Rare variants in the sodium-dependent phosphate transporter gene SLC34A3 explain missing heritability of urinary stone disease

Omid Sadeghi-Alavijeh, Melanie M.Y. Chan, Shabbir H. Moochhala, Genomics England Research Consortium, Sarah Howles, Daniel P. Gale, and Detlef Böckenhauer

1Department of Renal Medicine, University College London, London, UK; 2Genomics England, Queen Mary University of London, London, UK; and 3Nuffield Department of Surgical Sciences, University of Oxford, Oxford, UK

Urinary stone disease (USD) is a major health burden affecting over 10% of the United Kingdom population. While stone disease is associated with lifestyle, genetic factors also strongly contribute. Common genetic variants at multiple loci from genome-wide association studies account for 5% of the estimated 45% heritability of the disorder. Here, we investigated the extent to which rare genetic variation contributes to the unexplained heritability of USD. Among participants of the United Kingdom 100,000-genome project, 374 unrelated individuals were identified and assigned diagnostic codes indicative of USD. Whole genome gene-based rare variant testing and polygenic risk scoring against a control population of 24,930 ancestry-matched controls was performed. We observed (and replicated in an independent dataset) exome-wide significant enrichment of monallelic rare, predicted damaging variants in the SLC34A3 gene for a sodium-dependent phosphate transporter that were present in 5% cases compared with 1.6% of controls. This gene was previously associated with autosomal recessive disease. The effect on USD risk of having a qualifying SLC34A3 variant was greater than that of a standard deviation increase in polygenic risk derived from GWAS. Addition of the rare qualifying variants in SLC34A3 to a linear model including polygenic score increased the liability-adjusted heritability from 5.1% to 14.2% in the discovery cohort. We conclude that rare variants in SLC34A3 represent an important genetic risk factor for USD, with effect size intermediate between the fully penetrant rare variants linked with Mendelian disorders and common variants associated with USD. Thus, our findings explain some of the heritability unexplained by prior common variant genome-wide association studies.

Kidney International (2023) 104, 975–984; https://doi.org/10.1016/j.kint.2023.06.019

KEYWORDS: genetics; genomics; polygenic risk score; urinary stone disease

Lay Summary

Kidney stones affect around 10% of the population worldwide, causing great suffering to patients and costs to healthcare systems. To improve patient management, an understanding of the causes is important. One’s genetic code is responsible for an estimated 45% of the cause of kidney stones, with environmental factors such as diet and obesity creating the rest of the risk. We looked at the entirety of the genetic code, using the latest computational techniques, in a cohort of people with versus without kidney stones and found that changes in 1 of 2 copies of the gene SLC34A3 lead to a significantly increased risk of developing kidney stones. This work will help identify more people who will go on to develop kidney stones, allowing for better-targeted treatments in the future, as well as providing an explanation for the cause to those already suffering from kidney stones.

Urinary stone disease (USD) is a significant clinical and societal health burden affecting roughly 10% of the population at some point in their lives. The prevalence is increasing, and now, over 80,000 hospital episodes occur per year in the UK. Consequently, the health economic burden is substantial, estimated to be around £250,000,000 in England per year for the initial stone treatment alone. In the US, the annual cost for USD in 2000 was calculated as being almost $3 billion and is estimated to reach $4 billion by 2030. Moreover, a strong association between kidney stones and the development of chronic kidney disease (CKD) further adds to the burden from USD.

The etiology of USD is multifactorial, with genetic and environmental factors implicated. A strong association exists between the affluence of a society and the prevalence of USD, likely reflecting Western lifestyle habits that include a high-salt and high–animal protein intake. Yet, a strong genetic contribution also plays a role; a family history is seen in up to 65% of patients with USD, with the heritability of stone disease estimated to be as high as 45%. Indeed, a strong
family history of kidney stone disease can confer a >50 times increased risk in an individual.11,12 At a polygenic level, multiple genome-wide association studies (GWASs) have been conducted in multi-ancestry populations, with >15 independent loci reported, accounting for roughly 5% of heritability.13 Moreover, the realization that the burden of monogenic causes of USD is considerable is increasing. In 2 recent studies, up to 20% of subjects with USD were considered to have a monogenic cause for their disease, although the rates are highly variable depending on whether a pediatric or adult population is used for analysis, and these studies may be subject to recruitment bias.7

Identification of underlying genetic factors is important, as it facilitates targeted treatment and specific prognostic and genetic counselling.7 The gap between the contribution of the known polygenic risk factors and the observed heritability suggests that important genetic contributors to USD remain to be identified.

