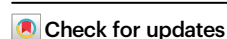


Trans-ancestry GWAS identifies 59 loci and improves risk prediction and fine-mapping for kidney stone disease

Received: 31 July 2024

Accepted: 27 March 2025

Published online: 11 April 2025



Xi Cao¹, Minghui Jiang², Yunlong Guan¹, Si Li¹, Chen Duan³, Yan Gong³, Yifan Kong¹, Zhonghe Shao¹, Hongji Wu¹, Xiangyang Yao³, Bo Li³, Miao Wang⁴✉, Hua Xu^{3,5}✉ & Xingjie Hao¹✉

Kidney stone disease is a multifactorial disease with increasing incidence worldwide. Trans-ancestry GWAS has become a popular strategy to dissect genetic structure of complex traits. Here, we conduct a large trans-ancestry GWAS meta-analysis on kidney stone disease with 31,715 cases and 943,655 controls in European and East Asian populations. We identify 59 kidney stone disease susceptibility loci, including 13 novel loci and show similar effects across populations. Using fine-mapping, we detect 1612 variants at these loci, and pinpoint 25 causal signals with a posterior inclusion probability >0.5 among them. At a novel locus, we pinpoint *TRIOBP* gene and discuss its potential link to kidney stone disease. We show that a cross-population polygenic risk score, $PRS-CS_{EAS\&EUR}$, exhibits superior predictive performance for kidney stone disease than other polygenic risk scores constructed in our study. Relative to individuals in the third quintile of $PRS-CS_{EAS\&EUR}$, those in the lowest and highest quintiles exhibit distinct kidney stone disease risks with odds ratios of 0.57 (0.51–0.63) and 1.83 (1.68–1.98), respectively. Our results suggest that kidney stone disease patients with higher polygenic risk scores are younger at onset. In summary, our study advances the understanding of kidney stone disease genetic architecture and improves its genetic predictability.

Kidney stone disease (KSD) is a common urological disease caused by abnormal accumulation of crystalline substances, most of which are calcium stones¹. KSD has an increasing prevalence and a recurrence rate of 50% within 5 years². KSD is considered as a multifactorial disease, involving environmental, dietary, hormonal, and genetic components³. A recent twin study revealed that KSD exhibits substantial heritability, with estimates of 0.46 in women and 0.57 in men⁴.

Genetic research on KSD holds the potential to uncover its pathogenesis and inform more effective prevention strategies.

In recent years, genome-wide association studies (GWAS) in different ancestries brought more diverse perspectives for the research on KSD genetic architecture. Several GWASs and meta-analyses have identified numerous genetic susceptibility loci, related to multiple potential mechanisms of KSD, such as calcium and phosphorus

¹Department of Epidemiology and Biostatistics, School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei, China. ²Department of Neurology; Center of excellence for Omics Research (CORe), Beijing Tiantan Hospital, Capital Medical University; China National Clinical Research Center for Neurological Diseases, Beijing, China. ³Department of Urology, Tumor Precision Diagnosis and Treatment Technology and Translational Medicine, Hubei Engineering Research Center, Zhongnan Hospital of Wuhan University, Wuhan, China. ⁴Department of Urology, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China. ⁵Taikang Center for Life and Medical Sciences, Wuhan University, Wuhan, Hubei, China. ✉e-mail: wangmiaotj@163.com; xuhuawhu@163.com; xingjie@hust.edu.cn

metabolism and kidney function^{5–8}. By assessing the heterogeneity of KSD associated loci effects in UK and Japanese populations, a study suggested that the KSD genetic architecture was similar between populations of EUR and EAS ancestry⁷. Although cross-population fine-mapping is an efficient tool to pinpoint the susceptibility loci, we found that the causal signals of KSD have not been comprehensively explored to date. In addition, polygenic risk score (PRS), constructed through calculating weighted scores of genome-wide risk alleles, could be used to identify individuals at high genetic risk for KSD and thus enable enhanced screening or preventive therapies⁹. While to our knowledge, all currently published KSD-PRS were constructed using GWAS summary statistics from EUR population only^{10–13}, which related to the imbalance of GWAS in populations of different ancestries. Considering that the genetic architecture of KSD exhibits both similarities and differences across different ancestries, the cross-population PRS might improve the predictive performance for KSD risk and enhance the transferability across populations.

In the present study, we conducted a large-scale trans-ancestry GWAS meta-analysis for KSD, including 31,715 KSD cases and 943,655 controls from EUR and EAS populations. We aimed to identify additional KSD susceptibility loci, pinpoint causal signals on them, and further explore the potential mechanisms underlying their associations with KSD. In addition, we constructed cross-population PRS for KSD to improve the predictive performance and pave the way for precision medicine.

Results

Trans-ancestry GWAS meta-analysis identified 13 Novel loci

As shown in Fig. 1, this study included several GWAS summary statistics for KSD from EUR and EAS populations. The EUR GWAS summary statistics were sourced from our previous study⁸, which contained 9372 KSD cases and 369,102 controls from UK Biobank (UKB)¹⁴ and 8597 cases with 333,128 controls from FinnGen¹⁵. For the EAS populations, we included GWAS summary statistics from the China Kadoorie Biobank (CKB)¹⁶ with 2047 cases and 74,398 controls and Biobank Japan (BBJ)¹⁷ with 11,699 cases and 167,027 controls. Further details on GWAS quality control, sample size, phenotype definition, and statistical analysis were provided in Supplementary Data 1.

By integrating EUR and EAS summary statistics (Supplementary Note 1), we performed a trans-ancestry GWAS meta-analysis using a fixed-effect inverse-variance weighted model with METAL. Among 5,960,489 variants shared between EUR and EAS meta-analyses, 84 independent significant signals were observed after clumping. Following genomic control corrections, the inflation factors for the meta-analysis were $\lambda = 1.124$ and $\lambda_{1000} = 1.002$ (Supplementary Fig. 1). KSD susceptibility loci were defined as non-overlapping genomic regions within 1000 kb of the lead SNPs. By merging lead SNPs within 1000 kb of each other, 59 loci were finally identified across these signals, including 13 novel loci (Fig. 2A, Table 1, Supplementary Data 2). In the replication study using GWAS summary statistics from Million Veteran Program (MVP)¹⁸, all 13 lead SNPs exhibited similar effect sizes with the same sign direction between our study and MVP (Supplementary Fig. 2). Among them, ten SNPs reached nominal significance, and eight SNPs remained significant after Bonferroni correction ($P < 3.85 \times 10^{-3}$, Table 1). An additional trans-ancestry meta-analysis was performed using MR-MEGA. Employing the same clumping and merging strategies, we identified six novel KSD loci, all of which were also detected by METAL (Supplementary Data 3). Most lead SNPs identified by METAL also exhibited genome-wide significance in MR-MEGA, including all variants with nominal significant ancestry-heterogeneity (Supplementary Fig. 3, Supplementary Note 2). These findings collectively reinforce the reliability of our study.

Among these novel loci, the lead variant rs7974476 at *GXYLT1* was the most significant signal ($P = 7.94 \times 10^{-11}$) and the rs76271768 at *CPNE8* exhibited the largest effect, with an odds ratio (OR) of 1.11 (95%

confidence interval (CI): 1.07–1.15) for allele A. Several genes of novel loci suggested potential mechanisms linked to KSD. For example, *SLC17A3*, linked to KSD with the lead SNP rs972087, encodes human sodium phosphate transporter 4 (hNPT4), which facilitates the excretion of intracellular urates and organic anions from the blood into renal tubule cells^{19,20}. On chromosome 16, we found that rs9925265 at downstream region of *SLC12A3* was associated with KSD (OR = 1.05 for risk allele A, 95% CI: 1.03–1.07, $P = 1.19 \times 10^{-8}$). *SLC12A3* encodes the thiazide-sensitive $\text{Na}^+\text{-Cl}^-$ cotransporter (NCC)²¹, which plays an important role in sodium excretion and maintenance of salt homeostasis. In addition, rs6503517 at *CDK12*, which was negatively associated with the estimated glomerular filtration rate^{22,23}, showed the protective effect with OR (95% CI) of 0.94 (0.92–0.96) for allele A.

Similar effects in European and East Asian populations

The trans-ancestry GWAS meta-analysis helped identify loci that were not significant in population-specific analyses (Fig. 2B). Of the 59 KSD loci, only nine lead variants were significant in both EUR and EAS population samples (Fig. 2C). Nevertheless, we observed consistent effect directions and similar effect sizes for most lead variants across EUR and EAS population samples, with a Lin's concordance correlation coefficient (ρ_c) of 0.94 (95% CI: 0.89–0.96). Only two variants, rs1936824 and rs35310650, exhibited significant heterogeneity in the meta-analysis using METAL after Bonferroni corrections (Supplementary Data 2). In MR-MEGA analysis, these two variants, along with an additional variant, rs2776288, were also observed with significant ancestry-heterogeneity (Supplementary Note 2). Among them, rs1936824 exhibited the largest difference in effect size (β) on KSD between the two population, with a stronger effect of 0.177 for allele A observed in the EUR population samples compared to 0.090 in the EAS population samples.

