



Derivation and validation of genome-wide polygenic score for urinary tract stone diagnosis

Ishan Paranjpe^{1,2,3,4}, Noah Tsao⁵, Renae Judy⁵, Manish Paranjpe⁶, Kumardeep Chaudhary^{1,3,4}, Derek Klarin^{7,8,9}, Iain Forrest^{1,14}, Ross O'Hagan^{1,3,4}, Arjun Kapoor², John Pfail², Suraj Jaladanki^{1,3,4}, Fayzan Chaudhry⁴, Akhil Vaid^{1,4}, Phan Q. Duy¹⁰, Charles Bronfman Institute of Personalized Medicine Genomics Team¹, Regeneron Genomics Team¹¹, John Cijiang He^{12,13}, Benjamin S. Glicksberg^{4,14}, Steven G. Coca¹², Mantu Gupta², Ron Do^{1,3,14}, Scott M. Damrauer⁵ and Girish N. Nadkarni^{1,3,4,12,13}

¹The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, New York, USA; ²Department of Urology, Icahn School of Medicine at Mount Sinai, New York, New York, USA; ³The BioMe Phenomics Center, Icahn School of Medicine at Mount Sinai, New York, New York, USA; ⁴The Hasso Plattner Institute for Digital Health at Mount Sinai, Icahn School of Medicine at Mount Sinai, New York, New York, USA; ⁵Department of Surgery, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA; ⁶Division of Health Sciences and Technology, Harvard Medical School, Boston, Massachusetts, USA; ⁷Malcolm Randall Veterans Affairs Medical Center, Gainesville, Florida, USA; ⁸Division of Vascular Surgery and Endovascular Therapy, University of Florida College of Medicine, Gainesville, Florida, USA; ⁹Broad Institute of Harvard and Massachusetts Institute of Technology, Cambridge, Massachusetts, USA; ¹⁰Medical Scientist Training Program, Yale University School of Medicine, New Haven, Connecticut, USA; ¹¹Regeneron Genetics Center, Tarrytown, New York, USA; ¹²Division of Nephrology, Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, New York, USA; ¹³Renal Program, James J Peters Veterans Affairs Medical Center at Bronx, New York, New York, USA; and ¹⁴Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, New York, USA

Urinary tract stones have high heritability indicating a strong genetic component. However, genome-wide association studies (GWAS) have uncovered only a few genome wide significant single nucleotide polymorphisms (SNPs). Polygenic risk scores (PRS) sum cumulative effect of many SNPs and shed light on underlying genetic architecture. Using GWAS summary statistics from 361,141 participants in the United Kingdom Biobank, we generated a PRS and determined association with stone diagnosis in 28,877 participants in the Mount Sinai BioMe Biobank. In BioMe (1,071 cases and 27,806 controls), for every standard deviation increase, we observed a significant increment in adjusted odds ratio of a factor of 1.2 (95% confidence interval 1.13-1.26). In comparison, a risk score comprised of GWAS significant SNPs was not significantly associated with diagnosis. After stratifying individuals into low and high-risk categories on clinical risk factors, there was a significant increment in adjusted odds ratio of 1.3 (1.12-1.6) in the low- and 1.2 (1.1-1.2) in the high-risk group for every standard deviation increment in PRS. In a 14,348-participant validation cohort (Penn Medicine Biobank), every standard deviation increment was associated with a significant adjusted odds ratio of 1.1 (1.03 – 1.2). Thus, a genome-wide PRS is associated with urinary tract stones overall and in the absence of known clinical risk factors and illustrates their complex polygenic architecture.

Kidney International (2020) **98**, 1323–1330; <https://doi.org/10.1016/j.kint.2020.04.055>

Correspondence: Girish N. Nadkarni, One Gustave Levy Place, Box 1243, New York, New York 10029, USA. E-mail: girish.nadkarni@mountsinai.org

Received 21 December 2019; revised 30 March 2020; accepted 16 April 2020; published online 12 June 2020

KEYWORDS: genomics; nephrolithiasis; personalized medicine; polygenic risk score; urinary tract stone

Copyright © 2020, International Society of Nephrology. Published by Elsevier Inc. All rights reserved.

Urinary tract stones are highly prevalent in the American population, with up to 10% of men and 7% of women reporting at least 1 stone during their lifetime.¹ Their composition is extremely heterogeneous with 80% of stones being made up of a mixture of calcium oxalate and calcium phosphate and the majority of the remainder composed of uric acid, struvite, and cystine.^{2,3} In addition to causing acute discomfort, stones are also associated with increased risk for infection, hydronephrosis, and chronic kidney disease.^{3–5} Patients are often diagnosed after presenting with symptoms or incidentally on imaging studies. Current kidney risk prediction tools are limited to recurrence of urinary tract stones rather than prediction of stone formers in the general population.^{6,7}

Much progress has been made in understanding the clinical and environmental risk factors predisposing to stone formation. However, these factors do not completely explain the highly heritable nature of urinary tract stones and individuals with no clinical risk factors can still develop urinary tract stones. Prior research has identified several monogenic causes of stones. These monogenic causes are rare, although they are likely underdiagnosed.⁸ Although future work may uncover highly pathogenic variants associated with stone formation, it is likely that many urinary stones are of a polygenic nature, due to the combined effect of many genetic factors, each with a small individual effect size.