The 100,000 Genome Project (100KGP) is a pilot project to assess the utility of whole-genome sequencing (WGS) in rare disease diagnosis in routine healthcare.18 This project’s research arm provides an opportunity to correlate genomic information from participants with their clinical phenotype. We therefore aimed to investigate the contribution to USD of rare genetic variants (which have not been ascertained by previous GWASs) by performing whole-genome gene-based rare-variant studies in participants with human phenotype ontology (HPO) codes for nephrocalcinosis and/or USD, to identify and quantify genetic contributors to the missing heritability of stone disease.

METHODS
We utilized the Genomics England dataset (version 15),19 which contains WGS data, encoded clinical phenotypes using HPO terms,20 and National Health Service (NHS) hospital records, collected for more than 90,259 cancer and rare-disease patients (see Data Statement), as well as their unaffected relatives, to generate the cohorts. Ethical approval for the 100KGP was granted by the Research Ethics Committee for East of England—Cambridge South (REC Ref4/EE/1112). Supplementary Figure S1 details the study workflow, and a full description of the cohort creation can be found in Supplementary Methods S1–S4, Supplementary Table S1, and Supplementary Figure S2. After quality control, relatedness filtering, and ancestry matching, we were left with 374 cases and 24,930 controls for analysis.

Rare-variant gene-based analysis
Single-variant association testing is underpowered when variants are rare, and a collapsing approach, which aggregates variants by gene, can be adopted to boost power. We extracted coding single-nucleotide variants and indels with minor allele frequency <0.01% in the Genome Aggregation Database (gnomAD),21 annotated with one of the following—missense, in-frame insertion, in-frame deletion, start loss, stop gain, frameshift, splice donor, splice acceptor for each gene—and further filtered them by combined annotation-dependent depletion (CADD)version 1.5 score, using a threshold of ≥20, corresponding to the top 1% of all predicted deleterious variants in the genome. Variant quality control and the selection criteria for noncoding gene-based analyses are detailed in Supplementary Methods S5.

We employed the scalable and accurate implementation of generalized mixed model (SAIGE-GENE) (version 0.42.123 to ascertain whether rare coding variation was enriched in cases on a per-gene basis exome-wide, using SKAT-O (full details are given in Supplementary Methods S6). Self-reported sex and the top 10 principal components were included as fixed effects when fitting the null model (full details of workflow are provided in data statement below). A Bonferroni adjusted P value of $2.58 \times 10^{-6}$ (0.05/19,364 genes) was used to determine the exome-wide significance threshold. Binary odds ratios and 95% confidence intervals were calculated for exome-wide significance genes, by extracting the number of cases and controls carrying qualifying variants per gene in the collapsing analysis and applying a Fisher’s test in R (R Foundation). Power calculations can be found in Supplementary Methods S7. Quality control metrics for the rare variant association and cohort can be found in Supplementary Figures S3–S5.

Meta-analysis of the rare-variant association tests
Rare-variant meta-analysis was performed on a per-gene basis, using the Fast Region-Based Association Tests on Summary Statistics (sumFREGAT) package in R.24 The omnibus aggregated Cauchy association test (ACATO)25 was used to combine summary statistics from the AstraZeneca/UK Biobank (UKBB) and 100KGP analyses on a per-gene basis. Files were prepared in advance using the default settings. For details on the AstraZeneca/UKBB cohort analysis, see Supplementary Methods S8.

Polygenic risk scoring
In those cases without a clear genetic diagnosis from the 100KGP clinical pipeline, or a statistically significant gene association from the rare-variant burden analysis, a polygenic risk score (PRS) was applied from a prevalidated, multi-ancestry PRS of USD.26 The PRS was applied to 336 cases and 24,541 controls. Further details on the PRS quality control (Supplementary Figures S6 and S7), scoring, statistical methodology, and use in a logistic model to give liability-adjusted heritability within the 100KGP can be found in Supplementary Methods S9.

