Supplementary Material (online only)

Schoepf IC, Thorball CW et al.: *CAD-associated and Longevity-associated Genome-wide Polygenic Risk Scores for Prediction of Coronary Artery Disease Events in Swiss Persons Living with HIV*

Supplementary Methods

Genotyping, Quality Control. Single nucleotide polymorphisms (SNPs) were mapped to the human genome build GRCh37, with the correct strand orientation ensured using BCFTOOLS (v1.8). SNPs were removed from the dataset if they had a larger than 20% minor allele frequency (MAF) deviation from the 1000 Genomes Phase 3 EUR reference panel. The remaining genotypes were phased with EAGLE2[3] and missing genotypes were imputed using positional Burrows-Wheeler transform (PBWT)[2] at the Sanger Imputation Service,[1] using the 1000 Genomes Project Phase 3 panel as reference. Following imputation, only high-quality SNPs with an imputation information score (INFO > 0.8) were retained, after which separate genotyping batches were combined. Calculation of population structure and principal components was performed with EIGENSTRAT (v6.1.4)[4], together with the HapMap3 reference panel.[5] Only individuals clustering with the European HapMap3 samples were kept for subsequent analyses. The cohort was furthermore screened with KING (v2.1.3)[6] to verify that no duplicate or cryptic related samples were included. Finally, SNPs and samples with excessive missingness (above 10%), low MAF (below 1%) or excessive deviation from Hardy-Weinberg Equilibrium (PHWE < 1e-6) were removed prior to generating polygenic risk scores.[7]

Calculation of the meta-PRS (i.e. CAD-PRS plus longevity-PRS). We calculated the meta-PRS following the same principles described previously by Inouye *et al*,[8] thus obtaining a weighted average of the standardized PRS scores for CAD and longevity, while accounting for the correlation between the scores due to potential overlapping genetic signals caused by shared genetic loci.

$$metaPRS_i = \frac{\beta_1 Z_{i1} + \beta_2 Z_{i2}}{\sqrt{\beta_1^2 + \beta_2^2 + 2\beta_1\beta_2\rho_{1,2}}}$$

Prior to combination, the scores were standardized as z-scores, with Z_{i1} representing the CAD-PRS and Z_{i2} the longevity-PRS for the *i*th individual. β_1 and β_2 represents the obtained odds ratio per standard deviation from logistic regressions with the CAD- and longevity-PRS, respectively.

Supplementary Results

CAD Variability Explained by Traditional and HIV-related Risk Factors. CAD variability explained by traditional risk factors was 15.03% and 14.17% (continuous model vs. model with quintiles of traditional risk). CAD Variability explained by HIV-associated risk factors was 16.94% and 18.16% (continuous model vs. model with quintiles of HIV-associated risk). The combination of traditional and HIV-associated risk factors ("clinical risk") resulted in CAD variability explained being 30.54% and 30.86% (continuous model vs. model that included quintiles of clinical risk).

Supplementary Table 1: Single nucleotide polymorphisms included in the longevity-PRS

rsID	Gene	effect_allele	reference_allele	Beta_GWAS	SE_GWAS
rs7412	*APOE*	C	T	0.2452	0.0367
rs6859	*NECTIN2*	G	A	-0.1124	0.02
rs429358	*APOE*	T	C	0.5098	0.0322
rs405509	*APOE*	G	T	-0.1299	0.0199

Supplementary Table 2: Coronary Artery Disease (CAD) Odds Ratio (OR) According to Clinical Risk Factors. Univariable and Multivariable Analysis.