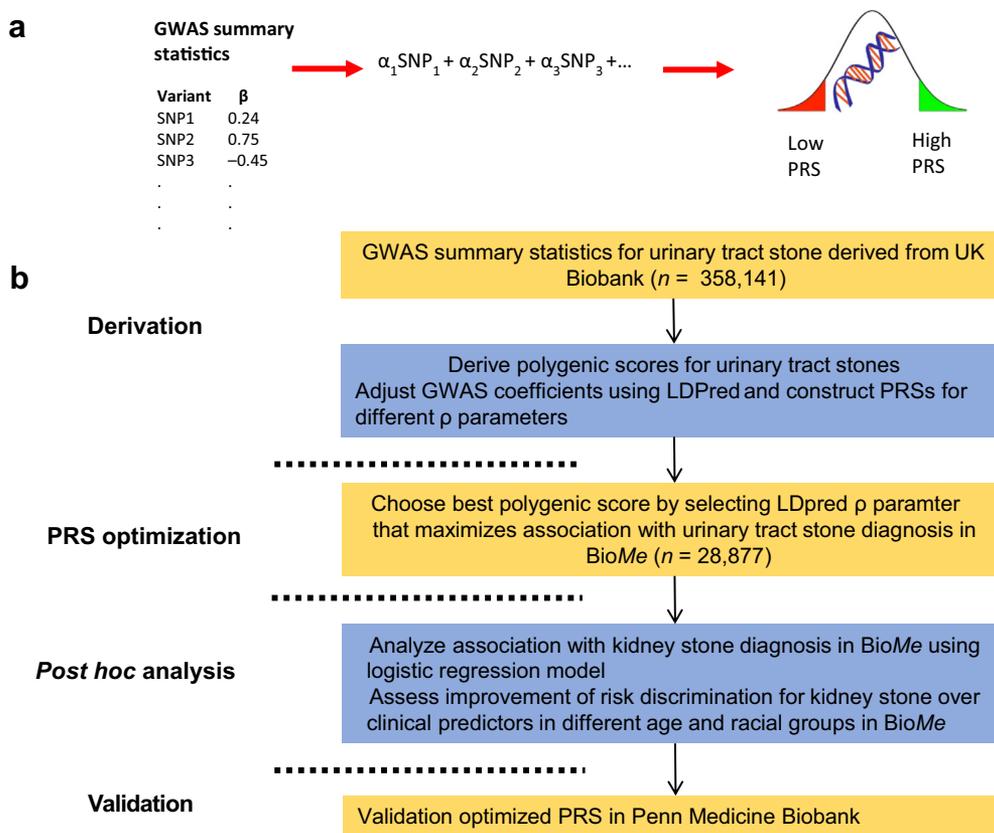


Figure 1 | Flowchart of polygenic risk score (PRS) generation and validation. (a) Generation of PRS from genome-wide area study (GWAS) summary statistics. (b) Overall scheme of PRS development and application in BioMe and Penn Medicine Biobank cohorts. SNP, single nucleotide polymorphism.

Genome-wide association studies (GWAS) have generated summary statistics data for millions of single nucleotide variations from large studies and can be used to derive polygenic risk scores (PRSs). In other heritable diseases such as schizophrenia and coronary artery disease, these risk score distributions have been applied to identify subgroups (i.e., those with a high polygenic burden) that are at similar risk for developing disease as patients with monogenic mutations but are much more common.^{9–11} A patient’s electronic health record (EHR) is the set of all clinical data available including laboratory measurements, imaging, clinical notes, vitals, and diagnostic and procedure codes. Thus the availability of EHR-linked biobanks allow for rapid assessment of PRSs with many comorbid conditions.

We sought to derive a polygenic score for kidney stones from the UK Biobank and validate it in a large multiethnic biobanked cohort (BioMe Biobank cohort). We tested polygenic scores with increasing number of single nucleotide variations. Finally, we assessed whether the best performing polygenic score was associated with urinary tract stones in a subgroup of individuals without clinical risk factors. Finally, we validated this polygenic score in an independent cohort of individuals.

RESULTS

The overall schema of the study is shown in Figure 1.

Study population

In this study, we included participants with genotyping data ($n = 28,877$) in the Mount Sinai BioMe cohort. We identified 1,071 participants (4%) with a diagnosis of urinary tract stones (cases) and 27,806 participants with no previous diagnosis of urinary tract stones (controls) for all case-control analyses. Participants with urinary tract stones were significantly older (60 vs. 57 years) and had higher body mass index (BMI) (29 vs. 28 kg/m²) as compared to controls ($P < 0.001$) (Table 1). In terms of self-reported race, cases had a significantly greater proportion of Hispanic individuals (44% vs. 32%) and fewer African American (17% vs. 24%) and European individuals (30% vs. 33%) than controls did ($P < 0.001$) (Table 1). Cases also had higher proportion of all comorbidities than did controls ($P < 0.001$) (Table 1).

Optimization of PRSs

We used summary statistics from an online repository of GWAS results performed on a subset of the UK Biobank cohort of European ancestry (Supplementary Table S1). Using these summary statistics, we varied the ρ parameter in LDpred¹² and computed PRSs for each ρ (assumed proportion of casual single nucleotide polymorphisms [SNPs]) using an additive model. We then selected the PRS corresponding to ρ , which maximized the association with

Table 1 | Baseline demographic and clinical characteristics of urinary tract stone cases and controls

Characteristics	Case (n = 1071)	Control (n = 27,806)	P value
Male sex	526 (49)	11,825 (43)	<0.001
Age, yr	60 ± 14	57 ± 18	<0.001
Race			<0.001
African American	180 (17)	6670 (24)	
European American	323 (30)	9248 (33)	
Hispanic American	473 (44)	8918 (32)	
Other	95 (9)	2992 (10.8)	
Body mass index, kg/m ²	29 ± 7	28 ± 7	<0.001
Clinical comorbidities			
Hypertension	723 (68)	13,147 (47)	<0.001
Coronary artery disease	345 (31)	5980 (22)	<0.001
Type 2 diabetes	341 (34)	5714 (21)	<0.001
Obesity	415 (39)	8314 (30)	<0.001
Gout	64 (6)	744 (3)	<0.001

Values are n (%) or mean ± SD.

urinary tract stone diagnosis in BioMe. We observed the greatest association with stone diagnosis for $\rho = 0.001$ (Supplementary Figure S1). This score included 7,670,833 single nucleotide variations and was used for all downstream analyses and validation.