Burden heritability regression (BHR) for rare variants
BHR was applied to both the 100KGP and the UKBB datasets, using the recommended default settings altered to match the input settings for the SAIGE-GENE analyses whereby, within each gene, variants were stratified into 2 allele frequency bins (minor allele frequency < $1 \times 10^{-5}$ and $1 \times 10^{-5}$ – $1 \times 10^{-4}$). The model was conditioned on the genome-wide burden model and fixed for effects of SLC34A3 given the SLC34A3 association. Heritability estimates were liability transformed, per the PRS methodology. Full details can be found in Supplementary Methods S10.

RESULTS
Participants
After quality control, genome build, and ancestry matching, we identified 374 unrelated probands with USD (244 recruited to 100KGP under “nephrocalcinosis/nephrolithiasis” as their primary diagnosis, and an additional 130 participants with Hospital Episode Statistics (HES) codes indicating USD) and 24,930 controls recruited to the UK 100KGP. Table 1 details their demographics and clinical characteristics.

Of the 244 primary recruited cases, 26 were previously solved by 100KGP, with the relevant genetic diagnoses being
reported back to the participants, representing a diagnostic yield of 10.7% (see Table 2 for full breakdown). A total of 21 of 244 (8.6%) had a primary diagnosis in keeping with stone-forming disease, and 5 of 244 (2%) had other secondary diagnoses that were delivered to them via the clinical reporting pipeline that did not account for their stone disease. All disease-causing genes followed their established modes of inheritance.

Rare-variant burden analysis of stone disease reveals significant enrichment in **SLC34A3**

Two genes showed statistically significant enrichment of rare and predicted damaging variation in USD cases, compared with controls, as follows: **SLC34A3** (\(P = 2.61 \times 10^{-07}\), odds ratio \(OR = 3.75\), 95% confidence interval [CI] 2.27–5.91); and **OR9K2**, encoding an olfactory receptor (\(P = 2.03 \times 10^{-06}\), \(OR = 8.47\), 95% CI 3.23–18.81; Figure 1). Full results of the association analysis can be found in Supplementary Table S2. No other genes were significantly enriched in the other tested collapsing tags: intronic, 5-untranslated region or 3-untranslated region, synonymous or splice site (see Supplementary Table S2).

**Replication of results in the UKBB**

Association of USD with a rare variation in **SLC34A3**, but not **OR9K2**, was replicated in publicly available analyses of whole-exome sequencing data from 3147 cases and 255,496 controls within the UKBB, an independent dataset (https://azphewas.com/).27 In this analysis, multiple rare-variants collapsing models were applied on a per-gene basis and were analyzed with SKAT-O across all listed UKBB phenotypes. Under the “fl exnonsynmtr” model, which equates to nonsynonymous variants with a minor allele frequency <0.01% in both gnomAD and the UKBB, with missense variants also having to fall within a region constrained for missense variation, USD was most strongly associated with **SLC34A3** (\(P = 3.67 \times 10^{-10}\), \(OR = 2.01\); see Figure 2). None of the cases in UKBB were homozygous for their qualifying **SLC34A3** variants (full list of variants is given in Supplementary Table S3).

**Meta-analysis of the UKBB and Genomics England data**

Meta-analysis of the 2 datasets confirmed a significant association in **SLC34A3** (\(P = 1.94 \times 10^{-18}\)). No other genome-
wide significant associations were detected (full results are given in Supplementary Table S2).

Phenotype/genotype analysis of SLC43A3 cases
Qualifying variants were found in 6% of the ascertained stone population in the study (21 of 374), compared to 1.6% (389 of 24,930) in the controls. Of the 21 cases with qualifying variants, 14 were recruited with stone disease, 1 with congenital anomalies of the kidney or urinary tract, 4 with cystic kidney disease, and 2 with intellectual disability. A total of 19 cases were heterozygous, and 2 were compound heterozygous for qualifying SLC43A3 variants with both cases’ variants being confirmed in trans (Table 2; a full breakdown of the cases is given in Supplementary Table S4). Qualifying variants in the control population were all heterozygous. Excluding the 2 compound heterozygous patients from the analysis and re-running the association led to a smaller, but still significant, association ($P = 1.47 \times 10^{-6}$).

