Supplementary (online-only) material

Dietrich LG et al. *Rapid Progression of Kidney Dysfunction in Swiss People Living with HIV: Contribution of Polygenic Risk Score and D:A:D Clinical Risk Score.*

Supplementary Methods.

Genotyping, Quality Control. First, all variants were flipped to their correct strand orientation using the human genome build GRCh37 and BCFTOOLS (v1.8). Variants with a larger than 20% minor allele frequency deviation from the 1000 Genomes Phase 3 EUR reference panel were removed. The filtered genotypes were then imputed with the Sanger Imputation Service[1] and the 1000 Genomes Project Phase 3 reference panel, using PBWT[2] as well as EAGLE2[3] for phasing. Prior to merger of all samples, only variants with a high imputation quality (INFO > 0.8) were retained.

Principal component analysis (**Supplementary Figure 1**) was performed on the merged cohort with EIGENSTRAT (v6.1.4)[4] together with the HapMap project[5], to include only individuals of European ancestries for subsequent analyses. The cohort was furthermore screened using KING (v2.1.3)[6] to ensure that no cryptic related or duplicate samples were included. The combined cohort was then prepared for analyses by removing variants with an excessive missingness (>10%), low minor allele frequency (<1%) or a large deviation from Hardy-Weinberg Equilibrium (P_{HWE} < 1e-6) with PLINK (v2.00a2.3)[7].

Exploratory GWAS. The imputed and filtered genotypes were tested for association with the case/control definition of rapid progression of kidney dysfunction described in the statistical analyses paragraph. The standard genome-wide significance threshold of P < 5e-8 was set to infer statistical significance. The GWAS was performed with PLINK (v2.00a2.3)[7], using logistic regression with age, sex and the first 5 principal components as covariates. The genomic inflation factor, lambda, was 0.99, indicating an absence of either systemic inflation or deflation of the association results. With

our sample size we had 80% power to detect variants with a minor allele frequency of 10% and an odds ratio of 2.23[8].

Replication of variants found previously associated with baseline eGFR values was performed by first clumping the summary statistics using an LD threshold of r^2=0.1, 1Mb windows and using our cohort as LD reference to extract all genome-wide significant variants also found or tagged in our cohort. A variant was considered replicated if it had a nominally significant p-value ($P < 0.05$) as well as an identical effect direction. With this significance threshold, we had 80% power to detect previously associated variants with a minor allele frequency of 10% and an odds ratio of 1.47[8].

Supplementary Results.

Supplementary Table 1. Hazard Ratio for Rapid Progression, Multivariable Analysis *allowing for interaction between Polygenic Risk Score and D:A:D CKD Risk Score*

	Hazard Ratio	**95% Confidence Interval**
Polygenic risk score, 1st Quartile	reference	
Polygenic risk score, 2nd Quartile	1.47	0.83-2.62
Polygenic risk score, 3rd Quartile	1.63	0.93-2.87
Polygenic risk score, 4th Quartile	1.71	0.97-3.02
D:A:D CKD Risk Score, Low risk	reference	
D:A:D CKD Risk Score, Medium risk	1.51	0.79-2.87
D:A:D CKD Risk Score, High risk	1.07	0.39-2.89
2nd Quartile*Medium risk	0.54	0.22-1.32
2nd Quartile*High risk	2.43	0.74-7.97
3rd Quartile*Medium risk	0.76	0.32-1.74
3rd Quartile*High risk	1.12	0.32-3.96
4th Quartile*Medium risk	1.15	0.52-2.58
4th Quartile*High risk	2.17	0.67-6.97

Supplementary Table 2. Multivariable model selection according to Akaike's information criterion and Bayesian information criterion

	Akaike Information Criterion	**Bayesian Information Criterion**
Polygenic Risk Score alone	3579.2	3598.3
D:A:D CKD Risk Score alone	3581.5	3594.2
Without interaction	3572.7	3604.6
With interaction	3575	3645.1

Supplementary Table 3. Hazard Ratios (95% confidence interval) for Rapid Progression according to D:A:D CKD Risk Score, Polygenic Risk Score, and Antiretroviral Therapy: Main study population *plus 2 sensitivity analyses*

