Title: Genotype-first analysis in an unselected health system-based population reveals variable phenotypic severity of *COL4A5* variants

Authors:

McKenzie Zellers[1], Kaushal Solanki[2], Kyle Retterer[3], Karyn M Murphy[3], Melissa A Kelly[4,5], H. Les Kirchner[6], Ion Dan Bucaloiu[2,6], Tooraj Mirshahi[5], Bryn Moore[5], Alexander R Chang[2,6]

[1]Geisinger Commonwealth School of Medicine, Scranton, PA
[2]Department of Nephrology, Geisinger, Danville, PA
[3]Geisinger, Danville, PA
[4]Autism and Developmental Medicine Institute, Geisinger, Danville, PA
[5]Department of Genomic Health, Geisinger, Danville, PA
[6]Department of Population Health Sciences, Geisinger, Danville, PA

Corresponding author:
Alexander R. Chang, MD, MS
Center for Kidney Health Research
Department of Nephrology
Geisinger Medical Center
100 North Academy Avenue,
Danville, PA 17822

Abstract word count: 300

Manuscript word count: ***

Abstract

Introduction: Our knowledge of X-linked Alport Syndrome [AS] comes mostly from selected cohorts with more severe disease.

Methods: We examined the phenotypic spectrum of X-linked AS in males and females with a genotype-based approach using data from the Geisinger MyCode DiscovEHR study, an unselected health system-based cohort with exome sequencing and electronic health record data. Patients with *COL4A5* variants reported as pathogenic (P) or likely pathogenic (LP) in ClinVar, or protein-truncating variants (PTVs), were each matched with up to 5 controls without *COL4A3/4/5* variants by sociodemographics, diabetes diagnosis, and year of first outpatient encounter. AS-related phenotypes included dipstick hematuria, bilateral sensorineural hearing loss (BSHL), proteinuria, decreased eGFR, and ESKD.

Results: Out of 170,856 patients, there were 30 hemizygous males (mean age 52.4 [SD 19.8] years) and 56 heterozygous females (mean age 58.5 [SD 19.4]) with a *COL4A5* P/LP variant, including 48 with the hypomorphic variant p.Gly624Asp. Overall, penetrance (having any AS phenotypic feature) was highest for non-p.Gly624Asp P/LP variants (males: 89%, females: 86%), intermediate for p.Gly624Asp (males: 77%, females: 69%), compared to controls (males: 32%; females: 50%). The proportion with ESKD was highest for males with P/LP variants (41%), intermediate for males with p.Gly624Asp (15%) and females with P/LP variants (10%), compared to controls (males: 3%, females 2%). Only 33% of males and 11% of females had a known diagnosis of Alport syndrome or thin basement membrane disease. Only 47% of individuals with *COL4A5* had completed albuminuria screening, and a minority were taking renin-angiotensin aldosterone system (RAAS) inhibitors.

Conclusion: In an unselected cohort, we show increased risks of AS-related phenotypes in men and women compared to matched controls, while showing a wider spectrum of severity than has been described previously and variability by genotype. Future studies are needed to determine whether early genetic diagnosis can improve outcomes in Alport Syndrome.

Keywords: Alport Syndrome; Chronic Kidney Disease; Clinical Epidemiology; Kidney Failure; Genetic Renal Disease, ESKD

Introduction:

Alport syndrome (AS) is the second most common genetic cause of end stage kidney disease behind polycystic kidney disease, affecting approximately one in 2320 individuals[1,2]. Nearly 85% of total cases of AS in the past have been due to variants in the *COL4A5* gene found on the X chromosome, with the rest of individuals with AS having variants in the *COL4A3* and *COL4A4* genes responsible for autosomal recessive and dominant inheritance patterns[1]. Individuals with AS can have phenotypic manifestations related to abnormalities in the basement membranes of the glomerulus, cochlea of the ear, and lens of the eye, resulting in microscopic hematuria, sensorineural hearing loss, and ocular disturbances. Each collagenous IV chain contains a collagenous domain made of Gly Xaa Yaa triplet repeats, resulting in a rigid protein structure[3]. Thus, glycine missense variants disrupting the collagenous domain as well as protein-truncating variants (PTVs) are likely to be deleterious. Some but not all studies suggest a worse prognosis for PTVs than missense variants[4–8]. One glycine missense variant, p.Gly624Asp, is adjacent to a non-collagenous region and known to have a more mild phenotype.[9]

Although females are more likely to carry *COL4A5* variants due to the nature of the X-linked inheritance pattern, males are more severely impacted, with reported risk of progression to ESKD to be nearly 100% in *COL4A5* males compared to 25% in similarly affected females[10,11]. An important limitation to the existing literature is that participants in these studies have been selected from families known to be affected by AS, likely leading to an overestimation of penetrance and phenotypic severity. As availability of genetic testing increases and costs of testing decrease, there is an urgent need to understand the penetrance and phenotypic spectrum in the general population to inform families about the implications of a genetic diagnosis of AS.

Using a genotype-first approach, we aimed to examine the phenotypic spectrum of X-linked AS due to *COL4A5* variants using an unselected health system-based population. We hypothesized that both males and females with *COL4A5* variants would have higher risk of kidney phenotypes though the penetrance and phenotypic severity of disease would be lower than previously reported in more selected cohorts.

Methods:

Study Population

There were 170,856 research participants in MyCode who had whole exome sequencing data available after quality control. These research participants are patients in the Geisinger Health System at large who consented to participate in the MyCode Community Health Initiative and were sequenced through the Geisinger-Regeneron DiscovEHR collaboration[12]. Patients with ClinVar P/LP or PTVs, as defined below, were classified as cases. We queried ClinVar [date accessed 4/24/2023] for *COL4A5* (NM_033380.3) variants previously classified as pathogenic (P) or likely pathogenic (LP) by >1 submitter and no conflicts, or conflicting variants with a ratio of ≥6 P/LP to other classifications[13]. Predicted PTVs were identified from those called as high impact variants by variant effect predictor v109. All variants underwent manual review to ensure technical confidence and clinical relevance.[14] Variants not meeting technical QC were orthogonally confirmed.

Phenotypic outcomes

Phenotypic data sources included International Classification of Diseases (ICD) 9 and 10 codes, blood and urine laboratory data, and chart review. ICD codes were used for hypertension, diabetes, hearing loss, bilateral sensorineural hearing loss, AS, focal segmental glomerulosclerosis (FSGS), and ESKD (**Supplemental Table 1**). We used the 2021 CKD-EPI equation to calculate eGFR, and we imputed eGFR of 5 ml/min/1.73m^2 for patients with kidney transplant.[15] Hematuria on dipstick was defined as having trace blood or greater on at least half of all urinalyses, excluding urinalyses positive for leukocyte esterase or nitrites.[5] Age at time of ESKD was determined using linkage to the United States Renal Data System,[16]

and if these data were not available, by chart review. Presence of albuminuria was defined as ever having albumin to creatinine ratio (ACR) ≥ 30 mg/g, and presence of proteinuria by dipstick was defined requiring at least 2 dipsticks. Any phenotypic feature of AS was defined as hematuria on urinalysis or ICD code, albuminuria, eGFR <60 ml/min/1.73m^2, FSGS ICD code, ESKD ICD code, or bilateral sensorineural hearing loss. Patients with most recent eGFR data at age ≥18 years were categorized into Kidney Disease Improving Global Outcomes (KDIGO) risk categories using estimated glomerular filtration rate (eGFR) and ACR (**Supplemental Table 2**).[5,17] If ACR was not available, urine dipstick protein data was used for KDIGO risk categorization with dipstick protein 1+ twice classified as ACR 30-299 mg/g, and dipstick protein 2+ or greater twice classified as ACR 300+ mg/g. Chart review was performed on all cases to further examine diagnosis of AS or TBMD, kidney biopsies, treatment with angiotensin-converting enzyme inhibitors (ACEIs) or angiotensin receptor blockers (ARBs), family history, and genetic testing.

Statistical analysis

Patients with *COL4A5* P/LP variants were matched to up to 5 controls (patients without *COL4A3*, *COL4A4*, and *COL4A5* variants at allele frequency<0.01), with exact matching on age (within 1 year) at time of most recent eGFR, sex, race, diabetes diagnosis, and year of first outpatient encounter. We compared characteristics between patients with *COL4A5* P/LP variants and matched controls using chi-square tests for categorical outcomes and paired t-tests for continuous outcomes overall, and by sex and variant group. We used Poisson regression models with robust variance as the primary analysis to estimate prevalence ratios to determine AS-related phenotypic outcomes, stratified by sex and genotype (P/LP variant or the known hypomorphic p.Gly624Asp variant), using each subgroups' matched controls. [18,19] We also conducted time-to-event analyses with ESKD as the outcome using Kaplan-Meier survival curves with age as the time variable. Age at time of first instance of ESKD, and for those who did not experience ESKD, we used age at the time of last eGFR. The proportional hazards assumption was graphically assessed, and Cox proportional hazards models were adjusted for age in sex-stratified analyses. As prior *COL4A5* studies have suggested genotype-phenotype correlations with the worst outcomes for nonsense and frameshift variants, intermediate outcomes for splice variants, and milder phenotypes for missense and in-frame

deletion variants,[6,20,21] we also estimated risk of ESKD using Cox models for nonsense/frameshift variants, splice variants, the hypomorphic p.Gly624Asp variant, and other missense or in-frame deletion variants.

Several sensitivity analyses were performed. Additional analyses using Firth logistic regression were done, including only one individual from each family. An analysis excluding 9 individuals with borderline quality control metrics (allele balance <0.30, depth <30, genotype quality <90) was also conducted. We considered p values <0.05 to be significant. Analyses were conducted using STATA/MP 15.1.

