BMC Genomics

RESEARCH Open Access

Whole exome sequencing shows novel COL4A3 and COL4A4 variants as causes of Alport syndrome in Rio Grande do Norte, Brazil

Washington Candeia de Araújo[1,2†], Raul Maia Falcão[3†], Raquel Araujo Costa Uchoa[1], Carlos Alexandre Garcia[1], Arthur Quintiliano Bezerra da Silva[4], Kesia Larissa Medeiros Quirino[1], Francisco Paulo Freire-Neto[1], Genilson Pereira Gurgel[2], Paulo Ricardo Porfirio Nascimento[1], Leonardo Capistrano Ferreira[1,2], Priya Duggal[5], Jorge Estefano S. de Souza[3] and Selma M. B. Jeronimo[1,2,6*]

Abstract

Background Alport syndrome is a progressive and hereditary nephropathy characterized by hematuria and proteinuria as well as extra renal manifestations as hearing loss and eye abnormalities. The disease can be expressed as autosomal recessive or autosomal dominant at COL4A3 and COL4A4 *loci*, respectively, or X-linked at the COL4A5 locus. This study investigated two unrelated families with nephropathy from Brazil with the aim to identify the mutations involved with the disease.

Methods Whole Exome Sequencing was performed for 4 people from each pedigree (case, parents and a sibling). DNA sequences were mapped against the human genome (GRCh38/hg38 build) to identify associated mutations.

Results Two novel deleterious variants in COL4A3 and COL4A4 loci on chromosome 2 were identified. The variants were detected in the probands with mutant alleles in the homozygous state, a premature stop codon at position 481 of COL4A3 protein and a frameshift mutation leading to a stop codon at position 786 of COL4A4 protein. For both Alport cases the putative variants were surrounded by broad Runs of Homozygosity as well as genes associated with other hereditary nephropathies. Genotyping for COL4A3 validated the exome findings.

Conclusions Two novel variants were identified in two unrelated families from northeast of Brazil. The two deleterious variants identified are located on ROH´s locus with all variants in a homozygous state.

Keywords Alport syndrome, Collagen, Exome

†Washington Candeia de Araújo and Raul Maia Falcão contributed equally to this work.

*Correspondence:
Selma M. B. Jeronimo
selma.jeronimo@ufrn.br
[1]Institute of Tropical Medicine of Rio Grande do Norte, Federal University of Rio Grande do Norte, Natal, RN, Brazil
[2]Department of Biochemistry, Biosciences Center, Federal University of Rio Grande do Norte, Natal, RN, Brazil
[3]Biome, Bioinformatics Multidisciplinary Environment, Federal University of Rio Grande do Norte, Natal, RN, Brazil
[4]Division of Nephrology, Department of Medicine, Federal University of Rio Grande do Norte, Natal, RN, Brazil
[5]Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA
[6]National Institute of Science and Technology of Tropical Diseases, Natal, Rio Grande do Norte, Brazil

Background

Alport syndrome is a hereditary Mendelian disease that progresses with hematuria and loss of kidney function, often leading to kidney transplantation and it is usually accompanied by sensory neural involvement, resulting in deafness and loss of eyesight [1]. The syndrome is caused by defects in the biogenesis of the glomerular basement membrane due to structural defects in type IV collagen alpha chains [2, 3].

Type IV collagen is encoded by six distinct genes (COL4A1 to COL4A6), which form six alpha chains (α1-α6) that assemble in heterotrimers [4, 5]. These heterotrimers coiled together form protomers, crucial for the structure of the glomerular basement membrane. Mutations in the α3, α4, or α5 chains result in defective type IV collagen networks, impairing the proper structure of the glomerular basement membrane and leading to the clinical manifestations of Alport syndrome [6].

Autosomal recessive Alport syndrome is associated with changes in the α3(IV) and α4(IV) chains (genes: COL4A3 and COL4A4), while X-linked Alport syndrome is related to mutations in the α5(IV) chain (gene: COL4A5). There are several known variants for Alport syndrome, including COL4A3 (MIM: 120070), COL4A4 (MIM: 120131), and X-linked COL4A5 (MIM: 303630). Mutations in α3, α4, and α5 chains affect the biogenesis of type IV collagen, leading to a reduction or absence of these molecules in the glomerular basement membrane, resulting in renal failure [6–8]. Of importance, there is a great phenotypic heterogeneity in subjects that carry one mutant allele within the same nuclear family with progressing with end stage chronical disease [9].

Given the well-documented impact of consanguinity on genetic disorders [10–12], our study focused on identifying potential mutations associated with Alport syndrome in two unrelated nuclear families from a community in northeast Brazil. We performed whole exome sequencing on these families to uncover variants that might explain the clinical presentation of Alport syndrome in these probands. This research addressed the genetic underrepresentation of certain populations, potentially offering new insights into the genetic landscape of Alport syndrome in Brazil and improving diagnostic and therapeutic strategies for affected individuals in similar communities.

Methods

Study participants

Two unrelated families with Alport syndrome from the state of Rio Grande do Norte state, Brazil, were evaluated. Family 1 is from the Western region of the state and had four individuals affected with Alport syndrome within the nuclear family, but there were five additional people in the extended family. Family 2 had one case of Alport syndrome within three siblings, but an additional four cases were identified in the family after reviewing family records and the list of subjects under hemodialysis. Families were recruited after the probands had already manifested loss of kidney function and they were under hemodialysis or transplated (Fig. 1).