All 21 patients had gone through the Genomics England clinical pipeline, with 4 cases receiving genetic diagnoses, including PKD2-associated cystic kidney disease (2 patients), Kabuki syndrome (KMT2D), and biallelic SLC34A3-associated USD. In the solved biallelic SLC34A3 case, both variants were annotated as being (likely) pathogenic by Clinvar (rs199690076 and rs762710288), whereas in the unsolved biallelic SLC34A3 case, the evidence for a clinical-grade diagnosis was weaker (rs369400414 and rs560440785); although both variants met our inclusion criteria for the collapsing rare-variant association, they did not meet the benchmark for clinically reportable results (see Table 3 for a full breakdown of variant annotation).

For both biallelic SLC34A3 cases, not enough phenotype data were available to ascertain whether they met clinical diagnostic criteria for hereditary hypophosphatemic rickets with hypercalciuria, but both had the hypercalciuria HPO code on record (a full breakdown of HPO codes associated with each case is given in Supplementary Table S4). The top 10 HPO codes associated with the patients with SLC34A3-associated USD are found in Figure 3, with a full list of associated phenotype codes given in Supplementary Table S5.
elevation in USD PRS was seen, compared with the controls ($P = 3.1 \times 10^{-04}$). Initial analysis including a cohort of the cases with qualifying SLC34A3 variants did identify statistically significant differences among the 3 cohorts ($P = 4.4 \times 10^{-04}$), but this signal was driven by the difference between unsolved cases and controls (Figure 4). The difference between the control group and the SLC34A3 cases did not reach statistical significance given the small number of SLC34A3 cases. Adjusted odds of a USD diagnosis increased by a factor of 1.22 (95% CI 1.10–1.36; $P = 0.003$) per SD of PRS in an adjusted model including self-reported sex and the first 10 principal components. The area under the curve (AUC) was 0.62 (95% CI 0.60–0.66).

The relationship between polygenic risk and monoallelic SLC34A3 variants is likely to be independent

In our model, a significant association was seen between phenotype and both PRS ($P = 3.8\times10^{-04}$) and the presence of a monoallelic SLC34A3 variant ($P = 2.72 \times 10^{-08}$). However, no log-additive (multiplicative) interaction occurred between PRS and the SLC34A3 binary with the phenotype ($P = 0.77$), although likely the study was underpowered to detect such an interaction. Of the other covariates, sex ($P = 1.34 \times 10^{-05}$) and the fourth principal component ($P = 4.72 \times 10^{-08}$) also were strongly associated with the phenotype. The presence of an SLC34A3 variant increased the frequency of USD within 100KGP when plotted against polygenic risk score (Figure 5). The addition of the SLC34A3 variant binary to the linear model including PRS led to a significant rise in the estimated variance explained by the model (liability-adjusted pseudo-$R^2$ rising from 5.1% to 14.2%) and a modest increase in the model’s predictive capability (AUC 0.64, 95% CI 0.61–0.66), thereby implying that SLC34A3 had a 9.1% contribution to the heritability model.

BHR analysis confirms the contribution of SLC34A3 in both cohorts

To confirm the heritability of the rare variants, as well as the contribution of SLC34A3, using orthogonal methodology, we applied the BHR tool to both the 100KGP and the UKBB datasets. The liability-adjusted gene-wise burden heritability of rare and ultra-rare predicted loss-of-function and damaging missense variants explained 10.8% (95% CI 7.7%–13.9%) of phenotypic variance within the 100KGP dataset, with variants in SLC34A3 making up 7.6% (95% CI 5.64%–

Table 3 | Qualifying variants making up the SLC34A3 association in the rare-variant burden analysis using SAIGE-GENE

<table>
<thead>
<tr>
<th>Variant type</th>
<th>Case count (%)</th>
<th>Control count (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Missense</td>
<td>17 (75)</td>
<td>346 (89)</td>
</tr>
<tr>
<td>Frameshift</td>
<td>1 (4)</td>
<td>14 (4)</td>
</tr>
<tr>
<td>Splice donor variant</td>
<td>1 (4)</td>
<td>20 (5)</td>
</tr>
<tr>
<td>Start lost</td>
<td>1 (4)</td>
<td>5 (1.1)</td>
</tr>
<tr>
<td>Stop gained</td>
<td>3 (13)</td>
<td>4 (0.9)</td>
</tr>
</tbody>
</table>

SAIGE-GENE, scalable and accurate implementation of generalized mixed model.

