Platform-dependent effects of genetic variants on plasma APOL1 and their implications for kidney disease

Qingbo S. Wang[1], Jinguo Huang[1], Leanne Chan[1], Nicole Haste[1], Niclas Olsson[1], Aleksandr Gaun[1,2], Fiona McAllister[1], Deepthi Madhireddy[1], Amos Baruch[1], Eugene Melamud[1], Anastasia Baryshnikova[1,3]

1. Calico Life Sciences LLC, South San Francisco, CA, USA
2. Present address: Flagship Pioneering, Cambridge, MA, USA
3. Corresponding author. E-mail: abaryshk@gmail.com

Author contributions are listed at the end of this document.

Abstract

Mutations in apolipoprotein L1 (APOL1) are strongly associated with an increased risk of kidney disease in individuals of African ancestry, yet the underlying mechanisms remain largely unknown. Plasma proteomics provides opportunities to elucidate mechanisms of disease by studying the effects of disease-associated variants on circulating protein levels. Here, we examine the genetic drivers of circulating APOL1 in individuals of African and European ancestry from four independent cohorts (UK Biobank, AASK, deCODE and Health ABC) employing three proteomic technologies (Olink, SomaLogic and mass spectrometry). We find that disease-associated *APOL1* G1 and G2 variants are strong pQTLs for plasma APOL1 in Olink and SomaLogic, but the direction of their effects depends on the proteomic platform. We identify an additional *APOL1* missense variant (rs2239785), common in Europeans, exhibiting the same platform-dependent directional discrepancy. Similarly, variants in the kallikrein-kinin pathway (*KLKB1*, *F12*, *KNG1*) and their genetic interactions exhibit strong *trans*-pQTL effects for APOL1 measured by Olink, but not SomaLogic. To explain these discrepancies, we propose a model in which *APOL1* mutations and the kallikrein-kinin pathway influence the relative abundance of two distinct APOL1 forms, corresponding to APOL1 bound to trypanolytic factors 1 and 2, which are differentially recognized by different proteomic platforms. We hypothesize that this shift in relative abundance of APOL1 forms may contribute to the development of kidney disease.

Introduction

Chronic kidney disease (CKD) is a complex condition characterized by gradual loss of kidney function with multiple genetic and environmental influences (Obrador *et al.*, 2017). Individuals of

West African descent exhibit a disproportionately high risk of developing CKD, leading to increased rates of kidney failure and mortality (Ferguson, 1991). This ancestry-based disparity is largely attributed to coding variants in the *APOL1* gene, termed G1 and G2, which are common in African populations but virtually absent elsewhere (**Table 1**) (Genovese *et al.*, 2010; Tzur *et al.*, 2010). The G1/G2 variants have been linked to markedly increased odds ratio (OR) for several renal pathologies, including hypertension-associated end-stage renal disease (H-ESRD; OR = 7–10), focal segmental glomerulonephrosis (FSGS; OR = 10–17) and HIV-associated nephropathy (HIVAN; OR = 29–89) (Friedman and Pollak, 2020).

The *APOL1* gene encodes apolipoprotein L1, a primate-specific protein expressed in multiple tissues and circulating in plasma as a component of high-density lipoprotein (HDL) particles (Shiflett *et al.*, 2005). In addition to APOL1, these HDL particles, termed trypanolytic factors 1 (TLF-1) and 2 (TLF-2), include apolipoprotein A-I (APOA1), haptoglobin-related protein (HPR) and, in the case of TLF-2, IgM molecules (Raper et al., 1999; Weckerle et al., 2016). Circulating APOL1 provides protection against African sleeping sickness by forming cytotoxic ion channels in the disease-causing parasite *Trypanosoma brucei*, leading to its destruction (Friedman and Pollak, 2020). The G1 and G2 variants in *APOL1*, resulting in two amino acid substitutions (referred to as G1M and G1G, in near perfect linkage disequilibrium) and a two amino acid deletion, respectively (**Table 1**), confer enhanced protection against a parasite subspecies, *T. brucei rhodesiense*, by preventing the binding of an APOL1-inhibiting factor (Friedman and Pollak, 2020). This enhanced protection likely provided a survival advantage to G1/G2 carriers in regions where the parasite is prevalent (Friedman and Pollak, 2011).

While the roles of APOL1 and its G1/G2 mutations in *Trypanosoma* resistance are well-established, the mechanisms by which G1/G2 increase CKD risk are less clear. A prevailing hypothesis suggests that, similar to its function in parasite defense, APOL1 acts as an ion channel in human podocytes. The G1/G2 variants are thought to dysregulate APOL1's channel function, causing ionic imbalance and kidney cell death (Lan *et al.*, 2014; Olabisi *et al.*, 2016; Giovinazzo *et al.*, 2020). This hypothesis is supported by promising results from preclinical and early clinical studies indicating that blocking APOL1 channels prevents podocyte injury, reduces proteinuria and preserves kidney function (Egbuna *et al.*, 2023; Vasquez-Rios, De Cos and Campbell, 2023). While most research has focused on APOL1 expressed in the kidneys, the role of circulating APOL1 has been less extensively studied, likely due to early reports of no association between plasma APOL1 levels and kidney disease (Bruggeman *et al.*, 2014; Kozlitina *et al.*, 2016).

Recent advancements in proteome profiling methods have greatly expanded our capacity to investigate the contribution of circulating proteins to disease. In particular, high throughput proteomic methods combined with genetic analyses of large collections of human plasma samples have driven the discovery of disease-relevant protein quantitative trait loci (pQTLs), i.e. genetic variants associated with both disease risk and altered protein levels in the plasma (Zhang *et al.*, 2022; Sun *et al.*, 2023). Large-scale plasma pQTL studies are enabled primarily by two high throughput and high sensitivity technologies: the proximity extension assay (PEA) commercialized by Olink, which uses target-specific antibodies tagged with molecular barcodes (Fredriksson *et al.*, 2002), and the DNA aptamer technology commercialized by SomaLogic,

which relies on target-specific single-stranded oligonucleotides (Gold *et al.*, 2010). While both technologies provide quantitative (though relative) and internally reproducible measurements of protein abundance, direct comparisons have revealed a wide range of correlations between the two platforms (Eldjarn *et al.*, 2023).

In this study, we present a comprehensive analysis of APOL1 plasma pQTLs in individuals of African descent, comparing data from the UK Biobank, generated with Olink's technology, and the African American Study of Kidney Disease and Hypertension (AASK), generated with SomaLogic's technology. Our findings reveal a key difference: while the G1/G2 mutations have strong pQTL effects on both platforms, they are associated with increased APOL1 plasma levels when measured by Olink and decreased levels when measured by SomaLogic. By comparing pQTLs in European individuals across both platforms, we confirmed that this discrepancy is specific to protein-altering variants in the *APOL1* gene itself, rather than non-coding *cis*-regulatory variants. Furthermore, we find that genes in the kallikrein-kinin pathway (*KLKB1*, *F12*, and *KNG1*) have strong *trans*-pQTL effects on APOL1, but these effects are only observed in the Olink data. We hypothesize that these platform-specific effects reflect differential affinity of the Olink and SomaLogic reagents to two distinct forms of APOL1, associated with TLF-1 and TLF-2 particles, respectively (**Fig. 7**). The relative abundance of these APOL1 complexes, influenced by both *cis* and *trans* APOL1 pQTLs, may impact APOL1 and IgM uptake into the kidneys and thus provide valuable insights into the mechanisms underlying G1/G2-associated kidney disease.

Results

G1 and G2 mutations increase Olink-based estimates of plasma APOL1 levels ($APOL1_{Olink}$) in UKBB

To explore a potential relationship between circulating APOL1 and kidney disease, we analyzed the association between the G1/G2 risk variants and plasma APOL1 levels in 828 individuals of African descent (AFR) from the UK Biobank (Sudlow *et al.*, 2015). As the UK Biobank study used Olink's PEA technology for relative protein quantification (Fredriksson *et al.*, 2002), we denote these measurements as $APOL1_{Olink}$ to distinguish them from data obtained via other platforms.

Due to concerns regarding the accuracy of imputation for the G1 variant (Methods, **Fig. S1**), we utilized both exome sequences and imputed genotypes. This analysis revealed a strong association signal at the *APOL1* locus ($p < 10^{-100}$; **Fig. 1A**), with multiple variants correlating with elevated plasma $APOL1_{Olink}$. Statistical fine-mapping of the region provided strong evidence that the G1 and G2 mutations are the causal determinants of this association, with a posterior inclusion probability (PIP) exceeding 0.99 for G2 and 0.76 for G1 (sum of 0.44 for G1M and 0.32 for G1G; **Fig. 1A**). Due to the near-perfect linkage between G1G and G1M, we used G1M as a proxy for both variants in subsequent analyses. Notably, due to the linkage between G1 and G2 (Pearson R = –0.27), the effect of G1 on $APOL1_{Olink}$ levels is only detected after adjusting for G2

($p = 0.23$ and $p = 1.3\times10^{-19}$ before and after adjustment, respectively). Adjusting for both G1 and G2 abolished the association signals of all other variants ($p < 10^{-100}$ and $p > 5\times10^{-8}$ before and after adjustment, respectively), confirming that G1 and G2 are the main *cis*-genetic determinants of plasma $APOL1_{Olink}$ levels in the AFR population (**Fig. 1A**).

While most previous studies have suggested a recessive inheritance model for the G1/G2 mutations in kidney disease (where two risk alleles are required for the effect), our analysis of plasma $APOL1_{Olink}$ levels favored an additive model (**Fig. 1B–C**). Specifically, each copy of the G1 or G2 risk allele contributed to increasing $APOL1_{Olink}$, such that heterozygous individuals exhibited intermediate $APOL1_{Olink}$ levels relative to those with zero or two copies of the variant (**Fig. 1B–C**). Notably, the effect of G2 in the absence of G1 ($\beta_{G2|G1=0} = 1.212$) was considerably stronger than the effect of G1 without G2 ($\beta_{G1|G2=0} = 0.287$; **Fig. 1B–C; Fig. S2A**). Both the additive nature of the risk alleles and the stronger influence of G2 relative to G1 align with the most recent study of CKD in West Africans which reported higher odds of disease for G2 compared to G1 and an intermediate disease risk for individuals heterozygous for either mutation (Gbadegesin *et al.*, 2024).

The presence of G2 appeared to further attenuate the effect of G1 ($\beta_{G1|G2=1} = 0.066$; **Fig. 1C**), indicating that G1 and G2 share a genetic interaction ($p = 0.0094$ for difference between $\beta_{G1|G2=0}$ and $\beta_{G1|G2=1}$; Methods). Similarly, G2 showed a genetic interaction with the APOL1 M1 variant (rs73885316, also known as p.N264K; **Table 1**), which did not exhibit an effect on its own but attenuated the effect of G2 ($\beta_{G2|(G1=0\&M1>0)} = 0.985$; $p = 0.032$ for difference between $\beta_{G2|(G1=0\&M1=0)}$ and $\beta_{(G2|G1=0\&M1>0)}$; **Fig. S2F**). This attenuation is consistent with a prior report (Adamson *et al.*, 2024) and supports a protective effect of M1 on G2-associated risk of kidney disease (Gupta *et al.*, 2023).

<u>The effects of G1 and G2 mutations on plasma APOL1 levels are platform-specific</u>

We sought to replicate the observed pQTL effects of G1/G2 mutations on plasma $APOL1_{Olink}$ in an independent cohort of similar size and genetic ancestry: the African American Study of Kidney Disease and Hypertension (AASK) (Appel *et al.*, 2003). This study includes 461 African American participants with plasma APOL1 levels measured using SomaLogic's aptamer technology (Surapaneni *et al.*, 2022). We denote these measurements as $APOL1_{SomaLogic}$. Analysis of pQTL summary statistics from AASK revealed a significant association between the G2 mutation and $APOL1_{SomaLogic}$ levels ($p = 3.2\times10^{-8}$). However, in contrast to our UK Biobank findings, this mutation was associated with an apparent decrease in APOL1 concentration ($\beta = -0.45$; **Fig. 2A**; **Table 2**). The direction of the effect of G1 was consistent with G2 ($\beta = -0.10$; **Fig. 2A**; **Table 2**), although it did not reach genome-wide significance likely due to linkage with G2.

To address the conflicting pQTL effects observed between the UK biobank (Olink) and AASK (SomaLogic), we sought further evidence from a third independent cohort: the Health, Aging and Body Composition (Health ABC) study (Newman *et al.*, 2023). We used Olink's technology and mass spectrometry to measure plasma APOL1 levels in 560 and 1,094 Health ABC

participants (including 194 and 390 individuals of African descent), respectively (Methods). As genotype data were not directly available, we inferred G1/G2 mutation status using peptide sequence analysis (Methods; **Fig. S3A**). Our data confirmed and further extended the platform-dependent G1/G2 effects (**Table 2**). Despite a robust correlation of APOL1 measurement between Olink and mass spectrometry ($R = 0.39$, $p = 4.2\times10^{-20}$; **Fig. S3C**), G1/G2 mutations were associated with increased APOL1 levels measured by Olink ($\beta = 0.59$, $p = 9.1\times10^{-4}$), but had no significant effect on APOL1 levels measured by mass spectrometry ($\beta = 0.069$, $p = 0.16$; $p = 3.5\times10^{-3}$ for heterogeneity of effect size; **Fig. 2B**). The lack of G1/G2 effect with mass spectrometry is consistent with previous studies which used enzyme-linked immunosorbent assay (ELISA), Western blots and mass spectrometry, and also reported no association between G1/G2 mutations and plasma APOL1 levels (Bruggeman *et al.*, 2014; Weckerle *et al.*, 2016; Andrews *et al.*, 2022).

These findings collectively suggest that, while G1/G2 mutations do not impact total APOL1 levels (as measured by mass spectrometry), they affect Olink- and SomaLogic-based quantification, resulting in an apparent anti-correlation of pQTL effects. This intriguing phenomenon may provide insights into how these mutations influence APOL1 function and thus merits further investigation.

<u>Platform-specific *trans*-pQTLs do not support an “epitope effect” hypothesis</u>

One possible explanation for the observed differences in G1/G2 pQTL effects between Olink and SomaLogic is that the reagents employed by each technology have differential sensitivity to protein conformational changes induced by the mutations – the so-called “epitope effect” (Ferkingstad *et al.*, 2021; Eldjarn *et al.*, 2023). Specifically, Olink’s antibodies may have higher affinity for the mutated APOL1 than for the reference protein, whereas SomaLogic’s aptamers may have lower affinity (**Fig. 2C**). While this is a plausible hypothesis, we are unable to test it directly due to the proprietary nature of Olink’s reagents. Instead, we reasoned that, if the epitope effect is indeed responsible for the observed differences, then the G1/G2 pQTL effects, and potentially those of other coding variants in APOL1, should be the sole differentiating factor between platforms, as only coding variants are expected to impact protein conformation. In contrast, *trans*-pQTL effects should remain consistent across platforms, as they reflect pathways regulating APOL1 abundance, independently of its conformation. To test this hypothesis, we examined APOL1 pQTLs in European populations genome-wide, taking advantage of their larger sample size and greater power to identify *trans*-pQTLs.

We conducted a genome-wide pQTL analysis of plasma APOL1 levels in UK Biobank participants with European ancestry ($n = 38{,}442$), measured using the Olink platform, and compared the results to pQTL summary statistics from the deCODE study ($n = 35{,}559$ Icelanders), which used SomaLogic’s technology (Ferkingstad *et al.*, 2021). Importantly, a previous analysis directly comparing Olink and SomaLogic measurements in 1,500 Icelandic plasma samples demonstrated a high correlation between the two platforms for overall APOL1 levels (Spearman $\rho = 0.69$) (Eldjarn *et al.*, 2023). This correlation suggests that any potential

platform-specific pQTLs would not be due to global inconsistencies between the platforms, but rather to differences in how the platforms interact with genetic variants.

Our comparison identified a shared *cis*-pQTL signal at the *APOL1* locus (chromosome 22) and a shared *trans*-pQTL signal at the *HPR* locus (chromosome 16), as well as three Olink-specific *trans*-pQTLs at the *KNG1*, *KLKB1* and *F12* loci (chromosomes 3, 4 and 5, respectively; **Fig. 3A**). Consistent with the epitope hypothesis, the effects of the shared *cis*-pQTLs exhibited a strong negative correlation between Olink and SomaLogic measurements (**Fig. 3B**), whereas the effects of shared *trans*-pQTLs were positively correlated (**Fig. 3C**). However, the Olink-specific *trans*-pQTLs at the *KNG1*, *KLKB1* and *F12* loci cannot be readily explained by the epitope hypothesis, suggesting that a more complex mechanism may underlie the observed differences between the platforms. To investigate such potential mechanisms, we examined each pQTL signal in greater detail.

<u>An APOL1 missense variant common in Europeans shows functional similarity to G1/G2</u>

The APOL1 pQTL signal on chromosome 22 localized to the *APOL1* gene region (**Fig. 4A**), and was strong ($p < 10^{-320}$) despite the virtual absence of G1/G2 mutations in European populations. Statistical fine-mapping of this region in the UK Biobank dataset identified three variants as potentially causal for the association: a common missense variant (rs2239785, hereafter referred to as V1; PIP = 0.75; **Table 1**), a common 5'-UTR variant (rs6000220, hereafter referred to as V2; PIP > 0.99; **Table 1**) and a rare missense variant (rs148296684; PIP > 0.99; **Table 1**). Due to its low frequency (MAF < 0.001, explaining less than 0.2% of the total variance in $APOL1_{Olink}$), we could not further characterize the rare missense variant. Adjusting for the two common variants substantially diminished the association signal of all other variants in the region ($p < 10^{-320}$ and $p > 10^{-20}$ before and after adjustment, respectively; **Fig. 4A**).

Comparing the effects of the missense variant (V1) and the 5'-UTR variant (V2) between the UK Biobank (Olink) and deCODE (SomaLogic) revealed a notable difference. While the 5'-UTR variant consistently showed a negative association with APOL1 in both datasets, the V1 variant exhibited opposite directional effects (**Fig. 3B, 4B**). This difference was also evident in the AFR populations from the UK Biobank and AASK, where both variants are common and independent from G1/G2 (**Fig. 4B**; **Table 2**). Specifically, the V1 G allele, which is most common in AFR populations (**Table 1**), was associated with higher APOL1 levels as measured by Olink (posterior $\beta = 0.38$ in EUR and 0.18 in AFR; Methods) and lower levels as measured by SomaLogic ($\beta = -0.38$ in EUR and -0.33 in AFR), mirroring the pattern observed for the G1/G2 mutations, although with smaller effect sizes (**Fig. 4B**; **Table 2**). Furthermore, consistent with the G1/G2 results, the effect of V1 was replicated in the Health ABC data measured by Olink and was significantly weaker when measured by mass spectrometry ($p = 7.7\times10^{-3}$ for heterogeneity of effect size; **Fig. 4C**). In Health ABC, V1 genotypes were inferred from peptide sequence analysis (Norris Bradley *et al.*, 2020) and tested for association with APOL1 levels using a combined sample of EUR and AFR individuals to increase power (Methods).

Despite being located ~200 amino acids apart in the linear protein sequence, the V1 and G1/G2 mutations are in close spatial proximity within the predicted three-dimensional structure of APOL1, as determined by AlphaFold (Jumper *et al.*, 2021) (**Fig. 4D**). This structural proximity, coupled with similar platform-dependent pQTL effects, is consistent with the epitope effect hypothesis, whereby both G1/G2 and V1 induce conformational changes in APOL1 that differentially impact the binding of Olink and SomaLogic reagents. If so, V1 and G1/G2 may share a functional role, with V1 serving as a valuable proxy for investigating the biological effects of G1/G2 in European populations, where these mutations are virtually non-existent.

