RetFit: A Novel Deep Learning-Based Biomarker of Cardiorespiratory Fitness Derived From the Retina

Authors | Ian Quintas[1,2,&], Dennis Bontempi[1,2,&], Sacha Bors[1,2], Olga Trofimova[1,2], Leah Böttger[1,2], Ilaria Iuliani[1,2], Sofía Ortín Vela[1,2], Bart Liefers[3,4], Jose Vargas-Quiros[3,4], Caroline C.W. Klaver[3,4,5,6], Ilenia Meloni[7,8], Adham Elwakil[7,8], Ciara Bergin[7,8], Mattia Tomasoni[7,8], VascX Consortium[$], Sven Bergmann[1,2,7,†], David M. Presby[1,2,†,*]

Affiliations | [1] Department of Computational Biology, University of Lausanne, Lausanne, Switzerland. [2] Swiss Institute of Bioinformatics, Lausanne, Switzerland. [3] Department of Ophthalmology, Erasmus University Medical Center, Rotterdam, The Netherlands. [4] Department of Epidemiology, Erasmus University Medical Center, Rotterdam, The Netherlands. [5] Department of Ophthalmology, Radboud University Medical Center, Nijmegen, The Netherlands. [6] Institute of Molecular and Clinical Ophthalmology, University of Basel, Switzerland. [7] Department of Ophthalmology, University of Lausanne, Fondation Asile des Aveugles, Jules Gonin Eye Hospital, Lausanne, Switzerland. [8] Platform for Research in Ocular Imaging, Fondation Asile des Aveugles, Jules Gonin Eye Hospital, Lausanne, Switzerland. [9] Department of Integrative Biomedical Sciences, University of Cape Town, Cape Town, South Africa.

&: joint first authors (ian.quintas@unil.ch and dennis.bontempi@unil.ch)

†: joint senior authors (david.presby@unil.ch and sven.bergmann@unil.ch)

$: see Author information for full list of co-authors.

***: Corresponding author |** David Presby, Ph.D., Department of Computational Biology; Genopode, office 2025.1 - CH1015 Lausanne, Switzerland.

e-mail: david.presby@unil.ch

Abstract

Cardiorespiratory fitness (CRF) is a strong predictor of cardiovascular events and all-cause mortality, often outperforming traditional risk factors. However, its clinical assessment remains limited due to the need for specialized equipment, personnel, and time demands. Because CRF is closely tied to vascular health, surrogate measures that capture vascular features may provide a practical alternative for its estimation. Retinal Color Fundus Images (CFIs) provide a non-invasive window into systemic vascular health and have already proven useful in predicting cardiovascular risk factors and diseases. However, CFIs have yet to be explored for their potential to predict CRF. In this study, we introduce RetFit, a novel CRF estimator derived from CFIs by leveraging state-of-the-art vision transformers. We evaluated RetFit's clinical relevance by analyzing its associations with cardiovascular risk factors and disease outcomes, and exploring its genetic architecture, benchmarking it against a submaximal-exercise-test CRF (SETCRF) estimate. RetFit was prognostic of both cardiovascular events (hazard ratios as low as 0.668, 95%CI 0.617–0.723, p<0.001) and overall mortality (hazard ratios as low as 0.780, 95%CI 0.754–0.801, p<0.001), and significantly associated with the majority of disease states and risk factors explored, with these effects being consistent across two external and independent cohorts. Although RetFit and SETCRF shared a moderate phenotypic correlation (*r*=0.45), their significant genetic associations were disjoint. Interpretability analyses suggest a role for retinal vasculature in RetFit's predictions, with attention maps emphasizing vascular regions and segmentation analyses showing arterial bifurcation count as the strongest associated feature (β=0.287, 95% CI 0.263–0.311, p<0.001). These findings highlight the potential of retinal imaging as a scalable, cost-effective, and accessible alternative for CRF estimation, supporting its use in large-scale screening and risk stratification in both clinical and public health contexts.

Introduction

Cardiorespiratory fitness (CRF) is a well-established, powerful, and independent predictor of cardiovascular events, mortality, and a wide range of disease outcomes[1,2]. Notably, CRF's predictive capacity often surpasses that of traditional risk factors such as smoking, hypertension, hypercholesterolemia, and type 2 diabetes, establishing CRF as one of the most powerful markers for clinical risk stratification[1,3–6]. Importantly, CRF is modifiable through lifestyle interventions, such as increased physical activity, and improvements in CRF are associated with reduced risks of cardiovascular events and overall mortality[7,8]. Therefore, efficiently and accurately estimating CRF can play a crucial role in the assessment of overall health status and disease risk. However, current gold-standard methods for measuring CRF are resource-intensive, requiring specialized equipment and trained personnel, which limits their feasibility in routine clinical and population-level settings.

One of the most widely used and informative measures of CRF is maximal oxygen uptake (VO_2max), which is defined as the highest rate at which an individual can consume oxygen during intense exercise. The gold standard for measuring VO_2max is the maximal cardiopulmonary exercise test (CPET)[1,9,10], which provides an objective and precise assessment of aerobic capacity by continuously measuring respiratory gas exchange during exercise to volitional exhaustion. However, despite its precision, CPET remains underutilized due to significant practical and clinical limitations: it requires expensive equipment, trained personnel, and a maximal volitional effort from the subject, and is contraindicated in subjects with certain cardiovascular predispositions or known conditions (e.g., active endocarditis, decompensated heart failure, unstable angina, physical/mental disability impairing movement)[10–12]. Consequently, submaximal- and non-exercise methods for estimating VO_2max have been developed and have demonstrated diagnostic and prognostic value for various health outcomes[13]. Submaximal exercise tests estimate CRF by modelling the relationship between heart rate (HR) response and work rate[14], whereas non-exercise

82 estimates of CRF may be derived from phenotypic, demographic, and anthropometric
83 variables, providing a rapid and cost-effective alternative suited for public health and clinical
84 settings[13,15]. In particular, estimates that utilize cardiovascular metrics, like resting heart rate,
85 demonstrate improved performance over estimates without cardiovascular measures,
86 potentially due to the cardiorespiratory system having a central, rate-limiting role in
87 determining VO_2max[16,17]. Therefore, identifying novel cardiovascular proxies may lead to
88 improvements in the accuracy and accessibility of CRF estimates.

89 Color fundus images (CFIs) represent a potentially promising source for such novel
90 cardiovascular proxies, as CFIs have been shown to provide valuable insights into the
91 cardiovascular system and reflect systemic vascular health[18,19]. Studies have shown that CFIs,
92 whether analyzed directly via end-to-end deep learning pipelines or through extracted features
93 quantifying retinal vascular properties, can reliably predict multiple cardiovascular risk factors
94 (e.g., age, sex, smoking status, and hypertension)[20,21] and diseases (e.g., myocardial
95 infarction, ischaemic stroke, and heart failure)[22,23]. Importantly, CFIs are noninvasive and
96 relatively easy to acquire, making them an appealing resource to estimate CRF. Building on
97 this rationale, we develop RetFit, a novel estimator of CRF from retinal images, based on a
98 state-of-the-art Vision Transformer (**Fig. 1a**). To determine the clinical relevance of RetFit, we
99 assessed its associations with genetics, risk factors, and diseases, and directly benchmarked
100 the diagnostic and prognostic capacities of this novel biomarker against an established
101 measure of submaximal exercise testing CRF (SETCRF)[24] (**Fig. 1b**). Furthermore, we
102 assessed RetFit's robustness and generalizability by externally validating the biomarker in two
103 external and independent cohorts[25,26] in which CRF was not measured (**Fig. 1c**). By showing
104 consistent results in the risk factor and disease association analyses, and strong prognostic
105 power through the survival analyses, we prove that RetFit captures clinically meaningful
106 information even in the absence of fitness-based measurements. Ultimately, this work aims to
107 provide a scalable and accessible method for estimating CRF, overcoming the key barriers of
108 conventional testing.

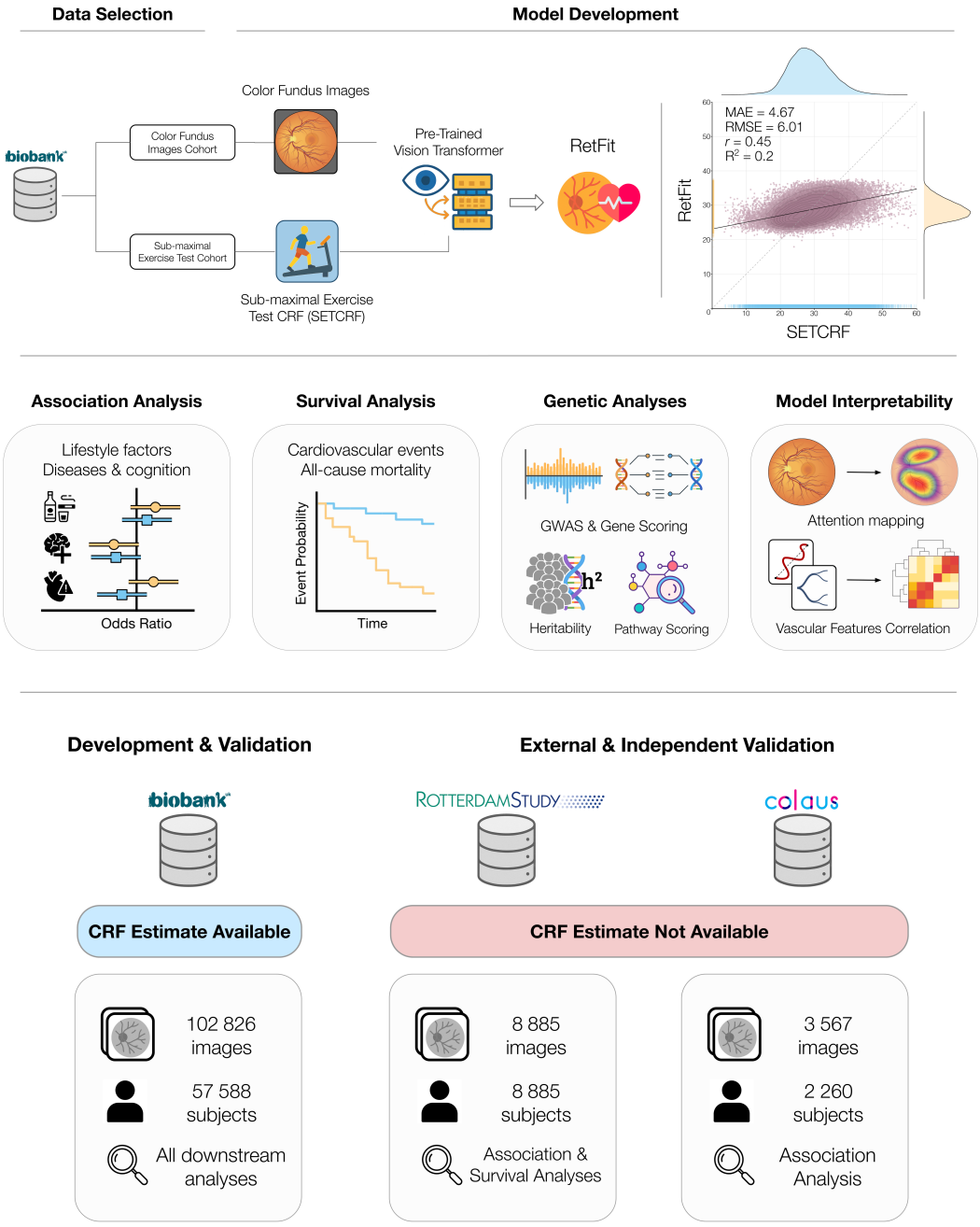


Figure 1 | Study Overview. a) We selected the subset of genotyped subjects from the UK Biobank who underwent a cardiorespiratory fitness (CRF) assessment via sub-maximal exercise testing (SETCRF) and had colour fundus images (CFIs) taken, as well as access to available health outcome information and risk factors, for a total of N=57 588 subjects and 102 826 images. We fine-tuned a state-of-the-art Vision Transformer to predict CRF from CFIs and obtained a CFI-based estimation of CRF, namely RetFit (16 outlier points not shown on scatter plot for displaying purposes). **b)** Using SETCRF as a benchmark in the UK Biobank dataset, we studied the association of RetFit with common lifestyle factors and diseases and assessed its prognostic power using survival analysis. Furthermore, we investigated the genetic architecture of both RetFit and SETCRF using GWAS, gene enrichment analyses, pathway scoring, and heritability estimation. We examined associations between retinal vascular properties and RetFit/SETCRF using attention mapping. **c)** Finally, we externally validated RetFit in two independent cohorts for which no CRF estimate was collected, the Rotterdam Study (N=8 885, 8 885 images; The Netherlands) and the CoLaus study (N= 2 260, 3 567 images; Switzerland), by running association and survival analyses.

