Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Locus-aware transcriptome-wide Mendelian randomization prioritizes candidate loci and expression traits for chronic kidney disease

Wenda Wu
Shanxi Medical University
Lei Wang
Shanxi Bethune Hospital, Shanxi Academy of Medical Sciences, Third Hospital of Shanxi Medical University
Jinming Guo
Shanxi Medical University
Xinfang Cao
Shanxi Medical University
Xiaomin Yang
Shanxi Medical University
Yu Liu
yuliu@sxmu.edu.cn

Shanxi Medical University

Article

Keywords: Mendelian randomization, chronic kidney disease, expression quantitative trait loci, colocalization, major histocompatibility complex

Posted Date: October 22nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11185479/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Locus-aware transcriptome-wide Mendelian randomization prioritizes candidate loci and expression traits for chronic kidney disease

Wenda Wu[1,+], Lei Wang[2,+], Jinming Guo[1], Xinfang Cao[1], Xiaomin Yang[1], and Yu Liu[1,*]

[1]Department of Pharmacology, Shanxi Medical University, Xinjiannanlu 56, Taiyuan, Shanxi Province 030001, China
[2]Department of Geriatric Medicine, Shanxi Bethune Hospital, Shanxi Academy of Medical Sciences, Third Hospital of Shanxi Medical University, Tongji Shanxi Hospital, No 99 Longcheng street, Taiyuan 030032, Shanxi Province, China
[*]Corresponding: yuliu@sxmu.edu.cn
[+]these authors contributed equally to this work

Abstract

Background

Transcriptome-wide Mendelian randomization (MR) can identify expression traits associated with disease, but correlated cis-regulation may cause multiple neighboring traits to reflect the same underlying locus. We investigated chronic kidney disease (CKD) using a locus-aware framework that interpreted genetically connected expression traits at the locus level rather than as independent gene-level findings.

Methods

We tested 15,285 blood cis-eQTL-defined expression traits for association with CKD in FinnGen. Single-instrument traits were analyzed using the Wald ratio, while traits with multiple instruments were analyzed using multiplicative random-effects inverse-variance weighted MR (MRE-IVW). Traits meeting the FDR threshold underwent locus mapping, colocalization, assessment in an independent CKD dataset, and integration of renal transcriptomic, kidney eQTL, and single-nucleus evidence.

Results

The primary analysis identified 28 FDR-significant gene-expression traits, 12 of which also met the Bonferroni threshold. These traits were grouped into nine operationally defined candidate loci, with 19 mapping to a single complex locus within the major histocompatibility complex (MHC). Strong colocalization was observed for *ANKUB1* and *RSAD2* at distinct non-MHC loci and for *IER3* within the MHC. After excluding the MHC, nine FDR-significant traits remained across eight non-MHC loci, including two Bonferroni-significant traits and two with strong colocalization. Among the five traits eligible for multi-instrument sensitivity analyses, none showed a significant MR-Egger intercept. One TUBB leave-one-out omission met the prespecified influence criterion. In the primary exact-rsID CKDGen analysis, 14 traits were estimable and 11 showed direction-concordant effects, although none remained significant after replication-wide FDR correction. Identifier harmonization enabled all 28 traits to be evaluated and identified four FDR-significant, direction-concordant traits, all within the MHC. These findings strengthened support for the MHC locus without resolving gene attribution within the region.

Conclusions

Transcriptome-wide MR identified 28 CKD-associated expression traits across nine candidate loci. *ANKUB1* and *RSAD2* were prioritized as distinct non-MHC candidates with strong colocalization support by colocalization, whereas *IER3* remained embedded within an unresolved complex MHC locus despite strong single-signal colocalization. These findings support targeted functional follow-up without establishing any candidate as a validated causal gene or therapeutic target.

Keywords: Mendelian randomization, chronic kidney disease, expression quantitative trait loci, colocalization, major histocompatibility complex

Introduction

Chronic kidney disease (CKD) is a major contributor to morbidity and premature mortality worldwide. Although several therapies can slow disease progression, substantial residual risk remains [1–3]. Large genome-wide association studies have identified numerous loci associated with kidney function and CKD-related traits. However, translating these locus-level associations into specific genes and biological mechanisms remains challenging [4, 5].

Expression quantitative trait loci (eQTLs) link genetic variation to gene expression and can therefore serve as genetic proxies for transcriptional regulation in transcriptome-wide Mendelian randomization and related association analyses [6–10]. However, an expression trait associated with disease does not necessarily represent an independently resolved gene. Neighboring genes may share regulatory variants, while linkage disequilibrium can cause multiple expression traits to track the same disease-associated haplotype [11]. Colocalization can assess whether gene-expression and disease-association signals are consistent with a shared underlying variant, but evidence from a single-signal model does not by itself establish that one gene is the unique effector at the locus [12–15].

This issue is particularly important in the MHC, where high gene density, extensive polymorphism, long-range linkage disequilibrium, and coordinated regulation of multiple genes complicate the interpretation of association signals [16, 17]. Expression associations arising from the same MHC region may therefore reflect correlated regulatory architecture rather than independent gene-level effects. Treating these associations as separate discoveries can inflate the apparent number of findings and obscure the attribution of disease-associated signals to specific genes.

In this study, we performed a transcriptome-wide cis-eQTL Mendelian randomization analysis of CKD and applied a consistent downstream evaluation to all expression traits that met the FDR threshold. Significant traits were grouped into genetically connected loci, and the contribution of the MHC to the overall discovery pattern was assessed

explicitly. Colocalization, an independent CKD outcome dataset, renal transcriptomic evidence, kidney eQTL data, and single-nucleus expression profiles were then integrated to refine locus-level interpretation. The study was designed to distinguish correlated cis-regulatory associations from more clearly resolved candidate loci and expression traits, while retaining negative, unavailable, and directionally discordant evidence. The overall aim was to prioritize a limited set of candidates for functional investigation without interpreting correlated trait-level associations as independent causal genes.

Methods

Study design and data sources

The study was conducted and reported in accordance with STROBE-MR recommendations [18]. Blood cis-eQTL summary statistics were obtained from the eQTLGen Consortium. In the source dataset, cis-eQTL pairs were defined as variants located within 1 Mb of the gene centre and tested in at least two participating cohorts [6]. Blood eQTLs were selected as the primary exposure resource because they enabled a consistent transcriptome-wide analysis across a large set of expression traits. Kidney eQTL datasets were used to provide complementary tissue-specific regulatory context rather than as formal replication datasets, because their smaller sample sizes and available summary-statistic formats did not permit a comparable transcriptome-wide analysis.

The primary outcome was the registry-defined chronic kidney disease endpoint N14_CHRONKIDNEYDIS from FinnGen release 13, which included 14,402 cases and 477,611 controls. The independent CKDGen dataset comprised participants of European ancestry and defined CKD primarily as an estimated glomerular filtration rate below 60 mL/min/1.73 m². This dataset included 41,395 cases and 439,303 controls [4, 19]. An overview of the study workflow and the characteristics of the data sources is provided in **Figure 1** and **Table 1**.

Instrument selection, harmonization and primary MR analysis

Genetic instruments were selected at $P < 5 \times 10^{-8}$ and clumped separately for each expression trait using the European 1000 Genomes reference panel, with an LD threshold of $r^2 < 0.001$ within a 10-Mb window. Instrument strength was assessed using $F = (beta_{exposure}/SE_{exposure})^2$, and all retained instruments had $F > 10$ [20]. After harmonization, the analytical dataset contained 28,915 gene-instrument rows representing 15,285 expression traits.

eQTLGen provided Z scores rather than effect estimates and standard errors. We therefore reconstructed the standard error as $SE = [2p(1-p)(N+Z^2)] \wedge -1/2$ and calculated the effect estimate as $beta = Z \times SE$, where p represented the assessed-allele frequency and N represented the SNP-gene sample size. Allele-frequency records were aligned to the assessed allele before reconstruction. The reported AlleleB frequency was retained when AlleleB corresponded to the assessed allele. When AlleleA corresponded to the assessed allele, the frequency was converted to $1 - AF_B$. Reconstruction and allele-orientation checks were required to agree within numerical tolerance before the variant was retained.

FinnGen outcome variants were matched first by rsID. When an rsID match was unavailable, matching was attempted using a unique lifted genomic coordinate together with a compatible allele pair. Direct, swapped, strand-complement, and strand-complement-swapped orientations were resolved before effect alleles were aligned. Palindromic variants with intermediate allele frequencies were excluded under action 2 harmonization. A variant was retained only when a single compatible FinnGen outcome record could be identified. Variants with missing outcome data, incompatible alleles, or multiple compatible outcome records were recorded and excluded rather than resolved through arbitrary row selection.

The 15,285 Ensembl v71 expression traits represented 15,285 unique gene identifiers and 15,245 gene symbols used for

display. Repeated symbols therefore corresponded to distinct Ensembl identifiers rather than duplicated analytical traits. No probe-level selection or aggregation was performed (Supplementary Table S10b).

Traits represented by a single instrument were analyzed using the Wald ratio. For traits with at least two instruments, multiplicative random-effects inverse-variance weighted MR was used. The residual dispersion factor was constrained to a minimum value of one so that residual underdispersion could not produce a standard error smaller than the corresponding fixed-effect estimate. The standard error was calculated as $SE_{MRE} = SE_{FE} \times \sqrt{\max(1, Q/df)}$.

This approach allows residual heterogeneity to increase the standard error but does not allow underdispersion to reduce it [21]. Benjamini-Hochberg false-discovery rates were calculated across all 15,285 expression traits. The Bonferroni-corrected significance threshold was $P < 3.27 \times 10^{-6}$. Additional implementation-level validation and computational cross-checks are described in the Supplementary Methods.

