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

OPEN

Aleksandra M. Żurowska[1,2,20], Olga Bielska[2,20], Patrycja Daca-Roszak[3], Maciej Jankowski[4], Maria Szczepańska[5], Dagmara Roszkowska-Bjanid[6], Elżbieta Kuźma-Mroczkowska[7,21], Małgorzata Pańczyk-Tomaszewska[7], Anna Moczulska[8], Dorota Drożdż[8], Despina Hadjipanagi[9], Constantinos Deltas[9], Danuta Ostalska-Nowicka[10], Alina Rabiega[10], Janina Taraszkiewicz[11], Katarzyna Taranta-Janusz[12], Anna Wieczorkiewicz-Plaza[13], Katarzyna Jobs[14], Judyta Mews[14], Kinga Musiał[15], Anna Jakubowska[15], Hanna Nosek[16], Anna E. Jander[17], Constantina Koutsofti[9], Anna Stanisławska-Sachadyn[18], Dominka Kuleszo[4], Ewa Ziętkiewicz[3] and Beata S. Lipska-Ziętkiewicz[1,19]

[1]Rare Diseases Centre, Medical University of Gdańsk, Gdańsk, Poland; [2]Department of Pediatrics, Nephrology and Hypertension, Faculty of Medicine, Medical University of Gdańsk, Gdańsk, Poland; [3]Institute of Human Genetics, Polish Academy of Sciences, Poznań, Poland; [4]Department of Biology and Medical Genetics, Faculty of Medicine, Medical University of Gdańsk, Gdańsk, Poland; [5]Department of Pediatrics, Faculty of Medical Sciences in Zabrze, Medical University of Silesia in Katowice, Zabrze, Poland; [6]Pediatric Nephrology Ward With Dialysis Division for Children, Public Clinical Hospital, Zabrze, Poland; [7]Department of Pediatrics and Nephrology, Medical University of Warsaw, Warsaw, Poland; [8]Department of Pediatric Nephrology and Hypertension, Jagiellonian University Medical College, Cracow, Poland; [9]Center of Excellence in Biobanking and Biomedical Research, Molecular Medicine Research Center, University of Cyprus Medical School, Nicosia, Cyprus; [10]Department of Pediatric Cardiology and Hypertensiology, Poznan University of Medical Sciences, Poznań, Poland; [11]Department of Pediatric Nephrology, Team of Municipal Hospitals in Chorzów, Chorzów, Poland; [12]Department of Pediatrics and Nephrology, Medical University of Białystok, Białystok, Poland; [13]Department Pediatric Nephrology, Medical University Lublin, Lublin, Poland; [14]Department of Pediatrics, Pediatric Nephrology and Allergology, Military Institute of Medicine, Warsaw, Poland; [15]Department of Pediatric Nephrology, Wrocław Medical University, Wrocław, Poland; [16]Department of Pediatrics, Gastroenterology and Nutrition, University of Warmia and Mazury, Olsztyn, Poland; [17]Department of Pediatrics, Immunology and Nephrology, Polish Mothers Memorial Hospital Research Institute, Łódź, Poland; [18]Department of Molecular Biotechnology and Microbiology, Gdańsk University of Technology, Gdańsk, Poland; and [19]Clinical Genetics Unit, Department of Biology and Medical Genetics, Faculty of Medicine, Medical University of Gdańsk, Gdańsk, Poland

A study of 269 children enrolled into a National Registry for children with persistent glomerular hematuria identified 131 individuals with genetically confirmed X-linked Alport Syndrome. A single variant c.1871G>A p.Gly624Asp (G624D) in *COL4A5* was predominant and accounted for 39% of X-linked Alport Syndrome in unrelated Polish families (44 of 113). To evaluate its origins, the genetic variation in a 2.79 Mb segment encompassing the *COL4A5 locus* on chromosome X was assessed. All G624D alleles were found on the same rare haplotype background, indicating a founder effect dating back to the 12-13th century. The phenotypic data of 131 children with X-linked Alport Syndrome and their 195 affected adult relatives revealed that the G624D variant was associated with a significantly milder clinical course in comparison to other pathogenic *COL4A5* variants. Furthermore the clinical course of this genetically uniform cohort was milder than that observed in individuals with other *COL4A5* missense mutations. In spite of the benign clinical manifestation throughout childhood and early adulthood, the G624D variant confers significant risk for both kidney failure and deafness in males, albeit 20-30 years later than that observed in individuals with other *COL4A5* pathogenic variants (50% cumulative risk of starting dialysis at 54 years (95% confidence interval: 50-62) v. 26 years (95% confidence interval: 22-30)). Thus, males with G624D are candidates for existing and emerging therapies for Alport Syndrome.

Correspondence: *Aleksandra M. Żurowska, Rare Diseases Centre, Department of Pediatrics, Nephrology and Hypertension, Faculty of Medicine, Medical University of Gdańsk, Debinki 7, 80-211 Gdańsk, Poland. E-mail: aleksandra.zurowska@gumed.edu.pl; and Beata S. Lipska-Zi tkiewicz, Rare Diseases Centre, Department of Biology and Medical Genetics, Clinical Genetics Unit, Faculty of Medicine, Medical University of Gdańsk, Debinki 1, 80-211 Gdańsk, Poland. E-mail: b.lipska@gumed.edu.pl*

[20]AMŻ and OB contributed equally.

[21]EK-M died February 11, 2020.

