

---

**Supplementary information**

---

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

---

In the format provided by the  
authors and unedited

# Supplementary information for

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

Xin Sheng<sup>1,2,3#</sup>, Yuting Guan<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>, Steven Vitale<sup>1,2,3</sup>, Zhen Miao<sup>1,2,3</sup>, 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>, Todd L. Edwards<sup>12</sup>, Jacklyn N. Hellwege<sup>12</sup>, Adriana M. Hung<sup>13</sup>, Mingyao Li<sup>8</sup>, Benjamin 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>

Correspondence: [ksusztak@pennmedicine.upenn.edu](mailto:ksusztak@pennmedicine.upenn.edu)

### **This PDF file includes:**

Supplementary Note

Figs. S1 to S36

### **Other Supplementary information for this manuscript include the following:**

Tables S1 to S29 (Excel)

### **Source Data:**

Source Data Supplementary Fig. 21

Source Data Supplementary Fig. 29



## Supplementary Note

### Methods

#### Genotype and RNA-seq data production

388 individuals were genotyped using the Affymetrix Axiom Biobank genotyping array. Quality control steps were conducted using PLINK v1.9<sup>1</sup>. In total, 21 samples were excluded due to age or missing values, nine samples due to elevated genotypic heterozygosity (inbreeding coefficient cutoff: heterozygosity rate  $\pm$  3-fold standard deviations from the mean), 6 samples were excluded, because of high PIHAT values ( $> 0.2$ ). Monomorphic SNPs and SNPs with a minor allele frequency (MAF)  $< 5\%$ , Hardy-Weinberg equilibrium  $P < 1 \times 10^{-6}$ , and call rate  $< 90\%$  were excluded. SNPs associated with chemistry plate ID and SNPs located on sex chromosomes were excluded. After quality control, 367 individuals and 399,455 SNPs were used for further analysis (**Supplementary Table 28**).

Principal Component Analysis (PCA) was implemented using EIGENSTRAT<sup>2</sup> on the genotype data of 367 participants with additional 2,504 genotype data from the 1,000 Genomes Project Phase 3<sup>3</sup>. Participants in our study clustered with their corresponding ancestral samples from 1,000 Genomes Projects (**Supplementary Fig. 36**). SNPs that were absent or had incompatible alleles with the 1,000 Genomes dataset were removed. Genotype data were subsequently pre-phased by SHAPEIT<sup>4,5</sup> and imputed by IMPUTE2<sup>6</sup>, using the multiethnic panel reference from 1,000 Genome Phase 3 (NCBI build 37, release date October 2014). Imputed SNPs with MAF  $< 5\%$ , Hardy-Weinberg Equilibrium  $p < 1E-06$ , missing rate  $< 95\%$  for best-estimated genotypes at posterior probability  $> 0.9$ , and imputation confidence score, INFO (a measure of  $r^2$ )  $< 0.4$  (estimated by SNPTEST<sup>7</sup>) were

excluded. As a result, 6,155,299 autosomal SNPs with high quality were used for the eQTL analysis.

Our RNA-seq data constituted of libraries sequenced single-end reads (85 and 100 bp) and paired end reads (150 bp). Sequence quality was assessed with FastQC (v0.11.2). Adaptors and low-quality bases were trimmed using Trim-galore (v0.4.4). RNA-seq reads were then aligned to the human genome (hg19/GRCh37) using STAR (v2.4.1d)<sup>8</sup> based on GENCODE v19 annotations<sup>9</sup>. Gene-level read count quantifications were estimated using RSEM (v1.3.1)<sup>10</sup>. The estimated read counts of each sample were normalized using DESeq2 (v1.22.2)<sup>11</sup> and then log-transformed with an offset of 1. Genes, with log-transformed value > 1 in > 5% of samples, were retained. The resulting read counts were then centered and unit-normalized. Next, we conducted hierarchical clustering on the resulting matrix based on average and cosine distance. We calculated a two-sided  $\chi^2$  P-value based on Mahalanobis distance in a 20-dimensional PC space. 5 samples (3 for tubule and 2 for glomerular) were identified as outliers ( $\geq 60\%$  samples with Bonferroni-adjusted p-values < 0.05) and excluded. Sex compatibility checks were conducted based on X and Y chromosomal genes. 5 samples, with inconsistent sex, were removed from the downstream analysis. 15 samples with age under 20 were further excluded. After quality control, 359 individuals (356 tubule and 303 glomerular samples) with high quality RNA-seq and genotype data and full phenotype records were retained for further analysis. The estimated read counts were normalized across samples using TMM methods implemented in edgeR (v3.24.3)<sup>12</sup>, and gene expression levels were estimated as transcripts per million (TPM). Only genes with at least 0.1 TPM in at least 20% samples for each kidney

compartment were considered as expressed and used for the downstream analysis.

### **Cell fraction estimation in bulk RNA-seq data**

To estimate the cell fractions in the corresponding kidney compartments, we used CIBERSORTx<sup>13</sup>, a reference guided deconvolution tool for bulk tissue transcriptomes. First, we compared the deconvolution methods (Bisque<sup>14</sup> and CIBERSORTx<sup>13</sup>) by using 3 different input single cell RNA-seq data, including human kidney scRNA-seq data<sup>15</sup>, human kidney snRNA-seq data<sup>16</sup> and mouse kidney scRNA-seq data<sup>17</sup>, as references. Next, we conducted a systematic analysis of the input datasets and analytical methods. We found that more abundant cell types, such as proximal tubules, endothelial cells, loop of Henle cells, and macrophages showed highly consistent results (Pearson's correlation ~0.9). Immune cells were poorly captured by single nuclear RNA-seq, so there was limited available information for replication. The current dataset and CIBERSORTx showed the most consistent results between datasets and deconvolution methods (**Supplementary Fig. 2**) and this was used for downstream analyses<sup>18</sup>. Common cell types and those with limited co-linearity were selected in each kidney compartment for downstream eQTL identification.

### **CRISPR/Cas9 mediated genome editing**

HEK293 cell stably expressing Cas9 was a gift from Dr. Liling Wan (University of Pennsylvania). The sgRNA expression plasmids were generated as previously described<sup>19</sup> and sgRNA sequences are listed in **Supplementary Table 29**. Annealed sgRNA oligos were subcloned into pLKO5.sgRNA.EFS. GFP with the BsmB1 site. All constructs were verified by Sanger sequencing. sgRNA expression

plasmids were transfected using lipofectamine 3000. After 72hrs, cells were harvested, and *ACE*, *ABR* and *FGF5* expression levels were determined by QRT-PCR. At the same time, DNA was isolated, and sgRNAs target region deletion was confirmed by Sanger sequencing. The sequencing data for each deleted genomic region are shown in **Supplementary Table 9**.

### **Meta-analysis of multiple-tissue eQTL mapping**

We downloaded the eQTL summary results of 48 other human tissues from GTEx (v7; <https://gtexportal.org/home/>). To define tubule-specific and glomerulus-specific eGenes (**Fig. 1f and g**), a meta-analysis was performed on all lead SNP-eGene pairs observed in tubuli and glomeruli using METASOFT (v2.0.1)<sup>20</sup>. We first extracted all Beta values (effect sizes) and standard errors (SE) of the lead SNP-eGene pairs in tubuli and glomeruli from the summary results of 48 other human tissues from GTEx. A random effect model (RE2)<sup>20</sup> in METASOFT was adopted, and the M-value, that represents the posterior probability that an eQTL effect in a given tissue was estimated for each SNP-gene pair and tissue tested. We defined tissue-specific or study-specific eGene effect as the eQTL with M-value > 0.9 only in one tissue or study. Otherwise, eGenes with M-value > 0.9 in multiple tissues or studies were defined as shared eGenes. For the eQTL pairs, rs4292-*ACE* (**Supplementary Fig. 26**) and rs6687360-*AGT* (**Supplementary Fig. 31**), we excluded adipose visceral omentum and sun exposed lower leg skin eQTL data due to the high heterogeneity.

### **Motif enrichment with chromVAR**

TF motif-associated differences across cell clusters were estimated using chromVAR (v.1.5.0)<sup>21</sup>. Counts per DAR peak per cell matrix was first calculated and then used as an input to chromVAR. GC bias was adjusted based on hg19

reference using *addGCBias* function available in chromVAR package. The human motifs from the JASPAR2018 package<sup>22</sup> were used to match the motifs contained in each cell type specific DARs by *matchMotifs* function available in motifmatchr package (v1.4.0)<sup>23</sup>. The deviation Z-scores for each TF motif in each cell were calculated using *computeDeviations* and *deviationScores* functions available in chromVAR package. We also used findMotifsGenome.pl package available in Homer (v4.11)<sup>24</sup> to identify cell cluster-specific motifs with cell type specific DARs. We manually compared the top variable TF motifs identified by chromVAR and Homer<sup>24</sup>, and then identified the consistent top variable TF motifs as enriched motifs for each cell type.

### **Transcription factor footprinting**

We conducted transcription factor (TF) footprinting analysis as described previously<sup>25</sup>. First, we determined the genome-wide stranded positions of TF motifs within peaks using *matchMotifs* function available in motifmatchr package (v1.4.0)<sup>23</sup>. Next, we extended +/- 250 bp centered at the motif. To estimate the insertions around these sites, we overlapped the Tn5-corrected insertions, generated by the Cell Ranger ATAC (version 1.1.0), with the extended motif regions using bedtools<sup>26</sup> and calculated the number of Tn5-corrected insertions that covered per bp around the motif center. We followed the definition of the 3 key relative positions: (i) the footprint base was defined as + 5 bp from the motif center, (ii) the footprint flank was defined as the region between the end of footprint base and 50 bp away from the motif center, and (iii) the footprint background was defined as the region between 200 and 250 bp away from the motif center. The accessibility (footprinting) was calculated as the fold change between the number of Tn5-corrected insertions and the mean accessibility of the footprint background.

### **Enrichment test of functional categories**

Human kidney ChIP-seq data for histone H3K4me1, H3K4me3, H3K27ac, H3K36me3, H3K9ac, and H3K9me3 (GSM621634, GSM621648, GSM621651, GSM670025, GSM772811, and GSM1112806) were used to annotate chromatin states using ChromHMM<sup>27</sup>. To estimate the distribution under the null hypothesis (there is no enrichment), we first generated 100 background datasets for the open chromatin peaks called from human kidney snATAC-seq data for each cell type (total: 8×100 background datasets) by matching sequence length, GC content and repeat fractions of the positive set using *genNullSeqs* function available in gkm-SVM package (v0.80)<sup>28</sup>. Next, we calculated the Odds Ratio to estimate the enrichment of the open chromatin regions for each cell type in different chromatin states (enhancer, promoter, transcribed and inactive) of adult kidney comparing to the matched regions in background dataset.

### **Mouse Kidney scRNA-seq data**

We used RNA-seq count matrix of 43,745 kidney cells profiled from healthy mice (GSE107585)<sup>17</sup> and processed and normalized using R package Seurat v2.3.4<sup>29</sup>, batch effect was adjusted using ComBat<sup>30</sup>, while preserving zero expression values. We then normalized and log transformed expression data by the *NormalizeData* function in Seurat with scale factor = 1E04. Cluster assignment information was extracted from Park et al.<sup>17</sup>.

### **Genetic variants effects on chromatin accessibility**

To predict the effects of non-coding variants on the chromatin accessibility, we first extended the open chromatin peak summit by +/- 250 base pairs. The top 20,000 peaks with the highest significance (MACS2 q-value) for each cell cluster were

used to generate background sequences by matching the sequence length, GC and repeat fractions of the positive set using *genNullSeqs* function available in gkm-SVM package (v0.80)<sup>28</sup>. Gkm-SVM (weighted gapped k-mer Support Vector Machine) is a sequence-based method for predicting and detecting the regulatory vocabulary encoded in the functional regions. Given the large dataset, we used LS-GKM<sup>31</sup>, which provided additional advanced gapped *k-mer* based kernel functions, to train the SVM model for each cell cluster. We used a 10-fold cross-validation scheme to evaluate the classification accuracy of each cell cluster model. We observed that all our models for each cell cluster could achieve AUROC above 0.9, while using the 'gkmrbf' kernel function to train our model. We then used *gkmpredict* available in LS-GKM to predict the effects of 11 mers based on the trained model for each cell cluster. For each SNP measured in our eQTL analysis, we created 51 bp sequences (25 bp flanking either side of the variant base). We then used 'deltasvm.pl' script to calculate the accessibility changes (deltaSVM scores) caused by the one allele difference (of the queried variant) of the 51 mers sequence for each cell type. Next, we built a null distribution by randomly permuting the effects of 11 mers and re-calculated deltaSVM scores, and used the parameters of this null distribution to calculate a Z-score for each variant. We calculated a two-sided P-value based on the Z-score for each variant and then calculated the q-value<sup>32</sup> to identify the significant variants at FDR < 0.5.