MR sensitivity analyses

Cochran's Q was calculated for traits with at least two instruments. MR-Egger, weighted-median, and leave-one-out analyses were performed for traits with at least three instruments, while MR-PRESSO was applied to traits with at least four instruments [22–24]. A leave-one-out deletion was prespecified as influential if removal of a single variant reversed the direction of the primary effect estimate or changed the beta by more than 50% relative to the estimate obtained using all retained instruments.

Steiger directionality analysis was performed exploratorily for all FDR-significant traits [25]. SNP-specific correlations with the exposure were reconstructed from the exposure effect estimate, standard error, and sample size. Correlations with the outcome were estimated using `get_r_from_lor` with 14,402 cases, 477,611 controls, and assumed population prevalences of 0.05, 0.10, and 0.15. Because this approach uses a prevalence-informed latent-logit

approximation rather than a conventional liability-scale or probit-scale transformation, the resulting directionality evidence was considered supportive rather than definitive. SNP-specific squared correlations were summed across the retained instruments before comparison using `mr_steiger`. Analyses that did not meet the required instrument count were classified as `NOT_APPLICABLE`.

Colocalization analysis

All FDR-significant expression traits underwent `coloc.abf` analysis within the same gene-centred ±500-kb window. Both the eQTL and CKD datasets were supplied using effect estimates and their corresponding variances. For the quantitative eQTL model, we used the aligned eQTLGen effect estimate, standard error, assessed-allele frequency, and variant-specific sample size, with `sdY = 1`. For the case-control CKD model, we used the FinnGen effect estimate, standard error, and alternate-allele frequency, with a total sample size of 492,013 and a case fraction of $14{,}402/492{,}013$.

The prior probabilities were initially specified as $p_1 = p_2 = 1 \times 10^{-4}$ and $p_{12} = 1 \times 10^{-5}$ [12]. In regions containing large numbers of matched variants, the effective values of p_1 and p_2 were capped at $0.99/m$, where m was the number of matched variants. The effective value of p_{12} was further constrained so that it did not exceed either p_1 or p_2. Colocalization was classified as strong when $\text{PP.H4} \geqslant 0.80$, moderate when $0.50 \leqslant \text{PP.H4} < 0.80$, and insufficiently supported when $\text{PP.H4} < 0.50$. Prior sensitivity analyses varied the requested value of p_{12} from 1×10^{-6} to 1×10^{-4} while applying the same validity constraints.

SuSiE-based colocalization was attempted only for regions in which aligned local linkage disequilibrium information covered at least 95% of the required variants [14]. Regions that did not meet this coverage threshold were not analyzed with SuSiE. Analyses that failed to converge were recorded as unavailable rather than interpreted as evidence against

colocalization.

Locus definition and MHC sensitivity analysis

The 28 FDR-significant expression traits were grouped into connected candidate loci after completion of the primary MR analysis. Two traits were considered connected if their gene coordinates were within 1 Mb of each other, if they shared an MR instrument, or if their lead eQTLs were in linkage disequilibrium at $r^2 \geqslant 0.1$ in the available European reference panel. Transitive connections were then used to define operational locus groups for interpretation. These groups were not considered to represent conditionally independent causal signals.

The reference MHC interval was defined as GRCh37 chr6:28,477,797–33,448,354. Its local architecture and the broader extended-MHC context were interpreted with reference to established MHC maps and regulatory studies [16, 17]. An MHC-excluded analysis was prespecified as a sensitivity analysis to assess the extent to which this complex region influenced the overall discovery pattern. It was interpreted alongside, rather than as a replacement for, the MHC-inclusive primary analysis.

The robustness of the locus definition was further examined by reducing the physical-distance threshold to 500 kb, varying the lead-eQTL linkage-disequilibrium threshold to $r^2 = 0.05$ and $r^2 = 0.2$, and applying a definition based only on shared instruments or linkage disequilibrium. For each alternative definition, we recorded the total number of connected components and assessed whether the key candidate loci remained stable.

Independent CKD outcome analysis

CKDGen summary statistics were used for independent outcome assessment. The primary analysis used exact rsID matching without linkage-disequilibrium proxies or post hoc variant substitution. Of the 50 requested variants, 22 were matched, allowing 14 expression traits to be estimated. A prespecified sensitivity analysis additionally allowed unique

coordinate-and-allele matching, which recovered 27 variants and left one unavailable. No proxy or nearest-variant substitution was used.

Trait-level effects were estimated using the same Wald-ratio and MRE-IVW procedures as in the primary analysis. Directional concordance with FinnGen, nominal significance, and FDR were evaluated separately. The exact-rsID analysis remained primary, while the identifier-harmonized analysis was treated as a sensitivity analysis.

Renal transcriptomic analysis

Renal transcriptomic datasets were identified from Gene Expression Omnibus metadata before inspection of candidate-level results [26, 27]. Twenty-two GEO Series contributed 61 prespecified disease and compartment contrasts. One contrast did not meet the eligibility criterion, leaving 60 contrasts for analysis. The renal-expression analysis evaluated prespecified individual contrasts, including GSE66494 as the primary dataset and GSE104954 as an independent dataset.

For each contrast, an approximate standardized mean difference was derived from the limma moderated t statistic as $d = t_{mod}\sqrt{(1/n_{case} + 1/n_{control})}$. Hedges' g was then calculated as $g = Jd$ where $J = 1 - 3/(4df - 1)$ and $df = n_{case} + n_{control} - 2$. Sampling variance was estimated using the prespecified large-sample approximation $Var(g) = (n_{case} + n_{control})/(n_{case}\, n_{control}) + g^2/(2df)$ for its sampling variance [28]. Because these effect sizes were derived from limma moderated t statistics rather than directly from group means and standard deviations, they were treated as approximate standardized effects. They were used for cross-study visualization and biological context rather than as causal estimates.

Within groups defined by disease, tissue compartment, and dataset independence, the dataset with the largest sample size was retained. When sample sizes were equal, the dataset with the smaller standard error was selected. Renal meta-

analysis was conducted separately from the individual contrast analysis and required at least two independent groups within the same prespecified disease and compartment stratum. Random-effects models were fitted using `metafor::rma.uni` with restricted maximum likelihood estimation and z-based inference [29, 30]. Global FDR correction was applied across the 284 estimable strata. Representative strata were selected according to cohort count, total sample size, and pooled standard error without consideration of statistical significance. IER3 was prespecified as the anchor expression trait for L03. Raw expression measurements from different platforms were not pooled.

Kidney eQTL and single-nucleus localization

Kidney regulatory evidence was assessed using GTEx v10 kidney cortex data and the Human Kidney eQTL Atlas meta-analysis, tubule, and glomerular datasets [31–34]. Because the available Atlas files were significance-truncated, they were used only to assess eGene status and allele-resolved overlap with MR instruments and were not used for colocalization. Single-nucleus localization was evaluated using 19,985 nuclei from five normal adult kidney-cortex donors [35]. Sparse count matrices were processed in batches of 500 nuclei and summarized at the donor and author-defined cell-type levels. No differential-expression analysis was performed.

Trait- and locus-level evidence integration

The main evidence landscape integrated eight genetic and contextual domains. Each domain was classified as supported, no support detected, unavailable, direction-discordant, or not applicable. MR support was defined as $\mathrm{FDR} < 0.05$, and Bonferroni support required $P < 3.27 \times 10^{-6}$. Colocalization support required $\mathrm{PP.H4} \geq 0.50$, with strong and moderate colocalization retained as separate categories in the underlying evidence table. Independent CKD support required an estimable effect that was direction-concordant with the primary FinnGen estimate and nominally significant at $P < 0.05$.

Criteria for renal expression, renal meta-analysis, kidney eQTL evidence, and single-nucleus localization are provided in the Supplementary Methods and Supplementary Table S22. IER3 was used consistently as the anchor expression trait for L03. Evidence from other MHC traits was not combined with IER3 to generate gene-specific support.

Evidence from multiple expression traits within the same locus was not summed. The main evidence landscape was therefore descriptive and did not generate an overall numerical ranking. Alternative weighting schemes and different approaches to unavailable evidence were examined only in supplementary sensitivity analyses.

Results

Transcriptome-wide Mendelian randomization identified 28 FDR-significant expression traits

The primary analysis included 15,285 cis-eQTL-defined expression traits. Of these, 28 were associated with CKD at a Benjamini–Hochberg false-discovery rate (FDR) below 0.05, and 12 also met the Bonferroni-corrected significance threshold (**Figure 2**; Supplementary Table S2). Instrument availability was limited among the significant traits. Thirteen were represented by a single instrument, 10 by two instruments, and only five by three or more instruments. When these 28 expression traits were grouped according to the prespecified locus definition, they formed nine operationally defined candidate loci. Nineteen traits mapped to a single complex MHC locus, indicating that the 28 significant expression traits did not represent 28 independent gene discoveries.

Primary MR estimates

Among the FDR-significant non-MHC traits, *ANKUB1* was estimated using a single-instrument Wald ratio and was associated with higher CKD odds ($\mathrm{OR} = 2.488$, 95% CI 1.701-3.638, $\mathrm{P} = 2.59\times10^{-6}$, $\mathrm{FDR} = 0.00330$, $\mathrm{nSNP} = 1$). *RSAD2* was also estimated using a single-instrument Wald ratio and was associated with lower CKD odds ($\mathrm{OR}=0.506$, 95% CI 0.360-0.712, $\mathrm{P} = 9.10\times10^{-5}$, $\mathrm{FDR} = 0.0497$, $\mathrm{nSNP} = 1$) (**Figure 3** and **Figure 4**). Within the

complex MHC locus, *IER3* was estimated using MRE-IVW and was associated with higher CKD odds ($OR = 1.133$, 95% CI 1.070-1.198, $P = 1.61 \times 10^{-5}$, $FDR = 0.0154$, $nSNP = 5$). The *IER3* estimate represents an expression trait within the MHC and does not resolve gene attribution within this complex locus.

