nature medicine

ARTICLES

https://doi.org/10.1038/s41591-022-01869-1

Genome-wide polygenic score to predict chronic kidney disease across ancestries

Atlas Khan[1], Michael C. Turchin[2], Amit Patki[3], Vinodh Srinivasasainagendra[3], Ning Shang[1], Rajiv Nadukuru[4], Alana C. Jones[5], Edyta Malolepsza[6], Ozan Dikilitas[7], Iftikhar J. Kullo[7], Daniel J. Schaid[7], Elizabeth Karlson[8], Tian Ge[9], James B. Meigs[10,11,12], Jordan W. Smoller[9], Christoph Lange[13], David R. Crosslin[14], Gail P. Jarvik[15], Pavan K. Bhatraju[16], Jacklyn N. Hellwege[17], Paulette Chandler[18], Laura Rasmussen Torvik[19], Alex Fedotov[20], Cong Liu[21], Christopher Kachulis[6], Niall Lennon[6], Noura S. Abul-Husn[2,22,23], Judy H. Cho[4], Iuliana Ionita-Laza[24], Ali G. Gharavi[1], Wendy K. Chung[25], George Hripcsak[21], Chunhua Weng[21], Girish Nadkarni[4], Marguerite R. Irvin[5], Hemant K. Tiwari[3], Eimear E. Kenny[2,22,26], Nita A. Limdi[27] and Krzysztof Kiryluk[1]✉

Chronic kidney disease (CKD) is a common complex condition associated with high morbidity and mortality. Polygenic prediction could enhance CKD screening and prevention; however, this approach has not been optimized for ancestrally diverse populations. By combining *APOL1* risk genotypes with genome-wide association studies (GWAS) of kidney function, we designed, optimized and validated a genome-wide polygenic score (GPS) for CKD. The new GPS was tested in 15 independent cohorts, including 3 cohorts of European ancestry ($n = 97{,}050$), 6 cohorts of African ancestry ($n = 14{,}544$), 4 cohorts of Asian ancestry ($n = 8{,}625$) and 2 admixed Latinx cohorts ($n = 3{,}625$). We demonstrated score transferability with reproducible performance across all tested cohorts. The top 2% of the GPS was associated with nearly threefold increased risk of CKD across ancestries. In African ancestry cohorts, the *APOL1* risk genotype and polygenic component of the GPS had additive effects on the risk of CKD.

CKD affects 10–16% of the general population and has high morbidity and mortality[1,2]. In the USA, CKD disproportionally affects African Americans (16.3%) when compared to European Americans (12.7%), Asian Americans (12.9%) or Hispanic Americans (13.6%)[3]. CKD stage 3 or greater is defined by a chronic loss of glomerular filtration rate (GFR) to below 60 ml min 1.73 m^{-2}. Because this definition is based on estimated kidney function rather than markers of specific kidney injury, it captures an etiologically heterogeneous set of primary and secondary kidney disorders. As expected for a highly heterogeneous trait, CKD has a complex determination with both genetic and environmental contributions. The observational heritability of CKD in the largest analysis of medical records ranged from 25 to 44%[4]. These estimates were generally consistent with smaller family-based studies of CKD and GFR[5–7].

High heritability of CKD is attributed to both monogenic[8,9] and polygenic causes[10,11]. Moreover, in individuals of African ancestry, two common risk alleles (G1 and G2) in the *APOL1* gene convey a large effect on the risk of kidney disease[12,13]. While heterozygotes for the G1 or G2 alleles appear to be protected from *Trypanosoma* sleeping sickness, kidney disease risk is conveyed under a recessive model in carriers of two risk alleles (G1G1, G2G2 or G1G2). Because of the selective pressure exerted by endemic trypanosomal species in certain parts of eastern and western Africa, G1 and

[1]Division of Nephrology, Department of Medicine, Vagelos College of Physicians & Surgeons, Columbia University, New York, NY, USA. [2]Institute for Genomic Health, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [3]Department of Biostatistics, School of Public Health, University of Alabama at Birmingham, Birmingham, AL, USA. [4]The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [5]Department of Epidemiology, School of Public Health, University of Alabama at Birmingham, Birmingham, AL, USA. [6]Broad Institute of MIT and Harvard, Cambridge, MA, USA. [7]Department of Cardiovascular Diseases, Mayo Clinic, Rochester, MN, USA. [8]Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA. [9]Department of Psychiatry, Massachusetts General Hospital, Boston, MA, USA. [10]Division of General Internal Medicine, Department of Medicine, Massachusetts General Hospital, Boston, MA, USA. [11]Harvard Medical School, Boston, MA, USA. [12]Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, USA. [13]Department of Biostatistics, Harvard T. H. Chan School of Public Health, Boston, MA, USA. [14]Division of Biomedical Informatics and Genomics, John W. Deming Department of Medicine, Tulane University School of Medicine, New Orleans, LA, USA. [15]Departments of Medicine (Medical Genetics) and Genome Sciences, University of Washington, Seattle, WA, USA. [16]Division of Pulmonary, Critical Care and Sleep Medicine, Department of Medicine, University of Washington, Seattle, WA, USA. [17]Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA. [18]Division of Preventive Medicine, Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA. [19]Department of Preventive Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA. [20]Irving Institute for Clinical and Translational Research, Columbia University, New York, NY, USA. [21]Department of Biomedical Informatics, Vagelos College of Physicians & Surgeons, Columbia University, New York, NY, USA. [22]Division of Genomic Medicine, Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [23]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [24]Department of Biostatistics, Columbia University, New York, NY, USA. [25]Department of Pediatrics, Columbia University Irving Medical Center, New York, NY, USA. [26]Division of General Internal Medicine, Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [27]Department of Neurology, School of Medicine, University of Alabama at Birmingham, Birmingham, AL, USA. ✉e-mail: kk473@columbia.edu

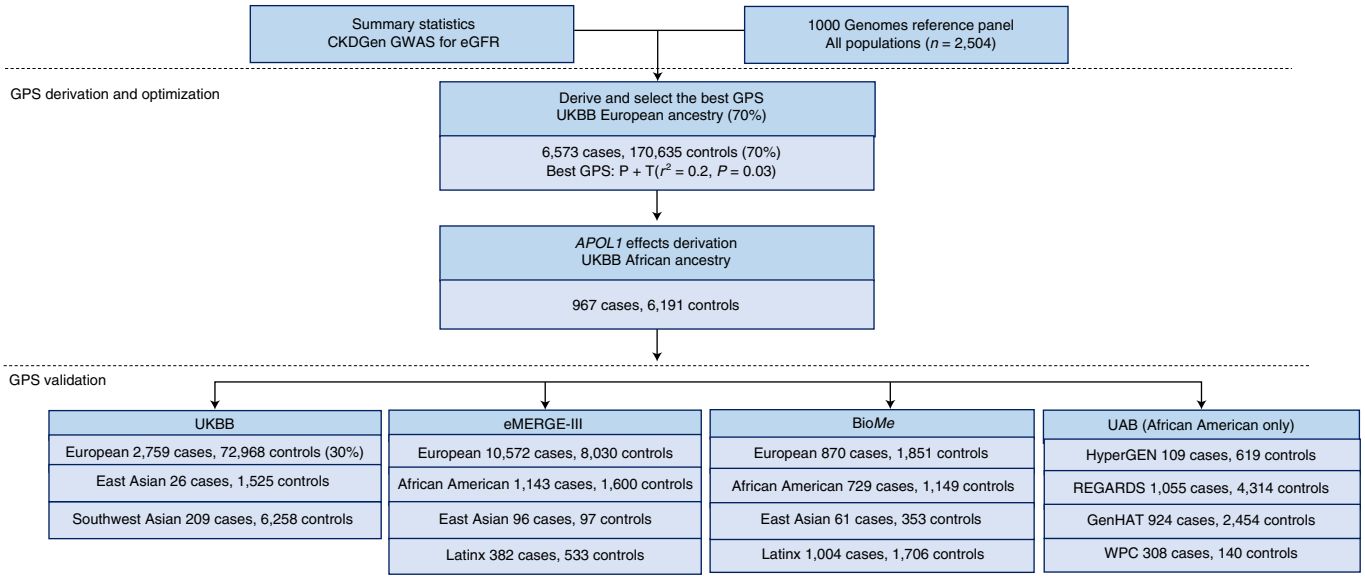


Fig. 1 | Overview of the study design. The CKD GPS was designed based on the CKDGen GWAS summary statistics for eGFR and a cosmopolitan LD reference panel of 1000 Genomes (all populations); optimization was performed in two stages using UKBB participants of European (optimization 1) and African (optimization 2) ancestries. GPS performance validation was conducted in 15 additional independent testing cohorts of diverse ancestries.

G2 alleles are observed almost exclusively in individuals whose ancestry can be linked to those areas[14,15]. In the US population, the frequency of *APOL1* risk genotypes is estimated at approximately 15% in African Americans, 0.5–2% in Hispanic Americans and <0.01% in Europeans[16]. These differences may be contributing to the higher prevalence of CKD in African Americans; however, additional non-*APOL1* genetic risk factors have not yet been elucidated.

GPS have emerged as promising tools for genetic risk stratification that can enhance traditional risk models for complex diseases. This approach has been applied to a variety of traits, including heart disease[17,18], diabetes[18,19], hypertension[20,21], obesity[22], schizophrenia[23–25] and cancers[26–31]. One of the major limitations of the GPS approach is that existing GWAS are based predominantly on European cohorts; thus, most GPS do not perform well in more diverse cohorts or in individuals with admixed ancestry[32]. Similar to other complex traits, GWAS for kidney function involved predominantly European cohorts. The latest study involved 765,348 participants, 75% of which were European, 23% East Asian, 2% African American and <1% Hispanic[11]. Notably, this study did not capture the effects of *APOL1* risk variants because of their recessive inheritance and very low frequencies in non-African populations.

The objective of the present study was to test if the existing knowledge on polygenic contributions to kidney function is sufficient to build a clinical risk predictor for moderate-to-advanced CKD with adequate performance across diverse ancestral groups. We specifically aimed to design, optimize and test a new GPS for clinical risk prediction of kidney disease that maximizes performance across ancestries. We combined information on *APOL1* risk genotypes with the latest GWAS for kidney function to formulate a GPS that can reliably discriminate moderate-to-advanced CKD (stage 3 or greater) from population controls. In our approach, we took advantage of the power of the existing GWAS for a quantitative biomarker of kidney function (serum creatinine-based estimated GFR (eGFR)) to predict a disease state. To demonstrate transferability across different ancestral groups, we performed rigorous testing of our GPS in 15 independent and ancestrally diverse case-control cohorts according to ClinGen standards[33].

Results

GPS optimization. The flowchart summary of our overall strategy is provided in Fig. 1. In the GPS optimization step, a total of 19 candidate scores were generated using 1000 Genomes (all populations) linkage disequilibrium (LD) reference and summary statistics from GWAS for eGFR[11]. We then used a large optimization dataset consisting of 70% of European UK Biobank (UKBB) participants to select the best performing model (Table 1 and Supplementary Table 1). The best model was based on the *P* value thresholding (P + T) method and involved 41,426 markers with nonzero weights selected based on $r^2 = 0.2$ and $P = 0.03$. The score was standardized to zero mean and unit variance based on ancestry-matched population controls. In the optimization dataset, the polygenic component of the score explained 4% of the variance (R^2), with 1 s.d. of the score increasing CKD risk by 86% (odds ratio (OR) = 1.86, 95% confidence interval (CI) = 1.83–1.89, $P < 1.00 \times 10^{-300}$) after controlling for age, sex, diabetes, center, genotyping/imputation batch and genetic ancestry (Supplementary Table 1).

The second optimization step involved testing for independent contributions of *APOL1* risk genotypes and included 7,158 UKBB participants of genetically defined African ancestry (967 cases and 6,191 controls). In the model adjusted for age, sex, diabetes, batch and principal components (PCs) of ancestry, we observed statistically significant independent effects of the polygenic component (OR per s.d. = 1.16, 95% CI = 1.09–1.25, $P = 1.00 \times 10^{-4}$) and the recessive *APOL1* risk genotype (OR = 1.19, 95% CI = 1.01–1.38, $P = 4.00 \times 10^{-2}$), but no significant multiplicative interactions between the two predictors (*P* interaction = 0.29) (Supplementary Table 2). Given these findings, we subsequently modeled *APOL1* risk as additive to the polygenic component, assuming that the *APOL1* risk genotype effects are approximately equivalent to 1 s.d. of the standard-normalized polygenic risk score (PRS) (the weight of 1 was used because the β per s.d. of the PRS and the β for the *APOL1* risk genotype were comparable in magnitude).

