DNA variant databases improve test accuracy and phenotype prediction in Alport syndrome

The International Alport Mutation Consortium; ¹Judy Savige, ²Elisabet Ars, ³Richard GH Cotton, ⁴David Crockett, ¹Hayat Dagher, ⁵Constantinos Deltas, ⁶Jie Ding, ⁷Frances Flinter, ⁴Genevieve Pont-Kingdon, ⁸Nizar Smaoui, ⁹Roser Torra, ¹⁰Helen Storey

¹The University of Melbourne, Department of Medicine (Northern Health), Epping, VIC, Australia; ² Nephrology Department, Fundacio Puigvert, Barcelona, Spain; ³Human Variome Project, The University of Melbourne, Parkville, VIC, Australia; ⁴ARUP Institute of Clinical and Experimental Pathology, ARUP laboratories, Salt Lake City, UT, USA; ⁵Molecular Medicine Research Center and Laboratory of Molecular and Medical Genetics, Department of Biological Sciences, University of Cyprus, Nicosia, Cyprus; ⁶Pediatric Department, Peking University First Hospital, Beijing, China; ⁷Department of Clinical Genetics, Guy's and St Thomas NHS Foundation Trust, London, UK; ⁸GeneDx, Gaithersburg, MD, USA; ⁹Molecular Biology Laboratory, Fundacio Puigvert, Universitat Autònoma de Barcelona, Barcelona, Spain; and ¹⁰GSTS Pathology, Guy's and St Thomas NHS Foundation Trust, London, UK.

Address for Correspondence:

Judy Savige

The University of Melbourne (AH/NH)

The Northern Hospital

Cooper St

Epping VIC 3076

Australia;

TEL: + 613 8405 8823

FAX: + 613 8405 8724

Email: jasavige@unimelb.edu.au

ABSTRACT

X-linked Alport syndrome is a form of progressive renal failure caused by pathogenic variants in the *COL4A5* gene. More than 700 variants have been described and a further 400 are estimated to be known to individual laboratories but are unpublished.

The major genetic testing laboratories for X-linked Alport syndrome worldwide have established a web-based database for published and unpublished *COL4A5* variants (https://grenada.lumc.nl/LOVD2/COL4A/home.php?select_db=COL4A5). This conforms with the recommendations of the Human Variome Project: it uses the Leiden Open Variation Database (LOVD) format, describes variants according to the human reference sequence with standardised nomenclature, indicates likely pathogenicity and associated clinical features, and credits the submitting laboratory. The database includes non-pathogenic and recurrent variants, and is linked to another *COL4A5* mutation database and relevant bioinformatics sites. Access is free.

Increasing the number of *COL4A5* variants in the public domain helps patients, diagnostic laboratories, clinicians and researchers. The database improves the accuracy and efficiency of genetic testing because its variants are already categorised for pathogenicity. The description of further *COL4A5* variants and clinical associations will improve our ability to predict phenotype and our understanding of collagen IV biochemistry. The database for X-linked Alport syndrome represents a model for databases in other inherited renal diseases.

MAIN TEXT

Alport syndrome

Alport syndrome is an inherited cause of renal failure that affects at least one in 50,000 individuals [1]. Eighty-five % of families have X-linked disease (MIM 301050) [2], and mutations in the *COL4A5* gene which codes for the $\alpha 5$ chain of collagen type IV [3].

Affected males with X-linked Alport syndrome have hematuria from infancy, and later develop renal failure, sensorineural hearing loss, and sometimes lenticonus, retinopathy or corneal dystrophy [4-7]. Those with end-stage renal failure before the age of 30 years ('juvenile or early onset ') are more likely to also have extra-renal features [8,9]. However milder phenotypes are recognised increasingly [10]. Females with X-linked disease have persistent hematuria, and sometimes hearing loss and retinopathy, and 15% develop renal failure by late middle age [11].