Sensitivity analyses of eligible multi-instrument traits

Sensitivity analyses were performed only when the required number of instruments was available (Supplementary Figures S2–S6 and Supplementary Table S7). Scatter plots comparing MR methods, leave-one-out plots, funnel plots, and SNP-level forest plots were generated for the five traits with at least three retained instruments: *C4A*, *IER3*, *PSORS1C3*, *RP11-137H2.4*, and *TUBB*. Cochran's Q could be calculated for 15 traits and showed nominal evidence of heterogeneity only for IER3 ($Q = 10.80$, $df = 4$, $P = 0.0289$). MR-Egger analysis was possible for all five traits with at least three instruments, and none of the intercept tests reached $P < 0.05$. For *IER3*, the intercept P value was 0.797. Because these analyses were based on only three to five instruments, the MR-Egger intercept tests had limited power, and non-significant results cannot exclude directional pleiotropy.

Weighted-median estimates were directionally consistent with the primary estimates for all five eligible traits, including *IER3* ($OR = 1.131$). MR-PRESSO could be applied to three traits with at least four instruments. None of the global tests was significant, and no outliers were identified. For *IER3*, the global P value was 0.354. Exploratory Steiger analyses favoured the exposure-to-outcome direction for all 28 traits under each prevalence assumption examined. These findings were treated as supportive rather than definitive because the analysis relied on a latent-logit approximation rather than a conventional liability- or probit-scale transformation.

Leave-one-out analysis was available for five traits. No single omission reversed the direction of effect. *TUBB* was the only trait that met the prespecified influence rule, as removing rs149543464 changed the primary beta by more than

50%. No omission met this criterion for *IER3*. Funnel-plot symmetry was not interpreted for the two traits with only three instruments. Overall, the sensitivity analyses provided no statistical evidence of directional pleiotropy or dependence of the *IER3* estimate on a single instrument. However, the limited number of instruments meant that these analyses could not establish uniform robustness across all multi-instrument traits.

Significant expression traits mapped to nine operationally defined candidate loci

Accordingly, the 28 FDR-significant expression traits represented nine candidate loci rather than 28 independent genes or causal signals (**Table 2**). L01 contained *RSAD2*, L02 contained *ANKUB1*, and L03 represented the complex MHC locus. The remaining loci comprised *KIAA1919* (L04), *BICC1* (L05), *RP11-137H2.4* (L06), *CLNS1A* (L07), *SKOR1* (L08), and *AGPHD1/PSMA4* (L09). *AGPHD1* and *PSMA4* shared the same MR instrument and were therefore grouped within the same locus.

Nineteen of the 28 significant expression traits were located within the MHC locus. These were *GNL1*, *TUBB*, *IER3*, *PSORS1C3*, *HCP5*, *LTA*, *APOM*, *DDAH2*, *HSPA1L*, *NEU1*, *EHMT2*, *C4A*, *TNXA*, *C4B*, *CYP21A2*, *ATF6B*, *PRRT1*, *RNF5*, and *TAP1*. The primary result should therefore be interpreted as 28 FDR-significant expression traits distributed across nine candidate loci rather than as 28 independent genes or causal signals.

The overall nine-locus structure was unchanged when the lead-eQTL LD threshold was set to either $r^2 = 0.05$ or $r^2 = 0.2$. Reducing the physical-distance threshold to 500 kb increased the number of loci to 10 and placed 18, rather than all 19, MHC traits within the component containing *IER3*. When the physical-distance criterion was removed entirely, the number of connected components increased to 25 because most MHC lead variants were either unavailable or only weakly linked in the reference panel. Despite these changes, *ANKUB1* and *RSAD2* remained separate singleton loci under every definition examined, while *AGPHD1* and *PSMA4* consistently remained grouped because they shared

the same MR instrument (Supplementary Figure S15 and Supplementary Table S11).

Excluding the MHC reduced trait-level discovery counts

The MHC-inclusive primary analysis identified 28 FDR-significant traits, 12 Bonferroni-significant traits, and three traits with strong colocalization. Excluding the MHC reduced these counts to nine FDR-significant traits, two Bonferroni-significant traits, and two strong-colocalization traits across eight non-MHC loci (Supplementary Figure S7 and Supplementary Table S9). The primary MR estimates were not recalculated. The reduction therefore reflected exclusion of MHC-associated traits rather than re-estimation of the primary effect estimates.

Colocalization supported two distinct non-MHC candidates and one unresolved MHC signal

All 28 FDR-significant expression traits underwent `coloc.abf` analysis, with at least 2,268 matched variants available in each region (**Figure 3** and **Figure 4**). Strong colocalization was observed for *ANKUB1* (PP.H4 = 0.969), *RSAD2* (PP.H4 = 0.953), and *IER3* (PP.H4 = 0.997). Moderate colocalization was found for *CLNS1A*, *KIAA1919*, and four MHC traits, including *C4A*, *C4B*, *CYP21A2*, and *TNXA*. The remaining 19 traits showed insufficient evidence of colocalization.

ANKUB1 and *RSAD2* showed strong colocalization at two distinct non-MHC loci. *IER3* also showed strong colocalization under the single-signal model, but its signal was located within the complex MHC locus. Because this region contained several correlated significant expression traits and no convergent multi-signal fine-mapping result was available, *IER3* was not considered an independently resolved causal gene. Two regions met the linkage-disequilibrium coverage requirement for SuSiE analysis, but neither analysis converged. These analyses were therefore recorded as unavailable rather than interpreted as evidence for or against colocalization.

Independent CKD outcome assessment

In the primary exact-rsID CKDGen analysis, 14 of the 28 traits were estimable, and 11 showed the same direction of effect as in FinnGen (**Figure 5**). *ANKUB1* was the only trait with nominal evidence of association ($\mathrm{OR} = 1.388$, 95% CI 1.049-1.836, $\mathrm{P} = 0.0216$), although none remained significant after replication-wide FDR correction. *IER3*, *C4A*, and *SKOR1* showed effects in the opposite direction. The other 14 traits were unavailable under the exact-rsID matching rule, and this missingness was not interpreted as evidence against association.

The identifier-harmonized sensitivity analysis recovered 27 uniquely matched coordinate-and-allele aliases, leaving only one requested variant unavailable. This allowed all 28 traits to be evaluated (Supplementary Figure S20 and Supplementary Tables S24–S25). Twenty-one estimates were direction-concordant with the FinnGen results, five reached nominal significance, and four remained significant after FDR correction in the sensitivity analysis. These were *APOM* ($\mathrm{OR} = 0.786$, 95% CI 0.717-0.861, $\mathrm{FDR} = 6.77 \times 10^{-6}$), *ATF6B* ($\mathrm{OR} = 1.144$, 95% CI 1.070-1.223, $\mathrm{FDR} = 0.00123$), *NEU1* ($\mathrm{OR} = 0.803$, 95% CI 0.717-0.899, $\mathrm{FDR} = 0.00135$), and *PRRT1* ($\mathrm{OR} = 1.215$, 95% CI 1.071-1.379, $\mathrm{FDR} = 0.0179$). All four traits were located within L03. These findings provided additional locus-level support for L03 but did not resolve a single causal gene within the region and did not replace the primary exact-rsID analysis.

Renal and kidney regulatory evidence was heterogeneous

The renal analysis yielded 1,415 eligible standardized contrast rows, of which 1,124 independent rows were retained across 676 disease–compartment strata. Random-effects models could be fitted for 284 strata, while 392 strata were represented by a single group and were retained for descriptive purposes only. For the locus-aware display, one representative renal stratum was selected for each of seven evaluable loci according to the predefined hierarchy of cohort count, total sample size, and pooled standard error. Statistical significance was not used in this selection process

(**Figure 6**).

Where available, the representative strata included three or four independent cohorts. *IER3*, which served as the anchor trait for L03, showed higher expression in the representative focal segmental glomerulosclerosis (FSGS) glomerular stratum (Hedges $g = 0.770$, 95% CI 0.403-1.137, global FDR $= 0.000846$). In contrast, the representative strata for *ANKUB1* and *RSAD2* were not statistically significant. *IER3* also showed FDR-significant effects opposite to the MR direction in several other disease and compartment strata, including tubulointerstitial contrasts in IgA nephropathy, minimal-change disease, and lupus nephritis. L03 was therefore classified as having mixed renal evidence rather than uniformly concordant or discordant evidence. All evaluated strata, including the diabetic-kidney-disease glomerular stratum for *ANKUB1*, are reported in the Supplementary Information without significance-based selection. Because disease-associated expression may reflect upstream mechanisms, downstream injury responses, treatment effects, or changes in cell composition, these renal findings were interpreted as contextual evidence rather than causal replication.

In the kidney eQTL analyses, six of the 28 traits were eGenes in GTEx kidney cortex. The Kidney eQTL Atlas identified 17 eGenes in the meta-analysis, 14 in tubule, and 17 in glomeruli. Allele-resolved instrument overlaps concordant with the MR direction were observed for *C4A* and *PSORS1C3* in GTEx, for *KIAA1919* in the Atlas meta-analysis and glomerular dataset, and for *RP11-137H2.4* in the Atlas meta-analysis. *BICC1* showed direction-discordant overlaps across all three Atlas resources. Because the Atlas datasets include related samples, these observations were not treated as independent replications.