Population differences in GPS distributions. We next examined the distributions of the polygenic risk component (without *APOL1*) and the final combined GPS (with *APOL1*) in the reference

Table 1 | Summary of study cohorts used for GPS optimization and testing

Study	Sub-cohort	CKD cases ($n = 27,787$)	Controls ($n = 280,423$)	Female (%)	Diabetes (%)	Mean age (years)
UKBB	Optimization 1: European ancestry (70%)	6,573	170,635	54	5	56.65
	Optimization 2: African ancestry	967	6,191	58	12	51.77
	Testing 1: European ancestry (30%)	2,759	72,968	54	5	56.64
	Testing 2: East Asian ancestry	26	1,525	68	5	52.37
	Testing 3: South Asian ancestry	209	6,258	46	18	53.32
eMERGE	Testing 4: European ancestry	10,572	8,030	52	35	71.23
	Testing 5: African ancestry	1,143	1,600	70	40	66.76
	Testing 6: Latinx admixed ancestry	382	533	64	38	66.77
	Testing 7: East Asian ancestry	96	97	59	27	72.81
UAB	Testing 8: WPC: African ancestry	308	140	58	50	61.48
	Testing 9: REGARDS: African ancestry	1,055	4,314	63	31	62.30
	Testing 10: GenHAT: African ancestry	924	2,454	58	45	65.74
	Testing 11: HyperGEN: African ancestry	109	619	62	31	52.23
Bio*Me*	Testing 12: European ancestry	870	1,851	38	14	61.87
	Testing 13: African ancestry	729	1,149	56	32	61.65
	Testing 14: Latinx admixed ancestry	1,004	1,706	38	33	62.04
	Testing 15: East Asian ancestry	61	353	41	15	55.75

UAB, University of Alabama.

populations of 1000 Genomes. We detected significant differences in the mean polygenic risk across populations (Fig. 2; analysis of variance (ANOVA) $P = 3.40 \times 10^{-154}$), with a notable shift toward higher average risk in African ancestry compared to all other populations ($P = 4.92 \times 10^{-163}$). This shift was even more pronounced after including *APOL1* risk genotype information in the combined GPS ($P = 1.58 \times 10^{-168}$). These results suggest that the PRS for CKD is considerably higher in African compared to non-African populations independent of *APOL1*.

Given that the weights of the score equation are fixed, we hypothesized that the observed distributional differences were driven by a higher frequency of CKD risk alleles in African genomes. Therefore, we examined the overall frequency spectrum of the CKD risk alleles included in the GPS between European and African reference populations (Extended Data Fig. 1). First, we observed a greater number of risk alleles in the African compared to European populations of 1000 Genomes at the extremes of the frequency spectrum (risk allele frequencies (RAFs) < 0.01 or > 0.99). This observation was expected due to the minor allele frequency (MAF) filter of 1% used in the European GWAS discovery cohorts. Second, across all variants included in the score, the mean difference in RAF between African and European populations was positive (that is, greater than the expected mean of 0), indicating a higher average frequency of risk alleles in African genomes. We further observed that the risk alleles with the largest weights (effect sizes in the GWAS) had a significantly higher frequency in African genomes compared to those with low ($P = 0.025$) or intermediate effect sizes ($P = 0.018$) (Extended Data Fig. 1d). Thus, it appears that the observed GPS distributional shifts between European and African populations are driven predominantly by frequency differences of large effect risk alleles.

GPS testing in cohorts of European ancestry. We next tested the final GPS in 3 European cohorts, including the remaining 30% of the UKBB (2,759 cases and 72,968 controls) and 2 large US-based European ancestry cohort—Electronic Medical Records and Genomics Phase III (eMERGE-III) (10,572 cases and 8,030 controls) and Bio*Me* (870 cases and 1,851 controls). In the combined meta-analysis, the GPS exhibited highly reproducible performance, with pooled OR per s.d. = 1.46, 95% CI = 1.43–1.48, $P < 1.00 \times 10^{-300}$ (Supplementary Table 3). While the UKBB testing cohort had nearly identical performance metrics to the optimization cohort, the effect sizes were attenuated slightly in the US-based cohorts. The frequency of the *APOL1* risk genotype was extremely low (Supplementary Table 4); thus, its effect was negligible in the European cohorts.

GPS testing in cohorts of African ancestry. The GPS was tested in 6 independent African ancestry cohorts, including eMERGE-III (1,143 cases and 1,600 controls), Bio*Me* (729 cases and 1,149 controls), Hypertension Genetic Epidemiology Network (HyperGEN) (109 cases and 619 controls), REasons for Geographic and Racial Differences in Stroke (REGARDS) (1,055 cases and 4,314 controls), Genetics of Hypertension Associated Treatments (GenHAT) (924 cases and 2,454 controls) and Warfarin Pharmacogenomics Cohort (WPC) (308 cases and 140 controls). In the combined meta-analysis, the GPS had pooled an OR per s.d. of 1.32, 95% CI = 1.26–1.38, $P = 1.78 \times 10^{-33}$ (Table 2 and Supplementary Table 5). The inclusion of the *APOL1* risk genotype considerably enhanced CKD risk prediction across all African ancestry cohorts, substantially improving tail cutoff discrimination, that is, the risk for the top 2% of individuals was approximately 1.80-fold higher in the model without *APOL1* and 2.70-fold higher in the model with *APOL1* compared to the remaining 98% of individuals (Table 3). The effects of the GPS stratified by *APOL1* risk genotype across all six African ancestry cohorts are depicted in Fig. 3a.

GPS testing in admixed ancestry Latinx cohorts. The GPS was also tested in 2 admixed ancestry Latinx cohorts from eMERGE-III (382 cases and 533 controls) and Bio*Me* (1,004 cases and 1,706 controls). The combined meta-analysis of these cohorts resulted in a pooled OR per s.d. = 1.42, 95% CI = 1.29–1.57, $P = 4.56 \times 10^{-12}$ (Supplementary Table 6). Similar to African ancestry cohorts, the inclusion of *APOL1* risk genotypes in the GPS improved risk prediction in these admixed cohorts (Table 3).

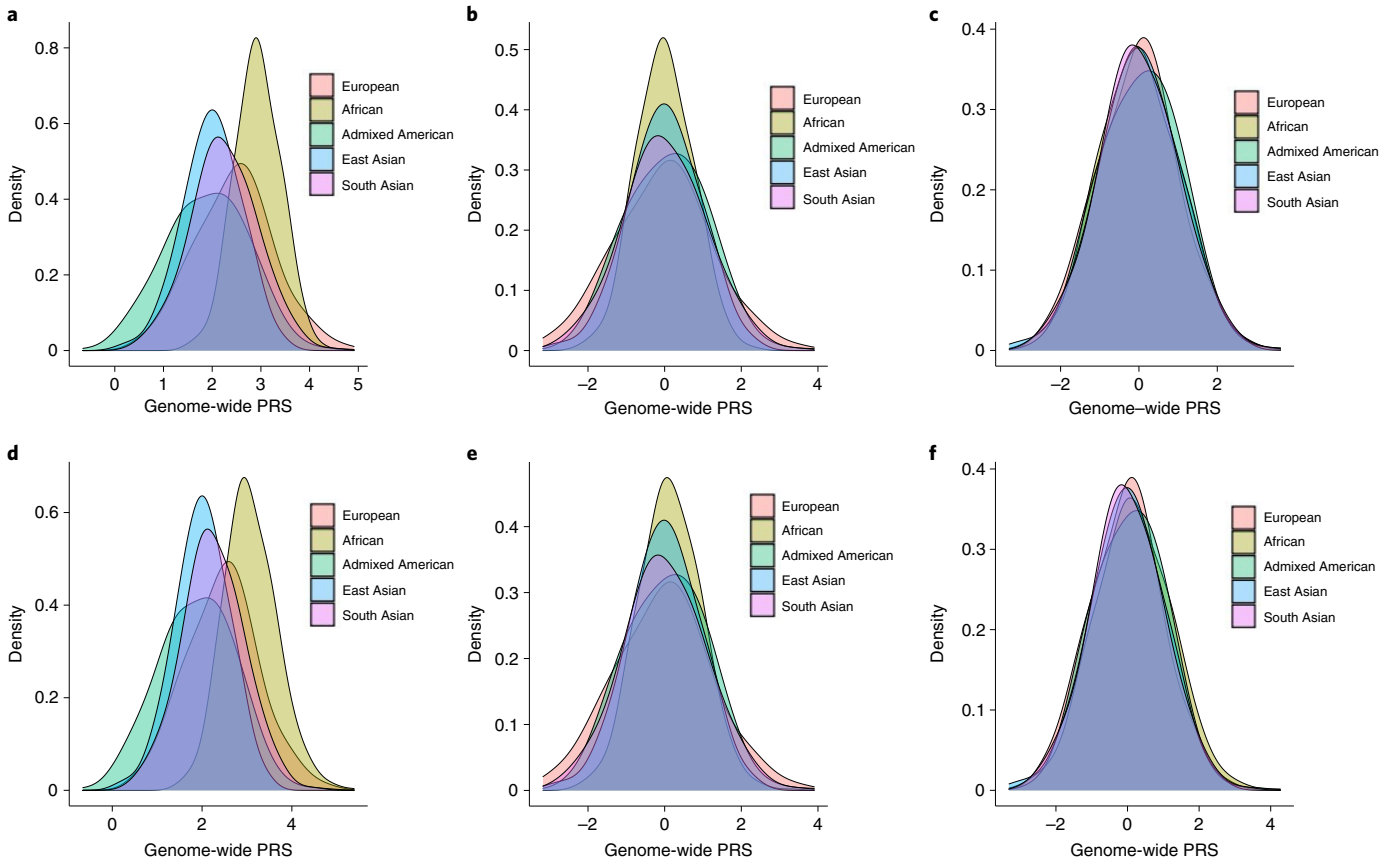


Fig. 2 | Risk score distributions in five 1000 Genomes populations. a, Raw PRS without *APOL1*. **b**, Ancestry-adjusted PRS without *APOL1* (method 1: mean only). **c**, Ancestry-adjusted PRS without *APOL1* (method 2: mean and variance). **d**, Raw combined GPS with *APOL1*. **e**, Ancestry-adjusted combined GPS with *APOL1* (method 1). **f**, Ancestry-adjusted combined GPS with *APOL1* (method 2).

GPS testing in cohorts of Asian ancestry. We tested GPS in 4 diverse Asian cohorts including UKBB southwest Asian (209 cases and 6,258 controls), UKBB East Asian (26 cases and 1,525 controls), eMERGE-III East Asian (96 cases and 97 controls) and Bio*Me* East Asian (61 cases and 353 controls) cohorts. The combined meta-analysis resulted in a pooled OR per s.d. = 1.68, 95% CI = 1.45–2.06, $P = 7.11 \times 10^{-13}$ (Supplementary Table 7). *APOL1* risk genotypes were absent in the Asian cohorts; thus, the modeled risk was entirely attributable to the polygenic component.

Tail discrimination performance by ancestry. For each individual testing cohort, we derived risk estimates comparing extreme tails of the risk score distribution to all other cohort members and estimated sensitivity and specificity for a range of tail cutoffs (20, 10, 5, 2 and 1%). These metrics were meta-analyzed by ancestry and are summarized in Table 2, Fig. 3b and Supplementary Tables 3 and 5–7. Depending on ancestry, the top 2% tail of the risk score distribution was associated with 2.66–4.93-fold higher risk of CKD than for the remaining 98% of individuals, including in European (OR = 3.60, 95% CI = 3.11–4.17, $P = 4.26 \times 10^{-66}$), African (OR = 2.66, 95% CI = 2.01–3.51, $P = 4.93 \times 10^{-12}$), admixed Latinx (OR = 4.93, 95% CI = 2.46–9.89, $P = 6.69 \times 10^{-6}$) and Asian ancestry (OR = 3.81, 95% CI = 1.91–7.59, $P = 1.35 \times 10^{-4}$) cohorts. We consider this cutoff as clinically meaningful because this degree of risk is approximately equivalent to the risk reported for a family history of kidney disease[34]. In Supplementary Table 8, we summarize various metrics of diagnostic performance for this cutoff by ancestry, including sensitivity, specificity, and prevalence-adjusted positive and negative predictive values. For comparison, we provide similar metrics for the top 5% of the risk score distribution.

Ancestry adjustments and calibration. We next compared the effect of two different ancestry adjustment methods on the GPS distributions in the 1000 Genomes, eMERGE-III and UKBB testing cohorts (Fig. 2 **and** Extended Data Fig. 2). Adjusting for mean only (method 1; Methods) eliminated major distributional shifts by ancestry, but did not fully resolve the observed tail differences. The ancestry adjustment method 2 (adjusting for both mean and variance) resulted in comparable shapes of the GPS distributions by ancestry and facilitated the selection of a single trans-ancestry tail cutoff. Both methods resulted in comparably good risk score calibration when applied to the combined multiethnic eMERGE-III dataset (Extended Data Fig. 3). However, as a trade-off, the ancestry adjustments reduced tail discrimination at extreme cutoffs as summarized in Supplementary Table 9. This trade-off appeared most pronounced for method 2, as well as for more admixed cohorts (African American and Latinx).