DNA extraction

DNA was extracted from 10 ml of anticoagulated whole blood (EDTA) by erythrocyte lysis in 70 µg of $NH_4H_2CO_3$/ml,7.0 mg of NH4Cl/ml followed by lysis of

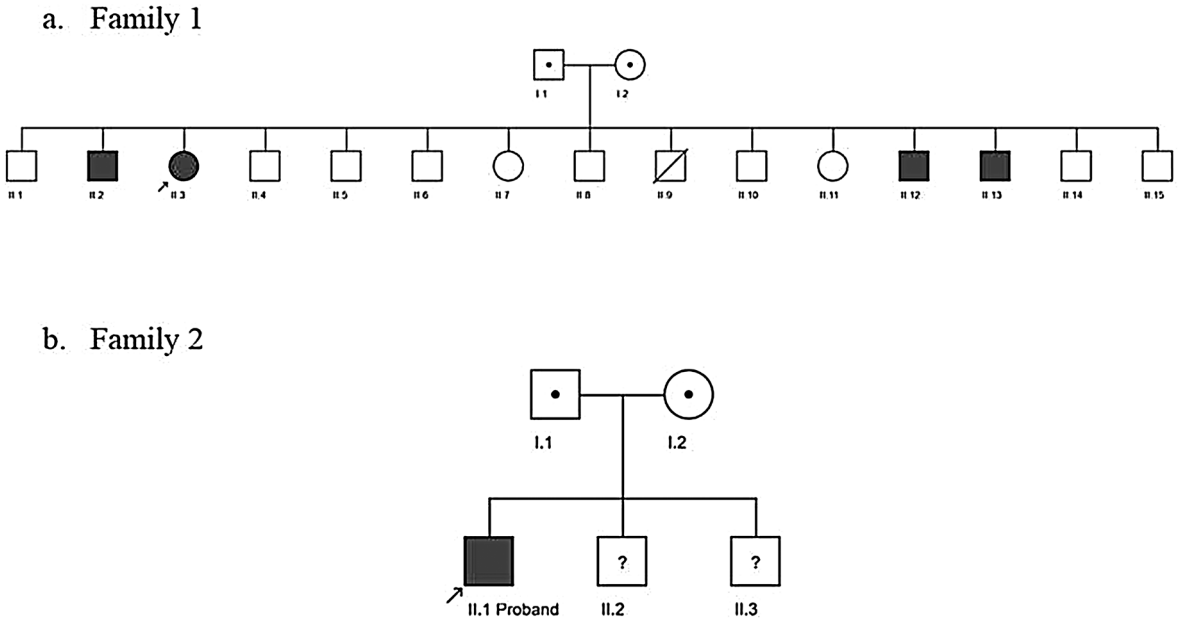

Fig. 1 Pedigree of the studied families. (**a**) Family 1, (**b**) Family 2

leukocytes in 1% sodium dodecyl sulfate, 100 mM EDTA, plus 200 mM Tris (pH 8.5) and precipitation in isopropanol. All DNA samples underwent quality control, which included gel electrophoresis, quantification, and determination of integrity by a fluorometer (qubit, Thermo Fisher, USA.) [13].

Whole exome sequencing

In each family the father, mother, unaffected sibling, and proband ($n = 4$) had their exomes sequenced. Whole exome sequencing (WES) was performed at the University of Iowa Genomics Division using manufacturer recommended protocols. Briefly, 3 µg of genomic DNA were sheared using the Covaris E220 sonicator. The sheared DNA was used to prepare indexed whole exome sequencing libraries using the Agilent SureSelect XT Human all exon v6 + UTR kit (Agilent Technologies, Santa Clara, CA, Cat. No. 5190–8883). The molar concentrations of the indexed libraries were measured using the Fragment Analyzer (Agilent Technologies) and combined equally into pools for sequencing. The pooled libraries were loaded on an Illumina HiSeq 4000 genome sequencer using the 2 × 150 bp paired-end sequencing by synthesis (SBS) chemistry.

Raw reads pre-processing and reference genome mapping

FastQC (v0.11.4) [14] software was used to quantify the sequences and analyze the sequencing quality. Trim Galore (v0.4.1) [15] was used to remove adapters by using the option to execute Cutadapt (v1.8.3) [16], with default values. Reads were mapped using BWA (v0.7.12-r1039) [17] software, with mem mode against to GRCh38/hg38 build [18]. SAMtools (v1.7) [19] were used to calculate total reads mapped and exome means coverage. The means coverage normalization was determined by dividing the average coverage per base by the individual number of reads sequenced then multiplying by the average of the reads sequenced of all individuals. MPileup, of SAMtools, was used to generate input to VarScan2 (v2.3.9) software [20].

Variant calling and data refinement

We used the best practices from Genome Analysis ToolKit (v3.4-46) with HaplotypeCaller [21]. For Base Quality Score Recalibration (BQSR) we used the following known sites: The Single Nucleotide Polymorphism Database (dbSNP 155); Mills & 1000G gold standard indels for hg38 and 1000G phase1 SNPS high confidence sites for hg38. The Vcftools (v0.1.13) [33] were used to filter low coverage variants and maintain records with at least one sample with variant call data. Variants with parameter deep (minDP < 20) and missing values (max-missing > 0.125) were eliminated from the analysis. The Vcfstats package was used to summarize basic variant calling statistics [22].

Variant annotation and screening for deleterious variants

Databases such as Exome Aggregation Consortium [23], 1000 Genomes Project and dbSNP 155 were integrated and variants were annotated by ANNOVAR (v2018Apr16) [24] and SnpEff (v 4.1) [25] software, respectively. To identify deleterious variants (genetic alteration that increases an individual's susceptibility or predisposition to a disease or disorder), we conducted two screening stages. In the first stage, we identified mutations in frameshift status and stop codon gain. In the second stage, we identified missense mutation variants. The two trials of the analysis focused on: (1) identifying homozygous deleterious variants exclusive in proband cases with allelic frequencies < 1% in the Exome Aggregation Consortium database and The Genome Aggregation Database (gnomAD) database; (2) keeping variants with potential pathogenic status in at least two predictors (SIFT, MutationTaster and PolyPhen-2) and remove variants that is classified as benign or likely-benign from ClinVar database; (3) Gene expression (> 10 FPKM) in the kidney, cochlea, and eye tissues samples from Illumina Body Map, Genotype-Tissue Expression (GTex) Project database and protein expression (> 5 TPM) from Human Protein Atlas (HPA) database; (4) genes related to Alport KEGG pathway (https://www.genome.jp/entry/H00581).