Twenty-two of the 28 candidate traits were detected in the normal-kidney single-nucleus atlas. These data provided information on expression across the author-defined kidney cell populations but were not used to assess differential expression or causal mediation.

Locus-aware evidence landscape distinguished genetic from contextual evidence

The nine-row evidence landscape distinguished genetic evidence from contextual renal evidence and did not assign an overall rank (**Figure 7**). L02, represented by *ANKUB1*, showed FDR- and Bonferroni-significant MR evidence, strong colocalization, nominal direction-concordant evidence in the exact-rsID CKDGen analysis, and localization in the single-nucleus dataset. L03 was represented by *IER3* throughout the main display. Its strong single-signal colocalization was accompanied by direction-discordant evidence in the exact-rsID CKDGen analysis, MHC-level support in the identifier-harmonized sensitivity analysis, mixed renal meta-analytic evidence, kidney eQTL support, and single-nucleus localization. L03 therefore combined strong genetic evidence with mixed contextual renal evidence.

The supplementary scoring analyses were sensitive to both domain weighting and the treatment of unavailable evidence, indicating that a single numerical ranking was not well supported. Regardless of the scoring scheme, *ANKUB1* and *RSAD2* remained the distinct non-MHC traits with strong colocalization, whereas *IER3* remained an anchor expression trait within the unresolved MHC locus.

Discussion

This study identified 28 CKD-associated expression traits that grouped into nine operationally defined candidate loci. Much of the apparent discovery burden was driven by a single complex MHC locus. The primary MR estimates themselves did not change, but the locus-aware interpretation substantially changed what could reasonably be inferred from them. The findings therefore support a set of associated expression traits and candidate loci rather than 28 independent causal genes. The nine locus groups were used as practical units for interpretation and should not be regarded as nine conditionally independent CKD causal signals.

ANKUB1 and *RSAD2* provided the clearest strong-colocalization evidence outside the MHC. *ANKUB1* was prioritized

because it combined Bonferroni-significant MR evidence, location in a distinct non-MHC locus, strong colocalization, and nominal direction-concordant association in the primary exact-rsID CKDGen analysis. This prioritization was not based on the magnitude of its odds ratio or on a presumed biological mechanism. Its CKDGen association did not survive replication FDR correction, and its renal expression estimates varied across disease–compartment strata. *ANKUB1* also remains poorly characterized in kidney biology. Its genetic association should therefore guide functional investigation rather than substitute for it. *RSAD2* also showed strong colocalization, but its CKDGen estimate was only direction-concordant without nominal significance, and its representative renal meta-analytic estimate was direction-discordant. Existing studies linking *RSAD2* to renal and interferon-related processes provide biological plausibility but do not establish that the same pathway mediates the MR association. Both *ANKUB1* and *RSAD2* were supported by a single instrument. MR-Egger, weighted-median, funnel, and leave-one-out analyses were therefore not applicable, and instrument-specific horizontal pleiotropy cannot be formally excluded despite the strong regional colocalization. These findings justify focused functional follow-up but do not validate either gene as causal or therapeutic.

BICC1 provides a useful example of why MR significance alone was not sufficient for prioritization. Its primary MR association exceeded the Bonferroni threshold, yet colocalization support was insufficient and both kidney regulatory evidence and renal meta-analytic direction were discordant. This pattern does not invalidate the MR estimate. It does, however, weaken a gene-specific interpretation of the locus and illustrates why candidates should not be ranked solely by MR P values.

Gene-level association and locus-level causal interpretation are not equivalent. Cis-eQTL MR estimates the association between genetically proxied expression and disease, but correlated regulation within the same region can generate significant results for several neighboring expression traits. This was particularly evident in the MHC, where 19 significant traits mapped to one complex locus. Trait-level significance was therefore not interpreted as evidence for 19

independent causal genes.

The MHC findings also highlight the limitations of single-signal colocalization. *IER3* had the highest PP.H4, while *C4A*, *C4B*, *CYP21A2*, and *TNXA* also showed moderate colocalization within the same gene-dense region. Extensive linkage disequilibrium and multi-gene regulatory effects can produce similar disease-associated patterns across several expression traits [15–17]. The defensible interpretation is therefore that *IER3* showed strong colocalization under the single-signal model within the complex MHC locus. Resolving a unique effector would require reliable conditional or multi-signal fine-mapping together with tissue-relevant functional evidence.

The independent outcome and renal analyses further narrowed the interpretation of the primary findings. In the primary exact-rsID CKDGen analysis, only one of the 14 estimable traits showed nominal evidence of association, and none survived replication FDR correction. Identifier harmonization increased coverage and identified four FDR-significant, direction-concordant traits, but all four were located within the complex MHC locus. This added support at the level of L03 without resolving gene attribution within the region. Renal expression analyses provided additional biological context but showed both concordant and discordant patterns. Such variation is consistent with disease-state transcription reflecting not only upstream mechanisms but also tissue injury, treatment effects, and changes in cellular composition. Kidney eQTL and single-nucleus data added information on regulatory activity and anatomical localization, but neither provided evidence of causal mediation. Earlier CKD MR studies often interpreted significant expression traits, proteins, or selected transporters at the trait or gene level, whereas an integrative *DACH1* study combined transcriptomic evidence with detailed functional experiments [36–41]. The present study instead applies a consistent locus-aware interpretation across a transcriptome-wide screen. Its contribution lies in the interpretive framework rather than in a new MR estimator or in experimental validation.

Experimental studies have linked *IER3* to the transition from acute kidney injury to CKD and have implicated *RSAD2* in interferon responses and podocyte differentiation [42–44]. These observations provide hypotheses for follow-up rather than confirmation of the present genetic findings. Kidney multi-omics and glomerular proteomic studies have also implicated MHC-related complement components, although they do not resolve the causal gene within the locus identified here [45, 46].

Several features strengthen the study. The primary MR analysis was finalized before the downstream analyses were repeated, and every FDR-significant trait underwent the same colocalization workflow. The complete set of 28 traits, including all 19 MHC traits, was retained in the supplementary material rather than selectively reported. The CKDGen sensitivity analysis preserved the primary exact-rsID analysis while separately evaluating uniquely recoverable identifier aliases. Locus mapping, the prespecified MHC-excluded sensitivity analysis, and explicit evidence states reduced the risk of counting correlated traits or unavailable data as independent support. Sensitivity analyses of locus definitions and evidence weighting also showed why a single numerical ranking would not be appropriate.

Several limitations should be considered. Blood cis-eQTL effects may not reflect regulation within specific kidney compartments or disease states. Available kidney eQTL datasets provided useful regulatory context but could not support a uniform transcriptome-wide replication analysis comparable with eQTLGen because of their smaller sample sizes and restricted summary-statistic structures. Thirteen of the 28 significant traits relied on single-instrument Wald ratios and therefore could not undergo multi-instrument pleiotropy or influence analyses. Standard MR assumptions remain relevant, including the possibility of horizontal pleiotropy, winner selection, exposure–outcome sample overlap, and limited generalizability beyond predominantly European or Finnish populations.

Clumping and locus connectivity were based on European 1000 Genomes linkage disequilibrium. This reference may

not fully capture Finnish founder-population LD and could therefore influence instrument independence or the connections used to define loci. MR-Egger analyses were available for only five traits, each with three to five instruments, which limited power to detect directional pleiotropy. A non-significant intercept should therefore not be interpreted as evidence that pleiotropy is absent. The exploratory Steiger analysis used a prevalence-informed latent-logit approximation rather than a conventional liability- or probit-scale transformation, so its directionality results were considered supportive rather than definitive. IER3 also showed nominal heterogeneity by Cochran's Q, although leave-one-out analysis did not indicate dependence on a single instrument.

Colocalization results depended on the choice of priors and on the assumptions of the single-signal model. Available linkage-disequilibrium data were also insufficient for convergent SuSiE inference. The FinnGen registry-defined CKD endpoint and the predominantly eGFR-based CKDGen case definition capture related but non-identical disease constructs, which may contribute to differences between datasets. The identifier-harmonized CKDGen sensitivity analysis depended on unique coordinate-and-allele matching and strengthened inference only at the MHC-locus level. Residual cohort overlap also cannot be excluded.

The renal analyses were heterogeneous with respect to disease, tissue compartment, platform, and control-group size. Standardized effect sizes derived from moderated t statistics were approximate and were used only for visualization and contextual synthesis. Bulk-tissue expression also reflects differences in cellular composition. Kidney eQTL sample sizes were relatively modest, and significance-truncated files from the Kidney eQTL Atlas were not suitable for colocalization. The single-nucleus atlas included only five normal kidney-cortex donors. Finally, no perturbation experiment was performed to test gene function or therapeutic potential directly.

Conclusion

In conclusion, transcriptome-wide MR identified 28 CKD-associated expression traits that mapped to nine operationally defined candidate loci. *ANKUB1* and *RSAD2* were prioritized as distinct non-MHC candidates with strong colocalization support, whereas the MHC findings remained unresolved at the gene level within a single complex locus. This locus-aware framework offers a transparent basis for prioritizing candidate loci and expression traits for functional follow-up without conflating correlated trait-level associations with validated causal genes.

Declarations

Ethics approval and consent to participate

This study involved secondary analyses of publicly available de-identified or summary-level human genetic and transcriptomic data. The authors did not recruit participants, collect human specimens, conduct interventions, or access identifiable individual-level information. No additional institutional ethics approval or informed consent was therefore required for the present analyses. Ethics approval, participant consent, and data-governance procedures were obtained by the original studies and consortia in accordance with their respective protocols.