Sensitivity analyses. The use of race in clinical predictive models has been scrutinized and a new Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) 2021 equation without a race variable has recently been proposed[35]. Therefore, we performed sensitivity analyses to examine the effect of the new equation on GPS performance in eMERGE-III, our largest and most diverse testing dataset. The CKD-EPI 2021 equation without race[35] was applied to redefine cases and controls across all eMERGE-III participants. As expected,

Table 2 | The performance metrics of the GPS in the testing cohorts meta-analyzed by ancestry

Meta-analysis	Cases/controls	OR per s.d. (95% CI), *P*	AUC (crude)	PRS threshold	OR (95% CI), *P*
European (3 cohorts)	14,201/82,849	1.46 (1.43–1.48), $P<1.00\times10^{-300}$	0.81 (0.62)	Top 20% versus other 80%	2.30 (2.17–2.44), $P=1.65\times10^{-174}$
				Top 10% versus other 90%	2.59 (2.40–2.78), $P=1.27\times10^{-142}$
				Top 5% versus other 95%	2.92 (2.65–3.21), $P=2.64\times10^{-104}$
				Top 2% versus other 98%	3.60 (3.11–4.17), $P=4.26\times10^{-66}$
				Top 1% versus other 99%	4.46 (3.66–5.44), $P=7.82\times10^{-50}$
African (6 cohorts)	4,268/10,276	1.32 (1.26–1.38), $P=1.78\times10^{-33}$	0.78 (0.57)	Top 20% versus other 80%	1.65 (1.49–1.82), $P=1.17\times10^{-22}$
				Top 10% versus other 90%	1.84 (1.61–2.09), $P=9.26\times10^{-20}$
				Top 5% versus other 95%	2.06 (1.72–2.47), $P=2.11\times10^{-15}$
				Top 2% versus other 98%	2.66 (2.01–3.51), $P=4.93\times10^{-12}$
				Top 1% versus other 99%	3.51 (2.37–5.22), $P=4.21\times10^{-10}$
Latinx (2 cohorts)	1,386/2,239	1.42 (1.29–1.57), $P=4.56\times10^{-12}$	0.88 (0.62)	Top 20% versus other 80%	1.88 (1.50–2.37), $P=5.46\times10^{-8}$
				Top 10% versus other 90%	2.26 (1.66–3.06), $P=1.56\times10^{-7}$
				Top 5% versus other 95%	2.67 (1.75–4.07), $P=4.96\times10^{-6}$
				Top 2% versus other 98%	4.93 (2.46–9.89), $P=6.69\times10^{-6}$
				Top 1% versus other 99%	6.61 (2.46–17.75), $P=1.77\times10^{-4}$
Asian (4 cohorts)	392/8,233	1.68 (1.45–2.06), $P=7.11\times10^{-13}$	0.91 (0.61)	Top 20% versus other 80%	2.42 (1.81–2.27), $P=4.39\times10^{-9}$
				Top 10% versus other 90%	2.95 (2.06–4.20), $P=2.43\times10^{-9}$
				Top 5% versus other 95%	3.56 (2.26–5.60), $P=4.09\times10^{-8}$
				Top 2% versus other 98%	3.81 (1.91–7.59), $P=1.35\times10^{-4}$
				Top 1% versus other 99%	8.46 (3.70–19.3), $P=4.00\times10^{-7}$
All 15 cohorts	20,247/103,597	1.44 (1.42–1.47), $P<1.00\times10^{-300}$	0.81 (0.61)	Top 20% versus other 80%	2.23 (2.11–2.35), $P=8.02\times10^{-195}$
				Top 10% versus other 90%	2.31 (2.17–2.45), $P=2.75\times10^{-168}$
				Top 5% versus other 95%	2.58 (2.39–2.79), $P=2.02\times10^{-123}$
				Top 2% versus other 98%	3.26 (2.89–3.67), $P=3.37\times10^{-84}$
				Top 1% versus other 99%	4.61 (3.84–5.53), $P=1.14\times10^{-60}$

For performance testing in individual cohorts, please refer to Supplementary Tables 4–7. The OR for the model was adjusted for age, sex, diabetes, PCs of ancestry and genotyping array or clinical site. The s.d. represents the s.d. of the GPS distribution in controls. The AUC for the model was adjusted for age, sex, diabetes, PCs of ancestry and genotyping array or clinical site. Crude signifies the AUC for GPS alone without any covariates.

the new equation reduced the number of CKD cases (and increased the numbers of controls) in the European, Latinx and East Asian cohorts, while the opposite effect was observed in the African ancestry cohort. Importantly, across all ancestral groups, GPS had comparable performance between the new and old phenotype definitions, with a notable trend for improved performance in the African ancestry cohort (Supplementary Table 10a).

We additionally tested the effect of using self-reported race/ethnicity versus genetic (PC-based) ancestry to define our testing cohorts. Despite smaller sample sizes of the cohorts defined by self-report, we observed no major performance differences between self-report and genetic ancestry-defined cohorts. Similarly, the use of the 2021 CKD-EPI eGFR equation to define CKD in self-report-based cohorts resulted in comparable performance to the 2009 equation, with a similar trend for improved performance in African ancestry cohorts (Supplementary Table 10b).

Lastly, using our multiethnic eMERGE-III cohorts, we compared the performance of our GPS to an alternative score recently proposed by Yu et al.[36] (Supplementary Table 11). This score was based on a GWAS with a higher proportion of Europeans, did not include the *APOL1* risk genotype and was not optimized for trans-ethnic performance. Given these differences, the score by Yu et al.[36] performed better in cohorts of European ancestry but was less predictive in cohorts of African or Asian ancestry. Notably, the African ancestry distribution of the GPS by Yu et al.[36] was also shifted toward higher values compared to other ancestries, confirming that the observed shift is independent of a specific method used to design the score (Extended Data Fig. 4).

Discussion

We developed a GPS for CKD that captures polygenic determinants of kidney function emerging from recent GWAS studies and predicts CKD across four ancestry groups. Our score was designed and validated for individual-level risk prediction according to the ClinGen guidelines[33]. The score had consistent performance despite heterogenous genotyping platforms and imputation methods employed in our testing studies. We also developed continuous ancestry adjustment methods to allow for cross-ancestry standardization of the score. Importantly, our testing studies demonstrated that extreme tails of the risk score distribution (top 2%) conveyed approximately a threefold increase in the disease risk across all ancestries. This magnitude of risk is equivalent to a positive family history of kidney disease[34].

In this study, we were unable to assess if the GPS conveys kidney disease risk independently of positive family history, mainly because family history is poorly captured in electronic health records and thus it is not routinely available for large electronic health record-linked biobanks[37]. However, prior cohort studies in

Table 3 | Added value *APOL1* risk genotype to PRS components in predicting CKD using the extreme tail (98th percentile) of the risk score distribution in African American (4,268 cases and 10,276 controls) and admixed Latinx (1,386 cases and 2,239 controls) cohorts

Cohorts	*APOL1* risk genotype OR (95% CI), *P*	Top 2% PRS without *APOL1* OR (95% CI), *P*	Top 2% PRS with *APOL1* OR (95% CI), *P*
African ancestry			
eMERGE	1.64 (1.42–1.86), $P=2.00\times10^{-5}$	2.10 (1.46–2.74), $P=2.00\times10^{-2}$	2.60 (1.38–4.90), $P=3.10\times10^{-3}$
Bio*Me*	1.38 (1.28–1.48), $P=3.30\times10^{-10}$	2.70 (1.93–3.47), $P=1.00\times10^{-2}$	5.75 (4.96–6.54), $P=1.00\times10^{-5}$
UAB HyperGEN	1.71 (0.93–3.12), $P=8.20\times10^{-2}$	2.22 (0.59–8.44), $P=2.40\times10^{-1}$	1.64 (0.43–6.20), $P=4.65\times10^{-1}$
UAB REGARDS	1.35 (1.08–1.77), $P=6.90\times10^{-3}$	1.26 (0.76–2.07), $P=3.60\times10^{-1}$	1.56 (0.97–2.59), $P=6.52\times10^{-2}$
UAB GenHAT	1.43 (1.12–1.81), $P=3.20\times10^{-3}$	2.80 (1.64–4.77), $P=1.50\times10^{-4}$	4.38 (2.56–7.50), $P=6.80\times10^{-8}$
UAB WPC	1.93 (1.07–3.49), $P=2.90\times10^{-2}$	-	1.59 (0.28–8.73), $P=5.96\times10^{-1}$
Meta-analysis	1.46 (1.38–1.54), $P=2.70\times10^{-19}$	1.76 (1.41–2.20), $P=5.90\times10^{-7}$	2.66 (2.01–3.51), $P=4.93\times10^{-12}$
Latinx admixed ancestry			
eMERGE	16.5 (5.70–48.1), $P=7.10\times10^{-4}$	1.41 (0.47–4.24), $P=5.40\times10^{-1}$	6.89 (1.60–29.07), $P=9.78\times10^{-3}$
Bio*Me*	1.17 (1.10–1.24), $P=2.40\times10^{-6}$	2.72 (1.97–3.47), $P=1.00\times10^{-2}$	4.48 (3.69–5.27), $P=2.10\times10^{-4}$
Meta-analysis	1.18 (1.09–1.27), $P=4.40\times10^{-5}$	2.21 (1.19–4.10), $P=1.10\times10^{-2}$	4.93 (2.46–9.89), $P=6.69\times10^{-6}$

All effect estimates were adjusted for age, sex, diabetes and PCs of ancestry.

cardiovascular disease[38–41] and breast cancer[42] clearly demonstrate that PRS for these conditions capture independent information from family history or traditional risk factors. For CKD, this specific question will be addressed by a prospective eMERGE-IV study testing the performance of our score against family history collected using MeTree software[43]. Additional studies are also needed to test if our GPS modifies penetrance of monogenic kidney disorders, similar to the effects reported for tier 1 genetic disorders[44].

Beyond enhanced disease screening of asymptomatic individuals, other potential applications of the GPS may include improved risk stratification of potential living kidney donors or enhanced quality assessment of deceased donor organs in the setting of kidney transplant[45]. The hypothesis that a donor polygenic risk is relevant to kidney allograft outcomes is to be tested but is supported by the fact that most candidate causal genes from GWAS for kidney function map to kidney cell types[46,47]. The clinical practice and local guidelines for genetic screening of living donors continue to evolve rapidly, but presently only monogenic causes of kidney disease and *APOL1* risk genotypes are being considered[48]. Our results stress an urgent need to test the utility donor GPS in this setting to better assess its impact on living donation risks as well as allograft outcomes.

Although our score is based on a multiethnic GWAS for eGFR, the allelic effect estimates are heavily biased by the predominant European-Asian composition of the discovery GWAS that included 75% European, 23% East Asian, 2% African and only <1% admixed ancestry Latinx participants. Because there are currently no studies of similar size performed in African and admixed ancestry participants, we were unable to improve the accuracy of effect estimates for these populations and our model assumed fixed effects across ancestries. We used a diverse LD reference to improve the trans-ethnic performance of the score and we further enhanced the model by including African ancestry-specific recessive *APOL1* risk genotypes known to have large effects on the risk of kidney disease. We demonstrated that *APOL1* risk genotypes (coded under a recessive model) have an additive effect with the polygenic component and significantly enhance case-control discrimination in African and admixed ancestry Latinx cohorts.

Several important limitations of this work need to be discussed. First, we are most limited by the lack of large-scale GWAS for kidney function in non-European populations, as well as the small sizes of existing cohorts that could be used for performance optimization in non-Europeans[49]. Thus, the largest cohorts presently available for robust risk score optimization are also of predominantly European ancestry. The assumption of fixed allelic effects across different ancestral groups is likely inaccurate because many disease-related lifestyle factors and environmental exposures correlated with ancestry could modify allelic effects. Accordingly, the overall tail discrimination of the score was lower in African than in European or Asian ancestry cohorts with notably lower sensitivity for the top 2% GPS cutoff. Although it is not possible to overcome this limitation in the present study, our GPS approach could be refined by including larger non-European GWAS studies for eGFR or CKD once available in the future.

Second, the performance comparisons between different ancestral groups could be biased by differences in genotyping platforms and the ascertainment methods employed by various biobanks. For example, the UKBB represents a population-based cohort recruiting European participants in the 40–60 age group, while the eMERGE and Bio*Me* case-control cohorts are ascertained in older patients with more comorbidities receiving care at tertiary US medical centers. The inclusion of older participants in our testing cohorts might lead to some case misclassification due to age-related kidney function decline. To mitigate this issue, we excluded CKD stage 2 from all performance tests in our study. However, stage 3a may also be prone to residual misclassification[50], resulting in risk underestimation for cohorts consisting of older participants.

Third, the ancestry definitions varied in our testing cohorts. While in eMERGE and UKBB ancestry was defined agnostically using genetic approaches, all other testing cohorts relied on self-report. Despite these differences, the risk score performance was similar across all cohorts and our sensitivity analyses demonstrated no major change in performance when cohorts were defined by genetics versus self-report. Notably, the African ancestry cohorts included in this study were predominantly of West African descent. Due to the lack of relevant genetic cohorts, we were unable to test GPS performance in other African populations, such as from East or South Africa. These populations have a relatively low frequency of *APOL1* risk alleles, which could potentially dampen the score effects; however, follow-up studies are needed for these populations. Additional GPS validation is also needed in Native American and Pacific Islander populations that were not represented in this study.

