# Journal Pre-proof

Quantifying *APOL1,* Human Leukocyte Antigen, and Other Genetic Contributions to Unexplained Kidney Failure

Omid Sadeghi-Alavijeh, MB, PhD, Melanie M.Y. Chan, MB, PhD, Konstantinos Tzoumkas, MSc, Gabriel T. Doctor, MB, Daniel P. Gale, MB, PhD

PII: S0272-6386(25)01016-9

DOI: https://doi.org/10.1053/j.ajkd.2025.07.005

Reference: YAJKD 58447

To appear in: *American Journal of Kidney Diseases*

Received Date: 23 October 2024

Revised Date: 10 July 2025

Accepted Date: 18 July 2025

Please cite this article as: Sadeghi-Alavijeh O, Chan MM, Tzoumkas K, Doctor GT, Gale DP, Quantifying *APOL1,* Human Leukocyte Antigen, and Other Genetic Contributions to Unexplained Kidney Failure, *American Journal of Kidney Diseases* (2025), doi: https://doi.org/10.1053/j.ajkd.2025.07.005.

# Quantifying *APOL1,* Human Leukocyte Antigen, and Other Genetic Contributions to Unexplained Kidney Failure

Omid Sadeghi-Alavijeh MB, PhD[1*], Melanie MY Chan MB, PhD[11,2*], Konstantinos Tzoumkas MSc[1], Gabriel T Doctor MB[1], Daniel P Gale MB, PhD[1†].

[1] Centre for Genetics and Genomics, UCL Department of Renal Medicine, UCL Medical School, Rowland Hill Street, London NW3 2PF

[2]Medical Research Council Laboratory of Medical Sciences, Hammersmith Campus, Du Cane Road, Imperial College London, W12 0NN.

[†] Corresponding author: Daniel P. Gale, Centre for Genetics and Genomics, UCL Department of Renal Medicine, UCL Medical School, Rowland Hill Street, London NW3 2PF, d.gale@ucl.ac.uk

* OS-A and MC contributed equally to this work

**Abstract**

**Rationale & Objective:** Unexplained kidney failure (uKF) affects 15% of individuals needing kidney replacement therapy. The lack of a clear diagnosis creates uncertainty about recurrence, familial risk, and trial eligibility. This study sought to identify genetic variants underlying uKF.

**Study Design:** Genomic analyses were conducted using whole genome sequencing (WGS) that were reviewed by a multidisciplinary team who identified candidate pathogenic variants. A case-control study was implemented for single and structural variants to perform gene-based and polygenic risk score association testing.

**Setting & Participants:** The study recruited 218 patients with uKF onset before age 50 from the UK's 100,000 Genomes Project. Association analysis was performed in 180 uKF cases who remained unsolved after clinical analysis and constituted the non-monogenic uKF cohort (NM-uKF). 26,373 controls were derived from the unaffected relatives of non-renal probands.

**Exposures:** Candidate variants in 537 genes were assessed at a structural and single variant level in the 218 recruited patients as were high-risk *APOL1* genotypes and polygenic risk scores for chronic kidney disease and various glomerulonephritides.

**Outcome:** The primary outcomes were establishing a genetic diagnosis and the associations between genetic findings, age, family history, and ancestry.

**Analytical Approach:** Candidate variants were reviewed for pathogenicity. Gene-based and structural variant analyses and high-risk *APOL1* genotype assessments were performed. Polygenic risk scores and post-hoc HLA associations were also investigated.

**Results:** Monogenic diagnoses were made in 38 of 218 patients (17%) using WGS via the clinical arm of the 100,000 Genomes Project. Median uKF onset was 36 years. Diagnoses were less frequent in patients aged 36 or older, irrespective of family history. Three older patients

without a family history had pathogenic variants in type IV collagen genes. Among individuals with recent African ancestry, high-risk *APOL1* genotypes were significantly more common in those with uKF (52% vs. 8.4% in those without uKF, P<0.001). An elevated steroid sensitive nephrotic syndrome (SSNS) polygenic risk score was observed in those with high-risk *APOL1* genotypes and uKF, partly due to differences at *HLA-DQB1*03:19*.

**Limitations:** Potential limitations include the small sizes of subgroups and use of short-read WGS.

**Conclusions:** WGS yielded a monogenic diagnosis in 17% of patients with uKF, with no additional solved cases arising from the case-control analysis. These findings underscore *APOL1*'s role in those with recent African ancestry and suggest a genetic architecture distinct from common chronic kidney disease.

Keywords: whole-genome sequencing, kidney failure, unexplained kidney failure, CKD, renal replacement therapy, *APOL1*, genetics, genomics, HLA

**Plain-Language Summary**

Our study was motivated by the difficulty in understanding why some people develop kidney failure without a clear cause. Many patients face uncertainty about their future health and treatment because doctors cannot always find an explanation for their condition. To address this challenge, we examined the complete genetic blueprint of individuals with unexplained kidney failure. We looked for genetic clues that might reveal hidden risks. Our work uncovered specific genetic factors that appear to contribute to kidney failure, especially among people with African heritage. These insights are important because they may help explain why kidney failure happens in some cases and inform personalized diagnosis and treatment.

## Introduction

Chronic kidney disease (CKD) is often asymptomatic until damage to the kidneys is advanced. Consequently, patients (especially adults) frequently present with non-specific glomerulosclerosis and tubulointerstitial fibrosis on kidney biopsy which is uninformative as to the underlying cause of their disease. As a result, around 15-20% of adults who require chronic dialysis or kidney transplantation (kidney replacement therapy, KRT) have kidney failure that is unexplained (uKF)[1]. Rare diseases, such as monogenic kidney diseases, make a contribution to KF that is disproportionate to their frequency among patients with earlier stages of CKD and may account for a significant proportion of uKF[2].

Over 450 genes have been identified that cause monogenic kidney disease with recent technological advances making gene panels, whole exome (WES), or whole genome sequencing (WGS), available to large numbers of patients[3]. These technologies have shown that monogenic causes make a substantial contribution to the overall burden of kidney disease, identifying a molecular diagnosis in 9.3% of over 3300 patients with kidney disease/failure including 17% of those with nephropathy of unknown origin[4]. Exome and genome sequencing carried out in smaller cohorts of uKF report a diagnostic yield between 12-47% depending on the population studied [5–9].