The diagnosis of X-linked Alport syndrome is often problematic. It is suspected when the characteristic clinical features are present, there is a family history of renal failure, the glomerular basement membrane is lamellated [12], or when the $\alpha 5(\text{IV})$ collagen chain is absent from the glomerular or epidermal membrane [13,14]. Other causes of familial haematuria and renal failure include Thin basement membrane nephropathy, IgA disease and C3 nephropathy [15]. The 'gold standard' for the diagnosis of X-linked Alport syndrome is the demonstration of a mutation in the *COL4A5* gene (Table 1). A *COL4A5* mutation also confirms X-linked inheritance, which is important because it determines the risk of renal failure in other family members.

Genetic mutations in Alport syndrome

The *COL4A5* gene is very large, encompassing 250 kb of genomic DNA and 51 exons encoding a 6.5 kb transcript [16]. Currently more than 250 patients with suspected X-linked Alport syndrome undergo genetic testing in a dozen laboratories around the world annually. The speed and scale with which new variants are identified is unprecedented [17,18]. To date, more than 700 *COL4A5* pathogenic and normal variants have been reported, with the majority of disease-associated variants being found in only one family each.

All types of pathogenic *COL4A5* variants have been described in X-linked Alport syndrome. Fifty % are large deletions or rearrangements, nonsense or splicing mutations [19]. Forty % are missense variants, which often result in the substitution of a glycine with a larger or a more highly charged amino acid [8,20-22]. Some missense mutations interfere with structural or functional sites in the type IV collagen $\alpha 5$ chain, similar to the changes described with *COL1A1* mutations in Osteogenesis imperfecta [23, 24].

COL4A5 variants are also common in normal individuals and any variant in a patient with X-linked Alport syndrome must be examined for likely pathogenicity. A DNA substitution that results in a stop codon, changes the mRNA reading frame, affects a canonical splice site, or results in a large rearrangement probably causes disease [25]. Other variants must generally fulfil several criteria. They typically change the amino acid, substitute for a highly conserved residue, affect a major structural or ligand-binding site, co-segregate with disease within an affected family, and are not present in 100 normal chromosomes. Nevertheless, some disease-associated variants do not change the amino acid sequence, the structural and functional sites in the $\alpha 5(\text{IV})$ collagen chain are poorly characterised,

segregation studies are not feasible in families that are small or have atypical disease, and the comparison group of 100 normal chromosomes may not include DNA samples from the patient's ethnicity. Distinguishing between a pathogenic and non-pathogenic variant is usually expensive, prone to error, time-consuming, and results in delays for the patient.

The most efficient strategy to determine a DNA variant's pathogenicity is to refer to a database, where variants have already been classified as pathogenic or otherwise, and that provides the evidence supporting this assessment or at least contact details for the laboratory that undertook the relevant studies. For such a database to be useful, its variant collection must be current and complete, and its classifications accurate. However, many *COL4A5* variants identified in individual laboratories have not been published. The major gene testing laboratories for Alport syndrome worldwide estimate they have more than 400 variants in their records that are unreported. This happens because of the effort required for publication and the lack of interest from journals in small series without a new phenotype or insight. Nevertheless getting correctly classified and validated variants, even non-pathogenic variants, into the public domain is important because it means that other laboratories do not have to repeat the studies that determined pathogenicity.

Human Variome Project

The Human Variome Project has highlighted the potential significance of the large number of unreported DNA variants in disease-associated genes, and has

provided guidelines for their collection and curation in web-based databases (Table 2) [26-28].

International Alport Mutation Consortium and COL4A5 database

An international consortium of 8 genetic testing laboratories for X-linked Alport syndrome (London, Boston, Beijing, Salt Lake City, Barcelona, Cyprus, Maryland, Melbourne), have agreed to submit their unpublished variants to the COL4A5 database that their members curate

(https://grenada.lumc.nl/LOVD2/COL4A/home.php?select_db=COL4A5) and that conforms to the Human Variome Project recommendations.

