

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10
- 11
- 12
- 13
- 14
- 15
- 16
- 17
- 18

Jing Zhang, Yue Guo, Luyu Gong, Limei Xia, Qiaoqiao Liu, Kangchun Wang, Qi Wang,
Zhaojun Liu, Zhaohui Qin, Shaolin Shi, Jingping Yang
Corresponding author: shaolinshi1001@yahoo.com; jpyang@nju.edu.cn

Materials and Methods
Figs. S1 to S8
References

19 **Materials and Methods**

20 **GWAS data collection and multiple independent associations analysis**

21 GWAS datasets of serum urate, eGFR_{crea}, BUN, UACR, gout and CKD used in this study
22 are downloaded from CKDGen Consortium (<https://ckdgen.imbi.uni-freiburg.de>). GWAS
23 datasets of leptin measurement, crohn disease, multiple sclerosis and intelligence are
24 downloaded from GWAS Catalog ([http://ftp.ebi.ac.uk/pub/databases/gwas/](http://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/)
25 [summary_statistics/](http://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/)). To obtain the independent associations of GWAS loci, we used
26 FUMA (1) (<https://fuma.ctglab.nl/snp2gene>) with default parameters and r^2 threshold to
27 define independent significant SNPs is < 0.8 . LocusZoom plots of the interest regions were
28 visualized with LocusZoom (<http://locuszoom.org>).

29 **LDSC regression analysis**

30 We downloaded chromatin segmentation of different tissues from EpiMap (2), and active
31 enhancer regions were used as functional categories for LDSC (3). LDSC regression
32 analysis was performed with LDSC version 1.0.1 by following the tutorial
33 (<https://github.com/bulik/ldsc/wiki>). The results of LDSC regression coefficient Z-scores
34 were visualized with the R package pheatmap version 1.0.12.

35 **Locus-specific stratification strategy and functional variants identification**

36 Previous studies reported a total of 267 unique lead SNPs in three GWASs on serum urate
37 level (4, 5). For each locus within 500kb of the reported lead SNP, we quartered the
38 significance of variant-trait associations from highest to lowest. Variants with the top
39 quartile associations were extracted as rank 1 high-risk variants and were used for
40 subsequent analysis. Open chromatin regions of kidney cell types obtained from our kidney
41 scATAC-seq data for functional annotation. Functional SNPs were identified by
42 overlapping high-risk variants itself or its LD extension with regulatory elements of kidney
43 cell types. For those high-risk variants that did not overlap with regulatory elements by
44 itself, we performed LD extension of the variants. Proxy SNPs were extracted with $r^2 > 0.8$
45 using LDlinkR version 1.1.2 (6). A total of 1328 functional high-risk variants were
46 identified.

47 **Other fine-mapping methods**

For LD extension, proxy SNPs were determined as SNPs within high LD ($r^2 > 0.8$) with lead SNPs. Bayesian fine-mapping was performed as described previously (7). Specially, for each locus, we first pre-filtered variants in at least low LD ($r^2 > 0.1$) with the lead SNPs. We calculated approximate Bayes factors (aBF) for each variant using the effect estimates (β) and standard errors (SE), assuming prior variance $w=0.04$. We calculated the posterior probability of association (PPA) by dividing the aBF for each variant by the sum of aBFs for all variants included in the signal. We then defined the 99% credible set as the smallest set of variants that added up to 99% PPA. The SNPs extracted by LD extension and Bayesian fine-mapping were then annotated using open chromatin regions of kidney cell types. Functionally annotated variants were categorized as functional high-risk variants in these two methods. Colocalization SNPs of urate-associated loci with kidney eQTLs were obtained from published manuscript and considered as the functional high-risk variants in colocalization (4).

Comparison of other fine-mapping methods to LSS

GWAS results of serum urate (EA Tin) were used to compare all other fine-mapping methods to LSS. Interpretable loci are defined as loci in which functional high-risk variants can be identified. For enrichment analysis of genetic contribution for functional high-risk variants, a random equal number of non-significant variants in the serum urate GWAS summary were used as control. The odds ratio was calculated as the enrichment of disease association from the functional high-risk variants compared to that of random control variants.

Annotation of the variants

PheWAS plot is generated from <https://atlas.ctglab.nl/PheWAS>. CADD score is obtained from <https://cadd.gs.washington.edu>.

scATAC-seq and scRNA-seq analysis

For the pre-processing of the 10x-based kidney scATAC-seq data, reads were aligned to the hg38 human genome by using the "cellranger-atac count" function in cellranger-atac version 2.0.0. Quality control measurements and filtering were conducted using the ArchR version 1.0.2 (8) with default parameters. We re-analyzed the kidney scRNA-seq dataset

by using scRNA-seq pipeline in Seurat version 3.0.0 (9) and annotating with markers reported in previous manuscript (10).

