

1 **Mapping the genetic architecture of human traits to cell types in the kidney identifies**  
2 **mechanisms of disease and potential treatments**

3

4 Xin Sheng<sup>1,2,3</sup>, Ziyuan Ma<sup>1,2,3</sup>, Junnan Wu<sup>1,2,3</sup>, Hongbo Liu<sup>1,2,3</sup>, Chengxiang Qiu<sup>1,2,3</sup>, Zhen Miao<sup>1,2,3</sup>,  
5 Matthew J. Seasock<sup>1,2,3</sup>, Matthew Palmer<sup>4</sup>, Myung K. Shin<sup>5</sup>, Kevin L. Duffin<sup>6</sup>, Steven S. Pullen<sup>7</sup>,  
6 Todd L. Edwards<sup>12</sup>, Jacklyn N. Hellwege<sup>12</sup>, Adriana M. Hung<sup>13</sup>, Mingyao Li<sup>8</sup>, Benjamin  
7 Voight<sup>2,3,9,10</sup>, Thomas Coffman<sup>11</sup>, Christopher D. Brown<sup>3</sup>, Katalin Susztak<sup>1,2,3\*</sup>

8

9 <sup>1</sup>Department of Medicine, Renal Electrolyte and Hypertension Division, University of  
10 Pennsylvania, Philadelphia, PA, USA

11 <sup>2</sup>Institute of Diabetes Obesity and Metabolism, University of Pennsylvania, Philadelphia, PA,  
12 USA

13 <sup>3</sup>Department of Genetics, University of Pennsylvania, Philadelphia, PA, USA

14 <sup>4</sup>Department of Pathology and Laboratory Medicine at the Hospital of the University of  
15 Pennsylvania, Philadelphia, PA, USA

16 <sup>5</sup>Merck & Co., Inc., Kenilworth, NJ, USA

17 <sup>6</sup>Eli Lilly and Company Lilly Corporate Center, Indianapolis, IN, USA

18 <sup>7</sup>Department of Cardiometabolic Diseases Research, Boehringer Ingelheim Pharmaceuticals,  
19 Ridgefield, Connecticut, USA

20 <sup>8</sup>Department of Epidemiology and Biostatistics, University of Pennsylvania, Perelman School of  
21 Medicine, Philadelphia, PA, USA

22 <sup>9</sup>Department of Systems Pharmacology and Translational Therapeutics, University of  
23 Pennsylvania, Philadelphia, PA, USA

24 <sup>10</sup>Institute of Translational Medicine and Therapeutics, University of Pennsylvania, Philadelphia,  
25 PA, USA

26 <sup>11</sup>Cardiovascular and Metabolic Disorders Program, Duke-NUS Medical School, Singapore  
27 Division of Nephrology, Duke University School of Medicine, Durham NC, USA

28 <sup>12</sup>Division of Epidemiology, Department of Medicine, Vanderbilt Genetics Institute, Vanderbilt  
29 University Medical Centre, Nashville, TN, USA

30 <sup>13</sup>Division of Nephrology and Hypertension, Department of Medicine, Vanderbilt University  
31 Medical Center, Nashville, TN, USA

32

33 \*Correspondence: [ksusztak@penncmedicine.upenn.edu](mailto:ksusztak@penncmedicine.upenn.edu)

34 Katalin Susztak, MD, PhD

35 Perelman School of Medicine, University of Pennsylvania

36 3400 Civic Center Blvd

37 12-123 Smilow Translational Building

38 Philadelphia, PA 19104

39 215 898 2009

40

41

1 **Abstract**

2

3 The functional interpretation of GWAS remains challenging due to cell-type dependent influences  
4 of genetic variants.

5 Here, we generated comprehensive maps of expression quantitative trait loci (eQTL) for 659  
6 microdissected human kidney samples and identified cell-type eQTLs by mapping interactions  
7 between cell type abundance and genotype. Separately, we generated single cell open chromatin  
8 maps (by snATAC-seq) for human kidney samples. We highlight critical enrichment of proximal  
9 tubules in kidney function and endothelial cells and distal tubule segments in blood pressure by  
10 partitioning heritability using stratified LD-score regression to integrate GWAS with scRNA-seq  
11 and snATAC-seq data. Bayesian colocalization analysis nominated more than 200 genes for  
12 kidney function and hypertension. Our study clarifies the mechanism of the most commonly used  
13 antihypertensive and renal protective drugs and identifies drug repurposing opportunities for  
14 kidney disease.

15

16 **One Sentence Summary:** We define causal cell types, genes and mechanism for kidney  
17 dysfunction.

18

## 1 **Main Text**

2

3 Chronic kidney disease (CKD) affects more than 800 million people worldwide (1). CKD causes  
4 toxic mineral, metabolite, salt, water accumulation (2), in addition to anemia, bone disease and  
5 hypertension leading to premature death (3). Therapeutic options for CKD have been limited.  
6 Guidelines mandate the use of agents blocking the renin-angiotensin-aldosterone-system (RAAS);  
7 such as angiotensin converting enzyme inhibitors (ACEi), and angiotensin receptor blockers  
8 (ARBs). Despite ACEi and ARBs were introduced more than 30 years ago, their mechanisms of  
9 action have not been fully clarified. As kidney disease is still responsible for 1 in 60 deaths  
10 worldwide, new therapeutics are desperately needed for CKD. Nelson et al. concluded, that  
11 pipeline drug targets with human genetic evidence of disease association are twice as likely to lead  
12 to approval (4), highlighting the critical importance of understanding disease genetics.

13

14 Kidney function is highly heritable, with estimates of 30-50% (5). Large population-based genome  
15 wide association studies (GWAS) have identified around 300 loci with statistically significant  
16 reproducible association with kidney function (eGFR; estimated glomerular filtration rate) (5).  
17 Despite the remarkable success of genetic signal mapping in GWAS, the functional interpretation  
18 of GWAS remains challenging. First, it is unclear in which tissues and cell types these variants  
19 active, and how they disrupt specific biological networks to impact disease risk. Second, more than  
20 90% of the identified variants are located in the non-coding regions, complicating the precise  
21 identification of CKD risk target genes. Furthermore, extensive linkage disequilibrium (LD) in the  
22 human genome confounds efforts to pinpoint causal variants.

23

24 A commonly used approach to map variants to causal genes utilizes expression quantitative trait  
25 loci (eQTLs) (6, 7). The Genotype-Tissue Expression (GTEx) project has been a highly valuable  
26 resource to catalogue eQTLs across multiple human tissues, including kidney cortex with a limited  
27 sample size (n=73) (6, 8). Our and other groups have generated bulk glomerular and tubule specific  
28 eQTL data for about 120 kidney samples. This dataset was able to prioritize disease-causing genes  
29 for close to 24 of the GWAS identified loci (7). As both GWAS and eQTL have significant  
30 limitations on defining causal variants due to the linkage disequilibrium (9), the integration of  
31 these datasets focused around causal gene (rather causal variant) identification (10, 11). Functional

1 genomics or epigenetics can provide an orthogonal information for GWAS interpretation by  
2 prioritizing likely causal variants (12, 13). Functionally important gene regulatory regions are  
3 mostly nucleosome-free and can be identified by transposase insertion (ATAC), DNase  
4 hypersensitivity (14) and histone tail modifications (15). Datasets generated from whole human  
5 kidney tissue highlighted a modest enrichment of the GWAS signals in kidney specific enhancers  
6 (12, 13).

7  
8 Another critical limitation of the eQTL and open chromatin analysis has been the cell type  
9 heterogeneity of the kidney. Recent single cell RNA-seq (scRNA-seq) studies have highlighted at  
10 least 30 different cell types in the kidney (16, 17). Studies analyzing a variety of blood cells  
11 identified (often rare) single disease driving cell types that was hard to read out using bulk datasets  
12 (18). Few approaches have been developed to identify disease-causal cell types. For example, cell-  
13 type-specific eQTLs have been generated using sorted (bulk) cell types (19, 20), an approach that  
14 is feasible for immune cells. It has been hard to generate cell-type specific eQTL data using  
15 scRNA-seq, as most current scRNA-seq methods suffer from uneven cell capture efficiency and  
16 low accuracy to define gene expression (16). Bulk tissue analysis coupled with computational  
17 cellular deconvolution have recently been developed for *in silico* estimation of cell type eQTLs,  
18 however, the accuracy of the computational methods has not been validated (21-25). Single cell  
19 epigenome studies, such as single-nucleus ATAC-seq (snATAC-seq), could provide potential  
20 insights into causal variants identification by overlapping GWAS variants with cell-type specific  
21 open chromatin regions (26-28).

22  
23 Here, we combined several orthogonal approaches to prioritize causal cell types, variants, and  
24 genes for complex kidney disease-associated phenotypes. First, we identified expression of  
25 quantitative trait loci for 659 kidney samples by bulk eQTL analysis. We developed cell fraction  
26 adjusted SNP-cell type interaction (eQTL(ci)) analysis using estimated cell fractions, using *in*  
27 *silico* deconvolution (29). Separately, we generated single nuclear ATAC-seq (snATAC-seq) maps  
28 for the human kidney and assessed enrichment of GWAS heritability in each cell type, obtained  
29 from scRNA-seq and snATAC-seq data, using stratified LD-score regression. Our study not only  
30 highlighted critical cell type convergence of kidney associated traits, but also nominated causal

1 variants, cell types and regulatory mechanisms for renal disease and defines mechanisms for  
2 commonly used antihypertensive and renoprotective drugs.

3

## 4 **Results**

5

### 6 **Cell fraction adjusted eQTL analysis of microdissected human kidney samples**

7

8 While GWAS efficiently describe the genotype association with specific traits and diseases,  
9 understanding the genotype effect on gene expression (eQTL) has been essential to nominate target  
10 genes for GWAS variants. Prior eQTL studies have been limited, as they used whole organ tissue,  
11 comprised of many cell types, and small sample sizes (7, 30). Here we collected a large number of  
12 human kidneys from subjects with different ethnicities, disease status and disease stages (**Tables**  
13 **S1 and S2**). To reduce cell heterogeneity, we manually microdissected each tissue sample into  
14 glomerular and tubule compartments, and performed RNA-sequencing and genotyping on 659  
15 samples (356 tubules and 303 glomeruli) (**Fig. 1A**). We initially performed eQTL analysis  
16 separately in glomerular and tubule samples following the GTEx v7 pipeline (6). The analysis took  
17 age, gender, collection site, RNA integrity (RIN), sequencing parameters (batch effect, read depth,  
18 and sequencing types), genotyping principle components (PCs), and kidney disease severity into  
19 consideration. This first cis-eQTL analysis identified 3,599 and 5,871 eGenes (genes that showed  
20 association with at least one SNP at a  $q$ -value $<0.05$ ) in tubule and glomerular samples, respectively  
21 (**Fig. S1A and B**).

22

23 Since we observed heterogeneity in the relative distributions of cell populations across the RNA-  
24 seq samples (**Fig. 1B and C**), we hypothesized that considering the cell population distributions  
25 of each sample would improve the quality of eQTL inferences by reducing heterogeneity and  
26 increasing our power to detect associations within tissues and/or cell types. We used kidney single  
27 cell gene expression matrix (29) and an *in silico* deconvolution method implemented in  
28 CIBERSORTx (22, 25) to estimate cell fractions in each microdissected sample (**Fig. 1A,**  
29 **Methods**). As expected, we observed that proximal tubules (PT) represented the major cell type  
30 in the human kidney tubule samples (**Fig. 1B**) and endothelial cells in glomeruli (**Fig. 1C**). Next,  
31 we performed a second eQTL analysis by further adjusting for cell population differences across

1 samples. We found that comparing to our initial eQTL analysis model without considering the cell  
2 fractions, the inclusion of cell fractions has improved the eGene identification by 28% (1,256  
3 genes) and 18% (1,194 genes) in tubule (**Fig. S1A**) and glomerular (**Fig. S1B**) compartments,  
4 respectively.