Identifying the molecular cause of kidney disease in a patient can confirm the diagnosis, inform prognosis, enable predictive testing of family members, and facilitate transplant and reproductive decisions. Efforts to broaden availability of genomic testing are underway in many countries across the world. However, interpretation of genetic variants in known monogenic kidney disease genes can be challenging because of the frequent occurrence of rare (non – disease-causing) variants in individuals and so a high level of certainty is needed that the variant

identified is responsible for a patient's disease in order for it to be reported back to them. Variant interpretation is therefore a significant obstacle to delivery of medically actionable information from genetic tests. The overall likelihood that a medical test result is interpreted correctly is heavily dependent on the prior probability of that outcome, so information that allows prior probability to be estimated accurately is enormously valuable, assisting both the interpretation of candidate variants and policy decisions on which patient groups (as defined by clinical features) are most likely to benefit from genetic testing.

Here we use clinically accredited WGS data from individuals with uKF recruited to the UK's 100,000 Genomes Project to examine which clinical features are associated with an underlying monogenic disorder. In addition, we combine rare variant studies with analyses of the burden of known common genetic risk factors for different types of kidney disease (polygenic risk scores and APOL1-related kidney disease) to gain insight into the contribution of non-Mendelian diseases to uKF. This work has driven the adoption of uKF as an indication for WGS as part of England's National Health Service (NHS) Genomic Medicine Service[10], meaning any patient who meets the inclusion criteria can have WGS sent by their clinician.



## Methods

### The 100,000 Genomes Project

The Genomics England dataset (version 10) includes WGS data, clinical phenotypes encoded using Human Phenotype Ontology (HPO) codes, and NHS hospital records for 89,139 individuals recruited with cancer, rare disease, and their unaffected relatives. Patients were recruited by their treating clinician locally if they met the nationally agreed recruitment criteria. Ethical approval for the 100KGP was granted by the Research Ethics Committee for East of England – Cambridge South (REC Ref: 14/EE/1112). Informed consent was obtained at each

recruiting centre using the 100,000 Genomes Project centralised consent form. The recruitment window was between 2015-2018.

**Unexplained kidney failure in young people**

Unrelated probands recruited to the 'Unexplained kidney failure in young people' cohort in the 100,000 Genomes Project (100KGP) were identified. These patients were recruited by their local clinicians who applied the inclusion and exclusion criteria locally before recruiting patients into the 100KGP study. These are individuals who required KRT before the age of 50 years in the absence of an identified cause for their KF. Prior to recruitment, any patients with a personal or family history of gout before the age of 30 in the absence of CKD stage 3, 4 or 5, or diabetes, were expected to be tested for *UMOD* and *HNF1B* variants, respectively.

The exclusion criteria were: likely or proven diabetic nephropathy; likely or proven renovascular disease; identified glomerular disorder on kidney biopsy (other than glomerulocystic disease, ischaemic changes or secondary glomerulosclerosis); evidence of autoimmune, infectious, malignant, metabolic or other systemic disorder likely to be responsible for kidney disease; renal sarcoidosis or tuberculosis; paraproteinaemia (unless kidney biopsy shows no evidence of renal monoclonal deposition); exposure to nephrotoxin suspected of causing renal dysfunction; obstructive uropathy; significant proteinuria (>1g/day; uPCR >100 mg/mmol) at presentation unless presentation was with kidney failure; identified tubular/electrolyte/acid base disorder; >5 kidney cysts (not attributable to acquired cystic disease of CKD); nephrolithiasis; structural kidney and urinary tract malformations.

DNA preparation and extraction, WGS, alignment and variant calling is described in detail in Item S1 and S2.

**Identification of pathogenic and likely pathogenic variants**

An expert-curated virtual gene panel of 183 (see Table S1 for the full gene list) known kidney disease genes(https://panelapp.genomicsengland.co.uk/panels/678/ - version 3.0) was applied to the WGS data of patients. This panel was created using panelapp[11], a publicly available, expert curated platform that iteratively analyses publicly available datasets to provide clinical grade gene panels. This results in panels with three levels of confidence of pathogenicity assigned by a traffic light system, green genes being those that can be used for clinical grade genome interpretation.

Candidate variants were assessed against American College of Medical Genetics and Genomics (ACMG) criteria to identify pathogenic or likely pathogenic variants[12]. These well-defined criteria integrate variant information including population frequency, computational predictions of deleteriousness, functional domain localisation, putative mechanism of disease and previous associations with phenotypes in reputable databases.

For the majority of patients, variants were discussed in a multidisciplinary team including clinical and molecular geneticists and the recruiting clinician. Outcomes were recorded as "solved," "partially solved" or "not solved" by the local clinical multidisciplinary team based on their consensus after assessment of the variant(s) identified in each individual against the ACMG/AMP criteria and outcomes were collected in an *Exit Questionnaire* returned to Genomics England. This dataset was queried for patients from the uKF cohort and those taken as "solved" with respect to their kidney disease were placed into the solved category.

"Not solved" patients were redefined as "non-monogenic" uKF patients (NM-uKF). Given the initial assessment by Genomics England in 2019 used version 3 of the uKF panel comprising 183 genes we reanalysed the NM-uKF cohorts' WGS data looking for high impact variants (e.g. likely loss-of function) and de novo moderate impact variants (e.g. missense) using an updated

panel of 537 genes (in total) from version 12.8 of the uKF panel (Table S2 and https://panelapp.genomicsengland.co.uk/panels/678/).

**Case control creation for association analyses**

Taking the uKF cohort we created an ancestrally matched control cohort to conduct downstream analyses. To generate the controls, we started with the unaffected relatives of probands recruited to the 100KGP with non-renal conditions. We then further depleted the cohort by removing those with any HPO or Hospital Episode Statistics (HES) pertaining to kidney disease. We then ensured there was no relatedness within or between cases and controls. The controls were then ancestry matched to the cases to generate a final cohort of 218 cases and 26,373 controls (Figure S1). Full details of the cohort creation and collapsing analyses can be found in Item S3-S5. This cohort was used for the *HLA*, *APOL1,* polygenic risk score and collapsing variant analyses.

