Deep learning aging marker from retinal images unveils sex-specific clinical and genetic signatures

Olga Trofimova[1,2]*, Leah Böttger[1,2], Sacha Bors[1,2], Yating Pan[3,4], Bart Liefers[5,6], Jose D. Vargas Quiros[5,6], Victor A. de Vries[5,6], Michael J. Beyeler[1,2], David M. Presby[1,2], Dennis Bontempi[1,2], Janna Hastings[4,7,2], Caroline C. W. Klaver[5,6,8,9], and Sven Bergmann[1,2,10]*

[1] Department of Computational Biology, University of Lausanne, Switzerland
[2] Swiss Institute of Bioinformatics, Lausanne, Switzerland
[3] Department of Computational Linguistics, University of Zurich, Switzerland
[4] Institute for Implementation Science in Health Care, Faculty of Medicine, University of Zurich, Switzerland
[5] Department of Ophthalmology, Erasmus University Medical Center, Rotterdam, the Netherlands
[6] Department of Epidemiology, Erasmus University Medical Center, Rotterdam, the Netherlands
[7] School of Medicine, University of St. Gallen, Switzerland
[8] Department of Ophthalmology, Radboud University Medical Center, Nijmegen, the Netherlands
[9] Institute of Molecular and Clinical Ophthalmology, University of Basel, Switzerland
[10] Department of Integrative Biomedical Sciences, University of Cape Town, South Africa

* Corresponding authors: olga.trofimova@unil.ch and sven.bergmann@unil.ch

Abstract

Retinal fundus images offer a non-invasive window into systemic aging. Here, we fine-tuned a foundation model (RETFound) to predict chronological age from color fundus images in 71,343 participants from the UK Biobank, achieving a mean absolute error of 2.85 years. The resulting retinal age gap, i.e. the difference between predicted and chronological age, was associated with cardiometabolic traits, inflammation, cognitive performance, all-cause mortality, dementia, cancer, and incident cardiovascular disease. Genome-wide analyses identified genes related to longevity, metabolism, neurodegeneration, and age-related eye diseases. Sex-stratified models revealed consistent performance but divergent biological signatures: males had younger-appearing retinas and stronger links to metabolic syndrome, while in females, both model attention and genetic associations pointed to a greater involvement of retinal vasculature. Our study positions retinal aging as a biologically meaningful and sex-sensitive biomarker that can support more personalized approaches to risk assessment and aging-related healthcare.

Introduction

Aging is strongly linked to disease susceptibility and mortality risk, yet its biological effects manifest at different rates across individuals—some people experience earlier onset of age-related decline, while others remain relatively healthy even at old age. The ability to measure biological age, distinct from chronological age, has therefore become a growing area of interest in biomedical research [1,2]. Estimators of biological age can offer valuable insights into disease susceptibility, predict health outcomes, and help identify individuals who may benefit from early interventions, such as medications or lifestyle changes [1,2].

Previous research has shown that deep learning models can estimate chronological age from retinal color fundus images (CFIs) with high accuracy [3–13]. Attention maps of such models reveal that the predictive signal draws on a range of anatomical features, including blood vessels [4,7,11,12], the macula [4,7,12], and the optic disc [4,11,12], suggesting that age-relevant information is distributed across multiple retinal regions. Importantly, the difference between predicted age and true age, known as retinal age gap (RAG), has emerged as a promising marker of biological aging. Several studies have reported associations between RAG and cardiovascular diseases (CVD) [5,12,14], chronic obstructive pulmonary disease [8], Parkinson's disease [15], cancer [12], and all-cause mortality [7,12,16], independent of traditional risk factors. Unlike other biological age markers that require more invasive procedures or costly laboratory assays, retinal age can be estimated cheaply and non-invasively from fundus photographs. This makes it a highly accessible and scalable tool for risk stratification in both clinical and population-level health settings.

As interest in biological age estimation grows, so does the need to understand how these estimates may vary across demographic groups, particularly between sexes. Research in neuroscience offers a relevant example: studies using brain imaging to estimate biological age have often found that female brains appear younger than male brains on average [17–19]. However, the findings are not fully consistent, as one study reports no sex differences [20], another finds female brains to appear older [21], and yet another shows region-specific variations in age estimates [22]. Despite such inconsistencies, the overall evidence supports the importance of stratifying analyses by sex, especially given that the brain age gap is differentially associated with disease patterns and risk factors in females and males [19–22]. In contrast, retinal age studies to date have largely relied on models trained on combined male and female samples, leaving it unclear whether, and how, RAG and its clinical correlates differ between sexes.

Traditionally, deep learning in biomedical imaging has relied on convolutional neural networks (CNNs), but recent advances have introduced Vision Transformers as viable alternatives. While there is no consensus on which architecture performs best overall, evidence suggests performance is task- and data-dependent. For instance, Vision Transformers often outperform CNNs in classification tasks when large datasets and pretraining are available, whereas CNN-based models like nnU-Net continue to excel in 3D medical image segmentation, particularly when rigorously validated and carefully configured [23,24]. RETFound, a foundation model for retinal images pretrained using a masked autoencoder with a large Vision Transformer backbone, has shown strong performance when fine-tuned for disease classification tasks, outperforming other contrastive learning-based approaches [25]. However, RETFound has not yet been applied to age prediction or explored in the context of retinal aging.

In this study, we aimed to extend and deepen the current understanding of retinal aging through several complementary analyses (**Fig. 1**). First, we benchmarked the performance of RETFound in age prediction from CFIs against conventional CNNs. Second, we trained sex-specific models—using separate cohorts of females and males—to investigate whether modeling each sex independently enhanced performance or revealed differences in aging patterns. We then explored the sex-specific value of predicted retinal age and RAG, particularly with respect to disease associations and genetic correlates. Third, we sought to interpret the internal representations of the model by analyzing attention maps and comparing the model's latent space to explicit image features. Through these analyses, we aim to uncover the retinal correlates of biological aging, assess their clinical utility, and understand how they may differ between females and males.

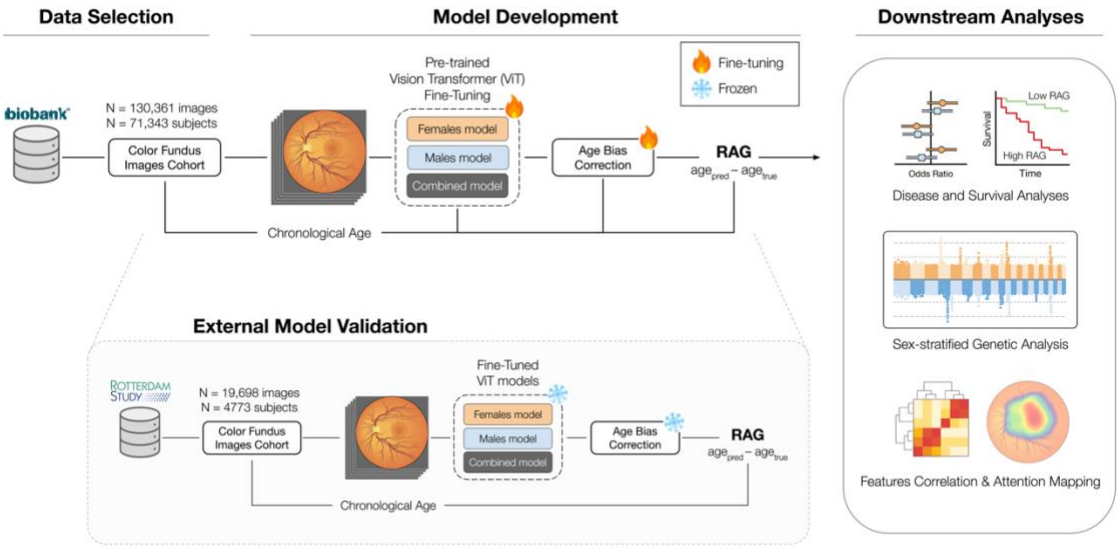

Figure 1. Study overview. We selected a subset of UK Biobank participants with retinal color fundus images that passed quality control. Using 5-fold cross-validation, we fine-tuned RETFound to predict age in a combined sample as well as in male- and female-specific samples. After correcting for regression-to-the-mean bias, we computed the retinal age gap (RAG). We conducted downstream analyses including disease associations, survival analysis, genome-wide association study, correlations between latent and explicit features, and model attention mapping, in both the combined and sex-specific samples. We then tested model performance and replication of genetic associations in an external dataset from the Rotterdam Study.

Results

RETFound predicts age with high accuracy and generalizes to an external dataset

Fine-tuning RETFound for age prediction yielded consistent performance across the five models trained via five-fold cross-validation (see **Table S1** for details). After pooling out-of-sample predictions across all folds (n = 130,360 fundus images from the UK Biobank (UKB)), the mean absolute error (MAE) was 2.99 years, Pearson's correlation (r) between predictions and chronological ages was 0.88, and the coefficient of determination (R^2) was 0.78 (**Table 1**). When averaging predictions across all available images from each participant at a given time point, the MAE improved to 2.85 years, with an r of 0.90 and an R^2 of 0.80. The model trained on a female-specific cohort achieved an MAE of 3.05 years, an r of 0.88, and an R^2 of 0.78 before averaging across images, which improved to an MAE of 2.92, an r of 0.89, and an R^2 of 0.80 after averaging. Similarly, the male-specific model showed an MAE of 3.09, an r of 0.88, and an R^2 of 0.77 before

averaging and improved to an MAE of 2.96, an *r* of 0.89, and an R^2 of 0.79 after averaging. The MAE difference between female and male models was not significant (z = -0.007, p = 0.99 for both image- and subject-level MAE).

To assess the generalizability of our models beyond the UKB, we tested their performance on independent sets of CFIs from the Rotterdam Study (RS). The image-level MAE was 3.58 years in the combined sample, 3.30 in females, and 3.86 in males (see **Table 1**). To quantify variability across models, we calculated the mean per-image standard deviation (SD) of the five model predictions, which was 0.99 years in the combined sample (28% of the MAE), 1.19 years in females (36% of the MAE), and 1.12 years in males (29% of the MAE). The subject-level MAE was 3.35 years in the combined sample, 3.07 years in females, and 3.61 years in males. The MAE was not significantly different between sexes, neither at the image level (z = -0.097, p = 0.92), nor at the subject level (z = -0.101, p = 0.92). While the squared correlation between true and predicted ages was 0.72 for the three models, the R^2 values were notably lower (0.61–0.70), reflecting a systematic bias in age prediction. The mean RAG values were -2.51 years in males, -0.88 years in females, and -1.94 years in the combined sample, indicating that the model consistently underestimated age in the RS, especially in males.

In both the UKB and the RS, predicted age tended to be overestimated in younger individuals and underestimated in older individuals (**Fig. S1a**), a regression-to-the-mean pattern commonly observed in aging biomarker studies [26]. Therefore, we applied the bias correction method proposed by Cole and colleagues [18], which rendered predicted age error orthogonal to chronological age (see **Fig. S1b**).