Construction of transcriptional regulatory network

We leveraged the integrated kidney scRNA-seq and scATAC-seq data by ArchR version 1.0.2 to compute the linkage score between chromatin accessible regions and gene expression ('peak-to-gene links') by using the "getPeak2GeneLinks" function in ArchR with default parameters. We further filtered out the linkages of peak-to-gene with high linkage score with $\text{abs}(\text{linkage score}) > 0.6$ and with genes expressed in not less than 5% cells.

Gene regulatory prioritization score

In order to prioritize the risk genes, we calculated a GRPS for each candidate gene by considering cumulative regulatory factors including the risk significance of its linked causal variants, the regulation strength from its linked causal variants, and the polygenic effects of multiple independent variants as follow:

$$\text{GRPS}_g = \sum_{k=1}^n R_k^2 \cdot (-\log_{10} P_k)$$

where GRPS_g is the cumulative regulatory prioritization score of gene g , n is the number of linked risk-peaks (peaks with high-risk variants identified by LSS) for gene g ; R is the peak-to-gene linkage score for peak k on gene g ; P is the highest disease association of high-risk variant in peak k . GRPS value were used to prioritize 160 candidate genes identified by LSS. The top 25% of ranked genes by GRPS prioritization were visualized with the ggplot2 version 3.1.1.

Other gene prioritization methods

For H-MAGMA (11), exonic and promoter SNPs were assigned to the genes in which they reside, while intronic and intergenic SNPs were coupled to their regulated genes based on high linkage score detected in peak-to-gene links. Then we ran H-MAGMA version 1.08 with the default parameter and obtained the gene prioritization. The priority by Open Targets Genetics (12) is downloaded by searching for 'urate measurement' at Open Targets

(<https://genetics.opentargets.org>). The priority by ABC-Max with keyword 'UA' (uric acid) is downloaded from Supplementary Table 10 of previous manuscript (13).

Performance of gene prioritization methods in predicting potential drug targets

Drug targets for treatment of hyperuricemia and gout were obtained from the Supplementary Table 1 of previous manuscript (14). To compare the performance of different gene prioritization methods in predicting potential drug targets, the priority scores obtained by each method are first subjected to quantile normalization using the R package preprocessCore version 1.60.2. Predicted probability and odds ratio for putative drug targets of priority genes by different strategies was calculated using logistic regression from “glm” in R package stats version 4.2.2. The predictor of the logistic regression model was the normalized priority score of gene, and the response of the model was whether the gene was drug target (0: no; 1: yes).

Gene enrichment analysis

For identified candidate causal genes, gene ontology and pathway enrichment analysis and DisGeNET enrichment analysis were performed with Metascape version 3.5 (<http://metascape.org>). Supporting evidence for candidate genes came from three sources: genes associated with hyperuricemia or gout in DISEASE (<https://diseases.jensenlab.org/>), genes were the strongest association with urate-related phenotypes in PheWAS (<https://atlas.ctglab.nl/PheWAS>), or genes for which mechanisms were explored in previous studies (15-17). The evidence score is sum score of above supports.

CRISPR activation

CRISPRa experiments were performed as previously described procedure (18-20). Lenti dCAS-VP64_Blast (Addgene, #61425) and lenti MS2-P65-HSF1_Hygro (Addgene, #61426) were packed as lentivirus and transfected to HEK293T cells to generate stable synergistic activation mediator (SAM) cell lines under 200 µg/mL Hygromycin (Thermo Fisher Scientific, 10687010) and 5 µg/mL Blasticidin (Thermo Fisher Scientific, A1113903) selection for 14 days. sgRNA of rs1165183-harboring-CRE was designed by IDTdna (<https://sg.idtdna.com/site/order/designtool/>) and the sgRNA of non-human-genome-targeting was derived from human GeCKOv2 library (18) as the negative control

(Table S4). The sgRNA were then inserted into lenti sgRNA (MS2) Puro backbone (Addgene, #73795) using BsmBI (NEB, R0739S) sites. The lentivirus of sgRNA was prepared and transfected into stable SAM cell lines under 2 µg/mL Puromycin (Thermo Fisher Scientific, A1113803) selection for 48 hours.

Quantitative real-time Polymerase Chain Reaction (qRT-PCR)

Total mRNA was extracted using TaKaRa MiniBEST Universal RNA Extraction Kit (Takara, 9767) following manufacturer's instruction. The RNA obtained (500 ng) was reverse-transcribed using PrimeScript™ RT Master Kit (Takara, RR036A) following the manufacturer's instruction. mRNA levels for *SLC17A4* were measured by qRT-PCR using Power SYBR™ Green PCR (Applied Biosystems™, 4367659) according to the manufacturer's instruction (Table S4).

***SLC17A4* overexpression**

The mammalian expression vector pcDNA3.1-GFP and pcDNA3.1-SLC17A4 was purchased from GenePharma. Transfection of negative control plasmid and overexpressing plasmid were performed using Lipofectamine 3000 (Invitrogen, L3000015) according to the manufacturer's instruction. For transfection, HEK293T cells were seeded in 24-well plates at ~50% confluency, and then transfected with 500 ng plasmids for each well. After 24 hours, RNA was extracted, then reverse-transcribed and *SLC17A4* gene expression was measured by qRT-PCR.