Availability of data and materials

All analyses were based on publicly available summary statistics and transcriptomic resources. Dataset accession numbers, versions, source locations, and relevant access or redistribution restrictions are provided in Supplementary Table S1. Non-restricted derived data supporting the reported analyses are provided in the Supplementary Information and accompanying source-data files. Original datasets subject to source-specific access or redistribution terms are not redistributed by the authors. Analysis code and supplementary materials are publicly available at https://doi.org/10.5281/zenodo.22852132.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the Fundamental Research Program of Shanxi Province (Grant No. 202203021211068 to L.W.) and the Traditional Chinese Medicine Science and Technology Project of the Health Commission of Shanxi Province (Grant No. 2025ZYYB079 to L.Y.).

Authors' contributions

W.W. and L.W. performed the data analyses and drafted the manuscript. J.G., X.C., and X.Y. contributed to data visualization and the preparation of figures and tables. Y.L. supervised the study and coordinated the overall research project. All authors reviewed the manuscript, approved the final version, and agreed to be accountable for the work.

Acknowledgments

We want to acknowledge the participants and investigators of the FinnGen study. We also thank the participants and investigators of the eQTLGen Consortium, CKDGen Consortium, GTEx Project, Gene Expression Omnibus studies, Human Kidney eQTL Atlas, and the kidney single-nucleus transcriptomic resource used in this study.

References

1. GBD Chronic Kidney Disease Collaboration. Global, regional, and national burden of chronic kidney disease, 1990-2017: A systematic analysis for the global burden of disease study 2017. Lancet. 2020;395:709–33. https://doi.org/10.1016/S0140-6736(20)30045-3.

2. Heerspink Hiddo J L, Stefánsson Bergur V, Correa-Rotter Ricardo, Chertow Glenn M, Greene Tom, Hou Fan-Fan, et al. Dapagliflozin in patients with chronic kidney disease. N Engl J Med. 2020;383:1436–46.

https://doi.org/10.1056/NEJMoa2024816.

3. The EMPA-KIDNEY Collaborative Group, Herrington William G, Staplin Natalie, Wanner Christoph, Green Jennifer B, Hauske Sibylle J, et al. Empagliflozin in patients with chronic kidney disease. N Engl J Med. 2023;388:117–27. https://doi.org/10.1056/NEJMoa2204233.

4. Wuttke Matthias, Li Yong, Li Man, Sieber Karsten B, Feitosa Mary F, Gorski Mathias, et al. A catalog of genetic loci associated with kidney function from analyses of a million individuals. Nat Genet. 2019;51:957–72. https://doi.org/10.1038/s41588-019-0407-x.

5. Stanzick Kira J, Li Yong, Schlosser Pascal, Gorski Mathias, Wuttke Matthias, Thomas Laurent F, et al. Discovery and prioritization of variants and genes for kidney function in >1.2 million individuals. Nat Commun. 2021;12:4350. https://doi.org/10.1038/s41467-021-24491-0.

6. Võsa Urmo, Claringbould Annique, Westra Harm-Jan, Bonder Marc Jan, Deelen Patrick, Zeng Biao, et al. Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. Nat Genet. 2021;53:1300–10. https://doi.org/10.1038/s41588-021-00913-z.

7. Hemani Gibran, Zheng Jie, Elsworth Benjamin, Wade Kaitlin H, Haberland Valeriia, Baird Denis, et al. The MR-base platform supports systematic causal inference across the human phenome. Elife. 2018;7:e34408. https://doi.org/10.7554/eLife.34408.

8. Zhu Zhihong, Zhang Futao, Hu Han, Bakshi Andrew, Robinson Matthew R, Powell Joseph E, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. Nat Genet. 2016;48:481–7. https://doi.org/10.1038/ng.3538.

9. Porcu Eleonora, Rüeger Sina, Lepik Kaido, eQTLGen Consortium, BIOS Consortium, Santoni Federico A, et al. Mendelian randomization integrating GWAS and eQTL data reveals genetic determinants of complex and clinical traits. Nat Commun. 2019;10:3300. https://doi.org/10.1038/s41467-019-10936-0.

10. Gusev Alexander, Ko Arthur, Shi Huwenbo, Bhatia Gaurav, Chung Wonil, Penninx Brenda W J H, et al. Integrative approaches for large-scale transcriptome-wide association studies. Nat Genet. 2016;48:245–52. https://doi.org/10.1038/ng.3506.

11. Wainberg Michael, Sinnott-Armstrong Nasa, Mancuso Nicholas, Barbeira Alvaro N, Knowles David A, Golan David, et al. Opportunities and challenges for transcriptome-wide association studies. Nat Genet. 2019;51:592–9. https://doi.org/10.1038/s41588-019-0385-z.

12. Giambartolomei Claudia, Vukcevic Damjan, Schadt Eric E, Franke Lude, Hingorani Aroon D, Wallace Chris, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet. 2014;10:e1004383. https://doi.org/10.1371/journal.pgen.1004383.

13. Hormozdiari Farhad, van de Bunt Martijn, Segrè Ayellet V, Li Xiao, Joo Jong Wha J, Bilow Michael, et al. Colocalization of GWAS and eQTL signals detects target genes. Am J Hum Genet. 2016;99:1245–60. https://doi.org/10.1016/j.ajhg.2016.10.003.

14. Wang Gao, Sarkar Abhishek, Carbonetto Peter, Stephens Matthew. A simple new approach to variable selection in regression, with application to genetic fine mapping. J R Stat Soc Series B Stat Methodol. 2020;82:1273–300. https://doi.org/10.1111/rssb.12388.

15. Wallace Chris. Eliciting priors and relaxing the single causal variant assumption in colocalisation analyses. PLoS Genet. 2020;16:e1008720. https://doi.org/10.1371/journal.pgen.1008720.

16. Horton Roger, Wilming Laurens, Rand Vikki, Lovering Ruth C, Bruford Elspeth A, Khodiyar Varsha K, et al. Gene map of the extended human MHC. Nat Rev Genet. 2004;5:889–99. https://doi.org/10.1038/nrg1489.

17. D'Antonio Matteo, Reyna Joaquin, Jakubosky David, Donovan Margaret Kr, Bonder Marc-Jan, Matsui Hiroko, et al. Systematic genetic analysis of the MHC region reveals mechanistic underpinnings of HLA type associations with disease. Elife. 2019;8:e48476. https://doi.org/10.7554/eLife.48476.

18. Skrivankova Veronika W, Richmond Rebecca C, Woolf Benjamin A R, Yarmolinsky James, Davies Neil M, Swanson Sonja A, et al. Strengthening the reporting of observational studies in epidemiology using mendelian randomization: The STROBE-MR statement. JAMA. 2021;326:1614–21. https://doi.org/10.1001/jama.2021.18236.

19. Kurki Mitja I, Karjalainen Juha, Palta Priit, Sipilä Timo P, Kristiansson Kati, Donner Kati M, et al. FinnGen provides genetic insights from a well-phenotyped isolated population. Nature. 2023;613:508–18. https://doi.org/10.1038/s41586-022-05473-8.

20. 1000 Genomes Project Consortium, Auton Adam, Brooks Lisa D, Durbin Richard M, Garrison Erik P, Kang Hyun Min, et al. A global reference for human genetic variation. Nature. 2015;526:68–74. https://doi.org/10.1038/nature15393.

21. Burgess Stephen, Butterworth Adam, Thompson Simon G. Mendelian randomization analysis with multiple genetic variants using summarized data. Genet Epidemiol. 2013;37:658–65. https://doi.org/10.1002/gepi.21758.

22. Bowden Jack, Davey Smith George, Burgess Stephen. Mendelian randomization with invalid instruments: Effect estimation and bias detection through egger regression. Int J Epidemiol. 2015;44:512–25. https://doi.org/10.1093/ije/dyv080.

23. Bowden Jack, Davey Smith George, Haycock Philip C, Burgess Stephen. Consistent estimation in mendelian randomization with some invalid instruments using a weighted median estimator. Genet Epidemiol. 2016;40:304–14. https://doi.org/10.1002/gepi.21965.

24. Verbanck Marie, Chen Chia-Yen, Neale Benjamin, Do Ron. Detection of widespread horizontal pleiotropy in causal

relationships inferred from mendelian randomization between complex traits and diseases. Nat Genet. 2018;50:693–8. https://doi.org/10.1038/s41588-018-0099-7.

25. Hemani Gibran, Tilling Kate, Davey Smith George. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. PLoS Genet. 2017;13:e1007081. https://doi.org/10.1371/journal.pgen.1007081.

26. Barrett Tanya, Wilhite Stephen E, Ledoux Pierre, Evangelista Carlos, Kim Irene F, Tomashevsky Maxim, et al. NCBI GEO: Archive for functional genomics data sets--update. Nucleic Acids Res. 2013;41 Database issue:D991-5. https://doi.org/10.1093/nar/gks1193.

27. Davis Sean, Meltzer Paul S. GEOquery: A bridge between the gene expression omnibus (GEO) and BioConductor. Bioinformatics. 2007;23:1846–7. https://doi.org/10.1093/bioinformatics/btm254.

28. Ritchie Matthew E, Phipson Belinda, Wu Di, Hu Yifang, Law Charity W, Shi Wei, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43:e47. https://doi.org/10.1093/nar/gkv007.

29. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. Statistics in Medicine. 2002;21:1539–58. https://doi.org/10.1002/sim.1186.

30. Viechtbauer W. Conducting meta-analyses in *R* with the **metafor** package. J Stat Soft. 2010;36. https://doi.org/10.18637/jss.v036.i03.

