Renal compartment-specific genetic variation analyses identify new pathways in chronic kidney disease

Chengxiang Qiu1,7, Shizheng Huang1,7, Jihwan Park1, YoSon Park2, Yi-An Ko1,2, Matthew J. Seasock1, Joshua S. Bryer1, Xiang-Xi Xu3, Wen-Chao Song4, Matthew Palmer5, Jon Hill6, Paolo Guarnieri6, Julie Hawkins6, Carine M. Boustany-Kari6, Steven S. Pullen6, Christopher D. Brown2 and Katalin Susztak1,2*

Chronic kidney disease (CKD), a condition in which the kidneys are unable to clear waste products, affects 700 million people globally. Genome-wide association studies (GWASs) have identified sequence variants for CKD; however, the biological basis of these GWAS results remains poorly understood. To address this issue, we created an expression quantitative trait loci (eQTL) atlas for the glomerular and tubular compartments of the human kidney. Through integrating the CKD GWAS with eQTL, single-cell RNA sequencing and regulatory region maps, we identified novel genes for CKD. Putative causal genes were enriched for proximal tubule expression and endolysosomal function, where DAB2, an adaptor protein in the TGF-β pathway, formed a central node. Functional experiments confirmed that reducing Dab2 expression in renal tubules protected mice from CKD. In conclusion, compartment-specific eQTL analysis is an important avenue for the identification of novel genes and cellular pathways involved in CKD development and thus potential new opportunities for its treatment.
has previously generated histone chromatin immunoprecipitation sequencing (ChIP-seq) data to annotate kidney-specific regulatory regions and performed eQTL analyses on 96 human whole-kidney tissue samples. As a result of this whole-tissue-based analysis, we identified putative disease-associated genes for 5 of the 83 previously identified CKD-associated loci identified through GWAS. Candidate gene and animal model studies indicated that UMOD and SHROOM3 are the likely causal genes for CKD GWAS variants. Causal genes and pathways for the remaining 76 loci remain unknown to date.

Here we argue that the cell-type heterogeneity of the eQTL data set and the sample size limitation are the key contributors to the low yield of identified causal genes for CKD using the GWAS-eQTL integration approach. Our recent single-cell transcriptome analysis highlighted important cell-type convergence, indicating that diseases that present with similar phenotypes originate from the same cell types. We propose that diseases are not organ-specific but rather are cell-type-specific; therefore, genetic variants are localized to cell-type-specific regulatory regions and influence gene expression changes only in disease-causing cell types.

As a first step toward identifying CKD-associated genes, we performed a compartment-based eQTL analysis of human kidney tissue samples using manual microdissection of the glomerulus and tubule, which are two key compartments of this organ. This microdissection markedly reduces cell heterogeneity, as each compartment is composed of around only five cell types. We aimed to define genotype-driven gene expression changes in the glomerular and tubular compartments of human kidneys, identifying genetic variants that influence the expression of genes. Here, we call genetic variants that influence gene expression eVariants and their target genes eGenes.

Subsequently, we integrated this information with genotype and phenotype association studies (that is, GWAS hits) to identify genes which are differentially expressed between the kidneys of individuals with different alleles of GWAS-identified variants (Supplementary Fig. 1a). We show that compartment-based eQTL data markedly improves identification of genes for which expression is regulated by GWAS-identified variants. Furthermore, we integrated the kidney eQTL data with epigenomic data and transcriptome analysis from single-cell RNA sequencing (RNA-seq) to study the regulatory mechanism of the cell-type-specific eQTL effects of disease-causing genetic variants. Finally, we performed cell-type-specific gene expression manipulations in animal models and specifically demonstrated that DAB2 is likely a causal gene for CKD development. Our study provides a new genetic framework for CKD development, as it defines key cell types and new mechanisms involved in the disease.

Results

Compartment-based eQTLs in the human kidney. We separated human kidney tissue compartments, in particular glomeruli and tubules, by manual microdissection; this was followed by RNA-seq of each compartment (Supplementary Fig. 1b). The expression of tubule-epithelial-specific markers, such as SLC12A1 and SLC34A1, was significantly greater in tubules (P < 2.2 x 10^-16 and P = 3.59 x 10^-11, respectively; two-sided Student's t-test), whereas glomerulus-epithelial-specific genes were almost exclusively expressed in glomeruli (Supplementary Fig. 1c). Well-known genes associated with nephrotic syndrome showed preferential expression in the glomerular compartment, and proximal tubulopathy genes were expressed in tubules (Supplementary Fig. 1d). We validated that the fraction of each cell type was similar in the kidney samples included in the analysis using in silico cell deconvolution analysis that estimates cell-type proportions on the basis of latent variable modeling (Supplementary Fig. 1e).

Furthermore, tissue samples underwent careful clinical and histological evaluation, and we included samples only without obvious structural and functional changes in the analysis to minimize nongenetically-driven fluctuations in gene expression (Supplementary Table 1).

Using these stringent criteria, we selected 151 kidneys for the analysis, including 121 tubule samples and 119 glomerulus samples used to identify compartment-based cis-eQTLs (further referred to as eQTLs) (Supplementary Fig. 1f). We performed eQTL analysis separately for the two compartments by using a linear model and a ±1-Mb window around the transcription start site (TSS). In the tubular compartment, we identified 4,081 genes with eQTLs at a 5% false discovery rate (FDR) (hereafter referred to as eGenes) and 389,454 significant SNP–gene pairs; in the glomerular compartment, we identified 4,913 eGenes and 467,994 significant SNP–gene pairs. We made this dataset available via our searchable website (http://susztaklab.com/eqtl/). We also performed meta-analysis on tubules and glomeruli eQTLs and defined tubule-compartment-specific eGenes (n = 417), glomerulus-compartment-specific eGenes (n = 674) and compartment-shared eGenes (n = 3,493) (Fig. 1a).

The number of eGenes identified is in line with expectations (Supplementary Fig. 2a) from previous publications examining 44 human tissues in the GTEx Consortium. Furthermore, many of the significant SNP–gene pairs identified by our tubule eQTL analysis can be replicated in other tissues (thoracic (r = 0.79) and transformed fibroblasts cells (r = 0.77); r refers to the proportion of true positives). As expected, the greatest overlap for tubule was observed in glomeruli (r = 0.91) (Supplementary Fig. 2b).

By performing meta-analysis of kidney compartments and 44 GTEx tissue samples, we could identify 589 tubule-specific eGenes (m > 0.9 in less than five tissues, including tubule; m refers to the posterior probability that the effect is shared in each tissue) and 594 glomerulus-specific eGenes. We found 7,050 shared eGenes in the tubules (m > 0.9 in more than 40 tissues, including tubule) and 7,090 shared eGenes in the glomerulus. These results are in line with previous observations indicating that most eGenes were shared across tissues (Fig. 1b).

Our compartment-based eQTL study identified many new eQTLs that were previously not detected by bulk-tissue, whole-kidney eQTL analysis (Supplementary Table 2). For example, LRRC3 is associated with the rs2838917 genotype only in the tubular compartment, but not in whole kidney or the glomerulus (Supplementary Fig. 2c,d). In contrast, ANXA2 is associated with the rs3068 genotype only in the glomerular compartment (Supplementary Fig. 2e,f). Of note, we found that one gene, NUCB1, has an opposite eQTL effect direction between tubule and glomerulus (Supplementary Fig. 2g).

These examples highlight the utility of compartment-based eQTL analysis to identify a large number of new eQTLs.

