JCI insight

Protein-truncating variant in APOL3 increases chronic kidney disease risk in epistasis with APOL1 risk alleles

David Y. Zhang, … , Marylyn D. Ritchie, Daniel J. Rader

Clinical Medicine   Genetics   Nephrology

Two coding alleles within the *APOL1* gene, G1 and G2, found almost exclusively in individuals genetically similar to West African populations, contribute substantially to the pathogenesis of chronic kidney disease (CKD). The *APOL* gene cluster on chromosome 22 contains a total of 6 *APOL* genes that have arisen as a result of gene duplication.

Using a genome-first approach in the Penn Medicine BioBank, we identified 62 protein-altering variants in the 6 *APOL* genes with a minor allele frequency of >0.1% in a population of participants genetically similar to African reference populations and performed population-specific phenome-wide association studies.

We identified rs1108978, a stop-gain variant in *APOL3* (p.Q58*)*,* to be significantly associated with increased CKD risk, even after conditioning on *APOL1* G1/G2 carrier status. These findings were replicated in the Veterans Affairs Million Veteran Program and the *All of Us* Research Program. *APOL3* p.Q58* was also significantly associated with a number of quantitative traits linked to CKD, including decreased kidney volume. This truncating variant contributed the most risk for CKD in patients monoallelic for *APOL1* G1/G2, suggesting an epistatic interaction and a potential protective effect of wild-type APOL3 against *APOL1*-induced kidney […]

Find the latest version:

https://jci.me/181238/pdf

JCI insight

Protein-truncating variant in APOL3 increases chronic kidney disease risk in epistasis with APOL1 risk alleles

David Y. Zhang,[1,2] Michael G. Levin,[3,4] Jeffrey T. Duda,[5] Latrice G. Landry,[1] Walter R. Witschey,[6] Scott M. Damrauer,[4,7] Marylyn D. Ritchie,[1,8] and Daniel J. Rader[1,2]

[1]Department of Genetics, [2]Department of Medicine, and [3]Division of Cardiovascular Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA. [4]Corporal Michael J. Crescenz VA Medical Center, Philadelphia, Pennsylvania, USA. [5]Penn Image Computing and Science Laboratory, University of Pennsylvania, Philadelphia, Pennsylvania, USA. [6]Department of Radiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA. [7]Department of Surgery, University of Pennsylvania, and [8]Institute for Biomedical Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA.

BACKGROUND. Two coding alleles within the *APOL1* gene, G1 and G2, found almost exclusively in individuals genetically similar to West African populations, contribute substantially to the pathogenesis of chronic kidney disease (CKD). The *APOL* gene cluster on chromosome 22 contains a total of 6 *APOL* genes that have arisen as a result of gene duplication.

METHODS. Using a genome-first approach in the Penn Medicine BioBank, we identified 62 protein-altering variants in the 6 *APOL* genes with a minor allele frequency of >0.1% in a population of participants genetically similar to African reference populations and performed population-specific phenome-wide association studies.

RESULTS. We identified rs1108978, a stop-gain variant in *APOL3* (p.Q58*), to be significantly associated with increased CKD risk, even after conditioning on *APOL1* G1/G2 carrier status. These findings were replicated in the Veterans Affairs Million Veteran Program and the *All of Us* Research Program. *APOL3* p.Q58* was also significantly associated with a number of quantitative traits linked to CKD, including decreased kidney volume. This truncating variant contributed the most risk for CKD in patients monoallelic for *APOL1* G1/G2, suggesting an epistatic interaction and a potential protective effect of wild-type APOL3 against *APOL1*-induced kidney disease.

CONCLUSION. This study demonstrates the utility of targeting population-specific variants in a genome-first approach, even in the context of well-studied gene-disease relationships.

FUNDING. National Heart, Lung, and Blood Institute (F30HL172382, R01HL169378, R01HL169458), Doris Duke Foundation (grant 2023-2024), National Institute of Biomedical Imaging and Bioengineering (P41EB029460), and National Center for Advancing Translational Sciences (UL1-TR-001878).

Introduction

Chronic kidney disease (CKD) and end-stage renal disease (ESRD) are significantly more common in individuals genetically similar to African reference populations (AFR) compared with those genetically similar to European reference populations (EUR) (1, 2). While there are multiple attributable factors, one of the most impactful is the common genetic variation in the *APOL1* gene (3–5). Two coding alleles within the *APOL1* gene, G1 and G2, are found almost exclusively in individuals genetically similar to West African populations and contribute substantially to the pathogenesis of nondiabetic kidney disease, focal segmental glomerulosclerosis, and HIV-associated nephropathy (6, 7). The G1 allele (minor allele frequency [MAF] of ~23% in individuals genetically similar to AFR used by gnomAD) comprises 2 missense variants in near-perfect linkage disequilibrium (LD), G1^G (p.S342G) and G1^M (p.I384M). The G2 allele (MAF of ~14% in individuals genetically similar to AFR used by gnomAD) is a 6-base pair in-frame deletion

(p.NYK388-389K). The high allele frequency of these variants, particularly in West African populations, is caused by a recent positive selective sweep due to the protective effects they confer against *Trypanosoma brucei* infections, the cause of African sleeping sickness (8). In fact, evidence suggests that G1 and G2 are toxic gain-of-function variants in *APOL1*, a gene shown to play roles in programmed cell death and pathogen immunity (9). The G1 and G2 alleles arose independently on separate chromosomes and are too close in proximity to have undergone a recombination event that would allow a single haplotype to carry both G1 and G2. Therefore, risk of CKD is modeled on a scale of 0 to 2 by the total number of G1/G2 alleles an individual carries. Individuals with 2 G1/G2 risk alleles are considered to be "high risk" for CKD. Recently, a missense variant, p.N264K, in *APOL1* was shown to exert protective effects in high-risk individuals carrying the *APOL1* variants by inhibiting APOL1 pore-forming function and ion channel conduction (10).

The *APOL* gene cluster on chromosome 22 contains a total of 6 *APOL* genes that have arisen as a result of gene duplication (11). The physiological functions of the APOL proteins are poorly understood. We hypothesized that there may be other protein-altering variants in the *APOL* genes that are associated with health and disease in the AFR population. Adopting a genome-first approach, we leveraged the Penn Medicine BioBank (PMBB), a large medical biobank with whole-exome sequence data linked to electronic health records (EHRs) (12), to study the phenotypes associated with protein-altering variants in the 6 *APOL* genes with a focus on the AFR population. We identified an AFR-specific protein-truncating variant in *APOL3* (MAF of ~22% in individuals genetically similar to AFR used by gnomAD) that was significantly associated with increased risk of CKD and primarily increased CKD risk in monoallelic carriers of the *APOL1* G1/G2 alleles.

Results

Phenome-wide association studies for protein-altering variants in APOL genes. Of the 43,731 consented individuals in the PMBB with whole-exome sequencing (WES), we extracted 841 protein-altering variants across all 6 *APOL* genes, of which 100 were predicted loss of function (pLOF), 4 were in-frame insertions/deletions, and 737 were missense. With a specific focus on variants in individuals genetically similar to AFR reference populations, we filtered our variant set down to 62 variants with a MAF of 0.1% in the PMBB AFR population (n = 11,198), for which we show that statistical power to detect an association was sufficient (Supplemental Figure 1; supplemental material available online with this article; https://doi.org/10.1172/jci.insight.181238DS1). This set of variants included 6 pLOFs, 1 in-frame deletion, and 55 missense mutations, including both *APOL1* G1 and G2 AFR-specific risk alleles for kidney disease (Supplemental Table 1). For each of these 62 variants, we performed a phenome-wide association study (PheWAS) in the AFR population in PMBB against 1,222 binary phenotypes defined as Phecodes derived from EHR data (Figure 1), followed by additional downstream variant-specific analyses (Supplemental Figure 2).

