

Population Genetics Analysis of *SLC3A1* and *SLC7A9* Revealed the Etiology of Cystine Stone May Be More Than What Our Current Genetic Knowledge Can Explain

Chen-Han Wilfred Wu*,¹ Jad Badreddine,¹ Joshua Chang,¹ Yu-Ren Mike Huang,¹ Fernando J. Kim,² Trevor Wild,² Anne Tsai,³ Naomi Meeks,⁴ Rodrigo Donalisio Da Silva,² Wilson R. Molina,⁵ Fredrick R. Schumacher⁶

¹*Department of Genetics and Genome Sciences, Department of Urology, Case Western Reserve University School of Medicine and University Hospitals, Cleveland, OH*

²*Division of Urology, Department of Surgery, University of Colorado School of Medicine, Aurora, CO*

³*Section of Genetics, Department of Pediatrics, University of Oklahoma Health Sciences Center, Oklahoma Children's Hospital, Oklahoma City, OK*

⁴*Division of Clinical Genetics and Metabolism, Department of Pediatrics, Children's Hospital Colorado, University of Colorado, Aurora, CO*

⁵*Department of Urology, University of Kansas School of Medicine, Kansas City, KS*

⁶*Department of Population and Quantitative Health Sciences, Case Western Reserve University School of Medicine, Cleveland, OH*

*Correspondence: wilfred.wu@case.edu

Abstract:

Background: Cystine stone is a Mendelian genetic disease caused by *SLC3A1* or *SLC7A9*. In this study, we aimed to estimate the genetic prevalence of cystine stones and compare it with the clinical prevalence to better understand the disease etiology.

Methods: We analyzed genetic variants in the general population using the 1000 Genomes project and the Human Gene Mutation Database to extract all *SLC3A1* and *SLC7A9* pathogenic variants. All variants procured from both databases were intersected. Pathogenic allele frequency, carrier rate, and affected rate were calculated and estimated based on Hardy-Weinberg equilibrium.

Results: We found that 9 unique *SLC3A1* pathogenic variants were carried by 26 people and 5 unique *SLC7A9* pathogenic variants were carried by 12 people, all of whom were heterozygote carriers. No homozygote, compound heterozygote, or double heterozygote was identified in the 1000 Genome database. Based on the Hardy-Weinberg equilibrium, the calculated genetic prevalence of cystine stone disease is 1 in 30,585.

Conclusion: The clinical prevalence of cystine stone has been previously reported as 1 in 7,000, a notably higher figure than the genetic prevalence of 1 in 30,585 calculated in this study. This suggests that the etiology of cystine stone is more complex than what our current genetic knowledge can explain. Possible factors that may contribute to this difference include novel causal genes, undiscovered pathogenic variants, alternative inheritance models, founder effects, epigenetic modifications, environmental factors, or other modifying factors. Further investigation is needed to fully understand the etiology of cystine stone.

Introduction:

Cystine stones are the product of a metabolic defect, cystinuria, characterized by the excessive excretion of cystine in the urine [1]. This metabolic disorder is a rare heritable condition caused by a defective amino acid transport system that prevents the reabsorption of cystine and other dibasic amino acids in the kidneys and intestinal epithelium [2]. Given their low solubility at physiologic pH, continuous accumulation of cystine in the urine promotes the formation of kidney stones and increases the risk of renal damage [3]. Cystinuria is responsible for approximately 6-8% of renal calculi in pediatrics and accounts for 1-2% of all kidney stone cases. Patients with cystine stones are usually identified at a young age as the vast majority present with their first stone episode during adolescence or early adulthood [4, 5].

The prevalence of cystinuria is shown to be highly variable among different countries ranging from 1 in 2,000 in Great Britain to 1 in 100,000 in Sweden [1, 6]. Efforts to uncover the genetic etiologies of cystine stones have identified two causal genes that are involved in the pathophysiology of the disease: *SLC3A1*, which is located on chromosome 2p16, and *SLC7A9*, which is located on chromosome 19q13 [7, 8]. These genes were found to encode for the two subunit proteins that make up the amino acid transporter [9].

Cystinuria was traditionally classified into three types based on the urinary excretion patterns of obligate heterozygotes. However, this phenotypic classification was shown to be inaccurate as it poorly correlated with the levels of aminoaciduria seen in heterozygote individuals carrying the same pathogenic variant.[8, 10]. Following the identification of *SLC3A1* and *SLC7A9* genes, a new genotypic classification system was introduced dividing cystinuria into several subtypes: type A if pathogenic variants are found in both *SLC3A1* alleles, type B if pathogenic variants are found in both *SLC7A9* alleles, and putative type AB if mutations exist in both genes [11].

Genotyping patients with cystine stones is not routinely performed as it currently has no implications for their treatment and clinical course. Reports investigating the genetic epidemiology of cystine stones are relatively scarce in the literature [12]. Our study aims to determine the genetic prevalence of cystine stones caused by type A and type B cystinuria through analyzing genetic information in population databases. We believe that understanding the frequency and distribution of these genetic variants will enable future efforts to develop a more personalized approach to diagnose and manage cystine stones.