Received 13 March 2020; revised 23 October 2020; accepted 30 October 2020; published online 10 December 2020

Kidney International (2021) **99,** 1451–1458; https://doi.org/10.1016/j.kint.2020.10.040

KEYWORDS: Alport syndrome; chronic kidney disease; *COL4A5*; prediction; genetic

The typical clinical features of Alport syndrome (AS) are persistent microscopic hematuria, nonnephrotic range proteinuria, end-stage kidney failure, hearing loss, and often a family history of hematuria or renal failure. The vast majority (85%) of subjects with AS have an X-linked disease (XLAS; Online Mendelian Inheritance in Man number #301050; Orphanet rare disease nomenclature [ORPHA]: 88917) and harbor pathogenic variants in the *COL4A5* gene, encoding collagen type IV α 5 chain.[1] To date, several hundreds of causative variants, by and large private, have been identified.[2–4] The type of gene defect has been correlated with the severity and progression of clinical symptoms.[2,5,6] Large chromosomal rearrangements and frameshift and nonsense mutations correlate with a more severe clinical course and faster progression of both renal and extrarenal symptoms, with a 90% risk of end-stage renal disease (ESRD) by the age of 30 years. Missense mutations are associated with a milder clinical course, with a 50% risk of ESRD by the age of 30 years.[5,6] The early recognition of a poor future outcome in subjects with AS is important as interventional therapy is offered to those at risk of renal damage.[7] Here, we report the phenotype of, the largest to date, genetically homogeneous cohort of XLAS patients with a single causative pathogenic variant, p.Gly624Asp.

RESULTS

Clinical characteristics of the study cohort

The cohort was established through next-generation sequencing–based genetic testing of *COL4A3-4-5* genes in subjects enrolled into a Polish National Registry for children with persistent glomerular hematuria, comprising a total of 269 unrelated index cases and >200 family members. A total of 131 children from 113 families had a *COL4A5* pathogenic variant. The NM_000495.5:c.1871G>A substitution (*rs104886142*; p.Gly624Asp, also referred to as G624D) was found in 52 children from 44 families. Of these, one female was a compound heterozygote, with p.Gly624Asp accompanied by c.1913G>A (p.Gly638Asp); no homozygous females were identified. Altogether, p.Gly624Asp accounted for 39% of the genetically confirmed XLAS, thus representing a predominant variant in the national cohort.

The comparison of the clinical manifestation in 51 pediatric subjects (43 probands and 8 affected siblings) with p.Gly624Asp and 79 children (69 probands and 10 siblings) with other pathogenic *COL4A5* variants (further referred to as non-p.Gly624Asp) is presented in Table 1. The age at phenotype assessment and sex distribution were similar in both groups. A significantly higher proportion of p.Gly624Asp subjects had only low-grade and/or intermittent hematuria ($P < 0.01$), with no gross hematuria incidents in their medical history. A significantly lower proportion of p.Gly624Asp children reported proteinuria ($P < 0.001$) or albuminuria ($P < 0.01$). Hearing loss was less common in p.Gly624Asp boys than in non-p.Gly624Asp boys ($P < 0.05$). The incidence of decreased glomerular filtration rate (6.1% vs. 13.2%) and hypertension (2.0% vs. 7.8%) was low in both pediatric cohorts.

Subsequently, clinical data of affected adult relatives of the probands for whom pedigree and phenotypic data were consistent with a clinical diagnosis of AS was collected

Table 1 | Phenotypic characteristics in childhood (aged <18 years) of subjects carrying p.Gly624Asp and other (non-p.Gly624Asp) *COL4A5* pathogenic variants

Characteristic	p.Gly624Asp			Other *COL4A5* pathogenic variants		
	Boys	Girls	Total	Boys	Girls	Total
No.	22	29	51	36	43	79
Age at assessment, median (IQR) (yr)	11 (7–15)	9 (6–13)	10 (6–13)	9 (6–13.5)	8 (5–12)	9 (6–13)
Hematuria (<10/11–50/>50 RBCs/HPF)	3/12/7 (13.6/54.6/31.8)	9/15/5 (31.0/51.7/17.3)[a]	12/27/12 (23.5/53.0/23.5)[b]	1/15/20 (2.8/41.7/55.5)	3/25/15 (7.0/58.1/34.9)	4/40/35 (5.1/50.6/44.3)
Incidents of gross hematuria	3/20 (15.0)	3/26 (11.5)	6/46 (13.0)[a]	14/35 (40.0)	11/37 (29.7)	25/72 (34.7)
Intermittent hematuria	7/22 (31.8)[b]	4/29 (13.8)	11/51 (21.6)[b]	1/36 (2.8)	3/43 (7.0)	4/79 (5.1)
Proteinuria	5/20 (25.0)[c]	5/27 (18.5)[c]	10/47 (21.3)[c]	27/35 (77.1)	26/42 (61.9)	53/77 (68.8)
Albuminuria	7/20 (35.0)[b]	10/27 (37.0)[a]	17/47 (36.2)[b]	21/28 (75.0)	25/40 (62.5)	46/68 (67.6)
ACEI/ARB therapy	1/20 (5.0)[c]	3/26 (11.5)[a]	4/46 (8.7)[c]	19/35 (54.3)	15/41 (36.6)	34/76 (44.7)
eGFR (FAS) <90 ml/min per 1.73 m^2	3/21 (14.3)	0/28 (0.0)[a]	3/49 (6.1)	3/34 (8.8)	7/42 (16.7)	10/76 (13.2)
Hypertension	0/22 (0.0)	1/29 (3.4)	1/51 (2.0)	2/35 (5.7)	4/42 (9.5)	6/77 (7.8)
Hearing loss	1/14 (7.1)[a]	2/19 (10.5)	3/33 (9.1)	12/30 (40.0)	3/32 (9.4)	15/62 (24.2)
Family history of hematuria	—	—	40/42 (95.2)	—	—	58/67 (86.6)
Family history of ESRD	—	—	9/42 (21.4)[c]	—	—	44/67 (65.7)
Family history of hearing loss	—	—	10/42 (23.8)[b]	—	—	37/67 (55.2)

ACEI, angiotensin-converting enzyme; ARB, angiotensin receptor blocker; eGFR, estimated glomerular filtration rate; ESRD, end-stage renal disease; FAS, full age spectrum; HPF, high-power field; IQR, interquartile range; RBC, red blood cell.