		UK Biobank					**Rotterdam Study**				
Sex	**Predictions**	**N**	**r**	**MAE**	**R^2**	**RAG mean (SD)**	**N**	**r**	**MAE**	**R^2**	**RAG mean (SD)**
Combined	**Image-level**	130,360	0.88	2.99	0.78	0.19 (3.86)	10,354	0.83	3.58	0.61	-2.07 (4.02)
	Subject-level	71,343	0.90	2.85	0.80	0.18 (3.66)	4757	0.85	3.35	0.66	-1.94 (3.73)
Female	**Image-level**	58,646	0.88	3.05	0.78	0.01 (3.94)	5849	0.83	3.30	0.66	-1.03 (4.08)
	Subject-level	32,918	0.89	2.92	0.80	0.002 (3.76)	2688	0.85	3.07	0.70	-0.88 (3.80)
	Male-trained model, subject-level	32,918	0.88	3.07	0.77	0.64 (3.89)	2688	0.84	3.47	0.62	-2.01 (3.88)

	Male-trained model, subject-level, corrected	32,918	0.88	3.36	0.73	0.73 (4.27)	2688	0.84	3.50	0.61	0.67 (4.38)
Male	**Image-level**	58,512	0.88	3.09	0.77	0.09 (4.00)	4505	0.83	3.86	0.56	-2.68 (4.08)
	Subject-level	32,918	0.89	2.96	0.79	0.07 (3.81)	2069	0.85	3.61	0.61	-2.51 (3.76)
	Female-trained model, subject-level	32,918	0.88	3.08	0.77	-0.66 (3.89)	2069	0.86	3.19	0.70	-1.51 (3.69)
	Female-trained model, subject-level, corrected	32,918	0.88	3.54	0.71	-0.83 (4.41)	2069	0.86	3.53	0.62	-0.81 (4.36)

Table 1. Model performance in the UK Biobank out-of-sample test sets and in the Rotterdam Study (external validation). Performance for female retinas evaluated on male-trained models, and vice versa, is also shown after applying age bias correction, to account for potential age distribution differences between females and males. MAE: mean absolute error, r: Pearson's correlation coefficient, R^2: coefficient of determination, RAG: retinal age gap, SD: standard deviation.

Retinal age gap is consistently lower in males

In the UKB combined cohort, males had a lower RAG than females on average (Δ = 0.28, t = 8.9, p = 7e-19, Cohen's d = 0.069; see **Fig. S2a**), indicating that the model estimated male retinas as appearing younger, though the effect was small. Additionally, cross-application of sex-specific models revealed that male retinas evaluated using the female-trained model had a mean RAG of -0.83, whereas female retinas evaluated using the male-trained model had a mean RAG of 0.73 after applying age bias correction (see **Table 1**), consistent with the observation that the models tended to estimate male retinas as appearing younger than female retinas. Given this somewhat unexpected sex difference, opposite to what might be anticipated based on lifespan trends, we explored whether it could be explained by anatomical differences, such as overall ocular size. Since body height is known to correlate with ocular dimensions [27], we regressed standing height out of the RAG. This adjustment attenuated the sex difference in RAG, which nonetheless remained significant (Δ = 0.16, t = 5.0, p = 6e-07, Cohen's d = 0.040; see **Fig. S2b**). Height and height-squared were therefore included as covariates in all subsequent analyses to account for potential confounding.

In the RS, we observed a similar pattern: males had lower mean RAG than females, indicating younger-appearing retinas (Δ = 0.37, t = 2.9, p = 0.004, Cohen's d = 0.085). When height was regressed out of the RAG, the sex difference was again reduced but not removed (Δ = 0.30, t = 2.4, p = 0.015, Cohen's d = 0.071). Cross-application of sex-specific models yielded

consistent results, with male retinas evaluated using the female-trained model showing a mean RAG of -0.81, and female retinas evaluated using the male-trained model showing a mean RAG of 0.67. These findings replicate the UKB results and further support the observation that retinal age predictions tended to estimate male retinas as appearing younger than female retinas.

Retinal age gap is associated with metabolic, inflammatory, and ocular health indicators

After adjusting for socio-demographic variables, higher RAG was significantly associated with increased pack-years of smoking, glycated hemoglobin (HbA1c), waist-to-hip ratio, body mass index (BMI), leukocyte count, C-reactive protein (CRP), glucose, triglycerides, systolic blood pressure (SBP), and slower reaction time in the combined sample (**Fig. 2a** and **Table S2** for detailed results). Conversely, RAG was negatively associated with VO_2max, insulin-like growth factor 1 (IGF-1), telomere length, low-density lipoprotein (LDL), high-density lipoprotein (HDL), and total cholesterol. Sex-specific associations generally mirrored the combined results, although males exhibited larger effect sizes across most biomarkers except CRP. Associations found in females but not males comprised neutrophil-to-lymphocyte ratio, LDL, total cholesterol, telomere length, and age at menopause, while male-specific associations included testosterone, SBP, triglycerides, and HDL. Absolute standardized effect sizes ranged from 0.01 to 0.09, reflecting the change in SD units of biomarker levels per SD increase in RAG.

For binary health outcomes, in the combined sample, higher RAG was associated with diabetes medication use, a history of cataract, diabetes, glaucoma, macular degeneration, cataract surgery, current or previous smoking status, use of cholesterol or blood pressure medications, metabolic syndrome, cancer (ever diagnosed), and hypertension (**Fig. 2b** and **Table S3**). Negative associations were observed with having children, visual and prospective memory performance, and alcohol consumption. Among females, postmenopausal status was associated with higher RAG, whereas hormone replacement therapy correlated with lower RAG. Males showed stronger associations than females with metabolic traits, including a male-specific association with metabolic syndrome, hypertension, and blood pressure medication. Odds ratios ranged from 0.94 to 2.06 per SD increase in RAG.

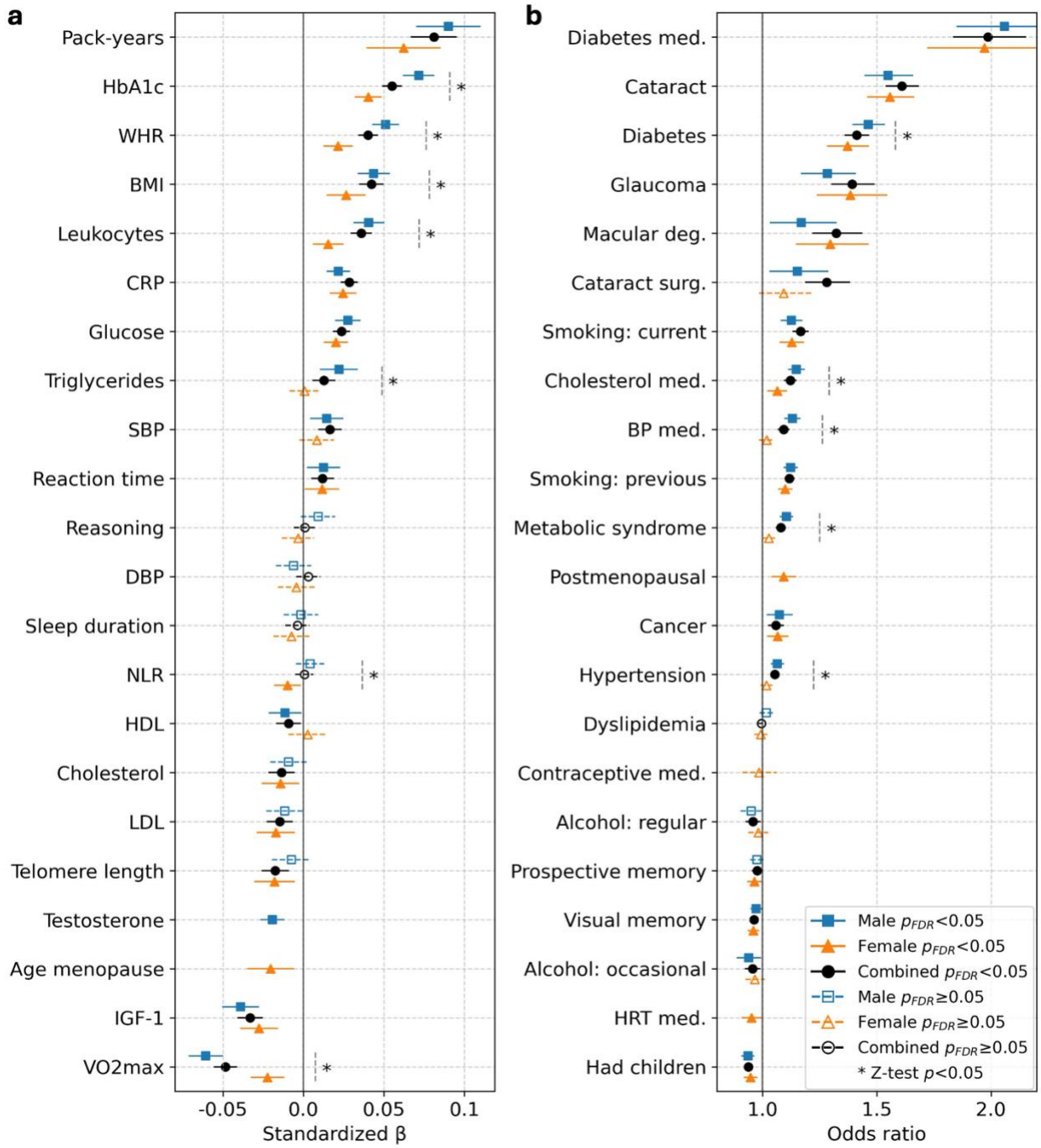

Figure 2. Standardized beta coefficients and odds ratios from models linking retinal age gap to risk factors and diseases. (a) Ordinary least square linear regression and **(b)** logistic regression models were adjusted for socio-demographic factors. Outcomes are ordered by average effect size between the three sex categories. Error bars represent the 95% confidence interval. Female- and male-specific effects were compared using a z-test. Sample sizes and detailed results are shown in **Tables S2** and **S3**. BMI: body mass index, BP: blood pressure, CRP: C-reactive protein, DBP: diastolic blood pressure, deg.: degeneration, FDR: false discovery rate, HDL: high-density lipoprotein, HRT: hormone replacement therapy, IGF-1: insulin-like growth

factor 1, LDL: low-density lipoprotein, med.: medication, NLR: neutrophil-to-lymphocyte ratio, SBP: systolic blood pressure, surg.: surgery, WHR: waist-to-hip ratio.

Retinal age gap predicts mortality and incident disease with sex-specific patterns

In survival analyses adjusted for socio-demographics, common cardiovascular risk factors, and previous events of the same category, RAG was significantly associated with all-cause mortality as well as incidence of CVD, chronic obstructive pulmonary disease, stroke, dementia, thrombosis, ischemic heart disease, and cancer in the combined sample (see **Fig. 3** and **Table S4**). When stratified by sex, thrombosis was associated with RAG in females but not males, and hazard ratios (HRs) for CVD and stroke were larger in females, although not significantly. In contrast, RAG predicted dementia and cancer incidence uniquely in males and was more strongly associated with all-cause mortality in males, though confidence intervals were large and the sex difference was only significant for cancer. HRs ranged from 1.03 to 1.20 per SD increase in RAG.