Materials and Methods:

SLC3A1 and *SLC7A9* variants in the 1000 Genomes (1KG) database were identified (Table 1 and Supplementary Tables 1 & 2). Related individuals were excluded. The 1000 Genomes database is a comprehensive database of genomes obtained from healthy individuals in multiple populations around the world, and it identifies 98% of alleles whose frequency is more than 1% in these populations [13]. Lower frequency alleles are also identified in these populations. Known *SLC3A1* and *SLC7A9* pathogenic variants associated with cystine stones were procured from the Human Gene Mutation Database (HGMD) (HGMD® Professional, <http://www.hgmd.cf.ac.uk/ac/index.php>, accessed in 2016). Variants included missense variants, nonsense variants, insertions, and deletions (Table 2). Variants present in both databases were compared and intersected, resulting in 14 disease-causing variants present in both the HGMD and the 1KG database (Table 3). The references associated with these variants were manually reviewed to ensure that they were correctly associated with cystine stone disease. The presence of homozygotes, compound heterozygotes, multiple variants in cis or trans, double homozygotes, and double heterozygotes in the dataset was examined (Figure 1).

Allele frequencies were then calculated from the 1KG database. By summing allele frequencies over all pathogenic variants, we obtained the cumulative frequency of pathogenic variants in *SLC3A1* and *SLC7A9*. The expected carrier and affected rates were estimated using Hardy–Weinberg equilibrium $p^2 + 2pq + q^2 = 1$, where q is the frequency of pathogenic variants, Table 4 [14]. The calculated prevalence of cystine stones caused by *SLC3A1* and *SLC7A9* pathogenic variants was compared with the observed prevalence of phenotypes in several populations.

Results

SLC3A1 and *SLC7A9* variants in the HGMD and 1KG database were identified. The 1KG project comprehensively documents common human genetic variants by sequencing the genomes of 2,504 non-related healthy individuals from 26 populations worldwide [15]. Our results are demonstrated in Table 1 and reveal a total of 1,686 unique variants for *SLC3A1* and 1,274 unique variants for *SLC7A9* (Table 1 and Supplemental Tables 1 & 2). The vast majority of unique variants in 1KG were found to be single nucleotide variations (SNV) affecting non-coding sequences in introns (84.7% for *SLC3A1* and 88.2% for *SLC7A9*) (Table 1). Further, nucleotide insertions and deletions were exclusively observed in the intron regions and constitute 1-4 % of the total number of unique variants in both genes (Table 1). Interestingly, no nonsense variants were identified in the 1KG database for both *SLC3A1* and *SLC7A9* genes (Table 1), suggesting that nonsense variants are strictly deleterious and unnatural to have in healthy individuals.

HGMD represents a wide-ranging collection of pathogenic variants that underlie or closely correlate with human diseases [16]. A total of 110 and 85 unique pathogenic variants for *SLC3A1* and *SLC7A9*, respectively, were identified. These include missense variants, nonsense variants, insertions, deletions, and complex substitutions (Table 2). In contrast to 1KG variants, almost all reported disease-causing variants in HGMD were observed in coding regions with very little documentation of variants occurring in intron or untranslated regions (Table 2). Non-synonymous missense variants constituted the great majority of variants (73.6% for *SLC3A1* and 51.7% for *SLC7A9*). On the other hand, synonymous variants were only observed in *SLC7A9* (8.2%). Moreover, nucleotide insertions and deletions were

found to contribute to more frameshift than in-frame variants in both genes (Table 2). Two complex substitutions were exclusively identified in *SLC3A1* in which multiple consecutive nucleotides were significantly altered (Table 2).

After intersecting the variants identified in the HGMD and 1KG database, only 14 unique disease-causing variants were present in both datasets (Figure 1, Table 3). Of the pathogenic variants present in HGMD, only 9 unique *SLC3A1* pathogenic variants were found to be carried by 26 people and 5 unique *SLC7A9* pathogenic variants were found to be carried by 12 people from the 2,504 healthy non-related individuals in 1KG (Figure 1, Table 3). Notably, all identified individuals were found to be heterozygote carriers. There were no homozygotes, compound heterozygotes, multiple pathogenic variants in cis or trans, double homozygotes, or double heterozygotes. Detailed characterization of these 14 unique variants including position, nucleotides and amino acids change, nomenclature, loci, and the number of people carrying each variant are summarized in Table 3. Remarkably, all the intersected variants were missense variants.

Based on these findings, the disease-causing alleles were found to have a frequency of 0.52% for *SLC3A1*, and 0.24% for *SLC7A9* (Table 3, Table 4). The missense variants c.797T>C (p.Phe266Ser) for *SLC3A1* and c.544G>A (p.Ala182Thr) for *SLC7A9* were found to be the most common cystine stone causing variants in 1KG database with frequencies of 0.14% and 0.1%, respectively (Table 3). Using the Hardy Weinberg Equilibrium ($p^2 + 2pq + q^2 = 1$), where q is the frequency of disease-causing variant, the expected carrier rates for cystine stones caused by *SLC3A1* and *SLC7A9* pathogenic variants were 1 in 96 and 1 in 209, respectively (Table 4). The combined expected carrier rate for both genes is 1 in 66 (Table 4). Similarly, the expected affected rates (q^2) summarized in table 4 were shown to be 1 in 37,100 for *SLC3A1* pathogenic variants, 1 in 174,167 for *SLC7A9* pathogenic variants, and 1 in 30,585 for the cumulative rate of combining both genes (Table 4).

