JASN® Journal of the American Society of Nephrology

# Prevalence Estimates of Predicted Pathogenic COL4A3 – COL4A5 Variants in a Population Sequencing Database and Their Implications for Alport Syndrome



| | |
|---|---|
| Journal: | *Journal of the American Society of Nephrology* |
| Manuscript ID | JASN-2020-07-1065.R3 |
| Manuscript Type: | Original Article - Basic Research |
| Date Submitted by the Author: | 29-Apr-2021 |
| Complete List of Authors: | Gibson, Joel; The University of Melbourne Department of Medicine Austin Health and Northern Health<br>Fieldhouse, Rachel; Garvan Institute of Medical Research, Kinghorn Centre for Clinical Genomics<br>Chan, Melanie; University College London, Department of Renal Medicine; Queen Mary University of London, Genomics England<br>Sadeghi-Alavijeh, Omid; University College London, Department of Renal Medicine; Queen Mary University of London, Genomics England<br>Burnett, Leslie; Garvan Institute of Medical Research, Kinghorn Centre for Clinical Genomics<br>Izzi, Valerio; University of Oulu, Center for Cell-Matrix Research and Biocenter Oulu<br>Persikov, Anton; Princeton University Princeton Center for Quantitative Biology Lewis-Sigler Institute for Integrative Genomics<br>Gale, Daniel; University College London, Department of Renal Medicine; Queen Mary University of London, Genomics England<br>Storey, Helen ; Guy's Hospital, Molecular Genetics, Viapath Laboratories<br>Savige, Judy; The University of Melbourne Department of Medicine Austin Health and Northern Health |
| Keywords: | Alport-s syndrome, Thin basement membrane nephropathy, Haematuria, COL4A5, COL4A3, COL4A4, collagen IV, genetic renal disease, glomerular basement membrane, Gly substitutions |
| | |

**Authors:** Gibson, Joel; Fieldhouse, Rachel; Chan, Melanie; Sadeghi-Alavijeh, Omid; Burnett, Leslie; Izzi, Valerio; Persikov, Anton; Gale, Daniel; Storey, Helen ; Savige, Judy

**Title:** Prevalence Estimates of Predicted Pathogenic COL4A3 â€“ COL4A5 Variants in a Population Sequencing Database and Their Implications for Alport Syndrome

**Running title:** Prevalence of COL4A3-COL4A5 variants

**Manuscript Type:** Original Article - Basic Research

**Manuscript Category**: Genetics and development

**Funders:** FUNDREF :No data available.

**Financial Disclosure:** No D. Gale reports Honoraria from Alexion, Otsuka; and Other Interests/Relationships as a Trustee for AlportUK. J. Savige reports Other Interests/Relationships with Alport Foundation Australia, and the PKD Australia Scientific Board. L. Burnett reports Research Funding from Takeda (Shire): Investigator Initiated Grant for genomic screening of Gaucher disease Scientific Advisor or Membership as a Member of Genetics Advisory Committee, Royal College of Pathologists of Australasia (RCPA), Member of Genetics Services Executive Committee, NSW Health Genetics Network, Member of Community Genetics NSW Advisory Committee, These roles are honorary; Other Interests/Relationships as a Member of KidGen Renal Genetics Flagship of Australian Genomics, Member of CIRCA Clinical Immunogenomics Research Consortium Australasia, Non-Executive Director, AACB Services Pty Ltd, Australasian Association of Clinical Biochemists, These roles are honorary. The remaining authors have no disclosures.

**Study Group/Organization Name:** CUST_STUDY_GROUP/ORGANIZATION_NAME :No data available.

**Study Group Members' Names:** CUST_STUDY_GROUP_MEMBERS :No data available.

**Total number of words:** 3800

**Abstract:** Background: The prevalence of Alport syndrome varies from one in 5,000 to one in 53,000. This study estimated the frequencies of predicted pathogenic COL4A3â€“ COL4A5 variants in sequencing databases of populations without known kidney disease.

Methods: Predicted pathogenic variants were identified using filtering steps based on the ACMG/AMP criteria that considered collagen IV Î±3â€“Î±5 position 1 Gly to be critical domains. The population frequencies of predicted pathogenic COL4A3â€“COL4A5 variants were then determined per mean number of sequenced alleles. Population frequencies for compound heterozygous and digenic combinations were calculated from the results for heterozygous variants.

Results: COL4A3â€“COL4A5 variants resulting in position 1 Gly substitutions were confirmed associated with haematuria (p each <0.0001). Predicted pathogenic COL4A5 variants were found in at least one in 2,320 individuals. p.(Gly624Asp), represented nearly half (16/33, 48%) the variants in Europeans. Most COL4A5 variants (54/59, 92%) had a biochemical feature that potentially mitigated clinical impact.

Predicted pathogenic heterozygous COL4A3 and COL4A4 variants affected one in 106 of the population, consistent with the finding of Thin basement membrane nephropathy in normal donor kidney biopsies. Predicted pathogenic compound heterozygous variants occurred in one in 88,866 individuals and digenic variants in at least one in 44,793.

Conclusions: The population frequencies for Alport syndrome are suggested by the frequencies of predicted pathogenic COL4A3â€“COL4A5 variants but must be adjusted for the disease penetrance of individual variants, as well as the likelihood of already diagnosed disease and non-Gly substitutions. Disease penetrance may depend on other genetic and environmental factors.

## Significance Statement

The population frequencies of Alport syndrome vary greatly in different reports. This study examined a population sequencing database of individuals not known to have kidney disease using filtering steps corresponding to the ACMG/AMP criteria for 'predicted pathogenic' variants in *COL4A3-COL4A5* that considered collagen chain position 1 Gly residues 'critical domains.' Predicted pathogenic *COL4A5* variants occurred in at least one in 2,320 individuals. Heterozygous *COL4A3 or COL4A4* variants affected one in 106; compound heterozygous *COL4A3* or *COL4A4* variants affected one in 88,866. The actual prevalences are even greater since they also include already-diagnosed disease and other variants not examined here. The high frequency of predicted pathogenic *COL4A3- COL4A5* variants suggests that other genetic and environmental factors mitigate the corresponding clinical manifestations of disease.

# Prevalence Estimates of Predicted Pathogenic *COL4A3 – COL4A5* Variants in a Population Sequencing Database and their Implications for Alport Syndrome

[1]Joel Gibson, [2]Rachel Fieldhouse, [3,4]Melanie MY Chan, [3,4]Omid Sadeghi-Alavijeh, [2]Leslie Burnett, [5]Valerio Izzi, [6]Anton V Persikov, [4]Genomics England Research Consortium, [3,4]Daniel P Gale, [7]Helen Storey, [1]Judy Savige

[1]The University of Melbourne Department of Medicine, Melbourne Health and Northern Health, Royal Melbourne Hospital, Parkville VIC 3050 AUSTRALIA; [2]Kinghorn Centre for Clinical Genomics, Garvan Institute of Medical Research, Darlinghurst, NSW, 2010, AUSTRALIA; [3]Department of Renal Medicine, University College London, London, UK; [4]Genomics England, Queen Mary University of London, London, UK; [5]Center for Cell-Matrix Research and Biocenter Oulu, University of Oulu, FINALND;[6]Lewis-Sigler Institute for Integrative Genomics, Princeton University, New Jersey, 08544, USA; [7]Molecular Genetics, Viapath Laboratories, 5th floor Tower Wing, Guy's Hospital, London, SE1 9RT, UK.

**Short running title:** Population Frequency of Alport Syndrome

**Address for Correspondence:**
Prof Judy Savige
The University of Melbourne
Department of Medicine (Melbourne Health and Northern Health)
Royal Melbourne Hospital
Parkville VIC 3050
AUSTRALIA
Email: jasavige@unimelb.edu.au
Tel: + 613 8344 3260

## Abstract

**Background:** The prevalence of Alport syndrome varies from one in 5,000 to one in 53,000. This study estimated the frequencies of predicted pathogenic *COL4A3*–*COL4A5* variants in sequencing databases of populations without known kidney disease.

**Methods:** Predicted pathogenic variants were identified using filtering steps based on the ACMG/AMP criteria that considered collagen IV □3–□5 position 1 Gly to be critical domains. The population frequencies of predicted pathogenic *COL4A3*–*COL4A5* variants were then determined per mean number of sequenced alleles. Population frequencies for compound heterozygous and digenic combinations were calculated from the results for heterozygous variants.

**Results:** *COL4A3*–*COL4A5* variants resulting in position 1 Gly substitutions were confirmed associated with haematuria (p each <0.0001). Predicted pathogenic *COL4A5* variants were found in at least one in 2,320 individuals. p.(Gly624Asp), represented nearly half (16/33, 48%) the variants in Europeans. Most *COL4A5* variants (54/59, 92%) had a biochemical feature that potentially mitigated clinical impact.

Predicted pathogenic heterozygous *COL4A3* and *COL4A4* variants affected one in 106 of the population, consistent with the finding of thin basement membrane nephropathy in normal donor kidney biopsies. Predicted pathogenic compound heterozygous variants occurred in one in 88,866 individuals and digenic variants in at least one in 44,793.

**Conclusions:** The population frequencies for Alport syndrome are suggested by the frequencies of predicted pathogenic *COL4A3*–*COL4A5* variants, but must be adjusted for the disease penetrance of individual variants as well as for the likelihood of already diagnosed disease and non-Gly substitutions. Disease penetrance may depend on other genetic and environmental factors.

## Introduction

Alport syndrome is an inherited renal disease characterised by progressive kidney failure, hearing loss, lenticonus and fleck retinopathy[1]. X-linked (XL) inheritance results from pathogenic variants in *COL4A5*[2], and autosomal recessive (AR) disease from two pathogenic variants in *COL4A3* or *COL4A4 in trans*[3]. The *COL4A3* – *COL4A5* genes code for the collagen IV □□, □□ and □□ chains that trimerise to form a network that represents the major component of the basement membranes of the glomerulus, cochlea, lens and retina.

Individuals with heterozygous pathogenic *COL4A3* or *COL4A4* variants are carriers for AR Alport syndrome and sometimes diagnosed with Thin basement membrane nephropathy[4, 5], even without a renal biopsy being performed. The term 'autosomal dominant' (AD) Alport syndrome is also used but affected individuals typically have persistent haematuria and not the hearing loss, ocular abnormalities, or typical glomerular basement membrane (GBM) lamellation[6]. In contrast to XL and AR disease, the risk of renal failure for heterozygous pathogenic *COL4A3* or *COL4A4* variants is uncertain because most reported series comprise only hospital-based patients[7]. Digenic Alport syndrome occurs with heterozygous pathogenic variants in two different *COL4A3* – *COL4A5* genes[8], and clinical features are more severe than with only a heterozygous *COL4A3* or *COL4A4* variant [8]. In addition, any variant or combination of variants in *COL4A3* – *COL4A5* may result in focal and segmental glomerulosclerosis probably secondary to a defective GBM and podocyte loss[9, 10].

Alport syndrome has been considered rare because it affects fewer than one in 2,000 individuals[11], and the US has less than 200,000 cases[7]. Its exact prevalence is not known [12]. Population-based studies have suggested frequencies of one in 5,000 in Utah[13]; one in 17,000 in southern Sweden[14]; and one in 53,000 in Finland[15]. However persistent haematuria and GBM thinning that correspond to heterozygous *COL4A3* or *COL4A4* variants occur in 1- 2.5% of children and adults on population screening[16, 17], and, depending on the definition of thinning, in 4 % to 7% of normal donor kidney transplant biopsies[18] respectively.

Alport syndrome is suspected where persistent haematuria occurs together with a family history of haematuria or renal impairment; or an abnormal GBM appearance or collagen IV chain composition[19]. Haematuria is highly penetrant[20, 21] but many affected individuals are undiagnosed. These are often men with X-linked disease and late onset kidney failure, or where there is no family history or the kidney biopsy is too scarred to demonstrate the characteristic features. Although women with X-linked disease are affected twice as often as men their recognition is often even more problematic[22], because they often have only haematuria and a thinned rather than lamellated GBM[22].

The most sensitive method for the diagnosis of Alport syndrome is genetic testing[19, 23] with Massively Parallel Sequencing, which detects causative variants in up to 90% of cases[23]. Identifying likely pathogenic variants in the *COL4A3-COL4A5* genes in a dataset of individuals not known to have kidney disease, such as gnomAD, gnomAD controls, Exome Variant Server (EVS) or TOPMed, indicates more accurately the population frequencies of undiagnosed *COL4A5*, heterozygous *COL4A3* or *COL4A4* variants, and hence compound heterozygous and digenic disease. However these databases share some datasets and include no clinical information. Conversely, the

Genomics England 100,000 Genomes Project is an independent dataset that includes clinical information but its recruitment of individuals with inherited kidney disease precludes its use in determining population frequencies.

Genomic variants are usually assessed for pathogenicity using the American College of Medical Genetics and Genomics/Association for Molecular Pathology (ACMG/AMP) criteria[24]. Nonsense, canonical splice site variants and frameshift changes in the *COL4A3 – COL4A5* genes are likely to be pathogenic since loss of function is a recognised disease mechanism for these genes[24]. However the commonest variants are missense (databases.lovd.nl/ shared/genes/). The interpretation of missense variants from a population not known to have kidney disease is complicated by the absence of clinical, segregation, and functional data **(Suppl Table 1),** and compounded by the lack of expert panel assertions for most variants (https://clinvarminer.genetics.utah.edu/).

However affecting a 'critical domain' also represents one of the ACMG/AMP criteria for pathogenicity (PM1)[24], and in osteogenesis imperfecta, position 1 Gly residues in the collagen I □1 chain Gly-Xaa-Yaa repeats are considered critical because of their disruptive effect on the molecule[25]. Gly is the smallest amino acid, position 1 Gly residues are found within the interior of the collagen triple helix, and their replacement by any larger residues disturbs the helix formation. The commonest reported missense variants in *COL4A3 - COL4A5* result in position 1 Gly substitutions in the intermediate domain of the collagen IV □3, □4 and □5 chains[12, 25], and the recent Chandos House meeting of the Alport Variant Consortium recommended that these are generally considered pathogenic[26].

There are nevertheless also important differences in the biochemistry of collagen IV and I. The collagen IV chains retain their non-collagenous amino and carboxy termini, and include multiple non-collagenous interruptions, in the Gly-Xaa-Yaa repeats of the intermediate domains **(Suppl Table 2)**[27-29]. Gly substitutions in the amino and carboxy termini, the interruptions and positions 2 and 3 in the Gly-Xaa-Yaa- repeats[29] are not considered critical[25, 30].

This study used the UK 100,000 Genomes Project to confirm that predicted pathogenic position 1 Gly substitutions in the collagen IV □3, □4 and □5 chains were associated with haematuria. It then identified predicted pathogenic *COL4A5* variants in the large, unbiased gnomAD2.1.1 dataset and confirmed the strategy using the gnomad control, EVS and TOPMed cohorts. The amount of overlap between the datasets was determined from the number of shared variants. A similar approach has been used previously to determine the population frequency of polycystic kidney and liver disease[31].

This study then identified predicted pathogenic *COL4A3* and *COL4A4* variants in gnomAD, and used these to calculate the frequencies of compound heterozygous and digenic variants. These population frequencies provide an approximate guide to the prevalence of XL, AR (compound heterozygous) and digenic Alport syndrome and Thin basement membrane nephropathy but the precise prevalence must also take into account the penetrance of individual variants as well as the occurrence of already diagnosed disease and missense variants due to non-Gly substitutions.

# Methods

## Datasets

**gnomAD**. The Genome Aggregation Database (gnomADv2.1.1, www.gnomAD.broadinstitute.org)[32] is a large dataset that includes *COL4A3 – 5* alleles from unrelated adults with equal numbers of males and females who have undergone exomic (in about 90%) or genomic sequencing in disease and population studies. It includes individuals with adult-onset diseases such as diabetes, cardiac disease and psychiatric conditions, and excludes those with severe paediatric diseases and their close family members. Participants were not known to have inherited kidney disease and individuals with any severe disease were expected less frequently than in the general population. gnomAD indicates variants that are homozygous but variants are not otherwise linked to an individual and it is not possible to identify a different second variant in *COL4A3 – COL4A5* and hence compound heterozygous or digenic variants directly. Ancestries are available but the clinical phenotypes are not recorded.

Three other datasets were also examined for predicted pathogenic variants in *COL4A5*. These were the **gnomAD controls** (vs2.1.1, www.gnomAD.broadinstitute.org**)** which comprises alleles from the normal individuals, recruited as controls for gnomAD studies; **Exome Variant Server (EVS)** (www.evs.gs.washington.edu/EVS/) which includes alleles from exomic sequencing of individuals recruited for mainly cardiac and pulmonary studies; and **TOPMed** (www.bravo.sph.umich.edu/freeze8/hg38/) with alleles from genomic sequencing of individuals recruited for heart, lung, blood and sleep studies. Although participants in these cohorts were not known to have inherited kidney disease, they had not been specifically phenotyped for renal disease and it was uncertain how many had underlying renal disease or impaired kidney function. The gnomAD dataset includes all the 'gnomAD control' cohort, and some of the studies in the other cohorts. The amount of overlap was assumed to include the *COL4A3 – COL4A5* variants randomly and was quantitated from the number of shared variants with the same gender.

In contrast, **the Genomics England 100,000 Genomes Project** database (www.genomicsengland.co.uk/)[33] includes whole genome sequencing data from individuals with rare diseases including inherited kidney disease and their unaffected relatives. This dataset also includes clinical phenotypes including haematuria noted in the hospital records. This database was used to examine whether substitutions affecting position 1 Gly substitutions in the collagen IV □3, □4 and □5 chains were associated with haematuria.

Ethics approval was obtained at the time of recruitment into participation in each of the constituent studies. The databases are freely-accessible online and were examined between 24 December 2019 and 1 February 2021.

## *COL4A3-5* variants

*COL4A5* variants were described using the GRCh37/hg19 sequence with 53 exons and the reference sequence for the collagen IV □5 chain (NM_033380.2, LRG_232t2). *COL4A3* and *COL4A4* variants were described using the GRCh37/hg19 sequence and the reference sequences for collagen IV α3 and α4 chains (NM 000091.4, LRG_230t1) and (NM 000092.4, LRG_231t1,) respectively.

The locations of the non-collagenous termini and interruptions in the collagenous domains of each chain were noted **(Suppl Table 2)[27-29].**

**Filtering steps to determine predicted pathogenic variants based on the ACMG/AMP criteria**

Variants were filtered to exclude variants that failed any step in order to be confident of the corresponding population frequencies. Variants were filtered according to their effect on the canonical transcripts using the following steps and ACMG criteria **(Figure 1, Suppl Table 1)[34]**. The term 'predicted pathogenic' was used for variants that were not excluded, to differentiate them from the 'pathogenic' and 'likely pathogenic' classes derived from the ACMG/AMP classification in individuals with clinical features of disease.

- Synonymous variants and intronic variants other than those affecting a +/-1 or 2 splice site were excluded.
- Variants that resulted in a protein-truncating change, frameshift, or canonical splice site change were graded **PVS1** because loss of function is a known disease mechanism for *COL4A3 – COL4A5*-associated disease[24]**.**
- Missense variants resulting in position 1 Gly substitutions within the intermediate collagenous domains of the collagen IV □□□□□□ chains since missense variants are a common mechanism for *COL4A3 – COL4A5*-associated disease with a low rate of benign missense variation **(PP2).** Variants that resulted in a position 1 Gly substitution were considered to affect a critical and well-established functional domain **(PM1)**.
- Missense variants were examined with individual computational tools: Polyphen-2 ( PP2, score >0.8, www.genetics.bwh.harvard.edu/pph2/); SIFT (Sorting Intolerant from Tolerant, Deleterious or Polymorphism, www.sift.bii.a-star.edu.sg/); and Mutation Taster, (MT, Disease-causing, or Polymorphism, www.mutationtaster.org/). Missense variants were also examined for conservation among vertebrates (www.asia.ensembl.org/) **(PP3)**.
- Variants occurred fewer than an arbitrary cut off of 30 times in gnomAD **(PM2)[34]**. The formula used to determine the cutoff is normally based on the disease prevalence and contribution of the gene which were not known[35]. While the cut-off varies for different genes, this level has been used for other heterogeneous genes resulting in AD kidney disease.
- Variants were assessed for previous reports as Likely Pathogenic or Pathogenic in a reputable source such as ClinVar (www.ncbi.nlm.nih.gov/clinvar/) or LOVD (www.databases.lovd.nl/shared/genes/COL4A3-5 **(PP5)** despite this practice being controversial[36, 37]. Variants in ClinVar reported as Benign, Likely Benign, Variant of Uncertain Significance or Conflicting, were excluded. Variants in LOVD reported as Benign or Likely Benign were excluded. Variants that resulted in the same amino acid change as a previously-established pathogenic variant regardless of nucleotide change in ClinVar (www.ncbi.nlm.nih.gov/clinvar/) or LOVD (www.databases.lovd.nl/shared/genes/COL4A3-5 were also noted (**PS1).** In these cases the submitters' evaluations were accepted since their assessments included the associated clinical data, that we were not able to access. Anyway all variants were assessed independently for pathogenicity using an objective online tool (Varsome).

- Variants were examined to determine if they resulted in a novel missense change at an amino acid where a different change was previously reported as pathogenic **(PM5)**.

All *COL4A3-COL4A5* variants were also assessed by Varsome (www.varsome.com/) using the ACMG/AMP criteria[38], and *COL4A5* variants in gnomAD were assessed in Alamut (www.interactive-biosoftware.com/alamut-visual/). These semi-automated online tools were chosen because they are objective, not subject to individual laboratory practice, and are widely available. They used slightly different criteria from our filtering steps, for example, numerous computational tools, GERP scores for conservation, and weighting of scores. Variants assessed as Benign, Likely Benign or a Variant of Uncertain Significance (VUS) by Varsome or Alamut were excluded.

**Predicted pathogenic *COL4A3 – COL4A5* variants resulting in position 1 Gly substitutions in the 100,000 Genomes Project database and haematuria**

Individuals in the 100,000 Genomes Project database with haematuria were identified from haematuria-related Human Phenotype Ontology terms after excluding those with cancer of the kidney or bladder, or kidney stones. Predicted pathogenic variants resulting in a position 1 Gly substitution in *COL4A3 - COL4A5* were identified from all Gly substitutions in these genes using our filtering strategy, and the total number of individuals with a predicted pathogenic variant were compared in the cohorts with haematuria and without (chi square with Yates' correction, two tailed, Graphpad).

**Prevalence of predicted pathogenic variants in *COL4A5, COL4A3* and *COL4A4***

The population frequencies of predicted pathogenic variants were calculated based on the mean number of alleles examined. For *COL4A5*, which is located on the X chromosome, the corresponding number of individuals was determined from the number of alleles after correcting for the number of males since they have only one allele for *COL4A5*. The total number of individuals recruited in the whole cohort was greater than the number calculated from the number of alleles examined because some regions of a gene were not sequenced or failed quality testing.

**Prevalence of variants in different ancestries and founder variants**

The frequencies of predicted pathogenic *COL4A5, COL4A3* and *COL4A4* variants were noted in people of different ancestries (African, Latino, Ashkenazi Jew, East Asian, European Finnish, European non-Finnish, South Asians and Others, who did not unambiguously cluster with any of the major populations by Principal Component Analysis). Any founder variants (arbitrarily set at $\geq$ 18 alleles) were noted.

**Biochemical features likely to mitigate clinical phenotypes for predicted pathogenic *COL4A5* variants**

The cohort with predicted pathogenic *COL4A5* variants were then assessed for any biochemical features associated with clinically milder phenotypes that might explain the lack of previous diagnosis. Mitigating features included female gender;[22] location within the 20 amino terminal exons including the amino non-collagenous domain (204 residues)[39]; position adjacent to a non-collagenous domain or interruption [30]; and Gly substitutions with a less disruptive residue, such as Ala, Ser or Cys [40, 41].

**Calculation of population frequencies for heterozygous, compound heterozygous and digenic *COL4A3* and *COL4A4* variants**

The population frequencies of predicted pathogenic heterozygous *COL4A3* and *COL4A4* variants were derived from the sum of the allele frequencies for each variant assuming that they were inherited independently and that no or very few individuals had inherited two.

The population frequency of AR Alport syndrome represents the sum of the frequencies of predicted pathogenic homozygous and compound heterozygous variants *in trans* in *COL4A3* plus *COL4A4*. Previous reports suggest homozygous inheritance represents 30% of all AR Alport syndrome[42]. Homozygous variants are indicated in gnomAD and usually result from relationships that are consanguineous or consanguineous by descent. However in Alport syndrome homozygous predicted pathogenic variants result in early onset kidney failure which would have precluded recruitment into gnomAD. The likelihood of individual homozygous *COL4A3 or COL4A4* variants occurring by chance was very small and not calculated here.

It was not possible to identify compound heterozygous and digenic variants from the databases because variants were not linked to any individual. In addition compound heterozygous and digenic variants were likely to be underrepresented because they too are associated with early onset kidney failure. Thus the population frequencies of predicted pathogenic compound heterozygous variants in *COL4A3* or *COL4A4* were calculated simply from the likelihood of the occurrence of each allele occurring independently.