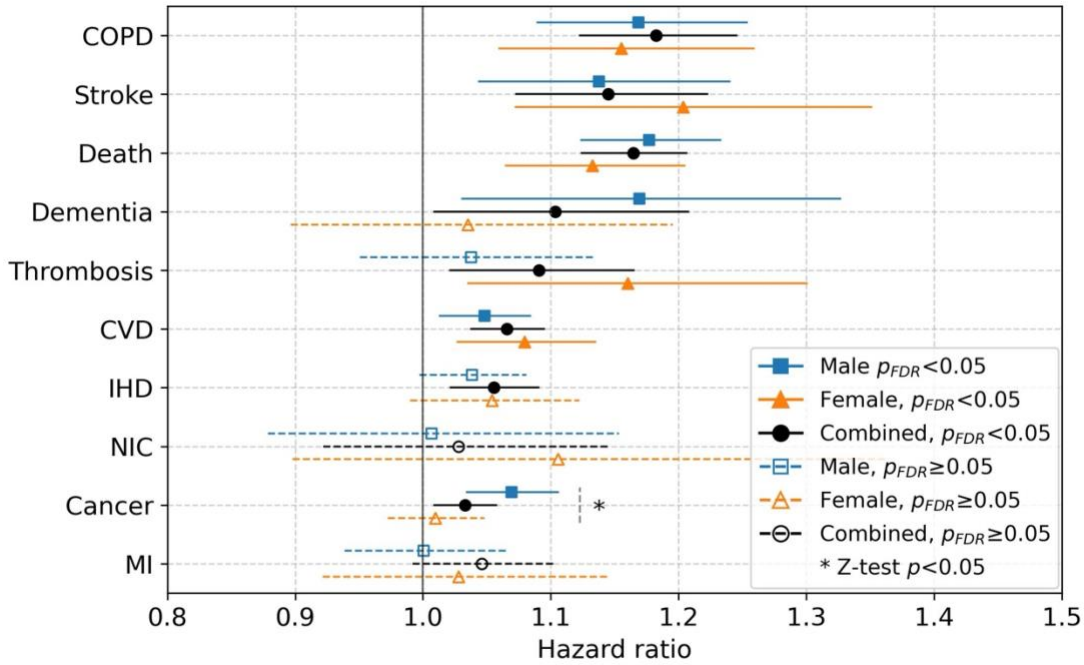

Figure 3. Hazard ratios (HR) from Cox proportional hazards regression analyses linking retinal age gap to mortality and diseases. Models were adjusted for socio-demographic factors, common cardiovascular risk factors, and previous events from the same category. Outcomes are ordered by average HR between the three sex categories. Error bars represent the 95% confidence interval. Female- and male-specific effects were compared using a z-test. Sample sizes, number of events, and detailed results are shown in **Table S4**. COPD: chronic obstructive pulmonary disease, CVD: cardiovascular disease, FDR: false discovery rate, IHD: ischemic heart disease, MI: myocardial infarction, NIC: non-ischemic cardiomyopathy.

Sex-stratified GWAS reveals shared and distinct genetic associations for retinal age gap

The genome-wide association study (GWAS) of RAG in the UKB identified 40 genome-wide significant loci in the combined sample (**Fig. 4a** and **Table S5**), 10 loci in females (**Fig. 4b**, top and **Table S6**), and 14 loci in males (**Fig. 4b**, bottom and **Table S7**). Gene-wise analysis using PascalX [28] revealed 75 genes across 23 independent linkage disequilibrium (LD) blocks in the combined sample, 18 genes in 5 LD blocks in females, and 16 genes in 5 LD blocks in males, with an overlap of 5 genes between females and males (see **Tables S8–S10**).

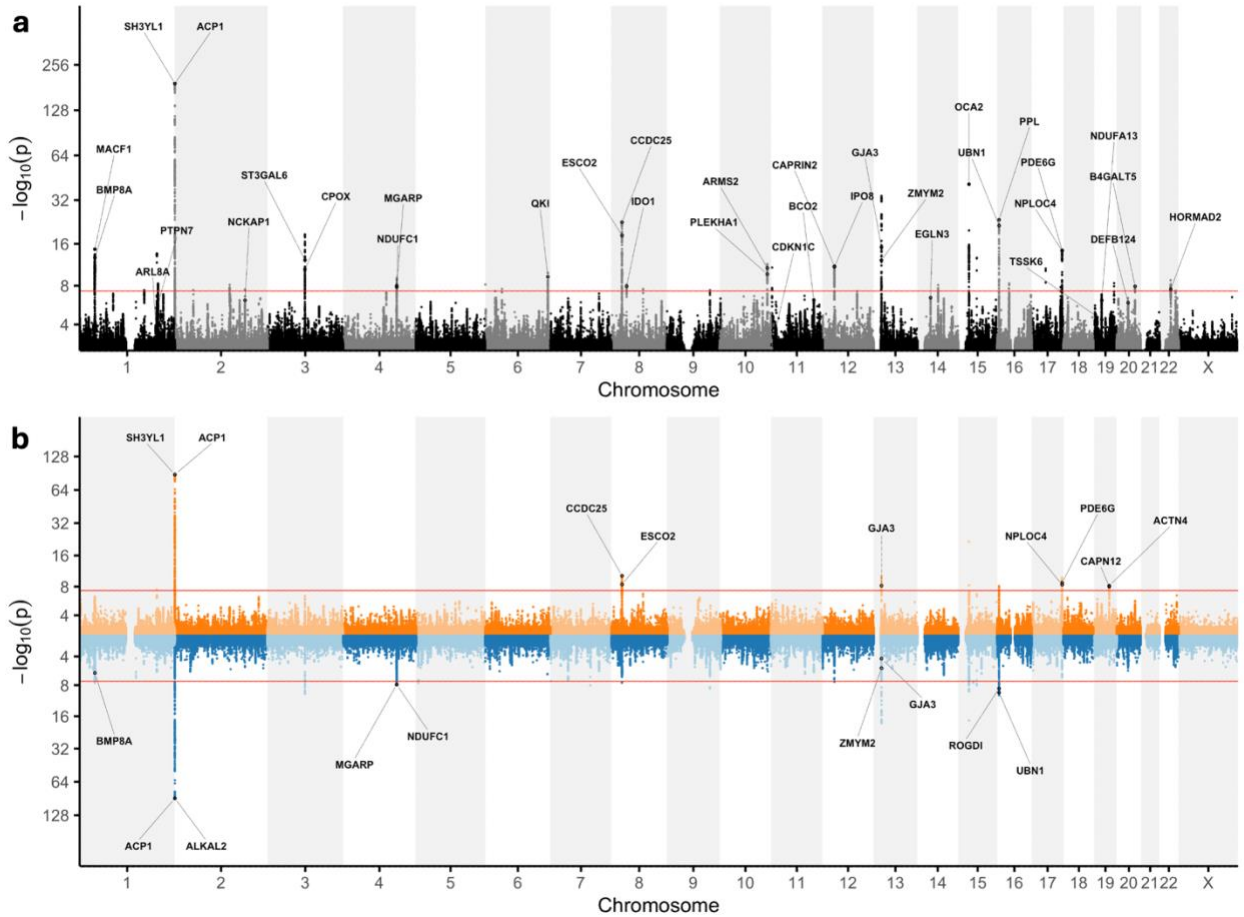

Figure 4. Manhattan plots from the genome-wide association study (GWAS) of retinal age gap, annotated with top genes. (a) Manhattan plot of the combined-sample GWAS (n=68,137). **(b)** Miami plot showing the female-specific GWAS (top, orange; n=30,698) and male-specific GWAS (bottom, blue; n=30,958). $-\log_{10}(p)$ values are plotted on a $\log_2$ scale. The red lines indicate genome-wide significance (5×10^{-8}). The annotated genes correspond to those identified as significant by the PascalX gene-based analysis. When more than two genes were identified at the same cytogenetic location, only the top two signals are labeled for clarity (the full gene lists are shown in **Tables S8–S10**).

In the combined sample, the strongest SNP association was observed for rs2290911, mapping within *SH3YL1* on chromosome (chr) 2. Gene-level analysis further revealed prominent signals for *ACP1*, *ALKAL2*/*FAM150B*, and *FAM110C*, which are located in close proximity to *SH3YL1* within the 2p25.3 region. The next strongest signal was found for rs1800407 on chr 15, located within *OCA2*. This was followed by rs11840593 on chr 13, in proximity to *GJA3*. Other notable signals included rs75804753 on chr 16 near *UBN1* and *PPL*, rs13267051 on chr 8 near *CCDC25* and *ESCO2*, rs10935473 on chr 3 within *ST3GAL6*, rs61779328 on chr 1 near *BMP8A* and *MACF1*, and rs3432 near *PDE6G* and *NPLOC4*. The sex-specific analyses revealed distinct genetic associations not observed in the combined sample. In females, unique signals included variants rs12714397 (chr2) and rs8182571 (chr19), and genes *CAPN12* and *ACTN4* (both on chr 19), the latter featuring the variant rs8182571. In contrast, rs79665333 (chr5), rs145132834 (chr7), and rs28844699 (chr15) were identified exclusively in males. Complete results are reported in **Tables S2–7**, and quality control plots for the GWAS analyses are presented in **Fig. S3**. The genomic control factor (λGC) was 1.14, 1.06, and 1.07 for the combined, female, and male GWAS, respectively, indicating modest inflation in the combined analysis and well-calibrated test statistics in the sex-stratified analyses.

The SNP-based heritability (h^2_{SNP}) of RAG, estimated using GREML, was 0.277 (± 0.009) for the combined sample, 0.278 (± 0.019) for females, and 0.247 (± 0.018) for males. The sex difference in h^2_{SNP} was not significant ($z = 1.15$, $p = 0.25$).

In the RS, 7 of 29 SNPs (11 unavailable) and 25 of 71 genes (4 unavailable) identified in the UKB combined model analysis also reached significance in the replication analysis (**Fig. S4**, **Tables S11–12**). The top SNPs in the RS, rs2290911 (chr2) and rs1800407 (chr15), were likewise the strongest signals in the UKB. Similarly, the top three genes, *ACP1*, *ALKAL2*/*FAM150B*, and *SH3YL1*, were identical across both cohorts and remained genome-wide significant in the RS. The remaining signals showed consistent directions of effect (see **Fig. S4c**), with enrichment beyond chance even when not significant, supporting robust genetic replication of RAG in the independent RS cohort.

