

# Contribution of common and rare variants to Asian neovascular age-related macular degeneration subtypes

Received: 29 May 2022

Accepted: 28 August 2023

Published online: 11 September 2023


 Check for updates

Qiao Fan<sup>1,2,19</sup>  , Hengtong Li<sup>3,4,5,19</sup>, Xiaomeng Wang<sup>2,6,19</sup>, Yih-Chung Tham<sup>2,3,4,5</sup>, Kelvin Yi Chong Teo<sup>2,3</sup>, Masayuki Yasuda<sup>7</sup>, Weng Khong Lim<sup>8,9,10,11</sup>, Yuet Ping Kwan<sup>2</sup>, Jing Xian Teo<sup>8</sup>, Ching-Jou Chen<sup>6</sup>, Li Jia Chen<sup>12</sup>, Jeeyun Ahn<sup>13</sup>, Sonia Davila<sup>8</sup>, Masahiro Miyake<sup>14</sup>, Patrick Tan<sup>8,10,15</sup>, Kyu Hyung Park<sup>16</sup>, Chi Pui Pang<sup>12</sup>, Chiea Chuan Khor<sup>15</sup>, Tien Yin Wong<sup>17</sup>, Yasuo Yanagi<sup>18</sup>, Chui Ming Gemmy Cheung<sup>2,3</sup> & Ching-Yu Cheng<sup>2,3,4,5</sup>  

Neovascular age-related macular degeneration (nAMD), along with its clinical subtype known as polypoidal choroidal vasculopathy (PCV), are among the leading causes of vision loss in elderly Asians. In a genome-wide association study (GWAS) comprising 3,128 nAMD (1,555 PCV and 1,573 typical nAMD), and 5,493 controls of East Asian ancestry, we identify twelve loci, of which four are novel ( $P < 1.19 \times 10^{-8}$ ). Substantial genetic sharing between PCV and typical nAMD is noted ( $r_g = 0.666$ ), whereas collagen extracellular matrix and fibrosis-related pathways are more pronounced for PCV. Whole-exome sequencing in 259 PCV patients revealed functional rare variants burden in collagen type I alpha 1 chain gene (*COL1A1*;  $P = 1.05 \times 10^{-6}$ ) and potential enrichment of functional rare mutations at AMD-associated loci. At the GATA binding protein 5 (*GATA5*) locus, the most significant GWAS novel loci, the expressions of genes including laminin subunit alpha 5 (*Lama5*), mitochondrial ribosome associated GTPase 2 (*Mtg2*), and collagen type IX alpha 3 chain (*Col9A3*), are significantly induced during retinal angiogenesis and subretinal fibrosis in murine models. Furthermore, retinoic acid increased the expression of *LAMA5* and *MTG2* in vitro. Taken together, our data provide insights into the genetic basis of AMD pathogenesis in the Asian population.

Age-related macular degeneration (AMD) is one of the leading causes of irreversible vision loss among people aged 50 years and above. By 2040, more than 170 million people are estimated to be affected by this debilitating eye disease, with over half of the incidence in Asia alone, largely due to its rapidly aging population<sup>1</sup>. Neovascular AMD (nAMD) is a late-stage AMD, characterized by exudation, subretinal hemorrhage, and scarring, which often lead to central vision loss<sup>2,3</sup>. In Asian populations, polypoidal choroidal vasculopathy (PCV) is a common subtype of overall nAMD. PCV accounted for up to 55% of all

nAMD in Asian populations<sup>4-6</sup>, compared to less than 10% of counterparts in white populations<sup>7,8</sup>. Although they share many common features, PCV is a distinct phenotype of typical nAMD. Clear differences in morphological and pathogenic features exist: PCV develops in eyes with “pachychoroid” and is not necessarily associated with drusen, subretinal pigment epithelial deposits<sup>3,9</sup>. On the other hand, drusen are the most salient features of typical nAMD. Clinically, PCV patients are generally younger and may present with recurrent episodes of sub-macular hemorrhage, possibly with less scarring and

A full list of affiliations appears at the end of the paper.  e-mail: [qiao.fan@duke-nus.edu.sg](mailto:qiao.fan@duke-nus.edu.sg); [chingyu.cheng@duke-nus.edu.sg](mailto:chingyu.cheng@duke-nus.edu.sg)

variations in their response to anti-VEGF therapy compared to typical nAMD patients<sup>10,11</sup>. Hence, a deep understanding of the biology and genetic signatures of PCV and typical nAMD is important for the identification of new drug targets that are more effective in treating Asian patient populations.

Some earlier studies have identified common and rare variants that contribute to the genetic architecture of AMD. However, most large-scale genome-wide scans were conducted in European populations, and data for the Asian AMD patient population, in particular for PCV, are scarce. Genome-wide association studies (GWAS) have identified ~51 loci associated with early or advanced AMD<sup>12–15</sup>. The association of several of these loci with PCV has been confirmed using candidate genes or meta-analysis approaches, such as *CFH*, *ARMS2/HTRA1*, *C2/CFB/SKIV2L*, *CETP*, and *VEGFA*, etc.<sup>16–19</sup>. Although these loci are shared between typical nAMD and PCV, the lead SNP may exhibit varying effect sizes. For instance, *ARMS2/HTRA1* variants (e.g., rs10490924, rs2672598) have a larger effect in typical nAMD than in PCV<sup>19,20</sup>. Furthermore, several rare functional variants, mainly involving complement factors-related genes such as *CFH* (e.g., p.Arg1210Cys, p.Arg303Gln), *C9* (e.g., p.Pro167Ser), *C3* (e.g., p.Lys155Gln), *CFI* (e.g., p.Gly119Arg)<sup>21–23</sup>, *CETP* (e.g., p.Arg303Gln)<sup>13</sup>, *SPEF2* (e.g., p.Arg201Gly)<sup>24</sup> and *UBE3D* (rs7739323,1039G > A)<sup>25</sup>, have been reported to be associated with nAMD. Additionally, an earlier whole-genome exome (WES) study showed that rare variants in *FGD6* (p.Lys329Arg) are associated with PCV in the Chinese population<sup>26</sup>. Thus, a comprehensive study is needed to further identify susceptibility genes for PCV and typical nAMD in the Asian population, and to differentiate potential genetic heterogeneity between the two conditions.

In this work, we seek to characterize the genetic architecture underlying PCV and typical nAMD through GWAS and WES approaches in East Asian samples from the Genetics of AMD in Asian (GAMA) Consortium. Our study, to our knowledge, is the most comprehensive GWAS meta-analysis comprising 1555 PCV, 1573 typical nAMD and 5493 healthy controls from Singapore, Hong Kong, Korea and Japan, and 259 PCV sequenced samples. We identify novel loci shared between PCV and typical nAMD, potential causal genes within these loci, as well as rare functional variant enrichment that possibly confers genetic susceptibility to Asian PCV and typical nAMD pathogenesis.

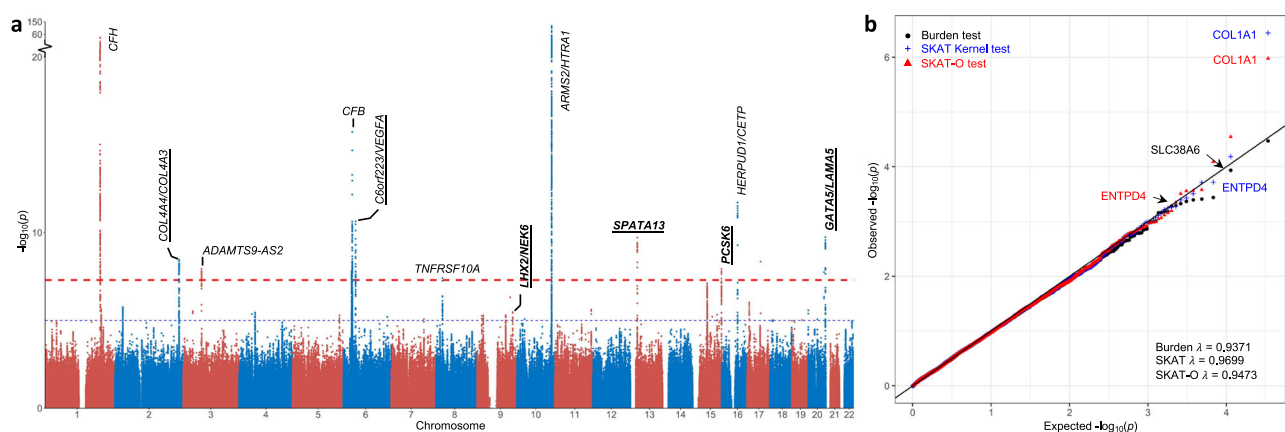
## Results

### Novel GWAS loci identified for Asian nAMD, PCV and typical AMD

The detailed information on patients of East Asian ancestry in participating studies from the GAMA Consortium is described in “Methods” and Supplementary Notes. Supplementary Fig. 1 depicts a flow chart of the participants recruited in each study and included for analysis, while Supplementary Data 1 represents information on AMD phenotyping and genotyping arrays. At the study level, stringent quality control (QC) was performed before imputation on the genotyped markers (see “Methods”). For each study, the genotype data were imputed using the 1000 Genomes Project multi-ethnic reference panels v3 on overlapped SNPs across genotyping arrays in cases and controls. Principal component analysis (PCA) showed that cases and controls were genetically well-matched (Supplementary Fig. 2).

For the discovery phase, meta-analyses were conducted in 8,560,176 variants in 1687 nAMD patients and 3437 controls of Chinese ancestry from Singapore and Hong Kong studies, using fixed-effect models implemented in the METAL software (<http://csg.sph.umich.edu/abecasis/Metal/download/>). The quantile-quantile (QQ) plots showed little evidence of inflation in test statistics due to the population stratification (Supplementary Fig. 3; Genomic Control  $\lambda_{GC}$ : 0.987). About 3180 variants with  $P$  values less than  $5 \times 10^{-5}$  at the discovery stage were carried over to the replication phase, for replicating in samples comprising 1441 nAMD patients and 2056 controls recruited from Korean and Japanese studies. Among these variants, 11 loci were identified with lead SNPs exceeding genome-wide significance (Supplementary Data 2A). We further performed a whole-genome meta-analysis of four cohorts for PCV and typical nAMD separately and identified an additional locus *NEK6/LHX2* for typical nAMD (Supplementary Data 2B). Our final sample for meta-analysis included 1555 PCV and 1573 typical nAMD patients, and 5493 healthy controls.

**GATA5/LAMA5 is the top novel locus for Asian nAMD.** Altogether, we identified 12 loci with genome-wide significance; of these four were novel, including *GATA5/LAMA5* at Chromosome 20q13.33, *SPATA13* at Chromosome 13q12.12, *PCSK6* at Chromosome 15q26.3, and *NEK6/LHX2* at Chromosome 9q33.3 (Manhattan plot, Fig. 1a; Regional plots, Supplementary Fig. 4). Except for the *NEK6/LHX2* locus, signals at the rest of the loci were shared between PCV and typical nAMD. The



**Fig. 1 | Manhattan plot of GWAS for neovascular AMD (nAMD) and QQ plot of whole-exome sequencing gene-level analysis for PCV.** **a** We conducted genome-wide single-variants association analyses for nAMD in 3128 patients (1555 PCV and 1573 typical nAMD) and 5493 controls. The Manhattan plot exhibits ten logarithms of  $P$  values for the association, annotating the nearest genes for each genetic locus that reaches genome-wide significance ( $P < 5 \times 10^{-8}$ ). The gene names of novel nAMD loci are highlighted in bold. Two loci (*COL4A4/COL4A3* and *C6orf223/VEGFA*) that showed significant association with PCV in our

study are underlined. The red dashed horizontal line represents genome-wide significance ( $P < 5 \times 10^{-8}$ ), and the blue line indicates the suggestive significance of  $P < 1 \times 10^{-20}$ . **b** We conducted three gene-based analyses: Burden test, Sequence Kernel Association Test (SKAT), and Combined test of burden test and SKAT (SKAT-O) for functional variants at MAF < 2% in 1019 Chinese subjects (259 PCV and 760 controls). Data points shown with distinct hues and shapes represent genes tested for three different tests. The genomic inflation factor ( $\lambda$ ) of each analysis is shown in the figure.

strongest signal was at *GATAS/LAMAS* for nAMD (lead SNP rs6121609;  $P=1.80 \times 10^{-10}$ ), whereas consistent effects were observed for AMD subtypes (rs6121609 C allele; PCV, OR=1.44,  $P=1.97 \times 10^{-7}$ ; typical nAMD, OR=1.45,  $P=1.34 \times 10^{-7}$ ; Table 1). Similarly, at the *PCSK6* and *SPATA13* loci, the magnitude of effects in all samples (*PCSK6* rs7402624 C allele; OR=1.22,  $P=1.19 \times 10^{-8}$ ; *SPATA13* rs4769312 A allele; OR=1.29,  $P=1.90 \times 10^{-10}$ ) were concordant across subtypes. The signal at the *NEK6/LHX2* locus, was mainly driven by typical nAMD (rs72759285 A allele; PCV, OR=1.10,  $P=6.84 \times 10^{-2}$ ; typical nAMD, OR=1.39,  $P=1.19 \times 10^{-8}$ ). Conditional analysis revealed two independent low-frequency variants at the known *CFH* locus (MAF  $\leq 6.3\%$ ; Supplementary Data 3). No additional novel loci were identified from the gene-based test using MAGMA (Supplementary Data 4).