Cross-population fine-mapping captured shared and ancestry-specific causal signals

Among the genomic regions surrounding 57 lead SNPs, MESuSiE identified at least one credible set in 54 regions, and SuSiE identified in 44 and 21 regions in EUR and EAS populations, respectively. Although the sizes of credible sets did not show significant difference ($P = 0.35$) (Fig. 3A), cross-population fine-mapping identified more causal signals with a PIP > 0.5 at these loci (Supplementary Data 4, Fig. 3B). Of the 25 causal signals identified by MESuSiE, 22 were classified as shared across populations, further supporting the genetic similarity between EUR and EAS populations for KSD. For example, as shown in Fig. 3C, the variant rs10051765 showed genome-wide significance in EUR, EAS and trans-ancestry GWAS meta-analyses. According to previous studies, it is an eQTL variant for *RGS14* gene^{24,25}, which has been reported to be associated with blood calcium level and KSD^{26,27}. MESuSiE detected rs10051765 as a causal signal shared across populations with PIP_{shared} of 1.000, while SuSiE neglected it in EAS population (PIP_{EAS} = 0.016). A total of 1609 variants were identified in 68 95% credible sets by MESuSiE, along with three additional isolated variants with a PIP > 0.5. These fine-mapped variants exhibited similar effects on KSD between EAS and EUR population samples, with ρ_c of 0.71 (0.69–0.73). When focusing on shared signals, and subsequently narrowing the comparison to shared signals with a genome-wide significant P value, the ρ_c further increased to 0.78 (0.76–0.80) and 0.93 (0.92–0.94), respectively (Supplementary Fig. 4). Nevertheless, rs11586977 at *ALPL*, rs12857280 at *CLDN10* and rs3886163 at *CYP24A1* were identified as EUR-specific causal signals with PIPs > 0.5 by both MeSuSiE and SuSiE, which reflected the ethnic differences in the predisposition to KSD (Supplementary Data 4).

Functional annotation for fine-mapped signals

Using the Variant Effect Predictor (VEP), we made a multifaceted functional annotation for the 1612 fine-mapped signals identified by

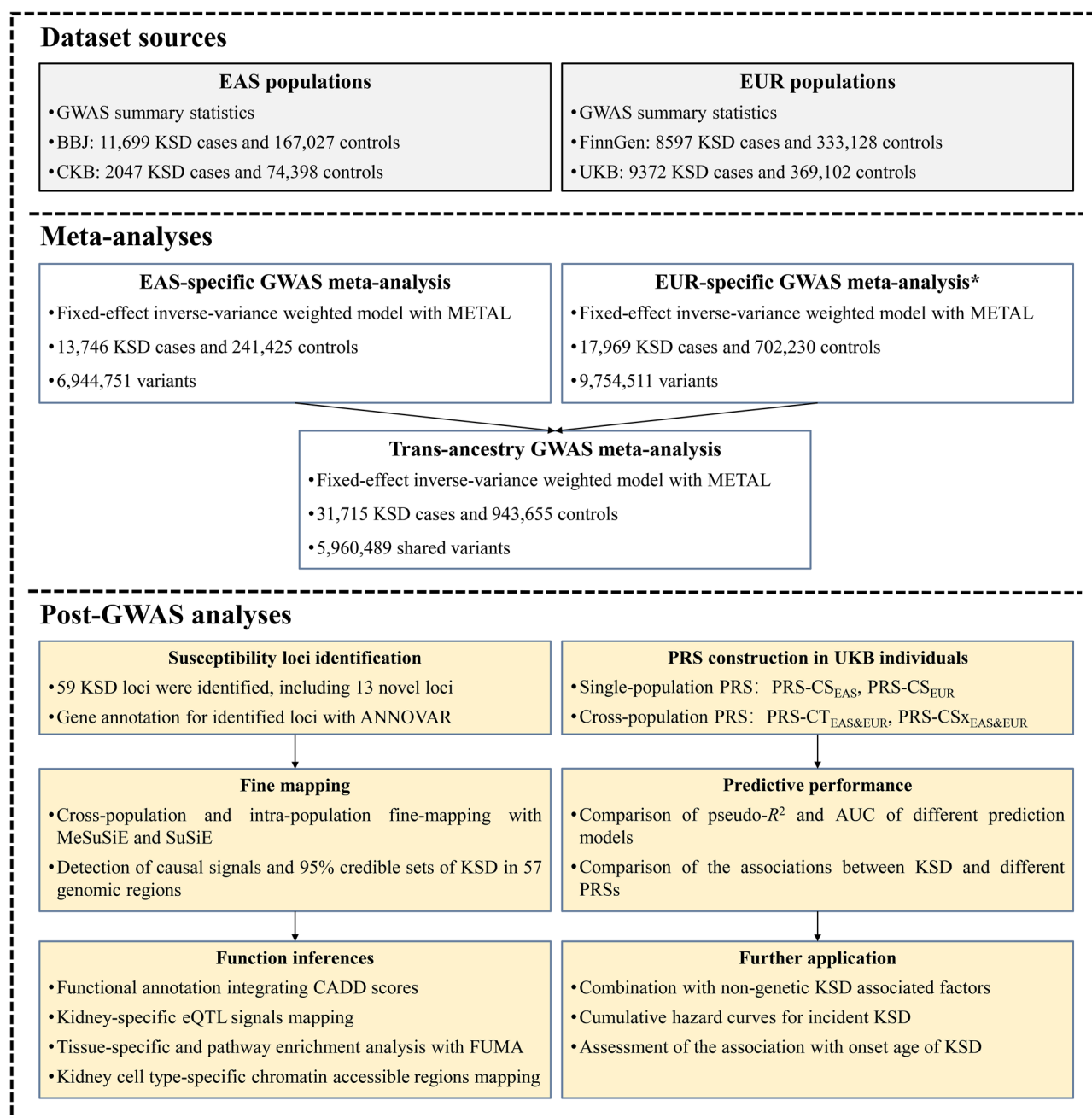


Fig. 1 | Study design and workflow. GWAS genome-wide association study, BBJ Biobank Japan, CKB China Kadoorie Biobank, UKB UK Biobank, CADD combined annotation-dependent depletion, AUC areas under the receiver operating curves,

PRS polygenic risk score, KSD kidney stone disease. *Referred that the EUR-specific meta-analysis was obtained in our previous study.

MESuSiE mentioned above, integrating deleterious scores from Combined Annotation Dependent Depletion (CADD)²⁸ (Supplementary Data 5). These fine-mapped variants were observed more likely among the top 10% of deleterious variants ranked by CADD compared to unmapped variants (Supplementary Data 6). Among them, missense variants with a PIP > 0.5 or mapped to a small credible set (number of SNPs < 5) were of particular concern for their strong support for the links between mapped genes and KSD. As shown in Supplementary Fig. 5, rs1260326 at *GCKR* and rs2231142 at *ABCG2*, were captured with CADD scores of 12.33 and 22.00, respectively, suggesting their potential deleterious effects. Among them, rs1260326 was the lead SNP of the corresponding susceptibility locus. Notably, two other missense variants, rs8140207 and rs12628603 at

TRIOBP, were detected at a novel susceptibility locus, with CADD scores of 13.25 and 18.18, respectively. We further observed that all signals within the credible set containing rs8140207 and rs12628603 were mapped as significant eQTL signals for the *TRIOBP* gene, including the lead SNP rs6000890 at this locus (Supplementary Fig. 6, Supplementary Data 5). These variants exhibited strong correlations with each other in both EUR and EAS populations, likely explaining their similarly low PIPs (Supplementary Fig. 7, Supplementary Fig. 8). Some other fine-mapped variants were also observed as eQTL signals for known KSD associated loci, such as *RGS14*, *UGT8* and *ALPL*. Compared to unmapped variants in cross-population fine-mapping, these fine-mapped variants were more likely to overlap with the kidney-specific eQTL signals (Supplementary Data 6), suggesting