Results:

Out of the 170,856 patients, there were 30 hemizygous males (mean age 52.4 [SD 19.8] years) and 56 heterozygous females (mean age 58.5 [SD 19.4]) with a *COL4A5* P/LP variant. The most common *COL4A5* P/LP variant was p.Gly624Asp (n=48), and there were 18 other unique variants, including 4 other missense or in-frame deletion variants (n=16), 1 in-frame deletion variant (n=1), 7 splicing variants (n=14), 4 frameshift variants (n=4), and 2 stop-gained variants (n=3). Of the 86 individuals with *COL4A5* P/LP variants, the vast majority were matched to 5 controls (80 individuals matched to 5 controls, 2 to 3 controls, 2 to 2 controls, 2 to 1 control). Compared to matched controls, individuals with *COL4A5* P/LP variants were more likely to have urinalysis (95% vs 81%, p=0.001) and ACR (47% vs 26%, p<0.001) results available in addition to having a hypertension diagnosis (67% vs 50%, p=0.004) whereas 100% of patients in both groups had evaluation of eGFR (**Table 1**). The median (interquartile interval) year of first outpatient visit was 2004 [2001-2009] for individuals with *COL4A5* and matched controls.

<u>Prevalence of AS Phenotypes by Sex</u>

Overall, males with *COL4A5* P/LP variants had the highest risk of AS-related phenotypes, followed by males with the hypomorphic p.Gly624Asp variant and females with *COL4A5* P/LP variants, with lowest risk in females with p.Gly624Asp (**Table 2, Figure 1**). The p.Gly624Asp variant was

associated with a milder phenotype than other P/LP variants for both males and females. Males with p.Gly624Asp demonstrated lower penetrance (n=13; 0% AS ICD, 15% ESKD ICD, 8% bilateral sensorineural hearing loss ICD, 23% eGFR <30 ml/min/1.73m^2) compared to other hemizygous *COL4A5* P/LP males (n=17; 29% AS ICD, 41% ESKD ICD, 29% bilateral sensorineural hearing loss ICD, 41% eGFR <30 ml/min/1.73m^2). KDIGO risk categorization demonstrated that 33% of p.Gly624Asp males (mean age 56.9 y [SD 17.9] had a high-risk category or greater as compared to 75% of other hemizygous *COL4A5* P/LP males (mean age 49.9 y [SD 21.8]) (**Table 3, Figure 1**). Similarly, the prevalence of high-risk KDIGO category or greater in p.Gly624Asp females (mean age 60.3 y [SD 14.8] was 24% compared to 62% of other P/LP females (mean age 55.6 y [SD 25.5]).

All groups had significantly higher risk of AS phenotypic features compared to matched controls with prevalence ratios (PR) of 2.88 (95% CI: 2.00, 4.16) for males with P/LP variants, 2.20 (95% CI: 1.39, 3.48) for males with p.Gly624Asp, 1.75 (95% CI: 1.34, 2.29) for females with P/LP variants, and 1.35 (95% CI: 1.03, 1.76) for females with p.Gly624Asp (**Table 3**, **Figure 2)**. Risks of the most common AS manifestation, dipstick hematuria, were similar for males with P/LP variants (PR 2.36, 95% CI: 1.07, 5.23), males with p.Gly624Asp (PR 3.34, 95% CI: 1.80, 6.21), females with P/LP variants (PR 2.55, 95% CI: 1.74, 3.74), and females with p.Gly624Asp (PR 1.86, 95% CI: 1.28, 2.70). Males and females with p.Gly624Asp tended to have milder phenotypic presentation than their counterparts with other P/LP variants. For example, prevalence ratios for ESKD were 35.00 (95% CI: 4.55, 269.06) for males with P/LP variants, 3.08 (95% CI: 0.56, 16.80) for males with p.Gly624Asp, 9.33 (95% CI: 0.88, 99.21) for females with P/LP variants, and 1.61 (95% CI: 0.17, 15.07) for females with p.Gly624Asp. Males with P/LP and p.Gly624Asp variants had significantly increased risk of hearing loss ICD diagnosis whereas females with P/LP variants or p.Gly624Asp did not have significantly increased risk of hearing loss ICD. Sensitivity analyses excluding 9 patients with borderline quality control metrics were overall consistent, as were analyses restricted to one individual per family (**Supplemental Tables 3, 4**). Variant-level phenotypic characteristics, stratified by sex, are shown in **Supplemental Table 5**.

Time-to-event Analyses for ESKD by Genotype (nonsense/frameshift variants, missense or in-frame indels, splice variants, hypomorphic p.Gly624Asp)

Kaplan-Meier survival curves for ESKD-free survival followed similar patterns (**Figure 3**). Overall, there was no difference in deaths in patients with *COL4A5* variants and controls (7% vs. 7%; p=0.95) with mean age of death of *COL4A5* patients 70.8 years (range 54.3-80.7). For males, all variant groups were associated with increased risk of ESKD with lowest risk for hypomorphic p.Gly624Asp (HR 5.61, 95% CI: 1.02, 30.78; p=0.047). Risk of ESKD by variant groups was numerically highest for nonsense/frameshift variants (HR 74.02, 95% CI: 3.54, 1549.74; p=0.006), then missense or in-frame variants (HR 23.33, 95% CI: 5.76, 94.44; p<0.001), and splice variants (HR 14.23, 95% CI: 2.38, 85.19; p=0.004) though 95% intervals were wide and overlapped between variant groups (**Table 4)**. For females, missense or in-frame variants and splice acceptor/donors (HR 9.84, 95% CI: 1.06, 91.20; p=0.048) and splice variants (HR 11.75, 95% CI: 1.18, 116.94; p=0.04) were associated with ESKD but not hypomorphic p.Gly624Asp variants (HR 2.07, 95% CI: 0.23, 18.60; p=0.5); 0/5 of females with nonsense/frameshift had ESKD.

Mean age of ESKD onset in males was 36.9 y (SD 15.0) males with P/LP variants and 50.0 (SD 5.7) for males with Gly624Asp. Mean age of ESKD onset was 51.7 (SD 12.6) for females with no difference between those with P/LP or Gly624Asp variants.

Individuals with a *COL4A5* P/LP variant and a second *COL4A3/4/5* variant

There were 15 individuals who had an additional rare (AF<0.01) variant in *COL4A3/4/5* with 4 (27%) having ESKD (**Supplemental Table 6)**. The individuals with ESKD included a male with *COL4A5* splice variant c.438+2T>G and *COL4A4* p.Pro352Leu; a male with *COL4A5* p.Leu1655Arg and *COL4A3* p.Val950Ile variant; a male with *COL4A5* p.Leu1655Arg and *COL4A3* p.Thr1489Ile, and a female with *COL4A5* c.4964T>G and *COL4A3* p.Ile1659Val.

Diagnosis of AS

AS or thin basement membrane disease was documented in only 10/17 (59%) of males with P/LP variants, 0/13 of males with Gly624Asp, 4/21 (19%) of females with P/LP variants, and 2/35 (6%) of females with Gly624Asp. A total of 18 individuals (21%) had documentation of a native kidney biopsy performed, and none had undergone a skin biopsy to diagnose AS. Of those with native kidney biopsies, 10 (56%) had a biopsy diagnosis of AS, hereditary nephritis or thin basement membrane disease; FSGS was reported in 6 (33%) biopsies (**Table 5**). One patient had 2 biopsies as a child that were inconclusive (1st biopsy: insufficient sample, thin basement membrane disease; 2nd biopsy: IgM nephropathy, FSGS) and has yet to be diagnosed with AS. Only 2 (11%) had specific mention of staining for the alpha5 chain of type IV collagen. In terms of genetic diagnosis, only 5 (6%) received genetic testing and 2 (2%) additional patients reported having a family member who had genetic testing confirming diagnosis.

Receipt of Recommended Care for AS

Only 40 (47%) of *COL4A5* patients had ever had ACR testing completed though 82 (95%) had urinalysis data. Overall, only 28 (33%) of *COL4A5* patients were receiving RAAS inhibitors. Among the 37 individuals with ACR ≥30 mg/g or 1+ protein twice on dipstick (and without ESKD), only 18 (49%) were taking RAAS inhibitors.

Discussion:

In this study using an unselected cohort, we provide more representative quantification of risks associated with X-linked AS, by sex and genotype. Risks of ESKD was highest among males with P/LP variants (41%), intermediate for males with p.Gly624Asp (15%) and females with P/LP

variants (10%), and not significantly increased for females with p.Gly624Asp (3%). Overall, we show a wider spectrum of disease severity in both men and women with *COL4A5* variants than described in prior studies that focused on families known to be afflicted by the disease. For example, prior studies suggested males with X-linked AS having a 100% chance of progression to ESKD[10]. In our cohort, there were 2 men above the age of 70 years who had eGFR ≥60 ml/min/1.73m^2, and only 75% of *COL4A5 P/LP* males and 33% of *COL4A5* p.Gly624Asp males were categorized as having KDIGO high-risk category or greater. This finding of more variable phenotypic severity has important implications in terms of providing more accurate information for families. The broader phenotypic spectrum using a genotype-first approach in an unselected cohort is consistent with other similar genotype-based studies for conditions ranging from autosomal dominant polycystic kidney disease to monogenic diabetes.[5,22,23] We also uncovered gaps in care related to AS. Despite a median follow up time of approximately 14 years, allowing ample opportunity for various forms of testing, less than half of individuals had albuminuria screening completed and very few individuals with *COL4A5* had a diagnosis of AS.

With our genotype-based approach, we are also able to study genotype-phenotype correlations in a less biased fashion than prior studies. Prior studies have suggested that nonsense/frameshift variants confer the worst prognosis, followed by splice variants, and then the mildest phenotypes for missense and in-frame deletion variants.[6,20] In our study, we did not observe this genotype-phenotype correlation. None of the 5 females with nonsense/frameshift variants (mean age 63.8 y [SD 27.0], mean eGFR 75.5 ml/min/1.73m^2 [SD 28.3]) had ESKD or even eGFR <30 ml/min/1.73m^2. It is possible that lumping variants into various groups may not reflect differential severity of individual variants within each group. Beyond p.Gly624Asp, which was associated with better outcomes for both males and females, we were unable to make firm conclusions about individual variants. Verifying genotype-phenotype correlations will require larger, additional cohort studies with unselected populations.