<u>HPR, an APOL1 binding partner, is a strong *trans*-pQTL for plasma APOL1 levels</u>

The APOL1 pQTL signal on chromosome 16 localized to the *HPR* gene region (**Fig. 3A**). *HPR* encodes haptoglobin-related protein, an essential component of the TLF particles that carry APOL1 in the bloodstream (Vanhollebeke and Pays, 2010). HPR mediates the entry of APOL1-TLF complexes into *Trypanosoma* by interacting with a parasite cell surface receptor (Vanhollebeke and Pays, 2010). Consistent with this close functional relationship, plasma levels of HPR and APOL1, as measured by mass spectrometry in the Health ABC dataset, were highly correlated ($R = 0.79$, $p = 1.2\times10^{-234}$; **Fig. S4A**). While statistical fine-mapping of the *HPR* region in the UK Biobank dataset identified three intronic variants with PIP > 0.9 (**Fig. S4B**), their functional interpretation remains challenging. However, the near-perfect correlation of variant effect sizes between the UK Biobank (Olink) and deCODE (SomaLogic) datasets ($R = 0.98$; **Fig. 3C**) strongly supports a consistent effect of HPR variants on plasma APOL1 levels, regardless of the proteomic platform. This platform-independent effect of HPR *trans*-pQTLs on APOL1 further supports the epitope effect hypothesis, providing additional evidence that only protein-altering APOL1 variants, but not regulatory effects, are differentially recognized by Olink and SomaLogic.

Interestingly, previous studies have identified copy number variation (CNV) in *HPR*, unique to AFR populations and associated with *Trypanosoma* resistance (Hardwick *et al.*, 2014). However, our fine-mapping analysis did not support an effect of *HPR* CNV on plasma $APOL1_{Olink}$ levels ($p = 1.3\times10^{-3}$, PIP = 0; **Fig. S4C–D**).

<u>Olink-specific *trans*-pQTLs identify the kallikrein-kinin pathway as a novel regulator of plasma APOL1</u>

In contrast to the *HPR trans*-pQTL effects, those associated with the *KNG1*, *KLKB1* and *F12* loci (**Fig. 3A**) directly challenge the epitope effect hypothesis. These pQTLs, despite having no probable impact on APOL1 conformation, are differentially detected by Olink and SomaLogic platforms, indicating that factors other than structural changes may also contribute to the observed discrepancies.

KNG1, *KLKB1*, and *F12* encode components of the kallikrein-kinin system, which plays a crucial role in regulating multiple physiological processes, including blood coagulation, inflammation,

and blood pressure control (Wisniewski, Gangnus and Burckhardt, 2024) (**Fig. 5A**). *KLKB1* and *F12* encode the serine proteases pre-kallikrein and factor XII, respectively, which are synthesized primarily in the liver (Motta, Juliano and Chagas, 2023). Factor XII activates pre-kallikrein via proteolytic cleavage, converting it to the active enzyme kallikrein (Motta, Juliano and Chagas, 2023) (**Fig. 5A**). *KNG1* encodes high-molecular-weight kininogen (HMWK), which serves as a cofactor and substrate for kallikrein in several branches of the kallikrein-kinin cascade (Motta, Juliano and Chagas, 2023) (**Fig. 5A**). Intriguingly, although KLKB1 and other hemostatic factors have been previously implicated in kidney disease (Yu *et al.*, 1998, 2000; Tin *et al.*, 2015), the biological mechanisms underlying the connection remain unclear.

To understand how the kallikrein-kinin pathway might influence plasma $APOL1_{Olink}$ levels, we performed statistical fine-mapping of the *KLKB1*, *F12* and *KNG1* gene regions (**Fig. 5B–D**). The association signal at the *F12* locus was largely attributed to a single 5'-UTR variant (rs1801020, PIP = 0.76; **Table 1**; **Fig. 5B**). The A allele at this variant introduces an alternative translation starting site for *F12*, leading to decreased production of functional factor XII protein (**Fig. S5A** and (Calafell *et al.*, 2010)).

The *KLKB1* locus displayed a more complex association pattern with multiple variants in high linkage disequilibrium (**Fig. 5C**). While the top two associated variants are intronic (rs139276089 and rs12509935, both with PIP = 0.14), we hypothesize that the third-ranking variant (rs3733402, PIP = 0.10), which is a missense variant, may be the true causal driver. This hypothesis is supported by prior evidence that the G allele at rs3733402 impairs the ability of kallikrein to bind and cleave HMWK, effectively acting as a partial loss-of-function (LOF) mutation (**Fig. S5A** and (Katsuda *et al.*, 2007)). Consistent with their partial LOF effect, both the *F12* variant (rs1801020-A) and the *KLKB1* variant (rs3733402-G) have been linked to a significant decrease in the plasma concentration of active renin (Biswas *et al.*, 2016), a key component of the renin-angiotensin system, which is interconnected with the kallikrein-kinin pathway in the regulation of blood pressure (**Fig. 5A**).

The mechanistic interpretation of the *KNG1* association signal (**Fig. 5D**) is less straightforward, as the two putatively causal variants fall into an intronic region (rs5030062; PIP = 0.66; **Table 1**) and the 3'-UTR (rs76438938; PIP = 0.29; **Table 1**). However, the A allele at rs5030062, along with the non-functional alleles in *F12* and *KLKB1* described above, has been linked to lower plasma thrombin levels (Pietzner *et al.*, 2020). Thrombin (factor II) is a blood coagulation protease acting downstream of factor XII and kallikrein (Chee, 2014) (**Fig. 5A**). This observation further supports the hypothesis that all three variants (rs1801020-A in *F12*, rs3733402-G in *KLKB1* and rs5030062-A in *KNG1*) are associated with decreased activity of the kallikrein-kinin pathway.

Importantly, the effects of the *F12*, *KLKB1* and *KNG1* variants on biomarkers of the kallikrein-kinin system are directionally consistent with their effects on plasma $APOL1_{Olink}$ levels, independent of blood pressure-related factors. Specifically, alleles associated with decreased thrombin are also associated with decreased $APOL1_{Olink}$, and vice versa (**Fig. S5A–B**). This finding suggests that the kallikrein-kinin pathway plays an active role in regulating circulating

APOL1 levels as measured by Olink, potentially by influencing the TLF particles that mediate APOL1 transport.

Genetic interaction analysis strengthens the association between the kallikrein-kinin pathway and plasma $APOL1_{Olink}$

While traditional pQTL analysis has proven effective in identifying individual variants associated with protein levels, it offers limited insight into the complex interplay of these variants and their combined influence on protein regulation. To overcome this limitation and move beyond the single-variant perspective, genetic interaction analysis can be employed to investigate the synergistic or antagonistic effects of variant combinations and to infer functional relationships between them. Despite its widespread use in dissecting molecular pathways in model organisms (Baryshnikova *et al.*, 2013), the application of this approach to quantitative traits in human populations remains largely unexplored.

We used genetic interaction analysis to investigate whether pQTLs associated with plasma $APOL1_{Olink}$ levels and, by extension, their cognate genes (*APOL1* itself, *HPR*, *F12*, *KLKB1* and *KNG1;* **Fig. 3A**) act independently or in concert. Our analysis revealed strong evidence for genetic interactions between the kallikrein-kinin pathway genes *KLKB1*, *F12* and *KNG1*, and, to a lesser extent, between *HPR* and *APOL1* (**Fig. 6A; Fig. S6**). Specifically, the interaction between *KLKB1* and *F12* variants was such that the association of each variant with plasma $APOL1_{Olink}$ levels was nearly abolished when the other gene carried a non-functional allele (**Fig. 6B**). This epistatic effect indicates that both *F12* and *KLKB1* must be functional for proper regulation of $APOL1_{Olink}$, mirroring their interdependent roles in the kallikrein-kinin system (**Fig. 5A**). Similarly, non-functional alleles in *F12* and *KLKB1* nearly eliminated the association of *KNG1* with plasma $APOL1_{Olink}$ levels; however, the reciprocal effect, consistent with the intronic nature of *KNG1* alleles, was much weaker (**Fig. S6C–D**). In contrast, variants in *KLKB1* and *HPR* combined additively to influence $APOL1_{Olink}$ levels (**Fig. 6C**), indicating that they act independently.

The genetic interactions between *KLKB1*, *F12* and *KNG1* variants, complementing their individual effects on plasma $APOL1_{Olink}$, strengthen the evidence for the role of the kallikrein-kinin pathway in APOL1 regulation and further solidify this newly discovered functional relationship.

A new model for plasma APOL1 regulation reconciles the platform specificity of both *cis*- and *trans*-pQTL effects

As we proposed earlier, platform-dependent *cis*-pQTL effects of *APOL1* coding variants (G1, G2, and V1) could be potentially explained by differential affinity of Olink's and SomaLogic's reagents for the mutated protein (**Fig. 2C**). However, such an epitope effect does not account for the observed platform-dependent *trans*-pQTL effects of the kallikrein-kinin system. A more comprehensive hypothesis that can explain both phenomena is that the proteomic platforms

differentially recognize two distinct forms of APOL1, whose relative abundance in plasma is influenced by both *cis* and *trans*-pQTLs.

APOL1 circulates in the blood as part of two similar but distinct HDL particles, termed trypanolytic factors (TLFs) because of their essential role in delivering APOL1 into *Trypanosoma* and ensuring its destruction (**Fig. 7**) (Rifkin, 1978; Hajduk *et al.*, 1989). Both TLF types include APOL1, APOA1, and HPR proteins, but only one type also includes IgM molecules (Raper *et al.*, 1999; Weckerle *et al.*, 2016). The IgM-free TLF (also known as TLF-1) is generally more abundant than its IgM-bound counterpart (TLF-2) (Verdi *et al.*, 2020). However, the relative proportions of TLF-1 and TLF-2 particles appear to be dynamically regulated by factors such as infections (Verdi *et al.*, 2020).

We hypothesize that SomaLogic and Olink technologies preferentially recognize TLF-1 and TLF-2 particles, respectively, leading to inversely correlated measurements of APOL1 when the relative abundance of these particles shifts (**Fig. 7**). Factors influencing the shift may include G1, G2, and V1 mutations in APOL1, as well as changes in IgM abundance driven by infection and, notably, activation of the kallikrein-kinin system (**Fig. 7**). Natural IgMs are known to inhibit the factor XII-initiated coagulation cascade (Obermayer *et al.*, 2021) and may therefore respond to changes in pathway activity resulting from non-functional alleles in *F12*, *KLKB1* and *KNG1*.

Importantly, our TLF shift hypothesis predicts that mass spectrometry, which likely measures total APOL1, would be insensitive to TLF-1/TLF-2 fluctuations. This prediction aligns with our findings that G1, G2, and V1 are not pQTLs for mass spectrometry-measured plasma APOL1 levels (**Fig. 2B, 4C**). The hypothesis is further supported by the observations that the APOL1 5'-UTR variant (V2) and the HPR variants (**Table 1**) show positively correlated effects across platforms (**Fig. 3B–C**), as they are more likely to affect total APOL1 than specific TLF forms.

Although our findings suggest that disease-causing APOL1 variants alter the TLF-1/TLF-2 ratio but not total plasma APOL1 levels, other factors may exert a stronger influence, affecting total APOL1 and the TLF-1/TLF-2 ratio, as well as disease risk. Interestingly, we observed that higher plasma APOL1 levels, as measured by both mass spectrometry (Health ABC) and Olink (UK Biobank), are associated with a reduced risk of all-cause mortality, independent of APOL1 genotype, age and genetic ancestry (**Fig. S7**). Furthermore, in the UK Biobank, higher plasma APOL1 levels correlated with better kidney function and a lower risk of CKD (**Fig. S7**). Though seemingly counterintuitive, these observations may indicate that higher plasma APOL1 levels reflect increased liver secretion or decreased cellular uptake from circulation (both potentially indicative of overall health), provided that the TLF-1/TLF-2 ratio is maintained.

The proposed model motivates us to consider the pathophysiology of APOL1-mediated kidney disease in the context of IgM levels. One possibility is that an excess of TLF-2 increases the flow of IgM into the kidneys (**Fig. 7**). While TLF uptake from the bloodstream into the kidneys has not yet been observed, the role of IgMs in nephropathy, particularly FSGS, is well-established (Gephardt *et al.*, 1986; Strassheim *et al.*, 2013; Panzer *et al.*, 2015).

Discussion

Previous research has primarily focused on the function of APOL1 within kidney cells, leaving the role of circulating APOL1 in kidney pathogenesis largely underexplored. In this study, we provide evidence that circulating APOL1, resistance to *Trypanosoma*, and APOL1-mediated CKD risk share a common genetic basis. Specifically, we show that the G1 and G2 variants, which confer resistance to *T. b. rhodesiense* and increase CKD risk (Friedman and Pollak, 2020), function as causal pQTLs for plasma APOL1, as measured by Olink and SomaLogic (**Fig. 1, 2A**). Furthermore, the effect of G2 on plasma APOL1 exceeds the effect of G1 (**Fig. 1B–C**) and is modulated by the p.N264K (M1) variant (**Fig. S2F**), mirroring their effects on CKD risk (Gupta *et al.*, 2023; Gbadegesin *et al.*, 2024; Sula Karreci *et al.*, 2024) and susceptibility to *Trypanosoma* infections (Cuypers *et al.*, 2016; Cooper *et al.*, 2017). These findings are consistent with the role of circulating APOL1 in trypanolysis, but also raise intriguing questions about its relationship to CKD. Does plasma APOL1 and CKD share a common genetic cause but remain functionally independent? Or does a causal relationship exist, where circulating APOL1 contributes to CKD development or, alternatively, acts as a downstream biomarker of kidney damage? Given that the majority of circulating APOL1 is secreted by the liver rather than the kidney (Shukha *et al.*, 2017), this latter biomarker hypothesis appears relatively less likely.

The predominant focus on the G1 and G2 APOL1 risk variants, found exclusively in individuals of African ancestry, has inadvertently limited our understanding of APOL1 function in other populations. Our findings demonstrate that the V1 mutation, common in both African and European ancestries, mirrors the effects of G1 and G2 on plasma $APOL1_{Olink}$ and $APOL1_{SomaLogic}$ levels, albeit with a reduced magnitude. This weaker pQTL effect of V1, coupled with challenges in detecting its association with CKD (**Fig. S8**), may stem from V1's subtler impact on APOL1 structure and function. Another contributing factor could be haplotype context. Recent evidence suggests that the V1 G allele (K150E) exerts a more pronounced cytotoxic effect *in vitro* when expressed within the EIK haplotype compared to the EMR haplotype, which are defined by the rs136175 (I228M) and rs136176 (K255R) variants (Gupta *et al.*, 2024). As our current analysis cannot differentiate between EIK and EMR, the observed V1 effect represents an aggregate of both haplotypes, potentially underestimating its true impact on plasma APOL1 and CKD risk. Although the clinical significance of V1 requires further investigation, its prevalence and pQTL effect in European populations suggest that APOL1's influence on kidney health may extend beyond populations of African ancestry and impact a wider range of individuals.

To account for the observed platform-specific effects of both *cis*- and *trans*-pQTLs for plasma APOL1, we propose a model where *APOL1* coding mutations (G1, G2 and V1) and the kallikrein-kinin pathway shift the balance of circulating TLF particles towards a higher TLF-2 to TLF-1 ratio (**Fig. 7**). This shift would be perceived differently by the Olink and SomaLogic platforms if they had differential affinity for APOL1 bound to TLF-1 versus TLF-2, analogous to the differential affinity of SomaLogic aptamers for free versus TCC-bound complement C7 (Cervia-Hasler *et al.*, 2024). We hypothesize that a higher proportion of circulating TLF-2 may promote increased uptake of these particles into the kidneys. As TLF-2 particles carry IgM molecules, enhanced uptake could lead to an accumulation of both APOL1 and IgM within the renal tissue, contributing to ionic imbalance, inflammation and injury. This proposed

TLF-mediated mechanism complements the current understanding of APOL1-mediated kidney disease, which primarily focuses on podocyte-resident APOL1 and its function as an ion channel in plasma membranes (Vasquez-Rios, De Cos and Campbell, 2023). Our model suggests that circulating APOL1, delivered to the kidneys via TLF particles, may contribute to CKD pathophysiology in addition to endogenously expressed APOL1. Furthermore, this model aligns with observations that infections associated with elevated IgM levels, including HIV and COVID-19, act as important secondary triggers for CKD development in individuals carrying APOL1 G1/G2 risk variants (Freedman and Skorecki, 2014).

Beyond the TLF-mediated hypothesis, alternative mechanisms may also explain the observed platform-specific pQTL effects. Our findings, along with prior observations (Weckerle *et al.*, 2016), provide evidence that APOL1 coding mutations induce significant conformational changes in the protein, impacting the binding of both Olink antibodies and SomaLogic aptamers. This observation may offer valuable insights into APOL1's structure-function relationship and its role in kidney disease. Additionally, the kallikrein-kinin pathway may regulate plasma APOL1 more directly. Given the proteolytic activity of KLKB1 and F12, and their association with plasma levels of multiple protease substrates and inhibitors in the UK Biobank (**Fig. S9A–B**), it is possible that APOL1 itself is a direct substrate for cleavage by these enzymes, resulting in protein isoforms that are differentially recognized by different proteomic platforms.

In conclusion, our analysis of the genetic architecture of plasma APOL1 underscores the power of a comprehensive approach that integrates multiple proteomic technologies (Olink, SomaLogic, and mass spectrometry), diverse genetic ancestries (European and African) and an understanding of endophenotypes (plasma APOL1 levels) in the context of disease outcomes (CKD). Plasma proteomics provided enhanced phenotypic resolution and enabled the discovery of subtle genetic associations that elude traditional disease-focused analyses. By integrating multiple proteomic technologies, each capable of recognizing distinct proteoforms, we achieved a more nuanced understanding of genetic variant effects and illustrated an approach that can be generalized and applied in a wider range of biological contexts. The adoption of this approach will help illuminating the complex interplay between genetic variation and disease processes, ultimately paving the way for a deeper understanding of human physiology and the development of more effective therapies.

Methods

Study populations

The UK Biobank (UKB) is a prospective cohort of over 500,000 participants aged 40–69 at recruitment with extensive genotype and phenotype data, as detailed at https://biobank.ndph.ox.ac.uk/showcase/. In the Pharma Proteomics Project (PPP), plasma samples from ~50,000 participants were analyzed using the Olink Explore 3072 platform to measure relative abundance of ~3,000 circulating proteins (Sun *et al.*, 2023). The UKB-PPP

cohort was used for pQTL association tests, whereas the entire UKB cohort was used for disease association tests. All participants provided informed consent.

We used the Pan-UKB consortium dataset (Karczewski *et al.*, 2024) to assign genetic ancestry to individuals in the UKB-PPP cohort, focusing on those of European and African descent. To minimize relatedness within each sample set, we randomly excluded one individual from each pair with a kinship coefficient (Pi_hat) greater than 0.125. This relatedness filter, while less stringent than standard approaches, maximized sample size, yielding to 47,178 and 948 individuals of European and African ancestry, respectively. Nearly 100% of these individuals (47,090 and 947) had both genotype data and APOL1 measurements, whereas 45,444 (96%) and 893 (94%) had both exome sequences and APOL1 measurements.

The Health, Aging, and Body Composition (Health ABC) study is an observational cohort study investigating the functional decline of healthy aging individuals (Newman *et al.*, 2023). The cohort includes 3,075 participants aged 70–79, with 33% men and 45% women self-reported as African-American. Mass spectrometry and the Olink assay were performed on a random subset of 1,094 and 560 Health ABC participants, respectively **(Table S1**).