For those previously identified GWAS loci for advanced AMD in European populations<sup>12,14</sup>, Supplementary Data 5 listed the association between index SNPs and nAMD in Asian samples. Among 37 genetic loci, lead variants at 4 loci reached genome-wide significance (10.8%;  $P < 5 \times 10^{-8}$ ); 17 had nominal significance (45.9%;  $5 \times 10^{-8} < P < 0.05$ ), and 16 did not show an association (43.2%;  $P > 0.05$ ). For those lead variants not showing genome-wide significance, the genetic effects were moderate, in general with an odds ratio (OR) less than 1.20. Across all these variants, the correlation between beta effects from our datasets and those reported from European samples was 0.831 ( $P < 1.90 \times 10^{-10}$ ).

**Independent variant in *COL4A4* and *C6orf223* genes associated with PCV.** The following five genome-wide loci in decreasing order of significance for PCV are also known loci for nAMD: *ARMS2/HTRA1*, *CFH*, *COL4A4/COL4A3*, *C6orf223/VEGFA*, *C2/CFB*, and *CETP*. For the *COL4A4/COL4A3* locus at chromosome 2, we identified a lead SNP rs56033528 in gene *COL4A4* for PCV ( $P=7.01 \times 10^{-10}$ ), with an attenuated effect for typical nAMD ( $P=2.63 \times 10^{-3}$ ). SNP rs56033528 is independent ( $r^2=0.002$ ) of the previously identified lead SNP rs11884770<sup>12</sup> at the *COL4A3* locus for advanced AMD in European populations (Table 1). Previously, we identified the *C6orf223/VEGFA* locus for nAMD from our GAMA GWAS data<sup>13</sup>. In this study, the lead SNP rs73733647 at *C6orf223/VEGFA* reached genome-wide significance for PCV ( $P=8.33 \times 10^{-10}$ ). SNP rs73733647 is independent of the lead SNP rs943080 at the *VEGFA* locus identified in the European population ( $r^2 < 0.001$ ).

**Transethnic-replication in European populations.** Next, we evaluated the transethnic replication of the above 6 loci in individuals of European ancestry from the International AMD Genomics Consortium (IAMGDC) dataset (16,144 cases and 17,832 controls). Lead SNPs at *GATAS/LAMAS*, *PCSK6*, and *COL4A4* showed potential association with AMD in this dataset ( $P < 0.05$ ; Supplementary Data 6). Overall, the directions of effect sizes were largely similar.

### WES pinpoints potential genes of rare variants' enrichment for PCV

To reveal any potential functional variants for PCV that alter peptide sequences (missense), or lead to stop-gain, stop-loss, frameshifting, or splice-site disruption, we performed exome sequencing in 259 PCV patients and 760 controls of the Chinese ancestry population in Singapore. Following the standard GATK pipeline (<https://github.com/broadinstitute/gatk/releases/tag/4.0.3.0>), we detected 637,799 single-nucleotide variants (SNVs), of which 166,045 were functional. In both patient samples and controls, there was substantial concordance between genotypes called by exome sequencing and the directly genotyped data from GWAS, for variants with MAF at the following ranges: 5–50% (concordance >99%), 1–5% (concordance >98%), <1% (concordance >95%; Supplementary Fig. 5). We conducted gene-based tests of the kernel (SKAT) and burden, and omnibus SKAT-O combining both, for PCV in 154,195 functional variants with MAF  $\leq 2\%$ . We used the SKAT v2.0.1R package (<https://cran.r-project.org/web/packages/SKAT>)

for the analyses, accounting for age, gender, and the top four principal components. No inflation was observed in test statistics as seen in QQ plots ( $\lambda_{GC}$  ranged from 0.931 to 0.970; Fig. 1b); population stratification had a negligible effect on the gene-based association tests for rare variants.

**Rare variants at nAMD-associated GWAS loci.** First, we sought to pinpoint rare variants burden at the identified loci for nAMD from our Asian population GWAS by assessing the functional variants in adjacent genes. The *ENTPD4* gene at GWAS locus 8p21.3, from which GWAS lead SNP rs13278062 (291 bp upstream from *TNFRSF10A*) is 204 kb downstream (SKAT-O:  $P=4.49 \times 10^{-4}$ ; SKAT:  $P=1.91 \times 10^{-4}$ ; Table 2 and Supplementary Data 7), was identified as the top gene for the functional variant enrichment for PCV. The  $P$  value from the kernel SKAT test, but not omnibus test SKAT-O, remained significant after multiple testing corrections for all 190 genes tested at the GWAS loci ( $P < 2.63 \times 10^{-4}$ ). *ENTPD4*, ~30 kb consisting of 14 exons, encodes the apyrase protein, which is an enzyme catalyzing the hydrolysis of nucleotide diphosphates and triphosphates. It comprises 12 missense variants, with mutation counts ranging from 1 to 18. The rare variant burden was not in the shade of LD with the GWAS signals, as suggested from the analysis conditional on the lead GWAS SNP (SKAT test: conditional  $P < 4.34 \times 10^{-4}$ ). Using protein-prediction algorithms (see “Method”), there were 3 rare variants predicted to be damaging (p.L197V, p.P202S, and p.R591H) residing in exons 6 and 13 (Supplementary Data 8). Variant p.P202S exhibited the strongest single-variant association (Supplementary Data 9). *OR4F15* gene at our newly identified GWAS locus 15q26.3 (~326 kb adjacent to GWAS lead SNP *PCSK6* rs7402624) and *LOXL2* gene at GWAS locus 8p21.3 (~71 kb from GWAS lead SNP rs13278062) showed marginal significance for kernel tests. Besides, previously identified rare variants in *FGD6* from WES<sup>26</sup>, which is also a known GWAS locus<sup>13</sup>, exhibited nominal significance (SKAT-O,  $P=9.41 \times 10^{-3}$ ; Supplementary Data 7). Single-variant analysis of functional variants for PCV at the GWAS loci identified several rare variants in genes *OR4F15* (p.S20L,  $P=1.09 \times 10^{-3}$ ), and *SELENOS* (p.R54G,  $P=2.15 \times 10^{-3}$ ) at GWAS locus 15q26.3, and two variants in *ENTPD4* (p.P202S,  $P=2.54 \times 10^{-3}$ ; p.S597L,  $P=6.32 \times 10^{-3}$ ; Supplementary Data 9). The significance for the variant p.S20L in *OR4F15* remained at  $P=1.56 \times 10^{-3}$ , after multiple testing corrections.

***COL1A1* exhibits exome-wide significance for PCV.** Across the whole exome, gene *COL1A1* at chromosome 17q21.33 had significant enrichment of rare variants for PCV (SKAT-O,  $P=1.05 \times 10^{-6}$ ; Table 2 and Supplementary Data 10), exceeding the exome-wide gene-level significance at  $2.98 \times 10^{-6}$ . No signals were seen from the common variants nearby, suggesting that the rare variants' burden in *COL1A1* was independent of GWAS hits. *COL1A1* encodes the pro- $\alpha 1$  chains of type I collagen found in most connective tissues. The majority of 27 functional variants within *COL1A1* were missense variants (24/27), largely residing from exons 15 to 51 (Supplementary Data 11). Ten rare variants were predicted to be damaging, with 8 being singleton variants, indicating the extremely rare functional variants were responsible for the association signals. *ALASI* at chromosome 3 encoding the mitochondrial enzyme (SKAT-O,  $P=2.84 \times 10^{-5}$ ), and *CCDC175* at chromosome 14 encoding coiled-coil domain-containing protein 175 (SKAT-O,  $P=8.10 \times 10^{-5}$ ) were found to be the top genes at nominal significance. The sensitivity analyses that included variants with MAF <1% showed consistent results; the association with *COL1A1* remained the strongest (SKAT-O,  $P=1.51 \times 10^{-7}$ ; Supplementary Data 10), while genes *ALASI* and *CCDC175* showed nominal significance.

**Heritability and genetic sharing between PCV and typical nAMD** Based on common variants (MAF >1%), the SNP-heritability of PCV and typical nAMD was estimated at 0.417 (SE = 0.121) and 0.386 (SE = 0.147), respectively (Table 3). For the East Asian general



**Table 1 | Association results for novel genetic loci for Asian nAMD and subtypes**

Lead variant	rs72759285*	rs4769312	rs6121609						
Chr:Position	9:126985763	13:24623220	20:61038322						
Nearest gene	NEK6/LHX2	SPATA13	GATA5/LAMA5						
EA/OA	A/G	A/G	C/T						
EAf (PCV/hypical nAMD/control)	0.80/0.84/0.80	0.76/0.77/0.72	0.16/0.16/0.13						
	OR (95% CI)	Het $f^2$	OR (95% CI)	Het $f^2$	P	Het $f^2$	P	Het $f^2$	P
Discovery (n = case/control: 1687/3437)	1.20 (1.06, 1.34)	1.26 (1.14, 1.41)	1.21 (1.10, 1.32)	1.47 (1.27, 1.70)	$3.69 \times 10^{-5}$	$3.69 \times 10^{-5}$	$1.94 \times 10^{-7}$	$1.94 \times 10^{-7}$	$1.94 \times 10^{-7}$
Replication (1441/2056)	1.26 (1.11, 1.43)	1.32 (1.17, 1.48)	1.24 (1.11, 1.37)	1.38 (1.16, 1.63)	$7.90 \times 10^{-5}$	$7.90 \times 10^{-5}$	$1.94 \times 10^{-4}$	$1.94 \times 10^{-4}$	$1.94 \times 10^{-4}$
Meta-analysis for nAMD (3128/5493)	1.23 (1.13, 1.34)	1.29 (1.19, 1.39)	1.22 (1.14, 1.31)	1.43 (1.28, 1.60)	$1.19 \times 10^{-8}$	$1.19 \times 10^{-8}$	$1.80 \times 10^{-10}$	$1.80 \times 10^{-10}$	$1.80 \times 10^{-10}$
Meta-analysis for PCV (1555/5493)	1.10 (0.99, 1.23)	1.24 (1.12, 1.37)	1.24 (1.14, 1.35)	1.44 (1.26, 1.66)	$1.44 \times 10^{-6}$	$1.44 \times 10^{-6}$	$1.97 \times 10^{-7}$	$1.97 \times 10^{-7}$	$1.97 \times 10^{-7}$
Meta-analysis for typical nAMD (1573/5493)	1.39 (1.24, 1.56)	1.34 (1.21, 1.48)	1.20 (1.10, 1.30)	1.45 (1.26, 1.66)	$5.48 \times 10^{-5}$	$5.48 \times 10^{-5}$	$1.34 \times 10^{-7}$	$1.34 \times 10^{-7}$	$1.34 \times 10^{-7}$
Independent variants for PCV at known loci									
Lead variant	Chr: Position	Nearest gene	PCV (1555/5493)	typical nAMD (1573/5493)	EA/OA	EAf			
rs56033528	2:227865660	COL4A4/COL4A3	OR (95% CI) P	OR (95% CI) P	A/G	0.20/0.17/0.15			
rs73733647	6:43961015	C6orf223/VEGFA	1.42 (1.27, 1.59) 7.01 $\times 10^{-10}$	1.19 (1.06, 1.34) 2.63 $\times 10^{-3}$	A/G	0.83/0.82/0.79			
			1.49 (1.31, 1.69) 8.33 $\times 10^{-10}$	1.32 (1.16, 1.49) 1.33 $\times 10^{-5}$	G/A				

Chr chromosome, EA effect allele, OA other allele, EAF effect allele frequency, OR (95% CI) odds ratio (95% confidence interval), Het  $f^2$  fixed-effect meta-analysis heterogeneity I-square statistics, nAMD neovascular AMD, PCV polypoidal choroidal vasculopathy. Physical positions and nearest genes are based on NCBI build 37 of the human genome. Discovery stage: Chinese samples from Singapore and Hong Kong studies; replication stage: samples from Korea and Japan studies. We conducted a genome-wide single-variant association analysis using the first bias-corrected likelihood-ratio model, incorporating the top four principal components as covariates.  $\chi^2$  statistics with one degree-of-freedom was used for generating P values. We performed a meta-analysis under an inverse-weighted fixed-effects model across all cohorts for each phenotype.

