RetFit: A Novel Deep Learning Biomarker based on 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], Sven Bergmann[1,2,3,†], David Presby[1,2,†]

Affiliations | [1] Department of Computational Biology, University of Lausanne, Lausanne, Switzerland. [2] Swiss Institute of Bioinformatics, Lausanne, Switzerland. [3] 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)

Ian Quintas
Dennis Bontempi
Sacha Bors
Olga Trofimova
Leah Böttger
Ilaria Iuliani
Sofia Ortín Vela
Sven Bergmann
David Presby

***: 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 powerful predictor of cardiovascular events and overall mortality, often surpassing traditional risk factors in prognostic value. However, its clinical use remains limited because current assessments rely on specialized equipment, trained personnel, and lengthy procedures that are often impractical for broad or routine application, especially in at-risk populations. Because CRF is closely tied to vascular health, surrogate measures that capture vascular features may offer a practical alternative for its estimation. Retinal Color Fundus Images (CFIs) provide a non-invasive window into systemic vascular health and have already demonstrated their utility in predicting cardiovascular risk factors and diseases, yet CFIs have yet to be explored for their potential to predict CRF. In this study, we develop RetFit, a novel CRF estimator derived from CFIs by leveraging state-of-the-art vision transformers. RetFit enables a non-invasive, easy-to-acquire CRF proxy, addressing some of the limitations inherent to standard CRF measures and linking retinal imaging to the cardiovascular system. We evaluated RetFit's clinical relevance by analyzing its associations with cardiovascular risk factors, disease outcomes, and exploring its genetic architecture, benchmarking it against a submaximal-exercise-test CRF (SETCRF) estimate. We find that RetFit is prognostic of both cardiovascular events (hazard ratios as low as 0.878, 95%CI 0.856 to 0.901, p<0.001) and overall mortality (hazard ratios as low as 0.780, 95%CI 0.754 to 0.801, p<0.001) and significantly associates with the majority of disease states and risk factors explored within our analysis. Although RetFit and SETCRF shared a moderate phenotypic correlation with each other (*r*=0.45), their significant genetic associations were disjoint. Interpretability analyses revealed that RetFit's predictions are driven by the retinal vasculature, with the number of arterial bifurcations showing the strongest association with RetFit (β=0.287, 95%CI 0.263 to 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 estimates of CRF may be derived from phenotypic, demographic, and anthropometric variables, providing a rapid and cost-effective alternative suited for public health and clinical settings[13,15]. In particular, estimates that utilize cardiovascular metrics, like resting heart rate, demonstrate improved performance over estimates without cardiovascular measures, potentially due to the cardiorespiratory system having a central, rate-limiting role in determining VO_2max[16,17]. Therefore, identifying novel cardiovascular proxies may lead to improvements in the accuracy and accessibility of CRF estimates.

Color fundus images (CFIs) represent a potentially promising source for such novel cardiovascular pro xies, as CFIs have been shown to provide valuable insights into the cardiovascular system and reflect systemic vascular health[18,19]. Studies have shown that CFIs, whether analyzed directly via end-to-end deep learning pipelines or through extracted features quantifying retinal vascular properties, can reliably predict multiple cardiovascular risk factors (e.g., age, sex, smoking status, and hypertension)[20,21] and diseases (e.g., myocardial infarction, ischaemic stroke, and heart failure)[22,23]. Importantly, CFIs are noninvasive and relatively easy to acquire, making them an appealing resource to estimate CRF. Building on this rationale, we develop RetFit, a novel estimator of CRF from retinal images, based on a state-of-the-art Vision Transformer. To determine the clinical relevance of RetFit, we assessed its associations with genetics, risk factors, and diseases, and directly benchmarked the diagnostic and prognostic capacities of this novel biomarker against an established measure of submaximal exercise testing CRF (SETCRF)[24] (**Fig. 1**) Ultimately, this work aims to provide a scalable and accessible method for estimating CRF, overcoming the key barriers of conventional testing.

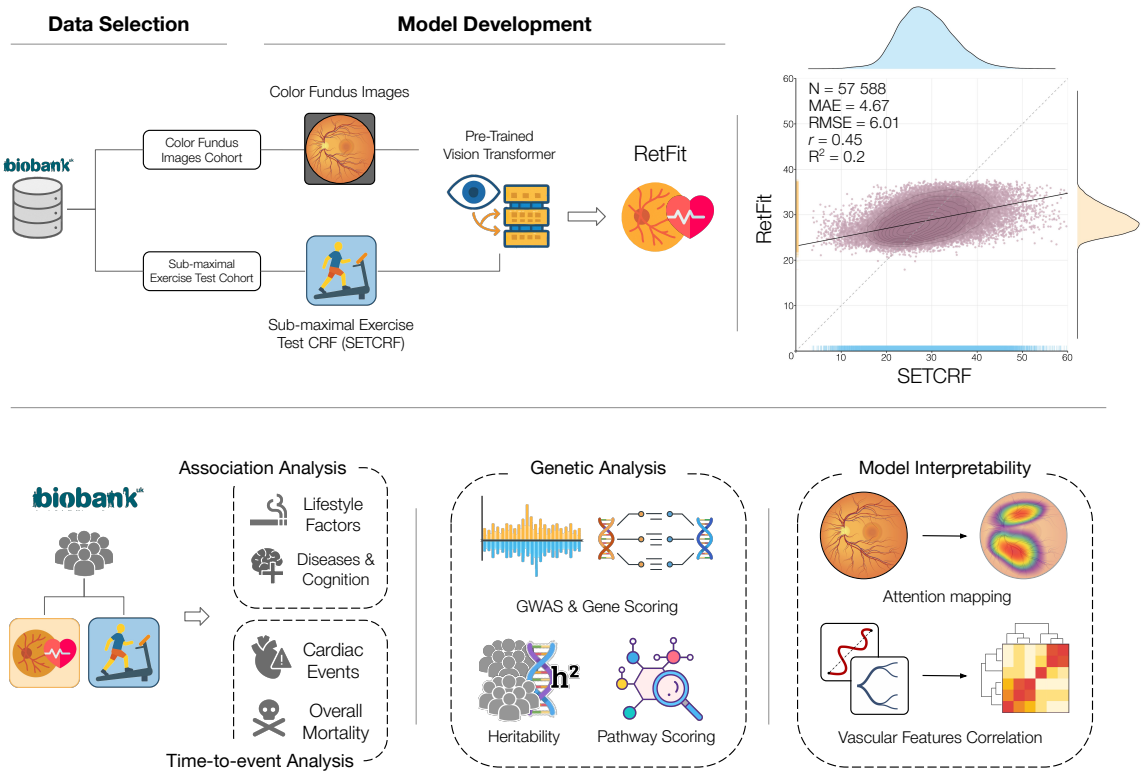

Figure 1 | Study Overview. We selected the subset of genotyped subjects from the UK Biobank who underwent a cardiorespiratory fitness (CRF) assessment via sub-maximal exercise testing (SETCRF) and had colour fundus images (CFIs) taken, as well as access to available health outcome information and risk factors, for a total of N=57 588 subjects. 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). Using SETCRF as a benchmark, 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. Finally, we examined associations between retinal vascular properties and RetFit/SETCRF based on attention maps.