Association of PRS with urinary tract stone diagnosis in BioMe

We then computed the association of the PRSs derived from UK Biobank with urinary tract stone status in BioMe. To account for population stratification, we performed our analysis stratified by self-reported race (Figure 2a) and then combined the results with inverse variance-weighted meta-analysis. We observed a significant association of PRSs with urinary tract stone case or control status in a logistic regression model adjusted for 10 genetic principal components (PCs), sex, age, BMI, and traditional risk factors including a history of gout, hypertension, and type 2 diabetes. For every SD increase in PRS, we observed increased odds for stone diagnosis of a factor of 1.2 (95% confidence interval [CI]: 1.1–1.3; $P < 0.001$) (Table 2). The odds of urinary tract stone prevalence increased monotonically as the PRS increased (Figure 2b). The distribution of PRSs in BioMe is provided in Figure 2c.

Comparison of genome-wide polygenic score to score composed of only GWAS-significant SNPs

To assess whether a genome-wide polygenic score performs better than one containing GWAS-significant SNPs, we then compared the association of the optimized PRSs with that of a PRS (PRS_{GWAS}) constructed using the 48 SNPs that met genome-wide significance ($P < 5 \times 10^{-8}$) for association with urinary tract stones in UK Biobank. PRS_{GWAS} was not significantly associated with stone diagnosis in BioMe (odds ratio [OR] = 0.98; 95% CI: 0.92–1.04) adjusted for age, sex, BMI, 10 genetic PCs, and history of gout, hypertension, and type 2 diabetes.

PRS associates with urinary tract stones in absence of traditional risk factors

We then stratified our cohort into a low-risk group without any clinical risk factors during their EHR history. Thus, the low-risk group did not have any diagnoses codes for hypertension, type 2 diabetes, or gout and had an average BMI ≤ 25 kg/m². There were 6,136 BioMe participants in the low-risk group (136 stone cases and 6,000 controls). We then defined a high-risk group that had 1 or more of the risk factors mentioned above. There were 22,741 BioMe participants (935 stone cases and 21,806 controls) in the high-risk group.

In the low-risk group, the adjusted odds of stone diagnosis increased by a factor of 1.3 (95% CI: 1.1–1.6; $P = 0.001$) per SD of PRS. In the high-risk group, we similarly observed an increased adjusted odds of stone diagnosis of 1.2 (95% CI: 1.1–1.3; $P < 0.001$) per SD of PRS (Table 3). We observed a monotonic increase in kidney stone diagnosis with PRS in both the low- and high-risk groups (Figure 3).

Validation of PRS in external cohort

We then further validated the PRS in a cohort from the Penn Medicine Biobank (PMBB). The PMBB cohort included 9,973 individuals of European genetic ancestry and 5,423 individuals of African genetic ancestry (Supplementary Table S2). Because the validation cohort had only a small number of individuals with low clinical risk, we were unable to stratify by clinical risk due to low statistical power. In a logistic regression model adjusted for age, sex, BMI, 10 genetic PCs, gout, hypertension, and type 2 diabetes, we found a significant association between PRS and urinary tract stone diagnosis in European Americans (adjusted OR = 1.16 per SD; 95% CI: 1.06–1.27; $P < 0.001$) but not in African Americans (adjusted OR = 1.02 per SD; 95% CI: 0.89–1.16; $P = 0.07$) (Figure 4a, Table 4). However, in a transethnic meta-analysis, we found a significant association with an adjusted OR of 1.12 (95% CI: 1.05–1.2; $P < 0.001$) for every SD increase in PRS (Table 4). Prevalence of stone diagnosis increased monotonically as a function of PRS in European Americans (Figure 4b) but not in African Americans (Figure 4c).

Improvement in risk discrimination with addition of PRS

We then assessed the improvement in discriminative ability of the PRS in BioMe (because PMBB had only high-risk individuals). The bootstrapped area under the curve ([AUC], 1000 iterations) of a logistic regression model incorporating only age, sex, and clinical covariates was 0.63 (95% CI: 0.62–0.65). After adding PRS, AUC improved to 0.68 (95% CI: 0.6–0.70). This represents a modest, but significant improvement in prediction.

DISCUSSION

Utilizing the combined availability of linked genotypic and clinical diagnosis data, we developed and validated a genome-wide PRS for urinary tract stones in 2 independent cohorts.