	Main Analysis: **SHCS participants with baseline eGFR >80 mL/min/1.73m^2 (n=4345)** *(NB: these are the numbers that are shown in **Fig. 3**)*		**Sensitivity Analysis:** **SHCS participants with baseline eGFR >70 mL/min/1.73m^2 (n=4700)**		**Sensitivity Analysis,** **SHCS participants with baseline eGFR >90 mL/min/1.73m^2 (n=3423)**	
Variable	**Univariable analysis**	**Multivariable analysis**	**Univariable analysis**	**Multivariable analysis**	**Univariable analysis**	**Multivariable analysis**
D:A:D CKD Risk score medium risk vs. low risk	**1.26** **(0.96-1.65)**	**1.30** **(0.99-1.71)**	1.16 (0.88-1.52)	1.19 (0.90-1.57)	1.39 (1.07-1.82)	1.44 (1.10-1.89)
D:A:D CKD Risk score high risk vs. low risk	**2.21** **(1.42-3.43)**	**1.82** **(1.28-2.60)**	1.39 (1.00-1.95)	1.50 (1.07-2.10)	2.12 (1.02-4.40)	2.24 (1.07-4.70)
Polygenic risk score 2nd quartile vs. 1st quartile	**1.35** **(0.91-1.99)**	**1.39** **(0.94-2.06)**	1.25 (0.86-1.83)	1.28 (0.87-1.86)	1.30 (0.85-1.97)	1.35 (0.88-2.05)
Polygenic risk score 3rd quartile vs. 1st quartile	**1.46** **(0.99-2.16)**	**1.52** **(1.04-2.24)**	1.36 (0.94-1.98)	1.40 (0.96-2.03)	1.56 (1.04-2.35)	1.64 (1.09-2.46)
Polygenic risk score 4th quartile vs. 1st quartile	**1.96** **(1.36-2.83)**	**2.04** **(1.41-2.94)**	1.77 (1.24-2.53)	1.83 (1.28-2.61)	2.04 (1.38-3.00)	2.13 (1.45-3.15)
Recent TDF exposure	**1.37** **(1.06-1.76)**	**1.36** **(1.06-1.76)**	1.04 (0.88-1.24)	1.37 (1.06-1.76)	1.26 (1.02-1.55)	1.40 (1.07-1.84)
Recent ATV/r exposure	**1.36** **(0.99-1.87)**	**1.30** **(0.95-1.79)**	0.99 (0.78-1.25)	1.36 (0.99-1.87)	1.10 (0.85-1.43)	1.30 (0.93-1.82)

Recent LPV/r exposure	**1.06 (0.71-1.59)**	**1.11 (0.74-1.67)**	0.94 (0.75-1.18)	1.05 (0.69-1.60)	1.21 (0.95-1.53)	1.17 (0.76-1.78)

51

52 **Supplementary Table 4. Rapid Progression of Kidney Dysfunction according to D:A:D CKD**
53 **Risk Score, Polygenic Risk Score, and Antiretroviral Therapy (Hazard Ratios (95%**
54 **confidence interval):**

55 ***Bivariable and Trivariable Models including tertiles of baseline eGFR***

56

	Bivariable Analysis, eGFR Tertiles plus D:A:D CKD Risk Score	**Bivariable Analysis, eGFR Tertiles plus Polygenic Risk Score**	**Trivariable Analysis, eGFR Tertiles plus D:A:D CKD Risk Score plus Polygenic Risk Score**
Baseline eGFR			
1st tertile	0.77 (0.53-1.1)	1.05 (0.77-1.43)	0.76 (0.53-1.09)
2nd tertile	Reference	Reference	Reference
3rd tertile	1.04 (0.77-1.41)	1.01 (0.75-1.37)	1.07 (0.79-1.44)
D:A:D medium vs. low risk	1.39 (1.04-1.85)	--	1.39 (1.04-1.85)
D:A:D high vs. low risk	2.31 (1.48-3.61)	--	2.31 (1.48-3.61)
Polygenic risk score 2nd vs. 1st quartile	--	1.42 (0.96-2.11)	1.42 (0.96-2.11)
Polygenic risk score 3rd vs. 1st quartile	--	1.54 (1.04-2.27)	1.54 (1.04-2.27)
Polygenic risk score 4th vs. 1st quartile	--	2.05 (1.42-2.96)	2.05 (1.42-2.96)

57

Recent LPV/r exposure	**1.06 (0.71-1.59)**	**1.11 (0.74-1.67)**	0.94 (0.75-1.18)	1.05 (0.69-1.60)	1.21 (0.95-1.53)	1.17 (0.76-1.78)

51

52 **Supplementary Table 4. Rapid Progression of Kidney Dysfunction according to D:A:D CKD**
53 **Risk Score, Polygenic Risk Score, and Antiretroviral Therapy (Hazard Ratios (95%**
54 **confidence interval):**