31. Qiu Chengxiang, Huang Shizheng, Park Jihwan, Park YoSon, Ko Yi-An, Seasock Matthew J, et al. Renal compartment-specific genetic variation analyses identify new pathways in chronic kidney disease. Nat Med. 2018;24:1721–31. https://doi.org/10.1038/s41591-018-0194-4.

32. GTEx Consortium. The GTEx consortium atlas of genetic regulatory effects across human tissues. Science. 2020;369:1318–30. https://doi.org/10.1126/science.aaz1776.

33. Sheng X, Guan Y, Ma Z, Wu J, Liu H, Qiu C, et al. Mapping the genetic architecture of human traits to cell types in the kidney identifies mechanisms of disease and potential treatments. Nat Genet. 2021;53:1322–33. https://doi.org/10.1038/s41588-021-00909-9.

34. Liu H, Doke T, Guo D, Sheng X, Ma Z, Park J, et al. Epigenomic and transcriptomic analyses define core cell types, genes and targetable mechanisms for kidney disease. Nat Genet. 2022;54:950–62. https://doi.org/10.1038/s41588-022-01097-w.

35. Muto Yoshiharu, Wilson Parker C, Ledru Nicolas, Wu Haojia, Dimke Henrik, Waikar Sushrut S, et al. Single cell transcriptional and chromatin accessibility profiling redefine cellular heterogeneity in the adult human kidney. Nat Commun. 2021;12:2190. https://doi.org/10.1038/s41467-021-22368-w.

36. Liang Run-Sen, Su Jin-Qi, Wu Xiang-Qi, Wang Qi, Cai Yong-Mei, Su Hong-Yong, et al. Identification of

therapeutic targets for chronic kidney disease through mendelian randomization analysis of druggable genes. Sci Rep. 2025;15:10779. https://doi.org/10.1038/s41598-025-94761-0.

37. Zhao Pin, Li Zhenhao, Xue Shilong, Cui Jinshan, Zhan Yonghao, Zhu Zhaowei, et al. Proteome-wide mendelian randomization identifies novel therapeutic targets for chronic kidney disease. Sci Rep. 2024;14:22114. https://doi.org/10.1038/s41598-024-72970-3.

38. Lanktree Matthew B, Perrot Nicolas, Smyth Andrew, Chong Michael, Narula Sukrit, Shanmuganathan Meera, et al. A novel multi-ancestry proteome-wide mendelian randomization study implicates extracellular proteins, tubular cells, and fibroblasts in estimated glomerular filtration rate regulation. Kidney Int. 2023;104:1170–84. https://doi.org/10.1016/j.kint.2023.08.025.

39. Ueda Masatoshi, Fukui Kenji, Kamatani Naoyuki, Kamitsuji Shigeo, Matsuo Akira, Sasase Tomohiko, et al. GLUT9 as a potential drug target for chronic kidney disease: Drug target validation by a mendelian randomization study. J Hum Genet. 2023;68:699–704. https://doi.org/10.1038/s10038-023-01168-8.

40. Gill Dipender, Zagkos Loukas, Gill Rubinder, Benzing Thomas, Jordan Jens, Birkenfeld Andreas L, et al. The citrate transporter SLC13A5 as a therapeutic target for kidney disease: Evidence from mendelian randomization to inform drug development. BMC Med. 2023;21:504. https://doi.org/10.1186/s12916-023-03227-5.

41. Doke Tomohito, Huang Shizheng, Qiu Chengxiang, Liu Hongbo, Guan Yuting, Hu Hailong, et al. Transcriptome-wide association analysis identifies DACH1 as a kidney disease risk gene that contributes to fibrosis. J Clin Invest. 2021;131:e141801. https://doi.org/10.1172/JCI141801.

42. Dou Qian, Tong Hang, Chen Juan, Yi Xiangling, Bai Lihua, He Junling, et al. IER3 drives the transition from sepsis-associated AKI to CKD by suppressing the mitochondrial translocation of PRDX5. Cell Mol Life Sci. 2025;82:439. https://doi.org/10.1007/s00018-025-05963-8.

43. Batten I, Robinson MW, White A, Walsh C, Fazekas B, Wyse J, et al. Investigation of type I interferon responses in ANCA-associated vasculitis. Sci Rep. 2021;11:8272. https://doi.org/10.1038/s41598-021-87760-4.

44. Li Zhigui, Yin Hongqiang, Hao Shuang, Wang Lifeng, Gao Jing, Tan Xiaoyue, et al. miR-200 family promotes podocyte differentiation through repression of RSAD2. Sci Rep. 2016;6:27105. https://doi.org/10.1038/srep27105.

45. Si Shucheng, Liu Hongyan, Xu Lu, Zhan Siyan. Identification of novel therapeutic targets for chronic kidney disease and kidney function by integrating multi-omics proteome with transcriptome. Genome Med. 2024;16:84. https://doi.org/10.1186/s13073-024-01356-x.

46. Sethi Sanjeev, Palma Lilian Monteiro P, Theis Jason D, Fervenza Fernando C. Proteomic analysis of complement proteins in glomerular diseases. Kidney Int Rep. 2023;8:827–36. https://doi.org/10.1016/j.ekir.2023.01.030.

Figures

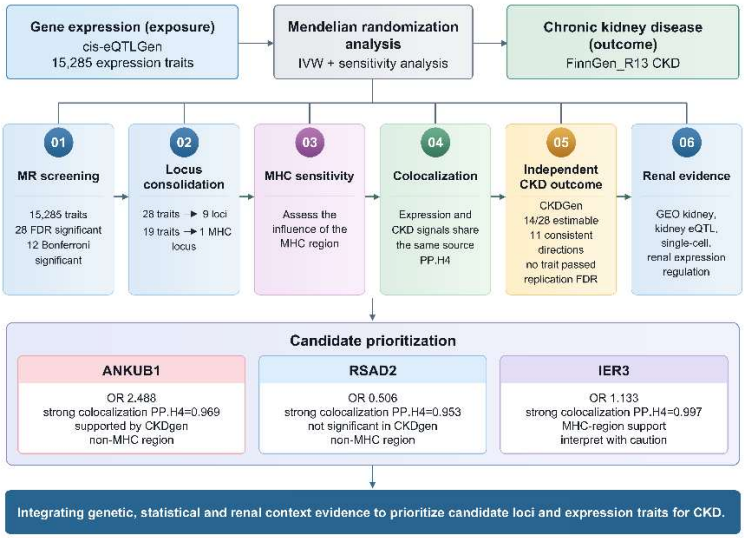


Figure 1. Locus-aware transcriptome-wide MR workflow. The primary analysis tested 15,285 expression traits and identified 28 FDR-significant traits, including 12 that met the Bonferroni threshold. These traits mapped to nine operationally defined candidate loci, including one complex MHC locus containing 19 traits. An MHC-excluded sensitivity analysis and downstream colocalization, independent CKD outcome assessment, and renal evidence integration were subsequently performed.

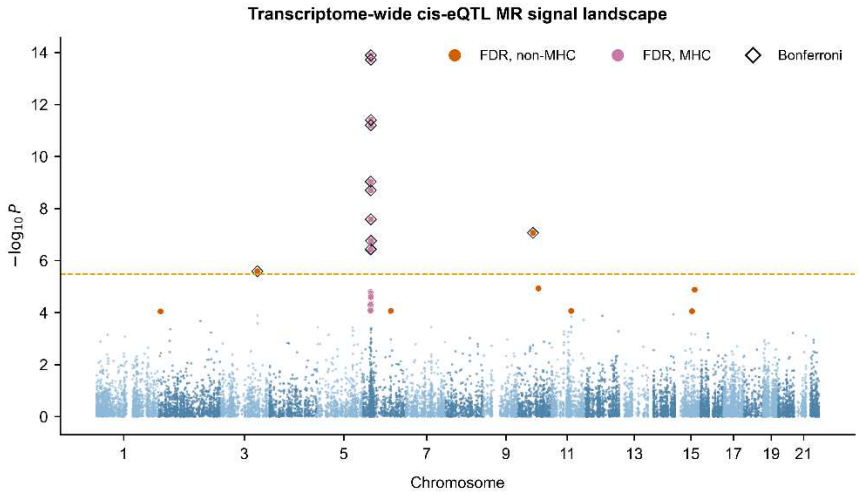


Figure 2. Transcriptome-wide cis-eQTL MR signal landscape. Chromosomal position and -$\log_{10}$ P are shown for all 15,285 primary MR results. Colour distinguishes FDR-significant MHC and non-MHC traits, open diamonds indicate Bonferroni-significant traits, and the horizontal line marks the Bonferroni threshold. Nineteen significant traits mapped to a single complex MHC locus.