Determination of urate levels

The negative control and *SLC17A4*-overexpressing HEK293T cells were treated with 80 µg/mL uric acid sodium salt (Sigma, U2875). The urate content of the supernatants and broken cells was detected using the urate kit (Nanjing Jiancheng Bioengineering Institute, C012-2-1) after incubation for 24 hours in a 37°C incubator. Three biological replications were carried out for each group.

Statistical Analysis

The following statistical tests were performed by two-tailed Student's t test (Fig. 5C, Fig. 5D, Fig. 5E, Fig. S8B). Data on Fig. 5C, Fig. 5D, Fig. 5E, Fig. S8B are mean ± SD.

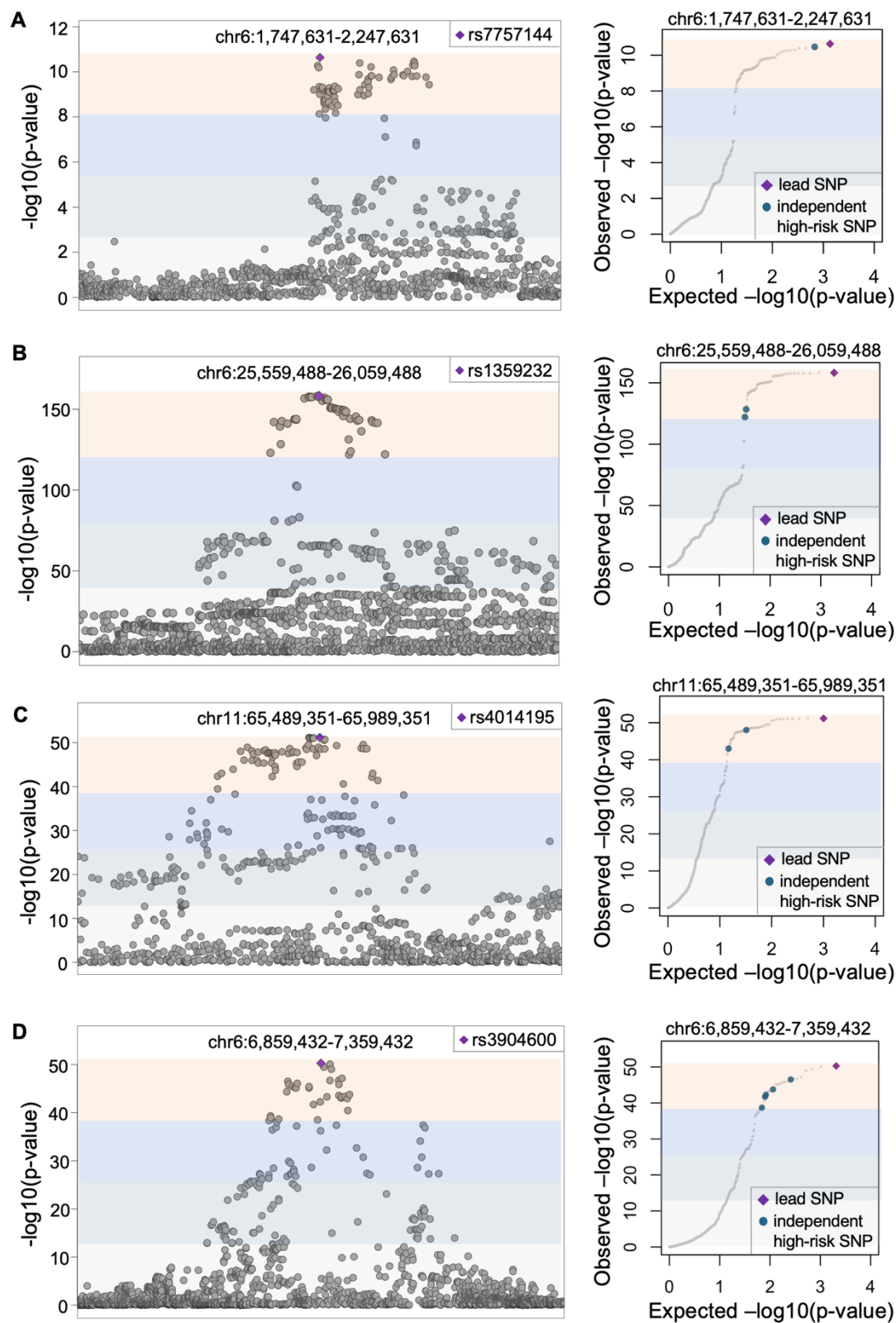


Fig. S1 Loci of serum urate GWAS with multiple independent associations.

LocusZoom plots (left) and quantile-quantile plots (right) for GWAS results in loci with multiple independent associations, where chr6:1,747,631-2,247,631 with lead SNP rs7757144 (**A**), chr6:25,559,488-26,059,488 with lead SNP rs1359232 (**B**), chr11:65,489,351-65,989,351 with lead SNP rs4014195 (**C**) and chr6:6,859,432-7,359,432 with lead SNP rs3904600 (**D**), respectively.