	Univariable analysis	Multivariable analysis
Never smoking	(reference)	(reference)
Current smoking, n (%)	OR 1.67; 95% CI 1.11–2.52; p= 0.02	OR 1.57; 95% CI 0.94–2.63; p= 0.08
Past smoking, n (%)	OR 1.32; 95% CI 0.86–2.04; p= 0.20	OR 0.85; 95% CI 0.50–1.46; p= 0.56
Family History of CAD, n (%)	OR 1.54; 95% CI 1.01–2.34; p= 0.05	OR 1.43; 95% CI 0.83–2.49; p= 0.20
Diabetes mellitus, n (%)	OR 3.30; 95% CI 1.99–5.45; p<0.01	OR 5.01; 95% CI 2.64–9.51; p<0.01
Hypertension, n (%)	OR 1.10; 95% CI 0.78–1.55; p= 0.59	OR 0.90; 95% CI 0.59–1.38; p= 0.62
Dyslipidemia, n (%)	OR 2.17; 95% CI 1.59–2.96; p<0.01	OR 1.83; 95% CI 1.25–2.66; p<0.01
Age at matching date (per year older) (IQR)	OR 1.28; 95% CI 1.16–1.41; p<0.01	OR 1.33; 95% CI 1.18–1.51; p<0.01
Ever cocaine use, n (%)	OR 1.09; 95% CI 0.62–1.93; p= 0.75	
HIV RNA <50 copies/mL at matching date	(reference)	(reference)
HIV RNA ≥50 copies/mL at matching date, n (%)	OR 0.74; 95% CI 0.48–1.15; p= 0.19	OR 0.94; 95% CI 0.54–1.64; p= 0.84
CD4 nadir (cells/µL), median (IQR)	OR 1.94; 95% CI 1.27–2.97; p<0.01	OR 0.97; 95% CI 0.56–1.69; p= 0.91
Currently on Abacavir, n (%)	OR 1.82; 95% CI 1.30–2.55; p<0.01	OR 1.98; 95% CI 1.28–3.08; p<0.01
Lopinavir/ritonavir, exposure ≥1 year, n (%)	OR 2.08; 95% CI 1.42–3.05; p<0.01	OR 1.86; 95% CI 0.99–2.61; p= 0.05
Indinavir, exposure ≥1 year, n (%)	OR 3.19; 95% CI 2.05–4.96; p<0.01	OR 2.23; 95% CI 1.31–3.78; p<0.01
Darunavir, exposure ≥1 year, n (%)	OR 2.05; 95% CI 1.31–3.21; p<0.01	OR 1.93; 95% CI 1.11–3.36; p= 0.02

Stavudine, exposure ≥1 year, n (%)	OR 4.89; 95% CI 3.28–7.28; p<0.01	OR 3.82; 95% CI 2.40–6.08; p<0.01
Hepatitis C Seropositivity, n (%)	OR 1.51; 95% CI 1.03–2.20; p= 0.03	OR 0.99; 95% CI 0.50–1.97; p= 0.98
CMV Seropositivity, n (%)	OR 1.70; 95% CI 1.10–2.61; p= 0.02	OR 1.78; 95% CI 1.05–3.04; p= 0.03
IDU	OR 1.41; 95% CI 0.92–2.14; p= 0.11	OR 1.86; 95% CI 0.84–4.13; p= 0.13

Abbreviations. CAD, coronary artery disease; CMV, cytomegalovirus; IDU, intravenous drug use; IQR, interquartile range;

51

References

1. McCarthy S, Das S, Kretzschmar W, et al. A reference panel of 64,976 haplotypes for genotype imputation. Nat. Genet. **2016**; 48:1279–1283.

2. Durbin R. Efficient haplotype matching and storage using the positional Burrows-Wheeler transform (PBWT). Bioinformatics. **2014**; 30:1266–1272.

3. Loh PR, Danecek P, Palamara PF, et al. Reference-based phasing using the Haplotype Reference Consortium panel. Nat. Genet. **2016**; 48:1443–1448.

4. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. **2006**; 38:904–909.

5. The International HapMap 3 Consortium, Altshuler DM, Peltonen L, et al. Integrating common and rare genetic variation in diverse human populations. Nature. **2010**; 467:52–58.

6. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen W-M. Robust relationship inference in genome-wide association studies. Bioinformatics. **2010**; 26:2867–2873.

7. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. **2015**; 4:7.

8. Inouye M, Abraham G, Nelson CP, et al. Genomic Risk Prediction of Coronary Artery Disease in 480,000 Adults: Implications for Primary Prevention. J Am Coll Cardiol. 2018; 72(16):1883-1893.