---

Figure 2 | Gene-based Manhattan plot of the UK Biobank analysis obtained from the AstraZeneca PhEWAS portal. Each symbol represents variants in a gene that are predicted to be non-synonymous with a minor allele frequency (MAF) < 0.1% in both Genome Aggregation Database (gnomAD) and the UK Biobank. The only exome-wide significant association was SLC34A3 ($P = 3.67 \times 10^{-10}$, odds ratio = 2.01). The multiple symbols under SLC34A3 (the "build-up") represent different analyses with respect to the predicted severity of the included variants. The lower dashed vertical line indicates exome-wide significance, and the upper dashed line indicates exome-wide significance corrected for the ~1500 different phenotypes analyzed. bp, base pairs.
9.6%) of this signal in total (\(\lambda = 1.076\)). In the UKBB analysis, the liability-adjusted gene-wise burden heritability of rare and ultra-rare predicted loss-of-function and damaging missense variants explained 5.4% (95% CI 3.3%–8.4%) of the phenotypic variance, with variants in \(\text{SLC34A3}\) making up 3.7% (95% CI 1.3%–6.1%) of this signal (\(\lambda = 1.018\)).

**DISCUSSION**

We identified rare variants in \(\text{SLC34A3}\) and \(\text{OR9K2}\) as being significantly associated with USD among 100KGP participants. \(\text{SLC34A3}\) encodes the sodium-dependent phosphate transport protein 2C expressed in the proximal tubule (NaPi-IIc). Data on the chemical composition of the stones in participants in the 100KGP are not available, so different (perhaps stronger) associations could have been observable in subgroups defined by stone composition. Although the association with \(\text{SLC34A3}\) was independently replicated in the UKBB, a signal was not observed in this dataset at \(\text{OR9K2}\), so the possibility exists that this finding is a type 1 error, which is well recognized with olfactory receptor genes, owing to their enrichment for loss-of-

**Figure 3** | The top 10 human phenotype ontology (HPO) codes associated with cases with qualifying SLC34A3 variants. Nephrolithiasis makes up the most common associated clinical code, followed by hypercalciuria and nephrocalcinosis.

**Figure 4** | Violin and boxplot comparing polygenic risk-score distribution across cohorts. Violin and boxplot showing the polygenic risk score distributions between controls (those with qualifying SLC34A3 variants removed), cases with qualifying SLC34A3 variants, and unsolved patients who have neither a reportable nor a qualifying variant in SLC34A3. The means of the 3 polygenic risk scores were compared with a Kruskal–Wallis test \((P = 4.04 \times 10^{-04})\) with the signal being driven by the difference between unsolved cases and controls (paired Wilcoxon = 3.6 \(\times\) 10\(^{-04}\)). NS, not significant. ***Statistical significance.
function variation without clinical consequences. However, similar olfactory gene associations have not been observed in other studies using the 100KGP dataset analyzed with similar methodology, and recognition that olfactory receptors regulate transport processes in many organ systems is increasing. —OR9K2 is expressed in the intestine, and conceivably, it could be involved in the regulation of absorption of substrates with relevance to stone formation, such as oxalate or calcium. Therefore, further studies are needed to assess the relevance of OR9K2 in USD.

Our results highlight the importance of both SLC34A3 as a contributor to USD, with more than 5% of patients in this cohort from the 100KGP harboring predicted damaging variants, and independent replication of this association in the UKBB dataset. Important to note is that the odds ratio for stone disease with rare, predicted damaging variants in SLC34A3 was comparable to that of the polygenic risk score (PRS). Further, the association was strongest with models that included predicted damaging missense variation, and it was weakened if the filter was constrained to only those variants predicted to cause protein truncation, suggesting that any predicted damaging variants in this gene can contribute to the risk of USD.

Our findings therefore highlight the importance of monoallelic variants in SLC34A3 for USD. SLC34A3 was reported in 2006 as a recessive disease gene for the rare disorder hereditary hypophosphatemic rickets with hypercalciuria (HHRH). Although the original publication indicated recognition that heterozygous carriers in the affected families were frequently affected by hypercalciuria, it remains listed as a recessive disease gene in the Online Mendelian Inheritance in Man (OMIM) dataset (*609826). Yet, good evidence has been obtained for the impact of monoallelic variants; an investigation in a cohort of affected families showed that the risk of USD was 46%, 16%, and 6% in subjects with biallelic, monoallelic, or no causative variants, respectively. This finding is consistent with a paradigm in which identification of rare, deleterious monoallelic SLC34A3 variants can be regarded as a risk factor for stone disease but is not a diagnostic finding.