Fourth, by design, our score models polygenic effects from GWAS for kidney function as approximated by eGFR from serum creatinine (filtration marker) rather than CKD itself. We recognize

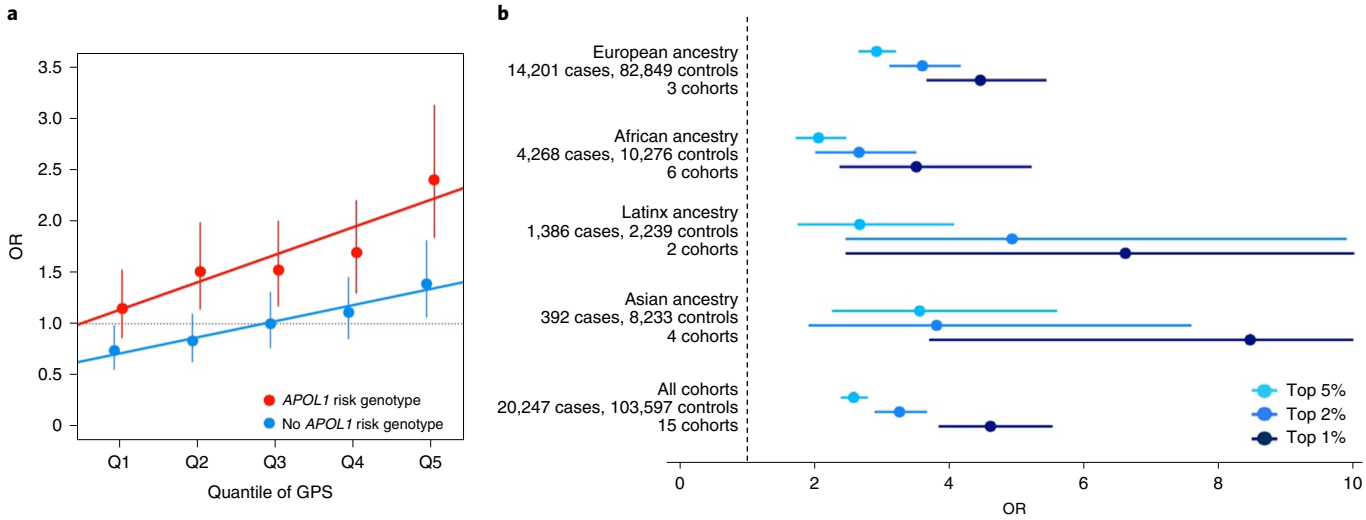


Fig. 3 | Effects of the GPS for CKD. a, **GPS quantile effects stratified by the *APOL1* risk genotype ($n = 2{,}020$ with and $n = 12{,}526$ without the *APOL1* risk genotype, shown in red and blue, respectively)**. The *x* axis depicts each quantile of the GPS ordered from the first (Q1) to the last (Q5) quantile. The *y* axis depicts the ORs of CKD for each of the quantile-defined subgroups in reference to the middle quantile (Q3) of those without the *APOL1* risk genotype. The effect estimates (dots) and 95% CIs (vertical bars) were derived based on a fixed-effects meta-analysis across all 6 African ancestry testing cohorts and adjusted for age, sex, diabetes and PCs of ancestry. Regression lines were fitted for each group defined by the presence of the *APOL1* risk genotype. **b**, GPS tail effects by ancestry. The *x* axis depicts the OR of CKD, the *y* axis depicts the testing cohort meta-analysis by ancestry (with numbers of cases, controls and cohorts). The effect estimates (dots) and 95% CIs (horizontal bars) are provided for the top 5% versus bottom 95% (sky blue), top 2% versus bottom 98% (cobalt blue) and top 1% versus bottom 99% (navy blue). All effect estimates are adjusted for age, sex, diabetes and PCs of ancestry.

multiple limitations to the use of eGFR as a phenotype in GWAS, including the fact that serum creatinine level is influenced by the rate of creatinine production and metabolism in addition to kidney clearance. Accordingly, to capture a clinically meaningful disease state, we designed the score to predict moderately advanced CKD (stage 3 and above) rather than a mild degree of kidney dysfunction. Notably, our risk score does not incorporate available information on the polygenic determination of albuminuria[51] or primary kidney diseases[52,53]. However, current GWAS for these traits remain several orders of magnitude smaller in sample size compared to the GWAS for eGFR; thus, incorporation of such data must await more powerful studies.

Fifth, we observed significant differences in the mean and variance of the GPS distributions by ancestry. The observed shift in the mean GPS toward higher values in individuals of African ancestry is independent of *APOL1* and is driven by a higher average RAF in the African genomes. Interpopulation RAF differences are greatest for the risk alleles with the largest effects. This pattern may be consistent with polygenic adaptation but the effects of uncorrected population stratification in the discovery GWAS may also potentially explain this phenomenon[54]. Therefore, based on this observation alone we were unable to determine if the observed shift contributes to the higher prevalence of CKD in individuals of African ancestry.

We note that the observed differences in the GPS distributions by ancestry represent a significant challenge for the clinical implementation of PRS. The key problem is that it is not possible to select a single GPS threshold for all ancestries that results in the similar magnitude of risk. Therefore, we explored several approaches that could be used to overcome this issue. One approach involves classifying individuals undergoing GPS testing into one of the four ancestry groups based on self-report, then using ancestry-specific cutoffs to interpret the results. However, because of a potential for racial bias, the use of race in clinical algorithms has been discouraged.[55] One could also classify a tested individual based on genetic ancestry inferred from single-nucleotide polymorphism (SNP) data with subsequent application of ancestry-specific cutoffs. This approach still categorizes individuals into distinct groups and can be inaccurate especially for those individuals with admixed genomes. Therefore, we tested two different regression-based ancestry correction methods that model a continuous spectrum of genetic ancestry based on the diverse reference panel of the 1000 Genomes. We demonstrated that the reference population-based correction for both mean and variance can best align distribution tails for selection of a single trans-ancestry GPS cutoff. However, this results in some performance trade-offs, especially in admixed populations. Although still imperfect, this ancestry adjustment may be helpful in improving risk score standardization for clinical use in diverse populations.

Lastly, we used the 2009 CKD-EPI equation (with a race coefficient) because no alternatives were available at the time of our analyses. We do not expect this equation to affect GPS performance because our analyses were stratified by genetic ancestry and the race coefficient was uniformly applied to all African ancestry cohorts. Importantly, our sensitivity analyses confirmed a comparable performance of the GPS when the case-control status was redefined using the newly proposed 2021 CKD-EPI equation without a race variable[35,56].

In summary, we derived, optimized and tested a new GPS for CKD across major ancestries and proposed new methods for its trans-ethnic GPS standardization. We demonstrated that the polygenic component and *APOL1* risk genotypes had additive effects on the risk of CKD. Our study showed that individuals in the highest 2% of the risk score distribution had nearly a threefold increase in disease risk, the degree of risk equivalent to a positive family history. The key advantage of the GPS over traditional screening is that it can identify at-risk individuals before the onset of any disease manifestations. Timely communication of high-risk status may lead to adoption of protective lifestyle changes and improved adoption of the recommended screening guidelines. Because the cost of SNP arrays is no longer prohibitive, and multiple PRS can be determined using a single array, a population-based genetic screening approach for common diseases (for example, individuals older than 40) may prove to represent a cost-effective public health strategy. While our

study represents only the first step in this direction, prospective studies are needed to test the clinical utility and cost-effectiveness of this approach. The prospective eMERGE-IV study is specifically designed to test this strategy in a newly recruited population-based cohort of over 20,000 volunteers.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41591-022-01869-1.

Received: 9 November 2021; Accepted: 11 May 2022;
Published online: 16 June 2022

References

1. Coresh, J. et al. Prevalence of chronic kidney disease in the United States. *JAMA* **298**, 2038–2047 (2007).
2. Naghavi, M. et al. Global, regional, and national age-sex specific mortality for 264 causes of death, 1980–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet* **390**, 1151–1210 (2017).
3. *Chronic Kidney Disease in the United States* (Centers for Disease Control and Prevention, 2022); https://www.cdc.gov/kidneydisease/publications-resources/ckd-national-facts.html
4. Shang, N. et al. Medical records-based chronic kidney disease phenotype for clinical care and "big data" observational and genetic studies. *NPJ Digit. Med.* **4**, 70 (2021).
5. Fox, C. S. et al. Genomewide linkage analysis to serum creatinine, GFR, and creatinine clearance in a community-based population: the Framingham Heart Study. *J. Am. Soc. Nephrol.* **15**, 2457–2461 (2004).
6. Langefeld, C. D. et al. Heritability of GFR and albuminuria in Caucasians with type 2 diabetes mellitus. *Am. J. Kidney Dis.* **43**, 796–800 (2004).
7. Satko, S. G. & Freedman, B. I. The familial clustering of renal disease and related phenotypes. *Med. Clin. North. Am.* **89**, 447–456 (2005).
8. Groopman, E. E. et al. Diagnostic utility of exome sequencing for kidney disease. *N. Engl. J. Med.* **380**, 142–151 (2019).
9. Lata, S. Whole-exome sequencing in adults with chronic kidney disease: a pilot study. *Ann. Intern. Med.* **168**, 100–109 (2018).
10. Köttgen, A. et al. Multiple loci associated with indices of renal function and chronic kidney disease. *Nat. Genet.* **41**, 712–717 (2009).
11. Wuttke, M. et al. A catalog of genetic loci associated with kidney function from analyses of a million individuals. *Nat. Genet.* **51**, 957–972 (2019).
12. Genovese, G. et al. Association of trypanolytic ApoL1 variants with kidney disease in African Americans. *Science* **329**, 841–845 (2010).
13. Parsa, A. et al. *APOL1* risk variants, race, and progression of chronic kidney disease. *N. Engl. J. Med.* **369**, 2183–2196 (2013).
14. Thomson, R. et al. Evolution of the primate trypanolytic factor APOL1. *Proc. Natl Acad. Sci. USA* **111**, E2130–E2139 (2014).
15. Ko, W.-Y. et al. Identifying Darwinian selection acting on different human *APOL1* variants among diverse African populations. *Am. J. Hum. Genet.* **93**, 54–66 (2013).
16. Nadkarni, G. N. et al. Worldwide frequencies of *APOL1* renal risk variants. *N. Engl. J. Med.* **379**, 2571–2572 (2018).
17. Gladding, P. A., Legget, M., Fatkin, D., Larsen, P. & Doughty, R. Polygenic risk scores in coronary artery disease and atrial fibrillation. *Heart Lung Circ.* **29**, 634–640 (2020).
18. Khera, A. V. et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat. Genet.* **50**, 1219–1224 (2018).
19. Läll, K., Mägi, R., Morris, A., Metspalu, A. & Fischer, K. Personalized risk prediction for type 2 diabetes: the potential of genetic risk scores. *Genet. Med.* **19**, 322–329 (2017).
20. Hoffmann, T. J. et al. Genome-wide association analyses using electronic health records identify new loci influencing blood pressure variation. *Nat. Genet.* **49**, 54–64 (2017).
21. Ehret, G. B. et al. The genetics of blood pressure regulation and its target organs from association studies in 342,415 individuals. *Nat. Genet.* **48**, 1171–1184 (2016).
22. Khera, A. V. et al. Polygenic prediction of weight and obesity trajectories from birth to adulthood. *Cell* **177**, 587–596.e9 (2019).
23. Weinberger, D. R. Polygenic risk scores in clinical schizophrenia research. *Am. J. Psychiatry* **176**, 3–4 (2019).
24. Reginsson, G. W. et al. Polygenic risk scores for schizophrenia and bipolar disorder associate with addiction. *Addict. Biol.* **23**, 485–492 (2018).
25. Power, R. A. et al. Polygenic risk scores for schizophrenia and bipolar disorder predict creativity. *Nat. Neurosci.* **18**, 953–955 (2015).
26. Aly, M. et al. Polygenic risk score improves prostate cancer risk prediction: results from the Stockholm-1 cohort study. *Eur. Urol.* **60**, 21–28 (2011).
27. Fritsche, L. G. et al. Association of polygenic risk scores for multiple cancers in a phenome-wide study: results from the Michigan Genomics Initiative. *Am. J. Hum. Genet.* **102**, 1048–1061 (2018).
28. Jeon, J. et al. Determining risk of colorectal cancer and starting age of screening based on lifestyle, environmental, and genetic factors. *Gastroenterology* **154**, 2152–2164.e19 (2018).
29. Huyghe, J. R. et al. Discovery of common and rare genetic risk variants for colorectal cancer. *Nat. Genet.* **51**, 76–87 (2019).
30. Mavaddat, N. et al. Polygenic risk scores for prediction of breast cancer and breast cancer subtypes. *Am. J. Hum. Genet.* **104**, 21–34 (2019).
31. Seibert, T. M. et al. Polygenic hazard score to guide screening for aggressive prostate cancer: development and validation in large scale cohorts. *BMJ* **360**, j5757 (2018).
32. Martin, A. R. et al. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat. Genet.* **51**, 584–591 (2019).
33. Wand, H. et al. Improving reporting standards for polygenic scores in risk prediction studies. *Nature* **591**, 211–219 (2021).
34. Zhang, J., Thio, C. H. L., Gansevoort, R. T. & Snieder, H. Familial aggregation of CKD and heritability of kidney biomarkers in the general population: the Lifelines Cohort Study. *Am. J. Kidney Dis.* **77**, 869–878 (2021).
35. Inker, L. A. et al. New creatinine- and cystatin C-based equations to estimate GFR without race. *N. Engl. J. Med.* **385**, 1737–1749 (2021).
36. Yu, Z. et al. Polygenic risk scores for kidney function and their associations with circulating proteome, and incident kidney diseases. *J. Am. Soc. Nephrol.* **32**, 3161–3173 (2021).
37. Polubriaginof, F., Tatonetti, N. P. & Vawdrey, D. K. An assessment of family history information captured in an electronic health record. *AMIA Annu. Symp. Proc.* **2015**, 2035–2042 (2015).
38. Tada, H. et al. Risk prediction by genetic risk scores for coronary heart disease is independent of self-reported family history. *Eur. Heart J.* **37**, 561–567 (2016).
39. Timmerman, N. et al. Family history and polygenic risk of cardiovascular disease: independent factors associated with secondary cardiovascular events in patients undergoing carotid endarterectomy. *Atherosclerosis* **307**, 121–129 (2020).
40. Hindy, G. et al. Genome-wide polygenic score, clinical risk factors, and long-term trajectories of coronary artery disease. *Arterioscler. Thromb. Vasc. Biol.* **40**, 2738–2746 (2020).
41. Inouye, M. et al. Genomic risk prediction of coronary artery disease in 480,000 adults: implications for primary prevention. *J. Am. Coll. Cardiol.* **72**, 1883–1893 (2018).
42. Lee, A. et al. BOADICEA: a comprehensive breast cancer risk prediction model incorporating genetic and nongenetic risk factors. *Genet. Med.* **21**, 1708–1718 (2019).
43. Orlando, L. A. et al. Development and validation of a primary care-based family health history and decision support program (MeTree). *N. C. Med. J.* **74**, 287–296 (2013).
44. Fahed, A. C. et al. Polygenic background modifies penetrance of monogenic variants for tier 1 genomic conditions. *Nat. Commun.* **11**, 3635 (2020).
45. Zanoni, F. & Kiryluk, K. Genetic background and transplantation outcomes: insights from genome-wide association studies. *Curr. Opin. Organ Transpl.* **25**, 35–41 (2020).
46. Hellwege, J. N. et al. Mapping eGFR loci to the renal transcriptome and phenome in the VA Million Veteran Program. *Nat. Commun.* **10**, 3842 (2019).
47. Sheng, X. et al. Mapping the genetic architecture of human traits to cell types in the kidney identifies mechanisms of disease and potential treatments. *Nat. Genet.* **53**, 1322–1333 (2021).
48. Neugut, Y. D., Mohan, S., Gharavi, A. G. & Kiryluk, K. Cases in precision medicine: *APOL1* and genetic testing in the evaluation of chronic kidney disease and potential transplant. *Ann. Intern. Med.* **171**, 659–664 (2019).
49. Sirugo, G., Williams, S. M. & Tishkoff, S. A. The missing diversity in human genetic studies. *Cell* **177**, 26–31 (2019).
50. Delanaye, P. et al. CKD: a call for an age-adapted definition. *J. Am. Soc. Nephrol.* **30**, 1785–1805 (2019).
51. Teumer, A. et al. Genome-wide association meta-analyses and fine-mapping elucidate pathways influencing albuminuria. *Nat. Commun.* **10**, 4130 (2019).
52. Kiryluk, K. et al. Discovery of new risk loci for IgA nephropathy implicates genes involved in immunity against intestinal pathogens. *Nat. Genet.* **46**, 1187–1196 (2014).

