology. *Nat. Commun.* *12*, 1609.
20. Glastonbury, C.A., Pulit, S.L., Honecker, J., Censin, J.C., Laber, S., Yaghootkar, H., Rahmioglu, N., Pastel, E., Kos, K., Pitt, A., et al. (2020). Machine Learning based histology phenotyping to investigate the epidemiologic and genetic basis of adipocyte morphology and cardiometabolic traits. *PLoS Comput. Biol.* *16*, e1008044.
21. Gharahkhani, P., Jorgenson, E., Hysi, P., Khawaja, A.P., Pendergrass, S., Han, X., Ong, J.S., Hewitt, A.W., Segre, A.V., Rouhana, J.M., et al.; NEIGHBORHOOD consortium; ANZRAG consortium; Biobank Japan project; FinnGen study; UK Biobank Eye and Vision Consortium; GIGA study group; and 23 and Me Research Team (2021). Genome-wide meta-analysis identifies 127 open-angle glaucoma loci with consistent effect across ancestries. *Nat. Commun.* *12*, 1258.
22. Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L.T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., et al. (2018). The UK Biobank resource with deep phenotyping and genomic data. *Nature* *562*, 203–209.
23. Craig, J.E., Han, X., Qassim, A., Hassall, M., Cooke Bailey, J.N., Kinzy, T.G., Khawaja, A.P., An, J., Marshall, H., Gharahkhani, P., et al.; NEIGHBORHOOD consortium; and UK Biobank Eye and Vision Consortium (2020). Multitrait analysis of glaucoma identifies new risk loci and enables polygenic prediction of disease susceptibility and progression. *Nat. Genet.* *52*, 160–166.
24. McCarthy, S., Das, S., Kretschmar, W., Delaneau, O., Wood, A.R., Teumer, A., Kang, H.M., Fuchsberger, C., Danecek, P., Sharp, K., et al.; Haplotype Reference Consortium (2016). A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* *48*, 1279–1283.
25. Walter, K., Min, J.L., Huang, J., Crooks, L., Memari, Y., McCarthy, S., Perry, J.R.B., Xu, C., Futema, M., Lawson, D., et al.; UK10K Consortium (2015). The UK10K project identifies rare variants in health and disease. *Nature* *526*, 82–90.
26. MacGregor, S., Ong, J.S., An, J., Han, X., Zhou, T., Siggs, O.M., Law, M.H., Souzeau, E., Sharma, S., Lynn, D.J., et al. (2018). Genome-wide association study of intraocular pressure uncovers new pathways to glaucoma. *Nat. Genet.* *50*, 1067–1071.
27. Raina, P.S., Wolfson, C., Kirkland, S.A., Griffith, L.E., Oremus, M., Patterson, C., Tuokko, H., Penning, M., Balion, C.M., Hogan, D., et al. (2009). The Canadian longitudinal study on aging (CLSA). *Can. J. Aging* *28*, 221–229.
28. Raina, P., Wolfson, C., Kirkland, S., Griffith, L.E., Balion, C., Cossette, B., Dionne, I., Hofer, S., Hogan, D., van den Heuvel, E.R., et al. (2019). Cohort Profile: The Canadian Longitudinal Study on Aging (CLSA). *Int. J. Epidemiol.* *48*, 1752–1753j.
29. Bonnemaier, P.W.M., Leeuwen, E.M.V., Iglesias, A.I., Gharahkhani, P., Vitart, V., Khawaja, A.P., Simcoe, M., Höhn, R., Cree, A.J., Igo, R.P., et al.; International Glaucoma Genetics Consortium; NEIGHBORHOOD consortium; and UK Biobank Eye and Vision Consortium (2019). Multi-trait genome-wide association study identifies new loci associated with optic disc parameters. *Commun. Biol.* *2*, 435.
30. He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778.
31. Krizhevsky, A., Sutskever, I., and Hinton, G.E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems 25*, F. Pereira, C.J.C. Burges, L. Bottou, and K.Q. Weinberger, eds. (Curran Associates, Inc.), pp. 1097–1105.
32. Howard, J., and Gugger, S. (2020). Fastai: A Layered API for Deep Learning. *Information* *11*, 108.
33. Smith, L.N. (2017). Cyclical Learning Rates for Training Neural Networks. In *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 464–472.