The underlying mechanism is thought to be hypophosphatemia-mediated suppression of fibroblast growth factor–23, with consequent activation of the 1-a hydroxylase and increased 1,25 dihydroxy vitamin D levels, which in turn stimulates intestinal calcium absorption. The same mechanism is thought to apply in infantile hypercalcinemia due to biallelic loss-of-function variants in SLC34A1. Although the role of monoallelic SLC34A1 variants in hypercalciuria has been controversial, large genome-wide studies have demonstrated a significant association between both coding and noncoding variants of SLC34A1 and USD, consistent with the concept that a reduction in proximal tubular phosphate transport does increase the risk for kidney stones.

In the AstraZeneca UKBB rare-variant collapsing analysis that used 12 different sets of qualifying variant filters (models: 10 dominant models; 1 recessive model; and 1 synonymous “control” variant model), SLC34A3 was the gene most significantly associated with USD. Further, the association was strongest with models that include predicted damaging missense variation, and it was weakened if the filter was constrained to only those variants predicted to cause protein truncation, suggesting that any predicted damaging variants in this gene can contribute to the risk of USD.

Our study provides evidence of clinical relevance for coding variants in SLC34A3, with a significant enrichment of rare and predicted deleterious variants in USD patients, compared with controls, among participants in both the 100KGP and UKBB datasets. Although the 100KGP did not specifically encourage enrollment of patients with a family history of the respective disorders, a recruitment bias is possible, which could have inflated the percentage of SLC34A3-related disease. Nevertheless, the additional identification of rare SLC34A3 variation as the strongest rare-variant association in UKBB participants provides independent replication and raises the question of whether identification of these risk variants in individual patients would provide utility in clinical practice. Although the modest risk
effect precludes predictive use of such a test, the above
pathophysiological mechanism suggests that phosphate sup-
plementation may be a suitable treatment to stimulate
fibroblast growth factor–23 and thereby suppress 1-a hy-
droxylation of vitamin D in patients at risk of SLC34A3-
related kidney stone disease. Indeed, successful use of this
treatment has been reported. However, clinical trial
data would be needed to support such an intervention,
because large doses of phosphate supplementation carry a risk
of increasing the urinary phosphate concentrations, with
consequent increased risk of calcium phosphate precipitation.
Indeed, nephrocalcinosis has been associated with phosphate
supplementation in patients with PHEX-associated hypo-
phosphatemic rickets, although these patients typically
received enormous doses. Thus, more data are needed before embarking on routine phosphate supplementation in
SLC34A3-associated USD.

Limitations
This study has several notable limitations. First, the study
was underpowered in the discovery cohort to discover novel
gene variants either with a weaker effect on risk or of greater
rarity (see the Supplementary Methods for a more detailed
description), due to our small case number. Those who were
recruited clearly may have had more severe USD, with poten-
tial ascertainment bias toward genes more likely to be
involved with more-severe disease. Following from this, the
addition of SLC34A3 into the logistic model with PRS risks a
“winner’s curse” bias whereby its effect is overstated. This
possibility is supported by the BHR scores being higher in
the 100KGP versus the UKBB cohort, although the CIs do
overlap for the analysis. However, the fact that the associa-
tion between rare and predicted damaging variants in
SLC34A3 and USD is replicated in an independent cohort (as
the “top gene”), and that this signal is enriched on meta-
analysis, is reassuring. The heritability and effect size of the
PRS in the 100KGP are similar to the known common
variant contribution to USD, implying a similar underlying
genetic architecture between cohorts. In terms of pheno-
typing, we are limited by the depth of information available
for our cohort in the 100KGP, which does not include the
biochemical stone properties for most cases. As with all
diseases that can be present silently, the chance exists that
our control population has been misclassi
ed, with a pro-
portion of them having USD. Although all efforts were taken
to remove any potential cause of USD, as well as hospital
codes directly pertaining to USD, we cannot completely
exclude this possibility. Although the statistical evidence of
enrichment of rare genetic variants in SLC34A3 is strong and
reproducible, without functional analysis of each missense
variant, determination of which of the observed variants play
a causal role in disease is not possible.