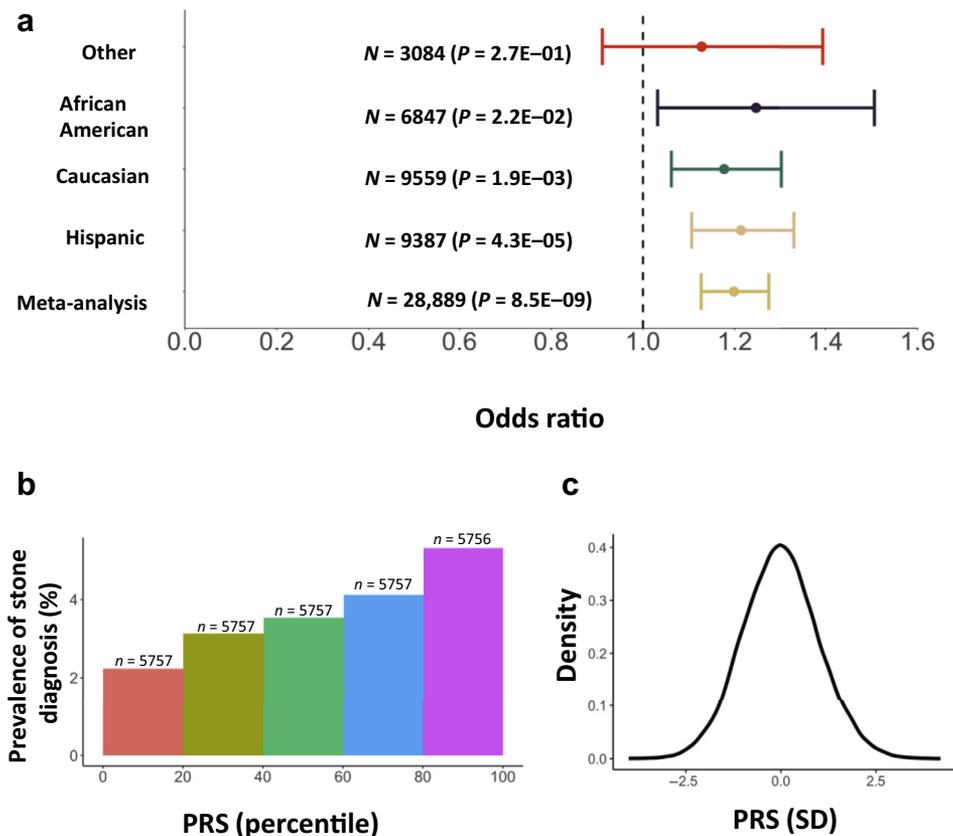


Figure 2 | Clinical impact of polygenic risk scores (PRSs) on urinary tract stone diagnosis. (a) Association of PRS with urinary tract stone diagnosis stratified by race in BioMe. Odds ratio was computed using a logistic regression adjusted for age, sex, 10 genetic principal components, history of hypertension, gout, and type 2 diabetes. (b) Prevalence of urinary tract stone diagnosis for different levels of PRS. (c) Distribution of normalized PRS in BioMe.

The advent of GWAS and large-scale population-based genetic studies such as the UK Biobank^{13,14} have led to the identification of the polygenic nature of complex disorders such as hypertension,¹⁵ coronary artery disease,¹¹ stroke,¹⁶ and schizophrenia.¹⁷ Because the genome does not change over one’s lifetime, polygenicity can be leveraged for better understanding of the genetic architecture of complex disease. In this study, we demonstrate that using genome-wide information summarized as a polygenic score significantly correlates with a stone diagnosis. We also show that individuals with a PRS in the top 10th percentile are at almost

3-fold odds of a urinary tract stone diagnosis as compared to individuals in the bottom 10th percentile of PRSs, indicating these individuals are at higher baseline genetic risk.

Urinary tract stone formation is a complex process with known risk factors, including family history, obesity, gout, type 2 diabetes, as well as dietary factors such as excessive animal protein and salt intake and insufficient water intake.⁶ In comparing cases with controls, we recapitulated known epidemiologic associations, with greater incidence of obesity, hypertension, type 2 diabetes, gout, and proportion of males in cases.^{18–21} However, we also demonstrate that our genome-wide PRS is significantly associated with urinary stone diagnosis in low-risk individuals with no obesity, hypertension, type 2 diabetes, or gout. While this is not a complete set of clinical risk factors and does not consider family history, it represents the first attempt at applying polygenic scores in individuals with no evident clinical risk factors. This indicates that urinary tract stones may have a genetic component that is additive to known clinical risk.

We also validated the PRS in an external cohort and show that the PRS is associated with stone diagnosis in European Americans but not African Americans. This failure to replicate in African Americans is likely because the GWAS

Table 2 | Association of PRSs with urinary tract stone diagnoses adjusted for age, sex, BMI, 10 genetic PCs, history of hypertension, gout and type 2 diabetes stratified by race

Group	Cases (n)	Controls (n)	Adjusted OR (95% CI)	P value
European American	323	9248	1.18 (1.06–1.3)	0.002
Hispanic American	473	8918	1.21 (1.11–1.33)	<0.001
African American	180	6670	1.25 (1.03–1.51)	0.02
Other	95	2992	1.13 (0.91–1.39)	0.2
Meta-analysis	1071	27,828	1.20 (1.13–1.27)	<0.001

BMI, body mass index; CI, confidence interval; OR, odds ratio; PC, principal component; PRS, polygenic risk score.