53. Xie, J. et al. The genetic architecture of membranous nephropathy and its potential to improve non-invasive diagnosis. *Nat. Commun.* **11**, 1600 (2020).
54. Sohail, M. et al. Polygenic adaptation on height is overestimated due to uncorrected stratification in genome-wide association studies. *Elife* **8**, e39702 (2019).
55. Vyas, D. A., Eisenstein, L. G. & Jones, D. S. Hidden in plain sight—reconsidering the use of race correction in clinical algorithms. *N. Engl. J. Med.* **383**, 874–882 (2020).
56. Delgado, C. et al. A unifying approach for GFR estimation: recommendations of the NKF-ASN Task Force on reassessing the inclusion of race in diagnosing kidney disease. *J. Am. Soc. Nephrol.* **32**, 2994–3015 (2021).

Methods

Ethics statement. The study was approved by the Columbia University institutional review board (IRB) (nos. IRB-AAAQ9205, IRB-AAAT8208 and IRB-AAAS3500). All participating studies were approved by their local IRBs, including all sites contributing human genetic and clinical data to the eMERGE-III consortium. Of note, BioVU operated on an opt-out basis until January 2015 and on an opt-in basis since. The phenotypic data in BioVU are all deidentified and the study was designated 'nonhuman participant' research by the Vanderbilt IRB. All other participants provided written informed consent to participate in genetic studies.

Study cohorts. *eMERGE.* The eMERGE network provides access to electronic health record information linked to GWAS data for 102,138 individuals recruited in 3 phases (eMERGE-I, II and III) across 12 participating medical centers from 2007 to 2019 (54% female, mean age 69 years, 76% European, 15% African American, 6% Latinx and 1% East or southeast Asian by self-report). All individuals were genotyped genome-wide; details on genotyping and quality control analyses have been described previously[4,57]. Briefly, all GWAS datasets were imputed using the multiethnic Haplotype Reference Consortium panel on the Michigan Imputation Server[58]. The imputation was performed in 81 batches. Post-imputation, we included only markers with a MAF $\geq$ 0.01 and $R^2 \geq 0.8$ in $\geq$75% of batches. A total of 7,529,684 variants were retained for the GPS analysis. For PC analysis (PCA), we used FlashPCA[59] on a set of 48,509 common (MAF $\geq$ 0.01) and independent variants (pruned in PLINK with the --indep-pairwise 500 50 0.05 command). The G1 and G2 alleles of *APOL1* were imputed separately using the Trans-Omics for Precision Medicine (TOPMed) imputation server[60]. The allelic frequencies of the G1 and G2 alleles were comparable to previous studies[16], as summarized in Supplementary Table 4. The analyses were performed using a combination of VCFtools v.0.1.13[61] and PLINK v.1.9 (ref. [62]).

UKBB. The UKBB is a prospective cohort based in the UK that enrolled individuals ages 40–69 across the UK from 2006 to 2010[63]. This cohort consisted of 488,377 individuals (54% female, mean age 57 years, 94% European, 2% East or southeast Asian and 2% African ancestry by self-report), genotyped with high-density SNP arrays and linked to electronic health record data. All individuals underwent genotyping with the UKBB Axiom array from Affymetrix and UK BiLEVE Axiom arrays (approximately 825,000 markers). Genotype imputation was carried out using a 1000 Genomes reference panel with the IMPUTE4 software, as described previously[63]. We then applied quality control filters similar to eMERGE-III, retaining 9,233,643 common (MAF $\geq$ 0.01) variants imputed with high confidence ($R^2 \geq 0.8$). For PCA by FlashPCA[59], we used a set of 35,226 variants that were common (MAF $\geq$ 0.01) and pruned them using the --indep-pairwise 500 50 0.05 command in PLINK v.1.9 (ref. [62]). The *APOL1* G1 and G2 alleles were imputed separately using the TOPMed imputation server[60].

BioMe biobank. The Bio*Me* biobank is an electronic health record-linked biorepository that has been enrolling participants nonselectively from across the Mount Sinai Health System in New York between 2007 and 2021. A total of 32,595 Bio*Me* participants were genotyped on the Illumina Global Screening Array (GSA) through a collaboration with the Regeneron Genetics Center and 11,953 on the Illumina Global Diversity Array (GDA) through a collaboration with Sema4. Population groups were determined by self-reported race/ethnicity as published previously, with 32% Hispanic/Latinx, 27% Europeans, 22% African Americans and 2.6% East and southeast Asian participants[64]. We removed participants under 40 years of age and those included in the CKDGen GWAS dataset[11]. We applied standard GWAS quality control analyses, removing participants with genotype missing rates >5% and variants with missing rates >5% or Hardy–Weinberg equilibrium violation per each ancestral group ($P < 1.00 \times 10^{-5}$ and $P < 1.00 \times 10^{-6}$ for the GSA and GDA arrays, respectively). We additionally removed individuals who were cryptically related (second degree or above) or whose genotype-inferred sex did not match the self-reported sex. Imputation (including the G1 and G2 variants in *APOL1*) was performed using the TOPMed Imputation Server with the TOPMed Freeze 8 reference. Post-imputation, variants with quality scores <0.7 were removed. After quality control, there were 9,154 Bio*Me* participants of European ancestry, 7,318 of African ancestry, 11,606 of admixed ancestry Latinx and 843 of East Asian ancestry included in the analysis.

REGARDS study. REGARDS is a population-based, longitudinal study of incident stroke and associated risk factors in over 30,000 adults aged 45 years or older between 2003 and 2007 from all 48 contiguous US states and the District of Columbia[65]. By design, participants were oversampled if they had African American ancestry. Genotyping was performed on 8,916 self-identified African American participants using Illumina MEGA-EX arrays and imputed using the National Heart, Lung, and Blood Institute (NHLBI) TOPMed reference panel (Freeze 8). Participants were excluded if they had call rates <95%, if they were internal duplicates, had sex mismatches or were outliers on PCA (outside of 6 s.d.). After quality control, there were 8,669 participants available for analysis (63% females, average age 62 years, 100% African American by self-report). Over 99% of the variants with a MAF > 1% had an imputation quality of 0.6 or higher; for GPS calculation we retained genotypes with a genotypic probability of 0.9 or higher. The *APOL1* G1 and G2 alleles were genotyped directly using TaqMan SNP Genotyping Assays (Thermo Fisher Scientific).

HyperGEN study. HyperGEN is a cross-sectional, population-based study and a component of the NHLBI Family Blood Pressure Program that was designed to identify genetic risk factors for hypertension and its complications[66]. The cohort was recruited from 1995 to 2000 and consisted of sibships with at least 2 siblings diagnosed with hypertension before age 60, their adult offspring and age-matched controls. The study was subsequently expanded to include additional siblings and offspring to a total of $n = 5{,}000$. African American participants (62% females, average age 52 years) underwent whole-genome sequencing (WGS) through the NHLBI WGS program. To harmonize WGS data with the array-based studies, we compiled a set of non-monomorphic and non-multiallelic SNPs with a MAF > 1% that were overlapping with our array-typed African American cohorts. This yielded a total of 2,204,415 SNPs that were used as fence post markers for imputation using the same TOPMed release 2 reference panel as for the REGARDS, GenHAT and WPC studies. Over 99% of the variants with a MAF > 1% had an imputation quality of 0.6 or higher. Only genotypes with genotypic probability >0.9 were retained for the risk score calculation. The *APOL1* genotypes were called directly from the WGS data. Individuals younger than 40 years were excluded; a total of 1,898 participants who self-identified as African American were retained in the testing cohort.

WPC. The WPC is a prospective cohort of first-time warfarin users aged 19 years or older starting anticoagulation for venous thromboembolism, stroke/transient ischemic attacks, atrial fibrillation, myocardial infarction and/or peripheral arterial disease[67]. The genotype data were generated using Illumina MEGA-EX arrays and 1 M duo arrays for 599 and 297 participants, respectively (58% females, average age 61 years, 100% African American by self-report). Imputation was performed using the TOPMed R2 reference panel (Freeze 8). More than 99% of the imputed variants with a MAF > 1% had an R^2 of 0.6 or higher; genotypes with a genotypic probability of 0.9 or higher were retained for PRS calculation. PCA was performed using EIGENSOFT v.6.1.4 based on 44,137 high-quality directly genotyped (missingness < 5%), common (MAF $\geq$ 5%) and independent ($R^2 < 0.05$) SNPs. The *APOL1* information was obtained from genotypic array data; rs143830837 was used as a proxy for rs71785313 because these SNPs represent the same G2 variant and were recently merged in the dbSNP. For this analysis, only participants aged 40 years or older were included, leaving a total of 448 self-identified African American participants.

GenHAT study. GenHAT is an ancillary study to the Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT)[68]. ALLHAT was a randomized, double-blind, multicenter clinical trial with over 42,000 high-risk individuals with hypertension, aged 55 years or older and with at least 1 additional risk factor for cardiovascular disease. ALLHAT is the largest antihypertensive treatment trial to date and was ethnically diverse, enrolling >15,000 African American participants[69]. Participants were randomized into four groups defined by the class of assigned antihypertensive medication including chlorthalidone, lisinopril, amlodipine and doxazosin. The original GenHAT study ($n = 39{,}114$) evaluated the effect of the interaction between candidate genetic variants and different antihypertensive treatments on the risk of cardiovascular outcomes[68]. In an ancillary study to the original study, genotyping using Illumina MEGA-EX arrays was performed on 7,546 African American participants. Samples with a low call rate (<95%), sex mismatches and outliers in the PCA (>6 s.d.) were excluded, which resulted in 6,919 participants with genotypes available for analysis (58% females, average age 66 years, 100% African American by self-report). Genotype imputation was carried out using the NHLBI TOPMed R2 reference panel (Freeze 8); only genotypes with a genotypic probability of 0.9 or higher were retained. The *APOL1* information was extracted from the genotypic array data, similar to the WPC cohort.