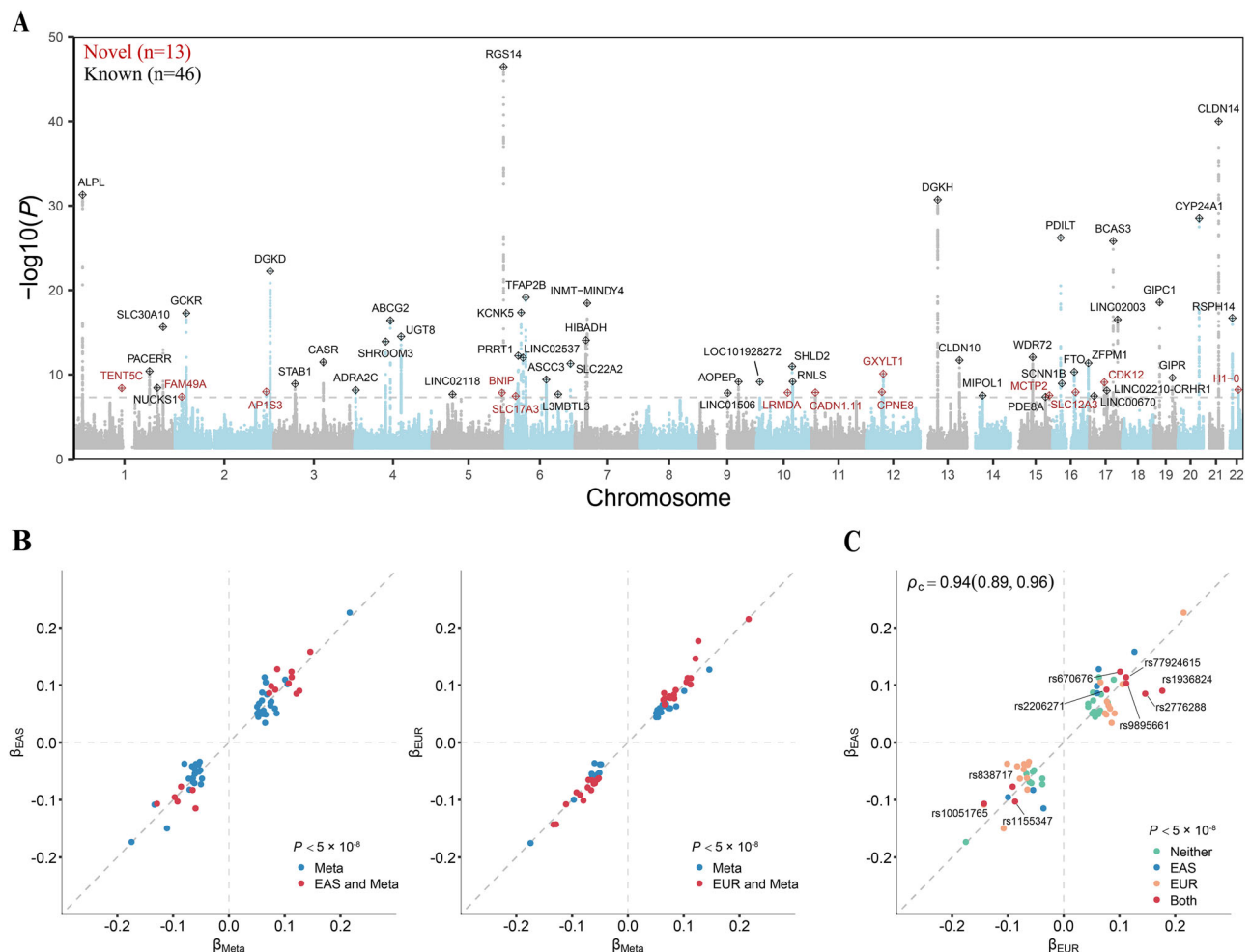


Fig. 2 | Trans-ancestry GWAS meta-analysis for KSD. A Manhattan plot for P -values of variants in trans-ancestry GWAS meta-analysis. The gray dashed line indicated genome-wide significance level of 5×10^{-8} . Red and black diamonds represented the novel and reported genetic susceptibility loci for KSD, respectively. **B** Comparison of lead variant effects at 59 identified loci between EAS-specific meta-analysis (left) and EUR-specific meta-analysis (right) with trans-ancestry meta-analysis. These points were colored based on their significances

from corresponding GWAS meta-analyses. Source data are provided as a Source Data file. **C** Comparison of lead variant effects at 59 identified loci between EAS-specific with EUR-specific meta-analyses. These points were colored based on their significances from corresponding GWAS meta-analyses. The ρ_c referred to Lin's concordance correlation coefficient. Variants that reached the genome-wide significance level in both EAS and EUR GWAS meta-analyses were labeled. Source data are provided as a Source Data file.

their potential genetic effects on KSD through gene expression regulation.

Additionally, we observed the largest number of fine-mapped variants in chromatin accessible regions specific to the kidney proximal tubule cell type (Supplementary Fig. 9). The tissue specificity analysis indicated differential expression of annotated genes for the fine-mapped variants in six tissues from the GTEx v.8 project, with significant upregulation observed in kidney cortex and stomach (Supplementary Fig. 10). Gene Ontology (GO) enrichment analysis suggested significant enrichment of annotated genes in biological processes related to the response to vitamin D and nutrient stimuli (Supplementary Data 7).

Incorporating trans-ancestry information improved PRS performance

The baseline information of UKB individuals used for PRS construction was presented in Supplementary Data 8. In the testing dataset, we constructed two single-population PRS (PRS-CS_{EUR} and PRS-CS_{EAS}) and two cross-population PRS (PRS-CT_{EAS&EUR} and PRS-CS_{EAS&EUR}) for 4996 KSD cases and 199,409 controls. As shown in Fig. 4A, all PRS

prediction models exhibited higher pseudo- R^2 and area under the receiver operating curves (AUC) values compared to the covariates-only model. Among these, the cross-population PRS-CS_{EAS&EUR} model showed the best predictive performance, with pseudo- R^2 of 0.047 and AUC of 0.677 (0.670–0.685). Delong tests confirmed that the differences in AUCs between PRS-CS_{EAS&EUR} model and other four models were statistically significant (Supplementary Data 9).

Of the four PRSs, PRS-CS_{EAS&EUR} caused the highest adjusted KSD risk per standard deviation increase with OR (95% CI) of 2.96 (2.75–3.19) (Table 2). No significant non-linear associations were observed between these PRSs and KSD (Supplementary Fig. 11). After grouping the testing dataset according to quintiles of each PRS, we observed the separation in KSD risk among individuals with PRS of different groups. As shown in Fig. 4B, PRS-CS_{EAS&EUR} caused the steepest step-wise change in KSD risk. Using Q3 group as reference, the ORs (95% CI) of Q1 and Q5 groups identified by the PRS-CS_{EAS&EUR} model were 0.57 (0.51–0.63) and 1.83 (1.68–1.98), respectively, indicating a 3.21-fold increase in KSD risk. For individuals with the top 5% PRS, the KSD risk identified by PRS-CS_{EAS&EUR} model was also the highest, with OR (95% CI) of 2.32 (2.07–2.59).

Table. 1 | 13 novel KSD susceptibility loci of cross-population GWAS meta-analysis

CHR	POS	Lead SNP	EA	NEA	EAF	OR (95% CI)	P value	Gene	Region	OR (95% CI) in MVP	P value in MVP
1	118174116	rs75417799	C	G	0.634	0.95 (0.93–0.96)	3.98×10^{-9}	TENT5C	intergenic (dist=3105)	0.96 (0.95–0.98)	4.0×10^{-6}
2	16733928	rs7552	A	G	0.544	1.05 (1.03–1.07)	4.20×10^{-8}	FAM49A	UTR3	1.04 (1.02–1.06)	1.3×10^{-5}
2	224679265	rs7600912	T	C	0.696	1.06 (1.04–1.08)	1.07×10^{-8}	AP1S3	intron	1.07 (1.05–1.09)	5.4×10^{-15}
5	172587274	rs255302	T	G	0.395	1.06 (1.04–1.08)	1.42×10^{-8}	BNIP1	intron	1.01 (0.99–1.02)	5.7×10^{-1}
6	25872579	rs972087	A	G	0.876	1.08 (1.05–1.11)	3.48×10^{-8}	SLC17A3	intron	1.01 (0.98–1.04)	5.9×10^{-1}
10	77690754	rs1359037	T	C	0.199	1.06 (1.04–1.09)	1.37×10^{-8}	LRMDA	intron	1.03 (1.00–1.05)	2.3×10^{-2}
11	10458596	rs1450270	T	C	0.402	1.05 (1.04–1.07)	1.31×10^{-8}	CAND1.11	intergenic (dist=6376)	1.02 (1.01–1.04)	3.8×10^{-3}
12	39306584	rs76271768	A	G	0.926	1.11 (1.07–1.15)	1.15×10^{-8}	CPNE8	intergenic (dist=7191)	1.04 (1.01–1.08)	1.1×10^{-2}
12	42542484	rs7974476	A	C	0.409	1.06 (1.04–1.08)	7.94×10^{-11}	GXYLT1	intergenic (dist=3808)	1.04 (1.02–1.06)	2.9×10^{-6}
15	94941051	rs8042590	A	G	0.583	0.95 (0.93–0.97)	2.86×10^{-8}	MCTP2	intron	0.97 (0.96–0.99)	1.2×10^{-3}
16	56950210	rs9925265	A	G	0.390	1.05 (1.03–1.07)	1.19×10^{-8}	SLC12A3	downstream	1.02 (1.00–1.03)	6.1×10^{-2}
17	37649687	rs6503517	A	G	0.783	0.94 (0.92–0.96)	8.06×10^{-10}	CDK12	intron	0.93 (0.92–0.95)	1.6×10^{-14}
22	38187301	rs6000890	T	C	0.609	0.95 (0.93–0.97)	6.22×10^{-9}	H1-O	intergenic (dist=13813)	0.95 (0.94–0.97)	6.5×10^{-8}

CHR chromosome, POS position, SNP single nucleotide polymorphism, EA effect allele, NEA non-effect allele, EAF effect allele frequency, OR odd ratio, CI confidence interval, MVP Million Veteran Program. The lead SNP represented the variant with the lowest P value of association test in the locus. SNPs located in the intergenic region use the nearest gene as corresponding gene, and the distance from the gene were showed in parentheses (unit: base pairs).

Based on previous studies and data accessibility, we included several known KSD related factors from UKB (Supplementary Methods), with their distribution among testing dataset individuals presented in Supplementary Data 8. Compared to these factors, adding PRS-CS_{EAS&EUR} to a basic model that included age, sex and the first ten principal components (PC1-10) brought larger AUC value increase (Fig. 4C). Adding PRS-CS_{EAS&EUR} to the prediction model containing age, sex, PC1-10 and those known KSD related factors significantly improved the reclassification for KSD, with a continuous net reclassification index (cNRI) of 0.326 (0.298–0.352). The net percentage of individuals with KSD correctly classified upward and those without KSD correctly classified downward was 0.165 (0.144–0.184) and 0.162 (0.151–0.172), respectively. Moreover, as illustrated in Fig. 4D, the adjusted cumulative hazard curves for incident KSD across different PRS-CS_{EAS&EUR} groups exhibited distinct separation over time ($P_{\text{log-rank}} < 0.001$).