34. Marsh, B.C., Cantor, L.B., WuDunn, D., Hoop, J., Lipyanik, J., Patella, V.M., Budenz, D.L., Greenfield, D.S., Savell, J., Schuman, J.S., and Varma, R. (2010). Optic nerve head (ONH) topographic analysis by stratus OCT in normal subjects: correlation to disc size, age, and ethnicity. *J. Glaucoma* *19*, 310–318.
35. Lee, R.Y., Kao, A.A., Kasuga, T., Vo, B.N., Cui, Q.N., Chiu, C.S., Weinreb, R.N., and Lin, S.C. (2013). Ethnic variation in optic disc size by fundus photography. *Curr. Eye Res.* *38*, 1142–1147.
36. Loh, P.-R., Tucker, G., Bulik-Sullivan, B.K., Vilhjálmsson, B.J., Finucane, H.K., Salem, R.M., Chasman, D.I., Ridker, P.M., Neale, B.M., Berger, B., et al. (2015). Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat. Genet.* *47*, 284–290.
37. Loh, P.-R., Kichaev, G., Gazal, S., Schoech, A.P., and Price, A.L. (2018). Mixed-model association for biobank-scale datasets. *Nat. Genet.* *50*, 906–908.
38. 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.
39. Bengtsson, B. (1976). The variation and covariation of cup and disc diameters. *Acta Ophthalmol. (Copenh.)* *54*, 804–818.
40. Willer, C.J., Li, Y., and Abecasis, G.R. (2010). METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* *26*, 2190–2191.
41. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., and Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* *81*, 559–575.
42. Bulik-Sullivan, B., Finucane, H.K., Anttila, V., Gusev, A., Day, F.R., Loh, P.-R., Duncan, L., Perry, J.R., Patterson, N., Robinson, E.B., et al.; ReproGen Consortium; Psychiatric Genomics Consortium; and Genetic Consortium for Anorexia Nervosa of the Wellcome Trust Case Control Consortium 3 (2015). An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* *47*, 1236–1241.
43. Bulik-Sullivan, B.K., Loh, P.-R., Finucane, H.K., Ripke, S., Yang, J., Patterson, N., Daly, M.J., Price, A.L., Neale, B.M.; and

- Schizophrenia Working Group of the Psychiatric Genomics Consortium (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* *47*, 291–295.
44. Morales, J., Welter, D., Bowler, E.H., Cerezo, M., Harris, L.W., McMahon, A.C., Hall, P., Junkins, H.A., Milano, A., Hastings, E., et al. (2018). A standardized framework for representation of ancestry data in genomics studies, with application to the NHGRI-EBI GWAS Catalog. *Genome Biol.* *19*, 21.
 45. Brown, B.C., Ye, C.J., Price, A.L., Zaitlen, N.; and Asian Genetic Epidemiology Network Type 2 Diabetes Consortium (2016). Transethnic Genetic-Correlation Estimates from Summary Statistics. *Am. J. Hum. Genet.* *99*, 76–88.
 46. Gusev, A., Ko, A., Shi, H., Bhatia, G., Chung, W., Penninx, B.W.J.H., Jansen, R., de Geus, E.J.C., Boomsma, D.I., Wright, F.A., et al. (2016). Integrative approaches for large-scale transcriptome-wide association studies. *Nat. Genet.* *48*, 245–252.
 47. Ratnapriya, R., Sosina, O.A., Starostik, M.R., Kwicklis, M., Kaphahn, R.J., Fritsche, L.G., Walton, A., Arvanitis, M., Gieser, L., Pietraszkiewicz, A., et al. (2019). Retinal transcriptome and eQTL analyses identify genes associated with age-related macular degeneration. *Nat. Genet.* *51*, 606–610.
 48. Zhu, Z., Zhang, F., Hu, H., Bakshi, A., Robinson, M.R., Powell, J.E., Montgomery, G.W., Goddard, M.E., Wray, N.R., Visscher, P.M., and Yang, J. (2016). Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* *48*, 481–487.
 49. de Leeuw, C.A., Mooij, J.M., Heskes, T., and Posthuma, D. (2015). MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.* *11*, e1004219.