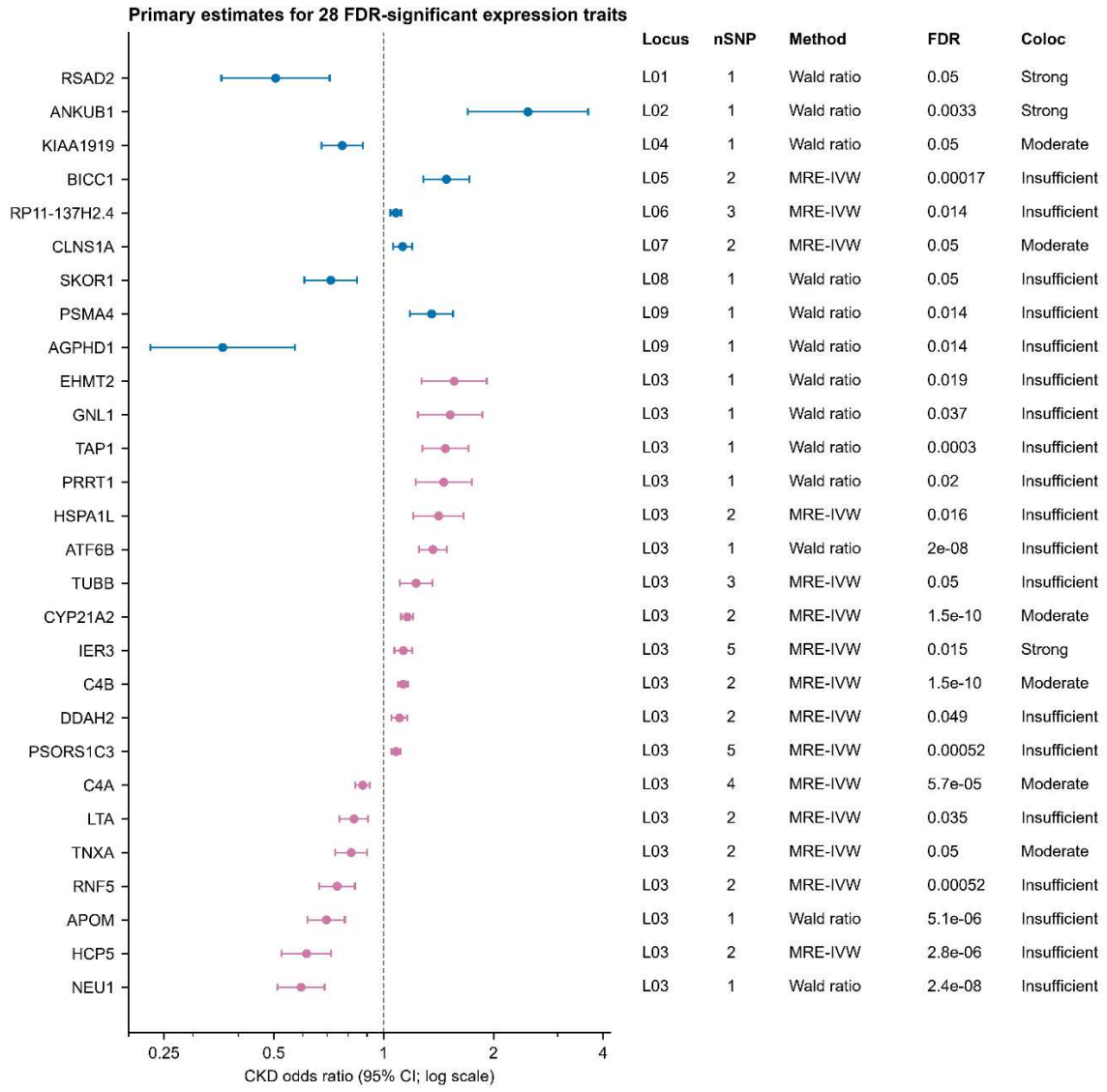


Figure 3. Primary MR estimates for the 28 FDR-significant expression traits. Odds ratios and 95% confidence intervals are shown with locus assignment, instrument count, MR method, and colocalization category. MHC and non-MHC traits are visually distinguished. The 28 traits mapped to nine candidate loci, including 19 traits within one complex MHC locus.

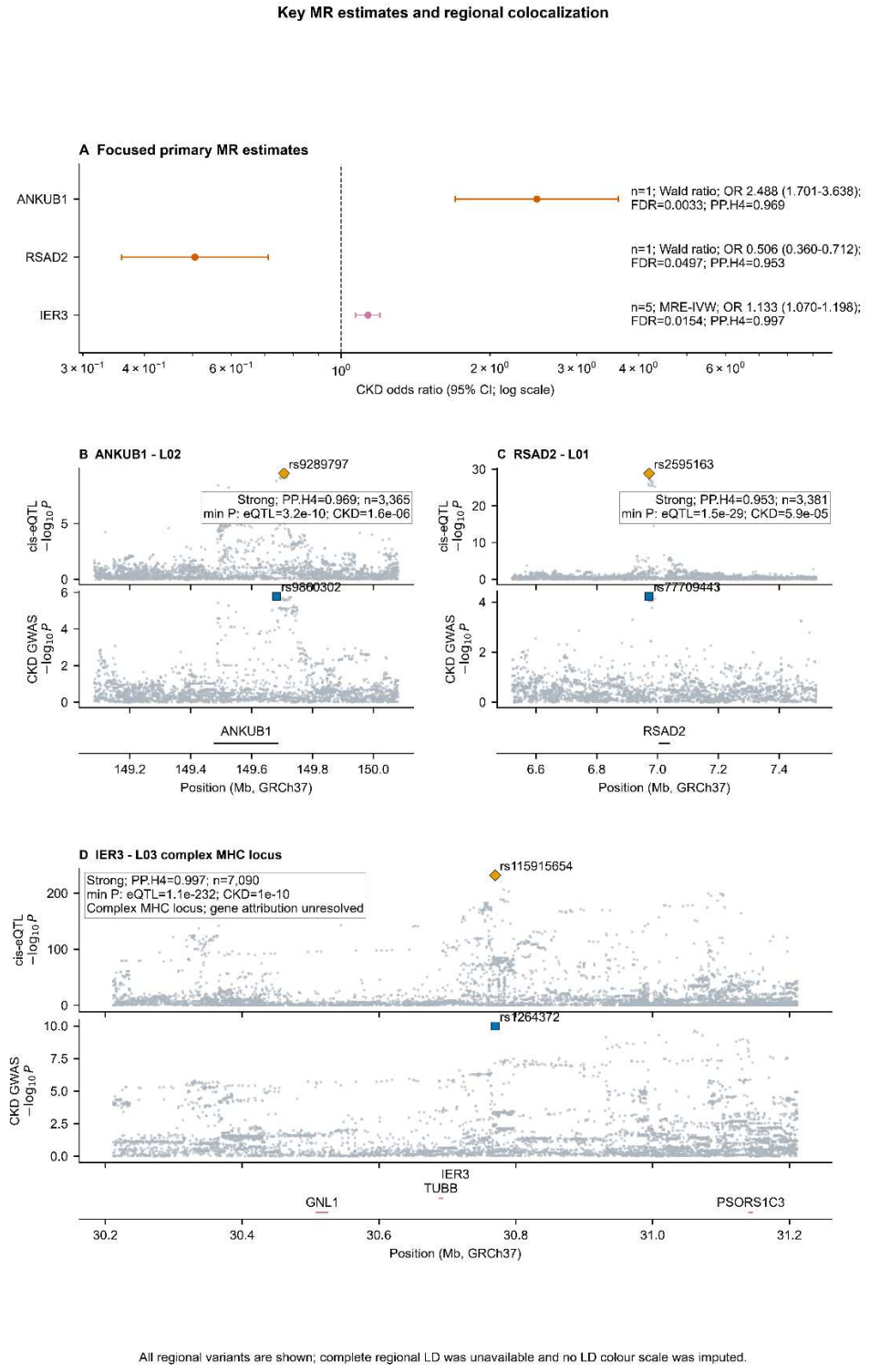


Figure 4. Key MR estimates and regional colocalization for prioritized loci. Panel A summarizes the primary MR and colocalization results for *ANKUB1*, *RSAD2*, and *IER3*. Panels B to D show regional cis-eQTL and CKD association patterns for the corresponding loci. IER3 lies within an unresolved complex MHC locus, and strong single-signal colocalization does not establish it as the unique effector.

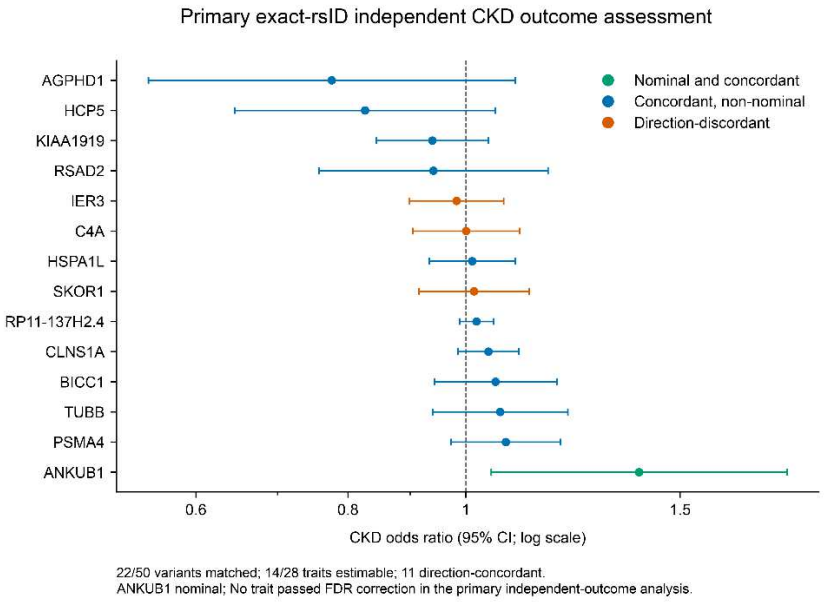


Figure 5. Primary exact-rsID independent CKD outcome assessment. Odds ratios and 95% confidence intervals are shown for the 14 traits estimable by exact-rsID matching. Color indicates direction concordance and nominal significance relative to the primary FinnGen results. Identifier-harmonized sensitivity results are presented separately in the Supplementary Information.