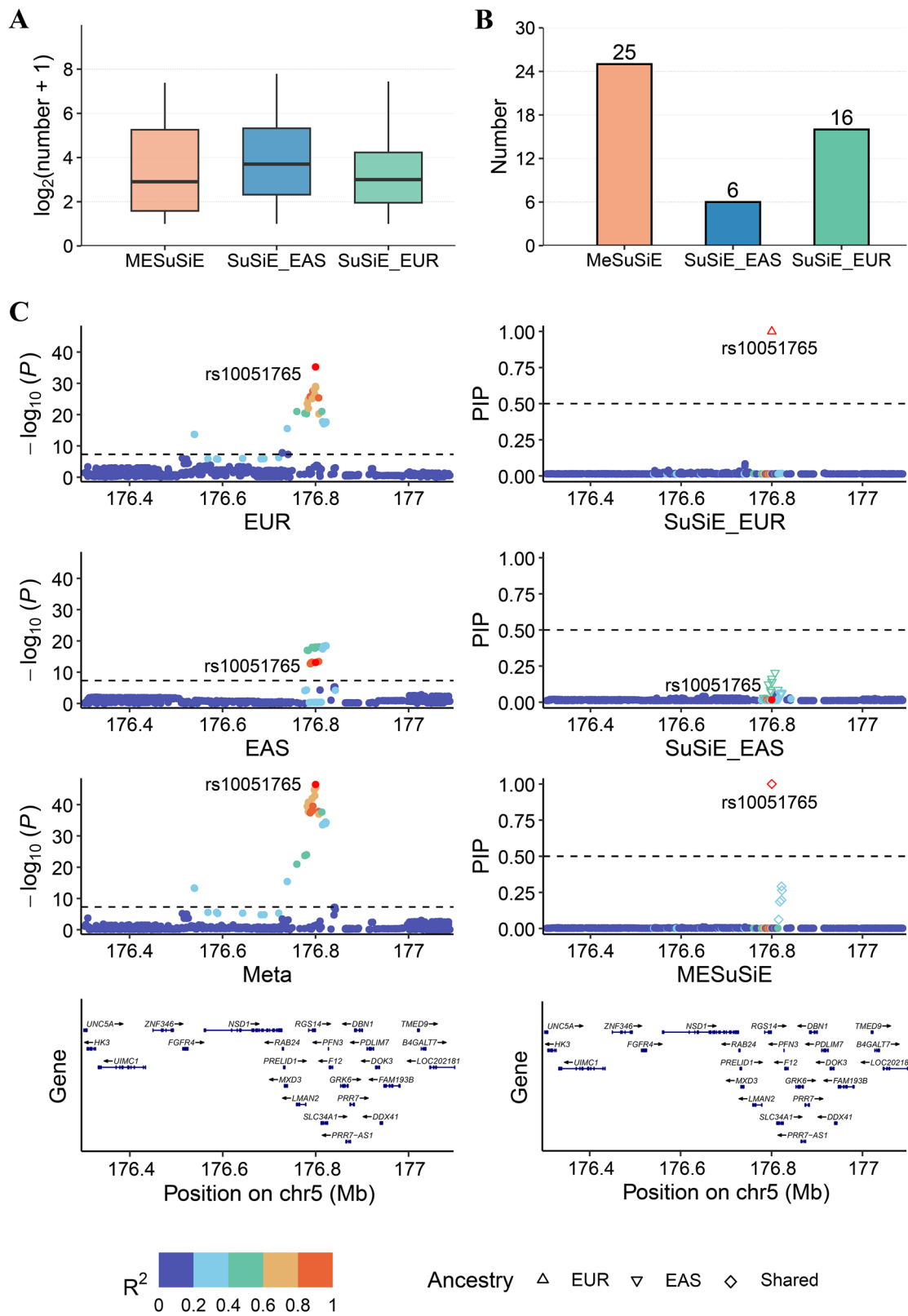
High PRS was associated with early onset age of KSD

To further explore the application of PRS in predicting KSD, we assessed the association between the onset age of KSD and PRSs. We observed that PRS-CS_{EUR}, PRS-CT_{EAS&EUR}, and PRS-CS_{EAS & EUR} were negatively associated with the onset age of KSD, and the change in onset age caused by PRS-CS_{EAS&EUR} was the largest for each standard deviation increment (Supplementary Data 10). After grouping these patients with the quintiles of each PRS, we observed significant results only for the two cross-population PRSs. Compared with patients of Q1 group, those of Q5 group identified by PRS-CT_{EAS&EUR} and PRS-CS_{EAS&EUR} showed an earlier onset age of KSD, respectively (Supplementary Data 10). Trend tests indicated that the decreasing trends of onset age with the increase of PRS-CT_{EAS&EUR} and PRS-CS_{EAS&EUR} were statistically significant ($P_{\text{trend}} = 0.005$ and 0.020 , respectively).

Discussion

In this study, we performed a very large-scale trans-ancestry GWAS meta-analysis and identified 59 susceptibility loci for KSD, including 13 novel loci. Through cross-population fine-mapping, we identified 25 causal signals at these loci. We pinpointed the *TRIOBP* gene at a novel locus associated with KSD. Moreover, we constructed a cross-population PRS-CS_{EAS&EUR} to improve the predictive performance and discrimination ability for KSD risk. Collectively, our study highlights the benefits of integrating genetic information from multiple populations for KSD genetic architecture research.

By increasing the sample size to 31,715 KSD cases and 943,655 controls, including populations of EUR and EAS ancestry, we improved the power of detecting susceptibility loci for KSD. In addition to repeating previously reported loci, we identified 13 novel loci. These novel susceptibility loci might be associated with KSD through multiple mechanisms, highlighting the diversity of genetic pathways. For example, the *SLC17A3* gene has been reported to be significantly associated with uric acid concentration and gout²⁹, and our study provided the direct evidence of its association with KSD. Gout is a recognized risk factor for stone formation, with exceeded uric acid concentrations in the urine of gout patients can lead to crystallization and the formation of uric acid stones^{30,31}. A mendelian randomization analysis, using genetic variants associated with systolic blood pressure at the *SLC12A3* gene and its regulatory regions as instrumental variable, showed that the genetic proxies of thiazide diuretics increase serum calcium while also reducing the KSD risk³². This study indirectly supported the association between *SLC12A3* and KSD, and suggested a possible pathway linking them. Additionally, a previous study reported that knockout of the *CDK12* gene leads to downregulation of Na-K-2Cl cotransporter 2 (NCKK2) in mouse kidneys, resulting in increased urine volume³³. However, more experiments and statistics are needed to verify the actual pathologies for them. Most identified KSD loci were observed to have similar effect sizes and consistent effect directions in



EUR and EAS population samples, which contained numerous shared causal signals detected by fine-mapping. Building on previous studies, our results further supported the similar KSD genetic architecture between EUR and EAS populations.

Cross-population fine-mapping identified more causal signals shared across populations and identified 95% credible sets in more genomic regions. Through functional annotation for fine-mapped

signals identified by MESuSiE, two missense variants, rs1260326 and rs2231142, were caught for their potential causal effects on KSD and deleterious effects according to CADD scores. The rs1260326 was also reported in a previous KSD GWAS as the lead SNP for *GCKR*, which was thought to promote KSD formation by regulating of metabolic traits³⁴. Notably, a recent KSD GWAS reported another variant rs2199936 at *ABCG2*, and suggested the need for further studies on *ABCG2* variants,

Fig. 3 | Fine-mapping results for KSD by using MeSuSiE and SuSiE. A Box plot of SNP number of credible sets identified by MeSuSiE and SuSiE. The central line, bounds of box and whiskers represented to the median value, quartiles and the minima and maxima, respectively. Source data are provided as a Source Data file. **B** Bar plot of the number of causal signals with PIP > 0.5 detected by MeSuSiE and SuSiE. Source data are provided as a Source Data file. **C** LocusZoom plot displayed the fine-mapping results on KSD from different methods in a genomic region on chromosome 5. The left column showed Manhattan plots of EUR, EAS and trans-ancestry GWAS meta-analyses results in this region. The genome-wide significance

threshold was 5×10^{-8} , indicated by the black dashed lines. The most significant signal in EUR and EAS populations was labeled. The right column showed LocusZoom plots, displaying the PIP of signals calculated by SuSiE in EUR and EAS populations and by MeSuSiE. The last row showed annotated genes in this genomic region. For the signals in the detected credible sets, we used an upper triangle to represent a European-specific signal, a lower triangle to represent an East Asian ancestry-specific signal and a diamond to represent a shared signal. PIP of 0.5 indicated a potential causal signal, whether shared or ancestry-specific. PIP posterior inclusion probability, EUR European, EAS East Asian.

given the known role of *ABCG* rs2231142 variant in regulating renal urate excretion^{35,36}. Our analysis successfully pinpointed rs2231142 as a key fine-mapped signal at *ABCG2*, demonstrating its significance and reliability. Interestingly, we observed strong support for the association between *TRIOBP* gene and KSD, although the lead SNP rs6000890 at this novel locus was annotated to *HI-O* gene in ANNOVAR due to the closer distance. The *TRIOBP* gene (also known as Tara) encodes Trio and F-actin binding protein³⁷. Trio protein is known to activate Rac1, a member of Rho family GTPases³⁸. Rac1 is highly expressed in the renal cortex and plays important roles in the structure and function of various kidney cells, including tubular epithelial cells, mesangial cells and podocytes³⁹. Its overactivation has been strongly linked to various nephropathies³⁹. The Trio and F-actin binding protein can bind to Trio and inhibit Rac1 signaling⁴⁰, which might be the mechanism linking *TRIOBP* gene to KSD. Of course, further studies are needed to fully elucidate the underlying genetic pathways. The risk allele T of rs6000890 was found to be associated with increased *TRIOBP* expression in the kidney-specific meta-eQTL analysis by Liu et al.²⁵ consistent with its protective effect against KSD observed in our study.