Renal compartment eQTLs are enriched for kidney-associated GWAS trait variants. By assessing overlap between the entire published GWAS catalog (downloaded 4 January 2017) and our kidney compartment eQTLs, we found that GWAS hits in genes encoding proteins that are related to regulating the levels various blood metabolites were significantly enriched for tubule eQTLs. Of the metabolite GWAS hits, 26% colocalized with tubule eQTLs, and its associated genes were significantly enriched for tubule eGenes (P = 3.04 x 10^-4, Fisher's exact test). Moreover, tubule eGenes were also significantly enriched for genes associated with kidney disease traits, such as IgA nephropathy (P = 2.12 x 10^-4). A similar pattern of kidney and metabolite trait enrichment was seen in the glomerulus eQTL dataset (Fig. 1c).

Next, we focused on our newly identified tubule-specific and glomerulus-specific eGenes by assessing overlap with the GWAS catalog. Tubule-specific eGenes, compared to nontubule-specific eGenes, were significantly enriched for CKD GWAS hits (P = 4.79 x 10^-3, Fisher's exact test) and glomerular filtration rate
GWAS hits ($P = 7.96 \times 10^{-3}$). In contrast, these tubule-shared eGenes were enriched in genes for inflammatory skin disease ($P = 5.94 \times 10^{-3}$) and Parkinson’s disease ($P = 2.95 \times 10^{-2}$) compared with nontubule-shared eGenes. Glomerulus-specific eGenes were not only enriched for CKD GWAS variants but also for blood pressure–associated genetic variants (Fig. 1d). In summary, these results...
Integrate CKD GWAS variants and compartment-based eQTLs to identify causal genes for kidney disease. Our goal was to annotate CKD-associated GWAS variants to identify genes and pathways responsible for CKD development. Through a literature survey, we identified a total of 83 replicated, CKD-associated GWAS loci11–12 (Supplementary Table 3). A direct overlap of loci that reached genome-wide significance in GWAS for CKD and in our eQTL data identified differences in the expression level of 27 genes associated with CKD GWAS loci. To test whether two traits share a causal variant at a given locus, we conducted colocalization analysis using colocal9. And, in a complementary analysis to assess whether a GWAS SNP tags the same functional variant as the eQTL, we calculated regulatory trait concordance (RTC) scores24 for each candidate SNP–gene pair. Most of these genetic regions showed a positive colocalization between the eQTL effect and CKD GWAS (posterior probability of colocalization between eQTL and GWAS effect in the given region (PP_H4) > 0.8), suggesting that these genes are strong causal candidates for kidney disease (Table 1). The functional role of MANBA, PGAP3 and CASP9 for kidney disease development has been supported by prior animal model studies, confirming the biological validity of our computational analysis34,45.

In summary, when compared to previous eQTL studies of whole-kidney tissue samples, the compartment-based eQTL analysis could identify a larger number of causal genes for CKD \( (n = 5 \text{ for bulk tissue and } n = 27 \text{ for specific compartment}) \). Furthermore, despite the significant eQTL sharing (\( \pi_e = 0.91 \)) between tubule and glomerulus, more causal genes identified in GWAS came from the smaller, newly discovered, compartment-specific eQTL subset.

Compartment-specific eQTLs are enriched in distal regulatory regions and show greater cell-type specificity. We hypothesized that integration of GWAS with compartment eQTLs can identify potential causal cell types for a given set of traits (Fig. 2a). To further understand whether the 27 genes with GWAS overlapping compartment-based eQTLs show cell-type specificity, we have examined the expression of putative GWAS target genes in kidney cells. We have previously performed single-cell RNA-seq on mouse kidney tissue samples, and our analysis has identified 16 distinct cell types in the kidney46. Using this unique expression dataset (only 13 well-annotated cell types), we next examined the cell types for which putative GWAS target genes are expressed. We were able to map the expression of 23 out of the 27 human target genes in our mouse kidney single-cell atlas. We found that most CKD-associated genes have a unique cell-type-enriched expression profile and that these transcripts are not broadly expressed in all kidney cells. Particularly, we found that renal proximal tubules show the greatest enrichment for GWAS-eQTL target genes, where 39% of genes (9 out of 23) are expressed in proximal tubule epithelial cells, which is markedly higher than what is expected if genes were randomly distributed in kidney cells (23 genes, 13 cell types; \( \sim \)1 gene per cell type) (Fig. 2b).

In our analysis, as observed by GTEx and most prior studies, the identified eQTLs were mostly enriched on promoters (Supplementary Fig. 3a). Upon examining for binding sites of transcription factors, we found enrichment for genes encoding SAP30, KDM5A, HDAC1, CREB1, ELK4 and ESRR, key transcription factors that are proposed to be involved in maintaining the high metabolic rate of proximal tubule cells (Supplementary Fig. 3b).

We reasoned that if compartment-based eQTLs are more informative for GWAS annotation, then compartment-specific eGenes must arise from variants that are localized to cell-type-specific regulatory regions, such as distant regulatory regions (e.g., enhancers), rather than proximal regulatory regions (i.e., promoters) that are mostly shared between cell types. When we compared shared and compartment-specific eGenes, we found that their eQTL variants were significantly further away from the TSS for compartment-specific eGenes compared to shared eGenes, indicating that compartment-specific eQTLs might be linked to more distal (i.e., enhancer) regulatory variants (Fig. 2c,d).

To further investigate this finding, we compared the genomic feature annotations for compartment-specific eGenes and compartment-shared eGenes using ChIP-seq data derived from adult human kidney. Compartment-shared eGenes were significantly more enriched in promoters than compartment-specific eGenes (\( (P < 2.2 \times 10^{-10}) \), Wilcoxon rank-sum test) (Fig. 2e). Moreover, we examined whether cell-type-specific enrichment would explain the identification of compartment-specific eGenes. We found that glomerulus-compartment-specific eGenes were significantly enriched for expression in podocytes (\( (P = 2.91 \times 10^{-7}) \), Fisher’s exact test), and tubule-compartment-specific eGenes were enriched for expression in tubule epithelial cells (\( (P = 6.14 \times 10^{-4}) \), Fisher’s exact test), indicating that disease-associated target genes are more likely to be cell-type-specific (Fig. 2f).

To gain further insight into pathogenesis of kidney disease, we performed functional annotation of the 27 genes identified by colocalization between kidney-compartment-based eQTLs and CKD GWAS. Using the pathway analysis described in genome-scale integrated analysis of gene network (GIANT)46, we found that these genes identified as tubule-compartment specific were significantly enriched for a single functional group: endolysosomal function (FDR = 1.5 \times 10^{-5}) (Fig. 2g). Combining results from both kidney compartments showed enrichment not only for endolysosomal genes, but also related pathways such as autophagy (FDR = 3.59 \times 10^{-5}) and the mitochondrial degradation pathway (FDR = 4.58 \times 10^{-5}) (Supplementary Fig. 3c). These results indicate that the expression of the target genes for kidney function are enriched for a limited number of functional groups. Specifically, we noticed that DAB2 had the greatest number of connections in the functional network and represented a central hub.

Identification of DAB2 as a kidney disease gene. To leverage our unique eQTL datasets to identify kidney-specific causal genes for kidney disease, we focused on the Complement C9 (C9)-Disabled homolog 2 (DAB2) locus for follow-up functional validation studies. The genotype of rs11959928 (at the DAB2-C9 locus) has been significantly associated with CKD in multiple CKD GWAS48,10,12. The DAB2 and C9 genes are located near rs11959928; however, it was unclear which gene is regulated by the disease-causing variant. In eQTL analysis at the whole-kidney level, the rs11959928 genotype was not significantly associated with either DAB2 or C9 level34. Compartment-based eQTL analysis identified an association between the rs11959928 genotype and the DAB2 level, but not the C9 level. Furthermore, this eQTL was only identified in tubule eQTLs, not in glomerulus eQTLs (Fig. 3a, b). The tubule-specific eQTL effect of rs11959928 on DAB2 can be replicated by a recent kidney cis-eQTLs study17 (Supplementary Fig. 4a). Although DAB2 is broadly expressed in multiple tissues and cell types, rs11959928 was not identified as a significant eQTL in 44 GTEx tissue samples35 (Fig. 4a). Meta-analysis35 of our kidney compartment and 44 GTEx tissues indicated that rs11959928 showed a tubule-specific eQTL effect only on DAB2 (\( m = 0.855 \)) and not on C9 (\( m = 0.407 \)) (Fig. 4b).