55 ***Bivariable and Trivariable Models including tertiles of baseline eGFR***

56

	Bivariable Analysis, eGFR Tertiles plus D:A:D CKD Risk Score	**Bivariable Analysis, eGFR Tertiles plus Polygenic Risk Score**	**Trivariable Analysis, eGFR Tertiles plus D:A:D CKD Risk Score plus Polygenic Risk Score**
Baseline eGFR			
1st tertile	0.77 (0.53-1.1)	1.05 (0.77-1.43)	0.76 (0.53-1.09)
2nd tertile	Reference	Reference	Reference
3rd tertile	1.04 (0.77-1.41)	1.01 (0.75-1.37)	1.07 (0.79-1.44)
D:A:D medium vs. low risk	1.39 (1.04-1.85)	--	1.39 (1.04-1.85)
D:A:D high vs. low risk	2.31 (1.48-3.61)	--	2.31 (1.48-3.61)
Polygenic risk score 2nd vs. 1st quartile	--	1.42 (0.96-2.11)	1.42 (0.96-2.11)
Polygenic risk score 3rd vs. 1st quartile	--	1.54 (1.04-2.27)	1.54 (1.04-2.27)
Polygenic risk score 4th vs. 1st quartile	--	2.05 (1.42-2.96)	2.05 (1.42-2.96)

57

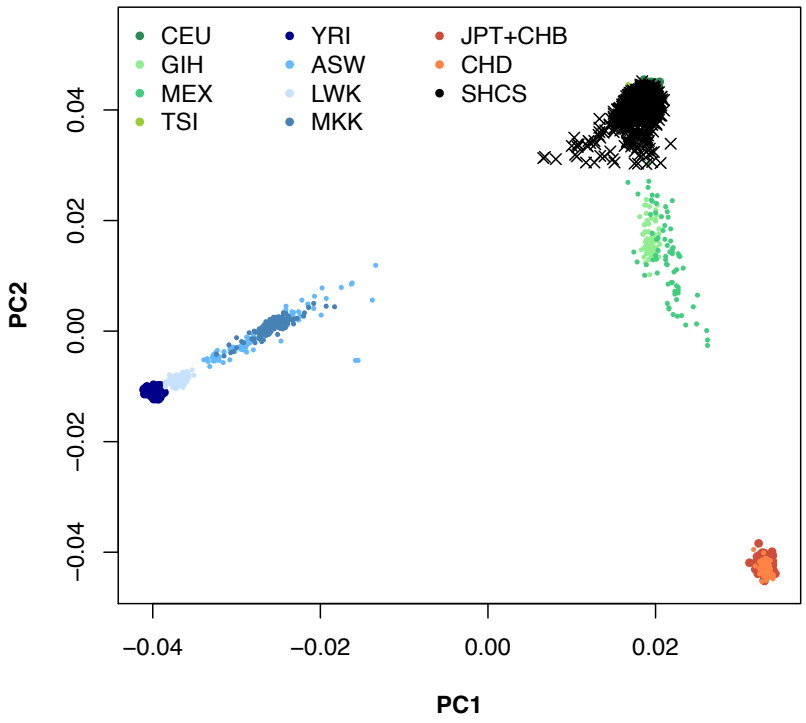
58

59

60

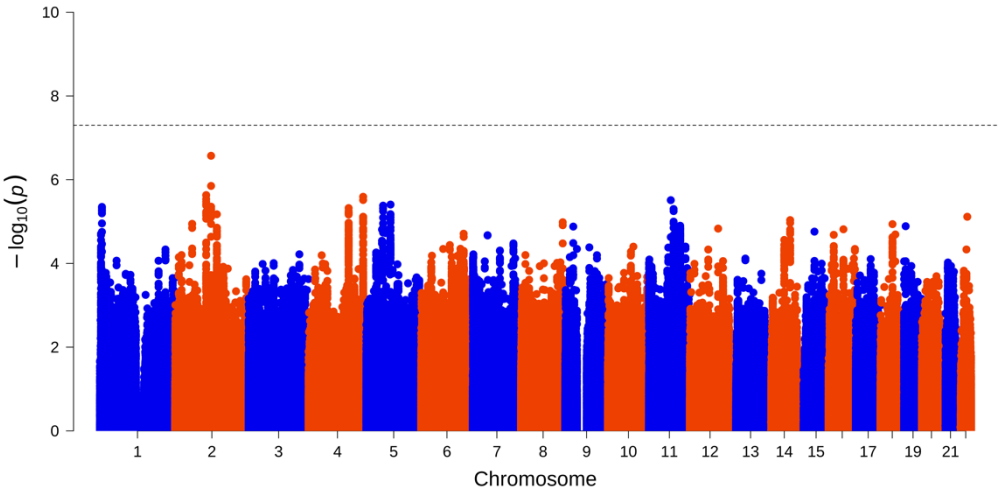
Supplementary Figure 1. Principal Component Analysis.