### **Peak-peak co-accessibility scores**

We used Cicero (v1.0.15)<sup>33</sup> to calculate peak-peak co-accessibility scores for each cell cluster. The extracted cells were further aggregated based on the 50 nearest neighbors from the UMAP reduced dimensions using the *make\_cicero\_cds* function. Co-accessibility scores were then calculated by setting the window size

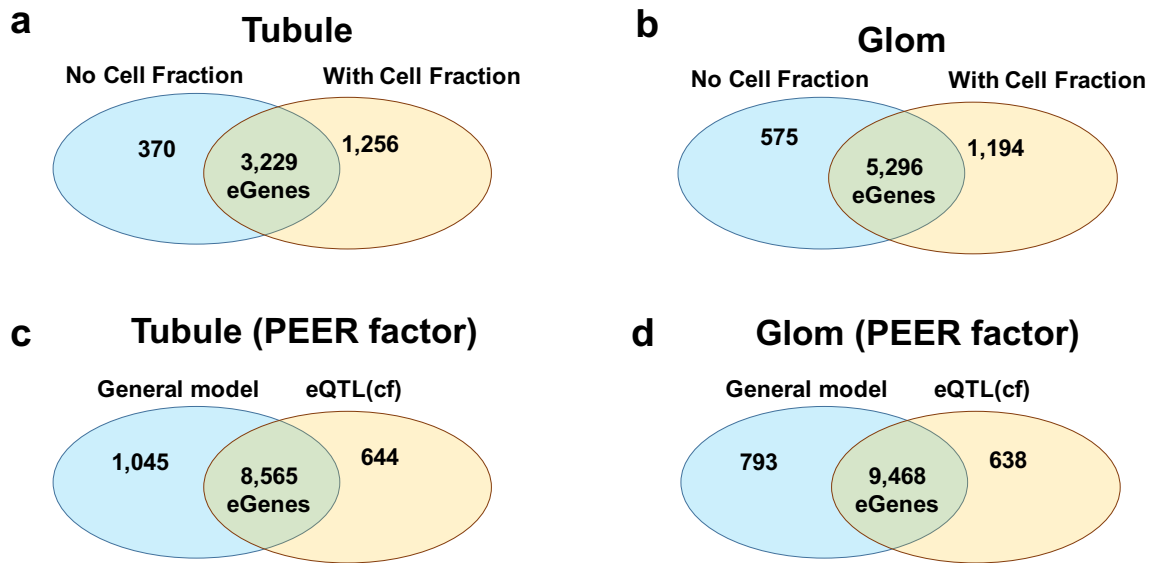
of 1 Mb and leaving the other parameters as the default setting using *run\_cicero* function available in the cicero. A heuristic cutoff of 0.25 was adopted to determine co-accessibility.

### **Fine mapping analysis**

To identify the causal variants that drive the gene expression changes in the tubule and glomerular compartments, we conducted an Iterative Bayesian Step-wise Selection (IBSS) based sum of single effect model for fine mapping analysis implemented in SuSiE (v0.9.26)<sup>34</sup> and followed the tutorial of SusieR (<https://stephenslab.github.io/susieR>). The SNP dosages of the variants located within +/- 1 Mb around the TSS site of the queried gene were used as the input X and the INT transformed gene expression was used as the Y to the *susie* function. In addition, we assumed that there were at most 10 causal variables for each gene, and set SuSiE prior variance to 0.1. The causal SNPs included in the 90% credible sets were identified as the causal SNPs that driving the gene expression variations in each kidney compartment.

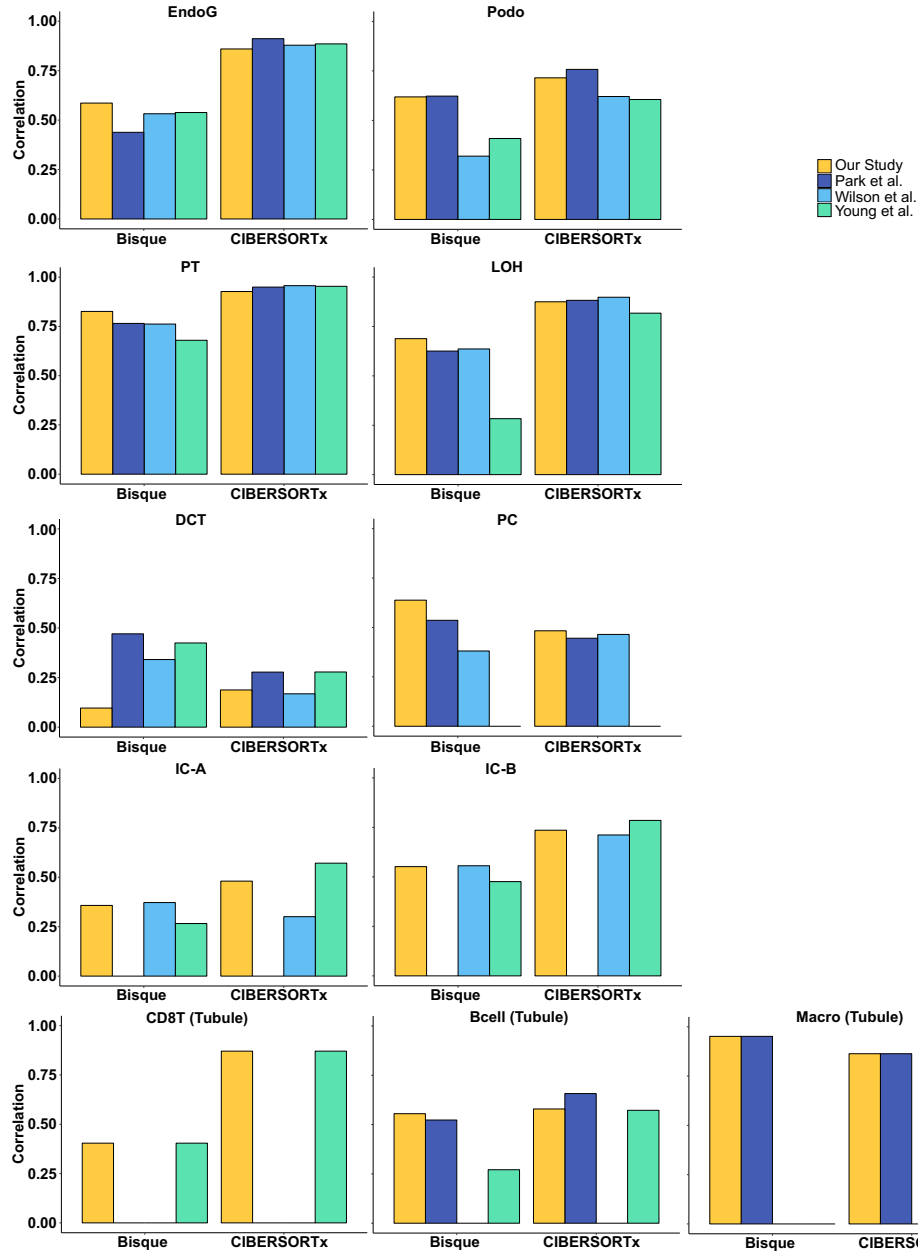


## Supplementary figure 1



**Fig. S1. The comparison of eGenes identified by different models.** a) Venn diagram illustrating the overlap of eGenes identified by the general eQTL model (without PEER factors) and the eQTL(cf) model (without PEER factors) in 356 human tubule samples. b) Venn diagram illustrating the overlap of eGenes identified by the general eQTL model (without PEER factors) and the eQTL(cf) model (without PEER factors) in 303 human glomerular samples. c) Venn diagram illustrating the overlap of eGenes identified by the general eQTL model and the eQTL(cf) model in 356 human tubule samples. d) Venn diagram illustrating the overlap of eGenes identified by the general eQTL model and the eQTL(cf) model in 303 human glomerular samples.

## Supplementary figure 2



**Fig. S2. Bulk data deconvolution analysis.** The mean Pearson's correlation coefficients value obtained for the pairwise comparison of deconvolution results using 4 single cell data sources<sup>15-17,35</sup> and 2 deconvolution methods (Bisque<sup>14</sup> and CIBERSORTx<sup>13</sup>) for 659 microdissected human kidney samples (356 tubules and 303 glomeruli).

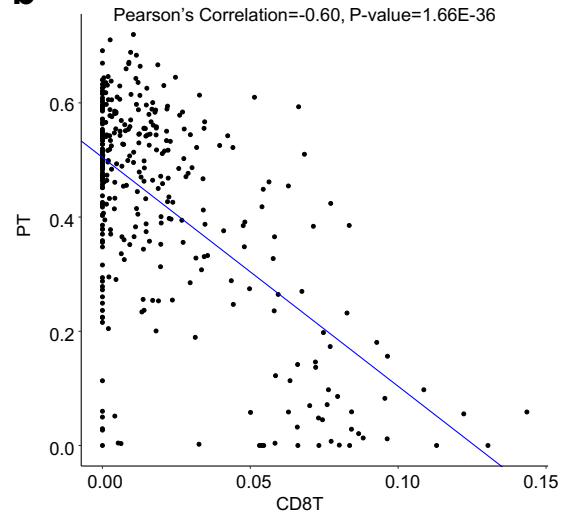
### Supplementary figure 3

**a**

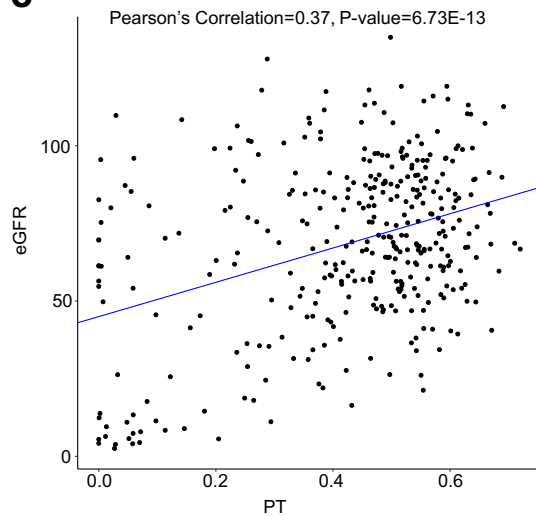
| Cell Type | Pearson's Correlation | Two-sided P-value |
|-----------|-----------------------|-------------------|
| CD8T      | -0.60                 | 1.66E-36          |
| Macro     | -0.68                 | 1.25E-48          |
| Treg      | -0.38                 | 2.62E-13          |
| Mono      | -0.33                 | 1.19E-10          |
| B         | -0.41                 | 9.57E-16          |

Two-sided P-value was calculated by t-test (df=354).

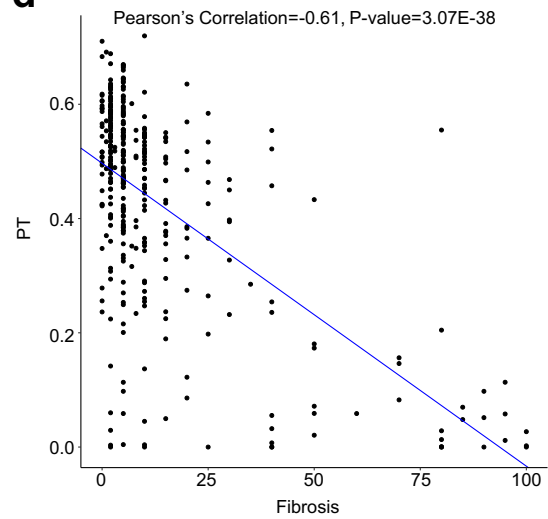
**b**



**c**



**d**



**Fig. S3. The relationship between cell fraction and disease states.** Pearson's correlation between PT cell fraction and a) immune cell fraction b) CD8T c) eGFR and d) fibrosis (N=356 tubule samples; Two-sided P-value was calculated by t-test with df=354).