Using single-cell RNA-seq of mouse kidney tissue samples, we found that in the kidney, Dab2 expression is restricted to the proximal tubules and to macrophages (Fig. 2a). Although one prior study reported a significant eQTL effect of rs11959928 for Dab2 in peripheral blood mononuclear cells (PBMCs)49, it was in the opposite direction (\( z = -32.69, P = 5.123 \times 10^{-125} \)) (Supplementary Fig. 4b), and this eQTL was not replicated in the
β whole-blood samples of the GTEx dataset (\( \beta = 0.032, P = 0.394 \)) (Supplementary Fig. 4c).

Given differences in eQTL effects, we examined the functional annotation of this candidate GWAS and eQTL locus in the human kidney and other tissues and cells by comparing the enrichment for histone H3K27ac, a marker that distinguishes an active enhancer from poised ones. We noticed a binding peak of H3K27ac that is specifically found in kidney, but not in macrophages or any other cells analyzed by ENCODE (Fig. 4c). The kidney-specific enhancer peak coincided with the CKD GWAS region, suggesting that this region might have a functional role in kidney tubule cells and likely explaining the existence of a kidney eQTL effect for a gene with a relatively broad expression profile.

Furthermore, we also noted that GTEx (GTex v7; https://www.gtexportal.org/) reported a kidney-specific DAB2 isoform when compared to other human tissues. This isoform contains an exon skipping (chr5: 39,382,720-39,383,373, hg19) and was highly expressed in the kidney. Using the LeafCutter\(^\text{10}\) method for isoform quantification, we confirmed the presence of this isoform in our human kidney tubule dataset (Supplementary Fig. 4d). Kidney-specific isoform expression can also contribute to the specific eQTL effect on DAB2 in tubule cells. In summary, we have identified the CKD-associated variant rs11959928 with tubule-specific effects on DAB2 expression levels. Our results indicate a kidney-specific isoform expression and a kidney-specific enhancer peak at this locus that is not present in other cell types and suggest genetic variation at this enhancer could lead to kidney diseases through higher DAB2 expression in renal tubules.

**Dab2 alters kidney disease development in mice.** To determine whether C9 or Dab2 plays a role in kidney disease development,
we performed in vivo validation using animal models with reduced gene dosage. Mice with global knockout of C9 and mice heterozygous for the gene were phenotypically normal, and we did not observe structural or functional changes in the kidney (Fig. 5a–c). We reasoned that C9 might alter injury response and still be responsible for kidney disease development. Therefore, we subjected wild-type (WT) and C9+/− mice to folic-acid-induced kidney injury (FAN; folic acid nephropathy) (Supplementary Fig. 5a). Transcript levels of kidney fibrosis markers including Fibronectin 1 ( Fn1), Collagen 1a1 (Col1a1) and Collagen 3a1 (Col3a1) were significantly higher in WT mice injected with folic acid. However, we found no differences between WT and C9+/− mice in this CKD model (Fig. 5a). Similar results were obtained when analyzing histological lesion on PAS-stained and Sirius Red–stained kidney sections (Fig. 5b,c).

Next, we studied the role of Dab2, as the CKD GWAS variant also influenced the expression of Dab2 in the kidney. We generated mice with tubule-epithelial-specific dose reduction of Dab2 using the well-characterized mice with tubule-specific expression of Cre recombinase under control of the Cdhl6 (Ksp) promoter (Ksp-Cre mice) and mice with Dab2 flanked by loxP (Dab2lox mice) (Supplementary Fig. 5b). The genetic and eQTL study integration suggested that the risk allele was associated with higher DAB2 levels in the kidney (Table 1). We did not observe any renal functional or histological abnormalities in Ksp-Cre and Dab2lox mice at baseline. Next, we subjected these animals to folic-acid-induced acute and chronic kidney injury. A significantly lower expression of fibrosis markers (Fn1, Col1a1 and Col3a1) at the mRNA level in Ksp-Cre and Dab2lox mice was observed after folic acid injection (Fig. 5d). We also found that the degree of disease was significantly

---

**Fig. 2 | Compartment-specific eQTLs show greater cell-type specificity and enrichment for distal regulatory elements. a.** Schematic of the integration of kidney eQTLs, GWAS, single-cell expression and regulatory region. b. Heatmap of cell-type-specific expression of identified CKD target genes. The blue and yellow colors correspond to the level of expression (z-score). c. Density plots of the best eVariants in tubule (top) and glomerulus (bottom) and the relationship to transcription start site (TSS). d. Distance of top eVariants (sorted into groups) from TSS. e. Odds ratios of the top eVariants (sorted into groups) on kidney promoter regions. The groups were compared to randomly selected variants matched by minor allele frequency (MAF) and distance to TSS (n = 5,000 randomly selected times). For d and e, center lines show the medians; box limits indicate the 25th and 75th percentiles; whiskers extend to the 5th and 95th percentiles; and outliers are represented by dots. f, Odds ratio (y axis) of eGenes from each group enriched by kidney-specific cell type expression. P was calculated by two-sided Fisher’s exact test. RTEC: PT, LOH and DCT; Myeloid: Fibro, Macro and Neutro; Lymphoid: B cell, T cell and NK. Tubule, tubule-specific-eGenes: n = 674 (d–f). g. GIANT visualization of tubule-specific candidate genes. The color of each link shows the relationship confidence, from green (0.01) to red (1).
lower in tubule-specific Dab2-knockout mice compared with folic-acid-treated WT mice (Fig. 5e, f), indicating that the directional-ity is consistent with the GWAS effect direction. In addition to the FAN model, we also found a lower level of interstitial fibrosis in mice with tubule-specific loss of Dab2 subjected to unilateral ureter obstruction (UUO) (Supplementary Fig. 5c–e).
To understand the role of Dab2 in kidney fibrosis development, we generated primary renal tubule epithelial cells (TECs) with lower Dab2 levels by infecting cells from Dab2<sup>−/−</sup> mice with Cre–eGFP adenovirus (Supplementary Fig. 6a). Injured TECs release tumor growth factor (TGF)-β (encoded by Tgfb1), which is one of the best-known profibrotic cytokines<sup>51,52</sup>. TGF-β plays an important role in TGF-β-induced profibrotic gene expression.

In summary, these results indicate that Dab2 in tubule cells is a likely causal gene for CKD development, as Dab2 plays an important role in TGF-β-induced profibrotic gene expression.

**Discussion**

A key limitation to disease understanding is that most human studies cannot go beyond descriptive and correlative analysis, and animal models often show limited relevance to human disease development as genotypes are established before disease development. GWASs provide valuable insight into disease mechanisms. Although GWAS has identified noncoding genetic variants associated with CKD, the underlying genes, cell types and mechanism still remain elusive. Here we generated new datasets and conducted a genome- and transcriptome-wide scan for genetic variants associated with kidney compartments: tubules and glomeruli. Although some cell-type-specific eQTL analyses have been performed in blood samples, we believe that this is the first study to directly characterize cell-type specificity.
of different compartments on solid organ samples. By integrating this data with prior GWASs, epigenome analysis and single-cell transcriptome analysis, we have identified putative causal genes for 24 of the 83 GWAS loci.