<u>UKB: Genotype data processing and quality control</u>

Genotyping, imputation and quality control of UKB samples were performed in the consortium pipeline (Bycroft *et al.*, 2018). We used version 3 of the imputed data, which was generated using the Haplotype Reference Consortium (HRC) and the UK10K haplotype resource as the imputation panel (described in the UK Biobank Category 100319). Variants with an INFO score below 0.6 were filtered out. Exome data were also processed through the UKB consortium pipeline and accessed via the UK Biobank Research Analysis Platform (UKB-RAP). Details of exome sequencing procedures are available on the UKB website (https://dnanexus.gitbook.io/uk-biobank-rap/science-corner/whole-exome-sequencing-oqfe-protocol) as well as related publications (Backman *et al.*, 2021; Okada and Wang, 2021). Within each sample set (defined above), variants with Minor Allele Count (MAC) below 10 were filtered out.

Genotypes from both imputed and genotyped data (hg19) were lifted over to hg38 using the Genome Analysis Toolkit (GATK) LiftoverVcf tool (https://gatk.broadinstitute.org/hc/en-us). Variants without unique mapping were filtered out. Exome data were already provided in the hg38 form.

Association tests were performed using primarily imputed genotype data, with the exception of the APOL1 G1 variant in the African population. Comparison of exome and imputed genotype data revealed discrepancies in G1M and G1G variant calls. Specifically, over 20% (73 of 351) of heterozygous G1M carriers and 5% (17 of 349) of heterozygous G1G carriers identified by exome sequencing were misidentified as homozygous in the imputed genotype data (**Fig. S1A-B**). Further supporting imputation inaccuracies, we observed a low correlation between G1G and G1M genotypes (60 samples showed discordant genotypes for G1G and G1M; **Fig. S1C**), excess linkage disequilibrium with G2 (56 samples appeared to be homozygous for G1M and heterozygous for G2, an unlikely genotype pattern given that G1 and G2 never occur on the

same haplotype; **Fig. S1D**) and deviation from Hardy-Weinberg equilibrium in the imputed data ($p < 10^{-14}$) compared to the exome data ($p > 0.05$; **Fig. S1E**). As a result of these discrepancies, we prioritized exome data when available. We confirmed that the association between G1/G2 and plasma APOL1 levels was similarly likely to be causal in the exome data alone, imputed genotype data alone, or a combination of both (**Fig. S1F-G; Fig. 1A**).

For visualization of individual variant effects (e.g., in violin plots or kernel density estimation plots), exome data were prioritized, unless the variant was absent (e.g., the non-coding HPR lead cis-variant).

UKB-PPP: pQTL analyses

In the UKB-PPP dataset, plasma protein levels are reported as Olink's Normalized Protein eXpression (NPX) values, which reflect relative change in protein abundance on a $\log_2$ scale following plate and batch normalization (Sun *et al.*, 2023).

Association tests between genotypes and protein levels were performed using tensorQTL v1.0.9 (https://github.com/broadinstitute/tensorqtl). Prior to analysis, protein NPX matrices were inverse-normal transformed within each study population. Age, sex, five genotype PCs and ten protein expression PCs were included as covariates. Genotype PCs were provided by the Pan-UKB consortium. Protein expression PCs were calculated using the PCA function of scikit-learn applied to the inverse-normal transformed protein NPX matrix.

Genetic effects were estimated using an additive model, where each copy of the effect allele contributes equally to the change in protein levels. To facilitate interpretation, variant effect sizes (β) are reported as per-allele changes in raw (rather than inverse-normal transformed) protein NPX values, unless explicitly stated. For variants in linkage disequilibrium with the APOL1 G2 variant, which exhibits a large effect on plasma APOL1 levels in the Olink-based measurements, we report posterior β estimates from the FINEMAP algorithm (Benner *et al.*, 2016). This approach accounts for the strong effect of G2 and provides a more accurate estimate of the independent effects of other variants in the region (**Fig. S2A–E**). Marginal and posterior β are explicitly distinguished throughout the manuscript. Effect sizes from the replication cohorts (AASK and deCODE; SomaLogic-based measurements) are reported as marginal β, since posterior β estimates are not available. Given that the effect of G2 in these studies is much smaller than in UKB, we do not anticipate marginal β to introduce substantial bias.

UKB: fine-mapping

Fine-mapping of each locus associated with protein expression was performed using FINEMAP v.1.3.1 (Benner *et al.*, 2016) and susieR v.0.11.43 (Wang *et al.*, 2020) with default parameter settings. For genes on the plus strand, the *cis*-locus was defined as the region spanning 0.2 Mb upstream of the transcription start site (TSS) to 0.2 Mb downstream of the transcription end site (TES). For genes on the minus strand, the *cis*-locus was defined as the region spanning 0.2 Mb upstream of the TES to 0.2 Mb downstream of the TSS. Input data included summary statistics from tensorQTL and the in-sample LD matrix adjusted for the same covariates used in the association tests. A maximum of ten causal variants was allowed.

UKB: *HPR* copy number analysis

HPR copy number (CN) data was obtained from the UKB-RAP, which provides CN estimations derived from whole-exome sequencing data analyzed by the Illumina DRAGEN platform (https://developer.illumina.com/dragen). For association tests and fine-mapping of the *HPR* region, HPR CN values were included as an additional variable, mean-centered and scaled like other genetic variants.

UKB: testing for genetic interactions

To identify potential genetic interactions between pairs of variants, we employed a linear regression model using the statsmodel package in Python. For each pair of variants (v_1 and v_2), we fit the following model: $y \sim v_1 + v_2 + v_1 \times v_2 + covariates$, where y represents the phenotype, and $covariates$ represents the same set of variables used in the association tests (see above). The significance of the genetic interaction was assessed using the p-value of the interaction term.

To assess the difference in the effect of variant v_1 conditioned on the presence of another variant (v_2), we employed a stratified analysis using separate linear regression models. Two models (M1 and M2) were fit: $y \sim v_1 + covariates$ when v_2 is absent (M1), and $y \sim v_1 + covariates$ when v_2 is present (M2). Let β_1 and β_2 denote the estimated effects of v_1 in models M1 and M2, respectively, with corresponding standard errors σ_1 and σ_2. We calculated the t-statistic for the difference in effects as $(\beta_1 - \beta_2)/\sqrt{\sigma_1^2 + \sigma_2^2}$. Under the null hypothesis of no difference in effects, this statistic is assumed to follow a t-distribution with $N_1 + N_2 - 4$ degrees of freedom, where N_1 and N_2 are the sample sizes in the respective strata.

When conditioning on the p.N264K variant, both the heterozygous (REF/ALT) and the homozygous alternative (ALT/ALT) genotypes for v_2 were included in model M2. When conditioning on the G1 or G2 variants, only the REF/ALT genotype was included in M2. This decision was based on the observation that G1 and G2 are never found on the same haplotype, such that, if one variant in ALT/ALT, the other is always REF/REF. As a result, including the ALT/ALT genotype in M2 would only reduce power.

Health ABC: proteomic analyses

Mass spectrometry analysis was performed on a random set of 1,111 Health ABC participants using samples collected at baseline and second year follow up visits. Data from the two visits were averaged to produce a single value per individual. Samples that failed during processing or data quality control were excluded from further examination, resulting in 1,094 participants available for downstream analysis.

Materials

HEPES and EPPS buffers were obtained from Boston BioProducts (Milford, MA, USA), and cOmplete ULTRA and PhosStop inhibitor tablets were obtained from Roche (Mannheim, Germany). Trypsin/Lys-C Mix was obtained from Promega Corporation (Madison, WI, USA). Tandem mass tag (TMT) 16-plex reagent sets were obtained from Thermo Fisher Scientific (Rockford, IL). Unless otherwise stated, all other chemicals were purchased from Sigma-Aldrich (St. Louis, MO, USA).

Sample preparation for liquid chromatography-mass spectrometry (LC-MS)

The samples were prepped using our in-house automated multiplexed proteome profiling platform (AutoMP3) protocol (Gaun *et al.*, 2021). Neat plasma (2 µL) was diluted with 38 µL of lysis buffer (75 mM NaCl, 50 mM HEPES pH 8.5, 3% sodium dodecyl sulfate (SDS) with cOmplete ULTRA and PhosSTOP tablets). The samples were reduced with 5 mM final concentration dithiothreitol (DTT) for 30 minutes at 25ºC, then alkylated with 15 mM final concentration iodoacetamide in the dark for 30 minutes at 25ºC. Excess iodoacetamide was quenched with 5 mM final concentration DTT and incubated in the dark for 15 minutes at 25ºC. Samples were cleaned up using a mixture of Sera-Mag beads (Cytiva Life Sciences, 44152105050350, 24152105050350; Marlborough, MA, USA). Beads were combined 1:1 (E3:E7) then washed 3 times with water. Samples were added directly to the beads, then acetonitrile was added to a final concentration of 75% and the mixture was incubated for 20 minutes at room temperature for protein binding. Beads were immobilized on a magnet then washed two times with 70% ethanol (200 µL) followed by two washes with 100% acetonitrile (ACN, 200 µL). The beads were resuspended in 40 µL of digestion buffer (50 mM EPPS pH 8.5, 10 mM CaCl2) then 10 µL of Trypsin/Lys-C (0.33 mg/mL) was added. The proteins were digested on-bead for 1 hour at 37ºC, 1000 rpm. Digested samples were immobilized on a magnet and the peptide supernatant was transferred to a new plate. To enable analysis across multiple plexes, a bridge sample was created by pooling a small amount of each peptide sample. Tandem mass tag (TMT) labeling was performed using 30 µg of peptides and adding 0.5 mg of TMT reagent. The TMT reaction was incubated for 1 hour at 25ºC then excessed TMT was quenched with 0.3% final concentration hydroxylamine. For each plex, samples were combined, mixed, then split into several wells (≤100 µL per well) for peptide cleanup. For peptide binding, acetonitrile was added to a final concentration of 95%. Beads were immobilized on a magnet then washed one time with 100% ACN (1 mL). Peptides were eluted using 3 rounds of 5% ACN (100 µL), then plexes were re-combined and dried down in a SpeedVac overnight at 25ºC. The combined samples were resuspended in 5% ACN / 5% formic acid (FA) then analyzed on a mass spectrometer.

Liquid chromatography-mass spectrometry

Peptides were analyzed on an Orbitrap Eclipse mass spectrometer coupled to an Ultimate 3000 (Thermo Fisher Scientific). Peptides were separated on an IonOpticks Aurora microcapillary column (75 µm inner diameter, 25 cm long, C18 resin, 1.6 µm, 120 Å, Gen2). The total LC-MS run length for each sample was 185 min including a 165 min gradient from 6 to 35% ACN in 0.1% formic acid. The flow rate was 300 nL / min, and the column was heated at 60° C. A Field Asymmetric Ion Mobility Spectroscopy device (FAIMS Pro) was connected to the Orbitrap

Eclipse mass spectrometer and a method with four experiments, where each experiment utilized a different FAIMS Pro compensation voltage: -40, -50, -60, and -70 volts was used. Each of the four experiments had a 1.25 second cycle time. A high resolution MS1 scan in the Orbitrap (m/z range 400-1,600, 120k resolution, standard Automatic Gain Control (AGC) target settings, "Auto" max injection time, ion funnel RF of 30%) was collected from which precursors were selected for MS2 followed by SPS MS3 analysis. For MS2 spectra, ions were isolated with the quadrupole mass filter using a 0.7 m/z isolation window. The MS2 product ion population was analyzed in the quadrupole ion trap (CID, "standard" AGC target, normalized collision energy 35%, "Custom" max injection time set to 35 ms) and the MS3 scan was analyzed in the Orbitrap (HCD, 50k resolution, normalized collision energy 45%, "custom" AGC target with normalized AGC target set to 200% and max injection time 200 ms). Up to ten fragment ions from each MS2 spectrum were selected for MS3 analysis using SPS. A human database was used for Real Time Search (RTS) (Schweppe *et al.*, 2020) and a maximum of two missed cleavages and one variable modification was allowed and FDR filtering enabled. The maximum search time was set to 35 ms and a Xcorr of 1, dCN 0.1 and precursor of 10 ppm for charge state 2, 3 and 4 was used.

Peptide Identification and Quantification

Mass spectrometry data were processed using a previously described software pipeline (Huttlin *et al.*, 2010). Raw files were converted to mzXML files and searched against a human database (downloaded on October 2nd, 2020), in forward and reverse orientations using the Comet algorithm (Huttlin *et al.*, 2010; Eng, Jahan and Hoopmann, 2013). Database searching matched MS/MS spectra with fully tryptic peptides from this composite dataset with a precursor ion tolerance of 20 ppm, fragment_bin_tol set to 1.0005 and fragment_bin_offset = 0.4 Da. TMTpro tags on peptide N-termini and lysines (+304.20 Da) were set as static modifications. Oxidation of methionine (+15.99 Da) was set as a variable modification. Linear discriminant analysis was used to filter peptide spectral matches to a 1% FDR (false discovery rate) as described previously (Elias and Gygi, 2007). Non-unique peptides that matched to multiple proteins were assigned to proteins that contained the largest number of matched redundant peptide sequences using the principle of Occam's razor. Quantification of TMT reporter ion intensities was performed by extracting the most intense ion within a 0.003 m/z window at the predicted m/z value for each reporter ion. Relative abundances of peptides were calculated as log2 ratio of sample to bridge, where bridge sample was a common control shared across all TMT plexes.

Olink Platform

Olink analysis was performed on 658 samples selected at random from those chosen for mass spectrometry. Samples were shipped to Olink Proteomics (Waltham, MA, USA) and processed by the manufacturer. Normalized protein expression (NPX) values were generated after quality control, normalization, and batch effect correction. Samples that failed during processing or data quality control were excluded from further examination, resulting in 560 samples available for downstream analysis.

<u>Health ABC: inference of G1, G2 and V1 genotypes from peptide sequence</u>

To infer APOL1 variant status, we analyzed peptide-level mass spectrometry data (**Fig. S3**).

For rs2239785 (V1), we utilized two peptides, SELEDNIRR and LEDNIRR, which overlap the V1 variant position and correspond to the REF (G) and ALT (A) alleles, respectively. A scatter plot of the two peptide measurements was used to infer genotypes (**Fig. S3B**). Samples with two copies of the reference allele (G/G) were expected to fall along the y=0 line, those with two copies of the alternative allele (A/A) along the x=0 line, and heterozygotes (G/A) along the y=x line. We classified samples as G/G, A/A or G/A, depending on whether they fall above, below or within a defined range ($x-0.5 < y < x+1$; **Fig. S3B**). Samples with a peptide measurement of exactly 0 were classified as homozygous for the corresponding allele, regardless of whether they fall within the defined range (**Fig. S3B**).

For G1/G2, we utilized the peptide LNILNNNYK, which overlaps the G1M/G2 variants and is theoretically only observed in individuals with a REF/REF or a REF/ALT genotype for both. This peptide's abundance is expected to be comparable to other APOL1 peptides in REF/REF individuals, but substantially lower in REF/ALT and ALT/ALT individuals. We trained a lasso model in European samples (virtually all REF/REF) to predict this peptide's abundance using other APOL1 peptides as features. The model was then applied to African samples to compare predicted and observed peptide measurements. We classified African samples as REF/ALT or ALT/ALT, depending on whether their observed values fell below $y < 5/3*x$ or $y < 5/3*x-1$ thresholds, respectively (where x is the predicted value; **Fig. S3A**). Samples with a peptide measurement of exactly 0 were additionally classified as ALT/ALT (**Fig. S3A**).

Comparison to AASK and deCODE data

Plasma pQTL summary statistics for the deCODE study were obtained from https://www.decode.com/summarydata/. Plasma pQTL summary statistics for the AASK study were obtained from (Surapaneni *et al.*, 2022). Throughout the manuscript, we report the effect sizes of the alternative alleles as defined in hg38, regardless of the allele frequencies, unless otherwise noted. For example, in the case of V1, we report the effect size of the reference allele (G) as it is most common in populations of African ancestry.

UKB: disease associations

The UKB field ID 42026, which reports the date of end-stage renal disease (ESRD) diagnosis, was used as evidence of disease. The field is based on a combination of multiple factors including health records, self-reports, and death certificates. We binarized the field, assigning a value of 1 to individuals with an ESRD diagnosis and 0 to those without a diagnosis (or with missing data). A relatedness filter, as defined by the Pan-UKB consortium, was applied. We then performed logistic regression using disease diagnosis as an outcome and including age, sex and 10 genotype PCs as covariates.

UKB: adjusting for blood pressure-related factors

To assess the impact of *KLKB1*, *F12* and *KNG1* variants on APOL1 levels while controlling for blood pressure-related factors, we performed linear regressions which included age, sex, five genotype PCs and the following covariates: diastolic blood pressure (UKB field 4079),

calculated as the mean of two repeated measurements; systolic blood pressure (UKB field 4080), calculated as the mean of two repeated measurements; history of hypertension diagnosis (UKB field 2966), coded as a binary variable (1 = diagnosis at any age, 0 = no diagnosis); history of blood pressure medication use (UKB fields 6153 and 6177), coded as a binary variable (1 = medication record present, 0 = no medication record). Missing values for any covariate were imputed using the sample mean. Linear regression models were fit without regularization.

Power calculation

To evaluate the statistical power to identify associations between risk alleles and ESRD phenotype in our UKB dataset, we conducted a power analysis using a two-proportion z-test (Sham and Purcell, 2014). The test compares the proportion of individuals with the high-risk genotype (under recessive model) in cases (individuals with ESRD) and controls (individuals without ESRD).

Given the number of cases and control (which we directly obtain in each population group as defined in the Pan-UKB), we calculated the probability of detecting a statistically significant difference in high-risk genotype frequencies between these groups with the type 1 error rate of 0.05. Within a population, the risk allele frequencies were calculated as (2 x number of individual with risk homozygous genotype + number of individuals with heterozygous genotype) / (2 x number of total individuals).

Statistical analysis

All the statistical tests were two-sided. Reported p-values are unadjusted unless otherwise noted. Error bars represent standard errors of the mean (SEM) unless otherwise noted.

Data and code availability

The UK Biobank (UKB) data were accessed following the processes described in https://www.ukbiobank.ac.uk/enable-your-research. The UKB Research Analysis Platform (RAP) was primarily used for UKB data analysis. The Health ABC data were accessed following the processes described in https://www.nia.nih.gov/healthabc-study. Plasma pQTL summary statistics for the deCODE study were obtained from https://www.decode.com/summarydata/. Plasma pQTL summary statistics for the AASK study were obtained from (Surapaneni *et al.*, 2022). Variant-level summary statistics generated in this manuscript are available upon request.

Software tools used for data analysis and visualization include: dx-toolkit (https://github.com/dnanexus/dx-toolkit); FINEMAP v.1.3.1 (http://www.christianbenner.com/); GATK v.4.1.9.0 LiftoverVcf (https://gatk.broadinstitute.org/); matplotlib v.3.3.4 (https://matplotlib.org); NumPy v.1.20.1 (https://numpy.org); pandas v.1.1.4 (https://pandas.pydata.org); scikit-learn v.0.24.1 (https://scikit-learn.github.io/stable); SciPy v.1.6.2 (https://scipy.org/); seaborn v.0.11.1 (https://seaborn.pydata.org); susieR v.0.11.43 (https://github.com/stephenslab/susieR); tensorQTL v.1.0.9 (https://github.com/broadinstitute/tensorqtl). Custom code is available upon request.

Acknowledgements

We would like to thank Katalin Susztak, Anil Raj, Martin Mullis, Qi Hao, Manoj Rathinaswamy, Bernd Wranik, Melissa Pilling, Robert Cohen and Nick van Bruggen for useful discussions and feedback. We also acknowledge Steven Cummings and the Health ABC study for enabling validation analyses. Research with the Health ABC data was supported by National Institute on Aging (NIA) Contracts N01-AG-6-2101; N01-AG-6-2103; N01-AG-6-2106; NIA grant R01-AG028050, and NINR grant R01-NR012459. This research was funded in part by the Intramural Research Program of the NIH, National Institute on Aging. Research with the UK Biobank Resource was conducted under application number 18448. This study was funded by Calico Life Sciences LLC.