In the NM-uKF patients (N=180) we attempted to ascertain whether there were missed genetic drivers at a population level by performing collapsing single and structural variant analyses. For this we used the same cohort creation method as above to generate a cohort of 180 cases and 26,373 ancestrally matched, unrelated controls. Full details of the cohort creation can be found in Item S1-S3.

**APOL1 ascertainment**

Using genetically predicted ancestries calculated by the 100KGP, we divided the uKF cohort into those predicted with greater than 90% confidence to be of African ancestry (n=27) and those not (n=191) and created a subset of the controls who were of African ancestry without uKF (n=608). The *APOL1* G1 (S342G and I384M) and G2 (del388N389Y) renal risk variants were bioinformatically ascertained from WGS data for the cases and controls. Patients with a G2 allele

who had the protective p.N264K missense variant [13], were reassigned as non-high risk *APOL1* genotype. The 169 people without a monogenic diagnosis and without a high risk *APOL1* genotype were labelled as "unsolved" unexplained kidney failure (unsolved-uKF) to differentiate them from the NM-uKF cohort of 180 non-monogenic patients.

**HLA ascertainment in the high risk APOL1 genotype cohort in 100KGP and UK Biobank**

Using HIBAG[14], HLA types were imputed at two field resolution for *HLA-A, HLA-C, HLA-B, HLA-DRB1, HLA-DQA1, HLA-DQB1*, and *HLA-DPB1* centrally by Genomics England[15]. These were then extracted for the cases and controls with high risk *APOL1* genotypes.

To increase numbers to detect whether HLA allotype modified *APOL1* related risk, we sought additional cases and controls from the UK Biobank (UKBB)[16]. Using WES data in UKBB we identified individuals carrying high-risk *APOL1* variants (G1/G1, G1/G2 and G2/G2). Cases (n=7) were defined as people with CKD stage 4, CKD stage 5 or requiring KRT before the age of 50, without known causes for CKD based on hospital inpatient diagnoses and operative and procedural records as taken from ICD-10 and OPCS [v4] codes. Controls (n=417) were people with the same age-range as cases without hypertension or CKD related codes. Imputed HLA allotypes (performed by UKBB using HLA*IMP:02[17]) were extracted for all individuals and the top two alleles of each HLA type were assigned based on the highest imputation probabilities. The UKBB and the 100KGP cohorts were then combined to create a case (n=21) control (n=468) cohort of patients of recent African ancestry with high risk *APOL*1 genotypes and their imputed *HLA* types.

To address confounding by population stratification, we generated principal components of the common variant genotype matrix of the cohort, to include as covariates. A list of high quality, biallelic, LD- and complex-region-pruned single nucleotide variants with a minor allele

frequency > 0.05, defined in the Genomics England dataset (see data availability) for ancestry estimation, were extracted. These were intersected with UKBB imputed genotypes (imputed by UKBB against TOPMed R2 panel[18]). Principal components analysis was performed on 63,523 variants using PLINK (version 1.9)[19]. These were used in the downstream analysis. Population statistics for specific HLA types were taken from the allele frequency net database[20].

**Polygenic risk scoring**

Four polygenic risk scores (PRS) were applied to the uKF cohort and ancestry matched controls. A multi-ancestry IgA nephropathy score encompassing 77 variants[21]; a European membranous nephropathy score encompassing 12 variants[22]; a European steroid sensitive nephrotic syndrome (SSNS) score encompassing 5 variants[23] and a multi-ancestry CKD score encompassing 471,316 variants[24]. The scores were lifted over from genome build 37 to 38 using the UCSC liftover tool[25] and applied to the cohorts using the "score" command in PLINK2[19]. PRS scores were standardized to controls using Z-score scaling.

The uKF cases were divided into those with high-risk *APOL1* genotypes, those with a monogenic diagnosis delivered by the clinical arm of 100KGP and those patients who were unsolved (labelled as unsolved-uKF). The three patients with both a monogenic diagnosis and a high risk *APOL1* genotype were reviewed and for this analysis classified as high risk *APOL1* only for the purposes of statistical analysis.

**Bioinformatic analysis of the non-monogenic uKF cohort**

In the 180 NM-uKF patients and 26,373 matched controls we attempted to ascertain whether there were missed genetic drivers by performing genome-wide rare variant collapsing analysis using SAIGE-GENE[26], collapsing variants across a number of "masks" or filters. The masks used for this analysis were a rare, damaging missense mask ("missense+"), a high confidence

loss of function mask ("LoF"), an intronic mask ("intronic"), a splice site mask, a 3-prime untranslated region mask (3'-UTR) and a 5-prime UTR mask (5'-UTR). We then applied an exome-wide structural variant analysis pipeline, using calls made by MANTA[27] and CANVAS[28] to look for structural and copy number variants that may account for uKF cases. We examined variants in all known coding genes.

Full details of the cohort creation and collapsing analyses can be found in Item S4-S5.

**Statistics**

***Clinical and demographic analysis of the whole uKF cohort who underwent WGS (n=218)***

Baseline characteristics are expressed as frequencies (n, %) and medians (IQR), as dictated by data type. A two-sided Fisher's exact test was used to compare clinical characteristics of those with and without a genetic diagnosis. A Wilcoxon-Mann-Whitney test was used to compare age of KF between groups. Multivariable logistic regression was performed to identify clinical characteristics associated with a positive genetic result using sex, age at kidney failure, family history, extra-renal features and consanguinity as covariates. Differences between PRS scores in cases and controls were compared using multivariable logistic regression using sex, age and ten principal components as covariates. All statistical analysis was performed in R (version 3.6.2). Two-sided $p$ values $< 0.05$ were considered statistically significant.