Digenic variants were calculated from two predicted pathogenic variants in *COL4A3* and *COL4A4 in cis* or *trans* and from a predicted pathogenic variant in *COL4A5* and either *COL4A3* or *COL4A4*.

## Results

### Predicted pathogenic *COL4A3- COL4A5* variants resulting in position 1 Gly substitutions in the 100,000 Genomes Project database and haematuria

This cohort comprised 2,221 individuals with documented haematuria and 37,200 with none. It included 19 predicted pathogenic *COL4A5* variants that affected a position 1 Gly and were associated with haematuria and 9 that were not, consistent with an association with haematuria (chi-square 95.83, p<0.0001). *COL4A3* and *COL4A4* variants resulting in predicted pathogenic position 1 Gly substitutions were also associated with haematuria (p <0.0001 for each, **Table 1, Suppl Tables 3 - 5)**.

## *COL4A5*

### Predicted pathogenic *COL4A5* variants in gnomAD

Fifty-four variants in *COL4A5* (frameshift, nonsense and canonical splice site variants, position 1 Gly substitutions) were identified for further assessment **(Table 2)**. These included one deletion resulting in a termination codon and 6 canonical splice site variants, all predicted pathogenic and found in 8 individuals. There were also 47 position 1 Gly substitutions, but 13 were excluded on filtering (including p.(Gly953Val), Suppl Table 3), resulting in 34 Gly substitutions in 51 individuals **(Table 2)**.

Thus, gnomAD included 59 individuals with a predicted pathogenic *COL4A5* variant in a mean allele number of 170,190 (113,460 individuals) or one in 1,923 individuals. If the population frequency were calculated from the total gnomAD cohort of 136,920 individuals at the time of examination, then the population frequency was one in 2,320.

Two of the predicted pathogenic variants found in gnomAD (p.(Gly624Asp and p.(Gly752Val)) were also present in the 100,000 Genomes Project database but none of the 4 individuals with these variants had haematuria.

### Population frequencies of predicted pathogenic *COL4A5* variants in different ancestries in gnomAD

The population frequencies of predicted pathogenic *COL4A5* variants differed in people of different ancestries **(Table 3)**. Predicted pathogenic *COL4A5* variants were more common in people of European (one in 1,800), African (1 in 1,733) or East Asian (1 in 2,310) ancestry. In general, *COL4A5* variants were less common in Ashkenazim (1 in 4,961) and not found in Finns.

The commonest predicted pathogenic variant, p.(Gly624Asp), was considered Pathogenic by ClinVar**and has been reported pathogenic many times[43]. It was present 16 times in 59,512 individuals of European ancestry corresponding to a prevalence of one in 2,479. Nearly half (16/33) of all Europeans with a predicted pathogenic *COL4A5* variant had p.(Gly624Asp) and it was not present in other ancestries. This variant is associated with late onset kidney failure in men [44, 45], and a 50% cumulative probability of end-stage kidney failure by the age of 54, which is nearly 30 years later than with other variants[46]. The variant is located immediately adjacent to an interruption in the collagenous sequence and was found in 12 females and only 4 males, which are both factors that helped explain its association with a milder phenotype and why it had not been detected previously in individuals recruited into gnomAD.

**Founder *COL4A5* variants in gnomAD**
Most *COL4A5* variants resulting in a Gly substitution were found once only but five were found more often, usually in a single ancestral group: p.(Gly624Asp) (n=16) and p.(Gly626Ser)(n=7) in Europeans; p.(Gly961Val)(n=5) and p.(Gly1074Ser)(n=7) in Latinos; and p.(Gly953Val) in South and East Asians. All these variants are located adjacent to a non-collagenous interruption, and are less likely to be disruptive than other position 1 Gly substitutions **(Suppl Table 2)**.

**Predicted pathogenic *COL4A5* variants in other cohorts**
In the normal control subset of gnomAD comprising a mean of 87,876 alleles or 58,584 individuals, 24 individuals had 17 predicted pathogenic variants, occurring in one in 2,441 individuals **(Table 2)**. All these variants were also found in the whole gnomAD cohort as expected.

In the EVS dataset comprising a mean of 10,563 alleles or 7,042 individuals, 3 individuals had predicted pathogenic variants, corresponding to a population frequency of one in 2,347 **(Suppl Table 6)**. Five of the 8 *COL4A5* variants found in the EVS dataset were also present with a consistent gender in gnomAD suggesting at most an overlap of 5/8 (63%) variants.

In the TOPMed cohort of a mean of 125,568 alleles or 83,713 individuals, 40 individuals had predicted pathogenic variants, corresponding to a population frequency of one in 2,093 **(Suppl Table 7)**. Only two of the 45 *COL4A5* variants found in the TOPMed dataset (p.(Gly524Asp) and p.(Gly1424Ser)) were also present with a consistent gender in gnomAD suggesting very little if any overlap in the datasets. The population frequency of predicted pathogenic variants *COL4A5* variants from TOPMed might therefore be considered independent and confirmatory of the gnomAD results.

Thus, the population frequencies of predicted pathogenic *COL4A5* variants varied from one in 2,320 (based on total participants recruited) or one in 1,923 (based on alleles examined) in gnomAD, one in 2,441 in the gnomAD normal controls, one in 2,347 in EVS, and one in 2,093 in TOPMed.

**Mitigating features for phenotypes of predicted pathogenic *COL4A5* variants in gnomAD**
Features that might mitigate the clinical features, for example, result in late onset kidney failure were found for 54 of the 59 *COL4A5* variants (92%). These included being identified in a woman (41/59, 69%); and for Gly substitutions: being located in the 20 amino exons (11/51, 22%) or adjacent to a non-collagenous interruption (23/51, 45%); or a Gly substitution with a less disruptive residue (19,37%)**(Suppl Table 8)**.

## *COL4A3*

**Predicted pathogenic *COL4A3* variants in gnomAD**
Five hundred and seventy-four variants (2.6%) were identified in *COL4A3* in gnomAD for further assessment. Fifteen position 1 Gly substitutions were excluded because evidence was not consistent with pathogenicity **(Suppl Table 9)**. Overall there were 559 predicted pathogenic variants in a mean of 245,889 (0.23%) alleles

corresponding to a population frequency of 0.45%. These variants were position 1 Gly substitutions (n=380, in 68% of variants, 0.16% of the population), frameshift (n=110, 20% of the variants, 0.04%), nonsense (n=33, 6% of the variants, 0.01%) or splicing variants (n=36, 6% of the variants, 0.01%). No predicted pathogenic *COL4A3* variant was found in the homozygous form.

In the total cohort of 136,930 individuals, the population frequency of a predicted pathogenic *COL4A3* variant was 0.41%.

**Population frequencies of predicted pathogenic *COL4A3* variants in different ancestries**

Most predicted pathogenic *COL4A3* variants were present in a single ancestry. Predicted pathogenic heterozygous *COL4A3* variants were commonest in people of Latino (0.80%) or East Asian (0.64%) ancestries, and least common in Askenazim (0.14%) and Finns (0.14%) **(Table 4)**.

## *COL4A4*

**Predicted pathogenic *COL4A4* variants in gnomAD**

Five hundred and ninety-nine variants were identified in *COL4A4* in gnomAD for further assessment. Twenty-two position 1 Gly substitutions were excluded because evidence was not consistent with pathogenicity **(Suppl Table 9)**. The variants p.(Gly545Ala) and p.(Gly999Glu) were considered Benign and are discussed in the footnote to Suppl Table 9.

In *COL4A4*, there were 577 predicted pathogenic variants in in a mean of 233,916 alleles (0.25% alleles) corresponding to a population frequency of 0.49%. Variants were position 1 Gly substitutions (n=458, 82% of all variants, in 0.20% of the population), frameshift (n=42, 0.02%), nonsense (n=48, 0.02%) or splicing variants (n=29, 0.01%). No homozygous predicted pathogenic *COL4A4* variant was found.

In the total cohort of 136,930 individuals, the population frequency of a predicted pathogenic *COL4A4* variant was 0.42%.

**Population frequencies of predicted pathogenic *COL4A4* variants in different ancestries**

Heterozygous predicted pathogenic *COL4A4* variants were commonest in people of Latino (0.78%) and East Asian (0.70%) ancestries, and least common in Finns (0.10%) and Ashkenazim (0.18%) **(Table 4)**.

Most founder variants were found once only in any ancestral group, but there were also 7 more common *COL4A4* founder variants, which were all position 1 Gly substitutions **(Suppl Table 10)**. These were mainly found in one population each including Africans (p.(Gly445Ala), Latinos (p.(Gly481Ser), East Asians (p.(Gly816Glu), non-Finnish Europeans (p.(Gly1178Ser), p.(Ser969Ter) and Others (p.(Gly774Arg)). No founder variants were found in Ashkenazim or Finns. p.(Ser969Ter) was common in Europeans, occurring in 18 alleles of the 280,956 examined or one in 15,608 individuals but otherwise no founder variants were frameshift or splicing variants.

**Population frequencies of heterozygous predicted pathogenic *COL4A3* and *COL4A4* variants**

The population frequency of heterozygous predicted pathogenic variants in *COL4A3* or *COL4A4* was 0.94% for all individuals **(Suppl Table 11)**.

Heterozygous predicted pathogenic *COL4A3* and *COL4A4* variants were commonest in people of Latino (in 1.58%), East Asian (1.34%) or African (1.28%) ancestries, and least common in Finns (0.24%) and Ashkenazim (0.32%).

**Association of heterozygous predicted pathogenic *COL4A3* and *COL4A4* variants in gnomAD with haematuria in the 100,000 Genomes Project dataset**

Seventeen *COL4A3* and 27 *COL4A4* predicted pathogenic Gly substitutions found in gnomAD were also present in the 100,000 Genomes database. Fourteen individuals of the 2,221 with haematuria had one of these variants and 78 of the 37,200 without haematuria also had one (chi squared 14.17, p=0.0002) which indicated that these variants were associated with haematuria. This confirmed that the heterozygous predicted pathogenic *COL4A3* and *COL4A4* variants resulting in Gly substitutions found in in gnomAD were associated with haematuria.

**Calculated population frequencies of compound heterozygous and digenic *COL4A3* and *COL4A4* variants**

The population frequency of 0.94% for heterozygous predicted pathogenic *COL4A3* or *COL4A4* variants corresponded to about 0.9% of the population or one in 106 people.

The population frequency of two compound heterozygous predicted pathogenic variants in *COL4A3* **or** *COL4A4* was $1.13 \times 10^{-5}$, or 0.001%, or 1 in 88,866 when the variants occurred independently.

The population frequency of a digenic predicted pathogenic *COL4A3* **and** a *COL4A4* variant was $2.23 \times 10^{-5}$ or 0.002% or 1 in 44,793 if each variant were *in cis* or *in trans* and the variants occurred independently. Digenic predicted pathogenic variants in *COL4A5* plus *COL4A3 or COL4A4* were much less common occurring in about one in 2,320 (for *COL4A5*) x one in 106 (for *COL4A3* or *COL4A4*) or one in 245,920 of the population.

## Discussion

Previous estimates of the population frequency of X-linked Alport syndrome have differed[13-15], but examination of a large variant database from people without known kidney disease predicted pathogenic *COL4A5* variants in at least one in 2,320 individuals and heterozygous pathogenic *COL4A3* or *COL4A4* variants of one in 106. The accuracy of population frequencies for these variants is suggested by the consistency of predicted pathogenic *COL4A5* variants in a control subset, an overlapping cohort, and an independent dataset. However the population frequencies of the corresponding diseases while directly related to the prevalence of the predicted pathogenic *COL4A3 – COL4A5* variants, must be corrected for the penetrance of clinical features for individual variants. Penetrance probably depends on other genetic and other factors, and we demonstrated for most *COL4A5* variants that there were mitigating features that explained why variants had not been detected previously.

The population frequency of one in 106 for heterozygous predicted pathogenic variants in *COL4A3* and *COL4A4* corresponded to a population frequency for compound heterozygous *COL4A3* **or** *COL4A4* variants of one in 88,866, and of digenic variants in each of *COL4A3* **and** *COL4A4* of one in 44,793, and a lesser contribution from *COL4A5* and *COL4A3* or *COL4A4.* The calculated frequency for compound heterozygous variants was less than expected for AR Alport syndrome[7] which previous reports have suggested affects one in 6 Alport families[47], or one in 40,000 of the population [7,48]. Recessive Alport syndrome results from compound heterozygous variants that occur by chance and from homozygous variants due to consanguinity or consanguinity by descent which account for a further 30% of cases[42]. Thus the overall population frequency for AR Alport syndrome must also take into account a further 30% from homozygous variants. Digenic *COL4A3 – COL4A5* variants occur together by chance and the calculated frequencies are more accurate.

These population frequencies of predicted pathogenic variants are still underestimates since they do not include individuals with already diagnosed disease, with the large deletions or deep intronic splicing variants that are not detected with Whole Exome Sequencing[23], or with pathogenic non- Gly substitutions, that were not evaluated [12, 49].

So what are the implications of these results for the prevalence of different modes of inheritance of Alport syndrome? In general, pathogenic *COL4A5* variants are highly penetrant for haematuria and renal failure in males[20], and for haematuria in females[21]. This study only examined high risk predicted pathogenic variants, that is, nonsense variants and position 1 Gly substitutions, but even so, many had biochemical features that potentially mitigated the clinical phenotype.

Interestingly, the population frequency of heterozygous predicted pathogenic *COL4A3* and *COL4A4* variants approximated the 1% estimated from the prevalence of persistent haematuria[16, 17] and Thin membrane nephropathy in normal donor kidney biopsies[18]. However the penetrance of persistent haematuria with *COL4A3* and *COL4A4* variants is about 70%[7], and the penetrance of a thinned GBM, FSGS and renal impairment are not known. This means that the population frequencies of Thin basement membrane nephropathy or AD Alport syndrome deduced from the prevalence of heterozygous predicted pathogenic *COL4A3* or *COL4A4* variants must be adjusted for the incomplete penetrance associated with these variants. The prevalence of autosomal recessive Alport syndrome due to compound heterozygous and of digenic Alport syndrome is probably affected less because haematuria and renal impairment are more likely with two *COL4A3 – COL4A5* variants [8, 50].

Genetic testing is the most sensitive method for the detection of *COL4A3 – 5* variants and hence for detection of the different forms of Alport syndrome[19]. Examination of data from the UK 100,000 Genomes Project confirmed that heterozygous predicted pathogenic position 1 Gly substitutions in *COL4A5*, *COL4A3* or *COL4A4* were each associated with haematuria and that these Gly residues were critical domains. While some position 1 Gly substitutions in *COL4A3 – COL4A5* were not considered pathogenic and assertions from collagen I variants in osteogenesis imperfecta suggest that nearby amino acids determine whether chain flexibility from a non-Gly substitution is tolerated[51].

The gnomAD and TOPMed datasets represented cohorts of individuals not known to have kidney disease that were used to determine the population frequencies of predicted pathogenic *COL4A3 – COL4A5* variants. Sequencing was not available for all study participants at each residue, and in general the population frequencies were calculated from the mean number of alleles examined and the deduced number of individuals. However we have cited the more conservative value of *COL4A5* variants in the whole cohort. There were also limitations in applying the ACMG/AMP criteria[24] to variants where clinical data were not available, but while our strategy may have excluded some variants that were pathogenic, these small numbers were unlikely to have significantly affected the overall frequencies.

The large number of undetected *COL4A5* variants identified in gnomAD may be explained by mitigating features that the associated clinical features were mild, such as haematuria or late onset kidney failure without extrarenal features. Thus, predicted pathogenic variants were found twice as often in women, which was consistent with XL inheritance, but different from the female representation in most previous reports[12-14]. Affected women often have only haematuria, but identifying affected women is still important because of their own risk of end-stage kidney failure, because half their sons develop kidney failure[22], and because treatment is effective[52]. Most other *COL4A5* variants identified in gnomAD were associated with a clinically-mitigating feature such as location in the amino terminus or adjacent to an interruption, or a Gly substitution with Ala, Ser or Cys.

This study also demonstrated that the population frequencies of predicted pathogenic *COL4A3 - COL4A5* variants varied in people of different ancestries. Predicted pathogenic variants in *COL4A5* were more common in people from a European or African background, present at a low level in Ashkenazim and absent from Finns. This was consistent with the previously-reported low population frequency in Finns [15]. In addition, heterozygous predicted pathogenic *COL4A3* and *COL4A4* variants affected more than 1% of people of Latino, African and East Asian ancestries, which was at least four times more common than in Ashkenazim and Finns. It may not be that *COL4A3 – COL4A5* variants are more common in some ancestries but rather that they are uncommon in others due to geographic and cultural isolation[53].

In conclusion, predicted pathogenic *COL4A5* variants occur at least one in 2,320 of the population but are often unrecognised. Heterozygous predicted pathogenic *COL4A3 or COL4A4* variants occur in at least one in 106 of the population. The true prevalence of potentially pathogenic *COL4A3 – COL4A5* variants is even greater since they must also include already-diagnosed disease, variants that were not detected with exomic sequencing as well as non-Gly substitutions. However the prevalence of the different forms of inheritance of Alport syndrome depends on the penetrance of haematuria and renal impairment with these variants. Future studies should address penetrance of these clinical features.

## Author contributions

JG undertook the initial analysis and made major contributions to the initial draft. RF, LB, VI, AP and HS devised the list of interruptions, provided associated disease phenotypes, and helped in the assessments of predicted pathogenicity. MMYC, OSA and DPG helped with the acquisition of data, analysis and interpretation of data, and with revisions. VI, AP and HS also contributed to the acquisition of data, analysis and interpretation of data. JS designed the study, undertook some of the analysis, and drafted and revised the manuscript. All authors contributed to the writing, have approved the final submission and agree to be accountable for the data. JG undertook this study while an undergraduate student at the University of Melbourne.

## Disclosures

D. Gale reports Honoraria from Alexion, Otsuka; and Other Interests/Relationships as a Trustee for AlportUK. J. Savige reports Other Interests/Relationships with Alport Foundation Australia, and the PKD Australia Scientific Board. L. Burnett reports Research Funding from Takeda (Shire): Investigator Initiated Grant for genomic screening of Gaucher disease Scientific Advisor or Membership as a Member of Genetics Advisory Committee, Royal College of Pathologists of Australasia (RCPA), Member of Genetics Services Executive Committee, NSW Health Genetics Network, Member of Community Genetics NSW Advisory Committee, These roles are honorary; Other Interests/Relationships as a Member of KidGen Renal Genetics Flagship of Australian Genomics, Member of CIRCA Clinical Immunogenomics Research Consortium Australasia, Non-Executive Director, AACB Services Pty Ltd, Australasian Association of Clinical Biochemists, These roles are honorary. The remaining authors have no disclosures.

## Acknowledgements

The authors would like to thank the Genome Aggregation Database (gnomAD) and the groups that provided exome and genome variant data to this resource. A full list of contributing groups can be found at http://gnomad.broadinstitute.org/about. The authors would like to thank the NHLBI GO Exome Sequencing Project and its ongoing studies that produced and provided exome variant calls for comparison: the Lung GO Sequencing Project (HL-102923), the WHI Sequencing Project (HL-102924), the Broad GO Sequencing Project (HL-102925), the Seattle GO Sequencing Project (HL-102926) and the Heart GO Sequencing Project (HL-103010). Finally, the authors would also like to thank the hosts of the many websites used to evaluate variants for predicted pathogenicity, Varsome in particular, and the LOVD database, its contributors and administrators.

This research was also 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.

## References


1. Gubler, M, Levy, M, Broyer, M, Naizot, C, Gonzales, G, Perrin, D, Habib, R: Alport's syndrome. A report of 58 cases and a review of the literature. *Am J Med,* 70**:** 493-505, 1981.
2. Barker, DF, Hostikka, SL, Zhou, J, Chow, LT, Oliphant, AR, Gerken, SC, Gregory, MC, Skolnick, MH, Atkin, CL, Tryggvason, K: Identification of mutations in the COL4A5 collagen gene in Alport syndrome. *Science,* 248**:** 1224-1227, 1990.
3. Mochizuki, T, Lemmink, HH, Mariyama, M, Antignac, C, Gubler, MC, Pirson, Y, Verellen-Dumoulin, C, Chan, B, Schroder, CH, Smeets, HJ, et al.: Identification of mutations in the alpha 3(IV) and alpha 4(IV) collagen genes in autosomal recessive Alport syndrome. *Nat Genet,* 8**:** 77-81, 1994.
4. Buzza, M, Wilson, D, Savige, J: Segregation of hematuria in thin basement membrane disease with haplotypes at the loci for Alport syndrome. *Kidney Int,* 59**:** 1670-1676, 2001.
5. Buzza, M, Dagher, H, Wang, YY, Wilson, D, Babon, JJ, Cotton, RG, Savige, J: Mutations in the COL4A4 gene in thin basement membrane disease. *Kidney Int,* 63**:** 447-453, 2003.
6. Savige, J: Should We Diagnose Autosomal Dominant Alport Syndrome When There Is a Pathogenic Heterozygous COL4A3 or COL4A4 Variant? *Kidney Int Rep,* 3**:** 1239-1241, 2018.
7. Savige, J, Rana, K, Tonna, S, Buzza, M, Dagher, H, Wang, YY: Thin basement membrane nephropathy. *Kidney Int,* 64**:** 1169-1178, 2003.
8. Mencarelli, MA, Heidet, L, Storey, H, van Geel, M, Knebelmann, B, Fallerini, C, Miglietti, N, Antonucci, MF, Cetta, F, Sayer, JA, van den Wijngaard, A, Yau, S, Mari, F, Bruttini, M, Ariani, F, Dahan, K, Smeets, B, Antignac, C, Flinter, F, Renieri, A: Evidence of digenic inheritance in Alport syndrome. *J Med Genet,* 52**:** 163-174, 2015.
9. Voskarides, K, Damianou, L, Neocleous, V, Zouvani, I, Christodoulidou, S, Hadjiconstantinou, V, Ioannou, K, Athanasiou, Y, Patsias, C, Alexopoulos, E, Pierides, A, Kyriacou, K, Deltas, C: COL4A3/COL4A4 mutations producing focal segmental glomerulosclerosis and renal failure in thin basement membrane nephropathy. *J Am Soc Nephrol,* 18**:** 3004-3016, 2007.
10. Wickman, L, Hodgin, JB, Wang, SQ, Afshinnia, F, Kershaw, D, Wiggins, RC: Podocyte Depletion in Thin GBM and Alport Syndrome. *PLoS One,* 11**:** e0155255, 2016.
11. Baldovino, S, Moliner, AM, Taruscio, D, Daina, E, Roccatello, D: Rare Diseases in Europe: from a Wide to a Local Perspective. *Isr Med Assoc J,* 18**:** 359-363, 2016.
12. Hertz, JM, Thomassen, M, Storey, H, Flinter, F: Clinical utility gene card for: Alport syndrome - update 2014. *Eur J Hum Genet,* 23, 2015.
13. Hasstedt, SJ, Atkin, CL: X-linked inheritance of Alport syndrome: family P revisited. *Am J Hum Genet,* 35**:** 1241-1251, 1983.
14. Persson, U, Hertz, JM, Wieslander, J, Segelmark, M: Alport syndrome in southern Sweden. *Clin Nephrol,* 64**:** 85-90, 2005.
15. Pajari, H, Kaariainen, H, Muhonen, T, Koskimies, O: Alport's syndrome in 78 patients: epidemiological and clinical study. *Acta Paediatr,* 85**:** 1300-1306, 1996.
16. Dodge, WF, West, EF, Smith, EH, Bruce, H, 3rd: Proteinuria and hematuria in schoolchildren: epidemiology and early natural history. *J Pediatr,* 88**:** 327-347, 1976.
17. Ritchie, CD, Bevan, EA, Collier, SJ: Importance of occult haematuria found at screening. *Br Med J (Clin Res Ed),* 292**:** 681-683, 1986.
18. Dische, FE, Anderson, VE, Keane, SJ, Taube, D, Bewick, M, Parsons, V: Incidence of thin membrane nephropathy: morphometric investigation of a population sample. *J Clin Pathol,* 43**:** 457-460, 1990.
19. Savige, J, Gregory, M, Gross, O, Kashtan, C, Ding, J, Flinter, F: Expert guidelines for the management of Alport syndrome and thin basement membrane nephropathy. *J Am Soc Nephrol,* 24**:** 364-375, 2013.
20. Jais, JP, Knebelmann, B, Giatras, I, De Marchi, M, Rizzoni, G, Renieri, A, Weber, M, Gross, O, Netzer, KO, Flinter, F, Pirson, Y, Verellen, C, Wieslander, J, Persson, U, Tryggvason, K, Martin, P, Hertz, JM, Schroder, C, Sanak, M, Krejcova, S, Carvalho, MF, Saus, J, Antignac, C, Smeets, H, Gubler, MC: X-linked Alport syndrome: natural history in 195 families and genotype- phenotype correlations in males. *J Am Soc Nephrol,* 11**:** 649-657, 2000.
21. Jais, JP, Knebelmann, B, Giatras, I, De Marchi, M, Rizzoni, G, Renieri, A, Weber, M, Gross, O, Netzer, KO, Flinter, F, Pirson, Y, Dahan, K, Wieslander, J, Persson, U, Tryggvason, K, Martin, P, Hertz, JM, Schroder, C, Sanak, M, Carvalho, MF, Saus, J, Antignac, C, Smeets, H, Gubler, MC: 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. *J Am Soc Nephrol,* 14**:** 2603-2610, 2003.
22. Savige, J, Colville, D, Rheault, M, Gear, S, Lennon, R, Lagas, S, Finlay, M, Flinter, F: Alport Syndrome in Women and Girls. *Clin J Am Soc Nephrol,* 11**:** 1713-1720, 2016.