Competing interests

QSW, JH, LC, NH, NO, FMA, DM, AB, EM and AB are employees of Calico Life Sciences LLC. AG was an employee of Calico Life Sciences LLC when this study was conducted.

Table 1. Key variants associated with plasma APOL1 levels.

RSID	Name	Variant ID (GRCh38)	Gene	Effect	Alternative allele frequency (gnomad)	
					EUR	AFR
rs73885319	G1G	22:36265860:A:G	*APOL1*	S342G	2.6×10^{-4}	0.22
rs60910145	G1M	22:36265988:T:G	*APOL1*	I384M	5.3×10^{-5}	0.23
rs1317778148 rs71785313 rs12106505	G2	22:36265995:AATAATT:A	*APOL1*	N388_Y389del	4.6×10^{-5}	0.14
rs73885316	M1	22:36265628:C:A	*APOL1*	N264K	1.4×10^{-4}	0.03
rs2239785	V1	22:36265284:G:A	*APOL1*	E150K	0.81	0.35
rs6000220	V2	22:36253920:C:T	*APOL1*	5'-UTR	0.11	0.16
rs148296684	–	22:36265308:C:T	*APOL1*	L158F	1.4×10^{-3}	1.3×10^{-4}
rs1801020	–	5:177409531:A:G	*F12*	5'-UTR	0.75	0.55
rs3733402	–	4:186236880:G:A	*KLKB1*	S143N	0.51	0.74

Table 2. Association of APOL1 coding variants with plasma APOL1 levels across multiple datasets and platforms. Upward and downward arrows indicate positive and negative associations of the variant most common in populations with African ancestry. The number of arrows represents the relative strength of the association (strongest to weakest) within each dataset. Dash lines ("–") indicate that data were not available; "n.s." indicates that the association was not significant.

				APOL1 coding variant		
Dataset	**Ancestral group**	**Platform**	**Sample size**	**G1**	**G2**	**V1**
UKB	AFR	Olink	828	↑↑	↑↑↑	↑
UKB	EUR	Olink	38,442	–	–	↑
AASK	AFR	SomaLogic	461	n.s.	↓↓↓	↓
deCODE	EUR	SomaLogic	35,559	–	–	↓
Health ABC	AFR	Olink	194	↑↑	↑↑↑	↑
Health ABC	EUR	Olink	366	–	–	↑
Health ABC	AFR	MS	390	n.s.	n.s.	n.s.
Health ABC	EUR	MS	704	–	–	n.s.

Figure 1

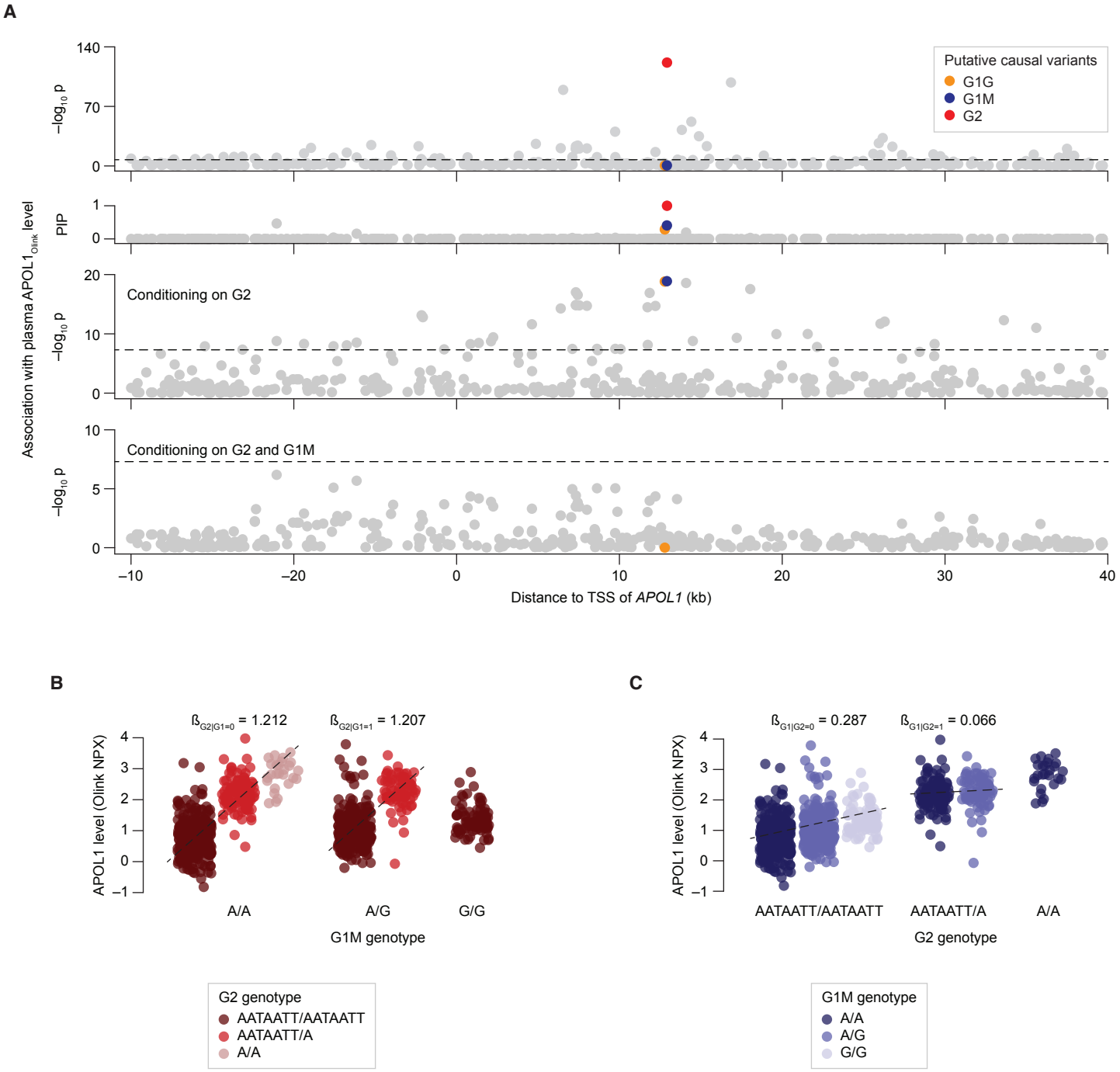


Figure 1. *APOL1* G1 and G2 variants are causal pQTLs for plasma APOL1$_{Olink}$ levels. (A) Manhattan plot showing the association of *APOL1 cis*-variants with plasma APOL1$_{Olink}$ levels in the UK Biobank population of African descent. Each dot represents a variant. The x-axis shows the variant's chromosomal position relative to the transcriptional start site (TSS) of *APOL1*. The y-axis shows the $-\log_{10}$ of the association p-value (row 1), the statistical fine-mapping posterior inclusion probability (PIP; row 2), the $-\log_{10}$ of the association p-value after adjusting for the effect of the G2 variant (row 3) and the $-\log_{10}$ of the association p-value after adjusting for the effects of both the G2 and G1M variants (row 4). Putative causal variants identified through fine-mapping are highlighted. Horizontal dashed lines represent the genome-wide significance threshold ($p = 5\times10^{-8}$). **(B)** Scatter plot showing the effect of the G1M variant on plasma APOL1$_{Olink}$ levels, stratified by G2 genotype. **(C)** Scatter plot showing the effect of the G2 variant on plasma APOL1$_{Olink}$ levels, stratified by the G1M genotype. **(B–C)** Each dot represents an individual. Within each genotype group, dots were jittered along the x-axis for visualization purposes.

Figure 2

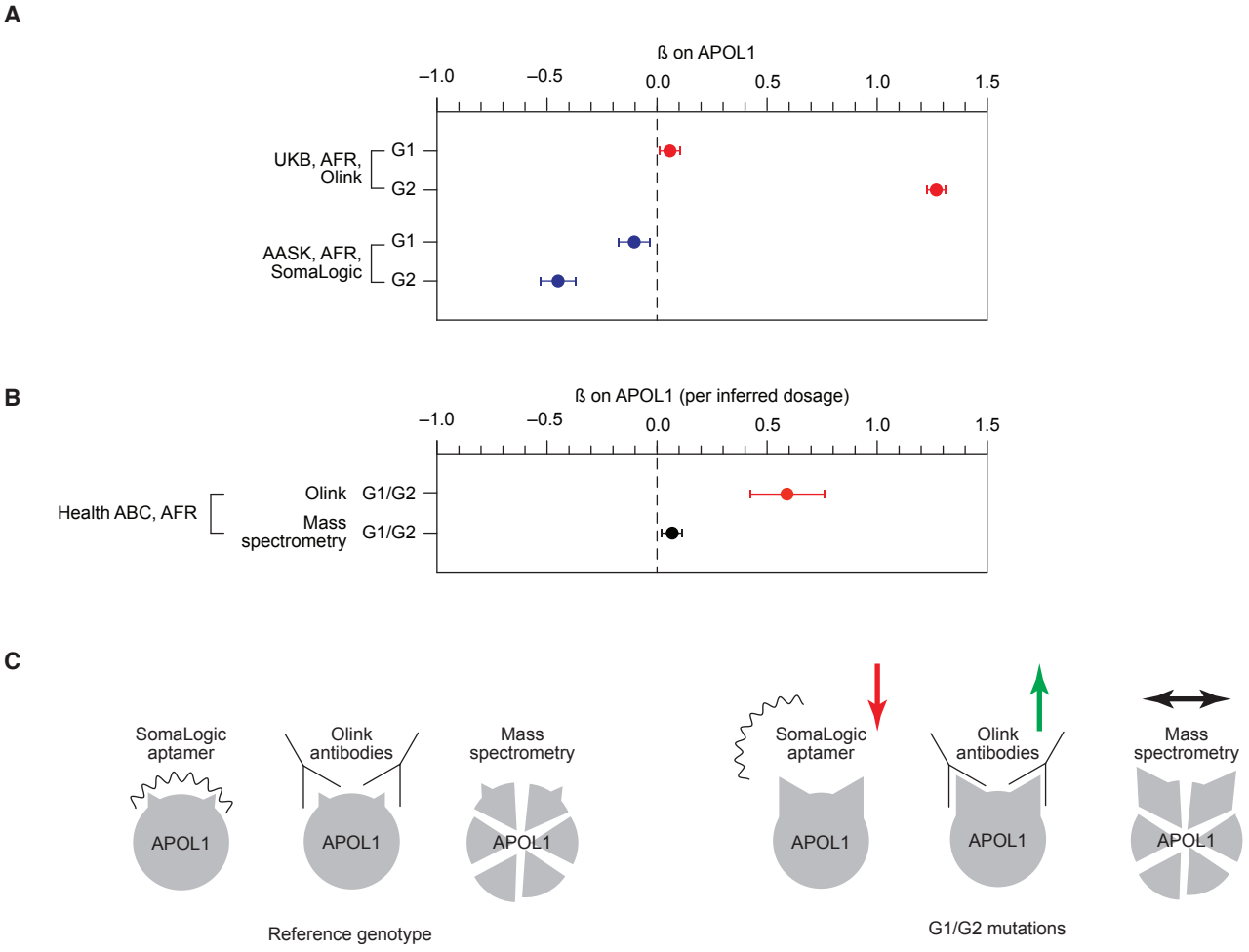


Figure 2. *APOL1* G1 and G2 variants show opposite effects on plasma APOL1 levels measured by Olink and SomaLogic platforms. (A) Association of *APOL1* G1 and G2 variants with plasma APOL1 levels in individuals of African ancestry from the UK Biobank (Olink platform) and AASK study (SomaLogic platform). Effect sizes are reported as posterior β (UK Biobank) and marginal β (AASK study). Error bars indicate standard errors of the mean. **(B)** Association of *APOL1* G1/G2 variants with plasma APOL1 levels measured by Olink's platform and mass spectrometry in individuals of African ancestry from Health ABC. Error bars indicate standard errors of the mean. **(C)** Schematic illustrating the epitope effect hypothesis that could potentially explain the observed differences in G1/G2 effects between platforms. Left: both SomaLogic aptamers and Olink antibodies bind to reference APOL1 protein. Right: G1/G2 mutations induce conformational changes that enhance the binding of Olink antibodies but hinder the binding of SomaLogic aptamers, resulting in an apparent increase and decrease in measured plasma APOL1 levels, respectively. Mass spectrometry-based measurements of APOL1 levels are largely unaffected by G1/G2 mutations as most peptides remain unchanged.

Figure 3

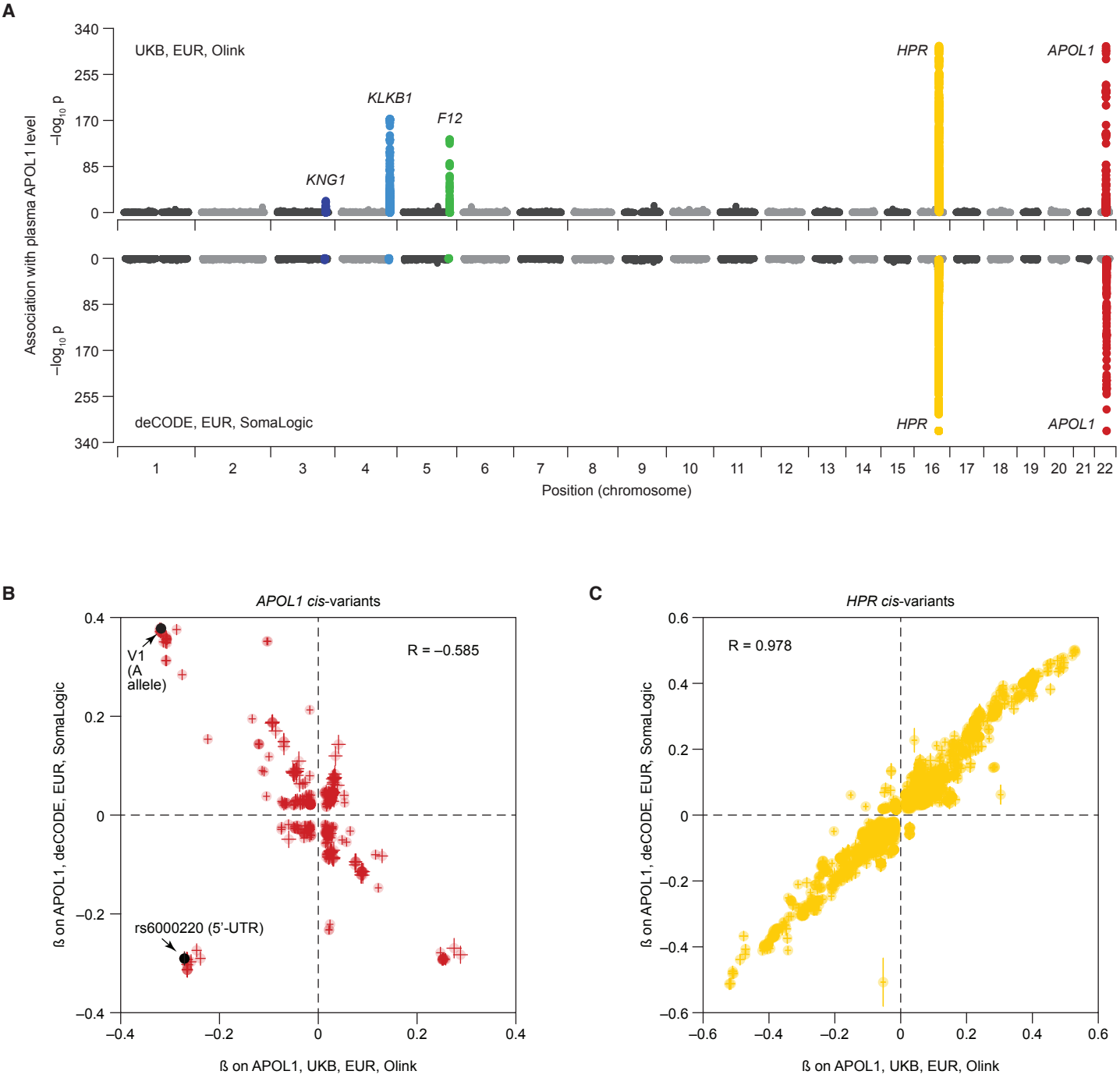


Figure 3. Platform-specific pQTLs influence plasma APOL1 levels through both *cis* and *trans* mechanisms. (A) Miami plot showing the genome-wide variant associations with plasma APOL1 levels in the UK Biobank (top, Olink) and deCODE (bottom, SomaLogic) populations of European ancestry (EUR). Each dot represents a variant. The x-axis shows chromosomal position. The y-axis shows the $-\log_{10}$ of the association p-value. The top five strongest associations are labeled with the names of the corresponding genes. **(B–C)** Scatter plots comparing the effect sizes (marginal β) of *cis*-APOL1 **(B)** and *cis*-HPR **(C)** variants on plasma APOL1 levels in the UK Biobank (x-axis) and deCODE (y-axis).

Figure 4

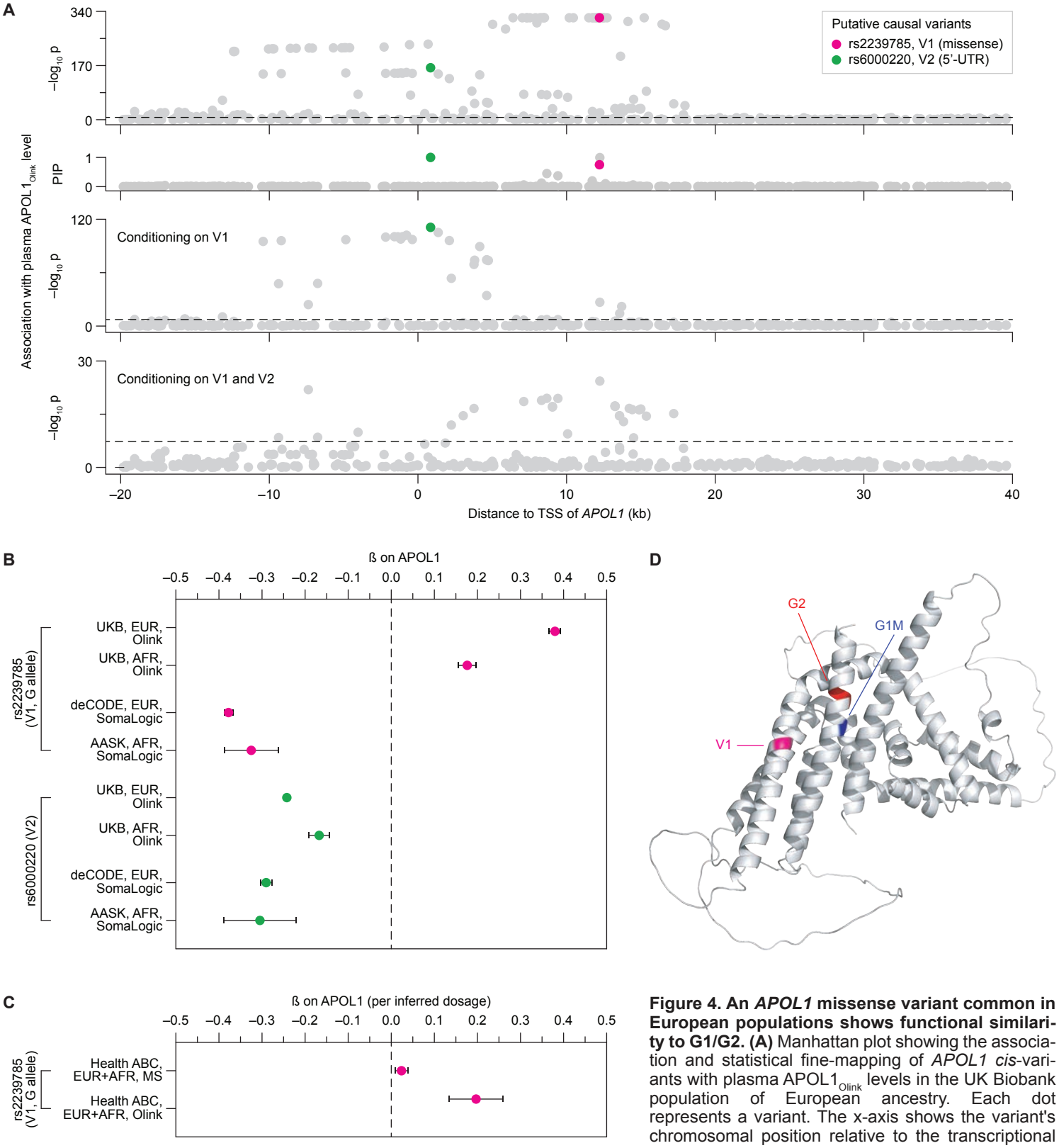


Figure 4. An *APOL1* missense variant common in European populations shows functional similarity to G1/G2. (A) Manhattan plot showing the association and statistical fine-mapping of *APOL1 cis*-variants with plasma $APOL1_{Olink}$ levels in the UK Biobank population of European ancestry. Each dot represents a variant. The x-axis shows the variant's chromosomal position relative to the transcriptional start site (TSS) of *APOL1*. The y-axis shows the $-\log_{10}$ of the association p-value (row 1), the fine-mapping posterior inclusion probability (PIP; row 2), the $-\log_{10}$ of the association p-value after adjusting for the effect of rs2239875 (V1; row 3) and the $-\log_{10}$ of the association p-value after adjusting for the effects for both V1 and rs6000220 (V2; row 4). Putative causal variants identified through fine-mapping are highlighted. Horizontal dashed lines represent the genome-wide significance threshold ($p = 5\times10^{-8}$). **(B)** Effects of the *APOL1* rs2239875 (V1) and rs6000220 variants on plasma $APOL1_{Olink}$ and $APOL1_{SomaLogic}$ levels in individuals of European ancestry from the UK Biobank and deCODE, respectively. Effect sizes are reported as posterior beta (UK Biobank, Olink) and marginal beta (deCODE, SomaLogic). Error bars indicate the standard error of the mean. **(C)** Association of the *APOL1* V1 variant with plasma APOL1 levels measured by Olink's platform and mass spectrometry in individuals from Health ABC. Error bars indicate standard errors of the mean. **(D)** Predicted APOL1 protein structure (AlphaFold 2.0, visualized by PyMOL). Positions of the G1, G2 and V1 variants are highlighted.