Results

Generation of a Fitness Marker from Fundus Images

We fine-tuned a transformer model specifically designed for CFIs, RETFound[22], on VO_2max estimates 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). We then conducted additional analyses to evaluate RetFit's diagnostic and prognostic value, examine its genetic architecture, and benchmark these associations against SETCRF (**Fig. 1**).

RetFit Informs Upon Risk Factors and Disease

We performed a series of linear regression analyses to evaluate the relative and independent contributions of RetFit and SETCRF to diseases and associated risk factors. 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. Of the 15 continuous risk factors examined, RetFit significantly informed on 14, whereas SETCRF informed on 13 factors (**Fig. 2a**). Of the 12 factors significantly associated with both RetFit and SETCRF, 10 demonstrated concordant directionality. Only LDL cholesterol and total cholesterol exhibited opposing associations, being positively associated with RetFit and negatively associated with SETCRF. We found three risk factors for which there was a difference in significance between RetFit and SETCRF: insulin-like growth factor 1 (IGF-1) and reasoning associated with RetFit but not SETCRF, whereas triglyceride levels were associated with SETCRF but not RetFit. Results from the log-likelihood ratio tests demonstrate that reasoning and triglycerides were the only two outcomes for which adding one predictor to a model already containing the other led to a significant improvement in model fit (right panel in **Fig. 2a**).

Of the 15 binary health outcomes examined, RetFit significantly informed upon 11 outcomes, whereas SETCRF significantly informed upon eight outcomes. RetFit and SETCRF informed upon five of the binary health outcomes in the same direction. Prior smoking status was the

only binary risk factor to exhibit opposing directions of association between RetFit and SETCRF, with RetFit showing a negative association and SETCRF showing a positive association. (**Fig. 2b**). Out of these 15 binary health outcomes, five were significant for RetFit but not for SETCRF (current smoking status, cataract, cataract surgery, prospective memory, and occasional alcohol consumption), whereas two associations were significant for SETCRF but not for RetFit (cancer and dyslipidemia). The LRT results for the binary outcomes showed a significant improvement in model fit when adding both RetFit and SETCRF for occasional alcohol consumption, cancer, cataract, cataract surgery, and prospective memory. Cancer diagnosis was the only outcome for which adding SETCRF to a model with RetFit improved the model, whereas for all the other outcomes, adding RetFit improved the model with SETCRF (see the right panel in **Fig. 2b)**.

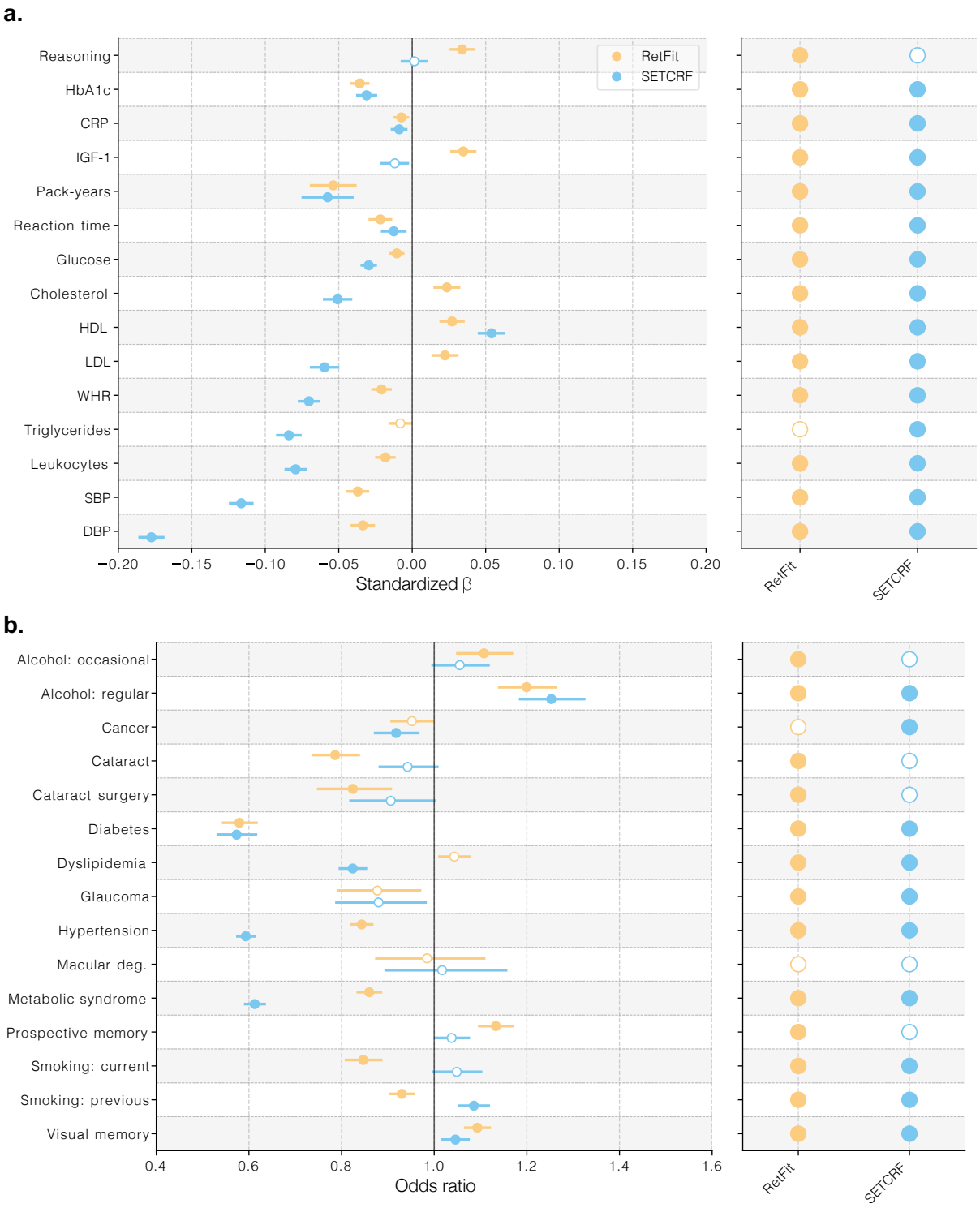

Figure 2 | Association of RetFit and SETCRF with risk factors and diseases. a. The left panel shows standardized coefficients for RetFit and SETCRF from independent models predicting continuous outcomes; the right panel shows results from likelihood ratio tests (LRTs), comparing models that included both RetFit and SETCRF with models including only one of the two CRF estimates. **b.** The left panel shows standardized coefficients for RetFit and SETCRF from independent models predicting binary outcomes; 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). HbA1c: Glycated hemoglobin; CRP: C-reactive protein; IGF-1: Insulin-like growth factor 1; HDL & LDL: high- and low-density lipoprotein; WHR: waist-to-hip ratio; 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. 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.

