



Research article

Plasma proteome analysis implicates novel proteins as potential therapeutic targets for chronic kidney disease: A proteome-wide association study

Yang Xiong^{a,1}, Tianhong Wang^{b,1}, Wei Wang^a, Yangchang Zhang^c, Fuxun Zhang^d, Jiuhong Yuan^a, Feng Qin^{a,**}, Xianding Wang^{a,e,*}

^a Department of Urology and Andrology Laboratory, West China Hospital, Sichuan University, Sichuan, 610041, China

^b Department of Anesthesiology, West China Hospital, Sichuan University, Sichuan, 610041, China

^c Department of Public Health, Capital Medical University, Beijing, 100000, China

^d Department of Urology, Tangdu Hospital, The Air Force Medical University, Xi'an, Shaanxi, 710000, China

^e Kidney Transplant Center, Transplant Center, West China Hospital, Sichuan University, Sichuan, 610041, China

ARTICLE INFO

Keywords:

Chronic kidney disease
Mendelian randomization
Plasma proteomes
Estimated glomerular filtration rate
Blood urea nitrogen

ABSTRACT

Chronic kidney disease (CKD) is prevalent globally with limited therapeutic drugs available. To systemically identify novel proteins involved in the pathogenesis of CKD and possible therapeutic targets, we integrated human plasma proteomes with the genome-wide association studies (GWASs) of CKD, estimated glomerular filtration rate (eGFR) and blood urea nitrogen (BUN) to perform proteome-wide association study (PWAS), Mendelian Randomization and Bayesian colocalization analyses. The single-cell RNA sequencing data of healthy human and mouse kidneys were analyzed to explore the cell-type specificity of identified genes. Functional enrichment analysis was conducted to investigate the involved signaling pathways. The PWAS identified 22 plasma proteins significantly associated with CKD. Of them, the significant associations of three proteins (INHBC, LMAN2, and SNUPN) were replicated in the GWASs of eGFR, and BUN. Mendelian Randomization analyses showed that INHBC and SNUPN were causally associated with CKD, eGFR, and BUN. The Bayesian colocalization analysis identified shared causal variants for INHBC in CKD, eGFR, and BUN (all $PP4 > 0.75$). The single-cell RNA sequencing revealed that the INHBC gene was sparsely scattered within the kidney cells. This proteomic study revealed that INHBC, LMAN2, and SNUPN may be involved in the pathogenesis of CKD, which represent novel therapeutic targets and warrant further exploration in future research.

1. Introduction

Chronic kidney disease (CKD) is a prevalent condition featured by persistent alterations of kidney structure and function [1]. The lasting kidney dysfunction can trigger edema, hypertension, metabolic dysfunction, and cardiovascular disease, leading to fatigue,

* Corresponding author. Department of Urology, West China Hospital, Sichuan University, 37 Guoxue Alley, Chengdu, Sichuan, 610041, China.

** Corresponding author. Department of Urology and Andrology Laboratory, West China Hospital, Sichuan University, 37 Guoxue Alley, Chengdu, Sichuan, 610041, China.

E-mail addresses: qinfeng@scu.edu.cn (F. Qin), xiandingwang@scu.edu.cn (X. Wang).

¹ Notes: Yang Xiong, and Tianhong Wang contributed equally to this work.

<https://doi.org/10.1016/j.heliyon.2024.e31704>

Received 13 January 2024; Received in revised form 13 May 2024; Accepted 21 May 2024

Available online 22 May 2024

2405-8440/© 2024 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

depression, and significantly decreased quality of life [2]. The global number of all-stage CKD cases had reached to 697.5 million, with a prevalence of 9.1 % in 2017 [3]. In 2016, 1.19 million deaths could be attributed to CKD, and it was estimated to increase to 3.09 million by 2040 [4]. The considerable disease burden requires a clear understanding of the pathogenesis of CKD and further developing effective intervention strategies.

In the last decade, large genome-wide association studies (GWASs) contributed to discovering the genetic variants of CKD [5]. Many genetic variants are in non-coding regions and influence phenotypes by regulating the expression of the target gene and the corresponding proteins [6,7]. However, clarifying the biological mechanisms linking genetic variants to CKD remains challenging for most of the loci, and only a few loci such as the *UMOD* locus was clearly described in the pathogenesis of CKD [8]. In addition, the causal variants driving the loci-phenotype associations are often obscured due to linkage disequilibrium [9]. The gap from association to function obstacles potential clinical applications, such as developing novel drugs by targeting candidate genes for variants [9]. Proteins are the ultimate product of transcripts and the major functional molecules in the onset and progress of CKD, especially for plasma proteins [10]. Plasma proteome is often dysregulated by CKD and highly amendable to drug targeting [10,11]. With the advance of high-throughput proteome sequencing technology, a number of studies have determined the protein quantitative trait loci (pQTL) for plasma proteome, linking genetic variants to abundances of plasma proteins [12]. The pQTLs can be further integrated with GWAS of CKD under the framework of proteome-wide association studies (PWAS) to systematically study the associations between plasma proteomics and CKD, possibly identifying novel therapeutic targets [12].

Identifying therapeutic targets is crucial for CKD, which may be beneficial to improve the prognosis of CKD patients. To discover potential drug targets for CKD, we first used PWAS to systematically explore the association between 1350 plasma proteins and CKD. The associations were further verified in GWASs of estimated glomerular filtration rate (eGFR), and blood urea nitrogen (BUN). The two traits are good clinical indicators of renal failure. Moreover, the *cis*-pQTLs of the identified proteins were extracted to perform Mendelian Randomization (MR) to confirm the causality. Bayesian colocalization was also applied to detect the causal variants driving the alteration of proteins and CKD, eGFR, and BUN. In addition, the cell-specific expression of significant gene was explored at the single-cell transcriptional level in the kidney of humans and mice. The identified genes were further searched in the Drug-Gene Interaction Database to check the druggability. These analyses may identify novel proteins closely associated with CKD and provide new drug targets for therapeutics.

2. Materials and methods

2.1. Human plasma proteomic and genetic data

Reference human plasma proteome for the PWAS was profiled from the plasma samples from the Atherosclerosis Risk in Communities (ARIC) study [12]. The ARIC study (third visit) recruited participants from European and African Americans across the US. To reduce the population architecture bias, we only used human proteomic and genetic data of European descent. After excluding participants without genotype data, a total of 7213 individuals with European ancestry were included in this study. The relative concentrations of plasma proteins were measured by a slow-off rate modified aptamer (SOMAmer)-based approach [13]. These aptamers can specifically tag the proteins and finally, 4657 SOMAmers measuring 4483 unique proteins were obtained in the original study [12]. The genotypes of included samples were sequenced by the Affymetrix 6.0 DNA microarray. To identify the *cis*-pQTL of detected plasma proteins, regression analyses were performed by adjusting for the age, gender, study sites, ten genetic principal components, and probabilistic estimation of expression residuals (PEER) factors. The *cis*-regions were defined as ± 500 kb of the transcriptional start sites, and 6,181,856 single nucleotide polymorphisms (SNPs, minor allele frequency >1 %) were analyzed in the *cis*-regions. Finally, 2004 significant SOMAmers were identified, who had at least one significant (false discovery rate (FDR) < 5 %) *cis*-pQTL near the encoding gene of the protein.