To date, GTEx has conducted eQTL analysis on 53 human tissues, 44 of which had sufficient sample size for further analyses (GTEx v6p). Kidney, however, was not included owing to the limited sample size. By comparing our results with GTEx, we showed that the majority of eQTLs are shared between multiple GTEx tissue types as well as two kidney compartments. Strikingly though, we observed that previously identified CKD GWAS loci were more likely to be associated with eQTLs that are compartment-specific,
that reside further away from promoters and that possibly act by modulating compartment-specific active enhancers compared to those identified by bulk-sequencing data analysis. Our study high- lights that cell-type-specific eQTLs and epigenome datasets are crucial components of post-GWAS prioritization and mechanism studies. Deeper insight into these cell-type-specific mechanisms may reveal that many diseases and conditions are perturbations of a specific cell type rather than specific organs.

Next, we investigated whether specific regulatory elements were enriched for compartment-specific or compartment-shared eQTLs. Functional annotation analysis indicated that compartment- and tissue-shared eQTLs were more likely to be located in promoter regions. On the other hand, by incorporating single-cell mouse kidney RNA-seq data, we demonstrate that disease-specific target gene regulation is likely to be cell-type specific, and the most likely causal GWAS variants are localized to cell-type-specific enhancers. We then focused on the DAB2 locus to illustrate such cell-type-specific eQTL and enhancer effects.

We used the comprehensive and new compartment-specific RNA-seq dataset and computational analysis as well as rigorous in vitro and in vivo assays to identify DAB2 as a new causal gene for CKD pathophysiology. The rs11959928 variant has been associated with CKD in multiple GWASs10–12. Yet, this variant was not associated with gene expression changes in GTEx25 or in bulk kidney eQTL datasets26. Furthermore, even though DAB2 is expressed across multiple cell types, we observed that this variant only displays tubule-specific effects on DAB2 expression. Single-cell RNA-seq data further refined the target tissues to two cell types (tubules and macrophages) with high DAB2 expression in the kidney. Of note, a prior report49 showed a macrophage-specific eQTL effect of DAB2 of this variant in the opposite direction as compared to the tubule-specific eQTL direction. However, this effect and direction were not replicated in the larger GTEx dataset49, and follow-up in vivo results failed to show consistent effects of macrophase-specific Dab2 reduction on kidney fibrosis development (unpublished data).

On the basis of the replicated eQTL effect, combination of strong colocalization between the tubule-specific eQTLs and CKD GWAS and the protection from kidney injury observed in Dab2-knockout mice, we conclude that the CKD risk allele at the DAB2 locus from GWAS is associated with higher expressed levels of DAB2 in tubule cells, leading to tubular fibrosis.

DAB2 is a central adaptor protein in several receptor-mediated pathways41. In vitro studies have shown that DAB2 is required for TGF-β-mediated signaling in epithelial cells38. TGF-β is one of the strongest inducers of kidney fibrosis and CKD31. The CKD-risk-increasing allele of the underlying causal variant likely modifies this enhancer function specifically in tubule cells, raising cytokine levels (including TGF-β), thus altering endocytosis and inducing downstream development of fibrosis through DAB2. We demonstrate the role of tubule-specific DAB2 in kidney disease development using cell-type-specific-knockout mice.

Through gene–gene network analysis, we also show that these 24 putative causal genes are enriched in a specific kidney cell type and highlight the endolysosomal pathway as a likely pathophysi- ological mechanism in the proximal tubules for kidney disease development. Proximal tubule epithelial cells are the most common cell type in the kidney, and dedifferentiation of these highly specialized cells has been linked to a reduction in kidney function and the development of kidney disease46,57. The proximal tubules play a key role in initial processing of the primary blood filtrate, as the renal glomerulus processes more than 144 L of primary filtrate a day. Further, the proximal tubules have one of the highest levels of endocytic activity in the body to reclaim critical nutrients from the primary filtrate46. Alterations in the endocytic, lysosomal and autophagy pathways of these cells have been previously observed in animal models and individuals with kidney disease66,67. Now, our genetic studies can help to refocus attention to this specific cell type and mechanisms in CKD development.

In summary, this is one of the first and most extensive post-GWAS annotations of CKD risk loci. We combined compartment-specific transcriptome, genome, epigenome and single-cell sequencing data to identify new genes, cell types and mechanisms contributing to the CKD pathophysiology. The integration of these datasets provides critically needed information for the field. Our data delivered new biological insight by the identification of DAB2 as kidney-disease-causing gene and the additional 26 putative candidate genes. We also show that renal tubule cells and the endolysosomal pathway play an important role in CKD, providing a new mechanism for disease development.

Online content
Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41591-018-0194-4.

Received: 15 November 2017; Accepted: 8 August 2018; Published online: 1 October 2018