23. Savige, J, Ariani, F, Mari, F, Bruttini, M, Renieri, A, Gross, O, Deltas, C, Flinter, F, Ding, J, Gale, DP, Nagel, M, Yau, M, Shagam, L, Torra, R, Ars, E, Hoefele, J, Garosi, G, Storey, H: Expert consensus guidelines for the genetic diagnosis of Alport syndrome. *Pediatr Nephrol*, 2018.
24. Richards, S, Aziz, N, Bale, S, Bick, D, Das, S, Gastier-Foster, J, Grody, WW, Hegde, M, Lyon, E, Spector, E, Voelkerding, K, Rehm, HL, Committee, ALQA: 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,* 17**:** 405-424, 2015.
25. Myllyharju, J, Kivirikko, KI: Collagens and collagen-related diseases. *Ann Med,* 33**:** 7-21, 2001.
26. Savige, J, Storey, H, Watson, E, Hertz, JM, Deltas, C, Renieri, A, Mari, F, Hilbert, P, Plevova, P, Byers, P, Cerkauskaite, A, Gregory, M, Cerkauskiene, R, Ljubanovic, DG, Becherucci, F, Errichiello, C, Massella, L, Aiello, V, Lennon, R, Hopkinson, L, Koziell, A, Lungu, A, Rothe, HM, Hoefele, J, Zacchia, M, Martic, TN, Gupta, A, van Eerde, A, Gear, S, Landini, S, Palazzo, V, Al-Rabadi, L, Claes, K, Corveleyn, A, Van Hoof, E, van Geel, M, Williams, M, Ashton, E, Belge, H, Ars, E, Bierzynska, A, Gangemi, C, Lipska-Zietkiewicz, BS: Consensus statement on standards and guidelines for the molecular diagnostics of Alport syndrome: refining the ACMG criteria. *Eur J Hum Genet*, 2021.
27. Mariyama, M, Leinonen, A, Mochizuki, T, Tryggvason, K, Reeders, ST: Complete primary structure of the human alpha 3(IV) collagen chain. Coexpression of the alpha 3(IV) and alpha 4(IV) collagen chains in human tissues. *J Biol Chem,* 269**:** 23013-23017, 1994.
28. Leinonen, A, Mariyama, M, Mochizuki, T, Tryggvason, K, Reeders, ST: Complete primary structure of the human type IV collagen alpha 4(IV) chain. Comparison with structure and expression of the other alpha (IV) chains. *J Biol Chem,* 269**:** 26172-26177, 1994.
29. Zhou, J, Hertz, JM, Leinonen, A, Tryggvason, K: Complete amino acid sequence of the human alpha 5 (IV) collagen chain and identification of a single-base mutation in exon 23 converting glycine 521 in the collagenous domain to cysteine in an Alport syndrome patient. *J Biol Chem,* 267**:** 12475-12481, 1992.
30. Demosthenous, P, Voskarides, K, Stylianou, K, Hadjigavriel, M, Arsali, M, Patsias, C, Georgaki, E, Zirogiannis, P, Stavrou, C, Daphnis, E, Pierides, A, Deltas, C, Hellenic Nephrogenetics Research, C: X-linked Alport syndrome in Hellenic families: phenotypic heterogeneity and mutations near interruptions of the collagen domain in COL4A5. *Clin Genet,* 81**:** 240-248, 2012.
31. Lanktree, MB, Haghighi, A, Guiard, E, Iliuta, IA, Song, X, Harris, PC, Paterson, AD, Pei, Y: Prevalence Estimates of Polycystic Kidney and Liver Disease by Population Sequencing. *J Am Soc Nephrol,* 29**:** 2593-2600, 2018.
32. Lek, M, Karczewski, KJ, Minikel, EV, Samocha, KE, Banks, E, Fennell, T, O'Donnell-Luria, AH, Ware, JS, Hill, AJ, Cummings, BB, Tukiainen, T, Birnbaum, DP, Kosmicki, JA, Duncan, LE, Estrada, K, Zhao, F, Zou, J, Pierce-Hoffman, E, Berghout, J, Cooper, DN, Deflaux, N, DePristo, M, Do, R, Flannick, J, Fromer, M, Gauthier, L, Goldstein, J, Gupta, N, Howrigan, D, Kiezun, A, Kurki, MI, Moonshine, AL, Natarajan, P, Orozco, L, Peloso, GM, Poplin, R, Rivas, MA, Ruano-Rubio, V, Rose, SA, Ruderfer, DM, Shakir, K, Stenson, PD, Stevens, C, Thomas, BP, Tiao, G, Tusie-Luna, MT, Weisburd, B, Won, HH, Yu, D, Altshuler, DM, Ardissino, D, Boehnke, M, Danesh, J, Donnelly, S, Elosua, R, Florez, JC, Gabriel, SB, Getz, G, Glatt, SJ, Hultman, CM, Kathiresan, S, Laakso, M, McCarroll, S, McCarthy, MI, McGovern, D, McPherson, R, Neale, BM, Palotie, A, Purcell, SM, Saleheen, D, Scharf, JM, Sklar, P, Sullivan, PF, Tuomilehto, J, Tsuang, MT, Watkins, HC, Wilson, JG, Daly, MJ, MacArthur, DG, Exome Aggregation, C: Analysis of protein-coding genetic variation in 60,706 humans. *Nature,* 536**:** 285-291, 2016.
33. The National Genomics Research and Healthcare Knowledgebase v5, Genomics England. 5, 2019.
34. van der Ven, AT, Connaughton, DM, Ityel, H, Mann, N, Nakayama, M, Chen, J, Vivante, A, Hwang, DY, Schulz, J, Braun, DA, Schmidt, JM, Schapiro, D, Schneider, R, Warejko, JK, Daga, A, Majmundar, AJ, Tan, W, Jobst-Schwan, T, Hermle, T, Widmeier, E, Ashraf, S, Amar, A, Hoogstraaten, CA, Hugo, H, Kitzler, TM, Kause, F, Kolvenbach, CM, Dai, R, Spaneas, L, Amann, K, Stein, DR, Baum, MA, Somers, MJG, Rodig, NM, Ferguson, MA, Traum, AZ, Daouk, GH, Bogdanovic, R, Stajic, N, Soliman, NA, Kari, JA, El Desoky, S, Fathy, HM, Milosevic, D, Al-Saffar, M, Awad, HS, Eid, LA, Selvin, A, Senguttuvan, P, Sanna-Cherchi, S, Rehm, HL, MacArthur, DG, Lek, M, Laricchia, KM, Wilson, MW, Mane, SM, Lifton, RP, Lee, RS, Bauer, SB, Lu, W, Reutter, HM, Tasic, V, Shril, S, Hildebrandt, F: Whole-Exome Sequencing Identifies Causative Mutations in Families with Congenital Anomalies of the Kidney and Urinary Tract. *J Am Soc Nephrol,* 29**:** 2348-2361, 2018.
35. Kelly, MA, Caleshu, C, Morales, A, Buchan, J, Wolf, Z, Harrison, SM, Cook, S, Dillon, MW, Garcia, J, Haverfield, E, Jongbloed, JDH, Macaya, D, Manrai, A, Orland, K, Richard, G, Spoonamore, K, Thomas, M, Thomson, K, Vincent, LM, Walsh, R, Watkins, H, Whiffin, N, Ingles, J, van Tintelen, JP, Semsarian, C, Ware, JS, Hershberger, R, Funke, B: Adaptation and validation of the ACMG/AMP variant classification framework for MYH7-associated inherited cardiomyopathies: recommendations by ClinGen's Inherited Cardiomyopathy Expert Panel. *Genet Med,* 20**:** 351-359, 2018.

36. Biesecker, LG, Harrison, SM, ClinGen Sequence Variant Interpretation Working, G: The ACMG/AMP reputable source criteria for the interpretation of sequence variants. *Genet Med,* 20**:** 1687-1688, 2018.
37. Richards, CS, Aziz, N, Bale, S, Bick, D, Das, S, Gastier-Foster, J, Grody, WW, Hegde, M, Lyon, E, Spector, E, Voelkerding, K, Rehm, HL, Group, AAIoSVW: Response to Biesecker and Harrison. *Genet Med,* 20**:** 1689-1690, 2018.
38. Abou Tayoun, AN, Pesaran, T, DiStefano, MT, Oza, A, Rehm, HL, Biesecker, LG, Harrison, SM, ClinGen Sequence Variant Interpretation Working, G: Recommendations for interpreting the loss of function PVS1 ACMG/AMP variant criterion. *Hum Mutat,* 39**:** 1517-1524, 2018.
39. Hashimura, Y, Nozu, K, Kaito, H, Nakanishi, K, Fu, XJ, Ohtsubo, H, Hashimoto, F, Oka, M, Ninchoji, T, Ishimori, S, Morisada, N, Matsunoshita, N, Kamiyoshi, N, Yoshikawa, N, Iijima, K: Milder clinical aspects of X-linked Alport syndrome in men positive for the collagen IV alpha5 chain. *Kidney Int,* 85**:** 1208-1213, 2014.
40. Savige, J, Storey, H, Il Cheong, H, Gyung Kang, H, Park, E, Hilbert, P, Persikov, A, Torres-Fernandez, C, Ars, E, Torra, R, Hertz, JM, Thomassen, M, Shagam, L, Wang, D, Wang, Y, Flinter, F, Nagel, M: X-Linked and Autosomal Recessive Alport Syndrome: Pathogenic Variant Features and Further Genotype-Phenotype Correlations. *PLoS One,* 11**:** e0161802, 2016.
41. Persikov, AV, Pillitteri, RJ, Amin, P, Schwarze, U, Byers, PH, Brodsky, B: Stability related bias in residues replacing glycines within the collagen triple helix (Gly-Xaa-Yaa) in inherited connective tissue disorders. *Hum Mutat,* 24**:** 330-337, 2004.
42. Storey, H, Savige, J, Sivakumar, V, Abbs, S, Flinter, FA: COL4A3/COL4A4 mutations and features in individuals with autosomal recessive Alport syndrome. *J Am Soc Nephrol,* 24**:** 1945-1954, 2013.
43. Slajpah, M, Gorinsek, B, Berginc, G, Vizjak, A, Ferluga, D, Hvala, A, Meglic, A, Jaksa, I, Furlan, P, Gregoric, A, Kaplan-Pavlovcic, S, Ravnik-Glavac, M, Glavac, D: Sixteen novel mutations identified in COL4A3, COL4A4, and COL4A5 genes in Slovenian families with Alport syndrome and benign familial hematuria. *Kidney Int,* 71**:** 1287-1295, 2007.
44. Macheroux, EP, Braunisch, MC, Pucci Pegler, S, Satanovskij, R, Riedhammer, KM, Gunthner, R, Gross, O, Nagel, M, Renders, L, Hoefele, J: The Hypomorphic Variant p.(Gly624Asp) in COL4A5 as a Possible Cause for an Unexpected Severe Phenotype in a Family With X-Linked Alport Syndrome. *Front Pediatr,* 7**:** 485, 2019.
45. Zurowska, AM, Bielska, O, Daca-Roszak, P, Jankowski, M, Szczepanska, M, Roszkowska-Bjanid, D, Kuzma-Mroczkowska, E, Panczyk-Tomaszewska, M, Moczulska, A, Drozdz, D, Hadjipanagi, D, Deltas, C, Ostalska-Nowicka, D, Rabiega, A, Taraszkiewicz, J, Taranta-Janusz, K, Wieczorkiewicz-Plaza, A, Jobs, K, Mews, J, Musial, K, Jakubowska, A, Nosek, H, Jander, AE, Koutsofti, C, Stanislawska-Sachadyn, A, Kuleszo, D, Zietkiewicz, E, Lipska-Zietkiewicz, BS: 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*, 2020.
46. AM, Z: Mild X-linked Alport syndrome due to the COL4A5 G624D variant origtinating in the Middle Ages is predominant in Central/East Europe and causes kidney failure in mid-life. *Kidney Int*, 2021.
47. Feingold, J, Bois, E, Chompret, A, Broyer, M, Gubler, MC, Grunfeld, JP: Genetic heterogeneity of Alport syndrome. *Kidney Int,* 27: 672-677, 1985.
48. Lemmink, HH, Nillesen, WN, Mochizuki, T, Schroder, CH, Brunner, HG, van Oost, BA, Monnens, LA, Smeets, HJ: Benign familial hematuria due to mutation of the type IV collagen alpha4 gene. *J Clin Invest,* 98**:** 1114-1118, 1996.
49. Persikov, AV, Ramshaw, JA, Kirkpatrick, A, Brodsky, B: Amino acid propensities for the collagen triple-helix. *Biochemistry,* 39**:** 14960-14967, 2000.
50. Fallerini, C, Baldassarri, M, Trevisson, E, Morbidoni, V, La Manna, A, Lazzarin, R, Pasini, A, Barbano, G, Pinciaroli, AR, Garosi, G, Frullanti, E, Pinto, AM, Mencarelli, MA, Mari, F, Renieri, A, Ariani, F: Alport syndrome: impact of digenic inheritance in patients management. *Clin Genet,* 92**:** 34-44, 2017.
51. Makareeva, E, Mertz, EL, Kuznetsova, NV, Sutter, MB, DeRidder, AM, Cabral, WA, Barnes, AM, McBride, DJ, Marini, JC, Leikin, S: Structural heterogeneity of type I collagen triple helix and its role in osteogenesis imperfecta. *J Biol Chem,* 283**:** 4787-4798, 2008.
52. Gross, O, Licht, C, Anders, HJ, Hoppe, B, Beck, B, Tonshoff, B, Hocker, B, Wygoda, S, Ehrich, JH, Pape, L, Konrad, M, Rascher, W, Dotsch, J, Muller-Wiefel, DE, Hoyer, P, Study Group Members of the Gesellschaft fur Padiatrische, N, Knebelmann, B, Pirson, Y, Grunfeld, JP, Niaudet, P, Cochat, P, Heidet, L, Lebbah, S, Torra, R, Friede, T, Lange, K, Muller, GA, Weber, M: Early angiotensin-converting enzyme inhibition in Alport syndrome delays renal failure and improves life expectancy. *Kidney Int,* 81**:** 494-501, 2012.
53. Kaariainen, H, Muilu, J, Perola, M, Kristiansson, K: Genetics in an isolated population like Finland: a different basis for genomic medicine? *J Community Genet,* 8**:** 319-326, 2017.
54. Kashtan, CE: Alport Syndrome. Gene Reviews

In: Adam MP, Ardinger HH, Pagon RA, et al., editors. GeneReviews® [Internet]. Seattle (WA): University of Washington, Seattle; 1993-2020. Available from: https://www.ncbi.nlm.nih.gov/books/NBK1207/. 2001.

55. Pierides, A, Voskarides, K, Athanasiou, Y, Ioannou, K, Damianou, L, Arsali, M, Zavros, M, Pierides, M, Vargemezis, V, Patsias, C, Zouvani, I, Elia, A, Kyriacou, K, Deltas, C: Clinico-pathological correlations in 127 patients in 11 large pedigrees, segregating one of three heterozygous mutations in the COL4A3/COL4A4 genes associated with familial haematuria and significant late progression to proteinuria and chronic kidney disease from focal segmental glomerulosclerosis. *Nephrol Dial Transplant,* 24**:** 2721-2729, 2009.
56. Matthaiou, M, Poulii, T, Matthaiou, M, Matthaiou, M, Poulli, T, Deltas, C: Prevalence of clinical, pathological and molecular features of glomerular basement membrane nephropathy caused by *COL4A3* or *COL4A4* mutations: a systematic review. *Clinical Kidney Journal* 1-12, 2020.
57. Gillion, V, Dahan, K, Cosyns, JP, Hilbert, P, Jadoul, M, Goffin, E, Godefroid, N, De Meyer, M, Mourad, M, Pirson, Y, Kanaan, N: Genotype and Outcome After Kidney Transplantation in Alport Syndrome. *Kidney Int Rep,* 3**:** 652-660, 2018.
58. Groopman, EE, Marasa, M, Cameron-Christie, S, Petrovski, S, Aggarwal, VS, Milo-Rasouly, H, Li, Y, Zhang, J, Nestor, J, Krithivasan, P, Lam, WY, Mitrotti, A, Piva, S, Kil, BH, Chatterjee, D, Reingold, R, Bradbury, D, DiVecchia, M, Snyder, H, Mu, X, Mehl, K, Balderes, O, Fasel, DA, Weng, C, Radhakrishnan, J, Canetta, P, Appel, GB, Bomback, AS, Ahn, W, Uy, NS, Alam, S, Cohen, DJ, Crew, RJ, Dube, GK, Rao, MK, Kamalakaran, S, Copeland, B, Ren, Z, Bridgers, J, Malone, CD, Mebane, CM, Dagaonkar, N, Fellstrom, BC, Haefliger, C, Mohan, S, Sanna-Cherchi, S, Kiryluk, K, Fleckner, J, March, R, Platt, A, Goldstein, DB, Gharavi, AG: Diagnostic Utility of Exome Sequencing for Kidney Disease. *N Engl J Med,* 380**:** 142-151, 2019.
59. Fallerini, C, Dosa, L, Tita, R, Del Prete, D, Feriozzi, S, Gai, G, Clementi, M, La Manna, A, Miglietti, N, Mancini, R, Mandrile, G, Ghiggeri, GM, Piaggio, G, Brancati, F, Diano, L, Frate, E, Pinciaroli, AR, Giani, M, Castorina, P, Bresin, E, Giachino, D, De Marchi, M, Mari, F, Bruttini, M, Renieri, A, Ariani, F: Unbiased next generation sequencing analysis confirms the existence of autosomal dominant Alport syndrome in a relevant fraction of cases. *Clin Genet,* 86**:** 252-257, 2014.
60. Bullich, G, Domingo-Gallego, A, Vargas, I, Ruiz, P, Lorente-Grandoso, L, Furlano, M, Fraga, G, Madrid, A, Ariceta, G, Borregan, M, Pinero-Fernandez, JA, Rodriguez-Pena, L, Ballesta-Martinez, MJ, Llano-Rivas, I, Menica, MA, Ballarin, J, Torrents, D, Torra, R, Ars, E: A kidney-disease gene panel allows a comprehensive genetic diagnosis of cystic and glomerular inherited kidney diseases. *Kidney Int,* 94**:** 363-371, 2018.
61. Gross, O, Tonshoff, B, Weber, LT, Pape, L, Latta, K, Fehrenbach, H, Lange-Sperandio, B, Zappel, H, Hoyer, P, Staude, H, Konig, S, John, U, Gellermann, J, Hoppe, B, Galiano, M, Hoecker, B, Ehren, R, Lerch, C, Kashtan, CE, Harden, M, Boeckhaus, J, Friede, T, German Pediatric Nephrology Study, G, Investigators, EP-TA: A multicenter, randomized, placebo-controlled, double-blind phase 3 trial with open-arm comparison indicates safety and efficacy of nephroprotective therapy with ramipril in children with Alport's syndrome. *Kidney Int,* 97**:** 1275-1286, 2020.
62. Hertz, JM, Juncker, I, Persson, U, Matthijs, G, Schmidtke, J, Petersen, MB, Kjeldsen, M, Gregersen, N: Detection of mutations in the COL4A5 gene by SSCP in X-linked Alport syndrome. *Hum Mutat,* 18**:** 141-148, 2001.
63. Plant, KE, Green, PM, Vetrie, D, Flinter, FA: Detection of mutations in COL4A5 in patients with Alport syndrome. *Hum Mutat,* 13**:** 124-132, 1999.
64. King, K, Flinter, FA, Green, PM: A two-tier approach to mutation detection in the COL4A5 gene for Alport syndrome. *Hum Mutat,* 27**:** 1061, 2006.
65. Cruz-Robles, D, Garcia-Torres, R, Antignac, C, Forestier, L, de la Puente, SG, Correa-Rotter, R, Garcia-Lopez, E, Orozco, L: Three novel mutations in the COL4A5 gene in Mexican Alport syndrome patients. *Clin Genet,* 56**:** 242-243, 1999.
66. Bekheirnia, MR, Reed, B, Gregory, MC, McFann, K, Shamshirsaz, AA, Masoumi, A, Schrier, RW: Genotype-phenotype correlation in X-linked Alport syndrome. *J Am Soc Nephrol,* 21**:** 876-883, 2010.
67. Mohammad, M, Nanra, R, Colville, D, Trevillian, P, Wang, Y, Storey, H, Flinter, F, Savige, J: A female with X-linked Alport syndrome and compound heterozygous COL4A5 mutations. *Pediatr Nephrol,* 29**:** 481-485, 2014.
68. Knebelmann, B, Breillat, C, Forestier, L, Arrondel, C, Jacassier, D, Giatras, I, Drouot, L, Deschenes, G, Grunfeld, JP, Broyer, M, Gubler, MC, Antignac, C: Spectrum of mutations in the COL4A5 collagen gene in X-linked Alport syndrome. *Am J Hum Genet,* 59**:** 1221-1232, 1996.
69. Inoue, Y, Nishio, H, Shirakawa, T, Nakanishi, K, Nakamura, H, Sumino, K, Nishiyama, K, Iijima, K, Yoshikawa, N: Detection of mutations in the COL4A5 gene in over 90% of male patients with X-linked Alport's syndrome by RT-PCR and direct sequencing. *Am J Kidney Dis,* 34**:** 854-862, 1999.

70. Zhang, H, Ding, J, Wang, F, Zhao, D: Mutation detection of COL4An gene based on mRNA of peripheral blood lymphocytes and prenatal diagnosis of Alport syndrome in China. *Nephrology (Carlton),* 16**:** 377-380, 2011.