Discussion:

Our study estimated the prevalence of cystine stones caused by *SLC3A1* and *SLC7A9* pathogenic variants using a genetic approach that intersected *SLC3A1* and *SLC7A9* gene variants from HGMD and 1KG databases. Among 110 *SLC3A1* and 85 *SLC7A9* unique pathogenic variants in the HGMD (Table 2), a total of 14 unique variants (9 unique variants for *SLC3A1* and 5 unique variants for *SLC7A9*), all missense mutations, were found to be present in 1KG database with a cumulative frequency of 0.75% (Table 4).

De Novo Mutations vs. Negative Selection

Only 9 out of 110 unique *SLC3A1* pathogenic variants were found in the 1KG. Similarly, only 5 out of 85 unique *SLC7A9* pathogenic variants were identified. This suggests that *SLC3A1/SLC7A9* pathogenic variants are not commonly carried by the general population. This phenomenon can be explained by either de novo mutations that have not yet risen to high frequencies or a negative selection against people carrying these pathogenic variants. Longitudinal studies evaluating pathogenic variants and their prevalence across time are required to better understand this phenomenon. An increase in the number of pathogenic variants or the clinical prevalence of cystine stones with time would suggest that multiple de novo mutations in *SLC3A1/SLC7A9* have occurred in the general population. This may point towards a possible mutation hotspot in *SLC3A1/SLC7A9* genes or the presence of a triggering environmental mutagen. On the contrary, if the observed clinical prevalence of cystine stones stays the same, or even trends down with time, then this would support the notion of a negative selection occurring in the general population against individuals carrying *SLC3A1/SLC7A9* pathogenic variants. This would require in-depth research to understand the potential disadvantages and detriments conferred by these pathogenic variants on their carriers. Additionally, defects of *SLC3A1/SLC7A9* will cause leaks of other dibasic amino acids through urine, such as lysine, arginine, and ornithine [2]. Although these compounds do not form kidney stones, the clinical significance of the loss may need to be further explored.

Discrepancy between Clinical and Genetic Prevalence

The worldwide prevalence of cystinuria is reported to be 1 in 7,000 individuals, however, variation by geographical location happens to be very significant (e.g. 1 per 2,000 in Spain, 1 per 4,000 in Australia, 1 per 18,000 in Japan) [17, 18]. The clinically observed prevalence in the general population is considerably higher than the genetic approach used in our study which estimated the overall affected rate of cystine stone disease to be 1 in 30,585. This indicates the etiology of cystine stone is far more than what our current genetic knowledge can explain. Possible causes for this discrepancy include undiscovered novel causal genes, pathogenic variants, alternative inheritance models, founder effect, or other modification factors.

Indeed, our current knowledge of cystine stone pathogenesis is far from complete. No pathogenic variants were detected in *SLC3A1* or *SLC7A9* in at least 5-10% of patients with clinically confirmed cystine stones [19, 20]. In other words, it is possible that genes other than *SLC3A1* and *SLC7A9* are implicated in the pathogenesis of cystine stones [21, 22].

Additionally, many identified variants in patients with cystine stones were uncharacterized or of unknown clinical significance as their role in the disease process was unclear [11]. This observation has been corroborated by several reports suggesting that our unawareness of many potential cystine stone-associated variants might account for the discrepancy between the genetic and clinical prevalence [23, 24]). It has been proposed that many of these unrecognized pathogenic variants escape detection as they lie deep in the intronic regions potentially affecting splice sites, 5' promoter regions, and 3' polyadenylation regions without actually altering the peptide sequence [22].

Another potential explanation for the difference between the genetic and clinical prevalence of cystine stone disease is the possibility of an alternate mode of inheritance. Although cystine stone disease is considered an autosomal recessive phenotype, cystinuria (without cystine stones) may exhibit an autosomal dominant pattern of inheritance with incomplete penetrance. These cases have been notably reported in heterozygote carriers of *SLC7A9* pathogenic variants [22]. This is suggested by the higher than normal urinary levels of cystine and late-onset disease observed in obligate heterozygotes [25]. Despite a low risk of nephrolithiasis, the wide variability in clinical presentation and disease course seen in these individuals warrant further research to better understand their risk profile.

The high clinical prevalence of cystinuria observed in some populations may be due to a founder effect, a reduced genetic diversity that happens when a community is founded by a small number of inhabiting ancestors and has not recruited newcomers [26]. This is exceptionally true in Libyan Jewish populations in which the clinical prevalence of cystinuria is as high as 1 in 2,500 [27].

Another feature of cystine stone disease that has been highlighted by many studies is the wide variability of phenotypes seen in individuals carrying the same genetic variant. This has been demonstrated by the different disease severities, frequency of stone episodes, and age of onset, even in families carrying the same pathogenic variant [23, 28]. For example, in the International Cystinuria Consortium cohort, it was noted that 15-30% of individuals started forming renal calculi before the age of 3. On the other hand, around 6% of the patients with homozygous pathogenic variants in *SLC3A1* and *SLC7A9* were stone-free [29]. This broad variable expressivity suggests that other factors such as epigenetic modification, environmental factors, or modifier genes might contribute to the presentation of the disease [30].

Strengths and Limitations

Our study has several strengths and limitations. Our biggest strength is utilizing a human-curated database, HGMD professional (<https://www.hgmd.cf.ac.uk/ac/index.php>), to assess the pathogenicity of each variant of two genes. Another strength is that we utilized the population database that provides individual-level genome data, 1KG (<https://www.genome.gov/27528684/1000-genomes-project>), to assess the carrier and allele frequencies. Compared to other databases that provide aggregate data, we were able to confirm that there are no homozygotes, no compound heterozygotes, or double heterozygotes.