Data are given as number (percentage) or number/total (percentage), unless otherwise indicated. Statistical analysis is performed separately for the subgroups: boys, girls, and total. Intermittent hematuria is defined as episodic nonhematuria urinalysis in a hematuric subject. Proteinuria is defined as urine protein-to-creatinine ratio >0.2 mg/mg or >0.1 g/m^2 per 24 hours. Albuminuria is defined as urine albumin-to-creatinine ratio >30 mg/g or >30 mg/24 h. eGFR is estimated using FAS equation. Hypertension is defined according to American Academy of Pediatrics (2017). Hearing loss is diagnosed by audiogram. Family history of hematuria, ESRD, and hearing loss is counted for index cases.

[a]$P < 0.05$.

[b]$P < 0.01$.

[c]$P < 0.001$ (Fisher exact probability test).

through telephone surveys. The data were available for 195 subjects: 71 from p.Gly624Asp families (median age, 42 [range, 21–85] years) and 124 from non-p.Gly624Asp families (median age, 39 [range, 18–69] years). A total of 29% (7 of 24) of affected males from p.Gly624Asp families developed ESRD at a median age of 51 (28, 36, 50, 51, 54, 60, and 62) years, compared with 91% (58 of 64) of males from non-p.Gly624Asp families, whose median age at dialysis onset was 24 (range, 14–52) years. A single p.Gly624Asp female developed ESRD in comparison to 11 (18%) females from non-p.Gly624Asp families. By the age of 40 years, only 2 of 17 males (12%) and a single female (1 of 26; 4%) from p.Gly624Asp families required renal replacement therapy, compared with 85% of males (51 of 60) and 22% (9 of 37) of females from non-p.Gly624Asp families ($P < 0.001$ and $P < 0.01$, respectively). The 50% cumulative risk of starting dialysis was reached by p.Gly624Asp males at the age of 54 years (95% confidence interval, 50–62 years) and by non-p.Gly624Asp males at the age of 26 years (95% confidence interval, 22–30 years) ($P < 0.0001$; Figure 1).

Hearing loss was reported by 24% of p.Gly624Asp family members compared with 55% of non-p.Gly624Asp families ($P < 0.01$). The audiogram data from adult family members were insufficient for a more detailed analysis of hearing disabilities.

A subgroup analysis comparing subjects harboring the p.Gly624Asp variant and 123 individuals (43 children and 80 adult relatives) with other pathogenic missense variants in the triple-helical region of collagen IV (the so-called Xaa-Yaa-Gly type variants) revealed a significantly milder phenotype of the p.Gly624Asp. Subjects with p.Gly624Asp variant less frequently presented proteinuria, albuminuria, and/or gross hematuria in childhood; and their affected adult relatives were less likely to report ESRD and/or hearing loss (Supplementary Table S1). The 50% cumulative risk of starting dialysis in the pGly624Asp cohort at the age of 54 years (95% confidence interval, 50–62 years) was reached by males with other Xaa-Yaa-Gly missense variants at a significantly younger age of 27 years (95% confidence interval, 22–31 years; $P < 0.001$) (Supplementary Figure S1).

Evaluation of the origins of the mutation

To shed light on the origin of the recurrent p.Gly624Asp variant, we characterized the genetic variation in a 2.79-Mb segment encompassing *COL4A5* locus (Figure 2), corresponding to the genetic distance of ~3.32 cM.[8] In addition to 44 subjects from p.Gly624Asp families, 2 Polish p.Gly624Asp males with subnephrotic proteinuria from the PodoNet Registry[9] and 8 Greek p.Gly624Asp males were included. The phase of the background haplotype was solved in 38 unrelated individuals, revealing at least 9 p.Gly624Asp-associated haplotypes, of which H1 observed in 14 of 38 phased p.Gly624Asp chromosomes was the most frequent and presumably represented the ancestral state (upper panel of Figure 2). The variability of the haplotype among 34 non-disease chromosomes examined in the patients and their

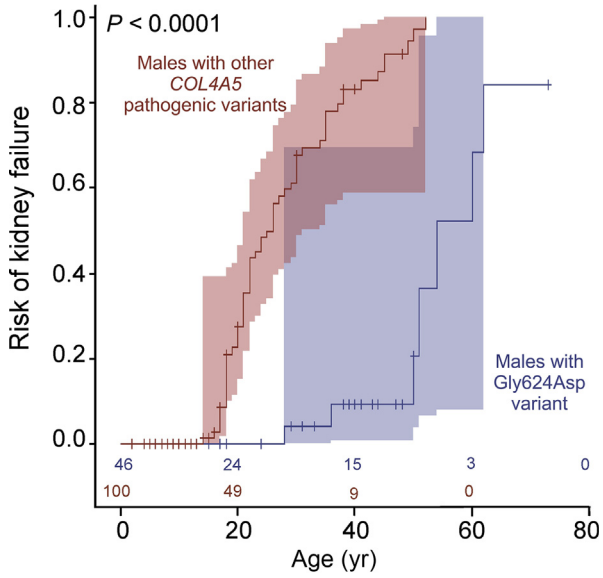

Figure 1 | Actuarial risk of kidney failure (end-stage renal disease [ESRD]) in 146 male subjects with Alport syndrome (AS; X-linked AS [XLAS]) due to p.Gly624Asp (46) and other *COL4A5* pathogenic variants (100). Kaplan-Meyer curve of probability of kidney failure (ESRD) with age, including number of subjects at risk and 95% Hall-Wellner bands. Males with p.Gly624Asp pathogenic variant in blue; XLAS males with other *COL4A5* pathogenic variants in red.

families was incomparably higher, with >25 different full-length (3.32-cM) haplotype variants, among which H1 linked with p.Gly642Asp was absent. The shorter core haplotype (AACC; 0.75 Mb or 0.89 cM), present in all p.Gly624Asp chromosomes, was rare (1%–2%), both in the examined control chromosomes from Polish population and in the reported general European population from the linkage disequilibrium database (ldlink.nci.nih.gov). These observations strongly suggest that all p.Gly624Asp chromosomes share a common ancestor.