2.2. GWAS data of eGFR, BUN, and CKD

The GWAS data of eGFR, BUN, and CKD were retrieved from the CKDGen consortium (round 4, <https://ckdgen.imbi.uni-freiburg.de/>) [5]. The CKDGen consortium collected GWAS data regarding kidney function from 121 cohorts in different countries and races. This consortium has expanded the sample size to a million individuals, which increases the power to identify genetic variants associated with CKD. To avoid the population architecture bias, we only used the data from European descent. The eGFR, BUN, and CKD GWASs included 567,460, 243,029, and 480,698 participants (41,395 cases and 439,303 controls), respectively. The genetic heritability for eGFR was 39 %. The estimation of eGFR was based on the blood creatinine, using the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation for participants aged >18 years, and Schwartz formula for participants aged ≤ 18 years. The cut-off for diagnosing CKD was eGFR below $60 \text{ ml min}^{-1} \text{ per } 1.73 \text{ m}^2$. The creatine and BUN were determined by enzymatic photometric assay, spectrophotometry, and modified kinetic Jaffe reaction in different cohorts. The identified SNPs for eGFR, BUN, and CKD were 8,834,748, 8,358,347, and 9,585,923 in the European ancestry, which were used as summary-level statistics in our analyses.

2.3. Proteome-wide association studies (PWAS)

Proteins are the ultimate product of transcripts and the major functional molecules in the onset and progress of CKD. To investigate the association of plasma proteins with CKD, PWAS was first performed using FUSION (<http://gusevlab.org/projects/fusion>) as shown by the schematic diagram in Fig. 1. The SNP-based heritability for the 2004 significant SOMAmers was estimated using the REML

algorithm in the GCTA [14]. Of them, 1350 SOMAmers were found to have significant non-zero *cis*-heritability ($P < 0.01$). Then, FUSION was used to estimate the SNP effects on the concentrations of SOMAmers. The protein imputation models were constructed using the top1, and enet methods. Models with the best performance in predicting protein abundance were selected. Then, FUSION combined the genetic effects of eGFR, BUN, and CKD (GWAS Z scores) with the protein weights by calculating the linear sum of Z scores \times weights for the independent SNPs at the locus to perform the PWAS [15]. Benjamini-Hochberg method was adopted to adjust for multiple hypothesis testing. Adjusted P value < 0.05 was seen as significant in statistics.

2.4. Protein interaction network and functional enrichment analyses

The identified proteins in PWAS were subjected to further protein-protein interaction network (PPI) analyses. The function interactions of these proteins were retrieved from the Search Tool for the Retrieval of Interacting Genes (STRING) database (<https://string-db.org>). The interactions with a combined score >0.4 were seen as statistically significant, which were then visualized by Cytoscape software (version 3.10.0). To identify the most closely connected modules in the PPI network, the Molecular Complex Detection (MCODE) tool in Cytoscape with defaulting parameters. The top ten hub genes in the PPI network were identified by the cytohubba tool according to the betweenness of genes. In addition, the “clusterProfiler” package in R (4.2.1) was used to perform functional enrichment analyses including Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses [16]. The most relevant GO terms associated with the genes encoding the proteins in PWAS were retrieved from the GO database (<http://geneontology.org/>). KEGG analysis was used to predict the possibly involved signaling pathways. The hypergeometric test was used to obtain P values for the enrichment analyses and the Benjamini-Hochberg method was adopted to adjust for multiple hypothesis testing. Adjusted P value < 0.05 was seen as significant in statistics.

2.5. Mendelian Randomization (MR) analyses

To further explore the causal associations between proteins and eGFR, BUN, and CKD, we extracted the SNPs of significant proteins

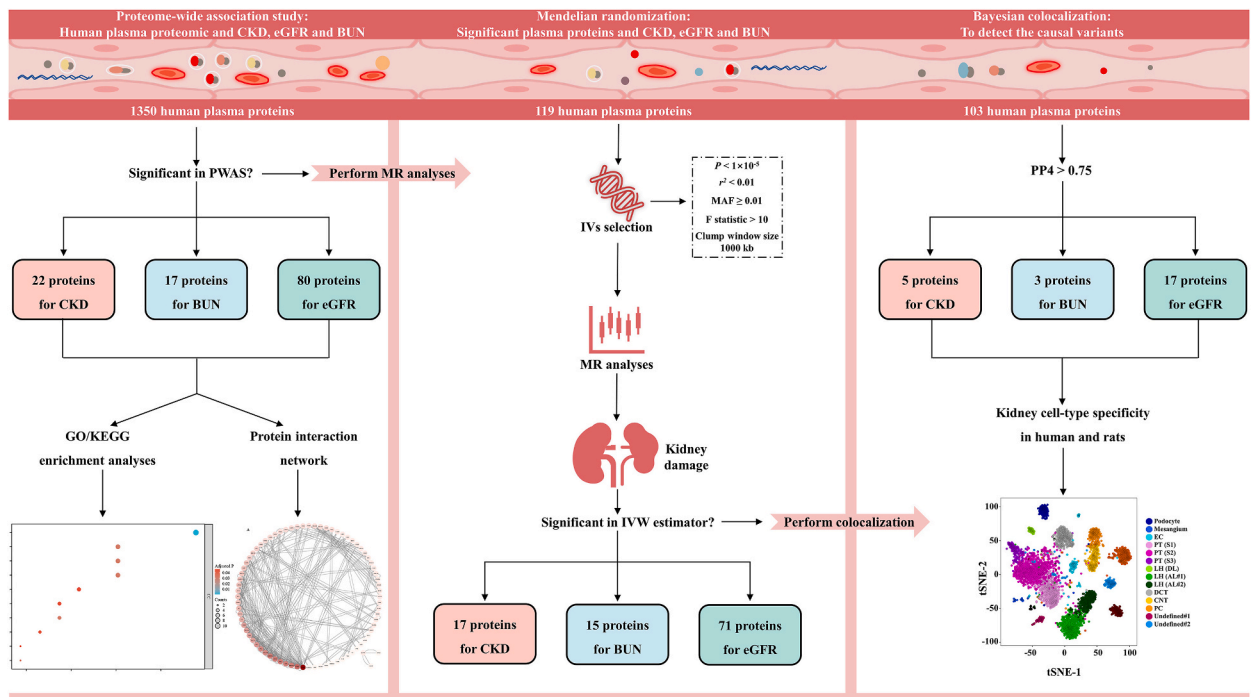
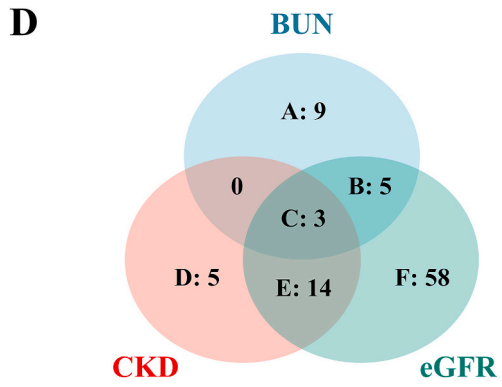
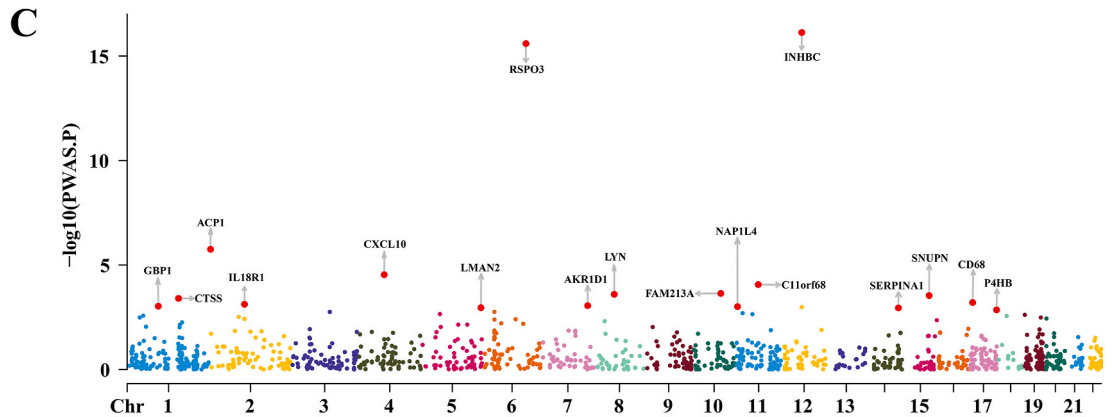
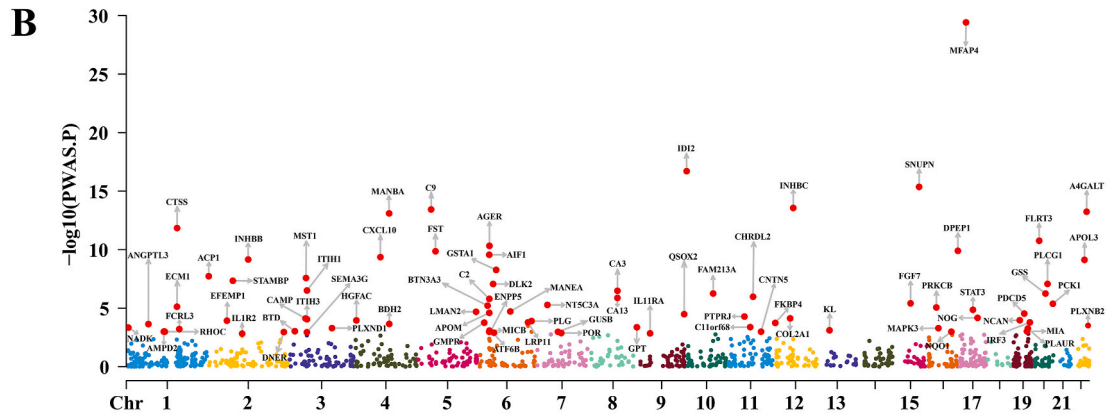
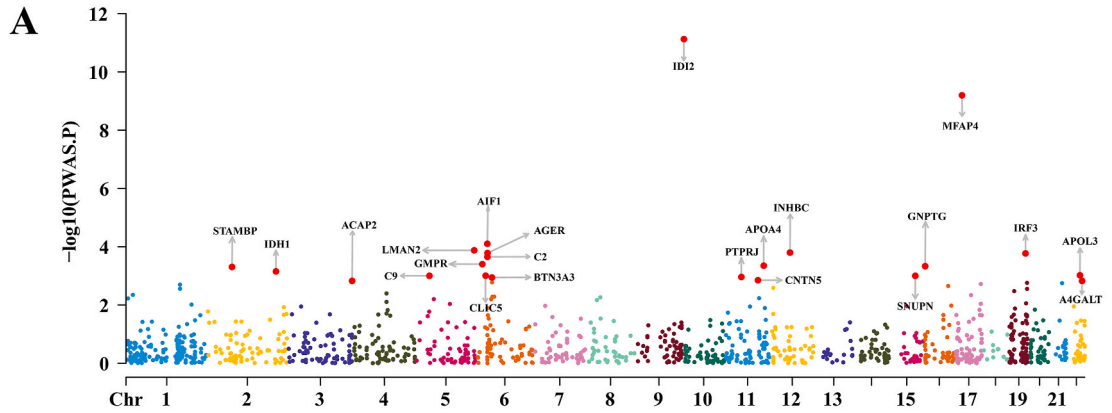