5  
6 Our final eQTL(cf) model included estimated latent variable adjustment by PEER factors in  
7 addition to cell fraction adjustment (**Methods**). This analysis identified 9,209 and 10,106 eGenes  
8 (**Fig. 1D and E**), including 6,821 and 7,501 protein coding genes, and 865,410 and 897,548  
9 significant SNP-gene pairs (at a q-value <0.05) (31), in tubular and glomeruli samples, respectively.  
10 It is important to note that this eQTL(cf) model replicated ~90% eGenes identified by the general  
11 eQTL model (only including PEER factors) (**Fig. S1C and D**), which is consistent with previous  
12 findings that PEER factors modeled cellular heterogeneity reasonably well (**Fig. S2**) (32).

13  
14 Our eQTL(cf) model identified more than twice as many eGenes as reported in our previous  
15 publication (7) (**Fig. S3**), which is likely due to increasing power of eGene discovery with a larger  
16 sample size and better adjustments for confounders. Several genes, such as *KIAA1841* (**Fig. S4A**)  
17 and *MRPL43* (**Fig. S4B**) did not reach significance in the previous eQTL analysis, but showed  
18 strong associations between gene expression and genotypes in our eQTL(cf) model, illustrating  
19 the power of the new dataset. The significant SNP-gene pairs identified by our previous study  
20 could be successfully replicated by the eQTL(cf) model with  $\pi_1=0.95$  in tubule and  $\pi_1=0.99$  in  
21 glomeruli, respectively (33). There was a 90% direct overlap between the newly and previously  
22 reported eGene list (**Fig. 1D and E**). Meta-analysis combining evidence from results with 48  
23 tissues in the published GTEx data showed a large number of shared eGenes (34) and also 1,309  
24 tubule-specific (**Fig. 1F**) and 1,458 glomerulus-specific (**Fig. 1G**) eGenes (M-value>0.9) in our  
25 eQTL(cf) study.

26 In summary, here we report bulk and cell fraction adjusted eQTL data for more than 600  
27 microdissected human kidney glomeruli and tubule samples. The increased sample size and  
28 improved cell fraction estimation have more than doubled the number of identified eGenes and  
29 revealed a large number of kidney compartment specific eGenes.

30  
31 **Defining cell-type dependent activities of genetic variants on gene expression**

1  
2 In order to identify eQTLs that modulate gene expression in a cell-type specific manner, we  
3 adopted a linear regression model that included an interaction term between cell fraction and  
4 genotype (**Methods**). We refer to these eQTLs as cell type interacting eQTLs (eQTL(ci)) (8) (**Fig.**  
5 **2A**). For example, kidneys of subjects with higher allele dosage of G (SNP rs12481710), showed  
6 a positive correlation between PT fractions and *SYCP2* (Synaptonemal complex protein 2)  
7 expression (**Fig. 2B**), indicating a cell-type specific genotype-gene expression interaction. In total,  
8 we identified 1,613 and 713 protein-coding eQTL(ci) genes in tubule (**Table S3**) and glomerular  
9 (**Table S4**) samples, respectively (**Fig. 2C**) (8), by evaluating the significance of the interaction  
10 effect between genotype and cell fraction (**Methods**) in 23 cell types (n=13 in tubule, n=10 in  
11 glomeruli) of the kidney. Kidney PT cells showed the highest number of eQTL(ci)s in the  
12 microdissected human kidney tubule compartment (**Fig. 2C**), while glomerular endothelial cells  
13 (EndoG) showed the highest number of eQTL(ci)s in glomeruli (**Fig. 2C**). The number of  
14 eQTL(ci)s for each cell type and sample size was in line with the previous publication (8).

15  
16 To characterize the cell type specificity of eQTL(ci)s, we examined the eQTL(ci) sharing across  
17 kidney cell types using a multivariate adaptive shrinkage method (*mash*) (35), which improves the  
18 effect size estimation for the cell fraction and genotype interaction. *Mash* leverages the correlations  
19 of effect sizes across all cell types to re-estimate eQTL(ci) effect size at each locus. Cell type  
20 specific eQTL(ci) was defined as the eQTL(ci) at LFSR<0.05 (the local false sign rate (36)) only  
21 in a single cell type. We found that using this stringent threshold, about 18% (n=3,544 of 20,146  
22 SNP-Gene pairs) eQTL(ci)s were cell type specific across the 23 highly correlated kidney and  
23 immune cell types (**Fig. S5**).

24  
25 Here, we show that lipoprotein A (*LPA*) expression significantly correlated with rs783153  
26 (eQTL(cf) P-value=5.76E-46) in tubule (**Fig 2D**). This SNP modified the relationship between PT  
27 cell fraction and *LPA* expression, such as *LPA* expression in subjects with increasing dosage of G  
28 allele is increased with increasing PT fractions. The *LPA* gene expression showed a greater  
29 increase with PT fractions in individuals with genotype G/G at rs783153, compared to individuals  
30 with genotype T/G and T/T, which is consistent with the bulk eQTL data. T-statistics for eQTL(ci)s  
31 in each cell type of the kidney tubules, which were estimated using *mash* and our eQTL(ci) linear

1 regression model, show consistent regulatory directions in each cell type. Interestingly, the eQTL  
2 effect observed in tubule was consistent with the PT specific regulatory eQTL(ci) effect (**Fig 2D**)  
3 (with LFSR<0.05 only in PT cells). These results indicate that for this eQTL pair (rs783153-*LPA*),  
4 the PT eQTL(ci) acted as the major driver of eQTL(cf) effect in the bulk tubule tissue. Similarly,  
5 we found that the eQTL pair (rs8023408-*SV2B* (Synaptic Vesicle Glycoprotein 2B)), which  
6 showed that individuals with genotype A/A had higher *SV2B* expression in the bulk glomerular  
7 tissue (**Fig S7**), was related to a EndoG specific eQTL(ci) effect. There was a greater increase in  
8 *SV2B* expression as the fraction of EndoG increased in subjects with A/A genotype. This was an  
9 EndoG specific eQTL(ci) (LFSR<0.05 only in EndoG cells) and the genotype did not significantly  
10 modify the cell fraction and gene expression relationship in other cell types. This indicated that  
11 the EndoG specific eQTL(ci) (rs783153-*LPA*) effect was the key driver of the eQTL(cf) (**Fig. S6**)  
12 in the bulk glomerular tissue. Taken together, using the eQTL(ci) model, we have defined a list of  
13 genotype-gene expression correlations that is likely driven by a specific cell type in the human  
14 kidney.

15

### 16 **Single cell resolution regulatory maps for the human kidney**

17

18 Cell-type specific regulatory elements have not been mapped for the human kidney, which is an  
19 important knowledge gap as functionally relevant GWAS variants are likely enriched in cell type  
20 specific regulatory regions (28). To characterize gene regulatory regions at a cellular resolution,  
21 we generated single nuclei accessible chromatin information for 12,720 human kidney cells using  
22 the 10x Genomics snATAC-seq platform (**Fig. 3A**). After initial quality control, we identified 8  
23 distinct cell clusters (**Fig. 3B**). By comparing the read density of the promoter regions in each  
24 cluster, we found that the initial clustering showed a clear match with previously identified key  
25 kidney gene expression markers (16, 37), such as *KDR* for endothelial cells (Endo), *NPHS2* for  
26 podocytes (Podo), *LRP2* for proximal tubule cells (PT), *UMOD* for loop of Henle (LOH),  
27 *SLC12A3* for distal convoluted tubule (DCT), *AQP2* for principal cell (PC), *ATP6V1G3* for  
28 intercalated cells (IC) of the collecting duct (CD) and *LST1* for immune cells (Immune),  
29 respectively (**Fig. 3C and Fig. S7**), indicating the robustness of the independently identified  
30 clusters. This analysis identified a total of 359,019 accessible chromatin peaks in the entire dataset  
31 for the human kidney (**Methods**). Single cell open chromatin peaks were enriched in human kidney



1 ChromHMM defined promoter and enhancer regions (**Fig. S8**). For example, 38% of open  
2 chromatin regions observed in PT cells overlapped with enhancer regions and 17% of peaks  
3 overlapped with promoter regions in adult bulk human kidney samples (**Table S5**).

4  
5 Next, we generated a list of differentially accessible chromatin regions (DARs) by comparing open  
6 chromatin regions across clusters and identified 60,661 cell-type specific peaks (**Methods**). To  
7 identify transcription factors (TFs) that predominantly occupy cell type specific open chromatin  
8 regions, we performed motif enrichment analysis of differentially accessible regions (DARs) for  
9 each cell type using chromVAR (38) and Homer (39). This analysis identified HNF4A motif  
10 enrichment in proximal tubules, ESRRB in loop of Henle, and ERG in endothelial cells (40) (**Fig.**  
11 **S9**). Furthermore, TF footprinting analysis across the 8 major cell types (**Fig. 3D and Fig. S10**)  
12 showed cell type specific enrichment of TF binding to HNF4A (**Fig. 3D**) and HNF4G (**Fig. S10**)  
13 in proximal tubule, and tubule compartment specific motif enrichment of TF HNF1B (**Fig. 3D**).  
14 Overall, these results provide a reference landscape for chromatin accessibility in the adult human  
15 kidneys at single cell resolution.

16  
17 The cell-type specificity of the eQTL(ci) has not been experimentally validated yet. We next  
18 estimated the enrichment of eQTL(ci) variants in the associated cell type specific open chromatin  
19 regions (by snATAC-seq) (41) (**Methods**). We found that for most common kidney cell types  
20 (EndoG, Endo, Podo, PT, ALOH and IC-B), eQTL(ci) SNPs showed the strongest enrichment in  
21 the open chromatin regions of the corresponding cell type (**Fig. 3E**), indicating consistency  
22 between cell type interaction eQTLs and the corresponding cell type specific regulatory regions.  
23 The similarity of open chromatin in DCT and LOH (**Fig. S9**) cells likely explained the overlap we  
24 observed for these cells.

25  
26 To summarize, we generated a single cell resolution regulatory map of genetic regulation of gene  
27 expression for the human kidney. We found that our cell-type specific eQTL(ci)s enriched in open  
28 chromatin regions of the corresponding cell type using snATAC-seq data, indicating accuracy of  
29 the computational eQTL(ci)s identification.

30  
31 **Single cell annotation highlights cell-type convergence of kidney relevant endophenotypes**

1  
2 To prioritize potential driver cell types for specific kidney dysfunction associated endophenotypes  
3 (such as glomerular filtration rate, albuminuria, and hypertension), we first calculated and ranked  
4 the expression specificity of each gene for each cell type from the single cell RNA-seq data (16).  
5 For a cell type, higher gene expression specificity deciles were associated with increasing levels  
6 of cell type expression specificity (**Methods**). Next, we assessed the partitioning heritability  
7 enrichment from GWAS summary statistics in genes of each decile (**Fig. 4A**) using linkage  
8 disequilibrium score regression (LDSC) (42). We observed that the enrichment of eGFR SNP  
9 heritability increased with increasing gene expression specificity for PT cells (**Fig. 4B**). Similar  
10 positive correlation was observed between SBP and DCT (**Fig. 4C**), PC (**Fig. S11**) and endothelial  
11 cells (**Fig. 4D**), suggesting the importance of these cell types for the tested endophenotypes.