Saliency and attention maps highlight the optic disc, macular area, and blood vessels

To better understand what the models rely on to predict age, we first computed saliency maps from the fine-tuned models using RELPROP [29], as in the original RETFound publication [25]. Saliency maps showed the strongest activation in the macula and the temporal half of the optic

disc, with sex-specific models displaying patterns highly similar to those of the combined-sample models (**Fig. 5a**, top row). A comparison between the youngest and oldest age quartiles revealed distinct patterns: in younger retinas, saliency was concentrated primarily in the macula (**Fig. 5b**, top row), whereas in older retinas, both the macula and the temporal optic disc contributed to the model's predictions (**Fig. 5c**, top row). Saliency maps were highly consistent across model folds and showed strong symmetry between left and right eyes (**Fig. S5**).

Attention maps extracted from attention blocks of the Vision Transformer were partially consistent with the saliency maps, highlighting the temporal optic disc across all groups, especially in the oldest quartile. However, unlike the saliency maps, attention maps emphasized the area surrounding the macula rather than the macula itself (**Fig. 5**, bottom rows). Additionally, the main vascular arches emerged as prominent regions of attention, particularly in female participants in the oldest quartile (**Fig. 5**, middle column, bottom row), suggesting that the model may rely more heavily on vascular features when predicting age in older female retinas. As observed with saliency maps, attention maps exhibited strong left–right eye symmetry, although they showed greater variability across model folds (see **Fig. S6**). Detailed maps from the 16 individual attention heads of the last block are shown in **Figures S7–S9**.

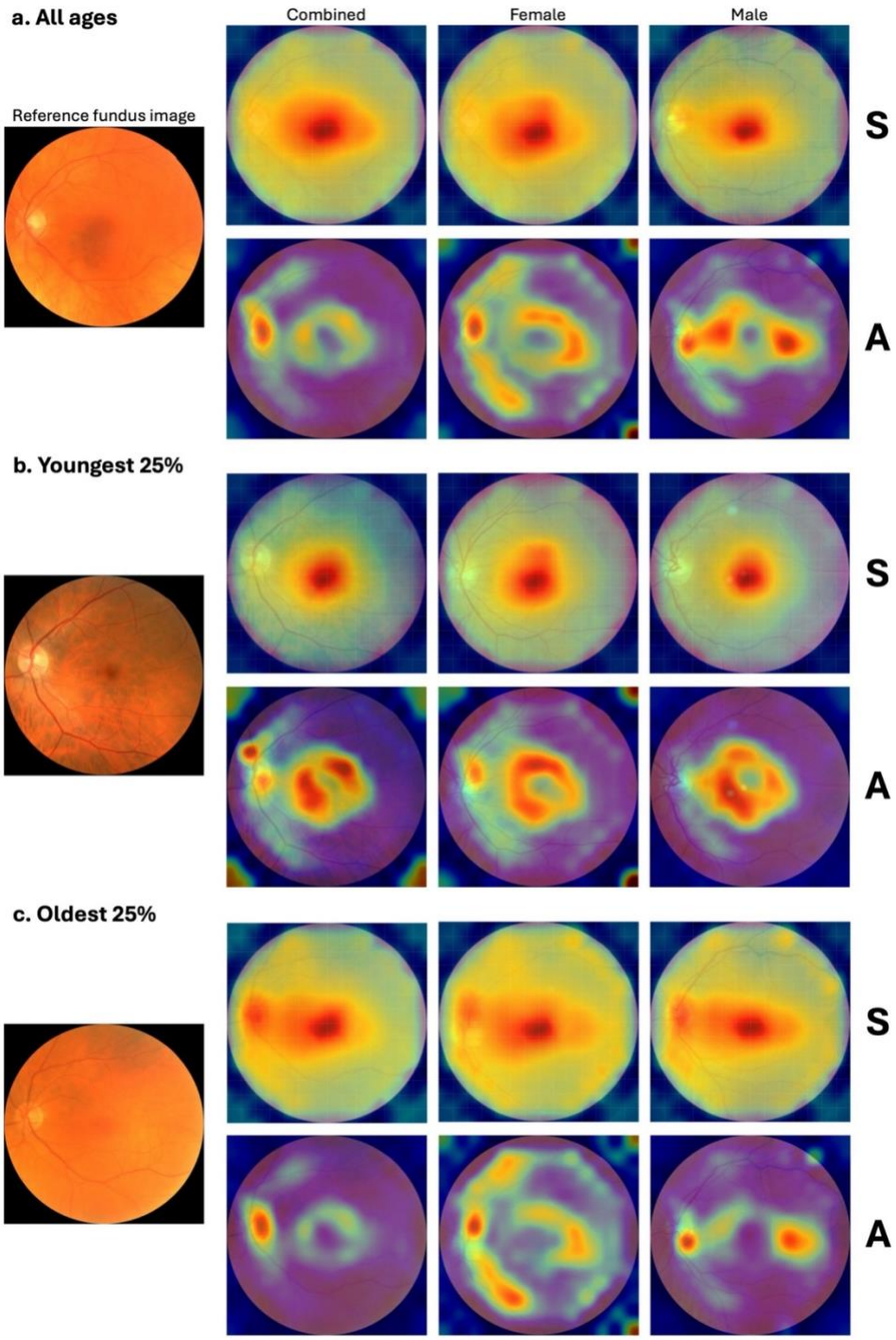

Figure 5. Group-average saliency and attention maps of the fine-tuned models on left eye images. Smoothed pixel-level saliency (S, upper rows) and attention (A, lower rows) maps of fold 1 models, obtained by bicubic interpolation from patch-level maps, are projected on a reference fundus image for **(a)** all ages (n=13,078 images combined sample, 5897 females and 5838 males), **(b)** the youngest quartile (n=3270 combined, 1475 females and 1460 males), and **(c)** the oldest quartile (n=3269 combined, 1474 females and 1460 males).

Model latent variables reflect vascular density, diameter, and bifurcations

We then examined the correlations between the model's latent variables (LVs) and interpretable image features, including vascular morphology, optic disc area, and RGB statistics. Most LVs were correlated with vascular density and bifurcation count (mean $|r| \in [0.26, 0.34]$). Notable correlations were also observed with central retinal arterial and venous equivalents (mean $|r| = 0.17$ and 0.13), artery diameter SD (mean $|r| = 0.14$), and to a lesser extent, median vascular diameters, and green channel mean and SD (mean $|r| \in [0.06, 0.14]$) (**Fig. 6**). The correlation pattern was highly consistent across all 15 models (5 folds × 3 groups), with correlation matrices showing strong agreement across folds and sex groups (**Figs. S10–S23**).

Interestingly, the correlation between image features and chronological age (shown for each feature in the rightmost column of **Fig. 6**) closely mirrored their correlation with the LVs. Vascular features such as density and bifurcations showed stronger correlations with age than with the LVs, while the opposite pattern was seen for RGB-based features, which correlated more with LVs than with age. This suggests that although the model's internal representations reflect vascular aging to some extent, they also encode additional non-age-related variation, particularly in color distribution and intensity, that may contribute to its age estimates.

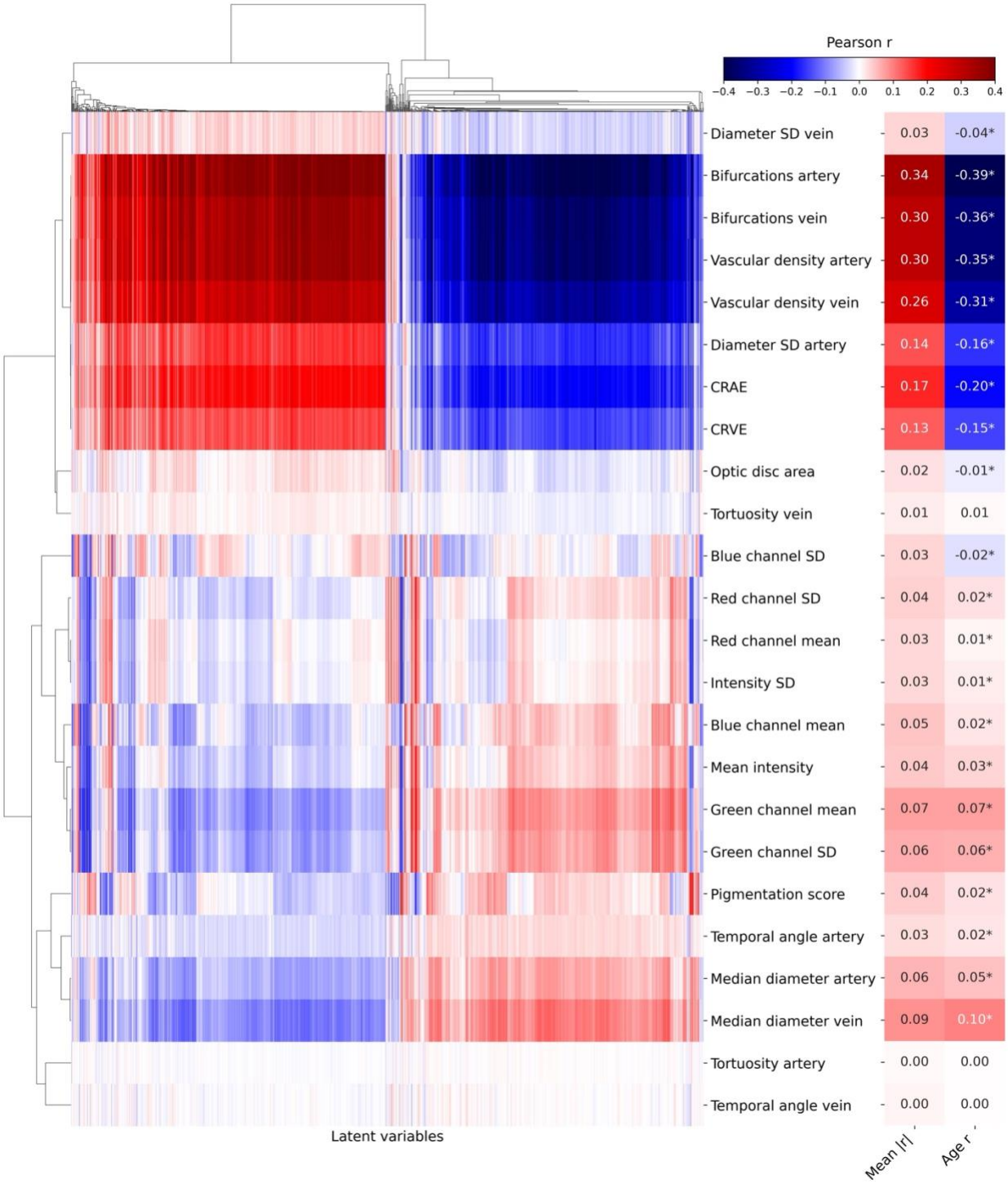

Figure 6. Correlation map between the model's 1024 latent variables and interpretable image features, in the combined sample for fold 1 (n=24,044 images). For every image feature, the mean absolute correlation with latent variables and the correlation with chronological age are shown on the right side of the figure. CRAE: central retinal arterial equivalent, CRVE: central retinal venous equivalent, SD: standard deviation, *: $p<0.05$.