Runs of homozygosity (ROH) ratio

ROH was determined by the algorithm H3M2 using a filter of long ROH fragment lengths (1.5 Mb) [26]. The reference length of ROH was determined by applying H3M2 in 20 randomly collected samples of individuals of European ancestry (CEU- Utah Residents, with Northern and Western European ancestry) sequenced by the 1000 Genomes Project [27]. Thus, as a measure of the consanguinity/endogamy, the homozygous ratio was calculated by dividing the total ROH length of each family member by the average ROH length from CEU samples.

Homozygosity mapping and haplotype phasing

To detect differences between the homozygous and heterozygous signals Homozygous Stretch Identifier (HomSI, v2.1) software was used with default parameters values [28]. Missing data inference and haplotype phasing were performed using Beagle (v5.1) software [29].

Sanger segregation analysis.

Follow up study of the nuclear families

The two nuclear families were revisited and clinical history were collected from all members and blood and urine samples were collected allowing laboratory examinations of renal function. Biochemical markers of renal

function (hemogram, hemoglobin, urea, creatinine, microalbuminuria, and urinary analysis) were performed.

Genotyping of COL4A3

DNA extracted from whole blood was genotyped for the putative causal mutation for COL4A3 identified in the exome study. COL4A3 genotype was performed using the TaqMan allelic discrimination assay on the QuantStudio-5 (Life Technologies, Foster City, CA, USA) and genotypes were identified using automated software (SDS 2.3; Life Technologies). Primers and TaqMan-MGB probe sequences for COL4A3 were designed using Custom TaqMan® Assay Design Tool (Thermo Fisher Scientific, Waltham, MA, USA). The general thermal profiling conditions were as follows: 60 °C×30″ (pre-read stage), 95 °C×5′ (hold), [95 °C×15″ 60 °C×1′] ×40 cycles (PCR), 60 °C×30″ (read).

Ethical considerations

The study protocol was reviewed and approved by the Federal University Ethical Committee (CEP-UFRN 50 – 01) and by the Brazilian Ethical Committee (CONEP-4569). All participants and-or legal guardians read, approved, and signed the informed consent. Subjects with medical conditions identified during the study were treated by the study team or referred to a clinic. Two subjects with Alport syndrome had kidney transplants at the time of recruitment. All subjects with Alport syndrome were followed at the University Hospital for their medical care.

Results

Characterization of Alport families

Family 1 had 4 Alport cases out of 15 siblings; of these 2 were kidney transplanted at the time of recruitment, with the parents being their donors (Fig. 1a). Later two other siblings developed Alport and one received a kidney from a sibling and the fourth Alport case received the organ from a cadaver. The first two cases were transplanted 23 years ago and the other two 15 years ago. The Alport case from family 2 (Fig. 1b) received a transplant from his mother, but rejected after 8 years, and he is currently undergoing hemodialysis, and he is on waiting list for new transplant. Figure 2 shows the pedigrees withvariant alleles co-segregating for both nuclear families.

Whole exome sequencing and genome mapping report

We sequenced 8 individuals (3 females and 5 males) from 2 families, each containing one proband. After normalization, the exome target region exhibited 98.6% base