**Table 1: Association of predicted pathogenic *COL4A3 – COL4A5* variants resulting in position 1 Gly substitutions with haematuria in 100,000 Genomes Project database**

| Gene | Number of variants resulting in position 1 Gly substitutions in 100,000 Genomes Project | Number of variants resulting in position 1 Gly substitutions excluded after filtering | Number of individuals with a predicted pathogenic position 1 Gly substitution | | Chi-square (with Yate's correction), two-tailed p value |
|---|---|---|---|---|---|
| | | | In total with haematuria (n=2,221) | In total without haematuria (n=37,200) | |
| *COL4A5* | 20 | 1 (5%) | 12 | 9 | 95.83, p<0.0001 |
| *COL4A3* | 87 | 14 (16%) | 21 | 85 | 37.54, p<0.0001 |
| *COL4A4* | 84 | 13 (15%) | 19 | 95 | 24.14, p<0.0001 |

**Table 2: Assessment of *COL4A5* variants in the gnomAD database and gnomAD control subset**

| Total gnomAD database | | | | | | | | | | | gnomAD control subset | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Position hg19 and Protein consequence** | **Transcript** | **Affected individuals gender, total alleles** | **PP2** | **SIFT** | **MT** | **Conserved** | **Previously reported; alternative pathogenic residue (LOVD)** | **Clin Var** | **Pathogenicity (Varsome)** | **Patho-genicity (Alamut)** | **Evidence supports patho-genicity** | **Affected individuals gender, total alleles** |
| 107783018 p.(Gly42Ser) | c.124G>A | 1M, 21,780 | 0 | Tol | DC | Gly | NO | Not found | Likely pathogenic, PM1, PM2, PP2,PP3 | Weak | NO | |
| 107783019 p.(Gly42Ala) | c.125G>C | 1F, 182,802 | 0.003 | Tol | DC | Gly | NO | Not found | Likely pathogenic, PM1,PM2,PP2,PP3 | Weak | NO | 1F, 80,278 |
| 107783019 p.(Gly42Asp) | c.125G>A | 1F, 182,802 | 0.739 | Tol | DC | Gly | NO | Not found | Likely pathogenic, PM1,PM2,PP2,PP3 | Weak | NO | |
| 107802357 p.(Gly69Arg) | c.205G>C | 1F, 183,072 | 0.01 | Del | DC | Gly | NO | Not found | Likely pathogenic, PM1,PM2,PP2,PP3 | Strong | YES (1F) | |
| 107807131 p.(Gly84Glu) | c.251G>A | 1F, 182,367 | 1 | Del | DC | Gly | NO | Not found | Likely pathogenic, PM1,PM2,PM5,PP2, PP3 | Strong | YES (1F) | |
| 107811896 p.(Gly105Ala) | c.314G>C | 1M, 183,404 | 1 | Del | DC | Gly | NO | Not found | Likely pathogenic, PM1,PM2,PP2,PP3 | Moderate | YES (1M) | 1M, 80,333 |
| 107812008 p.(Gly114Ala) | c.341G>C | 1F, 183,513 | 1 | Del | DC | Gly | NO, but p.(Gly114Arg) reported in LOVD | Not found | Likely pathogenic, PM1,PM2,PM5,PP2, PP3 | Moderate | YES (1F) | 1F, 80,345 |
| 107812044 p.(Gly126Glu) | c.377G>A | 1F, 21,871 | 1 | Del | DC | Gly | NO | Not found | Likely pathogenic, PM1,PM2,PP2,PP3 | Moderate | YES (1F) | 1F, 7600 |
| 107814670 p.(Gly138Ser) | c.412G>A | 1F, 183,271 | 1 | Tol | DC | Gly | YES Fallerini, 2014 [59]; and p(Gly138Asp) | Pathogenic (no asserti | Likely pathogenic, PM1, PM2,PM5,PP2,PP3, PP% | Weak | **NO** | |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | in LOVD | on criteria) | | | | |
| 107819195 p.(Gly201Ala) | c.602G>C | 1M, 168,800 | 0.999 | Del | DC | Gly | NO, but p(Gly201Val), Hertz 2001 [62] | Not found | Likely pathogenic, PM1,PM2,PM5,PP2,PP3 | Moderate | YES (1M) | |
| 107823808 p.(Gly276Ser) | c.826G>A | 1F, 183,230 | 1 | Tol | DC | Gly | NO | Not found | Likely pathogenic, PM1,PM2,PP2,PP3 | Weak | | 1F, 80,319 |
| 107823912 p.(Gly279Arg) | c.835G>C | 1M, 183,017 | 1 | Tol | DC | Gly | NO | Not found | Likely pathogenic, PM1,PM2,PP2,PP3 | Moderate | NO | |
| 107823913 p.(Gly279Ala) | c.836G>C | 1M, 183,034 | 1 | Del | DC | Gly | NO | Not found | Likely pathogenic, PM1,PM2,PP2,PP3 | Weak | YES (1M) | 1M, 80,323 |
| p.(Gly283Val) | c.848G>T | 1F, 183,010 | 1 | Del | DC | Gly | NO | Not found | Likely pathogenic, PM1,PM2,PP2,PP3 | Strong | YES (1F) | |
| p.(Gly286Cys) | c.856G>T | 1F, 182,986 | 1 | Del | DC | Gly | NO | Not found | Likely pathogenic, PM1,PM2,PP2,PP3 | Strong | YES (1F) | 1F, 80,324 |
| 107829879 p.(Gly356Ala) | c.1067G>C | 1F, 21,220 | 0.997 | Del | DC | Gly | NO | Not found | Likely pathogenic, PM1,PM2,PP2,PP3 | Strong | YES (1F) | |
| 107829879 p.(Gly356Glu) | c.1067G>A | 1M, 183,308 | 0.99 | Del | DC | Gly | NO | Not found | Likely pathogenic, PM1,PM2,PP2,PP3 | Strong | YES (1M) | |
| p.(Gly512Glu) | c.1535G>A | 1M, 145,435 | 1 | Del | DC | Gly | NO | Not found | Likely pathogenic, PM1,PM2,PP2,PP3 | Strong | YES (1M) | |
| 107840625 p.(Gly536Ser) | c.1606G>A | 1F, 183,226 | 1 | Del | DC | Gly | NO, but p(Gly536Asp) p(Gly536Val), in LOVD | Not found | Likely pathogenic, PM1,PM2, PM5,PP2,PP3 | Strong | YES (1F) | |
| 107844640 p.(Gly624Asp) | c.1871G>A | 4M,12F, 182,998 | 1 | Del | DC | Gly | YES Slajpah, 2007[43] | **Pathogenic **** | Likely pathogenic PM1,PP2,PP3 and PP5 | Strong | YES (4M,12F) | 2M, 6F, 80,268 |
| 107842028 p.(Gly626Ser) | c.1876G>A | 2M,5F, 204,674 | 1 | Del | DC | Gly | YES Hertz, 2001[62] | **Likely benign *** | Likely pathogenic, PM1,PM2,PM5,PP2,PP3, BP6 | Strong | **NO** | 1M, 4F, 87,809 |
| 107844640 | c.1966G | 1M, | 1 | Del | DC | Gly | NO | Not | Likely pathogenic, | Strong | YES | |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| p.(Gly656Ser) | >A | 181,478 | | | | | | found | PM1,PM2,PP2,PP3 | | (1M) | |
| 107845177 p.(Gly702Ser) | c.2104G>A | 1F, 129,179 | 1 | Del | DC | Gly | NO, but p(Gly702Val), in LOVD (VUS) | Not found | Likely pathogenic, PM1,PM2,PP2,PP3 | Strong | YES (1F) | 1M, 55,168 |
| 107846274 p.(Gly743Ser) | c.2227G>A | 2F, 148,007 | 1 | Del | DC | **Gly** | NO, but p.(Gly743Asp) Plant, 1999 [63] | Not found | Likely pathogenic, PM1,PM2, PM5,PP2,PP3 | Weak | NO | 2F, 63,637 |
| 107849982 p.(Gly752Val) | c.2255G>T | 2M, 183,227 | 1 | Del | DC | Gly | NO | Not found | Likely pathogenic, PM1,PM2,PP2,PP3 | Strong | YES (2M) | |
| 107850087 p.(Gly787Ala) | c.2360G>C | 1F, 183,169 | 1 | Del | DC | Gly | NO, but p(Gly787Val), King 2006 [64] | Not found | Likely pathogenic, PM1,PM2, PM5,PP2,PP3 | Strong | YES (1F) | |
| 107858210 p.(Gly822Glu) | c.2465G>A | 1M, 182,957 | 1 | Del | DC | Gly | NO, but p(Gly822Arg), Cruz-Robles, 1999 [65] | Not found | Likely pathogenic, PM1,PM2, PM5,PP2,PP3 | Strong | YES (1M) | |
| 107865033 p.(Gly893Val) | c.2678G>T | 1F, 180,196 | 1 | Del | DC | Gly | YES Bekheirnia, 2010 [66], and p(Gly893Arg) Mohammad, 2014 [67] | Pathogenic (no assertion criteria) | Likely pathogenic, PM1,PM2, PM5,PP2,PP3,PP5 | Weak | YES (1F) | |
| p.(Gly953Val) | c.2858G>T | 249M, 442F, 204,819 | 1 | Del | P | Gly | YES Knebelmann 1996 [68] | Conflicting interpretation, B, LB, VUS | Benign, PM1,PP2,PP3,BS1,BS2 | Benign | NO | 125M,230F, 87,876 |
| 107866020 p.(Gly961Val) | c.2882G>T | 2M,3F, 182,593 | 0.001 | Del | P | **Gly** | NO | Not found | Variant of Uncertain Significance, PM1,PP2,BP4 | Weak | NO | 1M, 1F, 80,233 |
| 107866037 p.(Gly967Arg) | c.2899G>A | 1F, 181,344 | 1 | Del | DC | Gly | NO | Not found | Likely pathogenic,PM1,P | Moderate | YES (1F) | |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | M2,PP2,PP3 | | | |
| 107869494 p(Gly1054Asp) | c.3161G>A | 1M, 183,204 | 1 | Del | DC | Gly | LOVD | Not found | Likely pathogenic, PM1,PM2,PP2,PP3 | Moderate | YES (1M) | |
| 107869553 p(Gly1074Arg) | c.3220G>C | 1M, 183,246 | 0.073 | Del | DC | Gly | NO | Not found | Likely pathogenic, PM1,PM2,PP2,PP3 | Weak | NO | |
| 107869553 p.(Gly1074Ser) | c.3220G>A | 4M,3F, 204,841 | 0.129 | Del | DC | Gly | NO | Not found | Variant of Uncertain Significance, PM1,PP2,PP3, BS2 | Weak | NO | 1M,2F, 80,317 |
| p.(Gly1134Cys) | c.3400G>T | 1F, 180,336 | 1 | Del | DC | Gly | NO | Not found | Likely pathogenic, PM1,PM2,PP2,PP3 | Weak | YES (1F) | 1F, 79,040 |
| p.(Gly1170Ser) | c.3508G>A | 1F, 174,685 | 1 | Del | DC | Gly | YES Inoue, 1999 [69] | Patho-genic (no assertion criteria) | Pathogenic, PS1,PM1,PM2,PM5,PP2,PP3,PP5 | Weak | YES (1F) | 1F, 76,459 |
| 107909824 p.(Gly1185Ser) | c.3553G>A | 1F, 159,615 | 0.99 | Del | DC | Gly | NO | Not found | Pathogenic,PVS1, PM1,PM2,PP2,PP3 | Weak | YES (1F) | |
| 107911689 p.(Gly1249Arg) | c.3745G>A | 1F, 174,771 | 0.99 | Del | DC | Gly | NO | Not found | Likely pathogenic, PM1,PM2,PP2,PP3 | Weak | YES (1F) | |
| 107920766 p.(Gly1282Glu) | c.3845G>A | 1M, 183,125 | 1 | Del | DC | Gly | NO, but p(Gly1282Val), in LOVD | Not found | Likely pathogenic, PM1,PM2,PP2,PP3 | Weak | YES (1M) | |
| 107920774 p.(Gly1285Ser) | c.3853G>A | 1F, 183,121 | 0.026 | Tol | DC | Gly | NO | Not found | Likely pathogenic, PM1,PM2,PP2,PP3 | Weak | NO | 1F, 80,288 |
| 107920820 p.(Gly1300Ala) | c.3899G>C | 1F, 182,971 | 0.999 | Del | DC | Gly | NO | Not found | Likely pathogenic, PM1,PM2,PP2,PP3 | Weak | YES (1F) | |
| 107920829 p.(Gly1303Ala) | c.3908G>C | 1M, 182,951 | 1 | Del | DC | Gly | NO | Not found | Likely pathogenic, PM1,PM2,PP2,PP3 | Weak | YES (1M) | |
| 107923928 | c.3962G | 1F, | 1 | Del | DC | Gly | NO | Not | Likely pathogenic, | Weak | YES | |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| p.(Gly1321Val) | >T | 182,976 | | | | | | found | PM1,PM2,PP2,PP3 | | (1F) | |
| 107923937 p.(Gly1324Glu) | c.3971G>A | 1F, 182,990 | 1 | Del | DC | Gly | NO | Not found | Likely pathogenic, PM1,PM2,PP2,PP3 | Weak | YES (1F) | 1F, 80,251 |
| p.(Gly1333Cys) | c.3997G>T | 1M, 182,997 | 1 | Del | DC | Gly | NO, but p(Gly1333Ser), Plant, 1999 [63, 70] | Not found | Pathogenic, PVS1,PM1,PM2,PM5,PP2,PP3 | Weak | YES (1M) | 1M, 80,243 |
| 107925082 p.(Gly1394Cys) | c.4180G>T | 2F, 182,474 | 1 | Del | DC | Gly | NO | Not found | Pathogenic and predicted to result in termination codon, PP3 | Moderate | YES (2F) | |
| 107929314 p.(Gly1424Ser) | c.4270G>A | 1F, 182,592 | 1 | Del | DC | Gly | YES LOVD, and p(Gly1424Glu) Zhang, 2011 | Not found | Likely pathogenic, PM2,PM3,PM5,PP2,PP3 | Weak | YES (1F) | |
| 107802385 | c.231+2T>C | 1M, 182,806 | | | | | NO | Not found | Pathogenic, PVS1,PM2,PP3 | Exon 3 skip likely, likely abolition of donor site (3/4 tools) | YES (1M) | 1F, 80,152 |
| 107821614 | c.780+1G>T | 1F, 183,138 | | | | | NO | Not found | Pathogenic, PVS1,PM2,PP3 | Exon 13 skip likely, likely abolition of donor site (4/4 tools) | YES (1F) | 1M, 80,312 |
| 107863487 | c.1340-1G>A | 1F, 183,216 | | | | | NO | Likely pathogenic* | Pathogenic, PVS1,PM2,PP3,PP5 | Exon21 skip likely, likely abolition of acceptor site (4/4 tools) | YES (1F) | 1M, 80,317 |
| 107863487 | c.2510-2A>G | 1F, 181,241 | | | | | NO | Pathogenic (no assertion criteria) | Pathogenic, PVS1,PM2,PP3,PP5 | Exon 31 skip likely, likely abolition of acceptor site (3/4 tools) | YES (1F) | 1M, 80,268 |
| 107917984 | c.3808+ | 1F, | | | | | NO | Not | Pathogenic, | Exon 43 skip likely, | YES | |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1G>C | 169,376 | | | | | | found | PVS1,PM2,PP3,PP5 | likely abolition of donor site (3/4) | (1F) | |
| 107929360 | c.4315+1G>A | 2F, 200,378 | | | | | NO | Pathogenic* | Pathogenic, PVS1,PM2,PP3,PP5 | Exon 48 skip likely, likely abolition of donor site (3/4 tools) | YES (2F) | |
| 107898559 p.(Gly1083Ter) | c.3247_3248 delGG | 1F, 182,859 | | | | | NO | Not found | Pathogenic, PVS1,PM2,PP3 | Likely abolition of acceptor site (4/4 tools) interrupts reading frame prematurely | YES ) | 1F, 80,169 |
| **54 variants assessed** | | 88 individuals potentially affected (29M, 59F) | 47 variants patho-genic | 47 variants patho-genic | 52 variants patho-genic | 45 missense variants con-served | 16 previous reports of pathogenicity | 9 decisions ; 7 pathogenic | 51 variants predicted pathogenic in Varsome | 30 variants moderately or strongly pathogenic | 9 decisions; 7 pathogenic | 23 predicted pathogenic alleles (9M, 14F) in mean 73,849 alleles examined |

**Evidence for pathogenicity is highlighted in red and evidence against pathogenicity is in blue. M male, F female. PP2 (Polyphen-2), SIFT (Sorting Intolerant from Tolerant), Mutation taster are computational tools to assess pathogenicity, and scores of >0.80, Del (for deleterious) or DC (disease-causing) are consistent with pathogenicity respectively. Conservation of Gly was examined to vertebrates (in birds). ClinVar – uses P pathogenic, LP Likely pathogenic, VUS – variant of uncertain significance, B – benign, LB likely benign, and star system used for quality of assertion * to ****. Varsome uses ACMG/AMP grading of P, LP,VUS,LB and B and criteria of PVS, PM, PP, BS etc. While individual tools may be consistent with a pathogenic or benign classification these are not individually sufficient to assert variant pathogenicity. The tools used here have been to principally exclude variants where the evidence is not totally supportive of pathogenicity.**

**Table 3: Prevalence of predicted pathogenic variants in *COL4A5* in people of different ancestries in gnomAD**

| Ethnicity | Number of individuals tested | Number of predicted pathogenic variants | Prevalence of predicted pathogenic variants in *COL4A5* |
|---|---|---|---|
| African | 12,132 | 7 | One in 1,733 |
| Ashkenazi Jewish | 4,961 | 1 | One in 4,961 |
| East Asian | 9,243 | 4 | One in 2,310 |
| European | 59,512 | 33 | One in 1,800 |
| Finnish | 12,242 | 0 | None |
| Latino | 18,105 | 6 | One in 3,018 |
| South Asian | 12,948 | 4 | One in 3,237 |
| Other | 3,140 | 4 | One in 785 |
| Total | 136,920 | 59 | One in 2,320 |

**Table 4: Prevalence of predicted pathogenic variants in *COL4A3* and *COL4A4* in people of different ancestries in gnomAD**

| ***COL4A3*** | **Probably pathogenic variants/mean total alleles examined** | **Prevalence** | **African** | **Latino** | **Ashkenazim** | **East Asian** | **European Finnish** | **European non-Finnish** | **Other** | **South Asians** |
|---|---|---|---|---|---|---|---|---|---|---|
| **Position 1 Gly substitutions** | 380/234,398 | 0.16% | 32/15,280 | 96/32,139 | 5/9400 | 46/16,842 | 7/20,376 | 140/106,288 | 8/5739 | 46/28,333 |
| **Frameshift variants** | 110/280,838 | 0.04% | 3/15,431 | 26/35,368 | 0/10,360 | 2/17,966 | 6/25,004 | 48/128,652 | 5/7,148 | 18/30,602 |
| **Nonsense variants** | 33/245,590 | 0.01% | 8/18,361 | 4/32,413 | 2/9,487 | 2/17,393 | 1/21,567 | 12/111,937 | 0/6,117 | 4/28,315 |
| **Splicing variants** | 36/264,528 | 0.01% | 4/14,588 | 5/30,169 | 0/8,840 | 4/15,827 | 1/19,207 | 13/99,841 | 2/5,421 | 7/26,560 |
| **% predicted pathogenic alleles/mean total alleles** | 559/245,889 | 0.23% | 47/15,755 (0.30%) | 131/32,713 (0.40%) | 7/9,425 (0.07%) | 54/16,829 (0.32%) | 15/22,228 (0.07%) | 213/111,253 (0.19%) | 15/6,166 (0.24%) | 75/28,711 (0.26%) |
| **% individuals with *COL4A3* variant** | | 0.45% | 0.60% | 0.80% | 0.14% | 0.64% | 0.14% | 0.38% | 0.48% | 0.52% |
| ***COL4A4*** | | | | | | | | | | |
| **Position 1 Gly substitutions** | 458/234,586 | 0.20% | 44/15,805 | 114/31,850 | 6/9,292 | 52/16,793 | 9/20,406 | 174/106,590 | 7/5,759 | 52/28,090 |
| **Frameshift variants** | 42/223,504 | 0.02% | 5/14,880 | 2/30,446 | 1/8,878 | 3/16,014 | 1/19,435 | 24/101,495 | 0/5,482 | 6/26,870 |
| **Nonsense** | 48/228,799 | 0.02% | 2/15,211 | 2/31,170 | 1/9,098 | 2/16,386 | 0/19,838 | 29/104,040 | 3/5,609 | 4/27,444 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **variants** | | | | | | | | | | |
| **Splicing variants** | 29/246,884 | 0.01% | 2/15,273 | 5/34,351 | 0/10,009 | 3/17,838 | 0/21,428 | 14/111,706 | 0/6,006 | 5/30,271 |
| **% predicted pathogenic alleles/mean total alleles** | 577/233,916 | 0.25% | 53/15,675 (0.34%) | 123/31,917 (0.39%) | 8/9,216 (0.09%) | 60/16,758 (0.35%) | 10/20,309 (0.05%) | 241/106,073 (0.23%) | 10/5,714 (0.18%) | 67/28,104 (0.24%) |
| **% individuals with *COL4A4* variant** | | 0.49% | 0.68% | 0.78% | 0.18% | 0.70% | 0.10% | 0.46% | 0.36% | 0.48% |
| **% with COL4A3 or *COL4A4* variant** | | 0.94% | 1.28% | 1.58% | 0.32% | 1.34% | 0.24% | 0.84% | 0.84% | 1.00% |

**These population frequencies are based on the mean total alleles examined for each variant and not on the total number of individuals examined.**

## Figure Legend

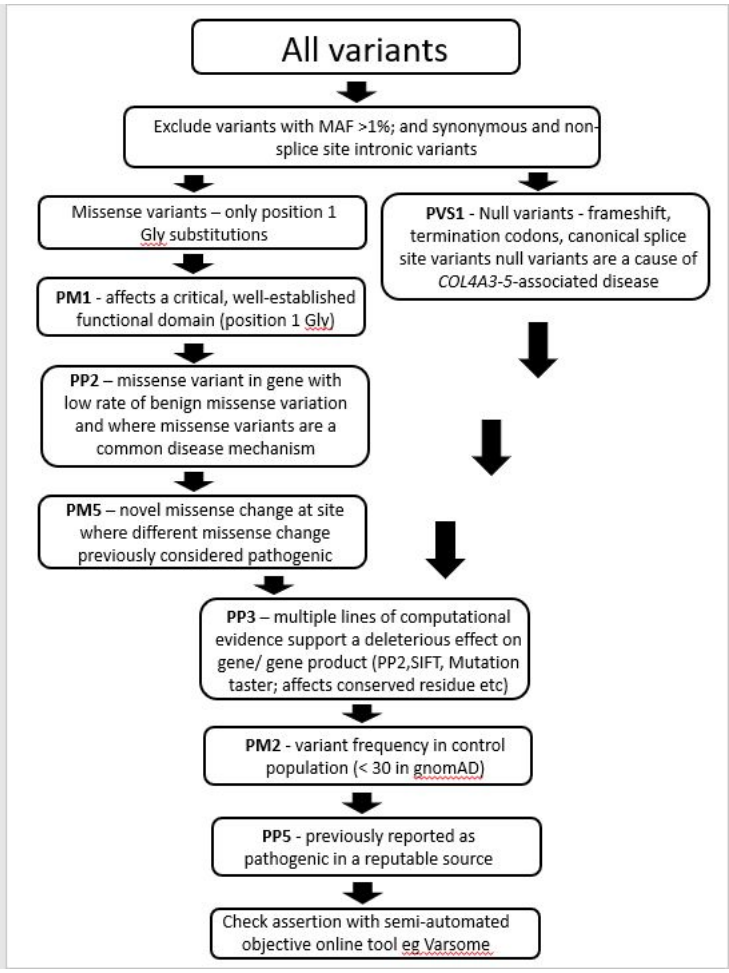


**Figure 1: Filtering steps in assessment of individual *COL4A3 – COL4A5* variants using ACMG/AMP criteria. PVS1 – very strong evidence of pathogenicity; PM1, PM2, PM5 - Moderate evidence of pathogenicity; PP2, PP3, PP5 – Supporting evidence of pathogenicity[24]. MAF - Minor Allele Frequency**

**Supplementary Tables**

**Suppl Table 1: Applicability of ACMG/AMP criteria for *COL4A3 – COL4A5* variants in population frequency databases** [24]

**Suppl Table 2: Non-collagenous domains and interruptions in collagen IV □5, □3 and □4 chains**[28, *26, 27*]

**Suppl Table 3: Assessment of *COL4A5* variants in the 100,000 Genomes Project database and correlation with haematuria**

**Suppl Table 4: Assessment of *COL4A3* variants in the 100,000 Genomes Project database and correlation with haematuria**

**Suppl Table 5: Assessment of *COL4A4* variants in the 100,000 Genomes Project database and correlation with haematuria**

**Suppl Table 6: Predicted pathogenic variants in *COL4A5* in the Exome Variant Server database**

**Suppl Table 7: *COL4A5* predicted pathogenic variants in the TOPMed database**

**Suppl Table 8: Mitigating features for the clinical effects of *COL4A5*predicted pathogenic variants resulting in position 1 Gly substitutions in gnomAD: associations with gender, location, and replacement residues**

**Suppl Table 9: Variants in *COL4A3* and *COL4A4* in gnomAD with inconsistent assessments subsequently excluded from population frequency studies**

**Suppl Table 10: Predicted pathogenic founder variants in *COL4A3* and *COL4A4* in different ancestries in gnomAD**

**Suppl Table 11: Estimated population frequencies for predicted pathogenic heterozygous, compound heterozygous, and digenic variants**

**Supplementary Tables**

**Suppl Table 1: Applicability of ACMG/AMP criteria for *COL4A3 – COL4A5* variants in population frequency databases [24]**

**Suppl Table 2: Non-collagenous domains and interruptions in collagen IV α5, α3 and α4 chains[28, 26, 27]**

**Suppl Table 3: Assessment of *COL4A5* variants in the 100,000 Genomes Project database and correlation with haematuria**

**Suppl Table 4: Assessment of *COL4A3* variants in the 100,000 Genomes Project database and correlation with haematuria**

**Suppl Table 5: Assessment of *COL4A4* variants in the 100,000 Genomes Project database and correlation with haematuria**

**Suppl Table 6: Predicted pathogenic variants in *COL4A5* in the Exome Variant Server database**

**Suppl Table 7: *COL4A5* predicted pathogenic variants in the TOPMed database**

**Suppl Table 8: Mitigating features for the clinical effects of *COL4A5*predicted pathogenic variants resulting in position 1 Gly substitutions in gnomAD: associations with gender, location, and replacement residues**

**Suppl Table 9: Variants in *COL4A3* and *COL4A4* in gnomAD with inconsistent assessments subsequently excluded from population frequency studies**

**Suppl Table 10: Predicted pathogenic founder variants in *COL4A3* and *COL4A4* in different ancestries in gnomAD**

**Suppl Table 11: Estimated population frequencies for predicted pathogenic heterozygous, compound heterozygous, and digenic variants**

**Suppl Table 1: Applicability of ACMG/AMP criteria for *COL4A3 – COL4A5* variants in population frequency databases [24]**

| Pathogenic criteria | Description | Applicability to population frequency databases for *COL4A3 – COL4A5* |
|---|---|---|
| **Very strong evidence of pathogenicity (PVS)** | | |
| PVS1 | Null variant (nonsense, frameshift, *canonical +/- 1 or 2 splice site, initiation codon, single or multi-exon deletion) | YES |
| **Strong evidence of pathogenicity (PS)** | | |
| PS1 | Same amino acid change as described previously regardless of nucleotide change | YES |
| PS2 | *De novo* variant in a patient with the disease and no family history | NO |
| PS3 | Well-established *in vitro* or *in vivo* functional studies supporting a damaging effect on the gene or gene product | No (not available for *COL4A3 – COL4A5)* |
| PS4 | Prevalence of variant in affected individuals is significantly increased compared with prevalence in controls | NO |
| **Moderate evidence of pathogenicity (PM)** | | |
| PM1 | Located in a mutational hotspot and/or critical and well-established functional domain | YES** |
| PM2 | Absent from controls (or low prevalence) | YES |
| PM3 | For AR diseases, detected *in trans* with a pathogenic variant | NO |
| PM4 | Protein length changes due in in-frame deletions/insertions or stop loss variants | YES |
| PM5 | Novel amino acid change in an amino acid residue where a different missense change determined to be pathogenic has been seen before | YES |
| PM6 | Assumed *de novo*, but without confirmation of paternity and maternity | NO |
| **Supporting evidence of pathogenicity (PP)** | | |
| PP1 | Co-segregation with disease in multiple affected family members in causative gene | NO |
| PP2 | Missense variant where these are a common mechanism of disease | YES |
| PP3 | Multiple lines of computational evidence support a deleterious effect (conservation, evolutionary, splicing impact etc) | YES |
| PP4 | Patient's phenotype or family history is highly specific for a disease with a single genetic aetiology | NO |
| PP5 | Reputable source recently reports variant as pathogenic but the evidence is not available to the laboratory | YES |
| **Benign criteria** | | |

| | | |
|---|---|---|
| **Stand alone evidence of benign impact (BA)** | | |
| BA1 | Allele frequency is above 5% in EVS, 1000 genomes or ExAC | NO |
| **Strong evidence of benign impact (BS)** | | |
| BS1 | Allele frequency is greater than expected for disorder | YES |
| BS2 | Observed in a healthy adult individual for X-linked disorder with full penetrance expected at an early age | NO |
| BS3 | Well-established *in vitro* or *in vivo* functional studies shows no damaging effect on protein function or splicing | NO |
| BS4 | Lack of segregation in affected members of a family | NO |
| **Supporting evidence of benign impact (BP)** | | |
| BP1 | Missense variant in a gene for which primarily truncating variants are known to cause disease | NO |
| BP2 | Observed *in trans* with a pathogenic variant for a fully penetrant dominant gene/disorder, or *in cis* with a pathogenic variant in any inheritance pattern | NO |
| BP3 | In-frame deletions/insertions in a repetitive region without a known function | YES |
| BP4 | Multiple lines of computational evidence suggest no impact on gene or gene product (conservation, evolutionary, splicing impact etc) | YES |
| BP5 | Variant found in a case with an alternate molecular basis for disease | NO |
| BP6 | Reputable source recently reports variant as benign but the evidence is not available for an independent evaluation | YES |
| BP7 | A synonymous variant for which splicing prediction algorithms predict no impact to the splice consensus sequence nor the creation of a new splice site AND the nucleotide is not highly conserved | YES |

PVS1, PS1, PM1 etc are explained in the Table

*A recent publication has warned against the use of PP3 for canonical splice-site variants but these are still used by Varsome[36], and we did not alter Varsome assessments.