***Case-control analyses***

The burden of high risk *APOL1* genotypes (G1/G1, G1/G2 and G2/G2) were compared across cohorts (case vs control, *APOL1* cases versus African ancestry controls with and without high risk *APOL1*) using a one-sided Fisher's exact test and this variable (the presence or absence of a high-risk genotype) was used as part of the logistic regression model alongside various polygenic risk scores, the first ten principal components, age and sex.

To test the significance between the PRS of the three case cohorts and controls we applied an ANOVA test followed by a Tukey's HSD test to differentiate the source of statistical significance in R [29].

The HLA types in the combined UKBB/100KGP cohort with high risk *APOL1* genotypes was analysed using a logistic regression model as implemented within the HIBAG tool using HLA-type and five principal components as covariates. A chi-squared test was used for significance testing using a Bonferroni corrected P-value of 0.007 (7 HLA classes analysed, $\alpha$=0.05).

Collapsing rare variant analysis was performed using SAIGE-GENE which uses logistic mixed modelling to look for genetic associations between uKF cases and controls. Ten principal components and age were used as covariates in this analysis. For collapsing structural variant analysis, the difference between the burden of variants in cases and controls was calculated with a two-sided Fisher's exact test.

All statistical analysis and plotting was performed with R[29].

**Results**

Figure 1 gives the full study workflow and key results which are elaborated on below.

**Demographic details and outcomes from the clinical WGS arm of the uKF analysis**

We analysed WGS data from 218 probands with uKF: 62% were male, 40% had an affected first-degree relative, 53% had extra-renal manifestations, and 6% had self-reported parental consanguinity. 97 of the 218 (44.4%) patients had undergone a kidney biopsy prior to recruitment to the study; in those patients where the biopsy data was available the results only indicated non-specific histopathological changes associated with chronic kidney damage. In the 89 patients where prior genetic testing data was available 5 (5.6%) had undergone prior panel

genetic testing (3 for tubulopathy and 2 for *MUC1*) – all of which were negative. Data on age at KF was available for 190/218 (87%) individuals with the median age at KF calculated as 36 years (IQR 16).

Using the initial virtual panel of 183 genes, 38/218 (17%) had a pathogenic or likely pathogenic variant affecting a known kidney disease gene, with 31/38 (82%) solved cases having the age at KF available (Figure 2, Table S3). Re-analysing those patients who were unsolved using the extended virtual panel of 537 genes did not yield any additional diagnoses. This cohort of patients where a monogenic variant was not found were labelled as the "non-monogenic" uKF cohort (NM-uKF).

Pathogenic copy number variants (CNVs) were seen in 2/218 (0.9%) individuals: A homozygous 110kb whole gene deletion of *NPHP1* was detected in a male with microscopic haematuria, proteinuria and gout who developed KF in his 30s (30-35). In addition, a 1.9Mb 17q12 duplication (encompassing *HNF1B*) was detected in a female with small kidneys who developed KF in her 30s (30-35). This duplication segregated with her affected mother.

**Clinical predictors of a positive genetic diagnosis in the whole WGS uKF cohort**

We next investigated whether there were any differences between patients with and without a genetic diagnosis to ascertain who might benefit most from genetic testing (Table 1). Clinical and demographic features were similar between the two groups and although median age at kidney failure was lower in those with a genetic diagnosis, this was not statistically significant (Table 1). Multivariable logistic regression did not identify any significant predictors of a positive genetic diagnosis in this relatively small cohort (Table S4 and Figure S2) and there was no association seen with specific extra-renal HPO terms (eye, ear, autoimmune, haematological, diabetes, endocrine, neurological, dermatological, gout, congenital malformations).

The diagnostic yield in those with and without a family history was compared stratifying by age at KF (Table 2). In those with a family history, significantly fewer genetic diagnoses were made in those ≥ 36 years compared to those < 36 years (35% vs 11%, P=0.03) with the diagnostic yield in individuals ≥ 36 years without a family history just 5% (Figure S3).

Three individuals with no family history who developed KF ≥36 years old received a genetic diagnosis. Interestingly, all three were heterozygous for type IV collagen variants: a likely pathogenic missense C*OL4A3*:c.3760G>C;p.(Gly1254Arg) variant, previously reported in a patient with Alport syndrome, was detected in a female with diabetes mellitus who developed KF in their late 30s (36-40); a pathogenic frameshift *COL4A4*:c.282_283del;p.(Asp96ProfsTer13) variant was identified in a male who developed KF in their 50s (50-55), who had microscopic haematuria, proteinuria, and gout; and a likely pathogenic *COL4A5*:c.367G>A;p.(Gly123Arg) variant was seen in a male with haematuria and proteinuria who reached KF in their 30s (36-40), consistent with a diagnosis of X-linked Alport syndrome. Although none of the individuals had a documented family history, identification of these type IV collagen variants initiated additional screening of family members.

We next looked for enrichment of rare variation in the NM-uKF cohort (n=180) compared to controls (n=26,373) using genome-wide collapsing rare variant analysis to look for novel candidate genes. However, this collapsing gene-based analysis did not reveal any significant genes at single nucleotide or structural variant level including collapsing variants that were in intronic, UTR or splicing domains (Figure S4).

**Case control analysis of APOL1 genotypes in people of African ancestry**

High-risk *APOL1* genotypes are observed in 13% of individuals with recent African ancestry and have been associated with an increased risk of kidney disease. We therefore hypothesized

that these renal risk alleles might contribute to uKF in this cohort. Within the cohort, 27 probands were of African ancestry. 16 had biallelic *APOL1 G1/G2* risk alleles but two uKF patients carried the protective p.N264K variant alongside the G2 allele, leaving 14 cases with high-risk *APOL1* genotype. High-risk *APOL1* genotypes were significantly enriched in individuals with uKF (14/27; 52%) when compared to controls of African ancestry (51/608; 8.4%; *P<0.001*). Three uKF cases with high-risk *APOL1* genotype had been given a monogenic diagnosis (*NPHP1*, *EVC, COL4A4*). There was no significant difference in the age at KF between the high-risk *APOL1* and low-risk *APOL1* uKF cases of African ancestry (median 37 vs 39 years; *P=0.7*).