Using a strict Bonferroni's P value correction adjusting for every single variant-phenotype association performed, 58 significant associations ($P < 6.63 \times 10^{-7}$) were observed across 5 variants (Table 1) as were 23 unique phenotypes, all of which were related to renal disease (Supplemental Table 2). Three of the 5 significant variants were in *APOL1,* including the G1 risk allele (rs73885319 [p.S342G] and rs60910145 [p.I384M]) as well as a third missense variant rs2239785 (p.E150K). All 3 variants had strong associations with end-stage renal disease (ESRD) with odds ratios (ORs) of 1.70 (95% CI = 1.52–1.90, $P = 6.92 \times 10^{-21}$), 1.68 (95% CI = 1.50–1.88, $P = 5.83 \times 10^{-20}$), and 1.50 (95% CI = 1.34–1.68, $P = 1.06 \times 10^{-12}$), respectively. A missense variant in *APOL2*, rs7285167 (p.R182C), was significantly associated with ESRD with an OR of 1.43 (95% CI = 1.29–1.59, $P = 1.29 \times 10^{-11}$). Finally, a stop-gain variant in *APOL3*, rs11089781 (p.Q58*), was significantly associated with ESRD, with an OR of 1.39 (95% CI = 1.24–1.56, $P = 5.18 \times 10^{-9}$). Of note, the *APOL1* G2 allele alone did not meet our strict significance threshold but was also found to be strongly associated with ESRD with an OR of 1.33 (95% CI = 1.16–1.53, $P = 2.45 \times 10^{-5}$). Similarly, the *APOL1* p.N264K protective variant also did not meet our significance threshold but was found to be nominally associated with decreased risk of nephritis, nephrosis, and renal sclerosis, with an OR of 0.43 (95% CI = 0.23–0.81, $P = 4.63 \times 10^{-3}$), and decreased risk of ESRD, with an OR of 0.64 (95% CI = 0.41–0.99, $P = 2.18 \times 10^{-2}$). We repeated the analyses in the EUR population and meta-analyzed under a fixed-effects model and observed similar significant findings (Supplemental Figure 3 and Supplemental Table 3). ESRD had a prevalence of 8.4% in the PMBB AFR population. The average age of individuals with ESRD was 55.8 years old compared with 51.2 years old for individuals acting as controls (t test, $P = 1.86 \times 10^{-22}$). 58.3% of individuals in the case group were male compared with 34.8% of individuals acting as controls (χ^2, $P = 3.31 \times 10^{-45}$).
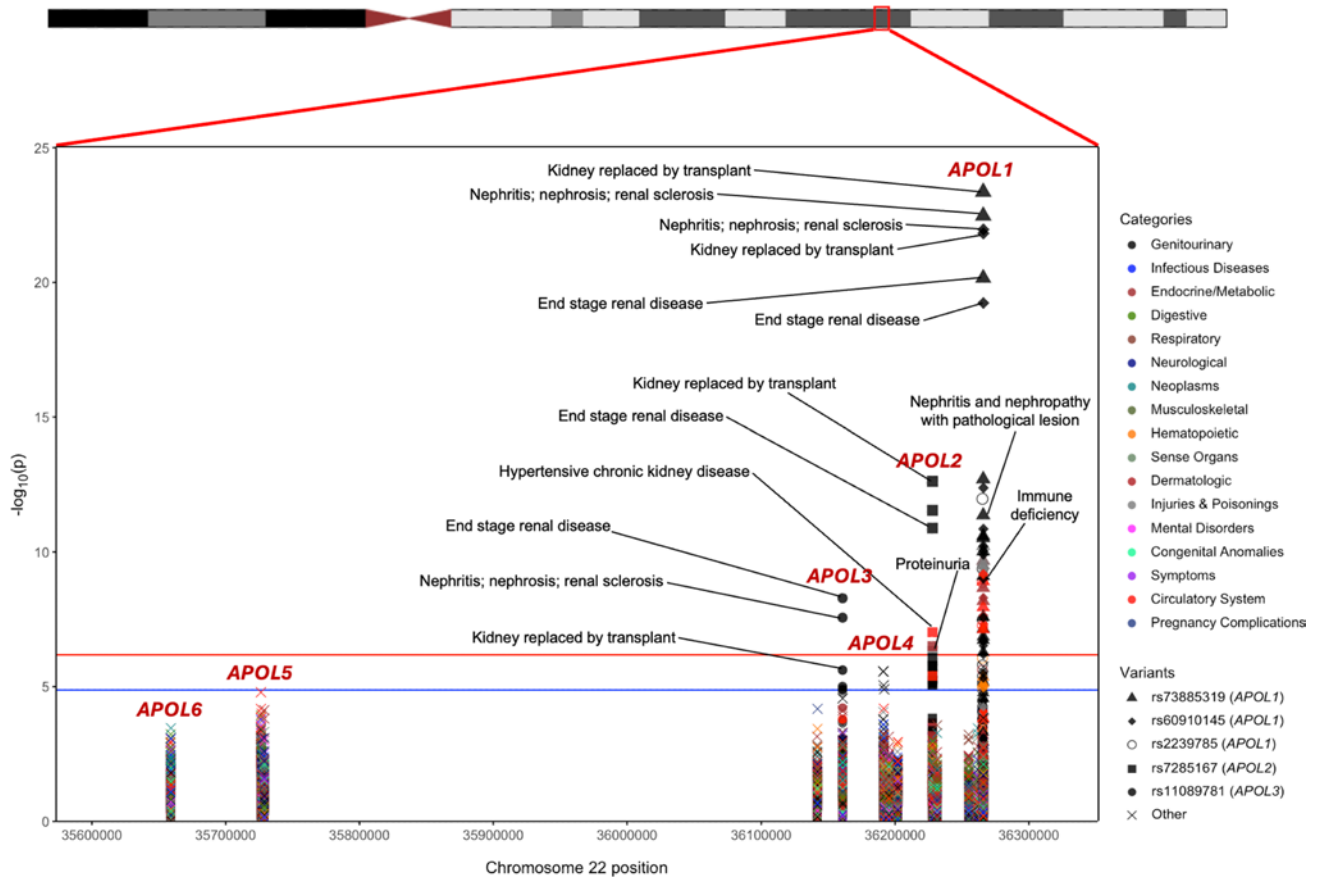
Figure 1. PheWAS for 62 protein-altering variants in *APOL* genes. Associations were performed in the PMBB AFR population. The red line represents the Bonferroni-adjusted *P* value significance threshold of 6.63×10^{-7}. The blue line represents a suggestive *P* value threshold of 1.33×10^{-5}. Variants with at least one phenotype association above the significance threshold are listed in the legend and detailed in Supplemental Table 2.