Discussion

In this study, we fine-tuned RETFound to predict chronological age from retinal images, demonstrating robust generalization to an independent dataset with improved accuracy through averaging predictions from both eyes. We trained sex-specific models and found differences in RAG and its associations with clinical outcomes and genetic variants between females and males. RAG was associated with key health outcomes including CVD, cancer, and death, as well as eye disease, metabolic and inflammatory biomarkers. Genetic analyses identified novel and previously reported associations involved in aging-relevant pathways. Attention maps and latent space analyses pointed to the temporal optic disc, macular area, and blood vessels as important regions, and revealed correlations with vascular features.

With an MAE of 2.85 years and an R^2 of 0.80, RETFound fine-tuned to predict age on UKB data outperformed previous CNN-based models, which have reported MAE between 3.08 and 4.5 years and R^2 between 0.51 and 0.77 [3,6,7,9,10]. Although foundation models require considerable computational resources to pretrain, our results suggest they offer a measurable benefit in downstream regression tasks such as age prediction. When applied to the independent RS dataset, the model maintained a high correlation between predicted and true age but exhibited a consistent downward bias, especially in males. This systematic underestimation may be partially explained by the slightly older age distribution in the RS cohort (on average 0.3 years older than the UKB), combined with regression-to-the-mean effects that tend to produce larger underestimations at the upper end of the age range prior to bias correction. Furthermore, the lower age predictions in the RS could be due to technical differences between datasets, such as the camera model or image acquisition settings, or it may reflect genuine differences in participants' health, whereby RS retinas may appear biologically younger than UKB retinas.

Consistent with previous findings [5,9], we show that averaging predictions across images from both eyes improves accuracy, likely by reducing noise arising from variations in image quality or other image parameters. While this strategy may not be ideal in clinical datasets, as it may obscure pathologies affecting only one eye, or with asymmetrical disease progression, it is well-suited to address image quality variability in population-level datasets like the UKB. Therefore, the choice to average or not the predictions coming from both eyes should be guided by the specific research question and context of a given study.

We observed a modest but consistent sex difference in predicted retinal age across cohorts, with male retinas estimated as appearing younger. This stands in contrast to brain age models, which often estimate younger brain age in females [17–19], although findings vary across studies [20–22]. Both brain and retinal age predictions may be influenced by overall body or organ size. Brain age studies typically account for intracranial volume by using corrected image-derived phenotypes, whereas in our study, age was predicted directly from raw retinal images without adjusting for potential effects of eye size. Given that females and older individuals tend to have smaller bodies and possibly eye sizes on average, we hypothesize that this may have confounded the model, contributing to the observed overestimation of age in females. Supporting this, the sex difference in RAG was attenuated, though not eliminated, after adjusting for height. Further research is needed to explore anatomical scaling, physiological, hormonal or metabolic factors as potential drivers of the observed sex difference in retinal age predictions.

Interestingly, while cross-sectional associations between RAG and health indicators were generally stronger in males, RAG predicted incident CVD similarly in males and females, after adjusting for conventional risk factors such as BMI, smoking, SBP, cholesterol, and diabetes. In fact, HRs were higher in females for stroke, thrombosis, and overall CVD—though not significantly so, likely due to limited statistical power (**Table S4**). While males have a higher overall risk of CVD [29], our findings suggest that this difference may be largely captured by traditional risk factors included in the models. In contrast, the additional variance captured by RAG may be more informative in females, potentially reflecting sex-specific biological aging processes not well represented by traditional risk metrics. For example, women exhibit higher baseline inflammation, including elevated CRP and interleukin-6, which has been linked to greater CVD risk compared to men, even after adjusting for risk factors like smoking and BMI [30]. Hormonal dynamics and sex-specific metabolic changes in females may contribute to patterns of aging that are not well captured by conventional risk metrics, but are more readily reflected in RAG.

Conversely, RAG predicted future cancer and dementia risk in males but not in females. The association with cancer may reflect known biological differences: male cells experience greater oxidative stress and accumulate more somatic mutations than female cells [31], both of which increase susceptibility to malignant transformation and represent hallmarks of aging that may be captured by RAG. For dementia, we pooled all dementia subtypes due to limited case numbers, but we hypothesize the observed association may have been driven primarily by vascular dementia, given its shared link with metabolic syndrome and cardiovascular risk factors [32].

Genetic analyses of RAG revealed both novel loci and replication of previously reported associations [9,10,33]. Novel loci had been previously associated with corneal resistance factor (rs1200105) [34], educational attainment (rs79073127 and rs6477724) [35], and skin and eye cancer (rs12203592 and rs2413887) [36,37]. At the gene level, unlike traditional approaches that link only the most significant SNP to the nearest gene, our method aggregates signals across all SNPs within each gene region while accounting for LD, thereby improving sensitivity to detect biologically meaningful associations [28]. The genes we identified have been implicated in metabolism, lifespan, and retinal vessel density (*ACP1* [38,39], *SH3YL1* [39–41], *ALKAL2* [39,42], *MACF1-BMP8A* [43]), in age-related eye diseases and retinal vessel diameter variability (*GJA3* [44], *ARMS2* [45], *NPLOC4-PDE6G* [39,46]), cellular senescence (*UBN1*) [47], Alzheimer's disease (*CCDC25-ESCO2*) [48], and immune function (*PTPN7*) [49] among others. This is consistent with the clinical traits linked to RAG, supporting its biological relevance. Like previous retinal age studies [9,10,33], we also identified *OCA2*, a gene involved in eye pigmentation, suggesting that image features that are unlikely to be age-relevant may influence the model prediction. Among the female-specific associations, we identified two genes previously linked to retinal vessel tortuosity (*CAPN12* and *ACTN4*) [39,50,51], in line with our observation that model attention maps highlighted blood vessels more prominently in females.

Saliency maps consistently showed prominent model activation in the macula and temporal half of the optic disc, while attention maps partially aligned by highlighting the optic disc and surrounding macular regions. Crucially, attention maps also emphasized the main vascular arches in females—particularly in older retinas—mirroring our latent-variable analysis, which showed strong correlations with vessel density, bifurcation count, and vessel diameter. We hypothesize that the number of visible small vessels, which typically decreases with age [52], plays a role in model predictions. Interestingly, a recent study reported higher macular vascular density in males compared to females, with sex differences varying between foveal and parafoveal areas and across age groups [53]. This anatomical variation could help explain our observation that saliency maps focused on the fovea while attention maps emphasized the surrounding macular region. These findings suggest that the parafoveal vasculature may carry both age-relevant and sex-discriminative information, contributing to the model's ability to capture distinct biological aging patterns between sexes.

Finally, our findings must be interpreted within the context of the study's limitations. Although RETFound was pre-trained on data with some ethnic diversity, both the UKB and RS cohorts analyzed here were predominantly composed of individuals of White European ancestry. As such,

the generalizability of our results to other populations remains to be evaluated. Broader representation in training datasets will be essential to ensuring equitable performance across ancestries. Furthermore, both cohorts represent relatively healthy, aging populations; as such, the findings may not generalize to clinical populations with higher disease burden or atypical ocular presentations. The limited number of incident dementia cases, particularly within sex-specific strata, restricted our ability to examine subtype-specific associations with sufficient statistical power. Finally, despite leveraging saliency and attention maps, as well as correlations with interpretable image features, the precise biological mechanisms underlying the age-related signals captured by the models remain incompletely understood.

Overall, our study demonstrates that RAG offers a non-invasive, scalable biomarker that complements traditional indicators of biological aging. While training sex-specific models did not improve predictive accuracy, the distinct genetic and clinical correlates observed between females and males underscore the value of sex-stratified analyses in aging research. Incorporating sex-specific perspectives will therefore be essential for refining aging biomarkers and advancing their application in precision medicine.

Methods

Imaging data

We used color fundus images (CFIs) from UK Biobank (UKB) instances 0 and 1. These single-field, 45° color photographs, centered to capture the optic disc and macula, were acquired using a Topcon 3D OCT-1000 Mark II digital ophthalmic camera. To ensure image quality, we applied a quality control (QC) procedure based on the method described by Zekavat et al. [41], excluding the lowest-quality 25% of images. Briefly, a quality score was assigned to each image based on a deep learning model trained to identify poor-quality scans, and images below the 25th percentile of the score distribution were removed. This resulted in a final dataset of 130,360 images from 71,343 participants. We cropped black borders from all retained images [54]. Participants were between 39.2 and 79.1 years old (mean age = 57.7 ± 8.2 years). For the sex-specific analyses, we created age-matched cohorts of males and females, each comprising 32,918 individuals with a mean age of 57.9 ± 8.3 years and an age range of 40.1 to 75.8 years.

For the external validation of our models, we used data from the Rotterdam Study (RS), a prospective, population-based cohort study conducted in Ommoord, a district of the city of Rotterdam, The Netherlands [55]. It comprises four cohorts, initiated in 1989, 2000, 2006, and 2015, with a total of 17,931 participants. Participants undergo examinations approximately every five years, which include retinal imaging using various imaging modalities and camera systems across study rounds. For the model replication, CFIs acquired with the Topcon 3D OCT-2000 FA plus camera with a macula-centered field-of-view were used, as this modality was deemed most comparable to the UKB imaging while offering broad availability within the RS. In total, 37,771 fundus images were available from 7480 participants (14,912 eyes across 8813 visits). After filtering by age (40 < age ≤ 70 years, a range encompassing 97% of the UKB sample) and image quality (excluding the lowest 25%), 10,354 images from 4757 participants were retained for analysis (mean age = 58.0 ± 7.2 years). For the genetic analysis replication, we aimed to maximize sample size by selecting one assessment per participant according to the following rules: preferentially retaining Topcon 3D OCT-2000 FA macula-centered images when available, and otherwise using the closest available modality. When multiple assessments were present, we chose the visit with the highest average image quality. We applied the same age and quality filters. This resulted in a dataset of 11,368 participants (mean age = 60.6 ± 6.4 years) with 88,789 fundus images.

RETFound fine-tuning

RETFound is a Vision Transformer-based foundation model pre-trained on a large collection of unlabeled retinal images using self-supervised learning [25]. Originally designed for classification tasks, we adapted the architecture for regression to predict chronological age from CFIs. We employed a five-fold cross-validation framework, where participants were randomly divided into five non-overlapping groups. For each fold, we trained a separate model using 60% of participants for training, 20% for validation, and 20% for testing, with training, validation, and test sets rotating across folds. We used all available images, ensuring that, for a given fold, images from the same participant were assigned to the same set to avoid leakage between sets.

We resized images to 224 × 224 pixels. For data augmentation during training, we applied random crops and rescaling with a scale range of (0.8, 1.0) and aspect ratio range of (0.9, 1.1), along with random rotations up to 20°. We normalized all images using the ImageNet RGB mean and SD.