12  
13 Next, we adopted a generalized gene set enrichment analysis established in MAGMA (43) to  
14 evaluate whether gene-level genetic association with kidney relevant phenotypes increased  
15 linearly with cell type specificity (**Methods**). We have analyzed 24 different GWAS traits (**Fig.**  
16 **4E and Fig. S12**) including three independent eGFR GWAS (5, 44, 45), two systolic blood  
17 pressure (SBP) GWAS (46), serum urate (47) (Urate), urinary albumin-to-creatinine ratio (UACR)  
18 (48), urinary sodium (Na<sup>+</sup>) and potassium (K<sup>+</sup>) excretion (49), and cardiometabolic traits. As  
19 established earlier, multiple sclerosis showed significant enrichment in lymphocytes also in our  
20 kidney dataset (50). We found the strongest enrichment for kidney function (eGFR) and serum  
21 urate level in proximal tubules (51) (**Fig. 4E**). Genetic signals for blood pressure showed  
22 enrichment in endothelial cells.

23  
24 Simultaneously, we adopted a third independent analysis by implementing LDSC using the single  
25 cell open chromatin information. We note that snATAC-seq LDSC evaluates the heritability of  
26 GWAS variants under snATAC-seq peaks, which is distinct from the measured gene expression  
27 output of scRNA-seq. We found a highly consistent enrichment of multiple eGFR GWAS signals  
28 in PT cells. Serum urate signals were the most enriched in PT cells comparing to other cells (**Figs.**  
29 **4E and Fig. S12**). We observed enrichment of SBP genetic signals in endothelial cells, tubule cells  
30 in the distal nephron (DCT and PC cells) (**Fig. 4E and Fig. S12**).

31

1 To summarize, using a variety of orthogonal datasets and methods, we observed important cell  
2 type convergence for several tested endophenotypes. The strongest of all was the enrichment  
3 between eGFR associated genetic signals and proximal tubule cells (**Fig. 4B and E**) and SBP  
4 associated genetic signals in endothelial (**Fig. 4E**) and DCT cells (**Figs. 4C, E and S11**).

## 5 6 **Comprehensive gene prioritization provides new mechanistic insights into kidney function** 7 **and blood pressure regulation**

8  
9 Next, we prioritized genes that likely mediate phenotypic variations for specific renal  
10 endophenotypes. Using the Bayesian test for colocalization (implemented in coloc) (52) to  
11 integrate 6 kidney related traits including eGFR, SBP, serum urate, UACR, urinary Na<sup>+</sup> and K<sup>+</sup>,  
12 and kidney microdissected compartment eQTL(cf)s data, we identified 240 colocalized protein  
13 coding genes in the tubule (**Table S6**) and 230 genes in the glomerular compartment (**Table S7**).  
14 At these regions, causal genetic variants driving gene expression changes and the variations of  
15 these 6 renal endophenotypes were shared. The analysis prioritized 182 genes for kidney function  
16 (eGFR) (**Fig. 5A and Table S8**) (5). This number was markedly increased compared to the 27  
17 genes prioritized by earlier studies (7). Similar colocalization analysis prioritized 88 genes for the  
18 340 GWAS loci (46) associated with SBP (**Fig. 5B and Table S9**), where the causal genetic  
19 variants associated with blood pressure and gene expression were shared. The distribution of the  
20 eGFR and blood pressure prioritized genes in the glomerular and tubular compartments across the  
21 genome is illustrated in **Fig. 5C and Fig. S13**. Only 10 genes were prioritized for both SBP and  
22 eGFR, indicating that genetic architecture of these two traits are likely distinct, which is consistent  
23 with the LDSC analysis estimating the genetic correlation of these 6 GWAS traits (**Fig. S14**) (9).  
24 One of the loci and genes that was associated with both traits was *FGF5* (Fibroblast Growth Factor  
25 5). Allelic dosage of G at the SNP rs3733336 position was associated with a higher *FGF5*  
26 expression levels in glomeruli (**Fig. S15A**). There was a strong correlation between the eQTL and  
27 GWAS effect sizes of SNPs located within ±100 kb of this locus for both SBP (**Fig. S15B**) and  
28 eGFR GWAS (**Fig. S15C**).

29  
30 To understand whether the eQTL(ci) data could be used to improve GWAS target identification  
31 and also to nominate causal cell types, we repeated the Bayesian colocalization analysis using

1 GWAS summary data of the 6 kidney related traits and the eQTL(ci)s data. We prioritized 61  
2 potentially causal protein coding genes, where variants for the 6 renal endophenotypes and cell-  
3 type specific gene regulation were shared (**Table S10**). We again observed a significant  
4 enrichment of PT eQTL(ci)s and eGFR GWAS, such as 23 of the 40 showed colocalization with  
5 interaction PT eQTL(ci)s. For SBP GWAS, we observed significant enrichment for EndoG eQTL  
6 (ci)s, where 9 of the 28 identified variants showed cell type enrichment (**Fig. 5D, Figs. S16 and**  
7 **S17**).

8  
9 Functional annotation (Gene Ontology) of prioritized causal genes for kidney function showed  
10 enrichment for cellular metabolic processes, including organic substrate, nitrogen compound,  
11 sulfur and macromolecule metabolism (**Fig. S18A and Table S11**). The enriched pathways were  
12 consistent with the functional role of the kidney in detoxification and metabolism. Gene ontology  
13 annotation prioritized biological processes for HTN shown enrichment for angiogenesis and  
14 circulation (**Fig. S18B and Table S12**).

15  
16 Taken together, kidney compartment-based eQTL(cf)s colocalization annotation of renal disease  
17 endophenotypes increased the prioritized genes by 10-fold (230 from the previous 27 (7)).  
18 Functional annotation highlighted the likely role of metabolism in kidney function regulation and  
19 angiogenesis in blood pressure genetics. Furthermore, eQTL(ci) enabled the discovery of driver  
20 cell types and genes previously masked by bulk tissue analysis.

21  
22 **Multi-omics integrative annotation highlights important therapeutics targets for CKD and**  
23 **HTN**

24  
25 At present, the most widely used therapies for preventing kidney disease progression are inhibitors  
26 of the renin-angiotensin-aldosterone system. About 70-80% of patients with CKD take ACEi or  
27 ARBs. These drugs also effectively reduce blood pressure, and are one of the most prescribed  
28 classes of anti-hypertensive medications taken by roughly 30% of patients with HTN worldwide.  
29 Despite their remarkable efficacy and common use over the last 3 decades, the molecular  
30 mechanism of protection has not been identified as both genes are expressed by multiple cell types  
31 and genetic studies have not elucidated why these treatments are effective.

1  
2 A recent SBP GWAS (46) identified a genetic signal on chromosome 17, rs4292, which barely  
3 passed genome wide significance ( $P\text{-value}_{\text{GWAS}}=5.88\text{E-}09$ ). Our eQTL analysis indicated that  
4 the same SNP, rs4292 was significantly associated with *ACE* expression in microdissected kidney  
5 tubule samples (**Fig. S19A**). Interestingly, we observed that the allelic dosage of C or T is markedly  
6 changed the relationship between PT fractions and *ACE* expression in human kidney proximal  
7 tubules (**Fig. 6A**). Meta-analysis of this eQTL association (rs4292 and *ACE*) (kidney  
8 compartments and 46 GTEx tissue (7), **Methods**) indicated that rs4292 had the most significant  
9 eQTL effect on *ACE* ( $M\text{-value}=1$ ) in kidney, especially in the tubule fraction (**Fig. 6B** and **Fig.**  
10 **S20**). Furthermore, the signal colocalized with PT eQTL(ci) with a high confidence (**Fig. 6C**). An  
11 Iterative Bayesian Step-wise Selection based sum of single effect model for fine mapping analysis  
12 implemented in Susie (53) indicated that across the  $\sim 2,500$  SNPs located within  $\pm 1$  Mb around the  
13 transcription start site of *ACE*, rs4292 has the highest posterior inclusion probability ( $\text{PIP}=55.5\%$ )  
14 (**Fig. S21**) for being causal. Single cell open chromatin data indicated that rs4292 was located on  
15 a cell type specific open chromatin region only in PT around the promoter region of *ACE* (**Fig.**  
16 **6D**). Single cell open chromatin accessibility analysis highlighted the strong correlation between  
17 open chromatin regions of *ACE* and rs4292 and other significant GWAS SNPs in PT cells (**Fig.**  
18 **6D** and **Fig. S22A**). Finally, using gapped *k*-mer support vector machine based methods (gkm-  
19 SVM and deltaSVM) (54, 55), we were able to computationally quantify the likely causal  
20 regulatory effect of this variant on the transcription factor binding activity (of the 50 bp region  
21 around rs4292) significantly increased, when the C allele was replaced by allele T (deltaSVM=2.16,  
22  $P\text{-value}=6.75\text{E-}07$ ) of rs4292 (**Fig. 6E**). This observation is consistent with the association  
23 between rs4292 and *ACE* gene expression, for individuals with the higher dosage of allele T at the  
24 rs4292 locus, *ACE* expression levels are significantly higher in tubules (**Fig. S19A**).

25  
26 To our surprise, we found that our analysis identified not only *ACE*, but also *AGT*, encoding the  
27 primary substrate for production of angiotensin II. The SBP GWAS (46) have identified a  
28 significant signal at rs6687360 ( $P_{\text{GWAS}(\text{rs6687360})}=4.72\text{E-}13$ ). Kidney tubule eQTL data highlighted  
29 a strong association with *AGT* expression ( $P_{\text{eQTL}}=4.83\text{E-}44$ ) (**Fig. S19B**). Allelic dosage of C at  
30 the rs6687360 variant, significantly reduced the correlation between PT fraction and *AGT*  
31 expression with  $P\text{-value}=1.22\text{E-}05$  (**Fig. 6F**). Interestingly, although *AGT* is expressed in multiple

1 tissues and cell types (34), meta-analysis of the eQTL rs6687360-*AGT* in kidney compartments  
2 studies (7) and 46 GTEx tissues (**Methods**) indicated that rs6687360 was most significantly  
3 associated with *AGT* (M-value=1) in the kidney, especially in the tubule compartment (**Fig. 6G**  
4 **and Fig. S23**). We did not observe colocalization with the genetic signal using bulk eQTL data  
5 from tubule (PP4=0.00), but found a strong colocalization signal between PT eQTL(ci) and the  
6 genetic signal, highlighting the shared genetic casual variants between the GWAS and PT eQTL(ci)  
7 (PP4=0.73) (**Fig. 6H**). We only observed a single open chromatin region of this area in the intronic  
8 region of *AGT* in PT cells (**Fig. 6I**). Single-cell open chromatin co-accessibility analysis showed  
9 an interaction between opening regions of *AGT* and rs6687360 (and other significant GWAS SNPs  
10 in PT cells (**Fig. 6I and Fig. S22B**)). In addition, the deltaSVM analysis indicated the TF binding  
11 activity of  $\pm 25$  bp region around rs6687360 increased as allele C changed to T on rs6687360 in  
12 PT cells (**Fig. 6J**). The observation was consistent with tubule eQTL effect (**Fig. S19B**), samples  
13 with the higher dosage of T at rs6687360 locus had higher *AGT* expression.

14  
15 The discovery of more than 200 prioritized genes for kidney function and hypertension allowed us  
16 to identify targets for potential repurposing of drugs. Using the Drug Gene Interaction Database  
17 (DGIdb v3.0, <https://www.dgldb.org>), we identified 54 genes that can be targeted by already FDA-  
18 approved drugs (**Table S15**). In addition, to the large number of drugs that targeted *ACE* and *AGT*  
19 with clinically approval drugs, several drugs directly target prioritized genes, such as cilastatin that  
20 inhibits the kidney dipeptidase (*DPEPI*) (**Table S15**). *DPEPI* was nominated as a kidney disease  
21 gene based on our eQTL annotation. The *DPEPI* inhibitor cilastatin showed protection in  
22 preliminary studies from acute toxic kidney injury in animal models (56). Caspase9 has also been  
23 prioritized as a CKD gene, with already available inhibitors, such as emricasan and nivocasan  
24 (**Table S15**). Several genes from Solute Carrier Family (SLC) have been identified in our study  
25 that might be particularly amenable for pharmacological targeting, such as *SLC2A4* (GLUT4),  
26 *SLC47A1* and *SLC15A2*(57), in addition to the key regulator of the Notch signaling; *APHIA*,  
27 hedgehog signaling; *SMO* and *TGFBI* (**Table S15 and S16**).