The enrichment of annotated genes for fine-mapped variants in the proximal tubule open chromatin regions supported the important role of proximal tubules in KSD development, which has been detailed elsewhere⁸. Additionally, GO enrichment analysis suggested the association between vitamin D-related biological processes and KSD. Among the overlapped genes (Supplementary Data 7), *ALPL* and *PTGS2* are strongly associated with multiple bone metabolic processes^{41–43}. *CASR* plays a pivotal role in systemic calcium metabolism by regulating parathyroid hormone secretion and urinary calcium excretion⁴⁴. *CYP24A1*, encoding vitamin D-24-hydroxylase, regulates the catabolism of 1,25-(OH)₂D⁴⁵. Therefore, we believed that this link is likely mediated by bone metabolism and the calcium-parathyroid hormone-vitamin D axis⁴⁶.

In recent years, genetically-based disease prediction has become increasingly widespread. PRS can quantify an individual's susceptibility to certain diseases, offering significant potential for early screening and prevention. Rely on the growing number of KSD-GWAS in different populations and ongoing advancements in PRS construction methods, we constructed the cross-population PRS-CS_{EAS&EUR} for KSD. This PRS benefited from larger discovery sample size and integrated genetic information from multiple populations. Since UKB individuals were required for PRS construction and testing, we removed the UKB GWAS summary statistics from the discovery dataset. Although this might lead to some loss of statistical power due to the reduced discovery sample size, we believed that it could effectively avoid potential bias of the prediction results caused by the kinship or overlap between the target and discovery samples. Moreover, the use of large-scale validation and testing datasets also contributed to the robustness of our prediction results. Our study clearly illustrated that the cross-population PRS-CS_{EAS&EUR} was able to exhibit superior predictive performance and discrimination ability for KSD risk by incorporating genetic information from EAS population. On the other hand, it implied the application potential of KSD-PRS in EAS population, highlighting the importance of reducing disparities in access to genetic resources for non-European populations.

There are some limitations in this study. First, for the EAS population samples, we used an external LD reference in both cross-population and intra-population fine-mappings, although we performed recommended cleansing to address LD mismatched variants before conducting the analyses. Second, considering the differences in baseline age and gender structure might exist between the GWAS discovery sample for PRS derivation and the target sample for PRS validation and testing, as well as the different environmental conditions, the accuracy and applicability of PRS performance might be affected. Third, due to data limitations, the cross-population PRS-CS_{EAS&EUR} only included population samples of EAS and EUR ancestry. Its generalization in other populations requires to be supplemented by additional GWASs. Additionally, we only evaluated the performance of PRS models in EUR population samples. The transferability of PRS in EAS populations requires further investigation. Finally, the inclusion of GWAS summary statistics from multiple countries may potentially affect diagnoses and lead to classification bias between cases and controls, considering the heterogeneity in different health systems.

Methods

Ethics

We obtained genotype and phenotype data from the UKB, which received ethics approval from the North West Centre for Research Ethics Committee. The application number of UKB in our study was 88159.

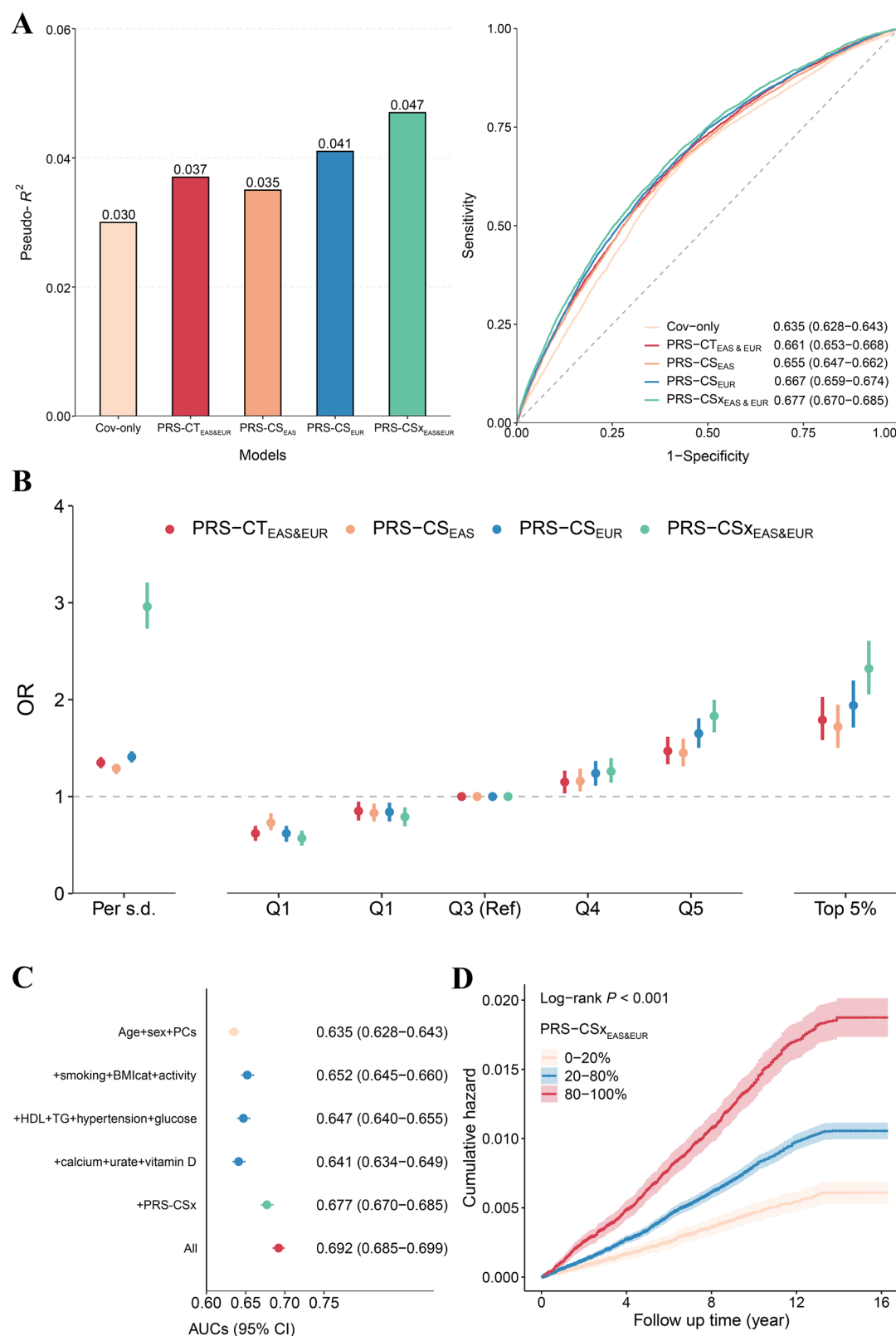
GWAS summary data source

The study included several GWAS summary statistics for KSD from EUR and EAS populations. The GWAS summary statistics in EUR populations was obtained in our previous GWAS meta-analysis for KSD⁸, which included 9372 KSD cases and 369,102 controls from UK Biobank (UKB)¹⁴ and 8597 cases with 333,128 controls FinnGen¹⁵. Additionally, we included EAS GWASs from China Kadoorie Biobank (CKB)¹⁶ with 2047 cases and 74,398 controls and Biobank Japan (BBJ)¹⁷ with 11,699 cases and 167,027 controls. More details of these GWAS analyses were described in Supplementary Methods. Each study obtained informed consent from participants and approval from the relevant institutional review board.

GWAS meta-analysis

GWAS meta-analyses were conducted using a fixed-effect inverse-variance weighted model with METAL⁴⁷. Firstly, we performed an EAS-specific GWAS meta-analysis combining summary data from BBJ and CKB. Subsequently, by integrating previously conducted KSD GWAS summary data of EUR populations, we conducted the trans-ancestry GWAS meta-analysis using METAL. Genomic control corrections were applied to both input files and results of these meta-analyses, and heterogeneity tests were conducted to evaluate heterogeneities between studies. We excluded the variants exclusive to single dataset from the EAS-specific and trans-ancestry meta-analyses, respectively.

To further assess the potential impact of heterogeneity across ancestries, including intra-population studies, we utilized MR-MEGA⁴⁸ to conduct an additional trans-ancestry GWAS meta-analysis. MR-MEGA uses trans-ethnic meta-regression to account for heterogeneity



in allelic effects correlated with ancestry⁴⁸. The results from METAL and MR-MEGA analyses were compared to assess their robustness.

Susceptibility loci identification

KSD associated variants were defined by genome-wide significance level (5×10^{-8}). We then performed linkage disequilibrium (LD) clumping for these variants with PLINK v.1.9⁴⁹ (command: --clump -p1

5e-8 --clump-p2 5e-8 --clump-r2 0.1 --clump-kb 10000) and merged those loci with lead SNPs within 1000 kb of each other to obtain the final independently significant loci. For EAS-specific GWAS meta-analysis, we used the LD reference panel of EAS individuals from the 1000 genome project phases 3 (1KGP3)⁵⁰. For the trans-ancestry meta-analysis, given that individuals of European ancestry made up about three-quarters of the total sample size, we used LD reference panel of 5000

Fig. 4 | Trans-ancestry PRS-CSx improved predictive performance and discrimination ability for KSD risk. **A** Comparison of pseudo- R^2 and AUCs for cov-only model and four PRS models. **B** Comparison of KSD risk by four PRSs. The central points and error bars represented ORs and the 95% confidence intervals, respectively. Source data are provided as a Source Data file. **C** Predictive performance of PRS in relation to known non-genetic KSD related factors. The central points and error bars represented AUC values and 95% confidence intervals, respectively. Source data are provided as a Source Data file. **D** Cumulative hazard

curve of KSD according to stratified PRS-CSx_{EAS&EUR}. The models constructed for Figure A and Figure B were adjusted for age, sex, and PC1-10, with an additional adjustment for non-genetic KSD-related factors for Figure D. All analyses were completed in testing dataset, including 4996 KSD cases and 199,409 controls. AUC area under receiver operating curve, cov-only covariates-only prediction model, OR odd ratio, HDL high-density lipoprotein, TG triglycerides. The glucose referred to the fasting glucose, and the calcium, urate, and vitamin D referred to the blood calcium, urate and vitamin D concentrations.