To estimate the time of the purported founder effect, we applied an approach[10] based on the proportion of chromosomes with the nonrecombined, full-length p.Gly624Asp haplotype (see Methods). Assuming that the prevalent haplotype variant H1 represented the ancestral state, the age of the p.Gly624Asp haplotype was estimated at 30 generations. This places the date of the founder effect between ~1110 and ~1260 AD (assuming 30 and 25 years per generation, respectively).

DISCUSSION

The classically described natural history of XLAS due to *COL4A5* mutations, associated with progressive renal disease and hearing loss, early onset of ESRD in males, and a later onset in a significantly lower proportion of females, is based on descriptions of families from West Europe and North America.[2,5,6] The clinical manifestation of XLAS in childhood is not characteristic, with hematuria being the only detected symptom until proteinuria and a progressive decrease in

Sample	Sex	rs1285715	Msat	rs699669	rs5973851	rs6622333	G624D	rs1801164	rs62595622	rs12008095	rs839378	rs839407	rs5942853	Haplotype name
		←											→	
				1.82 cM						1.50 cM				
					0.75 cM									
AL214	M	G	10--6	A	A	C	A	C	C	A	C	G	C	H1
AL130	M	G	10--6	A	A	C	A	C	C	A	C	G	C	H1
AL208	M	G	10--6	A	A	C	A	C	C	A	C	G	C	H1
AL317	F,M	G	10--6	A	A	C	A	C	C	A	C	G	C	H1
AL113	M	G	10--6	A	A	C	A	C	C	A	C	G	C	H1
AL122	M	G	10--6	A	A	C	A	C	C	A	C	G	C	H1
AL135	M	G	10--6	A	A	C	A	C	C	A	C	G	C	H1
AL084	M	G	10--6	A	A	C	A	C	C	A	C	G	C	H1
AL009	M	G	10--6	A	A	C	A	C	C	A	C	G	C	H1
AL172	M	G	10--6	A	A	C	A	C	C	A	C	G	C	H1
AL194	M	G	10--6	A	A	C	A	C	C	A	C	G	C	H1
Cy-2	M	G	10--6	A	A	C	A	C	C	A	C	G	C	H1
Cy-7	M	G	10--6	A	A	C	A	C	C	A	C	G	C	H1
Cy-8	M	G	10--6	A	A	C	A	C	C	A	C	G	C	H1
Cy-1	M	T	10--6	A	A	C	A	C	C	A	C	G	C	H1-1
Cy-5	M	G	11--5	A	A	C	A	C	C	A	C	G	C	H1-2
AL222	M	G	10--6	A	A	C	A	C	C	A	C	G	T	H2
AL225	M	G	10--6	A	A	C	A	C	C	A	C	G	T	H2
Cy-3	M	G	10--6	A	A	C	A	C	C	A	C	G	T	H2
Cy-4	M	G	10--6	A	A	C	A	C	C	A	C	G	T	H2
AL201	M	G	10--6	A	A	C	A	C	C	A	C	G	T	H2
HEI029	M	G	10--6	A	A	C	A	C	C	A	C	G	T	H2
AL116*	F	G	10--6/11-5	A	a	c	A	c	C	a/g	C	G	T	H2 or H2-1
AL200*	F	G	10--6	a	A	C	A	C	C	A	c/g	g/c	T	H2 or H3 or H4
AL289	M	G	10--6	A	A	C	A	C	C	A	G	C	T	H3
GDA027	M	G	10--6	A	A	C	A	C	C	A	G	C	T	H3
AL308	F,M	G	10--6	A	A	C	A	C	C	A	G	C	T	H3
AL316	F,M	G	10--6	A	A	C	A	C	C	A	G	C	T	H3
AL195*	F	g/t	10--6	A	A	C	A	C	C	A	G	C	T	H3 or H3-2
AL003	M	G	10--6	A	A	C	A	C	C	A	G	C	C	H3-1
AL171	M	G	10--6	A	A	C	A	C	C	A	C	C	T	H4
AL057	F	G	10--6	A	A	C	A	C	C	A	C	C	T	H4
AL087	M	G	10--6	A	A	C	A	C	C	A	C	C	T	H4
AL360	F,M	G	10--6	A	A	C	A	C	C	A	C	C	T	H4
AL338	F,M	T	11--5	A	A	C	A	C	C	A	C	C	T	H4-1
AL191*	F	g/t	11--5	A	A	C	A	C	C	A	c/g	C	t/c	H4-1 or H4-2
AL111	F,M	G	10--6	A	A	C	A	C	T	A	C	C	T	H5
AL202*	F	G	10--6	a	a	c	A	c	T	a/g	C	c/g	T	H5 or H5-1
Cy-6	M	G	NA	A	A	NA	A	NA	C	A	NA	G	C	Unsolved
AL164	F	G	10--6	A	a	c	A	c	c/t	A	C	G	C	Unsolved
AL161	F	G	10--6	a	A	C	A	C	C	A	C	G	c/t	Unsolved
AL326	F	G	10--6	a	a	c	A	c	c/t	a/g	C	G	c/t	Unsolved
AL344	F	G	10--6	a	a	c	A	c	c/t	A	c/g	G	c/t	Unsolved
AL067	F	G	10-6/11-5	a	a	c	A	c	c/t	A	c/g	G	c/t	Unsolved
AL022	F	G	10-6/11-5	A	a	c	A	c	c/t	a/g	c/g	G	c/t	Unsolved
AL336	F	G	10--6	a	a	c	A	c	C	A	c/g	g/c	c/t	Unsolved
AL269	F	G	10--6	A	A	C	A	C	C	A	c/g	G	c/t	Unsolved
AL145	F	G	10--6	a	a	c	A	c	C	A	c/g	G	c/t	Unsolved
AL066	F	G	10--6	A	a	c	A	c	C	a/g	c/g	G	c/t	Unsolved
AL157	F	G	10-6/10-7	a	a	c	A	c	C	A	c/g	G	c/t	Unsolved
AL078	F	G	10-6/11-5	A	a	c	A	c	C	A	c/g	G	c/t	Unsolved
AL288	F	G	10-6/12-5	A	a	c	A	c	c/t	a/g	C	g/c	c/t	Unsolved
AL346	F	G	10-6/11-5	a	A	C	A	c	c/t	a/g	C	g/c	c/t	Unsolved
AL001	F	g/t	10-6/11-5	A	a	c	A	c	c/t	A	c/g	g/c	c/t	Unsolved