28 Overall, our combined eQTL(ci) and snATAC-based GWAS analysis had defined the genetic and  
29 mechanistic basis for ACEi and ARBs; the most commonly used anti-hypertensive and kidney

1 protective drugs, and prioritized a large number of mostly approved compounds that can be  
2 repurposed for the treatment of kidney disease and hypertension.

3

4

## 5 **Discussion**

6 Here, we generated a comprehensive multi-omics dataset and used orthogonal analytical  
7 approaches to annotate renal endophenotypes. Our analysis prioritized more than 182 likely causal  
8 genes for kidney function and 88 for hypertension, potentially defining the core genes for these  
9 conditions in the kidney. We showed the critical cell type convergence of endophenotypes, such  
10 as kidney function and metabolite heritability enrichment in PT and SBP heritability enrichment  
11 in endothelial cells and distal tubule segments. Functional annotation indicated the likely role of  
12 metabolism in kidney disease and angiogenesis in hypertension. Finally, our analysis clarified the  
13 key genetic underpinning of action for the most commonly used antihypertensive and renal  
14 protective drugs, as well as highlighted a large number of additional drugs that could potentially  
15 be effective for HTN and CKD.

16

17 While GWAS studies have been extremely powerful in disease associated loci discovery, signals  
18 at these sites have rarely been translated into molecular mechanisms and therapeutics. One of the  
19 most important computational methods used to annotate GWAS signals is eQTL analysis. Limited  
20 availability of kidney eQTL datasets has been an important limitation of renal related traits GWAS  
21 annotation. With the increased sample size provided in this study, we were able to double the  
22 number of identified eQTLs.

23

24 Another key limitation of the eQTL studies has been the heterogenous cell types in the bulk tissue  
25 samples. We found that while PEER factors have been able to account for cell fraction  
26 heterogeneity, additional adjustment using computational deconvolution could further improve the  
27 model. We took this a step further and identified eQTL(ci) that define genotype-cell interaction  
28 (cell-type) effects. We cataloged SNP-gene pairs by examining the interaction of cell fraction and  
29 genetic variations in a large number of kidney compartment samples. We also noted some  
30 limitations of the eQTL(ci) approach. First, the approach is limited by the quality of the single cell  
31 gene expression data. While human kidney single cell RNA-seq data are now available, we found

1 that the higher quality mice dataset generated more reliable information on cell fractions (16, 58).  
2 Second, the co-linearity of cell fractions in tissue samples, such as the loss of kidney parenchyma,  
3 is associated with a relative increase in proportion of mesenchymal and immune cells, and so these  
4 methods are unable to distinguish between such closely related cell fractions. Finally, the analysis  
5 appeared to be more robust for the abundant cell types of each tissue. While eQTL(ci) is a new  
6 and exciting method, the model has not been experimentally validated. Here, we used snATAC-  
7 seq data to understand whether eQTL(ci) could be linked to cell-type specific open chromatin  
8 regions. Our data supports the results from eQTL(ci)s as they showed enrichment for cell type  
9 specific regulatory regions.

10

11 Our combined analyses that used LDSC (for scRNA-seq and snATAC-seq data), MAGMA and  
12 eQTL(ci) coloc have provided several meaningful insights into driver cell types for kidney disease-  
13 associated endophenotypes. We demonstrated that eGFR heritability is primarily derived from  
14 eQTL influence in PT cells where metabolite levels, such as serum uric acid, are also regulated by  
15 PT functions. In contrast, SBP was associated with genes expressed specifically in endothelial  
16 cells and DCT cells. The cell type enrichment of endophenotypes appears to be robust and highly  
17 consistent using different methods and datasets.

18

19 Finally, our results provide insights into previously cryptic disease mechanisms and therapeutic  
20 targeting opportunities. Inhibitors of the renin-angiotensin-aldosterone system, such as the ACEi  
21 and ARBs, are the most commonly used medication class for SBP and preventing kidney disease  
22 progression. Yet, we lacked human genetic data supporting the therapeutic benefits of ACEi and  
23 ARB in CKD or hypertension. The *ACE* locus passed genome wide significance in a large SBP  
24 GWAS and our kidney eQTL annotation provided critical information for the annotation of this  
25 locus and highlighted its association with *ACE* expression in PTs. Our orthogonal analyses  
26 including snATAC and deltaSVM nominated rs4292 for this effect.

27

28 It is also important to note that GWAS variants at the *AGT* locus were robustly associated with  
29 SBP, but the associations of genetic variants with gene expression using bulk eQTL datasets were  
30 not sufficient to prioritize the causal gene for this locus. It was really the interaction effect between  
31 genetic variants and PT cell fraction variations identified from PT eQTL(ci) and snATAC data that



1 had helped with the target gene and cell type prioritization. Mouse model experiments are  
2 consistent with our results, notably, a linear correlation between *AGT* levels and blood pressure  
3 has been previously established in mouse models (59). *AGT* made in PT is shed into the tubular  
4 fluid (60), where it co-localizes with *ACE*, which is highly expressed on the brush border on the  
5 tubular surface of proximal epithelia. Overall, our study highlights the critical role of the kidney  
6 in blood pressure regulation and raises the possibility that these variants might be useful for future  
7 precision therapeutic approaches, such as to understand differences in response to ACEi and ARB.

8

9 In summary, we present an integrated multi-omics study for functional interpretation of common  
10 variants associated with renal traits using kidney compartment-based eQTL and single-cell  
11 resolution open chromatin identification. The application of multiple orthogonal computational  
12 methods was critical to prioritize more than 200 genes for eGFR and SBP, highlighted the key  
13 disease-associated variants, cell types and disease pathomechanism, clarified mechanism of widely  
14 used therapeutics and highlighted new potential drugs for these kidney related diseases.

15

16

## 1 References and Notes

- 2 1. K. J. Jager *et al.* (Oxford University Press, 2019).
- 3 2. R. Z. Alicic, M. T. Rooney, K. R. Tuttle, Diabetic Kidney Disease Challenges, Progress,  
4 and Possibilities. *Clinical Journal of the American Society of Nephrology* **12**, 2032-2045  
5 (2017).
- 6 3. A. C. Webster, E. V. Nagler, R. L. Morton, P. Masson, Chronic kidney disease. *The lancet*  
7 **389**, 1238-1252 (2017).
- 8 4. M. R. Nelson *et al.*, The support of human genetic evidence for approved drug indications.  
9 *Nature genetics* **47**, 856-860 (2015).
- 10 5. M. Wuttke *et al.*, A catalog of genetic loci associated with kidney function from analyses  
11 of a million individuals. *Nature genetics* **51**, 957 (2019).
- 12 6. G. Consortium, The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene  
13 regulation in humans. *Science* **348**, 648-660 (2015).
- 14 7. C. Qiu *et al.*, Renal compartment-specific genetic variation analyses identify new  
15 pathways in chronic kidney disease. *Nature medicine* **24**, 1721 (2018).
- 16 8. S. Kim-Hellmuth *et al.*, Cell type-specific genetic regulation of gene expression across  
17 human tissues. *Science* **369**, (2020).
- 18 9. B. K. Bulik-Sullivan *et al.*, LD Score regression distinguishes confounding from  
19 polygenicity in genome-wide association studies. *Nature genetics* **47**, 291-295 (2015).
- 20 10. C. Giambartolomei *et al.*, A Bayesian framework for multiple trait colocalization from  
21 summary association statistics. *Bioinformatics* **34**, 2538-2545 (2018).
- 22 11. A. Gusev *et al.*, A transcriptome-wide association study of high-grade serous epithelial  
23 ovarian cancer identifies new susceptibility genes and splice variants. *Nature genetics* **51**,  
24 815-823 (2019).
- 25 12. M. Kato, R. Natarajan, Diabetic nephropathy—emerging epigenetic mechanisms. *Nature*  
26 *Reviews Nephrology* **10**, 517 (2014).
- 27 13. S. Tonna, A. El-Osta, M. E. Cooper, C. Tikellis, Metabolic memory and diabetic  
28 nephropathy: potential role for epigenetic mechanisms. *Nature Reviews Nephrology* **6**, 332  
29 (2010).
- 30 14. Y. Sun, N. Miao, T. Sun, Detect accessible chromatin using ATAC-sequencing, from  
31 principle to applications. *Hereditas* **156**, 1-9 (2019).
- 32 15. N. D. Heintzman *et al.*, Distinct and predictive chromatin signatures of transcriptional  
33 promoters and enhancers in the human genome. *Nature genetics* **39**, 311-318 (2007).
- 34 16. J. Park *et al.*, Single-cell transcriptomics of the mouse kidney reveals potential cellular  
35 targets of kidney disease. *Science* **360**, 758-763 (2018).
- 36 17. A. Ransick *et al.*, Single-cell profiling reveals sex, lineage, and regional diversity in the  
37 mouse kidney. *Developmental cell* **51**, 399-413. e397 (2019).
- 38 18. R. E. Gate *et al.*, Genetic determinants of co-accessible chromatin regions in activated T  
39 cells across humans. *Nature genetics* **50**, 1140-1150 (2018).
- 40 19. S. Kasela *et al.*, Pathogenic implications for autoimmune mechanisms derived by  
41 comparative eQTL analysis of CD4+ versus CD8+ T cells. *PLoS genetics* **13**, e1006643  
42 (2017).
- 43 20. K. Ishigaki *et al.*, Polygenic burdens on cell-specific pathways underlie the risk of  
44 rheumatoid arthritis. *Nature genetics* **49**, 1120 (2017).
- 45 21. S. Huang, X. Sheng, K. Susztak, The kidney transcriptome, from single cells to whole  
46 organs and back. *Current opinion in nephrology and hypertension* **28**, 219-226 (2019).