**Polygenic risk scoring for CKD and glomerular disease in the case control cohort**

We next explored whether individuals with uKF had a genetic susceptibility to other non-monogenic kidney diseases applying PRSs for IgA nephropathy, SSNS, membranous nephropathy and CKD across the uKF cohort and controls (218 cases and 26,373 controls). A multivariable logistic regression model was built for uKF against each PRS as well as the presence of high-risk *APOL1* genotype, using population covariates (top ten principal components, age, and sex). Only high-risk *APOL1* genotype was strongly associated with uKF (Figure 3; *P<0.001*; OR=9.15; 95% CI 4.11-2.03) uKF cases had a lower IgA nephropathy PRS than controls (*P=0.04*; OR=0.87; 95%CI 0.76-0.99). There was no difference between uKF cases and controls when applying PRSs for SSNS (*P=0.05*), membranous nephropathy (*P=0.6*) or CKD (*P=0.6*).

**Examination of the interaction between uKF, APOL1 genotypes and polygenic risk scores in defined subgroups**

The uKF cohort was divided into people without a monogenic diagnosis and without a high risk *APOL1* genotype (unsolved-uKF, n=169), people with high risk *APOL1* genotypes (n=14),

and those with a monogenic diagnosis (n=35). Of note the 3 patients with a monogenic diagnosis and high risk *APOL1* were counted as being in the *APOL1* group for this analysis hence the drop in the solved cohort number from 38 to 35. The SSNS PRS was significantly higher in the cohort with high-risk *APOL1* genotype without a monogenic diagnosis (*P<0.001*) (Figure 4).There was no difference across the three cohorts in the remaining glomerular or CKD PRSs. The group of 3 individuals with both a monogenic diagnosis and *APOL1* risk genotype was not analysed separately owing to its small size and included within in the *APOL1* cohort.

The SSNS PRS remained significantly elevated when comparing the high-risk *APOL1* genotype uKF cohort (n=14) and African ancestry controls with (*P*=0.05, n=51) and without (*P*=0.03, n=557) high-risk *APOL1* genotypes.

**Analysis of the interaction between uKF, high risk APOL1 genotypes and HLA across the 100KGP and UK Biobank**

Three of the five loci contributing to the SSNS PRS are in *HLA-DQB1*. To determine whether an HLA-risk allotype could be identified, we conducted a joint analysis of imputed HLA types across the UKBB and 100KGP in 21 (14 from 100KGP and 7 from the UKBB) individuals with early-onset KF (< 50 years) and 468 controls, all with high-risk *APOL1* genotype. This revealed significant enrichment of *HLA-DQB1*03:19* in the early-onset KF group (*P*=0.001; OR=23.62; 95%CI 3.09-180.38, Table S5).

Using the allele frequency net database[20] we ascertained that *HLA-DQB1**03:19 has a high frequency in Gambia (28%) with the next highest population frequency seen in Tanzania (5.1%)[20]. A full list of population allele frequencies can be found in Table 3.

**Discussion**

The monogenic diagnosis rate of 17% we observed using WGS in 218 patients with young-

onset uKF is similar to the 17.1% reported using exome sequencing in the largest cohort investigated to date (n=281)[4]. Diagnostic rates of 12-47% have been reported in other exome-based studies focusing on patients with uKF and/or CKD[5,7,8] and a recent WGS study in 100 individuals with CKD5 and median age at KF of 32 years reported a diagnostic yield of 25%[30]. Each study has differing proportions of familial disease and extra-renal features and used different gene panels, likely explaining some of this variability. WGS enables ascertainment of almost all types of genomic variation in both coding and non-coding regions and provides more uniform coverage across the genome offering enhanced detection of single-nucleotide variants, structural variants and those found in homologous pseudogenes, as in the case of *PKD1*[24,31]. Our results suggest prioritizing genetic testing for those with uKF occurring before the age of 36 years and/or a family history of kidney disease, and this is now reflected in England's NHS Genomic Medicine Service eligibility criteria recommending WGS for patients with uKF under the age of 36 years.

Beyond diagnostic yield, WGS also enables analysis of other types of genetic contributors to risk, including application of PRSs and modelling these alongside *APOL1* genotype and monogenic causes of KF. A PRS for CKD was not elevated in uKF cases, highlighting differences in genetic architecture in those who present with young-onset KF versus a general CKD population with progressive decline in GFR [27,36]. Further studies that leverage rare variants in the PRS and are disease specific may yield better insights into underlying genetic architecture of these diseases.

The glomerular disease PRS results should be interpreted within the context of how they were originally derived. For example, the IgA nephropathy PRS was developed in patients who are likely to have presented before CKD stage 5. This suggests that undiagnosed IgA

nephropathy is unlikely to explain a large proportion of this unsolved uKF cohort. Membranous nephropathy has a median age at diagnosis of 56.4 years and typically presents with nephrotic syndrome and preserved kidney function[32]. It is therefore unsurprising that the PRS for membranous nephropathy was not elevated in this uKF cohort in which the upper age limit for KF was 50 years.

The *APOL1*-associated uKF patients did show enrichment for the SSNS PRS. Although derived from a genome-wide association study in childhood SSNS[33], this PRS is also elevated in individuals with gene test negative steroid resistant nephrotic syndrome and adult-onset nephrotic syndrome and minimal change disease, suggesting that the score identifies genetic predisposition to autoimmune podocytopathy[14]. One possible explanation for the enrichment in individuals with *APOL1*-associated uKF is that the combination of *APOL1* high-risk genotype with autoimmune podocytopathy is a particularly strong risk factor for developing KF. However, none of the individuals included in the uKF cohort had a preceding diagnosis of nephrotic syndrome (or indeed proteinuric kidney disease) so data to support this hypothesis is lacking at present[33]. A potential role for autoimmune podocytopathy is further supported by our finding of a strong HLA association at *DQB1* in those patients with KF and *APOL1* high-risk genotypes across both the UKBB and 100KGP cohorts. This HLA allele is seen at high frequency in West African populations and its enrichment in those with KF may result from population differences that were not adequately corrected by the genomic control measures we implemented. Interestingly, while previous studies have shown associations between SSNS and HLA-*DQA1*[34] and *DQB1*[39], this is the first time enrichment in these HLA alleles has been reported in individuals with a high-risk *APOL1* genotype and KF.