\*SNP rs72759285 at locus NEK6/LHX2 is genome-wide significant ( $1.19 \times 10^{-6}$ ) in GWAS for typical nAMD.

population, assuming late AMD prevalence at 0.5%<sup>27,28</sup>, the liability-scaled heritability was 0.287 (SE = 0.083) for PCV and 0.264 (SE = 0.101) for typical nAMD. The genetic correlation between PCV and typical nAMD was estimated at 0.666 (SE = 0.187), suggesting substantial sharing of the genetic basis of common variants for these subtypes.

To evaluate whether genetic effect sizes were consistent at varying degrees of MAF, we compared effect sizes for each subtype at different MAF ranges and P values (Fig. 2). We grouped these variants by allele frequencies (1 – 10%, 10 – 25%, and 25 – 50%) and P values ( $P \leq 0.05$ ,  $P \leq 1 \times 10^{-5}$  and  $P \leq 5 \times 10^{-8}$ ). The effect sizes and direction were largely consistent for common or high-frequency variants but exhibited variations to some extent for low-frequency variants. To identify those exhibiting the largest difference in the effects, we performed GWAS for PCV vs. typical nAMD. The QQ plot indicated a genomic inflation factor  $\lambda_{GC}$  at 0.957, with no evidence of potential population substructures (Supplementary Fig. 6). The ARMS2 rs61871744 variant remained the most significant signal ( $P_{diff} = 1.91 \times 10^{-8}$ ; Supplementary Data 12 and Supplementary Fig. 7), corroborating previous findings<sup>29</sup>. Thirteen loci showed marginal significance at P value less than  $1 \times 10^{-5}$ , with opposite genetic effects observed for each subtype. However, none of these loci showed genome-wide significance in subtype-specific analysis.

Supplementary Data 13 summarizes the implicated genes for AMD, PCV, and typical nAMD, from previous large-scale genome studies<sup>12–20,22,24–26</sup>, along with ours. A large body of AMD genes is discovered in European samples; the most significant genes showing the strongest effect size for European advanced AMD are shared in Asian nAMD. Novel genes identified for PCV are mainly from Asian GWAS or WES.

### Fibrosis and collagen-related pathways for PCV

We further conducted pathway analysis on 15,481 gene sets. After correction for multiple testing, fibroblast senescence, which elevates fibroblast numbers leading to tissue fibrosis and stiffening in the elderly<sup>30</sup>, and collagen extracellular matrix (ECM) assembly, were identified as top pathways for PCV (LY\_AGING\_OLD\_UP, M8910,  $P = 2.28 \times 10^{-7}$ ; go\_collagen\_type\_iv\_trimer, M25627,  $P = 1.61 \times 10^{-5}$ , respectively; Supplementary Data 14). HTRA1, along with several genes (e.g., CST6, FMOD, COMP) are involved in fibroblast senescence. The collagen ECM assembly pathway requires type IV collagen triple helices to form three-dimensional networks within the basement membrane and involves COL4A4 and COL4A3 genes. The corresponding genes involved in the key pathways associated with PCV are listed in Supplementary Data 15. These two pathways, in contrast, had much weaker signals for typical nAMD ( $P = 0.017$  and 0.032 respectively).

### eQTL analyses identify multiple potential target genes at the novel loci

To identify potential genes at the four novel loci—GATA5/LAMA5, PCSK6, SPATA13 and NEK6/LHX2, we performed GWAS SNP expression quantitative trait locus (eQTL) association using GTEx data across 53 human tissues. GATA5 was the most significant novel locus identified in our study. Within the GATA5 locus, comprised of about 30 genes in a 500 kb region, SNP rs6121609 was associated with the expression of eQTL genes including GATA5, LAMA5 and MTG2 ( $P < 9.43 \times 10^{-4}$ ; Supplementary Data 16). CABLES2 showed nominal association at  $P = 9.44 \times 10^{-4}$ . Several other genes (e.g., RBBP8NL, TCFL5 and COL9A3) showed marginal significance ( $P < 5 \times 10^{-3}$ ); among these, MTG2, TCFL5 and COL9A3 were about 100 kb away from the top peak. For the other 3 loci, the genes where the top hits reside, or the nearest genes, were found to be the top eQTL gene (e.g., NEK6/LHX2, eQTL gene: NEK6; SPATA13, eQTL gene: SPATA13;  $P < 9.43 \times 10^{-4}$ ).

**Table 2 | Association of top genes with the rare variant burden for PCV identified from whole-exome gene-based analyses**

Gene	Chr	Functional variants (No.)	AA change (minor allele counts)	Gene-based analysis					
				Kernel test	Burden test	SKAT-O test	SKAT-O test		
Top gene at exome-wide genome				Gene-based analysis on pathogenic variants					
COL1A1	17	27	p.A1421T (35), p.S1417Rfs*29 (2), p.E1401K (1), p.A1387S (15), p.V1289L (5), p.A1256T (3), p.R1252S (1), p.E1243K (1), p.S1239N (1), p.A1149S (7), p.V1057I (1), p.R1036C (1), p.M1000I (3), p.R918H (1), c.2452-1G>T (1; splicing), c.2452-2A>C (1; splicing), p.P817A (1), p.A798V (1), p.P739L (1), p.P651S (3), p.Q576H (1), p.S441N (22), p.Q358E (1), p.D325G (1), p.M217L (1), p.E77K (1)	3.57 × 10 <sup>-7</sup>	9.88 × 10 <sup>-4</sup>	1.05 × 10 <sup>-6</sup>	3.14 × 10 <sup>-7</sup>	1.02 × 10 <sup>-4</sup>	8.37 × 10 <sup>-7</sup>
Top gene at AMD-associated GWAS loci				Conditional on lead GWAS SNP rs13278062 at GWAS locus 8p21.3					
ENTPD4	8	12	p.S597L (13), p.R591H (5), p.V524I (1), p.D521N (4), p.M444V (14), p.Q424R (7), p.T406R (1), p.R396G (10), p.R387Q (7), p.P360L (2), p.P202S (18), p.L197V (1)	1.91 × 10 <sup>-4</sup>	6.81 × 10 <sup>-2</sup>	4.49 × 10 <sup>-4</sup>	4.34 × 10 <sup>-4</sup>	1.18 × 10 <sup>-1</sup>	9.78 × 10 <sup>-4</sup>

259 PCV and 760 controls of Chinese descent were included in the whole-exome sequencing (WES) analyses. Functional variants include missense, stop-gain, stop-loss, frameshifting, or splice-site disruption variants with MAF <2%. Rare variants predicted to be damaging were listed in Supplementary Data 8 and 11.

AA change amino-acid change; for two splicing variants, cDNA change indicated. SKAT-O sequence kernel association optimal test to combine both kernel and burden tests, \*fs frameshifting, PCV polypoidal choroidal vasculopathy.

**Table 3 | SNP-heritability estimation for Asian typical nAMD and PCV, and genetic correlation**

Phenotype	SNP- <i>h</i> <sup>2</sup>		Genetic correlation				
	Observed scale		Liability scale		<i>r</i> <sub>g</sub>	SE	<i>P</i>
	<i>h</i> <sup>2</sup>	SE	<i>h</i> <sup>2</sup>	SE			
PCV	0.417	0.121	0.287	0.083	0.666	0.187	4.00 × 10 <sup>-4</sup>
Typical nAMD	0.386	0.147	0.264	0.101			

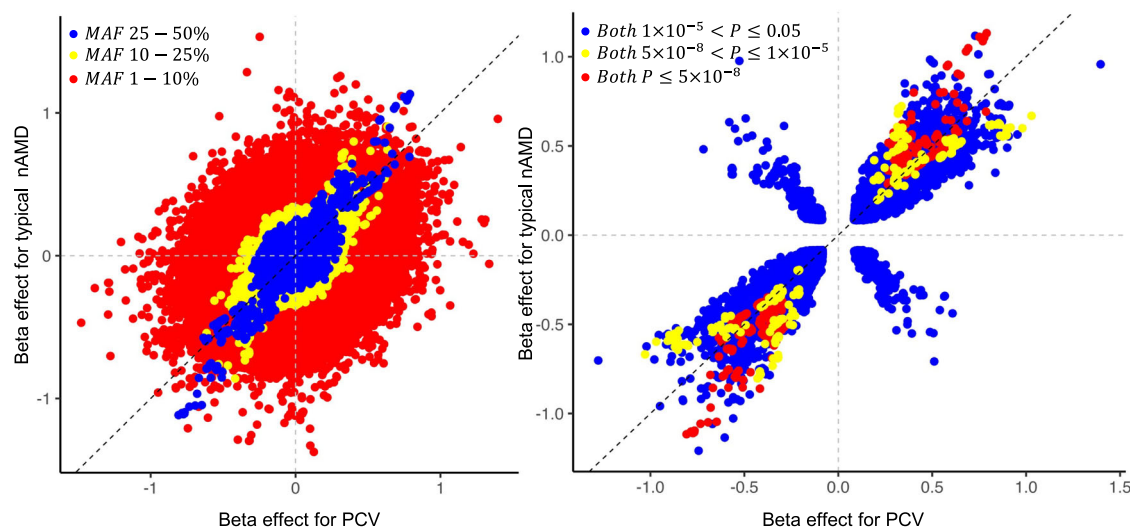
We conducted heritability and genetic correlation analyses using the LD score regression method, applying GWAS summary statistics from 1555 cases with PCV, 1573 cases with typical nAMD and 5493 controls. We retained variants that, after merging with a Hapmap3 SNP list, had a minor allele frequency ≥ 1% and an imputation quality info >0.8. We calculated both observed and liability scale SNP-*h*<sup>2</sup>, taking into account the 0.5% AMD prevalence in Asian populations<sup>27,28</sup>. nAMD neovascular AMD, PCV polypoidal choroidal vasculopathy, *h*<sup>2</sup> heritability estimate, SE standard error, *r*<sub>g</sub> genetic correlation estimate.

**Functional characterization of target genes at *GATA5* locus and of *COL1A1* gene**

We employed preclinical models of ocular angiogenesis and subretinal fibrosis to examine the gene expression of potential eQTL genes identified at the *GATA5* locus and *COL1A1* gene identified by WES. Genes that showed low expression at the RNA level, or no expression at the protein level in eyes in the Human Protein Atlas database (HPA; <https://www.proteinatlas.org/>), were omitted in further analyses. Four genes—*LAMA5*, *MTG2*, *CABLES2*, and *COL9A3*—along with *COL1A1*, were included for functional studies. *GATA5*, the nearest gene adjacent to the lead variants, had low expression in the neuroretina and RPE/choroid complex in normal mice (threshold cycle value >30). We thus examined the expression of *GATA5* only in the mouse model of laser-induced choroidal neovascularization and human RPE cells for retinoic acid treatment.