## Supplementary figure 4

a

### Tubuli

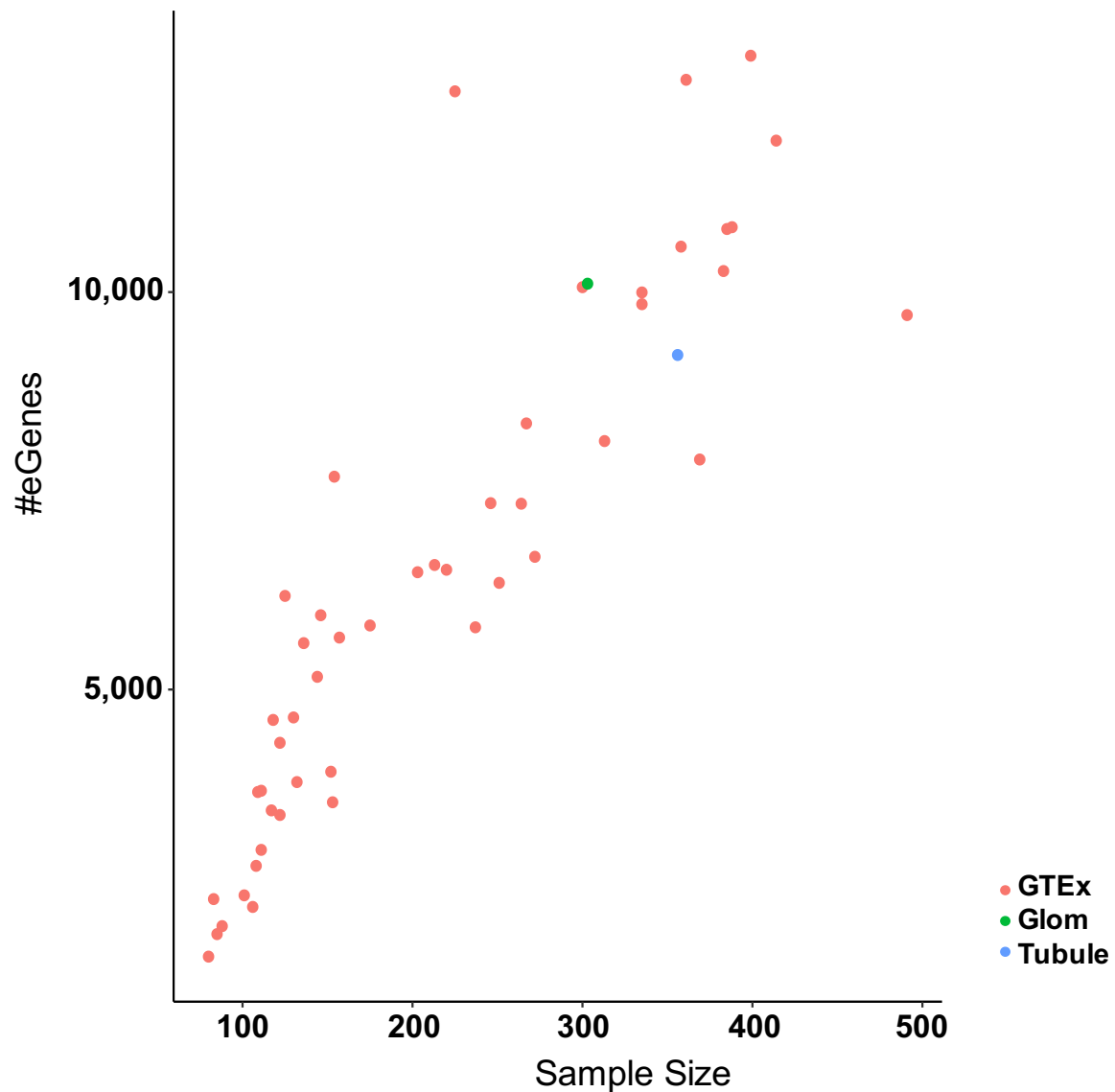
| PT     | ALOH  | DCT   | IC-B  | CNT   | DC    | Granul | Treg  | NK    | Th17  | CD8+ T | CD4+ T | Macro | Fibrosis | Batch | Site  | Paired-end | Age   | Gender | RIN   | PC1   | PC2   | PC3   | PC4   | PC5   |
|--------|-------|-------|-------|-------|-------|--------|-------|-------|-------|--------|--------|-------|----------|-------|-------|------------|-------|--------|-------|-------|-------|-------|-------|-------|
| PEER1  | 0.37  | -0.14 | -0.17 | -0.07 | -0.29 | -0.43  | 0.07  | -0.02 | -0.20 | -0.21  | -0.14  | -0.23 | -0.17    | -0.18 | -0.15 | -0.19      | -0.05 | 0.06   | -0.31 | 0.13  | 0.14  | -0.24 | -0.31 | -0.03 |
| PEER2  | 0.09  | -0.01 | 0.00  | -0.18 | 0.09  | -0.02  | -0.15 | -0.14 | -0.12 | -0.11  | -0.16  | -0.02 | -0.10    | -0.18 | 0.67  | 0.49       | 0.08  | 0.03   | -0.16 | 0.06  | -0.36 | -0.42 | -0.24 | -0.31 |
| PEER3  | 0.43  | 0.01  | -0.26 | -0.56 | 0.06  | -0.27  | -0.17 | -0.24 | -0.38 | -0.56  | -0.52  | -0.03 | -0.36    | -0.30 | -0.09 | 0.10       | 0.10  | 0.14   | 0.02  | 0.21  | -0.09 | -0.06 | -0.06 | 0.03  |
| PEER4  | 0.48  | -0.03 | -0.08 | -0.30 | -0.24 | -0.50  | -0.24 | -0.24 | -0.37 | -0.40  | -0.21  | -0.45 | -0.36    | -0.43 | 0.58  | 0.24       | 0.33  | 0.03   | -0.03 | 0.12  | -0.21 | -0.25 | -0.14 | -0.19 |
| PEER5  | 0.48  | 0.40  | 0.26  | -0.48 | -0.63 | -0.56  | -0.30 | -0.29 | -0.51 | -0.22  | -0.44  | -0.43 | -0.52    | -0.44 | 0.10  | 0.05       | -0.03 | -0.15  | 0.13  | -0.03 | -0.06 | -0.07 | -0.11 | 0.00  |
| PEER6  | 0.74  | 0.08  | -0.09 | -0.64 | -0.44 | -0.74  | -0.34 | -0.39 | -0.68 | -0.56  | -0.56  | -0.46 | -0.65    | -0.67 | 0.15  | 0.09       | -0.18 | -0.11  | -0.06 | -0.05 | -0.14 | -0.13 | -0.10 | -0.05 |
| PEER7  | -0.69 | -0.19 | -0.10 | 0.57  | 0.59  | 0.71   | 0.30  | 0.42  | 0.70  | 0.39   | 0.61   | 0.35  | 0.63     | 0.55  | -0.23 | -0.13      | -0.14 | 0.16   | 0.03  | -0.02 | 0.14  | 0.13  | 0.11  | 0.03  |
| PEER8  | -0.27 | 0.05  | 0.14  | 0.10  | 0.06  | 0.14   | 0.22  | 0.12  | 0.17  | 0.22   | 0.02   | 0.20  | 0.16     | 0.18  | -0.18 | 0.10       | 0.08  | -0.04  | -0.11 | -0.10 | -0.13 | -0.10 | -0.06 | -0.04 |
| PEER9  | -0.12 | -0.20 | 0.02  | 0.20  | 0.05  | 0.12   | 0.08  | 0.15  | 0.20  | 0.12   | 0.35   | 0.02  | 0.18     | 0.00  | 0.03  | 0.00       | -0.18 | 0.00   | 0.03  | 0.09  | -0.17 | -0.14 | -0.06 | -0.05 |
| PEER10 | 0.34  | 0.37  | -0.02 | -0.24 | -0.45 | -0.23  | -0.18 | -0.11 | -0.24 | -0.16  | -0.15  | -0.01 | -0.23    | -0.22 | 0.05  | -0.04      | -0.07 | -0.13  | -0.10 | 0.11  | -0.08 | -0.09 | -0.07 | 0.03  |
| PEER11 | -0.32 | 0.40  | 0.39  | 0.18  | -0.13 | 0.08   | 0.16  | -0.02 | 0.14  | 0.31   | 0.13   | 0.01  | 0.13     | 0.20  | 0.14  | 0.06       | 0.23  | 0.14   | -0.05 | 0.12  | 0.04  | 0.00  | 0.04  | -0.07 |
| PEER12 | 0.49  | -0.59 | -0.37 | -0.13 | 0.09  | -0.25  | -0.12 | 0.03  | -0.07 | -0.48  | -0.03  | -0.22 | -0.08    | -0.15 | 0.18  | 0.11       | 0.05  | 0.07   | -0.01 | 0.05  | -0.12 | -0.11 | -0.06 | -0.04 |
| PEER13 | -0.37 | 0.24  | 0.57  | 0.16  | -0.12 | 0.10   | 0.26  | -0.03 | 0.07  | 0.53   | 0.07   | 0.12  | 0.12     | 0.03  | 0.07  | 0.07       | 0.05  | -0.11  | 0.07  | 0.14  | 0.01  | -0.02 | 0.00  | -0.07 |
| PEER14 | 0.31  | -0.24 | -0.39 | -0.08 | 0.01  | -0.12  | -0.02 | 0.02  | -0.29 | -0.13  | 0.04   | -0.10 | 0.09     | -0.04 | 0.04  | -0.02      | -0.02 | 0.13   | -0.07 | 0.13  | 0.10  | 0.04  | 0.03  | -0.04 |
| PEER15 | -0.49 | -0.38 | 0.35  | 0.49  | 0.17  | 0.50   | 0.27  | 0.39  | 0.55  | 0.51   | 0.42   | 0.20  | 0.49     | 0.44  | -0.05 | -0.01      | 0.20  | -0.07  | -0.06 | 0.03  | 0.14  | 0.10  | 0.03  | -0.02 |
| PEER16 | 0.39  | 0.08  | -0.15 | -0.37 | -0.21 | -0.38  | -0.07 | -0.27 | -0.37 | -0.31  | -0.42  | -0.12 | -0.28    | -0.26 | 0.13  | 0.05       | 0.08  | 0.21   | -0.09 | -0.19 | 0.06  | 0.04  | -0.10 | -0.06 |
| PEER17 | -0.28 | -0.21 | -0.09 | 0.34  | 0.37  | 0.33   | 0.23  | 0.09  | 0.19  | 0.25   | 0.25   | 0.23  | 0.29     | 0.43  | -0.16 | -0.24      | -0.14 | -0.07  | 0.03  | -0.09 | 0.15  | 0.14  | 0.09  | 0.05  |
| PEER18 | 0.44  | -0.24 | -0.41 | -0.23 | 0.07  | -0.27  | -0.17 | -0.27 | -0.19 | -0.51  | -0.18  | -0.12 | -0.20    | -0.20 | 0.09  | 0.16       | 0.15  | -0.06  | 0.12  | -0.15 | 0.04  | 0.05  | 0.06  | 0.01  |
| PEER19 | 0.07  | 0.03  | -0.17 | -0.14 | -0.14 | -0.05  | 0.00  | 0.03  | -0.09 | -0.14  | -0.23  | 0.06  | -0.12    | -0.17 | -0.26 | -0.17      | -0.33 | 0.00   | -0.15 | -0.01 | -0.02 | 0.06  | 0.12  | 0.08  |
| PEER20 | 0.30  | -0.31 | -0.34 | -0.12 | 0.15  | -0.14  | -0.09 | -0.15 | -0.17 | -0.33  | -0.18  | -0.15 | -0.16    | -0.05 | 0.04  | 0.22       | 0.29  | -0.03  | -0.15 | -0.11 | -0.22 | -0.18 | -0.05 | -0.04 |
| PEER21 | -0.06 | 0.16  | -0.04 | 0.09  | -0.04 | -0.04  | -0.07 | 0.08  | 0.07  | -0.11  | 0.10   | -0.15 | -0.02    | -0.03 | 0.12  | 0.07       | 0.11  | -0.04  | -0.14 | 0.07  | -0.15 | -0.17 | -0.08 | -0.14 |
| PEER22 | -0.22 | -0.31 | 0.40  | -0.19 | -0.05 | 0.10   | 0.19  | -0.15 | 0.07  | 0.26   | 0.13   | 0.16  | 0.13     | 0.07  | 0.22  | 0.11       | 0.39  | -0.20  | -0.11 | 0.13  | 0.02  | 0.03  | 0.05  | 0.02  |
| PEER23 | -0.07 | -0.11 | 0.03  | 0.04  | 0.24  | 0.03   | -0.02 | -0.06 | -0.03 | 0.18   | 0.08   | -0.07 | -0.03    | 0.08  | -0.17 | -0.01      | -0.21 | -0.04  | 0.15  | -0.07 | -0.01 | -0.04 | -0.06 | -0.03 |
| PEER24 | 0.01  | -0.25 | -0.06 | 0.16  | 0.15  | 0.20   | 0.11  | 0.19  | 0.21  | 0.00   | 0.18   | 0.14  | 0.13     | 0.12  | -0.09 | -0.18      | -0.37 | -0.03  | 0.04  | -0.02 | 0.22  | 0.19  | 0.10  | 0.04  |
| PEER25 | 0.39  | 0.09  | -0.20 | -0.32 | -0.29 | -0.23  | -0.17 | 0.01  | -0.20 | -0.42  | -0.29  | -0.09 | -0.23    | -0.25 | 0.26  | 0.07       | 0.29  | 0.05   | -0.06 | -0.11 | 0.16  | 0.12  | 0.05  | 0.02  |
| PEER26 | 0.35  | -0.12 | -0.39 | -0.20 | 0.00  | -0.24  | -0.28 | -0.14 | -0.12 | -0.47  | -0.12  | -0.11 | -0.11    | -0.17 | -0.05 | 0.08       | -0.10 | 0.12   | 0.05  | -0.17 | -0.11 | -0.18 | -0.18 | -0.13 |
| PEER27 | -0.31 | -0.08 | 0.12  | 0.43  | 0.09  | 0.32   | 0.19  | 0.27  | 0.44  | 0.25   | 0.36   | 0.16  | 0.34     | 0.30  | 0.04  | -0.01      | 0.19  | -0.05  | -0.11 | -0.02 | 0.05  | 0.04  | 0.05  | 0.01  |
| PEER28 | 0.02  | -0.07 | 0.06  | -0.05 | -0.02 | -0.16  | -0.06 | -0.30 | -0.14 | 0.15   | -0.05  | -0.09 | -0.13    | 0.06  | -0.04 | -0.05      | -0.15 | 0.04   | -0.22 | 0.03  | 0.27  | 0.25  | 0.14  | 0.07  |
| PEER29 | -0.06 | 0.01  | -0.02 | -0.05 | -0.02 | 0.01   | -0.02 | -0.10 | -0.02 | 0.00   | -0.11  | -0.05 | -0.11    | 0.05  | -0.21 | -0.19      | 0.25  | 0.05   | -0.05 | -0.05 | -0.05 | -0.05 | -0.05 | -0.05 |
| PEER30 | 0.11  | -0.23 | -0.05 | 0.03  | -0.10 | 0.04   | -0.01 | -0.03 | 0.10  | -0.07  | -0.02  | 0.07  | 0.05     | -0.04 | 0.04  | 0.05       | 0.28  | 0.04   | -0.18 | -0.04 | -0.11 | -0.06 | 0.00  | -0.08 |
| PEER31 | -0.50 | 0.20  | 0.34  | 0.44  | 0.22  | 0.40   | 0.23  | 0.11  | 0.33  | 0.55   | 0.41   | 0.18  | 0.31     | 0.36  | -0.04 | -0.22      | -0.08 | -0.15  | 0.15  | 0.13  | 0.25  | 0.23  | 0.11  | 0.08  |
| PEER32 | 0.47  | 0.18  | -0.11 | -0.46 | -0.29 | -0.40  | -0.31 | -0.32 | -0.50 | -0.28  | -0.43  | -0.14 | -0.53    | -0.39 | 0.22  | 0.17       | 0.16  | 0.02   | 0.11  | -0.10 | -0.16 | -0.16 | -0.04 | -0.02 |
| PEER33 | -0.21 | -0.08 | 0.03  | 0.06  | 0.13  | 0.12   | 0.12  | 0.04  | -0.01 | 0.17   | 0.06   | 0.10  | 0.03     | 0.11  | -0.13 | 0.01       | -0.01 | 0.01   | -0.11 | 0.06  | -0.01 | 0.02  | 0.06  | 0.03  |
| PEER34 | 0.30  | -0.32 | -0.27 | -0.06 | 0.06  | -0.04  | -0.07 | -0.03 | 0.02  | -0.34  | -0.03  | -0.04 | -0.06    | -0.05 | 0.17  | 0.19       | 0.60  | -0.04  | 0.04  | -0.11 | 0.03  | 0.04  | 0.06  | 0.01  |
| PEER35 | -0.05 | 0.20  | 0.14  | -0.03 | -0.15 | 0.12   | -0.17 | -0.09 | 0.06  | 0.08   | 0.04   | -0.05 | -0.05    | -0.06 | 0.18  | 0.26       | 0.57  | -0.05  | -0.18 | -0.04 | -0.17 | -0.12 | -0.03 | -0.14 |