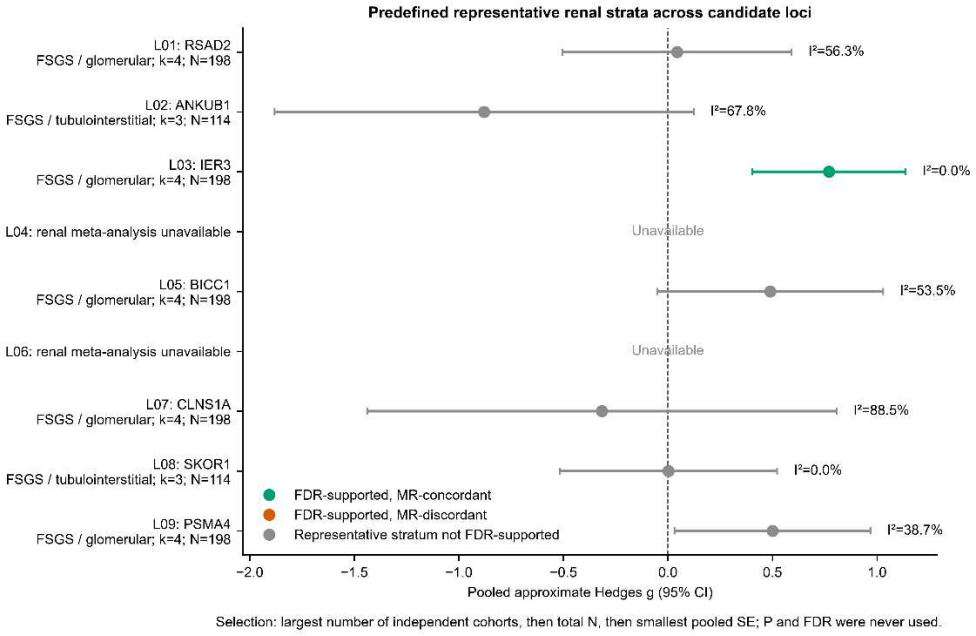


Figure 6. Locus-aware renal multi-cohort synthesis. The representative renal stratum is shown for each of seven evaluable loci. Selection was based on cohort count, total sample size, and pooled standard error without consideration of statistical significance. *IER3* served as the L03 anchor, and renal expression findings were interpreted as contextual rather than causal evidence.

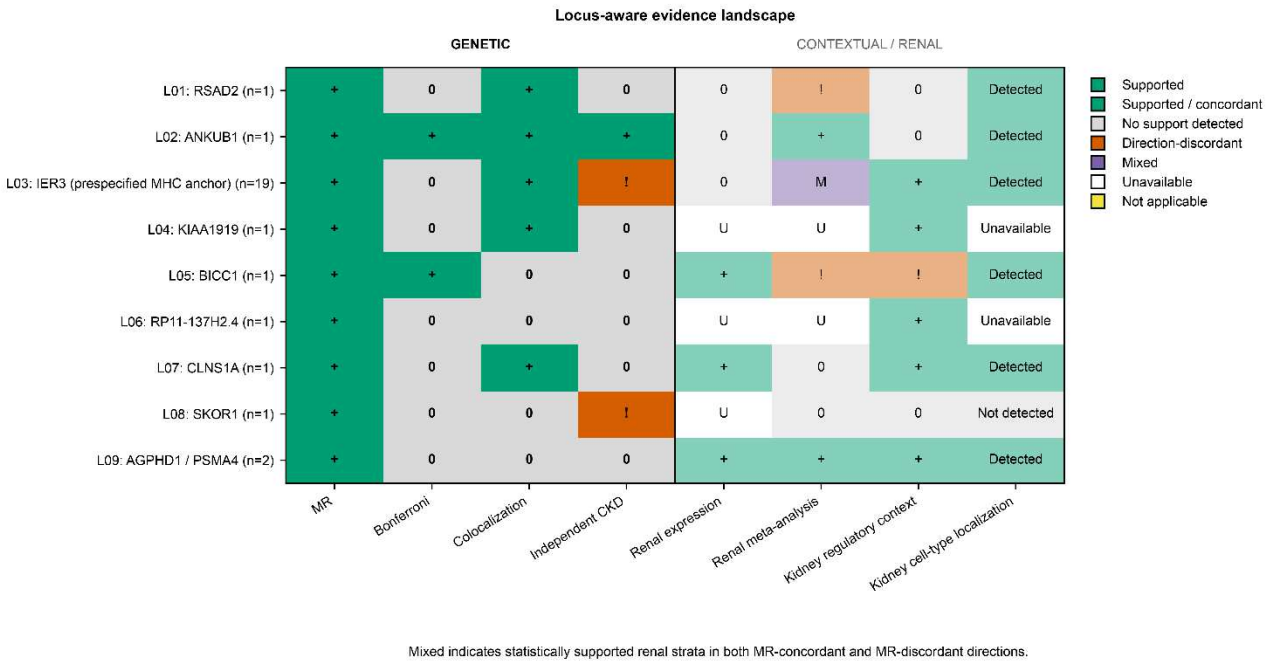


Figure 7. Locus-aware evidence landscape. Genetic and contextual evidence is summarized across nine operationally defined candidate loci without assigning a composite score or rank. L03 is represented by *IER3* as the prespecified MHC anchor. Mixed evidence indicates significant renal findings in both MR-concordant and MR-discordant directions.

Tables

Table 1. Data sources and study characteristics

Resource	Analysis role	Population/tissue	Phenotype / data type	Sample size	Ancestry	Genome build	Reference
eQTLGen	Discovery cis-eQTL exposures	Predominantly European whole blood	Blood gene expression cis-eQTLs	Up to 31,684	Predominantly European	GRCh37/hg19	6
FinnGen release 13	Primary CKD outcome	Finnish biobank participants	Registry-defined CKD	492,013 (14,402 cases, 477,611 controls)	Finnish	GRCh38	19
CKDGen	Independent CKD outcome	European-ancestry	CKD, primarily defined as eGFR <60 mL/min/1.73 m^2	480,698 (41,395 cases, 439,303 controls)	European	GRCh37	4
GEO renal datasets	Renal expression context	Kidney disease tissues compartments	60 prespecified disease compartment contrasts	Study-specific	Mixed, recorded by Series	Platform-specific	26, 27, study-specific references in Supplementary Table S17
GTEx V10	Kidney eQTL context	Kidney cortex	Bulk kidney cortex eQTLs	103	Not ancestry-stratified in the public resource used	GRCh38	32
Human Kidney eQTL Atlas	Kidney regulatory context	Whole kidney, tubule, and glomerular compartments	Kidney eQTL meta-analysis and compartment-specific eQTLs	686 meta-analysis, 356 tubules, 303 glomeruli	Predominantly European	Source-specific	33, 34
Muto kidney atlas	Single-nucleus localization	Normal adult kidney cortex	Single-nucleus RNA sequencing	19,985 nuclei from 5 donors	Reported by source	GRCh38	35

Note: CKD, chronic kidney disease; eGFR, estimated glomerular filtration rate; eQTL, expression quantitative trait locus; GEO, Gene Expression Omnibus; GTEx, Genotype-Tissue Expression.

Table 2. Genetic evidence across nine candidate loci

Locus	Representative trait(s)	FDR / Bonferroni traits	Primary MR estimate	Best colocalization	Independent CKD outcome
L01	*RSAD2*	1 / 0	OR 0.506 (0.360–0.712), 1 SNP, Wald ratio	Strong, PP.H4 = 0.953 (RSAD2)	Concordant, P ≥ 0.05
L02	*ANKUB1*	1 / 1	OR 2.488 (1.701–3.638), 1 SNP, Wald ratio	Strong, PP.H4 = 0.969 (ANKUB1)	Concordant, P < 0.05, FDR ≥ 0.05
L03	*IER3* (MHC anchor)	19 / 10	OR 1.133 (1.070–1.198), 5 SNPs, MRE-IVW	Strong, PP.H4 = 0.997 (IER3)	Exact-rsID discordant, harmonized sensitivity supports L03
L04	*KIAA1919*	1 / 0	OR 0.770 (0.676–0.877), 1 SNP, Wald ratio	Moderate, PP.H4 = 0.707 (KIAA1919)	Concordant, P ≥ 0.05
L05	*BICC1*	1 / 1	OR 1.487 (1.286–1.720), 2 SNPs, MRE-IVW	Insufficient, PP.H4 = 0.378 (BICC1)	Concordant, P ≥ 0.05
L06	*RP11-137H2.4*	1 / 0	OR 1.080 (1.043–1.118), 3 SNPs, MRE-IVW	Insufficient, PP.H4 = 0.178 (RP11-137H2.4)	Concordant, P ≥ 0.05
L07	*CLNS1A*	1 / 0	OR 1.128 (1.062–1.198), 2 SNPs, MRE-IVW	Moderate, PP.H4 = 0.761 (CLNS1A)	Concordant, P ≥ 0.05
L08	*SKOR1*	1 / 0	OR 0.717 (0.607–0.847), 1 SNP, Wald ratio	Insufficient, PP.H4 = 0.445 (SKOR1)	Discordant
L09	*AGPHD1 / PSMA4*	2 / 0	Trait-specific Wald ratios, 1 SNP each	Insufficient, PP.H4 = 0.171 (AGPHD1)	Concordant, P ≥ 0.05

Note: FDR / Bonferroni traits indicate the number of expression traits meeting each significance threshold within the locus. Primary MR estimates were obtained using the Wald ratio or MRE-IVW as appropriate. L03 contains 19 expression traits within one complex MHC locus, with *IER3* used as the prespecified anchor trait. Best colocalization denotes the highest single-signal PP.H4 within each locus and the corresponding expression trait. Independent CKD outcome refers to the primary exact-rsID CKDGen analysis unless otherwise specified. *AGPHD1* and *PSMA4* share the same MR instrument and therefore remain grouped within L09.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementaryfiguresall.pdf
- SupplementaryTables.xlsx
- SupplementaryInformation.pdf