The database describes variants according to the human reference sequence using standardised nomenclature, and provides the corresponding clinical data and contact details for the submitting laboratory in case more clinical information is required. The database includes non-pathogenic variants and, for epidemiological purposes, variants that have been reported previously in a different family. The database does not yet include patients with similar but atypical phenotypes where no mutation has been found and that potentially identify new genetic loci. The COL4A5 database is curated by 'experts' - a renal physician and laboratory scientist, who are both involved in genetic testing for Alport syndrome. It uses the Leiden Open Variation Database (LOVD) format [29], and is hosted and supported by a server at the University of Leiden, the single largest host of human variant databases worldwide. The LOVD site also provides access to a program to verify mutation nomenclature (Mutalyser), and is linked to both the major US and

European Bioinformatics sites at NCBI and EBI. Access to the LOVD *COL4A5* variant database is freely available.

All variants submitted to the LOVD database will be shared with an independent , public and also expertly-curated *COL4A5* database hosted by the University of Utah (www.arup.utah.edu/database/) [30]. The two databases cross-reference each other. Why have two databases? This is a pragmatic solution to the current circumstance of two databases that were established separately and after a lot of effort. Sharing information and a link means that no laboratory, clinician or patient will be disadvantaged by consulting a less up- to-date source. In addition, continuing with two independent but collaborative databases provides a means of cross-checking for completeness and accuracy, and increases the likelihood of sustainability. Only one of these databases includes clinical information.

Many web-based non-LOVD-associated collections of DNA variants are associated with descriptions of the corresponding gene or protein, but include only published variants and not those from laboratories' own collections. Others are incomplete, lack clinical features, and, frequently, their descriptions incorporate errors (5–40%) [31]. This is partly because their curators are responsible for many variant collections rather than being experts in the gene of interest.

The Human Variome Project has also provided guidelines on the ethical issues involved in establishing and maintaining a database and, in particular, those posed by sharing patient information across national borders. Its recommendations suggest the need for balance between the value of sharing clinical details and the recognition of an individual's right to privacy [32]. The Human Variome Project acknowledges that different cultures have different views that will inform national law. However, many of these issues have already been resolved previously with

publication of mutations and the corresponding phenotypes in journal articles. Often permission for genetic testing includes permission to publish variants and de-identified clinical information in scientific manuscripts and online. Consumer members of the Human Variome Project emphasise that individuals with inherited disease who undergo genetic testing overwhelmingly support their de-identified genetic information being used to help alleviate another family's suffering or potentially identify a cure for their disease. Nevertheless some databases choose to make clinical information only accessible to registered users, usually other health professionals. The *COL4A5* database provides clinical data to all users and also the submitting laboratory's contact details for further information.

Uses of the *COL4A5* database

1. Distinction between pathogenic and non-pathogenic variants

The distinction between a pathogenic and non-pathogenic *COL4A5* variant is critical for genetic testing laboratories, because the diagnosis of Alport syndrome implies the affected individual and subsequent generations of their family will develop renal failure. Interpreting a normal variant as pathogenic results in the individual being misinformed about their prognosis. Conversely, misinterpreting a disease-causing variant as normal means that the patient does not have the opportunity for treatment to retard renal failure progression and for reproductive counselling. In addition, family members cannot determine their disease status.

The distinction between pathogenicity and non-pathogenicity requires the documentation of disease-causing and normal variants, from genetic diagnostic laboratories. To date, fewer than 100 non-pathogenic *COL4A5* variants have been

described by diagnostic laboratories. While intronic changes are likely to be more abundant because of the larger intron size, they are not detected because laboratory testing strategies focus on the DNA coding region. Non-pathogenic variants are also less well-documented in non-Europeans, which makes variant interpretation in Chinese, African and Indian people particularly problematic. Thus, identifying novel non-pathogenic *COL4A5* variants within the coding regions from non-Europeans is a priority for our databases. The '1000 Genomes' and 'Hapmap' Projects have identified relatively few normal coding region *COL4A5* variants but exomic and next generation sequencing are likely to do so.

2. Using mutations to predict clinical phenotype and severity

In X-linked Alport syndrome, some pathogenic DNA variants are consistently associated, in different families, with early onset renal failure and extrarenal features, such as hearing loss, lenticonus and retinopathy [8,33]. These associations have been sufficiently strong to develop rules for genotype-phenotype correlations that have been confirmed in different populations. Thus reference to a DNA variant database potentially indicates to the clinician, from the time of presentation, the likelihood of early onset renal failure and the possible need for renin-angiotensin system blockade even before the onset of proteinuria [34,35].