One limitation we have is that the HGMD professional version is proprietary, and is not available to the general public. In this study, we used the disease-causing variants that were documented in HGMD back in 2016. Therefore, novel pathogenic variants that were discovered after 2016 were not included. However, as discussed, longitudinal studies

across different time points are needed to evaluate potential de novo mutations or negative selection processes. Our findings mark a reference point that future epidemiological studies can use to test our proposed hypotheses behind the observed discrepancy between the genetic and clinical prevalence of cystine stone disease.

Conclusion:

Our study aims to estimate the genetic prevalence of cystine stones disease. Among the 110 and 85 pathogenic variants for *SLC3A1* and *SLC7A9*, respectively, only 14 unique disease-causing variants were carried in the general population in the 1KG database. The observed clinical prevalence of cystine stone disease (1 in 7,000) is much higher than the overall genetic prevalence caused by disease-causing variants in *SLC3A1* and *SLC7A9* (1 in 30,585), indicating the etiology of cystine stone is far more than what our current genetic knowledge can explain. This difference may be explained by novel causal genes, undiscovered pathogenic variants, alternative inheritance models, founder effect, epigenetic modification, environmental factors, modifier genes, or other modification factors. There could also be undiscovered de novo mutations or negative selection processes for *SLC3A1* and *SLC7A9*. Our results shed light on this knowledge gap and call for further research to identify further factors that are implicated in the pathogenesis of cystine stones.

Figure Legends

Figure 1. Flowchart to calculate pathogenic allele frequencies, carrier rates, and affected rates of *SLC3A1* and *SLC7A9* for cystine stone

1KG: 1000 Genomes Project Phase 3

HGMD: Human Gene Mutation Database

light blue background: specific to *SLC3A1*

light green background: specific to *SLC7A9*

Funding Disclosures: None

Acknowledgements: We gratefully acknowledge the invaluable and unwavering support provided by the University of Colorado School of Medicine, Harvard Medical School, and Case Western Reserve University and University Hospitals. Their resources have been instrumental in the advancement of our research, enabling us to pursue discoveries with confidence and rigor.

Competing interests: Dr. Wilson Molina received grants from Boston Scientific and consulting fees from Boston scientific, IPG, Olympus, Johnson&Johnson in the past 36 months.

Ethics approval statement:

Disclosure of potential conflicts of interest:

No funding was received to assist with the preparation of this manuscript. The authors have no competing interests to declare that are relevant to the content of this article.

Research involving Human Participants and/or Animals:

This is an observational study. The data used is de-identified and publicly available. The Case Western Reserve University/ University Hospital institutional review board has confirmed that no ethical approval is required.

Informed consent:

Given the deidentified nature of the data, the Case Western Reserve University/ University Hospital institutional review board determined that this observational study did not constitute human participants research and was thus exempted from review and the need for informed consent, in accordance with 45 CFR §46.

Contributorship Statement:

C.W W: Project development, Data Collection, Manuscript writing

J B: Data collection, Manuscript writing

J C : Data analysis, Manuscript editing

YR H: Visualization

FJ K: Manuscript editing

T W: Manuscript editing

A T: Manuscript editing

N M: Data collection

RD S: Manuscript editing

WR M: Manuscript editing

FR S: Data analysis

References:

- 1 Palacín M, Borsani G, Sebastio G. The molecular bases of cystinuria and lysinuric protein intolerance. *Curr Opin Genet Dev* 2001;**11**:328–35.
- 2 Yahyaoui R, Pérez-Frías J. Amino Acid Transport Defects in Human Inherited Metabolic Disorders. *Int J Mol Sci* 2019;**21**:119.
- 3 Milliner DS, Murphy ME. Urolithiasis in pediatric patients. *Mayo Clin Proc* 1993;**68**:241–8.
- 4 Knoll T, Zöllner A, Wendt-Nordahl G, Michel MS, Alken P. Cystinuria in childhood and adolescence: recommendations for diagnosis, treatment, and follow-up. *Pediatr Nephrol* 2005;**20**:19–24.
- 5 Friedlander JJ, Antonelli JA, Canvasser NE, Morgan MSC, Mollengarden D, Best S, Pearle MS. Do Urinary Cystine Parameters Predict Clinical Stone Activity? *J Urol* 2018;**199**:495–9.
- 6 Wartenfeld R, Golomb E, Katz G, Bale SJ, Goldman B, Pras M, Kastner DL, Pras E. Molecular analysis of cystinuria in Libyan Jews: exclusion of the SLC3A1 gene and mapping of a new locus on 19q. *Am J Hum Genet* 1997;**60**:617–24.
- 7 Pras E, Arber N, Aksentijevich I, Katz G, Schapiro JM, Prosen L, Gruberg L, Harel D, Liberman U, Weissenbach J. Localization of a gene causing cystinuria to chromosome 2p. *Nat Genet* 1994;**6**:415–9.
- 8 Feliubadaló L, Font M, Purroy J, Rousaud F, Estivill X, Nunes V, Golomb E, Centola M, Aksentijevich I, Kreiss Y, Goldman B, Pras M, Kastner DL, Pras E, Gasparini P, Bisceglia L, Beccia E, Gallucci M, de Sanctis L, Ponzzone A, Rizzoni GF, Zelante L, Bassi MT, George AL, Manzoni M, De Grandi A, Riboni M, Endsley JK, Ballabio A, Borsani G, Reig N, Fernández E, Estévez R, Pineda M, Torrents D, Camps M, Lloberas J, Zorzano A, Palacín M, International Cystinuria Consortium. Non-type I cystinuria caused by mutations in SLC7A9, encoding a subunit (bo,+AT) of rBAT. *Nat Genet* 1999;**23**:52–7.
- 9 Andreassen KH, Pedersen KV, Osther SS, Jung HU, Lildal SK, Osther PJS. How should patients with cystine stone disease be evaluated and treated in the twenty-first century? *Urolithiasis* 2016;**44**:65–76.
- 10 The metabolic and molecular bases of inherited disease seventh edition: Edited by C R Scriver, A L Beaudet, W S Sly and D Valle. P 4605. McGraw-Hill, New York. 1995. ISBN 0-07-909826-6 - Vella - 1996 - Biochemical Education - Wiley Online Library. <https://onlinelibrary.wiley.com/doi/abs/10.1016/S0307-4412%2896%2980019-7> (accessed 3 Aug2022).
- 11 Di Perna M, Louizou E, Fischetti L, Dedoussis GVZ, Stanziale P, Michelakakis H, Zelante L, Pras E, Bisceglia L. Twenty-four novel mutations identified in a cohort of 85 patients by direct sequencing of the SLC3A1 and SLC7A9 cystinuria genes. *Genet Test* 2008;**12**:351–5.
- 12 Calonge MJ, Gasparini P, Chillarón J, Chillón M, Gallucci M, Rousaud F, Zelante L, Testar X, Dallapiccola B, Di Silverio F. Cystinuria caused by mutations in rBAT, a gene involved in the transport of cystine. *Nat Genet* 1994;**6**:420–5.
- 13 An integrated map of genetic variation from 1,092 human genomes. *Nature* 2012;**491**:56–65.
- 14 Hardy GH. MENDELIAN PROPORTIONS IN A MIXED POPULATION. *Science* 1908;**28**:49–50.
- 15 Auton A, Abecasis GR, Altshuler DM, Durbin RM, Abecasis GR, Bentley DR, Chakravarti A, Clark AG, Donnelly P, Eichler EE, Flicek P, Gabriel SB, Gibbs RA, Green ED, Hurles ME, Knoppers BM, Korbel JO, Lander ES, Lee C, Lehrach H, Mardis ER, Marth GT, McVean GA, Nickerson DA, Schmidt JP, Sherry ST, Wang J, Wilson RK, Gibbs RA, Boerwinkle E, Doddapaneni H, Han Y, Korchina V, Kovar C, Lee S, Muzny D, Reid JG, Zhu Y, Wang J, Chang Y, Feng Q, Fang X, Guo X, Jian M, Jiang H, Jin X, Lan T, Li G, Li J, Li Y, Liu S, Liu X, Lu Y, Ma X, Tang M, Wang B, Wang G, Wu H, Wu R, Xu X, Yin Y, Zhang D, Zhang W, Zhao J, Zhao M, Zheng X, Lander ES, Altshuler DM, Gabriel SB, Gupta N, Gharani N, Toji LH, Gerry NP, Resch AM, Flicek P, Barker J, Clarke L, Gil L, Hunt SE, Kelman G, Kulesha E, Leinonen R, McLaren WM,