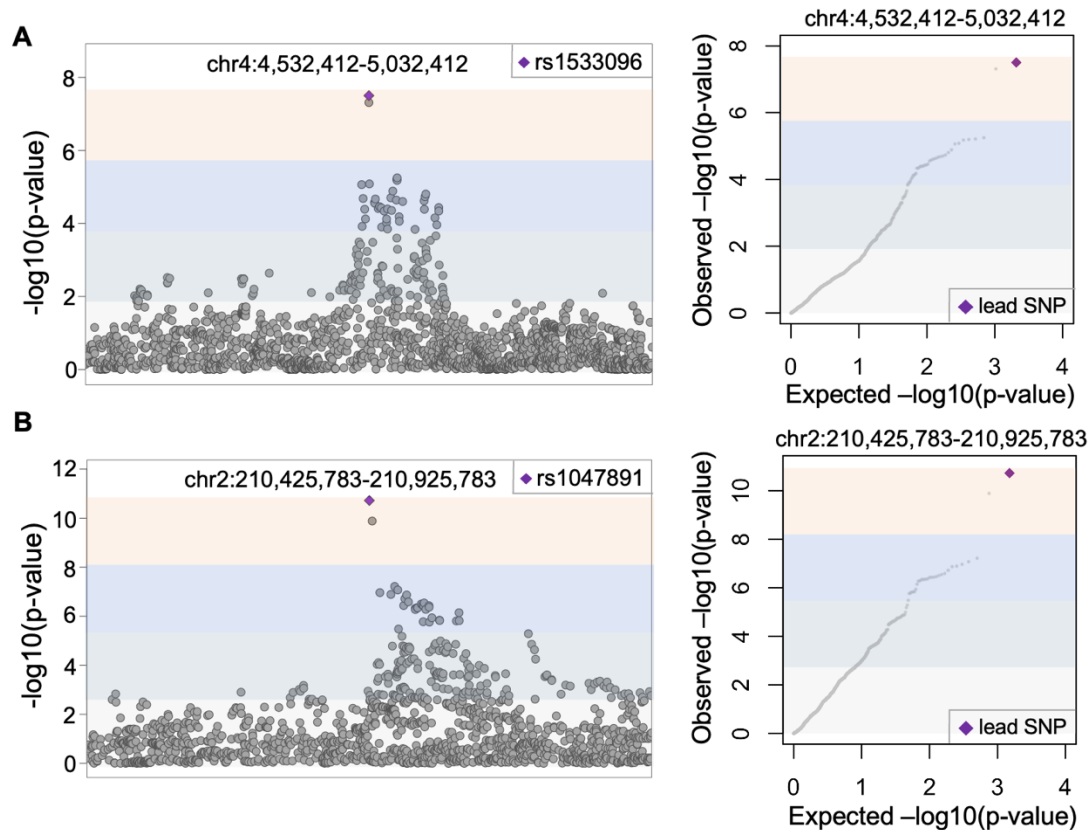


Fig. S2 Loci of serum urate GWAS with single association. LocusZoom plots (left) and quantile-quantile plots (right) for GWAS results at loci with single association, where chr4:4,532,412-5,032,412 with lead SNP rs1533096 (A), chr2:210,425,783-210,925,783 with lead SNP rs1047891 (B), respectively.