Fig. 1. Overview of the study design and analysis strategy.

PWAS was first performed by FUSION to identify the associations between plasma proteins, CKD, eGFR and BUN. Protein interaction network and functional enrichment analyses were carried out for the identified genes. To extend the observation association to causal association, the causal associations, significant plasma proteins in PWAS were further subjected to MR analyses. Filtered SNPs that meet the criteria are considered as instrumental variables (IVs) and undergo sensitivity analyses, heterogeneity and pleiotropy evaluation. In this step, inverse variance weighting and MR-Egger methods were used. To assess the probability of a shared causal variant driving the alteration of proteins and eGFR, BUN, and CKD, Bayesian colocalization analyses were further conducted for the significant *cis*-regulated plasma proteins. Single-cell RNA sequencing data of healthy human and mouse kidneys were used to check the expression and characteristics of INHBC. PWAS: proteome-wide association study; CKD: chronic kidney disease; MR: Mendelian Randomization; eGFR: estimated glomerular filtration rate; BUN: blood urea nitrogen; MAF: minor allele frequency; IV: instrumental variable.



- A:** RSPO3, LYN, CD68, IL18R1, AKR1D1, GBP1, NAP1L4, SERPINA1, P4HB
- B:** ACP1, CXCL10, C11orf68, FAM213A, CTSS
- C:** LMAN2, INHBC, SNUPN
- D:** APOA4, GNPTG, IDH1, CLIC5, ACAP2
- E:** IDI2, MFAP4, AIF1, AGER, IRF3, C2, GMPR, STAMBP, APOL3, BTN3A3, C9, PTPRJ, CNTN5, A4GALT

(caption on next page)

Fig. 2. Manhattan plots for the PWAS results in CKD, eGFR, and BUN.

The results of PWAS in CKD, eGFR, BUN were displayed in panel A–C, respectively. Panel D showed the Venn plot for the intersection of significant genes identified in CKD, eGFR, and BUN. The red dot indicates a false discovery rate <0.05. PWAS: proteome-wide association study; CKD: chronic kidney disease; MR: Mendelian Randomization; eGFR: estimated glomerular filtration rate; BUN: blood urea nitrogen; Chr: chromosome.

in PWAS as instrumental variables (IVs) to perform MR analyses (Fig. 1). To include ample SNPs and increase the statistical power, the significant threshold of IVs was set at $P < 1 \times 10^{-5}$. In addition, the included SNPs were further clumped at a window size of 1000 kb. SNPs with $r^2 \geq 0.01$ were pruned to obtain independent IVs [17]. Moreover, we also assessed the strength of left SNPs to avoid weak IVs bias. The F statistics were calculated as follows: $F \text{ statistics} = R^2 \times (N-k-1)/((1-R^2) \times k)$. Generally, F statistics <10 indicate the presence of weak IVs bias [18]. In this study, all the F statistics of IVs were >20. Thus, they were retained in the final MR analyses.

To combine the genetic effects between exposures (significant proteins in PWAS) and outcomes (eGFR, BUN, and CKD), the Wald ratio (1 IV), and inverse variance weighting (IVW, >1 IV) estimators were used. Besides, the MR-Egger method was selected as a complementary method to verify the findings. The distance between zero and the intercept of Egger regression was used to quantify horizontal pleiotropy. When the intercept was away from zero, it indicated the existence of pleiotropy. Additionally, the Q statistics of the IVW method were used to quantify the heterogeneity. The leave one out analyses were performed to explore the influential SNPs in MR analyses, which may bias the results. In different stages, packages including “TwoSampleMR”, “MR-PRESSO”, and “MendelianRandomization” were used. Benjamini-Hochberg method was adopted to adjust for multiple hypothesis testing. Adjusted P value < 0.05 was seen as significant in statistics.

2.6. Bayesian colocalization analyses and kidney cell-type specificity in human and rats

To assess the probability of a shared causal variant driving the alteration of proteins and eGFR, BUN, and CKD, we used the Bayesian test for colocalization analyses. The “coloc” package in R (4.2.1) was used with defaulting parameters ($p1 = 1 \times 10^{-4}$, $p2 = 1 \times 10^{-4}$, $p12 = 1 \times 10^{-5}$) [19]. Here, p1 represents the prior probability that a variant is associated with eGFR, BUN, or CKD, p2 is the prior probability that a variant is associated with proteins, and p12 is the prior probability that a variant is associated with both traits. Based on the summary-level statistics of GWASs, the Approximate Bayes Factor was calculated to produce the posterior probability (PP) for the following five hypotheses: H0) no association with either trait (PP0); H1) association with eGFR, BUN, or CKD, not with proteins (PP1); H2) association with pQTL, not with eGFR, BUN, or CKD (PP2); H3) two independent SNPs associated with GWASs and pQTL (PP3); H4) one shared SNP associated with GWASs and pQTL (PP4). In this study, PP4 >0.75 signals for colocalization as previous studies reported [20].