Figure 2 | *COL4A5* c.1871G>A p.Gly624Asp-associated haplotype variants. Top panel: The phase of the background haplotype was directly read in 26 hemizygous males (including 7 Greeks from Cyprus [Cy]), whereas in 7 heterozygous female samples, its phase (the distribution of polymorphic alleles among 2 X chromosomes) was inferred from family data. Nine distinct p.Gly624Asp-associated (continued)

estimated glomerular filtration rate develop in teenagers. The p.Gly624Asp variant broadens the phenotypic spectrum of *COL4A5* mutations with that of a clearly milder clinical course. In childhood until early adulthood, symptoms resemble those observed in thin basement membrane disease, although notably 36% of children with p.Gly624Asp variant demonstrate albuminuria and 21% demonstrate proteinuria at a median age of 10 years. Progression of disease is slow, but in midlife, affected males are at high risk of developing renal failure, albeit ~27 years later than that observed in classic descriptions of XLAS,[1,4–6] as shown in our study and in previous casuistic reports of patients from Greece, Slovenia, and Czech Republic.[11–13] Male subjects have a 50% cumulative probability of reaching ESRD by the age of 54 years. They should therefore benefit from existing treatment recommendations for XLAS and may be candidates for other emerging treatment options for slowing progression of renal damage.[14,15] In females, even in elderly age, kidney survival does not appear to be significantly affected. Nevertheless, the outcome data show some interfamilial variation. A single adult male has been identified in the presented cohort with negligible urine findings at the age of 30 years. Another unrelated male subject has not reached ESRD by the age of 70 years. On the other hand, 2 males and a single female from our cohort required dialysis before the age of 40 years; and, together with a report of a severe phenotype in a German family[16] with p.Gly624Asp, are notable exceptions to the generally observed mild phenotype. Although extensive genetic testing comprising whole-exome and whole-genome sequencing of 2 of these cases showed no additional pathogenic variants, one cannot exclude the possibility of an additive effect of modifying variants and/or environmental factors responsible for the acceleration of the disease in a vulnerable subset of patients.[16,17]

Nevertheless, symptoms observed in most p.Gly642Asp patients are mild, and proteinuria, hypertension, and deterioration of renal function as well as hearing problems manifest late, in midlife.

Glycine, the smallest amino acid with a hydrogen atom instead of a side chain, is characterized by the largest conformational freedom, allowing for the widest angle turn of the polypeptide chain. The presence of glycine in Xaa-Yaa-Gly repeats is essential for the formation of the characteristic structure of collagen, where 3 protomers interleave to form a linear, strong triple helix. Substitution of glycine in this conserved position by other amino acids introduces a steric hindrance, rendering helical conformation impossible, and results in a bulge or kink in the linear collagen molecule. In membranous collagens, such as collagen IV assemblies, such interruptions are naturally present. Collagen IV molecules are therefore less stable but more flexible; and instead of linear rope-like fibrils, they form mesh, creating collagenous sheets that are crucial in the construction of renal glomerular basement membrane.[18,19] In the amino acid sequence of *COL4A5*, the p.Gly624 residue (underlined) is adjacent to a natural interruption (Phe-Gly marked in bold) in the string of Xaa-Yaa-Gly triplets: Asn-Ile-Gly Pro-Met-Gly Pro-Pro-Gly **Phe-Gly** Pro-Pro-Gly Pro-Val-Gly Glu-Lys-Gly. Although there are no structural or functional data supporting such a hypothesis, one may postulate that substitution of glycine in this position widens the naturally occurring interruption in the Xaa-Yaa-Gly triplet cadence, which further weakens an already more flexible fragment. It may therefore constitute a less drastic change than breaking a naturally undisrupted long patch of Xaa-Yaa-Gly triplets in the remaining more rigid regions.[20] As a result, the consequence of p.Gly642Asp substitution for the structure of the glomerular basement membrane may be rather moderate.