All 5 significant variants were substantially more common in the AFR population compared with the EUR population in PMBB, and the MAFs in PMBB were similar to the MAFs in the gnomAD v.4.0.0 database (Table 1). We examined the LD structure between these 5 variants (as well as the *APOL1* G2 allele and *APOL1* p.N264K variant) in the PMBB AFR population (Figure 2). As expected, the 2 missense variants that comprise the G1 allele were in virtually complete LD, and neither were in any LD with the G2 allele. The *APOL1* rs2239785 (p.E150K) allele had an r^2 value of 0.16 with the G1 allele. The *APOL2* rs7285167 (p.R182C) allele was in slightly more LD with the G1 allele ($r^2 = 0.26$). Importantly, the *APOL3* stop-gain variant, rs11089781 (p.Q58*), was in weak LD with the G1 allele ($r^2 = 0.11$) as well as with *APOL1* p.N264K ($r^2 = 0.001$). The *APOL3* variant is also highly specific for the AFR population with a PMBB AFR (*n* = 11,198) MAF of 0.211 and a PMBB EUR (*n* = 30,324) MAF of 6.43×10^{-4}. Given that this variant appeared to have an independent signal for association with renal disease, we performed a single variant PheWAS restricted to AFR individuals, showing a strong clustering of significant kidney-associated phenotypes (Figure 3). We replicated this observed association between the *APOL3* p.Q58* variant and CKD/ESRD using identical methods in the AFR population in the Million Veteran Program (MVP) (*n* = 120,839) and found an OR of 1.16 (95% CI = 1.10–1.22, *P* = 1.01×10^{-8}). We then meta-analyzed the AFR and EUR results (*n* = 577,021) to obtain an OR of 1.18 (95% CI = 1.12–1.24, *P* = 2.37×10^{-10}). Performing the same variant-phenotype association in the *All of Us* Research Program, we found an AFR-specific (*n* = 36,262) OR of 1.24 (1.09–1.40, *P* = 3.72×10^{-4}) and an AFR with EUR meta-analyzed (*n* = 139,019) OR of 1.23 (95% CI = 1.09–1.40, *P* = 8.13×10^{-4}).

To isolate the independent effects of the 3 significant non-*APOL1* G1/G2 variants on renal disease, we assessed conditional associations for the 3 variants in the PMBB AFR population for all phenotypes originally found to be significantly associated during assessment of conditioning for *APOL1* G1/G2, which was risk modeled using APOL's well-known recessive inheritance pattern (Table 2). The *APOL1* rs2239785 (p.E150K)

Table 1. Significant variants from AFR PheWAS analysis

Marker ID (chr:pos:ref:alt)	rsID	Gene	Major	Minor	Function	AA change	PMBB AFR MAF	PMBB EUR MAF	gnomAD v.4 AFR MAF	gnomAD v.4 NFE MAF
22:36265988:T:G	rs60910145	APOL1 (G1)	T	G	Nonsynonymous	p.I384M	0.215	2.97×10^{-4}	0.227	5.34×10^{-5}
22:36265860:A:G	rs73885319	*APOL1* (G1)	A	G	Nonsynonymous	p.S324G	0.219	3.30×10^{-4}	0.229	5.85×10^{-5}
22:36265284:G:A	rs2239785	APOL1	A	G	Nonsynonymous	p.E150K	0.632[A]	0.209	0.645[A]	0.158
22:36227874:G:A	rs7285167	APOL2	G	A	Nonsynonymous	p.R182C	0.296	0.084	0.308	0.069
22:36160720:G:A	rs11089781	APOL3	G	A	Stop-gain	p.Q58*	0.211	6.43×10^{-4}	0.218	

Annotations for variants with at least one significant phenotype association in AFR analysis. All significant phenotype associations are detailed in Supplemental Table 2. MAFs calculated both in PMBB and from allele counts in gnomAD v.4.0.0 are shown. NFE indicates the non-Finnish European population in gnomAD. [A]Allele frequency >0.5 for the minor allele, as determined by gnomAD. rsID, reference SNP cluster ID; SNP, single nucleotide polymorphism.

allele and the *APOL2* rs7285167 (p.R182C) allele were no longer significantly associated with renal phenotypes. However, the *APOL3* stop-gain variant rs11089781 (p.Q58*) remained nominally significantly associated with increased risk for renal disease. Conditioning *APOL3* p.Q58* on *APOL1* p.N264K made minimal difference compared with the unconditional analysis, and conditioning on both *APOL1* G1/G2 and p.N264K resulted in minimal change compared with only conditioning on *APOL1* G1/G2 (Supplemental Table 4). This persistence of the association signal indicates that this stop-gain variant in *APOL3* has some independent effect on CKD risk.

Interrogation of APOL3 p.Q58 association with ESRD.* To investigate potential gene dosage effects of *APOL3* p.Q58* on ESRD risk, we compared the prevalence of individuals with ESRD among different carrier statuses of rs11089781 in the AFR population in our biobank. 7.3% of noncarriers of *APOL3* rs11089781 were diagnosed with ESRD (cases, $n = 501$; controls, $n = 6,386$) compared with 9.4% of heterozygote carriers (cases, $n = 332$; controls, $n = 3,206$), and 15.5% of homozygote carriers (cases, $n = 84$; controls, $n = 459$). Using Fisher's exact tests, the prevalence of ESRD was significantly different across all 3 carrier groups ($P < 1 \times 10^{-3}$), suggesting a gene dosage effect on renal disease. Furthermore, we compared the age of onset for ESRD in our cohort among patients with different carrier statuses of p.Q58*. The average age of onset was 55.0 years in patients who do not carry the variant, 53.1 years in patients who carry 1 copy of the variant, and 49.6 years in patients who carry 2 copies. Using Student's *t* tests, the difference between noncarriers and monoallelic carriers was near significant ($P = 0.056$), the difference between monoallelic and biallelic carriers was nominally significant ($P = 0.042$), and the difference between noncarriers and biallelic carriers was most significant ($P = 0.002$).

Building upon our PheWAS results, we analyzed relevant EHR-derived laboratory measurements and kidney imaging traits to perform quantitative associations with this *APOL3* stop-gain variant. Focusing on renal and kidney-related hematological lab values, we computed the maximum, median, and minimum values for each trait for each individual, performed inverse-normal transformation, and again ran all associations in the PMBB AFR population. The same analysis was carried out for relevant kidney imaging traits, where we curated clinically available CT scans from patients in PMBB, segmented the left and right kidneys, extracted quantitative imaging traits, and computed the maximum, median, and minimum values for each trait for each individual followed by normalization. We found that the *APOL3* variant was strongly associated with decreased minimum estimated glomerular filtration rate (eGFR; $P = 2.17 \times 10^{-7}$, $n = 10,435$) and increased maximum creatinine ($P = 3.85 \times 10^{-6}$, $n = 10,435$), consistent with increased risk for renal disease (Supplemental Table 5). We also identified nominally significant associations with decreased minimum red blood cell counts ($P = 3.49 \times 10^{-3}$, $n = 10,020$), consistent with decreased erythropoietin production, as well as decreased minimum lymphocyte percentage ($P = 5.57 \times 10^{-3}$, $n = 10,118$). We replicated quantitative associations for eGFR and creatinine in MVP and identified concordant significant associations for decreased mean eGFR ($P = 2.42 \times 10^{-13}$, $n = 110,674$) and increased mean creatinine ($P = 1.63 \times 10^{-8}$, $n = 116,531$). We observed similar associations in *All of Us* for decreased minimum eGFR ($P = 7.97 \times 10^{-3}$, $n = 25,572$) and increased maximum creatinine ($P = 8.13 \times 10^{-3}$, $n = 25,572$). Finally, using our analysis of quantitative kidney-derived CT imaging traits, carriers of the *APOL3* p.Q58* were found to have significantly decreased minimum kidney volume ($P = 2.49 \times 10^{-3}$, $n = 1,767$) as well as decreased minimum kidney surface area ($P = 5.14 \times 10^{-3}$, $n = 1,768$) (Supplemental Table 5). Similar association results were

Figure 2. LD heatmap between significant variants in *APOL* genes. The 5 variants with significant phenotype associations as well as the *APOL1* G1/G2 risk alleles and the *APOL1* p.N264K variant. Metrics calculated using haplotypes in the PMBB AFR population (*n* = 11,198). LD metrics computed using haplotypes from the 1000 Genomes Project can be found in Supplemental Figure 4.

observed when the AFR population results were meta-analyzed under a fixed-effects model with results from the EUR population (Supplemental Table 6). Thus, in addition to its association with diagnosis codes reflecting CKD, *APOL3* p.Q58* is associated with a number of quantitative traits concordant with CKD across 3 different cohorts enriched in participants genetically similar to AFR reference populations.