Ancestry definitions. In the UKBB and eMERGE-III datasets, the ancestry subcohorts were defined based on PCA clustering. We grouped all individuals into four major continental ancestry clusters by projecting each sample onto the reference PCs calculated from the 1000 Genomes reference panel[70]. Briefly, we merged our UKBB and eMERGE genotypes with 1000 Genomes data and kept only SNPs in common with 1000 Genomes. The markers were then pruned using PLINK --indep-pairwise 500 50 0.05. The numbers of pruned variants for the UKBB and eMERGE were 35,091 and 43,080, respectively. We then calculated the PCs for 1000 Genomes using FlashPCA and projected each of our samples onto those PCs. Ancestry assignments were then performed by coclustering of the reference populations. Ancestry memberships were verified by visual inspection of PCA plots and projections of self-reported race and ethnicity labels on the genetically defined ancestral clusters (Extended Data Fig. 5). Ancestry in Bio*Me*, REGARDS, HyperGEN, WPC and GenHAT was determined by self-reported race/ethnicity; PCA was subsequently performed for ancestry verification and to exclude outliers.

CKD phenotyping and case-control definitions. For phenotyping, we used the computable CKD phenotype extensively validated by the eMERGE-III network[4]. We defined cases as having CKD stage 3 or above based on a 2009 CKD-EPI[71]

eGFR < 60 ml min 1.73m^{-2} on at least 2 serum creatinine measurements 3 months apart, or patients on chronic dialysis or after a kidney transplant. Controls were defined by an eGFR > 90 ml min 1.73m^{-2} based on the latest serum creatinine in the absence of CKD-related International Classification of Diseases (ICD) codes[4]. The exclusion of individuals with CKD stage 2 (eGFR = 60–90 ml min 1.73m^{-2}) from case-control cohorts aimed to minimize potential case-control misclassification due to age-related decline in eGFR. The CKD definition of eGFR < 60 ml min 1.73m^{-2} is thought to reflect <50% of kidney function in adults, has been associated with increased morbidity and mortality and has been endorsed as a clinically meaningful threshold by the Kidney Disease: Improving Global Outcomes 2012 clinical practice guideline for the evaluation and management of CKD[72]. Only individuals 40 years of age or older were included across all datasets for consistency with the UKBB ascertainment strategy. Additional covariates used in the predictive models included age, sex, diabetes and PCs of ancestry. The diagnosis of diabetes, an established risk factor for kidney failure, was defined based on ICD codes as previously published[18]. The diagnosis of hypertension was not added as a covariate to avoid over-adjustment because the cause–effect relationship of hypertension to CKD is difficult to establish based on electronic health record data and CKD itself represents the most common cause of secondary hypertension.

PRS design and optimization. We used 70% Europeans of the UKBB (6,573 cases and 170,635 controls) to optimize the polygenic component of the GPS. Optimization was performed by selecting the best model between two commonly used methods and a range of input parameters (Table 1 and Supplementary Table 1). We used the summary statistics for 8.2 million SNPs from the CKDGen consortium GWAS for eGFR[11] in combination with the LD reference panel from the phase 3 1000 Genomes Project (all populations, $n = 2{,}504$)[70]. We first computed 7 candidate GPS using the LDPred algorithm[73] across the following range of rho (fraction of casual variants): 1.00, 1.00×10^{-1}, 1.00×10^{-2}, 1.00×10^{-3}, 3.00×10^{-1}, 3.00×10^{-2} and 3.00×10^{-3}. We also generated 12 pruning and thresholding (P + T) scores with $r^2 = 0.2$ and P thresholds of 1.0, 1.00×10^{-2}, 1.00×10^{-3}, 1.00×10^{-4}, 1.00×10^{-5}, 1.00×10^{-6}, 1.00×10^{-7}, 1.00×10^{-8}, 3.00×10^{-2}, 3.00×10^{-3}, 3.00×10^{-4} and 3.00×10^{-5}. Based on the above parameters, each GPS was expressed as a weighted sum of alleles with weights based on the GWAS for the eGFR study:

$$\mathrm{GPS} = -\sum_{j=1}^{M} \beta_j \times \left(\mathrm{dosage}_j \text{ or } \mathrm{genotype}_j\right)$$

where M is number of variants in the model and β_j is the weight based on the GWAS summary statistics and the negative sign reflects an inverse relationship between eGFR and CKD.

Each of the 19 scores derived above was subsequently assessed for discrimination of CKD cases from controls in the first UKBB optimization dataset after adjustment for age, sex, diabetes status and 4 PCs of ancestry. The score with the best performance was defined by the maximal area under the receiver operator curve (AUC) and the largest fraction of variance explained. The best performing score was based on the P + T method ($r^2 = 0.2$, $P = 0.03$) and consisted of 471,316 variants, 41,426 of which had nonzero weights (Supplementary Table 1). This score was normal-standardized (by subtracting the control mean and dividing by the control s.d.) and advanced for testing in the second UKBB optimization cohort of African ancestry.

Modeling the effects of the *APOL1* risk genotypes. To optimize trans-ethnic performance, our final score was further optimized using the second UKBB optimization dataset of African ancestry (967 cases and 6,191 controls). We aimed to assess if adding the *APOL1* risk genotype (under a recessive model) enhanced CKD risk prediction. For this purpose, we first removed any variants in the *APOL1* region from the GPS equation to avoid duplicate scoring of this region. Next, we tested the GPS and *APOL1* risk genotype jointly for association with CKD in this dataset. The GPS (without the *APOL1* region) and recessive *APOL1* risk genotypes both represented independently significant predictors of CKD before and after adjustment for age, sex, diabetes and four PCs of ancestry. The risk effects of *APOL1* and GPS were additive, with 1 s.d. unit of the standard-normalized GPS conveying the risk that was approximately equivalent to the *APOL1* risk genotype (Supplementary Table 2). We also tested for effect modification of the *APOL1* risk genotype by the polygenic component in CKD prediction but found no significant interactive effects (P interaction = 0.29). Therefore, to account for the additive effect of the *APOL1* risk genotypes, we updated the GPS for each individual using the following equation:

$$\mathrm{GPS} = \begin{cases} \text{Standardized GPS} + 1, & \text{if } \textit{APOL1} \text{ risk genotype is present} \\ \text{Standardized GPS} + 0, & \text{otherwise} \end{cases}$$

Predictive performance in independent testing datasets. The predictive performance of the final GPS formulation was assessed in 15 ancestrally diverse testing datasets, including 3 cohorts of European ancestry (14,201 cases and 82,849 controls), 6 cohorts of African ancestry (4,268 cases and 10,276 controls), 4 cohorts of Asian (East and southwest) ancestry (392 cases and 8,233 controls) and 2 admixed ancestry Latinx cohorts (1,386 cases and 2,239 controls)). We calculated a full set of standardized performance metrics according to the ClinGen guidelines[33]. Logistic regression models were used for predicting case-control status with adjustment for age, sex, diabetes, center and genotype/imputation batch (if relevant) and four PCs of ancestry using the glm function in R v.3.6.3.

We used the pROC R package v.1.18.0 to calculate the AUC. We calculated variance explained using the Nagelkerke's pseudo-R^2, including for the full model (GPS plus covariates), for the covariates-only model and for the GPS component alone expressed as the R^2 difference between the full and covariates-only models. We also expressed the effect of the standardized risk score as ORs (with 95% CIs) per s.d. unit of the control standard-normalized risk score distribution in each of the validation cohorts. We examined the risk score discrimination at tail cutoffs corresponding to the top 20, 10, 5, 2 and 1% of the GPS distribution by deriving the ORs of disease for each tail of the distribution compared to all other individuals in each cohort. We also calculated sensitivities and specificities for each cutoff point in each cohort.

The performance metrics were meta-analyzed across the testing cohorts using an inverse variance weighed fixed-effects method to derive pooled performance metrics for each ancestral grouping[74]. Finally, we calculated prevalence-adjusted positive and negative predictive values for each GPS cutoff based on pooled estimates of sensitivity and specificity and known CKD prevalence in the US population by ancestry. Statistical analyses were conducted using R v.3.6.3.

Comparing GPS distributions in the 1000 Genomes reference populations. To assess differences in the distributions of GPS by ancestry, we computed risk scores for the multiethnic reference of all 1000 Genomes phase 3 participants using our final equation:

$$\mathrm{GPS} = -\sum_{j=1}^{M} \beta_j \times \left(\mathrm{dosage}_j\right)$$

where M is the total number of variants included the model, β_j is the optimized weight based on the GWAS summary statistics for each marker included in the score and dosage_j refers to the effect allele dosage (0, 1, 2) for each variant j in the 1000 Genomes samples. Distributions were examined visually in the form of histograms and distributional differences by ancestry were tested using ANOVA.

Post-hoc ancestry adjustment. To express GPS effects on the same scale across ancestrally diverse individuals and facilitate selection of a single cutoff for clinical implementation, we adjusted for differences in the first two moments of the GPS distributions by ancestry. Using multiethnic eMERGE cohorts, we tested two different regression-based ancestry adjustment strategies that utilize the 1000 Genomes (all populations) reference: method 1, which adjusts for differences in mean, and method 2, which adjusts for both differences in mean and variance.

For method 1, we first regressed the GPS of 1000 Genomes participants against the first five PCs as proposed previously[75]:

$$\mathrm{GPS} \approx \alpha_0 + \sum_{i=1}^{5} \alpha_i \times \mathrm{PC}_i$$

Fitting the model to the 1000 Genomes reference panel allows us to find α values and generate residuals. Next, we used the estimated α values to calculate the adjusted score for any individual projected onto the same PCA space:

$$\text{Adjusted } z \text{ score (method 1)} = \frac{\sum_{j=1}^{M} w_J \times D_J - \left(\alpha_0 + \sum_{i=1}^{5} \alpha_i \times PC_i\right)}{\delta}$$

where $\sum_{j=1}^{M} w_J \times D_J$ is the raw GPS, $\alpha_0 + \sum_{i=1}^{5} \alpha_i \times PC_i$ is the predicted (ancestry-adjusted) mean and δ is the residual s.d. from the 1000 Genomes model (all populations).

To adjust for ancestral differences in both mean and variance (method 2), we used the same method as above but we also modeled residual variance (δ^2) as a function of the PCs of ancestry:

$$\delta^2 \approx \beta_0 + \sum_{i=1}^{5} \beta_i \times \mathrm{PC}_i$$

Next, we used the estimated α and β values to calculate the adjusted z-score:

$$\text{Adjusted } z \text{ score (method 2)} = \frac{\sum_{j=1}^{M} w_j \times D_j - \left(\alpha_0 + \sum_{i=1}^{5} \alpha_i \times \mathrm{PC}_i\right)}{\sqrt{\beta_0 + \sum_{i=1}^{5} \beta_i \times \mathrm{PC}_i}}$$

where $\beta_0 + \sum_{i=1}^{5} \beta_i \times \mathrm{PC}_i$ is the predicted (ancestry-adjusted) residual variance.

The distributional transformations achieved by these methods were examined visually. We then compared the effects of these adjustments for the top percentile cutoffs in the eMERGE-III cohorts.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The final formulation of the GPS for CKD along with the standardized metrics of performance have been deposited in the PGS catalog at https://www.pgscatalog.org/publication/PGP000269/. The UKBB genotype and phenotype data are available through the UKBB web portal at https://www.ukbiobank.ac.uk/. The eMERGE-III imputed genotype and phenotype data are available through the database of Genotypes and Phenotypes (dbGAP) under accession no. phs001584.v2.p2. The Bio*Me* genotype datasets used in this study were generated by Regeneron and are not publicly available. However, the data will be made available for the purposes of replicating the results by contacting the corresponding author and through appropriate collaboration and/or data sharing agreements. The WPC and REGARDS imputed genotype and phenotype data are available through dbGAP under accession nos. phs000708.v1.p1 and phs002719.v1.p1, respectively. The GenHAT cohort is also available on dbGAP under accession no. phs002716.v1.p1. The HyperGEN cohort has been sequenced by the TOPMed consortium; WGS data along with phenotype data are available through dbGAP under accession no. phs001293.v3.p1. Minimum testing datasets with the GPS, CKD outcome, and a set of essential clinical covariates for each cohort are also available when consistent with the consent given by the participants and can be requested directly from the corresponding author with a 2–4-week response time frame. Because these datasets contain clinical data, access to them may require a data use agreement.

Code availability

The CKD phenotype software is available from the Phenotype Knowledge Database at https://phekb.org/phenotype/chronic-kidney-disease. The CKD GPS score equation is available through the PGS catalog at https://www.pgscatalog.org/publication/PGP000269/ and through our laboratory website at http://www.columbiamedicine.org/divisions/kiryluk/study_GPS_CKD.php.