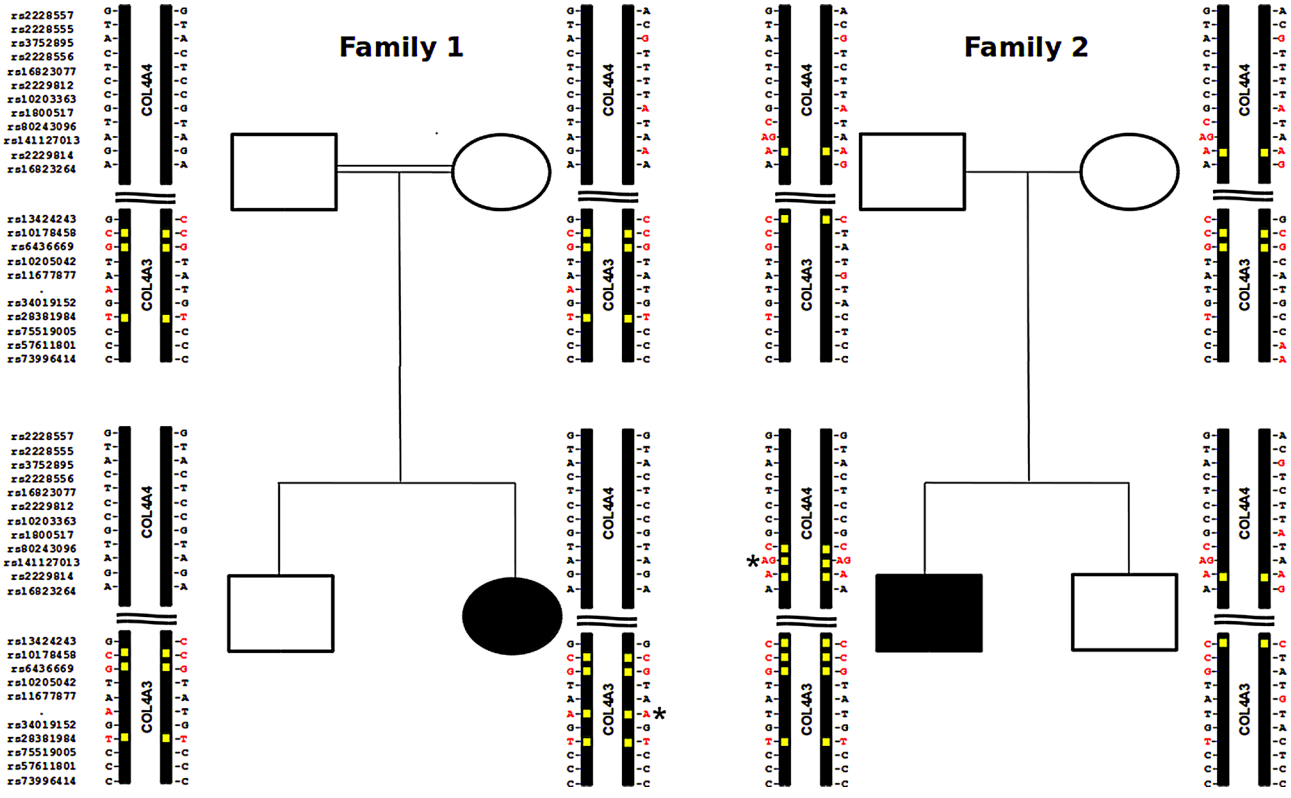

Fig. 2 COL4A4 and COL4A3 variants alleles co-segregating. This figure shows the variants of chromosome 2 for COL4A4 and COL4A3 of all individuals in families, in homozygosity (non-synonymous variant in red, homozygous non-synonymous variant in yellow). The probands are in black. Black "*" defines a stop codon gain in proband of Family 1 and a frame-shift in proband of Family 2

coverage with a minimum depth of 20x and an average depth of ~ 190x, despite variability in read counts among individuals (see Table S1 in the Supplementary Material for a detailed analysis). The total number of mapped reads ranged from 157,896,820 to 204,975,749 in Family 1 and from 153,688,254 to 212,010,435 in Family 2.

Variant calling, data refinement and variant annotation

A total of 130,227 variants were identified and after vcftools filtering for minimum coverage 106,726 variants were kept. From those variants, there were 18,293 synonymous, 16,830 missense, 124 stop codons, 262 frameshifts, and 71,217 in other regions (UTR regions, intronic, intragenic).

Screening for deleterious variants

In the exome screening, a total of 32 variants of high impact were identified (one stop-codon, one frameshift, thirty non synonymous) with the alternative allele in the homozygous state exclusively in the probands (see Table S2 in the Supplementary Material for comprehensive analysis) and allele frequency was less than 1% in the Exome Aggregation Consortium and gnomAD database. Of these, 12 variants remained after filtering for pathogenicity in at least two pathogenicity predictors and ensuring no benign classification in ClinVar. Following expression filtering in kidney, eye, and cochlea tissues, 9 variants remained. Of these, only 2 variants are associated with genes involved in the Alport syndrome pathway.

Those two variants were classified as possibly deleterious and found within COL4A3 (p.Tyr481*, NC_000002.12:g.227267027T > A) and COL4A4 (p.Val741fs, NC_000002.12:g.227059572dup) genes directly linked to the production of collagen type IV and expressed in the kidney, eye, and cochlea. There was no evidence of deleterious variants in the COL4A5 gene. Figure 3 shows distribution of deleterious and homozygous missense variants in COL4A4, COL4A3 identified in this study.

Runs of homozigosity ratio homozygous regions in COL4A genes

Chromosome 2 showed a common ROH locus in both probands (Fig. 4). In Family 1, the proband has a homozygous locus of 20 Mb with 140 genes. In Family 2, the proband has a homozygous locus of 75 Mb with 364 genes. By analyzing this region, we observed that only the COL4A3 and COL4A4 genes had deleterious variants. The total ROH length for all autosomal chromosomes and homozygous ratio for each member of the family (see Table S5 in the Supplementary Material for comprehensive analysis) showed a high level of consanguinity, with a high homozygous ratio in the offspring (parameterized by the CEU population ROH average, see Table S6 in the Supplementary Material for comprehensive analysis).

Homozygosity mapping haplotype phasing and segregation analysis in COL4A genes

Eleven variants in the COL4A3 gene occurred at the triple helix region domain of collagen type IV. We observed one copy of a premature stop codon (p.Tyr481*, NC_000002.12:g.227267027T > A) allele in every member of Family 1, except the proband. The proband was homozygous or had two copies of the deleterious allele. Family 2 did not have this variant.

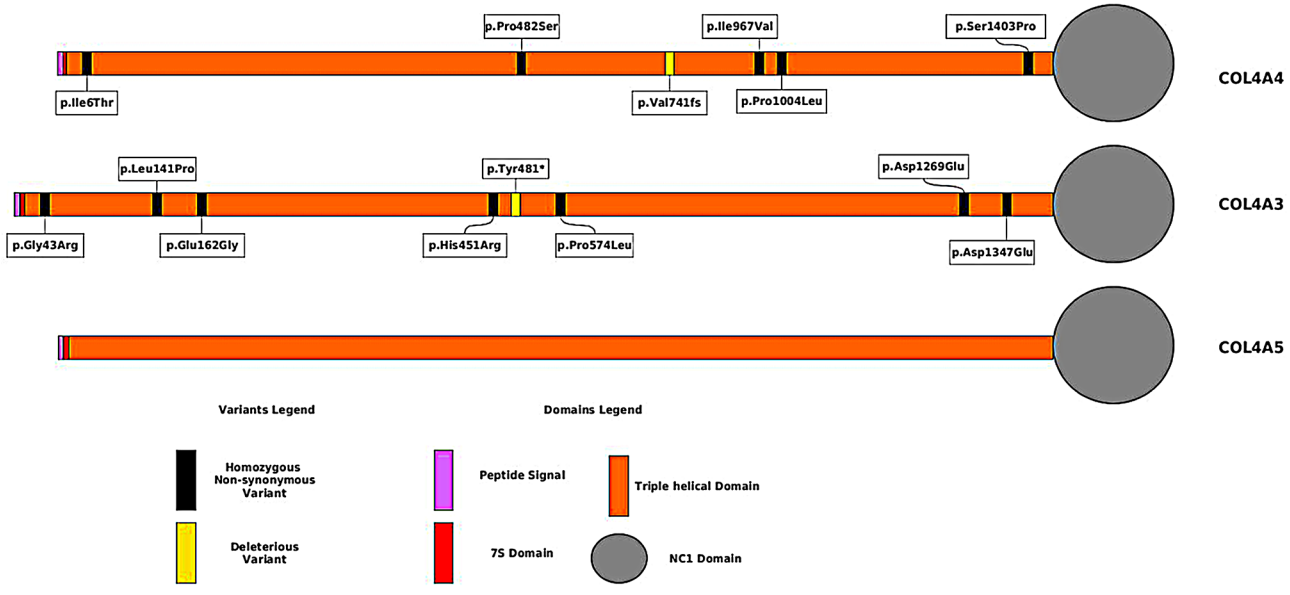

Fig. 3 Collagenous helix domain. Distribution of deleterious and homozygous missense variants in COL4A4, COL4A3 and COL4A5 genes