b

### Glomeruli

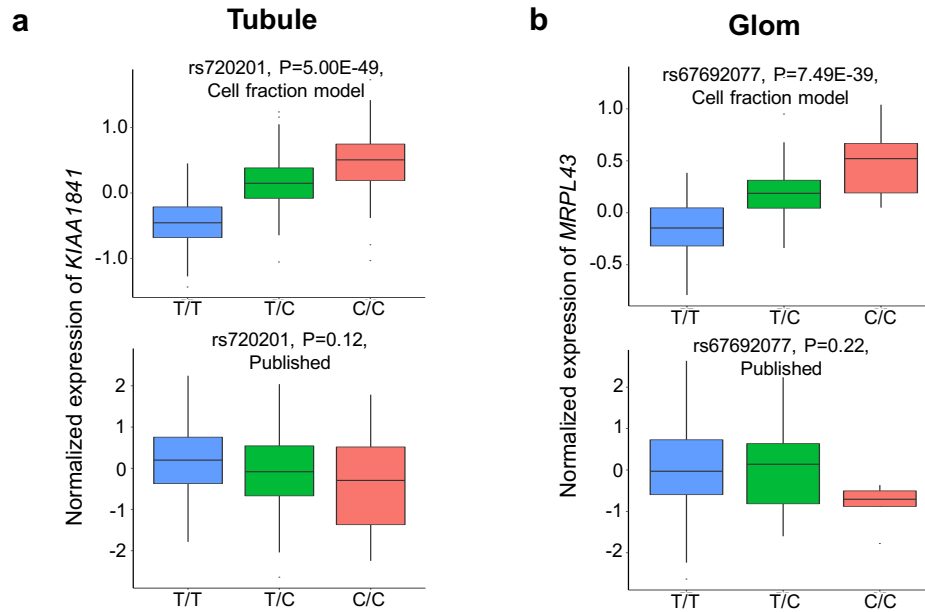
|        | EndoG | Endo  | Podo  | Granul | Treg  | NK    | Th17  | CD8+ T | CD4+ T | Macro | Sclerosis | Batch | Paired-end | Age   | Gender | RIN   | Site  | PC1   | PC2   | PC3   | PC4   | PC5   |
|--------|-------|-------|-------|--------|-------|-------|-------|--------|--------|-------|-----------|-------|------------|-------|--------|-------|-------|-------|-------|-------|-------|-------|
| PEER1  | 0.45  | -0.31 | 0.22  | -0.30  | -0.03 | 0.14  | 0.01  | -0.06  | 0.04   | -0.24 | -0.03     | 0.28  | -0.16      | 0.04  | -0.10  | 0.43  | 0.48  | -0.35 | -0.42 | -0.24 | -0.32 | 0.27  |
| PEER2  | 0.23  | -0.10 | 0.16  | -0.16  | 0.04  | 0.30  | 0.04  | -0.11  | 0.08   | 0.06  | 0.18      | -0.35 | -0.38      | 0.01  | -0.06  | 0.24  | -0.05 | 0.02  | 0.03  | 0.02  | 0.01  | 0.08  |
| PEER3  | 0.58  | -0.34 | 0.17  | -0.21  | 0.19  | 0.32  | -0.25 | -0.23  | 0.21   | -0.13 | 0.11      | -0.45 | -0.11      | 0.21  | -0.19  | 0.04  | 0.29  | -0.14 | -0.13 | -0.06 | -0.05 | 0.04  |
| PEER4  | 0.02  | 0.01  | 0.18  | 0.18   | -0.27 | -0.59 | -0.28 | 0.29   | -0.05  | -0.19 | -0.24     | -0.20 | -0.24      | -0.10 | 0.02   | 0.27  | -0.03 | -0.03 | 0.01  | 0.06  | 0.02  | -0.01 |
| PEER5  | -0.04 | 0.15  | -0.23 | -0.03  | 0.27  | -0.04 | -0.05 | 0.11   | 0.04   | 0.11  | -0.13     | -0.17 | -0.06      | -0.07 | -0.02  | -0.31 | 0.13  | -0.08 | -0.09 | -0.09 | -0.04 | -0.02 |
| PEER6  | -0.55 | 0.61  | -0.62 | 0.38   | 0.16  | -0.20 | -0.06 | 0.02   | 0.09   | 0.72  | 0.23      | 0.01  | 0.10       | -0.06 | 0.16   | -0.21 | -0.34 | 0.14  | 0.19  | 0.18  | 0.12  | -0.07 |
| PEER7  | -0.25 | 0.26  | -0.42 | 0.22   | 0.10  | 0.11  | 0.10  | -0.07  | -0.11  | 0.53  | 0.29      | -0.31 | -0.02      | 0.15  | 0.06   | -0.36 | -0.20 | 0.39  | 0.40  | 0.25  | 0.19  | -0.08 |
| PEER8  | -0.06 | 0.08  | -0.05 | -0.22  | 0.16  | 0.29  | 0.19  | -0.11  | 0.06   | 0.12  | 0.16      | 0.18  | 0.10       | 0.07  | 0.14   | -0.02 | 0.01  | 0.08  | 0.05  | -0.05 | 0.02  | 0.02  |
| PEER9  | -0.63 | 0.56  | -0.33 | 0.44   | -0.12 | -0.62 | -0.04 | 0.41   | -0.09  | 0.38  | -0.07     | 0.07  | 0.13       | -0.11 | 0.11   | -0.09 | -0.31 | 0.09  | 0.13  | 0.13  | 0.09  | -0.04 |
| PEER10 | -0.11 | 0.12  | 0.09  | 0.11   | -0.14 | -0.16 | 0.01  | 0.01   | -0.02  | -0.07 | -0.04     | -0.03 | -0.07      | -0.05 | -0.05  | -0.07 | 0.07  | -0.06 | -0.07 | -0.01 | -0.06 | -0.01 |
| PEER11 | 0.41  | -0.17 | -0.05 | 0.02   | -0.07 | 0.01  | -0.23 | -0.08  | 0.10   | -0.12 | -0.08     | -0.27 | -0.13      | 0.12  | 0.03   | -0.20 | 0.21  | -0.16 | -0.17 | -0.11 | -0.05 | -0.09 |
| PEER12 | 0.32  | -0.08 | -0.16 | -0.17  | 0.00  | 0.10  | -0.03 | 0.03   | -0.13  | -0.10 | 0.08      | -0.12 | -0.16      | -0.09 | 0.06   | 0.04  | 0.00  | 0.11  | 0.10  | 0.00  | 0.11  | -0.04 |
| PEER13 | 0.03  | 0.16  | -0.33 | 0.24   | -0.08 | -0.21 | -0.27 | 0.02   | 0.17   | 0.14  | -0.08     | -0.21 | -0.20      | -0.13 | 0.24   | -0.04 | -0.03 | 0.16  | 0.15  | 0.04  | 0.10  | 0.02  |
| PEER14 | -0.04 | 0.05  | -0.26 | 0.32   | 0.01  | -0.09 | -0.23 | -0.15  | 0.27   | 0.10  | -0.08     | 0.04  | 0.00       | -0.07 | 0.02   | 0.03  | 0.07  | 0.10  | 0.10  | 0.01  | 0.07  | -0.03 |
| PEER15 | -0.19 | 0.11  | 0.10  | 0.15   | -0.04 | -0.14 | -0.07 | 0.09   | 0.06   | 0.15  | 0.00      | -0.10 | -0.34      | -0.03 | 0.00   | 0.03  | -0.04 | 0.04  | 0.00  | -0.08 | 0.00  | 0.03  |
| PEER16 | -0.14 | -0.01 | 0.06  | -0.08  | 0.06  | 0.08  | 0.13  | -0.09  | -0.01  | 0.04  | 0.13      | 0.21  | 0.23       | 0.05  | 0.12   | 0.07  | 0.02  | 0.08  | 0.06  | -0.06 | 0.05  | 0.01  |
| PEER17 | 0.56  | -0.39 | 0.41  | -0.26  | -0.15 | 0.13  | -0.12 | -0.02  | -0.02  | -0.53 | -0.15     | -0.46 | -0.50      | 0.13  | -0.07  | 0.08  | 0.12  | -0.21 | -0.18 | -0.10 | -0.01 | -0.03 |
| PEER18 | 0.06  | 0.13  | -0.46 | -0.23  | 0.33  | 0.06  | 0.27  | 0.08   | -0.20  | 0.15  | 0.09      | 0.13  | 0.43       | -0.04 | 0.13   | -0.08 | -0.04 | 0.00  | -0.03 | 0.05  | -0.09 | -0.05 |
| PEER19 | -0.44 | 0.15  | 0.05  | 0.14   | -0.08 | 0.04  | 0.22  | -0.11  | 0.12   | 0.10  | 0.09      | 0.25  | 0.16       | -0.07 | -0.15  | -0.01 | -0.11 | 0.13  | 0.11  | 0.05  | 0.02  | -0.02 |
| PEER20 | -0.49 | 0.34  | -0.26 | 0.22   | 0.16  | -0.14 | 0.02  | 0.09   | -0.01  | 0.49  | 0.02      | 0.26  | 0.19       | -0.14 | -0.01  | 0.01  | -0.17 | -0.03 | -0.05 | -0.07 | -0.04 | -0.04 |
| PEER21 | 0.06  | 0.05  | -0.13 | 0.16   | 0.05  | -0.04 | -0.08 | 0.04   | 0.00   | 0.01  | -0.12     | -0.01 | 0.08       | -0.14 | -0.28  | -0.12 | 0.01  | -0.04 | 0.01  | 0.06  | 0.04  | -0.03 |
| PEER22 | 0.09  | 0.00  | -0.04 | 0.05   | 0.03  | -0.12 | -0.13 | 0.04   | 0.08   | -0.06 | -0.12     | -0.02 | 0.04       | -0.23 | 0.14   | -0.07 | -0.04 | 0.11  | 0.16  | 0.10  | 0.16  | -0.11 |
| PEER23 | 0.06  | -0.17 | 0.26  | -0.01  | -0.20 | 0.06  | -0.08 | 0.11   | -0.09  | -0.04 | 0.08      | -0.24 | -0.17      | 0.10  | 0.27   | -0.12 | 0.00  | -0.19 | -0.20 | -0.16 | -0.07 | 0.02  |
| PEER24 | 0.01  | 0.03  | -0.02 | -0.02  | 0.05  | 0.02  | -0.13 | 0.07   | 0.01   | 0.16  | 0.07      | -0.11 | -0.22      | 0.07  | 0.29   | -0.01 | 0.00  | -0.20 | -0.20 | -0.09 | -0.10 | -0.03 |
| PEER25 | -0.10 | 0.00  | 0.24  | 0.06   | -0.23 | -0.05 | -0.11 | 0.03   | 0.06   | -0.03 | 0.08      | -0.19 | -0.31      | 0.10  | 0.22   | 0.00  | -0.04 | -0.22 | -0.21 | -0.13 | -0.08 | 0.00  |
| PEER26 | 0.41  | -0.35 | 0.39  | -0.07  | -0.24 | -0.01 | -0.31 | 0.02   | 0.09   | -0.37 | -0.09     | -0.33 | -0.37      | -0.10 | -0.02  | 0.08  | 0.18  | -0.16 | -0.12 | -0.11 | -0.02 | 0.10  |
| PEER27 | -0.01 | -0.02 | 0.14  | -0.02  | 0.04  | 0.11  | 0.03  | -0.01  | 0.00   | 0.03  | 0.03      | -0.07 | -0.09      | 0.07  | -0.26  | 0.07  | -0.06 | 0.04  | -0.04 | -0.05 | 0.10  | 0.01  |
| PEER28 | -0.11 | 0.11  | -0.04 | -0.04  | -0.01 | -0.07 | 0.01  | -0.12  | -0.05  | -0.06 | -0.09     | -0.16 | 0.30       | -0.03 | 0.04   | 0.05  | 0.06  | -0.21 | -0.13 | -0.03 | -0.03 | 0.03  |
| PEER29 | 0.10  | 0.04  | 0.05  | 0.04   | 0.08  | 0.02  | -0.08 | 0.01   | 0.02   | 0.10  | -0.09     | -0.04 | 0.09       | -0.31 | -0.04  | -0.06 | 0.04  | -0.08 | 0.01  | 0.02  | 0.15  | 0.04  |
| PEER30 | -0.12 | 0.08  | -0.10 | -0.10  | 0.07  | -0.13 | 0.09  | -0.04  | -0.03  | 0.18  | 0.13      | 0.11  | 0.00       | -0.09 | -0.16  | -0.11 | -0.05 | 0.07  | 0.17  | 0.28  | 0.09  | 0.02  |
| PEER31 | -0.07 | 0.07  | -0.15 | 0.03   | 0.04  | 0.00  | 0.04  | 0.04   | -0.06  | 0.06  | 0.11      | 0.11  | 0.30       | 0.22  | -0.03  | -0.19 | -0.05 | 0.32  | 0.25  | 0.12  | 0.02  | -0.03 |
| PEER32 | 0.06  | 0.04  | 0.07  | 0.02   | -0.13 | -0.10 | -0.01 | -0.04  | 0.05   | -0.16 | -0.13     | -0.02 | -0.05      | -0.15 | -0.01  | 0.11  | 0.01  | -0.04 | 0.04  | 0.07  | 0.09  | 0.02  |
| PEER33 | -0.07 | 0.16  | -0.21 | 0.12   | 0.20  | -0.06 | -0.01 | -0.01  | 0.03   | 0.13  | -0.01     | 0.19  | 0.34       | -0.17 | -0.01  | 0.03  | -0.05 | 0.29  | 0.28  | 0.24  | 0.08  | -0.05 |
| PEER34 | 0.25  | -0.18 | 0.13  | 0.04   | -0.07 | 0.04  | -0.16 | -0.02  | 0.06   | -0.23 | -0.10     | -0.11 | 0.08       | -0.20 | -0.25  | 0.11  | 0.16  | -0.29 | -0.20 | -0.05 | 0.02  | 0.12  |
| PEER35 | -0.08 | 0.01  | 0.07  | 0.02   | -0.18 | -0.07 | 0.03  | -0.09  | 0.07   | -0.20 | -0.08     | -0.10 | -0.27      | -0.09 | -0.25  | -0.06 | -0.01 | 0.13  | 0.12  | 0.07  | 0.05  | 0.05  |