Overall, RetFit demonstrated better performance than SETCRF in the prognosis of cardiovascular events before and after correcting for cardiovascular risk factors. In a continuous setting, RetFit showed a stronger protective association, as indicated by a lower hazard ratio (HR) per standard deviation increase (**Fig. 3a**; see **Suppl. Fig. 3a** for the HR of CRF estimates expressed in METs). In the fully adjusted model, RetFit was significantly associated with cardiovascular events (HR=0.936, 95%CI 0.903 to 0.970, $p<0.001$) while SETCRF was not (HR=0.979, 95%CI 0.958 to 1.00, $p=0.0653$). Furthermore, 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**). These findings were consistent in the group analysis: Kaplan-Meier curves demonstrated good stratification of increasing cardiovascular event risk with decreasing fitness score (both RetFit and SETCRF; **Fig. 3b**) and, in the fully adjusted model, the third, fourth, and fifth RetFit quintiles were significantly associated with fewer cardiovascular events, while for SETCRF, none were (**Fig. 3c**). 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 χ^2=13.4, p<0.01) while the opposite was not true (LRT, 4 df χ^2=4.78, p=0.311; see **Suppl. Table 6a**) - suggesting, once again, that RetFit captured prognostic information not accounted for by SETCRF.

When using overall mortality as the outcome in a continuous univariate setting, the performance of the two estimates followed the trend observed with cardiovascular events, with RetFit outperforming SETCRF. However, in the model adjusted for demographics, the two estimates yielded similar results, and in the fully adjusted model, SETCRF showed a stronger protective association (RetFit: HR=0.945, 95%CI 0.902 to 0.989, p<0.05; SETCRF: HR=0.892, 95%CI 0.868 to 0.917, p<0.001; **Fig. 3a**; see **Suppl. Fig. 3a** for the HR of CRF estimates expressed in METs). When both estimates were included alongside all covariates, each estimate significantly improved model fit (see **Suppl. Table 5b**). In the group analysis, we observed similar results, starting with the unadjusted Kaplan-Meier analysis (**Fig. 3b** and **Suppl. Fig. 3b**), although RetFit's performance degraded after adjusting for demographics (**Suppl. Fig. 3c**) and the other cardiovascular risk factors, with SETCRF showing better performance than RetFit at stratifying risk (**Fig. 3c**). 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**.