- 1 22. A. M. Newman *et al.*, Robust enumeration of cell subsets from tissue expression profiles.  
2 *Nature methods* **12**, 453 (2015).
- 3 23. X. Wang, J. Park, K. Susztak, N. R. Zhang, M. Li, Bulk tissue cell type deconvolution with  
4 multi-subject single-cell expression reference. *Nature communications* **10**, 1-9 (2019).
- 5 24. D. Aran, Z. Hu, A. J. Butte, xCell: digitally portraying the tissue cellular heterogeneity  
6 landscape. *Genome biology* **18**, 220 (2017).
- 7 25. A. M. Newman *et al.*, Determining cell type abundance and expression from bulk tissues  
8 with digital cytometry. *Nature biotechnology* **37**, 773-782 (2019).
- 9 26. K. M. Sullivan, K. Susztak, Unravelling the complex genetics of common kidney diseases:  
10 from variants to mechanisms. *Nature Reviews Nephrology*, 1-13 (2020).
- 11 27. J. Chiou *et al.*, Single cell chromatin accessibility reveals pancreatic islet cell type-and  
12 state-specific regulatory programs of diabetes risk. *BioRxiv*, 693671 (2019).
- 13 28. M. R. Corces *et al.*, Single-cell epigenomic identification of inherited risk loci in  
14 Alzheimer's and Parkinson's disease. *bioRxiv*, (2020).
- 15 29. J. Park *et al.*, Renal proximal tubule cell state and metabolism are coupled by nuclear  
16 receptors. *bioRxiv*, (2020).
- 17 30. Y.-A. Ko *et al.*, Genetic-variation-driven gene-expression changes highlight genes with  
18 important functions for kidney disease. *The American Journal of Human Genetics* **100**,  
19 940-953 (2017).
- 20 31. J. D. Storey, A direct approach to false discovery rates. *Journal of the Royal Statistical*  
21 *Society: Series B (Statistical Methodology)* **64**, 479-498 (2002).
- 22 32. C. A. Glastonbury, A. C. Alves, J. S. E.-S. Moustafa, K. S. Small, Cell-type heterogeneity  
23 in adipose tissue is associated with complex traits and reveals disease-relevant cell-specific  
24 eQTLs. *The American Journal of Human Genetics* **104**, 1013-1024 (2019).
- 25 33. J. D. Storey, R. Tibshirani, Statistical significance for genomewide studies. *Proceedings of*  
26 *the National Academy of Sciences* **100**, 9440-9445 (2003).
- 27 34. G. Consortium, Genetic effects on gene expression across human tissues. *Nature* **550**, 204-  
28 213 (2017).
- 29 35. S. M. Urbut, G. Wang, P. Carbonetto, M. Stephens, Flexible statistical methods for  
30 estimating and testing effects in genomic studies with multiple conditions. *Nature genetics*  
31 **51**, 187-195 (2019).
- 32 36. M. Stephens, False discovery rates: a new deal. *Biostatistics* **18**, 275-294 (2017).
- 33 37. X. Zhang *et al.*, CellMarker: a manually curated resource of cell markers in human and  
34 mouse. *Nucleic acids research* **47**, D721-D728 (2019).
- 35 38. A. N. Schep, B. Wu, J. D. Buenrostro, W. J. Greenleaf, chromVAR: inferring transcription-  
36 factor-associated accessibility from single-cell epigenomic data. *Nature methods* **14**, 975-  
37 978 (2017).
- 38 39. S. Heinz *et al.*, Simple combinations of lineage-determining transcription factors prime  
39 cis-regulatory elements required for macrophage and B cell identities. *Molecular cell* **38**,  
40 576-589 (2010).
- 41 40. Z. Miao *et al.*, Single cell resolution regulatory landscape of the mouse kidney highlights  
42 cellular differentiation programs and renal disease targets. *bioRxiv*, (2020).
- 43 41. E. M. Schmidt *et al.*, GREGOR: evaluating global enrichment of trait-associated variants  
44 in epigenomic features using a systematic, data-driven approach. *Bioinformatics* **31**, 2601-  
45 2606 (2015).

- 1 42. H. K. Finucane *et al.*, Heritability enrichment of specifically expressed genes identifies  
2 disease-relevant tissues and cell types. *Nature genetics* **50**, 621-629 (2018).
- 3 43. C. A. de Leeuw, J. M. Mooij, T. Heskes, D. Posthuma, MAGMA: generalized gene-set  
4 analysis of GWAS data. *PLoS Comput Biol* **11**, e1004219 (2015).
- 5 44. J. N. Hellwege *et al.*, Mapping eGFR loci to the renal transcriptome and phenome in the  
6 VA Million Veteran Program. *Nat Commun* **10**, 3842 (2019).
- 7 45. A. P. Morris *et al.*, Trans-ethnic kidney function association study reveals putative causal  
8 genes and effects on kidney-specific disease aetiologies. *Nature communications* **10**,  
9 (2019).
- 10 46. A. Giri *et al.*, Trans-ethnic association study of blood pressure determinants in over  
11 750,000 individuals. *Nature genetics* **51**, 51-62 (2019).
- 12 47. A. Tin *et al.*, Target genes, variants, tissues and transcriptional pathways influencing  
13 human serum urate levels. *Nature genetics*, 1-16 (2019).
- 14 48. A. Teumer *et al.*, Genome-wide association meta-analyses and fine-mapping elucidate  
15 pathways influencing albuminuria. *Nature communications* **10**, 1-19 (2019).
- 16 49. R. Pazoki *et al.*, GWAS for urinary sodium and potassium excretion highlights pathways  
17 shared with cardiovascular traits. *Nature communications* **10**, 1-11 (2019).
- 18 50. I. M. S. G. Consortium\*†, ANZgene, IIBDGC, WTCCC2, Multiple sclerosis genomic map  
19 implicates peripheral immune cells and microglia in susceptibility. *Science* **365**, eaav7188  
20 (2019).
- 21 51. Y. Li *et al.*, Integration of GWAS Summary Statistics and Gene Expression Reveals Target  
22 Cell Types Underlying Kidney Function Traits. *Journal of the American Society of*  
23 *Nephrology*, (2020).
- 24 52. C. Giambartolomei *et al.*, Bayesian test for colocalisation between pairs of genetic  
25 association studies using summary statistics. *PLoS genetics* **10**, e1004383 (2014).
- 26 53. G. Wang, A. K. Sarkar, P. Carbonetto, M. Stephens, A simple new approach to variable  
27 selection in regression, with application to genetic fine-mapping. *bioRxiv*, 501114 (2019).
- 28 54. M. Ghandi *et al.*, gkmSVM: an R package for gapped-kmer SVM. *Bioinformatics* **32**,  
29 2205-2207 (2016).
- 30 55. D. Lee, LS-GKM: a new gkm-SVM for large-scale datasets. *Bioinformatics* **32**, 2196-2198  
31 (2016).
- 32 56. A. Lau *et al.*, Renal immune surveillance and dipeptidase-1 contribute to contrast-induced  
33 acute kidney injury. *The Journal of clinical investigation* **128**, 2894-2913 (2018).
- 34 57. L. Lin, S. W. Yee, R. B. Kim, K. M. Giacomini, SLC transporters as therapeutic targets:  
35 emerging opportunities. *Nature reviews Drug discovery* **14**, 543-560 (2015).
- 36 58. M. D. Young *et al.*, Single-cell transcriptomes from human kidneys reveal the cellular  
37 identity of renal tumors. *Science* **361**, 594-599 (2018).
- 38 59. H.-S. Kim *et al.*, Genetic control of blood pressure and the angiotensinogen locus.  
39 *Proceedings of the National Academy of Sciences* **92**, 2735-2739 (1995).
- 40 60. H. Kobori, L. M. Harrison-Bernard, L. G. Navar, Urinary excretion of angiotensinogen  
41 reflects intrarenal angiotensinogen production. *Kidney international* **61**, 579-585 (2002).
- 42 61. B. Han, E. Eskin, Interpreting meta-analyses of genome-wide association studies. *PLoS*  
43 *Genet* **8**, e1002555 (2012).
- 44
- 45

1 **Acknowledgements**

2

3 **Funding**

4

5 This work in the Susztak lab has been supported by the National Institute of Health NIH R01  
6 DK105821, R01 DK087635, R01 DK076077 and by the Foundation of the NIH Type 2 diabetes  
7 Accelerated Medicine Partnership Project. The authors thank the Molecular Pathology and  
8 Imaging Core (P30-DK050306) and Diabetes Research Center (P30-DK19525) at the University  
9 of Pennsylvania for their services.

10

11 **Author contributions**

12

13 KS and XS conceived, planned and oversaw the study, and wrote the manuscript. ZMa and JW  
14 conducted the human kidney snATAC-seq experiment. XS analyzed data with the help of HL, CQ,  
15 ZMi. MJS, MP, MKS, KD, SSP, TLE, JNH, AMH, ML, BV, TC, CDB, KS assisted with data  
16 generation and manuscript revision.

17

18 **Competing interests**

19

20 The laboratory of Dr. Susztak receives funding from GSK, Regeneron, Gilead, Merck Sharp &  
21 Dohme Corp., a subsidiary of Merck & Co., Inc., Kenilworth, NJ, USA, Boehringer Ingelheim,  
22 Bayer and Novo Nordisk. The funders had no influence on the data analysis. Dr. Susztak serves  
23 on the SAB of Jnana pharmaceuticals.

24

25 **Data and materials availability**

26 The RNA-seq, eQTL data and human kidney snATAC-seq data is publicly available at the diabetes  
27 epigenome atlas [www.diabetesepigenome.org](http://www.diabetesepigenome.org). No consent was obtained to share individual-level  
28 genotype data.

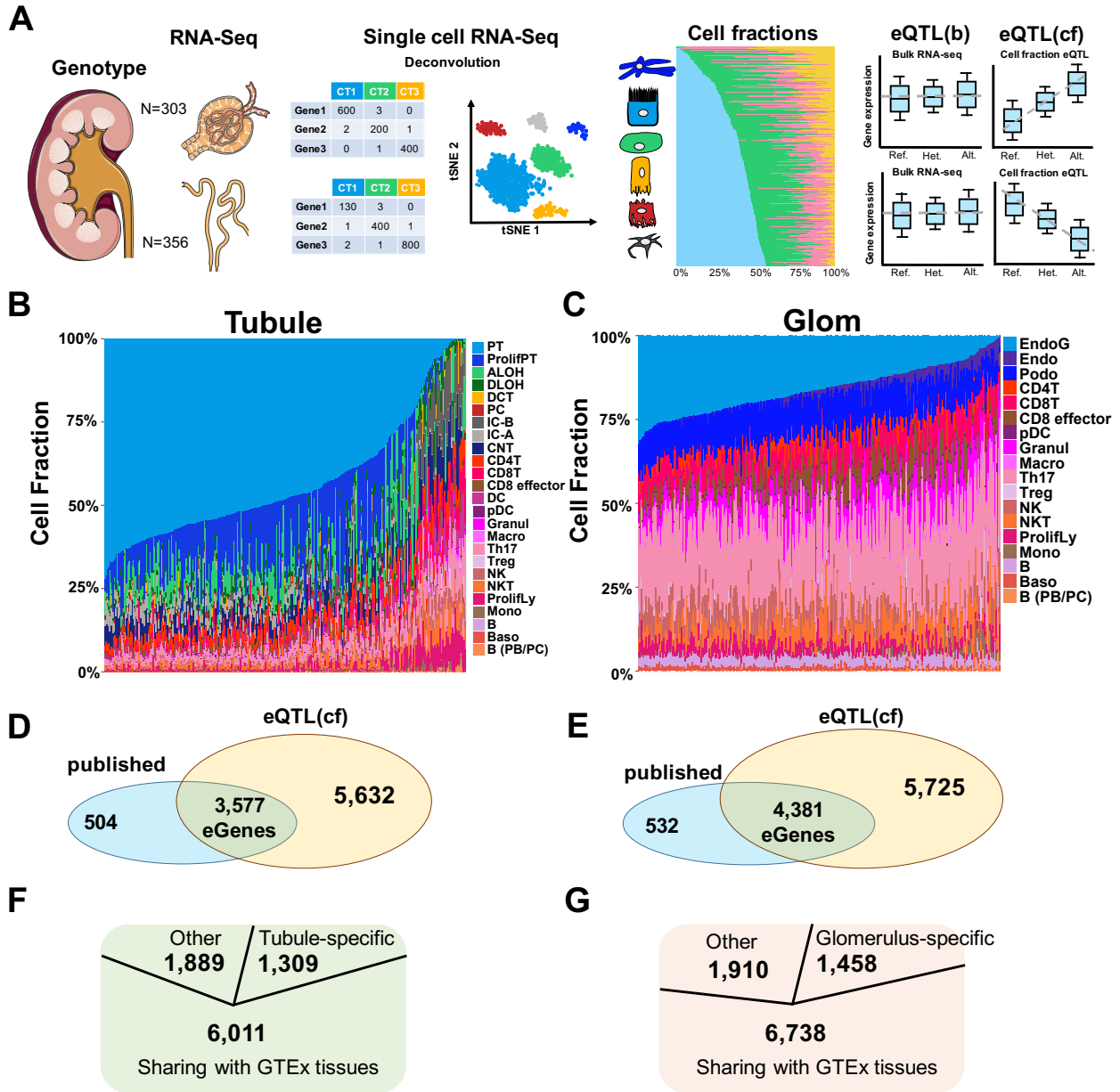
29

30

31

- 1 **Supplementary Materials**
- 2 Materials and Methods
- 3 Figs. S1 to S28
- 4 Tables S1 to S16
- 5 References (1-64)
- 6