Our study and others have shown that p.Gly624Asp, thought to have originated in Central and Eastern Europe due to a founder effect between the 12th and 13th century, confers a milder form of disease and clinical course which may be explained by the location of this variant within the type IV

collagen gene[24,25]. The p.Gly624Asp variant is found adjacent to a non-collagenous interruption within the collagenous domain and thus has been termed to be located at a non-collagenous boundary. It has been hypothesized that since p.Gly624Asp occurs adjacent to such an interruption, it likely increases the flexibility of an already flexible region, explaining why it results in a milder form of disease, as compared to variants inducing flexibility in the more rigid parts of the collagenous domain[25]. However, despite being associated with a milder presentation, males with p.Gly624Asp variants still had a 6-fold risk for ESKD, albeit many years later than has been reported for other P/LP variants[25].

In the past and still today per documentation found in our chart review, females are often denoted as simply being “carriers” of X-linked AS. Although the manifestations of disease are on average less severe as compared to males, risks are still present. An earlier study of families with X-linked AS by Jais et al. had reported a 30% risk of ESKD in females with *COL4A5* by the age of 60, with higher risk in those with concurrent proteinuria and hearing loss[11]. In our study, females with P/LP variants also tended to be at increased risk of ESKD though overall risk was lower with 10% developing ESKD by the age of 65y, likely due to the unselected nature of our cohort. Regardless, the diagnosis of AS in women has importance implications for affected women as well as for reproductive decision making and family planning.

Prior research by Gross et al. has demonstrated a role for RAAS inhibitors in delaying onset of renal failure and improving life expectancy in AS[26]. Their study largely focused on AS in males with sibling pairs showing a significantly delayed onset of renal failure in the younger sibling thought to be due to earlier diagnosis and initiation of medical therapy. This delayed onset of renal failure with RAAS blockade was similarly seen in a study examining females with AS[27]. Even in males and females with hypomorphic Gly624Asp variants, RAAS inhibitors have been shown to be associated with reduced risk of end stage renal failure in an observational study by Boeckhaus et al[28]. Despite these studies, only 50% of patients were on RAAS inhibitors even when albuminuria was detected on laboratory studies, emphasizing the importance of early genetic testing and

screening to identify such patients. This finding raises the question of whether or not diagnosing these individuals with AS would result in better risk awareness leading to improved adherence to medical therapy, and potentially better patient outcomes.

The major strength of our study is the utilization of an unselected health population and exome sequencing data to inform our diagnosis of patients with X-linked AS, allowing for an in-depth exploration of the phenotypic spectrum of disease. Additionally, due to the availability of extensive longitudinal EHR data, we were able to examine a wide breadth of different phenotypic outcomes. However, in the same vein, the use of the existing electronic health record was at times limited, particularly for traits such as ophthalmologic and audiometry data. Another limitation of our study is that all patients with *COL4A5* in our cohort were of European ancestry. In conclusion, we were able to show in a large unselected cohort, the wide spectrum of phenotypic presentation of males and females with *COL4A5* variants. X-linked AS was underdiagnosed in both males and females with opportunities to improve monitoring of disease progression and better use of guideline-recommended RAAS inhibitors. Further studies are needed to determine whether early genetic diagnosis can improve outcomes in AS.

Disclosures

A.C. has served as a consultant for Novartis, Reata, and Amgen, and has received research funding from Novartis, Novo Nordisk, and Bayer.

Acknowledgements

We are grateful to the many patients who contributed to the MyCode Community Health Initiative by providing genomic and electronic health information. We thank the Regeneron Genetics Center for providing funding for patient enrollment and exome sequencing for the DiscovEHR study, and we would like to acknowledge the Geisinger-Regeneron DiscovEHR Collaboration for the genotypic and phenotypic data. This work was supported by NIGMS grant GM111913 to T.M.

We thank the USRDS for their support. The data reported here have been supplied by the United States Renal Data System (USRDS). The interpretation and reporting of these data are the responsibility of the author(s) and in no way should be seen as an official policy or interpretation of the U.S. government.

Data availability: The data supporting the findings of this study are available within the article and its Supplementary Data files. Additional information for reproducing the results described in the article is available upon reasonable request and subject to a data use agreement.

Author contributions:
Study design (ARC, HLK), data collection (MZ, KS, KR, BM, IDB, TM), analysis (ARC), interpretation of results (MZ, IDB, TM, ARC), drafted manuscript (MZ, ARC), revised critically for important intellectual content (all authors). All authors have approved the version to be published.

Conflicts of interest: No potential conflicts of interest relevant to this article were reported.

References

1. Gibson J, Fieldhouse R, Chan MMY, et al. Prevalence Estimates of Predicted Pathogenic COL4A3-COL4A5 Variants in a Population Sequencing Database and Their Implications for Alport Syndrome. *J Am Soc Nephrol*. 2021;32(9):2273-2290. doi:10.1681/ASN.2020071065
2. Kashtan CE, Ding J, Garosi G, et al. Alport syndrome: a unified classification of genetic disorders of collagen IV α345: a position paper of the Alport Syndrome Classification Working Group. *Kidney Int*. 2018;93(5):1045-1051. doi:10.1016/J.KINT.2017.12.018
3. Hudson BG. The molecular basis of goodpasture and alport syndromes: Beacons for the discovery of the collagen IV family. *Journal of the American Society of Nephrology*. 2004;15(10):2514-2527. doi:10.1097/01.ASN.0000141462.00630.76
4. Savige J, Ariani F, Mari F, et al. Expert consensus guidelines for the genetic diagnosis of Alport syndrome. *Pediatr Nephrol*. 2019;34(7):1175-1189. doi:10.1007/S00467-018-3985-4
5. Solanki K V., Hu Y, Moore BS, et al. The Phenotypic Spectrum of COL4A3 Heterozygotes. *Kidney Int Rep*. 2023;8(10):2088-2099. doi:10.1016/j.ekir.2023.07.010
6. Gibson JT, Huang M, Shenelli Croos Dabrera M, et al. Genotype–phenotype correlations for COL4A3–COL4A5 variants resulting in Gly substitutions in Alport syndrome. *Scientific Reports 2022 12:1*. 2022;12(1):1-13. doi:10.1038/s41598-022-06525-9
7. Gibson JT, de Gooyer M, Huang M, Savige J. A Systematic Review of Pathogenic COL4A5 Variants and Proteinuria in Women and Girls With X-linked Alport Syndrome. *Kidney Int Rep*. 2022;7(11):2454-2461. doi:10.1016/J.EKIR.2022.08.021
8. Kim JH, Lim SH, Song JY, et al. Genotype-phenotype correlation of X-linked Alport syndrome observed in both genders: a multicenter study in South Korea. *Sci Rep*. 2023;13(1). doi:10.1038/S41598-023-34053-7
9. Żurowska AM, Bielska O, Daca-Roszak P, et al. Mild X-linked Alport syndrome due to the COL4A5 G624D variant originating in the Middle Ages is predominant in Central/East Europe and causes kidney failure in midlife. *Kidney Int*. 2021;99(6):1451-1458. doi:10.1016/J.KINT.2020.10.040
10. Kashtan CE, Ding J, Garosi G, et al. Alport syndrome: a unified classification of genetic disorders of collagen IV α345: a position paper of the Alport Syndrome Classification Working Group. *Kidney Int*. 2018;93(5):1045-1051. doi:10.1016/j.kint.2017.12.018
11. Jais JP, Knebelmann B, Giatras I, et al. X-linked Alport syndrome: Natural history and genotype-phenotype correlations in girls and women belonging to 195 families: A “European Community Alport Syndrome Concerted Action” study. *Journal of the American Society of Nephrology*. 2003;14(10):2603-2610. doi:10.1097/01.ASN.0000090034.71205.74
12. Carey DJ, Fetterolf SN, Davis FD, et al. The Geisinger MyCode community health initiative: an electronic health record–linked biobank for precision medicine research. *Genetics in Medicine*. 2016;18(9). doi:10.1038/gim.2015.187
13. Landrum MJ, Lee JM, Benson M, et al. ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res*. 2016;44(D1). doi:10.1093/nar/gkv1222
14. Kelly MA, Leader JB, Wain KE, et al. Leveraging population-based exome screening to impact clinical care: The evolution of variant assessment in the Geisinger MyCode research project. *Am J Med Genet C Semin Med Genet*. 2021;187(1):83-94. doi:10.1002/AJMG.C.31887
15. Inker LA, Eneanya ND, Coresh J, et al. New Creatinine- and Cystatin C-Based Equations to Estimate GFR without Race. *N Engl J Med*. 2021;385(19):1737-1749. doi:10.1056/NEJMOA2102953
16. U.S. Renal Data System. 2022 USRDS annual data report: Epidemiology of kidney disease in the United States. National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD, 2022.
17. Stevens PE, Ahmed SB, Carrero JJ, et al. KDIGO 2024 Clinical Practice Guideline for the Evaluation and Management of Chronic Kidney Disease. *Kidney Int*. 2024;105(4):S117-S314. doi:10.1016/J.KINT.2023.10.018
18. Talbot D, Mésidor M, Chiu Y, Simard M, Sirois C, Denis Talbot C. An Alternative Perspective on the Robust Poisson Method for Estimating Risk or Prevalence Ratios Running head: Alternative Perspective on Robust Poisson Method. doi:10.48550/arXiv.2112.00547