We trained models for 50 epochs, including 10 warm-up epochs. The learning rate followed a half-cycle cosine decay schedule: it increased linearly during the warm-up phase, then decreased following a cosine curve toward a minimum value. Specifically, the learning rate was initialized as 0.005×batch size/256, and decayed to a minimum of 1×10^{-6}. We used mean absolute error (MAE) as the loss function. The batch size for training was 16. Additional training hyperparameters included a drop path rate of 0.2, weight decay of 0.05, and layer-wise learning rate decay of 0.65. With the exception of the adjustments required for regression, we maintained the original RETFound training configuration as closely as possible.

We repeated the same five-fold cross-validation procedure separately for female and male participants using sex-specific, age-matched samples. This resulted in a total of 15 models: five for the combined sample, five for females, and five for males. This cross-validation setup ensured that each image had an out-of-sample prediction, allowing for robust performance assessment across the entire dataset. We evaluated the model performance using the MAE, Pearson's correlation (r) and coefficient of determination (R^2) between the predictions and the chronological ages. For external validation in the RS, we applied the five combined-model weights to the entire RS dataset and averaged the predicted ages to produce a single ensemble prediction per image. To evaluate overall performance, we computed the MAE, r, and R^2 between the ensemble predictions and the true ages. To assess the variability across individual models, we calculated the SD of the five model predictions for each image and reported the mean of these per-image SDs across the dataset. We repeated this procedure using the sex-specific models on the corresponding female and male RS samples.

Retinal age gap

To correct for regression-to-the-mean bias in predicted age, we applied a post hoc adjustment following the procedure described by Cole et al. [18]. Specifically, we fit a linear regression of predicted age on chronological age to estimate the slope (α) and intercept (β), then adjusted predicted age by subtracting β and dividing by α.

We then defined the retinal age gap (RAG) as the difference between the corrected predicted age and the participant's chronological age. To assess sex differences, we compared mean RAG values between females and males using independent-sample t-tests and quantified the effect size with Cohen's d. We performed this analysis in both the UKB and RS datasets, while subsequent disease associations were conducted using UKB data only.

Association with health and lifestyle factors

To investigate the relationship between RAG and various health outcomes, we conducted a cross-sectional analysis using data from the UKB. For participants with imaging data available at multiple time points, we used the first available instance.

Continuous outcomes included body mass index (BMI; data-field (DF) 21001), waist-to-hip ratio (WHR) defined as waist circumference (DF 48) divided by hip circumference (DF 49), systolic blood pressure (SBP; DF 4080 averaged across measurements), diastolic blood pressure (DBP; DF 4079 averaged across measurements), log-transformed pack-years of smoking (DF 20161), glycated hemoglobin (HbA1c; DF 30750), glucose (DF 30740), white blood cell (leukocyte) count (DF 30000), C-reactive protein (CRP; 30710), neutrophil (DF 30140) to lymphocyte (DF 30120) ratio (NLR), insulin-like growth factor 1 (IGF-1; DF 30770), low-density lipoprotein (LDL) cholesterol (DF 30780), high-density lipoprotein (HDL) cholesterol (DF 30760), triglycerides (DF 30870), total cholesterol (DF 30690), the VO_2max cardiorespiratory fitness estimate (DF 30038), sleep duration (DF 1160), reasoning score (also called fluid intelligence score; DF 20016), reaction time (DF 20023), and log-transformed telomere length (DF 22192). For participants who took antihypertensive medication, we added 10 mmHg to SBP and 5 mmHg to DBP values. Similarly, for participants who took cholesterol-lowering medication, we added 1.4 mmol/L to LDL, 0.4 mmol/L to triglycerides, 1.6 mmol/L to total cholesterol, and subtracted 0.1 mmol/L from HDL. Sex-specific associations additionally included testosterone (DF 30850) for males, and age at menopause (DF 3581) for females.

Binary outcomes included the use of diabetes, cholesterol, or BP medication (DFs 6153 and 6177), diabetes—defined as glucose >= 11.1 and/or HbA1c >= 48 and/or self-reported diagnosis by a doctor (based on DFs 2443 and 2976) and/or diabetes medication use—, hypertension—SBP > 140 and/or DBP > 90 and/or BP medication use—, dyslipidemia—cholesterol >= 5 and/or HDL <= 1 and/or LDL >=4.1 and/or triglycerides >= 2.2 and/or cholesterol medication use—, eye diseases (DF 6148) including cataract, glaucoma, and macular degeneration, cataract surgery (DF 5324), cancer history (DF 2453), metabolic syndrome—defined according to the International Diabetes Federation consensus [56] as having three or more of the following: large waist circumference (>= 80 cm for female or >= 94 cm for males), triglycerides > 1.7 or cholesterol medication, low HDL (< 1.29 for females or < 1.03 for males), high BP (SBP >= 130 and/or DBP >=85 and/or BP medication), and high glucose which we substituted with the above definition of diabetes since fasting glucose was not available—, smoking status (current/previous vs. never;

DF 20116), alcohol intake (occasional/regular vs. never; DF 1558), having children (binarized from DFs 2734 and 2405), visual memory (>= median in the pairs matching task; DF 399), and prospective memory (correct recall in first attempt; DF 20018). Additionally for females, we included the menopausal status (DF 2724), contraceptive pill and hormone replacement therapy (HRT; DF 6153).

All models controlled for age (derived from DFs 34, 52, and 53), age-squared, sex (DF 31), height (DF 50), height-squared, self-reported ethnic background (DF 21000, binarized as White/British vs. other given the low number of non-White participants), the first 10 genetic principal components (PCs; DF 22009), educational attainment (DF 6138)—categorized as low ("CSEs or equivalent", "NVQ or HND or HNC or equivalent", "Other professional qualifications eg: nursing, teaching", or "None of the above"), medium ("A levels/AS levels or equivalent", "O levels/GCSEs or equivalent"), or high ("College or University degree") [57]—, household income (DF 738)—categorized as low (< 18,000 £), medium (between 18,000 and 51,999 £), or high (>= 52,000 £)—, and Townsend Deprivation Index (DF 22189). We imputed missing covariate data using the group mean (for continuous covariates) or mode (for categorical covariates).

We standardized (z-scored) continuous outcomes and covariates, and excluded values that lay beyond 4 SD from the group mean. We applied linear regression models for standardized continuous outcomes and logistic regression for binary traits, with RAG entered as a standardized predictor. We conducted sex-stratified analyses to explore differential associations, and reported effect sizes as standardized beta coefficients or odds ratios. We corrected p-values for multiple testing (122 tests) using false discovery rate (FDR) [58] with an alpha threshold of 0.05.

Survival analysis

To assess the prospective associations between RAG and major health outcomes, we conducted Cox proportional hazards regression analyses using follow-up data spanning approximately 15 years from the UKB. We performed time-to-event analyses for all-cause mortality (DF 40000) and incident cases of chronic obstructive pulmonary disease (DFs 42016, 131486, 131488, 131490, and 131492), cancer (DF 40005), dementia (DFs 42018, 130836, 130838, 130840, and 130842), and cardiovascular disease (CVD), which encompassed stroke (DFs 42006, 131366, 131362, and 131364), thrombosis (DFs 131308, 131388, and 131400), myocardial infarction (DFs 131298, 131300, 131302, and 42000), ischemic heart disease (DFs 131296, 131304, and 131306), and non-ischemic cardiomyopathy (DFs 131338, 131340, 131288, and 131292). We adjusted models

for age, sex, height, ethnicity, education, income, Townsend, SBP, HDL, total cholesterol, smoking status, BMI, diabetes, and BP medication. We conducted analyses in the full sample and stratified by sex to examine potential sex-specific associations. We corrected p-values for multiple testing (30 tests) using FDR [58] with an alpha threshold of 0.05.

Genome-wide association study

We performed a genome-wide association study (GWAS) of RAG using Regenie [59], which accounts for relatedness and population structure via a two-step ridge regression approach. Before conducting the analyses, we applied filters for minor allele frequency (MAF ≥ 0.01), minor allele count (MAC ≥ 100), genotype missingness (≤ 0.1), and Hardy–Weinberg equilibrium (HWE, $p \geq 1\times10^{-15}$). Because quality control of the X chromosome remains debated given its biological specificities, we implemented additional steps [60]. We handled the X chromosome separately in each sex, using a relaxed missingness threshold for males (≤ 0.2) and excluding all ampliconic regions. We did not apply HWE filtering to the X chromosome, as the HWE test does not accommodate hemizygous males and may otherwise eliminate valid variants. Prior to the analysis, we applied a rank-inverse normal transformation to the RAG phenotype. The covariates included in the GWAS were age, sex (DF 31)—excluding cases of aneuploidy (DF 22019) or when genetic sex (DF 22001) did not match recorded or self-reported sex (DF 31)—, age-squared, age-by-sex, age-squared-by-sex, height, height-squared, instance, assessment centre (DF 21003), genotype measurement batch (DF 22000), spherical power (DFs 5084 and 5085), spherical power-squared, cylindrical power (DFs 5086 and 5087), cylindrical power-squared, and the first 20 genetic PCs. For sex-specific GWAS, the covariates were the same excluding sex.

We computed gene-level association scores using PascalX [28], which aggregates single nucleotide polymorphism (SNP)-level statistics into gene scores using the approximate “saddle” method. We scored protein-coding genes, including all SNPs with MAF ≥ 0.05, located within a 50 kb window around each gene. We used the UK10K reference panel [61] for gene scoring on the autosomes while we built the X chromosomal panel independently for females and males using 5000 unrelated white individuals per sex from the UKB.

We estimated SNP-based heritability (h^2_{SNP}) of RAG using the genomic-relatedness-based restricted maximum-likelihood (GREML) method implemented in the GCTA (Genome-wide complex trait analysis) software package [62]. We fitted a genetic relationship matrix (GRM)

constructed from common (MAF > 0.01) and directly genotyped variants to estimate the proportion of variance explained in RAG by all measured SNPs.

For the replication analysis, because imputation and QC were conducted independently in the four RS sub-cohorts, we first ran GWAS within each cohort and subsequently combined the results in a meta-analysis. For the individual GWAS, we applied linear regression models in PLINK 2.0 [63], adjusting for age, sex, age-squared, age-by-sex, age-squared-by-sex, height, height-squared, spherical power, spherical power-squared, cylindrical power, cylindrical power-squared, and the first 10 genetic PCs. We then meta-analyzed results across cohorts using an inverse-variance weighted fixed-effects approach implemented in METAL [64]. We derived association p-values from the corresponding *z*-statistics. We computed gene scores using PascalX [28] with the parameters described above. To assess replication of genetic findings, we adopted a candidate approach, applying FDR correction with a threshold 0.05 to p-values from SNPs, respectively genes, that were genome-wide significant in the UKB. We also computed Pearson's correlation between SNPs effect sizes (β) in the UKB and in the RS to assess consistency of results regardless of significance.