1 **Figure legends**



2

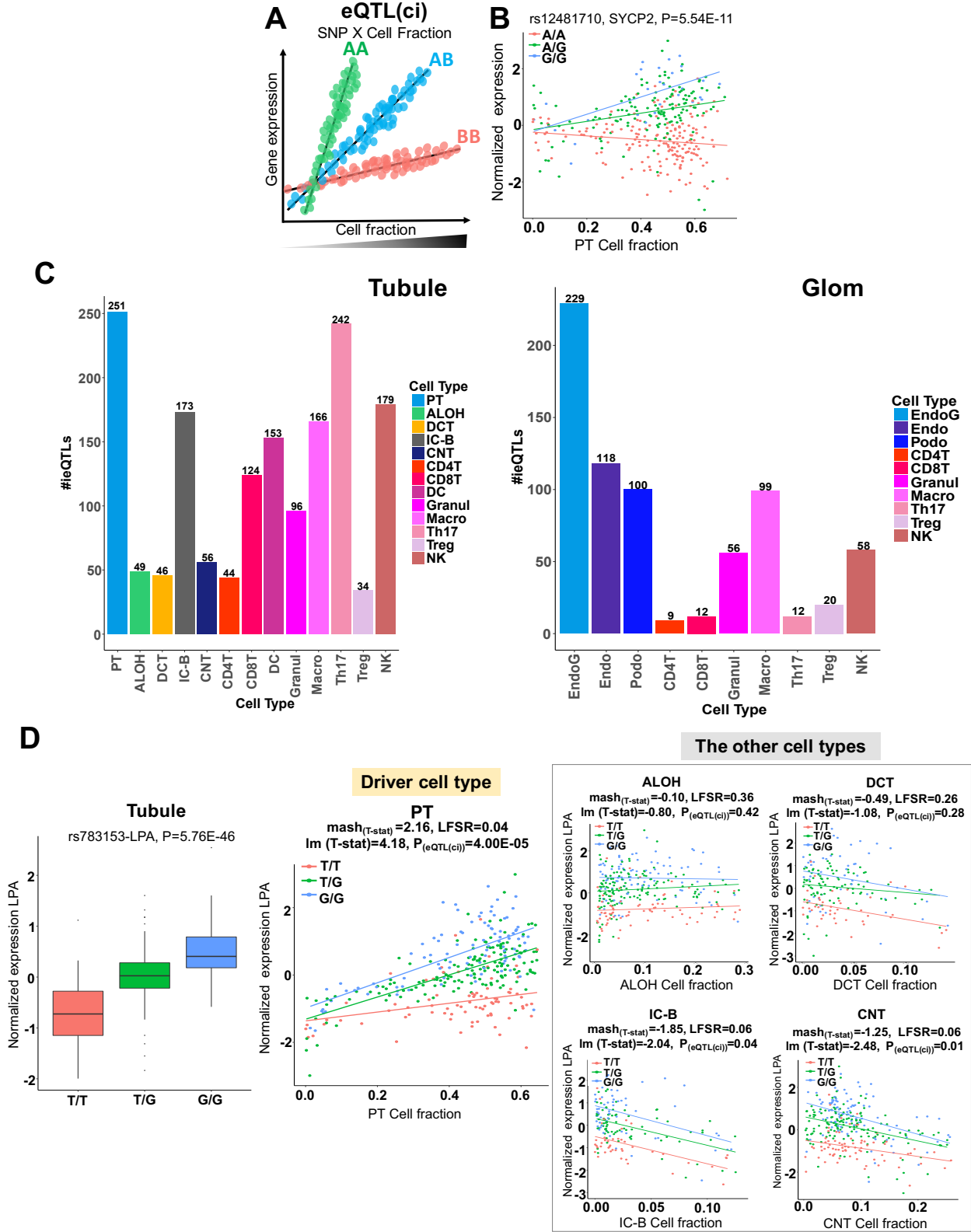
3 **Figure 1. Cell fraction adjusted expression quantitative trait analysis of microdissected**  
 4 **human kidney samples**

5 A. Experimental scheme. We collected high quality genotype and RNA-seq data for 303  
 6 human microdissected kidney glomeruli and 359 tubule samples. Cell fractions were  
 7 estimated by deconvolution method using scRNA-seq data (29) as the reference. eQTLs  
 8 were then identified using bulk and cell fraction adjusted models.

9 B. The distribution of the relative cell fractions across 359 microdissected human kidney

- 1 tubule samples. PT: proximal tubule, ProlifPT: proliferating PT cells, ALOH: ascending  
2 loop of Henle, DLOH: descending loop of Henle, DCT: distal convoluted tubule, PC:  
3 collecting duct principal cells, IC-B: beta intercalated cells, IC-A: alpha intercalated cells,  
4 CNT: connecting tubule, CD4T: CD4 T cells, CD8T: CD8 T cells, CD8 effector: CD8  
5 effector cells, DC: CD11b+ dendritic cells, pDC: plasmacytoid DC, Granul: granulocyte,  
6 Macro: macrophage, Th17: T helper 17 cells, Treg: regulatory T cells, NK: natural killer  
7 cells, NKT: natural killer T cells, ProlifLy: proliferating lymphocyte cells, Mono:  
8 monocyte, B: B lymphocyte, Baso: Basophil, B (PB/PC): memory B cells from the  
9 peripheral blood or plasma cells.
- 10 C. The distribution of the relative cell fractions across 303 microdissected human kidney  
11 glomerular samples. EndoG: glomerular endothelial cells, Endo: endothelial cells, Podo:  
12 podocyte, CD4T: CD4 T cells, CD8T: CD8 T cells, CD 8effector: CD8 effector cells, pDC:  
13 plasmacytoid DC, Granul: granulocyte, Macro: macrophage, Th17: T helper 17 cells, Treg:  
14 regulatory T cells, NK: natural killer cells, NKT: natural killer T cells, ProlifLy:  
15 proliferating lymphocyte cells, Mono: monocyte, B: B lymphocyte, Baso: Basophil, B  
16 (PB/PC): memory B cells from the peripheral blood or plasma cells.
- 17 D. Venn diagram illustrating the direct overlap of previously published (7) and newly  
18 identified significant eGenes in microdissected human kidney tubules
- 19 E. Venn diagram illustrating the direct overlap of previously published (7) and newly  
20 identified significant eGenes in microdissected human kidney glomeruli
- 21 F. Diagram illustrating the number of tubule-specific eGenes identified by meta-analysis of  
22 eQTLs from 49 human tissues (48 GTEx v7 (34) tissues plus tubule).
- 23 G. Diagram illustrating the number of glomerulus-specific eGenes identified by meta-analysis  
24 of eQTLs from 49 human tissues (48 GTEx v7 (34) tissues plus glomerulus).
- 25





1

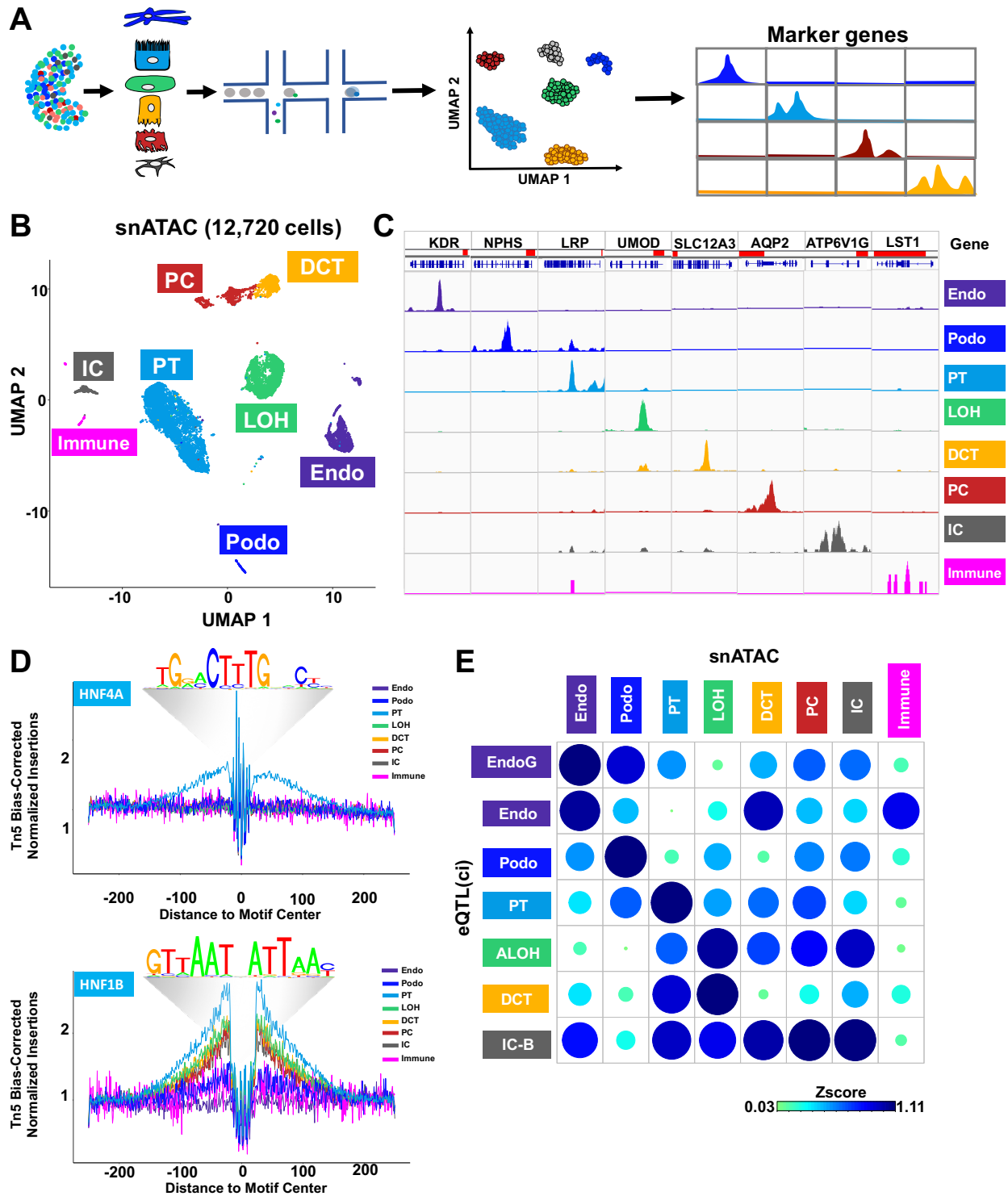
2 **Figure 2. Defining cell-type dependent activities of genetic variants on gene expression**

1 A. Scheme of cell type interaction eQTL (eQTL(ci)). For a specific SNP, the correlation between  
2 gene expression variations and cell fraction changes are significantly varying across individuals  
3 with different genotypes on this locus.

4 B. The y-axis is the relative expression of *SYCP2* in human kidney tubules (N=356 samples), the  
5 x-axis is the PT cell fraction colored by subject (rs12481710) genotype A/A (red), A/G (green)  
6 and G/G (blue). P value was calculated by eQTL(ci) model. Each data point represents a single  
7 subject. The y-axis is the normalized expression of *SYCP2* in human tubules (N=356 samples),  
8 the x-axis is the cell fraction colored by subject genotype A/A (red), A/G (green) and G/G (blue)  
9 at the rs12481710 locus.