Principal component analysis (PCA) of the included patients in the study with population references from the HapMap project. The included samples from the SHCS (black crosses) all colocalized with the European HapMap reference samples from CEU (Northern Europeans from Utah, USA) and TSI (Tuscans from Italy). Other included reference populations include Gujarati Indians in Houston, Texas (GHI), Mexicans from Los Angeles (MEX), Yoruba from Nigeria (YRI), Africans from Southwest USA (ASW), Luhya from Webuye, Kenya (LWK), Maasai from Kinyawa, Kenya (MKK), Japanese from Tokyo, Japan (JPT) plus Han Chinese from Beijing, China (CHB) and Chinese from Denver, Colorado (CHD).



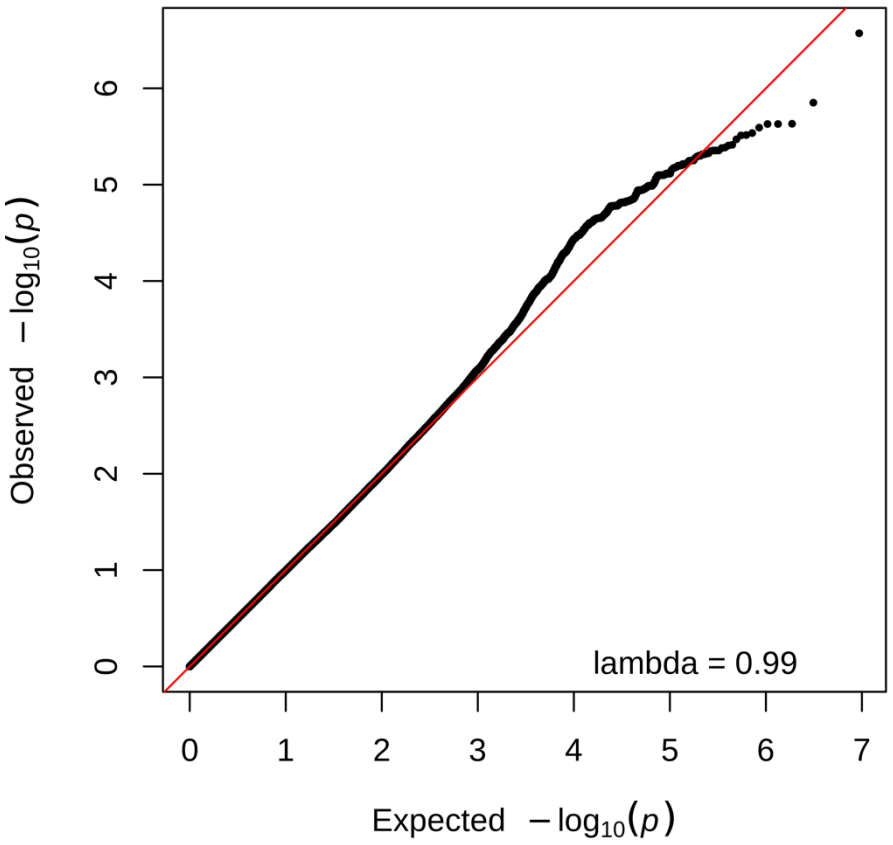
Supplementary Figure 2. Exploratory Genome-Wide Association Study.

Manhattan plot with association p-values (-log10(P)) per genetic variant, plotted by genomic position. The threshold for genome-wide significance (P = 5e-8) is indicated by the dashed line. No genetic variants for found to be genome-wide significant in the logistic regression.



Supplementary Figure 3. Exploratory Genome-Wide Association Study.

Quantile-Quantile Plot for the exploratory GWAS of rapid progression of kidney dysfunction. The plot shows the observed -log10(p-values) (black dots, y-axis) versus expected -log10(p-values) under the null hypothesis (red line). The genomic inflation factor, lambda, is indicated on the plot. Values ~1 implies the absence of genomic inflation due to confounding factors.



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 P-R, 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. Johnson JL, Abecasis GR. GAS Power Calculator: web-based power calculator for genetic association studies. bioRxiv **2017**; doi.org/10.1101/164343.