Results

Generation of a Fitness Marker from Fundus Images

We fine-tuned a transformer model specifically designed for CFIs, RETFound[22], on 102 826 images taken from 57 588 participants in the UK Biobank (UKBB) who had performed a sub-maximal exercise test to estimate their CRF (SETCRF). The CFI-derived estimates of CRF, termed RetFit, demonstrated a moderate, positive correlation with SETCRF (r=0.45, p<0.001; R^2=0.20; MAE=4.67; RMSE=6.01; **Fig. 1a**). We then conducted additional analyses to evaluate RetFit's diagnostic and prognostic value, examine its genetic architecture, and benchmark these associations against SETCRF (**Fig. 1b**). The fine-tuned weights of the transformer model were then used to predict RetFit in two different cohorts, the Rotterdam Study (RS) and the CoLaus Study (CoLaus) (**Fig. 1c**).

RetFit Informs on Risk Factors and Diseases

We performed a series of regression analyses to evaluate the relative and independent contributions of RetFit and SETCRF to diseases and associated risk factors in the UKBB. First, we assessed their independent associations by building separate models with either RetFit or SETCRF and extracting their standardized coefficients. All the models were adjusted for sex, age, body mass index (BMI), and socioeconomic status to account for confounding effects. We then used log-likelihood ratio tests (LRT) to compare models that included both RetFit and SETCRF to those that included just one or the other to determine whether each predictor offered a unique, non-redundant contribution. To assess the generalizability of RetFit, we ran similar linear regression analyses on the two external and independent cohorts.

Of the 21 binary health outcomes examined, RetFit significantly informed upon 20 of them, whereas SETCRF significantly informed upon 15 outcomes (**Fig. 2a**). RetFit and SETCRF informed upon all binary health outcomes in the same direction except for the current smoking status (against non-smokers) and dyslipidemia. Furthermore, we found five associations were

154 significant for RetFit but not for SETCRF (cataract, cataract surgery, current smoking status,
155 atherosclerosis, and prospective memory). Macular degeneration was the only outcome for
156 which neither of the biomarkers was significantly associated. The LRT results for the binary
157 outcomes showed a significant improvement in model fit when adding both RetFit and
158 SETCRF for cataract, cataract surgery, cancer, atherosclerosis, and prospective memory. As
159 for the logistic regression models, macular degeneration was the only disease for which
160 adding either of the two biomarkers did not result in a significant gain of information. Cancer
161 diagnosis was the only outcome for which adding SETCRF to a model with RetFit improved
162 the model, whereas for all the other outcomes, adding RetFit improved the model with
163 SETCRF (**Fig. 2b**).

164 Of the 15 continuous risk factors examined in the UKBB, we found that RetFit was significantly
165 associated with all of them, while SETCRF showed significant associations with all but
166 reasoning (**Fig. 2c**). Of the 14 factors significantly associated with both RetFit and SETCRF,
167 10 demonstrated concordant directionality. Only insulin-like growth factor 1 (IGF-1), low-
168 density lipoprotein cholesterol (LDL), and total cholesterol exhibited opposing associations,
169 being positively associated with RetFit and negatively associated with SETCRF. These results
170 are consistent with results obtained for dyslipidemia in the binary setting, as we also found
171 opposing directions. Results from the log-likelihood ratio tests demonstrate that reasoning and
172 triglycerides were the only two outcomes for which adding one predictor to a model already
173 containing the other led to a significant improvement in model fit (**Fig. 2d**).

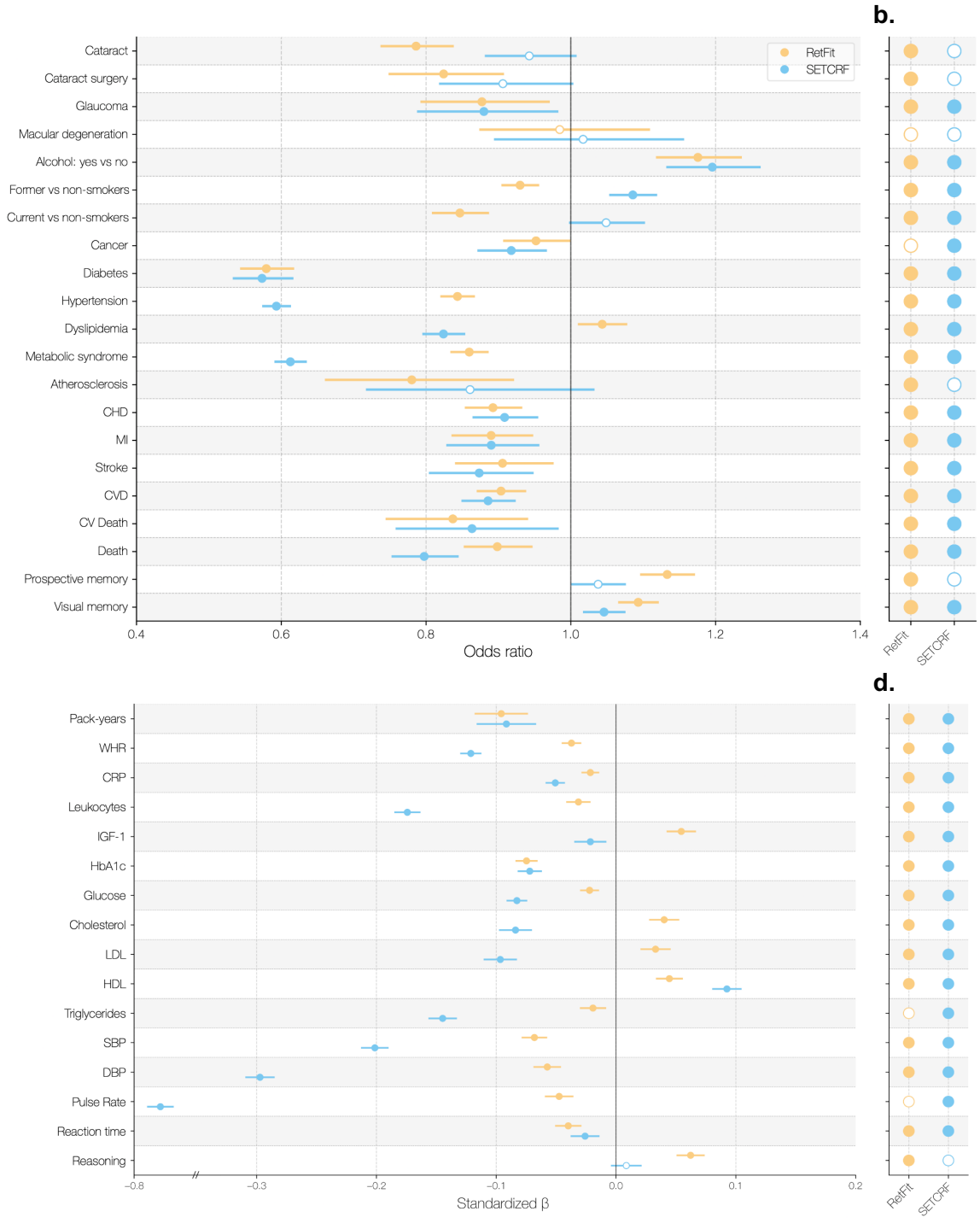
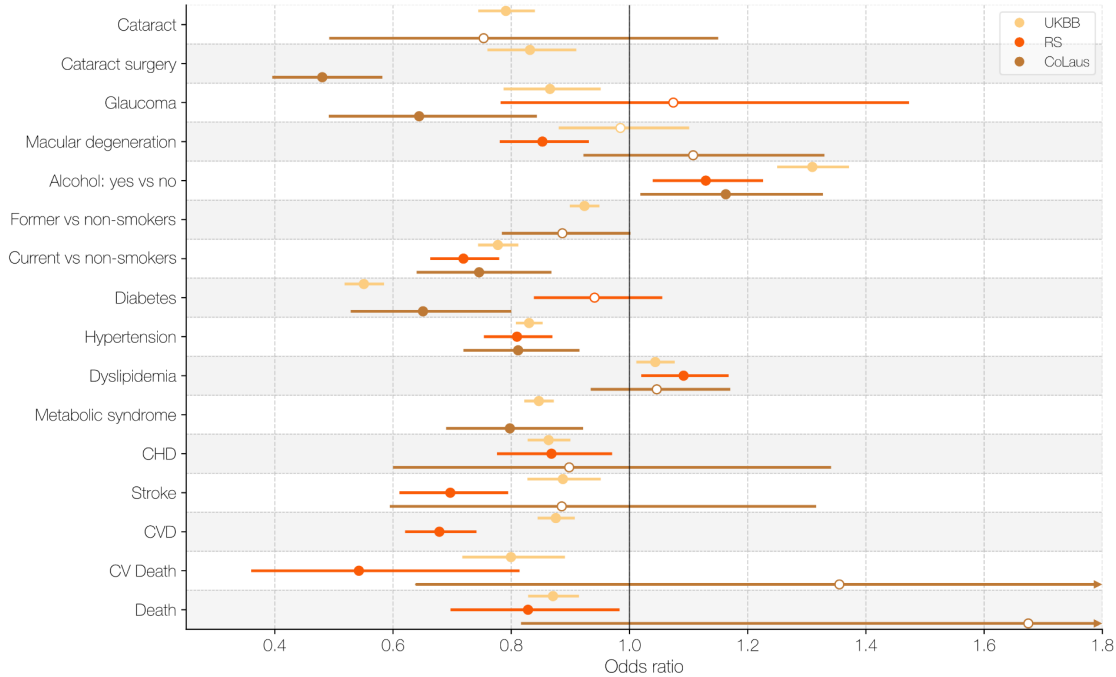