19. Chen W, Shi J, Qian L, Azen SP. Comparison of robustness to outliers between robust poisson models and log-binomial models when estimating relative risks for common binary outcomes: A simulation study. *BMC Med Res Methodol*. 2014;14(1). doi:10.1186/1471-2288-14-82
20. Kim JH, Lim SH, Song JY, et al. Genotype–phenotype correlation of X-linked Alport syndrome observed in both genders: a multicenter study in South Korea. *Scientific Reports 2023 13:1*. 2023;13(1):1-10. doi:10.1038/s41598-023-34053-7
21. JAIS JP, KNEBELMANN B, GIATRAS I, et al. X-linked Alport Syndrome. *Journal of the American Society of Nephrology*. 2000;11(4):649-657. doi:10.1681/ASN.V114649
22. Chang AR, Moore BS, Luo JZ, et al. Exome Sequencing of a Clinical Population for Autosomal Dominant Polycystic Kidney Disease. *JAMA*. 2022;328(24):2412-2421. doi:10.1001/JAMA.2022.22847
23. Mirshahi UL, Colclough K, Wright CF, et al. Reduced penetrance of MODY-associated HNF1A/HNF4A variants but not GCK variants in clinically unselected cohorts. *Am J Hum Genet*. 2022;109(11):2018-2028. doi:10.1016/J.AJHG.2022.09.014
24. Gibson JT, Huang M, Shenelli Croos Dabrera M, et al. Genotype–phenotype correlations for COL4A3–COL4A5 variants resulting in Gly substitutions in Alport syndrome. *Sci Rep*. 2022;12(1). doi:10.1038/s41598-022-06525-9
25. Żurowska AM, Bielska O, Daca-Roszak P, et al. Mild X-linked Alport syndrome due to the COL4A5 G624D variant originating in the Middle Ages is predominant in Central/East Europe and causes kidney failure in midlife. *Kidney Int*. 2021;99(6):1451-1458. doi:10.1016/j.kint.2020.10.040
26. Gross O, Licht C, Anders HJ, et al. Early angiotensin-converting enzyme inhibition in Alport syndrome delays renal failure and improves life expectancy. *Kidney Int*. 2012;81(5):494-501. doi:10.1038/ki.2011.407
27. Temme J, Peters F, Lange K, et al. Incidence of renal failure and nephroprotection by RAAS inhibition in heterozygous carriers of X-chromosomal and autosomal recessive Alport mutations. *Kidney Int*. 2012;81(8):779-783. doi:10.1038/ki.2011.452
28. Boeckhaus J, Hoefele J, Riedhammer KM, et al. Lifelong effect of therapy in young patients with the COL4A5 Alport missense variant p.(Gly624Asp): a prospective cohort study. *Nephrology Dialysis Transplantation*. 2022;37(12):2496-2504. doi:10.1093/ndt/gfac006

Table 1. Characteristics of *COL4A5* P/LP X-linked Alport Syndrome patients and matched controls

	COL4A5			**Matched Controls**		
	Overall (n=86)	Males (n=30)	Females (n=56)	Overall (n=412)	Males (n=145)	Females (n=267)
Age	56.4 (19.7)	52.4 (19.8)	58.5 (19.4)	56.9 (19.3)	52.8 (19.6)	59.1 (18.8)
Year of first outpatient visit	2004 (2001-2009)	2005 (2001-2009)	2003 (2001-2009.5)	2004 (2001-2009)	2004 (2001-2009)	2003 (2001-2010)
Hypertension	58 (67)	21 (70)	37 (66)	208 (50)	77 (53)	131 (49)
Diabetes	18 (21)	6 (20)	12 (21)	74 (18)	27 (19)	47 (18)
eGFR available	86 (100)	31 (100)	56 (100)	412 (100)	145 (100)	267 (100)
Urinalysis available	82 (95)	28 (93)	54 (96)	334 (81)	105 (72)	229 (86)
ACR available	40 (47)	16 (53)	24 (43)	107 (26)	35 (24)	72 (27)

Cases were matched 1:up to 5 without replacement by age (1-year increments), sex, ICD diagnosis of diabetes, year of first outpatient encounter at Geisinger.

Table 2. Phenotypic Features of *COL4A5* P/LP X-linked Alport Syndrome by Sex and Genotype

	Males			Females		
	P/LP variants (n=17)	Hypomorphic variant (Gly624Asp) (n=13)	Controls (n=145)	P/LP variants (n=21)	Hypomorphic variant (Gly624Asp) (n=35)	Controls (n=267)
Age	49.9 (21.8)	56.9 (17.9)	52.8 (19.6)	55.6 (25.5)	60.3 (14.8)	59.7 (18.6)
Any AS phenotypic feature below	16 (89)	10 (77)	47 (32)	18 (86)	24 (69)	134 (50)
ICD Code-based diagnoses						
Alport Syndrome	5 (29)	0	0	4 (19)	2 (6)	0
Hematuria ICD code (%)	3 (18)	0	5 (3)	2 (10)	4 (11)	23 (9)
FSGS ICD (%)	0	0	0	1 (5)	2 (6)	1 (0.4)
ESKD ICD (%)	7 (41)	2 (15)	4 (3)	2 (10)	1 (3)	4 (2)
Bilateral sensorineural hearing loss ICD (%)	5 (29)	1 (8)	3 (2)	2 (10)	3 (9)	15 (6)
Hearing loss ICD (%)	13 (76)	4 (31)	15 (10)	7 (35)	6 (18)	45 (17)
Lab-based diagnoses						
Hematuria on UA* (%)	7/16 (44)	9/12 (75)	21/103 (20)	16/20 (80)	20/33 (61)	69/215 (32)
1+ protein twice on UA	12/16 (75)	4/12 (33)	23/105 (22)	11/20 (55)	12/34 (35)	50/229 (22)
2+ protein twice on UA	9/16 (56)	4/12 (33)	15/105 (14)	8/20 (40)	8/34 (24)	23/229 (10)
ACR ≥30 mg/g (%)	7/11 (64)	4/5 (80)	7/35 (20)	10/10 (100)	8/14 (57)	16/72 (22)
ACR ≥300 mg/g (%)	4/11 (36)	2/5 (40)	0/35	4/10 (40)	2/14 (14)	4/71 (6)
eGFR<60 (%)	9 (53)	3 (23)	21 (14)	9 (45)	7 (21)	52 (19)
eGFR<30 (%)	7 (41)	3 (23)	4 (3)	2 (10)	2 (6)	7 (3)
KDIGO CKD Risk Category**						
N adults 18+ years w/eGFR and UA data	16	12	107	20	34	217
No CKD (%)	1 (6)	2 (17)	66 (62)	2 (10)	10 (29)	100 (46)
Hematuria alone (%)	0	4 (33)	10 (9)	2 (10)	8 (24)	40 (18)
Moderately increased risk (%)	3 (19)	2 (17)	15 (14)	3 (15)	8 (24)	45 (21)
High risk (%)	3 (19)	1 (8)	10 (9)	8 (40)	5 (15)	18 (8)
Very high risk (%)	2 (13)	0	2 (2)	3 (15)	1 (3)	7 (3)
Extremely high risk (%)	7 (44)	3 (25)	4 (4)	2 (10)	2 (6)	7 (3)

Cases were matched 1:5 to controls (without *COL4A3/4/5* variants at allele frequency<0.01) by age at time of most recent eGFR (1-year increments), sex, ICD diagnosis of diabetes, and year of first outpatient encounter at Geisinger.

This table does not use USRDS data.

*Defined as trace blood or greater on UA on at least 50% of urinalyses

**2 cases and 6 controls <18 years of age were excluded from this categorization

Table 3. Prevalence Ratios of Alport Syndrome Phenotypic Features by Sex and Genotype

	Male		Female	
	P/LP variants (other than Gly624Asp) Prevalence Ratio (95%)	Gly624Asp Prevalence Ratio (95%)	P/LP variants (other than Gly624Asp) Prevalence Ratio (95%)	Gly624Asp Prevalence Ratio (95%)
Any AS phenotypic feature below	2.88 (2.00, 4.16)***	2.20 (1.39, 3.48)**	1.75 (1.34, 2.29)***	1.35 (1.03, 1.76)*
ICD diagnosis-based				
ESKD ICD	35.00 (4.54, 269.06)**	3.08 (0.56, 16.80)	9.33 (0.88, 99.21)	1.61 (0.17, 15.07)
Bilateral sensorineural hearing loss ICD	12.50 (2.62, 59.64)**	4.62 (0.30, 70.41)	1.04 (0.24, 4.48)	2.41 (0.63, 9.23)
Hearing loss ICD	3.08 (1.51, 6.28)**	9.23 (1.87, 45.67)**	1.37 (0.69, 2.73)	1.49 (0.65, 3.44)
Lab-based diagnoses				
Trace blood or greater on UA*	2.36 (1.07, 5.23)*	3.34 (1.80, 6.21)***	2.55 (1.74, 3.74)***	1.86 (1.28, 2.70)**
1+ protein twice on UA	2.87 (1.67, 4.92)***	1.48 (0.57, 3.89)	2.63 (1.50, 4.64)**	1.57 (0.90, 2.73)
2+ protein twice on UA	4.50 (1.98, 10.24)***	2.04 (0.73, 5.71)	3.64 (1.64, 8.08)**	2.50 (1.12, 5.55)*
ACR ≥30 mg/g	2.42 (0.99, 5.89)	6.40 (1.57, 26.01)**	5.80 (2.59, 13.01)***	2.23 (1.12, 4.45)*
eGFR<60	3.75 (1.87, 7.50)***	1.54 (0.48, 4.95)	2.21 (1.16, 4.19)*	1.02 (0.49, 2.13)
eGFR<30	35.00 (4.55, 269.06)**	4.62 (1.04, 20.56)*	4.43 (0.66, 29.92)	1.87 (0.38, 9.31)

Each patient with *COL4A5* P/LP variant matched with up to 5 controls (without *COL4A3/4/5* variants at allele frequency<0.01). Poisson regression analyses with robust variance performed in sex-stratified analyses. For kidney transplant patients, eGFR imputed as 5 ml/min/1.73m^2

This table does not use USRDS data.