We then analyzed the association of *APOL3* p.Q58* with CKD after stratifying by *APOL1* G1/G2 carrier status. Given the possible recessive effects of our *APOL3* variant suggested by our gene dosage results, we performed the stratification analyses under both an additive and recessive model. Interestingly, we found that *APOL3* p.Q58* increases risk for ESRD most significantly under a recessive inheritance pattern in *monoallelic APOL1* G1/G2 risk allele carriers (Tables 3 and 4). We found the same result upon stratified analyses of this variant with eGFR and creatinine. This result suggested that this *APOL3* stop-gain variant may have an epistatic interaction with *APOL1* G1/G2 and increases risk of CKD most prominently in monoallelic carriers for either the *APOL1* G1 or G2 allele. Carrier counts for both *APOL1* G1/G2 and *APOL3* p.Q58* are specified in Supplemental Table 7. In these *APOL1* G1/G2 monoallelic individuals, we found that 7.0% of individuals who are low-risk (Q/Q or Q/*) for p.Q58* under its recessive inheritance pattern were diagnosed with ESRD (cases, *n* = 330; controls, *n* = 4,379) compared with 11.6% of individuals who are high-risk (*/*) for p.Q58* (cases, *n* = 29; controls, *n* = 221) with a significant Fischer's exact test (*P* = 0.011). In addition, the average age of onset for ESRD in the p.Q58* low-risk group was 56.9 years compared with 50.6 years in the p.Q58* high-risk group (*t* test, *P* = 0.036). Furthermore, we performed an interaction analysis between *APOL1* G1/G2, modeled under its well-known recessive inheritance pattern, and *APOL3* p.Q58*, also under a recessive model given the results of the stratified analyses. We found that the interaction between the variants was nominally significant (*P* = 0.02), with a negative association coefficient (β = –0.15), suggesting that although there is evidence of variant interaction, any risk conferred by *APOL3* p.Q58* is likely overwhelmed by biallelic *APOL1* G1/G2 risk.

Discussion

The growing scale of genetic association studies has powered the discovery of novel disease variants, increased our understanding of disease pathogenesis, and spurred the development of precision medicine therapeutics (13–15). Yet, of all the GWAS currently compiled in the GWAS catalog, approximately 95% of all GWAS participants are genetically similar to EUR reference populations (16). The lack of population diversity not only limits the study and discovery of non-EUR variants to those with high penetrance and large effect sizes (17–20), but also hinders the generalizability of any GWAS discoveries at risk of further compounding existing health disparities (21–23). Even though using increasingly diverse population cohorts
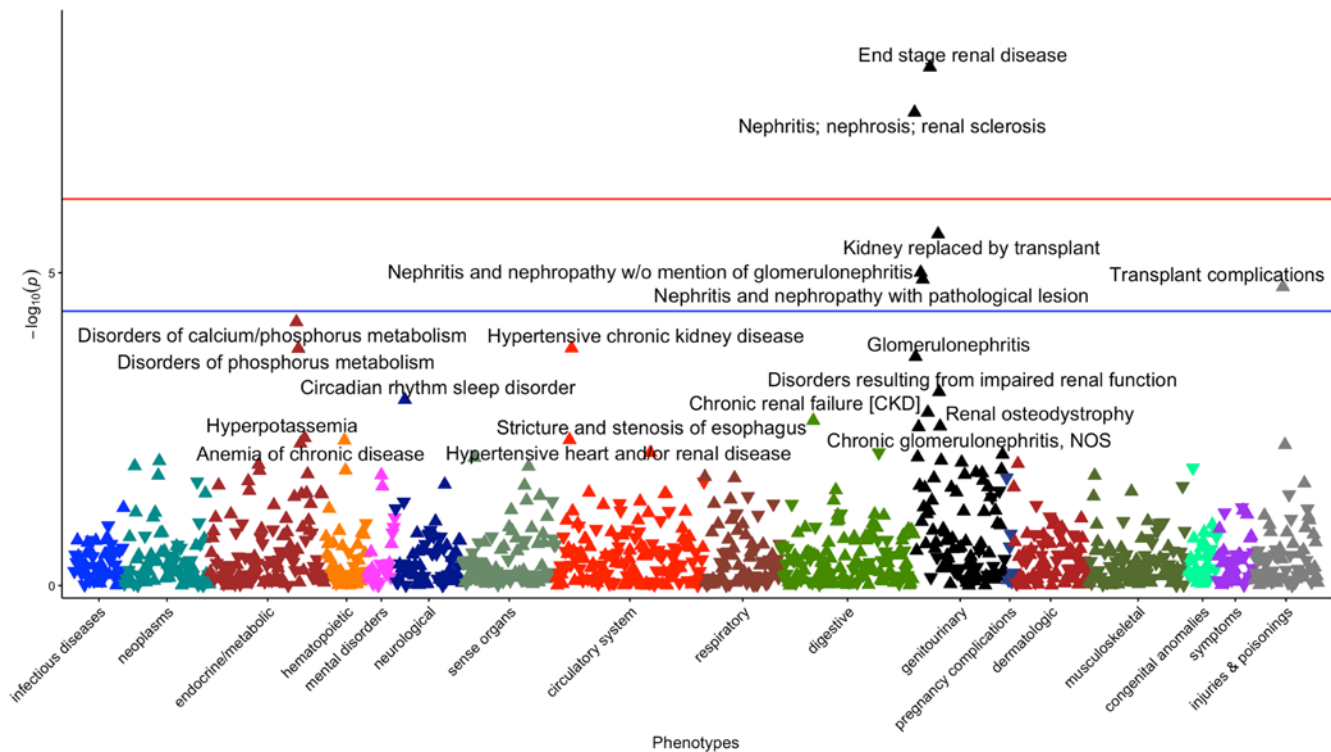
Figure 3. Single-variant PheWAS for *APOL3* variant rs11089781 in the PMBB AFR population. The red line represents the Bonferroni-adjusted *P* value significance threshold ($P = 6.63 \times 10^{-7}$) from the complete AFR associations of all 62 variants. The blue line represents the Bonferroni-adjusted *P* value significance threshold for the single-variant PheWAS of 4.11×10^{-5}.

will begin to mitigate these concerns, further enriching a genetic association study specifically for variants common in non-EUR populations may also enhance our ability to uncover new genetic variant associations. It is well-known that 2 coding alleles within the *APOL1* gene, G1 and G2, found almost exclusively in individuals genetically similar to West African populations, contribute substantially to risk for CKD. Taking a genome-first approach, we used a medical biobank enriched in participants genetically similar to AFR reference populations with whole-exome genomic data linked to rich phenotypic data to perform PheWAS of protein-coding variants in the 6 *APOL* genes. After correction for multiple testing, we identified several variants in the *APOL* gene family predominantly represented in the AFR population and significantly associated with CKD. Of particular interest, we identified a stop-gain variant, p.Q58*, in the *APOL3* gene, with an AFR MAF of 0.211 and EUR MAF of <0.001, that is significantly associated with CKD risk independent of *APOL1* G1/G2. These results highlight the value of combining targeted genome-first approaches with PheWAS in diverse patient biobanks for better understanding genetic risk for kidney disease.