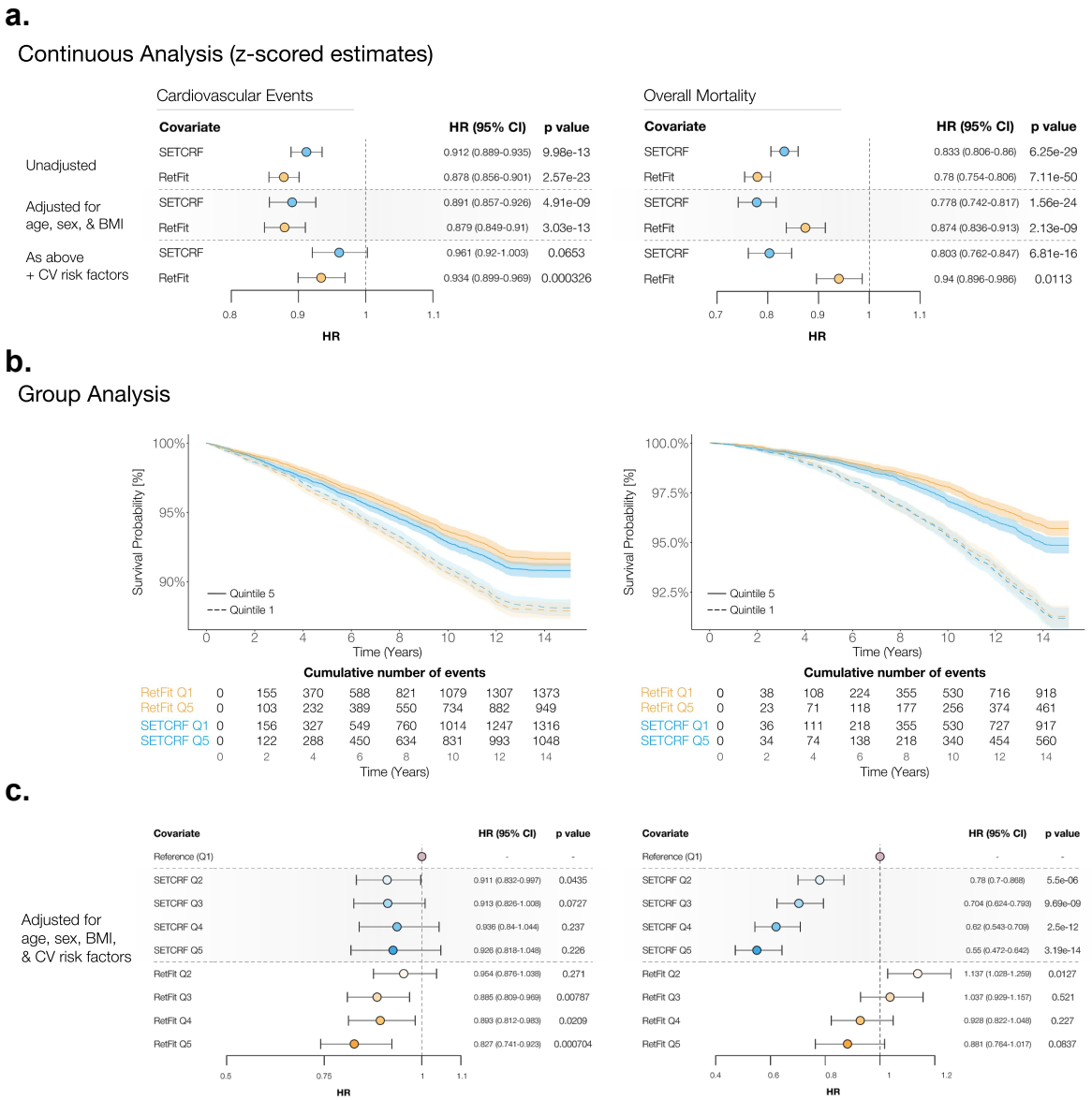

Figure 3 | Prognostic performance of RetFit and SETCRF. a. Forest plots of the CRF estimates as a continuous parameter for cardiovascular events (left) and overall mortality (right). **b.** Kaplan-Meier survival analysis of the RetFit and SETCRF (quintiles 1 and 5), and **c.** forest plots for all the quintiles. 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. Figure 4a & 4b** 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. 4;** see **Suppl. Tables 10a** and **10b** 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 11a** and **11b** for their description, and **Suppl. Tables 11c** and **11d** 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).

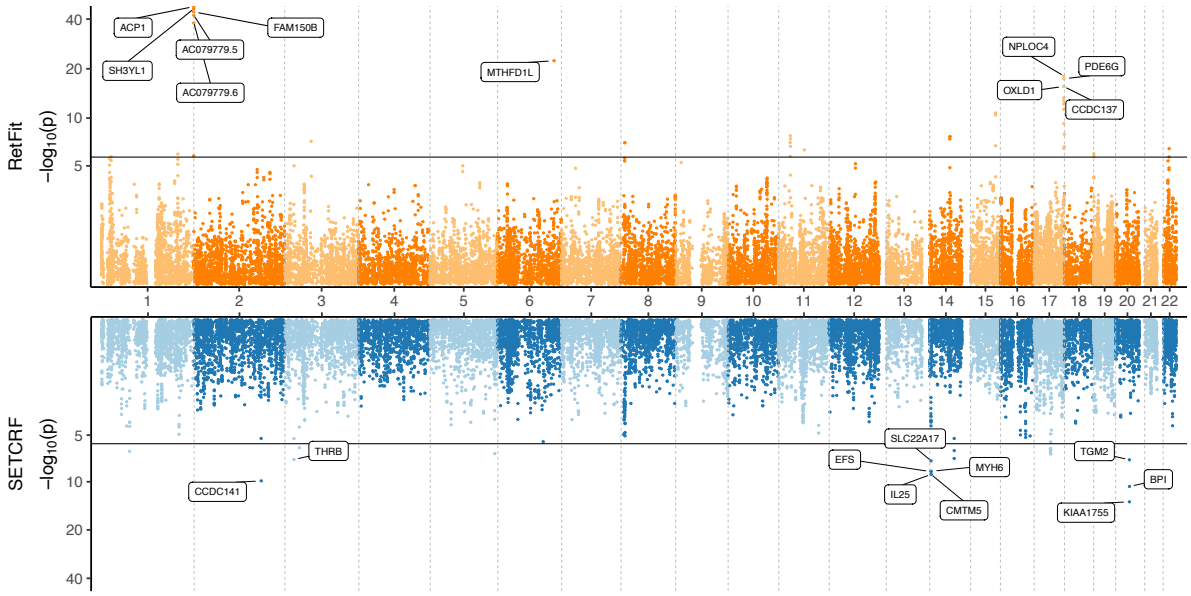

Figure 4 | Gene enrichment analysis results for RetFit and SETCRF. Miami plot showing all genes that were scored for the analysis using PascalX. Each dot corresponds to the p-value of one of the 24 922 genes scored. The x-axis shows the position of the gene in the genome, and the y-axis shows the significance 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. 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. 5a**). Furthermore, we utilized 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[25]. 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) (**Suppl. Fig. 5**). 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. 5b**). The number of A bifurcations remained the vascular property showing the strongest association with RetFit (β=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 (β=0.391, 95%CI 0.352 to 0.43, p<0.001) (**Fig. 5b**).

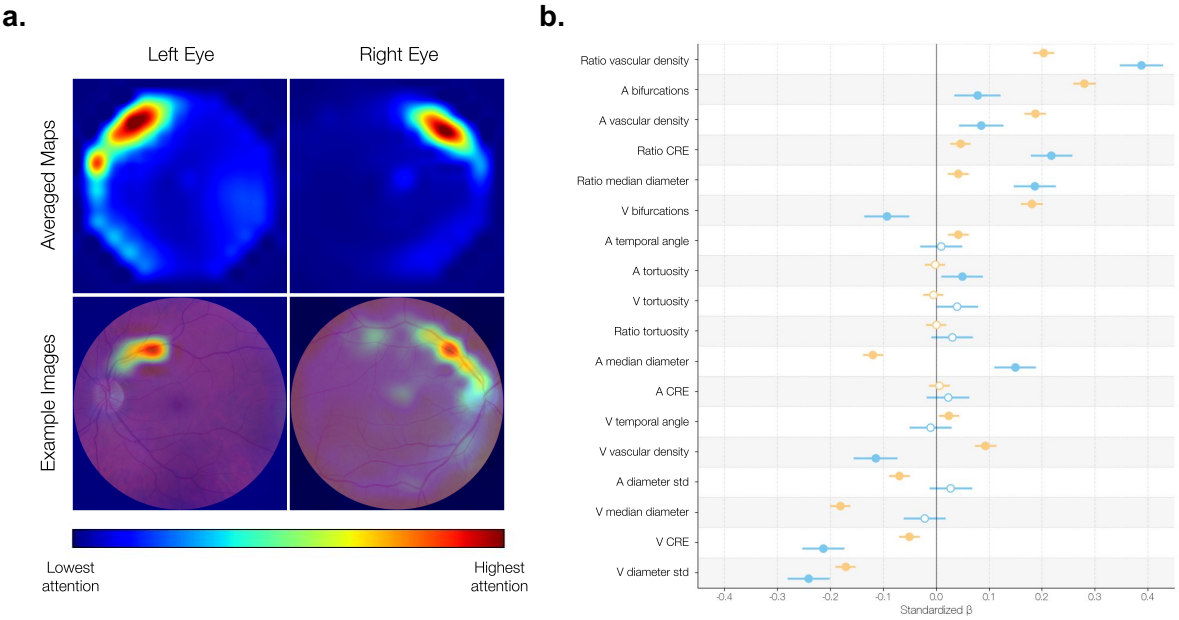

Figure 5 | Attention mapping of the RetFit vision transformer model and association of RetFit and SETCRF with retinal vasculature features. a. The top two images show the average attention maps across all images for the left and right eyes, respectively. The bottom two images present examples of attention 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[25] 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 CRF could be reliably estimated from retinal CFIs and to determine whether this estimator (that we call *RetFit)* correlates with current disease diagnoses or can predict future cardiovascular events. We found that while RetFit only captures a moderate fraction of CRF's variability (R^2=0.2), it nevertheless provided diagnostic and prognostic information regarding disease, in some cases outperforming the original CRF estimate. 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 measure for CRF, underscoring the need for continued exploration into the biological mechanisms linking retinal vascular characteristics with systemic health.