Figure 5

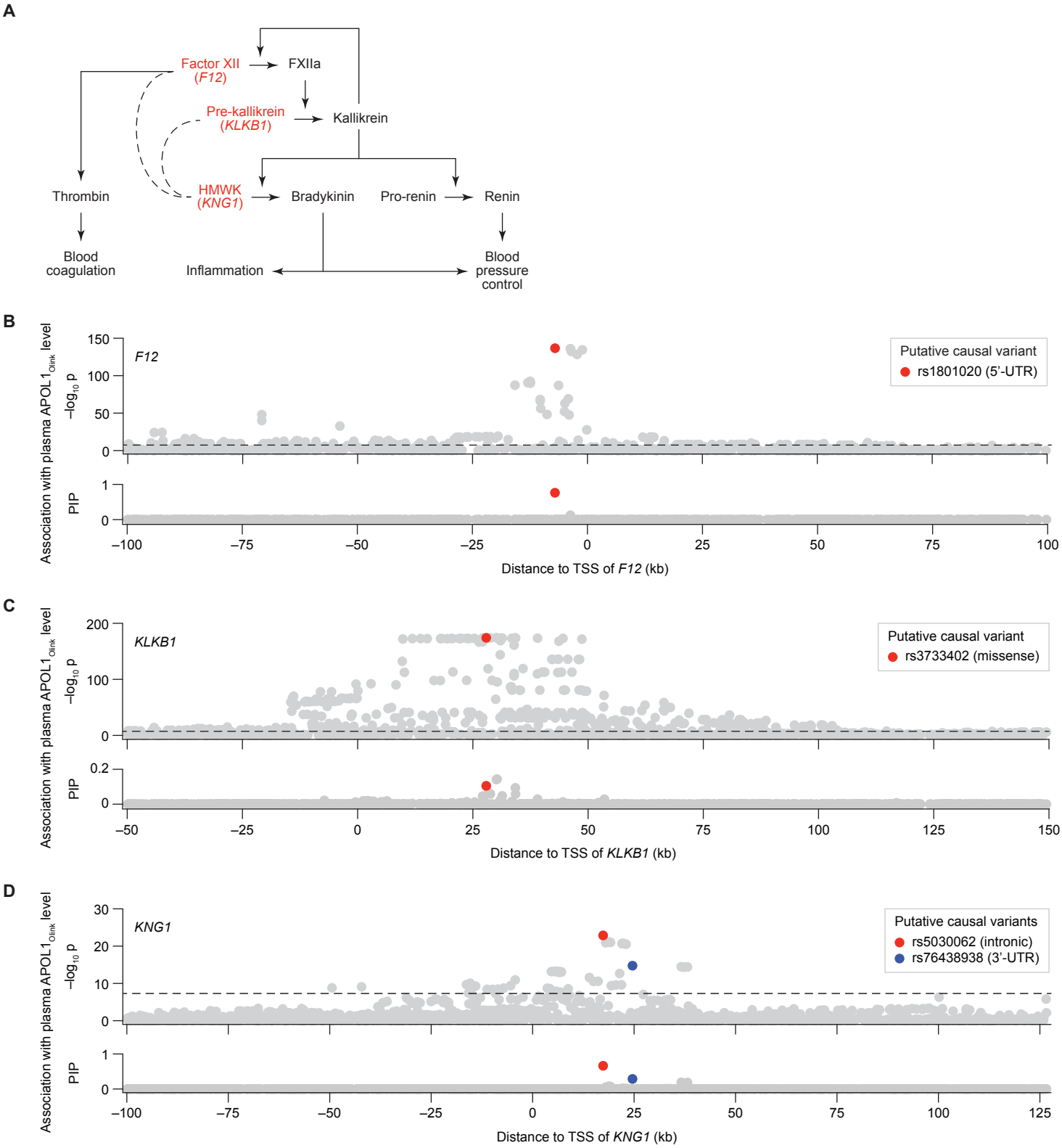


Figure 5. Variants in *F12*, *KLKB1* and *KNG1*, acting in the kallikrein-kinin pathway, are *trans*-pQTLs for plasma $APOL1_{Olink}$ levels. (A) A schematic illustration of the kallikrein-kinin system, highlighting the relative influence of factor XII (FXII, encoded by *F12*), pre-kallikrein (encoded by *KLKB1*) and high molecular weight kininogen (HMWK, encoded by *KNG1*) on blood coagulation, inflammation and blood pressure control mechanisms. Arrows indicate enzymatic reactions and regulation. Dashed lines indicate co-factor/enzyme relationships. **(B–D)** Manhattan plots showing the association and statistical fine-mapping of *cis-F12* **(B)**, *cis-KLKB1* **(C)** and *cis-KNG1* **(D)** variants with plasma $APOL1_{Olink}$ levels in the UK Biobank population of European ancestry. Each dot represents a variant. The x-axis show the variant's chromosomal position relative to the transcriptional start site (TSS) of the indicated gene (*F12*, *KLKB1* or *KNG1*). The y-axis shows the $-\log_{10}$ of the association p-value (row 1) and the fine-mapping posterior inclusion probability (PIP; row 2). Putative causal variants identified through fine-mapping are highlighted. Horizontal dashed lines represent the genome-wide significance threshold ($p = 5\times10^{-8}$).

Figure 6

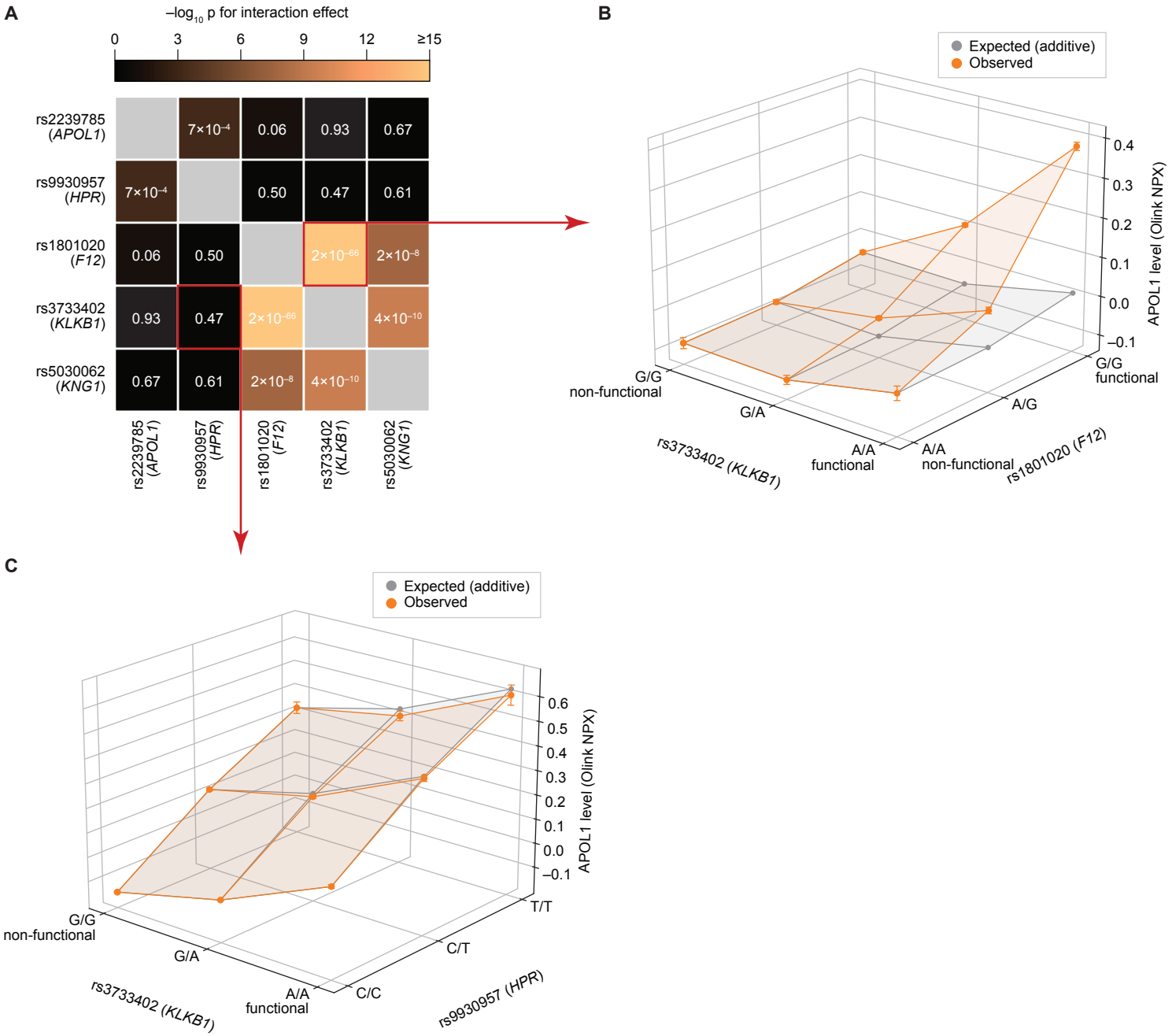


Figure 6. Variants in the kallikrein-kinin pathway genetically interact to influence plasma APOL1$_{Olink}$ levels. (A) Heatmap showing the significance of pairwise genetic interactions between the top five putative causal pQTLs for plasma APOL1$_{Olink}$ levels. Cell color indicates the –log$_{10}$ of the interaction p-value (Methods). **(B–C)** Examples of an interacting **(B)** and a non-interacting **(C)** variant pair. In each plot, the x- and y-axes show variant genotypes, and the z-axis shows observed (orange) and expected, assuming additivity (grey), mean plasma APOL1$_{Olink}$ levels for each genotype combination. Error bars indicate the standard error of the mean.

Figure 7

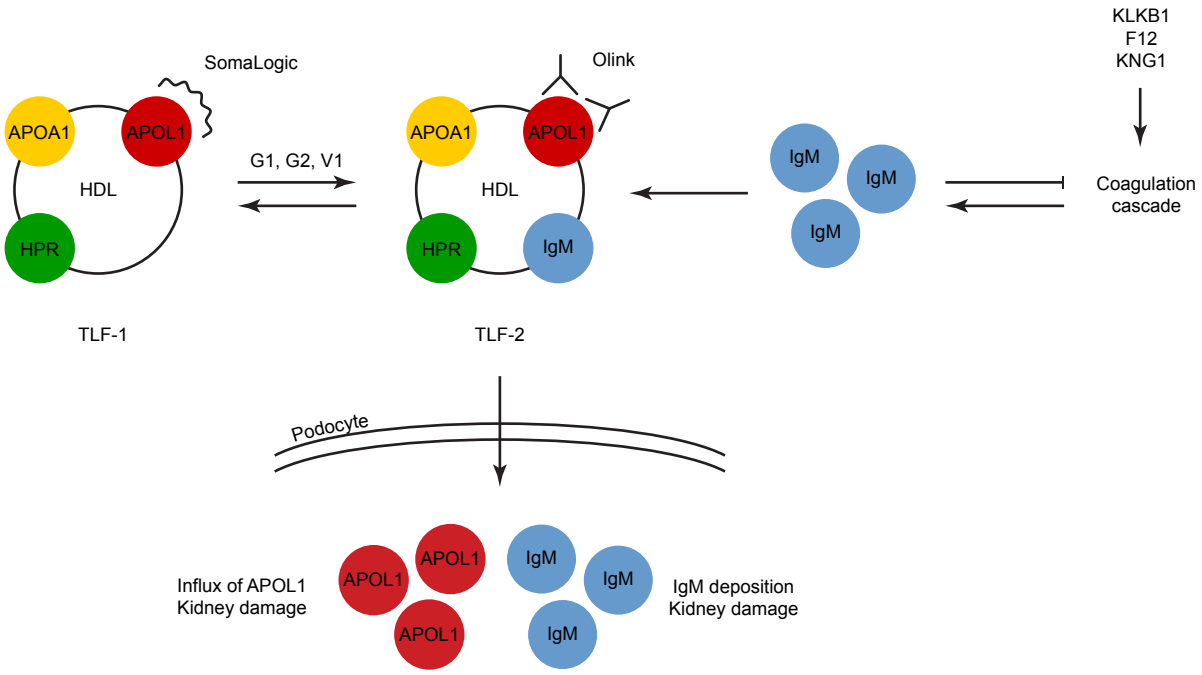


Figure 7. A model explaining platform-specific APOL1 pQTLs effects and their potential link to kidney disease. We propose that the Olink and SomaLogic platforms have different affinity for different HDL particles transporting APOL1 in plasma: the trypanosome lytic factor 1 (TLF-1) and 2 (TLF-2). *APOL1* G1/G2 variants, along with factors influencing IgM levels (e.g., the kallikrein-kinin system), can shift the balance between TLF-1 and TLF-2, leading to different APOL1 measurements across platforms. Unlike TLF-1, TLF-2 contains IgMs. Excess TLF-2 may promote IgM deposition in the kidney, a hallmark of APOL1-mediated nephropathies.

References:

Adamson, W.E. *et al.* (2024) 'APOL1 variants G1, G2 and N264K affect APOL1 plasma protein concentration: a UK Biobank study', *medRxiv*. Available at: https://doi.org/10.1101/2024.02.28.24303461.

Andrews, M. *et al.* (2022) 'Variant APOL1 protein in plasma associates with larger particles in humans and mouse models of kidney injury', *PloS one*, 17(10), p. e0276649.

Appel, L.J. *et al.* (2003) 'The rationale and design of the AASK cohort study', *Journal of the American Society of Nephrology : JASN*, 14(7 Suppl 2), pp. S166–72.

Backman, J.D. *et al.* (2021) 'Exome sequencing and analysis of 454,787 UK Biobank participants', *Nature*, 599(7886), pp. 628–634.

Baryshnikova, A. *et al.* (2013) 'Genetic interaction networks: toward an understanding of heritability', *Annual review of genomics and human genetics*, 14, pp. 111–133.

Benner, C. *et al.* (2016) 'FINEMAP: efficient variable selection using summary data from genome-wide association studies', *Bioinformatics (Oxford, England)*, 32(10), pp. 1493–1501.

Biswas, N. *et al.* (2016) 'Polymorphisms at the F12 and KLKB1 loci have significant trait association with activation of the renin-angiotensin system', *BMC medical genetics*, 17, p. 21.

Bruggeman, L.A. *et al.* (2014) 'Plasma apolipoprotein L1 levels do not correlate with CKD', *Journal of the American Society of Nephrology : JASN*, 25(3), pp. 634–644.

Bycroft, C. *et al.* (2018) 'The UK Biobank resource with deep phenotyping and genomic data', *Nature*, 562(7726), pp. 203–209.

Calafell, F. *et al.* (2010) 'Sequence variation and genetic evolution at the human F12 locus: mapping quantitative trait nucleotides that influence FXII plasma levels', *Human molecular genetics*, 19(3), pp. 517–525.

Cervia-Hasler, C. *et al.* (2024) 'Persistent complement dysregulation with signs of thromboinflammation in active Long Covid', *Science (New York, N.Y.)*, 383(6680), p. eadg7942.

Chee, Y.L. (2014) 'Coagulation', *The journal of the Royal College of Physicians of Edinburgh*, 44(1), pp. 42–45.

Cooper, A. *et al.* (2017) 'renal risk variants have contrasting resistance and susceptibility associations with African trypanosomiasis', *eLife*, 6. Available at: https://doi.org/10.7554/eLife.25461.

Cuypers, B. *et al.* (2016) 'Apolipoprotein L1 Variant Associated with Increased Susceptibility to Trypanosome Infection', *mBio*, 7(2), pp. e02198–15.

Egbuna, O. *et al.* (2023) 'Inaxaplin for Proteinuric Kidney Disease in Persons with Two Variants', *The New England journal of medicine*, 388(11), pp. 969–979.

Eldjarn, G.H. *et al.* (2023) 'Large-scale plasma proteomics comparisons through genetics and disease associations', *Nature*, 622(7982), pp. 348–358.

Elias, J.E. and Gygi, S.P. (2007) ‘Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry’, *Nature methods*, 4(3), pp. 207–214.

Eng, J.K., Jahan, T.A. and Hoopmann, M.R. (2013) ‘Comet: an open-source MS/MS sequence database search tool’, *Proteomics*, 13(1), pp. 22–24.

Ferguson, R. (1991) ‘End-stage renal disease and race: an overview and perspective’, *Journal of the National Medical Association*, 83(9), pp. 794–798.

Ferkingstad, E. *et al.* (2021) ‘Large-scale integration of the plasma proteome with genetics and disease’, *Nature genetics*, 53(12), pp. 1712–1721.

Fredriksson, S. *et al.* (2002) ‘Protein detection using proximity-dependent DNA ligation assays’, *Nature biotechnology*, 20(5), pp. 473–477.

Freedman, B.I. and Skorecki, K. (2014) ‘Gene-gene and gene-environment interactions in apolipoprotein L1 gene-associated nephropathy’, *Clinical journal of the American Society of Nephrology : CJASN*, 9(11), pp. 2006–2013.

Friedman, D.J. and Pollak, M.R. (2011) ‘Genetics of kidney failure and the evolving story of APOL1’, *The Journal of clinical investigation*, 121(9), pp. 3367–3374.