Our initial analysis identified 3 significant variants other than *APOL1* G1/G2 with significant kidney disease associations. rs2239785 is a missense variant in *APOL1* that has been previously reported to be linked with risk for nondiabetic nephropathy and focal segmental glomerulosclerosis (24, 25). Another missense variant, rs7285167, in *APOL2* has been shown to be weakly associated with all-cause ESRD (26); additionally, it has been identified in a *APOL2* protein-specific quantitative trait loci study as strongly associated with protein levels (27). However, this variant is in moderate linkage with *APOL1* G1/G2 ($r^2 = 0.264$). Furthermore, in our present study, the significant association signals for both variants disappeared when conditioning on *APOL1* G1/G2 recessive risk, decreasing our confidence that they play an independent role in renal disease risk. A third variant that we identified, rs11089781, is a stop-gain variant in *APOL3* at amino acid position 58 of 402 (p.Q58*), thereby truncating most of the peptide and likely inhibiting its wild-type function. This specific nonsense mutation is also documented as likely to trigger nonsense-mediated decay (28). rs11089781 has previously been identified to be weakly associated with nondiabetic nephropathy in small African American and Hispanic American populations, but the finding did not always replicate successfully (29, 30). This *APOL3* p.Q58* variant is in minimal LD with the APOL1 G1/G2 alleles and upon conditioning on *APOL1* G1/G2, remained nominally significantly associated with CKD. The average age of onset for ESRD in individuals who are homozygous for the *APOL3* variant was significantly younger

Table 2. Conditional associations on *APOL1* G1/G2 carrier status

	Phecode	OR	95% CI	*P* value	Phenotype
rs2239785 (*APOL1*)					
	585.32	0.9182	(0.819–1.029)	1.43×10^{-1}	ESRD
	587	1.0331	(0.884–1.208)	6.83×10^{-1}	Kidney replaced by transplant
rs7285167 (APOL2)					
	585.32	1.0702	(0.952–1.203)	2.55×10^{-1}	ESRD
	587	1.0777	(0.918–1.265)	3.59×10^{-1}	Kidney replaced by transplant
	580	1.1525	(0.992–1.339)	6.36×10^{-2}	Nephritis; nephrosis; renal sclerosis
	401.22	1.0777	(0.977–1.189)	1.36×10^{-1}	Hypertensive chronic kidney disease
rs11089781 (*APOL3*)					
	585.32	1.1567	(1.025–1.306)	9.35×10^{-3}	ESRD
	580	1.2071	(1.032–1.412)	9.32×10^{-3}	Nephritis; nephrosis; renal sclerosis

Conditional associations for all significant variant-phenotype associations conditioned on *APOL1* G1/G2 carrier status modeled under a recessive inheritance pattern. Significant phenotype associations for *APOL1* G1/G2 are not included to avoid conditioning on themselves.

than that of individuals who do not carry the p.Q58* variant. Furthermore, we replicated this significant association of *APOL1* p.Q58* with CKD in both the MVP and the *All of Us* Research Program. Finally, we found that *APOL3* p.Q58* was significantly associated with a number of quantitative traits associated with CKD, including increased creatinine and decreased eGFR, kidney volume, and surface area.

The APOL3 protein is thought to play a role in pathogen immunity, similar to APOL1, but specifically targeting intracellular pathogens by dissolving their anionic membranes (31). Our genetic data indicate that wild-type APOL3 may play a protective role in CKD, given that the *APOL3* p.Q58* risk variant is very likely to be loss of function and confers increased risk for renal disease. While association results solely in individuals with low-risk *APOL1* G1/G2 genotypes showed that our *APOL3* variant was still strongly associated with increased risk for CKD, our stratified analyses identified that *APOL3* p.Q58* contributes the most risk for renal disease in patients who carry *1 copy* of an *APOL1* G1 or G2 risk allele, suggesting a complex epistatic interaction with *APOL1* G1/G2. Of interest, there is an increasing body of evidence suggesting that *APOL1* G1 and G2 are toxic gain-of-function mutations, despite their observed recessive inheritance pattern (9, 32–34). A previous study on the interactions between APOL1 and APOL3 suggested that deletion of *APOL3* triggers intracellular actomyosin reorganization, increasing susceptibility to kidney disease through *APOL1*-induced podocyte dysfunction and kidney damage (35). The specific mechanisms by which *APOL1* risk alleles induce podocyte dysfunction are still being studied, with evidence that suggests the *APOL1* variant proteins increase endoplasmic reticulum stress, enhance inflammatory signaling within the cells, and interfere with endosomal trafficking (36–38). The fact that *APOL3* p.Q58* contributes no additional significant risk in individuals who carry 0 copies of *APOL1* G1/G2 suggests that the truncated APOL3 protein is not likely to have a toxic gain of function and further supports the presence of some epistatic interaction with the *APOL1* variants. The lack of additional significant risk in patients with 2 *APOL1* G1/G2 risk alleles may indicate that, in this high-risk situation, loss of APOL3 has little effect on further increasing CKD risk. Our interaction analysis also showed a nominally significant association between the interaction of *APOL1* G1/G2 and *APOL3* p.Q58* with CKD, further supporting the presence of some variant-variant interaction. The negative association coefficient indicates that the overall effect of the interaction term is less than the cumulative effects of the 2 separate *APOL1* and *APOL3* risk alleles, supporting the conclusion from the stratified analyses that this truncating *APOL3* variant confers little additional risk in the setting of the high-risk *APOL1* G1/G2 genotype. The concept that loss of *APOL3* promotes the toxic gain of function of a *single* copy of *APOL1* G1/G2 might partially explain the observed positive selection of this *APOL3* stop-gain variant in AFR populations, inferred from the regions of extended homozygosity around the truncating mutation (39). A potential explanation is that selection for the *APOL3* loss-of-function variant may increase the pathogenicity of *APOL1* G1/G2, increasing potential for efficacy against *Trypanosoma* infection, even in the presence of only 1 G1 or G2 allele. We suggest that the protective effects of wild-type APOL3 on CKD, most markedly observed in monoallelic *APOL1* G1/G2 carriers, contribute to the general lack of kidney injury induced by 1 copy of the *APOL1* risk alleles, but not by 2 copies of the *APOL1* risk

Table 3. Stratified associations by *APOL1* G1/G2 carrier status for *APOL3* p.Q58* against ESRD

| | ESRD | | | | | | | |
| | | | Additive | | | Recessive | | |
Stratify by	Cases	Controls	OR	95% CI	*P* value	OR	95% CI	*P* value
APOL1 G1/G2 - 0/1	639	8,837	1.1924	(1.0139–1.4024)	1.67×10^{-2}	1.3271	(1.0851–1.6231)	2.94×10^{-3}
APOL1 G1/G2 - 0	285	4,260	1.3087	(0.9969–1.7180)	5.27×10^{-2}	1.3711	(0.8373–^{2}.2454)	2.10×10^{-1}
APOL1 G1/G2 - 1	354	4,577	1.1235	(0.9347–1.3505)	2.15×10^{-1}	1.3272	(1.0558–1.6685)	7.66×10^{-3}
APOL1 G1/G2 - 2	272	1,160	1.0526	(0.8660–1.2793)	6.07×10^{-1}	1.0207	(0.8506–1.2249)	8.26×10^{-1}

Associations for *APOL3* variant rs11089781 against ESRD stratified by *APOL1* G1/G2 carrier status. Both an additive and a recessive model were used to represent the carrier status for rs11089781.