To check the expression and characteristics of identified genes (significant in PWAS, MR, and colocalization concurrently), we analyzed the single-cell RNA sequencing data of healthy human and mouse kidneys [21,22]. The single-cell RNA sequencing data were from the Kidney Interactive Transcriptomics database (<http://humphreyslab.com/SingleCell/>). Details of the data process can be accessed in the original studies.

Table 1
PWAS identified three significant genes across CKD, eGFR, and BUN.

Genes	Z-score/OR/ PP4	CKD			eGFR			BUN		
		Statistical estimates	P value	FDR	Statistical estimates	P value	FDR	Statistical estimates	P value	FDR
INHBC	Z-score in PWAS	3.777	1.59×10^{-4}	1.05×10^{-2}	-7.611	2.73×10^{-14}	1.58×10^{-11}	8.338	7.53×10^{-17}	1.02×10^{-13}
	OR (95 % CI) in MR	1.025 (1.009–1.041)	1.86×10^{-3}	2.69×10^{-3}	0.998 (0.997–0.998)	1.99×10^{-22}	1.22×10^{-20}	1.008 (1.006–1.009)	7.32×10^{-23}	9.00×10^{-21}
	PP4 in colocalization	0.761	-	-	0.918	-	-	0.953	-	-
	LMAN2	Z-score in PWAS	3.820	1.33×10^{-4}	8.96×10^{-3}	-4.261	2.04×10^{-5}	1.87×10^{-3}	3.261	1.11×10^{-3}
LMAN2	OR (95 % CI) in MR	1.283 (0.999–1.647)	5.03×10^{-2}	5.63×10^{-2}	0.989 (0.975–1.003)	1.30×10^{-1}	1.41×10^{-1}	1.032 (1.010–1.055)	4.38×10^{-3}	5.85×10^{-3}
	PP4 in colocalization	6.17×10^{-5}	-	-	1.23×10^{-6}	-	-	6.08×10^{-5}	-	-
	SNUPN	Z-score in PWAS	3.290	1.00×10^{-3}	4.00×10^{-2}	-8.129	4.32×10^{-16}	3.49×10^{-13}	3.623	2.91×10^{-4}
SNUPN	OR (95 % CI) in MR	1.114 (1.066–1.165)	1.74×10^{-6}	6.11×10^{-6}	0.988 (0.985–0.992)	1.23×10^{-8}	7.95×10^{-8}	1.015 (1.004–1.025)	5.03×10^{-3}	6.51×10^{-3}
	PP4 in colocalization	4.29×10^{-3}	-	-	9.92×10^{-17}	-	-	1.01×10^{-21}	-	-

PWAS: proteome-wide association study; CKD: chronic kidney disease; MR: Mendelian Randomization; eGFR: estimated glomerular filtration rate; BUN: blood urea nitrogen; OR: odds ratio; CI: confidence interval; FDR: false discovery rate; PP: posterior probability.

3. Results

3.1. Associations of plasma proteins with eGFR, BUN, and CKD

The PWAS identified 22 genes whose *cis*-regulated plasma proteins abundances were associated with CKD at an FDR of $P < 0.05$ (Fig. 2A, Table S1). To validate the findings, further PWAS was performed in eGFR and BUN. For eGFR, the PWAS identified 80 genes (FDR <0.05), and 17 of them were significant in CKD (Fig. 2B, Table S2). For BUN, the PWAS identified 17 proteins (FDR <0.05), and 12 of them were significant in CKD (Fig. 2C, Table S3). As shown in Fig. 2D, only three proteins (INHBC, LMAN2, and SNUPN) remained significant across CKD, eGFR, and BUN (all FDR <0.05). In Table 1, the plasma protein abundance of INHBC was associated with increased risks of CKD (Z-score = 3.777, $P = 1.59 \times 10^{-4}$), declined eGFR (Z-score = -7.611, $P = 2.73 \times 10^{-14}$), and elevated

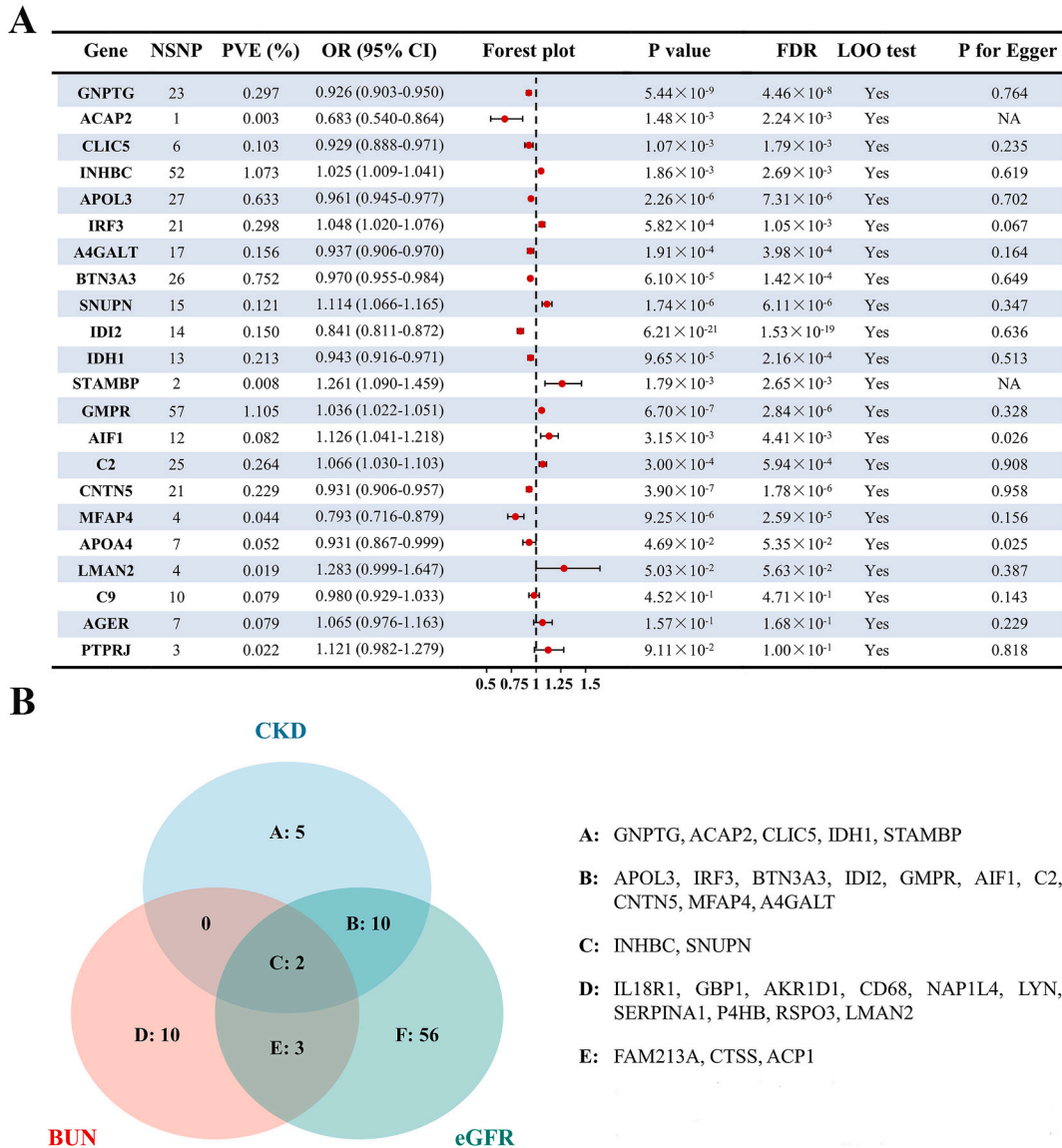


Fig. 3. Forest plot for the MR results in CKD, eGFR, and BUN. Panel A visualized the results of MR in CKD by the IVW estimator. Panel B showed the Venn plot for the intersection of significant genes identified in CKD, eGFR, and BUN. The significance is defined as false discovery rate <0.05. *P* for Egger represents the *P* value for the intercept in MR-Egger regression, which can be used for detecting pleiotropy. NSNP: number of single nucleotide polymorphism; CKD: chronic kidney disease; MR: Mendelian Randomization; eGFR: estimated glomerular filtration rate; BUN: blood urea nitrogen; OR: odds ratio; CI: confidence interval; PVE: proportion of variance explained; LOO: leave one out.