Both the nature and the location of a *COL4A5* mutation contribute to the likelihood of early onset renal failure and extra-renal manifestations in males [8,9,20,33]. Thus, large deletions and rearrangements, nonsense and splicing mutations are associated more often with renal failure before the age of 30 years, as well as hearing loss, lenticonus and retinopathy [8,9,20,33]. Missense mutations at the

collagen chain carboxy terminus, or where glycine is replaced with bulkier or charged amino acids such as arginine, glutamic acid or glutamine [8,22,33] also often result in severe disease. Mutations affecting major structural or ligand-binding sites, such as those for integrins, may result in a poorer outcome too [24,36]. Importantly, there are also disease-associated variants that produce a mild clinical phenotype with later onset renal failure, possibly because the variants are located near non-collagenous interruptions [10].

In females, the effect of a mutation on clinical phenotype is less consistent because of random X chromosome inactivation.

The more pathogenic and non-pathogenic variants reported, the greater our ability to use precedent to predict clinical phenotype. Even when a variant has not been described before, a variant database that includes clinical features increases our understanding of its effect on the collagen $\alpha 5(\text{IV})$ chain and thus phenotype.

3. Increasing our understanding of the biochemistry of the collagen IV $\alpha 3\alpha 4\alpha 5$ network in the glomerular basement membrane

The very first DNA variant database was established to explain the structure-function relationships of the globin chain, and this approach has proven useful for many proteins including type I collagen [24]. A linear protein map or 'interactome' of the collagen I $\alpha 1\alpha 1\alpha 2$ heterotrimer demonstrates a major cell interaction domain, that regulates integrin-mediated cell binding and fibril remodelling, and a matrix interaction domain that determines cross-linking between the chains, proteoglycan interactions and tissue mineralisation. Missense mutations affecting

the collagen I major ligand-binding sites result in more severe forms of the inherited disease, Osteogenesis imperfecta [24].

Interactomes have recently been published for the collagen IV $\alpha1\alpha1\alpha2$, $\alpha3\alpha4\alpha5$ and $\alpha5\alpha5\alpha6$ molecules [36]. These demonstrate functional domains for autoimmunity, tumor growth and inhibition, and infection, with multiple related ligands binding within these locations. The maps also suggest that missense mutations affecting major ligand binding sites result in distinctive phenotypes. Thus missense mutations affecting the collagen IV $\alpha1$ chain that produce Hereditary Angiopathy, Nephropathy Aneurysms and muscle Cramps (HANAC) syndrome are all located near the binding site for von Hippel Lindau binding protein, and HANAC syndrome shares clinical features with von Hippel Lindau disease. In addition, missense mutations affecting integrin binding sites in the collagen IV $\alpha5$ chain possibly result in X-linked Alport syndrome with early onset renal failure [23,36].

The better populated the collagen IV maps are with structural domains, ligand-binding sites and missense mutations, the more information they yield about the effect of mutations on clinical phenotype and about collagen IV biochemistry.

Databases available for variants in other inherited renal diseases

Databases for the *COL4A3* and *COL4A4* genes have also been established (http://grenada.lumc.nl/LOVD2/COL4A/home.php?select_db=COL4A3, COL4A4), to improve the diagnostic accuracy and genotype-phenotype correlations in autosomal recessive Alport syndrome and Thin basement membrane nephropathy. These can be used to correlate mutations with clinical phenotypes in autosomal recessive Alport syndrome, and to explain why some heterozygous mutations

result in Thin basement membrane nephropathy with normal renal function, and others produce autosomal dominant Alport syndrome with renal failure.