Table 3 | Association of PRSs with urinary tract stone diagnoses in clinically high-risk and low-risk subgroups in BioMe adjusted for age, sex, BMI, and 10 genetic PCs

Group	P value	OR	95% CI	Cases (n)	Controls (n)
High clinical risk					
Hispanic American	0.002	1.17	(1.06–1.29)	428	7753
African American	0.07	1.20	(0.98–1.45)	166	5861
European American	0.003	1.19	(1.06–1.33)	265	6151
Other	0.3	1.12	(0.89–1.4)	76	2041
Meta-analysis	<0.001	1.18	(1.10–1.26)	935	21,806
Low clinical risk					
Hispanic American	0.0004	1.72	(1.27–2.32)	45	1161
African American	0.096	1.84	(0.91–3.87)	14	806
European American	0.3	1.13	(0.88–1.44)	58	3085
Other	0.4	1.24	(0.72–2.18)	19	948
Meta-analysis	0.001	1.33	(1.12–1.58)	136	6000

BMI, body mass index; CI, confidence interval; OR, odds ratio; PC, principal component; PRS, polygenic risk score.

summary statistics used to generate the PRS were derived from a European cohort. This highlights a crucial equity issue. There is underrepresentation of ethnic minorities in genomic research and thus a need to conduct genomic analyses in multiethnic datasets.²²

Although we cannot make claims about clinical utility, with the increasing availability of genetic information linked to clinical phenotype information, polygenic scores could be utilized to identify individuals, especially those without traditional risk factors with a high risk of developing urinary tract stones.²³ However, a polygenic score should be considered as a biomarker, albeit constant and further studies using clinical utility of these scores need to be explored further to determine the feasibility of incorporating polygenic scores in clinical practice. “When the PRS was added to a logistic regression model with only clinical variables, the AUC for urinary tract stone prediction increased by 5%. This represents a moderate increase in discriminative ability.

However, because family history and diet were not available in this study, these variables must be considered as covariates when evaluating the added predictive ability of the PRS.

This study should be interpreted in the light of some limitations. First, we defined stone diagnosis utilizing International Classification of Diseases (ICD) codes. Although ICD codes have been previously used in epidemiological studies,^{24,25} and Semins *et al.*²⁶ previously reported a high positive predictive value of 96% for using ICD codes to identify urinary tract stones in EHR systems, the possibility of misclassification exists. The overall prevalence of urinary tract stone diagnosis was 3.7% in BioMe and 5.6% in PMBB, which is significantly lower than the population prevalence estimates of approximately 10%. Second, our study is limited by the scope of EHR data. Because EHR data is intended to serve a billing tool rather than a complete patient phenotype, we did not have access to records of dietary habits, environmental influences, or family history of urinary tract stones. Thus, we are unable to determine the relative predictive value of these nongenetic factors. However, even if a subset of the low-risk patients has a family history of kidney stones, our PRS can still be robustly applied in this setting because many patients may be unaware of their family history. In this case, applying the PRS would provide an unbiased estimate of kidney stone risk. While lack of complete phenotypic information is a restriction to using EHR data, by using this cohort, our approach shows utility of this approach in large multiethnic populations. Third, our validation cohort (PMBB) included only a small number of individuals with low clinical risk due to differences in recruitment strategies. Thus, due to low statistical power, we could not conduct stratification based on clinical risk as was done in the BioMe Biobank. Fourth, because our dataset is retrospective, we are unable to determine the clinical utility of a PRS as a screening tool to guide clinical management. Fifth, because we use millions of SNPs with small but nonzero effect sizes, and our

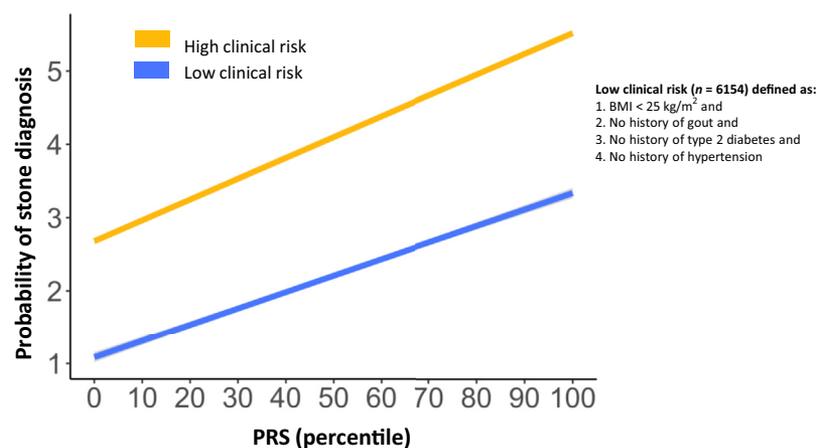


Figure 3 | Association of polygenic risk scores (PRSs) with urinary tract stone diagnoses in BioMe stratified by clinical risk adjusted for age, sex, body mass index (BMI), and 10 genetic principal components. Low clinical risk was defined as BMI <25 kg/m² and no history of gout, type 2 diabetes, and hypertension. High clinical risk included all other patients.