Friedman, D.J. and Pollak, M.R. (2020) ‘and Kidney Disease: From Genetics to Biology’, *Annual review of physiology*, 82, pp. 323–342.

Gaun, A. *et al.* (2021) ‘Automated 16-Plex Plasma Proteomics with Real-Time Search and Ion Mobility Mass Spectrometry Enables Large-Scale Profiling in Naked Mole-Rats and Mice’, *Journal of proteome research*, 20(2), pp. 1280–1295.

Gbadegesin, R.A. *et al.* (2024) ‘Bi- and Monoallelic Variants and Chronic Kidney Disease in West Africans’, *The New England journal of medicine* [Preprint]. Available at: https://doi.org/10.1056/NEJMoa2404211.

Genovese, G. *et al.* (2010) ‘Association of trypanolytic ApoL1 variants with kidney disease in African Americans’, *Science (New York, N.Y.)*, 329(5993), pp. 841–845.

Gephardt, G.N. *et al.* (1986) ‘Focal and segmental glomerulosclerosis. Immunohistologic study of 20 renal biopsy specimens’, *Archives of pathology & laboratory medicine*, 110(10), pp. 902–905.

Giovinazzo, J.A. *et al.* (2020) ‘Apolipoprotein L-1 renal risk variants form active channels at the plasma membrane driving cytotoxicity’, *eLife*, 9. Available at: https://doi.org/10.7554/eLife.51185.

Gold, L. *et al.* (2010) ‘Aptamer-based multiplexed proteomic technology for biomarker discovery’, *PloS one*, 5(12), p. e15004.

Gupta, N. *et al.* (2024) ‘Apolipoprotein L1 (APOL1) renal risk variant-mediated podocyte cytotoxicity depends on African haplotype and surface expression’, *Scientific reports*, 14(1), p. 3765.

Gupta, Y. *et al.* (2023) ‘Strong protective effect of the APOL1 p.N264K variant against G2-associated focal segmental glomerulosclerosis and kidney disease’, *Nature*

communications, 14(1), p. 7836.

Hajduk, S.L. *et al.* (1989) 'Lysis of Trypanosoma brucei by a toxic subspecies of human high density lipoprotein', *The Journal of biological chemistry*, 264(9), pp. 5210–5217.

Hardwick, R.J. *et al.* (2014) 'Haptoglobin (HP) and Haptoglobin-related protein (HPR) copy number variation, natural selection, and trypanosomiasis', *Human genetics*, 133(1), pp. 69–83.

Huttlin, E.L. *et al.* (2010) 'A tissue-specific atlas of mouse protein phosphorylation and expression', *Cell*, 143(7), pp. 1174–1189.

Jumper, J. *et al.* (2021) 'Highly accurate protein structure prediction with AlphaFold', *Nature*, 596(7873), pp. 583–589.

Karczewski, K.J. *et al.* (2024) 'Pan-UK Biobank GWAS improves discovery, analysis of genetic architecture, and resolution into ancestry-enriched effects'. Available at: https://doi.org/10.1101/2024.03.13.24303864.

Katsuda, I. *et al.* (2007) 'A new type of plasma prekallikrein deficiency associated with homozygosity for Gly104Arg and Asn124Ser in apple domain 2 of the heavy-chain region', *European journal of haematology*, 79(1), pp. 59–68.

Kozlitina, J. *et al.* (2016) 'Plasma Levels of Risk-Variant APOL1 Do Not Associate with Renal Disease in a Population-Based Cohort', *Journal of the American Society of Nephrology : JASN*, 27(10), pp. 3204–3219.

Lan, X. *et al.* (2014) 'APOL1 risk variants enhance podocyte necrosis through compromising lysosomal membrane permeability', *American journal of physiology. Renal physiology*, 307(3), pp. F326–36.

Motta, G., Juliano, L. and Chagas, J.R. (2023) 'Human plasma kallikrein: roles in coagulation, fibrinolysis, inflammation pathways, and beyond', *Frontiers in physiology*, 14, p. 1188816.

Newman, A.B. *et al.* (2023) 'The Health, Aging, and Body Composition (Health ABC) Study-Ground-Breaking Science for 25 Years and Counting', *The journals of gerontology. Series A, Biological sciences and medical sciences*, 78(11), pp. 2024–2034.

Norris Bradley, M. *et al.* (2020) 'Quality over quantity: A qualitative, targeted bottom-up proteomics approach to genotyping apolipoprotein L1', *Clinical biochemistry*, 82, pp. 58–65.

Obermayer, G. *et al.* (2021) 'Natural IgM antibodies inhibit microvesicle-driven coagulation and thrombosis', *Blood*, 137(10), pp. 1406–1415.

Obrador, G.T. *et al.* (2017) 'Genetic and environmental risk factors for chronic kidney disease', *Kidney international supplements*, 7(2), pp. 88–106.

Okada, Y. and Wang, Q.S. (2021) 'A massive effort links protein-coding gene variants to health', *Nature*, 599(7886), pp. 561–563.

Olabisi, O.A. *et al.* (2016) 'APOL1 kidney disease risk variants cause cytotoxicity by depleting cellular potassium and inducing stress-activated protein kinases', *Proceedings of the National Academy of Sciences of the United States of America*, 113(4), pp. 830–837.

Panzer, S.E. *et al.* (2015) 'IgM exacerbates glomerular disease progression in complement-induced glomerulopathy', *Kidney international*, 88(3), pp. 528–537.

Pietzner, M. *et al.* (2020) 'Genetic architecture of host proteins involved in SARS-CoV-2 infection', *Nature communications*, 11(1), p. 6397.

Raper, J. *et al.* (1999) 'Characterization of a novel trypanosome lytic factor from human serum', *Infection and immunity*, 67(4), pp. 1910–1916.

Rifkin, M.R. (1978) 'Identification of the trypanocidal factor in normal human serum: high density lipoprotein', *Proceedings of the National Academy of Sciences of the United States of America*, 75(7), pp. 3450–3454.

Schweppe, D.K. *et al.* (2020) 'Full-Featured, Real-Time Database Searching Platform Enables Fast and Accurate Multiplexed Quantitative Proteomics', *Journal of proteome research*, 19(5), pp. 2026–2034.

Sham, P.C. and Purcell, S.M. (2014) 'Statistical power and significance testing in large-scale genetic studies', *Nature reviews. Genetics*, 15(5), pp. 335–346.

Shiflett, A.M. *et al.* (2005) 'Human high density lipoproteins are platforms for the assembly of multi-component innate immune complexes', *The Journal of biological chemistry*, 280(38), pp. 32578–32585.

Shukha, K. *et al.* (2017) 'Most ApoL1 Is Secreted by the Liver', *Journal of the American Society of Nephrology : JASN*, 28(4), pp. 1079–1083.

Strassheim, D. *et al.* (2013) 'IgM contributes to glomerular injury in FSGS', *Journal of the American Society of Nephrology : JASN*, 24(3), pp. 393–406.

Sudlow, C. *et al.* (2015) 'UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age', *PLoS medicine*, 12(3), p. e1001779.

Sula Karreci, E. *et al.* (2024) 'Differing sensitivities to angiotensin converting enzyme inhibition of kidney disease mediated by APOL1 high-risk variants G1 and G2', *Kidney international*, 106(6), pp. 1072–1085.

Sun, B.B. *et al.* (2023) 'Plasma proteomic associations with genetics and health in the UK Biobank', *Nature*, 622(7982), pp. 329–338.

Surapaneni, A. *et al.* (2022) 'Identification of 969 protein quantitative trait loci in an African American population with kidney disease attributed to hypertension', *Kidney international*, 102(5), pp. 1167–1177.

Tin, A. *et al.* (2015) 'Hemostatic Factors, APOL1 Risk Variants, and the Risk of ESRD in the Atherosclerosis Risk in Communities Study', *Clinical journal of the American Society of Nephrology : CJASN*, 10(5), pp. 784–790.

Tzur, S. *et al.* (2010) 'Missense mutations in the APOL1 gene are highly associated with end stage kidney disease risk previously attributed to the MYH9 gene', *Human genetics*, 128(3), pp. 345–350.

Vanhollebeke, B. and Pays, E. (2010) 'The trypanolytic factor of human serum: many ways to

enter the parasite, a single way to kill', *Molecular microbiology*, 76(4), pp. 806–814.

Vasquez-Rios, G., De Cos, M. and Campbell, K.N. (2023) 'Novel Therapies in -Mediated Kidney Disease: From Molecular Pathways to Therapeutic Options', *Kidney international reports*, 8(11), pp. 2226–2234.

Verdi, J. *et al.* (2020) 'Inducible Germline IgMs Bridge Trypanosome Lytic Factor Assembly and Parasite Recognition', *Cell host & microbe*, 28(1), pp. 79–88.e4.

Wang, G. *et al.* (2020) 'A simple new approach to variable selection in regression, with application to genetic fine mapping', *Journal of the Royal Statistical Society. Series B, Statistical methodology*, 82(5), pp. 1273–1300.

Weckerle, A. *et al.* (2016) 'Characterization of circulating APOL1 protein complexes in African Americans', *Journal of lipid research*, 57(1), pp. 120–130.

Wisniewski, P., Gangnus, T. and Burckhardt, B.B. (2024) 'Recent advances in the discovery and development of drugs targeting the kallikrein-kinin system', *Journal of translational medicine*, 22(1), p. 388.

Yu, H. *et al.* (1998) 'Identification of human plasma kallikrein gene polymorphisms and evaluation of their role in end-stage renal disease', *Hypertension (Dallas, Tex. : 1979)*, 31(4), pp. 906–911.

Yu, H. *et al.* (2000) 'Genomic structure of the human plasma prekallikrein gene, identification of allelic variants, and analysis in end-stage renal disease', *Genomics*, 69(2), pp. 225–234.

Zhang, J. *et al.* (2022) 'Plasma proteome analyses in individuals of European and African ancestry identify cis-pQTLs and models for proteome-wide association studies', *Nature genetics*, 54(5), pp. 593–602.

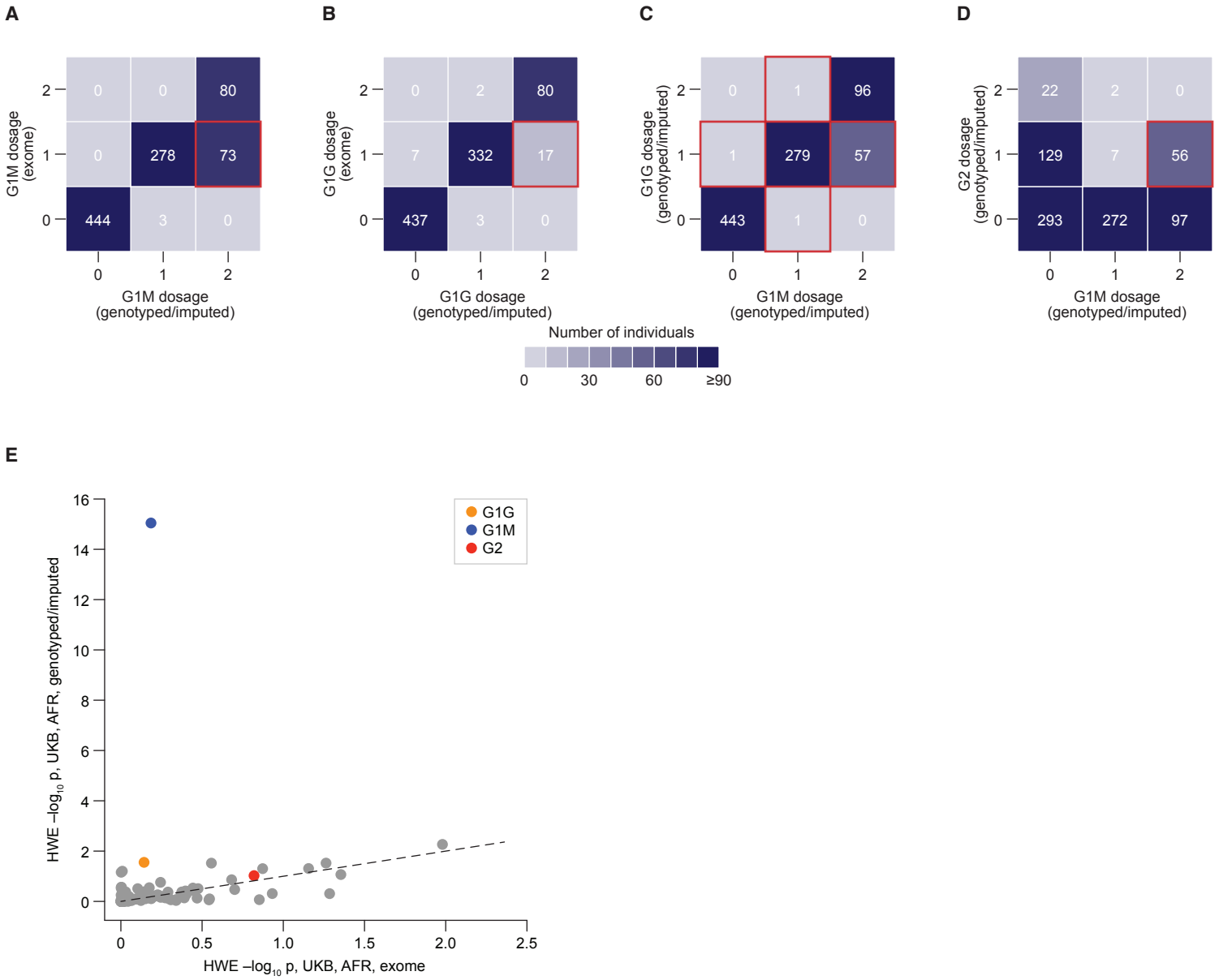
Supplementary Material

Platform-dependent effects of genetic variants on plasma APOL1 and their implications for kidney disease

Qingbo S. Wang[1], Jinguo Huang[1], Leanne Chan[1], Nicole Haste[1], Niclas Olsson[1], Aleksandr Gaun[1,2], Fiona McAllister[1], Deepthi Madhireddy[1], Amos Baruch[1], Eugene Melamud[1], Anastasia Baryshnikova[1,3]

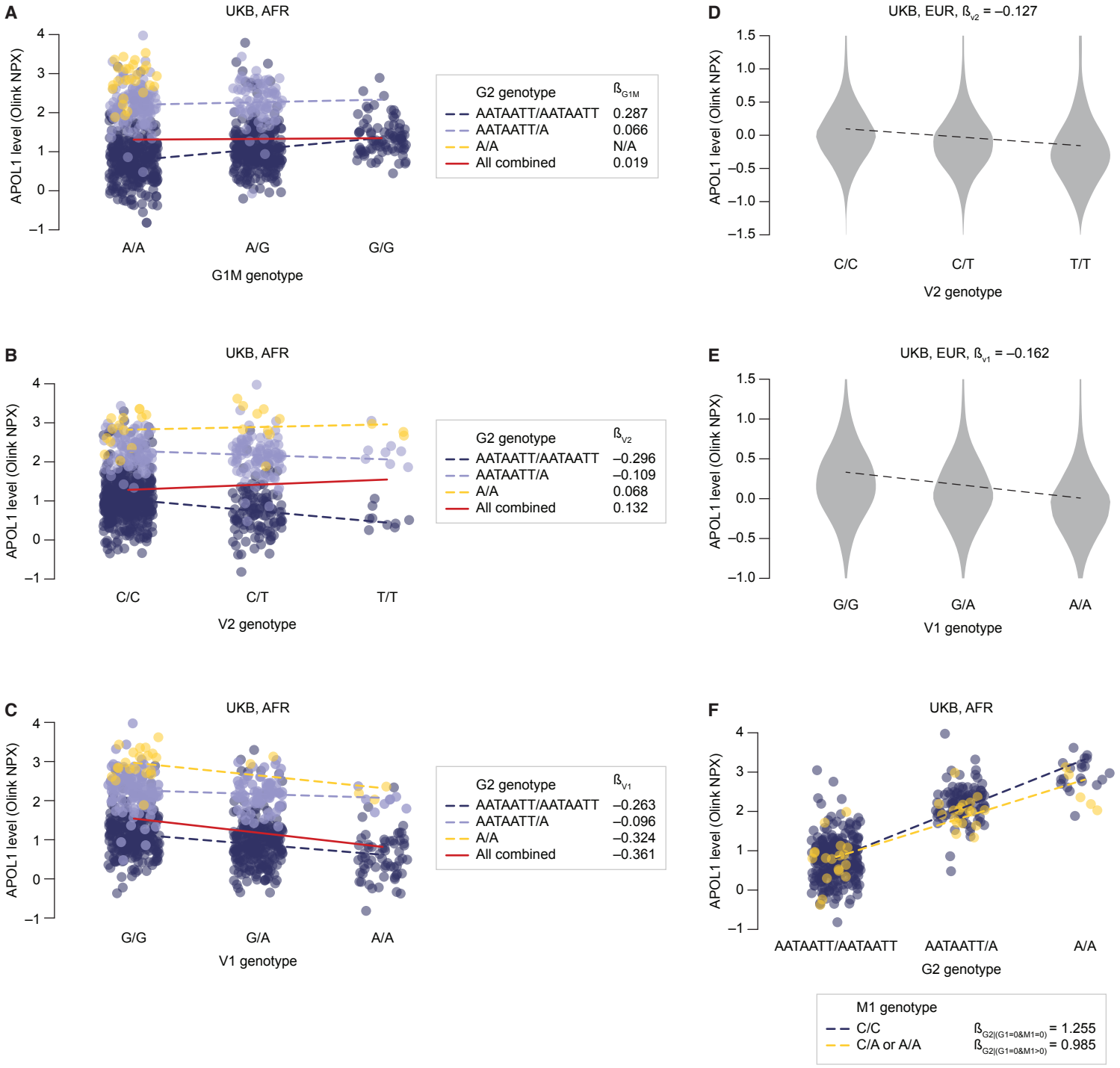
1. Calico Life Sciences LLC, South San Francisco, CA, USA
2. Present address: Flagship Pioneering, Cambridge, MA, USA
3. Corresponding author. E-mail: abaryshnikova@calicolabs.com