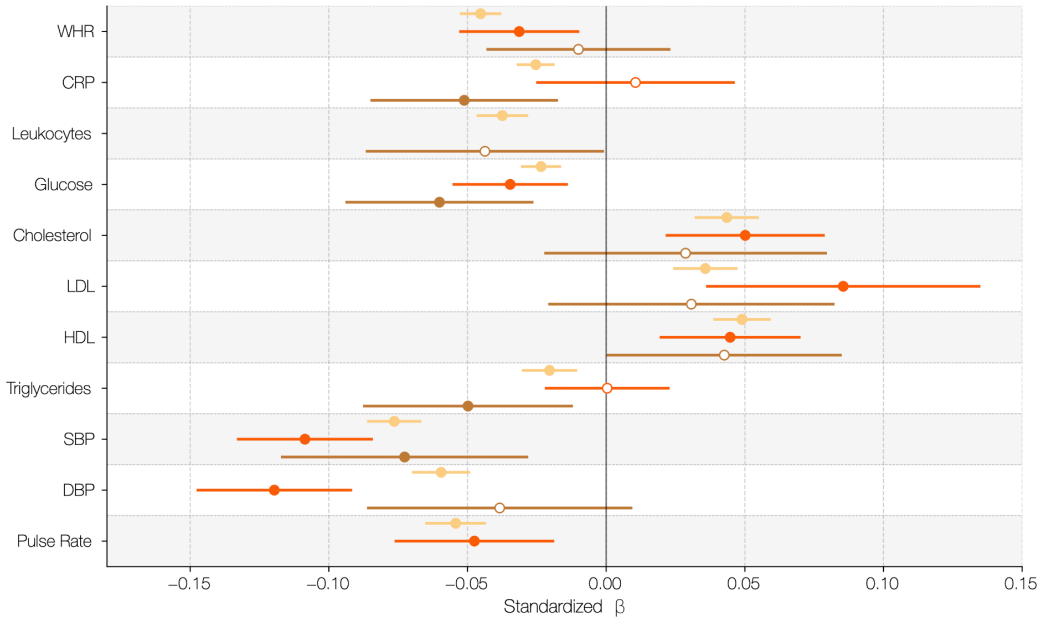
Figure 2 | Association of RetFit and SETCRF with diseases and risk factors in the UKBB dataset. **a)** The left panel shows standardized coefficients for RetFit and SETCRF from independent models predicting binary outcomes in the UKBB. **b)** The right panel shows results from likelihood ratio tests (LRTs), comparing models that included both RetFit and SETCRF, with models including only one of the two CRF estimates. **c)** The left panel shows standardized coefficients for RetFit and SETCRF from independent models predicting continuous outcomes. **d)** The right panel shows results from likelihood ratio tests (LRTs), comparing models that included both RetFit and SETCRF with models excluding either CRF estimate. Filled circles indicate significance after Benjamini-Hochberg correction (in both panels). CHD: Coronary Heart Disease; MI: Myocardial Infarction; CVD: Cardiovascular Disease; CV Death: Cardiovascular Death; WHR: waist-to-hip ratio; CRP: C-reactive protein; IGF-1: Insulin-like growth factor 1; HbA1c: Glycated hemoglobin; LDL & HDL: low- and high-density lipoprotein; SBP & DBP: systolic and diastolic blood pressure.

190 To further validate RetFit's ability to inform upon disease and investigate its performance in
191 populations without CRF testing available, we performed regression analyses using RetFit
192 estimates derived from the CFIs in the Rotterdam Study (RS) and the CoLaus study (N=8 885
193 and N = 2 260, respectively) using outcomes analogous to those in the initial analysis shown
194 in **Figure 2a & 2c** (**Fig. 3a & 3b**). Adjusting for sex, age, and BMI, we found that, for the binary
195 disease outcomes, the external validation confirmed most of the significant results obtained in
196 the UKBB in either the RS or in the CoLaus (10/12 and 7/15 significant results reproduced,
197 respectively), with cataract disease and former smokers failing to reach significance in the
198 external validation cohorts (**Fig. 3a**). Moreover, we identified a novel association in the RS
199 between RetFit and macular degeneration ($OR_{macular\ deg.}$=0.853, 95% CI 0.783 to 0.929,
200 p<0.001) that did not reach statistical significance in the UKBB. Regarding the continuous risk
201 factors (**Fig. 3b**), the effect sizes from the external validation reproduced those we obtained
202 in the UKBB in at least one of the two external and independent cohorts (8/10 and 4/10
203 significant results reproduced for the RS and the CoLaus respectively), except for leukocytes,
204 whose association with RetFit was trending towards being significant in the CoLaus (β=-0.044,
205 95%CI -0.086 to -0.001, p=0.077).

206 **a.**



207
208 **b.**



209

210 **Figure 3 | External validation of RetFit for the disease/risk factors association analyses in the**
211 **Rotterdam Study and the CoLaus cohort. a)** Standardized coefficients for RetFit from independent
212 models predicting binary outcomes in three independent cohorts (UKBB, RS, and CoLaus). **b)**
213 Standardized coefficients for RetFit and SETCRF from independent models predicting continuous
214 outcomes. All models were corrected for sex, age, and BMI. Filled circles indicate significance after
215 Benjamini-Hochberg correction (in both panels). UKBB: UK Biobank; RS: Rotterdam Study; CoLaus:
216 CoLaus Study; CHD: Coronary Heart Disease; CVD: Cardiovascular Disease; CV Death:
217 Cardiovascular Death; WHR: waist-to-hip ratio; CRP: C-reactive protein; LDL & HDL: low- and high-
218 density lipoprotein; SBP & DBP: systolic and diastolic blood pressure.

RetFit is an Independent Biomarker for Cardiovascular Diseases and Overall Mortality

To investigate RetFit's prognostic power, we evaluated its ability to predict future cardiovascular events and all-cause mortality using Cox Proportional Hazard (PH) models and benchmarked the results to SETCRF in the UKBB. We fitted unadjusted and adjusted models (see Methods) to determine how RetFit and SETCRF predict disease independently and jointly with common risk factors. Moreover, we used LRT to determine whether RetFit or SETCRF was better at improving the predictive capacity of models. Finally, we replicated the analysis in the Rotterdam Study cohort.

In the UKBB, both RetFit and SETCRF were significantly prognostic for cardiovascular events before and after correcting for cardiovascular risk factors. However, when we added to the model known cardiovascular disease risk factors alongside sex and BMI, only RetFit remained significantly prognostic (RetFit: HR=0.936, 95%CI 0.903 to 0.970, $p<0.001$; SETCRF: HR=0.979, 95%CI 0.958 to 1.00, $p=0.0653$; see **Fig. 4a** and **Suppl. Fig. 2a** for the HR of CRF estimates expressed in METs). Interestingly, we found that adding SETCRF to a model with RetFit showed no significant improvement (LRT, 1 degree of freedom (df) $\chi^2=2.68$, $p=0.102$) while adding RetFit to the model with SETCRF did (LRT, 1 df $\chi^2=12.2$, $p<0.001$; see **Suppl. Table 5a**), suggesting that RetFit adds distinct cardiovascular disease information that SETCRF fails to capture. Kaplan-Meier curves demonstrated good stratification of increasing cardiovascular event risk with decreasing fitness score (both RetFit and SETCRF; **Fig. 4b** and **Suppl. Fig. 2b**) and, in the fully adjusted Cox model, the third, fourth, and fifth RetFit quintiles were significantly associated with fewer cardiovascular events, while for SETCRF, none were (**Fig. 4c** and **Suppl. Fig. 2c**). When both biomarkers were included in the fully adjusted model, only RetFit quintiles remained significantly associated with cardiovascular endpoints. Adding RetFit quintiles to a model with SETCRF quintiles improved the model (LRT, 4 df $\chi^2=13.4$, $p<0.01$) while the opposite was not true (LRT, 4 df $\chi^2=4.78$, $p=0.311$; see **Suppl. Table 6a**), suggesting, once again, that RetFit captured prognostic information not accounted for by

SETCRF. RetFit's prognostic power for cardiovascular events was corroborated in the Rotterdam Study (RS) cohort, where we found it was significantly associated with such an endpoint before (HR=0.668, 95%CI 0.617 to 0.723, p<0.001) and after adjusting for all the risk factors (HR=0.813, 95%CI 0.733 to 0.901, p<0.001). RetFit's generalisation capabilities were further confirmed in the RS group analysis, where KM curves showed good stratification (**Fig. 4d**; **Suppl. Table 9** and **Suppl. Fig. 3a**) and, in the Cox analysis, all of the quintiles but the second were significantly associated with cardiovascular events before (HR as low as 0.306, 95%CI 0.221 to 0.424, p<0.001; see **Suppl. Fig. 3b–c**) and after adjusting for all the risk factors (HR as low as 0.608, 95%CI 0.420 to 0.880, p=0.00837; see **Fig. 4d**).

When using overall mortality as the outcome in a continuous univariate setting, both RetFit and SETCRF were again significantly prognostic before and after correcting for cardiovascular risk factors (**Fig. 4a**; see **Suppl. Fig. 2a** for the HR of CRF estimates expressed in METs). When the estimates were included alongside all covariates, both significantly improved model fit (see **Suppl. Table 5b**). In the Kaplan-Meier analysis, we observed good stratification power (**Fig. 4b** and **Suppl. Fig. 2b**), while in the Cox analysis RetFit's performance degraded after adjusting for demographics and the other cardiovascular risk factors (**Fig. 4c and Suppl. Fig. 2c**). Adding RetFit quintiles to a (fully adjusted) model with SETCRF quintiles significantly improved the model (LRT, 4 degrees of freedom χ^2=19.7, p<0.001), and so did adding SETCRF quintiles to a model with RetFit quintiles (LRT, 4 degrees of freedom χ^2=62.2, p<0.001) (see **Suppl. Table 6b**). Surprisingly, the second RetFit quintile showed an increased risk compared to the baseline, suggesting confounding effects might be at play. Additional details on the models' fits are available in **Suppl. Tables 7** and **8**.

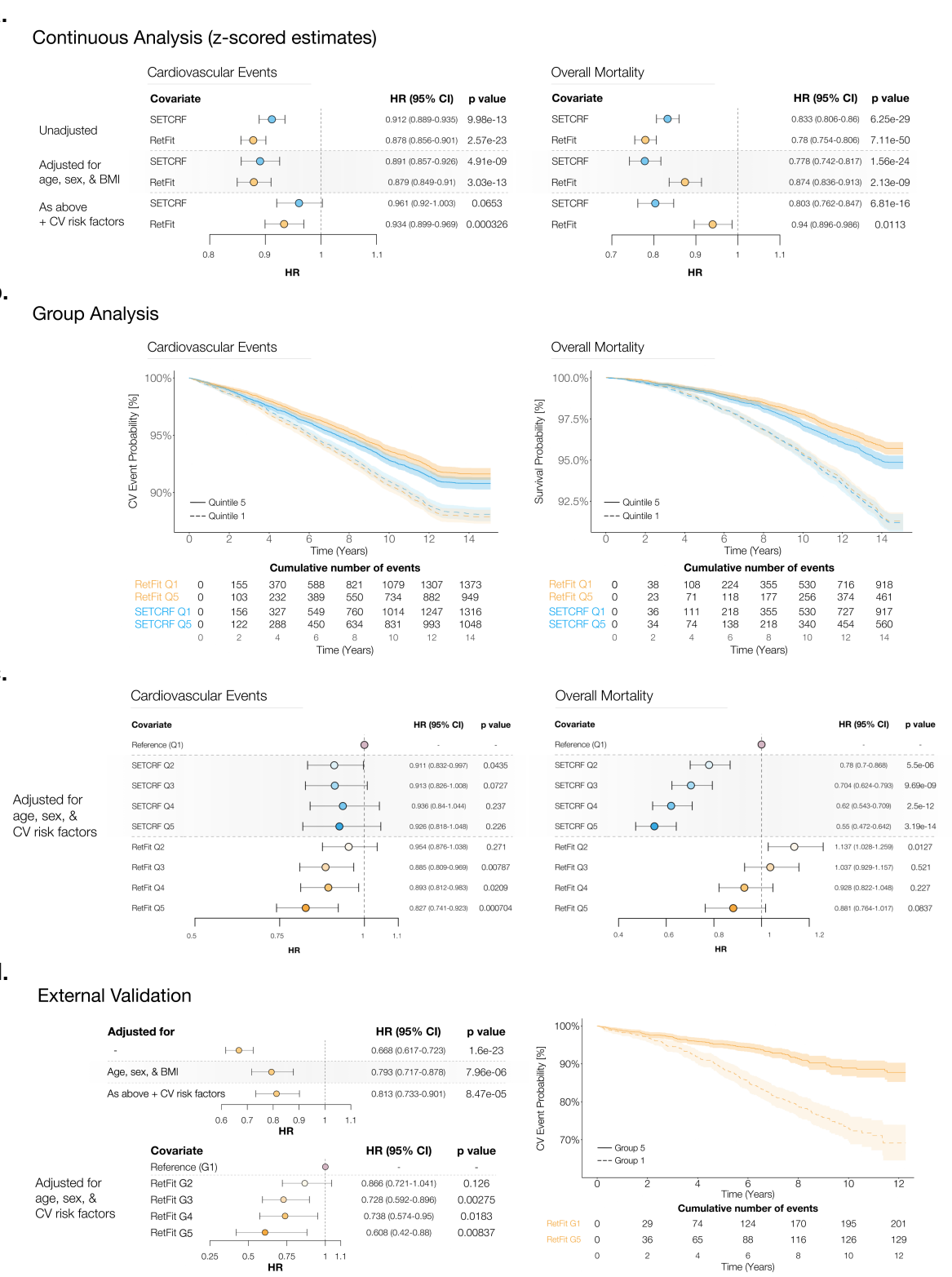


Figure 4 | Prognostic performance and external validation of RetFit. a) Forest plots of the CRF estimates as a continuous parameter for cardiovascular events (left) and overall mortality (right) in the UKBB. **b)** Kaplan-Meier survival analysis of the RetFit and SETCRF (quintiles 1 and 5), and **c)** forest plots for all the quintiles in the UKBB. **d)** External validation of RetFit in the Rotterdam Study cohort. On the left side, from top to bottom: continuous analysis (z-scored) and group analysis. On the right side, Kaplan-Meier survival analysis (groups 1 and 5). For the group analysis, all the group thresholds were computed independently on the UKBB. UKBB: UK Biobank; HR: hazard ratio; CI: confidence interval; CV: cardiovascular.