- Radhakrishnan R, Roa A, Smirnov D, Smith RE, Streeter I, Thormann A, Toneva I, Vaughan B, Zheng-Bradley X, Bentley DR, Grocock R, Humphray S, James T, Kingsbury Z, Lehrach H, Sudbrak R, Albrecht MW, Amstislavskiy VS, Borodina TA, Lienhard M, Mertes F, Sultan M, Timmermann B, Yaspo M-L, Mardis ER, Wilson RK, Fulton L, Fulton R, Sherry ST, Ananiev V, Belaia Z, Beloslyudtsev D, Bouk N, Chen C, Church D, Cohen R, Cook C, Garner J, Hefferon T, Kimelman M, Liu C, Lopez J, Meric P, O'Sullivan C, Ostapchuk Y, Phan L, Ponomarov S, Schneider V, Shekhtman E, Sirotkin K, Slotta D, Zhang H, McVean GA, Durbin RM, Balasubramaniam S, Burton J, Danecek P, Keane TM, Kolb-Kokocinski A, McCarthy S, Stalker J, Quail M, Schmidt JP, Davies CJ, Gollub J, Webster T, Wong B, Zhan Y, Auton A, Campbell CL, Kong Y, Marcketta A, Gibbs RA, Yu F, Antunes L, Bainbridge M, Muzny D, Sabo A, Huang Z, Wang J, Coin LJM, Fang L, Guo X, Jin X, Li G, Li Q, Li Y, Li Z, Lin H, Liu B, Luo R, Shao H, Xie Y, Ye C, Yu C, Zhang F, Zheng H, Zhu H, Alkan C, Dal E, Kahveci F, Marth GT, Garrison EP, Kural D, Lee W-P, Fung Leong W, Stromberg M, Ward AN, Wu J, Zhang M, Daly MJ, DePristo MA, Handsaker RE, Altshuler DM, Banks E, Bhatia G, del Angel G, Gabriel SB, Genovese G, Gupta N, Li H, Kashin S, Lander ES, McCarroll SA, Nemes J, Poplin RE, Yoon SC, Lihm J, Makarov V, Clark AG, Gottipati S, Keinan A, Rodriguez-Flores JL, Korbel JO, Rausch T, Fritz MH, Stütz AM, Flicek P, Beal K, Clarke L, Datta A, Herrero J, McLaren WM, Ritchie GRS, Smith RE, Zerbino D, Zheng-Bradley X, Sabeti PC, Shlyakhter I, Schaffner SF, Vitti J, Cooper DN, Ball EV, Stenson PD, Bentley DR, Barnes B, Bauer M, Keira Cheetham R, Cox A, Eberle M, Humphray S, Kahn S, Murray L, Peden J, Shaw R, Kenny EE, Batzer MA, Konkel MK, Walker JA, MacArthur DG, Lek M, Sudbrak R, Amstislavskiy VS, Herwig R, Mardis ER, Ding L, Koboldt DC, Larson D, Ye K, Gravel S, The 1000 Genomes Project Consortium, Corresponding authors, Steering committee, Production group, Baylor College of Medicine, BGI-Shenzhen, Broad Institute of MIT and Harvard, Coriell Institute for Medical Research, European Molecular Biology Laboratory EBI, Illumina, Max Planck Institute for Molecular Genetics, McDonnell Genome Institute at Washington University, US National Institutes of Health, University of Oxford, Wellcome Trust Sanger Institute, Analysis group, Affymetrix, Albert Einstein College of Medicine, Bilkent University, Boston College, Cold Spring Harbor Laboratory, Cornell University, European Molecular Biology Laboratory, Harvard University, Human Gene Mutation Database, Icahn School of Medicine at Mount Sinai, Louisiana State University, Massachusetts General Hospital, McGill University, National Eye Institute N. A global reference for human genetic variation. *Nature* 2015;**526**:68–74.
- 16 Stenson PD, Mort M, Ball EV, Evans K, Hayden M, Heywood S, Hussain M, Phillips AD, Cooper DN. The Human Gene Mutation Database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. *Hum Genet* 2017;**136**:665–77.
- 17 Leslie SW, Sajjad H, Nazzal L. Cystinuria. In: *StatPearls*. Treasure Island (FL): : StatPearls Publishing 2022. <http://www.ncbi.nlm.nih.gov/books/NBK470527/> (accessed 31 Aug2022).
- 18 Chillarón J, Font-Llitjós M, Fort J, Zorzano A, Goldfarb DS, Nunes V, Palacín M. Pathophysiology and treatment of cystinuria. *Nat Rev Nephrol* 2010;**6**:424–34.
- 19 Font-Llitjós M, Jiménez-Vidal M, Bisceglia L, Di Perna M, de Sanctis L, Rousaud F, Zelante L, Palacín M, Nunes V. New insights into cystinuria: 40 new mutations, genotype-phenotype correlation, and digenic inheritance causing partial phenotype. *J Med Genet* 2005;**42**:58–68.
- 20 Rhodes HL, Yarram-Smith L, Rice SJ, Tabakert A, Edwards N, Hartley A, Woodward MN, Smithson SL, Tomson C, Welsh GI, Williams M, Thwaites DT, Sayer JA, Coward RJM. Clinical and Genetic Analysis of Patients with Cystinuria in the United Kingdom. *Clin J Am Soc Nephrol* 2015;**10**:1235–45.
- 21 Gasparini P, Calonge MJ, Bisceglia L, Purroy J, Dianzani I, Notarangelo A, Rousaud F, Gallucci M, Testar X, Ponzzone A, Estivill X, Zorzano A, Palacín M, Nunes V, Zelante L. Molecular Genetics of Cystinuria: Identification of Four New Mutations and Seven Polymorphisms, and Evidence for Genetic Heterogeneity. *Am J Hum Genet* 1995;**57**:781–8.

- 22 Wong KA, Mein R, Wass M, Flinter F, Pardy C, Bultitude M, Thomas K. The genetic diversity of cystinuria in a UK population of patients. *BJU Int* 2015;**116**:109–16.
- 23 Alghamdi M, Alhasan KA, Taha Elawad A, Salim S, Abdelhakim M, Nashabat M, Raina R, Kari J, Alfadhel M. Diversity of Phenotype and Genetic Etiology of 23 Cystinuria Saudi Patients: A Retrospective Study. *Front Pediatr* 2020;**8**:569389.
- 24 Gaildrat P, Lebbah S, Tebani A, Sudrié-Arnaud B, Tostivint I, Bollee G, Tubeuf H, Charles T, Bertholet-Thomas A, Goldenberg A, Barbey F, Martins A, Saugier-veber P, Frébourg T, Knebelmann B, Bekri S. Clinical and molecular characterization of cystinuria in a French cohort: relevance of assessing large-scale rearrangements and splicing variants. *Mol Genet Genomic Med* 2017;**5**:373–89.
- 25 Rosenberg LE, Downing S, Durant JL, Segal S. Cystinuria: biochemical evidence for three genetically distinct diseases. *J Clin Invest* 1966;**45**:365–71.
- 26 Cystinuria - an overview | ScienceDirect Topics. <https://www.sciencedirect.com/topics/medicine-and-dentistry/cystinuria> (accessed 31 Aug2022).
- 27 Edvardsson VO, Goldfarb DS, Lieske JC, Beara-Lasic L, Anglani F, Milliner DS, Palsson R. Hereditary causes of kidney stones and chronic kidney disease. *Pediatr Nephrol* 2013;**28**:1923–42.
- 28 Yuen Y-P, Lam C-W, Lai C-K, Tong S-F, Li P-S, Tam S, Kwan EY-W, Chan S-Y, Tsang W-K, Chan K-Y, Mak W-L, Cheng C-W, Chan Y-W. Heterogeneous mutations in the SLC3A1 and SLC7A9 genes in Chinese patients with cystinuria. *Kidney Int* 2006;**69**:123–8.
- 29 Harnevik L, Fjellstedt E, Molbaek A, Denneberg T, Söderkvist P. Mutation analysis of SLC7A9 in cystinuria patients in Sweden. *Genet Test* 2003;**7**:13–20.
- 30 Sahota A, Tischfield JA, Goldfarb DS, Ward MD, Hu L. Cystinuria: Genetic aspects, mouse models, and a new approach to therapy. *Urolithiasis* 2019;**47**:57–66.