The prognostic power of objectively measured VO_2max in predicting future cardiovascular events is extensively documented across large-scale cohort studies[1–3,8,13]. A strong, inverse, and graded association exists between an individual's VO_2max and the incidence of coronary artery disease, myocardial infarction, stroke, and cardiovascular mortality in general[3,26]. Previous work has demonstrated that CRF remains an independent predictor of cardiovascular events after adjustment for established multivariable risk scores, such as the Framingham Risk Score, or major cardiovascular risk factors (such as sex, age, and blood pressure)[1,27,28]. In our analysis, this was true for RetFit (i.e., it remained significant after adjustment for age, sex, BMI, SBP, HDL cholesterol, total cholesterol, smoking status, diabetes diagnosis, hypertension treatment, and previous cardiovascular events) but, interestingly, not for the sub-maximal-exercise estimate SETCRF. This finding may reflect fundamental differences between the two estimates: whereas SETCRF is derived from exercise performance, which is closely linked to established cardiovascular risk factors, RetFit leverages retinal image features that may quantify additional, subclinical markers of vascular health not captured by conventional measurements. Consequently, RetFit could complement existing risk scores, offering incremental prognostic value to routinely acquired clinical data. It is worth noting that, in the study proposing SETCRF, Gonzales et al. (2021) reported that the estimate remained a significant predictor of cardiovascular events after adjustment for several risk factors[24]. In our analysis, significance was not reached (p=0.065) after adjusting for a slightly different set of covariates (including systolic blood pressure, HDL, and total cholesterol, and excluding several covariates from Gonzales et al.) and using a different definition of prior cardiovascular events. Differences in sample size (57 588 vs. 79 981, as we selected the subset of participants with retinal images, see **Suppl. Fig. 1**) and grouping strategy (quintiles 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 noticed a major difference between the association of the two traits with cognitive performance metrics such as reasoning or prospective memory. In particular, these appear to be better captured by RetFit than SETCRF, pointing to potential systemic relevance beyond cardiovascular outcomes. Interestingly, the standardized beta 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[29–31]. 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, reinforcing its broader clinical significance.

From the genetic analysis, we found that both RetFit and SETCRF yielded significant genetic associations; however, the genes and pathways identified were distinct, with no overlap observed between the two traits. The most significant hit we identified for RetFit, *ACP1* (p=5.84e-48), is a protein-encoding gene regulating cellular signaling by removing phosphate groups from tyrosine residues on proteins, known to be involved in coronary artery disease[32] (see **Supp. Tables 10a** and **10b**). For SETCRF, our findings indicate *MYH6* (p=2.63e–09) as one of the strongest hits; this gene encodes the alpha-myosin heavy chain, the most abundant protein in the heart and a central determinant of cardiac muscle contraction, with established

associations to cardiomyopathy[33]. Furthermore, we observed significant but distinct gene hits related to mitochondrial processes for both traits. In the case of RetFit, our analysis revealed a significant association with *MTHFD1L* (p=5.6e-23), a protein-encoding gene involved in the synthesis of tetrahydrofolate (THF) in the mitochondrion[34]. For SETCRF, we found a significant association with several genes such as *IL25* (p=1.05e-09), a cytokine promoting mitochondrial respiration[35], *TGM2* (p=6.04e-08), an enzyme regulating mitophagy[36], or *THRB* (p=5.91e-08), a thyroid hormone receptor[37], known to be involved in the mitochondrial pathway. Furthermore, we identified several significant pathways linking SETCRF with CVD when running the pathway analysis (e.g., abnormality of cardiac conduction or impaired conduction of cardiac impulse, see **Supp. Table 11b** for pathway description), but not for RetFit, where pathways related to the cardiovascular system failed to reach significance. However, as we showed in the survival analysis, RetFit demonstrated better prognostic power for CV events than SETCRF, highlighting the need to better understand the differentiation between diagnostic and prognostic leverage and genetic mechanisms. In contrast, we found that the disease and genetic analyses seem to converge on endpoints associated with neuronal development, specifically, reasoning, visual memory, reaction time, and prospective memory. RetFit showed significant associations with all these risk factors (whereas SETCRF did not for all of them), and multiple enriched pathways suggest that its genetic architecture influences processes related to the nervous system, neuronal development, and cerebral function (e.g., regulation of neuron differentiation, see **Supp. Table 11a**). Consequently, the associations we observed with reasoning, visual memory, reaction time, and prospective memory are consistent with the genetic evidence, reinforcing the hypothesis that RetFit is functionally related to neuronal processes.

Regarding heritability, we found that both RetFit and SETCRF showed higher estimates than other studies that used the UKBB cohort[38,39]. For example, the results of Cai et al. (2023), based on the same CRF estimate and a similar cohort, found a heritability estimate of 12.7% (SE=1.0%), which represents a 66% decrease compared to our estimate. This discrepancy is

possibly due to the difference in methods for heritability estimation: Cai et al. used LD score regression[39], while we used the Genomic Restricted Maximum Likelihood (GREML) method, which has been shown to yield more accurate results[40]. It is also important to note that UKBB h^2 estimates are lower than in other cohorts, such as in The Trøndelag Health Study, where the h^2 estimate for the total population was 37.5% (SE=4.6%)[41]. Furthermore, several twin studies and family studies have estimated CRF heritability to range from 40 to 70%[39,42], which are values substantially higher than what we obtained for either RetFit or SETCRF, although family studies tend to yield higher h^2 estimates than population-wide studies[43].

The physiological limit of CRF is fundamentally determined by the cardiovascular system's ability to transport oxygen to metabolically active tissues[44]. While cardiac output and arterial oxygen content are critical components, CRF may also hinge on the integrity and function of the systemic vasculature, where oxygen exchange occurs[45]. The retinal circulation provides a unique, non-invasive window to observe this vasculature network in vivo, and morphological changes within it are well-established markers for systemic vascular pathology[46]. For example, narrowing of retinal arterioles and widening of venules, often summarized as a reduced arteriolar-to-venular ratio (AVR), are robustly associated with hypertension and the long-term risk of coronary heart disease[47]. Beyond vessel geometry, the branching complexity of the vascular tree, quantified by its fractal dimension, reflects the efficiency of tissue perfusion; a sparser, less complex network (lower fractal dimension) is linked to vascular rarefaction and has been shown to predict adverse outcomes like stroke[48]. Because these retinal features capture systemic vascular health, they may also reflect physiological pathways that underlie CRF. In line with this observation, most TIFs showed significant associations with both RetFit and SETCRF, with some of them displaying stronger associations with SETCRF than RetFit, despite the latter being based on CFIs. In our view, this shows the potential for future research linking CRF to retinal imaging.