RetFit is Genetically Distinct from SETCRF

To understand whether RetFit reveals new genes that may be important for CRF, we performed Genome-Wide Association Studies (GWAS) on both RetFit and SETCRF (see **Suppl. Fig. 4a–b** for genome-wide SNPs Manhattan plots). We found that RetFit was associated with 41 genes and SETCRF was associated with 25 genes. No significant genes were shared between RetFit and SETCRFs (**Fig. 5;** see **Suppl. Tables 11a** and **11b** for full list of genes). We also performed pathway analyses using the gene scores and found that RetFit was associated with 39 pathways and SETCRF was associated with 158 pathways. The significant pathways found for RetFit were mostly related to abnormalities of the eye, cell development mechanisms, and the nervous system. The 158 significant pathways we found for SETCRF were mostly associated with the cardiovascular system (see **Tables 1a** and **1b** for the 10 most significant pathways, **Suppl. Tables 12a–b** for their description, and **Suppl. Tables 12c–d** for the full list of significant pathways). Heritability (h^2) estimates were similar for both traits (SETCRF h^2=21.1%, SE=1%; RetFit h^2=22.82%, SE=1%), and the genetic correlation of the two estimates was low (0.09, SE=0.03, p=0.008).

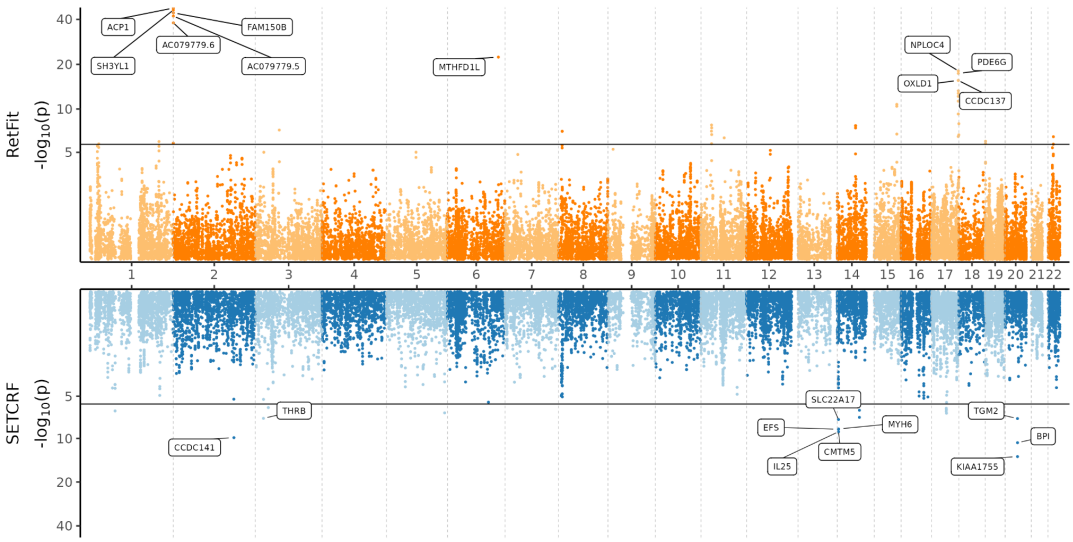


Figure 5 | Gene enrichment analysis results for RetFit and SETCRF in the UKBB dataset. Miami plot showing all genes that were scored for the analysis using PascalX. Each dot corresponds to the p-value of one of the 24 922 genes scored. The x-axis shows the position of the gene in the genome, and the y-axis shows the significance of every gene score after a $-\log_{10}$ transformation, plotted on a $\log_2$ scale. The black line represents the significance threshold after Bonferroni correction. The upper panel shows the gene scores for RetFit, and the lower panel shows the gene scores for SETCRF. The 10 most significant genes are labeled for each trait.

Table 1 | Pathway analysis of the two CRF estimates (UKBB dataset). List of 10 most significant pathways for **a)** RetFit and **b)** SETCRF. All the pathways available in MSIgDB v7.2 were scored using PascalX.

a.

Pathway name	p-value
HP_ABNORMALITY_OF_VISION	2.56e-07
LBP1_Q6	4.51e-07
HP_ABNORMAL_CONJUGATE_EYE_MOVEMENT	3.73e-06
HP_CYSTOID_MACULAR_EDEMA	5.25e-06
GO_REGULATION_OF_NEURON_DIFFERENTIATION	5.48e-06
GO_REGULATION_OF_CELL_MORPHOGENESIS_INVOLVED_IN_DIFFERENTIATION	5.64e-06
GO_REGULATION_OF_CELL_DEVELOPMENT	7.16e-06
HP_ABNORMALITY_OF_THE_OPTIC_DISC	1.21e-05
ZFHX3_TARGET_GENES	1.25e-05
ZNF596_TARGET_GENES	1.30e-05

b.

Pathway name	p-value
HP_CARDIAC_CONDUCTION_ABNORMALITY	7.55e-12
HP_HEART_BLOCK	2.50e-11
GO_HEART_PROCESS	4.48e-11
HP_SUPRAVENTRICULAR_ARRHYTHMIA	2.06e-10
HP_ATRIAL_ARRHYTHMIA	5.18e-10
GO_CARDIAC_MUSCLE_CONTRACTION	1.18e-09
GO_REGULATION_OF_STRIATED_MUSCLE_CONTRACTION	1.25e-09
GO_REGULATION_OF_BLOOD_CIRCULATION	1.47e-09
GO_CIRCULATORY_SYSTEM_PROCESS	2.08e-09
GO_STRIATED_MUSCLE_CONTRACTION	4.33e-09

Retinal Vasculature Associations with CRF

To better understand how RetFit might be predicting CRF from CFIs, we generated attention maps for RetFit to identify which image features are driving the CRF estimation. These attention maps qualitatively indicated that the model focused on the areas of the CFIs where the superior vascular beds typically lie (**Fig. 6a**). Furthermore, we used a state-of-the-art pipeline to extract quantitative vascular features from CFIs (hereafter referred to as Tangible Image Features, or TIFs) to assess their associations with both RetFit and SETCRF[28]. When uncorrected for covariates, all TIFs but three were associated with either RetFit or SETCRF (the Ratio tortuosity, the A diameter std, and the A median diameter for SETCRF and the A diameter std, the ratio median diameter and the ratio central retinal equivalent for RetFit; **Suppl. Fig. 5**). The number of arterial (A) bifurcations demonstrated the strongest correlation (Pearson *r* Coefficient) with both RetFit ($r=0.259$, 95%CI 0.251 to 0.267, $p<0.001$) and SETCRF ($r=0.155$, 95%CI 0.147 to 0.163, $p<0.001$). After accounting for age, sex, and BMI, 11 of the 18 TIFs remained associated with SETCRF, whereas 13 of the 18 TIFs were still associated with RetFit. (**Fig. 6b**). The number of A bifurcations remained the vascular property showing the strongest association with RetFit ($\beta=0.287$, 95%CI 0.263 to 0.311, $p<0.001$), but the vascular density ratio (arteries/veins) showed the strongest association with SETCRF ($\beta=0.391$, 95%CI 0.352 to 0.43, $p<0.001$) (**Fig. 6b**).

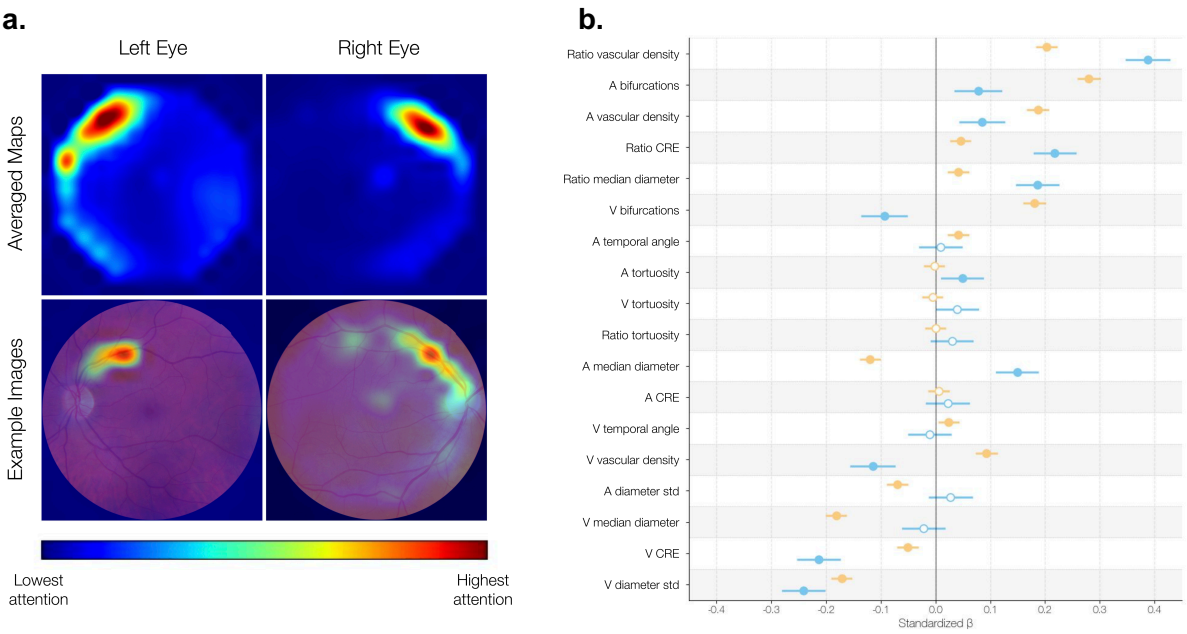


Figure 6 | Attention mapping of the RetFit vision transformer model and association of RetFit and SETCRF with retinal vasculature features in the UKBB. a. The top two images show the average attention maps across all images from the UKBB for the left and right eyes, respectively. The bottom two images present examples of attention maps overlaid on CFIs for the left and right eye, providing a clear illustration of the specific regions the model focuses on when making predictions. **b.** Forest plot showing associations between retinal vascular features extracted from CFIs using VascX models[28] and RetFit and SETCRF corrected for age, sex, and BMI. A: arteries; V: veins; CRE: central retinal equivalent; std: standard deviation.