References

57. Khan, A. et al. Medical records-based genetic studies of the complement system. *J. Am. Soc. Nephrol.* **32**, 2031–2047 (2021).
58. McCarthy, S. et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283 (2016).
59. Abraham, G. & Inouye, M. Fast principal component analysis of large-scale genome-wide data. *PLoS ONE* **9**, e93766 (2014).
60. Das, S. et al. Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016).
61. Danecek, P. et al. The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
62. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
63. Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
64. Belbin, G. M. et al. Toward a fine-scale population health monitoring system. *Cell* **184**, 2068–2083.e11 (2021).
65. Howard, V. J. et al. The reasons for geographic and racial differences in stroke study: objectives and design. *Neuroepidemiology* **25**, 135–143 (2005).
66. Williams, R. R. et al. NHLBI family blood pressure program: methodology and recruitment in the HyperGEN network. *Ann. Epidemiol.* **10**, 389–400 (2000).
67. Limdi, N. A. et al. Influence of kidney function on risk of supratherapeutic international normalized ratio-related hemorrhage in warfarin users: a prospective cohort study. *Am. J. Kidney Dis.* **65**, 701–709 (2015).
68. Arnett, D. K. et al. Pharmacogenetic approaches to hypertension therapy: design and rationale for the Genetics of Hypertension Associated Treatment (GenHAT) study. *Pharmacogenomics J.* **2**, 309–317 (2002).
69. Furberg, C. D. et al. Major cardiovascular events in hypertensive patients randomized to doxazosin vs chlorthalidone: the antihypertensive and lipid-lowering treatment to prevent heart attack trial (ALLHAT). ALLHAT Collaborative Research Group. *JAMA* **283**, 1967–1975 (2000).
70. Auton, A. et al. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
71. Levey, A. S. & Stevens, L. A. Estimating GFR using the CKD Epidemiology Collaboration (CKD-EPI) creatinine equation: more accurate GFR estimates, lower CKD prevalence estimates, and better risk predictions. *Am. J. Kidney Dis.* **55**, 622–627 (2010).
72. Kidney Disease: Improving Global Outcomes (KDIGO) Chronic Kidney Disease Work Group KDIGO 2012 clinical practice guideline for the evaluation and management of chronic kidney disease. *Kidney Int. Suppl.* **3**, 1–150 (2013).
73. Vilhjálmsson, B. J. et al. Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *Am. J. Hum. Genet.* **97**, 576–592 (2015).
74. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
75. Khera, A. V. et al. Whole-genome sequencing to characterize monogenic and polygenic contributions in patients hospitalized with early-onset myocardial infarction. *Circulation* **139**, 1593–1602 (2019).

Acknowledgements

This work was funded by the National Human Genome Research Institute eMERGE-IV grant nos. 2U01HG008680-05, 1U01HG011167-01 and 1U01HG011176-01. Additional sources of funding included grant nos. UG3DK114926 (K.K.), RC2DK116690 (K.K.), R01LM013061 (C.W., K.K.), K25DK128563 (A.K.), UL1TR001873 (A.K., K.K.), R01HL151855 (J.B.M.) and UM1DK078616 (J.B.M.). The parent REGARDS study was supported by cooperative agreement no. U01NS041588 cofunded by the National Institute of Neurological Disorders and Stroke and the National Institute on Aging, the National Institutes of Health (NIH) and the Department of Health and Human Services. The HyperGEN (R01HL055673), GenHAT (R01HL123782) and WPC (R01HL092173, K24HL133373) studies were all supported by the NHLBI. Parts of this study were conducted using the UKBB resource under UKBB project no. 41849. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Author contributions

A.K. and K.K. conceptualized the study. A.K., K.K., M.C.T., A.P., V.S., R.N., A.C.J., E.M., C.K., N.L., I.I.L., T.G., M.R.I., H.K.T. and E.E.K. devised the methodology and carried out the genetic data analysis. N.S., C.Li., G.H., C.W. and G.N. carried out the e-phenotyping. O.D., I.J.K., D.J.S., E.K., J.B.M., J.W.S., C.La., D.R.C., G.P.J., P.K.B., J.N.H., P.C., L.R.T., A.G.G., W.K.C., G.H. and C.W. provided the eMERGE-III data contributions. G.N., J.H.C., N.S.A-H. and E.E.K. provided the Bio*Me* data contributions. M.R.I., H.K.T. and N.A.L. provided the UAB data contributions. A.K. and A.F. managed the project. K.K. provided overall supervision. A.K. and K.K. wrote the original manuscript draft. A.K., N.S., E.M., A.G.G., T.G., J.B.M., D.R.C., J.N.H., I.I.L., G.H., M.R.I., H.K.T., E.E.K., N.A.L. and K.K. reviewed and edited the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

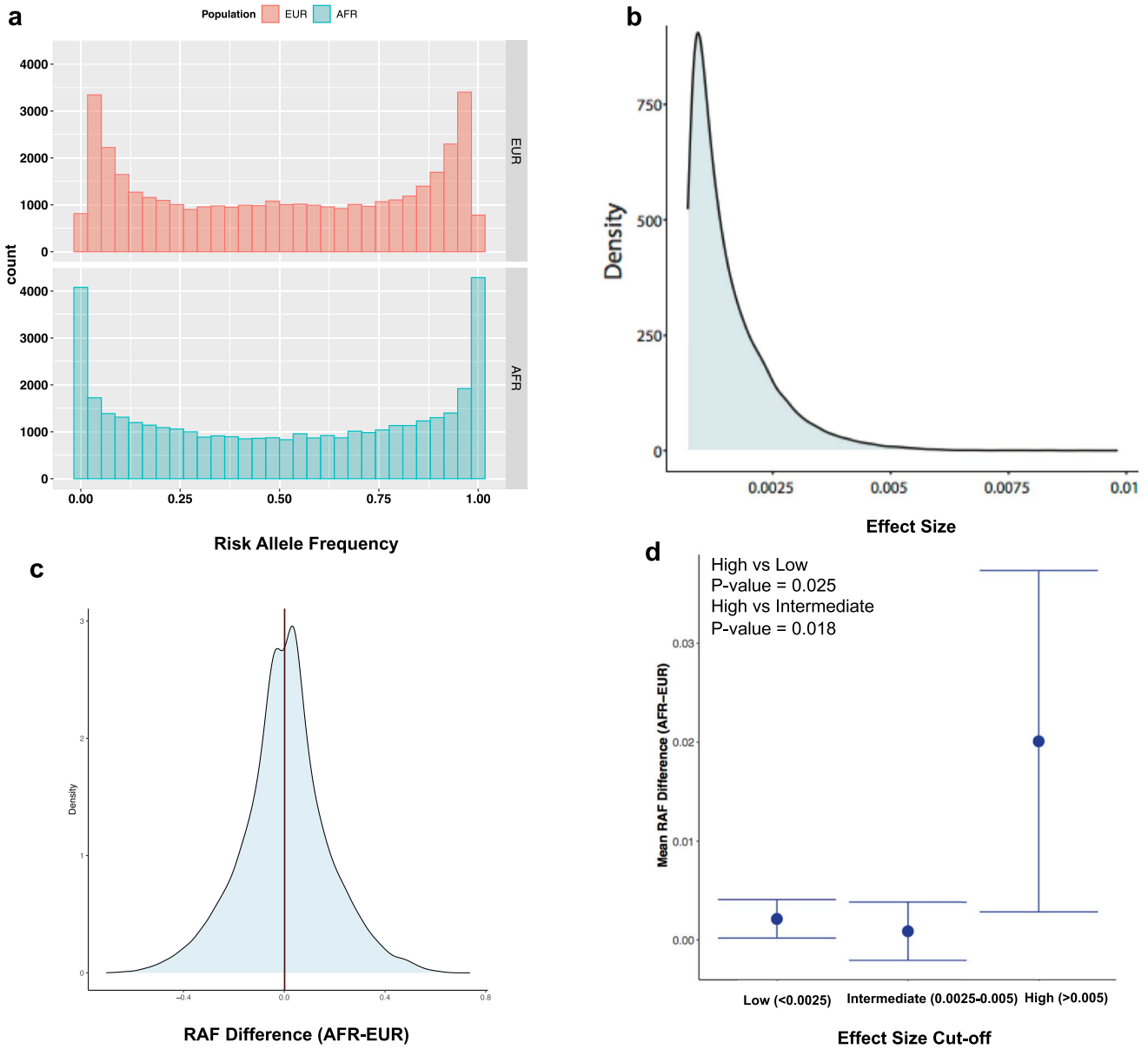
Extended data is available for this paper at https://doi.org/10.1038/s41591-022-01869-1.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41591-022-01869-1.

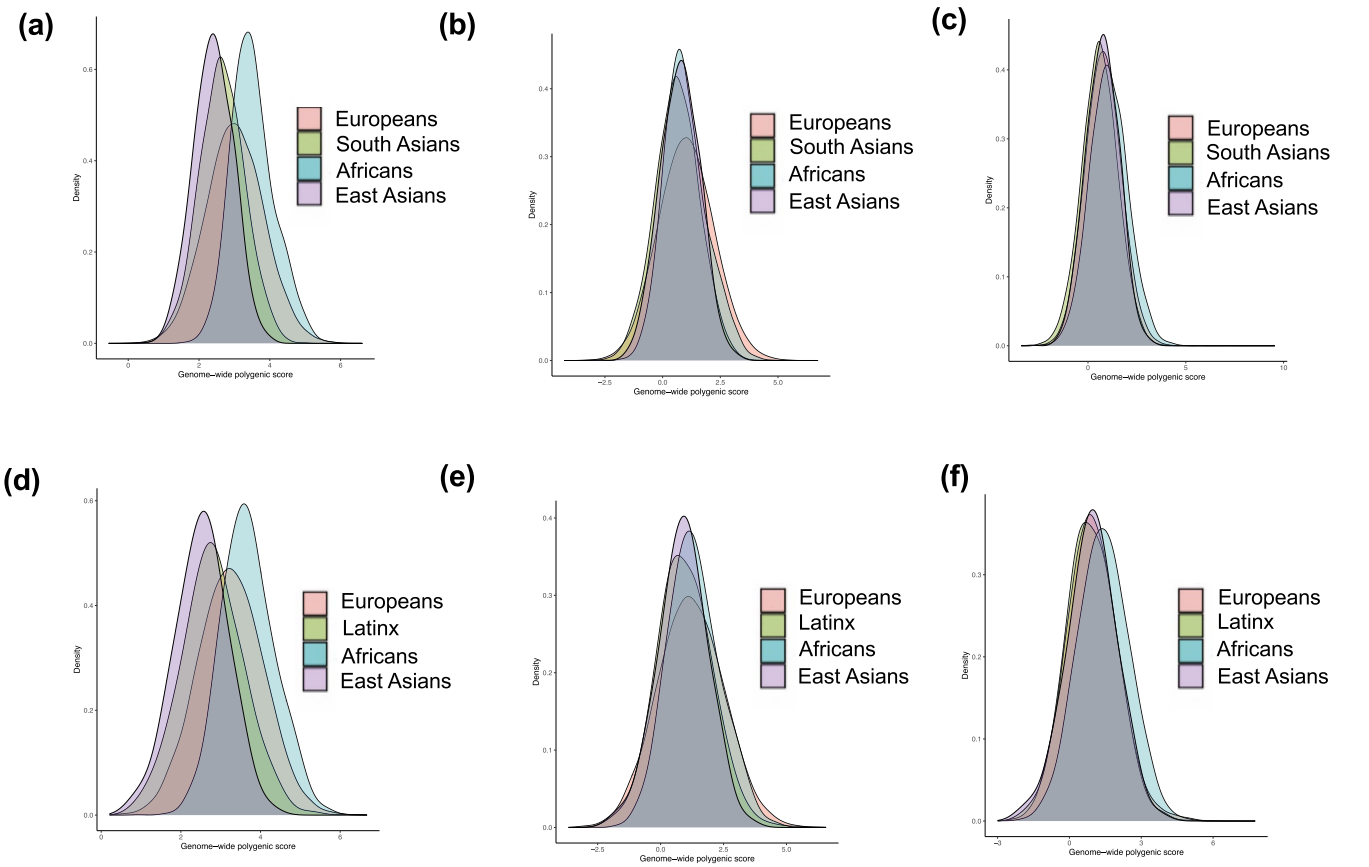
Correspondence and requests for materials should be addressed to Krzysztof Kiryluk.

Peer review information *Nature Medicine* thanks Charles Rotimi, Andrew Mallett and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editor: Anna Maria Ranzoni, in collaboration with the *Nature Medicine* team.