*P<0.05, **P<0.01, ***P<0.001

Table 4. Hazard Ratios for ESKD by Genotype in Males and Females

	HR (95% CI)	P value
Males		
Control (n=145)	Ref	
Missense or in-frame variant (n=8)	23.33 (5.76, 94.44)	<0.001
Hypomorphic Gly624Asp (n=13)	5.61 (1.02, 30.78)	0.047
Nonsense or frameshift (n=2)	74.02 (3.54, 1549.74)	0.006
Splice acceptor/donor (n=7)	14.23 (2.38, 85.19)	0.004
Females		
Control (n=267)	Ref	
Missense or in-frame variant (n=9)	9.84 (1.06, 91.20)	0.048
Hypomorphic Gly624Asp (n=35)	2.07 (0.23, 18.60)	0.5
Nonsense or frameshift (n=5)	-	-
Splice acceptor/donor (n=7)	11.75 (1.18, 116.94)	0.04

Cox proportional hazards models adjusted for age in sex-stratified analyses.

Table 5. Characteristics of *COL4A5* P/LP X-linked Alport Syndrome with native kidney biopsies

Study ID, variant	Age at time of biopsy, sex	CKD Stage, hematuria at time of biopsy	Last eGFR, age	Extra-renal features	Kidney Biopsy lab (year), diagnosis	Electron Microscopy findings			Immuno-fluorescence staining for collagen IV	IFTA	Global GS
						Main EM findings	Classic EM features of Alport's (GBM splitting, Lamination basketweaving)	Thin GBM			
Males											
ADPKD163904 p.Arg1680Ter	5-9y, M	G1A1, Y	eGFR 124 at 20-24y	HL	GMC (2005-2009): Alport syndrome.	Variable GBM thickness.	GBM splitting Lucent areas noted in GBM	Y	Scattered positive staining of collagen type IV along glomerular loops	Absent	NA
ADPKD4155 p.Gly1244Asp	15-19y, M	G1A2, Y	eGFR 126 at 15-19y	N	Nephropath (2014-2018): Thin glomerular basement membranes	Prominent global thinning of GBM (mean basement membrane thickness of 161nm). Mild segmental epithelial foot process effacement.	No	Y	Type IV alpha5 weak staining of GBM	Absent	2/26
ADPKD81593 p.Gly1244Asp	15-19y, M	G1A2, Y	eGFR 96 at 20-24y	N	Nephropath (2015-2019): Thin glomerular basement membranes	Prominent global GBM thinning Moderate foot process effacement.	NA	Y	NA	Minimal	1/21
ADPKD155265 p.Gly1244Asp	35-39y, M	G1A3, Y	eGFR 46 at 45-49y	HL	GMC (2015-2019): FSGS	Podocytes segmentally swollen and effaced with few small associated subepithelial and subendothelial dense deposits. No ultrastructural organization of dense deposits.	No	N	NA	Minimal	2/17
ADPKD98958 p.Leu1655Arg	45-49y, M	G2A3, Y	eGFR 41 at 50-54y	N	GMC (2015-2019): Specimen suboptimal for evaluation. Ultrastructural glomerular features suggesting hereditary nephritis.	Irregular GBM with nonspecific variations in thickness. Global podocyte foot proces effacement.	No	Irregular	NA	Absent	NA
ADPKD57778 p.Leu1655Arg	15-19y, M	NA	ESKD	HL	GMC (1990-1994): No biopsy report available. Nephrology note - renal biopsy results reported to patient were post-strep glomerulonephritis.	NA	NA	NA	NA	NA	NA
ADPKD78032 c.546+2dup	35-39y, M	NA	ESKD	N	Unknown (1995-1999): Hereditary Nephritis (Nephrology note) No biopsy report available.	NA	NA	NA	NA	NA	NA
ADPKD109651 p.Gln1481Ter	10-14y, M	G1A3, Y	ESKD	HL	GMC (2000-2004): Alport glomerulopathy, FSGS	Widespread irregularity of GBM with alternating thinning and thickening and foci of splitting and fragmentation of lamina densa. Focal effacement of foot processes.	Y GBM splitting	Y	NA	Mild	NA
ADPKD112266 c.438+2T>G	15-19y, M	NA	ESKD	HL	Unknown (2000-2004): hereditary nephritis, questionable	NA	NA	NA	NA	NA	NA

					membranoproliferative glomerulonephritis						
ADPKD121508 p.Gly624Asp	45-49y, M	G3bA3, Y	eGFR 14 at 55-59y	N	GMC (2010-2014): Suboptimal specimen for a definitive diagnosis.	Focal podocyte foot process effacement Mesangial sclerosis, and some electron-dense material. (1 glom)	No	NR	NA	Mild-moderate	NA
ADPKD88030 p.Gly624Asp	60-64y, M	G3aA3, Y	eGFR 20 at 80-84y	HL	Hershey Medical Center (2000-2004): FSGS (Nephrology note) No biopsy report available.	NA	NA	NA	NA	NA	NA
Females											
ADPKD83056, Ile355TyrfsTer 56	5-9y (2nd biopsy), F	G1A3, Y (2nd biopsy)	eGFR 125 at 20-24y	N	GMC (2005-2009): IgM nephropathy, FSGS CHOP (2000-2004): insufficient sample, thin basement membrane disease	GMC (2005-2009): GBM segmentally thickened w/segmental sclerotic features. Segment fusion of foot processes. CHOP (2000-2004): GBM focally appear thin (approx. 200nm) but no diffuse thinning. GBM, patchy effacement of epithelial foot processes	GMC (2005-2009): N CHOP (2000-2004): GBM appear focally thin	Focally thin on initial biopsy		Minimal	NR
ADPKD61311 c.546+2dup	5-9y, F	G1A1, Y	eGFR 139 at 15-19y	HL	GMC (2010-2014): Hereditary nephropathy.	Diffuse, thin GBM with focal variation in thickness.	No	Y	NA	Absent	Absent
ADPKD5080 c.546+2dup	5-9y, F	G1A1, Y	eGFR 94 at 20-24y	HL	GMC (2005-2009): No light microscopic or immunofluorescence abnormalities.	NA	NA	NA	NA	Absent	Absent
ADPKD15744 p.Gly624Asp	10-14y, F	NA	eGFR 97 at 35-39y	N	Unknown (1990-1994): No biopsy report available. Nephrology note - renal biopsy obtained in Philadelphia - Alport Syndrome diagnosis. No additional details available.	NA	NA	NA	NA	NA	NA
ADPKD155572 p.Gly624Asp	35-39y, F	G1A2, Y	eGFR 108 at 45-49y	N	GMC (2015-2019): Focal segmental proliferative glomerulonephritis with complement deposits, severe arteriosclerosis, minimal interstitial fibrosis (5%).	Segmental podocyte swelling and effacement; few small associated subepithelial and subendothelial dense deposits and endocapillary proliferation with membrane remodeling.	No	No	NA	Minimal	NA
ADPKD167136 p.Gly624Asp	Age NA, F	NA	eGFR 76 at 50-54y	N	Unknown: No biopsy report available. PCP notes had renal biopsy but provides no other information.	NA	NA	NA	NA	NA	NA
ADPKD77063 p.Gly624Asp	40-44y, F	G1A3, Y	eGFR 73 at 60-64y	N	GMC (2000-2004): Consistent with an early stage of focal segmental glomerulosclerosis with no significant interstitial fibrosis.	NA	NA	NA	NA	Absent	NA

Sample/Patient IDs used in Table 4 are unknown to anyone outside the research group. Ages and years are given as a range to preserve anonymity.

This table does not use any USRDS data.

Abbreviations: CKD (chronic kidney disease), : HL (hearing loss), NA (not available), FSGS (focal segmental glomerulosclerosis), GBM (glomerular basement membrane), GS (glomerulosclerosis), EM (electronic microscopy),

PCP (primary care provider), IFTA (interstitial fibrosis and tubular atrophy), GMC (Geisinger Medical Center); CKD staging, ACR, PCR, 24 hour urine protein obtained +/- 1 year of biopsy date; CKD-EPI Creatinine Equation (2021) to estimate GFR for adults, Bedside Schwartz equation (2009) used to estimate GFR for pediatrics

Figure 1. KDIGO risk categories by sex and variant type

Figure 1 Legend. KDIGO risk categories created using eGFR and ACR. If ACR not available, then urine dipstick protein data used with dipstick protein 1+ twice classified as ACR 30-299 mg/g, and dipstick protein 2+ or greater twice classified as ACR 300+ mg/g. KDIGO risk categories defined as: 0 - no CKD or hematuria; 1 - hematuria alone (eGFR ≥60, ACR <30); 2 – Moderately increased risk (eGFR 45-59 and ACR<30 or eGFR ≥60 with ACR 30-299); 3 – High risk (eGFR 30-44 with ACR<30 or eGFR 45-59 with ACR 30-299, or eGFR ≥60 with ACR 300+); 4 – Very high risk (eGFR 15-29 with ACR<300 or eGFR 30-44 with ACR ≥30 or eGFR 45-59 with ACR ≥300); 5 – Extremely high risk (eGFR<15 or eGFR 15-29 with ACR 300+; see Supplemental Table 2).