BUN (Z-score = 8.338, $P = 7.53 \times 10^{-17}$). In addition, the abundance of LMAN2 was also associated with increased risks of CKD (Z-score = 3.820, $P = 1.33 \times 10^{-4}$), declined eGFR (Z-score = -4.261, $P = 2.04 \times 10^{-5}$), and elevated BUN (Z-score = 3.261, $P = 1.11 \times 10^{-3}$). Modeled by the enet method, the PWAS also reported a significant association between the abundance of SNUPN and increased risks of CKD (Z-score = 3.290, $P = 1.00 \times 10^{-3}$), declined eGFR (Z-score = -8.129, $P = 4.32 \times 10^{-16}$), and elevated BUN (Z-score = 3.623, $P = 2.91 \times 10^{-4}$). The associations remained significant after correcting for multiple testing (all FDR < 0.05).

To explore the biological role of identified genes, the significant genes in PWAS were enriched by GO and KEGG. The cellular component (CC) involved collagen-containing extracellular matrix, secretory granule lumen, cytoplasmic vesicle lumen, and lysosomal lumen (Fig. S1A). The biological process (BP) terms showed that the identified genes were mainly enriched in immune response pathways (activation of the immune response, mononuclear cell migration, response to interleukin-1, and regulation of antigen

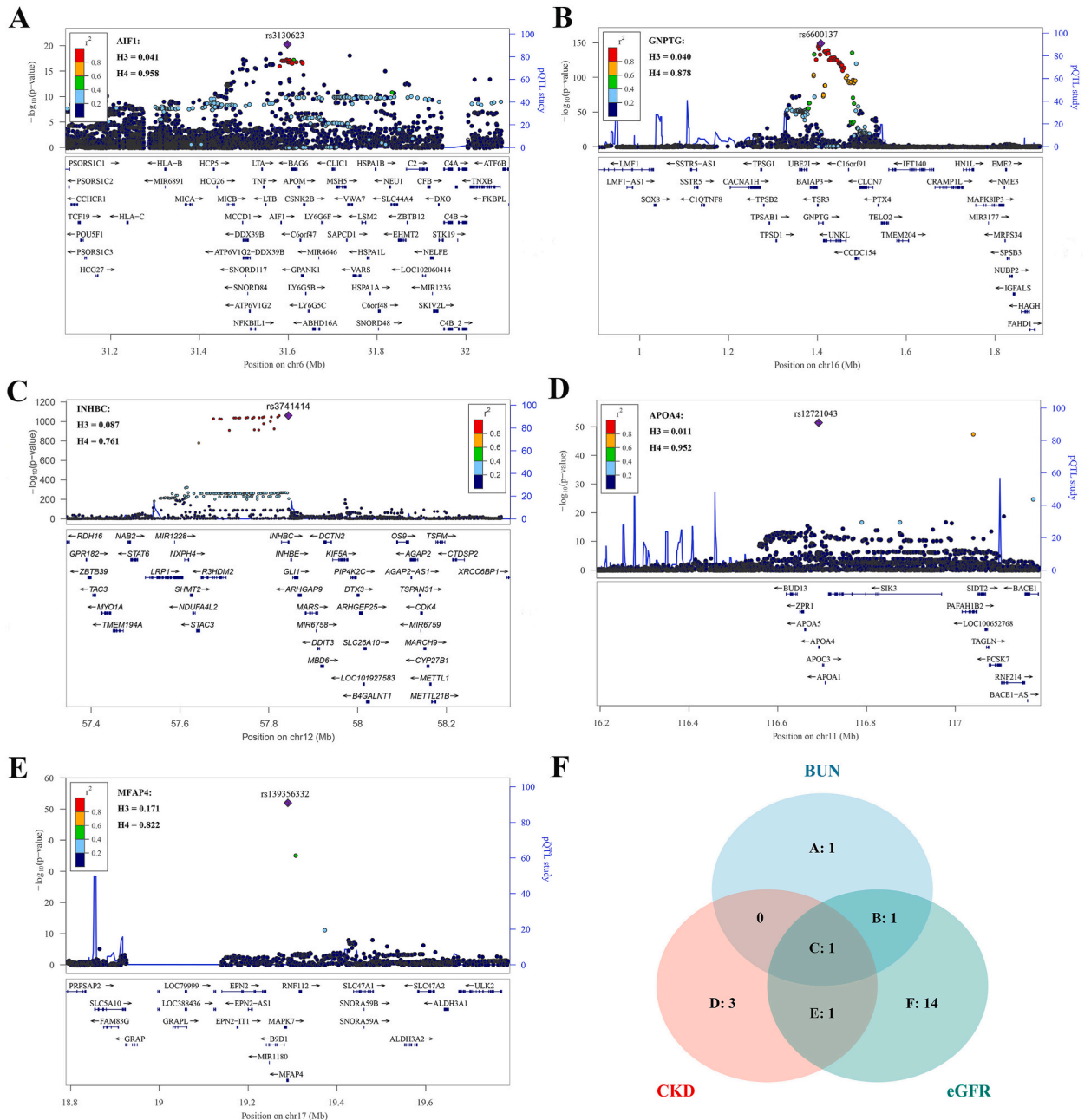


Fig. 4. Illustration of the colocalization results in CKD, eGFR, and BUN. The significant colocalization results in CKD were displayed in panel A–E. The CKD association and pQTL association were plotted at the AIF1, GNPTG, INHBC, APOA4, and MFAP4 loci. Panel F showed the Venn plot for the intersection of significant genes identified in CKD, eGFR, and BUN. CKD: chronic kidney disease; eGFR: estimated glomerular filtration rate; BUN: blood urea nitrogen; Chr: chromosome.

receptor-mediated signaling pathway), indicating the risk role of the immune system in the onset of CKD (Fig. S1B). The molecular function (MF) terms mainly concentrated on cytokine activity and receptor binding including growth factor and fibroblast growth factor (Fig. S1C). The KEGG analysis revealed that the TGF- β signaling pathway and complement and coagulation cascades were involved in CKD (Fig. S1D). The significant genes were further imported into the STRING database to construct the PPI network (Fig. S2A). The constructed network included 79 nodes and 322 edges. The MCODE tool in Cytoscape software identified six key modules in the PPI network (Figs. S2B–S2F). All the MCODE scores were ≥ 3 , and only the first key module had MCODE score ≥ 4 . Ranked by the Betweenness method, the cytoHubba tool detected the hub genes in the PPI network, and the top ten hub genes were MAPK3, VTN, FGF7, GPT, NOG, APOA4, PLG, EFEMP1, GUSB, and SERPINA1.