Mutation databases are already available for many other renal diseases including autosomal dominant polycystic kidney disease (*PKD1*, *PKD2* variants), autosomal recessive polycystic kidney disease (*PKHD1* variants) [37,38], focal and segmental glomerulosclerosis (*NPHS2* variants) [39] and nephronophthisis [40]. Databases for the commonest renal diseases are often curated by the corresponding major laboratories working on these diseases, and are complete and up to date. However many other databases, especially those for which there is a fee, are curated by non-experts, and include only published variants, rather than encouraging diagnostic laboratories to submit their still unpublished mutations. In addition, these databases rarely include clinical features or normal gene variants. However these are also common problems with mutation databases for diseases in other organ systems [41].

Conclusions

The Human Variome Project envisages a curated locus- or gene-specific variant database for each protein affected in human disease [42]. Pathogenic variants have been described in more than 120 different genes in inherited kidney disease [43], but collaborative efforts are required to develop variant databases that are current, complete, accurate and freely-available. These will benefit patients, diagnostic laboratories, clinicians and researchers.

Acknowledgements: We would like to thank the many patients and families with X-linked Alport syndrome whose mutations are published in the databases. We would also like to thank Dr Dev Batish for helpful discussions.

Competing Interests: The authors have no Conflicts of Interest to declare but DB, NZ and HS are all employed by independent commercial or publicly-listed laboratories that test for Alport syndrome.

Funding: We would also like to thank the US Alport Syndrome Foundation for providing establishment funding for the *COL4A5* database, and the Australian Alport Foundation for providing funding to determine normal variants in non-Caucasian ethnicities.

References

1. Levy M, Feingold J (2000) Estimating prevalence in single-gene kidney diseases progressing to renal failure. *Kidney Int* 58:925-943
2. Feingold J, Bois E, Chompret A, Broyer M, Gubler MC, Grunfeld JP (1985) Genetic heterogeneity of Alport syndrome. *Kidney Int* 27:672-677
3. Barker DF, Hostikka SL, Zhou J, Chow LT, Oliphant AR, Gerken SC, Gregory MC, Skolnick MH, Atkin CL, Tryggvason K (1990) Identification of mutations in the COL4A5 collagen gene in Alport syndrome. *Science* 248:1224-1227
4. Flinter FA, Cameron JS, Chantler C, Houston I, Bobrow M (1988) Genetics of classic Alport's syndrome. *Lancet* 2:1005-1007
5. Gubler M, Levy M, Broyer M, Naizot C, Gonzales G, Perrin D, Habib R (1981) Alport's syndrome. A report of 58 cases and a review of the literature. *Am J Med* 70:493-505
6. Hanson H, Storey H, Pagan J, Flinter F (2011) The value of clinical criteria in identifying patients with X-linked Alport syndrome. *Clin J Am Soc Nephrol* 6:198-203
7. Teekhasaene C, Nimmanit S, Wutthiphan S, Vareesangthip K, Laohapand T, Malasit P, Ritch R (1991) Posterior polymorphous dystrophy and Alport syndrome. *Ophthalmology* 98:1207-1215
8. 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 (2000) X-linked Alport syndrome:

natural history in 195 families and genotype- phenotype correlations in males. J Am Soc Nephrol 11:649-657

9. Tan R, Colville D, Wang YY, Rigby L, Savage J (2010) Alport retinopathy results from "severe" COL4A5 mutations and predicts early renal failure. Clin J Am Soc Nephrol 5:34-38

10. Demosthenous P, Voskarides K, Stylianou K, Hadjigavriel M, Arsali M, Patsias C, Georgaki E, Ziropiannis P, Stavrou C, Daphnis E, Pierides A, Deltas C (2011) X-linked Alport syndrome in Hellenic families: Phenotypic heterogeneity and mutations near interruptions of the collagen domain in COL4A5. Clin Genet 81:240-248

11. Dagher H, Buzza M, Colville D, Jones C, Powell H, Fassett R, Wilson D, Agar J, Savage J (2001) A comparison of the clinical, histopathologic, and ultrastructural phenotypes in carriers of X-linked and autosomal recessive Alport's syndrome. Am J Kidney Dis 38:1217-1228