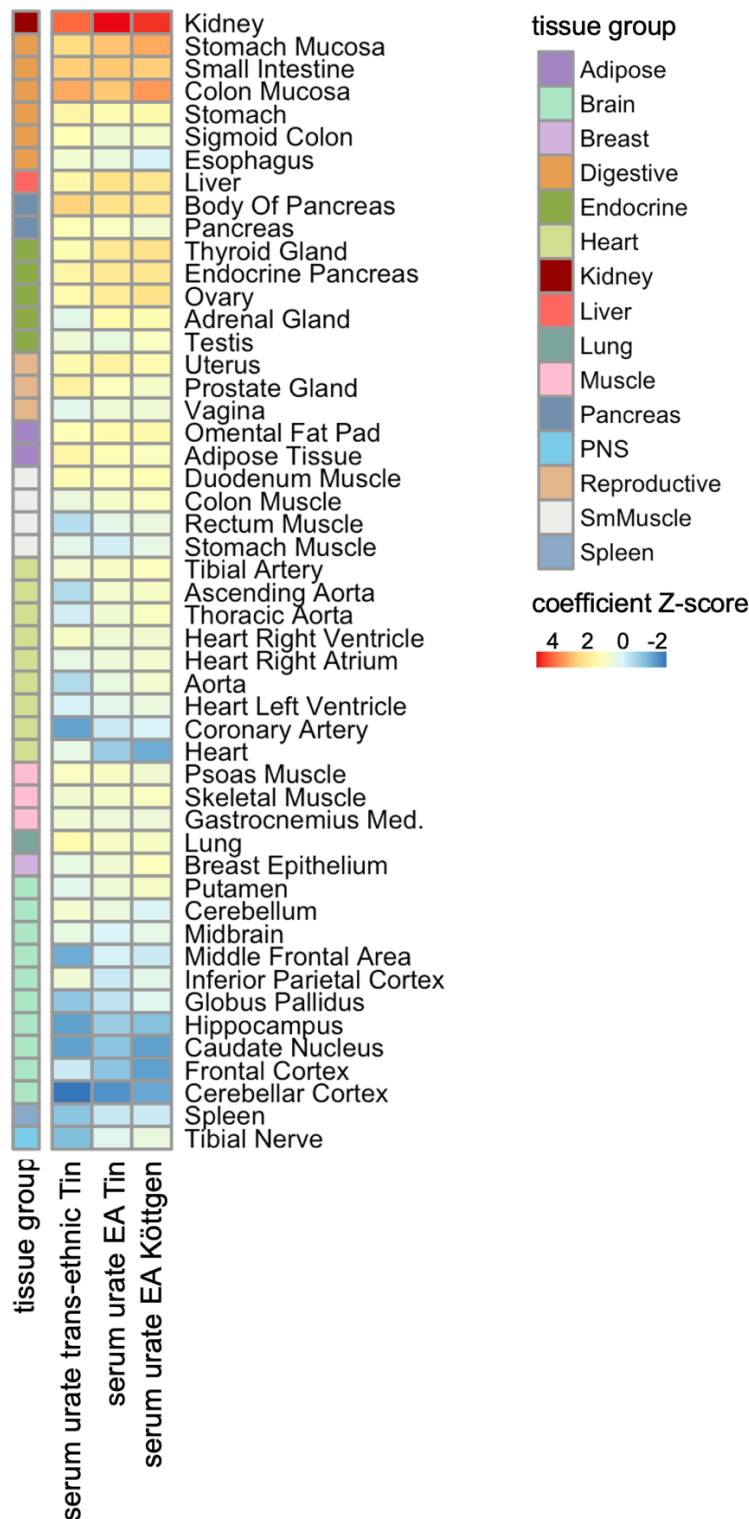


Fig. S3 Tissue enrichment for genetic risk of serum urate. A heatmap displaying LD regression coefficient Z-scores for human tissues across three serum urate GWAS studies analyzed. Data are annotated by groups of tissues.

Number of cell type specific CRE-variants

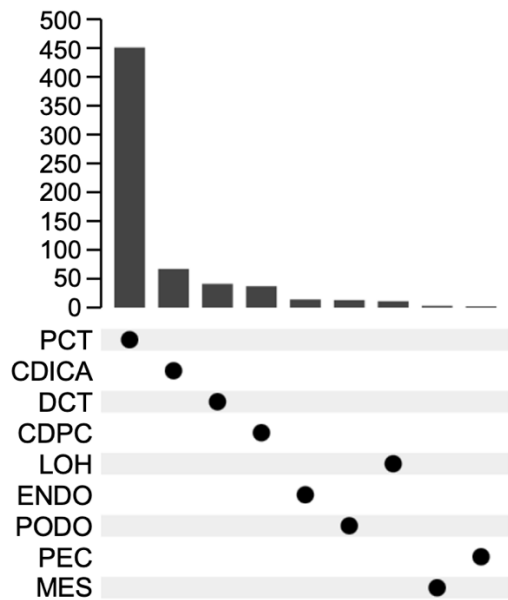


Fig. S4 Cell type contribution for genetic risk of serum urate. Number of high-risk variants overlapping with regulatory elements in unique kidney cell type.