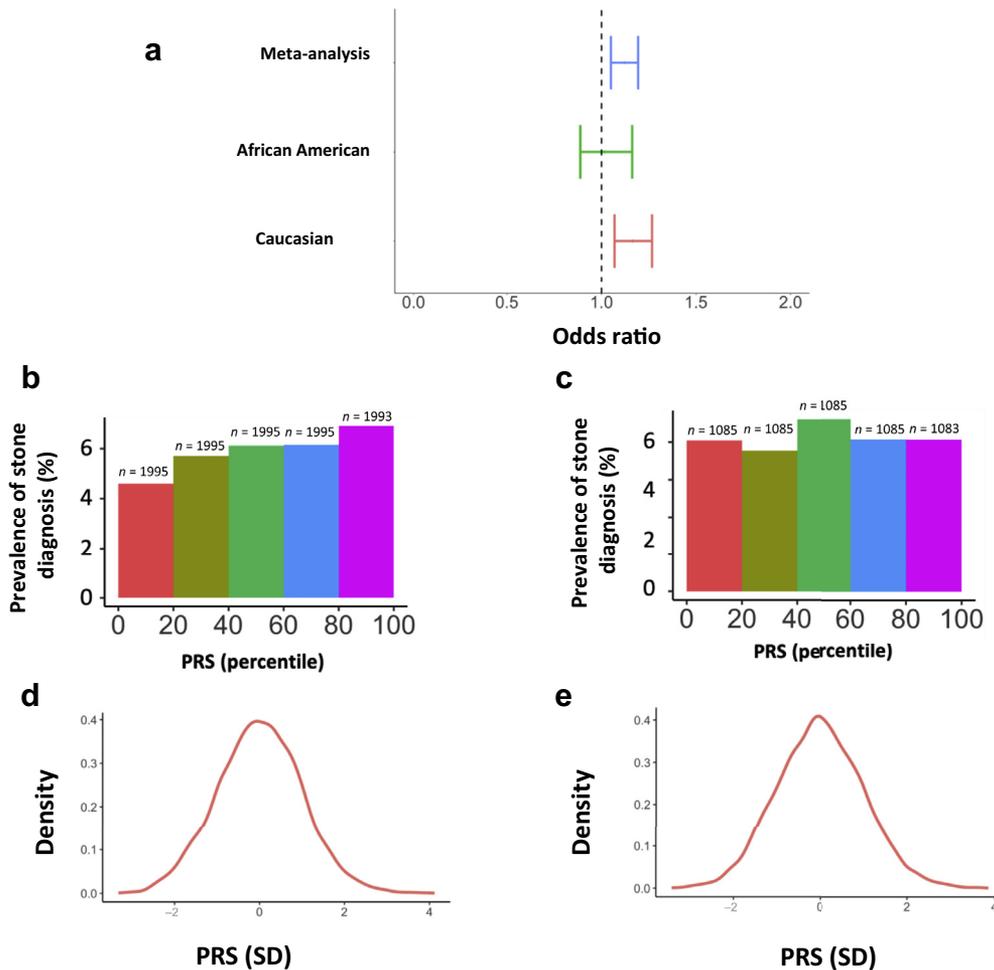


Figure 4 | Association of polygenic risk scores (PRSs) with urinary tract stone diagnosis in Penn Medicine Biobank. (a) Association of PRS with urinary tract stone diagnosis in Penn Medicine Biobank validation cohort stratified by self-reported race. Prevalence of urinary tract stone diagnosis for different levels of PRSs in (b) European Americans and (c) African Americans. Distribution of PRS in (d) European Americans and (e) African Americans.

analysis uses linear regression on a binary outcome, it is difficult to interpret the contribution of individual variants in isolation. However, this emphasizes the polygenic architecture of many complex diseases. Finally, we could not assess whether PRS associates with number, size, type, recurrence, and composition of kidney stones, because this information is not easily accessible in routinely collected EHR data. However, these scores can be validated in other existing cohorts of kidney stone patients with granular information about kidney stones.

In summary, we show that a genome-wide polygenic score derived from publicly available GWAS data and validated within 2 large, multiethnic population biobanks is significantly associated with urinary stone diagnosis. We show that in the absence of clinical risk factors, the PRS was significantly associated with urinary tract stones. Although further validation and implementation needs to be conducted in prospective longitudinal studies, this work highlights the highly polygenic nature of this complex disease.

Conclusions

A genome-wide PRS is significantly associated with urinary tract stone diagnosis independent of clinical risk factors.

METHODS

Derivation of polygenic scores

Polygenic scores measure the cumulative impact of common variants on the risk of certain disorders with the impact of individual variants

Table 4 | Association of PRSs with urinary tract stone diagnoses in external validation cohort (Penn Medicine Biobank)

Group	Cases (n)	Controls (n)	OR (95% CI)	P value
African American	223	5200	1.02 (0.89–1.16)	0.07
European American	588	9385	1.16 (1.07–1.27)	<0.001
Meta-analysis	811	14,585	1.12 (1.05–1.19)	<0.001

CI, confidence interval; OR, odds ratio; PRS, polygenic risk score. A logistic regression model adjusted for age, sex, body mass index, and 10 genetic principal components was fit in racial groups. A meta-analysis was performed using the inverse variance method.

assumed to be additive. Thus, for each individual, scores are computed by taking the sum of the dosage of risk variants weighted by their effect on the disease under consideration.

We derived all polygenic scores using data from the UK Biobank. Briefly, the UK Biobank consists of genotype, phenotype, and demographic data of more than 500,000 individuals recruited across the United Kingdom enrolled between ages 40 and 69 years. Each individual completed a questionnaire with medical history and demographics. Additionally, blood biochemistry assays were performed at time of enrollment. Individual genotypes were generated from either the Affymetrix Axiom (Santa Clara, CA) UK Biobank array (~450,000 individuals) or the UK BiLEVE (Affymetrix) array (~50,000 individuals), each containing ~0.8 million markers. More variants were then imputed using the Haplotype Reference Consortium combined with UK10K haplotype resource, leading to ~96 million variants available.

We obtained GWAS summary statistics from an online GWAS UK Biobank database²⁷ for the kidney stone, ureter stone, and bladder stone phenotype. This phenotype was identified using diagnostic codes. The summary statistics were generated by first restricting SNPs to those with minor allele frequency >0.1% and Hardy-Weinberg equilibrium P value $>1 \times 10^{-10}$. Following quality control, 10.8 million SNPs were included in downstream association analysis. Association analysis was conducted by fitting a logistic regression model with additive genotype coding (0, 1, or 2 copies of the minor allele) adjusted for sex and the first 10 genetic principal components.