Table 4. Stratified associations by *APOL1* G1/G2 carrier status for *APOL3* p.Q58* against creatinine and eGFR

| | Maximum creatinine | | | | | | |
| | | Additive | | | Recessive | | |
Stratify by	Counts	β	SE	*P* value	β	SE	*P* value
APOL1 G1/G2 - 0/1	9,062	0.0148	0.0169	3.79×10^{-1}	0.0578	0.0259	2.53×10^{-2}
APOL1 G1/G2 - 0	4,335	0.0353	0.0309	2.53×10^{-1}	0.0655	0.0600	2.75×10^{-1}
APOL1 G1/G2 - 1	4,727	0.0044	0.0220	8.42×10^{-1}	0.0591	0.0298	4.72×10^{-2}
APOL1 G1/G2 - 2	1,373	0.0423	0.0405	2.96×10^{-1}	0.0188	0.0380	6.21×10^{-1}
	Minimum eGFR						
APOL1 G1/G2 - 0/1	9,062	−0.0199	0.0166	2.31×10^{-1}	−0.0727	0.0254	4.17×10^{-3}
APOL1 G1/G2 - 0	4,335	−0.0394	0.0302	1.93×10^{-1}	−0.0699	0.0588	2.35×10^{-1}
APOL1 G1/G2 - 1	4,727	−0.0098	0.0208	6.38×10^{-1}	−0.0757	0.0282	7.20×10^{-3}
APOL1 G1/G2 - 2	1,373	−0.0479	0.0407	2.40×10^{-1}	−0.0197	0.0383	

Associations for *APOL3* variant rs11089781 against maximum creatinine and minimum eGFR stratified by *APOL1* G1/G2 carrier status. Both an additive and a recessive model were used to represent the carrier status for rs11089781.

alleles. This gives rise to the biological manifestation of a recessive inheritance pattern for the *APOL1* G1/G2 alleles while reconciling with their gain-of-function toxicity.

The genome-first approach used in our study to filter for AFR-specific protein-altering variants allowed us to identify and focus on variants that likely would have been omitted in larger association studies in EUR-dominated populations. Many of the variants included in our study are extremely rare in the EUR population (Supplemental Table 1), including this *APOL3* stop-gain variant, and are too rare to study in predominantly EUR cohorts, highlighting the necessity of study cohorts enriched for non-EUR populations. Furthermore, the threshold for significance used in our initial meta-analysis was calculated based on a strict Bonferroni's correction of the total number of genotype-phenotype associations performed, often recommended for PheWAS (40). However, it is evident that not all the variants studied in the *APOL* gene family are in perfect linkage equilibrium and not all phenotypes analyzed are independent of each other. Using an overly strict significance threshold gives us increased confidence in our findings.

We note that there are certain limitations in the context of our work. Phenotyping data derived from EHRs have intrinsic noise and imprecisions. We attempted to mitigate this by using Phecodes that group relevant ICD codes together and applying strict quality control steps on our quantitative clinical traits. We also recognize that there is some selection bias, in that patients with disease are more likely to get laboratory markers measured and to undergo CT imaging. This means that the quantitative traits derived from lab values and imaging metrics we have available may be enriched for individuals with pathological trait values. However, this problem is reduced by the diversity in phenotypes in the PMBB, such that a large proportion of patients have nonrenal conditions and may still have normal kidney-related lab values and imaging traits.

In addition, our LD calculations were based on unphased WES data estimated using maximum likelihoods on the haplotype frequency cubic equation instead of phased haplotype frequencies. Although the computed values may be inexact, we also obtained LD metrics calculated using phased haplotypes from the 1000 Genomes Project as validation. While the overall sample size of the AFR population in PMBB is limited, we replicated our observations in MVP and *All of Us*, both of which have substantial participation from individuals genetically similar to AFR reference populations. However, it is worthwhile to note that the reference populations used for defining population assignments differ between the biobanks, as noted in the Methods.

In conclusion, our study represents a targeted approach to studying population-specific genetic variation in a diverse medical biobank. Our approach identified multiple AFR-specific protein-altering variants in the *APOL* gene family implicated in kidney disease risk, including a stop-gain variant in *APOL3* that increases the risk of CKD primarily in persons carrying 1 *APOL1* G1/G2 risk allele. While it is imperative that population diversity is emphasized when recruiting patients for biobanks and building cohorts for association studies, our genome-first approach that filters for population-specific variants represents a step in that same direction in helping us understand disease risk in underrepresented populations.

Methods

Sex as a biological variable. Cisgender women and men were included the study.

Setting and study participants. All individuals who were recruited for the PMBB are patients of clinical practice sites of the University of Pennsylvania Health System (12). Replication analyses were conducted using genotype and imputed genetic data from the MVP (41) and whole genome sequencing data from the *All of Us* Research Program (42) as well as both of their respective corresponding phenotyping data derived from the EHR.

Exome sequencing. This study included 43,731 individuals in the PMBB with exome sequencing and corresponding EHR-derived traits. Genetic sequencing was performed by the Regeneron Genetics Center using protocols as described previously, and all sequences were mapped to GRCh38 (12). In our population-specific and subsequent meta-analyses, we identified individuals genetically similar to the AFR ($n = 11,198$) and EUR ($n = 30,324$) superpopulations using kernel density estimation as defined by HapMap3 (43). Of note, our WES data was unphased, so LD estimates were calculated by finding the maximum likelihood solution of the haplotype frequency cubic equation as implemented by plink 2.0 (44, 45). As validation, LD metrics were also obtained using phased haplotypes from the 1000 Genomes Project (46).

For replication studies in MVP, we interrogated an additional 121,177 AFR individuals and 449,042 EUR individuals with genotyped and imputed data, obtained as previously described (47, 48). Population assignments were based on genetic similarity using a random forest classifier to respective reference populations from the 1000 Genomes Project. In the *All of Us* Research Program, we used 50,080 AFR individuals and 125,860 EUR individuals with whole-genome sequencing data. Population assignments were determined based on genetic similarity using a random forest classifier to respective reference populations from the Human Genome Diversity Project and 1000 Genomes (https://www.internationalgenome.org/data-portal/data-collection/hgdp). Documentation on data quality and curation are as described previously (49, 50).

Variant annotations. Annotations for variants selected for the initial analyses in PMBB were obtained using ANNOVAR (51). Variants of interest were annotated as pLOF, missense, or in-frame insertion/deletion variants according to the NCBI Reference Sequencing database (https://www.ncbi.nlm.nih.gov/refseq/). pLOF variants were defined as frameshift substitution, stop-gain, or splicing variants. MAFs for each variant were calculated in the relevant PMBB population using their respective allele counts. Only variants with a MAF of >0.1% in the PMBB AFR population were considered in our study. MAFs for all studied variants were also compiled from the Genome Aggregation Database gnomAD (https://gnomad.broadinstitute.org/) (v4.0.0) (52).