The p.Gly624Asp variant, predominant in the analyzed cohort of XLAS Polish patients, has already been identified in clinical testing and listed as pathogenic in the ClinVar database (https://www.ncbi.nlm.nih.gov/clinvar/) by several laboratories (variation identifier: 24455). The rare haplotype background suggests that the repetitive occurrence of p.Gly624Asp is due to a founder effect rather than to recurrent mutations. The presented p.Gly624Asp subjects did not cluster to any specific region of Poland, indicating that this effect is not recent. p.Gly624Asp has primarily been reported in Central and East Europe (Poland, Czech Republic, Lithuania, Russia, Hungary, Slovenia, Croatia, and Bulgaria) but also anecdotally in Germans and Greeks (R. Cerkauskiene, 2020, personal communication; D. Galesic Ljubanovic, 2020, personal communication).[11–13,20] Analysis of the

Figure 2 | (continued) haplotypes were distinguished (5 main haplotypes and 4 subhaplotypes denoted with additional numbers in their names). H1, the most frequent haplotype, was considered ancestral. Fragments of the haplotypes, which underwent recombination, are indicated by shaded bold-lined boxes. Haplotypes of 5 more samples (marked with * following sample identifier) included in this panel could not be fully solved because of missing family data; this is indicated by double alleles at the ambiguous positions. Nevertheless, their recombinational decay compared with the ancestral H1 was clearly seen. Bottom panel: The phase of the full-lengths haplotypes in the remaining 16 samples (multiple heterozygote females with no paternal data, and 1 Greek male for whom full data were not available) could not be solved, and their relation to H1 could not be determined; these haplotypes were not used in the inference of the ancestral haplotype's age. In both panels, the 4-marker core haplotype AACC, encompassing ~0.79 Mb or ~0.75 cM, is indicated by a double-lined box. This haplotype was invariable in 35 p.Gly624Asp unequivocally phased chromosomes, and in the remaining ones, it could be inferred by applying the maximum parsimony rule (using the smallest number of steps to explain the observed state); positions where the phase of the core haplotype was not experimentally confirmed are indicated by single lowercase letters. Positions with the unsolved phase are indicated by double lowercase letters. F, female; F,M, female patients with haplotype solving supported by genotyping their fathers; M, male; Msat, a compound microsatellite with varying number of ACT and AAT repeats (overlapping *rs5962851*; alleles are described as a number of repeats); NA, data not available.

recombination-related decay of the background haplotype shared by Polish and Greek XLAS males suggests the Middle Ages for the time of p.Gly624Asp origin. The mutation most likely occurred between 12th and 13th century, and the variant was spread over several generations through demographic events affecting Central and East Europe. The relatively moderate clinical course of the disease, which by and large does not affect reproductive life, enabled the wide dissemination of this pathogenic variant.

On the basis of the reference Genome Aggregation Database, the present day frequency of p.Gly624Asp allele in the general European non-Finnish population is 0.0002. Taking into account the unequal contribution of the X chromosomes in both sexes and the data contents of the Genome Aggregation Database, the frequency of subjects having p.Gly624Asp allele (either heterozygous or hemizygous) in the general population is 1 in 3500 individuals, largely exceeding the reported XLAS frequency of 1 in 5000 to 10,000.[21] This mostly reflects contribution of heterozygous females, which is ~4 times higher than that of hemizygous males (1 in 2100 vs. 1 in 7900). Given the mild symptoms observed in p.Gly624Asp female subjects, one may speculate that p.Gly624Asp heterozygotes reported in the general population databases represent as yet undiagnosed XLAS cases with nonpersistent mild hematuria.

The study enabled us to identify and delineate the phenotype of the largest homogeneous group of XLAS patients, predominantly of Slavic ethnicity, who harbor a single identical genetic cause of the disease. Despite the strikingly benign disease course of this unique cohort in comparison to previous reports of X-linked AS, male subjects are at risk for late onset ESRD and may be candidates for both existing and emerging therapies for XLAS. Furthermore, any observed variability in the relatively uniform clinical course of the disease within the cluster may facilitate identification of modifier genes and/or predictors of disease progression.

METHODS

Study design

The research design was a multicenter open cohort study. The national character of the study guaranteed the appropriate size of the study group, which is representative of the Polish population. Bioethical Committee approval was obtained from Medical University Gdańsk (Gdańsk, Poland). Collection of clinical data and biological samples for DNA studies was performed in the years 2017 to 2019. Patients were recruited from 13 to 14 Polish pediatric nephrology centers covering the whole geographical area of the country. Studies were conducted following written consent to participate.

Inclusion criteria. Inclusion criteria are as follows: Isolated familial hematuria; isolated hematuria or hematuria and proteinuria with positive family history of kidney failure and hearing loss; chronic kidney disease with a history of previous hematuria and a positive family history of kidney disease; other clinical suspicion of AS.

Exclusion criteria. Exclusion criteria include the following: glomerular hematuria due to immunological disease; lack of written consent to participate in the study or to undergo genetic testing.

Family history, clinical data, and (if performed) kidney biopsy results were entered into a secure anonymized online database. Subjects were given an individual code for the scientific research, thus ensuring that data are anonymous. Biochemical blood and urine tests and other diagnostic procedures (i.e., audiological and ophthalmological evaluation, echocardiography, abdominal ultrasonography, 24-hour ambulatory blood pressure monitoring, and others) were performed in agreement with currently employed clinical standards of procedure, according to a set protocol. Proteinuria was defined as urine protein-to-creatinine ratio >0.2 mg/mg or >0.1 g/m^2 per 24 hours; albuminuria was defined as urine albumin-to-creatinine ratio >30 mg/g or >30 mg/24 h.

Hematuria, defined as presence of ≥5 red blood cells/high-power field on urinalysis, was stratified for severity according to number of red blood cells/high-power field, presence of intermittent or persistent hematuria, and presence of gross hematuria episodes; estimated glomerular filtration rate was estimated using full age spectrum equation[22,23]; and hypertension was defined according to American Academy of Pediatrics (2017).[24]

Profiles of patients enrolled into the database were updated on an annual basis with the above-mentioned results. Samples with genetic material were anonymized in compliance with article 29 of the Polish Medical Code of Ethics and personal data protection legislation.