** https://www.acgs.uk.com/media/11285/uk-practice-guidelines-for-variant-classification-2019-v1-0-3.pdf

**Suppl Table 2: Non-collagenous domains and interruptions in collagen IV α5, α3 and α4 chains**[28, 26, 27]

| Collagen IV α5 chain | Collagen IV α3 chain | Collagen IV α4 chain |
|---|---|---|
| Amino -NC domain (1-41) | Amino NC (1-42) | Amino NC (1-61) |
| I (160-167) | I (160-170) | I (176-183) |
| II (220-223) | II (222-223) | II (235-236) |
| III (243-257) | III (245-258) | III (258-269) |
| IV (282\|283)* | IV (283-287) | IV (294-295) |
| V (344-355) | V (345-353) | V (359-366) |
| VI (390-393) | VI (387-388) | VI (400-401) |
| VII (416-419) | VII (413-414) | VII (429-432) |
| VIII (442-453) | VIII (442-445) | VIII (457-462) |
| IX (479-487) | IX (476-483) | IX (493-496) |
| X (549-554) | X (544-547) | X (560-565) |
| XI (595-599) | XI (587-589) | XI (605-609) |
| XII (625) | XII (617-618) | XII (631-632) |
| XIII (657-662) | XIII (649-655) | XIII (666-673) |
| XIV (706) | XIV (698-699) | XIV (716-718) |
| XV (753-756) | XV (745-749) | XV (740-741) |
| XVI (818) | XVI (810-811) | XVI (763-764) |
| XVII (853-856) | XVII (848-849) | XVII (828-830) |
| XVIII (954-960) | XVIII (946-951) | XVIII (966-971) |
| XIX (1070-1073) | XIX (1060-1064) | XIX (1014) |
| XX (1189) | XX (1176-1179) | XX (1078-1081) |
| XXI (1245-8) | XXI (1234-1235) | XXI (1196-1197) |
| XXII (1373-78) | XXII (1263-1264) | XXII (1222-1223) |
| Carboxy-NC domain (1461-1691) | XXIII (1352-1357) | XXIII (1251-1257) |
| | Carboxy- NC (1439-1670) | XXIV (1285-1288) |
| | | XXV (1370-1379) |
| | | XXVI (1404-1405) |
| | | Carboxy-NC (1460-1690) |

*This is not an interruption but rather is formed from two Gly residues adjacent to each other.

**Suppl Table 3: Assessment of *COL4A5* variants in the 100,000 Genomes Project database and correlation with haematuria**

| Protein consequence | Transcript | PP2 | SIFT | M T | Cons | gnomAD (Hem/ Het, total alleles) | Previously reported; or substitution with other residue (LOVD) | ClinVar | Varsome | Predicted pathogenic | Number in 100,000 Genomes Project with haematuria | Number in 100,000 Genomes Project without haematuria |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| p.(Gly93Cys) | c.277G>T | 1 | 0 | 1 | Gly | None | NO | | P (PM1,PM2,PM5,PP2,PP3) | YES | 1 | 0 |
| p.(Gly108Glu) | c.323G>A | 1 | 0 | 1 | Gly | None | NO | | P (PM1,PM2,PM5,PP2,PP3) | YES | 1 | 0 |
| p.(Gly138Cys) | c.412G>T | 1 | 0 | 1 | Gly | None | NO; other - LP/P | LP* | LP (PM1,PM2,PM5,PP2, PP3, PP5) | YES | 0 | 1 |
| p.(Gly307Ser) | c.919G>A | 1 | 0 | 1 | Gly | None | YES- LP, other-P | | LP (PM1,PM2,PP2,PP3) | YES | 0 | 1 |
| p.(Gly325Arg) | c.973G>A | 0.999 | 0 | 1 | Gly | None | YES - P | P* | P (PS1,PM1,PM2,PM5,PP2, PP3,PP5) | YES | 1 | 0 |
| p.(Gly426Arg) | c.1276G>A | 1 | 0.001 | 1 | Gly | None | YES - P | | P (PS1,PM1,PM2,PP2,PP3) | YES | 1 | 0 |
| p.(Gly491Arg) | c.1471G>A | 0.999 | 0.003 | 1 | Gly | None | Other - P | | LP(PM1,PM2,PM5,PP2,PP3) | YES | 0 | 1 |
| p.(Gly533Arg) | c.1597G>A | 1 | 0 | 1 | Gly | None | Other - P | | LP (PM1,PM2,PM5,PP2,PP3) | YES | 1 | 0 |
| p.(Gly635Arg) | c.1903G>C | 1 | 0 | 1 | Gly | None | LP; other -P | | P (PM1,PM2,PM5,PP2,PP3) | YES | 1 | 0 |
| [GN] p.(Gly624Asp) | c.1871G>A | 1 | 0.021 | 1 | Gly | 16/ 182,998 | Slajpah (2007) | P** | P (PM1,PP2,PP3,PP5) | YES | 0 | 1 |
| p.(Gly653Val) | c.1958G>T | 1 | 0 | 1 | Gly | None | Other x2 - P | | LP (PM1, PM2, PM5, PP2,PP3) | YES | 1 | 0 |
| p.(Gly693Val) | c.2078G>T | 1 | | | Gly | None | Other x2- LP | | LP (PM1,PM2, PM5,PP2, PP3) | YES | 1 | 0 |
| p.(Gly707Val) | c.2120G>T | 1 | 0.002 | 1 | Gly | None | NO | | LP (PM1,PM2,PP2,PP3 | YES | 1 | 0 |
| p.(Gly743Ala) | c.2228G>C | 1 | 0 | 1 | Gly | None | Other - P | | LP (PM1,PM2,PM5,PP2,PP3) | YES | 0 | 1 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| GN p.(Gly752Val) | c.2255G>T | 1 | 0.031 | 1 | Gly | 2/183,277 | Other – P | | LP (PM1,PM2,PP2,PP3) | YES | 0 | 3 |
| p.(Gly953Val) | c.2858G>T | 1 | 0.001 | 1 | Gly | 705/204,819 | Knebelman (1996) | Conflicting* | VUS (PM1,PP2,PP3,PP5, BS1,BS2) | NO | 5 | 63 |
| p.(Gly1066Ser) | c.3196G>A | 1 | 0 | 1 | Gly | None | Other x 3 - P | P* | P (PS1, PM1,PM2, PM5,PP2,PP3,PP5) | YES | 1 | 0 |
| p.(Gly1388Ser) | c.4162G>A | 1 | 0.005 | 1 | Gly | None | NO | | LP (PM1,PM2,PP2,PP3) | YES | 0 | 1 |
| p.(Gly1421Asp) | c.4262G>A | 1 | 0 | 1 | Gly | None | Other - LP | | LP (PM1,PM2,PP2,PP3) | YES | 1 | 0 |
| p.(Gly1433Ser) | c.4297G>A | 1 | 0 | 1 | Gly | None | Other -3 in 4 - LP | | P (PVS1, PM1, PM2, PM5, PP2, PP3) | YES | 1 | 0 |
| Total number with a predicted pathogenic variant with and without haematuria respectively | | | | | | | | | | | 12 | 9 |
| Total number with predicted non-pathogenic variant with or without haematuria respectively | | | | | | | | | | | 5 | 63 |
| Total number with haematuria and without haematuria respectively | | | | | | | | | | | 2221 | 37,200 |

**GnomAD indicates that the variant is also found in gnomAD database. Data highlighted in blue are those with evidence against pathogenicity, and conclusion and number of individuals from 100,000 Genomes Project with this variant with or without haematuria are. PP2 (Polyphen-2), SIFT (Sorting Intolerant from Tolerant), Mutation taster are computational tools to assess pathogenicity, and scores of >0.80, Del (for deleterious) or DC (disease-causing) are consistent with pathogenicity respectively. Conservation of Gly was examined to vertebrates (in birds). ClinVar – uses P pathogenic, LP Likely pathogenic, VUS – variant of uncertain significance, B – benign, LB likely benign, and star system used for quality of assertion * to ****. Varsome uses ACMG/AMP grading of P, LP,VUS,LB and B and criteria of PVS, PM, PP, BS etc. The tools used here have been to principally exclude variants where the evidence is not supportive of pathogenicity.**

Of especial interest was p.(Gly953Val) which was not predicted pathogenic and excluded from the population frequency. It was assessed as pathogenic in only two of the three computational tools, and had a conflicting ClinVar assessment. It was present 691 times in the gnomAD cohort almost always in people of East or South Asian ancestry. Alamut considered it Likely Pathogenic based on 2 moderate (PM1-PM6) and > 4 supporting (PP1-PP5) criteria, but Varsome assessed it as Benign (PM1, PP2, PP3, BS1, BS2). This variant has been described previously as hypomorphic[41], but has also been detected on the same allele as other pathogenic *COL4A5* variants[42, 43]. A recent study with clinical data from several families concluded that it was not pathogenic[44].

**Suppl Table 4: Assessment of *COL4A3* variants in the 100,000 Genomes Project database and correlation with haematuria**

| Protein consequence | Transcript | PP2 | SIFT | M T | Cons | gnomAD (Hem/ Het, total alleles) | Previously reported; or substitution with other residue (LOVD) | ClinVar | Varsome | Predicted pathogenic | Number in 100,000 Genomes Project with haematuria | Number in 100,000 Genomes Project without haematuria |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| p.(Gly46Arg) | c.136G>A | 1 | 0 | 0.999 | Gly | 61/ 280,882 | LB | LB** | LP (PM1,PM2PP2,PP3,BP6) | NO | 0 | 10 |
| p.(Gly49Arg) | c.145G>C | 1 | 0.14 (Tol) | | Gly | None | VUS | | VUS (PM1,PM2, PP2) | NO | 0 | 1 |
| p.(Gly49Glu) | c.146G>A | 1 | 0.002 | 0.9999 | Gly | None | NO | | LP (PM1,PM2,PP2,PP3) | YES | 0 | 1 |
| p.(Gly58Arg) | c.172G>C | 1 | 0 | 0.9999 | Gly | 1/ 240,996 | Other x1 P | | LP (PM1,PM2,PM5,PP2,PP3) | YES | 0 | 1 |
| p.(Gly58Asp) | c.173G>A | 1 | 0 | 0.9999 | Gly | None | Other x1 P | | LP (PM1,PM2,PM5,PP2,PP3) | YES | 0 | 1 |
| p.(Gly58Ser) | c.172G>A | 1 | 0 | 0.9999 | Gly | 6/ 243,384 | Other x1 P | VUS* | LP (PM1,PM2,PM5,PP2,PP3) | YES | 0 | 6 |
| p.(Gly67Ter) | c.199G>T | | | | Gly | None | NO | | P (PVS1,PM2,PP3) | YES | 0 | 1 |
| p.(Gly70Arg) | c.208G>C | 1 | 0.002 | 0.9999 | Gly | None | NO | | LP (PM1,PM2,PP2,PP3) | YES | 0 | 1 |
| GNp.(Gly94Glu) | c.281G>A | 1 | 0 | 0.9999 | Gly | 1/ 249,140 | NO | | LP (PM1,PM2,PP2,PP3) | YES | 0 | 1 |
| p.(Gly94Asp) | c.281G>A | 1 | | | Gly | None | NO | | LP (PM1,PM2,PP2,PP3) | YES | 0 | 1 |
| p.(Gly97Ser) | c.292G>A | 1 | | | Gly | None | NO | | LP (PM1,PM2,PP2,PP3) | YES | 0 | 1 |
| p.(Gly97Cys) | c.292G>T | 1 | 0 | 0.9999 | Gly | None | NO | | LP (PM1,PM2,PP2,PP3) | YES | 0 | 1 |
| GNp.(Gly100Arg) | c.298G>A | 1 | 0 | 0.9999 | Gly | 1/ 249,204 | NO | | LP (PM1,PM2,PP2,PP3) | YES | 0 | 1 |
| p.(Gly112Ser) | c.334G>A | 1 | 0 | 1 | Gly | None | NO | | LP (PM1,PM2,PP2,PP3) | YES | 0 | 1 |
| p.(Gly115Asp) | c.341G>A | | 0 | 0.9999 | Gly | None | LP | P | LP (PM1,PM2,PP2,PP3,PP5) | YES | 0 | 1 |
| p.(Gly118Arg) | c.352G>A | 1 | 0 | 0.9121 | Gly | | NO | | LP (PM1,PM2,PP2,PP3) | YES | 1 | 1 |
| GNp.(Gly124Glu) | c.371G>A | 1 | 0 | 0.9999 | Gly | 1/ 249,394 | NO | | LP (PM1,PM2,PP2,PP3) | YES | 0 | 2 |
| p.(Gly139Arg) | c.415G>C | 1 | 0 | 0.9999 | Gly | None | P | VUS** | LP (PM1,PM2,PP2,PP3) | YES | 1 | 1 |
| p.(Gly148Val) | c.443G>T | 1 | 0.002 | 1 | Gly | 4/ 280,874 | LP, VUS | | LP (PM1,PM2,PP2,PP3) | YES | 1 | 1 |
| p.(Gly174Val) | c.521G>T | 1 | 0 | 1 | Gly | None | Other x2 P | | LP,PM1,PM2,PM5,PP2,PP3) | YES | 1 | 0 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| [GN]p.(Gly183Asp) | c.548G>A | 1 | 0.001 | 1 | Gly | 1/ 249,322 | Other x 3 P | | LP (PM1,PM2,PP2,PP3) | YES | 0 | 1 |
| p.(Gly201Glu) | c.602G>A | 1 | 0.002 | 1 | Gly | None | NO | | LP (PM1,PM2,PP2,PP3) | YES | 0 | 2 |
| p.(Gly204Val) | c.611G>T | 1 | 0.002 | 1 | Gly | None | NO | | LP (PM1,PM2,PP2,PP3) | YES | 0 | 1 |
| p.(Gly213Arg) | c.637G>A | 1 | 0.001 | 0.9999 | Gly | None | NO | | P (PS1,PM1,PM2,PP2,PP3) | YES | 0 | 1 |
| p.(Gly242Glu) | c.725G>A | 1 | 0.003 | 0.9999 | Gly | None | P | LP* | LP (PM1,PM2,PP2,PP3,PP5) | YES | 0 | 1 |
| p.(Gly291Glu) | c.872G>A | 1 | 0 | 1 | Gly | None | P | LP | LP (PM1,PM2,PP2,PP3,PP5) | YES | 0 | 1 |
| [GN]p.(Gly300Arg) | c.898G>A | 1 | 0 | 0.9999 | Gly | 5/ 249,426 | LP, VUS | | LP (PM1,PM2,PP2,PP3) | YES | 0 | 1 |
| p.(Gly372Ser) | c.1114G>A | 1 | 0.006 | 0.9998 | Gly | None | NO | | P (PVS1,PM1,PM2,PP2,PP3) | YES | 0 | 1 |
| p.(Gly378Glu) | c.1133G>A | 1 | 0.0002 | 0.9936 | Gly | None | LP | | LP (PM1,PM2,PP2,PP3) | YES | 1 | 0 |
| [GN]p.(Gly389Asp) | c.1166G>A | 1 | 0.002 | 0.9999 | Gly | 1/ 248,214 | NO | | LP (PM1,PM2,PP2,PP3) | YES | 0 | 1 |
| p.(Gly407Ala) | c.1220G>C | 1 | 0.008 | 1 | Ile | None | Other x 1 VUS | | LP (PM1,PM2,PM5,PP2,PP3) | NO | 1 | 0 |
| [GN]p.(Gly436Ala) | c.1307G>C | 1 | 0.01 | 0.9999 | Gly | 5/ 280,838 | NO | | LP (PM1,PM2,PP2,PP3) | YES | 0 | 2 |
| [GN]p.(Gly484Arg) | c.1450G>A | 1 | 0.018 | 0.9999 | Gly | 7/ 249,506 | NO | | P (PS1,PM1,PM2,PP2,PP3) | YES | 0 | 3 |
| p.(Gly487Val) | c.1460G>T | 1 | 0.012 | 0.9999 | Gly | None | Other x 1 P | | LP (PM1,PM2,PM5,PP2,PP3) | YES | 0 | 1 |
| p.(Gly487Ser) | c.1459G>A | 1 | 0.038 (Dam) | 0.7711 (Pol) | Gly | 31/ 280,912 | Other x 1 P | | LP (PM1,PM2,PM5,PP2,PP3) | NO | 0 | 4 |
| p.(Gly511Arg) | c.1531G>A | 1 | 0.001 | 0.9999 | Gly | None | Other x 1 LP | | LP (PM1,PM2,PP2,PP3) | YES | 0 | 1 |
| p.(Gly511Ala) | c.1532G>C | 1 | 0.003 | 0.9999 | Gly | None | Other x 1 LP | | LP (PM1,PM2,PP2,PP3) | YES | 0 | 1 |
| p.(Gly532Cys) | c.1594G>T | 1 | 0.002 | 1 | Gly | None | P or VUS; other x1 LP | P, LP** | P (PP5, PM1,PM2,PM5,Pp2,PP3) | YES | 0 | 1 |
| p.(Gly560Val) | c.1679G>T | 1 | 0 | 0.9999 | Gly | None | LP | | LP (PM1,PM2,PP2,PP3) | YES | 1 | 0 |
| p.(Gly575Arg) | c.1723G>A | 1 | 0 | 0.9999 | Gly | None | NO | | LP (PM1,PM2,PP2,PP3) | YES | 1 | 0 |
| p.(Gly575Glu) | c.1724G>A | 1 | 0 | 0.9999 | Gly | None | NO | | LP (PM1,PM2,PP2,PP3) | YES | 1 | 0 |
| p.(Gly575Val) | c.1724G>T | 1 | 0 | 1 | Gly | None | NO | | LP (PM1,PM2,PP2,PP3) | YES | 0 | 1 |
| p.(Gly602Asp) | c.1805G>A | 0.795 | 0.003 | 0.9999 | Gly | None | NO | | LP (PM1,PM2, PP2,PP3) | NO | 0 | 1 |
| p.(Gly611Glu) | c.1832G>A | 1 | 0.002 | 0.9999 | Gly | None | VUS | | LP (PM1,PM2,PM5,PP2,PP3) | YES | 0 | 1 |
| [GN].(Gly619Arg) | c.1855G>A | 1 | 0.002 | 0.9999 | Gly | 2/ 246,972 | LP, P, or VUS | | P (PS1,PM1,PM2,PP2,PP3) | YES | 1 | 0 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| [GN]p.(Gly637Arg) | c.1909G>A | 1 | 0 | 0.9999 | Gly | 2/ 246,114 | NO | | P (PS1,PM1,PM2,PP2,PP3) | YES | 0 | 1 |
| p.(Gly680AspfsTer70) | c.2031_2038dup | 1 | | | Gly | 2/ 244,894 P | P | | P (PVS1, PM2,PP3,PP5) | YES | 0 | 2 |
| p.(Gly695Glu) | c.2084G>A | 1 | 0 | 0.9999 | Gly | None | Other x 1 LP, P or VUS | | LP (PM1,PM2,PM5,PP2,PP3) | YES | 0 | 1 |
| p.(Gly695Arg) | c.2083G>A | 1 | 0 | 0.9999 | Gly | 31/ 278,152 | LP, P or VUS | Con-flicting | P (PS1, PM1,PM2,PP2,PP3) | NO | 3 | 18 |
| p.(Gly700Glu) | c.2099G>A | 1 | 0.001 | 0.9999 | Gly | None | NO | | LP (PM1,PM2,PP2,PP3) | YES | 0 | 1 |
| p.(Gly721ValfsTer26) | c.2162del | 1 | 0 | 1 | Gly | 3/ 243,442 | NO | P * | LP (PM1,PM2,PP2,PP3) | YES | 0 | 1 |
| p.(Gly730Glu) | c.2189G>A | 1 | 0 | 0.9999 | Gly | None | NO | VUS | LP (PM1,PM2,PP2,PP3) | YES | 0 | 1 |
| p.(Gly780Glu) | c.2339G>A | 1 | 0 | 1 | Gly | None | LP | | LP (PM1,PM2,PP2,PP3) | YES | 0 | 1 |
| p.(Gly792Glu) | c.2375G>A | 1 | 0 | 1 | Gly | None | P; other x1 LP | | LP (PM1,PM2,PP2,PP3) | YES | 0 | 1 |
| p.(Gly801Glu) | c.2402G>A | 1 | 0 | 1 | Gly | None | NO | | LP (PM1,PM2,PP2,PP3) | YES | 1 | 1 |
| p.(Gly807ArgfsTer28) | c.2417dup | 1 | 0.002 | 0.9999 | Gly | 1/ 172,936 | P | LP** | P (PVS1,PM2,PP3,PP5) | YES | 0 | 2 |
| p.(Gly812Ser) | c.2434G>A | 0.339 | 0.019 | 0.9986 (Pol) | Gly | 10/ 178,112 | VUS | | VUS (PM1,PM2,PP2,BP4) | NO | 0 | 6 |
| [GN]p.(Gly818Arg) | c.2452G>A | 1 | 0 | 1 | Gly | 2/ 174,872 | LP, VUS | P | P (PS1,PM1,PM2,PP2,PP3) | YES | 3 | 4 |
| p.(Gly824Glu) | c.2471G>A | 1 | 0 | 1 | Gly | None | NO | | LP (PM1,PM2,PP2,PP3) | YES | 0 | 1 |
| p.(Gly856Glu) | c.2567G>A | 0.997 | 0 | 1 | Gly | None | LP, P or VUS | | LP (PM1,PM2,PP2,PP3) | YES | 1 | 0 |
| p.(Gly874Ter) | c.2620G>T | | | 1 | Gly | None | NO | | P (PVS1,PM2,PP3 | YES | 0 | 1 |
| p.(Gly874AspfsTer9) | c.2621del | 1 | | | Gly | 10/ 280,726 | NO | LP* | P (PVS1,PM2,PP3,PP5) | YES | 1 | 6 |
| p.(Gly883Arg) | c.2647G>A | 1 | 0 | 1 | Gly | None | NO | | P (PS1,PM1,PM2,PP2,PP3) | YES | 0 | 1 |
| p.(Gly895Asp) | c.2684G>A | 1 | 0 | 1 | Gly | None | NO | LP | LP (PM1,PM2,PP2,PP3,PP5) | YES | 0 | 1 |
| p.(Gly943Arg) | c.2827G>A | 1 | 0 | 0.9999 | Gly | 2/ 249,332 | NO | | LP (PM1,PM2,PP2,Pp3) | YES | 0 | 4 |
| p.(Gly1128Ala) | c.3383G>C | 0.999 | 0 | 1 | Gly | None | NO | | LP (PM1,PM2,PP2,Pp3) | YES | 0 | 1 |
| p.(Gly1137Ser) | c.3409G>A | 1 | 0 | 1 | Gly | None | Other x 1 P | | LP (PM1,PM2,PM5,PP2,PP3) | YES | 0 | 1 |
| p.(Gly1152Val) | c.3455G>T | 1 | 0 | 1 | Gly | None | NO | | LP (PM1,PM2,PM5,PP2,PP3) | YES | 0 | 1 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| p.(Gly1158Arg) | c.3472G>C | 1 | 0 | 1 | Gly | 1/ 249,490 | VUS | Con-flicting * | LP (PM1,PM2,PP2,PP3) | NO | 0 | 1 |
| GNp.(Gly1167Arg) | c.3499G>A | 1 | 0 | 1 | Gly | 2/ 249,476 | LP, P, VUS | P,LP | P (PS1,PM1,PM2,PP2,PP3,PP5) | YES | 1 | 3 |
| p.(Gly1198Ser) | c.3592G>A | 1 | 0 | 1 | Gly | None | NO | LP | LP (PM1,PM2,PM5,PP2,PP3,PP5) | YES | 1 | 0 |
| GNp.(Gly1207Arg) | c.3619G>C | 1 | 0.532 | 0.9981 | Gly | 1/ 234,658 | Other x1, P | P | VUS (PM1PM2,PP2) | NO | 1 | 0 |
| p.(Gly1231Ser) | c.3691G>A | 1 | 0.006 | 1 | Gly | None | Other x1, P | VUS | LP (PM1,PM2,PP2,PP3) | YES | 0 | 3 |
| GNp.(Gly1254Arg) | c.3760G>C | 1 | 0.002 | 1 | Gly | 1/ 249,506 | P | | LP (PM1,PM2,PP2,PP3) | YES | 3 | 1 |
| p.(Gly1268Glu) | c.3803G>A | 1 | 0 | 1 | Gly | None | NO | | LP (PM1,PM2,PP2,PP3) | YES | 0 | 1 |
| GNp.(Gly1277Ser) | c.3829G>A | 1 | 0 | 1 | Gly | 102/ 280,750 | LP, P or VUS | VUS | P (PS1,PM1,PM2,PP2,PP3,PP5) | NO | 2 | 28 |
| p.(Gly1280Cys) | c.3838G>T | 1 | 0 | 1 | Gly | None | NO | | LP (PM1,PM2,PP2,PP3) | YES | 1 | 1 |
| p.(Gly1283Glu) | c.3848G>A | 1 | 0 | 1 | Gly | None | NO | | LP (PM1,PM2,PP2,PP3) | YES | 0 | 1 |
| p.(Gly1298Arg) | c.3892G>C | 1 | 0.001 | 1 | Gly | None | NO | | LP (PM1,PM2,PP2,PP3) | YES | 1 | 0 |
| p.(Gly1304Arg) | c.3910G>A | 1 | 0 | 1 | Gly | None | NO | | LP (PM1,PM2,PP2,PP3) | YES | 0 | 2 |
| p.(Gly1313Glu) | c.3938G>A | 1 | 0 | 1 | Gly | None | NO | | LP (PM1,PM2,PP2,PP3) | YES | 0 | 1 |
| p.(Gly1328Val) | c.3983G>T | 1 | 0.002 | 1 | Gly | 8/ 280,792 | NO | | LP (PM1,PM2,PP2,PP3) | YES | 0 | 2 |
| GNp.(Gly1367Arg) | c.4099G>C | 1 | 0 | 1 | Gly | 1/ 249,320 | NO | | LP (PM1,PM2,PP2,PP3) | YES | 0 | 1 |
| p.(Gly1376Arg) | c.4126G>C | 1 | 0.001 | 1 | Gly | None | NO | | LP (PM1,PM2,PP2,PP3) | YES | 0 | 1 |
| GNp.(Gly1397Glu) | c.4190G>A | 1 | 0.002 | 1 | Gly | 1/ 249,248 | NO | | LP (PM1,PM2,PP2,Pp3) | YES | 0 | 1 |
| GNp.(Gly1400Glu) | c.4199G>A | 1 | 0 | 1 | Gly | 2/ 249,256 | NO | | LP (PM1,PM2,PP2,PP3) | YES | 0 | 1 |
| p.(Gly1418Glu) | c.4253G>A | 1 | 0 | 1 | Gly | None | LP | | LP (PM1,PM2,PP2,PP3) | YES | 1 | 0 |
| Total number with a predicted pathogenic variant with haematuria or without respectively | | | | | | | | | | | 21 | 85 |
| Total number with haematuria or without haematuria respectively | | | | | | | | | | | 2,221 | 37,200 |