Figure S1



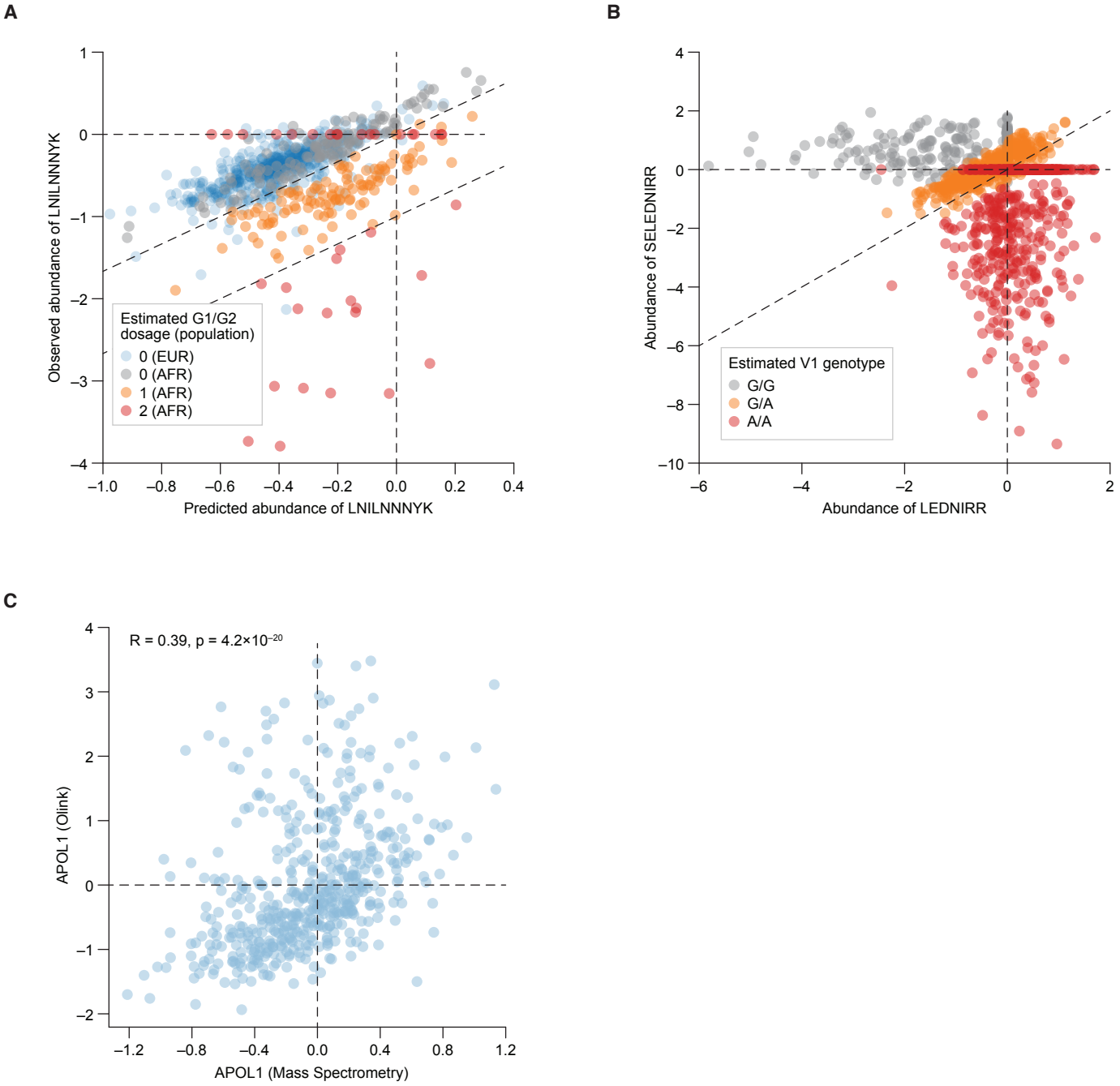
Supplementary Figure 1. Exome sequencing reveals discrepancies in *APOL1* G1M imputation in the UK biobank. Genotypes for the *APOL1* G1M, G1G and G2 variants were obtained for 878 UK Biobank-PPP participants with African ancestry (UKB, AFR) using both imputed genotype array data and exome sequencing. **(A–B)** Heatmaps comparing genotype calls for the G1M **(A)** and G1G **(B)** variants obtained from imputed (x-axis) and exome-derived data (y-axis). Each cell represents the number of individuals with the corresponding genotypes. Non-zero values off the diagonal represent conflicting genotype calls. For example, over 20% (73 of 351; **A**, red square) of heterozygous G1M carriers and 5% (17 of 349; **B**, red square) of heterozygous G1G carriers identified by exome sequencing are misidentified as homozygous in the imputed genotype data. **(C–D)** Heatmaps comparing genotype calls for the G1M (x-axis) and G1G (**C**; y-axis) or G2 (**D**; y-axis) variants obtained from imputed data. Each cell represents the number of individuals with the corresponding genotypes. Improbable genotype combinations are highlighted in red. For example, 60 individuals show discordant genotypes for G1G and G1M **(C)**. Similarly, 56 individuals appear to be homozygous for G1M and heterozygous for G2, an unlikely genotype pattern given that G1 and G2 never occur on the same haplotype **(D)**. **(E)** Scatter plot comparing deviation from Hardy-Weinberg Equilibrium (HWE) for *cis-APOL1* variants estimated from exome-derived data (x-axis) and imputed genotypes (y-axis). Each dot corresponds to a variant, with G1M being a clear outlier.

Figure S2



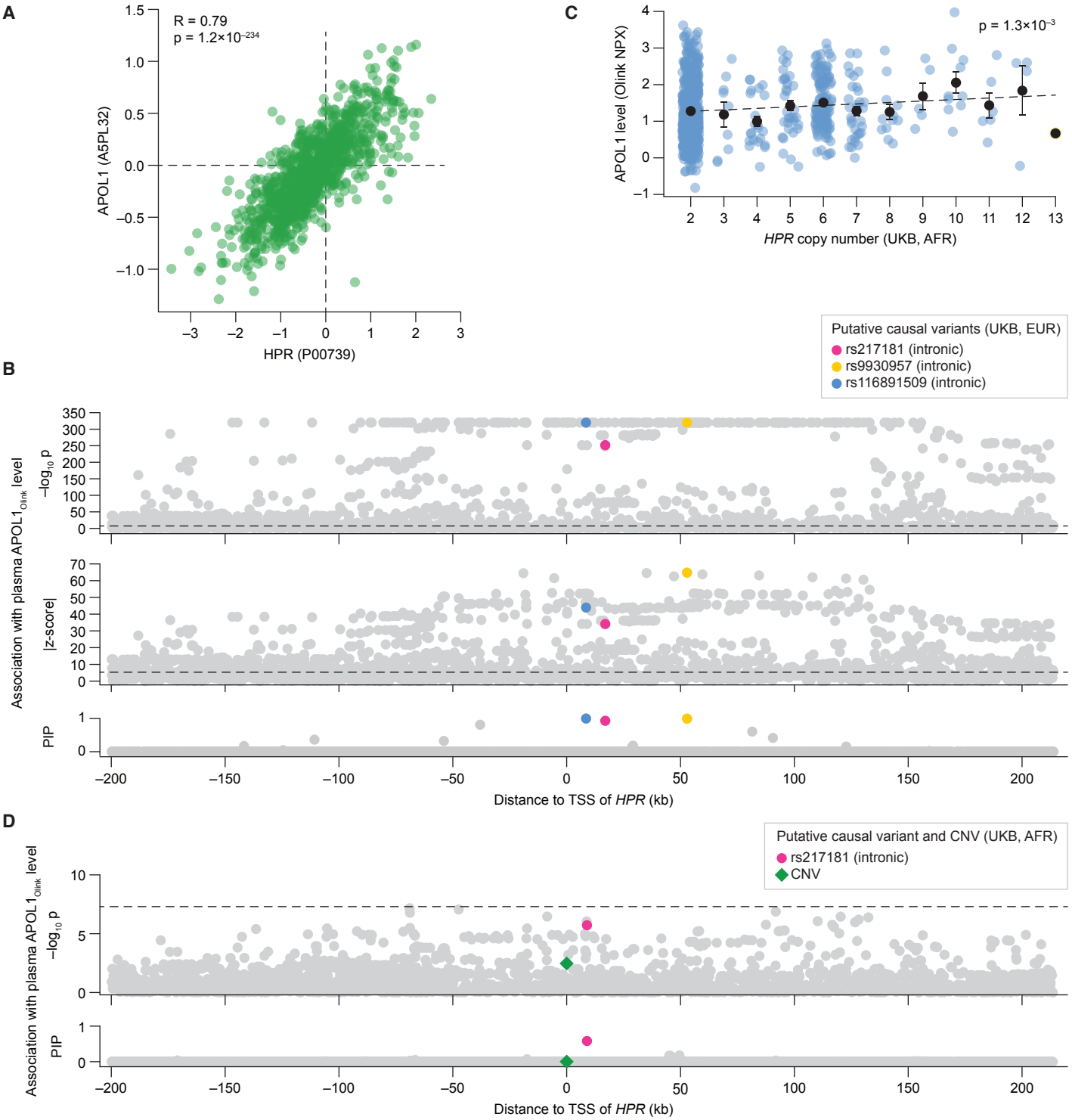
Supplementary Figure 2. Linkage disequilibrium with G2 and genetic interactions with M1 influence the estimation of effect sizes for *APOL1* cis-variants. (A–C) Scatter plots showing the effects of G1M **(A)**, V2 **(B)** or V1 **(C)** genotypes (x-axis) on plasma APOL1 levels (measured by Olink; y-axis) in UK Biobank participants with African ancestry (UKB, AFR), stratified by G2 genotype. Each data point represents an individual. Data points are colored according to G2 genotype: navy (REF/REF), light blue (REF/ALT) and yellow (ALT/ALT). Within each G1M **(A)**, V2 **(B)** or V1 **(C)** genotype group, data points are jittered along the x-axis for visualization purposes. Lines represent linear regression fits for each G2 genotype group (dashed, color-coded) and the overall trend (solid, red). **(D–E)** Violin plots showing the effects of V2 **(D)** or V1 **(E)** genotypes (x-axis) on plasma APOL1 levels (measured by Olink; y-axis) in UK Biobank participants with European ancestry (UKB, EUR). Dashed lines represent linear regression fits. **(F)** Scatter plot showing the effect of the G2 genotype (x-axis) on plasma APOL1 levels (measured by Olink; y-axis) in UK Biobank participants with African ancestry (UKB, AFR), stratified by M1 genotype. Each data point represents an individual. Data points are colored according to M1 genotype: navy (REF/REF), and yellow (REF/ALT or ALT/ALT). Within each G2 genotype group, data points are jittered along the x-axis for visualization purposes. Lines represent linear regression fits for each M1 genotype group (dashed, color-coded).

Figure S3



Supplementary Figure 3. In Health ABC, *APOL1* G1/G2 and V1 variant genotypes can be inferred from peptide sequence analysis. (A) Inference of G1/G2 ALT dosage from peptide sequence analysis in mass spectrometry (Methods). Each data point represents an individual. The dosage of the ALT alleles for G1M/G2 in each individual was estimated by comparing the predicted (x-axis) and observed (y-axis) abundance of the peptide LNILNNNYK, which overlaps both the G1M and G2 variants. Predictions were generated using a lasso model trained on G1/G2-independent peptides in individuals with European ancestry (virtually all REF/REF) and applied to individuals with African ancestry (Methods). Data points are colored according to their ancestry (blue for EUR, other colors for AFR) and inferred G1/G2 genotype (grey for REF/REF, orange for REF/ALT and red for ALT/ALT). Dashed lines indicate classification boundaries. **(B)** Inference of V1 ALT dosage from peptide sequence analysis in mass spectrometry (Methods). Each data point represents an individual. The dosage of the ALT allele for V1 in each individual was estimated by comparing the observed abundance of the LEDNIRR (x-axis) and SELEDNIRR (y-axis) peptides, which correspond to the V1 A (ALT) and G (REF) alleles, respectively. Data points are colored according to their inferred genotypes (grey for REF/REF, orange for REF/ALT and red for ALT/ALT). Dashed lines indicate classification boundaries. **(C)** Scatter plot comparing plasma APOL1 levels measured by mass spectrometry (x-axis) and Olink's platform (y-axis) in the same individuals from Health ABC.

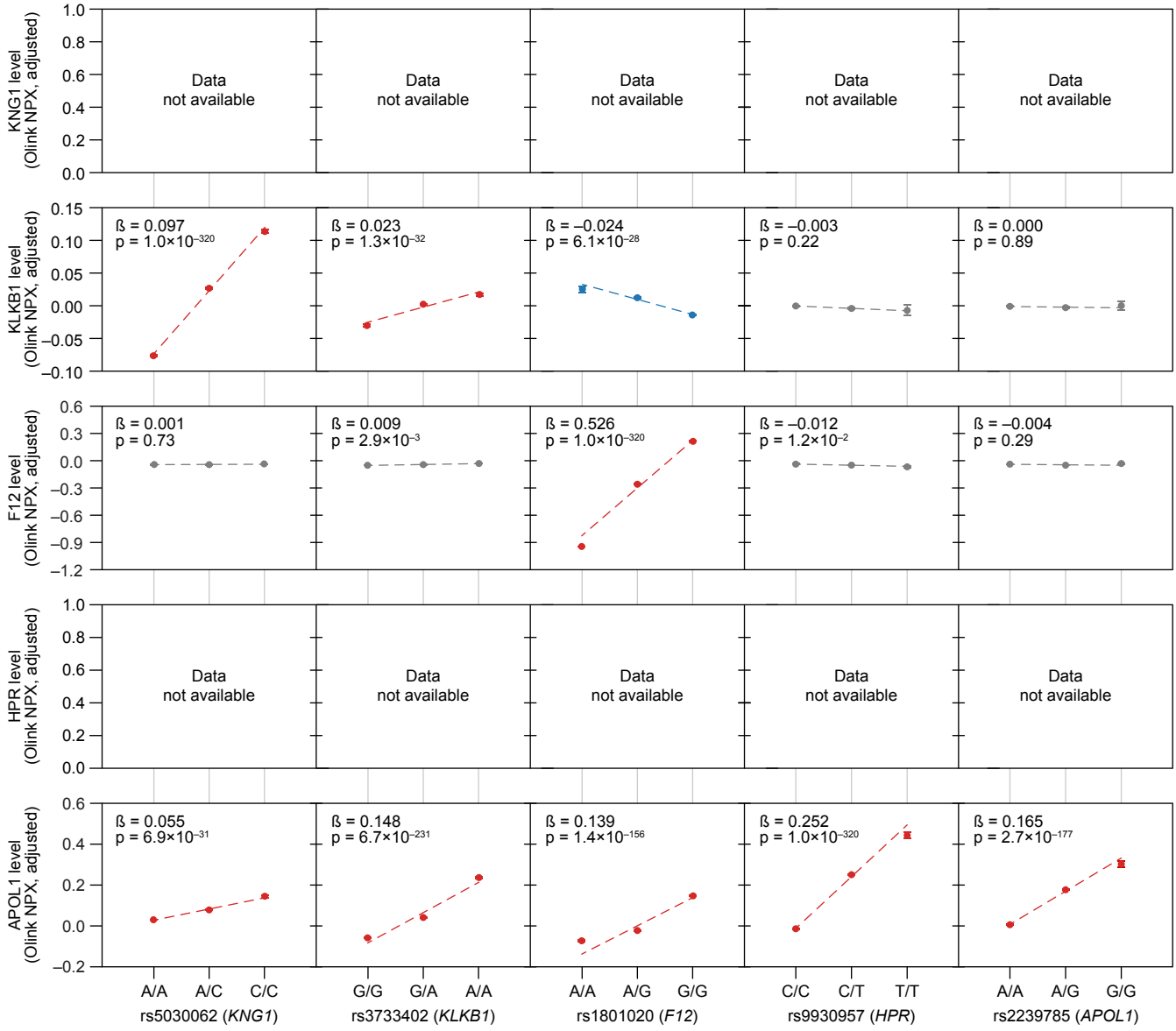
Figure S4



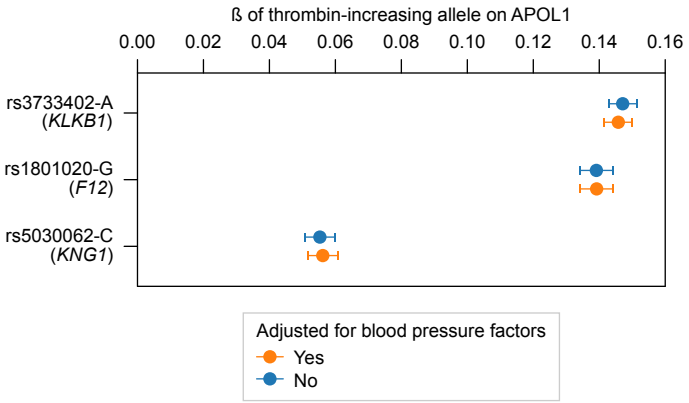
Supplementary Figure 4. HPR cis-variants, but not copy number variation, are associated with plasma $APOL1_{Olink}$ levels. (A) Scatter plot comparing plasma levels of HPR (x-axis) and APOL1 (y-axis) measured by mass spectrometry in the same individuals from Health ABC. **(B)** Manhattan plot showing the association of *HPR cis*-variants with plasma $APOL1_{Olink}$ levels in the UK Biobank population of European descent. Each dot represents a variant. The x-axis shows the variant's chromosomal position relative to the transcriptional start site (TSS) of *HPR*. The y-axis shows the $-\log_{10}$ of the association p-value (capped at 320; row 1), the absolute value of the association z-score (row 2) and the statistical fine-mapping posterior inclusion probability (PIP; row 3). Putative causal variants identified through fine-mapping are highlighted. Horizontal dashed lines represent the genome-wide significance threshold ($p = 5\times10^{-8}$). **(C)** Scatter plot comparing estimated *HPR* copy number (x-axis) and plasma $APOL1_{Olink}$ levels (y-axis) in the UK Biobank population of African descent. Each dot represents an individual. Within each copy number group, dots were jittered along the x-axis for visualization purposes. Black dots and corresponding error bars indicate the mean and standard error of the mean of each copy number group. The horizontal dashed line represents a linear regression fit. **(D)** Manhattan plot showing the association of *HPR cis*-variants and copy number variation (CNV) with plasma $APOL1_{Olink}$ levels in the UK Biobank population of African descent. Each dot represents a variant. The x-axis shows the variant's chromosomal position relative to the TSS of *HPR*, except for the CNV, which is plotted at the TSS directly. The y-axis shows the $-\log_{10}$ of the association p-value (row 1) and the statistical fine-mapping PIP (row 2). Putative causal variants identified through fine-mapping and the CNV are highlighted. Horizontal dashed line represents the genome-wide significance threshold ($p = 5\times10^{-8}$).