Table. 2 | Comparison of KSD risk by different PRSs

PRS	Continuous	Q1	Q2	Q3	Q4	Q5	Top 5%
PRS-CT_{EAS&EUR}							
Case/Control	–	619/ 40,262	839/ 40,042	985/39,896	1127/39,754	1426/39,455	433/9788
OR (95% CI)	1.35 (1.31–1.39)	0.62 (0.56–0.68)	0.85 (0.77–0.93)	1.00 (Ref)	1.15 (1.05–1.25)	1.47 (1.35–1.60)	1.79 (1.60–2.01)
P value	2.10×10 ^{−94}	2.17 × 10 ^{−20}	4.65 × 10 ^{−4}		1.57 × 10 ^{−3}	6.06 × 10 ^{−20}	3.96 × 10 ^{−23}
PRS-CS_{EAS}							
Case/Control	–	718/40,163	811/40,070	968/39,913	1117/39,764	1382/39,499	407/9814
OR (95% CI)	1.29 (1.25–1.32)	0.73 (0.67–0.81)	0.83 (0.76–0.91)	1.00 (Ref)	1.16 (1.07–1.27)	1.45 (1.33–1.58)	1.72 (1.52–1.93)
P value	9.35 × 10 ^{−70}	6.10 × 10 ^{−10}	1.35 × 10 ^{−4}		7.08 × 10 ^{−4}	3.65 × 10 ^{−18}	3.55 × 10 ^{−19}
PRS-CS_{EUR}							
Case/Control	–	588/40,293	791/40,091	940/39,940	1154/39,727	1523/39,358	440/9871
OR (95% CI)	1.41 (1.37–1.45)	0.62 (0.55–0.68)	0.84 (0.76–0.92)	1.00 (Ref)	1.24 (1.13–1.35)	1.65 (1.52–1.79)	1.94 (1.73–2.18)
P value	6.25 × 10 ^{−125}	6.80 × 10 ^{−20}	2.58 × 10 ^{−4}		2.04 × 10 ^{−6}	2.29 × 10 ^{−32}	4.20 × 10 ^{−29}
PRS-CSx_{EAS&EUR}							
Case/Control	–	534/40,347	731/40,150	924/39,957	1159/39,722	1648/39,233	514/9707
OR (95% CI)	2.96 (2.75–3.19)	0.57 (0.51–0.63)	0.79 (0.71–0.87)	1.00 (Ref)	1.26 (1.16–1.38)	1.83 (1.68–1.98)	2.32 (2.07–2.59)
P value	3.16 × 10 ^{−179}	7.39 × 10 ^{−25}	1.72 × 10 ^{−6}		2.19 × 10 ^{−7}	5.99 × 10 ^{−47}	4.22 × 10 ^{−50}

PRS polygenic risk score, OR odd ratio, CI confidence interval.
HRs and 95% CIs were calculated with the use of the Logistic regression model.
All models were adjusted for age- sex- and PC1-10.

random selected participants of European ancestry in UKB. The final lead variants from these independent loci associated with KSD were annotated with the closest gene and functional regions using ANNOVAR⁵¹. Furthermore, we compared the effect sizes and directions of these variants across different populations, assessing their consistency using Lin's concordance correlation coefficients (ρ_c).

To define the novel loci, we downloaded the known loci associated with KSD from the GWAS Catalog database⁵² by December 2023 and a recently published study⁸ not yet included in the GWAS Catalog (Supplementary Data 11). If necessary, the variant positions were flipped to GRCh37 using LiftOver⁵³ to align with our results. Loci identified in this study were classified as novel if they were located more than 1000 kb away from previously reported loci.

Replication in the Million Veteran Program

The Million Veteran Program (MVP) is a large-scale observational cohort study and mega-biobank within the American Department of Veterans Affairs health care system⁵⁴. Recently, MVP PheWeb (<https://phenomics.va.gov/pheweb>) published results from GWASs across 2068 traits in AFR, AMR, EAS, and EUR reference populations, as well as cross-population GWAS meta-analyses, which allows us to perform an external validation of the novel loci identified by our study¹⁸. The meta-analysis for KSD in MVP included 42,198 cases, defined as calculus of kidney (phecode 594.1), and 618,588 controls. Building on that, we conducted a lookup study for the lead variants from these loci within MVP and compared their effect sizes between the two studies.

Fine-mapping

MeSuSiE⁵⁵ and SuSiE⁵⁶ were respectively used to perform cross-population and intra-population fine-mapping for genomic regions surrounding lead SNPs (500 kb upstream and downstream) of identified KSD susceptibility loci. Before running fine mapping, we performed a series of recommended data preparation⁵⁵. We derived EAS and EUR summary statistics from EAS-specific GWAS meta-analysis and our previous KSD GWAS⁸. The relative ancestry-specific SNP-SNP correlation matrices were calculated using reference panels from 5000 EUR participants and 2594 EAS participants in UKB, respectively. After excluding SNPs with MAF < 0.001, strand ambiguous or multi-allelic and in the major histocompatibility complex regions (chr6: 25–34 Mb), only the SNPs that appeared in both the EAS and EUR GWAS summary statistics and ancestry-specific SNP correlation matrices were finally remained. Therefore, two loci of lead SNPs in major histocompatibility complex regions were not included in fine-mapping. Additionally, LD mismatch variants were examined and excluded before fine-mapping analysis of each genomic region as recommended by the MESuSiE guideline (Supplementary Methods), which is described in detail elsewhere (https://borangao.github.io/meSuSie_Analysis).

By calculating the posterior inclusion probability (PIP) for the remaining SNPs using MeSuSiE and SuSiE, we identified potential causal effect signals with a PIP threshold of 0.5, as recommended in the MESuSiE method guideline. They also constructed 95% credible sets in each genomic region, where the sum of PIPs for the variants exceeded 0.95. The 95% credible sets constructed by MESuSiE comprised SNPs with a non-zero effect in at least one ancestry. These SNPs were further

categorized as either “shared” or “ancestry-specific” based on their contribution to the credible set. SuSiE identifies population-specific signals by fine-mapping variants within each population. If a causal signal was detected in both populations, we considered it a shared signal identified by SuSiE. We used a two-side Kruskal-Wallis test to compare the sizes of the credible sets constructed by MeSuSiE and SuSiE, which was represented by the number of variants contained in each credible set. Additionally, regarding variants with a PIP > 0.5 or included in credible sets identified by MeSuSiE as fine-mapped variants, we also compared their genetic effects on KSD between EAS and EUR populations.

Function inferences

We made function inferences for the fine-mapped variants identified by MeSuSiE from several aspects. First, we performed functional annotation for them using VEP v.113³⁷ (CRCh37) with the “-everything” flag, integrating CADD scores through an additional plugin for CADD²⁸ to evaluate their deleteriousness. These variants were also annotated to significant eQTL genes based on a previous human kidney-specific meta-eQTL study²⁵. Among these fine-mapped signals, missense variants with a PIP > 0.5 or mapped to a small credible set (number of SNPs < 5) were specifically emphasized. We also annotated the unmapped variants within the 57 genomic regions used for fine-mapping and compared their CADD scores (categorized with CADD scores of 10 and 20, respectively) and eQTL signal overlaps with those of the fine-mapped signals using two-side Chi-square tests. Additionally, we explored the distribution of fine-mapped variants in different cell type-specific chromatin accessible regions identified by human kidney single-nucleus ATAC-seq (snATAC-seq) data from Sheng et al. (gene expression omnibus accession number GSE172008)³⁸. For genes annotated to fine-mapped signals in VEP, we used FUMA platform (v.1.5.2)⁵⁹ to evaluate tissue specificity across 54 tissues from the GTEx v.8 project⁶⁰ and to perform Gene Ontology (GO) enrichment analysis based on Molecular signatures database (MSigDB)⁶¹. The input genes were tested against pre-defined differential expression gene (DEG) sets using the hypergeometric test on the platform. Bonferroni corrections were applied to the analyses, and the adjusted *P*-value threshold of 0.05 was used to define significance.

PRS construction

PRS application cohorts. We extracted 408,8019 individuals of European ancestry who self-identified as white British and had very similar genetic ancestry based on the PCA of the genotypes from UKB, serving as the target sample for PRS validation and test⁸. Among them, 10,017 KSD cases were defined as calculus of kidney and ureter (field ID 132036, ICD-10 N20), and the 398,792 remaining individuals were controls. The diagnosis was obtained by using linkage with the death register, primary care, hospital inpatient records, and medical conditions self-reported at the baseline or subsequent UKB assessment center. To avoid the relatedness or overlap between individuals of discovery sample and target sample, we derived PRSs for KSD from the BBJ, CKB and FinnGen cohorts. Specifically, we used FinnGen to represent EUR populations and the meta-analysis of CKB and BBJ to represent EAS populations in the discovery sample. A trans-ancestry meta-analysis without UKB was conducted using METAL on the discovery sample for the subsequent PRS construction.

PRS derivation and parameter tuning. We considered several types of KSD-PRS as following:

Single-population PRS-CS_{EUR} and PRS-CS_{EAS} were constructed using the PRS-CS method, with GWAS summary data from EUR and EAS populations in discovery sample, respectively⁶².