**Variants with any evidence not consistent with pathogenicity are highlighted in blue. Otherwise notes are the same as for Suppl Table 3.**

## Suppl Table 5: Assessment of *COL4A4* variants in the 100,000 Genomes Project database and correlation with haematuria

| Protein consequence | Transcript | PP2 | SIFT | M T | Cons | gnomAD (Hem/ Het, total alleles) | Previously reported; or substitution with other residue (LOVD) | ClinVar | Varsome | Predicted pathogenic | Number in 100,000 Genomes Project with haematuria | Number in 100,000 Genomes Project without haematuria |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| p.(Gly1459Val) | c.4376G>T | 0.627 | 0.044 (Dam) | 0.9994 (Pol) | Gly | 2/ 274,180 | LP | | VUS (PM1,PM2,PP2, BP4) | NO | 0 | 1 |
| GNp.(Gly1430Arg) | c.4288G>A | 1 | 0.001 | 1 | Gly | 1/ 249,468 | NO | P* | LP (PM1,PM2PP2,PP3,PP5) | YES | 0 | 1 |
| p.(Gly1418Asp) | c.4253G>A | 1 | 0.003 | 1 | Gly | None | NO | | LP (PM1,PM2PP2,PP3) | YES | 0 | 1 |
| p.(Gly1401ProfsTer31) | c.4200_4201del | 1 | | | Gly | None | P | | P (PVS1,PM2,PP3) | YES | 0 | 1 |
| p.(Gly1380Asp) | c.4139G>A | 1 | 0.001 | 0.999 | Gly | None | NO | | LP (PM1,PM2,PP2,PP3) | YES | 0 | 2 |
| GNp.(Gly1346Val) | c.4037G>T | 1 | 0.002 | 1 | Gly | 4/ 249,206 | NO | | LP (PM1,PM2,PP2,PP3) | YES | 0 | 2 |
| GNp.(Gly1325Arg) | c.3973G>C | 1 | 0 | 0.9999 | Gly | 9/ 249,006 | NO | | P (PVS1,PM2,PP2,PP3) | YES | 0 | 4 |
| p.(Gly1319Arg) | c.3955G>A | 1 | 0 | 0.9999 | Gly | None | NO | | LP (PM1,PM2,PP2,PP3) | YES | 0 | 1 |
| p.(Gly1295Asp) | c.3884G>A | 1 | 0.002 | 0.9998 | Gly | None | NO | | LP (PM1,PM2,PP2,PP3) | YES | 0 | 1 |
| p.(Gly1292Asp) | c.3875G>A | 1 | 0 | 1 | Gly | None | LP | LP | LP (PM1,PM2,PP2,PP3,PP5) | YES | 1 | 0 |
| p.(Gly1258Asp) | c.3773G>A | 1 | 0.005 | 0.998 | Gly | None | NO | | LP (PM1,PM2,PP2,PP3) | YES | 0 | 2 |
| GNp.(Gly1248Glu) | c.3743G>A | 1 | 0.003 | 0.9999 | Gly | 17/ 280,842 | VUS | LP* | LP (PM1,PM2,PP2,PP3,PP5) | YES | 0 | 1 |
| p.(Gly1230Asp) | c.3689G>A | 1 | 0.003 | 0.9999 | Gly | None | NO | | LP (PM1,PM2,PP2,PP3) | YES | 0 | 1 |
| GNp.(Gly1201Asp) | c.3602G>A | 1 | 0 | 0.9999 | Gly | 1/ 249,040 | LP | | LP (PM1,PM2,PM5,PP2,PP3) | YES | 0 | 1 |
| GNp.(Gly1178Ser) | c.3532G>A | 1 | 0 | 0.9999 | Gly | 24/ 280,386 | LP, VUS | | LP (PM1,PM2,PP2,PP3) | YES | 0 | 3 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| [GN]p.(Gly1151Ala) | c.3452G>C | 1 | 0 | 1 | Gly | 10/249,432 | NO | | LP (PM1,PM2,PP2.PP3) | YES | 0 | 7 |
| [GN]p.(Gly1136Ala) | c.3407G>C | 1 | 0 | 0.9999 | Gly | 3/280,616 | NO | | LP (PM1,PM2,PP2,PP3) | YES | 0 | 1 |
| p.(Gly1124Val) | c.3371G>T | 1 | 0.001 | 1 | Gly | None | NO | | LP (PM1,PM2,PP2,PP3) | YES | 0 | 1 |
| [GN]p.(Gly1103Arg) | c.3307G>A | 1 | 0.001 | 0.9999 | Gly | 5/249,156 | P, VUS | LP* | LP (PM1,PM2,PP2,PP3,PP5) | YES | 0 | 2 |
| p.(Gly1082Val) | c.3245G>T | 1 | 0.002 | 1 | Gly | None | NO | | LP (PM1,PM2,PP2,PP3) | YES | 0 | 1 |
| p.(Gly1069SerfsTer4) | c.3204_3205insTCTT | 1 | | | Gly | None | NO | | P (PVS1,PM2,PP3) | YES | 2 | 0 |
| p.(Gly1018Arg) | c.3052G>C | 1 | 0 | 1 | Gly | 4/249,580 | NO | | LP (PM1,PM2,PP2,PP3) | YES | 0 | 2 |
| p.(Gly1015Arg) | c.3043G>A | 1 | 0.001 | 1 | Gly | None | Other x1 LP | | LP (PM1,PM2,PM5,PP2,PP3) | YES | 0 | 1 |
| p.(Gly1015Glu) | c.3044G>A | 1 | 0.002 | 1 | Gly | 14/280,944 | Other x1 LP | Con-flicting* | LP ( (PM1,PM2,PP2,PP3,PP5) | NO | 1 | 8 |
| [GN]p.(Gly996Arg) | c.2986G>A | 1 | 0 | 1 | Gly | 10/280,858 | P, VUS | P* | LP (PM1,PM2,PP2,PP3,PP5) | YES | 0 | 4 |
| [GN] p.(Gly990Asp) | c.2969G>A | 1 | 0 | 1 | Gly | 4/249,330 | NO | | LP (PM1,PM2,PP2,PP3) | YES | 0 | 1 |
| [GN] p.(Gly963Glu) | c.2888G>A | 0.997 | 0.272 (Tol) | 0.9999 | Gly | 2/249,566 | NO | | LP (PM1.PM2,PP2,PP3) | NO | 1 | 0 |
| [GN] p.(Gly960Arg) | c.2878G>A | 1 | 0.002 | 1 | Gly | 3/249,540 | NO | LP or P ** | P (PS1,PM1,PM2,PP2,PP3,PP5) | YES | 0 | 1 |
| p.(Gly948Ala) | c.2843G>C | 1 | 0 | 0.9999 | Gly | 3/280,952 | NO | | LP (PM1,PM2,PP2,PP3) | YES | 1 | 0 |
| p.(Gly939Ser) | c.2815G>A | 1 | 0 | 0.9999 | Gly | None | NO | | LP (PM1,PM2,PP2,PP3) | YES | 0 | 1 |
| p.(Gly927Glu) | c.2780G>A | 1 | 0 | 1 | Gly | None | NO | | LP(PM1,PM2,PP2,PP3) | YES | 0 | 1 |
| p.(Gly873Arg) | c.2617G>A | 1 | 0 | 1 | Gly | None | P, LP | Con-flicting* | P (PM1,PM2,PP2,PP3) | NO | 1 | 1 |
| [GN]p.(Gly864Arg) | c.2590G>A | 1 | 0 | 1 | Gly | 1/31,324 | LP, P | LP* | LP (PM1,PM2,PP2,PP3,PP5) | YES | 0 | 2 |
| p.(Gly861Glu) | c.2582G>A | 1 | 0.001 | 0.9999 | Gly | None | NO | | LP (PM1,PM2,PP2,PP3) | YES | 1 | 0 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| p.(Gly840Arg) | c.2518G>A | 1 | 0 | 1 | Gly | None | NO | | LP (PM1,PM2,PP2,PP3) | YES | 1 | 0 |
| p.(Gly837Arg) | c.2509G>C | 1 | 0 | 1 | Gly | None | Other x1, P | | LP (PM1,PM2,PM5,PP2,PP3) | YES | 0 | 2 |
| p.(Gly825Ala) | c.2474G>C | 0.125 | 0.048 | 0.9981 | Gly | None | NO | | LP (PM1,PM2,PP2,PP3) | NO | 0 | 1 |
| GNp.(Gly816Val) | c.2447G>T | 1 | 0 | 1 | Gly | 4/ 249,158 | NO | | LP (PM1,PM2,PP2,PP3) | YES | 0 | 2 |
| p.(Gly816Glu) | c.2447G>A | 1 | 0 | 1 | Gly | 24/ 249,158 | NO | | LP (PM1,PM2,PP2,PP3) | YES | 0 | 1 |
| GNp.(Gly801Ala) | c.2402G>C | 1 | 0 | 0.9999 | Gly | 1/ 249,424 | NO | | LP (PM1,PM2,PP2,PP3) | YES | 0 | 1 |
| p.(Gly801Glu) | c.2402G>A | 1 | 0 | 1 | Gly | None | NO | | LP (PM1,PM2,PP2,PP3) | YES | 0 | 1 |
| p.(Gly798Ser) | c.2392G>A | 1 | 0.005 | 0.9999 | Gly | 3/ 280,816 | NO | VUS* | LP (PM1,PM2,PP2,PP3) | YES | 0 | 3 |
| p.(Gly783Arg) | c.2347G>A | 1 | 0 | 1 | Gly | None | NO | VUS* | LP (PM1,PM2,PP2,PP3) | YES | 0 | 1 |
| GNp.(Gly774Arg) | c.2320G>C | 1 | 0 | 1 | Gly | 19/ 280,942 | LP,P, VUS | LP** | LP (PM1,PM2,PP2,PP3,PP5) | YES | 0 | 1 |
| GNp.(Gly765Val) | c.2294G>T | 1 | 0.009 | 1 | Gly | 9/ 249,576 | NO | | LP (PM1,PM2,PP2,PP3) | YES | 0 | 2 |
| p.(Gly751Arg) | c.2251G>A | 1 | 0 | 1 | Gly | None | NO | | LP (PM1,PM2,PP2,PP3) | YES | 0 | 1 |
| GNp.(Gly748Ser) | c.2242G>A | 0.954 | 0 | 1 | Gly | 9/ 280,758 | NO | LP** | LP (PM1,PM2,PP2,PP3,PP5) | YES | 2 | 3 |
| GNp.(Gly734Ser) | c.2200G>A | 1 | 0.009 | 0.9999 | Gly | 2/ 249,028 | NO | | LP (PM1,PM2,PP2,PP3) | YES | 0 | 1 |
| p.(Gly698Arg) | c.2092G>A | 1 | 0 | 1 | Gly | None | NO | VUS* | LP (PM1,PM2,PP2,PP3) | YES | 1 | 0 |
| p.(Gly660Asp) | c.1979G>A | 1 | 0.006 | 0.9999 | Gly | None | LP | | LP (PM1,PM2,PP2,PP3) | YES | 0 | 1 |
| p.(Gly651Cys) | c.1951G>T | 1 | 0 | 0.9999 | Gly | None | Other x1, LP | | LP (PM1,PM2,PP2,PP3) | YES | 0 | 1 |
| p.(Gly625Glu) | c.1874G>A | 1 | 0.002 | 0.9999 | Gly | None | NO | | LP (PM1,PM2,PP2,PP3) | YES | 1 | 1 |
| p.(Gly596Arg) | c.1786G>A | 1 | 0 | 1 | Gly | None | NO | | LP (PM1,PM2,PP2,PP3) | YES | 0 | 1 |
| p.(Gly551TrpfsTer8) | c.1649dup | 1 | | | Gly | None | LP | | LP (PM1,PM2,PP2,PP3) | YES | 1 | 0 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| p.(Gly545Asp) | c.1634G>A | 1 | 0 | 0.9999 | Gly | None | B,VUS | | LP (PM1,PM2,PP2,PP3) | NO | 0 | 1 |
| p.(Gly533Asp) | c.1598G>A | 0.999 | 0 | 0.9982 | Gly | None | LP, P | P* | LP (PM1,PM2,PP2,PP3,PP5) | YES | 2 | 1 |
| p.(Gly524Glu) | c.1571G>A | 0.576 | 0.002 | 0.9999 | Gly | 1/ 249,290 | LB | | LP (PM1,PM2,PP2,PP3) | NO | 0 | 1 |
| p.(Gly503Trpfs Ter12) | c.1505dup | 1 | | | Gly | None | NO | | LP (PM1,PM2,PP2,PP3) | YES | 0 | 1 |
| p.(Gly500Val) | c.1499G>T | 1 | 0.03 | 1 | Gly | None | NO | | LP (PM1,PM2,PP2,PP3) | YES | 0 | 1 |
| p.(Gly500Ala) | c.1499G>C | 1 | 0.186 Tol) | 0.9998 | Gly | None | NO | | LP (PM1,PM2,PP2,PP3) | NO | 0 | 1 |
| p.(Gly479Arg) | c.1435G>C | 0 | 0.466 (Tol) | 0.9975 (Pol) | Gly | 106/ 280,814 | | Con-flicting* | VUS (PM1,PM2,PP2, BP4) | NO | 2 | 8 |
| p.(Gly475Ala) | c.1424G>C | 0.991 | 0 | 0.9999 | Gly | None | NO | LP | LP (PM1,PM2,PM5,PP2,PP3) | YES | 0 | 1 |
| p.(Gly451Asp) | c.1352G>A | 1 | 0.002 | 0.9999 | Gly | None | Other x1 VUS | | LP (PM1,PM2,PP2,PP3) | YES | 1 | 0 |
| p.(Gly442Ser) | c.1324G>A | 0.997 | 0.06 | 0.9999 | Gly | None | NO | | LP (PM1,PM2,PP2,PP3) | YES | 0 | 1 |
| GNp.(Gly426Arg) | c.1276G>A | 1 | 0.001 | 0.9999 | Gly | 4/ 248,782 | NO | | LP (PM1,PM2,PP2,PP3) | YES | 0 | 3 |
| GNp.(Gly402Asp) | c.1205G>A | 1 | 0.003 | 0.9999 | Gly | 5/ 247,742 | NO | | LP (PM1,PM2,PP2,PP3) | YES | 0 | 3 |
| p.(Gly382Ala) | c.1145G>C | 0.999 | 0 | 0.9999 | Gly | None | NO | | LP (PM1,PM2,PP2,PP3) | YES | 1 | 0 |
| p.(Gly379Arg) | c.1135G>A | 1 | 0 | 0.9999 | Gly | None | NO | | LP (PM1,PM2,PP2,PP3) | YES | 1 | 0 |
| p.(Gly379Ala) | c.1136G>C | 0.999 | 0 | 0.9999 | Gly | None | NO | | LP (PM1,PM2,PP2,PP3) | YES | 0 | 1 |
| GNp.(Gly373Glu) | c.1118G>A | 1 | 0 | 0.9999 | Gly | 2/ 248,186 | LP, VUS | P* | LP (PM1,PM2,PP2,PP3,PP5) | YES | 2 | 2 |
| p.(Gly335Val) | c.1004G>T | 1 | 0 | 1 | Gly | None | NO | | LP (PM1,PM2,PP2,PP3) | YES | 1 | 0 |
| p.(Gly308Glu) | c.923G>A | 1 | 0 | 0.9999 | Gly | None | NO | | LP (PM1,PM2,PP2,PP3) | YES | 0 | 2 |
| p.(Gly291Glu) | c.872G>A | 1 | 0.004 | 1 | Gly | None | NO | | LP (PM1,PM2,PP2,PP3) | YES | 0 | 1 |
| p.(Gly252Ser) | c.754G>A | 1 | 0.009 | 0.9999 | Gly | None | Other x 1 VUS | | LP (PM1,PM2,PM5,PP2,PP3) | YES | 0 | 1 |
| GNp.(Gly190Arg) | c.568G>C | 1 | 0 | 1 | Gly | 1/ 249,424 | NO | | LP (PM1,PM2,PP2,PP3) | YES | 1 | 0 |
| p.(Gly164Val) | c.491G>T | 1 | 0 | 1 | Gly | None | NO | | LP (PM1,PM2,PP2,PP3) | YES | 0 | 1 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| p.(Gly161Val) | c.482G>T | 1 | 0 | 1 | Gly | 4/ 249,086 | VUS; Other x2 LP, VUS | LP* | LP (PM1,PM2,PM5,PP2,PP3,PP5) | YES | 0 | 2 |
| GNp.(Gly161Arg) | c.481G>C | 1 | 0 | 1 | Gly | 3/ 280,524 | NO | LP* | LP (PM1,PM2,PM5,PP2,PP3,PP5) | YES | 0 | 2 |
| p.(Gly143Asp) | c.428G>A | 1 | 0 | 1 | Gly | None | NO | | LP (PM1,PM2,PP2,PP3) | YES | 0 | 1 |
| GNp.(Gly137Ser) | c.409G>A | 1 | 0 | 1 | Gly | 1/ 248,804 | LP, VUS | | LP (PM1,PM2,PP2,PP3) | YES | 0 | 1 |
| p.(Gly122Ser) | c.364G>A | 1 | 0 | 0.9999 | Gly | None | NO | | LP (PM1,PM2,PP2,PP3) | YES | 0 | 1 |
| p.(Gly113Asp) | c.338G>A | 1 | 0 | 0.9999 | Gly | 14/ 257,452 | NO | VUS | LP (PM1,PM2,PP2,PP3) | YES | 0 | 2 |
| GNp.(Gly98Ser) | c.292G>A | 1 | 0 | 0.9999 | Gly | 1/ 247,318 | NO | | LP (PM1,PM2,PP2,PP3) | YES | 0 | 3 |
| p.(Gly92Glufs Ter2) | c.275del | 1 | 0 | 1 | Gly | None | NO | | LP (PM1,PM2,PP2,PP3) | YES | 0 | 1 |
| Total number with a predicted pathogenic variant with and without haematuria respectively | | | | | | | | | | | 19 | 95 |
| Total number with haematuria and without haematuria respectively | | | | | | | | | | | 2,221 | 37,200 |

**Variants with any evidence not consistent with pathogenicity are highlighted in blue. Otherwise notes are the same as for Suppl Table 3.**

**Suppl Table 6: Predicted pathogenic variants in *COL4A5* in the Exome Variant Server database**

| Protein consequence | Transcript | Affected individuals/ total allele count | PP2 | SIFT | Mutation taster | Conserved | Previously reported; or substitution with other residue (LOVD) | ClinVar | Varsome | Alamut | Predicted pathogenic |
|---|---|---|---|---|---|---|---|---|---|---|---|
| GNp.(Gly42Ser) | c.124G>A | 1M/ 10,563 | Benign (0.013) | Damaging (0.047) | Disease causing (0.99) | Gly | NO | | Likely pathogenic (PM1, PM2, PP2, PP3) | Weak | NO |
| GNp.(Gly624Asp) | c.1871G>A | 1F/ 10,563 | Prob damaging (0.999) | Damaging (0.119) | Disease causing (1.00) | Gly | Slajpah, 2007 | Pathogenic ** | Pathogenic (PM1, PP2, PP3, PP5) | Strong | YES |
| p.(Gly905Ser) | c.2713G>A | 1F/10,563 | Probably damaging 1.00 | Damaging (0.00) | Disease causing (1.00) | Gly | NO | | Likely pathogenic (PM1, PM2, PP2, PP3) | Strong | YES |
| GNp.(Gly953Val) | c.2858G>T | 2M/ 10,563 | Probably damaging 1.00 | Damaging (0.002) | Disease causing (1.00) | Gly | Knebelmann, 1996 | Conflicting B, LB, VUS | Benign (PM1, PP2, PP3,PP5, BS1, BS2) | Weak | NO |
| GNp.(Gly1074Arg) | c.3220G>C | 1M/ 10,563 | Benign (0.09) | Damaging (0) | Disease causing (0.999) | Gly | NO | | Likely pathogenic (PM1, PM2,PP2,PP3) | Weak | NO |
| GNp.(Gly1244Ser) | c.3730G>A | 1F/ 10,563 | Benign (0.315) | Damaging (0.06) | Disease causing (1.00) | Gly | NO | | Likely pathogenic (PM1, PM2, PM5, PP2,PP3) | | NO |
| GNp.(Gly1424Ser) | c.4270G>A | 1M/ 10,563 | Probably damaging 1.00 | Damaging (0) | Disease causing (1.00) | Gly | LOVD, p.(Gly1424Glu), Zhang, 2011 | | Likely pathogenic, PM2,PM3,PM5,PP2,PP3 | Weak | YES |

Evidence against pathogenicity is highlighted in blue. Three variants were predicted to be pathogenic in the cohort of 7042 individuals; p.(Gly624Asp), p.(Gly905Ser); and p.(Gly1424Ser). This cohort had no frameshift, termination codons or splice site variants. M male; F female. PP2 Polyphen-2, SIFT, Mutation taster computational tools assess pathogenicity. Conserved residue in vertebrates. ClinVar – star system used for quality of assertion * to ****. P Pathogenic, LP, Likely pathogenic, B – benign, LB likely benign, VUS – variant of uncertain significance. PVS, PM, PP, BS are ACMG/AMP criteria referring to criteria for P pathogenicity; and B benign nature; p.(Leu691Phefs*7)[1] where [1]read depth = 14 was excluded where all other read depths were > 30.