12. Churg J, Sherman RL (1973) Pathologic characteristics of hereditary nephritis. Arch Pathol 95:374-379

13. van der Loop FT, Monnens LA, Schroder CH, Lemmink HH, Breuning MH, Timmer ED, Smeets HJ (1999) Identification of COL4A5 defects in Alport's syndrome by immunohistochemistry of skin. Kidney Int 55:1217-1224

14. Yoshioka K, Hino S, Takemura T, Maki S, Wieslander J, Takekoshi Y, Makino H, Kagawa M, Sado Y, Kashtan CE (1994) Type IV collagen alpha 5 chain. Normal distribution and abnormalities in X-linked Alport syndrome revealed by monoclonal antibody. Am J Pathol 144:986-996

15. Deltas C, Gale D, Cook T, Voskarides K, Athanasiou Y, Pierides A (2013) C3 Glomerulonephritis/CFHR5 Nephropathy Is an Endemic Disease in Cyprus: Clinical and Molecular Findings in 21 Families. *Adv Exp Med Biol* 734:189-196
16. Zhou J, Leinonen A, Tryggvason K (1994) Structure of the human type IV collagen COL4A5 gene. *J Biol Chem* 269:6608-6614
17. Bockenhauer D, Medlar AJ, Ashton E, Kleita R, Lench N (2012) Genetic testing in renal disease. *Pediatr Nephrol* 27:873-883
18. Bale S, Devisscher M, Van Criekinge W, Rehm HL, Decouttere F, Nussbaum R, Dunnen JT, Willems P (2011) MutaDATABASE: a centralized and standardized DNA variation database. *Nat Biotechnol* 29:117-118
19. Hertz JM, Thomassen M, Storey H, Flinter F (2012) Clinical utility gene card for: Alport syndrome. *Eur J Hum Genet*. DOI:10.1038/EJHG.2011.237
20. 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 (2003) 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
21. Lemmink HH, Schroder CH, Monnens LA, Smeets HJ (1997) The clinical spectrum of type IV collagen mutations. *Hum Mutat* 9:477-499
22. Persikov AV, Pillitteri RJ, Amin P, Schwarze U, Byers PH, Brodsky B (2004) 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

23. Marini JC, Forlino A, Cabral WA, Barnes AM, San Antonio JD, Milgrom S, Hyland JC, Korkko J, Prockop DJ, De Paepe A, Coucke P, Symoens S, Glorieux FH, Roughley PJ, Lund AM, Kuurila-Svahn K, Hartikka H, Cohn DH, Krakow D, Mottes M, Schwarze U, Chen D, Yang K, Kuslich C, Troendle J, Dalglish R, Byers PH (2007) Consortium for osteogenesis imperfecta mutations in the helical domain of type I collagen: regions rich in lethal mutations align with collagen binding sites for integrins and proteoglycans. *Hum Mutat* 28:209-221
24. Sweeney SM, Orgel JP, Fertala A, McAuliffe JD, Turner KR, Di Lullo GA, Chen S, Antipova O, Perumal S, Ala-Kokko L, Forlino A, Cabral WA, Barnes AM, Marini JC, San Antonio JD (2008) Candidate cell and matrix interaction domains on the collagen fibril, the predominant protein of vertebrates. *J Biol Chem* 283:21187-21197
25. Harris PC, Rossetti S (2010) Molecular diagnostics for autosomal dominant polycystic kidney disease. *Nat Rev Nephrol* 6:197-206
26. Cotton RG (2000) Progress of the HUGO mutation database initiative: a brief introduction to the human mutation MDI special issue. *Hum Mutat* 15:4-6
27. Cotton RG, Auerbach AD, Beckmann JS, Blumenfeld OO, Brookes AJ, Brown AF, Carrera P, Cox DW, Gottlieb B, Greenblatt MS, Hilbert P, Lehvaslaiho H, Liang P, Marsh S, Nebert DW, Povey S, Rossetti S, Scriver CR, Summar M, Tolan DR, Verma IC, Vihinen M, den Dunnen JT (2008) Recommendations for locus-specific databases and their curation. *Hum Mutat* 29:2-5
28. Kaput J, Cotton RG, Hardman L, Watson M, Al Aqeel AI, Al-Aama JY, Al-Mulla F, Alonso S, Aretz S, Auerbach AD, Bapat B, Bernstein IT, Bhak J, Bleoo SL, Blocker H, Brenner SE, Burn J, Bustamante M, Calzone R, Cambon-Thomsen