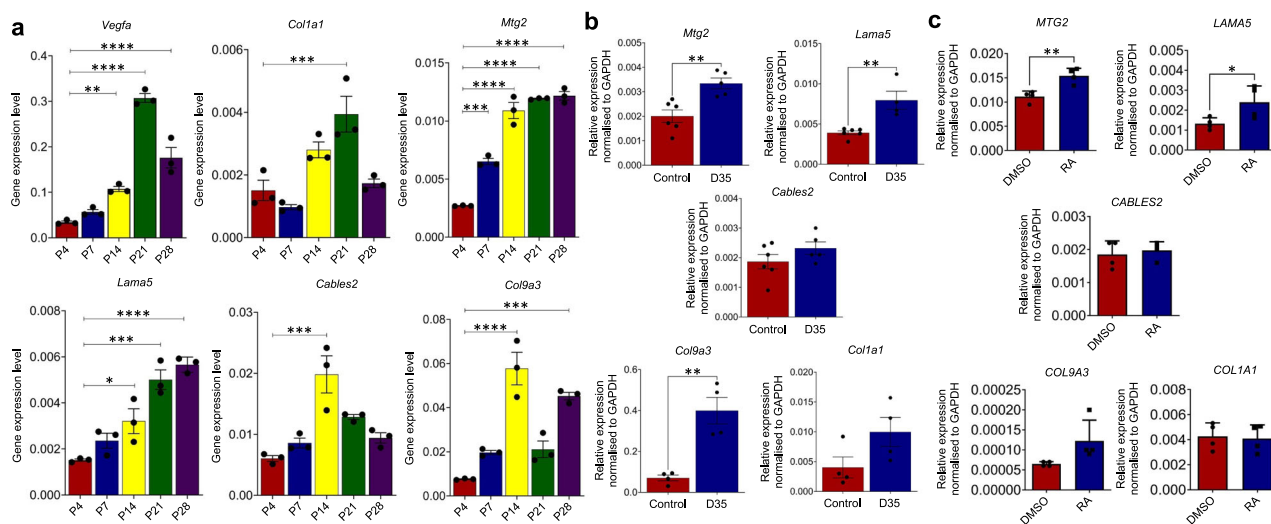
**mRNA expression patterns in vivo ocular angiogenesis models.** The murine retina provides an ideal model for studying sprouting angiogenesis<sup>30</sup>. To establish the association between lead gene candidates and developmental angiogenesis in the eye, the expression profiles of selected gene targets were analyzed in the neuroretina of mice at postnatal days P4, P7, P14, P21, and P28 (Fig. 3a). The superficial vascular network of the retina formed at P4 in our study suggests that all candidate genes were expressed at low levels in the P4 retinae. At P7, the superficial vascular plexus began to mature and branch vertically to reach the deep vascular plexus during the second week after birth<sup>31</sup>. At this stage, the vascular network fully matured throughout the retina. At P14, the intermediate vascular plexus began to form in the retinae, and at P21, the plexus developed deeper and the hyaloid vessels regressed. The maturation of the retinal vessels, especially the intermediate plexus, continued until P28. The expression of *Col1a1* in the retinae increased gradually after birth and reached its highest levels at P21 (2.6-fold increase compared to that at P4, *P* < 0.001) which mirrors the expression pattern of *Vegfa*, a master regulator of angiogenesis vascular endothelial growth factor (9-fold increase, *P* < 1 × 10<sup>-4</sup>). On the contrary, *Mtg2* and *Lama5* expressions continued to increase till P28, suggesting that these two genes may also be involved in retinal vessel remodeling (4.5-fold increase, *P* < 1 × 10<sup>-4</sup>; 3.7-fold increase, *P* = 1 × 10<sup>-4</sup>, respectively). The expression levels of *Cables2* and *Col9a3*, on the other hand, peaked at P14 when retinal angiogenesis reached its highest level (3.3-fold higher than those at P4, *P* = 5 × 10<sup>-4</sup>; 7.6-fold higher, *P* < 1 × 10<sup>-4</sup>, respectively), significantly decreasing thereafter at P21.

To understand the roles of the above genes in the development of choroidal vasculature, we further examined the expression levels in the mice RPE/Choroid complex at postnatal days P4, P14, and P28 respectively (Supplementary Fig. 8). Similar to *Vegfa*, *Lama5*



**Fig. 2 | Genetic effect size in PCV versus typical nAMD.** For the figure on the left, we grouped variants by minor allele frequency (1–10%, 10–25%, or 25–50%). Minor allele frequency was computed based on our GWAS data in 8621 subjects of East Asian descent. For the figure on the right, variants were grouped by different association  $P$  value thresholds ( $P \leq 5 \times 10^{-8}$ ,  $5 \times 10^{-8} < P \leq 1 \times 10^{-5}$ , or  $1 \times 10^{-5} < P \leq 0.05$ ) in GWAS for both PCV and typical nAMD. We conducted a single-variant test using the firth bias-corrected likelihood-ratio model, incorporating the top four principal components as covariates.  $\chi^2$  statistics with one degree-

of-freedom were used for the single-variant test. We then performed a meta-analysis under an inverse-weighted fixed-effect model, combining data from all cohorts for each phenotype. Each data point represents a variant out of the total 7,911,145 overlapped variants in PCV and typical nAMD GWAS. The black diagonal dashed line represents the same effect size between PCV and typical nAMD. The gray horizontal and vertical dashed lines at the value of 0 separate the figure into four quadrants. Variants in the first and the third quadrants have a consistent direction of effect between the two AMD subtypes.



**Fig. 3 | Comparative mRNA expression analysis of selected genes in the retinal tissue of developing mice, RPE cells following the laser-induced CNV in mice, and human retinal pigment epithelial (HRPE) cells upon retinoic acid treatment.** **a** mRNA expression of *COL1A1* and gene candidates at *GATA5* locus in the retinal tissue of developing mice. Each group included three mice as test subjects. The expression profiles of selected gene targets from the mouse retina were analyzed on postnatal days P4, P7, P14, P21, and P28, as shown on the y-axis of bar plots ( $n = 3$  animals). Significant differences in gene expression level were observed between P4 and P14, P21, and P28 for *Vegfa* ( $P$  values: 0.0041,  $< 0.0001$ , and  $< 0.0001$ , respectively). For *Col1a1*, the  $P$  value is  $< 0.001$  for P4 vs. P21. For *Mtg2*, significant differences were found for P4 vs. P7, P14, P21, and P28 ( $P$  value are 0.0001,  $< 0.0001$ ,  $< 0.0001$ , and  $< 0.0001$ , separately). For *Lama5*, significant differences were observed for P4 vs. P14 ( $P$  value is 0.0215), P4 vs. P21 ( $P$  value is 0.0001), and P4 vs. P28 ( $P$  value  $< 0.0001$ ). For *Cables2*, the  $P$  value is 0.0005 for P4 vs. P14. For *Col9a3*, significant differences were observed for P4 vs. P14 and P4 vs. P28 with  $P$  values  $< 0.0001$  and  $= 0.0001$ , respectively. **b** Target genes induced in

RPE cells following the laser-induced CNV in mice. qRT-PCR analysis of each gene in RPE cells isolated from mice at day 35 following laser-induced CNV. For *Col1a1* and *Col9a3*, experimental groups had four mice each. *Lama5*'s control group had six mice, and D35 group had four mice. For *Cables2* and *Mtg2*, the control group had six mice, and the D35 group had five mice. The  $P$  values for *Mtg2*, *Lama5*, *Col9a3*, *Col1a1*, and *Cables2* are 0.004, 0.003, 0.003, 0.095, and 0.203, respectively. **c** mRNA expression of gene candidates in human retinal pigment epithelial (HRPE) cells upon retinoic acid treatment, compared to induction using dimethyl sulfoxide (DMSO) solvent in four independent experiments.  $P$  values for gene candidates are: *MTG2* (0.016), *LAMA5* (0.048), *COL9A3* (0.070), *COL1A1* (0.854), and *CABLES2* (0.627). For all independent experiments, the mRNA expressions of lead gene candidates were measured using qRT-PCR method. Source data are provided as a Source Data file. Error bars were presented as mean  $\pm$  S.E.M. Statistical significance was determined by one-way ANOVA or two-tailed, unpaired Student's  $t$  test, \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ , and \*\*\*\* $P < 0.0001$ .



expression was higher at P14 compared to that at P4 ( $P = 0.021$ ). Other genes showed lower (*Cables2* and *Col9a3*) or similar expression levels (*Mtg2*) at P14 and P28, whereas *Col1a1* level was down-regulated at P28.

**Induction of *Lama5*, *Mtg2*, and *Col9a3* expression in vivo subretinal fibrosis models.** The mouse model of laser-induced choroidal neovascularization (CNV) serves as a good model to study subretinal fibrosis<sup>32</sup>, a characteristic feature of patients with advanced nAMD. The extent of subretinal fibrosis becomes apparent at day 35 post-laser treatment<sup>33</sup>. Retinal Pigment Epithelium (RPE) cells are key contributors to the abnormal accumulation of extracellular matrix (ECM) components in the subretinal space of nAMD patients<sup>34</sup>. To evaluate the association between novel target genes and subretinal fibrosis, the expression of gene candidates in the RPE/choroid tissues isolated from C57BL/6 mice at day 35 following the induction of CNV by laser photocoagulation was analyzed by qRT-PCR. Our study showed the following expression pattern in CNV eyes with subretinal fibrosis compared to those in non-lasered controls: 2-fold increase in *Lama5* ( $P = 0.003$ ), 1.7-fold increase in *Mtg2* ( $P = 0.004$ ) and 5.7-fold increase in *Col9a3* ( $P = 0.003$ ), and no significant change in *Cables2* (1.2-fold,  $P = 0.203$ ) in (Fig. 3b); *Col1a1* expression level was induced 2.5-fold, but the increase was not statistically significant ( $P = 0.095$ ). For *Gata5*, the expression level was low and there was no significant difference in *Gata5* expression in the RPE/choroid tissues at day 35 post-laser treatment (Supplementary Fig. 9A).

**Retinoic acid induces *LAMA5* and *MTG2* expression in Human RPE cells.** Our in silico analysis showed that the top 30 SNPs at the *GATA5* locus reside in the putative transcription factor binding sites (TFBS; Supplementary Fig. 10). Intriguingly, the major transcription factors (TFs) are responsive to retinoic acid and glucocorticoid receptor signaling. To establish the association between RA and the above gene candidates expressed in the eye, human retinal pigment epithelial (HRPE) cells were treated with retinoic acid, and the mRNA expression of each gene was measured using qRT-PCR. Our study showed that 24-h treatment with 10  $\mu\text{m}$  retinoic acid had no obvious impact on RPE cell morphology and viability (Supplementary Fig. 11). Both *LAMA5* and *MTG2* were significantly induced by retinoic acid in RPE cells (1.8 fold,  $P = 0.048$ ; 1.4 fold,  $P = 0.016$  respectively; Fig. 3c). On the other hand, dexamethasone could not induce the expression of any candidate genes in HRPE cells (Supplementary Fig. 12). For *GATA5*, the expression level was low in HRPE cells, and RA treatment had no impact on the *GATA5* expression (Supplementary Fig. 9B). Our data suggest that retinoic acid-responsive genes (e.g., *LAMA5* and *MTG2*) may be involved in the pathophysiology of nAMD. Further investigations are needed to confirm this hypothesis.

## Discussion

The current study is the most comprehensive meta-analysis of GWAS for Asian nAMD and PCV in East Asian populations, at both common and rare variants. We identified four novel loci (the strongest association at *GATA5* locus) for Asian nAMD, and rare variant mutations enriched gene (*COL1A1*) for PCV. Using human multi-tissue eQTL analysis and the murine model, we prioritized potential target genes, including *LAMA5*, *MTG2* and *COL9A3* at the *GATA5* locus. These genes were induced in mouse eyes with ocular angiogenesis and subretinal fibrosis, characteristic pathological features of nAMD. We further showed retinoic acid, critical for visual pigment recycling, increased the expression of *LAMA5* and *MTG2* in HRPE cells. While PCV and typical nAMD exhibited substantially high genetic correlation and shared top variants, collagen ECM assembly and fibrosis-related pathways were more pronounced for PCV.

Our study highlights the role of collagen-related genes associated with PCV. We found genome-wide significance at the *COL4A4*, and using the exome-sequencing approach, we identified functional

mutations in the *COL1A1* gene that are associated with PCV. Protein-altering variants result in altered residual changes in the collagen alpha 1 chain triple helical domain and C-terminal propeptide (COLF1)<sup>35</sup>; the latter controls procollagen intracellular assembly and the extracellular assembly of collagen fibrils. The rare mutations identified in *COL1A1* have not been previously reported as ClinVar, except p. A1256T for several Mendelian diseases, such as Ehlers-Danlos syndrome<sup>36</sup>, a connective tissue disorder that is characterized by joint hypermobility, tissue fragility, and skin abnormalities. *COL1A1* has also been reported as a keratoconus-associated gene<sup>37</sup>. In the eye, *COL1A1* represents one of the prominent ECM components driving subretinal fibrosis<sup>38–40</sup>; mutations in type I collagen have been reported to cause dysregulated TGF $\beta$  signaling<sup>41</sup>, a pathway that chiefly regulates tissue fibrosis by controlling TGF $\beta$  bioavailability. Intriguingly, pathway analysis based on common variants revealed that fibrosis and collagen ECM pathways are more pronounced for PCV, with the involvement of type 4 collagen genes (e.g., *COL4A4*). Furthermore, the non-collagenous (NC) domain of various collagens has been implicated in angiogenesis<sup>42,43</sup>. For example, the NC1 domain of  $\alpha 1$ ,  $\alpha 2$ ,  $\alpha 3$ , and  $\alpha 6$  chains of collagen IV could be released via proteolytic cleavage, which inhibited angiogenesis<sup>44–46</sup>. However, an antiangiogenic role for *Col1a1* or the protein it encodes has not been established to date. It will be interesting to study how PCV-associated functional mutations in *COL1A1* and *COL4A4* affect ECM rigidity, as well as fibrosis and angiogenesis pathways in the future.