10 C. The number of identified eQTL(ci)s in each cell types in human kidney tubules (left panel)  
11 and glomeruli (right panel). EndoG: glomerular endothelial cells, Endo: endothelial cells, Podo:  
12 podocyte, PT: proximal tubule, ALOH: ascending loop of Henle, DCT: distal convoluted tubule,  
13 IC-B: beta intercalated cells, CNT: connecting tubule, CD4T: CD4 T cells, CD8T: CD8 T cells,  
14 DC: CD11b+ dendritic cells, Granul: granulocyte, Macro: macrophage, Th17: T helper 17 cells,  
15 Treg: regulatory T cells, NK: natural killer cells.

16 D. A PT specific eQTL(ci) for *LPA*. The leftmost panel illustrates the association between  
17 genotype of SNP rs783153 and gene expression of *LPA* in bulk tubule samples (N=356 samples).  
18 The y-axis is the normalized expression of *LPA* in human kidney tubules, the x-axis is the  
19 genotype at rs783153 locus. Center lines show the medians; box limits indicate the 25<sup>th</sup> and 75<sup>th</sup>  
20 percentiles; whiskers extend to the 5<sup>th</sup> and 95<sup>th</sup> percentiles; outliers are represented by dots. P  
21 was calculated by linear regression eQTL(cf) model. The following panels show PT cell-type-  
22 specificity of this eQTL(ci) across 5 different kidney function cell types in human tubules  
23 (N=356 samples). The y-axis is the normalized expression of *LPA* in human tubules (N=356  
24 samples), the x-axis is the cell fraction colored by subject genotype T/T (red), T/G (green) and  
25 G/G (blue) at the rs783153 locus. lm (T-stat): T-statistics calculated from linear model of  
26 eQTL(ci), mash<sub>(T-stat)</sub>: T-statistics estimated by *mash*. PT: proximal tubule, ALOH: ascending  
27 loop of Henle, DCT: distal convoluted tubule, IC-B: beta intercalated cells, CNT: connecting  
28 tubule. Each data point represents one sample.



1

2 **Figure 3. Single cell resolution regulatory maps for the human kidney**

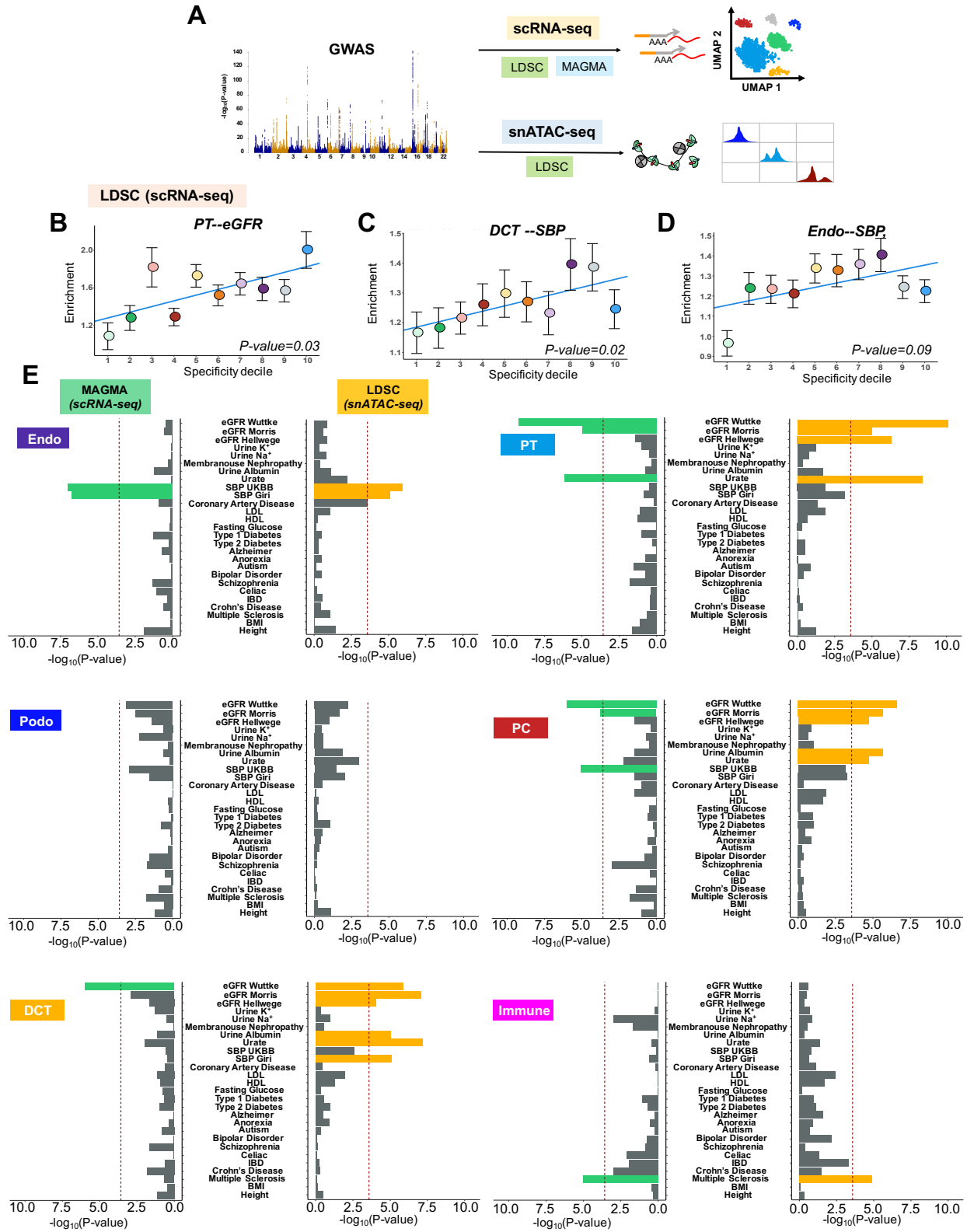
3

4 A. Experimental scheme of snATAC-seq. Single nuclei accessible chromatin information was  
 5 generated for 12,720 human kidney cells, using the 10X Genomics snATAC-seq platform.

5

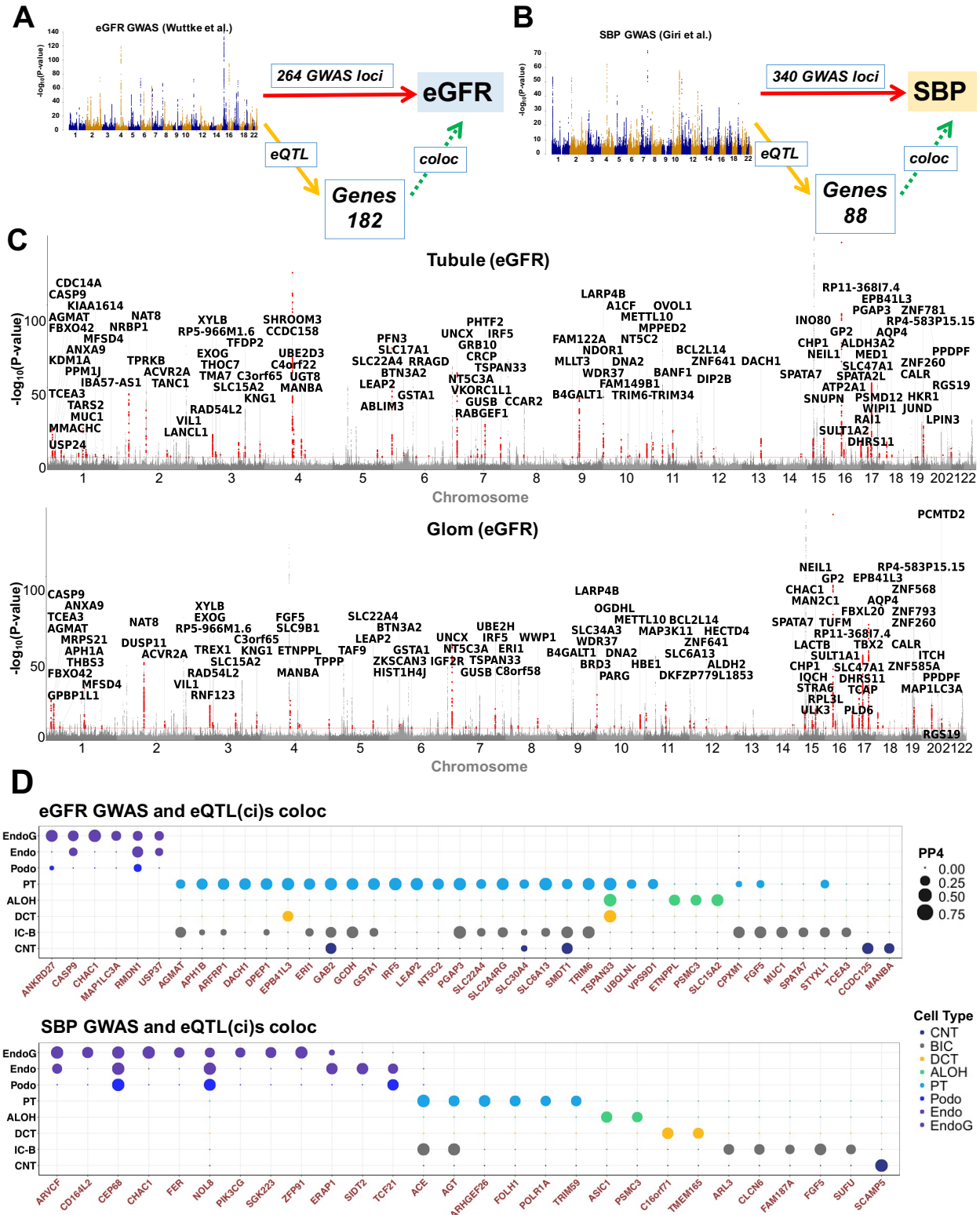
Kidney cell clusters were then identified and open chromatin peaks were called for each

- 1 cell cluster.
- 2 B. The UMAP of human kidney snATAC-seq data. 8 distinct cell clusters were identified  
3 using previously defined cell type marker genes (16). Endo: endothelial cells, Podo:  
4 podocyte, PT: proximal tubule, LOH: loop of Henle, DCT: distal convoluted tubule, PC:  
5 collecting duct principal cells, IC: collecting duct intercalated cells, Immune: immune cells.
- 6 C. Genome browser plots showing aggregate read density for cells within each cell cluster in  
7 the promoter region of each cell type marker gene: *KDR* (endothelial cells), *NPHS2*  
8 (podocyte), *LRP2* (proximal tubule), *UMOD* (LOH), *SLC12A3* (distal convoluted tubule),  
9 *AQP2* (collecting duct principal cells), *ATP6V1G3* (collecting duct intercalated cells) and  
10 *LST1* (immune cells). The promoter region for each gene is highlighted in red color.
- 11 D. Footprinting analysis of the HNF4A (top panel) and HNF1B (bottom panel) transcription  
12 factors across 8 major cell types. The motif logos are shown above.
- 13 E. Bubble plots of enrichment of eQTL(ci)s for each cell type (row) and open human kidney  
14 snATAC-seq chromatin regions of each cell cluster (Column). The size and the color of the  
15 bubble indicates enrichment (Z-score). EndoG: glomerular endothelial cells, Endo:  
16 endothelial cells, Podo: podocyte, PT: proximal tubule, LOH: loop of Henle, ALOH:  
17 ascending loop of Henle, DCT: distal convoluted tubule, PC: collecting duct principal cells,  
18 IC: collecting duct intercalated cells, IC-B: beta intercalated cells, Immune: immune cells.