References
1. Eckardt, K. U. et al. Evolving importance of kidney disease: from subspecialty to global health burden. Lancet 382, 158–169 (2013).
2. Webster, A. C., Nagler, E. V., Morton, R. L. & Masson, P. Chronic kidney disease. Lancet 389, 1238–1252 (2017).
3. Rhee, C. M. & Kovesdy, C. P. Epidemiology: spotlight on CKD deaths—increasing mortality worldwide. Nat. Rev. Nephrol. 11, 199–200 (2015).
4. Sud, M., Tangri, N., Pintilie, M., Levey, A. S. & Naimark, D. M. Progression to Stage 4 chronic kidney disease and death, acute kidney injury and hospitalization risk: a retrospective cohort study. Nephrol. Dial. Transplant. 31, 1122–1130 (2016).
5. Garrett, M. R., Pezzolesi, M. G. & Korstand, R. Integrating human and rodent data to identify the genetic factors involved in chronic kidney disease. J. Am. Soc. Nephrol. 21, 398–405 (2010).
6. Braun, D. A. et al. Mutations in nuclear pore genes NUP93, NUP205 and XPO5 cause steroid-resistant nephrotic syndrome. Nat. Genet. 48, 457–465 (2016).
7. Tory, K. et al. Mutation-dependent recessive inheritance of NPHS2-associated steroid-resistant nephrotic syndrome. Nat. Genet. 46, 299–304 (2014).
8. Fox, C. S. et al. Genomewide linkage analysis to serum creatinine, GFR, and creatinine clearance in a community-based population: the Framingham Heart Study. J. Am. Soc. Nephrol. 15, 2457–2461 (2004).
9. Pattaro, C. et al. Genetic associations at 53 loci highlight cell types and biological pathways relevant for kidney function. Nat. Commun. 7, 10023 (2016).
10. Köttgen, A. et al. New loci associated with kidney function and chronic kidney disease. Nat. Genet. 42, 376–384 (2010).
11. Okada, Y. et al. Meta-analysis identifies multiple loci associated with kidney function-related traits in east Asian populations. Nat. Genet. 44, 904–909 (2012).
12. Köttgen, A. et al. Multiple loci associated with indices of renal function and chronic kidney disease. Nat. Genet. 41, 712–717 (2009).
13. Witte, J. S., Visscher, P. M. & Wray, N. R. The contribution of genetic variants to disease depends on the ruler. Nat. Rev. Genet. 15, 765–776 (2014).
14. Chatterjee, N., Shi, J. & Garcia-Closas, M. Developing and evaluating polygenic risk prediction models for stratified disease prevention. Nat. Rev. Genet. 17, 392–406 (2016).
15. Pasaniuc, B. & Price, A. L. Dissecting the genetics of complex traits using summary association statistics. Nat. Rev. Genet. 18, 117–127 (2017).
16. Korstanje, R. & Paigen, B. From QTL to gene: the harvest begins. Nat. Genet. 31, 235–236 (2002).
17. Paschos, E. E. et al. Large, diverse population cohorts of hipscas and derived hepatocyte-like cells reveal functional genetic variation at blood lipid-associated loci. Cell Stem Cell 20, 558–570 (2017).
18. Gupta, R. M. et al. A genetic variant associated with five vascular diseases is a distal regulator of endothelin-1 gene expression. Cell 170, 522–533 (2017).
19. Musumur, K. et al. From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. Nature 466, 714–719 (2010).
20. Capellini, T. D. et al. Ancient selection for derived alleles at a GDF5 enhancer influencing human growth and osteoarthritis risk. Nat. Genet. 49, 1202–1210 (2017).
21. Soldner, F. et al. Parkinson-associated risk variant in distal enhancer of α-synuclein modulates target gene expression. Nature 533, 95–99 (2016).
22. Claussnitzer, M. et al. FTO obesity variant circuitry and adipocyte browning in humans. N. Engl. J. Med. 373, 895–907 (2015).
23. The ENCODE Project Consortium. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature 447, 799–816 (2007).
24. Roadmap Epigenomic Consortium et al. Integrative analysis of 111 reference human epigenomes. Nature 518, 317–330 (2015).
25. GTEx Consortium et al. Genetic effects on gene expression across human tissues. Nature 550, 204–213 (2017).
26. GTEx Consortium. The Genotype–Tissue Expression (GTEx) project. Nat. Genet. 45, 580–585 (2013).
27. GTEx Consortium. Human genomes. The Genotype–Tissue Expression (GTEx) pilot analysis: multiscale gene regulation in humans. Science 348, 648–660 (2015).
28. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature 489, 57–74 (2012).
29. Harrow, J. et al. GENCODE: the reference human genome annotation for the ENCODE Project. Genome Res. 22, 1760–1774 (2012).
30. Ko, Y. A. et al. Genetic-variation-driven gene-expression changes highlight genes with important functions for kidney disease. Am. J. Hum. Genet. 100, 940–953 (2017).
31. Yeo, N. C. et al. Shroom3 contributes to the maintenance of the glomerular filtration barrier integrity. Genome Res. 25, 57–65 (2015).
32. Trudu, M. et al. Common noncoding eQTLs as putative drivers of kidney disease. Nat. Genet. 45, 1120–1125 (2017).
33. Lee, J. W. et al. Polarization of the effects of autoimmune and neurodegenerative risk alleles in leukocytes. Science 344, 519–523 (2014).
34. Raj, T. et al. Polarization of the effects of autoimmune and neurodegenerative risk alleles in leukocytes. Science 344, 519–523 (2014).
35. Park, J. et al. Single-cell transcriptomics of the mouse kidney reveals potential cellular targets of kidney disease. Nature 541, 266–2677 (2015).
36. Newman, A. M. et al. Robust enumeration of cell subsets from tissue expression profiles. Nat. Methods 12, 453–457 (2015).
37. Ishigaki, K. et al. Polygenic burdens on cell-specific pathways underlie the risk of rheumatoid arthritis. Nat. Genet. 49, 1120–1125 (2017).
38. Lee, J. W., Chou, C. L. & Knepper, M. A. Deep sequencing in microdissected renal tubules identifies nephron segment–specific transcriptomes. J. Am. Soc. Nephrol. 26, 1486–1494 (2005).
39. Edeling, M., Ragi, G., Huang, S., Pavenstädt, H. & Susztak, K. Developmental signalling pathways in renal fibrosis: the roles of Notch, Wnt and Hedgehog. Nat. Rev. Nephrol. 12, 426–439 (2016).
40. Reidy, K., Kang, H. M., Hostetter, T. & Susztak, K. Molecular mechanisms of diabetic kidney disease. J. Clin. Invest. 124, 2333–2340 (2014).
41. Zhuo, J. L. & Li, X. C. Proximal nephron. Comp. Physiol. 3, 1079–1123 (2013).
42. Kimura, T. et al. Autophagy protects kidney proximal tubule epithelial cells from mitochondrial metabolic stress. Autophagy 9, 1876–1886 (2013).
43. Dickson, L. E., Wagner, M. C., Sandoval, R. M. & Moltóiris, B. A. The proximal tubule and albuminuria: really! J. Am. Soc. Nephrol. 25, 443–453 (2014).

Acknowledgements
The authors thank the Molecular Pathology and Imaging Core (P30-DK050306) and Diabetes Research Center (P30-DK19523) at the University of Pennsylvania for their services. This work in the Susztak lab has been supported by the National Institute of Health NIH R01 DK087635, DK076077 and DP3108220, Boehringer Ingelheim, the Eli Lilly Co. and the Juvenile Diabetes Research Foundation.

Author contributions
K.S., C.Q., S.H. and C.D.B. conceived, planned and oversaw the study. C.Q., S.H. and J.S.B. performed and analyzed experiments. Y.P., M.J.S. and J.S.B. assisted with data analysis. X.-X.X. provided Dab2 knockout mice. M.P. performed biodistribution analyses. X.-X.X. provided Dab2 knock floxed mice. W.-C.S. provided C9 knockout mice. M.P. performed pathology examination. M.J.S., J. Hill, P.G, J. Hawkins, C.M.B.-K. and S.S.P. are full-time employees of Boehringer Ingelheim Pharmaceuticals, Inc. and the Eli Lilly Co. The proximal tubule and albuminuria: really! J. Am. Soc. Nephrol. 25, 443–453 (2014).

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41591-018-0194-4.
Reprints and permissions information is available at www.nature.com/reprints.
Correspondence and requests for materials should be addressed to K.S.
Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
Methods

Sample procurement. Human kidney tissue collection was approved by the University of Pennsylvania Institutional Review Board. Kidney samples were obtained from surgical nephrectomies. Nephrectomies were de-identified, and the corresponding clinical information was collected through an honest broker; therefore, no consent was obtained from the subjects.

Collected tissue was immersed in RNAlater (Ambion no. AM7020) solution at 4°C for several hours before being stored at −80°C in RNAlater. Tissue was thawed on ice, placed in RNAlater, and manually microdissected for glomerular and tubular compartments. In general, 60–150 glomeruli that readily released from the capsule were collected and placed into RNeasy RNA Tissue Lysis Buffer Solution (per Qiagen RNeasy kit manufacturer instructions (Qiagen no. 74106)). We refer to the remaining compartment as tubule throughout the article.

Part of the tissue core was formalin fixed and paraffin embedded. These tissues included 16 samples because of their SNP data, which were ambiguous regarding sex. An additional two were excluded because of elevated heterozygosity (Supplementary Fig. 7a). To further identify potential sample contamination, identity-by-descent (IBD) was computed between all pairwise sample combinations, and eight samples were excluded due to their high IBD (≥0.2). To quantify population structure, principal component analysis (PCA) implemented in EIGENSTRAT was conducted on these final 408 samples, with additional genotype data from 395 HapMap2 samples (112 CEU, 84 CHB, 86 JPT, 113 YRI) and 2,504 samples from the 1,000 Genomes Project Phase 1 (release v3, 661 AFR, 347 AMR, 504 EAS, 503 EUR, 489 SAS), respectively.22,23. Plots of the two main principal components (PCs) for each dataset are shown in Supplementary Fig. 7b,c. Subsequently, genotypes of the 408 samples were phased with SHAPEIT2 (ref. 19) and imputed with IMPUTE2 (ref. 20), using multithreaded panel reference66 from 1000 Genomes Phase 1 v5.

Separated by two kidney compartments, RNA was isolated from a total of 329 tubule samples and 311 glomerulus samples from up to 359 individuals. RNA quality was assessed with the Agilent Bioanalyzer 2100. Thirty samples with RIN scores above 7 and a minimum total RNA of 100 ng were used for cDNA production. Strand-specific, polyA + selected RNA-Seq libraries were generated using the Illumina TruSeq protocol. Libraries were sequenced to a median depth of 35 million 100-bp single-end reads. After first assessing sequence quality using FastQC and then trimming the adaptor and lower-quality bases using Trimgalore, RNA-Seq reads were aligned to the human genome (hg19/GRCh37) using STAR (v2.4.1d) based on GENCODE v19 annotations.24-26. Gene-level expression was estimated on uniquely mapped reads as reads per kilobase of transcript per million mapped reads (RPKM) using HtSeq (0.9.1) and DESeq2 (ref. 27). Samples with less than 15 million mapped reads were excluded.