Suppl Table 7: Predicted pathogenic variants in *COL4A5* in the TOPMed database

| Protein consequence | Transcript | Affected individuals/ total allele count = 125,568 | PP2 | SIFT | Mutation taster | Conserved | Previous reports of variant, or other variants at this location (LOVD) | ClinVar | Varsome | Predicted pathogenicity |
|---|---|---|---|---|---|---|---|---|---|---|
| In gnomAD | c.231+2T>C | 1F | | | Disease-causing | | | Not found | Pathogenic (PVS1, PM2,PP3) | YES |
| | c.1024_1032+22delCCTGGACTTGTAAGTTTTTTTTTTTTAGTCT | 1F | | | Disease-causing | | | Not found | Pathogenic (PVS1,PM2,PP3) | YES |
| p.(Gly57Arg) | c.169G>C | 1F | 1.00 | Damaging (0) | Disease-causing | Gly | NO | Not found | Likely pathogenic (PM1,PM2,PP2,PP3) | YES |
| p.(Gly111Ser) | c.331G>A | 1F | 1.00 | Damaging (0) | Disease-causing | Gly | NO | Not found | Likely pathogenic (PM1,PM2,PP2,PP3) | YES |
| GNp.(Gly279Arg | c.835G>C | 2F | 0.999 | Damaging (0.003) | Disease-causing | Gly | NO | Not found | Likely pathogenic (PM1,PM2,PP2,PP3) | YES |
| p.(Gly282Ser) | c.844G>A | 1F | 1.00 | Damaging (0) | Disease-causing | Gly | NO | Not found | Likely pathogenic (PM1,PM2,PP2,PP3) | YES |
| GNp.(Gly512Glu) | c.1535G>A | 2F | 1.00 | Damaging | Disease-causing | Gly | NO | Not found | Likely pathogenic (PM1,PM2,PP2,PP3) | YES |
| GNp.(Gly624Asp) | c.2073G>A | 2M, 7F | 0.99 | Damaging (0) | Disease-causing | Gly | YES | Pathogenic ** | Pathogenic (PM1, PP2, PP3, PP5) | YES |
| GNp.(Gly752Val) | c.2255G>T | 1F | 1.00 | Damaging (0.031) | Disease-causing | Gly | NO | Not found | Likely pathogenic (PM1,PM2,PP2,PP3) | YES |
| p.(Gly769Glu) | c.2306G>A | 1F | 1.00 | Damaging (0) | Disease-causing | Gly | NO | Not found | Likely pathogenic (PM1,PM2,PM5,PP2,PP3) | YES |
| p.(Gly817Arg) | c.2449G>C | 2M, 4F | 1.00 | Damaging (0.001) | Disease-causing | Gly | NO | Not found | Likely pathogenic (PM1,PM2,PP2,PP3) | YES |
| p.(Gly905Ser) | c.2713G>A | 2F | 1.00 | Damaging (0.001) | Disease-causing | Gly | NO | Not found | Likely pathogenic (PM1,PM2,PM5,PP2,PP3) | YES |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| p.(Gly973Ser) | c.2917G>A | 1F | 1.00 | Damaging (0) | Disease-causing | Gly | NO | Not found | Pathogenic (PVS1, PM1,PM2,PM5, PP2,PP3) | YES |
| p.(Gly1069Val) | c.3206G>T | 1F | 1.00 | Damaging (0) | Disease-causing | Gly | VUS | Conflicting * | Likely pathogenic (PM1,PM2,PP2,PP3) | NO |
| p.(Gly1113Ala) | c.3338G>C | 1F | 1.00 | Damaging (0.01) | Disease-causing | Gly | NO | Not found | Likely pathogenic (PM1,PM2,PP2,PP3) | YES |
| p.(Gly1119Ala) | c.3356G>C | 1F | 0.999 | Damaging (0) | Disease-causing | Gly | NO | Not found | Likely pathogenic (PM1,PM2,PP2,PP3) | YES |
| p.(Gly1153Arg) | c.3457G>C | 3M, 2F | 0.077 | Tol (0.515) | Pol (0.9981) | Gly | NO | Not found | VUS (PM1,PP2,BP4) | NO |
| p.(Gly1303Cys) | c.3907G>T | 1M, 1F | 1.00 | 0 | Disease-causing | Gly | NO | Not found | Likely pathogenic (PM1,PM2,PP2,PP3) | YES |
| p.(Gly1330Ser) | c.3988G>A | 1F | 0.998 | 0 | Disease-causing | Gly | NO | Not found | Likely pathogenic (PM1,PM2,PP2,PP3) | YES |
| p.(Gly1339Ala) | c.4016>C | 1M, 1F | 1.00 | Damaging (0.007) | Disease-causing | Gly | NO | Not found | Pathogenic (PM1,PM2,PM5,PP2 ,PP3) | YES |
| [GN]p.(Gly1424Ser) | c.4270G>A | 3F | Probably damaging 0.998 | Damaging (0) | Disease causing (1.00) | Gly | LOVD, p.(Gly1424Glu), Zhang, 2011 | Not found | Likely pathogenic (PM1,PM2,PM5,PP2 ,PP3) | YES |
| | | 6M,34F in 83,712 = 1/2093 | | | | | | | | |

[GN] also in gnomAD

**Suppl Table 8: Mitigating features for the clinical effects of predicted pathogenic *COL4A5* variants resulting in position 1 Gly substitutions in gnomAD: associations with gender, location, and replacement residues**

| Mutation | Alleles | Gender | Exons 1-20 | Adjacent to non-collagenous domain or interruption | Substitution with Ala, Ser, Cys |
|---|---|---|---|---|---|
| p.(Gly69Arg) | 1 | F | Yes | | |
| p.(Gly84Glu) | 1 | F | Yes | | |
| p.(Gly105Ala) | 1 | M | Yes | | Yes |
| p.(Gly114Ala) | 1 | F | Yes | | Yes |
| p.(Gly126Glu) | 1 | F | Yes | | |
| p.(Gly201Ala) | 1 | M | Yes | | Yes |
| p.(Gly279Ala) | 1 | M | Yes | | Yes |
| p.(Gly283Val) | 1 | F | Yes | Yes | |
| p.(Gly286Cys) | 1 | F | Yes | | Yes |
| p.(Gly356Glu) | 1 | M | Yes | Yes | |
| p.(Gly356Ala) | 1 | F | Yes | Yes | Yes |
| **p.(Gly512Glu)*** | **1** | **M** | | | |
| p.(Gly536Ser) | 1 | F | | | Yes |
| p.(Gly624Asp) | 16 | M (n=4), F (n=12) | | Yes | |
| p.(Gly656Ser) | 1 | M | | Yes | Yes |
| p.(Gly702Ser) | 1 | F | | | Yes |
| p.(Gly752Val) | 2 | M (n=2) | | Yes | |
| p.(Gly787Ala) | 1 | F | | | Yes |
| **p.(Gly822Glu)*** | **1** | **M** | | | |
| p.(Gly893Val) | 1 | F | | | |
| p.(Gly967Arg) | 1 | F | | | |
| **p.(Gly1054Asp)*** | **1** | **M** | | | |
| p.(Gly1134Cys) | 1 | F | | | Yes |
| p.(Gly1170Ser) | 1 | F | | | Yes |
| p.(Gly1185Ser) | 1 | F | | | Yes |
| p.(Gly1249Arg) | 1 | F | | Yes | |
| **p.(Gly1282Glu)** | **1** | **M** | | | |
| p.(Gly1300Ala) | 1 | F | | | Yes |
| p.(Gly1303Ala) | 1 | M | | | Yes |
| p.(Gly1321Val) | 1 | F | | | |
| p.(Gly1324Glu) | 1 | F | | | |
| p.(Gly1333Cys) | 1 | M | | | Yes |
| p.(Gly1394Cys) | 2 | F (n=2) | | | Yes |
| p.(Gly1424Ser) | 1 | F | | | Yes |
| N=34 | 51 | 17M, 34F | N=11 | N=7 | N=18 |

M male, F female * variants where there were no biochemical features potentially mitigating clinical phenotype

**Suppl Table 9: Variants in *COL4A3* and *COL4A4* in gnomAD with inconsistent assessments subsequently excluded from population frequency studies**

| hg19 | rs ID | Variant | Transcript | PP2 | SIFT | MT | Conserved in vertebrates | Clin Var | Varsome | gnomad | Cohort size | A | L | AJ | EA | F | EUR | O | SA |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ***COL4A3*** | | | | | | | | | | | | | | | | | | | |
| 228102723 | rs13424243 | p.(Gly43Arg) | c.127G>A | 1 | DEL | P | Ser | Benign | VUS (PM1,PM2,PP2,BP4) | 3 | 280794 | 0 | 0 | 1 | 0 | 0 | 2 | 0 | 0 |
| 228102723 | rs13424243 | p.(Gly43Arg) | c.127G>C | 1 | DEL | P | Ser | Benign | Benign (PM1,PP2,BA1,BP4,BP6) | 80620 | 280794 | 7524 | 7749 | 3493 | 740 | 8390 | 46200 | 2242 | 4282 |
| 228102723 | rs13424243 | p.(Gly43Trp) | c.127G>T | 1 | DEL | P | Ser | Benign | VUS (PM1,PM2,PP2,BP4) | 80623 | 280794 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 228102724 | rs776294835 | p.(Gly43Glu) | c.128G>A | 1 | DEL | P | Ser | Not found | VUS (PM1,PM2,PP2,BP4) | 3 | 249528 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 228102732 | rs200866082 | p.(Gly46Arg) | c.136G>A | 1 | DEL | DC | Gly | Likely benign | Likely pathogenic (PM1,PM2,PP2,PP3,BP6) | 61 | 280882 | 57 | 3 | 0 | 0 | 0 | 0 | 1 | 0 |
| 228131759 | rs745472969 | p.(Gly487Arg) | c.1459G>C | 1 | DEL | 0.606 - P | Gly | VUS | Likely pathogenic (PM1,PM2,PP2,PP3) | 1 | 249514 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 228142243 | rs1331805495 | p.(Gly700Val) | c.2099G>T | 0.977 | TOL (0.06) | DC | Gly | Not found | Likely pathogenic (PM1,PM2,PP2,PP3) | 1 | 245134 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 228145668 | rs774838919 | p.(Gly812Ser) | c.2434G>A | 0.339 | TOL (0.21) | P | Cys | Not found | VUS (PM1, PM2,PP2,BP4) | 10 | 178112 | 0 | 4 | 0 | 0 | 0 | 5 | 1 | 0 |
| 228142227 | rs200287952 | p.(Gly695Arg) | c.2083G>A | 1 | TOL 0.08 | DC | Gly | Pathogenic | Pathogenic (PM1,PM2,PP2,PP3,PP5) | 31 | 278152 | 1 | 0 | 0 | 0 | 0 | 27 | 3 | 0 |
| 228104886 | rs184730597 | p.(Gly58Ser) | c.172G>A | 1 | DEL | DC | Gly | VUS | Likely pathogenic (PM1,PM2,PP2,PP3) | 6 | 243384 | 0 | 3 | 0 | 0 | 0 | 2 | 1 | 0 |
| 228112275 | rs775373641 | p.(Gly148Val) | c.443G>T | 1 | DEL | DC | Gly | VUS | Likely pathogenic (PM1,PM2,PP2,PP3) | 4 | 280874 | 0 | 0 | 1 | 0 | 0 | 3 | 0 | 0 |
| 228113159 | rs764451365 | p.(Gly157Arg) | c.469G>C | 1 | DEL | DC | Gly | VUS | Likely pathogenic (PM1,PM2,PP2,PP3) | 4 | 249350 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 2 |
| 228131759 | rs745472969 | p.(Gly487Ser) | c.1459G>A | 1 | DEL | DC | Gly | VUS | Likely pathogenic (PM1,PM2,PM5,PP2,PP3) | 31 | 280912 | 0 | 15 | 0 | 0 | 0 | 10 | 1 | 5 |
| 228149007 | rs1265432530 | p.(Gly943Arg) | c.2827G>A | 1 | DEL | DC | Gly | VUS | Likely pathogenic (PM1,PM2,PP2,PP3) | 2 | 249332 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 |
| 228168602 | rs372237167 | p.(Gly1328Ala) | c.3983G>C | 1 | DEL | DC | Gly | VUS | Likely pathogenic (PM1,PM2,PP2,PP3) | 1 | 249382 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| ***COL4A4*** | | | | | | | | | | | | | | | | | | | |
| 227984654 | rs1370340334 | p.(Gly110Ala) | c.329G>C | 1 | TOL | DC | Gly | Not found | Likely pathogenic (PM1,PM2, PP2,PP3) | 1 | 221310 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 227984645 | rs766085522 | p.(Gly113Asp) | c.338G>A | 1 | TOL | DC | Gly | Not found | Likely pathogenic (PM1,PM2, PP2,PP3) | 14 | 257452 | 9 | 5 | 0 | 0 | 0 | 0 | 0 | 0 |
| 227983440 | rs377511303 | p.(Gly137Asp) | c.410G>A | 1 | DEL | DC | Gly | VUS | Likely pathogenic (PM1,PM2, PP2,PP3) | 2 | 248802 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 227983392 | rs773360119 | p.(Gly153Val) | c.458G>T | 0.712 | DEL | DC | Gly | Not found | VUS (PM1,PM2,PP2,BP4) | 2 | 249180 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 |

| | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 227958987 | rs1026613471 | p.(Gly408Glu) | c.1223G>A | 1 | DEL | DC | Gly | Conflicting; Benign | Likely pathogenic (PM1,PM2,PP2,PP3) | 6 | 248574 | 0 | 0 | 5 | 0 | 0 | 1 | 0 | 0 |
| 227954608 | rs202210475 | p.(Gly479Arg) | c.1435G>C | 0 | TOL | P | Gly | Conflcting;VUS | VUS (PM1,PM2,PP2,BP4) | 106 | 280814 | 0 | 6 | 75 | 0 | 0 | 16 | 8 | 1 |
| 227946893 | rs1800516 | p.(Gly545Ala) | c.1634G>C | 1 | DEL | P | Gly | Benign | Benign (PM1,PP2, PP3,BS1,BS2,BP6) | 7694 | 280884 | 798 | 675 | 446 | 0 | 548 | 4724 | 262 | 241 |
| 227924914 | rs781014928 | p.(Gly701Val) | c.2102G>T | 0.82 | TOL | DC | Gly | Not found | Likely pathogenic (PM1,PM2, PP2,PP3) | 1 | 248786 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 227922308 | rs760803228 | p.(Gly798Ser) | c.2392G>A | 1 | DEL | DC | Gly | VUS | Likely pathogenic (PM1,PM2,PP2,PP3) | 3 | 280816 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 0 |
| 227920820 | rs752296059 | p.(Gly853Arg) | c.2557G>A | 0 | TOL | P | Gly | Not found | VUS (PM1,PM2,PP2,BP4) | 1 | 249324 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| 227915847 | rs13027659 | p.(Gly999Glu) | c.2996G>A | 1 | DEL | DC | Gly | Conflicting; benign; likely benign | VUS (PM1, PP2,PP3,PP5, BS1,BS2) | 3321 | 280922 | 68 | 278 | 27 | 0 | 228 | 2508 | 104 | 108 |
| 227915821 | rs371172166 | p.(Gly1008Arg) | c.3022G>A | 1 | DEL | DC | Gly | VUS | Likely pathogenic (PM1,PM2,PP2,PP3) | 9 | 280942 | 6 | 2 | 0 | 0 | 0 | 1 | 0 | 0 |
| 227915799 | rs764323652 | p.(Gly1015Glu) | c.3044G>A | 1 | DEL | DC | Gly | Conflicting; Likely pathogenic; VUS | Likely pathogenic (PM1,PM2,PP2,PP3) | 14 | 280944 | 0 | 0 | 0 | 0 | 0 | 14 | 0 | 0 |
| 227915754 | rs772699709 | p.(Gly1030Val) | c.3089G>T | 1 | DEL | DC | Gly | VUS | Likely pathogenic (PM1, PM2,PM5,PP2,PP3,PP5) | 1 | 249562 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 227983440 | rs377511303 | p.(Gly137Asp) | c.410G>A | 1 | DEL | DC | Gly | VUS | Likely pathogenic (PM1,PM2, PP2,PP3) | 2 | 248802 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 227876903 | rs779769090 | p.(Gly1443Arg) | c.4327G>A | 0.005 | DEL (0.22) | P | Gly | Not found | VUS (PM1,PM2,PP2,BP4) | 1 | 249498 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 227876903 | rs779769090 | p.(Gly1443Arg) | c.4327G>C | 0.005 | DEL (0.22) | P | Gly | Not found | VUS (PM1,PM2,PP2,BP4) | 13 | 249498 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 13 |
| 227875175 | rs1287040507 | p.(Gly1459Val) | c.4376G>T | 0.627 | DEL | P | Gly | Not found | VUS (PM1,PM2,PP2,BP4) | 2 | 274180 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 |

**Evidence supporting pathogenicity is highlighted in red; and not supporting pathogenicity is in blue. PP2, Polyphen-2, SIFT, MT Mutation taster. PP2 scores >0.8 indicate Pathgoenic or Likely Pathogenic. SIFT scores DEL for deleterious, TOL for tolerated. MT scores DC for disease-causing and P for polymorphism. Varsome online tool to assess ACMG criteria (varsome.com), gnomad variant database used to assess variant prevalence (approximate population n= 256, 562 for *COL4A3*, and 233,529 for *COL4A4*). Cohort size varied for different variants, A African (n = 15,944) L Latino (n=32,544), AJ Ashenazi Jewish (n=9528), EA East Asian (n=17,021), F Finnish (n=21,559), Eur European (non-Finnish, n=111,786), Other (n=6112), SE South Asian (n=28,471)**

Of especial interest were p.(Gly545Ala) and p.(Gly999Glu) in *COL4A4*, which were not predicted pathogenic and excluded from further analysis. The p.(Gly545Ala) variant was found 7,626 times in gnomAD, including 145 times in the homozygous form. It was considered pathogenic in two out of three computational tools, affected a conserved residue, but had Benign assessments from ClinVar and Varsome (based on one moderate (PM1) and 2 supporting criteria for pathogenic (PP2, PP3) and two strong (BS1,BS2) and one supporting criteria for benign (BP6)).

The p.(Gly999Glu) variant was found 3,259 times, including 33 times in the homozygous form. It was considered pathogenic in all three computational tools, and affected a conserved residue, but was considered a Variant of Uncertain Significance by both ClinVar and Varsome based on one moderate (PM1), and three supporting (PP2,PP3 and PP5) criteria for pathogenicity, and two strong criteria for benign (BS1 and BS2).

**Suppl Table 10: Predicted pathogenic founder variants in *COL4A3* and *COL4A4* in different ancestries in gnomAD**

| Gene | Variant | Transcript | Total | African | Latino | Ashkenazi | East Asian | Eur (Finnish) | Eur (non-Finnish) | Other | South Asian |
|---|---|---|---|---|---|---|---|---|---|---|---|
| ***COL4A3*** | None | | | | | | | | | | |
| ***COL4A4*** | p.(Gly445Ala) | c.1334G>C | **22** | **19** | | | | | **3** | | |
| | p.(Gly481Ser) | c.1441G>A | **21** | | **21** | | | | | | |
| | p.(Gly816Glu) | c.2447G>A | **24** | | | | **24** | | | | |
| | p.(Gly1005Glu) | c.3014G>A | **25** | **1** | **23** | | | | **1** | | |
| | p.(Gly1178Ser) | c.3532G>A | **24** | | **3** | | | **3** | **16** | **2** | |
| | p.(Gly774Arg) | c.2320G>C | **19** | | | | | **1** | **5** | **13** | |
| | p.(Ser969Ter) | c.2906C>G | **18** | **1** | | | | | **16** | **1** | |

**Variants were arbitrarily designated founder variants if they resulted in a position 1 Gly substitution, frameshift, nonsense or canonical splice variant that was predicted to be pathogenic and occurred ≥ 18 times in gnomAD**

**Suppl Table 11: Estimated population frequencies for predicted pathogenic heterozygous, compound heterozygous, and digenic variants**

| *COL4A3* | *COL4A4* | Classification | Frequency | Total population frequency* |
|---|---|---|---|---|
| +/+ | +/+ | Wildtype | $9.91 \times 10^{-1}$ | $9.91 \times 10^{-1}$<br>or 99 in 100 |
| +/+ | +/- | Heterozygous predicted pathogenic | $2.45 \times 10^{-3}$ | $9.41 \times 10^{-3}$<br>or one in 106 |
| +/+ | -/+ | Heterozygous predicted pathogenic | $2.45 \times 10^{-3}$ | |
| +/- | +/+ | Heterozygous predicted pathogenic | $2.26 \times 10^{-3}$ | |
| -/+ | +/+ | Heterozygous predicted pathogenic | $2.26 \times 10^{-3}$ | |
| +/- | +/- | Digenic predicted pathogenic variants | $5.58 \times 10^{-6}$ | $2.23 \times 10^{-5}$<br>or one in 44,793 |
| +/- | -/+ | Digenic predicted pathogenic variants | $5.58 \times 10^{-6}$ | |
| -/+ | +/- | Digenic predicted pathogenic variants | $5.58 \times 10^{-6}$ | |
| -/+ | -/+ | Digenic predicted pathogenic variants | $5.58 \times 10^{-6}$ | |
| +/+ | -/- | Compound heterozygous predicted pathogenic variants | $6.06 \times 10^{-6}$ | $1.13 \times 10^{-5}$<br>or one in 88,866 individuals |
| +/- | -/- | Compound heterozygous predicted pathogenic variants | $1.38 \times 10^{-8}$ | |
| -/+ | -/- | Compound heterozygous predicted pathogenic variants | $1.38 \times 10^{-8}$ | |
| -/- | +/+ | Compound heterozygous predicted pathogenic variants | $5.14 \times 10^{-6}$ | |
| -/- | +/- | Compound heterozygous predicted pathogenic variants | $1.27 \times 10^{-8}$ | |
| -/- | -/+ | Compound heterozygous predicted pathogenic variants | $1.27 \times 10^{-8}$ | |
| -/- | -/- | Compound heterozygous predicted pathogenic variants | $3.14 \times 10^{-11}$ | |

+ wildtype, - variant allele. The frequencies \have been calculated from the sum of the likelihood of the individual combinations of alleles. These calculations are based on the number of predicted pathogenic

variants in the mean alleles examined that is 559/245,889 for *COL4A3* and 577/233,916 for *COL4A4* and assume that *COL4A3* and *COL4A4* variants occur together independently. If all predicted pathogenic variants are pathogenic, then heterozygous predicted pathogenic variants correspond to Thin basement membrane nephropathy or AD Alport syndrome; digenic predicted pathogenic variants correspond to digenic Alport syndrome; and compound heterozygous pathogenic variants correspond to the compound heterozygous form of AR Alport syndrome. It is not possible to use this method to calculate how commonly homozygous forms of predicted pathogenic variants occur because they are usually the result of consanguineous relationships and do not occur by chance.