Phenotype data collection. ICD-9 and ICD-10 disease diagnosis codes and laboratory measurements were extracted from patient EHRs for the PMBB. Binary phenotypes for each individual were determined by mapping ICD-9 and ICD-10 codes to Phecodes as previously described (53). A rule of 2 was then applied where participants were determined as having a certain disease phenotype if they had the corresponding Phecode diagnosis on 2 or more dates, while phenotypic controls consisted of individuals who never had the Phecode. Individuals with a Phecode diagnosis on only one date were not considered in statistical analyses.

Quantitative laboratory traits were also extracted from patient EHRs for the PMBB. All units were converted to their respective clinical traditional units. After removing outliers (greater than 4 standard deviations from the mean), we recorded the minimum, median, and maximum measurements for each

laboratory measurement and each individual to use for subsequent association analyses. For our imaging-derived phenotypes, the kidney was segmented from abdominal and pelvic CT scans using Total-Segmentator (54). Of the 7,946 unique individuals with scans in which the entirety of the kidney was captured and labeled, subsequent quantitative kidney traits such as volume and surface area were derived using *PyRadiomics* (55). Outliers were then removed in a similar fashion and a minimum, median, and maximum value for each trait and individual was computed and used for downstream analysis.

Phecodes were classified in an identical fashion in the MVP cohort. Quantitative traits were extracted as previously described (48). Binary Phecodes and quantitative traits were computed in the *All of Us* cohort using the same methods as the PMBB cohort.

PheWAS. Within the AFR and EUR population groups in PMBB, we performed a PheWAS for each of the 62 variants of interest against 1,222 binary phenotypes with at least 20 cases in both the PMBB AFR and EUR population. We used the generalized linear-mixed model framework to account for participant relatedness and unbalanced case-control ratios with the SAIGE package (56). Directly genotyped variants were used for step 1 of SAIGE. Whole-exome sequenced variants were used for step 2 of SAIGE. Analyses were adjusted for sex, age, age^2, and 5 population-specific genetic principal components (PCs) in the AFR population and 10 PCs in the EUR population. A fixed-effects meta-analysis was then performed using the inverse variance method for pooling as implemented by the "meta" R package (57). A Bonferroni-adjusted significance threshold was then computed using the total number of associations performed in both genetically inferred population groups. A suggestive significance threshold was also calculated by adjusting 1 by the total number of associations.

Replication for specific variants and phenotypes of interests were also performed using SAIGE in MVP, adjusting for age, sex, and 10 population-specific PCs (48). Identical methods were used for replication in the *All of Us* cohort as in PMBB, including using SAIGE for association analyses.

Conditional and stratified analysis. Summary statistics for conditional associations were calculated from the conditional normal distribution implemented by the SAIGE package (56). Recognizing that a single haplotype would never carry both G1 and G2 (6), we collapsed the alleles into a single risk allele based on the total number of G1/G2 alleles an individual carries. We coded all monoallelic G1/G2 allele carriers as homozygous reference allele carriers given the recessive nature of the *APOL1* G1/G2 risk alleles. Identical variables were adjusted for as in the unconditioned associations.

The same single risk allele that represents *APOL1* risk was used to stratify the study population for our stratified analyses. Associations were also performed using SAIGE and adjusted for the same variables. Both an additive and a recessive inheritance pattern were used to model rs11089781 risk. Interaction analyses were performed in R, using the same recessive inheritance models and adjusting for the same covariates.

Statistics. Generalized linear-mixed models were used for all association studies, including conditional and stratified, to account for participant relatedness and unbalanced case-control ratios as implemented by the SAIGE package (56). Subsequent meta-analysis was performed under a fixed-effects model using the "meta" R package (57). Statistical tests including the 2-tailed Student's *t* test, Fischer's exact test, and the χ^2 test were performed in R. A strict Bonferroni's correction was applied to all PheWAS summary statistics to determine the threshold for significance and correct for multiple hypothesis testing. $P < 6.63 \times 10^{-7}$ was considered significant.

Study approval. Appropriate consent was obtained from each participant regarding storage of biological specimens, genetic sequencing, access to all available EHR data, and permission to recontact for future studies. The study was approved by the Institutional Review Board of the University of Pennsylvania and complied with the principles set out in the Declaration of Helsinki.

Data availability. All summary statistics for significant variant-phenotype associations in the PMBB, as well as significant replications from each replication cohort, are fully detailed in the main text and in Supplemental Tables 2–6. A list of all the single variants used in this study is provided in Supplemental Table 1. All other relevant data values can be found in the Supporting Data Values file. Individual-level data are not publicly available due to research participant privacy concerns; however, requests from accredited researchers for access to individual-level data relevant to this article can be made by contacting the corresponding author. Details on defining phenotypes in MVP are as previously described (48). This study also used data from the *All of Us* Research Program's Controlled Tier Dataset v7, available to authorized users on the Researcher Workbench. Code for defining phenotypes used in PMBB and *All of Us* can be found at https://github.com/davidz987/APOL3_CKD/commit/426ceff58f72e4368497a7f52c97aa5047d52969.

Author contributions

All authors reviewed and approved the submitted version of the manuscript. DYZ, MDR, and DJR conceived the idea, designed the project, acquired the data, and interpreted the results. DYZ performed all the analyses in the PMBB and *All of Us* and wrote the manuscript. MGL and SMD performed all analyses in the MVP. JTD and WRW acquired and processed all the CT imaging data. LGL provided critical guidance on interpreting and communicating results for population-specific findings.

Acknowledgments

Address correspondence to: Daniel J. Rader, 11-125 Smilow Center for Translational Research, 3400 Civic Center Blvd., Philadelphia, Pennsylvania 19104-5158, USA. Phone: 215.573.4176; Email: rader@pennmedicine.upenn.edu.