Among the four novel loci identified for nAMD, *GATA5* has emerged as an interesting AMD-associated locus. GWAS carried out on the Japanese population ( $r^2 = 0.931$ )<sup>47</sup> showed that rs6121609, the top SNP identified, is in LD with rs6061548, which is known to be associated with central serous chorioretinopathy (CSC). Both minor alleles were associated with a higher risk of AMD and CSC, in contrast to the opposite effect observed at *CFH* rs800292 for AMD and CSC<sup>48</sup>. Overall, our data provide further genetic evidence for the link between AMD and CSC.

It is challenging to assign the non-coding GWAS associations at GWAS loci to causal variants or genes. We fine-mapped the functional rare variants using exome sequencing data on genes identified at GWAS loci. We discovered the enrichment of rare mutations in the *ENTPD4* gene for PCV at the GWAS locus 8p21.3. *ENTPD4* encodes NTPDase4, a member of the apyrase protein family which belongs to the same family as CD39<sup>49</sup>, the latter tightly regulating ATP signaling as NTPDase<sup>50</sup>. High levels of extracellular ATP released by stressed cells can cause cytotoxic calcium influx<sup>51–53</sup> and increase the vulnerability of RPE cells<sup>54,55</sup>. *ENTPD4* is highly expressed in endothelial, neuronal and glial cells of the retina<sup>56,57</sup>. We speculate that NTPDase4-regulated ATP signaling has a potential role in contributing to unique vascular and RPE cell pathologies in PCV patients. The majority of genes harbored in GWAS loci, including the *GATA5* locus, did not exhibit significant functional rare variant enrichment; this is in part due to moderate effects of the rare variants, or distal target genes responsible for the signal.

We further characterized the lead SNPs located at -230 bp from the 3' end of *GATA5* gene. Our in silico analysis showed that the majority of variants reside in the putative binding sites of transcription factors that are responsive to retinoic acid and glucocorticoid receptor (GR) signalings. We have shown that the expression of *LAMA5* and *MTG2* were significantly induced by retinoic acid in human RPE cells; this is concordant with our in silico analysis, suggesting that retinoic acid signaling may serve as an important mechanism of SNP association at the *GATA5* locus in nAMD patients. Both in vitro and in vivo studies have shown all-trans-retinol acid to be a pro-angiogenic factor for nAMD<sup>58</sup>, which directly transactivates angiogenic factors such as VEGF in RPE cells via retinoic acid receptors/retinoid X receptors binding sites in their promoter regions<sup>59</sup>. *MTG2* and collagens were also previously shown to be regulated by retinoic acid in non-ocular

tissues<sup>60</sup>. Although endothelial *GATA5* was previously reported to regulate angiogenesis<sup>61</sup>, it is expressed at low levels in human and mouse ocular tissues as well as human RPE cells. In addition, *GATA5* expression was not induced in the RPE/choroid complex of mouse eyes subjected to laser-induced choroidal neovascularization and fibrosis. Our data do not support the involvement of *GATA5* in AMD pathologies. Further studies are needed to establish the direct association between lead SNPs at the *GATA5* locus and the expression of target genes in ocular cells, identify the role of retinoic acid-responsive transcription factors in regulating the expression of target genes, and explore the functional consequences of altered gene expression on AMD pathologies.

Furthermore, the subretinal fibrosis models in our study showed higher expression levels of the target genes at the *GATA5* locus including *LAMA5*, *MTG2*, and *COL9A3*. Clinically, fibrosis is known to be an important predictor of vision outcomes, following anti-VEGF treatment<sup>62–64</sup>. Interestingly, our study has suggested that the prevalence of fibrosis increased 3 folds from 13.0 to 37.8% between baseline and 12 months of AMD diagnosis in Singapore patients<sup>64</sup>. The prevalence is much higher than in the European sample. Further characterization of AMD genes responsible for fibrosis might shed light on alternative treatment strategies beyond anti-VEGF.

Besides the *GATA5* locus, three novel loci were identified including *NEK6/LHX2*, *SPATA13* and *PCSK6*. *NEK6/LHX2* locus was associated with typical nAMD, but exhibited a much weaker signal for PCV. *NEK6* expression was induced by hypoxia in a Hif1 $\alpha$ -dependent manner<sup>65</sup> and involved in key angiogenic pathways<sup>66</sup>. *LHX2* has been reported to activate mTOR and PDGF signaling, pathways associated with nAMD<sup>67,68</sup>. Both genes are highly expressed in muller glial cells of the eye<sup>56</sup>. Muller cells lose the linear shape and become migratory in typical nAMD and develop into a membrane in the region of the retina with the presence of CNV<sup>69,70</sup>. However, little attention has been given to muller cell changes in eyes with PCV. The *NEK6/LHX2*-mediated changes in muller glial cells might underlie the different pathogenesis of PCV and typical nAMD. For *SPATA13* gene, it has been reported to be associated with primary angle-closure glaucoma pathogenesis, by regulating nucleotide exchange factors activity and affecting homeostasis in the mitosis in the retina, RPE, cornea and lens etc.<sup>71</sup>. For *PCSK6*, it encodes the proprotein convertase subtilisin/kexin family; the key function of *PCSK6* has been reported to be involved in lipid metabolism, atherosclerotic plaques, ECM remodeling and cytokines<sup>72</sup>. How these genes-mediated changes in the eye underlie the different pathogenesis of nAMD warrants further investigation.

Although the underlying genetic effects of common variants were largely shared between PCV and typical nAMD, certain variants exhibited some degree of variation in the genetic effects. Through exome sequencing approaches, we identified rare variant burden for PCV (e.g., *COL1A1*), and replicated the previously reported gene *FGD6*<sup>26</sup>. For common variants, although largely concordant, some subtle differences were observed between subtypes. For example, PCV patients have different genetic effect sizes at the same index SNPs (i.e., *ARMS2/HTRA1*, *NEK6/LHX2*), or manifested the association at independent variants (i.e., *CO14A4/COL4A3*). Taken together, our study highlights the potential of employing common and rare variants identified through whole-genome scans to determine variants or genes for PCV in East Asians.

In summary, we have identified novel loci for nAMD and PCV in East Asians, through both common and rare variant association analyses. Additional functional studies have identified potential candidate target genes at the *GATA5* locus for nAMD. Our findings shed light on the underlying molecular mechanisms of nAMD in Asian populations, providing insights for the development of novel Asian-specific treatment strategies.

## Methods

All studies were performed with the approval of their Human Research and Ethics Committee, adhering to the Declaration of Helsinki principles. Written informed consent was obtained by the ethics committee of all the participating institutions, which include Singapore National Eye Center, National University Health System, Tan Tock Seng Hospital, Hong Kong Eye Hospital, Prince of Wales Hospital Eye Center, Sun Yat-sen University Cancer Center, Department of Ophthalmology at Kyoto University Hospital, Fukushima Medical University Hospital, Kobe City Medical Center General Hospital, Ozaki Eye Hospital, Mizoguchi Eye Clinic, Japanese Red Cross Otsu Hospital, Nagahama City Hospital, Seoul National University Bundang Hospital, Seoul National University Hospital, Kyungpook National University Hospital, Yeungnam University Hospital, Kosin University Hospital, and Busan Paik Hospital.

## Study samples and phenotyping

We performed GWAS and WES analyses for investigating nAMD associations in East Asians from the GAMA Consortium. The GAMA consortium includes data from four case-control studies from Singapore, Hong Kong, Korea and Japan, each of which contains: (1) cases with polypoidal choroidal vasculopathy (PCV), (2) cases with typical nAMD, and (3) hospital-based controls without any clinical signs of AMD, or general controls enrolled from population-based studies. Cases with other macular diseases such as central serous chorioretinopathy, high myopia, and angioid streaks were excluded. Details of the sample recruitment procedure are provided in Supplementary Notes. A total of 3128 cases (1555 PCV patients and 1573 typical nAMD patients) and 5493 controls were included in the analyses, and for the WES analysis, 1019 Chinese subjects (259 PCV and 760 controls) were recruited from Singapore (Supplementary Fig. 1). For all studies, the diagnosis of advanced AMD phenotyping followed the Age-Related Eye Disease Study (AREDS) or Wisconsin age-related maculopathy grading system based on dilated fundus photography (Supplementary Notes). For the presence of choroidal neovascularization, grading of fluorescein angiography was performed using a modification from the Macular Photocoagulation Study<sup>73</sup>. Indocyanine green angiography was performed to diagnose PCV using the Japanese Study Group guidelines<sup>74</sup>.

## Genotyping, quality control, and imputation

The DNA samples were genotyped on Illumina OmniExpress-12 and Human610-Quad Bead Chips. Details of genotyping platform, number of variants and phenotyping are provided in Supplementary Data 1.

The centralized quality control (QC) procedure for GWAS genotyping data was applied to each study. For variant-based quality control, genetic variants with minor allele frequency (MAF) <0.5%, genotype call rates <95%, or deviation from Hardy-Weinberg equilibrium (HWE) ( $P < 1 \times 10^{-6}$ ) were excluded. Here, we only included autosomal variants in the analysis. For sample-based quality control, participants with relatedness corresponding to halfway of second- and third-degree relatives or closer (Identity-by-descent [IBD] >0.1875), as well as samples with missing genotype calls >5% or an excess of heterozygosity (>6 SD) were excluded. For related samples in family clustering, one individual with the lowest missing rate of genotypes was retained. The analyses above were performed using PLINK 1.9.0 beta (<https://www.cog-genomics.org/plink/>) and R 3.5.3 (<https://cloud.r-project.org/>). Principal component analysis (PCA) was conducted to detect outlier samples using EIGENSOFT v7.2.1 (<https://github.com/DReichLab/EIG/archive/v7.2.1.tar.gz>). Reference data were based on HapMap 3 of Japanese, Han Chinese, Europeans, Yoruba, and Mexican populations and SEED data of Indian and Malay populations<sup>75</sup>. Prior to performing imputation, we excluded variants that were not present in 1000 Genomes Phase 3 of the East Asian population or had mismatched chromosomes, base pairs, or alleles using the HRC script. We removed palindromic variants (A/T or G/C) with MAF >0.4 and variants



with an allele frequency difference  $>0.2$  in our data and the 1000 Genomes Phase 3 reference panel of the East Asian population. All variants were aligned to the forward strand, and reference alleles were fixed according to the reference panel.

Genotyping was conducted using Illumina OmniExpress-12 v1 on 714,255 variants in the Japan study. After applying QC procedures, we retained 2113 Japanese participants (956 cases and 1157 controls) and 579,490 variants. In the Korean cohort, we utilized the same Illumina OmniExpress-12 v1 for genotyping and, after QC, retained 574,860 variants in 1384 participants (485 cases and 899 controls). For the Hong Kong study, we performed genotyping for Chinese participants using Illumina OmniExpress-12 v1 in cases on 714,255 variants and Human610-Quad BeadChip in controls on 573,871 variants. After applying QC, we included a total of 295,943 variants common for all samples (466 cases and 1006 controls) which was used for imputation.

For the Singapore study, the genotyping was performed in Chinese participants using Human610-Quad BeadChip v1 or Illumina OmniExpress-12 v1 on up to 714,255 variants. We conducted QC at the batch level because genotyping was done in subsequent batches for cases. QC was done in the following steps: (1) at each batch, variant-based quality control was employed using the same criteria as above. We ensured a forward strand for each batch based on 1000 Genomes Phase 3 reference panel of the East Asian population. (2) We conducted a Fisher exact test on the genotype counts for batch effects for cases. We tested the null hypothesis that a given batch has the same genotype frequencies as other batches combined. Variants with a  $P$  value less than  $1 \times 10^{-6}$  were excluded in all batches. (3) We combined all batches and re-performed variant-based quality control with the same criteria as in the first step. (4) Sample-based quality control was then performed as above. Finally, 283,411 post-QC variants common for 3652 Chinese participants (1221 cases and 2431 controls) were used for imputation.

We imputed whole-genome variants for each study separately based on the post-QCed genotypes, with the 1000 Genomes Phase 3 mixed populations as reference panel, using Minimac4 on the Michigan Imputation Server (<https://imputationserver.sph.umich.edu/index.html#!pages/home>). For the Hong Kong and Singaporean cohorts, we used the intersection of genotyped SNPs from different genotyping arrays in cases and controls for imputation<sup>76</sup>.