Our study is not without limitations. While we were able to study common and rare variants

in a relatively large cohort of patients with uKF the number of individuals is too small to detect weaker genetic effects typically observed with PRS, especially when subgroup analyses were performed. Therefore, the lack of PRS enrichment for CKD, IgA and membranous nephropathy does not exclude the possibility that these causes of kidney disease could be responsible for KF in a proportion of those presenting with uKF. In addition, we recognise that there are limitations to short-read WGS testing in both the detection of complex structural variants and the identification of variants in highly repetitive GC-rich regions, such as in the case of ADTKD-*MUC1,* and the use of long-read sequencing technologies in the future may improve the diagnostic yield in this patient group.

Our findings corroborate that monogenic causes account for a sizeable share of young-onset uKF and reveal an over-representation of high-risk *APOL1* genotypes in UK patients of recent African ancestry. These high-risk *APOL1* individuals carry both an elevated steroid-sensitive nephrotic-syndrome polygenic risk score and a marked enrichment of the class-II allele HLA-DQB1*03:19, suggesting an additional adaptive-immune 'hit'. Taken together, the data indicate that convergent innate (*APOL1*) and adaptive (HLA-restricted) podocyte-injury pathways can coexist within the same patient and ultimately present as clinically 'unexplained' kidney failure.

**Supplementary Material**

Supplementary File 1 (PDF)

Figure S1. Ancestry Matching

Figure S2. Multivariable logistic regression of factors associated with a positive genetic diagnosis

Figure S3. Number of patients, binned by age at kidney failure stratified according to whether they received a genetic diagnosis ('solved') or not ('unsolved').

Figure S4. Aggregated Manhattan Plot of multiple collapsing gene-based tests

Item S1. DNA preparation, and whole genome sequencing

Item S2. Genomic variant calling format file annotation and variant-level quality control

Item S3. Creation of ancestry matched unsolved cohort

Item S4. Rare variant selection for collapsing analysis.

Item S5. Structural and copy number variant analysis.

Supplementary File 2 (XLSX)
Table S1. Genes searched
Table S2. Expanded gene panel
Table S3. Solved cases
Table S4. Logistic regression analysis
Table S5. HLA associations

**Article Information**

**Authors' Contributions:** Study concept: DPG; analysis: OSA, MMYC, GTD, KT. Each author contributed important intellectual content during manuscript drafting or revision and agrees to be personally accountable for the individual's own contributions and to ensure that questions pertaining to the accuracy or integrity of any portion of the work, even one in which the author was not directly involved, are appropriately investigated and resolved, including with documentation in the literature if appropriate

**Support:** OSA was funded by an MRC Clinical Research Training Fellowship (MR/S021329/1). MMYC was funded by a Kidney Research UK Clinical Research Fellowship (TF_004_20161125). DPG is supported by the St Peter's Trust for Kidney, Bladder and Prostate Research. The funders were not involved in the study design, analysis, or the decision to submit for publication.

**Financial Disclosure:** The authors declare that they have no relevant financial interests.

**Acknowledgments:** This research was made possible through access to the data and findings generated by the 100,000 Genomes Project. The 100,000 Genomes Project is managed by Genomics England Limited (a wholly owned company of the Department of Health and Social Care). The 100,000 Genomes Project is funded by the National Institute for Health Research and NHS England. The Wellcome Trust, Cancer Research UK and the Medical Research Council have also funded research infrastructure. The 100,000 Genomes Project uses data provided by patients and collected by the National Health Service as part of their care and support. The authors gratefully acknowledge the participation of the patients, and their families recruited to the 100,000 Genomes Project.

**Data Sharing:** Details of the aggregated dataset used for the analysis can be found at: https://re-docs.genomicsengland.co.uk/aggv2/. The principal components were derived using the common, high quality independent SNPs derived by Genomics England: https://re-docs.genomicsengland.co.uk/principal_components/. Genomic and phenotype data from participants recruited to the 100,000 Genomes Project can be accessed by application to Genomics England Ltd (https://www.genomicsengland.co.uk/about-gecip/joining-research-community/). The SSNS polygenic risk score can be found at: https://www.pgscatalog.org/score/PGS003354/ The membranous nephropathy polygenic risk score can be found at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10116192/. The IgA polygenic risk score can be found at: https://www.columbiamedicine.org/divisions/kiryluk/resources.php. The CKD polygenic risk score can be found at: https://www.columbiamedicine.org/divisions/kiryluk/study_GPS_CKD.php. The HLA allele frequency tables are taken from the Allele Frequency Net Database: http://www.allelefrequencies.net/. Code used for the analyses in this paper can be found at: https://github.com/oalavijeh/unexplained_kf_paper. Details of the rare variant workflow can be found at: https://re-docs.genomicsengland.co.uk/avt/. Details of the common variant GWAS

workflow can be found at: https://re-docs.genomicsengland.co.uk/gwas/.
**Peer Review:** Received October 23, 2024. Evaluated by 2 external peer reviewers, with direct editorial input from a Statistics/Methods Editor, an Associate Editor, and the Editor-in-Chief. Accepted in revised form July 18, 2025.