Supplementary figure 5



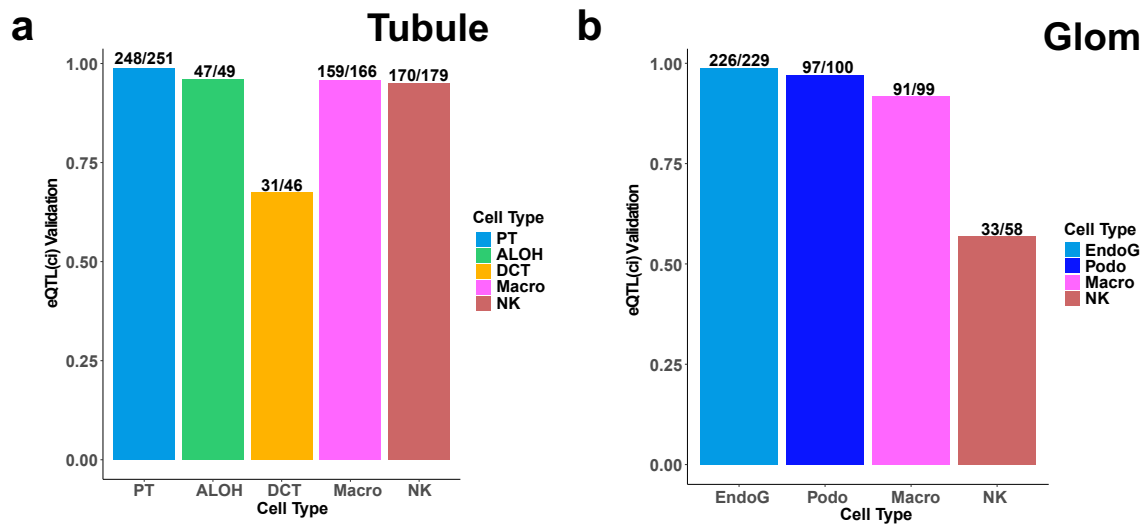
**Fig. S5. The number of identified eGenes in each kidney compartment.** Relationship between sample size (x-axis) and the number of identified eGenes (y-axis). Green dot shows the number of eGenes identified in glomeruli; blue dot shows the number of eGenes identified in tubule; red dots show the number of eGenes identified in 48 human tissues from GTEx (v7).

## Supplementary figure 6



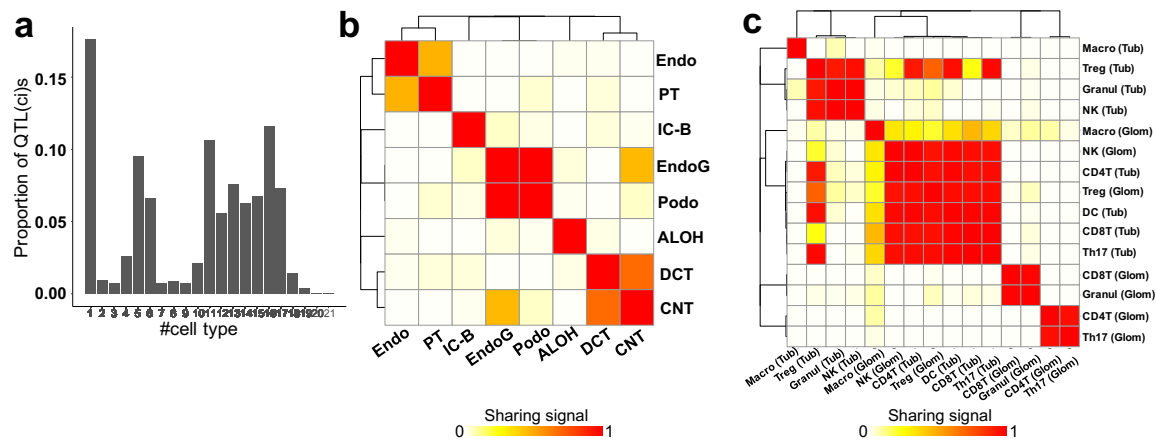
**Fig. S6. Comparing newly identified eQTLs using eQTL(cf) model to previously published eQTL data<sup>36</sup>.** a) The genotype of SNP rs720201 and the gene expression of *KIAA1841* association in human kidney tubules in new cell fraction adjusted eQTL(cf) data (top panel, N=356 samples; two-sided P-value was calculated by linear regression eQTL(cf) model) or the previously published tubule eQTL data<sup>36</sup> (bottom panel, N=121 samples; two-sided P-value was calculated by linear regression model). b) The genotype of SNP rs67692077 and the gene expression of *MRPL43* association in human kidney glomeruli in the cell fraction adjusted eQTL(cf) data (top panel, N=303 samples; two-sided P-value was calculated by linear regression eQTL(cf) model) or the previously published glomeruli eQTL data<sup>36</sup> (bottom panel, N=119 samples; two-sided P-value was calculated by linear regression model). Center lines show the medians; box limits indicate the 25<sup>th</sup> and 75<sup>th</sup> percentiles; whiskers extend to the 5<sup>th</sup> and 95<sup>th</sup> percentiles; outliers are represented by dots. P was calculated by linear regression model.

## Supplementary figure 7



**Fig. S7. The proportion cell type eQTL(cis) validated using healthy mouse kidney scRNA-seq data for deconvolution<sup>17</sup>** a) in tubule b) and glomerular fraction (Glom). Validation was defined as eQTL(cis) that has Akaike's information criterion (AIC) < 0 and nominal two-sided  $\chi^2$  p-value < 0.05. EndoG: glomerular endothelial cells, Podo: podocyte, PT: proximal tubule, ALOH: ascending loop of Henle, DCT: distal convoluted tubule, Macro: macrophage, NK: natural killer cells.

## Supplementary figure 8

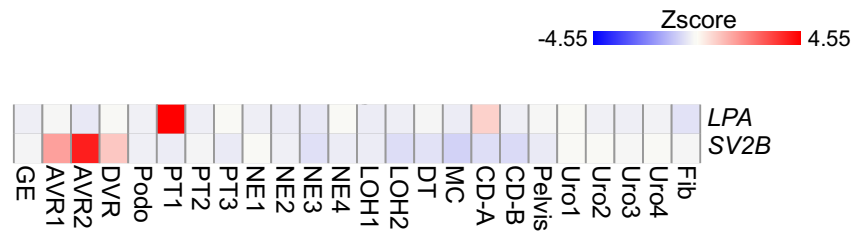


**Fig. S8. The eQTL(cis) sharing across cell types in tubuli and glomeruli.** a) About 18% of the identified eQTL(cis) showed cell type specificity across the 23 highly correlated kidney cell types. Cell type specific eQTL(cis) was defined as the eQTL(cis) with LFSR < 0.05 only in a single cell type. b) Pairwise eQTL(cis) sharing estimated by *mash*<sup>37</sup> of eQTL(cis) across kidney cell types. c) Pairwise eQTL(cis) sharing estimated by *mash*<sup>37</sup> of eQTL(cis) across immune cell types. The proportion of the shared (same sign and similar magnitude (effect within a factor of 0.5)) significant eQTL(cis) (LFSR<0.05) between two cell types. EndoG: glomerular endothelial cells, Endo: endothelial cells, Podo: podocyte, PT: proximal tubule, ALOH: ascending loop of Henle, DCT: distal convoluted tubule, IC-B: beta intercalated cells, CNT: connecting tubule, CD4T: CD4 T cells, CD8T: CD8 T cells, DC: CD11b+ dendritic cells, Granul: granulocyte, Macro: macrophage, Th17: T helper 17 cells, Treg: regulatory T cells, NK: natural killer cells. Tub: tubule, Glom: glomeruli.