Table 1. Classification of SLC3A1 and SLC7A9 unique variants in the 1KG database

Unique Variants in 1KG			SLC3A1 (type A)	SLC7A9 (type B)	
SNV	Exon	Coding	Missense	102 (6%)	69 (5.4%)
			Nonsense	0	0
			Synonymous	7 (0.4%)	7 (0.5%)
	UTR		47 (2.7%)	11 (0.8%)	
Intron			1429 (84.7%)	1124 (88.2%)	
Deletion	Exon	Coding	Inframe	0	0
			Frameshift	0	0
	UTR		0	0	
Intron			63 (3.7%)	40 (3.1%)	
Insertion	Exon	Coding	Inframe	0	0
			Frameshift	0	0
	UTR		0	0	
Intron			38 (2.2%)	23 (1.8%)	
Complex Substitutions			0	0	
Total number of Variants			1686 (100%)	1274 (100%)	

1KG: 1000 Genomes Project Phase 3; SNV: Single Nucleotide Variant;

UTR: Untranslated Region

Table 2. Classification of SLC3A1 and SLC7A9 unique variants in the HGMD database

Unique Pathogenic Variants in HGMD			SLC3A1 (type A)	SLC7A9 (type B)	
SNV	Exon	Coding	Missense	81 (73.6%)	44 (51.7%)
			Nonsense	10 (9%)	2 (2.3%)
			Synonymous	0	7 (8.2%)
	UTR		0	0	
Intron		0	0		
Deletion	Exon	Coding	Inframe	1 (0.9%)	3 (3.5%)
			Frameshift	11 (10%)	20 (23.5%)
		UTR		0	0
	Intron		0	2 (2.3%)	
Insertion	Exon	Coding	Inframe	0	2 (2.3%)
			Frameshift	5 (4.5%)	4 (4.7%)
		UTR		0	0
	Intron		0	1 (1.1%)	
Complex Substitutions			2 (1.8%)	0	
Total number of Pathogenic Variants			110 (100%)	85 (100%)	

HGMD: Human Gene Mutation Database; SNV: Single Nucleotide Variant;