DNA variant detection and annotation

Mutational screening of the genes encoding collagen IV, *COL4A3*, *COL4A4*, and *COL4A5* (all exons and adjacent intronic sequences), was performed using targeted gene panel SureMASTR Alport (Agilent, Santa Clara, CA) and run on MiSeq Illumina System (San Diego, CA). Alignment was made by Burrows-Wheeler Aligner. Downstream processing was performed with the Genome Analysis Toolkit, SAMtools, and Picard, following documented best practices (http://www.broadinstitute.org/gatk/guide/topic?name=bestpractices). Variant calls were made by both the Genome Analysis Toolkit Unified Genotyper and SAM-tools. The 2 calling techniques were merged using Genome Analysis Toolkit combined variant. All the annotation processes were based on the Ensembl database (version 74), Single-Nucleotide Polymorphism Database (version 135), Exome Variant Server (EVS; ESP6500SI-V2), 1000 Genomes Project (release date: May 21, 2011), and local single-nucleotide polymorphism database. The results were analyzed using Integrative Genomics Viewer, Illumina Variant Studio, and the Imagine Institute/Paris Descartes University PolyWeb online interface.

For detection of large deletions/duplications, the SALSA MLPA Probemixes P191 Alport-mix 1 and P192 Alport-mix 2 (MRC Holland, Amsterdam, the Netherlands) were used, according to manufacturer protocol.

Pathogenicity of the detected variants was assessed against previously reported mutations in Leiden Open Variation Database, ClinVar Database, ARUP ALPORT Database, and Human Gene Mutation Database; for novel variants, comprehensive *in silico* evaluation along with assessment of the intrafamily segregation of the variant with the phenotype were performed. Eventually, the variants were classified according to American College of Medical Genetics and Genomics (2015) criteria and the expert consensus guidelines for genetic diagnosis of AS.[25]

Statistical analyses

Frequencies were compared using χ^2 tests with continuity corrections or Fisher exact test with Freeman-Halton extension for 2 × 3

tables when applicable. For continuous variables, differences between groups were evaluated using Kruskal-Wallis test for global and Mann-Whitney *U* test for pairwise comparisons. Renal survival rates, presented as actuarial risk of ESRD, were calculated using Kaplan-Meier life table analysis. Because of the exploratory character of the analysis, no adjustment for multiple comparisons was done. Statistical analyses were performed using STATISTICA 13.3 (StatSoft, Tulsa, OK) and SAS 9.4 (SAS Institute, Cary, NC) data analysis software systems.

Genetic clock methods

When a disease mutation arises *de novo* or is introduced into a population by a carrier from another population, all descendant individuals initially carry the same disease-associated haplotype. The fraction of the mutation-carrying chromosomes harboring the full-length founder haplotype decreases over time, at a proportion $(1 - r)$ per generation, where r (cM) is the recombination rate over the genetic distance considered. After g generations, the expected proportion of chromosomes with the nonrecombined founder haplotype is as follows: $P = (1 - r)$.[10] At a small r, the age of the mutation origin (g) can be estimated from the following equation: $\ln P = -rg$.

Ten biallelic single-nucleotide polymorphisms with minor allele frequency ≥0.2 and 1 compound microsatellite (overlapping the biallelic *rs5962851*) were selected as haplotype markers, using LD-link database (ldlink.nci.nih.gov). The length of the analyzed segment encompassing *COL4A5* locus was 2.79 Mb (2789 kb), corresponding to the genetic distance of ∼3.32 cM.[8]

Analysis of the background haplotype surrounding p.Gly624Asp mutation

The full-length p.Gly624Asp haplotype was absent from 27 control (p.Gly624Asp-negative) Polish chromosomes examined. The 4-marker core haplotype surrounding p.Gly624Asp mutation (∼0.75 Mb; A-A-C-C), was invariable in 35 phased p.Gly624Asp chromosomes, and in the remaining ones (whose phase was not unambiguously solved) could be inferred applying the maximum parsimony rule. This short core haplotype was found in only 1 among 103 control Polish chromosomes (data not shown), which is consistent with its European frequency reported in the LD-link database, where it was reported in only 16 of 743 chromosomes. This indicates that the strong LD observed in the region surrounding p.Gly624Asp mutation is associated with the mutation rather than with the genomic region of interest.

DISCLOSURE

All the authors declared no competing interests.

ACKNOWLEDGMENTS

The coauthors wish to dedicate the article to our late colleague, Elżbieta Kuźma Mroczkowska, who passed away on February 11, 2020. This work has been financed by the Polish National Science Center grants 2017/25/N/NZ5/00466 and 2017/25/B/NZ2/00519. AMZ and BSL-Z are members of the European Reference Network for Rare Kidney Diseases (ERKNet). ERKNet is co-funded by the European Union within the framework of the Third Health Program "ERN 2016 Framework Partnership Agreement 2017–2021." Next-generation sequencing data processing was performed by Cecile Fourrage from Bioinformatic Platform, Paris Descartes Sorbonne Paris Cité University, Imagine Institute, Paris, France.

SUPPLEMENTARY MATERIAL

Supplementary File (PDF)

Figure S1. Actuarial risk of kidney failure (end-stage renal disease [ESRD]) in 87 male subjects with X-linked Alport syndrome (XLAS) due to missense pathogenic variants of the triple helical domain of collagen IV (46 due to p.Gly624Asp and 41 due to other Xaa-Yaa-Gly variants). Kaplan-Meyer curve of probability of kidney failure (ESRD) with age, including number of subjects at risk and 95% Hall-Wellner bands. Males with p.Gly624Asp pathogenic variant in blue; XLAS males with other Xaa-Yaa-Gly *COL4A5* pathogenic variants in red.