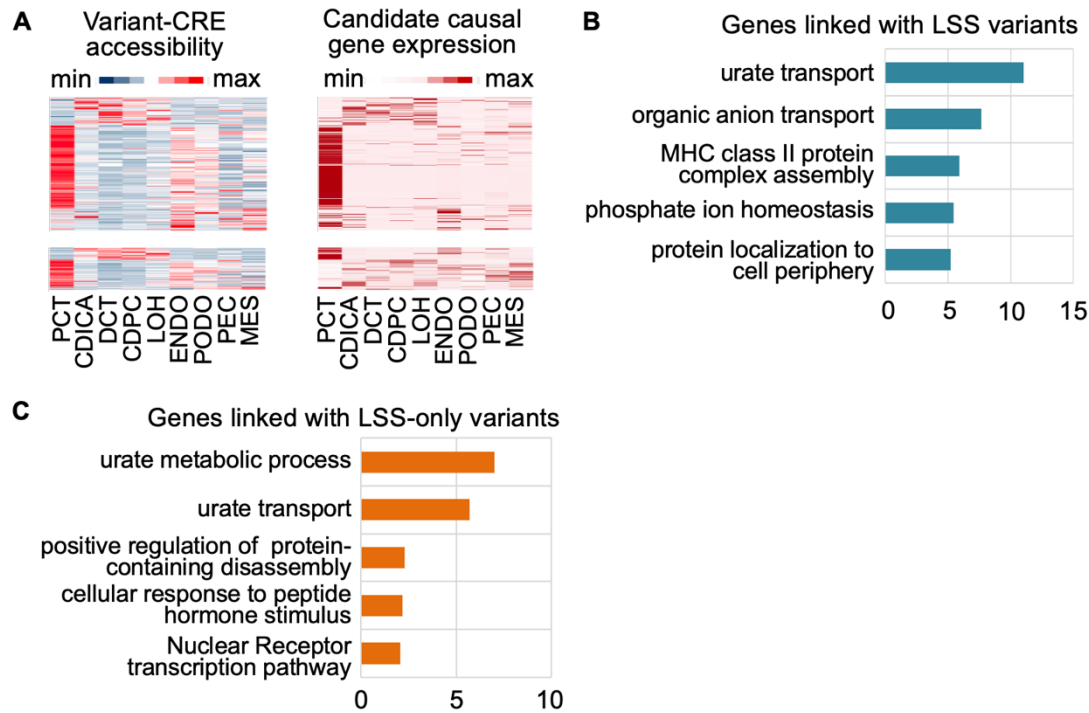


Fig. S5 Identification of candidate causal genes for functional variants. (A) Peak-to-gene links in kidney cell types including positive and negative regulation. Heatmaps showed the chromatin accessibility of regulatory elements and gene expression in peak-to-gene links, respectively. The color represents z-score. (B and C) Top 5 enriched gene ontology and pathway terms on candidate causal genes linked with LSS (B) or LSS-only (C) high-risk variants.

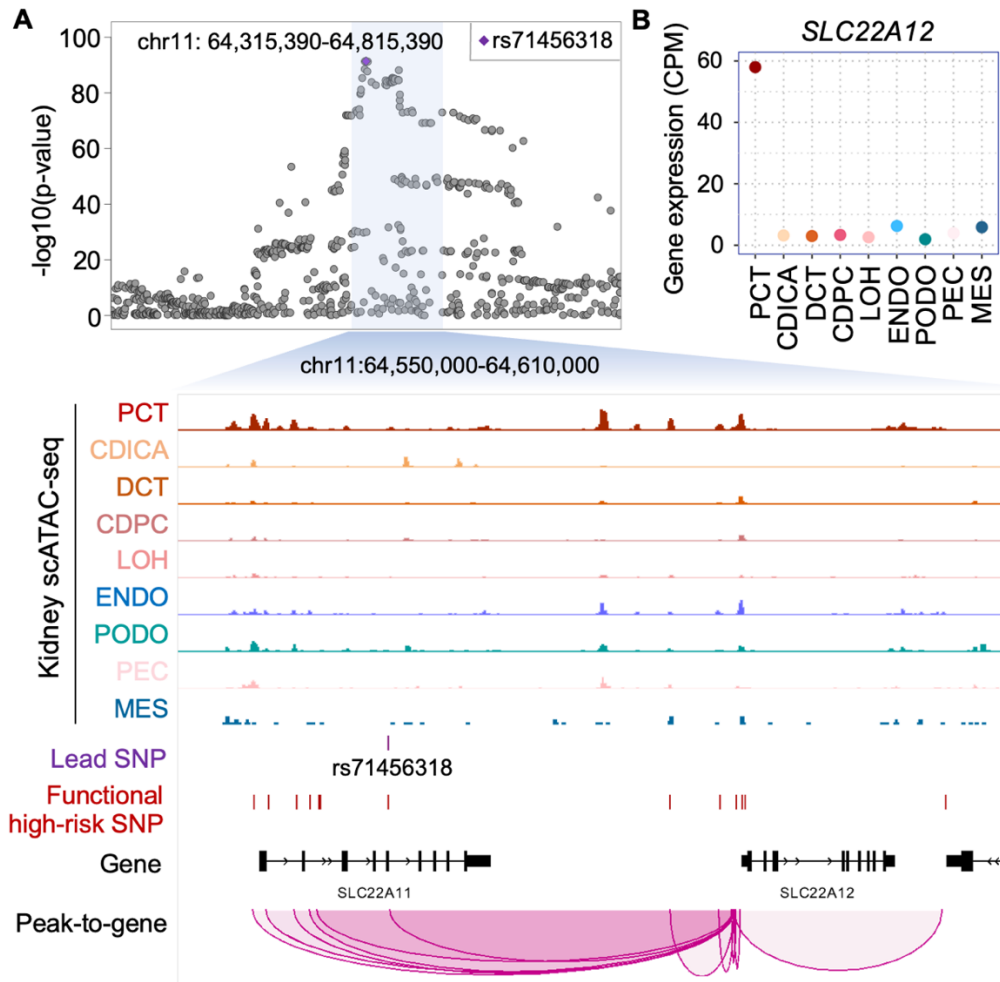


Fig. S6 The complex regulation between variants and candidate causal genes. (A) LocusZoom plot of GWAS result at locus chr11: 64,315,390-64,815,390 with lead SNP rs71456318 (top), and genome browser view of the highlighted region (bottom). The genome browser includes tracks for chromatin accessibilities in kidney cell types, the position of the lead SNP and functional high-risk SNPs, the location of genes, and the peak-to-gene linkage. **(B)** Gene expression of *SLC22A12* in kidney cell types.