Saliency and attention maps

To interpret the internal representations learned by RETFound and visualize the model's implicit focus during prediction, we applied two distinct methods.
First, we generated saliency maps using RELPROP [29], a method that applies layer-wise relevance propagation (LRP) tailored for Transformer-based architectures, as used in the original RETFound publication [25]. RELPROP reveals model focus by integrating relevance scores with gradient information and propagating relevance backward through the whole network while preserving the total relevance across layers, creating more accurate attribution than conventional attention visualization alone. To accurately capture RETFound's prediction process, we adapted the RELPROP method to handle the global pooling architecture by distributing relevance equally across all non-CLS tokens, ensuring proper attribution for the regression pathways. Furthermore, we generated group-average saliency visualizations that highlight image regions consistently attended to by the model across a large set of input images, revealing which anatomical structures drive the model's decision-making process at the population level. This approach is meaningful in our context, as we restricted the analysis to left eye images (right eye saliency maps are shown in **Fig. S4**), ensuring that key anatomical features such as the optic disc and macula were consistently located across individuals.

Second, we extracted attention maps from the 24 attention blocks of the Vision Transformer, and computed attention rollout [65]. Specifically, we captured the spatial attention scores of the class token directed towards the patch tokens ("CLS→patch attention"), a commonly used proxy for identifying important image regions in Transformer-based models [66]. We implemented a forward hook to extract the softmax-normalized attention weights from each block. These weights represent the flow of information within the Transformer, indicating how much the CLS token relies on other patches, and form 14×14 spatial attention maps for each of the 16 attention heads. We then averaged the attention across all heads and aggregated across all blocks to obtain a single attention map per image [65]. Finally, similarly to saliency maps, we averaged attention maps across the cohorts to produce population-level visualizations of the regions most consistently attended to by the model.

Saliency maps and attention maps represent different but complementary approaches to interpreting Transformer-based models. Saliency maps operate in a backward direction, tracing the model's decision back through the network using relevance scores combined with gradients. This yields causal attributions by accounting for non-linear interactions across all layers and attention heads. In contrast, attention maps are obtained in the forward pass, using the softmax-normalized weights from the attention layers. These reflect the model's internal information routing (i.e., how much each input token is attended to by the CLS token), but do not indicate actual influence on the output, making them associative rather than causal. Because saliency maps trace decision-making and attention maps reveal focus patterns, differences between them can be informative. For example, the model might consistently attend to certain patches for context, while relying on different regions to make the final prediction.

Linking latent representations to interpretable image features

To investigate which interpretable image features the models may rely on to predict age, we examined Pearson's correlations between the 1024 latent variables (LVs), derived from the encoder portion of the models, and a comprehensive set of handcrafted retinal features. Specifically, we computed correlations between each LV and 24 image-derived variables: mean image intensity defined as the mean of the red, green, and blue channels, SD of image intensity, mean and SD of the red, green, and blue channels separately, optic disc area, central retinal arterial equivalent, central retinal venous equivalent, median and SD of artery and vein diameters, artery and vein tortuosity, artery and vein bifurcation counts, artery and vein temporal angles,

artery and vein vascular density, and pigmentation score. All vascular features were extracted using the VascX software [67]. For the optic disc segmentation, we annotated a random set of 100 CFIs and used them to re-train a publicly available U-Net model [68]. We computed the pigmentation score using the pipeline published in [69], which was developed to decouple traditional demographic variables from clinical imaging characteristics and has been previously validated on the UKB data. We computed correlations separately for each of the 15 trained models (across five cross-validation folds and three sex groups). Additionally, for the 24 image features, we computed correlations with the chronological age.

Data and code availability

UK Biobank data are available upon successful application (https://www.ukbiobank.ac.uk/enable-your-research/apply-for-access). RETFound code is publicly available at https://github.com/rmaphoh/RETFound. The code used to generate the presented results will be made available on GitHub upon publication (https://github.com/ot710/retinal-age).

Acknowledgements

This research has been conducted using the UK Biobank Resource under Application Number 90947. This work was funded by the Swiss National Science Foundation grant no. CRSII5 209510 for the “VascX” Sinergia project.

References

1. Mathur A, Taurin S, Alshammary S. New insights into methods to measure biological age: a literature review. Front Aging. 2024;5: 1395649. doi:10.3389/fragi.2024.1395649

2. Haugg F, Lee G, He J, Johnson J, Zapaishchykova A, Bitterman DS, et al. Imaging biomarkers of ageing: a review of artificial intelligence-based approaches for age estimation. Lancet Healthy Longev. 2025; 100728. doi:10.1016/j.lanhl.2025.100728

3. Poplin R, Varadarajan AV, Blumer K, Liu Y, McConnell MV, Corrado GS, et al. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. Nat Biomed Eng. 2018;2: 158–164. doi:10.1038/s41551-018-0195-0

4. Kim YD, Noh KJ, Byun SJ, Lee S, Kim T, Sunwoo L, et al. Effects of hypertension, diabetes, and smoking on age and sex prediction from retinal fundus images. Sci Rep. 2020;10: 4623. doi:10.1038/s41598-020-61519-9

5. Gerrits N, Elen B, Van Craenendonck T, Triantafyllidou D, Petropoulos IN, Malik RA, et al. Age and sex affect deep learning prediction of cardiometabolic risk factors from retinal images. Sci Rep. 2020;10: 9432. doi:10.1038/s41598-020-65794-4

6. Rim TH, Lee G, Kim Y, Tham Y-C, Lee CJ, Baik SJ, et al. Prediction of systemic biomarkers from retinal photographs: development and validation of deep-learning algorithms. Lancet Digit Health. 2020;2: e526–e536. doi:10.1016/S2589-7500(20)30216-8

7. Zhu Z, Shi D, Guankai P, Tan Z, Shang X, Hu W, et al. Retinal age gap as a predictive biomarker for mortality risk. Br J Ophthalmol. 2023;107: 547–554. doi:10.1136/bjophthalmol-2021-319807

8. Peng Q, Rim TH, Soh ZD, Chee ML, Tham Y-C, Zhu Z, et al. Predictive potential of retina-based biological age in assessing chronic obstructive pulmonary disease risk. Clin Experiment Ophthalmol. 2025. doi:10.1111/ceo.14501

9. Ahadi S, Wilson KA, Babenko B, McLean CY, Bryant D, Pritchard O, et al. Longitudinal fundus imaging and its genome-wide association analysis provide evidence for a human retinal aging clock. Elife. 2023;12. doi:10.7554/eLife.82364

10. Le Goallec A, Diai S, Collin S, Vincent T, Patel CJ. Deep learning of fundus and optical coherence tomography images enables identification of diverse genetic and environmental factors associated with eye aging. medRxiv. medRxiv; 2021. doi:10.1101/2021.06.24.21259471

11. Wang J, Gao Y, Wang F, Zeng S, Li J, Miao H, et al. Accurate estimation of biological age and its application in disease prediction using a multimodal image Transformer system. Proc Natl Acad Sci U S A. 2024;121: e2308812120. doi:10.1073/pnas.2308812120

12. Nusinovici S, Rim TH, Yu M, Lee G, Tham Y-C, Cheung N, et al. Retinal photograph-based deep learning predicts biological age, and stratifies morbidity and mortality risk. Age Ageing. 2022;51. doi:10.1093/ageing/afac065

13. Grimbly MJ, Koopowitz S-M, Chen R, Sun Z, Foster PJ, He M, et al. Estimating biological age from retinal imaging: a scoping review. BMJ Open Ophthalmol. 2024;9: e001794. doi:10.1136/bmjophth-2024-001794

14. Zhu Z, Chen Y, Wang W, Wang Y, Hu W, Shang X, et al. Association of retinal age gap with arterial stiffness and incident cardiovascular disease. Stroke. 2022;53: 3320–3328. doi:10.1161/STROKEAHA.122.038809

15. Hu W, Wang W, Wang Y, Chen Y, Shang X, Liao H, et al. Retinal age gap as a predictive biomarker of future risk of Parkinson's disease. Age Ageing. 2022;51. doi:10.1093/ageing/afac062

16. Nusinovici S, Rim TH, Li H, Yu M, Deshmukh M, Quek TC, et al. Application of a deep-learning marker for morbidity and mortality prediction derived from retinal photographs: a

cohort development and validation study. Lancet Healthy Longev. 2024;5: 100593. doi:10.1016/S2666-7568(24)00089-8

17. Goyal MS, Blazey TM, Su Y, Couture LE, Durbin TJ, Bateman RJ, et al. Persistent metabolic youth in the aging female brain. Proc Natl Acad Sci U S A. 2019;116: 3251–3255. doi:10.1073/pnas.1815917116

18. Cole JH, Ritchie SJ, Bastin ME, Valdés Hernández MC, Muñoz Maniega S, Royle N, et al. Brain age predicts mortality. Mol Psychiatry. 2018;23: 1385–1392. doi:10.1038/mp.2017.62

19. Subramaniapillai S, Schindler LS, Redmond P, Bastin ME, Wardlaw JM, Valdés Hernández M, et al. Sex-dependent effects of cardiometabolic health and APOE4 on brain age: A longitudinal cohort study. Neurology. 2024;103: e209744. doi:10.1212/WNL.0000000000209744

20. Phyo AZZ, Fransquet PD, Wrigglesworth J, Woods RL, Espinoza SE, Ryan J. Sex differences in biological aging and the association with clinical measures in older adults. GeroScience. 2024;46: 1775–1788. doi:10.1007/s11357-023-00941-z

21. Smith SM, Vidaurre D, Alfaro-Almagro F, Nichols TE, Miller KL. Estimation of brain age delta from brain imaging. Neuroimage. 2019;200: 528–539. doi:10.1016/j.neuroimage.2019.06.017

22. Sanford N, Ge R, Antoniades M, Modabbernia A, Haas SS, Whalley HC, et al. Sex differences in predictors and regional patterns of brain age gap estimates. Hum Brain Mapp. 2022;43: 4689–4698. doi:10.1002/hbm.25983

23. Takahashi S, Sakaguchi Y, Kouno N, Takasawa K, Ishizu K, Akagi Y, et al. Comparison of vision transformers and convolutional neural networks in medical image analysis: A systematic review. J Med Syst. 2024;48: 84. doi:10.1007/s10916-024-02105-8

24. Isensee F, Wald T, Ulrich C, Baumgartner M, Roy S, Maier-Hein K, et al. NnU-net revisited: A call for rigorous validation in 3D medical image segmentation. Lecture Notes in Computer Science. Cham: Springer Nature Switzerland; 2024. pp. 488–498. doi:10.1007/978-3-031-72114-4_47

25. Zhou Y, Chia MA, Wagner SK, Ayhan MS, Williamson DJ, Struyven RR, et al. A foundation model for generalizable disease detection from retinal images. Nature. 2023;622: 156–163. doi:10.1038/s41586-023-06555-x

26. de Lange A-MG, Cole JH. Commentary: Correction procedures in brain-age prediction. NeuroImage Clin. 2020;26: 102229. doi:10.1016/j.nicl.2020.102229