Table S1. Phenotypic characteristics in childhood (age, <18 years) of subjects carrying p.Gly624Asp and other missense Xaa-Yaa-Gly *COL4A5* pathogenic variants. Statistical analysis is performed separately for the subgroups: boys, girls, and total. Intermittent hematuria is defined as episodic nonhematuria urinalysis in a hematuric subject. Proteinuria is defined as urine protein-to-creatinine ratio >0.2 mg/mg or >0.1 g/m^2 per 24 hours. Albuminuria is defined as urine albumin-to-creatinine ratio >30 mg/g or urine albumin >30 mg/24 h. Estimated glomerular filtration rate (eGFR) is estimated using full age spectrum (FAS) equation. Hypertension is defined according to American Academy of Pediatrics (2017). Hearing loss is diagnosed by audiogram. Family history of hematuria, end-stage renal disease (ESRD), and hearing loss is counted for index cases. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$ (Fisher exact probability test).

REFERENCES

1. Nozu K, Nakanishi K, Abe Y, et al. A review of clinical characteristics and genetic backgrounds in Alport syndrome. *Clin Exp Nephrol*. 2019;23:158–168.
2. Kashtan CE, Ding J, Garosi G, et al. Alport syndrome: a unified classification of genetic disorders of collagen IV alpha345: a position paper of the Alport Syndrome Classification Working Group. *Kidney Int*. 2018;93:1045–1051.
3. Savige J, Ars E, Cotton RGH, et al. DNA variant databases improve test accuracy and phenotype prediction in Alport syndrome. *Pediatr Nephrol*. 2014;29:971–977.
4. Morinière V, Dahan K, Hilbert P, et al. Improving mutation screening in familial hematuric nephropathies through next generation sequencing. *J Am Soc Nephrol*. 2014;25:2740–2751.
5. Bekheirnia MR, Reed B, Gregory MC, et al. Genotype–phenotype correlation in X-linked Alport syndrome. *J Am Soc Nephrol*. 2010;21:876–883.
6. Savige J, Storey H, Il Cheong H, et al. X-linked and autosomal recessive Alport syndrome: pathogenic variant features and further genotype-phenotype correlations. *PLoS One*. 2016;11:e0161802.
7. Gross O, Tönshoff B, Weber LT, et al. 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*. 2020;97:1275–1286.
8. Kong A, Gudbjartsson DF, Sainz J, et al. A high-resolution recombination map of the human genome. *Nat Genet*. 2002;31:241–247.
9. Trautmann A, Lipska-Zi tkiewicz BS, Schaefer F. Exploring the clinical and genetic spectrum of steroid resistant nephrotic syndrome: the PodoNet Registry. *Front Pediatr*. 2018;6:1–15.
10. Yotova V, Labuda D, Zietkiewicz E, et al. Anatomy of a founder effect: myotonic dystrophy in Northeastern Quebec. *Hum Genet*. 2005;117:177–187.
11. Pierides A, Voskarides K, Kkolou M, et al. X-linked, COL4A5 hypomorphic Alport mutations such as G624D and P628L may only exhibit thin basement membrane nephropathy with microhematuria and late onset kidney failure. *Hippokratia*. 2013;17:207–213.
12. Šlajpah M, Gorinšek B, Berginc G, et al. Sixteen novel mutations identified in COL4A3, COL4A4, and COL4A5 genes in Slovenian families with Alport syndrome and benign familial hematuria. *Kidney Int*. 2007;71:1287–1295.
13. Plevová P, Gut J, Janda J. Familial hematuria: a review. *Medicina (Kaunas)*. 2017;53:1–10.

14. Gross O, Licht C, Anders HJ, et al. Early angiotensin-converting enzyme inhibition in Alport syndrome delays renal failure and improves life expectancy. *Kidney Int*. 2012;81:494–501.
15. Torra R, Furlano M. New therapeutic options for Alport syndrome. *Nephrol Dial Transplant*. 2019;34:1272–1279.
16. Macheroux EP, Braunisch MC, Pucci Pegler S, et al. 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*. 2019;7:485.
17. Voskarides K, Papagregoriou G, Hadjipanagi D, et al. COL4A5 and LAMA5 variants co-inherited in familial hematuria: digenic inheritance or genetic modifier effect? *BMC Nephrol*. 2018;19:1–8.
18. Gelse K, Pöschl E, Aigner T. Collagens–structure, function, and biosynthesis. *Adv Drug Del Rev*. 2003;55:1531–1546.
19. Hudson BG, Reeders ST, Tryggvason K. Type IV collagen: structure, gene organization, and role in human diseases: molecular basis of Goodpasture and Alport syndromes and diffuse leiomyomatosis. *J Biol Chem*. 1993;268:26033–26036.
20. Demosthenous P1, Voskarides K, Stylianou K, et al., Hellenic Nephrogenetics Research Consortium. X-linked Alport syndrome in Hellenic families: phenotypic heterogeneity and mutations near interruptions of the collagen domain in COL4A5. *Clin Genet*. 2012;81:240–248.
21. Hertz JM, Thomassen M, Storey H, Flinter F. Clinical utility gene card for: Alport syndrome. *Eur J Hum Genet*. 2012;20:713.
22. Pottel H, Delanaye P, Schaeffner E, et al. Estimating glomerular filtration rate for the full age spectrum from serum creatinine and cystatin C. *Nephrol Dial Transplant*. 2017;32:gfw425.
23. Pottel H, Björk J, Bökenkamp A, et al. Estimating glomerular filtration rate at the transition from pediatric to adult care. *Kidney Int*. 2019;95:1234–1243.
24. Flynn JT, Kaelber DC, Baker-Smith CM, et al. Clinical practice guideline for screening and management of high blood pressure in children and adolescents. *Pediatrics*. 2017;140:1–72.
25. Savige J, Ariani F, Mari F, et al. Expert consensus guidelines for the genetic diagnosis of Alport syndrome. *Pediatr Nephrol*. 2019;34:1175–1189.