For eQTL analysis, only samples with European ancestry and absence of significant kidney structural changes (tubular fibrosis < 10%; glomerular sclerosis < 10%) were used. Variants were excluded from analysis if they: (i) had a call rate < 90%; (ii) had a MAF < 5%; (iii) deviated from Hardy–Weinberg Equilibrium (P < 10−8); and (iv) had an imputation info score < 0.4. Finally, 121 samples were used for tubule eQTL analysis, and 119 samples were used for glomerulus eQTL analysis. In addition, genotype-based PCA analysis was conducted again using the final set of tubule and glomerulus samples, and then the first six PCs were used as covariates (Supplementary Fig. 7d,e). We also generated a variant call format (VCF) file with dosages for alternative allele counts used as input for the Matrix eQTL and FastQTL software packages.28-30

Cis-eQTL mapping. We conducted cis-eQTL (further just referred to as eQTL) mapping with 121 samples of tubules and 119 samples of glomeruli. Only genes having at least 0.1 RPKM in two or more samples were considered significantly expressed and used for eQTL mapping. Within each compartment, quantile-normalization was performed on RPKMs, and then expression measurements for each gene were rank-based inverse normal transformed. To remove the effects of unobserved confounding variables, PEER was performed for 15 factors on gene expression with age, gender, collected site, sequencing batch, RNA integrity number (RIN) and RIN percentage (for tubule samples) or sclerosis percentage (for glomerulus samples) as covariates. The Pearson correlation r between known clinical variables and 15 PEER factors calculated by 121 tubule RNA-seq samples are shown in Supplementary Fig. 7f. The Pearson correlation r between known clinical variables and 15 PEER factors calculated by 119 glomerulus RNA-Seq samples are shown in Supplementary Fig. 7g. Finally, once the PEER factors were regressed out, the residuals of the analysis with rank-based inverse normal transformation were used as expression measurements in following eQTL mapping, and 32,596 genes were used for tubule analysis as well as 31,635 for glomerul eurine analysis.

The cis window was defined as 1 Mb up- and downstream of the transcriptional start site (± 1 Mb), and we tested about 1.25 × 106 gene–SNP pairs for each compartment. Nominal P values were calculated for each SNP–gene pair with FastQTL using linear regression with an additive effect model and were adjusted by six genotype PCs. Significance of the top associated variant per gene was estimated by adaptive permutation with the setting “–permute 10000” in FastQTL. Beta distribution–adjusted empirical P values were used to calculate q using Storey’s q method and a q threshold ≤0.05 was applied to identify eGenes (eQTL-containing genes).

Next, a genome-wide empirical P threshold, Pc, was defined as the empirical P of the gene closest to the 0.05 FDR threshold, P, was used to calculate a nominal P threshold for each gene based on the beta distribution model (from FastQTL) of the minimum P distribution (Pmin) obtained from the permutations for the gene. For each eGene, variants with a nominal P above the gene-level threshold were considered significant as eVariants.24

Meta-analysis of multiple-tissue eQTL mapping. The eQTL summary results of 44 other human tissues were downloaded from GTEx (v6p; www.gtexportal.org)31. To evaluate which of the GTEx eQTLs replicated the significant SNP–gene pair effects found in the kidney eQTLs, we calculated the fraction of true positive eQTLs, considering the formula π = 1 − χ, where χ is the estimated fraction of null eQTLs from the full distribution of P. For example, for tubule eQTL, eGenes and their top significant SNPs were selected. The nominal P values of these SNP–gene pairs in another GTEx eQTL were extracted, and then χ was calculated on these P values using Storey’s q method.24,25

To further define tubule-specific and glomerulus-specific eGenes, METASOFT, a meta-analysis method, was performed on all variant–gene pairs that were significant (FDR < 5%) in at least one of the 46 tissues (2 kidney compartments and 44 GTEx tissues) based on the single-tissue results from FastQTL. A random effects model in METASOFT (called RE2) was used, and the posterior probability that an eQTL effect exists in a given tissue (called m) was calculated for each SNP–gene pair and tissue tested. A significance cutoff of m = 0.9 was used to discover high-confidence eQTLs. On the basis of this meta-analysis result, tubule-specific eGenes (glomerulus-specific eGenes) were defined as having m > 0.9 in less than five tissues, including tubule (glomerulus). In contrast, tubule-shared eGenes (glomerulus-shared eGenes) were defined as having m = 0.9 in more than 40 tissues, including tubule (glomerulus) (Supplementary Fig. 8a).

To further test the sensitivity of the original results to these cut-offs, we changed the threshold numbers and re-estimated the enrichment results of tubule-specific eGenes and tubule-shared eGenes. Even though the P values are slightly inflated, we found that chronic kidney disease, glomerular filtration rate and metabolite traits are still the top traits enriched for tubule-specific eGenes, and tubule-shared eGenes are enriched for some common human diseases (Supplementary Fig. 8b–d).

To identify tubule-compartment-specific eGenes, glomerulus-compartment-specific eGenes and kidney-compartment-shared eGenes, meta-tissue, which was used to address effect-size heterogeneity in detection of eQTLs across multiple tissues, was performed on three eQTL datasets, tubule eQTLs, glomerulus eQTLs and our previous whole-kidney eQTLs, to remove the effect brought by considering multiple tissues from the same individuals.

Functional annotation. Traits and disease-associated genetic variants were extracted from GWAS Catalog v10.1 (accessed 4 January 2017; https://www.ebi.ac.uk/gwas/).14 These variants were used for enrichment analysis for kidney compartment eQTLs.

Human kidney-specific ChIP-seq data can be found at GEO accession numbers GSM621654, GSM670825, GSM621648, GSM772811, GSM661251, GSM1112806 and GSM1112807. Different histone markers were combined into chromatin states using ChromHMM.24-25. To quantify kidney eQTL enrichment in these regulatory regions, the top significant SNPs per eGenes were used as tested eQTL subset. To calculate enrichment significance, the data was compared between the tested eQTL subset and a randomly selected SNP set that was matched for MAF (< 0.01) and distance to the TSS of target genes (< 2.5kb). Transcription factor ChIP-seq (141 factors) was extracted from ENCODE project TFBS clusters (V3), and the enrichment analysis procedure was the same as formerly described.32

GWAS analysis. Leading SNPs of CKD GWAS, eGFR associated GWAS SNPs were collected from the CKDGen Consortium and several other studies to determine whether kidney disease traits in our eQTLs were enriched. This step yielded 83 leading SNPs.

Cocl. To estimate the posterior probabilities of whether two potentially related phenotypes—kidney disease and kidney compartment gene expression—share
common genetic causal variants in a given region, we performed colocalization analysis using coloc and summary data of GWAS and eQTL analyses. The eGFR-associated GWAS study from CKDGen Consortium was used as the CKD GWAS dataset. Then, summary data from both studies of SNPs within 100 kb of each leading SNP were used to calculate the posterior probability. In the coloc results, H3 represents the posterior probability that both traits are associated, but with different causal variants; H4 represents the posterior probability that both traits are associated and share a single causal variant. We used PP_H4 > 0.8 as the threshold of colocalization.

RTC. Regulatory trait concordance (RTC) score was used to assess whether a GWAS SNP tags the same functional variant as a regulatory variant. In short, if an eQTL variant and GWAS leading SNP were located in the same region between recombination hotspots, the eQTL phenotype (gene expression) was corrected for all the N variants within this region using linear regression, and then the residuals were used as N pseudophenotypes. The P values, obtained by testing for eQTL association between the eQTL variant and these N pseudophenotypes, were sorted (descending) and ranked. Then, the rank of the P value obtained from the pseudophenotype corrected by GWAS SNP was used to calculate the RTC score: \((N - \text{GWAS}\_\text{SNP}) / N\). The RTC score ranges from 0 to 1, and 1 represents higher likelihood of shared effect between eQTL and GWAS.