## References


1. Gubler, M, Levy, M, Broyer, M, Naizot, C, Gonzales, G, Perrin, D, Habib, R: Alport's syndrome. A report of 58 cases and a review of the literature. *Am J Med,* 70**:** 493-505, 1981.
2. Barker, DF, Hostikka, SL, Zhou, J, Chow, LT, Oliphant, AR, Gerken, SC, Gregory, MC, Skolnick, MH, Atkin, CL, Tryggvason, K: Identification of mutations in the COL4A5 collagen gene in Alport syndrome. *Science,* 248**:** 1224-1227, 1990.
3. Mochizuki, T, Lemmink, HH, Mariyama, M, Antignac, C, Gubler, MC, Pirson, Y, Verellen-Dumoulin, C, Chan, B, Schroder, CH, Smeets, HJ, et al.: Identification of mutations in the alpha 3(IV) and alpha 4(IV) collagen genes in autosomal recessive Alport syndrome. *Nat Genet,* 8**:** 77-81, 1994.
4. Buzza, M, Wilson, D, Savige, J: Segregation of hematuria in thin basement membrane disease with haplotypes at the loci for Alport syndrome. *Kidney Int,* 59**:** 1670-1676, 2001.
5. Buzza, M, Dagher, H, Wang, YY, Wilson, D, Babon, JJ, Cotton, RG, Savige, J: Mutations in the COL4A4 gene in thin basement membrane disease. *Kidney Int,* 63**:** 447-453, 2003.
6. Savige, J: Should We Diagnose Autosomal Dominant Alport Syndrome When There Is a Pathogenic Heterozygous COL4A3 or COL4A4 Variant? *Kidney Int Rep,* 3**:** 1239-1241, 2018.
7. Savige, J, Rana, K, Tonna, S, Buzza, M, Dagher, H, Wang, YY: Thin basement membrane nephropathy. *Kidney Int,* 64**:** 1169-1178, 2003.
8. Mencarelli, MA, Heidet, L, Storey, H, van Geel, M, Knebelmann, B, Fallerini, C, Miglietti, N, Antonucci, MF, Cetta, F, Sayer, JA, van den Wijngaard, A, Yau, S, Mari, F, Bruttini, M, Ariani, F, Dahan, K, Smeets, B, Antignac, C, Flinter, F, Renieri, A: Evidence of digenic inheritance in Alport syndrome. *J Med Genet,* 52**:** 163-174, 2015.
9. Voskarides, K, Damianou, L, Neocleous, V, Zouvani, I, Christodoulidou, S, Hadjiconstantinou, V, Ioannou, K, Athanasiou, Y, Patsias, C, Alexopoulos, E, Pierides, A, Kyriacou, K, Deltas, C: COL4A3/COL4A4 mutations producing focal segmental glomerulosclerosis and renal failure in thin basement membrane nephropathy. *J Am Soc Nephrol,* 18**:** 3004-3016, 2007.
10. Wickman, L, Hodgin, JB, Wang, SQ, Afshinnia, F, Kershaw, D, Wiggins, RC: Podocyte Depletion in Thin GBM and Alport Syndrome. *PLoS One,* 11**:** e0155255, 2016.
11. Baldovino, S, Moliner, AM, Taruscio, D, Daina, E, Roccatello, D: Rare Diseases in Europe: from a Wide to a Local Perspective. *Isr Med Assoc J,* 18**:** 359-363, 2016.
12. Hertz, JM, Thomassen, M, Storey, H, Flinter, F: Clinical utility gene card for: Alport syndrome - update 2014. *Eur J Hum Genet,* 23, 2015.
13. Hasstedt, SJ, Atkin, CL: X-linked inheritance of Alport syndrome: family P revisited. *Am J Hum Genet,* 35**:** 1241-1251, 1983.
14. Persson, U, Hertz, JM, Wieslander, J, Segelmark, M: Alport syndrome in southern Sweden. *Clin Nephrol,* 64**:** 85-90, 2005.
15. Pajari, H, Kaariainen, H, Muhonen, T, Koskimies, O: Alport's syndrome in 78 patients: epidemiological and clinical study. *Acta Paediatr,* 85**:** 1300-1306, 1996.
16. Dodge, WF, West, EF, Smith, EH, Bruce, H, 3rd: Proteinuria and hematuria in schoolchildren: epidemiology and early natural history. *J Pediatr,* 88**:** 327-347, 1976.
17. Ritchie, CD, Bevan, EA, Collier, SJ: Importance of occult haematuria found at screening. *Br Med J (Clin Res Ed),* 292**:** 681-683, 1986.
18. Dische, FE, Anderson, VE, Keane, SJ, Taube, D, Bewick, M, Parsons, V: Incidence of thin membrane nephropathy: morphometric investigation of a population sample. *J Clin Pathol,* 43**:** 457-460, 1990.
19. Savige, J, Gregory, M, Gross, O, Kashtan, C, Ding, J, Flinter, F: Expert guidelines for the management of Alport syndrome and thin basement membrane nephropathy. *J Am Soc Nephrol,* 24**:** 364-375, 2013.
20. Jais, JP, Knebelmann, B, Giatras, I, De Marchi, M, Rizzoni, G, Renieri, A, Weber, M, Gross, O, Netzer, KO, Flinter, F, Pirson, Y, Verellen, C, Wieslander, J, Persson, U, Tryggvason, K, Martin, P, Hertz, JM, Schroder, C, Sanak, M, Krejcova, S, Carvalho, MF, Saus, J, Antignac, C, Smeets, H, Gubler, MC: X-linked Alport syndrome: natural history in 195 families and genotype- phenotype correlations in males. *J Am Soc Nephrol,* 11**:** 649-657, 2000.
21. Jais, JP, Knebelmann, B, Giatras, I, De Marchi, M, Rizzoni, G, Renieri, A, Weber, M, Gross, O, Netzer, KO, Flinter, F, Pirson, Y, Dahan, K, Wieslander, J, Persson, U, Tryggvason, K, Martin, P, Hertz, JM, Schroder, C, Sanak, M, Carvalho, MF, Saus, J, Antignac, C, Smeets, H, Gubler, MC: 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. *J Am Soc Nephrol,* 14**:** 2603-2610, 2003.
22. Savige, J, Colville, D, Rheault, M, Gear, S, Lennon, R, Lagas, S, Finlay, M, Flinter, F: Alport Syndrome in Women and Girls. *Clin J Am Soc Nephrol,* 11**:** 1713-1720, 2016.

23. Savige, J, Ariani, F, Mari, F, Bruttini, M, Renieri, A, Gross, O, Deltas, C, Flinter, F, Ding, J, Gale, DP, Nagel, M, Yau, M, Shagam, L, Torra, R, Ars, E, Hoefele, J, Garosi, G, Storey, H: Expert consensus guidelines for the genetic diagnosis of Alport syndrome. *Pediatr Nephrol*, 2018.
24. Richards, S, Aziz, N, Bale, S, Bick, D, Das, S, Gastier-Foster, J, Grody, WW, Hegde, M, Lyon, E, Spector, E, Voelkerding, K, Rehm, HL, Committee, ALQA: 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*, 17**:** 405-424, 2015.
25. Myllyharju, J, Kivirikko, KI: Collagens and collagen-related diseases. *Ann Med*, 33**:** 7-21, 2001.
26. Mariyama, M, Leinonen, A, Mochizuki, T, Tryggvason, K, Reeders, ST: Complete primary structure of the human alpha 3(IV) collagen chain. Coexpression of the alpha 3(IV) and alpha 4(IV) collagen chains in human tissues. *J Biol Chem*, 269**:** 23013-23017, 1994.
27. Leinonen, A, Mariyama, M, Mochizuki, T, Tryggvason, K, Reeders, ST: Complete primary structure of the human type IV collagen alpha 4(IV) chain. Comparison with structure and expression of the other alpha (IV) chains. *J Biol Chem*, 269**:** 26172-26177, 1994.
28. Zhou, J, Hertz, JM, Leinonen, A, Tryggvason, K: Complete amino acid sequence of the human alpha 5 (IV) collagen chain and identification of a single-base mutation in exon 23 converting glycine 521 in the collagenous domain to cysteine in an Alport syndrome patient. *J Biol Chem*, 267**:** 12475-12481, 1992.
29. Demosthenous, P, Voskarides, K, Stylianou, K, Hadjigavriel, M, Arsali, M, Patsias, C, Georgaki, E, Zirogiannis, P, Stavrou, C, Daphnis, E, Pierides, A, Deltas, C, Hellenic Nephrogenetics Research, C: X-linked Alport syndrome in Hellenic families: phenotypic heterogeneity and mutations near interruptions of the collagen domain in COL4A5. *Clin Genet*, 81**:** 240-248, 2012.
30. Lanktree, MB, Haghighi, A, Guiard, E, Iliuta, IA, Song, X, Harris, PC, Paterson, AD, Pei, Y: Prevalence Estimates of Polycystic Kidney and Liver Disease by Population Sequencing. *J Am Soc Nephrol*, 29**:** 2593-2600, 2018.
31. Lek, M, Karczewski, KJ, Minikel, EV, Samocha, KE, Banks, E, Fennell, T, O'Donnell-Luria, AH, Ware, JS, Hill, AJ, Cummings, BB, Tukiainen, T, Birnbaum, DP, Kosmicki, JA, Duncan, LE, Estrada, K, Zhao, F, Zou, J, Pierce-Hoffman, E, Berghout, J, Cooper, DN, Deflaux, N, DePristo, M, Do, R, Flannick, J, Fromer, M, Gauthier, L, Goldstein, J, Gupta, N, Howrigan, D, Kiezun, A, Kurki, MI, Moonshine, AL, Natarajan, P, Orozco, L, Peloso, GM, Poplin, R, Rivas, MA, Ruano-Rubio, V, Rose, SA, Ruderfer, DM, Shakir, K, Stenson, PD, Stevens, C, Thomas, BP, Tiao, G, Tusie-Luna, MT, Weisburd, B, Won, HH, Yu, D, Altshuler, DM, Ardissino, D, Boehnke, M, Danesh, J, Donnelly, S, Elosua, R, Florez, JC, Gabriel, SB, Getz, G, Glatt, SJ, Hultman, CM, Kathiresan, S, Laakso, M, McCarroll, S, McCarthy, MI, McGovern, D, McPherson, R, Neale, BM, Palotie, A, Purcell, SM, Saleheen, D, Scharf, JM, Sklar, P, Sullivan, PF, Tuomilehto, J, Tsuang, MT, Watkins, HC, Wilson, JG, Daly, MJ, MacArthur, DG, Exome Aggregation, C: Analysis of protein-coding genetic variation in 60,706 humans. *Nature*, 536**:** 285-291, 2016.
32. The National Genomics Research and Healthcare Knowledgebase v5, Genomics England. 5, 2019.
33. van der Ven, AT, Connaughton, DM, Ityel, H, Mann, N, Nakayama, M, Chen, J, Vivante, A, Hwang, DY, Schulz, J, Braun, DA, Schmidt, JM, Schapiro, D, Schneider, R, Warejko, JK, Daga, A, Majmundar, AJ, Tan, W, Jobst-Schwan, T, Hermle, T, Widmeier, E, Ashraf, S, Amar, A, Hoogstraaten, CA, Hugo, H, Kitzler, TM, Kause, F, Kolvenbach, CM, Dai, R, Spaneas, L, Amann, K, Stein, DR, Baum, MA, Somers, MJG, Rodig, NM, Ferguson, MA, Traum, AZ, Daouk, GH, Bogdanovic, R, Stajic, N, Soliman, NA, Kari, JA, El Desoky, S, Fathy, HM, Milosevic, D, Al-Saffar, M, Awad, HS, Eid, LA, Selvin, A, Senguttuvan, P, Sanna-Cherchi, S, Rehm, HL, MacArthur, DG, Lek, M, Laricchia, KM, Wilson, MW, Mane, SM, Lifton, RP, Lee, RS, Bauer, SB, Lu, W, Reutter, HM, Tasic, V, Shril, S, Hildebrandt, F: Whole-Exome Sequencing Identifies Causative Mutations in Families with Congenital Anomalies of the Kidney and Urinary Tract. *J Am Soc Nephrol*, 29**:** 2348-2361, 2018.
34. Biesecker, LG, Harrison, SM, ClinGen Sequence Variant Interpretation Working, G: The ACMG/AMP reputable source criteria for the interpretation of sequence variants. *Genet Med*, 20**:** 1687-1688, 2018.
35. Richards, CS, Aziz, N, Bale, S, Bick, D, Das, S, Gastier-Foster, J, Grody, WW, Hegde, M, Lyon, E, Spector, E, Voelkerding, K, Rehm, HL, Group, AAIoSVW: Response to Biesecker and Harrison. *Genet Med*, 20**:** 1689-1690, 2018.
36. Abou Tayoun, AN, Pesaran, T, DiStefano, MT, Oza, A, Rehm, HL, Biesecker, LG, Harrison, SM, ClinGen Sequence Variant Interpretation Working, G: Recommendations for interpreting the loss of function PVS1 ACMG/AMP variant criterion. *Hum Mutat*, 39**:** 1517-1524, 2018.
37. Hashimura, Y, Nozu, K, Kaito, H, Nakanishi, K, Fu, XJ, Ohtsubo, H, Hashimoto, F, Oka, M, Ninchoji, T, Ishimori, S, Morisada, N, Matsunoshita, N, Kamiyoshi, N, Yoshikawa, N, Iijima, K: Milder clinical aspects of X-linked Alport syndrome in men positive for the collagen IV alpha5 chain. *Kidney Int*, 85**:** 1208-1213, 2014.

38. Savige, J, Storey, H, Il Cheong, H, Gyung Kang, H, Park, E, Hilbert, P, Persikov, A, Torres-Fernandez, C, Ars, E, Torra, R, Hertz, JM, Thomassen, M, Shagam, L, Wang, D, Wang, Y, Flinter, F, Nagel, M: X-Linked and Autosomal Recessive Alport Syndrome: Pathogenic Variant Features and Further Genotype-Phenotype Correlations. *PLoS One,* 11**:** e0161802, 2016.
39. Persikov, AV, Pillitteri, RJ, Amin, P, Schwarze, U, Byers, PH, Brodsky, B: Stability related bias in residues replacing glycines within the collagen triple helix (Gly-Xaa-Yaa) in inherited connective tissue disorders. *Hum Mutat,* 24**:** 330-337, 2004.
40. Storey, H, Savige, J, Sivakumar, V, Abbs, S, Flinter, FA: COL4A3/COL4A4 mutations and features in individuals with autosomal recessive Alport syndrome. *J Am Soc Nephrol,* 24**:** 1945-1954, 2013.
41. Lennon, R, Stuart, HM, Bierzynska, A, Randles, MJ, Kerr, B, Hillman, KA, Batra, G, Campbell, J, Storey, H, Flinter, FA, Koziell, A, Welsh, GI, Saleem, MA, Webb, NJ, Woolf, AS: Coinheritance of COL4A5 and MYO1E mutations accentuate the severity of kidney disease. *Pediatr Nephrol,* 30**:** 1459-1465, 2015.
42. Knebelmann, B, Breillat, C, Forestier, L, Arrondel, C, Jacassier, D, Giatras, I, Drouot, L, Deschenes, G, Grunfeld, JP, Broyer, M, Gubler, MC, Antignac, C: Spectrum of mutations in the COL4A5 collagen gene in X-linked Alport syndrome. *Am J Hum Genet,* 59**:** 1221-1232, 1996.
43. Inoue, YJ, Nishio, H, Shirakawa, T, Nakanishi, K, Nakamura, H, Sumino, K, Nishiyama, K, Iijima, K, Yoshikawa, N: Detection of mutations in the COL4A5 gene in over 90% of male patients with X-linked Alport's syndrome by RT-PCR and direct sequencing. *Am J Kidney Dis,* 34**:** 854-862, 1999.
44. Zhang, Y, Ding, J, Wang, S, Zhang, H, Zhong, X, Liu, X, Xu, K, Wang, F: Reassessing the pathogenicity of c.2858G>T(p.(G953V)) in COL4A5 Gene: report of 19 Chinese families. *Eur J Hum Genet,* 28**:** 244-252, 2020.
45. Slajpah, M, Gorinsek, B, Berginc, G, Vizjak, A, Ferluga, D, Hvala, A, Meglic, A, Jaksa, I, Furlan, P, Gregoric, A, Kaplan-Pavlovcic, S, Ravnik-Glavac, M, Glavac, D: Sixteen novel mutations identified in COL4A3, COL4A4, and COL4A5 genes in Slovenian families with Alport syndrome and benign familial hematuria. *Kidney Int,* 71**:** 1287-1295, 2007.
46. Macheroux, EP, Braunisch, MC, Pucci Pegler, S, Satanovskij, R, Riedhammer, KM, Gunthner, R, Gross, O, Nagel, M, Renders, L, Hoefele, J: The Hypomorphic Variant p.(Gly624Asp) in COL4A5 as a Possible Cause for an Unexpected Severe Phenotype in a Family With X-Linked Alport Syndrome. *Front Pediatr,* 7**:** 485, 2019.
47. Zurowska, AM, Bielska, O, Daca-Roszak, P, Jankowski, M, Szczepanska, M, Roszkowska-Bjanid, D, Kuzma-Mroczkowska, E, Panczyk-Tomaszewska, M, Moczulska, A, Drozdz, D, Hadjipanagi, D, Deltas, C, Ostalska-Nowicka, D, Rabiega, A, Taraszkiewicz, J, Taranta-Janusz, K, Wieczorkiewicz-Plaza, A, Jobs, K, Mews, J, Musial, K, Jakubowska, A, Nosek, H, Jander, AE, Koutsofti, C, Stanislawska-Sachadyn, A, Kuleszo, D, Zietkiewicz, E, Lipska-Zietkiewicz, BS: 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,* 2020.
48. AM, Z: **Mild X-linked Alport syndrome due to the COL4A5 G624D variant origtinating in the Middle Ages is predominant in Central/East Europe and causes kidney failure in mid-life**. *Kidney Int,* 2021.
49. Persikov, AV, Ramshaw, JA, Kirkpatrick, A, Brodsky, B: Amino acid propensities for the collagen triple-helix. *Biochemistry,* 39**:** 14960-14967, 2000.
50. Fallerini, C, Baldassarri, M, Trevisson, E, Morbidoni, V, La Manna, A, Lazzarin, R, Pasini, A, Barbano, G, Pinciaroli, AR, Garosi, G, Frullanti, E, Pinto, AM, Mencarelli, MA, Mari, F, Renieri, A, Ariani, F: Alport syndrome: impact of digenic inheritance in patients management. *Clin Genet,* 92**:** 34-44, 2017.
51. Feingold, J, Bois, E, Chompret, A, Broyer, M, Gubler, MC, Grunfeld, JP: Genetic heterogeneity of Alport syndrome. *Kidney Int,* 27**:** 672-677, 1985.
52. Lemmink, HH, Nillesen, WN, Mochizuki, T, Schroder, CH, Brunner, HG, van Oost, BA, Monnens, LA, Smeets, HJ: Benign familial hematuria due to mutation of the type IV collagen alpha4 gene. *J Clin Invest,* 98**:** 1114-1118, 1996.
53. Makareeva, E, Mertz, EL, Kuznetsova, NV, Sutter, MB, DeRidder, AM, Cabral, WA, Barnes, AM, McBride, DJ, Marini, JC, Leikin, S: Structural heterogeneity of type I collagen triple helix and its role in osteogenesis imperfecta. *J Biol Chem,* 283**:** 4787-4798, 2008.
54. Gross, O, Licht, C, Anders, HJ, Hoppe, B, Beck, B, Tonshoff, B, Hocker, B, Wygoda, S, Ehrich, JH, Pape, L, Konrad, M, Rascher, W, Dotsch, J, Muller-Wiefel, DE, Hoyer, P, Study Group Members of the Gesellschaft fur Padiatrische, N, Knebelmann, B, Pirson, Y, Grunfeld, JP, Niaudet, P, Cochat, P, Heidet, L, Lebbah, S, Torra, R, Friede, T, Lange, K, Muller, GA, Weber, M: Early angiotensin-converting enzyme inhibition in Alport syndrome delays renal failure and improves life expectancy. *Kidney Int,* 81**:** 494-501, 2012.
55. Kashtan, CE: Alport Syndrome. Gene Reviews

In: Adam MP, Ardinger HH, Pagon RA, et al., editors. GeneReviews® [Internet]. Seattle (WA): University of Washington, Seattle; 1993-2020. Available from: https://www.ncbi.nlm.nih.gov/books/NBK1207/. 2001.

56. Pierides, A, Voskarides, K, Athanasiou, Y, Ioannou, K, Damianou, L, Arsali, M, Zavros, M, Pierides, M, Vargemezis, V, Patsias, C, Zouvani, I, Elia, A, Kyriacou, K, Deltas, C: Clinico-pathological correlations in 127 patients in 11 large pedigrees, segregating one of three heterozygous mutations in the COL4A3/ COL4A4 genes associated with familial haematuria and significant late progression to proteinuria and chronic kidney disease from focal segmental glomerulosclerosis. *Nephrol Dial Transplant,* 24**:** 2721-2729, 2009.
57. Matthaiou, M, Poulii, T, Matthaiou, M, Matthaiou, M, Poulli, T, Deltas, C: Prevalence of clinical, pathological and molecular features of glomerular basement membrane nephropathy caused by *COL4A3* or *COL4A4* mutations: a systematic review. *Clinical Kidney Journal* 1-12, 2020.
58. Gillion, V, Dahan, K, Cosyns, JP, Hilbert, P, Jadoul, M, Goffin, E, Godefroid, N, De Meyer, M, Mourad, M, Pirson, Y, Kanaan, N: Genotype and Outcome After Kidney Transplantation in Alport Syndrome. *Kidney Int Rep,* 3**:** 652-660, 2018.
59. Groopman, EE, Marasa, M, Cameron-Christie, S, Petrovski, S, Aggarwal, VS, Milo-Rasouly, H, Li, Y, Zhang, J, Nestor, J, Krithivasan, P, Lam, WY, Mitrotti, A, Piva, S, Kil, BH, Chatterjee, D, Reingold, R, Bradbury, D, DiVecchia, M, Snyder, H, Mu, X, Mehl, K, Balderes, O, Fasel, DA, Weng, C, Radhakrishnan, J, Canetta, P, Appel, GB, Bomback, AS, Ahn, W, Uy, NS, Alam, S, Cohen, DJ, Crew, RJ, Dube, GK, Rao, MK, Kamalakaran, S, Copeland, B, Ren, Z, Bridgers, J, Malone, CD, Mebane, CM, Dagaonkar, N, Fellstrom, BC, Haefliger, C, Mohan, S, Sanna-Cherchi, S, Kiryluk, K, Fleckner, J, March, R, Platt, A, Goldstein, DB, Gharavi, AG: Diagnostic Utility of Exome Sequencing for Kidney Disease. *N Engl J Med,* 380**:** 142-151, 2019.
60. Fallerini, C, Dosa, L, Tita, R, Del Prete, D, Feriozzi, S, Gai, G, Clementi, M, La Manna, A, Miglietti, N, Mancini, R, Mandrile, G, Ghiggeri, GM, Piaggio, G, Brancati, F, Diano, L, Frate, E, Pinciaroli, AR, Giani, M, Castorina, P, Bresin, E, Giachino, D, De Marchi, M, Mari, F, Bruttini, M, Renieri, A, Ariani, F: Unbiased next generation sequencing analysis confirms the existence of autosomal dominant Alport syndrome in a relevant fraction of cases. *Clin Genet,* 86**:** 252-257, 2014.
61. Bullich, G, Domingo-Gallego, A, Vargas, I, Ruiz, P, Lorente-Grandoso, L, Furlano, M, Fraga, G, Madrid, A, Ariceta, G, Borregan, M, Pinero-Fernandez, JA, Rodriguez-Pena, L, Ballesta-Martinez, MJ, Llano-Rivas, I, Menica, MA, Ballarin, J, Torrents, D, Torra, R, Ars, E: A kidney-disease gene panel allows a comprehensive genetic diagnosis of cystic and glomerular inherited kidney diseases. *Kidney Int,* 94**:** 363-371, 2018.
62. Gross, O, Tonshoff, B, Weber, LT, Pape, L, Latta, K, Fehrenbach, H, Lange-Sperandio, B, Zappel, H, Hoyer, P, Staude, H, Konig, S, John, U, Gellermann, J, Hoppe, B, Galiano, M, Hoecker, B, Ehren, R, Lerch, C, Kashtan, CE, Harden, M, Boeckhaus, J, Friede, T, German Pediatric Nephrology Study, G, Investigators, EP-TA: A multicenter, randomized, placebo-controlled, double-blind phase 3 trial with open-arm comparison indicates safety and efficacy of nephroprotective therapy with ramipril in children with Alport's syndrome. *Kidney Int,* 97**:** 1275-1286, 2020.
63. Hertz, JM, Juncker, I, Persson, U, Matthijs, G, Schmidtke, J, Petersen, MB, Kjeldsen, M, Gregersen, N: Detection of mutations in the COL4A5 gene by SSCP in X-linked Alport syndrome. *Hum Mutat,* 18**:** 141-148, 2001.
64. Plant, KE, Green, PM, Vetrie, D, Flinter, FA: Detection of mutations in COL4A5 in patients with Alport syndrome. *Hum Mutat,* 13**:** 124-132, 1999.
65. King, K, Flinter, FA, Green, PM: A two-tier approach to mutation detection in the COL4A5 gene for Alport syndrome. *Hum Mutat,* 27**:** 1061, 2006.
66. Cruz-Robles, D, Garcia-Torres, R, Antignac, C, Forestier, L, de la Puente, SG, Correa-Rotter, R, Garcia-Lopez, E, Orozco, L: Three novel mutations in the COL4A5 gene in Mexican Alport syndrome patients. *Clin Genet,* 56**:** 242-243, 1999.
67. Bekheirnia, MR, Reed, B, Gregory, MC, McFann, K, Shamshirsaz, AA, Masoumi, A, Schrier, RW: Genotype-phenotype correlation in X-linked Alport syndrome. *J Am Soc Nephrol,* 21**:** 876-883, 2010.
68. Mohammad, M, Nanra, R, Colville, D, Trevillian, P, Wang, Y, Storey, H, Flinter, F, Savige, J: A female with X-linked Alport syndrome and compound heterozygous COL4A5 mutations. *Pediatr Nephrol,* 29**:** 481-485, 2014.
69. Inoue, Y, Nishio, H, Shirakawa, T, Nakanishi, K, Nakamura, H, Sumino, K, Nishiyama, K, Iijima, K, Yoshikawa, N: Detection of mutations in the COL4A5 gene in over 90% of male patients with X-linked Alport's syndrome by RT-PCR and direct sequencing. *Am J Kidney Dis,* 34**:** 854-862, 1999.
70. Zhang, H, Ding, J, Wang, F, Zhao, D: Mutation detection of COL4An gene based on mRNA of peripheral blood lymphocytes and prenatal diagnosis of Alport syndrome in China. *Nephrology (Carlton),* 16**:** 377-380, 2011.