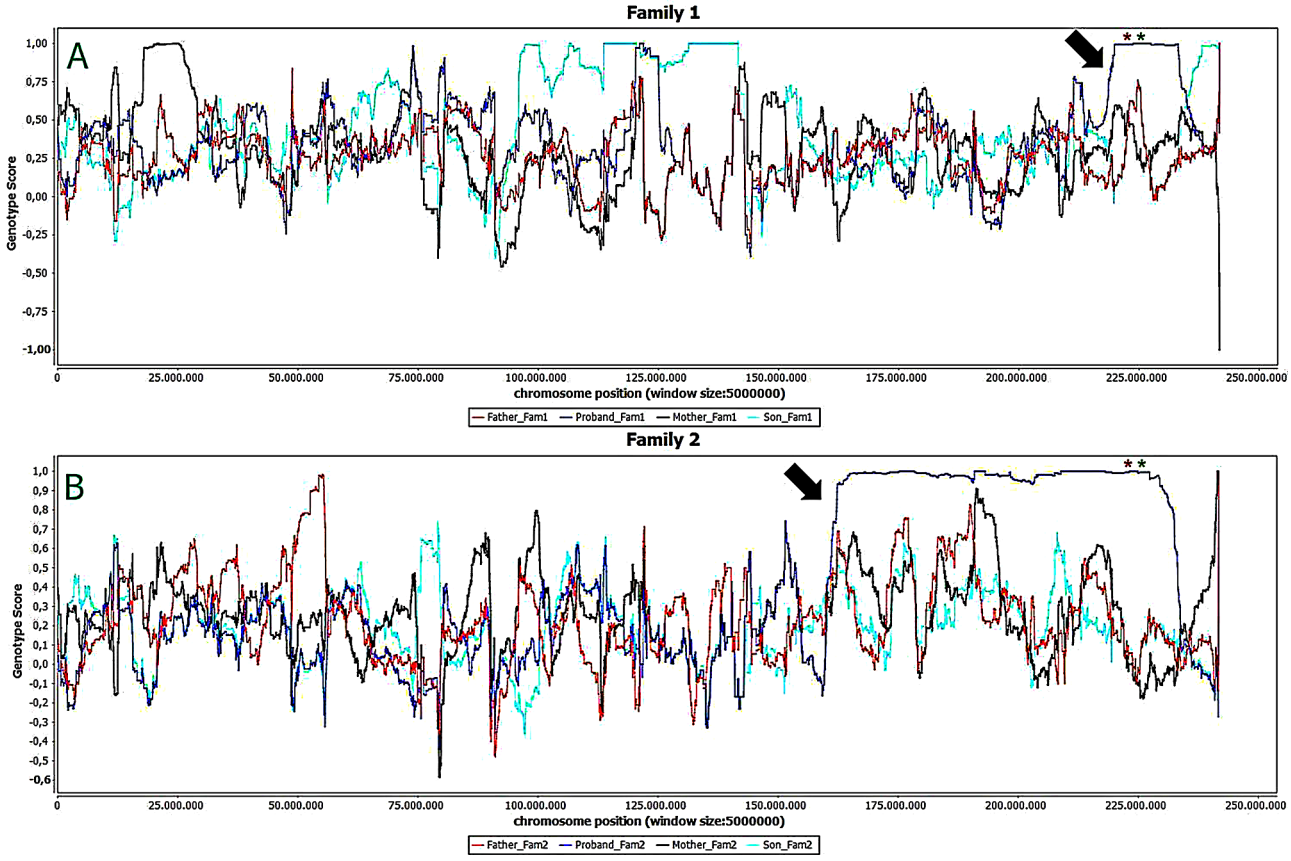

Fig. 4 Pedigree Differences between homozygous and heterozygous signals in complete chromosome 2 extension. (**A**) Family 1, (**B**) Family 2. Black arrow shows where homozygosity region begins in both probands (dark blue line). Red and green * define COL4A4 and COL4A3 genes respectively

Nine variants in the COL4A4 gene occurred at the triple helix region domain of collagen type IV. We observed a frameshift (p.Val741fs, NC_000002.12:g.227059572dup) mutation in all members of Family 2. The proband was homozygous for this frameshift leading to a truncated protein with 786 amino acids. Family 1 did not have this variant.

Also, after haplotype phasing both probands presented an excess of non-synonymous variants at a homozygous state at these loci. For Family 1, the proband had four variants in the COL4A3 gene, and all remaining members had the same three variants in COL4A3 (see Table S3 in the Supplementary Material for comprehensive analysis). Conversely, the Family 2 proband had four variants in the COL4A3 gene and three in COL4A4 loci, while all remaining members had at most three variants distributed in both COL4A genes (Supplementary Table S3). All variants of the genes encoding α4 and α3 chains of collagen type IV perfectly co-segregated with the disease in both families (Fig. 2). The sequencing coverage was 190x (see Table S4 in the Supplementary Material for comprehensive analysis) in these homozygous regions. The missense variant alleles found in COL4A independently showed high allelic frequency rates in the The 1000 Genomes Project (1KGP). However, their co-occurrence with haplotype frequency was rare in all cases. Co-occurrence of the deleterious variants presented in this study did not occur in any 1KGP individual.