Single-cell RNA-seq of mouse kidney. Single-cell RNA-seq libraries were generated using the 10X Chromium Single Cell Instrument and the 10X Chromium TM 3′ Library-3′ Library Kit (12 cells) as per the manufacturer’s protocol. We have generated single-cell suspensions from seven different mouse kidney samples. The single-cell sequencing libraries were sequenced on an Illumina HiSeq with a 2 × 150 paired-end kit. The sequencing reads were demultiplexed, aligned to the mouse genome (mm10) and processed to generate gene–cell data matrix using Cell Ranger 1.3 (http://10xgenomics.com). We have sequenced 57,979 mouse kidney cells.

Cells were clustered using dimension reduction. We identified three cell types in the glomerulus and seven kidney cell types in the tubule in addition to six immune cells. Of the entire 16 cell clusters identified by single-cell RNA-seq in mice, 13 are previously known cell types, and the other 3 are novel cell types.

Pathway analysis of candidate genes. The pathway analysis of candidate genes, identified by colocalization between eGFR GWAS and kidney compartment-based eQTLs, was performed by the web-based program GIANT (http://giant.princeton.edu).

In silico kidney segment estimation. CIBERSORT was developed to identify the relative levels of distinct cell types within a complex expression admixture. Here, we use RNA-seq data of microdissected rat kidney tubule segments (https://hpcwebapps.cit.nih.gov/ESBL) as signature matrix, to decompose our 121 tubule samples, identifying their segment composition.

Isoform detection using LeafCutter. LeafCutter was used to quantify RNA splicing variation using short-read RNA-seq data by leveraging reads that span each intron to quantify intron usage across samples (http://davidknowles.github.io/leafcutter/). We performed LeafCutter on 121 tubule RNA-seq samples, by which a specific exon deletion was identified on DAB2.

SMR. SMR summary-databased Mendelian randomization was used to test for potential causal effects of, for example, gene on complex trait, given a SNP as an instrumental variable, using summary-level data from GWAS and eQTL studies. We performed SMR on CKD GWAS and eQTLs (tubule and blood) separately to estimate the direction of expression changes of DAB2 with eGFR, using multiple variants in the cis-eQTL region of the gene.

Mice. C9-knockout mice purchased from Jackson Lab (Stock no. 022779) were kindly provided by W.-C. from University of Pennsylvania. Dab2fl/fl mice were kindly provided by X.-X. from University of Miami. Ksp-Cre mice were purchased from Jackson Lab (stock no. 012237). R- to 10-week-old mice were used in this study. Mice were injected with EA (250 mg per kg body weight, dissolved in 300 mM NaHCO3) intraperitoneally and were euthanized 1 week later. For the UUO model, mice underwent ligation of the left ureter and were euthanized 1 week later. Sham-operated mice were used as controls. Animal studies were approved by the Animal Care Committee of the University of Pennsylvania.

Quantitative real-time PCR. RNA was isolated using RNeasy Mini Kit per the manufacturer’s instructions. 1 μg RNA was reverse transcribed using cDNA archival kit (Applied Biosystems no. 4368813), and quantitative real-time PCR was conducted in the Viia 7 System (Applied Biosystems) machine using SYBRGreen Master Mix (Applied Biosystems no. A25742). The data were normalized and analyzed using the ∆∆Ct method. Primers used are listed in Supplementary Table 4.

Histological analysis. We used formalin-fixed, paraffin-embedded kidney sections stained with periodic acid Schiff (PAS) or Picrosirius red (Polyscience no. 24901). Slides were examined, and pictures were taken with an Olympus BX43 microscope and Olympus DP73 Diagnostic CCD camera. Quantification was performed on Sirius Red-stained kidney sections. For each section, five random fields were quantified in a unified manner using ImageJ. Immunohistochemistry images of LRRC3 (https://www.proteinatlas.org/ENSG00000160233-LRRC3/tissue/kidney#img) and ANXA2 (https://www.proteinatlas.org/ENSG00000182718-ANXA2/tissue/kidney#img) were downloaded from Human Protein Atlas (v18. proteinatlas.org).

Primary culture of renal tubule cells. Kidneys were collected from Dab2fl/fl mice (males, 3–5 weeks old). Cells were isolated using 2 mg/ml collagenase I (Worthington Biochemical Product no. CLS-1) digestion for 30 min at 37 °C with gentle stirring. Cells were then filtered through the 100-μm mesh to isolate single cells. Cell suspensions were cultured in RPMI 1640 (Corning no. 10-040-CM) supplement with 10% FBS (Atlantic Biologicals no. S1950), 20 μg/ml EGF (Peprotech no. AF-100-15), 1 × ITS (Cultbio no. S1500-056) and 1% penicillin-streptomycin (Corning no. 30-002-CI) at 5% CO2 and 37 °C. When cell confluence reached 70%, the medium was replaced with serum-free RPMI and infected with Ad5CMV-eGFP (Ad-GFP) or Ad5CMVCre-eGFP (Ad-Cre-eGFP) (University of Iowa Gene Transfer Vector Core, Iowa City, IA) at 4 × 10^6 plaque forming units per ml for 24 h. Infection efficiency was estimated under fluorescence microscope by the presence of GFP positive cells. After that, cells were treated with 5 μg/ml TGF-β (Peprotech no. 100-21) for 24 h.

Western blot analysis. Cell lysates were prepared with SDS lysis buffer containing protease and phosphatase inhibitor cocktail (Complete Mini, Roche no. 11836135001) and phosphatase inhibitor (PhosSTOP, Roche no. 4906837001). Proteins were resolved on 8–12% gradient gels, transferred on to polyvinylidene difluoride membranes and probed with antibodies as below: DAB2 (BD Transduction Laboratories no. 610464; 1:1,000), p-SMAD2 (Cell signaling no. 3108; 1:1,000), SMAD2 (Cell signaling no. 5339; 1:2,000), p-SMAD3 (Abcam no. ab52903; 1:1,000), SMAD3 (Abcam no. ab18740; 1:2,000), p-JNK (Cell signaling no. 4668; 1:2,000), JNK (Cell signaling no. 9252; 1:2,000), p-ERK1/2 (Cell signaling no. 4370; 1:2,000), ERK1/2 (Cell signaling no. 4695; 1:2,000), p-p38 (Cell signaling no. 4511; 1:2,000), p38 (Cell signaling no. 8690; 1:2,000) and fibronectin (Abcam no. ab32419; 1:2,000). Anti-rabbit (Cell signaling no. 7074) or anti-mouse (Cell signaling no. 7076) IgG horseradish peroxidase (HRP) was used as a secondary antibody. Blots were detected by enhanced chemiluminescence (Western Lightning-ECL, Thermo Scientific). The full scans of the western blots are showed in Supplementary Fig. 9.

Statistics. Statistical analyses for animal studies were performed using GraphPad Prism 6.0 software. All values are expressed as mean and s.d. Unpaired two-sided Student’s t test was used for comparisons between two groups. One-way ANOVA with post hoc Tukey test was used to compare multiple groups. P < 0.05 was considered statistically significant.

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

Data availability

The eQTL data is publicly available at http://susztaklab.com/eqtl. RNA-seq data have been deposited in the Gene Expression Omnibus (GEO) with the accession code GSE115098. As the samples were collected from de-identified kidney tissue samples, no consent was obtained to share individual-level genotype data.