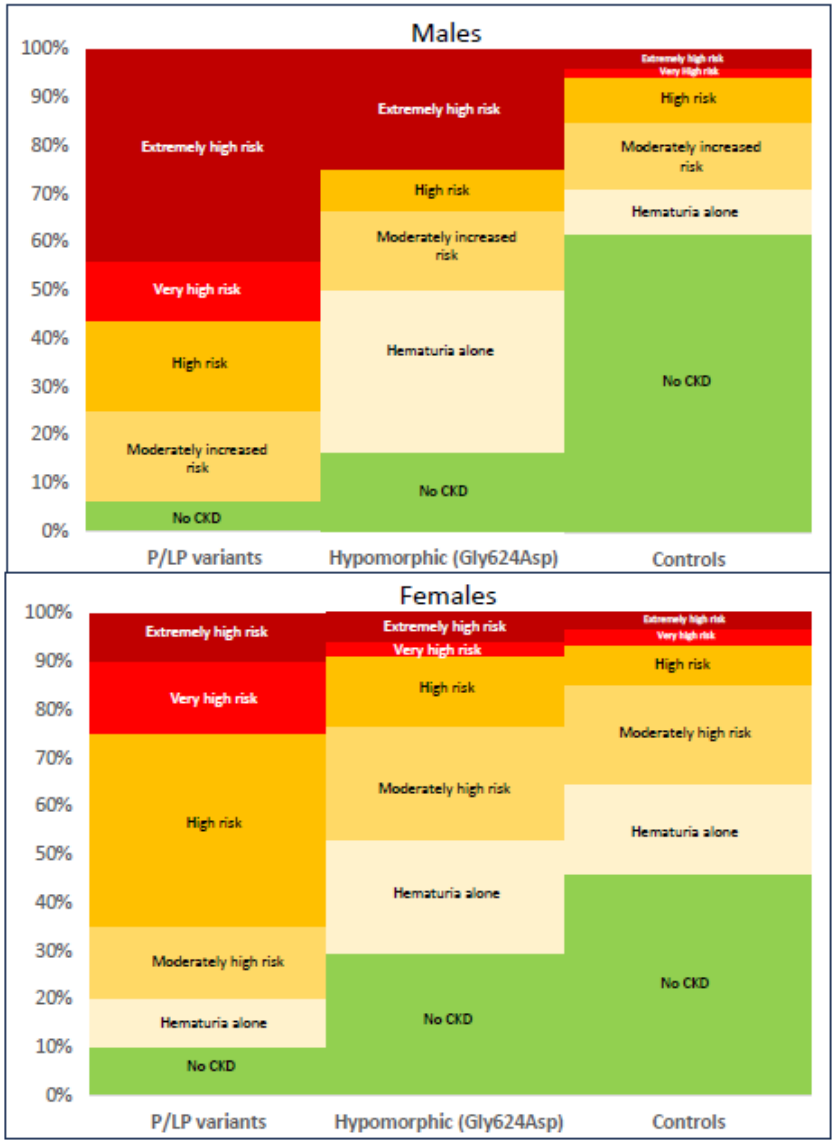


Figure 2. Prevalence Ratios for Alport Syndrome-related phenotypes

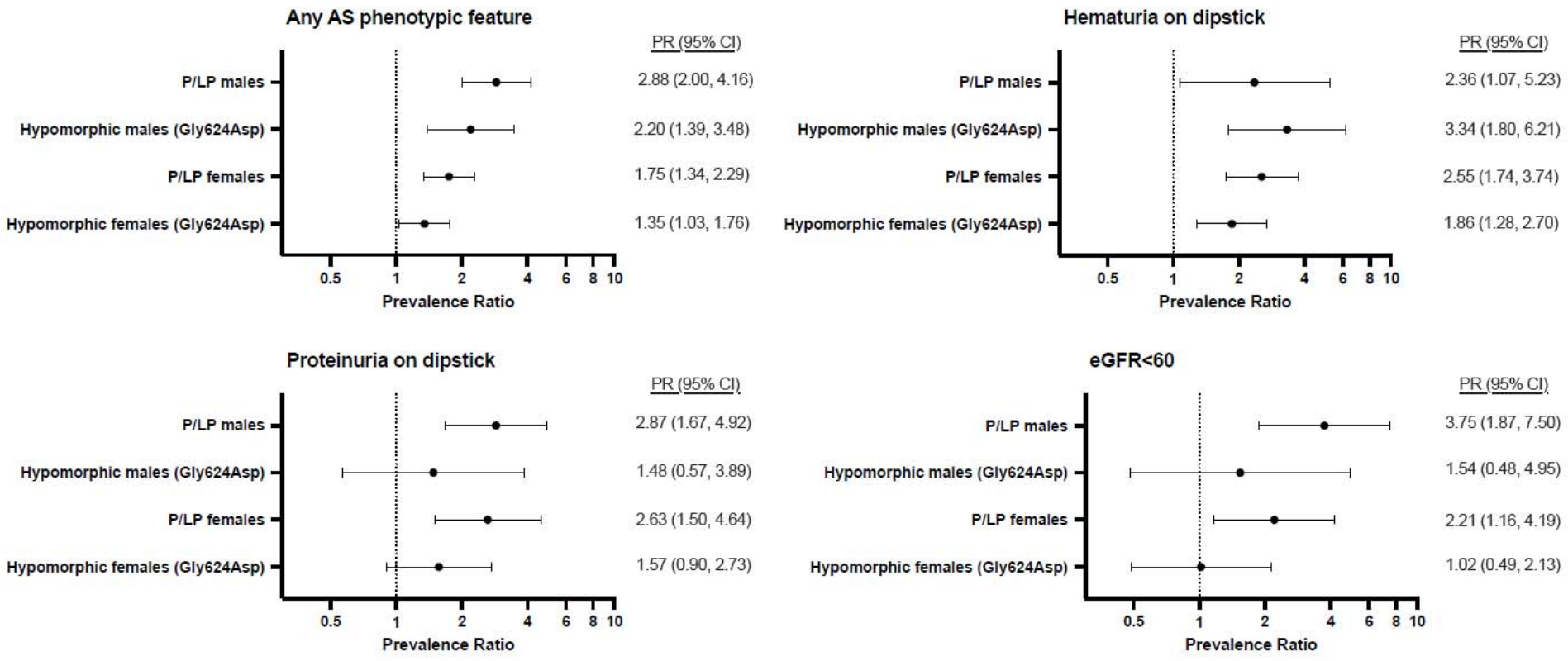


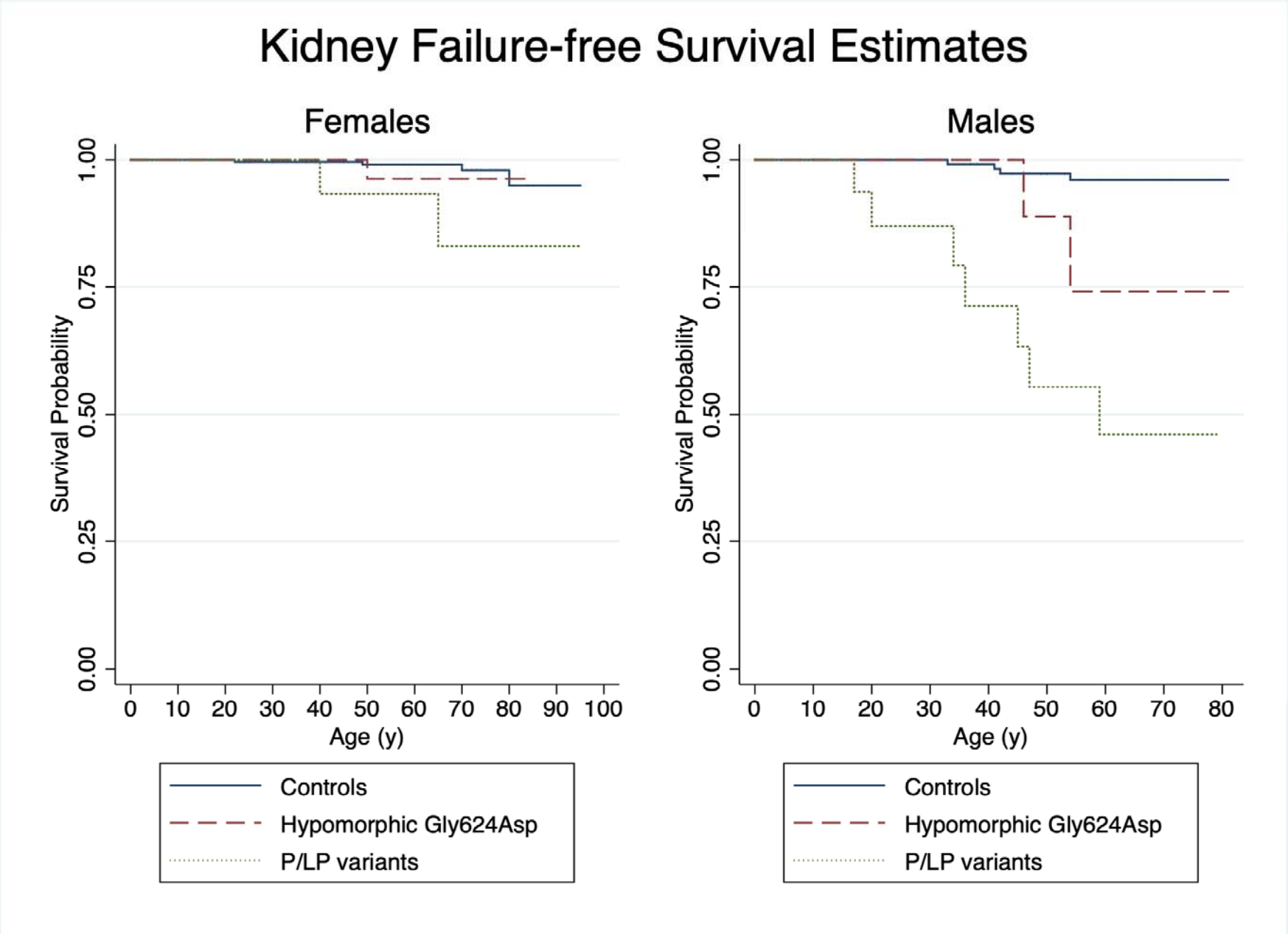
Figure 2 Legend.