UTR: Untranslated Region

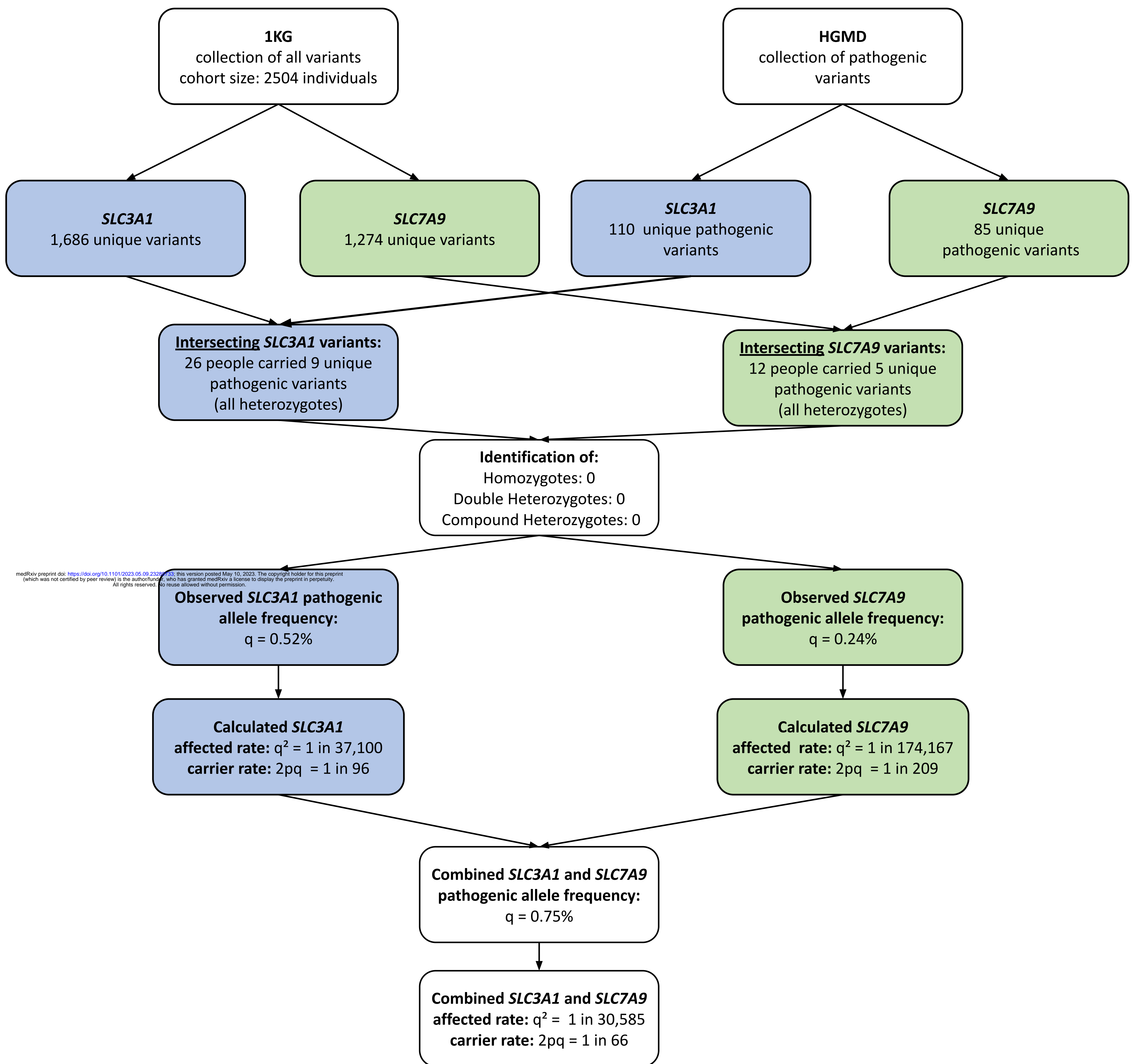
Table 3. Characteristics of intersected genetic variants (pathogenic variants) of SLC3A1 and SLC7A9 in 1KG and HGMD

	Chromosome	Position	Variant Identifier	Reference Allele	Variant Allele	C. nomenclature	P. nomenclature	Exon number	Type	Number of people carrying this variant (all heterogeneous)	Observed Allele Frequency (%)
SLC3A1	chr2	44507990	rs140317484	C	T	c.566C>T	p.Thr189Met	2/10	missense	6	6/5008 (0.12%)
	chr2	44513202	rs141587158	T	C	c.797T>C	p.Phe266Ser	4/10	missense	7	7/5008 (0.14%)
	chr2	44528215	rs121912697	G	A	c.1085G>A	p.Arg362His	6/10	missense	2	2/5008 (0.04%)
	chr2	44528224	rs567478582	G	A	c.1094G>A	p.Arg365Gln	6/10	missense	1	1/5008 (0.02%)
	chr2	44539726	rs187962930	T	C	c.1334T>C	p.Ile445Thr	8/10	missense	2	2/5008 (0.04%)
	chr2	44539746	rs201502095	C	T	c.1354C>T	p.Arg452Trp	8/10	missense	1	1/5008 (0.02%)
	chr2	44539758	rs139251285	C	T	c.1366C>T	p.Arg456Cys	8/10	missense	1	1/5008 (0.02%)
	chr2	44539773	rs144162964	T	C	c.1381T>C	p.Tyr461His	8/10	missense	4	4/5008 (0.08%)
	chr2	44539792	rs121912691	T	C	c.1400T>C	p.Met467Thr	8/10	missense	2	2/5008 (0.04%)
	SLC3A1 accumulation									26	26/5008 (0.52%)
SLC7A9	chr19	33321545	rs146815072	G	A	c.1445C>T	p.Pro482Leu	13/13	missense	1	1/5008 (0.02%)
	chr19	33353409	rs531029519	C	T	c.562G>A	p.Val188Met	5/13	missense	1	1/5008 (0.02%)
	chr19	33353427	rs79389353	C	T	c.544G>A	p.Ala182Thr	5/13	missense	5	5/5008 (0.1%)
	chr19	33355112	rs79987078	G	A	c.368C>T	p.Thr123Met	4/13	missense	2	2/5008 (0.04%)
	chr19	33355167	rs121908480	C	T	c.313G>A	p.Gly105Arg	4/13	missense	3	3/5008 (0.06%)
		SLC7A9 accumulation									12

Chromosome position based on Genome Reference Consortium Human Build 37 (GRCh37, hg19)

Table 4. Observed and Calculated allele frequencies using Hardy-Weinberg equilibrium

	Observed Carrier Frequency (%)	Observed Allele Frequency (%)	Calculated/Expected	
			Carrier Rate	Affected Rate
		q	2pq	q ²
SLC3A1 (type A)	26/2504 (1.03%)	26/5008 (0.52%)	1 in 96	1 in 37,100
SLC7A9 (type B)	12/2504 (0.47%)	12/5008 (0.24%)	1 in 209	1 in 174,167
Combined SLC3A1 + SLC7A9	38/2504 (1.51%)	38/5008 (0.75%)	1 in 66	1 in 30,585
				Clinical prevalance: 1 in 7,000



medRxiv preprint doi: <https://doi.org/10.1101/2023.05.09.23297733>; this version posted May 10, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.