Cross-population PRS-CT_{EAS&EUR} was constructed using the clumping and thresholding (C + T) method, with GWAS summary data from the trans-ancestry meta-analysis without UKB.

Cross-population PRS-CSx_{EAS&EUR} was constructed using the PRS-CSx method, with GWAS summary data from EUR and EAS populations in discovery sample⁶³.

The LD reference panels were obtained from EAS and EUR individuals from 1KGP3, respectively. Final PRSs for each individual were calculated using PLINK2 for different PRS methods, as the weighted sum of risk alleles carried by the variants used in each method⁶⁴. The weights corresponded to the effect sizes of variants for the C + T method and the posterior effect values of variants estimated by the PRS-CS or PRS-CSx methods.

Target sample individuals were randomly divided into training dataset (*N* = 204,404) with 5021 cases and testing dataset (*N* = 204,405) with 4996 cases. Parameter tuning was performed in training dataset. For PRS-CS method, we considered the global shrinkage parameter Φ , using grid research and the PRS-CS-auto method (a fully Bayesian approach that enables automatic learning of Φ from GWAS summary statistics), as 1, 10^{-2} , 10^{-4} , 10^{-6} and *auto*. For C + T method, we considered the LD panels for clumping (LD panels from 1KG EAS and EUR individuals) and the *P* value thresholds P_t (5×10^{-8} , 1×10^{-6} , 1×10^{-4} , 1×10^{-3} , 0.01, 0.1, 0.2, 0.5 and 1). For PRS-CSx method, in addition to the global shrinkage parameter Φ (1, 10^{-2} , 10^{-4} , 10^{-6} and PRS-CSx-*auto*) like PRS-CS method, we considered the weights w , corresponding to a specific Φ , of the PRS of EAS and EUR populations through the following formula:

$$y \sim w_{\Phi, EAS} PRS_{\Phi, EAS} + w_{\Phi, EUR} PRS_{\Phi, EUR} \quad (1)$$

where *y* is the KSD indicator (0 for control and 1 for case), with a specific Φ , the PRS_{EAS} and PRS_{EUR} are the standardized PRS constructed with PRS-CSx method for EAS and EUR populations, and w_{EAS} and w_{EUR} are the corresponding regression coefficients for them. We performed logistic regression of KSD on each PRS, adjusting sex, age, and PC1-10 as covariates. Through comparing the pseudo-*R*² values and AUCs of different prediction models, we selected the optimum parameters for each PRS construction method. Additionally, the final PRS-CSx_{EAS&EUR} was derived by the following formula:

$$PRS - CSx_{EAS\&EUR} = \hat{w}_{EAS} PRS_{EAS} + \hat{w}_{EUR} PRS_{EUR} \quad (2)$$

where the PRS_{EAS} and PRS_{EUR} are the standardized PRS for EAS and EUR populations on the optimum Φ , \hat{w}_{EAS} and \hat{w}_{EUR} were the estimations of *w* for them.

For the single-population PRS, we selected a global shrinkage parameter $\Phi = 10^{-6}$ for PRS-CS_{EAS} and used PRS-CS-auto method for PRS-CS_{EUR}. For PRS-CT, we applied $r^2 = 0.1$ to perform LD clumping with LD panels from 1KG EAS individuals and selected P_t of 10^{-6} for summary statistics from the non-UKB trans-ancestry meta-analysis. Lastly, we constructed PRS-CSx_{EAS&EUR} using the PRS-CSx-auto method, integrating EAS and EUR summary statistics, with the corresponding weight estimates being $\hat{w}_{EAS} = 0.169$ and $\hat{w}_{EUR} = 0.290$, respectively (Supplementary Data 12, Supplementary Data 13).

PRS predictive performance evaluation. The final PRSs for the testing dataset participants were constructed based on corresponding optimum parameters for each method selected from the training dataset. In the testing dataset, we compared the predictive performances of these PRSs in several ways. Firstly, we compared the pseudo-*R*² values and AUCs of the covariates-only model and four covariates-adjusted PRS prediction models. Covariates included age, sex and PC1-10. Delong tests were conducted to compare the AUC of the PRS-CSx model with those of the other four models. Then, the adjusted odd ratio (OR) and 95% confidence interval (CI) of KSD caused by each PRS per standard deviation were calculated through logistic regression model. We assessed whether non-linear associations existed between these PRSs and KSD risk using the restricted cubic spline

(RCS) method. After that, we grouped testing dataset based on the quintiles of each PRS (Q1–Q5 with PRS from low to high). Using Q3 group as reference, the ORs of KSD were calculated for participants in remaining four groups and additionally for those with the top 5% PRS.

To further assess the stability of our findings and the performance of PRS-CS_{EAS&EUR}, we included some known KSD related factors from UKB for target sample individuals (details in Supplementary Methods). We compared the discrimination ability of PRS-CS_{EAS&EUR} relative to these non-genetic factors for KSD risk using AUC, by adding them iteratively to a basic prediction model comprising age, sex and PC1–10. We also calculated the cNRI to assess the reclassification improvement for the prediction model with PRS-CS_{EAS&EUR}^{65,66}. In addition, based on survival analysis, we constructed the cumulative hazard curve for incident KSD over the follow-up time (median follow-up of 13.4 years), with stratification by PRS-CS_{EAS&EUR} (0–20%, 20–80%, and 80–100%) and adjustment for age, sex, PC1–10 and all non-genetic KSD related factors obtained.

Association between onset age of KSD and PRS. We extracted all 10,017 KSD cases of European ancestry in UKB. We excluded individuals with onset ages identified as outliers based on a 3-fold median absolute deviation (MAD) range, leaving 9533 individuals in the final sample. Then we employed multiple linear regression model to evaluate the association between the onset age of KSD and each PRS, adjusting for sex, PC1–10, and all nongenetic KSD-related factors obtained. We used each standardized PRS as both continuous variable and categorical variable for regression analyses, and additionally performed trend tests. The categorized PRSs were also derived based on their quintiles (Q1–Q5 with PRSs from low to high).

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The trans-ancestry GWAS summary statistics in METAL generated in this study have been deposited in the Zenodo database under accession code [14790324](https://doi.org/10.5281/zenodo.14790324). As described in Supplementary Methods, the original KSD GWAS summary statistics from BBJ, CKB, and FinnGen cohorts used in this study are publicly available online (BBJ: <https://pheweb.jp>; CKB: <https://pheweb.ckbiobank.org>; FinnGen: <https://r8.finnngen.fi>). The UKB GWAS summary statistics and EUR-specific GWAS meta-analysis summary statistics used in this study are also publicly available in the Zenodo database [<https://zenodo.org/records/10060271>]. Source data are provided with this paper. The source data for Fig. 4A are not included, as they are described in Supplementary Data 9. The source data for Supplementary Fig. 10 are not included neither, as the figure was directly generated using the FUMA platform (<https://fuma.ctglab.nl/>).

References

- Cappuccio, F. P. & MacGregor, G. A. The pathogenesis and treatment of kidney stones. *N. Engl. J. Med.* **328**, 444 (1993).
- Siener, R. Nutrition and kidney stone disease. *Nutrients* **13**, 1917 (2021).
- Singh, P., Harris, P. C., Sas, D. J. & Lieske, J. C. The genetics of kidney stone disease and nephrocalcinosis. *Nat. Rev. Nephrol.* **18**, 224–240 (2022).
- Goldfarb, D. S., Avery, A. R., Beara-Lasic, L., Duncan, G. E. & Goldberg, J. A Twin Study of Genetic Influences on Nephrolithiasis in Women and Men. *Kidney Int. Rep.* **4**, 535–540 (2019).
- Oddsson, A. et al. Common and rare variants associated with kidney stones and biochemical traits. *Nat. Commun.* **6**, 7975 (2015).
- Urabe, Y. et al. A genome-wide association study of nephrolithiasis in the Japanese population identifies novel susceptible Loci at 5q35.3, 7p14.3, and 13q14.1. *PLoS Genet.* **8**, e1002541 (2012).
- Howles, S. A. et al. Genetic variants of calcium and vitamin D metabolism in kidney stone disease. *Nat. Commun.* **10**, 5175 (2019).
- Hao, X. et al. Integrative genome-wide analyses identify novel loci associated with kidney stones and provide insights into its genetic architecture. *Nat. Commun.* **14**, 7498 (2023).
- Torkamani, A., Wineinger, N. E. & Topol, E. J. The personal and clinical utility of polygenic risk scores. *Nat. Rev. Genet.* **19**, 581–590 (2018).
- Jung, H. et al. Integration of risk factor polygenic risk score with disease polygenic risk score for disease prediction. *Commun. Biol.* **7**, 180 (2024).
- Paranjpe, I. et al. Derivation and validation of genome-wide polygenic score for urinary tract stone diagnosis. *Kidney Int.* **98**, 1323–1330 (2020).
- Privé, F. et al. Portability of 245 polygenic scores when derived from the UK Biobank and applied to 9 ancestry groups from the same cohort. *Am. J. Hum. Genet.* **109**, 12–23 (2022).
- Tanigawa, Y. et al. Significant sparse polygenic risk scores across 813 traits in UK Biobank. *PLoS Genet.* **18**, e1010105 (2022).
- Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
- Kurki, M. I. et al. FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature* **613**, 508–518 (2023).
- Walters, R. G. et al. Genotyping and population characteristics of the China Kadoorie Biobank. *Cell Genom.* **3**, 100361 (2023).
- Sakaue, S. et al. A cross-population atlas of genetic associations for 220 human phenotypes. *Nat. Genet.* **53**, 1415–1424 (2021).
- Verma, A. et al. Diversity and scale: Genetic architecture of 2068 traits in the VA Million Veteran Program. *Science* **385**, eadj1182 (2024).
- Jutabha, P. et al. Human sodium phosphate transporter 4 (hNPT4/SLC17A3) as a common renal secretory pathway for drugs and urate. *J. Biol. Chem.* **285**, 35123–35132 (2010).
- Lin, Z., Jayachandran, M., Haskic, Z., Kumar, S. & Lieske, J. C. Differences of uric acid transporters carrying extracellular vesicles in the urine from uric acid and calcium stone formers and non-stone formers. *Int. J. Mol. Sci.* **23**, 10010 (2022).
- Obermüller, N. et al. Expression of the thiazide-sensitive Na-Cl cotransporter in rat and human kidney. *Am. J. Physiol.* **269**, F900–F910 (1995).
- Lin, B. M. et al. Whole genome sequence analyses of eGFR in 23,732 people representing multiple ancestries in the NHLBI trans-omics for precision medicine (TOPMed) consortium. *EBioMedicine* **63**, 103157 (2021).
- Mahajan, A. et al. Trans-ethnic fine mapping highlights kidney-function genes linked to salt sensitivity. *Am. J. Hum. Genet.* **99**, 636–646 (2016).
- Qiu, C. et al. Renal compartment-specific genetic variation analyses identify new pathways in chronic kidney disease. *Nat. Med.* **24**, 1721–1731 (2018).
- Liu, H. et al. Epigenomic and transcriptomic analyses define core cell types, genes and targetable mechanisms for kidney disease. *Nat. Genet.* **54**, 950–962 (2022).
- Gattineni, J. et al. Regulation of renal phosphate transport by FGF23 is mediated by FGFR1 and FGFR4. *Am. J. Physiol. Ren. Physiol.* **306**, F351–F358 (2014).
- Chen, W. C. et al. The rs1256328 (ALPL) and rs12654812 (RGS14) polymorphisms are associated with susceptibility to calcium nephrolithiasis in a Taiwanese population. *Sci. Rep.* **9**, 17296 (2019).
- Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J. & Kircher, M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res.* **47**, D886–d894 (2019).