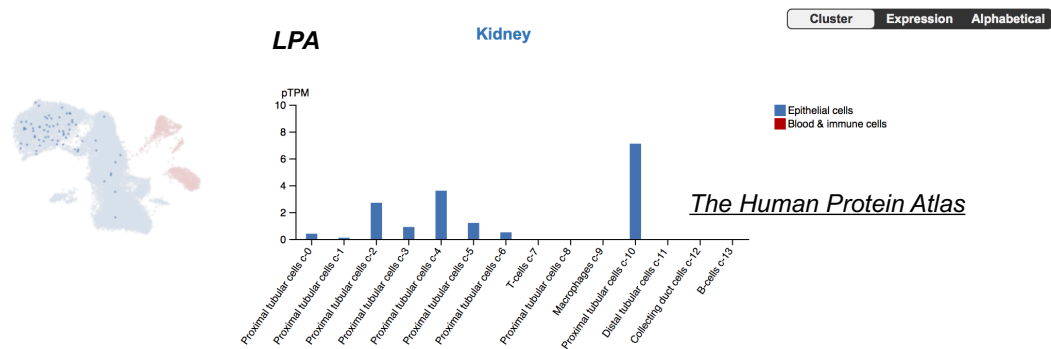


## Supplementary figure 9

**a**

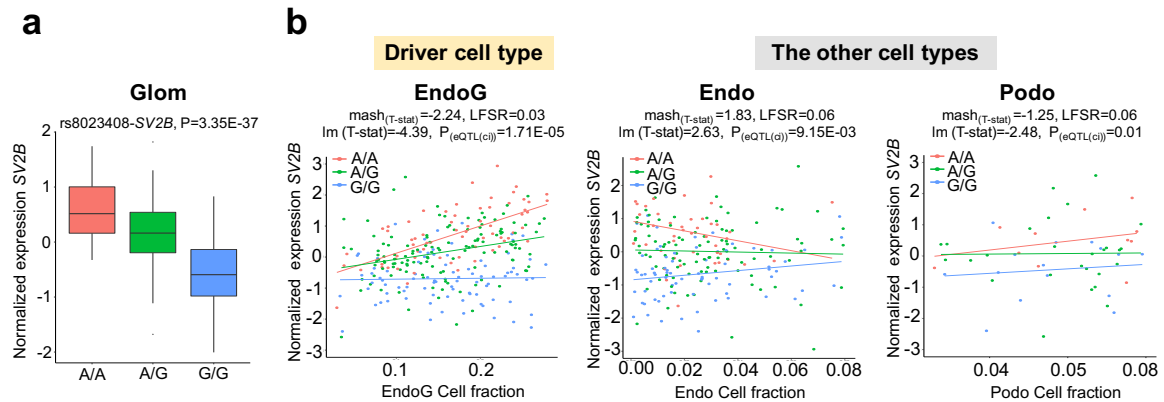


**b**



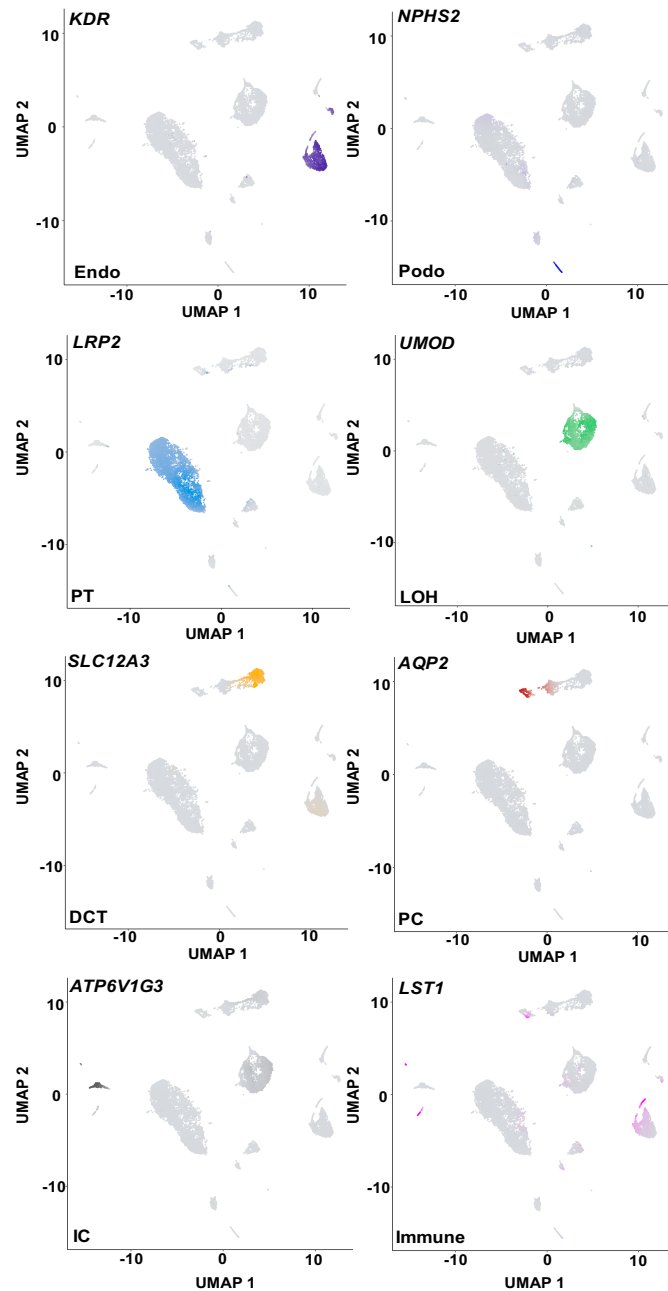
**Fig. S9. *LPA* and *SV2B* expression in a) human scRNA-seq data<sup>15</sup> b) and the Human Protein Atlas (<https://www.proteinatlas.org>).**

## Supplementary figure 10



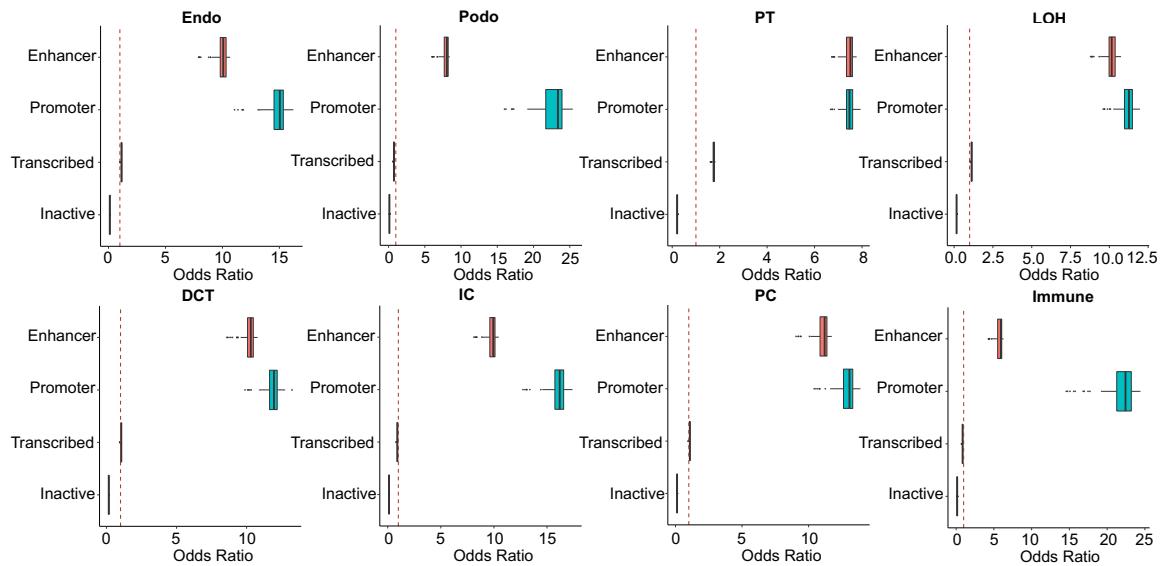
**Fig. S10. Defining cell-type dependent activities of genetic variants on gene expression.** a) The association between genotype of SNP rs8023408 and gene expression of *SV2B* in human glomeruli (N=303 samples). The y-axis is the normalized expression of *SV2B* in human kidney glomeruli, the x-axis is the genotype at rs8023408 locus. Center lines show the medians; box limits indicate the 25<sup>th</sup> and 75<sup>th</sup> percentiles; whiskers extend to the 5<sup>th</sup> and 95<sup>th</sup> percentiles; outliers are represented by dots. Two-sided P-value was calculated by linear regression eQTL(cf) model. b) EndoG cell-type-specificity of this eQTL(ci) across in human glomeruli (N=303 samples). The y-axis is the normalized expression of *SV2B* in human kidney glomeruli, the x-axis is the cell fraction for each specific kidney function cell type in subjects with genotype A/A (red), A/G (green) and G/G (blue) at rs8023408 locus. Each data point represents for an individual.  $lm(T-stat)$ : T-statistics calculated from linear model of eQTL(ci). Two-sided P-value was calculated by linear regression eQTL(ci) model.  $Mash_{(T-stat)}$ : T-statistics was estimated by *mash*. LFSR: local false sign rate estimated by *mash*. EndoG: glomerular endothelial cell, Endo: endothelial, Podo: podocyte.

## Supplementary figure 11



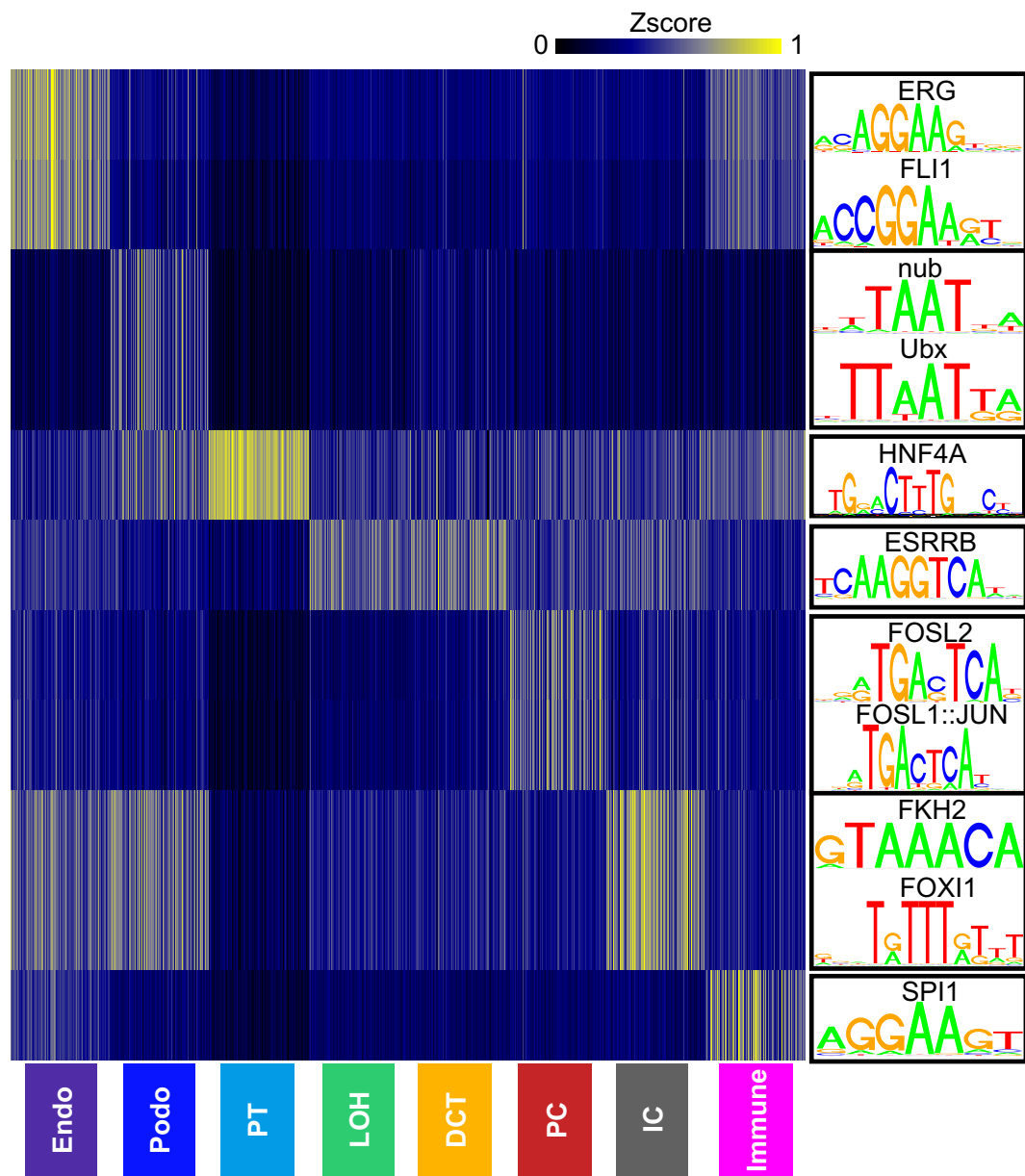
**Fig. S11. Feature plots of open chromatin information for known cell-type specific marker genes<sup>17,38</sup>.** Endo: endothelial cells, Podo: podocyte, PT: proximal tubule, LOH: loop of Henle, DCT: distal convoluted tubule, PC: collecting duct principal cells, IC: collecting duct intercalated cells, Immune: immune cells.

## Supplementary figure 12



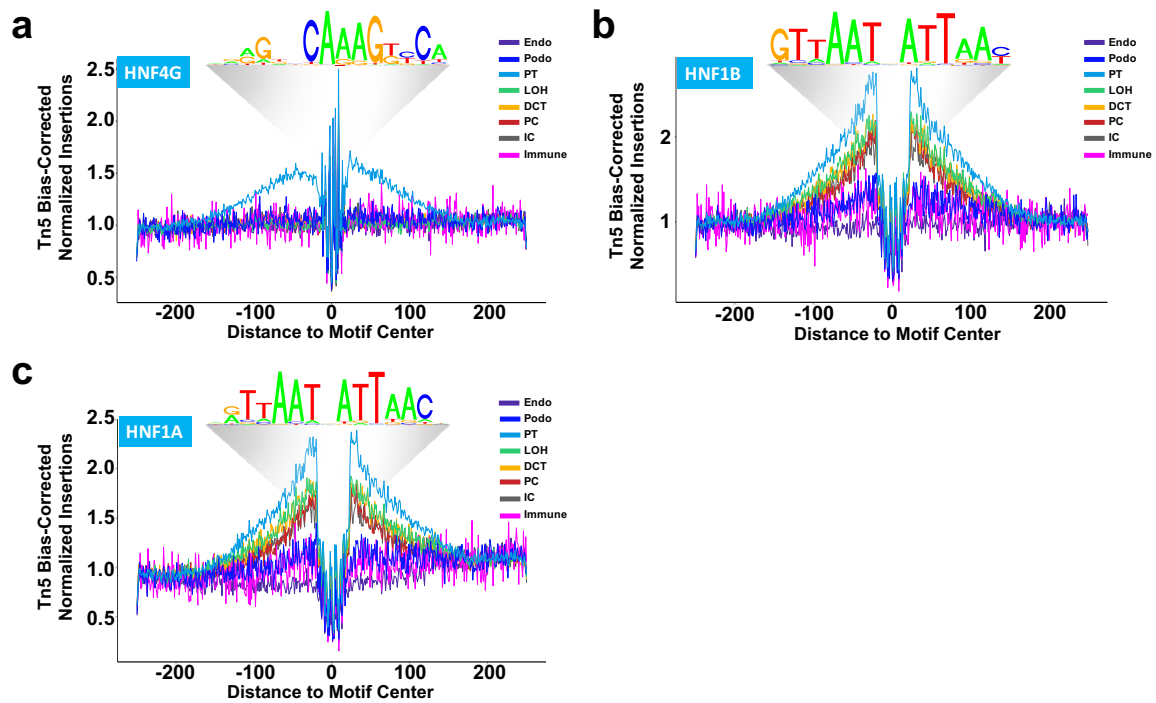
**Fig. S12. Functional annotation of human kidney snATAC-Seq data.** The enrichment (Odds Ratio, x-axis) of snATAC-Seq open chromatin peaks in promoters, enhancers, transcribed and inactive regions (y-axis). Center lines show the medians; box limits indicate the 25<sup>th</sup> and 75<sup>th</sup> percentiles; whiskers extend to the 5<sup>th</sup> and 95<sup>th</sup> percentiles; outliers are represented by dots. Endo: endothelial cells, Podo: podocyte, PT: proximal tubule, LOH: loop of Henle, DCT: distal convoluted tubule, PC: collecting duct principal cells, IC: collecting duct intercalated cells, Immune: immune cells. N=12,720 cells.