It is critical to interpret these findings in the context of the study cohort, which consists largely of individuals of White British ancestry, aged from 40 to 70. As a volunteer-based cohort, the UKBB is subject to healthy-volunteer and socio-economic selection biases: participants are typically healthier, less socioeconomically deprived, and report fewer health conditions than the general population[49]. However, the UKBB remains the largest and one of the only cohorts for which both CFIs and CRF estimates are available. In this regard, it is important to acknowledge that, since RetFit was modeled after SETCRF, which is itself a CRF estimate[24]. Notably, VO_2max was not objectively measured in any of the UKBB participants, and the validation cohort demonstrated higher exercise capacity than the average UKBB participant[24]. Such factors likely affect the accuracy of CRF estimation and, consequently, may also influence our analyses. Finally, it is important to note that, since RetFit is derived from a deep learning (DL) pipeline, its distribution may exhibit a reduced standard deviation compared to other CRF estimates (due to the tendency of DL models to produce outputs that regress toward the mean, thereby compressing the range of predicted values[50]), and therefore yield inflated HRs when the latter are reported in units of MET. Given this, we opted to report the HR in units of standard deviation rather than METs whenever comparing the predictive power of RetFit and SETCRF in a continuous setting. We nevertheless report the HR in units of MET in the supplementary material to allow for a direct comparison with other DL and non-DL derived markers from the literature.

In summary, in this work, we demonstrated that DL can generate a powerful and scalable biomarker from retinal images that reflects an individual's health and disease risk. While conceived as a surrogate for CRF, our findings reveal a more complex and compelling story: RetFit's superior prognostic power for future cardiovascular diseases, its comparable diagnostic capacity regarding disease outcomes and risk factors, combined with its distinct genetic architecture from SETCRF, establishes it not as a mere proxy but as a novel biological indicator. This reinforces the idea that the retinal microvasculature provides a unique window into a dimension of systemic health that is related to, but not fully captured by, traditional

physiological fitness. Overall, RetFit offers a major practical advantage as, unlike classical cardiorespiratory fitness tests, which require expensive equipment, trained personnel, and maximal physical effort, RetFit can be estimated from a simple retinal image, making it vastly more scalable and accessible 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[51]. 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[52]. 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[53]. 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

increasing from 0 to 5×10^{-4}), followed by a cosine annealing schedule (where *r* decreases from 5×10^{-4} to 1×10^{-6}). We set the batch size to 16, the dropout rate to 0.4, and the image size to 224×224 pixels. After each epoch, we evaluated the model on a validation dataset and then kept the model that had the lowest mean square error across all epochs. To ensure robust evaluation, we implemented a 5-fold cross-validation strategy with a leave-one-out design: for each fold, the model was trained on 80% of the dataset and evaluated on the remaining left-out split. All the results we report in the manuscript were obtained by aggregating the predictions across the five left-out partitions. In addition, we fine-tuned separate models for each laterality, yielding five cross-validated models for the right eye and five for the left eye. To obtain the final subject- and instance-level prediction, we averaged the outputs from the left and right eyes when both images passed quality control, or used the single available prediction otherwise. Pearson correlations, R^2, MAE, and RMSE were computed on the CRF averaged over both eyes.

Disease Association

For the disease association analyses, we included a set of disease and general systemic risk factors that had either a binary or a continuous outcome known to affect disease susceptibility. We collected individual-level disease and risk factor information from the UKBB. We prioritized information provided through health-related records (Category 100091), biological samples (Category 100078), and Physical measures (Category 100006) over those obtained through the assessment center Touchscreen questionnaire (Category 100025) or verbal interviews (Category 100071) as self-reported measures are often unreliable due to high reporting errors or inconsistencies[54,55]. We made exceptions for alcohol intake frequency and smoking status (i.e., never, previous, current). As studies have shown the prevalence of undiagnosed diabetic or pre-diabetic conditions and associated health complications[56,57], we determined diabetes status using a combination of variables: an official diagnosis based on the date E10 first

reported (insulin-dependent diabetes mellitus), HbA1C readings of larger or equal to 48 mmol/mol, or glucose levels of larger or equal to 11.1 mmol/l[58,59]. For participants who took antihypertensive medication, we added 10 mmHg to SBP and 5 mmHg to DBP values and for those 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. Participants who had cholesterol levels higher than 5, HDL lower than 1, LDL higher than 4.1, or triglycerides higher than 2.2 were considered to have dyslipidemia. We defined hypertension as having SBP > 140 or DBP > 90 or the use of blood pressure medication. Following the International Diabetes Federation consensus[60], we defined metabolic syndrome as having three or more of these phenotypes: 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. All the fields used for the analysis and their descriptions can be found in **Suppl. Table 3**.

We used multiple linear regression for continuous risk factors and multiple logistic regression for binary ones. All models were corrected for sex, age, BMI, income, educational level, and the Townsend deprivation index. Furthermore, all the independent variables that were used for the linear regressions were standardized. We also binarized all diseases collected as “date first diagnosed” for the logistic regressions. To determine the significance of regression coefficients, we computed p-values and corrected for multiple hypothesis testing using the Benjamini-Hochberg method[61], using a false discovery rate of 0.05. We also used the log-likelihood ratio test (LRT) to compare the addition of independent information for RetFit and SETCRF by fitting a nested model (with either RetFit or SETCRF and the rest of the covariates) and a model with both predictors with the same set of covariates.