Clinical follow up and validation of COL4A3 mutation

A clinical follow up of two families were performed after 20 years since enrolment. The clinical data is presented in Table 1. One of the Alport case from family 1 died of renal failure after 22 years of the transplant; and the other siblings are well (Table 1). Two of the 7 siblings tested were heterozygous for the COL4A3 mutation (Table 1). The examination of family two, mutation COL4A4, showed that one of the heterozygous siblings is evolving with Alport Syndrome. He developed kidney failure (increased creatinine, hematuria and microalbuminuria and anemia).

Discussion

Chronic kidney disease is common in Brazil, with about 30% of cases having unknown etiological causes. In this study, we investigated two families from northeast Brazil with history of renal disease, where some family members were diagnosed with Alport syndrome. We

Table 1 Demographic, clinical and biochemical characteristics of the subjects

Variables	Family 2				Family 1			
Subject id	2_3	2_1	2_2	2_4	1_13	1_1	1_2	1_4
Age								
Sex	Male	Male	Female	Male	Male	Male	Female	Female
Affection status at enrollment	Unaffec.	Unaffec.	Unaffec.	Alport	Unaffec.	Unaffec.	Unaffec.	Alport
Age at onset of symptoms	18	NA	NA	8	NA	NA	NA	NA
Transplant variables								
Hearing	altered	not altered	not altered	altered	not altered	not altered	NA	NA
Eye sight	altered	not altered	not altered	altered	not altered	not altered	NA	NA
Mutated gene	*COL4A4*	*COL4A4*	*COL4A4*	*COL4A4*	*COL4A3*	*COL4A3*	*COL4A3*	*COL4A3*
Zygosity	Heterozygous	Heterozygous	Heterozygous	Homozygous	Heterozygous	Heterozygous	Heterozygous	Homozygous
Laboratory analysis								
Hemoglobin, g/dL	9.6 *	12.1 *	12 *	9.3 *	15.4	14	NA	NA
Hematocrit, %	27.6 *	36.2 *	36.0 *	27.8 *	45.5	42.9	NA	NA
Leukocytes, cells/mm3	6230	6230	8700	4420	6280	4780	NA	NA
Platelets, cells/mm3	350,000	266,000	305,000	270,000	231,000	184,000	NA	NA
Urea, mg/dL	105 *	31	42	90 *	24	36	NA	NA
Creatinine, mg/dL	5.96 *	0.71	1,07 *	9.62 *	1.10	1.00	NA	NA
GFR, mL/min/1.73m2	11	104	58				NA	NA
Microalbuminuria, mg/g	2364.1 *	40 *	236.8 *	NA	13	3	NA	NA
Urinalysis								
General appearance								
Color	yellow	yellow	yellow	NA	yellow	yellow	NA	NA
Aspect	clear	clear	clear	NA	clear	clear	NA	NA
Density	1020	1020	1015	NA	1025	1010	NA	NA
pH	5	6	5	NA	6	6	NA	NA
Chemical analysis								
Urobilonegen	normal	normal	normal	NA	normal	normal	NA	NA
Protein	positive *	negative	positive *	NA	negative	negative	NA	NA
Glucose	positive *	negative	negative	NA	negative	negative	NA	NA
Ketonic bodies	negative	negative	negative	NA	negative	negative	NA	NA
Birrubin	negative	negative	negative	NA	negative	negative	NA	NA
Nitrites	negative	negative	negative	NA	negative	negative	NA	NA
Hemoglobin	positive *	negative	negative	NA	negative	negative	NA	NA

Table 1 (continued)

Variables	Family 2				Family 1			
Microscopic examination								
Epithelial cell casts	presence*	presence*	0	NA	negative	negative	NA	NA
Leuckocytes	2	1	1	NA	2	2	NA	NA
Red cells	5	0	0	NA	1	1	NA	NA
Granular casts	present*	negative	negative	NA	negative	negative	NA	NA
Uric acid	present*	present*	negative	NA	negative	negative	NA	NA

* Out of the reference values

identified two novel variants in the genes COL4A3 and COL4A4, present in broad regions of runs of homozygosity (ROH) in both probands (Fig. 4).

Previous studies indicated that the frequency of consanguinity in the Northeast is 15 times higher (6-9%) than in the Southeast of the country [30, 31]. The analysis of these two consanguineous families revealed homozygous segments characterized by ROHs in chromosomal segments [11, 12, 32]. Genes related to Alport syndrome and other hereditary nephropathies were identified in ROH islands. Szpiech et al. (2013) [33] demonstrated that a significantly larger fraction of predicted harmful homozygous variants is surrounded by ROH stretches compared to non-damaging homozygous variants. Long ROHs (>1.5 Mb) are known to be enriched with deleterious recessive mutations [34, 35]. In our study, the unaffected offspring had the highest value for the homozygous ratio (Table S5), did not have type IV collagen affected by a deleterious variant. A possible explanation for this excessive homozygosity in the individuals would be a large deletion occurring in the offspring leading to uniparental inheritance. However, the segregation analyses, chromosome phasing, and average coverage in ROH regions (Table S3 and Table S4) pointed to an allele inheritance and not to a deletion inheritance.

The elevated homozygosity in some individuals in these families can be attributed to the consanguinity present in their lineages. Family 1 had offspring resulting from a consanguineous marriage between the parents, while the second family's parents showed a third-degree genetic relationship (data not shown). This relationship can be observed through parental alleles in the exome or the inheritance pattern of alleles to their offspring.

Both families have documented cases of consanguinity in the lineages of the progenitors, grandparents, and great-grandparents, confirming higher ROH rates in the sequenced individuals compared to individuals from the 1000 Genomes Project (Tables S5 and Table S6).

Upon analysing the data and segregation of genetic variants in our study cases, we highlight several relevant points. Although ROH presents an increased risk, it does not appear to be the determining factor for Alport syndrome, as many healthy individuals in these families had elevated ROH indices (Table S5). The variants in the genes COL4A3 (p.Tyr481*) and COL4A4 (p.Val741fs) also appear in heterozygosity in the genotypes of healthy individuals (Fig. 2). The differentiating factor for the probands is the presence of the deleterious variants COL4A3 (p.Tyr481*) and COL4A4 (p.Val741fs) in homozygosity, resulting from the combination of deleterious variants with ROH.