References

61. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 4, 7 (2015).
62. The 1000 Genomes Project Consortium. A global reference for human genetic variation. Nature 526, 68–74 (2015).
63. Price, A. L. et al. Principal components analysiscorrects for stratification in genome-wide association studies. Nat. Genet. 38, 904–909 (2006).
64. Delanueva, O., Marchini, J. & Zagury, J. F. A linear complexity phasing method for thousands of genomes. Nat. Methods 9, 179–181 (2011).
65. Howie, B., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. 5, e1000529 (2009).
66. Howie, B., Marchini, J. & Stephens, M. Genotype imputation with thousands of genomes. PLoS Genet. 3, e1000109 (2007).
67. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21 (2013).
68. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics 31, 166–169 (2015).
69. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550 (2014).
70. Shabalin, A. A. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. Bioinformatics 28, 1353–1358 (2012).
71. Ongen, H., Buil, A., Brown, A. A., Dermitzakis, E. T. & Delaneau, O. Fast and efficient QTL mapper for thousands of molecular phenotypes. *Bioinformatics* **32**, 1479–1485 (2016).
72. Stegle, O., Parts, L., Piipari, M., Winn, J. & Durbin, R. Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. *Nat. Protoc.* **7**, 500–507 (2012).
73. Storey, J. D. A direct approach to false discovery rates. *J. R. Stat. Soc. Series B Stat. Methodol.* **64**, 479–498 (2002).
74. Sul, J. H., Han, B., Ye, C., Choi, T. & Eskin, E. Effectively identifying eQTLs from multiple tissues by combining mixed model and meta-analytic approaches. *PLoS Genet.* **9**, e1003491 (2013).
75. Bernstein, B. E. et al. The NIH Roadmap Epigenomics Mapping Consortium. *Nat. Biotechnol.* **28**, 1045–1048 (2010).
76. Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nat. Methods* **9**, 215–216 (2012).
77. Zhu, Z. et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* **48**, 481–487 (2016).
78. Tao, W., Moore, R., Smith, E. R. & Xu, X. X. Endocytosis and physiology: insights from disabled-2 deficient mice. *Front. Cell Dev. Biol.* **4**, 129 (2016).
79. Moore, R., Cai, K. Q., Tao, W., Smith, E. R. & Xu, X. X. Differential requirement for Dab2 in the development of embryonic and extra-embryonic tissues. *BMC Dev. Biol.* **13**, 39 (2013).
80. Uhlén, M. et al. Proteomics. Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).
Corresponding author(s): Katalin Susztak

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

- n/a
- Confirmed

☐ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement

☐ An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly

☐ Only common tests should be described solely by name; describe more complex techniques in the Methods section.

☐ A description of all covariates tested

☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons

☐ A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)

☐ For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted

☐ Give P values as exact values whenever suitable.

☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings

☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes

☐ Estimates of effect sizes (e.g. Cohen’s d, Pearson’s r), indicating how they were calculated

☐ Clearly defined error bars

☐ State explicitly what error bars represent (e.g. SD, SE, CI)

Software and code

Policy information about availability of computer code

Data collection

N/A

Data analysis

PLINK (v1.9), IMPUTE2 (v2.3.2), FastQTL (v2.165), Coloc (3.1), METASOFT (v2.0.1), LeafCutter (v2.0), STAR (v2.4.1d), Matrix eQTL (v2.1.0), ChromHMM (v1.15), FastQC (v0.11.2), HTSeq (v0.10.0), DESeq2 (v3.7), Meta-Tissue (v0.5), LocusZoom (v0.4.8), GIANT (the latest version, 2018), SMR (v0.710), CIBERSORT (v1.05), PEER (v1.3), EIGENSTRAT (v7.2.0), R (v3.2.2), Python (v2.7.5), GraphPad Prism (v6.0), Image J (v1.49)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.
Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

We have generated a publicly searchable eQTL database (http://susztaklab.com/eqtl), to present the detailed results including boxplots. RNA-sequencing data has been deposited in the Gene Expression Omnibus (GEO) with the accession code GSE115098. Since the samples were collected from de-identified kidney tissue samples, no consent was obtained to share individual-level genotype data.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

121 samples (tubule) and 119 samples (glomerulus) were used in our eQTL analysis, and this sample size was similar as what GTEx Consortium used. The number of identified eGenes was also in line with expectations. No other specific statistical methods were used to predetermine the sample size.

For animal study, a sample size of at least 3 mice per group were determined based on the means and variation of previous pilot and published experiments (PMID: 25419765, 20978353). We routinely exceeded the minimal number required for each experiment as indicated in the figure legends.

Data exclusions

For eQTL analysis, only samples with European ancestry and absence of significant kidney structural changes (tubule fibrosis < 10%, glomerular sclerosis < 10%) were used. The major thesis of the current work is to examine genotype and gene expression changes in kidney compartments (limited cell types) that are relevant to disease development. Diseased kidneys contain large amounts of fibroblasts and immune cells in variable proportions compared to control healthy kidneys.

No data were excluded from the animal study.

Replication

All attempts at replication were successful.

Randomization

Randomization was not used in human sample collection. Mice were randomly distributed prior to FA injection and UUO surgery.

Blinding

Blinding was not used in human sample collection. Staining images were evaluated by pathologists who were blinded to other data. Investigators were blinded to allocation during experiments and outcome assessments.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a Involved in the study

- Unique biological materials
- Antibodies
- Eukaryotic cell lines
- Palaeontology
- Animals and other organisms
- Human research participants

Methods

n/a Involved in the study

- ChIP-seq
- Flow cytometry
- MRI-based neuroimaging
Unique biological materials

Policy information about availability of materials

Obtaining unique materials Limited number of human kidney tissues and slices

Antibodies

Antibodies used

- Dab2 (BD Transduction Laboratories# 610464; 1:1000), Phospho-SMAD2 (Cell signaling#3108; 1:1000), SMAD2 (Cell signaling#5339; 1:2000), Phospho-SMAD3 (Abcam#ab40854; 1:2000), Phospho-JNK (Cell signaling#4668; 1:2000), JNK (Cell signaling#9252; 1:2000), Phospho-Erk1/2 (Cell signaling#4370; 1:2000), Erk1/2 (Cell signaling#4695; 1:2000), Phospho-p38 (Cell signaling#4511; 1:2000), p38 (Cell signaling#8690; 1:2000), Fibronectin (Abcam#ab32419; 1:2000)

Validation All the antibodies were previously validated by the manufacturer (online at BD Transduction Laboratories, Cell Signaling and Abcam websites).

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s) Primary TECs isolated from mouse kidneys were used, but no established cell lines were used in this study.

Authentication Primary TECs culture protocol was previously validated in our lab (PMID: 25419765, 26776520).

Mycoplasma contamination Primary TECs were not tested for mycoplasma contamination.

Commonly misidentified lines (See ICLAC register) N/A

Animals and other organisms

Policy information about studies involving animals: ARRIVE guidelines recommended for reporting animal research

Laboratory animals C9-knockout mice purchased from Jackson Lab (Stock No: 022779) were kindly provided by Dr. Wen-chao Song from University of Pennsylvania. Dab2 flox/flox mice were kindly provided by Dr. Xiang-Xi Xu from University of Miami. Ksp-Cre mice were purchased from Jackson Lab (Stock No: 012237). 8- to 10-week-old male mice were used in this study.

Wild animals This study did not involve wild animals.

Field-collected samples This study did not involve samples collected from the field.

Human research participants

Policy information about studies involving human research participants

Population characteristics We conducted expression quantitative trait loci (eQTL) analyses separately on the glomerular and tubular portions of healthy human kidney samples (absence of significant kidney structural changes (tubular fibrosis < 10%, glomerular sclerosis < 10%)) obtained from 151 subjects of European descent. The detailed information of human participants can be found in Supplementary Table 1.

Recruitment Kidney samples were obtained from surgical nephrectomies approved by the University of Pennsylvania Institutional Review Board. Nephrectomies were de-identified, and the corresponding clinical information was collected through an honest broker.