Conclusion
Our study highlights the substantial contribution of rare and
predicted damaging variants in SLC34A3 to the burden of
USD, helping to close the missing heritability gap and sup-
porting the idea of routine genetic testing in affected patients.

DISCLOSURE
All the authors declared no competing interests.

DATA STATEMENT
Details of the aggregated dataset used for the analysis can be found
at https://re-docs.genomicsengland.co.uk/aggv2/. Genomic and
phenotype data from participants recruited to the 100,000 Genomes
Project can be accessed by application to Genomics England Ltd at
https://www.genomicsengland.co.uk/about-gecip/joining-research-
community/. Details of the AstraZeneca workflow can be found at
https://azphewas.com/about. Details of the GeneBass workflow
work can be found at https://app.genebass.org/about. Code for the case–
control ancestry-matching algorithm can be found at https://github.
com/APLevine/PCA_Matching. Details of the rare-variant workflow
can be found at https://re-docs.genomicsengland.co.uk/avt/. Details
of the common-variant GWAS workflow can be found at https://re-
docs.genomicsengland.co.uk/gwas/.

SUPPLEMENTARY MATERIAL
Supplementary File (Excel)
Supplementary Table S1. List of terms used to create Hospital
Episode Statistics (HES)–derived cohorts; list of codes used to exclude
controls.
Supplementary Table S2. Full summary statistics from the rare-
variant burden analysis of stone cases versus controls using scalable
and accurate implementation of generalized mixed model (SAIGE-
GENE).
Supplementary Table S3. List of variants making up with UK
Biobank collapsing analysis for SLC34A3.
Supplementary Table S4. Full breakdown of the 21 cases that make
up the SLC34A3 association using scalable and accurate
implementation of generalized mixed model (SAIGE-GENE).
Supplementary Table S5. All human phenotype ontology (HPO)
codes associated with the urinary stone disease (USD) cohort (n = 374).
Any code with a frequency less than 10 has been filtered out for
patient identification purposes.
Supplementary File (Word)
Supplementary Methods S1. Cohort creation and rationale.
Supplementary Methods S2. Genomic Variant Call Format
annotation and variant-level quality control.
Supplementary Methods S3. Relatedness estimation and principal components analysis.

Supplementary Methods S4. Ancestry-matching of cases and controls.

Supplementary Methods S5. Rare-variant gene-based testing: variant-level quality control, further SLC34A3 and SLC34A1 testing, noncoding gene-based tags, and genomic inflation.

Supplementary Methods S6. Scalable and accurate implementation of generalized mixed model (SAIGE-GENE) methodology.

Supplementary Methods S7. Power calculation.

Supplementary Methods S8. Validation of rare-variant burden results in the UK Biobank.


Supplementary Figure S1. Study workflow. The flowchart shows the number of samples included at each stage of filtering and the analytical strategies employed. USD, urinary stone disease.

Supplementary Figure S2. Ancestry matching. Principal component analysis showing the first 8 principal components for matched cases (red) and controls (green) and unmatched controls (grey).

Supplementary Figure S3. Rare-variant analysis metrics. Quantile–quantile plot from the rare-variant scalable and accurate implementation of generalized mixed model (SAIGE-GENE) analysis of 374 cases versus 24,930 controls.

Supplementary Figure S4. Common-variant metrics. Quantile–quantile plot of the common-variant genome-wide association study (minor allele frequency [MAF] >0.5%) of the study cohort across 10,409,709 markers. The genomic inflation factor is 1.02, indicating that population stratification is well controlled.

Supplementary Figure S5. UK Biobank analysis metrics. Quantile–quantile plot of the rare-variant scalable and accurate implementation of generalized mixed model (SAIGE) analysis of the urinary stone disease in the AstraZeneca study cohort of 3147 cases and 255,496 controls.

Supplementary Figure S6. Polygenic risk across ancestries. Standardized polygenic risk-score distributions across the case–control cohort by genetically determined ancestry, as ascertained via principal component analysis.

Supplementary Figure S7. Polygenic risk score effect sizes across ancestries. Effect size urinary stone disease polygenic risk score by ancestry, per SD of stone disease.

REFERENCES