**References**

1. Kramer A, Pippias M, Noordzij M, et al. The European renal association - European dialysis and transplant association (ERA-EDTA) Registry Annual Report 2016: A summary. *Clin Kidney J*. 2019;12(5):702-720. doi:10.1093/ckj/sfz011
2. Wong K, Pitcher D, Braddon F, et al. Effects of rare kidney diseases on kidney failure: a longitudinal analysis of the UK National Registry of Rare Kidney Diseases (RaDaR) cohort. *Lancet*. Published online March 13, 2024. doi:10.1016/S0140-6736(23)02843-X
3. Connaughton DM, Hildebrandt F. Personalized medicine in chronic kidney disease by detection of monogenic mutations. *Nephrol Dial Transplant*. 2020;35(3):390-397. doi:10.1093/ndt/gfz028
4. Groopman EE, Marasa M, Cameron-Christie S, et al. Diagnostic Utility of Exome Sequencing for Kidney Disease. *N Engl J Med*. 2019;380(2):142-151. doi:10.1056/NEJMoa1806891
5. Lata S, Marasa M, Li Y, et al. Whole-exome sequencing in adults with chronic kidney disease: A pilot study. *Ann Intern Med*. 2018;168(2):100-109. doi:10.7326/M17-1319
6. Snoek R, van Jaarsveld RH, Nguyen TQ, et al. Genetics-first approach improves diagnostics of ESKD patients <50 years old. *Nephrol Dial Transplant*. 2022;37(2):349-357. doi:10.1093/ndt/gfaa363
7. Connaughton DM, Kennedy C, Shril S, et al. Monogenic causes of chronic kidney disease in adults. *Kidney Int*. 2019;95(4):914-928. doi:10.1016/j.kint.2018.10.031
8. Ottlewski I, Münch J, Wagner T, et al. Value of renal gene panel diagnostics in adults waiting for kidney transplantation due to undetermined end-stage renal disease. *Kidney Int*. 2019;96(1):222-230. doi:10.1016/j.kint.2019.01.038
9. Mallawaarachchi AC, Fowles L, Wardrop L, et al. Genomic Testing in Patients with Kidney Failure of an Unknown Cause: a National Australian Study. *Clin J Am Soc Nephrol*. Published online May 3, 2024. doi:10.2215/CJN.0000000000000464
10. Beal F, Forrester N, Watson E, et al. A targeted gene panel illuminates pathogenesis in young people with unexplained kidney failure. *J Nephrol*. 2024;37(5):1273-1284. doi:10.1007/s40620-024-01964-1
11. Martin AR, Williams E, Foulger RE, et al. PanelApp crowdsources expert knowledge to establish consensus diagnostic gene panels. *Nat Genet*. 2019;51(11):1560-1565. doi:10.1038/s41588-019-0528-2
12. Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*. 2015;17(5):405-424. doi:10.1038/gim.2015.30
13. Gupta Y, Friedman DJ, McNulty MT, et al. Strong protective effect of the APOL1 p.N264K variant against G2-associated focal segmental glomerulosclerosis and kidney disease. *Nat Commun*. 2023;14(1):7836. doi:10.1038/s41467-023-43020-9
14. Zheng X, Shen J, Cox C, et al. HIBAG--HLA genotype imputation with attribute bagging. *Pharmacogenomics J*. 2014;14(2):192-200. doi:10.1038/tpj.2013.18

15. Kousathanas A, Pairo-Castineira E, Rawlik K, et al. Whole-genome sequencing reveals host factors underlying critical COVID-19. *Nature*. 2022;607(7917):97-103. doi:10.1038/s41586-022-04576-6
16. Sudlow C, Gallacher J, Allen N, et al. UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLoS Med*. 2015;12(3):e1001779. doi:10.1371/JOURNAL.PMED.1001779
17. Dilthey A, Leslie S, Moutsianas L, et al. Multi-population classical HLA type imputation. *PLoS Comput Biol*. 2013;9(2):e1002877. doi:10.1371/journal.pcbi.1002877
18. Taliun D, Harris DN, Kessler MD, et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *bioRxiv*. Published online March 6, 2019. doi:10.1101/563866
19. Chang CC, Chow CC, Tellier LCAM, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience*. 2015;4(1):7. doi:10.1186/S13742-015-0047-8/2707533
20. Gonzalez-Galarza FF, McCabe A, Santos EJMD, et al. Allele frequency net database (AFND) 2020 update: gold-standard data classification, open access genotype data and new query tools. *Nucleic Acids Res*. 2020;48(D1):D783-D788. doi:10.1093/nar/gkz1029
21. Kiryluk K, Sanchez-Rodriguez E, Zhou XJ, et al. Genome-wide association analyses define pathogenic signaling pathways and prioritize drug targets for IgA nephropathy. *Nat Genet*. 2023;55(7):1091-1105. doi:10.1038/s41588-023-01422-x
22. Gupta S, Downie ML, Cheshire C, et al. A Genetic Risk Score Distinguishes Different Types of Autoantibody-Mediated Membranous Nephropathy. *Glomerular Dis*. 2023;3(1):116-125. doi:10.1159/000529959
23. Downie ML, Gupta S, Chan MMY, et al. Shared genetic risk across different presentations of gene test-negative idiopathic nephrotic syndrome. *Pediatr Nephrol*. Published online November 10, 2022. doi:10.1007/s00467-022-05789-7
24. Khan A, Turchin MC, Patki A, et al. Genome-wide polygenic score to predict chronic kidney disease across ancestries. *Nat Med*. 2022;28(7):1412-1420. doi:10.1038/s41591-022-01869-1
25. Hinrichs AS, Karolchik D, Baertsch R, et al. The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res*. 2006;34(Database issue). doi:10.1093/NAR/GKJ144
26. Zhou W, Bi W, Zhao Z, et al. SAIGE-GENE+ improves the efficiency and accuracy of set-based rare variant association tests. *Nat Genet*. 2022;54(10):1466-1469. doi:10.1038/s41588-022-01178-w
27. Chen X, Schulz-Trieglaff O, Shaw R, et al. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics*. 2016;32(8):1220-1222. doi:10.1093/BIOINFORMATICS/BTV710
28. Roller E, Ivakhno S, Lee S, Royce T, Bioinformatics ST., 2016 U. Canvas: versatile and scalable detection of copy number variants. *academic.oup.com*. https://academic.oup.com/bioinformatics/article-abstract/32/15/2375/1743834
29. Wickham H. *Ggplot2*. Springer New York doi:10.1007/978-0-387-98141-3
30. Turro E, Astle WJ, Megy K, et al. Whole-genome sequencing of patients with rare diseases in a national health system. *Nature*. 2020;583(7814):96-102. doi:10.1038/s41586-020-2434-2
31. Vy HMT, Coca SG, Sawant A, et al. Genome-Wide Polygenic Risk Score for CKD in Individuals with APOL1 High-Risk Genotypes. *Clin J Am Soc Nephrol*. 2024;19(3):374-376. doi:10.2215/CJN.0000000000000379
32. Dufek-Kamperis S, Kleta R, Bockenhauer D, Gale D, Downie ML. Novel insights in the genetics of steroid-sensitive nephrotic syndrome in childhood. *Pediatr Nephrol*.