3.2. Causal associations of identified proteins by PWAS with eGFR, BUN, and CKD

To explore the causal associations between proteins and the risk of CKD, the pQTLs were extracted for further MR analyses. For CKD, the IVW estimator confirmed that 17 out of the 22 significant genes in PWAS showed causal associations (FDR < 0.05 , Fig. 3A, Table S4). Among them, MR-Egger regression detected significant pleiotropy for the IVs of AIF1 and APOA4 ($P < 0.05$). In addition, the IVW method found that 71 out of 78 genes were causally associated with eGFR (FDR < 0.05 , Fig. 3B, Table S5). Two genes (cAMP, and FST) retrieved no qualified IVs, which were excluded from the MR analyses. For BUN, 15 out of the 17 significant genes in PWAS showed causal associations (FDR < 0.05 , Fig. 3B, Table S6).

In Fig. 3B and Table 1, the INHBC gene remained significant across CKD (OR = 1.025, FDR = 2.69×10^{-3}), eGFR (OR = 0.998, FDR = 1.22×10^{-20}), and BUN (OR = 1.008, FDR = 9.00×10^{-21}). Similarly, significant causal associations were also observed between the SNUPN gene and CKD (OR = 1.114, FDR = 6.11×10^{-6}), eGFR (OR = 0.988, FDR = 7.95×10^{-8}), and BUN (OR = 1.015, FDR = 6.51×10^{-3}). Of note, although the LMAN2 gene showed a causal association with BUN (OR = 1.032, FDR = 5.85×10^{-3}), it did not reach the significant threshold for CKD (OR = 1.283, FDR = 5.63×10^{-2}), and eGFR (OR = 0.989, FDR = 0.141).

3.3. Shared causal variants of proteins and CKD

To assess the probability of a shared causal variant driving the alteration of proteins and eGFR, BUN, and CKD, Bayesian colocalization analyses were performed. Of the 119 proteins significant in PWAS, there were 25 proteins sharing a causal variant with CKD, eGFR, and BUN (detailed in Table S7). As shown in Fig. 4A–E, five genes (AIF1, GNPTG, INHBC, APOA4, and MFAP4) disclosed evidence for genetic colocalization for CKD (all PP4 > 0.75). In addition, there were three genes (INHBC, SERPINA1, and ACP1) showing a shared variant with BUN (all PP4 > 0.80 , Fig. S3). For eGFR, a total of 17 genes had a PP4 > 0.75 (Fig. S4). Notably, among them, the INHBC showed evidence of genetic colocalization for CKD (PP4 = 0.761), eGFR (PP4 = 0.918), and BUN (PP4 = 0.953), concurrently (Fig. 4F). The LMAN2 and SNUPN had no shared variants with CKD, eGFR, and BUN (all PP4 < 0.01 , Table 1).

Combining the results from PWAS, MR, and Bayesian colocalization, the INHBC may play a significant role in CKD. Therefore, we further explored the kidney cell-type specificity of the INHBC gene in humans and mice. The single-cell RNA sequencing revealed that both in the healthy human and mouse kidneys, the INHBC gene was sparsely scattered within the kidney cells (Fig. S5).

4. Discussion

In this study, we used pQTL data of plasma proteins and GWAS of CKD to explore their functional associations. A total of 22 proteins were found to be closely associated with CKD, and three of them (encoded by INHBC, LMAN2, and SNUPN) remained significant across CKD, eGFR, and BUN. In addition, the MR analyses for CKD, eGFR, and BUN further provided causal evidence for INHBC and SNUPN, with marginal significance for LMAN2 ($P = 0.0503$). The Bayesian colocalization analyses detected a shared variant for INHBC in CKD, verified in eGFR, and BUN. These genes may be the therapeutic targets for CKD and should be further explored in future studies.

Identifying therapeutic targets is crucial for CKD. This systematic PWAS analysis implicates 22 novel genes in the pathogenesis of CKD, including genes reported previously, such as MFAP4, AIF1, and AGER, as well as new candidates like IDI2, IDH1, and SNUPN. Among them, three genes (INHBC, LMAN2, and SNUPN) were verified in CKD, eGFR, and BUN, which were more likely to be the targets for CKD. According to the Drug-Gene Interaction Database (DGIdb V.4.2.0, <https://www.dgldb.org/>), ten out of the 22 genes (MFAP4, INHBC, AGER, C2, GMPR, APOA4, BTN3A3, C9, IDH1, and PTPRJ) are druggable genes [23]. Notably, there are developed drugs for IDH1 like Ivosidenib [24]. It has been approved by the Food and Drug Administration of the US and shows significant clinical benefit for IDH1-mutated acute myeloid leukemia [25]. However, to date, no study has evaluated the potential of targeting IDH1 in treating CKD. Our results provide novel insights into the therapeutic strategy for CKD.

Given the consistent results of INHBC, LMAN2, and SNUPN in CKD, eGFR, and BUN, the three genes display higher likelihoods to be involved in the pathogenesis of kidney damage. INHBC (Inhibin Subunit Beta C) encodes a member of the TGF- β superfamily and has been shown to play a role in regulating diverse functions such as hormone secretion, cell development, and differentiation [26–28]. Similarly, Schlosser P et al. also found that INHBC was involved in the pathogenesis of CKD [29]. However, this association is not further verified by MR analysis. Our study found that elevated plasma INHBC could causally increase the risk of CKD and the elevation was driven by a causal variant in the *cis*-region of INHBC. It highlights the potential importance of INHBC as another factor contributing to the development of CKD. In addition, INHBC has been reported to be associated with diabetic nephropathy in humans [30]. Du et al. [31] also found that INHBC was over-expressed in the rat's kidney of diabetic nephropathy. However, the biological mechanism linking INHBC to kidney damage and further CKD remains unknown. Further research is needed to fully understand the mechanisms by which elevated plasma INHBC increases the risk of CKD. Given the druggability of INHBC in drug development, it is

worthwhile to investigate whether interventions aimed at reducing plasma INHBC levels could be effective in preventing or delaying the onset of CKD in individuals with elevated INHBC [32].

Our study found that LMAN2 in plasma is associated with a higher risk of CKD. LMAN2, also known as Mannose Binding 2, is a protein-coding gene. It encodes the type I transmembrane lectin. The type I lectin shuttles between the endoplasmic reticulum, Golgi apparatus, and plasma membrane and has been shown to play a role in the sorting, trafficking, and quality control of high mannose glycoproteins [33]. This gene can be expressed in the skeletal muscle and kidney [34]. Pešić I et al. [35] reported that the lectin in urine was significantly higher in the patients with endemic nephropathy, instead in patients with diabetic nephropathy and acute kidney injury. The lectin pathway of complement may be the mechanism responsible for the increased risk of CKD in our results. It was observed that mannose-binding lectin is closely associated with the activation of the complement system, possibly leading to immune dysfunction and further IgA nephropathy, diabetic nephropathy, and renal ischemia/reperfusion injury [36–39]. The association between LMAN2 and CKD also suggests that determining the concentration of plasma lectin may help identify the high-risk individuals of CKD, allowing for earlier intervention and treatment.

SNUPN encodes the Snurportin-1 protein that is involved in the m3G-cap-dependent nuclear import of U small nuclear ribonucleoprotein [40]. Snurportin-1 protein can also regulate a number of genes expression by RNA polymerase II in a chromatin context [41,42]. Previous studies have revealed that SNUPN is associated with osteoarthritis, and acute lymphoblastic leukemia [43,44]. To the best of our knowledge, this is the first study report that the genetic predisposition to SNUPN can causally increase the risk of CKD. Genetic testing for variants in the *cis*-region of SNUPN may be useful in identifying high-risk individuals with CKD. However, the exact mechanism by which elevated SNUPN levels increase the risk of CKD is not yet fully understood. Further research is needed to elucidate the underlying mechanisms and to determine whether targeting SNUPN could be effective in preventing or treating CKD.