To assess if other mutations could be segregating with the observed phenotype in individuals, we used a rigorous two-step global screening process to identify deleterious variants associated with Alport syndrome. Initially, we focused on frameshift mutations and stop codon gains, followed by the identification of missense mutations.

The targeted analysis sought homozygous deleterious variants exclusive to the probands, with allele frequencies below 1% in the Exome Aggregation Consortium (ExAC) and gene expression in renal, cochlear, and ocular tissues, based on data from the Illumina Body Map and the Genotype-Tissue Expression Project. Initially, we identified 32 high-impact functional variants (Table S2), but ten were discarded due the filters of biomolecular/clinical relevance described in material and methods. The remaining two variants, in the genes COL4A3 (p.Tyr481*) and COL4A4 (p.Val741fs), are directly linked to type IV collagen production and are expressed in tissues relevant to Alport syndrome. The absence of deleterious variants in the COL4A5 (Fig. 3) gene reinforces the importance of the findings in COL4A3 and COL4A4.

This analysis enhances our understanding of the genetic basis of Alport syndrome, highlighting the importance of considering allele frequency and tissue-specific gene expression in genetic screenings.

Another important observation is the excess co-occurrence of homozygous variants in the COL4A3 and COL4A4 genes in the probands. The proband from

Family 1 has 4 co-occurrences, while the proband from Family 2 has 7 (Fig. 2). Although most of these mutations are not deleterious, we did not find individuals in the 1000 Genomes Project with a similar pattern of high co-occurrence rate for these genes, demonstrating the rarity of this event.

All sequenced members of Family 1 were heterozygous for the nonsense variant (p.Tyr481*) in the COL4A3 gene (Fig. 2 and Table S3), while only the proband was homozygous for this variant. The COL4A3 nonsense phenotype is severe and associated with end-stage renal disease, deafness, and ocular lesions [7]. The observed stop codon falls in the collagenous domain of type IV collagen (Fig. 3), leading to a truncated protein and destabilizing the formation of the collagen IV network. Other members of Family 2 were heterozygous for the frameshift variant (p.Val741fs) in the COL4A4 gene, but only the proband had the variant allele in the homozygous state (Fig. 2 and Table S3).

Alport syndrome is a rare disease, and the identification of new specific genetic variants significantly contributes to the overall understanding of this condition and the development of personalized therapeutic strategies. The two Brazilian families mentioned in the study represent under-represented populations in genetic research, broadening the diversity considered and allowing for a more comprehensive understanding of the disease. The recent follow up clinical examination showed that one of the heterozygous (COL4A4) has kidney failure (Table 1). Subjects who are heterozygous for COL4A4 mutation have a risk of developing kidney failure and their early diagnosis of Alport syndrome will aid the initiation of treatment with angiotensin II antagonists [9]. Furthermore, the discovery of new genetic variants in regions like Brazil, where representation in genetic databases is historically low, has significant practical implications for genetic counselling, enabling more precise clinical management and better treatment decisions. Therefore, these findings represent substantial contributions to rare disease research, the inclusion of neglected populations, and the improvement of genetic counselling in under-represented regions.

Conclusions

Two novel variants potentially responsible for Alport cases in two unrelated families from northeast of Brazil. The two deleterious variants identified are located on ROH´s locus with all variants in a homozygous state. In addition, it is essential to identify people heterozygous for variants associated with Alport syndrome in areas with a high level of consanguinity allowing early genetic counselling. Of importance, late onset of Alport syndrome in heterozygous subject was identified and therefore early intervention may aid delaying the loss of kidney.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-025-11466-4.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

Supplementary Material 4

Supplementary Material 5

Supplementary Material 6

Acknowledgements

We thank Mr Wellington Luiz de Souza from the Municipal Health Secretariat of São Miguel for his help in the recruitment of the participants.

Author contributions

WWCA. Analyzed the data, wrote the first draft and reviewed the final version of the manuscript. RMF, Analyzed the data, reviewed and approved the final version of the manuscript. RACU. Recruited the subjects, assessed their clinical status and reviewed and approved the final version of the manuscript. CAG. Assessed the clinical status of the subjects and reviewed and approved the final version of the manuscript. AQBS. Recruited participants, reviewed and approved the final version of the manuscript. KLMQ. Recruited participants, assessed their clinical status, reviewed and approved the final version of the manuscript. FPFN, Ran the genotyping, reviewed and approved the final version of the manuscript. GPG, ran the clinical laboratory exams, reviewed and approved the final version of the manuscript.PRN. Recruited participants, Ran the genotyping, reviewed and approved the final version of the manuscript.LCF, Analyzed the data, reviewed and approved the final version of the manuscript. PD. Analyzed the data, reviewed and approved the final version of the manuscript. JESS. Supervised the data analysis, wrote the first draft, reviewed and approved the final version of the manuscript. SMBJ. Designed the study, Recruited the participants, wrote the first draft, obtained the funding, reviewed and approved the final version of the manuscript.

Funding

WCA received a fellowship from CAPES/FAPERN and RMD to CAPES and Rede de Pesquisa em Genômica Populacional Humana (RedePGH) supported by Universidade Federal do Pará (UFPA). The Supercomputer of UFRN (23001011020P6 Project) was used for the analysis. This work was supported by a grant from CNPq awarded to SMBJ.

Data availability

Sequence data that support the findings of this study have been deposited in Bioproject code PRJNA1226620 and it is opened accessed.