1  
 2 **Figure 4. Single cell annotation highlights cell-type convergence of kidney relevant**  
 3 **endophenotypes**

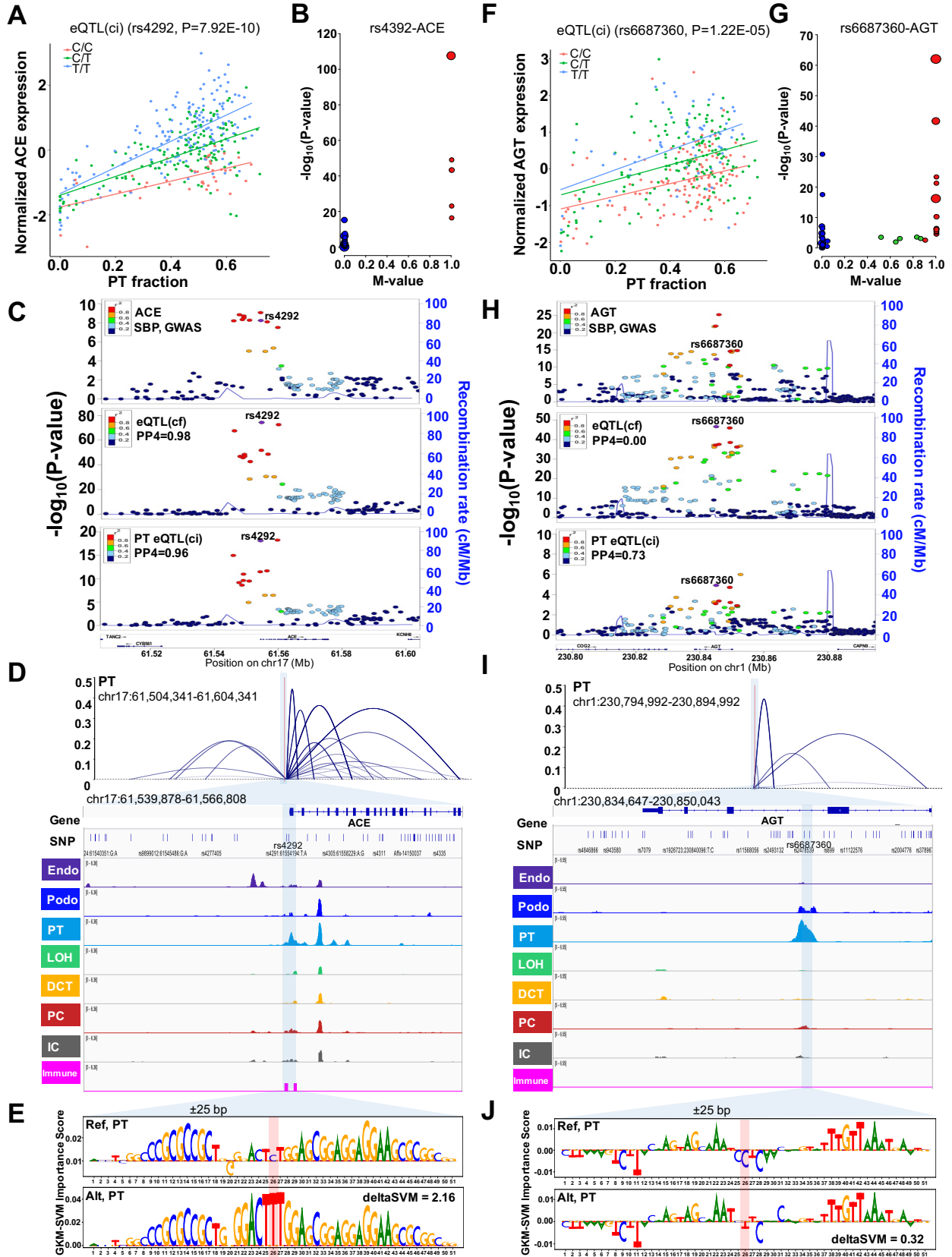
- 1 A. Experimental scheme of cell type-specific GWAS trait heritability enrichment analysis.  
2 Here we applied MAGMA to the scRNA-seq data and LDSC to the scRNA-seq and  
3 snATAC-seq data to assess the GWAS per-SNP heritability enrichment.
- 4 B. Enrichment of eGFR-SNP heritability in each of the specificity deciles for proximal tubule  
5 cells calculated using LDSC from kidney scRNA-seq data (16).
- 6 C. Enrichment of SBP-SNP heritability in each of the specificity deciles in distal convoluted  
7 tubule (DCT) cells.
- 8 D. Enrichment of SBP-SNP heritability in each of the specific deciles for endothelial cells. X-  
9 axis is the gene expression specificity decile, Y-axis is the enrichment value calculated by  
10 LDSC. Colors of dots represent different specificity deciles. Blue line shows the linear  
11 regression slope fitted to the enrichment values. Error bars indicate the 95% confidence  
12 intervals.
- 13 E. Evaluation of enrichment of common GWAS variants in each kidney cell type scRNA-seq  
14 data (16) using MAGMA (the left panels) and human kidney snATAC-seq data using LDSC  
15 (the right panels). X-axis represents 24 different GWAS traits and y-axis is the  $-\log_{10}(P$ -  
16 value). P-value was estimated by MAGMA (the left panels) and LDSC (the right panels).  
17 Endo: endothelial cells, PT: proximal tubule, Podo: podocyte, PC: collecting duct principal  
18 cells, DCT: distal convoluted tubule, Immune: T lymphocyte.  
19



1  
 2 **Figure 5. Comprehensive gene prioritization provides new mechanistic insights into kidney**  
 3 **function and blood pressure regulation**  
 4 A. Experimental scheme. We used Bayesian colocalization which combined information from

- 1 eQTL(cf)s to annotate 264 eGFR associated loci (5) and prioritized 182 causal genes for  
2 kidney disease, where the causal variants for gene expression and kidney function are  
3 shared.
- 4 B. Experimental scheme. We used Bayesian colocalization which combined information from  
5 eQTL(cf)s to annotate 340 SBP GWAS associated loci (46) and prioritized 88 causal genes  
6 for hypertension, where the causal variants for gene expression and HTN are shared.
- 7 C. Miami plot showing coloc (10) prioritized genes for eGFR GWAS in tubules and glomeruli  
8 by eQTL(cf)s. Red dots define prioritized protein coding genes. X-axis represents the  
9 chromosome. Y-axis represents the GWAS significance  $-\log_{10}(\text{P-value})$ .
- 10 D. Bubble plots showing coloc (10) prioritized cell type specific (eQTL(ci)) genes for eGFR  
11 and SBP GWAS loci. The size of the dot correlated with the posterior possibility of  
12 colocalization (PP4). EndoG: glomerular endothelial cells, Endo: endothelial cells, Podo:  
13 podocyte, PT: proximal tubule, ALOH: ascending loop of Henle, DCT: distal convoluted  
14 tubule, IC-B: beta intercalated cells, CNT: connecting tubule.  
15





1

2 **Figure 6. Multi-omics integrative annotation highlights important therapeutic targets for**

## 1 **CKD and HTN**

- 2 A. The y-axis is the normalized expression of *ACE* in human kidney tubules (N=356 samples),  
3 the x-axis is the PT cell fraction, each dot represents a single sample colored by their  
4 genotype C/C (red), C/T (green) and T/T (blue) at rs4292 locus. P was calculated by  
5 eQTL(ci) model.
- 6 B. eQTL meta-analysis showing the association between rs4292 and *ACE* in kidney  
7 compartments and of 46 human tissues from GTEx (v7). Each dot represents one tissue,  
8 and the size represents the M-value. Red dots: M-value $\geq$ 0.9, blue dots: M-value $\leq$ 0.1, green  
9 dots: 0.1<M-value<0.9. The y axis shows the meta-analysis P of association in each single  
10 tissue. The x axis shows the M-value; the posterior probability of the effect in each tissue  
11 estimated by METASOFT (61)
- 12 C. Locuszoom plots of SBP GWAS variants (top) (46) , kidney tubule *ACE* eQTL(cf)s, and  
13 PT *ACE* eQTL(ci)s. The x-axis shows the genomic region  $\pm$ 100 kb around rs4292. Each  
14 data point represents a variant and the color represents the  $r^2$  (the degree of linkage  
15 disequilibrium).
- 16 D. Top: PT cell Cicero-inferred co-accessibility between pairs of accessible chromatin sites at  
17 the *ACE* promoter. Each line is an SBP GWAS (46) significant SNP (P<5E-08). Single cell  
18 chromatin accessibility of human kidney single nuclear ATAC-seq data. Endo: endothelial  
19 cells, Podo: podocyte, PT: proximal tubule, LOH: loop of Henle, DCT: distal convoluted  
20 tubule, PC: collecting duct principal cells, IC: collecting duct intercalated cells, Immune:  
21 immune cells.
- 22 E. Gapped *k*-mer support vector machine (GKM-SVM) method (54) estimated importance  
23 score for each base in the  $\pm$ 25 bp region surrounding rs4292 for the Ref (C) and Alt (T)  
24 alleles from the gkm-SVM model corresponding to PT cell cluster. The SNP of interest is  
25 highlighted in red color. The deltaSVM value was calculated to represent the changes of  
26 the transcription factor binding activity of this 51 bp sequence after replacing allele C to  
27 allele T of rs4292.
- 28 F. The y-axis is the normalized expression of *AGT* in human kidney tubules (N=356 samples),  
29 the x-axis is the PT cell fraction, each dot represents a single sample colored by their  
30 genotype C/C (red), C/T (green) and T/T (blue) at rs6687360 locus. P was calculated by  
31 eQTL(ci) model.

- 1 G. Results of eQTL from meta-analysis showing the association between rs6687360 and *AGT*  
2 using eQTLs of kidney compartments and of 46 human tissues from GTEx (v7). Each dot  
3 represents one tissue, and the size represents the M-value. Red dots:  $M\text{-value} \geq 0.9$ , blue  
4 dots:  $M\text{-value} \leq 0.1$ , green dots:  $0.1 < M\text{-value} < 0.9$ . The y-axis shows the meta-analysis P of  
5 association in each single tissue. The x-axis shows the M-value; the posterior probability  
6 of the effect in each tissue estimated by METASOFT (61).
- 7 H. Locuszoom plots of SBP GWAS variants (46), kidney tubule eQTL(cf)s, and PT *AGT*  
8 eQTL(ci)s. The x-axis shows the genomic region  $\pm 100$  kb around rs6687360. Each data  
9 point represents a variant and the color represent the  $r^2$  (the degree of LD).
- 10 I. Top: PT cell Cicero-inferred co-accessibility between pairs of accessible chromatin sites at  
11 the *AGT* gene body region. Each line is an SBP GWAS (46) significant SNP ( $P < 5E-08$ ).  
12 Single cell chromatin accessibility of human kidney single nuclear ATAC-seq data. Endo:  
13 endothelial cells, Podo: podocyte, PT: proximal tubule, LOH: loop of Henle, DCT: distal  
14 convoluted tubule, PC: collecting duct principal cells, IC: collecting duct intercalated cells,  
15 Immune: immune cells. The showing genomic region is chr1:230,834,647-230,850,043.
- 16 J. Gapped  $k$ -mer support vector machine (GKM-SVM) method (54) estimated importance  
17 score for each base in the  $\pm 25$  bp region surrounding rs6687360 for the Ref and Alt alleles  
18 from the gkm-SVM model corresponding to PT cell cluster. The SNP of interest is  
19 highlighted in red color. The deltaSVM value was calculated to represent the changes of  
20 the transcription factor binding activity of this 51 bp sequence after replacing allele C to  
21 allele T of rs6687360.

22  
23