27. Nangia V, Jonas JB, Matin A, Kulkarni M, Sinha A, Gupta R. Body height and ocular dimensions in the adult population in rural Central India. The Central India Eye and Medical Study. Arbeitsphysiologie. 2010;248: 1657–1666. doi:10.1007/s00417-010-1448-0

28. Krefl D, Brandulas Cammarata A, Bergmann S. PascalX: a Python library for GWAS gene and pathway enrichment tests. Bioinformatics. 2023;39: btad296. doi:10.1093/bioinformatics/btad296

29. Chefer H, Gur S, Wolf L. Transformer Interpretability Beyond Attention Visualization. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE; 2021. doi:10.1109/cvpr46437.2021.00084

30. Kottilil S, Mathur P. The influence of inflammation on cardiovascular disease in women. Front Glob Womens Health. 2022;3: 979708. doi:10.3389/fgwh.2022.979708

31. Rubin JB, Lagas JS, Broestl L, Sponagel J, Rockwell N, Rhee G, et al. Sex differences in cancer mechanisms. Biol Sex Differ. 2020;11: 17. doi:10.1186/s13293-020-00291-x

32. Rasmussen IJ, Rasmussen K, Nordestgaard B, Tybjærg-Hansen A, Frikke-Schmidt R. Cardiovascular risk factors and risk of vascular-related dementia in high-risk and low-risk prospective cohort studies. Atherosclerosis. 2024;407: 120319. doi:10.1016/j.atherosclerosis.2024.118470

33. Huang Y, Syed MG, Chen R, Li C, Shang X, Wang W, et al. Genomic determinants of biological age estimated by deep learning applied to retinal images. GeroScience. 2025;47: 2613–2629. doi:10.1007/s11357-024-01481-w

34. He W, Han X, Ong J-S, Hewitt AW, Mackey DA, Gharahkhani P, et al. Association of novel loci with keratoconus susceptibility in a multitrait genome-wide association study of the UK Biobank database and Canadian Longitudinal Study on Aging. JAMA Ophthalmol. 2022;140: 568–576. doi:10.1001/jamaophthalmol.2022.0891

35. Okbay A, Wu Y, Wang N, Jayashankar H, Bennett M, Nehzati SM, et al. Polygenic prediction of educational attainment within and between families from genome-wide association analyses in 3 million individuals. Nat Genet. 2022;54: 437–449. doi:10.1038/s41588-022-01016-z

36. Liyanage UE, Law MH, Han X, An J, Ong J-S, Gharahkhani P, et al. Combined analysis of keratinocyte cancers identifies novel genome-wide loci. Hum Mol Genet. 2019;28: 3148–3160. doi:10.1093/hmg/ddz121

37. Verma A, Huffman JE, Rodriguez A, Conery M, Liu M, Ho Y-L, et al. Diversity and scale: Genetic architecture of 2068 traits in the VA Million Veteran Program. Science. 2024;385: eadj1182. doi:10.1126/science.adj1182

38. Lucarini N, Napolioni V, Magrini A, Gloria F. The Effect of ACP(1)-ADA(1) Genetic Interaction on Human Life Span. Hum Biol. 2012;84: 725–733.

39. Ortín Vela S, Beyeler MJ, Trofimova O, Iuliani I, Vargas Quiros JD, de Vries VA, et al. Phenotypic and genetic characteristics of retinal vascular parameters and their association with diseases. Nat Commun. 2024;15: 9593. doi:10.1038/s41467-024-52334-1

40. Choi GS, Min HS, Cha JJ. SH3YL1 protein as a novel biomarker for diabetic nephropathy in type 2 diabetes mellitus. Nutrition, metabolism, and cardiovascular diseases. NMCD. 2021;31: 498–505.

41. Zekavat SM, Raghu VK, Trinder M, Ye Y, Koyama S, Honigberg MC, et al. Deep learning of the retina enables phenome- and genome-wide analyses of the microvasculature. Circulation. 2022;145: 134–150. doi:10.1161/CIRCULATIONAHA.121.057709

42. Woodling NS, Aleyakpo B, Dyson MC, Minkley LJ, Rajasingam A, Dobson AJ, et al. The neuronal receptor tyrosine kinase Alk is a target for longevity. Aging Cell. 2020;19: e13137. doi:10.1111/acel.13137

43. Islam MR, The International Headache Genetics Consortium Ihgc, Nyholt DR. Genetic overlap analysis identifies a shared etiology between migraine and headache with type 2 diabetes. Genes (Basel). 2022;13: 1845. doi:10.3390/genes13101845

44. Tang X-J, Shentu X-C, Tang Y-L, Ping X-Y, Yu X-N. The impact of GJA3 SNPs on susceptibility to age-related cataract. Int J Ophthalmol. 2019;12: 1008–1011. doi:10.18240/ijo.2019.06.21

45. Fritsche LG, Loenhardt T, Janssen A, Fisher SA, Rivera A, Keilhauer CN, et al. Age-related macular degeneration is associated with an unstable ARMS2 (LOC387715) mRNA. Nat Genet. 2008;40: 892–896. doi:10.1038/ng.170

46. Choquet H, Melles RB, Anand D, Yin J, Cuellar-Partida G, Wang W, et al. A large multiethnic GWAS meta-analysis of cataract identifies new risk loci and sex-specific effects. Nat Commun. 2021;12: 3595. doi:10.1038/s41467-021-23873-8

47. Banumathy G, Somaiah N, Zhang R, Tang Y, Hoffmann J, Andrake M, et al. Human UBN1 is an ortholog of yeast Hpc2p and has an essential role in the HIRA/ASF1a chromatin-remodeling pathway in senescent cells. Mol Cell Biol. 2009;29: 758–770. doi:10.1128/MCB.01047-08

48. Blue EE, Thornton TA, Kooperberg C, Liu S, Wactawski-Wende J, Manson J, et al. Non-coding variants in MYH11, FZD3, and SORCS3 are associated with dementia in women. Alzheimers Dement. 2021;17: 215–225. doi:10.1002/alz.12181

49. Wang F, Wang X, Liu L, Deng S, Ji W, Liu Y, et al. Comprehensive analysis of PTPN gene family revealing PTPN7 as a novel biomarker for immuno-hot tumors in breast cancer. Front Genet. 2022;13: 981603. doi:10.3389/fgene.2022.981603

50. Tomasoni M, Beyeler MJ, Vela SO, Mounier N, Porcu E, Corre T, et al. Genome-wide Association Studies of retinal vessel tortuosity identify numerous novel loci revealing genes and pathways associated with ocular and cardiometabolic diseases. Ophthalmol Sci. 2023;3: 100288. doi:10.1016/j.xops.2023.100288

51. Jiang X, Hysi PG, Khawaja AP, Mahroo OA, Xu Z, Hammond CJ, et al. GWAS on retinal

vasculometry phenotypes. PLoS Genet. 2023;19: e1010583. doi:10.1371/journal.pgen.1010583

52. Wei Y, Jiang H, Shi Y, Qu D, Gregori G, Zheng F, et al. Age-related alterations in the retinal microvasculature, microcirculation, and microstructure. Invest Ophthalmol Vis Sci. 2017;58: 3804–3817. doi:10.1167/iovs.17-21460

53. Ghassemi F, Karimi M, Salari F, Bayat K. Exploring the impact of age and gender on retinal and choroidal thickness and vascular densities: a comprehensive analysis. Int J Retina Vitreous. 2025;11: 38. doi:10.1186/s40942-024-00619-4

54. Zhou Y, Wagner SK, Chia MA, Zhao A, Woodward-Court P, Xu M, et al. AutoMorph: Automated retinal vascular morphology quantification via a deep learning pipeline. Transl Vis Sci Technol. 2022;11: 12. doi:10.1167/tvst.11.7.12

55. Ikram MA, Brusselle G, Ghanbari M, Goedegebure A, Ikram MK, Kavousi M, et al. Objectives, design and main findings until 2020 from the Rotterdam Study. Eur J Epidemiol. 2020;35: 483–517. doi:10.1007/s10654-020-00640-5

56. Alberti KGMM, Zimmet P, Shaw J. Metabolic syndrome--a new world-wide definition. A Consensus Statement from the International Diabetes Federation. Diabet Med. 2006;23: 469–480. doi:10.1111/j.1464-5491.2006.01858.x

57. Chadeau-Hyam M, Bodinier B, Vermeulen R, Karimi M, Zuber V, Castagné R, et al. Education, biological ageing, all-cause and cause-specific mortality and morbidity: UK biobank cohort study. EClinicalMedicine. 2020;29-30: 100658. doi:10.1016/j.eclinm.2020.100658

58. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J R Stat Soc Series B Stat Methodol. 1995;57: 289–300. doi:10.1111/j.2517-6161.1995.tb02031.x

59. Mbatchou J, Barnard L, Backman J, Marcketta A, Kosmicki JA, Ziyatdinov A, et al. Computationally efficient whole-genome regression for quantitative and binary traits. Nat Genet. 2021;53: 1097–1103. doi:10.1038/s41588-021-00870-7

60. Böttger L, Bontempi D, Trofimova O, Beyeler MJ, Bors S, Iuliani I, et al. Sex-specific disease association and genetic architecture of retinal vascular traits. bioRxiv. 2025. doi:10.1101/2025.07.16.665150

61. Huang J, Howie B, McCarthy S, Memari Y, Walter K, Min JL, et al. Improved imputation of low-frequency and rare variants using the UK10K haplotype reference panel. Nat Commun. 2015;6: 8111. doi:10.1038/ncomms9111

62. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet. 2011;88: 76–82. doi:10.1016/j.ajhg.2010.11.011

63. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015;4: 7. doi:10.1186/s13742-015-0047-8

64. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010;26: 2190–2191. doi:10.1093/bioinformatics/btq340

65. Kashefi R, Barekatain L, Sabokrou M, Aghaeipoor F. Explainability of vision transformers: A comprehensive review and new perspectives. arXiv [cs.CV]. 2023. doi:10.48550/ARXIV.2311.06786

66. Caron M, Touvron H, Misra I, Jegou H, Mairal J, Bojanowski P, et al. Emerging properties in self-supervised vision transformers. 2021 IEEE/CVF International Conference on Computer Vision (ICCV). IEEE; 2021. doi:10.1109/iccv48922.2021.00951

67. Quiros JV, Liefers B, van Garderen K, Vermeulen J, Center, Eyened Reading, Consortium S, et al. VascX models: Model ensembles for retinal vascular analysis from color fundus images. arXiv [eess.IV]. 2024. doi:10.48550/ARXIV.2409.16016

68. Sevastopolsky A. Optic disc and cup segmentation methods for glaucoma detection with modification of U-Net convolutional neural network. Pattern Recognition and Image Analysis. 2017;27: 618–624. doi:10.1134/S1054661817030269

69. Rajesh AE, Olvera-Barrios A, Warwick AN, Wu Y, Stuart KV, Biradar MI, et al. Machine learning derived retinal pigment score from ophthalmic imaging shows ethnicity is not biology. Nat Commun. 2025;16: 60. doi:10.1038/s41467-024-55198-7