This study has some strengths and limitations. The major merit is that we adopted PWAS to implicate novel proteins for CKD, using the largest and comprehensive pQTL data of plasma protein and GWAS of CKD (approximately 500,000 participants). Moreover, the identified proteins were verified in the GWASs of eGFR and BUN. In addition, the associations in PWAS were further advanced to causal inferences using the MR analysis. To assess the probability of a shared causal variant driving the alteration of proteins and phenotypes, Bayesian colocalization analysis was used, confirming that INHBC was a pathogenetic protein for CKD, eGFR, and BUN, simultaneously. These endeavors add more credibility to the findings. However, this study still has some limitations. First, although 4483 plasma proteins were profiled in the pQTL study, the SOMAmer-based method cannot detect the whole plasma proteome. Other plasma proteins not detected in this study may also affect the onset and progress of CKD. Second, this study has strict restrictions on the descent of included samples (only the European descent). It reduces the bias from mixed races (population architecture bias) but also limits the generalizability of the conclusion to other races. Future studies in different nations with larger sample sizes should be considered. Lastly, the pQTL mapping cannot cover all the GWAS signals. At the protein level, we cannot comprehensively explore the pathogenesis of chronic kidney disease. Investigations enrolling methylation QTL, expression QTL, and single-cell sequencing data may provide more targeted therapeutics.

In conclusion, this study identified 22 plasma proteins associated with CKD, and three of them (INHBC, LMAN2, and SNUPN) were replicated in the GWASs of eGFR and BUN. For INHBC, further MR analysis verified the causal association and Bayesian colocalization analysis detected a common causal variant driving the alterations of protein and phenotypes in CKD, eGFR, and BUN. These proteins may be the new therapeutic targets for CKD and should be further explored in future studies.

Ethnics approval and consent to participant

Ethical review and approval can be accessed in the original studies. Informed consent was obtained from all subjects in the original genome-wide association studies. In this MR study, only summary-level statistics were used. No identifiable private information was contained in the GWAS datasets.

Consent for publication

Not applicable.

Data availability statement

The GWAS data of eGFR, BUN, and CKD were retrieved from the CKDGen consortium (round 4, <https://ckdgen.imbi.uni-freiburg.de/>). The pQTL data can be found at <http://nilanjanchatterjeeelab.org/pwas>.

CRediT authorship contribution statement

Yang Xiong: Writing – original draft, Visualization, Software, Formal analysis, Data curation, Conceptualization. **Tianhong Wang:** Conceptualization, Formal analysis, Methodology, Resources, Writing – original draft. **Wei Wang:** Writing – original draft, Formal analysis. **Yangchang Zhang:** Project administration, Methodology, Formal analysis. **Fuxun Zhang:** Validation, Methodology. **Jiuhong Yuan:** Writing – review & editing, Visualization, Formal analysis. **Feng Qin:** Writing – review & editing, Supervision, Conceptualization. **Xianding Wang:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the Natural Science Foundation of China (No. 82071639), 1.3.5 Project for Disciplines of Excellence-Clinical Research Incubation Project, West China Hospital, Sichuan University (No. 2021HXFH007), and Sichuan Science and Technology Program (No. 2023YFS0256). The authors sincerely thank the authors who shared the original dataset in this study.

List of abbreviations

ARIC	atherosclerosis risk in communities
BUN	blood urea nitrogen
CKD	chronic kidney disease
eGFR	estimated glomerular filtration rate
FDR	false discovery rate
GWAS	genome-wide association study
GO	gene ontology
IV	instrumental variables
KEGG	Kyoto encyclopedia of genes and genomes
MR	Mendelian Randomization
MCODE	molecular complex detection
PPI	protein-protein interaction
PWAS	proteome-wide association study
pQTL	protein quantitative trait loci
SNP	single nucleotide polymorphism
STRING	search tool for the retrieval of interacting genes

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.heliyon.2024.e31704>.

References

- [1] P. Romagnani, G. Remuzzi, R. Glassock, A. Levin, K.J. Jager, M. Tonelli, et al., Chronic kidney disease, *Nat. Rev. Dis. Prim.* 3 (2017) 17088.
- [2] B.R. Fletcher, S. Damery, O.L. Aiyegbusi, N. Anderson, M. Calvert, P. Cockwell, et al., Symptom burden and health-related quality of life in chronic kidney disease: a global systematic review and meta-analysis, *PLoS Med.* 19 (2022) e1003954.
- [3] GCKD Collaboration, Global, regional, and national burden of chronic kidney disease, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017, *Lancet* 395 (2020) 709–733.
- [4] K.J. Foreman, N. Marquez, A. Dolgert, K. Fukutaki, N. Fullman, M. Mcgaughey, et al., Forecasting life expectancy, years of life lost, and all-cause and cause-specific mortality for 250 causes of death: reference and alternative scenarios for 2016–40 for 195 countries and territories, *Lancet* 392 (2018) 2052–2090.
- [5] M. Wuttke, Y. Li, M. Li, K.B. Sieber, M.F. Feitosa, M. Gorski, et al., A catalog of genetic loci associated with kidney function from analyses of a million individuals, *Nat. Genet.* 51 (2019) 957–972.
- [6] F.W. Albert, L. Kruglyak, The role of regulatory variation in complex traits and disease, *Nat. Rev. Genet.* 16 (2015) 197–212.
- [7] H.J. Westra, M.J. Peters, T. Esko, H. Yaghootkar, C. Schurmann, J. Kettunen, et al., Systematic identification of trans eQTLs as putative drivers of known disease associations, *Nat. Genet.* 45 (2013) 1238–1243.
- [8] O. Devuyst, C. Pattaro, The UMOD locus: insights into the pathogenesis and prognosis of kidney disease, *J. Am. Soc. Nephrol.* 29 (2018) 713–726.
- [9] M.D. Gallagher, A.S. Chen-Plotkin, The post-GWAS era: from association to function, *Am. J. Hum. Genet.* 102 (2018) 717–730.
- [10] R.F. Dubin, E.P. Rhee, Proteomics and metabolomics in kidney disease, including insights into etiology, treatment, and prevention, *Clin. J. Am. Soc. Nephrol.* 15 (2020) 404–411.
- [11] B.B. Sun, J.C. Maranville, J.E. Peters, D. Stacey, J.R. Staley, J. Blackshaw, et al., Genomic atlas of the human plasma proteome, *Nature* 558 (2018) 73–79.
- [12] J. Zhang, D. Dutta, A. Köttgen, A. Tin, P. Schlosser, M.E. Grams, et al., Plasma proteome analyses in individuals of European and African ancestry identify cis-pQTLs and models for proteome-wide association studies, *Nat. Genet.* 54 (2022) 593–602.
- [13] S.A. Williams, M. Kivimaki, C. Langenberg, A.D. Hingorani, J.P. Casas, C. Bouchard, et al., Plasma protein patterns as comprehensive indicators of health, *Nat. Med.* 25 (2019) 1851–1857.
- [14] J. Yang, S.H. Lee, M.E. Goddard, P.M. Visscher, GCTA: a tool for genome-wide complex trait analysis, *Am. J. Hum. Genet.* 88 (2011) 76–82.
- [15] A. Gusev, A. Ko, H. Shi, G. Bhatia, W. Chung, B.W. Penninx, et al., Integrative approaches for large-scale transcriptome-wide association studies, *Nat. Genet.* 48 (2016) 245–252.
- [16] G. Yu, L.G. Wang, Y. Han, Q.Y. He, clusterProfiler: an R package for comparing biological themes among gene clusters, *OMICS* 16 (2012) 284–287.
- [17] Y. Xiong, F.X. Zhang, Y.C. Zhang, C.J. Wu, F. Qin, J.H. Yuan, Genetically predicted insomnia causally increases the risk of erectile dysfunction, *Asian J. Androl.* 25 (3) (2023) 421–425.
- [18] Y. Xiong, F. Zhang, Y. Zhang, W. Wang, Y. Ran, C. Wu, S. Zhu, F. Qin, J. Yuan, Insights into modifiable risk factors of erectile dysfunction, a wide-angled Mendelian Randomization study, *J. Adv. Res.* 58 (2024) 149–161.