## Supplementary figure 13



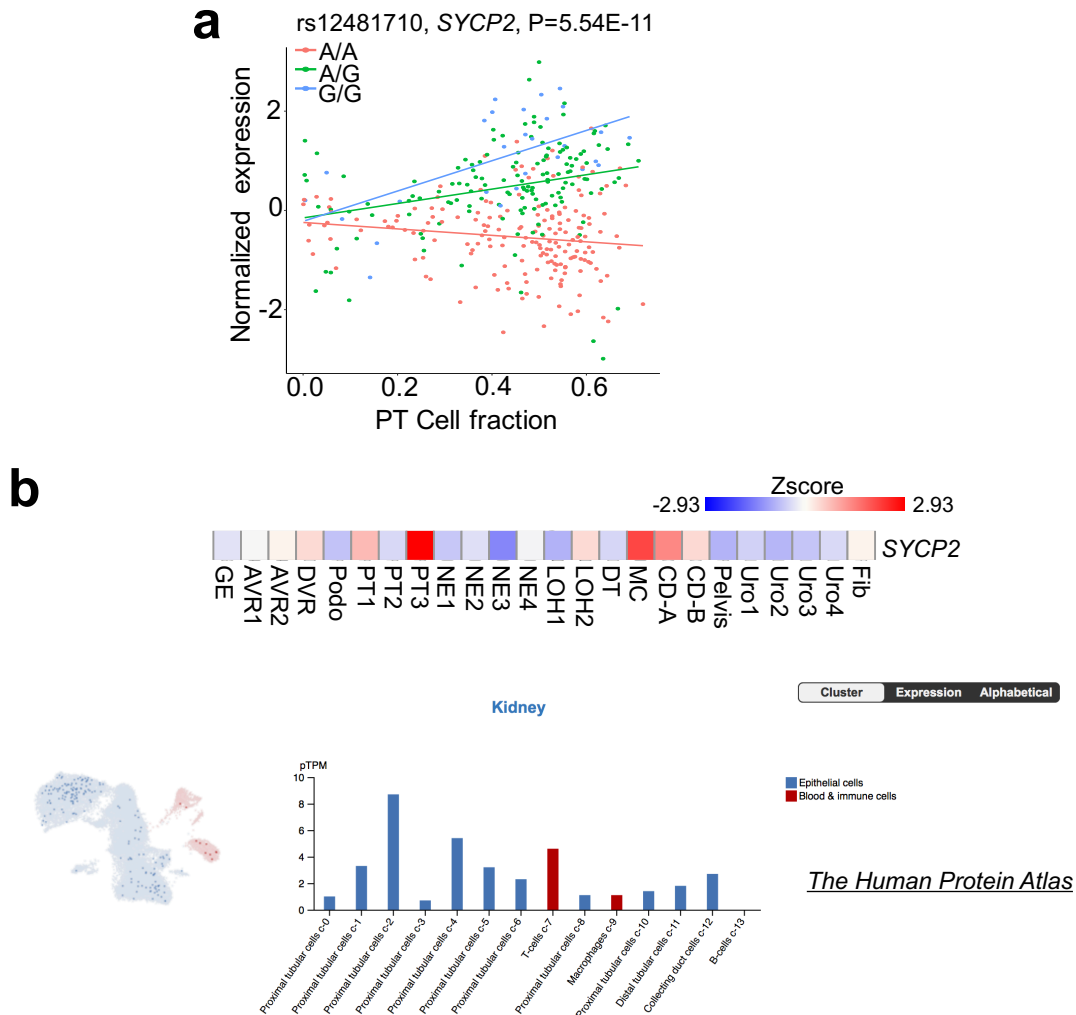
**Fig. S13. Motif enrichment analysis of the cell type-specific differential accessibility peaks (DAPs).** Each row represents the top enriched motif for each cell type. Each column represents 100 randomly selected single cells. Endo: endothelial cells, Podo: podocyte, PT: proximal tubule, LOH: loop of Henle, DCT: distal convoluted tubule, PC: collecting duct principal cells, IC: collecting duct intercalated cells, Immune: immune cells.

## Supplementary figure 14



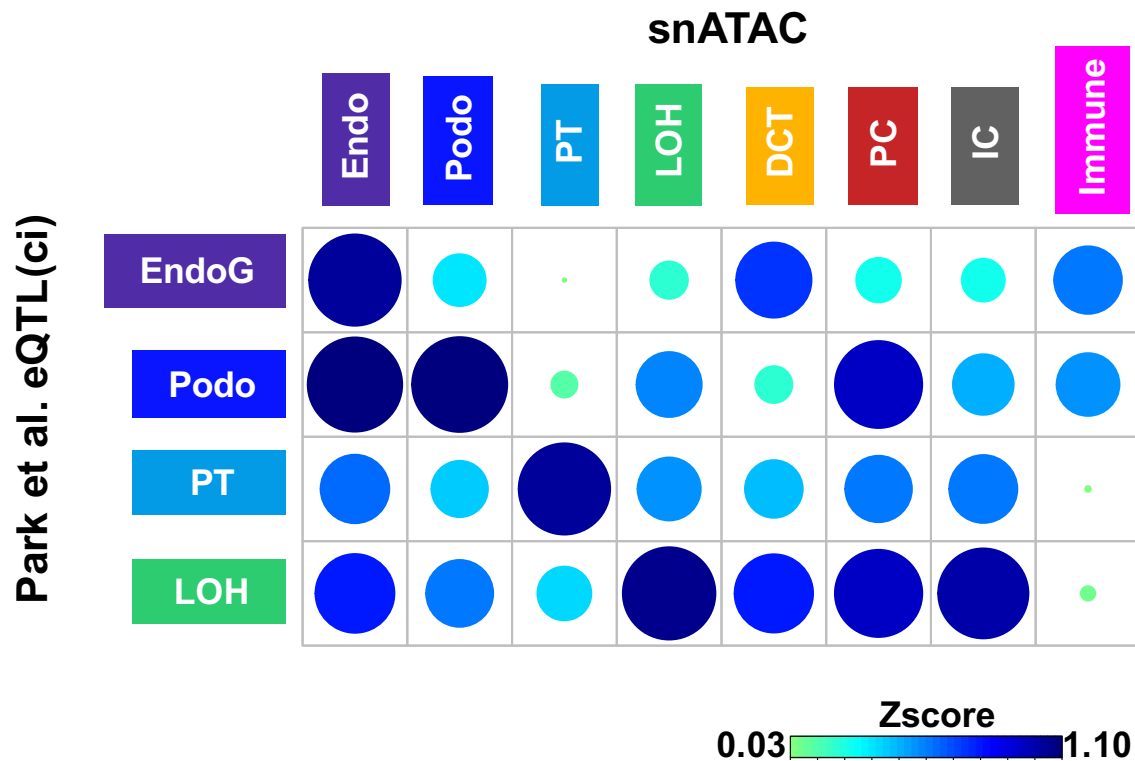
**Fig. S14. Footprinting analysis of the HNF4G (a), HNF1B (b) and HNF1A (c) transcription factors across the 8 major kidney cell types.** The motif logos are shown above. Endo: endothelial cells, Podo: podocyte, PT: proximal tubule, LOH: loop of Henle, DCT: distal convoluted tubule, PC: collecting duct principal cells, IC: collecting duct intercalated cells, Immune: immune cells.

## Supplementary figure 15



**Fig. S15. SYCP2 eQTL(ci) and gene expression information.** a) SYCP2 eQTL(ci) data. The y-axis is the relative expression of SYCP2 in human kidney tubules (N=356 samples), the x-axis is the PT cell fraction colored by subjects with (rs12481710) genotype A/A (red), A/G (green) and G/G (blue). Two-sided P-value was calculated by eQTL(ci) model. Each data point represents a single subject. The y-axis is the normalized expression of SYCP2 in human tubules (N=356 samples), the x-axis is the cell fraction colored by subjects with genotype A/A (red), A/G (green) and G/G (blue) at the rs12481710 locus. b) Relative SYCP2 expression in human scRNA-seq data<sup>15</sup> and the Human Protein Atlas (<https://www.proteinatlas.org>).

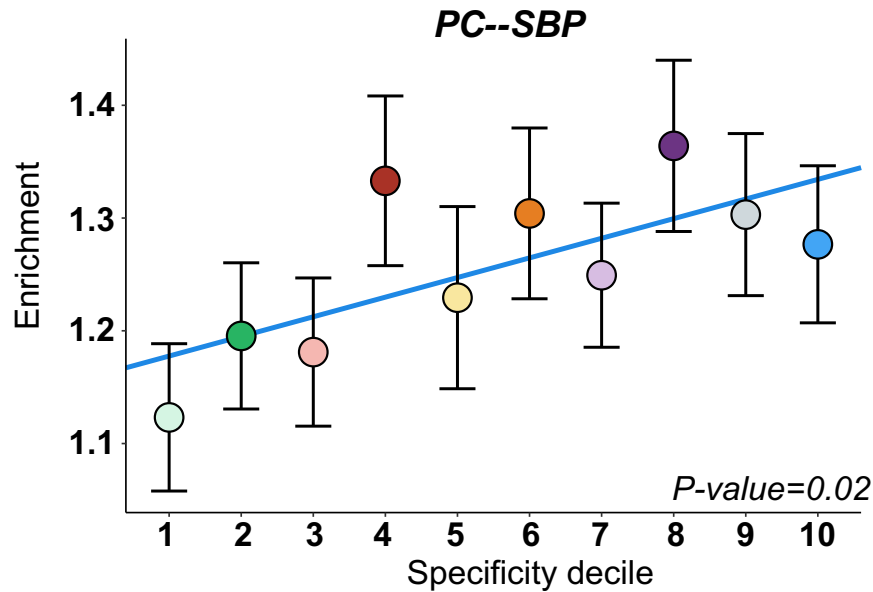
Supplementary figure 16



**Fig. S16. snATAC-Seq enrichment of the validation eQTL(ci)s data (using Park et al. as input single cell data)<sup>17</sup>.** Each column represents human kidney snATAC-seq chromatin data and each row represents eQTL(ci) using Park et al. as input single cell data. The size and the color of the bubble indicates enrichment (Z-score). EndoG: glomerular endothelial cells, Endo: endothelial cells, Podo: podocyte, PT: proximal tubule, LOH: loop of Henle, DCT: distal convoluted tubule, PC: collecting duct principal cells, IC: collecting duct intercalated cells, Immune: immune cells.