 50. Watanabe, K., Taskesen, E., van Bochoven, A., and Posthuma, D. (2017). Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* *8*, 1826.
 51. R Core Team (2017). R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing).
 52. Wolfs, R.C., Ramrattan, R.S., Hofman, A., and de Jong, P.T. (1999). Cup-to-disc ratio: ophthalmoscopy versus automated measurement in a general population: The Rotterdam Study. *Ophthalmology* *106*, 1597–1601.
 53. Harper, R., Reeves, B., and Smith, G. (2000). Observer variability in optic disc assessment: implications for glaucoma shared care. *Ophthalmic Physiol. Opt.* *20*, 265–273.
 54. Sisodia, D.S., and Nair, S. (2017). Diabetic retinal fundus images: Preprocessing and feature extraction for early detection of diabetic retinopathy. *Biomed. Pharmacol. J.* *10*. <https://doi.org/10.13005/bpj/1148>.
 55. Orlando, J.I., Prokofyeva, E., and del Fresno, M. (2017). Convolutional neural network transfer for automated glaucoma identification. In 12th International Symposium on Medical Information Processing and Analysis. <https://doi.org/10.1117/12.2255740>.
 56. Singh, A., Dutta, M.K., ParthaSarathi, M., Uher, V., and Burget, R. (2016). Image processing based automatic diagnosis of glaucoma using wavelet features of segmented optic disc from fundus image. *Comput. Methods Programs Biomed.* *124*, 108–120.
 57. Li, Z., He, Y., Keel, S., Meng, W., Chang, R.T., and He, M. (2018). Efficacy of a Deep Learning System for Detecting Glaucomatous Optic Neuropathy Based on Color Fundus Photographs. *Ophthalmology* *125*, 1199–1206.
 58. Cuellar-Partida, G., Lu, Y., Kho, P.F., Hewitt, A.W., Wichmann, H.E., Yazar, S., Stambolian, D., Bailey-Wilson, J.E., Wojciechowski, R., Wang, J.J., et al. (2016). Assessing the Genetic Predisposition of Education on Myopia: A Mendelian Randomization Study. *Genet. Epidemiol.* *40*, 66–72.
 59. Gelman, S., Cone, F.E., Pease, M.E., Nguyen, T.D., Myers, K., and Quigley, H.A. (2010). The presence and distribution of elastin in the posterior and retrolubar regions of the mouse eye. *Exp. Eye Res.* *90*, 210–215.
 60. Rebecca, M., Gayathri, R., Bhuvanasundar, R., SriPriya, K., Shantha, B., and Angayarkanni, N. (2019). Elastin modulation and modification by homocysteine: a key factor in the pathogenesis of Pseudoexfoliation syndrome? *Br. J. Ophthalmol.* *103*, 985–992.
 61. Beck, R.W., Messner, D.K., Musch, D.C., Martonyi, C.L., and Lichter, P.R. (1985). Is there a racial difference in physiologic cup size? *Ophthalmology* *92*, 873–876.
 62. Varma, R., Tielsch, J.M., Quigley, H.A., Hilton, S.C., Katz, J., Spaeth, G.L., and Sommer, A. (1994). Race-, age-, gender-, and refractive error-related differences in the normal optic disc. *Arch. Ophthalmol.* *112*, 1068–1076.
 63. Soh, Z.D., Chee, M.L., Thakur, S., Tham, Y.C., Tao, Y., Lim, Z.W., Mani, B., Wong, T.T., Aung, T., and Cheng, C.-Y. (2020). Asian-specific vertical cup-to-disc ratio cut-off for glaucoma screening: An evidence-based recommendation from a multi-ethnic Asian population. *Clin. Exp. Ophthalmol.* *48*, 1210–1218.
 64. Crowston, J.G., Hopley, C.R., Healey, P.R., Lee, A., Mitchell, P., and Blue Mountains Eye Study (2004). The effect of optic disc diameter on vertical cup to disc ratio percentiles in a population based cohort: the Blue Mountains Eye Study. *Br. J. Ophthalmol.* *88*, 766–770.
 65. Tham, Y.-C., Li, X., Wong, T.Y., Quigley, H.A., Aung, T., and Cheng, C.-Y. (2014). Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis. *Ophthalmology* *121*, 2081–2090.