To compute polygenic scores, we first adjusted summary statistics effect sizes using LDpred.¹² This tool adjusts effect sizes for linkage disequilibrium. We used the 1000 Genomes European population linkage disequilibrium reference panel. The key tuning parameter in LDpred is ρ , the assumed proportion of causal SNPs. We used effect sizes adjusted for several values of ρ between 1 (assumes all SNPs are causal) and 0.001. We then computed a polygenic score for each using an additive function such that $PRS = \sum iS_i \times G_i$, where S_i = adjusted β statistics for minor allele and G_i = genotype (0, 1, or 2). We chose an optimal ρ by selecting the ρ and the PRS that had a maximal association with urinary tract stone diagnosis in BioMe. We scaled the polygenic scores to have mean of 0 and SD of 1.

A score was computed with only GWAS-significant SNPs that met the P value threshold ($P < 5 \times 10^{-8}$) for association with urinary tract stones in UK Biobank. Weights from the GWAS summary statistics were directly used without LDpred adjustment.

BioMe discovery cohort

We utilized the BioMe Biobank at Mount Sinai. Briefly, the BioMe Biobank is an EHR-linked clinical care cohort composed of over 45,000 participants from diverse ancestries (African, Hispanic, Latino, European, and Other ancestries), with accompanying genome-wide genotyping data for 31,441 participants. Along with the genetic information, BioMe is linked to a wide array of biomedical traits, originating from Mount Sinai's system-wide Epic EHR. Enrollment of participants is predominantly through ambulatory care practices and is representative of Mount Sinai's larger patient population. BioMe participants ($n = 31,441$) were genotyped on the Global Screening Array platform (Illumina, San Diego, CA). Quality control and imputation of the Global Screening Array data is detailed in [Supplementary Methods](#).

PMBB validation cohort

The PMBB consists of 60,000 patients recruited from clinical sites across the University of Pennsylvania Health System who have

provided consent for access to all EHRs. This study included a subset of 9,973 European and 5,423 African ancestry patients that had undergone genotyping. Samples within PMBB were genotyped on the Illumina Quad Omni SNP Chip by Regeneron Pharmaceuticals (Tarrytown, NY). Following sequencing, standard quality control procedures were followed to remove rare and missing variants as well as variants in linkage disequilibrium. Variants with a minor allele frequency of <0.05, a missing rate of >0.05 and a Hardy-Weinberg P of $>10^6$ were removed. Furthermore, samples with a genotype missingness of >0.02 were also removed.

Identification of urinary tract stone patients

We defined urinary tract stone cases in BioMe and PMBB as any patient with 1 or more encounters with an ICD-10 diagnosis code of N20.0, N20.1, N20.2, or N20.9 or ICD-9 diagnosis code of 592.0, 592.1, or 592.9.

Statistical analysis

We compared categorical and continuous variables between cases and controls using the χ^2 test and Student's t test, respectively. To test for association between scaled PRSs and urinary tract stone diagnoses, we utilized a logistic regression model adjusting for age, sex, 10 PCs, and history of clinical comorbidities, gout, hypertension, and type 2 diabetes determined using previously validated phenotyping algorithms²⁸ and diagnostic codes. For all association analyses, we stratified participants by self-reported race or ethnicity or both (European American, African American, Hispanic American, and Other). Using race-specific results, we performed a meta-analysis using the inverse variance method as implemented in the *meta* R package.²⁹

We then sought to assess the utility of the best performing polygenic score in improving discrimination of stone diagnosis in the absence and presence of clinical risk factors, including gout, hypertension, type 2 diabetes, and BMI. A low-risk group was defined as individuals with BMI <25 kg/m² and no history of hypertension, type 2 diabetes, or gout. The high-risk group included all other individuals. We then fit a logistic regression model adjusted for age, sex, and 10 genetic PCs to determine the association between polygenic scores and kidney stone diagnoses in each subgroup separately stratified by race. We also used a bootstrap method with 1000 iterations to determine the AUC of logistic regression models with (i) clinical variables and (ii) clinical variables + PRS. The 95% CIs of AUC were computed from bootstrap results.

DISCLOSURE

All the authors declared no competing interests.

ACKNOWLEDGMENTS

PQD was supported by the National Institutes of Health Medical Scientist Training Program Training grant T32GM007205. SGC is supported by the following grants: R01DK106085, R01HL85757, R01DK112258, and U01OH011326. RD is supported by R35GM124836 from the National Institute of General Medical Sciences of the National Institutes of Health and R01HL139865 from the National Heart, Lung, Blood Institute of the National Institutes of Health and previously by American Heart Association Cardiovascular Genome-Phenome Discovery grant 15CVGSPD27130014. SMD is supported by the US Department of Veteran Affairs grant IK2-CX001780. GNN is supported by a career development award K23DK107908 from the National Institutes of Health and is also supported by R01DK108803, U01HG007278, U01HG009610, and 1U01DK116100.

The BioMe health care delivery cohort at Mount Sinai was established and maintained with a generous gift from the Andrea and

Charles Bronfman Philanthropies. This work was supported in part through the computational resources and staff expertise provided by the Department of Scientific Computing at the Icahn School of Medicine at Mount Sinai.

AUTHOR CONTRIBUTIONS

IP, GNN, JCH, SGC, RD, BSG, MG, and SMD designed the study. IP, NT, RJ, GNN, and RD acquired the data. IP, MP, RO'H, JP, AK, MG, KC, AV, FC, DK, NT, and RJ analyzed and interpreted the data. IP, PQD, SJ, MP, MG, GNN, SGC, RD, SMD, and JCH drafted and revised the manuscript.