Figure S5

A

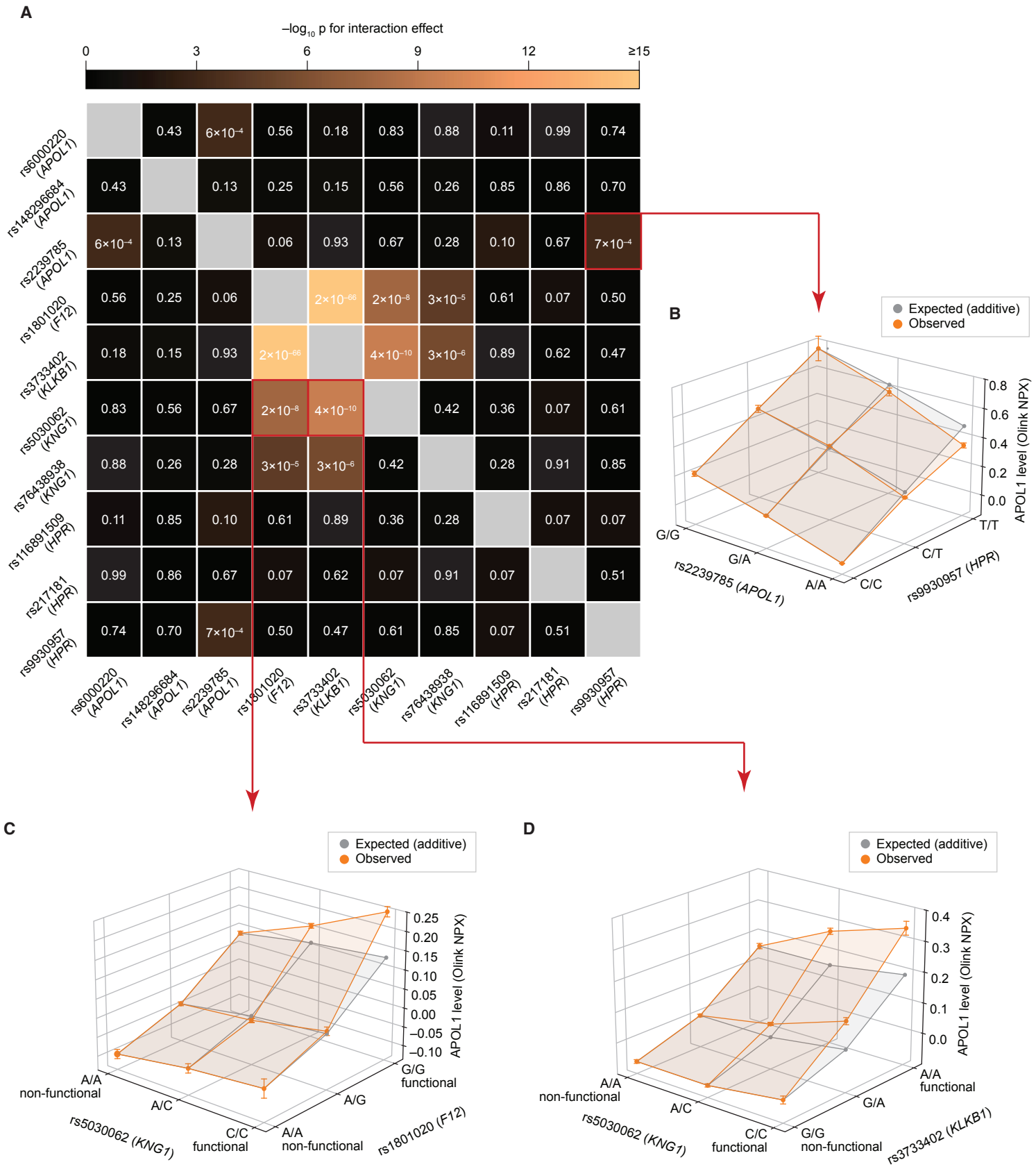


B



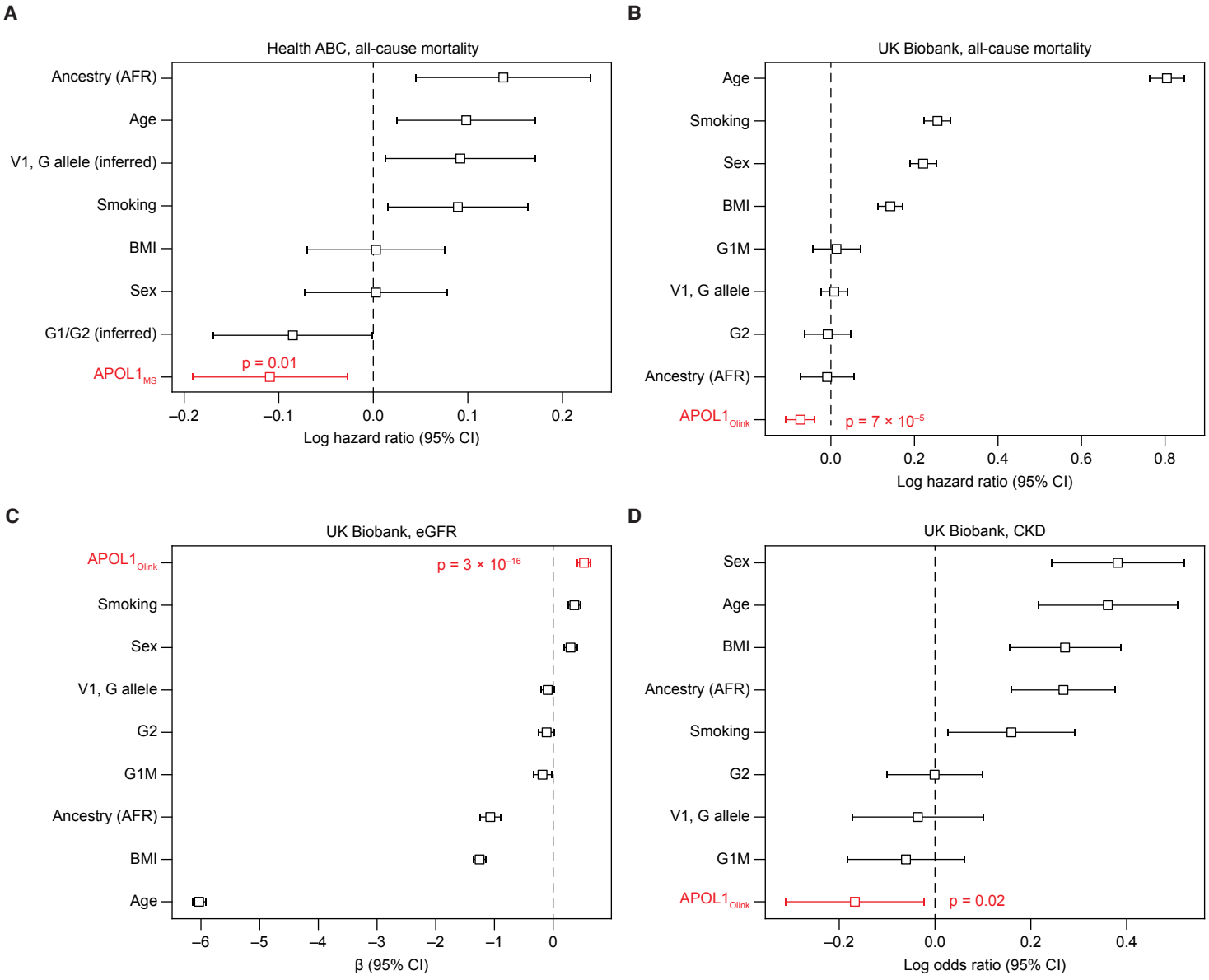
Supplementary Figure 5. Comparison of pQTL effects suggests a positive correlation between the activity of the kallikrein-kinin pathway and plasma APOL1$_{Olink}$ levels. (A) Dot plots showing the effects of KNG1, *KLKB1*, *F12*, *HPR*, and *APOL1* variants on covariate-adjusted plasma levels of KLKB1, F12, and APOL1$_{Olink}$ in individuals with European ancestry from the UK Biobank (Methods). Data for plasma KNG1 and HPR were not available as these proteins were not included in the Olink Explore 3072 platform used for the measurements. The mean protein level (y-axis) is shown for each variant genotype (x-axis). Error bars indicate standard errors of the mean. Associations with a significant positive or negative effect ($p < 5\times10^{-8}$) are colored in red and blue, respectively. **(B)** Forest plot showing the association of thrombin-increasing alleles of *KLKB1* (rs3733402-A), *F12* (rs1801020-G) and *KNG1* (rs5030062-C) with plasma APOL1$_{Olink}$ levels, both with and without adjustment for blood pressure-related factors (Methods). Effect sizes are reported as β on covariate adjusted APOL1$_{Olink}$ level (Methods). Error bars indicate standard errors of the mean.

Figure S6



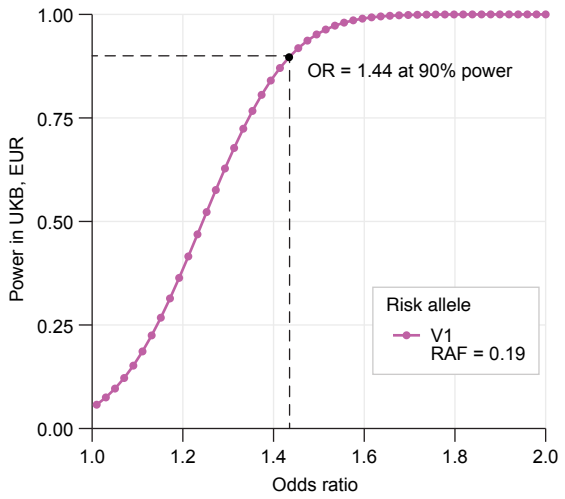
Supplementary Figure 6. Variants in *KNG1*x*F12*, *KNG1*x*KLKB1*, and *HPR*x*APOL1* genetically interact to influence plasma APOL1$_{Olink}$ levels. (A) Heatmap showing the significance of pairwise genetic interactions between all major putative causal pQTLs for plasma APOL1$_{Olink}$ levels. Cell color indicates the $-\log_{10}$ of the interaction p-value (Methods). **(B–D)** Examples of interacting variant pairs, in addition to those displayed in Fig. 6. In each plot, the x- and y-axes show variant genotypes, and the z-axis shows observed (orange) and expected, assuming additivity (grey), mean plasma APOL1$_{Olink}$ levels for each genotype combination. Error bars indicate the standard error of the mean.

Figure S7



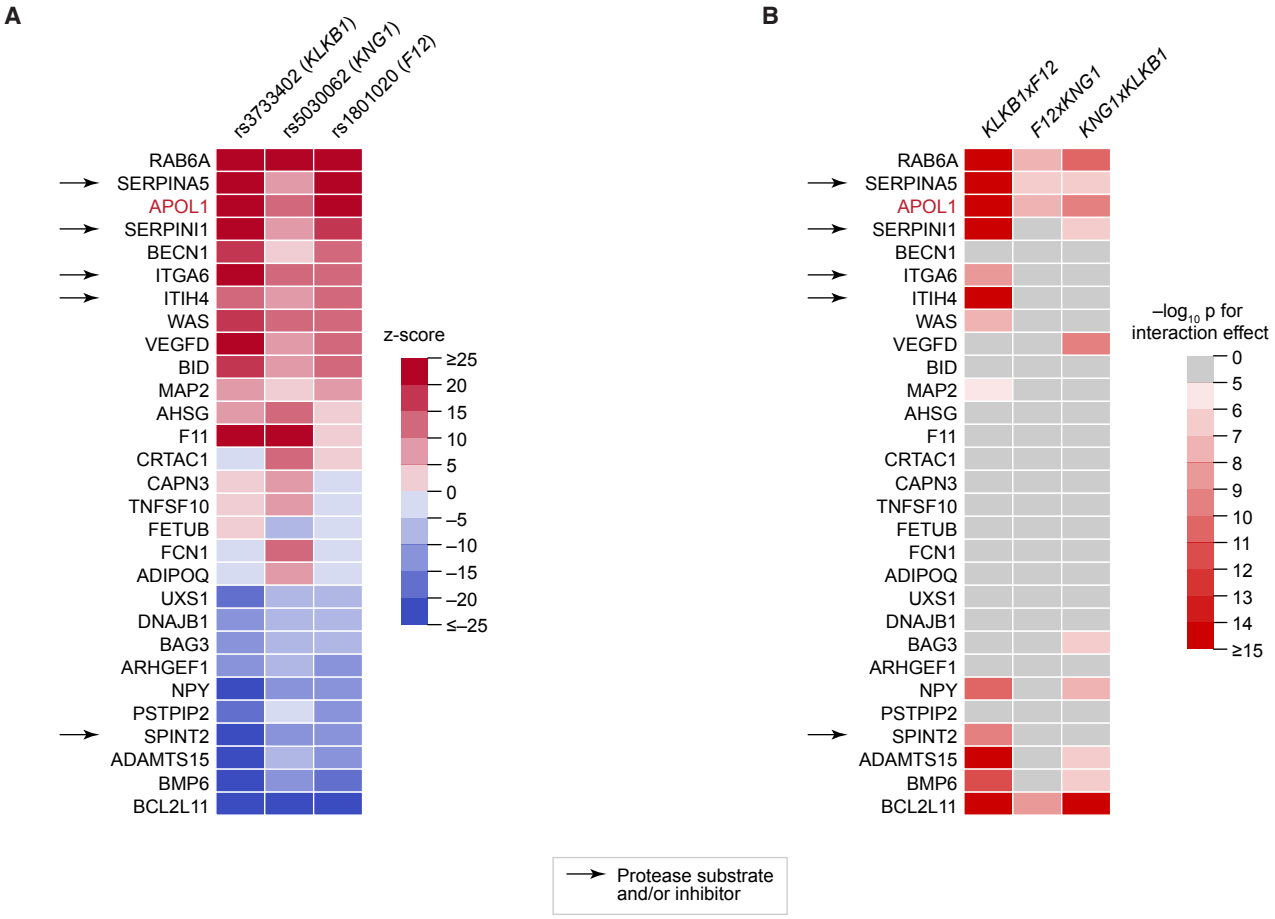
Supplementary Figure 7. Higher plasma APOL1 levels are associated with a reduced risk of all-cause mortality and kidney disease. (A–B) Association between plasma APOL1 levels and all-cause mortality in the Health ABC cohort **(A)** and the UK Biobank **(B)**. A Cox proportional hazards model was used to assess the effect of baseline plasma $APOL1_{MS}$ **(A)** and $APOL1_{Olink}$ levels **(B)** on future mortality events, adjusting for age at recruitment, sex (0 = female, 1 = male), genetic ancestry (0 = European, 1 = African), smoking status (0 = never, 1 = ever) and body mass index (BMI). The forest plots display the log hazard ratio (HR) and the 95% confidence interval for each variable. **(C)** Association between plasma $APOL1_{Olink}$ levels and estimated glomerular filtration rate (eGFR) in the UK Biobank. A linear regression model was used to assess the relationship between plasma $APOL1_{Olink}$ levels and eGFR, adjusting for age at recruitment, sex, genetic ancestry, smoking status and BMI. The forest plot displays the regression coefficients for each variable. eGFR was calculated using the 2021 CKD-EPI creatinine formula (Inker et al., 2021). **(D)** Association between plasma $APOL1_{Olink}$ levels and CKD diagnosis in the UK Biobank. A logistic regression model was used to assess the relationship between plasma $APOL1_{Olink}$ levels and CKD, adjusting for age at recruitment, sex, genetic ancestry, smoking status and BMI. The forest plot displays the regression coefficients for each variable. CKD diagnosis was determined using ICD-10 code N18.

Figure S8



Supplementary Figure 8. Power analysis suggests limited ability to detect the effect of rs2239785 (V1) on end-stage kidney disease (ESRD) in the UK Biobank. Power to detect the effect of V1 at a significance level of $\alpha = 0.05$ was calculated for a range of expected odds ratios, given the size of the UK Biobank population with European ancestry (UKB, EUR) and the prevalence of ESRD and risk allele frequency (RAF) in that population (Methods). Power (y-axis) is shown as a function of the odds ratio (x-axis). Dashed lines indicate that the minimal odds ratio detectable at 90% power is 1.44. In our analysis of ESRD in the UK Biobank population with European ancestry, we did not observe a significant effect of V1 ($p > 0.05$). This result is consistent with a study in the hospital-based Penn Medicine Biobank (Zhang et al., 2024), which also failed to detect a significant V1 ESRD effect, independent of G1/G2. In PMBB, G1/G2 are associated with ESRD with OR = 1.6–1.7. Given the much weaker pQTL effect of V1 relative to G1/G2, and assuming that its association with ESRD would be similarly weaker (i.e., much lower than the minimal detectable OR of 1.44), we conclude that the UK Biobank and PMBB are likely underpowered to detect the clinical impact of the V1 variant. In support of this hypothesis, previous case/control studies in African Americans with diabetic and non-diabetic nephropathies successfully reported a significant association between the V1 G allele and increased disease risk, independent of G1/G2 (Bostrom et al., 2012; Bailey et al., 2014), suggesting that the clinical relevance of V1 is more readily detected in disease-focused populations.

Figure S9



Supplementary Figure 9. Variants in *KLKB1*, *F12* and *KNG1* are associated with plasma levels of multiple protease substrates and inhibitors. (A) Heatmap showing the associations between rs3733402 (*KLKB1*), rs1801020 (*F12*), and rs5030062 (*KNG1*) and the plasma levels of 29 proteins in individuals with European ancestry from the UK Biobank (Methods). Cell color indicates the signed z-score of the association test. Only proteins ranking in the top 20 strongest associations for at least one of the three variants are shown. **(B)** Heatmap showing the significance of the rs3733402 (*KLKB1*) x rs1801020 (*F12*), rs1801020 (*F12*) x rs5030062 (*KNG1*), and rs5030062 (*KNG1*) x rs3733402 (*KLKB1*) genetic interactions for plasma levels of the 29 proteins from **(A)**. Cell color indicates the $-\log_{10}$ of the interaction p-value. Interactions with $p > 10^{-5}$ are shown in grey. **(A–B)** The F12 and KLKB1 proteins themselves are omitted from the heatmaps for simplicity.

Table S1. Demographic characteristics of the Health ABC sub-cohorts that were used for mass spectrometry and Olink analyses.

	Sub-cohort examined by mass spectrometry	Sub-cohort examined by Olink
Number of participants: n	1094	560
Age at baseline: mean (SD)	73.96 (3.21)	73.82 (3.14)
BMI: mean (SD)	27.40 (4.55)	27.72 (4.50)
Systolic blood pressure: mean (SD)	136.23 (20.46)	135.78 (20.04)
Cystatin C: mean (SD)	1.03 (0.30)	1.03 (0.29)
Vitamin D: mean (SD)	25.97 (11.44)	25.78 (10.85)
C-reactive protein (CRP): mean (SD)	6.93 (27.06)	8.36 (32.65)
Albumin/creatinine ratio in urine: mean (SD)	41.38 (168.60)	41.19 (187.01)
Creatinine: mean (SD)	1.05 (0.32)	1.04 (0.32)
Fasting glucose: mean (SD)	102.39 (29.66)	104.23 (31.02)
Hemoglobin A1C: mean (SD)	6.26 (0.96)	6.31 (1.02)
LDL: mean (SD)	121.62 (33.39)	122.40 (33.20)
Triglycerides: mean (SD)	140.25 (89.96)	144.02 (82.73)
HDL: mean (SD)	52.58 (15.82)	52.35 (16.46)
Gender (female): n (%)	554 (50.6)	286 (52.3)
Race (African): n (%)	390 (35.6)	194 (35.5)
Education level: n (%)		
Less than high school	262 (23.9)	144 (26.3)
High school graduate	344 (31.4)	176 (32.2)
Postsecondary	485 (44.3)	226 (41.3)
Former smoker: n (%)	550 (50.3)	268 (49.0)
Chronic kidney disease (CKD): n (%)	18 (1.6)	10 (1.8)
Diabetes: n (%)	147 (13.4)	79 (14.4)

References

Bailey, J.N.C. *et al.* (2014) ‘Analysis of coding variants identified from exome sequencing resources for association with diabetic and non-diabetic nephropathy in African Americans’, *Human genetics*, 133(6), pp. 769–779.

Bostrom, M.A. *et al.* (2012) ‘Genetic association and gene-gene interaction analyses in African American dialysis patients with nondiabetic nephropathy’, *American journal of kidney diseases : the official journal of the National Kidney Foundation*, 59(2), pp. 210–221.

Inker, L.A. *et al.* (2021) ‘New Creatinine- and Cystatin C-Based Equations to Estimate GFR without Race’, *The New England journal of medicine*, 385(19), pp. 1737–1749.

Zhang, D.Y. *et al.* (2024) ‘Protein-truncating variant in APOL3 increases chronic kidney disease risk in epistasis with APOL1 risk alleles’, *JCI insight*, 9(19). Available at: https://doi.org/10.1172/jci.insight.181238.

Author contributions

Qingbo S. Wang[2,3,5,6,8,9,10], Jinguo Huang[5,9,10], Leanne Chan[6,7], Nicole Haste[6,7], Niclas Olsson[6,7,9], Aleksandr Gaun[6,7], Fiona McAllister[6,7,9,11], Deepthi Madhireddy[6,7], Amos Baruch[6,7,9,11], Eugene Melamud[9,11], Anastasia Baryshnikova[1,2,8,9,10,11,12]

1. Conceptualization: ideas; formulation of the overarching research goals and aims
2. Methodology: development or design of methodology or creation of models
3. Software: programming; software development; designing computer programs; implementation of computer code or algorithms; testing code components
4. Validation: verification of the replication and reproducibility of results, experiments, or other research outputs
5. Formal analysis: application of statistical, mathematical, computational, or other techniques to analyse or synthesise data
6. Investigation: conducting the research and investigation process, specifically performing the experiments or data collection
7. Resources: provision of study materials, reagents, materials, patients, laboratory samples, animals, instrumentation, computing resources, or analysis tools
8. Writing—original draft: preparation, creation and/or presentation of the published work, specifically writing the initial draft
9. Writing—review & editing: critical review, commentary, or revision
10. Visualization: preparation, creation and/or presentation of the published work, specifically data presentation or visualisation
11. Supervision: oversight and leadership responsibilities, including mentorship
12. Project administration: coordination of the research activity planning and execution