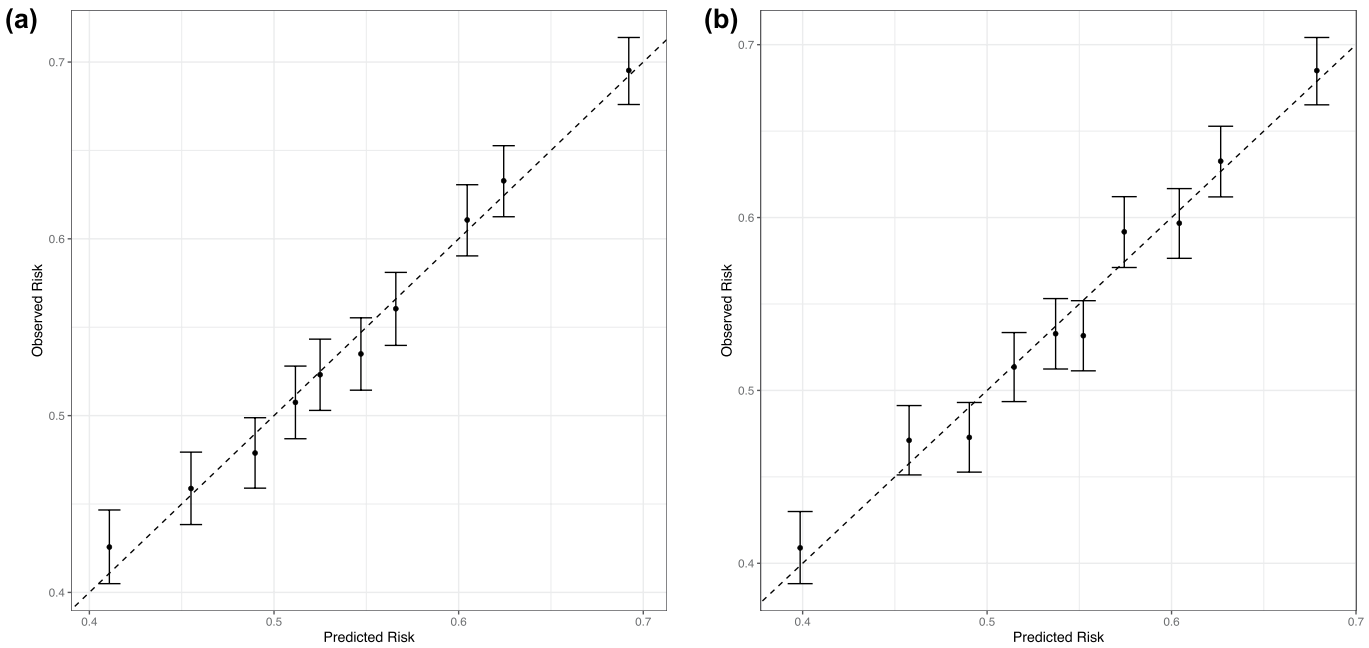
Reprints and permissions information is available at www.nature.com/reprints.



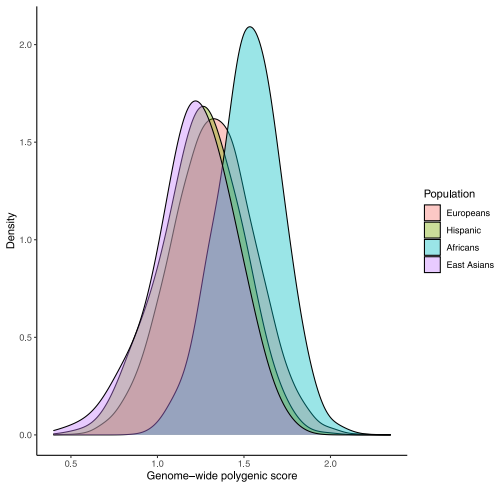
Extended Data Fig. 1 | Distribution of risk allele frequencies (RAF) and their effect sizes for the variants included in the GPS. Distribution of risk allele frequencies (RAF) and their effect sizes for the variants included in the GPS (a) comparison of RAF distributions for the risk variants included in the CKD GPS demonstrates higher frequency of rare (RAF < 0.01) and common (RAF > 0.99) risk alleles in African compared to European genomes (based on 1000 G reference populations); this may be explained by the exclusion of variants with MAF < 0.01 in European discovery GWAS; **(b)** highly skewed effect size (weight) distribution for the variants included in the GPS for CKD; **(c)** Distribution of RAF difference (AFR-EUR) demonstrating higher average frequency of risk alleles in African genomes (mean RAF difference = 0.002) and a slight rightward shift of the RAF difference distribution from the expected mean of 0; **(d)** Mean RAF difference (AFR-EUR) as a function of effect size binned into three categories (high, intermediate, and low) based on the observed distribution of effects sizes in panel b, demonstrating that the risk alleles with larger effect size have higher average frequency in African compared to European genomes. EUR: European (N = 503) and AFR: African (N = 661). The bars represent 95% confidence intervals around the mean RAF difference estimate for each bin; two-sided P-values were calculated using t-test.



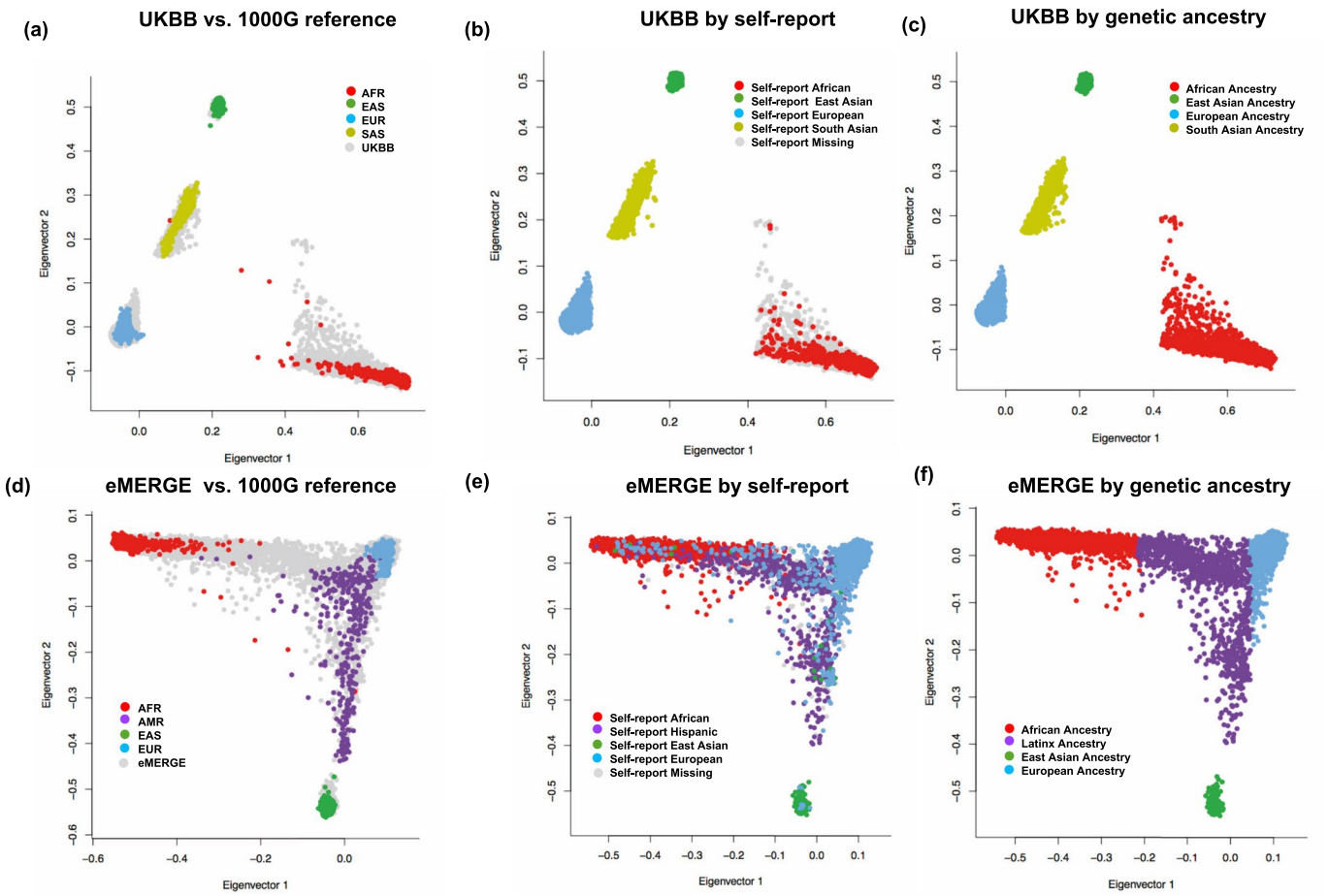
Extended Data Fig. 2 | Risk score distributions in eMERGE-III (N = 22,453) and UKBB (N = 77,584) validation datasets. Risk score distributions in eMERGE-III (N = 22,453) and UKBB (N = 77,584) validation datasets: (a) the distribution of raw polygenic score without *APOL1* in UKBB by ancestry; **(b)** the distribution of ancestry-adjusted polygenic score (method 1: mean-adjusted) in UKBB by ancestry; **(c)** the distribution of ancestry-adjusted polygenic score (method 2: mean and variance-adjusted) in UKBB by ancestry. Panels **(d), (e)** and **(f)** show the same analyses for the eMERGE-III dataset, respectively.



Extended Data Fig. 3 | Final GPS calibration analysis in eMERGE-III cohorts combined (N = 22,453). Final GPS calibration analysis in eMERGE-III cohorts combined (N = 22,453): predicted risk (X-axis) as a function of the observed risk (Y-axis) in the multiethnic eMERGE-III dataset after ancestry adjustment with **(a)** method 1 and **(b)** method 2. The bars represent 95% confidence intervals.



Extended Data Fig. 4 | Distributions of the raw (non-standardized) genome-wide polygenic score (GPS) by Yu et al. in the eMERGE-III validation datasets by ancestry. Distributions of the raw (non-standardized) genome-wide polygenic score (GPS) by Yu et al. in the eMERGE-III validation datasets by ancestry.



Extended Data Fig. 5 | PCA projections of the study participants from the UKBB (top) and eMERGE-III (bottom) against the 1000 G reference populations. PCA projections of the study participants from the UKBB (top) and eMERGE-III (bottom) against the 1000 G reference populations: (a) UKBB (N = 77,584) and (b) eMERGE-III (N = 22,453) participants plotted against the reference 1000 G populations (N = 2,504); (b, e) plotted by self-reported race/ethnicity; and (c, f) plotted by final ancestry group assignment. X-axis: PC1; Y-axis: PC2; AFR: African; AMR: Admixed American; EAS: East Asian; EUR: European; and SAS: South Asian.

nature research

Corresponding author(s): Krzysztof Kiryluk

Last updated by author(s): Apr 23, 2022

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Electronic CKD Phenotype software is publicly available from the Phenotype Knowledge Database (PheKB, https://phekb.org/phenotype/chronic-kidney-disease).
Data analysis	The data was analyzed using open source statistical software packages including: R version 3.6.3 (CRAN), LDPred, and PLINK v. 1.9, IMPUTE4, FlashPCA, EIGENSOFT version 6.1.4.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The final formulation of the GPS for CKD along with the standardized metrics of performance were deposited in the GPS catalogue: https://www.pgscatalog.org/publication/PGP000269/. The UK Biobank genotype and phenotype data are available through the UK Biobank web portal at https://www.ukbiobank.ac.uk/. The Electronic Medical Records and Genomics-III imputed genotype and phenotype data are available through dbGAP, accession number phs001584.v2.p2.
The BioMe genotype datasets used in this study were generated by Regeneron and are not publicly available. However, the data will be made available for purposes of replicating the results by contacting the corresponding author and appropriate collaboration and/or data sharing agreements. The Warfarin, REGARDS and GenHAT imputed genotype and phenotype data available through dbGAP, respective accession numbers phs000708.v1.p1, phs002719.v1.p1, and phs002716.v1.p1.

The HyperGEN cohort has been sequenced by the TopMED consortium and WGS data along with the phenotype data are available through dbGAP, accession number phs001293.v3.p1. Minimum testing datasets with the GPS, CKD outcome, and essential clinical covariates for each cohort are also available upon reasonable request directed to the corresponding author (KK) with a 2–4-week response timeframe; because these datasets contain clinical data, access to these datasets will require a data use agreement.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	We utilized all case-control cohorts available through the Electronic Health Records and Genomics consortium without excluding any cohorts or studies. The aim of the consortium is to validate polygenic scores in ancestrally diverse population, and thus we aimed to achieve a minimum of 1,000 CKD cases per each continental ancestry in the meta-analysis to assure adequate sample size for risk score performance testing. The final number of the testing datasets used involved all datasets with renal function and genome-wide genotype data available to the eMERGE consortium investigators that have not been previously included in the CKDGen GWAS meta-analyses for eGFR (i.e. we included only fully independent datasets).
Data exclusions	From our optimization and testing sets, we excluded all cohorts that were included in the published GWAS for eGFR by the CKDGen Consortium. This exclusion was necessary to assure that the optimization and testing cohorts are fully independent of the GWAS discovery study that was used to develop the risk score equation.
Replication	We have performed independent performance testing in 15 ancestrally diverse cohorts. We were not able to identify a larger number of independent datasets. With the available datasets we successfully demonstrated replication of our key findings across all ancestral groups as described in the manuscript.
Randomization	Randomization was not required, since our study design involved analysis of observational data rather than an intervention.
Blinding	Blinding was not required, since our study design involved analysis observational data.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	Multiple cohort studies were included in the analysis, and details of each study participants are provided in the manuscript with relevant references. Demographic characteristics relevant to this study are also summarized for each cohort in Table 1, including ancestry, age, sex, diagnosis of diabetes, and CKD case status.
Recruitment	No active recruitment took place specifically for this study. The analysis involved only existing genetic cohorts.
Ethics oversight	The study was approved by the Columbia University Institutional Review Board (IRB numbers IRB-AAAQ9205, IRB-AAAT8208, and IRB-AAAS3500). All participating studies were approved by their local institutional review boards, including all sites contributing human genetic and clinical data to the Electronic Medical Records and Genomics phase 3 (eMERGE-III) consortium. Of note, BioVU operated on an opt-out basis until January 2015 and on an opt-in basis since. The phenotypic

data in BioVU are all deidentified, and the study was designated “nonhuman subjects” research by the Vanderbilt Institutional Review Board. All other participants provided written informed consent to participate in genetic studies.

Note that full information on the approval of the study protocol must also be provided in the manuscript.