Discussion

Cardiorespiratory fitness (CRF) is among the most robust predictors of disease outcomes[1–6], yet it remains challenging to measure accurately at scale[9–12]. In this study, we aimed to assess whether SETCRF (a submaximal exercise test to estimate CRF) could be reliably estimated from CFIs and to determine whether this estimator (that we call *RetFit)* correlates with current disease diagnoses and can predict future cardiovascular events. We found that, while RetFit only captures a moderate fraction of SETCRF's variability (R^2=0.2), it nevertheless provided diagnostic and prognostic information regarding disease, equaling the performance of direct SETCRF estimates reported in the literature[27] and, in some cases, surpassing them. Interestingly, our GWAS conducted on both SETCRF and RetFit revealed no overlapping genetic associations, indicating distinct genetic underpinnings for these two fitness estimates. These findings highlight both the potential and limitations of retinal imaging as a surrogate

352 measure for CRF, underscoring the need for continued exploration into the biological
353 mechanisms linking retinal vascular characteristics with systemic health.

354 The prognostic power of objectively measured VO_2max in predicting future cardiovascular
355 events is extensively documented across large-scale cohort studies[1–3,8,13]. A strong, inverse,
356 and graded association exists between an individual's VO_2max and the incidence of coronary
357 artery disease, myocardial infarction, stroke, and cardiovascular mortality in general[3,29].
358 Previous work has demonstrated that CRF remains a significant predictor of cardiovascular
359 events after adjustment for established multivariable risk scores, such as the Framingham
360 Risk Score, or major cardiovascular risk factors (such as sex, age, and blood pressure)[1,30,31].
361 In our analysis, we made a similar observation for RetFit in the UKBB (i.e., it remained
362 significant after adjustment for age, sex, BMI, SBP, HDL cholesterol, total cholesterol, smoking
363 status, diabetes diagnosis, hypertension treatment, and previous cardiovascular events), but,
364 interestingly, not for SETCRF. This finding may reflect fundamental differences between the
365 two estimates: whereas SETCRF is derived from exercise performance, which is closely linked
366 to established cardiovascular risk factors, RetFit leverages retinal image features that may
367 quantify additional, subclinical markers of vascular health not captured by conventional
368 measurements. Consequently, RetFit could complement existing risk scores, offering
369 additional diagnostic and prognostic value to routinely acquired clinical data. It is worth noting
370 that, in the study proposing SETCRF, Gonzales et al. (2021) reported that the estimate
371 remained a significant predictor of cardiovascular events after adjustment for several risk
372 factors[24]. In our analysis, significance was not reached (p=0.065) after adjusting for a slightly
373 different set of covariates (including systolic blood pressure, HDL, and total cholesterol, and
374 excluding several covariates from Gonzales et al.) and using a different definition of prior
375 cardiovascular events. Differences in sample size (57 588 vs. 79 981, as we selected the
376 subset of participants with retinal images, see **Suppl. Fig. 1**) and grouping strategy (quintiles
377 vs. 1-MET increments) likely contributed to this discrepancy.

When we looked at disease associations, RetFit consistently showed a comparable diagnostic power to SETCRF across a broad range of analyses, with standardized beta coefficients and odds ratios often differing in magnitude but pointing in the same direction, underscoring its clinical relevance. We only noticed a major difference between the association of the two traits with a few traits, including cognitive performance metrics such as reasoning or prospective memory. In fact, the latter appear to be better captured by RetFit than SETCRF, pointing to potential systemic relevance beyond cardiovascular outcomes. Interestingly, the standardized effect size for LDL and total cholesterol differed in sign between RetFit and SETCRF, potentially reflecting the *cholesterol paradox,* where some studies have reported higher survival among older individuals with high levels of LDL and total cholesterol, possibly due to reverse causation or U-shaped associations[32–34]. As we previously discussed, RetFit showed superior prognostic power, independent of cardiovascular risk factors (e.g., HDL, total cholesterol, and SBP) for cardiovascular events. This was consistent with the results of the association analysis, where RetFit, after adjusting for age, sex, BMI, and socioeconomic status, showed a weaker correlation with SBP, DBP, HDL, LDL, total cholesterol, and triglycerides compared to SETCRF, indicating that RetFit is less tied to these factors. Beyond cardiovascular outcomes, RetFit also predicted overall mortality independently of conventional risk factors, underlining its broad clinical significance. We confirmed RetFit's diagnostic and prognostic capabilities in two external and independent cohorts, namely, the Rotterdam Study and the CoLaus study. Besides validating the generalizability of our CFI-based fitness estimate beyond the UK Biobank population, this is particularly significant because it confirms RetFit's potential as a highly scalable biomarker for CRF estimation in populations where maximal oxygen consumption or even submaximal exercise testing data are unavailable. Indeed, many large-scale population and clinical cohorts lack resource-intensive CRF measurements, but include CFI, in which case RetFit provides a novel, externally validated, and non-invasive method to retrospectively and prospectively estimate a robust, disease-relevant CRF-related measure, unlocking new opportunities for large-scale cardiovascular risk stratification and investigation of the molecular underpinnings of systemic health.

406 From the genetic analysis, we found that both RetFit and SETCRF yielded significant genetic
407 associations; however, the genes and pathways identified were distinct, with no overlap
408 observed between the two traits. The most significant hit we identified for RetFit, *ACP1*
409 (p=5.84e-48), is a protein-encoding gene regulating cellular signaling by removing phosphate
410 groups from tyrosine residues on proteins, known to be involved in coronary artery disease[35]
411 (see **Supp. Tables 10a** and **10b**). For SETCRF, our findings indicate *MYH6* (p=2.63e–09) as
412 one of the strongest hits; this gene encodes the alpha-myosin heavy chain, the most abundant
413 protein in the heart and a central determinant of cardiac muscle contraction, with established
414 associations to cardiomyopathy[36]. Furthermore, we observed significant but distinct gene hits
415 related to mitochondrial processes for both traits. In the case of RetFit, our analysis revealed
416 a significant association with *MTHFD1L* (p=5.6e-23), a protein-encoding gene involved in the
417 synthesis of tetrahydrofolate (THF) in the mitochondrion[37]. For SETCRF, we found a
418 significant association with several genes such as *IL25* (p=1.05e-09), a cytokine promoting
419 mitochondrial respiration[38], *TGM2* (p=6.04e-08), an enzyme regulating mitophagy[39], or *THRB*
420 (p=5.91e-08), a thyroid hormone receptor, known to be involved with the mitochondria[40].
421 Furthermore, we identified several significant pathways linking SETCRF with CVD when
422 running the pathway analysis (e.g., abnormality of cardiac conduction or impaired conduction
423 of cardiac impulse, see **Supp. Table 11b** for pathway description), but not for RetFit, where
424 pathways related to the cardiovascular system failed to reach significance. However, as we
425 showed in the survival analysis, RetFit demonstrated better prognostic power for CV events
426 than SETCRF, highlighting the need to better understand the differentiation between
427 diagnostic and prognostic leverage and genetic mechanisms. In contrast, we found that the
428 disease and genetic analyses seem to converge on endpoints associated with neuronal
429 development, specifically, reasoning, visual memory, reaction time, and prospective memory.
430 RetFit showed significant associations with each of these risk factors (whereas SETCRF did
431 not), and multiple enriched pathways suggest that its genetic architecture influences
432 processes related to the nervous system, neuronal development, and cerebral function (e.g.,
433 regulation of neuron differentiation, see **Supp. Table 11a**). Consequently, the associations we

434 observed with cognitive markers are consistent with the genetic evidence, supporting the
435 hypothesis that RetFit is functionally related to neuronal processes.

436 The physiological limit of CRF is fundamentally determined by the cardiovascular system's
437 ability to transport oxygen to metabolically active tissues[47]. While cardiac output and arterial
438 oxygen content are critical components, CRF may also hinge on the integrity and function of
439 the systemic vasculature, where oxygen exchange occurs[48]. The retinal circulation provides a
440 unique, non-invasive window to observe this vasculature network in vivo, and morphological
441 changes within it are well-established markers for systemic vascular pathology[49]. For example,
442 narrowing of retinal arterioles and widening of venules, often summarized as a reduced
443 arteriolar-to-venular ratio (AVR), are robustly associated with hypertension and the long-term
444 risk of coronary heart disease[50]. Beyond vessel geometry, the branching complexity of the
445 vascular tree, quantified by its fractal dimension, reflects the efficiency of tissue perfusion; a
446 sparser, less complex network (lower fractal dimension) is linked to vascular rarefaction and
447 has been shown to predict adverse outcomes like stroke[51]. Because these retinal features
448 capture systemic vascular health, they may also reflect physiological pathways that underlie
449 CRF. In line with this observation, most TIFs showed significant associations with both RetFit
450 and SETCRF, with some of them displaying stronger associations with SETCRF than RetFit,
451 despite the latter being based on CFIs.

452 It is critical to interpret these findings in the context of the study cohorts. We developed RetFit
453 on a subset of the UKBB, which consists largely of individuals of White British ancestry, aged
454 from 40 to 70. As a volunteer-based cohort, the UKBB is subject to healthy-volunteer and
455 socio-economic selection biases: participants are typically healthier, less socioeconomically
456 deprived, and report fewer health conditions than the general population[52]. However, the
457 UKBB remains the largest and one of the only cohorts for which both CFIs and CRF estimates
458 are available. In this regard, it is important to acknowledge that RetFit was modeled after
459 SETCRF, which is itself a CRF estimate[24]. Notably, VO_2max was not objectively measured in
460 any of the UKBB participants, and the validation cohort demonstrated higher exercise capacity

461 than the average UKBB participant[24]. Such factors likely affect the accuracy of CRF estimation
462 and, consequently, may also influence our analyses. It is also important to note that, since
463 RetFit is derived from a deep learning (DL) pipeline, its distribution may exhibit a reduced
464 standard deviation compared to other CRF estimates (due to the tendency of DL models to
465 produce outputs that regress toward the mean, thereby compressing the range of predicted
466 values[53]), and therefore yield inflated HRs when the latter are reported in units of MET. Given
467 this, we opted to report the HR in units of standard deviation rather than METs whenever
468 comparing the predictive power of RetFit and SETCRF in a continuous setting. We
469 nevertheless report the HR in units of MET in the supplementary material to allow for a direct
470 comparison with other DL and non-DL derived markers from the literature. Finally, the external
471 validation cohorts (i.e., the Rotterdam Study and the CoLaus study) are also largely composed
472 of individuals of White European ancestry.

473 In summary, in this work, we demonstrated that DL can generate a powerful and scalable
474 biomarker from retinal images that reflects an individual's health and disease risk. While
475 conceived as a surrogate for CRF, our findings reveal a more complex and compelling story:
476 RetFit's superior prognostic power for future cardiovascular diseases, its comparable
477 diagnostic capacity regarding disease outcomes and risk factors, combined with its distinct
478 genetic architecture from SETCRF, establishes it not as a mere proxy but as a novel biological
479 indicator. This reinforces the idea that the retinal microvasculature provides a unique window
480 into a dimension of systemic health that is related to, but not fully captured by, traditional
481 physiological fitness. As a purely image-derived estimate of CRF, RetFit offers a major
482 practical advantage as it is much less invasive than classical cardiorespiratory fitness tests
483 (i.e., VO_2max test), which require expensive equipment, trained personnel, and maximal
484 physical effort. Thus, RetFit can offer a more scalable alternative, in particular for frail and
485 elderly populations, for whom CFIs often exist already or can be readily acquired, making it a
486 promising tool for widespread clinical and population health use.

Methods

Data and quality control

The UK Biobank (UKBB) is a population-based cohort comprising approximately 488 000 participants, featuring extensive longitudinal phenotypic data, including comprehensive medical histories, with a median follow-up period of 10 years[54]. Cardiorespiratory fitness (CRF) was defined as maximal oxygen consumption or VO_2max and estimated from heart rate responses obtained during submaximal exercise testing[24]. Standard 45° retinal color fundus images (CFIs) were acquired using a Topcon 3D-OCT 1000 Mark II imaging system.