Associations between *COL4A5* P/LP variants and renal phenotypes examined using Poisson regression with robust variance in analyses in stratified analyses by genotype and sex.

Figure 3. Kaplan-Meier Survival Curves for ESKD-free Survival

Figure 3 Legend.

Kaplan-Meier curves comparing risk of ESKD by genotype for females and males with age as the time variable.



Supplemental Materials

Supplemental Methods.
Supplemental Table 1. Coding definitions for comorbidities

Supplemental Table 2. Modified KDIGO Risk Categories

Supplemental Table 3. Logistic Regression Analyses: Phenotypic Features of COL4A5 P/LP X-linked Alport Syndrome by Sex and Genotype

Supplemental Table 4. Association of Genotype with eGFR: Sex-stratified analyses

Supplemental Table 5. Characteristics by variants

Supplemental Table 6. Individuals with a *COL4A5* P/LP variant and an additional rare variant in *COL4A3/4/5*

Supplemental Table 1. Coding definitions for comorbidities

Comorbidities	Codes
Hematuria	ICD-10: R31.0, R31.1, R31.21, R31.29, R31.9; ICD-9: 599.7*
FSGS	ICD-10: N03.1, N04.1, N05.1, N06.1; ICD-9: 581.1
Alport syndrome	ICD-10: Q87.81
Bilateral sensorineural hearing loss	ICD-10: H90.3; ICD-9: 389.11, 389.12, 389.14, 389.18
Hearing loss	Insert codes
ESKD	Dialysis ICD codes 39.27, 39.42, 39.53, 39.54, 585.6, V45.11, V45.12, V56.1, V56.2, V56.31, V56.32, V56.8, V45.1, N18.6, Z91.15, N18.5+Z99.2 Transplant ICD codes 00.91, 00.92, 00.93, 55.53, 55.69, V42.0, 0TY****, Z94.0
Hypertension	ICD-9: 401 - 405 ICD-10: I10 - I16
Diabetes	ICD-9: 250* ICD-10: E10, E11, E13

Supplemental Table 2. Modified KDIGO Risk Categories

	ACR, mg/g		
eGFR	**<30**	**30-299**	**300+**
60+ without hematuria	No CKD	Moderately increased risk	High risk
60+ with hematuria	Hematuria alone	Moderately increased risk	High risk
45-59	Moderately increased risk	High risk	Very high risk
30-44	High risk	Very high risk	Very high risk
15-29	Very high risk	Very high risk	Extremely high risk
<15	Extremely high risk	Extremely high risk	Extremely high risk

Supplemental Table 3. Sensitivity Analysis excluding 9 patients with borderline quality control metrics - Firth Logistic Regression Analyses

	Male		Female	
	P/LP variants (other than Gly624Asp) OR (95%)	Gly624Asp OR (95%)	P/LP variants (other than Gly624Asp) OR (95%)	Gly624Asp OR (95%)
Any AS phenotypic feature below	21.40 (3.62, 126.46)	5.48 (1.14, 26.30)*	5.11 (1.51, 17.24)**	2.06 (0.96, 4.41)
ICD diagnosis-based				
ESKD ICD	62.05 (3.09, 1246.28)**	3.57 (0.59, 21.66)	8.33 (1.04, 67.03)*	2.07 (0.29, 14.50)
Bilateral sensorineural hearing loss ICD	20.37 (2.84, 145.97)**	4.65 (0.43, 50.28)	1.20 (0.27, 5.29)	2.71 (0.70, 10.48)
Hearing loss ICD	3.71 (1.10, 12.53)*	14.18 (1.76, 114.06)*	1.37 (0.48, 3.88)	1.66 (0.63, 4.38)
Lab-based diagnoses				
Trace blood or greater on UA*	2.70 (0.73, 9.99)	9.19 (1.73, 48.74)**	7.77 (2.46, 24.49)***	3.13 (1.44, 6.81)**
1+ protein twice on UA	11.21 (2.75, 45.72)**	2.25 (0.47, 10.77)	4.68 (1.69, 12.98)**	1.90 (0.85, 4.21)
2+ protein twice on UA	8.29 (2.09, 32.87)**	3.30 (0.65, 16.69)	5.39 (1.80, 16.12)**	2.98 (1.15, 7.75)*
ACR ≥30 mg/g	6.63 (1.14, 38.51)*	16.33 (1.25, 212.60)*	93.55 (4.73, 1849.01)**	3.70 (1.09, 12.55)*
ACR ≥300 mg/g	16.69 (0.75, 371.09)	23.00 (0.82, 642.70)	40.85 (1.95, 856.17)*	1.71 (0.32, 9.11)
eGFR<60	5.15 (1.47, 18.08)*	2.86 (0.61, 13.41)	3.16 (1.17, 8.52)*	1.07 (0.44, 2.61)
eGFR<30	62.05 (3.09, 1246.28)**	5.77 (1.05, 31.57)*	4.68 (0.75, 28.96)	2.16 (0.46, 10.10)

Supplemental Table 4. Firth Logistic Regression Analyses: Phenotypic Features of COL4A5 P/LP X-linked Alport Syndrome by Sex and Genotype

	Male		Female	
	P/LP variants (other than Gly624Asp) OR (95%)	Gly624Asp OR (95%)	P/LP variants (other than Gly624Asp) OR (95%)	Gly624Asp OR (95%)
Any AS phenotypic feature below	12.12 (2.92, 50.41)**	5.37 (1.44, 20.11)*	4.31 (1.26, 14.79)*	2.33 (0.99, 5.47)
ICD diagnosis-based				
ESKD ICD	38.54 (5.78, 257.14)***	3.51 (0.61, 20.04)	10.48 (1.29, 85.10)*	2.59 (0.37, 18.33)
Bilateral sensorineural hearing loss ICD	17.50 (3.44, 89.02)**	4.68 (0.45, 48.79)	1.52 (0.34, 6.77)	3.45 (0.88, 13.52)
Hearing loss ICD	4.74 (1.51, 14.83)**	10.89 (2.01, 59.14)**	1.63 (0.56, 4.71)	1.71 (0.61, 4.81)
Lab-based diagnoses				
Trace blood or greater on UA*	4.24 (1.26, 14.29)*	8.86 (2.20, 35.62)**	6.01 (1.87, 19.33)**	2.95 (1.28, 6.82)*
1+ protein twice on UA	11.70 (3.03, 45.17)***	1.73 (0.46, 6.46)	4.71 (1.60, 13.87)**	1.94 (0.82, 4.55)
2+ protein twice on UA	11.40 (3.10, 41.87)***	2.52 (0.65, 9.84)	7.76 (2.46, 24.45)***	3.24 (1.19, 8.84)*
ACR ≥30 mg/g	3.81 (0.81, 17.97)	16.20 (1.64, 159.68)*	84.64 (4.25, 1683.74)**	3.85 (1.06, 13.99)*
ACR ≥300 mg/g	27.00 (1.27, 572.18)*	22.14 (0.86, 571.28)	48.27 (2.26, 1029.82)*	1.12 (0.16, 7.94)
eGFR<60	6.66 (2.11, 21.06)**	1.77 (0.44, 7.14)	3.65 (1.28, 10.42)*	1.42 (0.57, 3.54)
eGFR<30	38.54 (5.78, 257.14)***	5.38 (1.06, 27.23)*	5.90 (0.94, 36.94)	2.74 (0.58, 12.90)

Firth logistic regression models, only including one individual per family. For kidney transplant patients, eGFR imputed as 5 ml/min/1.73m^2
This table does not use USRDS data.
*P<0.05, **P<0.01, ***P<0.001