### Genome-wide single-variant association analysis and meta-analysis

We conducted a single-variant test using the firch bias-corrected likelihood-ratio model implemented in EPACTS v3.3.0 (<https://genome.sph.umich.edu/wiki/EPACTS>). For each cohort, we performed case-control association tests under an additive model for each genotyped or imputed variant with all nAMD, PCV, and typical nAMD, separately. The number of risk alleles is an integer (0, 1, and 2) for directly genotyped SNPs, or allele dosage probability ranging from 0 to 2 for imputed SNPs. The top four principal components were included as covariates to account for the population substructure. We performed a meta-analysis under an inverse-weighted fixed-effects model across all cohorts for each phenotype, and filtered out variants with imputation quality  $<0.5$ . Genetic variants of imputation quality  $\geq 0.5$  in at least 3 studies were included in the meta-analysis. Inter-study heterogeneity was reported based on heterogeneity statistic  $I^2$ .  $P$  value of  $5 \times 10^{-8}$  was considered as genome-wide significance. We also perform a random-effect meta-analysis using GWAMA v2.2.2<sup>77</sup> ([https://www.genivaramu.ee/tools/GWAMA\\_v2.2.2.zip](https://www.genivaramu.ee/tools/GWAMA_v2.2.2.zip)), which estimated heterogeneity variance with the DerSimonian-Laird method.

To identify independently-associated variants, we performed the clumping analysis using FUMA v1.3.6a<sup>31</sup> (<https://fuma.ctglab.nl/>). The significant variants were identified ( $P < 5 \times 10^{-8}$ ) in each clump, which was formed of all other variants that were within 500 kb from the index variant ( $r^2 > 0.050$ ). The variant with the smallest associated  $P$  value in

each clump was considered the independent lead variant. A locus was identified by the independent lead variant with the regions flanking 500 kb on both sides. Single-variant annotation was performed using ANNOVAR (<https://annovar.openbioinformatics.org/en/latest/>).

Conditional analysis was performed using GCTA 1.92.4 beta version (<https://cns.genomics.com/software/gcta/#Download>)<sup>78</sup>. Adjusting for the lead variant, the variants in a 1 Mb region at each locus were extracted for the association analyses. Additional independent variants were defined as  $P \leq 5 \times 10^{-8}$  in original GWAS and conditional analysis. The LD reference data was the original GWAS data.

Furthermore, we also conducted GWAS by using 1555 PCV patients as cases and 1573 typical nAMD patients as controls. A single-variant firch regression was performed using EPACTS v3.3.0.

### Whole-genome gene-based and pathway analysis

The gene-based analysis tested the whole-genome variant associations using the meta-analyzed GWAS summary statistics. The variants were annotated to genes based on the base pairs, adopting a 10 kb window at both sides. Analyses were performed using MAGMA<sup>79</sup> as implemented in FUMA v1.3.6a<sup>80</sup>. The major histocompatibility complex (MHC) region was omitted from the analysis. A total of 18,830 protein-coding genes based on Ensemble v92 were included. A threshold  $P$  value of  $0.05/18,830 = 2.66 \times 10^{-6}$  was used to demonstrate statistical significance, after correcting for multiple testing using the Bonferroni method. For pathway analysis, gene-based  $P$  values were used to test 5497 curated gene sets (including canonical pathways) and 9984 GO terms from MsigDB v7.0<sup>81</sup>. A threshold  $P$  value of  $0.05/15,481 = 3.23 \times 10^{-6}$ , derived after Bonferroni correction, was considered statistically significant in the pathway analysis.

### Processing and QC of whole-exome sequencing data

Whole-exome sequencing was done for 1098 samples (301 cases and 797 controls) using the Nimblegen SeqCap EZ (Roche) exome kit. Samples were sequenced to high coverage (target  $50 \times$ ) on the Illumina HiSeq 1000 platform or NovaSeq 6000 using 150 bp or 100 bp paired-end sequencing.

For read mapping and variant analysis, the resulting reads were aligned to a human reference genome (GRCh37) using Burrows-Wheeler Aligner (BWA, v0.7.10; <https://sourceforge.net/projects/bio-bwa/files/>)<sup>82</sup>. The adapters and sites with a low Phred score ( $<20$ ) were filtered and data were processed with the Genome Analysis ToolKit (GATK v4.0.3.0) best practices workflow to obtain jointly genotyped VCF files<sup>83</sup>. Reads were locally realigned (GATK IndelRealigner), and their base qualities were recalibrated (BQSR; GATK Recalibration Table). We performed multi-sample joint calling to generate single-nucleotide variant (SNV) calls across all samples<sup>84</sup>. Genotypes were first filtered by the GATK variant quality score recalibration (VQSR). Additional filters excluded the sites with low depth rates ( $<10 \times$ ), high depth rates ( $>300 \times$ ), excessive heterozygosity ( $P < 1 \times 10^{-6}$ ), HWE  $P$  value  $< 1 \times 10^{-4}$ , or missingness rate  $>10\%$ . The media coverage of the target was  $32 \times - 52 \times$ . (Supplementary Fig. 12).

For data quality control, we excluded 37 controls with early AMD, three typical nAMD patients and one Malay case. The rest of the 1057 individuals were of Chinese ancestry based on the PCA results. PCA was performed by GCTA v1.92.4 beta. We further excluded 38 cases with a missingness rate of genotypes  $>10\%$ . After variant QC, we included 637,799 variants for 1019 individuals (259 PCV and 760 controls). Our data included two batches of cases (137 and 122 subjects) derived using 150 bp or 100 bp paired-end sequencing. Thus, we performed a Fisher exact test using genotype calls to check the batch effect. Seventeen markers that failed to pass the  $P$  value threshold of  $1 \times 10^{-6}$  were set to missing. Among the remaining 637,782 variants, 159,036 were exonic nonsynonymous, and 43,918 had a minor allele frequency (MAF)  $>0.1\%$ . Most of the detected SNPs were rare variants: 83.41% of them were found with MAF  $<1\%$ , and 88.50% were found with MAF  $<5\%$ .

### Association analysis of rare variants

We performed both single-variant and gene-based analyses to test the association of rare variants' burden on PCV. For single-variant analysis, a logistic regression model with the Firth correction was performed using EPCATS v3.3.0, adjusted for age, gender, and the top four principal components. Those variants with MAF >0.1% (or equivalently minor allele count [MAC] >2) were included in the single-variant analysis. For gene-based analysis, we applied the Burden test, SNP-set (Sequence) Kernel Association Test (SKAT), and SNP-set Kernel Association Optimal Test (SKAT-O) on variants with MAF <2%. In total, 145,829 variants used for the test were assigned to 16,794 genes. Analyses were implemented in an R package "SKAT" v2.0.1 (<https://cran.r-project.org/web/packages/SKAT>). The *P* value threshold was set at  $2.98 \times 10^{-6} = 0.05/16,794$ , to be considered statistically significant in the gene-based analysis. In addition, we conducted a sensitivity analysis on rare variants at MAF <1%. A total of 142,783 variants were mapped to 16,765 genes, with the *P* value threshold setting at  $0.05/16,765 = 2.98 \times 10^{-6}$ .

### Functional variants annotation

We annotated the variants which passed QC using ANNOVAR. The functional variants that alter peptide sequences (missense), or lead to stop-gain, stop-loss, frameshifting, or splice-site disruption were included in the association analysis. Functional effects of variants were predicted by three different algorithms: SIFT<sup>85</sup>, PolyPhen-2<sup>86</sup>, and Combined Annotation Dependent Depletion (CADD)<sup>87</sup>. Missense variants that had met the following criteria were computationally predicted to be damaging or deleterious: (1) probably damaging (D) in PolyPhen; (2) deleterious in SIFT (D) and (3) CADD >20. CADD C-score greater than 20 indicates variants ranking in the top 1% of most deleterious mutations in the human genome<sup>87</sup>.

### Heritability and genetic correlation

We assessed the SNP-heritability (SNP- $h^2$ ), as well as the genetic correlation between PCV and typical nAMD using LD score regression (LDSC v1.0.1; <https://github.com/bulik/ldsc>), based on 888,842 Hapmap3 common variants with MAF  $\geq 1\%$  and imputation quality info >0.8. The 1000 Genomes Phase 3 reference panel of the East Asian population was used to compute linkage disequilibrium scores. We reported SNP- $h^2$  and genetic correlation at the liability scale to account for AMD prevalence, assuming the prevalence of AMD is 0.5% in the East Asian general population<sup>27,28</sup>.

### Expression quantitative trait locus (eQTL) analysis

We conducted eQTL analysis for the non-coding/intronic SNPs at the four AMD-associated novel loci identified using FUMA<sup>80,88</sup>. Using the eQTL analysis, we assessed whether gene expressions were affected by the lead SNPs. Genetic variants at four novel loci showed a similar LD pattern between East Asians and Europeans (Supplementary Fig. 13). For cis-SNP-gene pairs up to 1 Mb apart, we performed association tests across 53 human tissues in the GTEx dataset v.8<sup>89</sup>. A *P* value =  $9.43 \times 10^{-4}$  was set to be statistically significant for the cis-eQTL association after multiple testing corrections (0.05/53).

### Functional studies in animal models

**Animals.** All experiments using animals were approved by the Institutional Animal Care and Use Committee of the Agency for Science, Technology, and Research (A\*STAR) (IACUC, Protocol number: 181334) and SingHealth Experimental Medicine Centre (SEMC) (IACUC, Protocol number: 2018/SHS/1449) in Singapore. Animal care and procedures were performed in accordance with the Guide for Care and Use of Laboratory Animals from the US National Institutes of Health and the Statement for the Use of Animals in Ophthalmic and Vision Research from the Association for Research in Vision and Ophthalmology (ARVO). C57BL/6J mice were purchased from InVivos Pte Ltd.,

Singapore. All mice were housed in an environmentally controlled room (22°C, 40 – 60% humidity, and a 12-h light cycle) with food and water available ad libitum. Eyes were enucleated at the experimental endpoints. The neuroretina or RPE/choroid tissue was dissected from the eye cup before being lysed for RNA extraction<sup>33</sup>.

**Quantitative real-time PCR (qRT-PCR).** We tested selected genes and only included those expressed in mouse ocular tissues in subsequent diseased models. Total RNA was extracted from homogenized ocular tissues using NucleoSpin RNA, Mini kit (Macherey-Nagel, Düren, Germany) following manufacturers' instructions. RNA concentrations were determined using a Nanodrop 2000C Spectrophotometer 19 (Thermo Fisher Scientific, Waltham, MA, USA). cDNA was synthesized using qScript cDNA Supermix (Quanta BioSciences, Beverly, MA, USA) according to the manufacturer's instructions. Quantitative real-time PCR was performed in a total volume of 20  $\mu$ l containing PrecisionFAST 2x qPCR Mastermix (with SYBR green and low ROX) (Primerdesign, Camberley, UK) using LightCycler<sup>®</sup> 480 System (Roche, Basel, Switzerland). Data were analyzed using the  $2^{-\Delta\Delta C(T)}$  method for quantification of the relative mRNA expression levels. The primer sequences used in this study were purchased from IDT and are listed in Supplementary Data 17. All gene expressions were normalized to that of GAPDH.

**In silico analysis of transcription factor binding sites.** The location of the top 30 SNPs sequences located at *GATAS* loci on chromosome 20 was analyzed using the PROMO v8.3 software ([https://algggen.lsi.upc.es/cgi-bin/promo\\_v3/promo/promoinit.cgi?dirDB=TF\\_8.3](https://algggen.lsi.upc.es/cgi-bin/promo_v3/promo/promoinit.cgi?dirDB=TF_8.3)), which incorporates TRANSFAC v6.4<sup>90,91</sup>, followed by analysis using FunciSNP (<https://www.bioconductor.org/packages//2.12/bioc/html/FunciSNP.html>)<sup>92</sup>. PROMO parameters were chosen for human TFBS. The SNP sequence carrying each allele was loaded as the query sequence. A schematic representation of the analysis output and the summary of TF where lead SNPs are located are shown in Supplementary Fig. 8.