For the present analysis, CRF estimates for 82 850 individuals measured as stated in[24] were used. Additionally, a total of 173 814 CFIs from 84 813 individuals were accessible for analysis. Given our objective of evaluating the predictive capability of CFIs for estimating CRF, we implemented a previously published quality control (QC) procedure to ensure image integrity[55]. Specifically, image quality was initially quantified by professional graders on a subset of 1 000 CFIs, after which a convolutional neural network (CNN) was trained to replicate this human quality assessment. The top 75% highest-quality images, as determined by the CNN, were included in subsequent analyses. Since CRF and CFIs could be acquired at either instance 0 or instance 1, we matched each CFI with the corresponding CRF measurement from the same instance, prioritizing instance 0 and using instance 1 only if instance 0 data were unavailable. In total, we had matching CFI and CRF for 57 588 individuals (see **Suppl. Fig. 1**).

Predicting CRF with RETFound

We utilized RETFound[22], a novel foundation model for CFIs that leverages Google's vision transformer large patch 16 architecture[56]. For predicting CRF, we modified some of the base hyperparameters established in the original RETFound manuscript[22] and trained the model over 50 epochs, with a "warm-up" in the 10 first epochs (with the learning rate *r* monotonically

511 increasing from 0 to 5×10^{-4}), followed by a cosine annealing schedule (where *r* decreases

512 from 5×10^{-4} to 1×10^{-6}). We set the batch size to 16, the dropout rate to 0.4, and the image

513 size to 224×224 pixels. After each epoch, we evaluated the model on a validation dataset and
514 then kept the model that had the lowest mean square error across all epochs. To ensure robust
515 evaluation, we implemented a 5-fold cross-validation strategy with a leave-one-out design: for
516 each fold, the model was trained on 80% of the dataset and evaluated on the remaining left-
517 out split. All the results we report in the manuscript were obtained by aggregating the
518 predictions across the five left-out partitions. In addition, we fine-tuned separate models for
519 each laterality, yielding five cross-validated models for the right eye and five for the left eye.
520 To obtain the final subject- and instance-level prediction, we averaged the outputs from the
521 left and right eyes when both images passed quality control, or used the single available
522 prediction otherwise. Pearson correlations, R^2, MAE, and RMSE were computed on the CRF
523 averaged over both eyes.

524 Disease Association

525 For the disease association analyses, we included a set of disease and general systemic risk
526 factors that had either a binary or a continuous outcome known to affect disease susceptibility.
527 We collected individual-level disease and risk factor information from the UKBB. We prioritized
528 information provided through health-related records (Category 100091), biological samples
529 (Category 100078), and Physical measures (Category 100006) over those obtained through
530 the assessment center Touchscreen questionnaire (Category 100025) or verbal interviews
531 (Category 100071), as self-reported measures are often unreliable due to high reporting errors
532 or inconsistencies[57,58]. We made exceptions for alcohol intake and smoking status (i.e., never,
533 previous, current). As studies have shown the prevalence of undiagnosed diabetic or pre-
534 diabetic conditions and associated health complications[59,60], we determined diabetes status
535 using a combination of variables: an official diagnosis based on the date E10 first reported
536 (insulin-dependent diabetes mellitus), HbA1C readings of larger or equal to 48 mmol/mol, or

537 glucose levels of larger or equal to 11.1 mmol/l[61,62]. For participants who took antihypertensive
538 medication, we added 10 mmHg to SBP and 5 mmHg to DBP values, and for those who took
539 cholesterol-lowering medication, we added 1.4 mmol/L to LDL, 0.4 mmol/L to triglycerides, 1.6
540 mmol/L to total cholesterol, and subtracted 0.1 mmol/L from HDL. Participants who had total
541 cholesterol levels higher than 5, HDL lower than 1, LDL higher than 4.1, triglycerides higher
542 than 2.2, or took cholesterol-lowering medication were considered to have dyslipidemia. For
543 blood pressure and pulse rate, we computed the average between (up to) the two automatic
544 measurements per subject (taken one after the other, on the same visit). We defined
545 hypertension as having SBP > 140 or DBP > 90 or the use of blood pressure medication.
546 Following the International Diabetes Federation consensus[63], we defined metabolic syndrome
547 as having three or more of these phenotypes: large waist circumference (>= 80 cm for female
548 or >= 94 cm for males), triglycerides > 1.7 or cholesterol medication, low HDL (< 1.29 for
549 females or < 1.03 for males), high BP (SBP >= 130 and/or DBP >=85 and/or BP medication),
550 and high glucose which we substituted with the above definition of diabetes since fasting
551 glucose was not available. We defined cardiovascular (CV) death by reviewing all ICD-10-
552 coded[64] causes of death and including those corresponding to cardiovascular conditions[65,66].
553 All the fields used for the analysis and their descriptions can be found in **Suppl. Tables 3a–c**.

554 We used multiple linear regressions for continuous risk factors and multivariate logistic
555 regressions for binary ones. All models were corrected for sex, age, BMI, income, educational
556 level, and the Townsend deprivation index (the latter three forming the socioeconomic status).
557 Furthermore, both the response and the independent variables that were used for the linear
558 regressions were standardized. We also binarized all diseases collected as “date first
559 diagnosed” for the logistic regressions. To determine the significance of regression
560 coefficients, we computed p-values and corrected for multiple hypothesis testing using the
561 Benjamini-Hochberg method[67], using a false discovery rate of 0.05. We also used the log-
562 likelihood ratio test (LRT) to compare the addition of independent information for RetFit and

563 SETCRF by fitting a nested model (with either RetFit or SETCRF and the rest of the
564 covariates) and a model with both predictors with the same set of covariates.

565 Survival Analysis

566 We defined two distinct clinical endpoints for the survival analysis: cardiovascular events
567 (obtained by aggregating Ischaemic Heart Disease (IHD), Myocardial Infarction (MI), stroke
568 events, Non-Ischaemic Cardiomyopathies (NIC), and thrombotic events), and overall
569 mortality. Time to event (TTE) was defined as the days between the date when the retinal
570 image was taken and the date of the reported event. If the subject experienced multiple events,
571 the first event after the imaging visit was used. We fitted Cox Proportional Hazard (PH) models
572 with three levels of covariate adjustment to assess the robustness of the CRF estimates. First,
573 we conducted univariate analyses including only the estimate of interest. Second, we adjusted
574 for age, sex, and BMI, as these are key determinants of CRF. Finally, we further adjusted for
575 additional cardiovascular risk factors that broadly matched those in the Framingham score[68],
576 i.e., age, sex, SBP, Total and HDL cholesterol, BMI, smoking status, diabetes diagnosis, and
577 hypertension treatment (i.e., medication), plus whether the subject had a previous
578 cardiovascular event. Hypertension treatment was defined by aggregating male- and female-
579 specific medication and filtering for "blood pressure medication". Diagnosis of previous
580 cardiovascular events was included due to an increased risk of recurrence[69]. All the fields used
581 for the analysis and their descriptions can be found in **Suppl. Tables 1–3**, while the distribution
582 of events and continuous risk factors can be found in **Suppl. Table 4a.** For both estimates
583 and for each clinical endpoint, we fit two independent models (i.e., one for RetFit and one for
584 SETCRF) using continuous measures as predictors. Additionally, we conducted a group
585 analysis by stratifying participants into quintiles based on their CRF estimates, reflecting a
586 common practice in the literature where CRF estimates are often categorized into quantiles or
587 fitness groups to assess dose-response relationships and facilitate clinical interpretation[11,6,70].
588 We applied the same stratification method to the RetFit score to compare its ability to stratify
589 risk with that of the estimated CRF. To compare the different CRF estimates and investigate

590 whether they improved the fit and the prognostic power of the different Cox models, we used
591 the LRT to compare nested models (i.e., to compare a model with a fixed set of covariates
592 and one of the estimates with a model with the same covariates and both estimates) (see
593 **Suppl. Table 5–8**). To maximise the sample size while avoiding sparse data at longer follow-
594 up times, we censored subjects at 15 years for both cardiovascular events and deaths.

595 Genome-wide association study (GWAS)

596 GWAS were conducted using REGENIE[71]. Before performing the analyses, the genotypic data
597 were processed following UKBB guidelines. Variants failing quality control thresholds (MAF <
598 0.01, MAC < 100, genotype missingness > 0.1, or Hardy-Weinberg equilibrium $p < 1\times10^{-15}$)
599 were excluded. The GWAS were corrected for a set of covariates known to influence
600 phenotypic variability (cardiovascular or eye-specific factors). The covariates were age at
601 assessment center visit, sex, BMI, age-squared, sex-by-age, sex-by-age-squared, spherical
602 power, cylindrical power, spherical power-squared, cylindrical power-squared, imaging
603 instance, assessment center, genotype measurement batch, and the first 20 genetic Principal
604 Components (PCs) (see **Suppl. Table 10** for corresponding UKBB data fields).

605 Gene and Pathway Scoring

606 We computed gene and pathway scores using PascalX, a Python library specialized in gene
607 and pathway scoring for GWAS summary statistics[72–74]. PascalX utilizes the sum of χ^2 method
608 to test for gene enrichment for gene scoring and a rank-based method for pathway scoring,
609 which unifies the gene p-values distribution and aggregates it to χ^2 distributed random
610 variables. We used the UK10K panel for gene scoring, and scored both protein-coding genes
611 and lincRNAs (long intergenic non-coding RNA) using the “saddle” method, taking into account
612 all SNPs within a 50kb window around each gene and with a MAF of 0.05. We scored all
613 pathways available in MSigDB v7.2 (n=31 120), fusing and rescoring any co-occurring genes

located less than 100 kb apart. We adjusted all significant gene and pathway scores for multiple testing using the Bonferroni and Benjamini-Hochberg methods, respectively.

Heritability and Genetic Correlation

We estimated SNP-based heritabilities for RetFit and SETCRF using the GCTA (Genome-wide complex trait analysis) software package[75]. We estimated the proportion of variance of the trait explained by the genetic data with the genomic-relatedness-based restricted maximum-likelihood (GREML) method[75,76]. We constructed genetic relationship matrices (GRM) for each chromosome using only directly genotyped SNPs with a MAF > 0.01 to reduce bias due to imputation quality and reduce computational burden. GRMs were then combined and used to estimate SNP-based heritability (h^2) and genetic correlations (GC).

Deriving Attention from RETFound

We generated attention maps from each of the five fine-tuned RETFound models to investigate which regions the model focuses on during prediction, and to gain insights into the representations learned by the model. Attention weights were extracted from the final attention block of the Vision Transformer, specifically capturing the softmax-normalized attention scores from the class token to the patch tokens ("CLS→patch attention")[77], resulting in 14 × 14 maps for each of the 16 attention heads. To account for the varying influence of individual attention heads, we performed a head sensitivity analysis. For each test image and each model fold, we ran 16 modified forward passes, each time muting a different attention head in the final block by zeroing the corresponding attention weights[78]. We used the squared absolute difference between the perturbed and unperturbed predictions as a proxy to quantify the head's contribution. We then used these differences to compute a weighted average of the attention maps across heads, emphasizing the heads that most strongly influenced the prediction. We averaged the resulting head-weighted attention maps across all individuals and

638 folds, separately for left and right eyes, to produce population-level maps. For visualization
639 purposes only, we upsampled such maps from 14 × 14 to the original image resolution using
640 bicubic interpolation.