A, Cargill M, Carrera P, Cavedon L, Cho YS, Chung YJ, Claustres M, Cutting G, Dagleish R, den Dunnen JT, Diaz C, Dobrowolski S, dos Santos MR, Ekong R, Flanagan SB, Flicek P, Furukawa Y, Genuardi M, Ghang H, Golubenko MV, Greenblatt MS, Hamosh A, Hancock JM, Hardison R, Harrison TM, Hoffmann R, Horaitis R, Howard HJ, Barash CI, Izagirre N, Jung J, Kojima T, Laradi S, Lee YS, Lee JY, Gil-da-Silva-Lopes VL, Macrae FA, Maglott D, Marafie MJ, Marsh SG, Matsubara Y, Messiaen LM, Moslein G, Netea MG, Norton ML, Oefner PJ, Oetting WS, O'Leary JC, de Ramirez AM, Paalman MH, Parboosingh J, Patrinos GP, Perozzi G, Phillips IR, Povey S, Prasad S, Qi M, Quin DJ, Ramesar RS, Richards CS, Savige J, Scheible DG, Scott RJ, Seminara D, Shephard EA, Sijmons RH, Smith TD, Sobrido MJ, Tanaka T, Tavtigian SV, Taylor GR, Teague J, Topel T, Ullman-Cullere M, Utsunomiya J, van Kranen HJ, Vihinen M, Webb E, Weber TK, Yeager M, Yeom YI, Yim SH, Yoo HS (2009) Planning the human variome project: the Spain report. *Hum Mutat* 30:496-510

29. Fokkema IF, den Dunnen JT, Taschner PE (2005) LOVD: easy creation of a locus-specific sequence variation database using an "LSDB-in-a-box" approach. *Hum Mutat* 26:63-68

30. Crockett DK, Pont-Kingdon G, Gedge F, Sumner K, Seamons R, Lyon E (2010) The Alport syndrome COL4A5 variant database. *Hum Mutat* 31:E1652-1657

31. Gout AM, Ravine D, Harris PC, Rossetti S, Peters D, Breuning M, Henske EP, Koizumi A, Inoue S, Shimizu Y, Thongnoppakhun W, Yenchitsomanus PT, Deltas C, Sandford R, Torra R, Turco AE, Jeffery S, Fontes M, Somlo S, Furu LM, Smulders YM, Mercier B, Ferec C, Burtey S, Pei Y, Kalaydjieva L, Bogdanova N, McCluskey M, Geon LJ, Wouters CH, Reiterova J, Stekrova J, San Millan JL,

Aguiari G, Del Senno L (2007) Analysis of published PKD1 gene sequence variants. *Nat Genet* 39:427-428

32. Povey S, Al Aqeel AI, Cambon-Thomsen A, Dagleish R, den Dunnen JT, Firth HV, Greenblatt MS, Barash CI, Parker M, Patrinos GP, Savage J, Sobrido MJ, Winship I, Cotton RG (2010) Practical guidelines addressing ethical issues pertaining to the curation of human locus-specific variation databases (LSDBs). *Hum Mutat* 31:1179-1184

33. Gross O, Netzer KO, Lambrecht R, Seibold S, Weber M (2002) Meta-analysis of genotype-phenotype correlation in X-linked Alport syndrome: impact on clinical counselling. *Nephrol Dial Transplant* 17:1218-1227

34. Gross O, Beirowski B, Koepke ML, Kuck J, Reiner M, Addicks K, Smyth N, Schulze-Lohoff E, Weber M (2003) Preemptive ramipril therapy delays renal failure and reduces renal fibrosis in COL4A3-knockout mice with Alport syndrome. *Kidney Int* 63:438-446