Declarations

Ethics approval and consent to participate

The protocol was reviewed and approved by the Federal University Ethical Committee. All participants or their legal guardian signed an informed consent.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 27 November 2023 / Accepted: 10 March 2025
Published online: 02 April 2025

References

1. Alport AC. HEREDITARY FAMILIAL CONGENITAL HAEMORRHAGIC NEPHRITIS. 1927.
2. Rinschen MM, Benzing T, Limbutara K, Pisitkun T. Proteomic analysis of the kidney filtration barrier-Problems and perspectives. Proteom - Clin Appl. 2015;9:1053–68.
3. Stokman MF, Renkema KY, Giles RH, Schaefer F, Knoers NVAM, Van Eerde AM. The expanding phenotypic spectra of kidney diseases: insights from genetic studies. Nat Rev Nephrol. 2016;12:472–83.
4. Pedchenko V, Kitching AR, Hudson BG. Goodpasture's autoimmune disease — A collagen IV disorder. Matrix Biol. 2018;71–72:240–9.
5. Fidler AL, Darris CE, Chetyrkin SV, Pedchenko VK, Boudko SP, Brown KL et al. Collagen Iv and basement membrane at the evolutionary dawn of metazoan tissues. Elife. 2017;6.
6. Nozu K, Nakanishi K, Abe Y, Udagawa T, Okada S, Okamoto T, et al. A review of clinical characteristics and genetic backgrounds in Alport syndrome. Clin Exp Nephrol. 2019;23:158–68.
7. Cosgrove D, Liu S, Collagen. IV diseases: A focus on the glomerular basement membrane in Alport syndrome. Matrix Biol. 2017;57–58:45–54. https://doi.org/10.1016/j.matbio.2016.08.005.
8. Kruegel J, Rubel D, Gross O. Alport syndrome - Insights from basic and clinical research. Nat Rev Nephrol. 2013;9:170–8.
9. Kashtan CE. Alport syndrome: achieving early diagnosis and treatment. Am J Kidney Dis. 2021;77:272–9. https://doi.org/10.1053/j.ajkd.2020.03.026.
10. Bittles AH, Black ML. Consanguinity, human evolution, and complex diseases. Proc Natl Acad Sci. 2010;107:1779–86.
11. McQuillan R, Leutenegger AL, Abdel-Rahman R, Franklin CS, Pericic M, Barac-Lauc L, et al. Runs of homozygosity in European populations. Am J Hum Genet. 2008;83:359–72.
12. Kirin M, McQuillan R, Franklin CS, Campbell H, Mckeigue PM, Wilson JF. Genomic runs of homozygosity record population history and consanguinity. PLoS ONE. 2010;5.
13. Grimberg J, Nawoschik S, Belluscio L, McKee R, Turck A, Eisenberg A. A simple and efficient non-organic procedure for the isolation of genomic DNA from blood. Nucleic Acids Res. 1989;17:8390.
14. Andrews S, Krueger F, Seconds-Pichon A, Biggins F, Wingett S. FastQC. A quality control tool for high throughput sequence data. Babraham Bioinf Babraham Inst. 2015;1:1.
15. Krueger F. A wrapper tool around cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files, with some extra functionality for MspI-digested RRBS-type (Reduced representation Bisulfite-Seq) libraries. 2023. https://github.com/FelixKrueger/TrimGalore
16. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet. 2011;17(1):10–2.
17. Li H. Aligning sequence Reads, clone sequences and assembly contigs with BWA-MEM. ArXiv Preprint arXiv. 2013;00:3.
18. Rosenbloom KR, Armstrong J, Barber GP, Casper J, Clawson H, Diekhans M, et al. The UCSC genome browser database: 2015 update. Nucleic Acids Res. 2015;43:D670–81.
19. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and samtools. Bioinformatics. 2009;25:2078–9.
20. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res. 2012;22:568–76.
21. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A et al. From fastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. 2013.
22. Garrison E, Yandell M, Shapiro M, Marth G, Durbin REE. VCFLIB: an ensemble of methods for variant manipulation and 338 population genetics. https://github.com/vcflib/vcflib
23. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016;536:285–91.
24. Wang K, Li M, Hakonarson H, ANNOVAR. Functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38.
25. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin). 2012;6:80–92.
26. Magi A, Tattini L, Palombo F, Benelli M, Gialluisi A, Giusti B, et al. H3M2: detection of runs of homozygosity from whole-exome sequencing data. Bioinformatics. 2014;30:2852–9.
27. Auton A, Abecasis GR, Altshuler DM, Durbin RM, Bentley DR, Chakravarti A, et al. A global reference for human genetic variation. Nature. 2015;526:68–74.
28. Görmez Z, Bakir-Gungor B, Sağıroğlu MŞ. HomSI: A homozygous stretch identifier from next-generation sequencing data. Bioinformatics. 2014;30:445–7.
29. Browning SR, Browning BL. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. Am J Hum Genet. 2007;81:1084–97.
30. de Farias AA, Nunes K, Lemes RB, Moura R, Fernandes GR, Melo US et al. Origin and age of the causative mutations in KLC2, IMPA1, MED25 and WNT7A unravelled through Brazilian admixed populations. Sci Rep. 2018;8.
31. Santos S, Da Silva Pequeno AA, Pessoa A, Galvão CRC, De Medeiros JLA, Mathias W, et al. Increased prevalence of inherited neuromuscular disorders due to endogamy in Northeast Brazil: the need of community genetics services. J Community Genet. 2014;5:199–203.
32. Nothnagel M, Lu TT, Kayser M, Krawczak M. Genomic and geographic distribution of Snpdefined runs of homozygosity in Europeans. Hum Mol Genet. 2010;19:2927–35.
33. Szpiech ZA, Xu J, Pemberton TJ, Peng W, Zöllner S, Rosenberg NA, et al. Long runs of homozygosity are enriched for deleterious variation. Am J Hum Genet. 2013;93(1):90–102.
34. Pemberton TJ, Szpiech ZA. Relationship between deleterious variation, genomic autozygosity, and disease risk: insights from the 1000 genomes project. Am J Hum Genet. 2018;102:658–75.
35. Sund KL, Rehder CW. Detection and reporting of homozygosity associated with consanguinity in the clinical laboratory. Hum Hered. 2014;77:217–24.

Publisher's note