**Cells and cell culture.** Primary human pigmented epithelial cells (HRPEs) were purchased from Angio-Proteomie (cAP-0037, Boston, MA, USA) and maintained in HRPEs Growth Medium (contains 10% serum and growth supplements, cAP-33) following the manufacturer's instructions. HRPE cells were plated in a 6-well plate and cultured for 24 h. Cells were washed once with PBS before being treated with MEM media (Lonza Bioscience) containing 10  $\mu$ M retinoic acid, dexamethasone, or vehicle control dimethyl sulfoxide (DMSO) for another 24 h<sup>93</sup>. Cells were washed three times with PBS before being lysed for total RNA extraction using NucleoSpin RNA, Mini Kit (Macherey-Nagel, Düren, Germany) following the manufacturer's instructions. cDNA was synthesized using qscript cDNA Supermix (Quanta BioSciences, Beverly, MA, USA) according to the manufacturer's instructions. Quantitative real-time PCR was performed in a total volume of 20  $\mu$ l containing PrecisionFast 2x qPCR Mastermix (with SYBR green and low ROX) (Primerdesign, Camberley, UK) using a lightCycler 480 (Roche).

**Laser-induced CNV and RNA collection from RPE/choroid complex of the eye.** Laser-induced CNV was performed as previously described<sup>94</sup>. In brief, a laser photocoagulator (HGM Medical Laser Systems Inc. Salt Lake City, USA) was used to induce CNV in mixed gender 6- to 8-week-old C57BL/6J mice. Six laser burns were generated in each eye -1 disc diameter from the optic nerve. The laser settings used were 250mv power, 0.1s duration, and 100  $\mu$ m spot size. The eyes were enucleated on day 35 following the laser injury. RNA was extracted from the RRE/choroid tissues using the NAEasy mini kit (Qiagen, Hilden, Germany) and snap-frozen for gene expression analysis.

**Statistical analyses.** Data are represented as mean  $\pm$  standard error (SEM). Statistical comparison of the results was performed via

two-tailed, unpaired Student's *t* test, one-way ANOVA, or Dunnett's multiple comparisons test, using Prism 8.0 (GraphPad, La Jolla, CA, USA; <https://www.graphpad.com/>). Statistical significance is denoted with asterisks as follows: \**P* < 0.05; \*\**P* < 0.01; \*\*\**P* < 0.001 and \*\*\*\**P* < 0.0001.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

The summary statistics of the top index variants were presented in the Supplementary Data, along with the data analyzed in this study. The summary statistics file from the meta-analysis for nAMD from GAMA is available on the Amazon Web Services (AWS) from [https://gamagenetics.s3.ap-southeast-1.amazonaws.com/AMD\\_Topvariants\\_EAS\\_NC2023.txt](https://gamagenetics.s3.ap-southeast-1.amazonaws.com/AMD_Topvariants_EAS_NC2023.txt). To protect patient privacy, individual-level whole-exome sequencing data access is restricted. Those interested in accessing the data are encouraged to contact corresponding authors to discuss potential collaborations and sign data use agreements. All requests for data access will be reviewed by our internal committee. Responses will be provided within 4 weeks. All datasets generated during the study can be obtained by contacting the corresponding authors. The IAMDGC AMD GWAS results are downloaded from the GWAS Catalog [http://ftp.ebi.ac.uk/pub/databases/gwas/summary\\_statistics/GCST003001-GCST004000/GCST003219/](http://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/GCST003001-GCST004000/GCST003219/) under study accession identifier GCST003219. The 1000 Genomes Phase 3 data on Genome Reference Consortium Human Build 37 (GRCh37) is available at <https://www.internationalgenome.org/>. The datasets used for eQTL mapping and gene-set analysis through the FUMA platform (<https://fuma.ctglab.nl/>) are available through: GTEx eQTL v8 (<https://www.gtexportal.org/home/datasets>) and MSigDB v7.0 gene-set file (<https://www.gsea-msigdb.org/gsea/msigdb>). Source data supporting our findings (Fig. 3 and Supplementary Figs. 8, 9 and 12) were provided in this paper as a Source Data file. Source data are provided with this paper.

### References

- Wong, W. L. et al. Global prevalence of age-related macular degeneration and disease burden projection for 2020 and 2040: a systematic review and meta-analysis. *Lancet Glob. Health* **2**, e106–e116 (2014).
- Laude, A. et al. Polypoidal choroidal vasculopathy and neovascular age-related macular degeneration: same or different disease? *Prog. Retin. Eye Res.* **29**, 19–29 (2010).
- Wong, C. W. et al. Age-related macular degeneration and polypoidal choroidal vasculopathy in Asians. *Prog. Retin. Eye Res.* **53**, 107–139 (2016).
- Cheung, C. M. G. et al. Polypoidal choroidal vasculopathy: definition, pathogenesis, diagnosis, and management. *Ophthalmology* **125**, 708–724 (2018).
- Yeung, L. et al. Angiographic subtypes of polypoidal choroidal vasculopathy in Taiwan: a prospective multicenter study. *Retina* **38**, 263–271 (2018).
- Maruko, I., Iida, T., Saito, M., Nagayama, D. & Saito, K. Clinical characteristics of exudative age-related macular degeneration in Japanese patients. *Am. J. Ophthalmol.* **144**, 15–22.e2 (2007).
- van Dijk, E. H. et al. European prevalence of polypoidal choroidal vasculopathy: a systematic review, meta-analysis, and forecasting study. *J. Clin. Med.* **11**, 4766 (2022).
- Anantharaman, G. et al. Polypoidal choroidal vasculopathy: pearls in diagnosis and management. *Indian J. Ophthalmol.* **66**, 896–908 (2018).
- Yanagi, Y., Foo, V. H. X. & Yoshida, A. Asian age-related macular degeneration: from basic science research perspective. *Eye* **33**, 34–49 (2019).
- Fenner, B. J. et al. Real-world treatment outcomes of age-related macular degeneration and polypoidal choroidal vasculopathy in Asians. *Ophthalmol. Retin.* **4**, 403–414 (2020).
- Cheung, C. M. G. et al. A prospective study of treatment patterns and 1-year outcome of Asian age-related macular degeneration and polypoidal choroidal vasculopathy. *PLoS ONE* **9**, e101057 (2014).
- Fritsche, L. G. et al. A large genome-wide association study of age-related macular degeneration highlights contributions of rare and common variants. *Nat. Genet.* **48**, 134–143 (2016).
- Cheng, C.-Y. et al. New loci and coding variants confer risk for age-related macular degeneration in East Asians. *Nat. Commun.* **6**, 1–10 (2015).
- Han, X. et al. Genome-wide meta-analysis identifies novel loci associated with age-related macular degeneration. *J. Hum. Genet.* **65**, 657–665 (2020).
- Winkler, T. W. et al. Genome-wide association meta-analysis for early age-related macular degeneration highlights novel loci and insights for advanced disease. *BMC Med. Genomics* **13**, 1–18 (2020).
- Ng, T. K. et al. Multiple gene polymorphisms in the complement factor h gene are associated with exudative age-related macular degeneration in Chinese. *Invest. Ophthalmol. Vis. Sci.* **49**, 3312–3317 (2008).
- Lee, K. Y. et al. Association analysis of CFH, C2, BF, and HTRA1 gene polymorphisms in Chinese patients with polypoidal choroidal vasculopathy. *Invest. Ophthalmol. Vis. Sci.* **49**, 2613–2619 (2008).
- Huang, L. et al. Different hereditary contribution of the CFH gene between polypoidal choroidal vasculopathy and age-related macular degeneration in Chinese Han people. *Invest. Ophthalmol. Vis. Sci.* **55**, 2534–2538 (2014).
- Fan, Q. et al. Shared genetic variants for polypoidal choroidal vasculopathy and typical neovascular age-related macular degeneration in East Asians. *J. Hum. Genet.* **62**, 1049–1055 (2017).
- Ng, T. K. et al. HTRA1 promoter variant differentiates polypoidal choroidal vasculopathy from exudative age-related macular degeneration. *Sci. Rep.* **6**, 1–9 (2016).
- Geerlings, M. J. et al. Geographic distribution of rare variants associated with age-related macular degeneration. *Mol. Vis.* **24**, 75–82 (2018).
- Seddon, J. M. et al. Rare variants in CFI, C3 and C9 are associated with high risk of advanced age-related macular degeneration. *Nat. Genet.* **45**, 1366–1370 (2013).
- Saksens, N. T. et al. Rare genetic variants associated with development of age-related macular degeneration. *JAMA Ophthalmol.* **134**, 287–293 (2016).
- Ratnapriya, R. et al. Family-based exome sequencing identifies rare coding variants in age-related macular degeneration. *Hum. Mol. Genet.* **29**, 2022–2034 (2020).
- Huang, L.-Z. et al. Whole-exome sequencing implicates UBE3D in age-related macular degeneration in East Asian populations. *Nat. Commun.* **6**, 1–6 (2015).
- Huang, L. et al. A missense variant in FGD6 confers increased risk of polypoidal choroidal vasculopathy. *Nat. Genet.* **48**, 640–647 (2016).
- Cheung, C. M. G. et al. Prevalence, racial variations, and risk factors of age-related macular degeneration in Singaporean Chinese, Indians, and Malays. *Ophthalmology* **121**, 1598–1603 (2014).
- Park, S. J. et al. Age-related macular degeneration: prevalence and risk factors from Korean National Health and Nutrition Examination Survey, 2008 through 2011. *Ophthalmology* **121**, 1756–1765 (2014).
- Grassmann, F., Heid, I. M., Weber, B. H. & Consortium, I. A. G. Recombinant haplotypes narrow the ARMS2/HTRA1 association