Survival Analysis

We defined two distinct clinical endpoints for the survival analysis: cardiovascular events (obtained by aggregating Ischaemic Heart Disease (IHD), Myocardial Infarction (MI), stroke events, Non-Ischaemic Cardiomyopathies (NIC), and thrombotic events), and overall mortality. Time to event (TTE) was defined as the days between the date when the retinal image was taken and the date of the reported event. If the subject experienced multiple events, the first event after the imaging visit was used. We fitted Cox Proportional Hazard (PH) models with three levels of covariate adjustment to assess the robustness of the CRF estimates. First, we conducted univariate analyses including only the estimate of interest. Second, we adjusted for age, sex, and BMI, as these are key determinants of CRF. Finally, we further adjusted for additional cardiovascular risk factors that broadly matched those in the Framingham score[62], i.e., age, sex, SBP, Total and HDL cholesterol, BMI, smoking status, diabetes diagnosis, and hypertension treatment (i.e., medication), plus whether the subject had a previous cardiovascular event. Hypertension treatment was defined by aggregating male- and female-specific medication and filtering for "blood pressure medication". Diagnosis of previous cardiovascular events was included due to an increased risk of recurrence[63]. All the fields used for the analysis and their descriptions can be found in **Suppl. Tables 1–4**, while the distribution of events and continuous risk factors can be found in **Suppl. Fig. 2.** For both estimates and for each clinical endpoint, we fit two independent models (i.e., one for RetFit and one for SETCRF) using continuous measures as predictors. Additionally, we conducted a group analysis by stratifying participants into quintiles based on their CRF estimates, reflecting a common practice in the literature where CRF estimates are often categorized into quantiles or fitness groups to assess dose-response relationships and facilitate clinical interpretation[11,6,64]. We applied the same stratification method to the RetFit score to compare its ability to stratify risk with that of the estimated CRF. To compare the different CRF estimates and investigate whether they improved the fit and the prognostic power of the different Cox models, we used the LRT to compare nested models (i.e., to compare a model with a fixed set of covariates

and one of the estimates with a model with the same covariates and both estimates) (see **Suppl. Table 5–8**).

Genome-wide association study (GWAS)

GWAS were conducted using REGENIE[65]. Before performing the analyses, the genotypic data were processed following UKBB guidelines. Variants failing quality control thresholds (MAF < 0.01, MAC < 100, genotype missingness > 0.1, or Hardy–Weinberg equilibrium $p < 1\times10^{-15}$) were excluded. The GWAS were corrected for a set of covariates known to influence phenotypic variability (cardiovascular or eye-specific factors). The covariates were age at assessment center visit, sex, BMI, age-squared, sex-by-age, sex-by-age-squared, spherical power, cylindrical power, spherical power-squared, cylindrical power-squared, imaging instance, assessment center, genotype measurement batch, and the first 20 genetic Principal Components (PCs) (see **Suppl. Table 9** for corresponding UKBB data fields).

Gene and Pathway Scoring

Gene and pathway scores were computed using PascalX, a Python library specialized in gene and pathway scoring for GWAS summary statistics[66–68]. PascalX utilizes the sum of χ^2 method to test for gene enrichment for gene scoring and a rank-based method for pathway scoring, which unifies the gene p-values distribution and aggregates it to χ^2 distributed random variables. The UK10K panel was used for gene scoring, and both protein-coding genes and lincRNAs (long intergenic non-coding RNA) were scored using the “saddle” method, taking into account all SNPs within a 50kb window around each gene and with a MAF of 0.05. All pathways available in MSigDB v7.2 were scored (n=31120), with any co-occurring genes located less than 100 kb apart being fused and rescored. All significant gene and pathway scores were adjusted for multiple testing using Bonferroni and Benjamini-Hochberg methods, respectively.

Heritability and Genetic Correlation

SNP-based heritabilities were estimated for RetFit and SETCRF using the GCTA (Genome-wide complex trait analysis) software package[69]. The proportion of variance of the trait explained by the genetic data was estimated with the genomic-relatedness-based restricted maximum-likelihood (GREML) method[69,70]. Genetic relationship matrices (GRM) for each chromosome were constructed 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")[71], 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[72]. 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 folds, separately for left and right eyes, to produce population-level maps. For visualization

purposes only, we upsampled such maps from 14 × 14 to the original image resolution using bicubic interpolation.

Vascular Associations

To explore the potential relationship between the retinal vasculature and both RetFit and CRF, we leveraged recently developed state-of-the-art deep learning models called *VascX* to extract 18 vascular traits from CFI[25]. *VascX* models work in a 2-step procedure, the first step being the segmentation of the CFI in order to separate retinal features (e.g., optic disc, vessels, and more precisely artery/vein), and the second step being the feature extraction from the masks that were output during the first step. In this case, we extracted 18 Tangible Image Features (TIFs) derived from the vessels' segmentation masks, and they comprise features such as tortuosity or vessel density (see **Suppl. Table 12** for the full list) and were all separated between arteries and veins. For a more detailed description of the algorithms to compute these features, see Ortin-Vela et al.[19] and Vargas Quiros et al.[19,25]. Once the TIFs were extracted, we used the same method to match them to RetFit/CRF estimates, meaning that values were averaged between eyes if both eyes were used to compute the average for RetFit. We then computed Pearson correlation coefficients and associated tests for each vascular trait independently. Multiple linear regressions with either Retfit or SETCRF as the response variable and each of the 18 TIFs as the dependent variable were also fitted, and all models were corrected for age, sex, and BMI. Multiple hypothesis testing was addressed using Benjamini-Hochberg, and all reported p-values have been corrected accordingly.

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 Atherosclerosis Score derived from Retinal Fundus Images. *Am J Ophthalmol* **217**, 121–130 (2020).
24. Gonzales, T. I. *et al.* Cardiorespiratory fitness assessment using risk-stratified exercise testing and dose-response relationships with disease outcomes. *Sci Rep* **11**, 15315 (2021).
25. Vargas Quiros, J., Liefers, B., van Garderen, K. A., Vermeulen, J. P. & Klaver, C. VascX Models: Deep Ensembles for Retinal Vascular Analysis From Color Fundus Images. *Transl Vis Sci Technol* **14**, 19 (2025).
26. Blair, S. N. *et al.* Physical fitness and all-cause mortality. A prospective study of healthy men and women. *JAMA* **262**, 2395–2401 (1989).
27. Cardiorespiratory Fitness Is an Independent Predictor of Cardiovascular Morbidity and Mortality and Improves Accuracy of Prediction Models. *Canadian Journal of Cardiology* **37**, 241–250 (2021).
28. Association of Cardiorespiratory Fitness With Coronary Heart Disease in Asymptomatic Men. *Mayo Clinic Proceedings* **90**, 1372–1379 (2015).
29. Kip, K. E., Diamond, D., Mulukutla, S. & Marroquin, O. C. Is LDL cholesterol associated with long-term mortality among primary prevention adults? A retrospective cohort study from a large healthcare system. *BMJ Open* **14**, e077949 (2024).
30. Ravnskov, U. *et al.* Lack of an association or an inverse association between low-density-lipoprotein cholesterol and mortality in the elderly: a systematic review. *BMJ Open* **6**, e010401 (2016).
31. Turusheva, A., Vaes, B., Degryse, J.-M. & Frolova, E. Low cholesterol levels are associated with a high mortality risk in older adults without statins therapy: An externally validated cohort study. *Arch Gerontol Geriatr* **90**, 104180 (2020).
32. Han, X. *et al.* Association between phosphatase related gene variants and coronary

artery disease: case-control study and meta-analysis. *Int J Mol Sci* **15**, 14058–14076 (2014).