29. Dehghan, A. et al. Association of three genetic loci with uric acid concentration and risk of gout: a genome-wide association study. *Lancet* **372**, 1953–1961 (2008).
30. Kramer, H. J., Choi, H. K., Atkinson, K., Stampfer, M. & Curhan, G. C. The association between gout and nephrolithiasis in men: The Health Professionals' Follow-Up Study. *Kidney Int.* **64**, 1022–1026 (2003).
31. Ferraro, P. M. & Curhan, G. C. Serum uric acid and risk of kidney stones. *Am. J. Kidney Dis.* **70**, 158–159 (2017).
32. Triozi, J. L. et al. Mendelian randomization analysis of genetic proxies of Thiazide diuretics and the reduction of kidney stone risk. *JAMA Netw. Open* **6**, e2343290 (2023).
33. Wang, B. et al. Tubular-specific CDK12 knockout causes a defect in urine concentration due to premature cleavage of the *slc12a1* gene. *Mol. Ther.* **30**, 3300–3312 (2022).
34. Tanikawa, C. et al. Novel risk Loci Identified in a Genome-Wide Association Study of Urolithiasis in a Japanese population. *J. Am. Soc. Nephrol.* **30**, 855–864 (2019).
35. Chen, W. C. et al. Identification of novel genetic susceptibility loci for calcium-containing kidney stone disease by genome-wide association study and polygenic risk score in a Taiwanese population. *Urolithiasis* **52**, 94 (2024).
36. Woodward, O. M. et al. Identification of a urate transporter, ABCG2, with a common functional polymorphism causing gout. *Proc. Natl Acad. Sci. USA* **106**, 10338–10342 (2009).
37. Seipel, K., O'Brien, S. P., Iannotti, E., Medley, Q. G. & Streuli, M. Tara, a novel F-actin binding protein, associates with the Trio guanine nucleotide exchange factor and regulates actin cytoskeletal organization. *J. Cell Sci.* **114**, 389–399 (2001).
38. Maier, M., Baldwin, C., Aoudjit, L. & Takano, T. The role of Trio, a Rho Guanine nucleotide exchange factor, in glomerular podocytes. *Int. J. Mol. Sci.* **19**, 479 (2018).
39. Steichen, C., Hervé, C., Hauet, T. & Bourmeyster, N. Rho GTPases in kidney physiology and diseases. *Small GTPases* **13**, 141–161 (2022).
40. Yano, T. et al. Tara up-regulates E-cadherin transcription by binding to the Trio RhoGEF and inhibiting Rac signaling. *J. Cell Biol.* **193**, 319–332 (2011).
41. Tournis, S., Yavropoulou, M. P., Polyzos, S. A. & Doulgeraki, A. Hypophosphatasia. *J. Clin. Med.* **10**, 5676 (2021).
42. Raisz, L. G. Potential impact of selective cyclooxygenase-2 inhibitors on bone metabolism in health and disease. *Am. J. Med* **110**, 43s–45s (2001).
43. Okada, Y., Pilbeam, C., Raisz, L. & Tanaka, Y. Role of cyclooxygenase-2 in bone resorption. *J. UOEH* **25**, 185–195 (2003).
44. Vahe, C. et al. Diseases associated with calcium-sensing receptor. *Orphanet J. Rare Dis.* **12**, 19 (2017).
45. Carpenter, T. O. CYP24A1 loss of function: Clinical phenotype of monoallelic and biallelic mutations. *J. Steroid Biochem. Mol. Biol.* **173**, 337–340 (2017).
46. Goltzman, D., Mannstadt, M. & Marcocci, C. Physiology of the CALCIUM-PARATHYROID HORMONE-Vitamin D Axis. *Front Horm. Res.* **50**, 1–13 (2018).
47. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
48. Mägi, R. et al. Trans-ethnic meta-regression of genome-wide association studies accounting for ancestry increases power for discovery and improves fine-mapping resolution. *Hum. Mol. Genet.* **26**, 3639–3650 (2017).
49. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
50. Auton, A. et al. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
51. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
52. MacArthur, J. et al. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res.* **45**, D896–d901 (2017).
53. Genovese, G. et al. BCFtools/liftover: an accurate and comprehensive tool to convert genetic variants across genome assemblies. *Bioinformatics* **40**, beta038 (2024).
54. Gaziano, J. M. et al. Million Veteran Program: A mega-biobank to study genetic influences on health and disease. *J. Clin. Epidemiol.* **70**, 214–223 (2016).
55. Gao, B. & Zhou, X. MESuSiE enables scalable and powerful multi-ancestry fine-mapping of causal variants in genome-wide association studies. *Nat. Genet.* **56**, 170–179 (2024).
56. Zou, Y., Carbonetto, P., Wang, G. & Stephens, M. Fine-mapping from summary data with the “Sum of Single Effects” model. *PLoS Genet.* **18**, e1010299 (2022).
57. McLaren, W. et al. The ensembl variant effect predictor. *Genome Biol.* **17**, 122 (2016).
58. Sheng, X. et al. Mapping the genetic architecture of human traits to cell types in the kidney identifies mechanisms of disease and potential treatments. *Nat. Genet.* **53**, 1322–1333 (2021).
59. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).
60. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).
61. Liberzon, A. et al. Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739–1740 (2011).
62. Ge, T., Chen, C. Y., Ni, Y., Feng, Y. A. & Smoller, J. W. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat. Commun.* **10**, 1776 (2019).
63. Ruan, Y. et al. Improving polygenic prediction in ancestrally diverse populations. *Nat. Genet.* **54**, 573–580 (2022).
64. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
65. Pencina, M. J., D'Agostino, R. B. Sr, D'Agostino, R. B. Jr & Vasan, R. S. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat. Med.* **27**, 157–172 (2008).
66. Pencina, M. J., D'Agostino, R. B. Sr. & Steyerberg, E. W. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Stat. Med.* **30**, 11–21 (2011).

Acknowledgements

National Natural Science Foundation of China (32470658 to XJH, 82270803 to HX, 82273443 to MW), Department of Science and Technology of Hubei Province (2024AFA019 to HX). Additionally, we acknowledge the contribution of participants and investigators of the UK Biobank, FinnGen, China Kadoorie Biobank and Biobank Japan, contributing to the GWAS summary statistics in our analysis. The computation is completed in the HPC Platform of Huazhong University of Science and Technology and the School of Public Health, Huazhong University of Science and Technology.

Author contributions

Conceptualization: Xingjie Hao; Minghui Jiang; Xi Cao Methodology: Xingjie Hao; Minghui Jiang; Xi Cao; Zhonghe Shao; Yifan Kong; Si Li; Yunlong Guan; Hongji Wu; Yan Gong Investigation: Chen Duan; Xiangyang Yao; Bo li; Visualization: Xi Cao; Xingjie Hao Funding acquisition: Xingjie Hao; Hua Xu Project administration: Xingjie Hao; Hua Xu; Miao Wang Supervision: Xingjie Hao; Hua Xu Writing – original draft: Xi Cao Writing – review & editing: Xi Cao; Xingjie Hao; Minghui Jiang; Hua Xu;

Miao Wang; Zhonghe Shao; Yifan Kong; Si Li; Yunlong Guan; Hongji Wu; Chen Duan; Xiangyang Yao; Bo li; Yan Gong.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-58782-7>.

Correspondence and requests for materials should be addressed to Miao Wang, Hua Xu or Xingjie Hao.

Peer review information *Nature Communications* thanks Cristian Pat-taro and the other, anonymous, reviewers for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025