Supplemental Table 5. Characteristics by variants

Variant	ClinVar	REVEL score or Splice AI for splice variants	Gnomad NFE AF	Sex	Age	Hematuria ICD code	Bilateral SNHL ICD code	Hearing loss ICD code	eGFR <60	eGFR< 30	ESKD*	Hematuria on dipstick	Proteinuria 1+ on dipstick twice	ACR ≥30 mg/g
NM_033380.3:c.231+1G>A (n=2 in 1 family)	P (1 star)	SpliceAI: AG 0, AL 0, DG 0.03, DL 0.96	Absent	Female: 2	77.0 (25.5)	1 (50)	0	0	1 (50)	0	0	1 (50)	0	NA
NM_033380.3:c.3017-2A>G (n=3 in 1 family)	P/LP; 2 star	SpliceAI: AG 0.87, AL 0.99, DG 0, DL 0	Absent	**Male: 2** Female: 1	**44.0 (48.1)** 45-49	**1 (50)** 0	**0** 0	**1 (50)** 0	**0** 0	**0** 0	**0** 0	**0** 0	**1 (50)** 0	**NA** NA
NM_033380.3:c.3553+1G>A (n=1 in 1 family)	NA	SpliceAI: AG 0, AL 0, DL 0.99, DG 0.86	Absent	Female: 1	80-84	0	0	0	1 (100)	0	0	1 (100)	0	1/1 (100)
NM_033380.3:c.438+2T>G (n=2 in 2 families)	NA	SpliceAI: AG 0.03, AL 0.20, DG 0, DL 0.03	Absent	**Male: 1** Female: 1	**45-49** 60-64	**0** 0	**1 (100)** 0	**1 (100)** 0	**1 (100)** 1 (100)	**1 (100)** 1 (100)	**1 (100)** 1 (100)	**1 (100)** 1 (100)	**1 (100)** 0	**0/1** NA
NM_033380.3:c.546+2dup (n=3 in 2 families)	P; 2 stars	SpliceAI: AG 0, AL 0.34, DG 0.01, DL 0.42	Absent	**Male: 1** Female: 2	**60-64** 18.5 (2.1)	**0** 0	**0** 0	**0** 2 (100)	**1 (100)** 0	**1 (100)** 0	**1 (100)** 0	**0** 2 (100)	**0** 1 (50)	**0/1** 1/1 (100)
NP_203699.1:p.Arg1680Ter (n=2 in 1 family)	P; 2 stars	NA	1.19E-06	**Male: 1** Female: 1	**20-24** 50-54	**0** 0	**1 (100)** 0	**1 (100)** 0	**0** 0	**0** 0	**0** 0	**1 (100)** 1 (100)	**1 (100)** 1 (100)	NA
NP_203699.1:p.Gln1481Ter (n=1 in 1 family)	P; 1 star	NA	Absent	**Male: 1**	**30-34**	**0**	**1 (100)**	**1 (100)**	**1 (100)**	**1 (100)**	**1 (100)**	**0**	**1 (100)**	**0**
NP_203699.1:p.Gly1122AspfsTer30 (n=1 in 1 family)	NA	NA	Absent	Female: 1	80-84	0	1 (100)	1 (100)	1 (100)	1 (100)	1 (100)	1 (100)	0	1/1 (100)
NP_203699.1:p.Gly1170Ser (n=1 in 1 family)	Conflicting (7P, 1 VUS); 1 star	0.985	1.19E-06	Female: 1	50-54	1 (100)	0	1 (100)	0	0	0	1 (100)	0	1/1 (100)
NP_203699.1:p.Gly1244Asp (n=5 in 4 families)	P; 2 stars	0.947	Absent	**Male: 3** Female: 2	**33.7 (22.0)** 53.0 (17.0)	**2 (67)** 0	**1 (33)** 0	**1 (33)** 0	**1 (33)** 1 (50)	**1 (33)** 0	**1 (33)** 0	**3 (100)** 2 (100)	**3 (100)** 2 (100)	**3/3 (100)** 1/1 (100)
NP_203699.1:p.Gly307Ser (n=2 in 2 families)	LP; 2 stars	0.984	2.50E-05	Female: 2	27.0 (4.2)	0	0	0	0	0	0	1 (50)	1 (50)	NA
NP_203699.1:p.Gly624Asp (n=47 in 37 families)	P/LP; 2 stars	0.91	6.15E-05	**Male: 13** Female: 34	**55.7 (17.1)** 60.0 (14.9)	**0** 4 (12)	**1 (8)** 3 (9)	**4 (31)** 6 (18)	**3 (23)** 7 (21)	**3 (23)** 2 (6)	**2 (15)** 1 (3)	**9/12 (75)** 19/32 (59)	**4/12 (33)** 12/33 (36)	**4/5 (80)** 8/14 (57)
NP_203699.1:p.Ile355TyrfsTer56 (n=1 in 1 family)	NA	NA	Absent	Female: 1	20-24	0	0	0	0	0	0	1 (100)	1 (100)	NA
NP_203699.1:p.Leu1655Arg (n=8 in 5 families)	P; 2 stars	0.991	Absent	**Male: 4** Female: 4	**54.0 (12.4)**	**0** 0	**1 (25)** 1 (25)	**3 (75)** 2 (50)	**4 (100)** 3 (75)	**3 (75)** 1 (25)	**3 (75)** 1 (25)	**2 (50)** 3 (75)	**3 (75)** 4 (100)	**2/3 (67)** 3/3 (100)

					64.5 (28.1)									
NP_203699.1:p.Pro1050LeufsTer102 (n=1 in 1 family)	NA	NA	Absent	Female: 1	85-89	0	0	0	0	0	0	NA	NA	NA
NP_203699.1:p.Pro765GlnfsTer27 (n=1 in 1 family)	P/LP; 2 stars	NA	Absent	Female: 1	70-74	0	0	0	1 (100)	0	0	1 (100)	1 (100)	1 (100)
NP_203699.1:p.Pro826_Gly828del (n=1 in 1 family)	P/LP; 2 stars	NA	Absent	**Male: 1**	**60-64**	**0**	**0**	**0**	**0**	**0**	**0**	**NA**	**NA**	**NA**
NM_033380.3:c.3791-2A>C (n=1 in 1 family)	NA	AG 0.04, AL 0.21, DG 0, DL 0.06	Absent	**Male: 1**	**70-74**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1 (100)**	**1/1 (100)**
NM_033380.3:c.3791-2A>G (n=2 in 2 families)	NA	AG 0.03, AL 0.20, DG 0, DL 0.03	1.2E-06	**Male: 2**	**71 (11.3)**	**0**	**0**	**0**	**1 (50) n3**	**0**	**0**	**0**	**1 (50)**	**1/1 (100)**

For variants with only 1 individuals, age given as a range to preserve anonymity.
Data from https://gnomad.broadinstitute.org/ were accessed on 3/28/24.
This table does not use USRDS data.
Abbreviations: ACMG (American College of Medical Genomics), NFE (non-Finnish European), P(pathogenic), LP (likely pathogenic), LB (likely benign), VUS (variant of unknown significance), NA (not available), BP (benign prediction), PP (pathogenic prediction), eGFR (estimated glomerular filtration rate), ESKD (end-stage kidney disease), ICD (international classification of diseases)

Study ID, sex	*COL4A5* P/LP variant	Additional heterozygous rare variant(s) in *COL4A3/4/5*	ClinVar	REVEL Score	Gnomad NFE AF	Phenotype
study_112266, male	NM_033380.3:c.438+2T>G	COL4A4 p.Pro352Leu	Conflicting (3 VUS, 1 benign), 1 star	0.569	0.0002610	ESKD, dipstick hematuria, bilateral sensorineural hearing loss
study_14580, male	NP_203699.1:p.Gly624Asp	COL4A4 p.Pro1241Arg	NA	0.433	8.996e-7	eGFR 112 at 35-39y, no dipstick proteinuria, no

Supplemental Table 6. Individuals with a *COL4A5* P/LP variant and an additional rare variant in *COL4A3/4/5*

		COL4A4 p.Ser1238Ala COL4A4 p.Pro1241Ala	NA NA	0.187 0.208	absent absent	hearing loss ICD
study_4155, male	NP_203699.1:p.Gly1244Asp	COL4A5 p.Val918Ile	NA	0.139	0.000004505	eGFR 126 at 15-19y, dipstick hematuria, ACR 300+, no hearing loss ICD
study_52460, male	NP_203699.1:p.Leu1655Arg	COL4A3 p.Val950Ile	Conflicting (3 uncertain, 1 LB), 1 star	0.332	0.0004797	ESKD, bilateral sensorineural hearing loss ICD, no dipstick hematuria, no dipstick proteinuria
study_57778, male	NP_203699.1:p.Leu1655Arg	COL4A3 p.Thr1489Ile	NA	0.68	0.00005000	ESKD, dipstick hematuria, ACR 300+, no hearing loss ICD
study_81593, male	NP_203699.1:p.Gly1244Asp	COL4A5 p.Val918Ile	NA	0.139	0.000004505	eGFR 97 at 20-24y, dipstick hematuria, ACR 300+, no hearing loss ICD
Family A*, study_49551, male	NM_033380.3:c.3017-2A>G	COL4A5 p.Ala430Asp	Benign/Likely Benign, 2 stars	0.202	0.005551	eGFR 150 at 10-14y, no dipstick hematuria, no hearing loss ICD
Family A*, study_77651, male	NM_033380.3:c.3017-2A>G	COL4A5 p.Ala430Asp	Benign/likely benign, 2 stars	0.202	0.005551	eGFR 77 at 75-59y, no dipstick hematuria, +dipstick proteinuria, no hearing loss ICD
Family A*, study_123841, female	NM_033380.3:c.3017-2A>G	COL4A5 p.Ala430Asp	Benign/Likely Benign, 2 stars	0.202	0.005551	eGFR 92 at 45-49y. No dipstick hematuria, no hearing loss ICD.
Family B*, study_140649, female	NP_203699.1:p.Gly1244Asp	COL4A5 p.Val918Ile	NA	0.139	0.000004505	eGFR 82 at 40-44y, ACR 300+ mg/g, no hearing loss ICD
Family B*, study_152355, female	NP_203699.1:p.Gly1244Asp	COL4A5 p.Val918Ile	NA	0.139	0.000004505	eGFR 46 at 65-59y, dipstick hematuria, 1+ proteinuria on dipstick, no hearing loss ICD
study_645, female	NP_203699.1:p.Gly624Asp	COL4A3 p.Glu269Lys COL4A3 p.GLy1277Ser	Benign, 2 stars Conflicting (4 LP, 12 VUS, 1 LB), 1 star	0.329 0.971	0.001730 0.0003288	eGFR 90 at 40-49y, dipstick hematuria, no dipstick proteinuria, no hearing loss ICD
study_7838, female	NM_033380.3:c.4964T>G	COL4A3 p.Ile1659Val	NA	0.432	0.00002119	ESKD, hearing loss ICD, ACR 300+, no dipstick hematuria
study_78385, female	NP_203699.1:p.Gly624Asp	COL4A4 p.Glu327Gly	NA	0.428	0.00004242	eGFR 103 at 65-69y, no dipstick hematuria, no dipstick proteinuria, no hearing loss ICD
study_132629, female	NP_203699.1:p.Gly1122AspfsTer30	COL4A4 p.Ile967Val	Benign, 2 stars	0.116	0.006885	eGFR 55 at 80-84y, ACR 30-299, bilateral sensorineural hearing loss ICD, no dipstick hematuria

Data from https://gnomad.broadinstitute.org/ were accessed on 3/28/24.
This table does not use USRDS data and Sample/Patient IDs are not known to anyone outside of the research group.
*Denotes related individuals
Abbreviations: ACMG (American College of Medical Genomics), NFE (non-Finnish European), P(pathogenic), LP (likely pathogenic), LB (likely benign), VUS (variant of unknown significance), NA (not available), BP (benign prediction), PP (pathogenic prediction), eGFR (estimated glomerular filtration rate), ESKD (end-stage kidney disease), ICD (international classification of diseases)