33. Chen, J.-H. *et al.* Identification of MYH6 as the potential gene for human ischaemic cardiomyopathy. *J Cell Mol Med* **25**, 10736–10746 (2021).
34. Pike, S. T., Rajendra, R., Artzt, K. & Appling, D. R. Mitochondrial C1-tetrahydrofolate synthase (MTHFD1L) supports the flow of mitochondrial one-carbon units into the methyl cycle in embryos. *J Biol Chem* **285**, 4612–4620 (2010).
35. Tsai, M.-L. *et al.* IL-25 Induced ROS-Mediated M2 Macrophage Polarization via AMPK-Associated Mitophagy. *Int J Mol Sci* **23**, (2021).
36. Rossin, F. *et al.* Transglutaminase 2 ablation leads to mitophagy impairment associated with a metabolic shift towards aerobic glycolysis. *Cell Death Differ* **22**, 408–418 (2015).
37. Wrutniak-Cabello, C., Casas, F. & Cabello, G. Thyroid Hormone Action: The p43 Mitochondrial Pathway. *Methods Mol Biol* **1801**, 163–181 (2018).
38. Hanscombe, K. B. *et al.* The genetic case for cardiorespiratory fitness as a clinical vital sign and the routine prescription of physical activity in healthcare. *Genome Med* **13**, 180 (2021).
39. Cai, L. *et al.* Causal associations between cardiorespiratory fitness and type 2 diabetes. *Nat Commun* **14**, 3904 (2023).
40. Evans, L. M. *et al.* Comparison of methods that use whole genome data to estimate the heritability and genetic architecture of complex traits. *Nat Genet* **50**, 737–745 (2018).
41. Klevjer, M. *et al.* Genome-Wide Association Study Identifies New Genetic Determinants of Cardiorespiratory Fitness: The Trøndelag Health Study. *Med Sci Sports Exerc* **54**, 1534–1545 (2022).
42. 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).
43. 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).

44. 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).
45. Hellsten, Y. & Gliemann, L. Peripheral limitations for performance: Muscle capillarization. *Scand J Med Sci Sports* **34**, e14442 (2024).
46. 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.
47. 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).
48. Liew, G. *et al.* Retinal Vasculature Fractal and Stroke Mortality. *Stroke* (2021) doi:10.1161/STROKEAHA.120.031886.
49. 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).
50. Systematic Bias of Machine Learning Regression Models and Correction. https://ieeexplore.ieee.org/document/10930829.
51. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
52. Zekavat, S. M. *et al.* Deep Learning of the Retina Enables Phenome- and Genome-Wide Analyses of the Microvasculature. *Circulation* (2022) doi:10.1161/CIRCULATIONAHA.121.057709.
53. Dosovitskiy, A. *et al.* An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. (2020).
54. Modin, H. E. *et al.* Pack-Year Cigarette Smoking History for Determination of Lung

Cancer Screening Eligibility. Comparison of the Electronic Medical Record versus a Shared Decision-making Conversation. *Annals of the American Thoracic Society* (2017) doi:10.1513/AnnalsATS.201612-984OC.

55. Schoeler, T., Pingault, J.-B. & Kutalik, Z. The impact of self-report inaccuracy in the UK Biobank and its interplay with selective participation. *Nat Hum Behav* (2024) doi:10.1038/s41562-024-02061-w.
56. Dankner, R., Geulayov, G., Olmer, L. & Kaplan, G. Undetected type 2 diabetes in older adults. *Age Ageing* **38**, 56–62 (2009).
57. Pierce, M. B., Zaninotto, P., Steel, N. & Mindell, J. Undiagnosed diabetes—data from the English longitudinal study of ageing. *Diabetic Medicine* **26**, 679–685 (2009).
58. Qiao, Z. *et al.* Estimation and implications of the genetic architecture of fasting and non-fasting blood glucose. *Nat Commun* **14**, 451 (2023).
59. Lugner, M., Rawshani, A., Helleryd, E. & Eliasson, B. Identifying top ten predictors of type 2 diabetes through machine learning analysis of UK Biobank data. *Sci Rep* **14**, 2102 (2024).
60. Alberti, K. G. M. M., Zimmet, P. & Shaw, J. Metabolic syndrome--a new world-wide definition. A Consensus Statement from the International Diabetes Federation. *Diabet Med* **23**, 469–480 (2006).
61. 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.* **57**, 289–300 (1995).
62. Wilson, P. W. *et al.* Prediction of coronary heart disease using risk factor categories. *Circulation* **97**, 1837–1847 (1998).
63. Lassenius, M. I. *et al.* Cardiovascular event rates increase after each recurrence and associate with poor statin adherence. *European journal of preventive cardiology* **28**, (2021).
64. Laukkanen, J. A., Kurl, S., Salonen, R., Rauramaa, R. & Salonen, J. T. The predictive

value of cardiorespiratory fitness for cardiovascular events in men with various risk profiles: a prospective population-based cohort study. *Eur Heart J* **25**, 1428–1437 (2004).

65. Mbatchou, J. *et al.* Computationally efficient whole-genome regression for quantitative and binary traits. *Nat Genet* **53**, 1097–1103 (2021).
66. Krefl, D. & Bergmann, S. Cross-GWAS coherence test at the gene and pathway level. *PLoS Comput Biol* **18**, e1010517 (2022).
67. Lamparter, D., Marbach, D., Rueedi, R., Kutalik, Z. & Bergmann, S. Fast and Rigorous Computation of Gene and Pathway Scores from SNP-Based Summary Statistics. *PLOS Computational Biology* **12**, e1004714 (2016).
68. Krefl, D., Brandulas Cammarata, A. & Bergmann, S. PascalX: a Python library for GWAS gene and pathway enrichment tests. *Bioinformatics* **39**, (2023).
69. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* **88**, 76–82 (2011).
70. Yang, J. *et al.* Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics* **42**, 565–569 (2010).
71. Caron, M. *et al.* Emerging properties in self-supervised vision transformers. in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)* (IEEE, 2021). doi:10.1109/iccv48922.2021.00951.
72. Are sixteen heads really better than one? https://dl.acm.org/doi/10.5555/3454287.3455544 doi:10.5555/3454287.3455544.