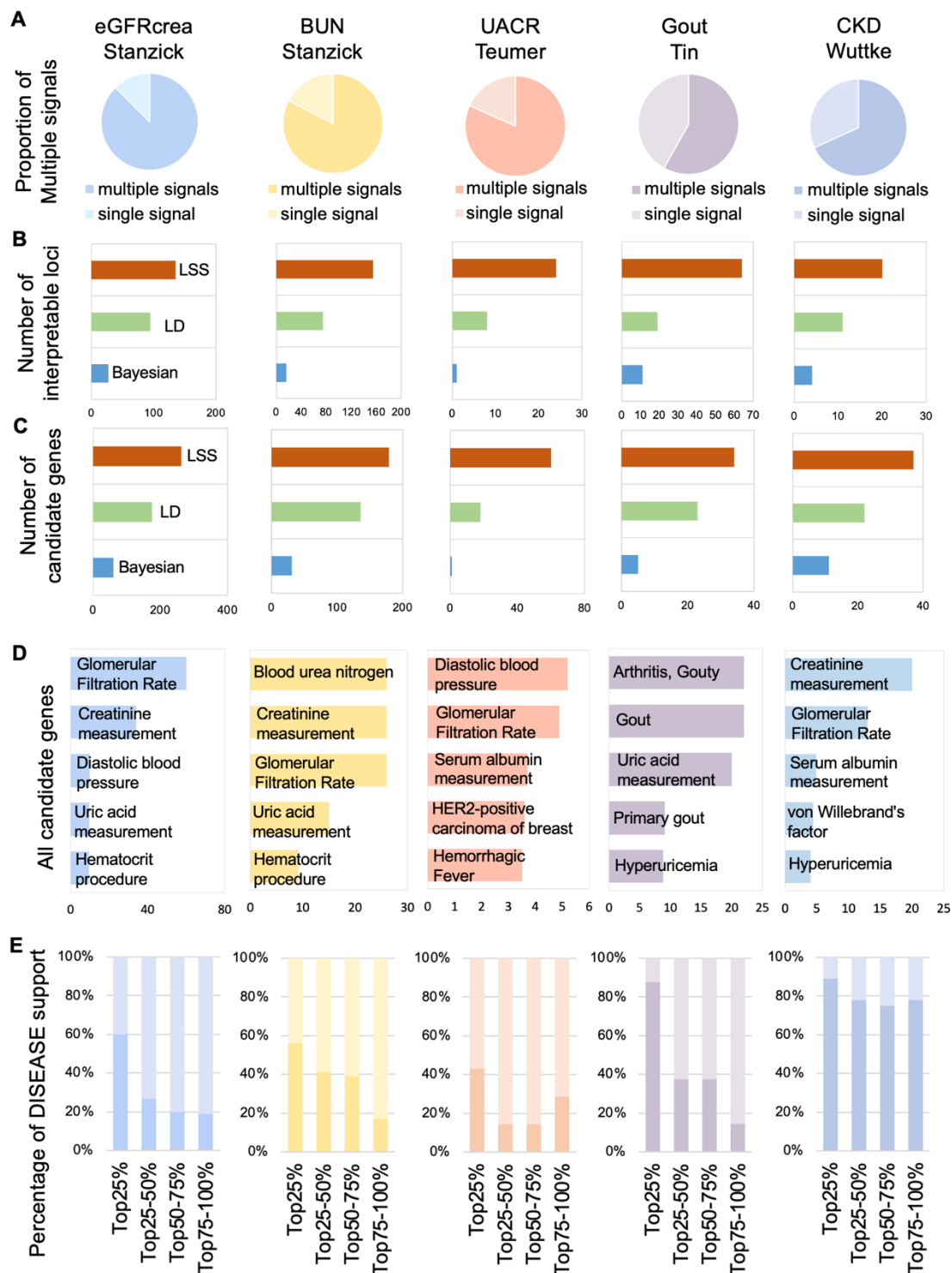


Fig. S7 Application of LSS and GRPS to other GWAS studies. (A) Proportion of loci with multiple independent associations in GWAS of eGFRcrea, BUN, UACR, gout and CKD. (B) Number of interpretable loci in five diseases and traits among different strategies. (C) Number of candidate risk genes identified by different strategies. (D) Top

205 5 enriched diseases terms on candidate causal genes regulated by functional variants. (E)
206 Risk genes are divided into 4 groups based on GRPS from highest to lowest, and the bar
207 chart shows the percentage of genes supported by disease association from DISEASE
208 displayed.
209