641 Vascular Associations

642 To explore the potential relationship between the retinal vasculature and both RetFit and CRF,
643 we leveraged recently developed state-of-the-art deep learning models called *VascX* to extract
644 18 vascular traits from CFI[28]. *VascX* models work in a 2-step procedure, the first step being
645 the segmentation of the CFI to separate retinal features (e.g., optic disc, vessels, and, more
646 precisely, artery/vein), and the second step being the feature extraction from the masks that
647 were output during the first step. In this case, we extracted 18 Tangible Image Features (TIFs)
648 derived from the vessels' segmentation masks, and they comprise features such as tortuosity
649 or vessel density (see **Suppl. Table 13** for the full list) and were all separated between arteries
650 and veins. For a more detailed description of the algorithms to compute these features, see
651 Ortin-Vela et al.[19] and Vargas Quiros et al.[19,28]. Once the TIFs were extracted, we used the
652 same method to match them to RetFit/CRF estimates, meaning that values were averaged
653 between eyes if both eyes were used to compute the average for RetFit. We then computed
654 Pearson correlation coefficients and associated tests for each vascular trait independently.
655 Multiple linear regressions with either Retfit or SETCRF as the response variable and each of
656 the 18 TIFs as the dependent variable were also fitted, and all models were corrected for age,
657 sex, and BMI. Multiple hypothesis testing was addressed using Benjamini-Hochberg, and all
658 reported p-values have been corrected accordingly.

659 External Validation

660 To investigate the generalizability of RetFit, we replicated the disease association and survival
661 analyses in two different cohorts, namely the Rotterdam Study (RS) and the CoLaus cohort.
662 We performed quality control on the CFIs in these two studies using the same QC method

663 mentioned under the *Data and quality control* section. We used the model weights fine-tuned
664 on UKBB data to predict RetFit for all the images from the two external cohorts deemed of
665 good enough quality. Given the 5-fold cross-validation strategy we adopted for the
666 development and validation of the model in the UKBB, we obtained 5 independent models for
667 each laterality. Therefore, we predicted five values of RetFitfor each image, which we
668 averaged to obtain the final estimate for each CFI. For the RS, a quality measure was available
669 for every image, so we selected the single image per visit and per subject with the highest
670 quality. For the CoLaus, when subjects had CFIs that passed QC for both lateralities, we
671 further took the mean between the two RetFit values. In the RS, when subjects had images
672 from multiple visits, we favored the instance with the highest proportion of non-missing values
673 for the risk factors and diseases we analysed. In the CoLaus, as CFIs were not acquired at
674 the same time as the risk factors and disease statuses, we selected the instance for the risk
675 factors and disease that was the closest to the date of imaging. After this phase, every subject
676 had only one CFI and one matching RetFit value that was further used for downstream
677 analyses.

678 For the continuous risk factor and binary disease variables in the RS and the CoLaus, we fitted
679 similar linear models (linear regression and logistic regression, respectively) as we did for the
680 UKBB. As neither of these two cohorts measured cardiorespiratory fitness estimates, we
681 focused on the replication of effect sizes and directionality of the associations, and on
682 associations with outcomes that were not available in the UKBB. We defined dyslipidemia and
683 corrected BP, cholesterol (total cholesterol, LDL, and HDL), and triglycerides as described
684 under the *Disease Association* section. All models were corrected for sex, age, and BMI, and
685 both independent variables and continuous response variables were standardized. We once
686 again computed p-values and corrected for multiple hypothesis testing using the Benjamini-
687 Hochberg method and using a false discovery rate of 0.05. The distribution of events and
688 continuous risk factors for both the RS and the CoLaus can be found in **Suppl. Table 4b–c.**

For the external validation of RetFit in a survival analysis setting, we defined the "cardiovascular events" endpoint by aggregating coronary heart disease, heart failure events, and stroke events. The time to event was defined as the number of days between the date the retinal image was taken and the date of the reported event. Similarly to the analysis in the UKBB, we fitted Cox Proportional Hazard (PH) models with three levels of adjustment to assess the generalisability of RetFit independent of several risk factors. In the univariate analysis, we included only RetFit; in the second step of adjustment we included age, sex, and BMI; in the complete model, we further adjusted for additional cardiovascular risk factors that broadly matched those in the Framingham score[68], i.e., age, sex, SBP, Total and HDL cholesterol, BMI, smoking status and hypertension treatment (i.e., medication). Different from the analysis in the UKBB dataset, we were not able to adjust for diabetes diagnosis due to such information being available for a very limited number of subjects. Furthermore, we did not include previous cardiovascular events as a risk factor, as the subjects in the Rotterdam study were cardiovascular-event-free at the time of enrollment. We defined the risk groups for the group analysis using the threshold we independently computed in the UKBB. Similarly to the UKBB, we limited the follow-up for cardiovascular events to 12 years to maximise sample size while avoiding sparse data at longer durations.

Author Information

VascX Consortium: Michael Beyeler [1,2]; Ciara Bergin [3]; Sven Bergmann [1,2,4]; Dennis Bontempi [1,2]; Sacha Bors [1,2]; Leah Böttger [1,2]; Bogdan Draganski [5,6,7]; Adham Elwakil [3,8]; Györgyi V. Hamvas [9,10]; Janna Hastings [11,12]; Ilaria Iuliani [1,2]; Caroline C.W. Klaver [13,14,15,16]; Ihor Kuras [6,7]; Bart Liefers [13,14]; Ilenia Meloni [3,8]; Sofia Ortin Vela [1,2]; David Presby [1,2]; Ian Quintas [1,2]; José Vargas Quiros [13,14,]; Marc Schindewolf [9,10]; Reinier O. Schlingemann [3,17]; Mattia Tomasoni [3,8]; Olga Trofimova [1,2].

[1] *Department of Computational Biology, University of Lausanne, Lausanne, Switzerland*

[2] *Swiss Institute of Bioinformatics, Lausanne, Switzerland*

[3] *Department of Ophthalmology, University of Lausanne, Fondation Asile des Aveugles, Jules Gonin Eye Hospital, Lausanne, Switzerland.*

[4] *Department of Integrative Biomedical Sciences, University of Cape Town, Cape Town, South Africa*

[5] *Department of Neurology, Max Planck Institute for Human Cognitive and Brain Sciences, Germany*

[6] *Insel University Hospital, Bern, Switzerland*

[7] *Department of Clinical Neuroscience, Lausanne University Hospital and University of Lausanne, Switzerland*

[8] *Platform for Research in Ocular Imaging, Fondation Asile des Aveugles, Jules Gonin Eye Hospital, Lausanne, Switzerland.*

[9] *Inselspital, Bern University Hospital, University of Bern, Switzerland.*

[10] *Department for BioMedical Research, Bern University Hospital, University of Bern, Switzerland.*

[11] *Institute for Implementation Science in Health Care, Faculty of Medicine, University of Zurich, Zürich, Switzerland*

[12] *School of Medicine, University of St Gallen, St. Gallen, Switzerland*

[13] *Department of Ophthalmology, Erasmus University Medical Center, Rotterdam, The Netherlands.*

[14] *Department of Epidemiology, Erasmus University Medical Center, Rotterdam, The Netherlands.*

[15] *Department of Ophthalmology, Radboud University Medical Center, Nijmegen, the Netherlands.*

[16] *Institute of Molecular and Clinical Ophthalmology, University of Basel, Switzerland.*

[17] *University Medical Centres, Amsterdam, The Netherlands.*

Data and Code Availability

The UK Biobank data used in this study are protected and not publicly available due to data privacy regulations. Access to UK Biobank data is granted to eligible researchers through an application process administered by the UK Biobank management team, subject to approval and payment of applicable access fees. Permission to access and analyse the UK Biobank data for the present study was granted under application number 90947.

Genome-wide association study (GWAS) summary statistics generated in this study will be made publicly available on Zenodo following peer-reviewed publication. Image-derived phenotypic data will be available exclusively through the UK Biobank cohort platform after publication, subject to restricted access. The analysis code supporting the findings of this study will be released on GitHub upon peer-reviewed publication of the manuscript.

Data from the Rotterdam Study and CoLaus studies can be made available to researchers upon reasonable request through a data transfer agreement.

Funding

This research was supported by the Swiss National Science Foundation grant no. CRSII5 209510 for the “VascX” Sinergia project.

Competing Interest

All the authors declare no competing interests, including both financial and non-financial interests.

References

1. Ross, R. *et al.* Importance of Assessing Cardiorespiratory Fitness in Clinical Practice: A Case for Fitness as a Clinical Vital Sign: A Scientific Statement From the American Heart Association. *Circulation* **134**, e653–e699 (2016).
2. Swainson, M. G., Ingle, L. & Carroll, S. Cardiorespiratory fitness as a predictor of short-term and lifetime estimated cardiovascular disease risk. *Scand J Med Sci Sports* **29**, 1402–1413 (2019).
3. Myers, J. *et al.* Exercise capacity and mortality among men referred for exercise testing. *N Engl J Med* **346**, 793–801 (2002).
4. Laukkanen, J. A. *et al.* Cardiovascular fitness as a predictor of mortality in men. *Arch Intern Med* **161**, 825–831 (2001).
5. Myers, J. New American Heart Association/American College of Cardiology guidelines on cardiovascular risk: when will fitness get the recognition it deserves? *Mayo Clin Proc* **89**, 722–726 (2014).
6. Kodama, S. *et al.* Cardiorespiratory fitness as a quantitative predictor of all-cause mortality and cardiovascular events in healthy men and women: a meta-analysis. *JAMA* **301**, 2024–2035 (2009).
7. Lee, D.-C., Artero, E. G., Sui, X. & Blair, S. N. Mortality trends in the general population: the importance of cardiorespiratory fitness. *J Psychopharmacol* **24**, 27–35 (2010).
8. Aker, A., Saliba, W., Bahouth, F., Naoum, I. & Zafrir, B. Cardiorespiratory Fitness and Risk of Cardiovascular Events and Mortality in Middle Age Patients without Known Cardiovascular Disease. *J Clin Med* **12**, (2023).
9. Dores, H. *et al.* Cardiopulmonary exercise testing in clinical practice: Principles, applications, and basic interpretation. *Rev Port Cardiol* **43**, 525–536 (2024).
10. Guazzi, M. *et al.* EACPR/AHA Scientific Statement. Clinical recommendations for cardiopulmonary exercise testing data assessment in specific patient populations. *Circulation* **126**, 2261–2274 (2012).