35. Kashtan CE, Ding J, Gregory M, Gross O, Heidet L, Knebelmann B, Rheault M, Licht C (2013) Clinical practice recommendations for the treatment of Alport syndrome: a statement of the Alport Syndrome Research Collaborative. *Pediatr Nephrol* 28:5-11

36. Parkin JD, San Antonio JD, Pedchenko V, Hudson B, Jensen ST, Savage J (2011) Mapping structural landmarks, ligand binding sites, and missense mutations to the collagen IV heterotrimers predicts major functional domains, novel interactions, and variation in phenotypes in inherited diseases affecting basement membranes. *Hum Mutat* 32:127-143

37. Bergmann C, Senderek J, Kupper F, Schneider F, Dornia C, Windelen E, Eggermann T, Rudnik-Schoneborn S, Kirfel J, Furu L, Onuchic LF, Rossetti S, Harris PC, Somlo S, Guay-Woodford L, Germino GG, Moser M, Buttner R, Zerres K (2004) PKHD1 mutations in autosomal recessive polycystic kidney disease (ARPKD). *Hum Mutat* 23:453-463
38. Gout AM, Martin NC, Brown AF, Ravine D (2007) PKDB: Polycystic Kidney Disease Mutation Database--a gene variant database for autosomal dominant polycystic kidney disease. *Hum Mutat* 28:654-659
39. Boute N, Gribouval O, Roselli S, Benessy F, Lee H, Fuchshuber A, Dahan K, Gubler MC, Niaudet P, Antignac C (2000) NPHS2, encoding the glomerular protein podocin, is mutated in autosomal recessive steroid-resistant nephrotic syndrome. *Nat Genet* 24:349-354
40. Hildebrandt F, Attanasio M, Otto E (2009) Nephronophthisis: disease mechanisms of a ciliopathy. *J Am Soc Nephrol* 20:23-35
41. Watson MS, Epstein C, Howell RR, Jones MC, Korf BR, McCabe ER, Simpson JL (2008) Developing a national collaborative study system for rare genetic diseases. *Genet Med* 10:325-329
42. Pennisi E (2007) Genetics. Working the (gene count) numbers: finally, a firm answer? *Science* 316:1113
43. Hildebrandt F (2010) Genetic kidney diseases. *Lancet* 375:1287-1295

Table 1: Uses of genetic testing in X-linked Alport syndrome

- To confirm the diagnosis of Alport syndrome
- To exclude the diagnosis of Thin basement membrane nephropathy in patients with persistent hematuria
- To indicate the mode of inheritance, and hence the risk of renal failure in other family members. To facilitate 'cascade testing' for at risk family members
- To help predict the risk of early or late onset renal failure, either because this has been reported previously for the same variant, or else, based on the variant's nature and location. Those individuals with a variant associated with early onset renal failure should undergo treatment with ACE inhibitors. Knowing the pathogenic variant also indicates the risk of renal failure in asymptomatic females who insist on donating a kidney to a family member
- To enable early prenatal diagnosis for women at risk of an affected pregnancy
- To facilitate preimplantation genetic diagnosis of embryos prior to use in IVF. Disease haplotypes from linkage studies can be used if the mutation is not known
- To help predict the likelihood of antglomerular basement membrane disease post-transplantation
- To determine whether a family member is unaffected and may be a renal donor
- To increase our understanding of the biochemistry of collagen IV through correlating the effect of pathogenic variants on the protein sequence and clinical phenotype

Table 2: Major recommendations of the Human Variome Project for Locus-Specific Databases and their curation [27]

- Variants are described according to the gene reference sequence and using standardised nomenclature
- Data fields are standardised
- The database is maintained by a team of expert curators with complementary skills
- The database is web-based and freely accessible
- Those individuals and laboratories who submit variants are credited
- The database indicates the number of times the same variant has been identified in different families
- The database includes non-pathogenic variants
- The database indicates the likely pathogenicity of a variant and the supporting evidence for this
- The database is linked to other relevant databases such as those with clinical and structural protein information
- The database is reviewed and updated regularly