SUPPLEMENTARY MATERIAL

Supplementary File (PDF)

Supplementary Methods.

Figure S1. Association with urinary tract stone diagnosis in BioMe adjusted for age, sex, BMI, and 10 genetic principal components for different values of LDpred p .

Table S1. Summary of UK Biobank cohort from which GWAS summary statistics were derived. Categorical and continuous variables were compared using the χ^2 test and t test, respectively.

Table S2. Summary of Penn Medicine Biobank dataset used for validation.

REFERENCES

1. Scales CD, Smith AC, Hanley JM, Saigal CS. Prevalence of kidney stones in the United States. *Eur Urol.* 2012;62:160–165.
2. Evan AP. Physiopathology and etiology of stone formation in the kidney and the urinary tract. *Pediatr Nephrol.* 2010;25:831–841.
3. Khan SR, Pearle MS, Robertson WG, et al. Kidney stones. *Nat Rev Dis Primers.* 2016;2:16008.
4. Rule AD, Bergstralh EJ, Melton LJ, et al. Kidney stones and the risk for chronic kidney disease. *Clin J Am Soc Nephrol.* 2009;4:804–811.
5. Monico CG, Milliner DS. Genetic determinants of urolithiasis. *Nat Rev Nephrol.* 2012;8:151–162.
6. Alelign T, Petros B. Kidney stone disease: an update on current concepts. *Adv Urol.* 2018;2018:3068365.
7. Punnoose AR, Golub RM, Lynn C. Kidney stones. *JAMA.* 2012;307:2557–2557.
8. Goldfarb DS. The search for monogenic causes of kidney stones. *J Am Soc Nephrol.* 2015;26:507–510.
9. Khera AV, Chaffin M, Aragam KG, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet.* 2018;50:1219–1224.
10. Fanous AH, Zhou B, Aggen SH, et al. Genome-wide association study of clinical dimensions of schizophrenia: polygenic effect on disorganized symptoms. *Am J Psychiatry.* 2012;169:1309–1317.
11. Inouye M, Abraham G, Nelson CP, et al. Genomic risk prediction of coronary artery disease in 480,000 adults: implications for primary prevention. *J Am Coll Cardiol.* 2018;72:1883–1893.
12. Vilhjálmsson BJ, Yang J, Finucane HK, et al. Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *Am J Hum Genet.* 2015;97:576–592.
13. Bycroft C, Freeman C, Petkova D, et al. Genome-wide genetic data on ~500,000 Biobank participants [preprint]. *bioRxiv.* July 20, 2017. <https://doi.org/10.1101/166298>.
14. Thompson SG, Willeit P. UK Biobank comes of age. *Lancet.* 2015;386:509–510.
15. Kraja AT, Cook JP, Warren HR, et al. New blood pressure-associated loci identified in meta-analyses of 475 000 individuals. *Circ Cardiovasc Genet.* 2017;10:e001778.
16. Pulit SL, Weng LC, McArdle PF, et al. Atrial fibrillation genetic risk differentiates cardioembolic stroke from other stroke subtypes. *Neurol Genet.* 2018;4:e293.
17. Ward J, Strawbridge RJ, Bailey MES, et al. Genome-wide analysis in UK Biobank identifies four loci associated with mood instability and genetic correlation with major depressive disorder, anxiety disorder and schizophrenia. *Transl Psychiatry.* 2017;7:1264.
18. Taylor EN, Stampfer MJ, Curhan GC. Diabetes mellitus and the risk of nephrolithiasis. *Kidney Int.* 2005;68:1230–1235.
19. Cappuccio FP, Strazzullo P, Mancini M. Kidney stones and hypertension: population based study of an independent clinical association. *BMJ.* 1990;300:1234–1236.
20. Taylor EN, Stampfer MJ, Curhan GC. Obesity, weight gain, and the risk of kidney stones. *JAMA.* 2005;293:455–462.
21. Yu TF. Urolithiasis in hyperuricemia and gout. *J Urol.* 1981;126:424–430.
22. Popejoy AB, Fullerton SM. Genomics is failing on diversity. *Nature.* 2016;538:161–164.
23. Frassetto L, Kohlstadt I. Treatment and prevention of kidney stones: an update. *Am Fam Physician.* 2011;84:1234–1242.
24. Tasian GE, Jemielita T, Goldfarb DS, et al. Oral antibiotic exposure and kidney stone disease. *J Am Soc Nephrol.* 2018;29:1731–1740.
25. Rule AD, Roger VL, Melton LJ 3rd, et al. Kidney stones associate with increased risk for myocardial infarction. *J Am Soc Nephrol.* 2010;21:1641–1644.
26. Semins MJ, Trock BJ, Matlaga BR. Validity of administrative coding in identifying patients with upper urinary tract calculi. *J Urol.* 2010;184:190–192.
27. Neale B. Rapid GWAS of thousands of phenotypes for 337,000 samples in the UK Biobank. The Neale Lab. September 20, 2017. Available at: <http://www.nealelab.is/blog/2017/7/19/rapid-gwas-of-thousands-of-phenotypes-for-337000-samples-in-the-uk-biobank>. Accessed August 12, 2019.
28. Nadkarni GN, Gottesman O, Linneman JG, et al. Development and validation of an electronic phenotyping algorithm for chronic kidney disease. *AMIA Annu Symp Proc.* 2014;2014:907–916.
29. Schwarzer G, Carpenter JR, Rücker G. An introduction to meta-analysis in R. In: *Meta-Analysis with R*. Basel, Switzerland: Springer Nature; 2015:3–17.