2021;36(8):2165-2175. doi:10.1007/s00467-020-04780-4
33. Adeyemo A, Esezobor C, Solarin A, et al. HLA-DQA1 and APOL1 as Risk Loci for Childhood-Onset Steroid-Sensitive and Steroid-Resistant Nephrotic Syndrome. *Am J Kidney Dis*. 2018;71(3):399-406. doi:10.1053/j.ajkd.2017.10.013
34. Barry A, McNulty MT, Jia X, et al. Multi-population genome-wide association study implicates immune and non-immune factors in pediatric steroid-sensitive nephrotic syndrome. *Nat Commun*. 2023;14(1):1-13. doi:10.1038/s41467-023-37985-w

**Table 1. Comparison of clinical and demographic features in patients with (solved) and without (unsolved) a genetic diagnosis.**

| | Solved (n=38) | Unsolved (n=180) | Total (n=218) | P |
|---|---|---|---|---|
| **Median age at KF (IQR)** | 31 (12) | 37 (16) | 36 (16) | 0.11 |
| **Male sex (%)** | 24 (63%) | 111 (62%) | 135 (62%) | 1 |
| **Family history (%)** | 18 (47%) | 69 (38%) | 87 (40%) | 0.36 |
| **Extra-renal features (%)** | 25 (66%) | 90 (50%) | 115 (53%) | 0.11 |
| **Consanguinity (%)** | 3 (8%) | 9 (5%) | 12 (6%) | 0.44 |

IQR, interquartile range; KF, kidney failure. P values calculated using a Mann Whitney Wilcoxon test (age at KF) or two-sided Fisher's exact test. Note that 31/38 solved cases and 159/180- unsolved cases had median age at KF available.

**Table 2. Diagnostic yield stratified by age at kidney failure and family history. Please note this table only represents the 190/218 patients where age at kidney failure data was available.**

| Family History | Age at KF < 36 years | Age at KF >=36 years |
|---|---|---|
| Yes | 9/26 (35%) | 5/45 (11%) |
| No | 12/64 (19%) | 3/55 (5%) |
| P | 0.17 | 0.46 |

KF, kidney failure. P values calculated using a two-sided Fisher's exact test.

**Table 3. Allele frequency of HLA DQB1*03:19 across the globe as taken from the Allele Frequency Net Database.**

| Population | | Allele Frequency (in decimals) | Sample Size |
|---|---|---|---|
| GMB | Gambia | 0.279 | 939 |
| TZA | Tanzania Maasai | 0.0507 | 336 |
| ARE | United Arab Emirates Abu Dhabi | 0.0385 | 52 |
| MEX | Mexico City Mestizo population | 0.014 | 143 |
| BRA | Brazil Paraná Caucasian | 0.0109 | 641 |
| MEX | Mexico City Mestizo population2 | 0.0107 | 234 |
| ESP | Spain | 0.009 | 4335 |
| MEX | Mexico Tixcacaltuyub Maya | 0.0075 | 67 |
| BRA | Brazil Puyanawa | 0.007 | 150 |
| SAU | Saudi Arabia pop 5 | 0.0063 | 158 |
| ZAF | South Africa Worcester | 0.006 | 159 |
| USA | USA San Diego | 0.004 | 496 |
| MYS | Malaysia Peninsular Chinese | 0.0026 | 194 |
| CZE | Czech Republic NMDR | 0.0019 | 5099 |
| PAN | Panama | 0.0016 | 462 |
| NLD | Netherlands Leiden | 0.001 | 1305 |
| ITA | Italy pop 5 | 0.001 | 975 |
| JPN | Japan pop 17 | 0.0003 | 3078 |
| IND | India North UCBB | 0.0001 | 5849 |
| IND | India West UCBB | 0.0001 | 5829 |
| MAR | Morocco Settat Chaouya | 0 | 98 |

NMDR - National Marrow Donors Registry, UCBB - University of Central Business and Biotechnology

**Figure 1 – Flow diagram of the analysis workflow and key findings.**
Green boxes donate clinical grade analyses; blue boxes donate research analyses and red boxes donate key results.

**Figure 2 - Breakdown of solved cases by gene category**
Breakdown of the solved uKF cases by gene and phenotypic category. All percentages represent a proportion of the total solved cases (n=38). ADTKD, autosomal dominant tubulointerstitial kidney disease; CAKUT, congenital anomalies of the kidneys and urinary tract.

**Figure 3 – Risks contributing to the chances of developing unexplained kidney failure**
Forest plot showing the outcome of the multivariable logistic regression model predicting the chance of developing unexplained kidney failure (uKF) using 218 cases and 26,373 controls when modelled against various polygenic risk scores, the absence or presence of a high risk APOL1 variant, age, sex and the first ten principal components. Results are presented as odds ratios (OR) with their 95% confidence intervals (CI). OR where the CI is above 1 indicate an increased risk of developing unexplained kidney failure, whereas a CI below 1 indicates a decreased risk of developing uKF. * Denotes statistical significance ($P < 0.05$)

**Figure 4 – Average polygenic risk score by cohort and PRS type**
Each point is the mean PRS with the bars representing the 95% confidence interval. 3 patients with both a monogenic diagnosis and a high risk APOL1 genotype were categorised as having APOL1 for the purposes of this analysis. Dotted line represents the normalized control PRS scores. * Denotes statistical significance. Unsolved-uKF – unsolved unexplained kidney failure, PRS – polygenic risk score, SSNS – steroid sensitive nephrotic syndrome, CKD – chronic kidney disease, IgA – Immunoglobulin A nephropathy