11. Fletcher, G. F. *et al.* Exercise standards for testing and training: a scientific statement from the American Heart Association. *Circulation* **128**, 873–934 (2013).
12. Balady, G. J. *et al.* Clinician's Guide to cardiopulmonary exercise testing in adults: a scientific statement from the American Heart Association. *Circulation* **122**, 191–225 (2010).
13. Ross, R. & Myers, J. Cardiorespiratory Fitness and Its Place in Medicine. *Rev Cardiovasc Med* **24**, 14 (2023).
14. Tanaka, H., Monahan, K. D. & Seals, D. R. Age-predicted maximal heart rate revisited. *J Am Coll Cardiol* **37**, 153–156 (2001).
15. Myers, J. *et al.* A Nonexercise Estimate of Cardiorespiratory Fitness Using a Symptom Questionnaire and Clinical Variables. *J Cardiopulm Rehabil Prev* **42**, 278–285 (2022).
16. Gonzales, T. I. *et al.* Resting heart rate is a population-level biomarker of cardiorespiratory fitness: The Fenland Study. *PLoS One* **18**, e0285272 (2023).
17. Neshitov, A. *et al.* Estimation of cardiorespiratory fitness using heart rate and step count data. *Sci Rep* **13**, 15808 (2023).
18. Li, L. Y. *et al.* Prediction of cardiovascular markers and diseases using retinal fundus images and deep learning: a systematic scoping review. *Eur Heart J Digit Health* **5**, 660–669 (2024).
19. Ortín Vela, S. *et al.* Phenotypic and genetic characteristics of retinal vascular parameters and their association with diseases. *Nature Communications* **15**, 1–17 (2024).
20. Poplin, R. *et al.* Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nature Biomedical Engineering* **2**, 158–164 (2018).
21. Nusinovici, S. *et al.* Retinal photograph-based deep learning predicts biological age, and stratifies morbidity and mortality risk. *Age Ageing* **51**, (2022).
22. Zhou, Y. *et al.* A foundation model for generalizable disease detection from retinal images. *Nature* **622**, 156–163 (2023).
23. Chang, J. *et al.* Association of Cardiovascular Mortality and Deep Learning-Funduscopic

808 Atherosclerosis Score derived from Retinal Fundus Images. *Am J Ophthalmol* **217**,
809 121–130 (2020).
810 24. Gonzales, T. I. *et al.* Cardiorespiratory fitness assessment using risk-stratified exercise
811 testing and dose-response relationships with disease outcomes. *Sci Rep* **11**, 15315
812 (2021).
813 25. Hofman, A. *et al.* The Rotterdam Study: objectives and design update. *Eur J Epidemiol*
814 **22**, 819–829 (2007).
815 26. Firmann, M. *et al.* The CoLaus study: a population-based study to investigate the
816 epidemiology and genetic determinants of cardiovascular risk factors and metabolic
817 syndrome. *BMC Cardiovasc Disord* **8**, 6 (2008).
818 27. Sui, X., LaMonte, M. J. & Blair, S. N. Cardiorespiratory Fitness as a Predictor of
819 Nonfatal Cardiovascular Events in Asymptomatic Women and Men. *American journal of*
820 *epidemiology* **165**, 1413 (2007).
821 28. Vargas Quiros, J., Liefers, B., van Garderen, K. A., Vermeulen, J. P. & Klaver, C. VascX
822 Models: Deep Ensembles for Retinal Vascular Analysis From Color Fundus Images.
823 *Transl Vis Sci Technol* **14**, 19 (2025).
824 29. Blair, S. N. *et al.* Physical fitness and all-cause mortality. A prospective study of healthy
825 men and women. *JAMA* **262**, 2395–2401 (1989).
826 30. Cardiorespiratory Fitness Is an Independent Predictor of Cardiovascular Morbidity and
827 Mortality and Improves Accuracy of Prediction Models. *Canadian Journal of Cardiology*
828 **37**, 241–250 (2021).
829 31. Association of Cardiorespiratory Fitness With Coronary Heart Disease in Asymptomatic
830 Men. *Mayo Clinic Proceedings* **90**, 1372–1379 (2015).
831 32. Kip, K. E., Diamond, D., Mulukutla, S. & Marroquin, O. C. Is LDL cholesterol associated
832 with long-term mortality among primary prevention adults? A retrospective cohort study
833 from a large healthcare system. *BMJ Open* **14**, e077949 (2024).
834 33. Ravnskov, U. *et al.* Lack of an association or an inverse association between low-
835 density-lipoprotein cholesterol and mortality in the elderly: a systematic review. *BMJ*

836 *Open* **6**, e010401 (2016).
837 34. Turusheva, A., Vaes, B., Degryse, J.-M. & Frolova, E. Low cholesterol levels are
838 associated with a high mortality risk in older adults without statins therapy: An externally
839 validated cohort study. *Arch Gerontol Geriatr* **90**, 104180 (2020).
840 35. Han, X. *et al.* Association between phosphatase related gene variants and coronary
841 artery disease: case-control study and meta-analysis. *Int J Mol Sci* **15**, 14058–14076
842 (2014).
843 36. Chen, J.-H. *et al.* Identification of MYH6 as the potential gene for human ischaemic
844 cardiomyopathy. *J Cell Mol Med* **25**, 10736–10746 (2021).
845 37. Pike, S. T., Rajendra, R., Artzt, K. & Appling, D. R. Mitochondrial C1-tetrahydrofolate
846 synthase (MTHFD1L) supports the flow of mitochondrial one-carbon units into the
847 methyl cycle in embryos. *J Biol Chem* **285**, 4612–4620 (2010).
848 38. Tsai, M.-L. *et al.* IL-25 Induced ROS-Mediated M2 Macrophage Polarization via AMPK-
849 Associated Mitophagy. *Int J Mol Sci* **23**, (2021).
850 39. Rossin, F. *et al.* Transglutaminase 2 ablation leads to mitophagy impairment associated
851 with a metabolic shift towards aerobic glycolysis. *Cell Death Differ* **22**, 408–418 (2015).
852 40. Wrutniak-Cabello, C., Casas, F. & Cabello, G. Thyroid Hormone Action: The p43
853 Mitochondrial Pathway. *Methods Mol Biol* **1801**, 163–181 (2018).
854 41. Hanscombe, K. B. *et al.* The genetic case for cardiorespiratory fitness as a clinical vital
855 sign and the routine prescription of physical activity in healthcare. *Genome Med* **13**, 180
856 (2021).
857 42. Cai, L. *et al.* Causal associations between cardiorespiratory fitness and type 2 diabetes.
858 *Nat Commun* **14**, 3904 (2023).
859 43. Evans, L. M. *et al.* Comparison of methods that use whole genome data to estimate the
860 heritability and genetic architecture of complex traits. *Nat Genet* **50**, 737–745 (2018).
861 44. Klevjer, M. *et al.* Genome-Wide Association Study Identifies New Genetic Determinants
862 of Cardiorespiratory Fitness: The Trøndelag Health Study. *Med Sci Sports Exerc* **54**,
863 1534–1545 (2022).

45. Schutte, N. M., Nederend, I., Hudziak, J. J., Bartels, M. & de Geus, E. J. C. Twin-sibling study and meta-analysis on the heritability of maximal oxygen consumption. *Physiol Genomics* **48**, 210–219 (2016).
46. Chen, X. *et al.* Dominant Genetic Variation and Missing Heritability for Human Complex Traits: Insights from Twin versus Genome-wide Common SNP Models. *Am J Hum Genet* **97**, 708–714 (2015).
47. Bassett, D. R., Jr & Howley, E. T. Limiting factors for maximum oxygen uptake and determinants of endurance performance. *Med Sci Sports Exerc* **32**, 70–84 (2000).
48. Hellsten, Y. & Gliemann, L. Peripheral limitations for performance: Muscle capillarization. *Scand J Med Sci Sports* **34**, e14442 (2024).
49. Cheung, C. Y.-L., Kamran Ikram, M., Sabanayagam, C. & Wong, T. Y. Retinal Microvasculature as a Model to Study the Manifestations of Hypertension. *Hypertension* (2012) doi:10.1161/HYPERTENSIONAHA.111.189142.
50. Wong, T. Y. *et al.* Retinal arteriolar narrowing and risk of coronary heart disease in men and women. The Atherosclerosis Risk in Communities Study. *JAMA* **287**, 1153–1159 (2002).
51. Liew, G. *et al.* Retinal Vasculature Fractal and Stroke Mortality. *Stroke* (2021) doi:10.1161/STROKEAHA.120.031886.
52. Fry, A. *et al.* Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population. *Am J Epidemiol* **186**, 1026–1034 (2017).
53. Systematic Bias of Machine Learning Regression Models and Correction. https://ieeexplore.ieee.org/document/10930829.
54. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
55. Zekavat, S. M. *et al.* Deep Learning of the Retina Enables Phenome- and Genome-Wide Analyses of the Microvasculature. *Circulation* (2022)
56. Dosovitskiy, A. *et al.* An Image is Worth 16x16 Words: Transformers for Image

892 Recognition at Scale. (2020).
893 57. Modin, H. E. *et al.* Pack-Year Cigarette Smoking History for Determination of Lung
894 Cancer Screening Eligibility. Comparison of the Electronic Medical Record versus a
895 Shared Decision-making Conversation. *Annals of the American Thoracic Society* (2017)
896 doi:10.1513/AnnalsATS.201612-984OC.
897 58. Schoeler, T., Pingault, J.-B. & Kutalik, Z. The impact of self-report inaccuracy in the UK
898 Biobank and its interplay with selective participation. *Nat Hum Behav* (2024)
899 doi:10.1038/s41562-024-02061-w.
900 59. Dankner, R., Geulayov, G., Olmer, L. & Kaplan, G. Undetected type 2 diabetes in older
901 adults. *Age Ageing* **38**, 56–62 (2009).
902 60. Pierce, M. B., Zaninotto, P., Steel, N. & Mindell, J. Undiagnosed diabetes—data from
903 the English longitudinal study of ageing. *Diabetic Medicine* **26**, 679–685 (2009).
904 61. Qiao, Z. *et al.* Estimation and implications of the genetic architecture of fasting and non-
905 fasting blood glucose. *Nat Commun* **14**, 451 (2023).
906 62. Lugner, M., Rawshani, A., Helleryd, E. & Eliasson, B. Identifying top ten predictors of
907 type 2 diabetes through machine learning analysis of UK Biobank data. *Sci Rep* **14**,
908 2102 (2024).
909 63. Alberti, K. G. M. M., Zimmet, P. & Shaw, J. Metabolic syndrome--a new world-wide
910 definition. A Consensus Statement from the International Diabetes Federation. *Diabet*
911 *Med* **23**, 469–480 (2006).
912 64. Hirsch, J. A. *et al.* ICD-10: History and Context. *AJNR: American Journal of*
913 *Neuroradiology* **37**, 596 (2016).
914 65. Raisi-Estabragh, Z. *et al.* Incident cardiovascular events and imaging phenotypes in UK
915 Biobank participants with past cancer. *Heart* **109**, 1007–1015 (2023).
916 66. Raisi-Estabragh, Z. *et al.* Cardiovascular disease and mortality sequelae of COVID-19
917 in the UK Biobank. *Heart* **109**, 119–126 (2022).
918 67. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and
919 Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Series B Stat. Methodol.* **57**,

920 289–300 (1995).
921 68. Wilson, P. W. *et al.* Prediction of coronary heart disease using risk factor categories.
922 *Circulation* **97**, 1837–1847 (1998).
923 69. Lassenius, M. I. *et al.* Cardiovascular event rates increase after each recurrence and
924 associate with poor statin adherence. *European journal of preventive cardiology* **28**,
925 (2021).
926 70. Laukkanen, J. A., Kurl, S., Salonen, R., Rauramaa, R. & Salonen, J. T. The predictive
927 value of cardiorespiratory fitness for cardiovascular events in men with various risk
928 profiles: a prospective population-based cohort study. *Eur Heart J* **25**, 1428–1437
929 (2004).
930 71. Mbatchou, J. *et al.* Computationally efficient whole-genome regression for quantitative
931 and binary traits. *Nat Genet* **53**, 1097–1103 (2021).
932 72. Krefl, D. & Bergmann, S. Cross-GWAS coherence test at the gene and pathway level.
933 *PLoS Comput Biol* **18**, e1010517 (2022).
934 73. Lamparter, D., Marbach, D., Rueedi, R., Kutalik, Z. & Bergmann, S. Fast and Rigorous
935 Computation of Gene and Pathway Scores from SNP-Based Summary Statistics. *PLOS*
936 *Computational Biology* **12**, e1004714 (2016).
937 74. Krefl, D., Brandulas Cammarata, A. & Bergmann, S. PascalX: a Python library for
938 GWAS gene and pathway enrichment tests. *Bioinformatics* **39**, (2023).
939 75. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide
940 complex trait analysis. *Am J Hum Genet* **88**, 76–82 (2011).
941 76. Yang, J. *et al.* Common SNPs explain a large proportion of the heritability for human
942 height. *Nature Genetics* **42**, 565–569 (2010).
943 77. Caron, M. *et al.* Emerging properties in self-supervised vision transformers. in *2021*
944 *IEEE/CVF International Conference on Computer Vision (ICCV)* (IEEE, 2021).
945 doi:10.1109/iccv48922.2021.00951.
946 78. Are sixteen heads really better than one?
947 https://dl.acm.org/doi/10.5555/3454287.3455544 doi:10.5555/3454287.3455544.