- signal for age-related macular degeneration. *Genetics* **205**, 919–924 (2017).
30. Stahl, A. et al. The mouse retina as an angiogenesis model. *Invest. Ophthalmol. Vis. Sci.* **51**, 2813–2826 (2010).
31. Fruttiger, M. Development of the retinal vasculature. *Angiogenesis* **10**, 77–88 (2007).
32. Grossniklaus, H. E., Kang, S. J. & Berglin, L. Animal models of choroidal and retinal neovascularization. *Prog. Retin. Eye Res.* **29**, 500–519 (2010).
33. Li, D., Zhang, J., Liu, Z., Gong, Y. & Zheng, Z. Human umbilical cord mesenchymal stem cell-derived exosomal miR-27b attenuates subretinal fibrosis via suppressing epithelial–mesenchymal transition by targeting HOXC6. *Stem Cell Res. Ther.* **12**, 1–17 (2021).
34. Ishikawa, K., Kannan, R. & Hinton, D. R. Molecular mechanisms of subretinal fibrosis in age-related macular degeneration. *Exp. Eye Res.* **142**, 19–25 (2016).
35. Gelse, K., Pöschl, E. & Aigner, T. Collagens—structure, function, and biosynthesis. *Adv. Drug Deliv. Rev.* **55**, 1531–1546 (2003).
36. Morlino, S. et al. COL1-related overlap disorder: a novel connective tissue disorder incorporating the osteogenesis imperfecta/Ehlers-Danlos syndrome overlap. *Clin. Genet.* **97**, 396–406 (2020).
37. Hardcastle, A. J. et al. A multi-ethnic genome-wide association study implicates collagen matrix integrity and cell differentiation pathways in keratoconus. *Commun. Biol.* **4**, 1–13 (2021).
38. Kimoto, K., Nakatsuka, K., Matsuo, N. & Yoshioka, H. p38 MAPK mediates the expression of type I collagen induced by TGF- $\beta$  2 in human retinal pigment epithelial cells ARPE-19. *Invest. Ophthalmol. Vis. Sci.* **45**, 2431–2437 (2004).
39. Yi, C. et al. Macrophage elastase (MMP12) critically contributes to the development of subretinal fibrosis. *J. Neuroinflammation* **19**, 78 (2022).
40. Jo, Y. J. et al. Establishment of a new animal model of focal subretinal fibrosis that resembles disciform lesion in advanced age-related macular degeneration. *Invest. Ophthalmol. Vis. Sci.* **52**, 6089–6095 (2011).
41. Grafe, I. et al. Excessive transforming growth factor-beta signaling is a common mechanism in osteogenesis imperfecta. *Nat. Med.* **20**, 670–675 (2014).
42. Pugh, C. W. & Ratcliffe, P. J. Regulation of angiogenesis by hypoxia: role of the HIF system. *Nat. Med.* **9**, 677–684 (2003).
43. Sottile, J. Regulation of angiogenesis by extracellular matrix. *Biochim. Biophys. Acta Rev. Cancer* **1654**, 13–22 (2004).
44. Karagiannis, E. D. & Popel, A. S. Identification of novel short peptides derived from the  $\alpha$ 4,  $\alpha$ 5, and  $\alpha$ 6 fibrils of type IV collagen with anti-angiogenic properties. *Biochem. Biophys. Res. Commun.* **354**, 434–439 (2007).
45. Sudhakar, A. & Boosani, C. S. Signaling mechanisms of endogenous angiogenesis inhibitors derived from type IV collagen. *Gene Regul. Syst. Biol.* **1**, 217–226 (2007).
46. Sudhakar, A. et al. Human tumstatin and human endostatin exhibit distinct antiangiogenic activities mediated by  $\alpha$ v $\beta$ 3 and  $\alpha$ 5 $\beta$ 1 integrins. *Proc. Natl Acad. Sci.* **100**, 4766–4771 (2003).
47. Hosoda, Y. et al. Genome-wide association analyses identify two susceptibility loci for pachychoroid disease central serous chorioretinopathy. *Commun. Biol.* **2**, 1–9 (2019).
48. Hosoda, Y. et al. CFH and VIPR2 as susceptibility loci in choroidal thickness and pachychoroid disease central serous chorioretinopathy. *Proc. Natl Acad. Sci.* **115**, 6261–6266 (2018).
49. Gorelik, A., Labriola, J. M., Illes, K. & Nagar, B. Crystal structure of the nucleotide-metabolizing enzyme NTPDase4. *Protein Sci.* **29**, 2054–2061 (2020).
50. Dwyer, K. M., Kishore, B. K. & Robson, S. C. Conversion of extracellular ATP into adenosine: a master switch in renal health and disease. *Nat. Rev. Nephrol.* **16**, 509–524 (2020).
51. Fletcher, E. L. Mechanisms of photoreceptor death during retinal degeneration. *Optom. Vis. Sci.* **87**, 269–275 (2010).
52. Galvao, J. et al. Adenosine A3 receptor activation is neuroprotective against retinal neurodegeneration. *Exp. Eye Res.* **140**, 65–74 (2015).
53. Aplin, F. P. et al. ATP-induced photoreceptor death in a feline model of retinal degeneration. *Invest. Ophthalmol. Vis. Sci.* **55**, 8319–8329 (2014).
54. Yang, D. et al. Activation of P2X receptors induces apoptosis in human retinal pigment epithelium. *Invest. Ophthalmol. Vis. Sci.* **52**, 1522–1530 (2011).
55. Fletcher, E. L. Advances in understanding the mechanisms of retinal degenerations. *Clin. Exp. Optom.* **103**, 723–732 (2020).
56. Uhlen, M. et al. Towards a knowledge-based human protein atlas. *Nat. Biotechnol.* **28**, 1248–1250 (2010).
57. Swamy, V. & McGaughey, D. Eye in a disk: eyeIntegrat human pan-eye and body transcriptome database version 1.0. *Invest. Ophthalmol. Vis. Sci.* **60**, 3236–3246 (2019).
58. Tan, X. et al. Excessive retinol intake exacerbates choroidal neovascularization through upregulated vascular endothelial growth factor in retinal pigment epithelium in mice. *Exp. Eye Res.* **131**, 77–83 (2015).
59. Iriyama, A. et al. A2E, a pigment of the lipofuscin of retinal pigment epithelial cells, is an endogenous ligand for retinoic acid receptor. *J. Biol. Chem.* **283**, 11947–11953 (2008).
60. Yang, M. et al. Inhibition of mitochondrial function by metformin increases glucose uptake, glycolysis and GDF-15 release from intestinal cells. *Sci. Rep.* **11**, 2529 (2021).
61. Ren, A. et al. Endothelial GATA5 positively regulates angiogenesis via cathepsin S-mediated Angpt2/Flk1 and MMP2/9 signaling pathways. *Biochem. Biophys. Res. Commun.* **609**, 111–118 (2022).
62. Hwang, J. C., Del Priore, L. V., Freund, K. B., Chang, S. & Iranmanesh, R. Development of subretinal fibrosis after anti-VEGF treatment in neovascular age-related macular degeneration. *Ophthalmic Surg. Lasers Imaging Retina* **42**, 6–11 (2011).
63. Bloch, S. B., Lund-Andersen, H., Sander, B. & Larsen, M. Subfoveal fibrosis in eyes with neovascular age-related macular degeneration treated with intravitreal ranibizumab. *Am. J. Ophthalmol.* **156**, 116–124.e1 (2013).
64. Cheung, C. M. G. et al. The evolution of fibrosis and atrophy and their relationship with visual outcomes in Asian persons with neovascular age-related macular degeneration. *Ophthalmol. Retina* **3**, 1045–1055 (2019).
65. Donato, M. D. et al. Nek6 and Hif-1 $\alpha$  cooperate with the cytoskeletal gateway of drug resistance to drive outcome in serous ovarian cancer. *Am. J. Cancer Res.* **5**, 1862–1877 (2015).
66. Yang, Z. M. et al. Predictive role of NEK6 in prognosis and immune infiltration in head and neck squamous cell carcinoma. *Front. Endocrinol.* **13**, 943686 (2022).
67. Zhao, C. & Vollrath, D. mTOR pathway activation in age-related retinal disease. *Aging* **3**, 346–347 (2011).
68. Zehetner, C. et al. Systemic upregulation of PDGF-B in patients with neovascular AMD. *Invest. Ophthalmol. Vis. Sci.* **55**, 337–344 (2014).
69. Edwards, M. M. et al. Idiopathic preretinal glia in aging and age-related macular degeneration. *Exp. Eye Res.* **150**, 44–61 (2016).
70. Edwards, M. M. et al. Subretinal glial membranes in eyes with geographic atrophy. *Invest. Ophthalmol. Vis. Sci.* **58**, 1352–1367 (2017).
71. Waseem, N. H. et al. Mutations in SPATA13/ASEF2 cause primary angle closure glaucoma. *PLoS Genet.* **16**, e1008721 (2020).

72. Rykaczewska, U. et al. PCSK6 is a key protease in the control of smooth muscle cell function in vascular remodeling. *Circ. Res.* **126**, 571–585 (2020).
73. Group, M. P. S. Laser photocoagulation of subfoveal neovascular lesions in age-related macular degeneration: results of a randomized clinical trial. *Arch. Ophthalmol.* **109**, 1220–1231 (1991).
74. Japanese Study Group of Polypoidal Choroidal Vasculopathy. Criteria for diagnosis of polypoidal choroidal vasculopathy. *Nippon Ganka Gakkai Zasshi* **109**, 417–427 (2005).
75. Majithia, S. et al. Cohort profile: the Singapore Epidemiology of Eye Diseases study (SEED). *Int. J. Epidemiol.* **50**, 41–52 (2021).
76. Johnson, E. O. et al. Imputation across genotyping arrays for genome-wide association studies: assessment of bias and a correction strategy. *Hum. Genet.* **132**, 509–522 (2013).
77. Mägi, R. & Morris, A. P. GWAMA: software for genome-wide association meta-analysis. *BMC Bioinforma.* **11**, 1–6 (2010).
78. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
79. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.* **11**, e1004219 (2015).
80. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).
81. Liberzon, A. et al. Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739–1740 (2011).
82. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
83. DePristo, M. A. et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).
84. Li, Y., Sidore, C., Kang, H. M., Boehnke, M. & Abecasis, G. R. Low-coverage sequencing: implications for design of complex trait association studies. *Genome Res.* **21**, 940–951 (2011).
85. Ng, P. C. & Henikoff, S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res.* **31**, 3812–3814 (2003).
86. Adzhubei, I. A. et al. A method and server for predicting damaging missense mutations. *Nat. Methods* **7**, 248–249 (2010).
87. Kircher, M. et al. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* **46**, 310–315 (2014).
88. Kundaje, A. et al. Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
89. Consortium, G. T. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).
90. Farré, D. et al. Identification of patterns in biological sequences at the ALGGEN server: PROMO and MALGEN. *Nucleic Acids Res.* **31**, 3651–3653 (2003).
91. Messeguer, X. et al. PROMO: detection of known transcription regulatory elements using species-tailored searches. *Bioinformatics* **18**, 333–334 (2002).
92. Coetzee, S. G., Rhie, S. K., Berman, B. P., Coetzee, G. A. & Nounshmehr, H. FunciSNP: an R/bioconductor tool integrating functional non-coding data sets with genetic association studies to identify candidate regulatory SNPs. *Nucleic Acids Res.* **40**, e139 (2012).
93. Zhang, D. et al. All-trans retinoic acid stimulates the secretion of TGF- $\beta$ 2 via the phospholipase C but not the adenylyl cyclase signaling pathway in retinal pigment epithelium cells. *BMC Ophthalmol.* **19**, 1–6 (2019).
94. Lambert, V. et al. Laser-induced choroidal neovascularization model to study age-related macular degeneration in mice. *Nat. Protoc.* **8**, 2197–2211 (2013).

## Acknowledgements

We thank the members of the GAMA consortiums of genome-wide association studies for making their data available. We gratefully appreciate all the participants or volunteers who participated in the studies. The computational work for this study was performed on resources of the National Supercomputing Centre, Singapore (<https://www.nsc.sg>). This work was supported by the Singapore National Medical Research Council (CSA/033/2012, SHF/PRISM003/2015, OFLCG/004/2018), EYE ACP—PRISM Precision Medicine Initiative—05/FY2018/EX/01-A31, and National Precision Medicine Programme (NPM) Phase II Funding (MOH-000588). Detailed funding information for each study is listed in the Supplementary Notes.

## Author contributions

Q.F. and C.Y.C. designed the study. Q.F., T.Y.W., C.M.G.C. and C.Y.C. obtained the funding. X.W., Y.C.T., M.Y., K.Y.C.T., W.K.L., J.X.T., C.J.C., L.J.C., J.A., S.D., M.M., P.T., K.H.P., C.P.P., C.C.K. and C.M.G.C. were involved in data collection and contributed to genotyping or sequencing. Q.F., H.L., X.W., Y.P.K., C.J.C., J.X.T. and W.K.L. were involved in data analysis. Q.F., H.L. and X.W. wrote the first draft of the paper and Y.C.T., L.J.C., T.Y.W., Y.Y., C.M.G.C. and C.Y.C. critically reviewed the manuscripts. All authors reviewed the manuscript draft. All authors read and approved the final manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-023-41256-z>.

**Correspondence** and requests for materials should be addressed to Qiao Fan or Ching-Yu Cheng.

**Peer review information** *Nature Communications* thanks Nobuhisa Mizuki, Kirstan Vessey and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023

<sup>1</sup>Center for Quantitative Medicine, Duke-NUS Medical School, National University of Singapore, Singapore, Singapore. <sup>2</sup>Ophthalmology & Visual Sciences Academic Clinical Program (Eye ACP), Duke-NUS Medical School, Singapore, Singapore. <sup>3</sup>Singapore Eye Research Institute, Singapore National Eye Centre, Singapore, Singapore. <sup>4</sup>Department of Ophthalmology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore. <sup>5</sup>Centre for Innovation and Precision Eye Health, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore. <sup>6</sup>Center for Vision Research, Duke-NUS Medical School, National University of Singapore, Singapore, Singapore. <sup>7</sup>Department of Ophthalmology, Tohoku University Graduate School of Medicine, Miyagi, Japan. <sup>8</sup>SingHealth Duke-NUS Institute of Precision Medicine, Singapore, Singapore. <sup>9</sup>SingHealth Duke-NUS Genomic Medicine Centre, Singapore, Singapore. <sup>10</sup>Cancer and Stem Cell Biology Program, Duke-NUS Medical School, Singapore, Singapore. <sup>11</sup>Laboratory of Genome Variation Analytics, Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore, Singapore. <sup>12</sup>Department of Ophthalmology and Visual Sciences, The Chinese University of Hong Kong, Hong Kong, China. <sup>13</sup>Department of Ophthalmology, Seoul National University College of Medicine, Seoul Metropolitan Government Seoul National University Boramae Medical Center, Seoul, Korea. <sup>14</sup>Department of Ophthalmology and Visual Sciences, Kyoto University Graduate School of Medicine, Kyoto, Japan. <sup>15</sup>Human Genetics, Genome Institute of Singapore, Singapore, Singapore. <sup>16</sup>Department of Ophthalmology, Seoul National University Hospital, Seoul, Korea. <sup>17</sup>Tsinghua Medicine, Tsinghua University, Beijing, China. <sup>18</sup>Department of Ophthalmology and Microtechnology, Yokohama City University, Yokohama, Japan. <sup>19</sup>These authors contributed equally: Qiao Fan, Hengtong Li, Xiaomeng Wang.

✉ e-mail: [qiao.fan@duke-nus.edu.sg](mailto:qiao.fan@duke-nus.edu.sg); [chingyu.cheng@duke-nus.edu.sg](mailto:chingyu.cheng@duke-nus.edu.sg)