1. Murphy D, et al. Trends in prevalence of chronic kidney disease in the United States. *Ann Intern Med*. 2016;165(7):473–481.
2. Coresh J, et al. Prevalence of chronic kidney disease in the United States. *JAMA*. 2007;298(17):2038–2047.
3. Genovese G, et al. Association of trypanolytic ApoL1 variants with kidney disease in African Americans. *Science*. 2010;329(5993):841–845.
4. Friedman DJ, et al. Population-based risk assessment of APOL1 on renal disease. *J Am Soc Nephrol*. 2011;22(11):2098–2105.
5. Tzur S, et al. Missense mutations in the APOL1 gene are highly associated with end stage kidney disease risk previously attributed to the MYH9 gene. *Hum Genet*. 2010;128(3):345–350.
6. Limou S, et al. APOL1 kidney risk alleles: population genetics and disease associations. *Adv Chronic Kidney Dis*. 2014;21(5):426–433.
7. Kopp JB, et al. APOL1 genetic variants in focal segmental glomerulosclerosis and HIV-associated nephropathy. *J Am Soc Nephrol*. 2011;22(11):2129–2137.
8. Smith EE, Malik HS. The apolipoprotein L family of programmed cell death and immunity genes rapidly evolved in primates at discrete sites of host-pathogen interactions. *Genome Res*. 2009;19(5):850–858.
9. McCarthy GM, et al. Recessive, gain-of-function toxicity in an APOL1 BAC transgenic mouse model mirrors human APOL1 kidney disease. *Dis Model Mech*. 2021;14(8):dmm048952.
10. Hung AM, et al. Genetic inhibition of APOL1 pore-forming function prevents APOL1-mediated kidney disease. *J Am Soc Nephrol*. 2023;34(11):1889–1899.
11. Monajemi H, et al. The apolipoprotein L gene cluster has emerged recently in evolution and is expressed in human vascular tissue. *Genomics*. 2002;79(4):539–546.
12. Verma A, et al. The Penn Medicine BioBank: towards a genomics-enabled learning healthcare system to accelerate precision medicine in a diverse population. *J Pers Med*. 2022;12(12):1974.
13. Ritchie MD. The success of pharmacogenomics in moving genetic association studies from bench to bedside: study design and implementation of precision medicine in the post-GWAS era. *Hum Genet*. 2012;131(10):1615–1626.
14. Ashley EA. Towards precision medicine. *Nat Rev Genet*. 2016;17(9):507–522.
15. Denny JC, et al. Phenome-wide association studies as a tool to advance precision medicine. *Annu Rev Genomics Hum Genet*. 2016;17(1):353–373.
16. Mills MC, Rahal C. The GWAS Diversity Monitor tracks diversity by disease in real time. *Nat Genet*. 2020;52(3):242–243.
17. Piel FB, et al. Sickle cell disease. *N Engl J Med*. 2017;376(16):1561–1573.
18. Freedman BI, et al. APOL1-associated nephropathy: a key contributor to racial disparities in CKD. *Am J Kidney Dis*. 2018;72(5 suppl 1):S8–S16.
19. Kilpeläinen TO, et al. Multi-ancestry study of blood lipid levels identifies four loci interacting with physical activity. *Nat Commun*. 2019;10(1):376.

20. Mason PJ, et al. G6PD deficiency: the genotype-phenotype association. *Blood Rev*. 2007;21(5):267–283.

21. Sirugo G, et al. The missing diversity in human genetic studies. *Cell*. 2019;177(1):26–31.

22. Popejoy AB, Fullerton SM. Genomics is failing on diversity. *Nature*. 2016;538(7624):161–164.

23. Martin AR, et al. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat Genet*. 2019;51(4):584–591.

24. Bostrom MA, et al. Genetic association and gene-gene interaction analyses in African American dialysis patients with nondiabetic nephropathy. *Am J Kidney Dis*. 2012;59(2):210–221.

25. Genovese G, et al. A risk allele for focal segmental glomerulosclerosis in African Americans is located within a region containing APOL1 and MYH9. *Kidney Int*. 2010;78(7):698–704.

26. Cooke Bailey JN, et al. Analysis of coding variants identified from exome sequencing resources for association with diabetic and non-diabetic nephropathy in African Americans. *Hum Genet*. 2014;133(6):769–779.

27. Battle A, et al. Genomic variation. Impact of regulatory variation from RNA to protein. *Science*. 2015;347(6222):664–667.

28. Yngvadottir B, et al. A genome-wide survey of the prevalence and evolutionary forces acting on human nonsense SNPs. *Am J Hum Genet*. 2009;84(2):224–234.

29. Skorecki KL, et al. A null variant in the apolipoprotein L3 gene is associated with non-diabetic nephropathy. *Nephrol Dial Transplant*. 2017;33(2):323–330.

30. Hawkins GA, et al. Re-sequencing of the APOL1-APOL4 and MYH9 gene regions in African Americans does not identify additional risks for CKD progression. *Am J Nephrol*. 2015;42(2):99–106.

31. Gaudet RG, et al. A human apolipoprotein L with detergent-like activity kills intracellular pathogens. *Science*. 2021;373(6552):eabf8113.

32. Limou S, et al. APOL1 toxin, innate immunity, and kidney injury. *Kidney Int*. 2015;88(1):28–34.

33. Granado D, et al. Intracellular APOL1 risk variants cause cytotoxicity accompanied by energy depletion. *J Am Soc Nephrol*. 2017;28(11):3227–3238.

34. Johnstone DB, et al. APOL1 null alleles from a rural village in India do not correlate with glomerulosclerosis. *PLoS One*. 2012;7(12):e51546.

35. Uzureau S, et al. APOL1 C-terminal variants may trigger kidney disease through interference with APOL3 control of actomyosin. *Cell Rep*. 2020;30(11):3821–3836.

36. Wen H, et al. APOL1 risk variants cause podocytes injury through enhancing endoplasmic reticulum stress. *Biosci Rep*. 2018;38(4):BSR20171713.

37. Wakashin H, et al. APOL1 renal risk variants exacerbate podocyte injury by increasing inflammatory stress. *BMC Nephrol*. 2020;21(1):371.

38. Beckerman P, et al. Transgenic expression of human APOL1 risk variants in podocytes induces kidney disease in mice. *Nat Med*. 2017;23(4):429–438.

39. Rausell A, et al. Common homozygosity for predicted loss-of-function variants reveals both redundant and advantageous effects of dispensable human genes. *Proc Natl Acad Sci U S A*. 2020;117(24):13626–13636.

40. Bastarache L, et al. Phenome-wide association studies. *JAMA*. 2022;327(1):75–76.

41. Gaziano JM, et al. Million veteran program: a mega-biobank to study genetic influences on health and disease. *J Clin Epidemiol*. 2016;70:214–223.

42. Denny JC, et al. The "All of Us" research program. *N Engl J Med*. 2019;381(7):668–676.

43. Altshuler DM, et al. Integrating common and rare genetic variation in diverse human populations. *Nature*. 2010;467(7311):52–58.

44. Chang CC, et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*. 2015;4:7.

45. Gaunt TR, et al. Cubic exact solutions for the estimation of pairwise haplotype frequencies: implications for linkage disequilibrium analyses and a web tool 'CubeX'. *BMC Bioinformatics*. 2007;8(1):428.

46. Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics*. 2015;31(21):3555–3557.

47. Hunter-Zinck H, et al. Genotyping array design and data quality control in the Million Veteran Program. *Am J Hum Genet*. 2020;106(4):535–548.

48. Verma A, et al. Diversity and scale: genetic architecture of 2068 traits in the VA Million Veteran Program. *Science*. 2024;385(6706):eadj1182.

49. Ramirez AH, et al. The All of Us research program: data quality, utility, and diversity. *Patterns (N Y)*. 2022;3(8):100570.

50. NIH. Data Methods: All of Us Reearch Hub. https://www.researchallofus.org/methods/. Accessed August 22, 2024.

51. Wang K, et al. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38(16):e164.

52. Karczewski KJ, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*. 2020;581(7809):434–443.

53. Wu P, et al. Mapping ICD-10 and ICD-10-CM codes to PheCodes: workflow development and initial evaluation. *JMIR Med Inform*. 2019;7(4):e14325.

54. Wasserthal J, et al. TotalSegmentator: robust segmentation of 104 anatomic structures in CT images. *Radiol Artif Intell*. 2023;5(5):e230024.

55. van Griethuysen JJM, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res*. 2017;77(21):e104–e107.

56. Zhou W, et al. Scalable generalized linear mixed model for region-based association tests in large biobanks and cohorts. *Nat Genet*. 2020;52(6):634–639.

57. Schwarzer G. crossnma: an R package to synthesize cross-design evidence and cross-format data using network meta-analysis and network meta-regression. *BMC Med Res Methodol*. 2007;24(1):169.