- [19] C. Giambartolomei, D. Vukcevic, E.E. Schadt, L. Franke, A.D. Hingorani, C. Wallace, et al., Bayesian test for colocalisation between pairs of genetic association studies using summary statistics, *PLoS Genet.* 10 (2014) e1004383.
- [20] C. Zhang, F. Qin, X. Li, X. Du, T. Li, Identification of novel proteins for lacunar stroke by integrating genome-wide association data and human brain proteomes, *BMC Med.* 20 (2022) 211.
- [21] H. Wu, K. Uchimura, E.L. Donnelly, Y. Kirita, S.A. Morris, B.D. Humphreys, Comparative analysis and refinement of human PSC-derived kidney organoid differentiation with single-cell Transcriptomics, *Cell Stem Cell* 23 (2018) 869–881.e868.
- [22] H. Wu, Y. Kirita, E.L. Donnelly, B.D. Humphreys, Advantages of single-nucleus over single-cell RNA sequencing of adult kidney: rare cell types and novel cell states revealed in fibrosis, *J. Am. Soc. Nephrol.* 30 (2019) 23–32.
- [23] S.L. Freshour, S. Kiwala, K.C. Cotto, A.C. Coffman, J.F. McMichael, J.J. Song, et al., Integration of the drug-gene interaction database (DGIdb 4.0) with open crowdsourced efforts, *Nucleic Acids Res.* 49 (2021) D1144–d1151.
- [24] G.K. Abou-Alfa, T. Macarulla, M.M. Javle, R.K. Kelley, S.J. Lubner, J. Adeva, et al., Ivosidenib in IDH1-mutant, chemotherapy-refractory cholangiocarcinoma (ClarIDHy): a multicentre, randomised, double-blind, placebo-controlled, phase 3 study, *Lancet Oncol.* 21 (6) (2020 Jun) 796–807.
- [25] P. Montesinos, C. Recher, S. Vives, E. Zarzycka, J. Wang, G. Bertani, et al., Ivosidenib and azacitidine in IDH1-mutated acute myeloid leukemia, *N. Engl. J. Med.* 386 (2022) 1519–1531.
- [26] J. Schmitt, G. Hötten, N.A. Jenkins, D.J. Gilbert, N.G. Copeland, J. Pohl, et al., Structure, chromosomal localization, and expression analysis of the mouse inhibin/activin beta C (Inhbc) gene, *Genomics* 32 (1996) 358–366.
- [27] A.J. Phipps-Green, M.E. Merriman, R. Topless, S. Altaf, G.W. Montgomery, C. Franklin, et al., Twenty-eight loci that influence serum urate levels: analysis of association with gout, *Ann. Rheum. Dis.* 75 (2016) 124–130.
- [28] E.C. Ottley, K.L. Reader, K. Lee, F.E. Marino, H.D. Nicholson, G.P. Risbridger, et al., Over-expression of activin-β(C) is associated with murine and human prostate disease, *Horm. Cancer* 8 (2017) 100–107.
- [29] P. Schlosser, J. Zhang, H. Liu, A.L. Surapaneni, E.P. Rhee, D.E. Arking, B. Yu, E. Boerwinkle, P.A. Welling, N. Chatterjee, K. Susztak, J. Coresh, M.E. Grams, Transcriptome- and proteome-wide association studies nominate determinants of kidney function and damage, *Genome Biol.* 24 (1) (2023 Jun 26) 150.
- [30] H. Yang, D. Lian, X. Zhang, H. Li, G. Xin, Key genes and signaling pathways contribute to the pathogenesis of diabetic nephropathy, *Iran J Kidney Dis* 13 (2019) 87–97.
- [31] X.Y. Du, B.T. Zheng, Y. Pang, W. Zhang, M. Liu, X.L. Xu, et al., The potential mechanism of INHBC and CSF1R in diabetic nephropathy, *Eur. Rev. Med. Pharmacol. Sci.* 24 (2020) 1970–1978.
- [32] C. Finan, A. Gaulton, F.A. Kruger, R.T. Lumbers, T. Shah, J. Engmann, et al., The druggable genome and support for target identification and validation in drug development, *Sci. Transl. Med.* 9 (2017) eaag1166.
- [33] D. Zhang, L. Ye, S. Hu, Q. Zhu, C. Li, C. Zhu, Comprehensive analysis of the expression and prognostic value of LMN2 in HER2+ breast cancer, *J Immunol Res* 2022 (2022) 7623654.
- [34] E.P. Neve, K. Svensson, J. Fuxe, R.F. Pettersson, VIPL, a VIP36-like membrane protein with a putative function in the export of glycoproteins from the endoplasmic reticulum, *Exp. Cell Res.* 288 (2003) 70–83.
- [35] I. Pešić, V. Stefanović, G.A. Müller, C.A. Marino, O. Jahn, et al., Identification and validation of six proteins as marker for endemic nephropathy, *J. Proteomics* 74 (2011) 1994–2007.
- [36] A. Roos, M.R. Daha, J. Van Pelt, S.P. Berger, Mannose-binding lectin and the kidney, *Nephrol. Dial. Transplant.* 22 (2007) 3370–3377.
- [37] M. Matsuda, K. Shikata, J. Wada, H. Sugimoto, Y. Shikata, T. Kawasaki, et al., Deposition of mannan binding protein and mannan binding protein-mediated complement activation in the glomeruli of patients with IgA nephropathy, *Nephron* 80 (1998) 408–413.
- [38] T.K. Hansen, L. Tarnow, S. Thiel, R. Steffensen, C.D. Stehouwer, C.G. Schalkwijk, et al., Association between mannose-binding lectin and vascular complications in type 1 diabetes, *Diabetes* 53 (2004) 1570–1576.
- [39] J.E. Jordan, M.C. Montalto, G.L. Stahl, Inhibition of mannose-binding lectin reduces posts ischemic myocardial reperfusion injury, *Circulation* 104 (2001) 1413–1418.
- [40] E. Paraskeva, E. Izaurre, F.R. Bischoff, J. Huber, U. Kutay, E. Hartmann, et al., CRM1-mediated recycling of snurportin 1 to the cytoplasm, *J. Cell Biol.* 145 (1999) 255–264.
- [41] S. Li, A.R. Almeida, C.A. Radebaugh, L. Zhang, X. Chen, L. Huang, et al., The elongation factor Spn1 is a multi-functional chromatin binding protein, *Nucleic Acids Res.* 46 (2018) 2321–2334.
- [42] V. Pujari, C.A. Radebaugh, J.V. Chodaparambil, U.M. Muthurajan, A.R. Almeida, J.A. Fischbeck, et al., The transcription factor Spn1 regulates gene expression via a highly conserved novel structural motif, *J. Mol. Biol.* 404 (2010) 1–15.
- [43] L. Hao, X. Shang, Y. Wu, J. Chen, S. Chen, Construction of a diagnostic m(7)G regulator-mediated scoring model for identifying the characteristics and immune landscapes of osteoarthritis, *Biomolecules* 13 (2023) 539.
- [44] M. Mata-Rocha, A. Rangel-López, E. Jiménez-Hernández, B.A. Morales-Castillo, C. González-Torres, J. Gaytan-Cervantes, et al., Identification and characterization of novel fusion genes with potential clinical applications in Mexican children with acute lymphoblastic leukemia, *Int. J. Mol. Sci.* 20 (2019) 2394.