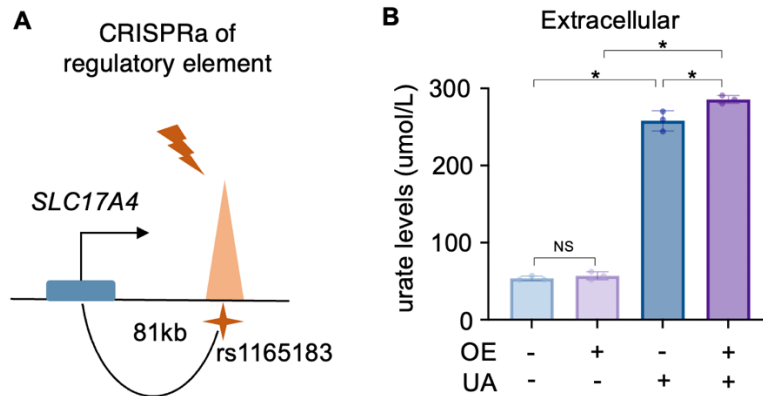


Fig. S8 Validating the regulatory and functional mechanisms of the risk gene. (A) Experimental scheme for CRISPRa with rs1165183-harboring-CRE. **(B)** Effects of *SLC17A4* overexpression on the extracellular urate levels of HEK293T cells (n=3, two-tailed Student's t test, P-value for negative control cells which are untreated or treated with uric acid (UA) is <0.0001; P-value for negative control vs *SLC17A4* OE cells which are all treated with UA is 0.0268, P-value for negative control cells vs *SLC17A4* OE cells which are all untreated with UA is 0.3738, P-value for *SLC17A4* OE cells which are untreated or treated with UA is <0.0001; * indicates P-value <0.05, NS indicates not significant).

References

1. K. Watanabe, E. Taskesen, A. van Bochoven, D. Posthuma, Functional mapping and annotation of genetic associations with FUMA. *Nat Commun* **8**, 1826 (2017).
2. C. A. Boix, B. T. James, Y. P. Park, W. Meuleman, M. Kellis, Regulatory genomic circuitry of human disease loci by integrative epigenomics. *Nature* **590**, 300-307 (2021).
3. P. W. Hook, A. S. McCallion, Leveraging mouse chromatin data for heritability enrichment informs common disease architecture and reveals cortical layer contributions to schizophrenia. *Genome Res* **30**, 528-539 (2020).
4. A. Tin *et al.*, Target genes, variants, tissues and transcriptional pathways influencing human serum urate levels. *Nat Genet* **51**, 1459-+ (2019).
5. A. Kottgen *et al.*, Genome-wide association analyses identify 18 new loci associated with serum urate concentrations. *Nat Genet* **45**, 145-154 (2013).
6. T. A. Myers, S. J. Chanock, M. J. Machiela, LDlinkR: An R Package for Rapidly Calculating Linkage Disequilibrium Statistics in Diverse Populations. *Front Genet* **11**, (2020).
7. K. Zhang *et al.*, A single-cell atlas of chromatin accessibility in the human genome. *Cell* **184**, 5985-6001 e5919 (2021).
8. J. M. Granja *et al.*, ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis (vol 53, pg 403, 2021). *Nat Genet* **53**, 935-935 (2021).
9. T. Stuart *et al.*, Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902 e1821 (2019).
10. P. C. Wilson *et al.*, The single-cell transcriptomic landscape of early human diabetic nephropathy. *Proc Natl Acad Sci U S A* **116**, 19619-19625 (2019).
11. N. Y. A. Sey *et al.*, A computational tool (H-MAGMA) for improved prediction of brain-disorder risk genes by incorporating brain chromatin interaction profiles. *Nat Neurosci* **23**, 583-593 (2020).
12. E. Mountjoy *et al.*, An open approach to systematically prioritize causal variants and genes at all published human GWAS trait-associated loci. *Nat Genet* **53**, 1527-+ (2021).
13. J. Nasser *et al.*, Genome-wide enhancer maps link risk variants to disease genes. *Nature* **593**, 238-243 (2021).
14. K. Trajanoska *et al.*, From target discovery to clinical drug development with human genetics. *Nature* **620**, 737-745 (2023).
15. A. K. Mandal, D. B. Mount, The molecular physiology of uric acid homeostasis. *Annu Rev Physiol* **77**, 323-345 (2015).
16. A. K. Mandal *et al.*, Genetic and Physiological Effects of Insulin-Like Growth Factor-1 (IGF-1) on Human Urate Homeostasis. *J Am Soc Nephrol* **34**, 451-466 (2023).
17. M. Sato *et al.*, Renal secretion of uric acid by organic anion transporter 2 (OAT2/SLC22A7) in human. *Biol Pharm Bull* **33**, 498-503 (2010).
18. J. Joung *et al.*, Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening. *Nat Protoc* **12**, 828-863 (2017).

- 266 19. X. Deng *et al.*, Long noncoding RNA PiHL regulates p53 protein stability
267 through GRWD1/RPL11/MDM2 axis in colorectal cancer. *Theranostics* **10**, 265-
268 280 (2020).
- 269 20. J. Joung *et al.*, Genome-scale activation screen identifies a lncRNA locus
270 regulating a gene neighbourhood. *Nature* **548**, 343-346 (2017).
271