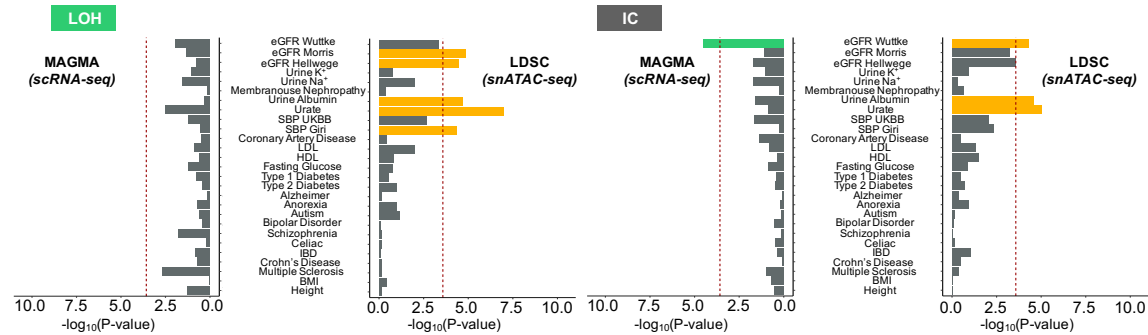


### Supplementary figure 17



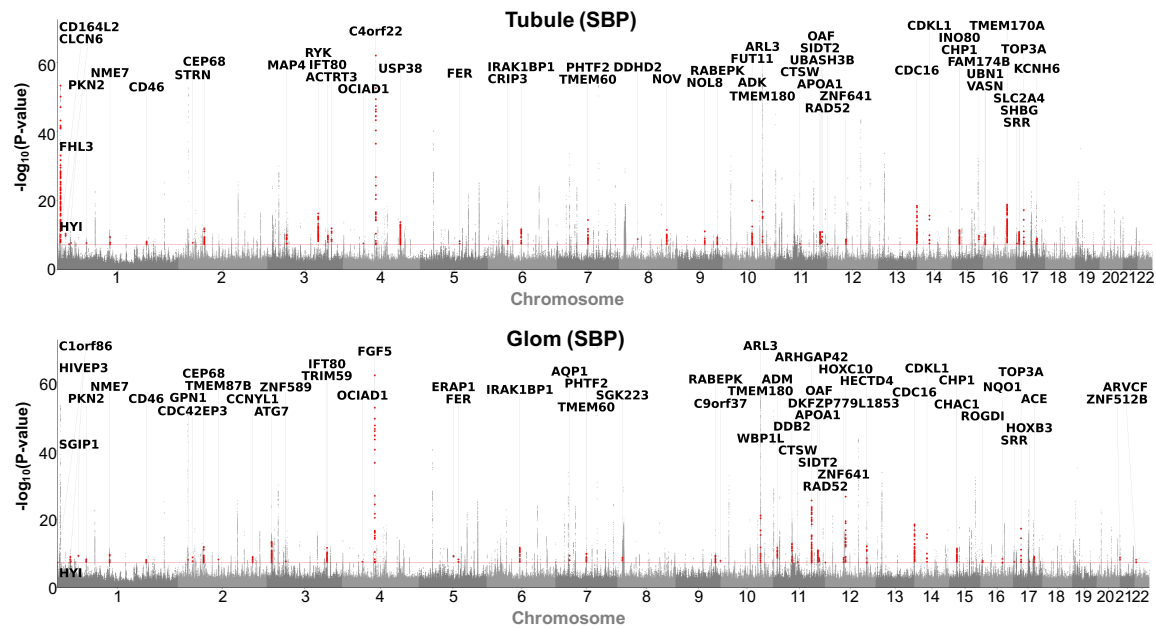
**Fig. S17. Enrichment of SBP-SNP heritability in each of the specificity deciles for collecting duct principal (PC) cells calculated using LDSC from kidney scRNA-seq data<sup>17</sup>.** X-axis is the gene expression specificity decile, y-axis is the enrichment value calculated by LDSC. Colors of dots represent different specificity deciles. Blue line shows the linear regression slope fitted to the enrichment values. Error bars indicate the 95% confidence intervals. Data are presented as mean values +/- SE. N=477,054 individuals for SBP GWAS; N=57,979 cells from healthy mouse kidneys.

## Supplementary figure 18



**Fig. S18. Enrichment of common GWAS variants in kidney scRNA-seq data<sup>17</sup>** using MAGMA (the left panels) and human kidney snATAC-seq data using LDSC (the right panels). Y-axis represents 24 different GWAS traits and x-axis is the  $-\log_{10}(\text{P-value})$ . Two-sided P-value was estimated by MAGMA (the left panels) and LDSC (the right panels). LOH: loop of Henle, IC: collecting duct intercalated cells. The left panels show MAGMA results of mouse kidney scRNA-seq data. Green: P-value <  $2.3\text{E}-04$  ( $0.05/(27 \times 8)$ , Bonferroni-corrected cutoff (27 GWAS data and 8 cell types)); grey: not significant. The right panels show LDSC results of human kidney snATAC-seq data. Yellow: P-value <  $2.3\text{E}-04$ .

## Supplementary figure 19



**Fig. S19. Miami plot showing coloc nominated target genes for the SBP GWAS associated loci identified by eQTL(cf) model in tubules and glomeruli across the genome. X-axis represents the chromosomal location. Y-axis represents the GWAS significance  $-\log_{10}(\text{P-value})$ . Two-sided P-value was obtained from the GWAS study<sup>39</sup>.**

Supplementary figure 20

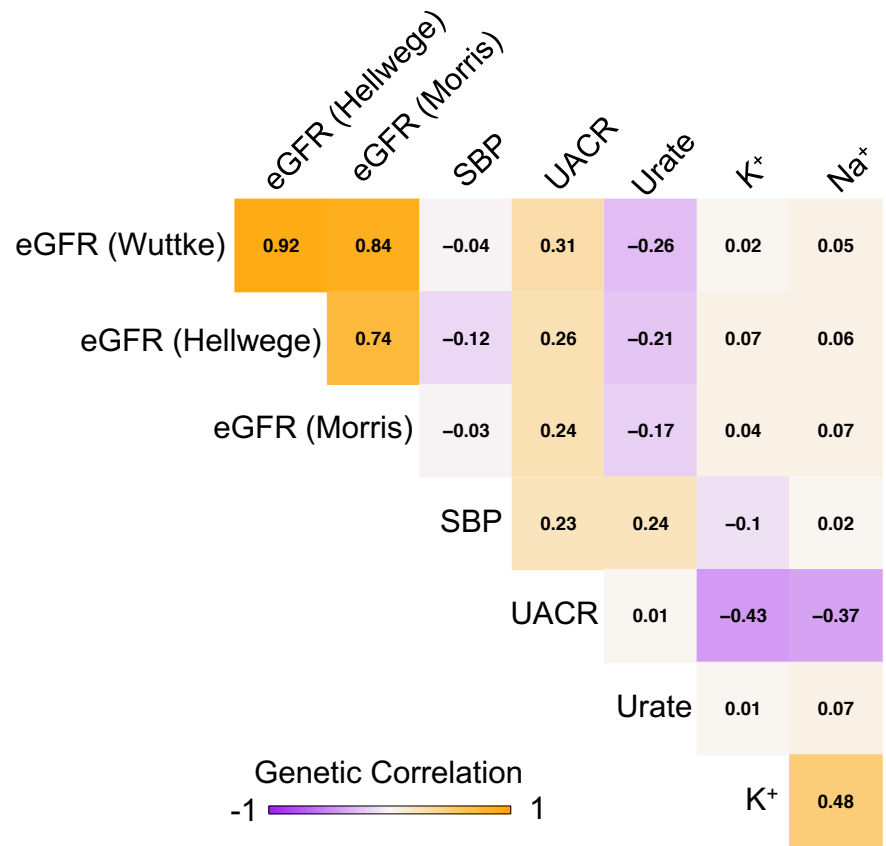
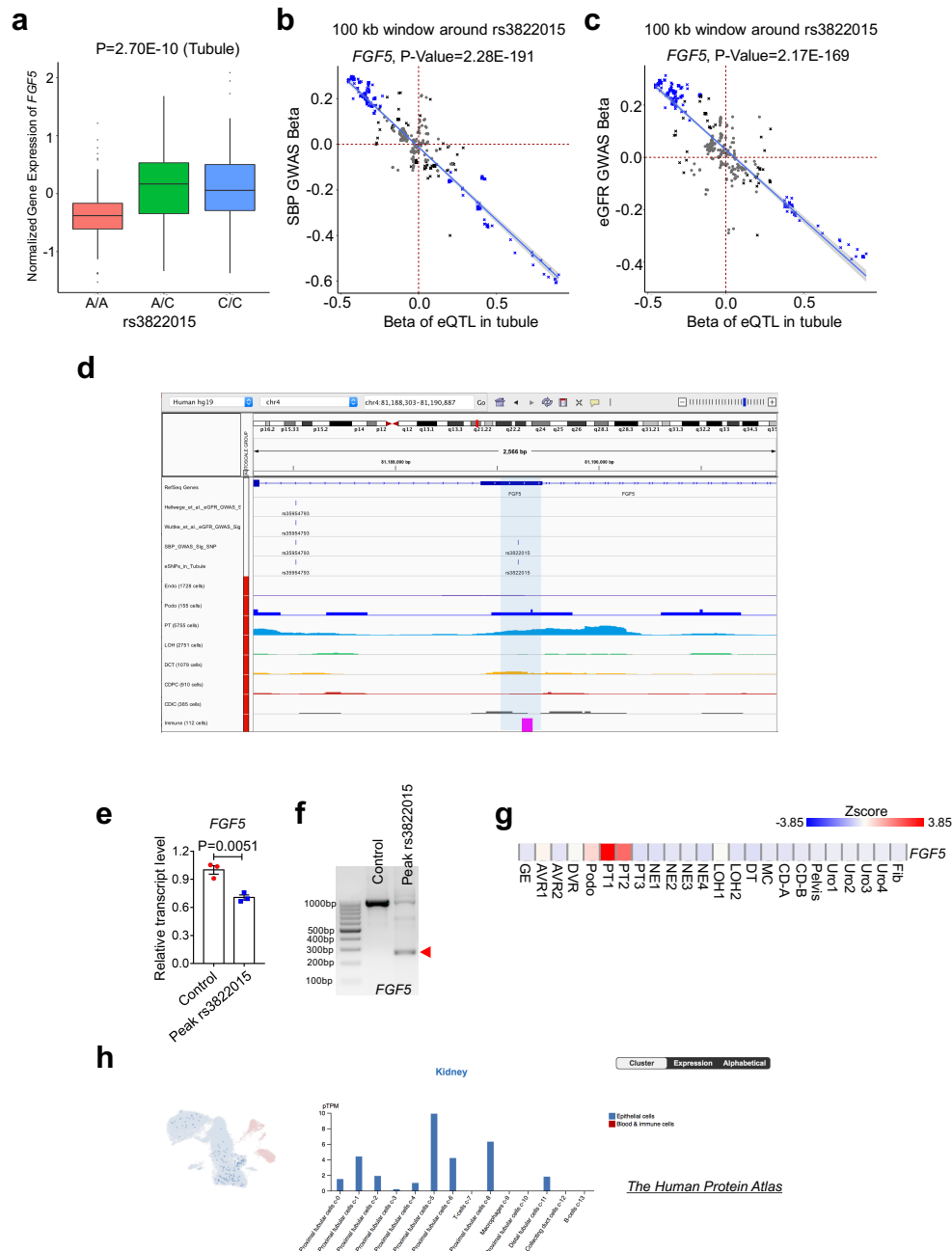


Fig. S20. Pairwise genetic correlation of 6 kidney related GWAS endophenotypes.

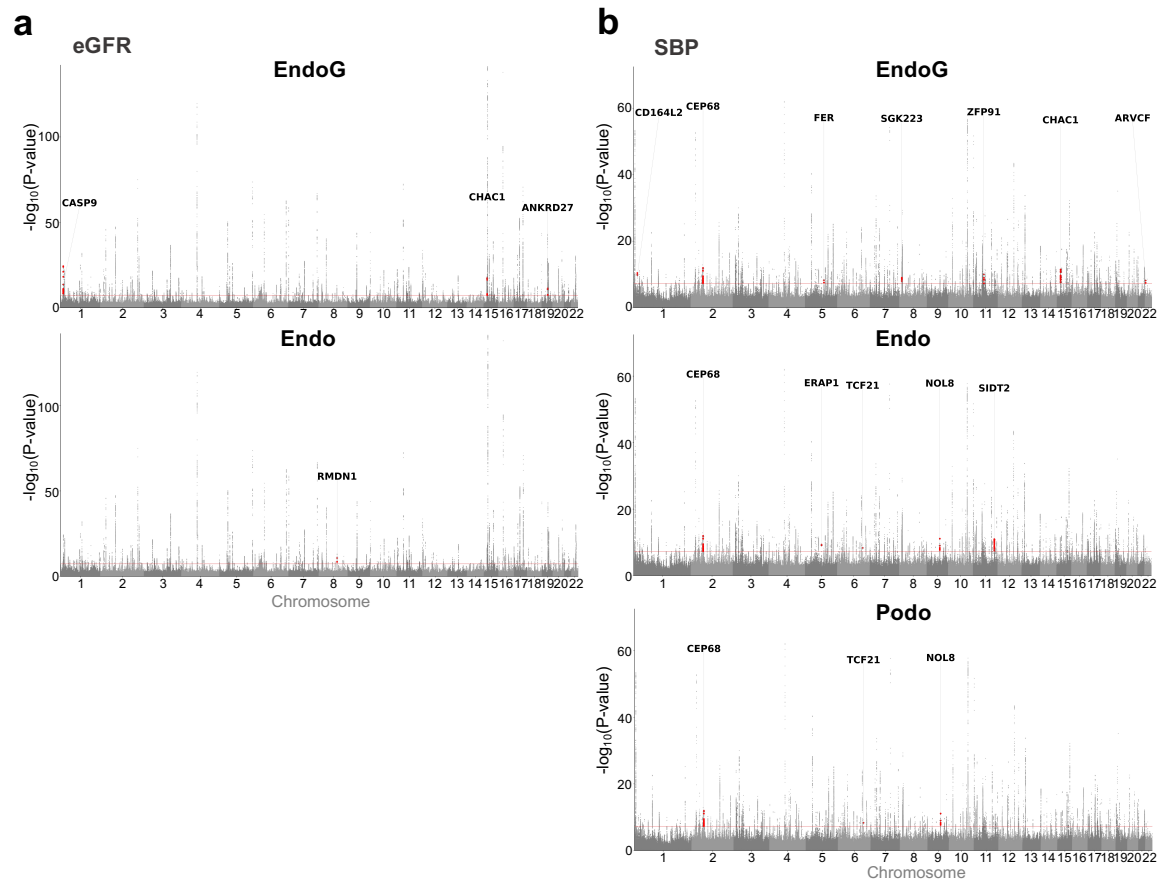
## Supplementary figure 21



**Fig. S21. *FGF5* was prioritized as a causal gene for kidney function and systolic blood pressure traits.** a) The association between rs3822015 genotype and *FGF5* expression in 356 human tubule samples. Two-sided P-value was calculated by linear regression eQTL(cf) model. Center lines show the medians; box limits indicate the 25<sup>th</sup> and 75<sup>th</sup> percentiles; whiskers extend to the 5<sup>th</sup> and 95<sup>th</sup>

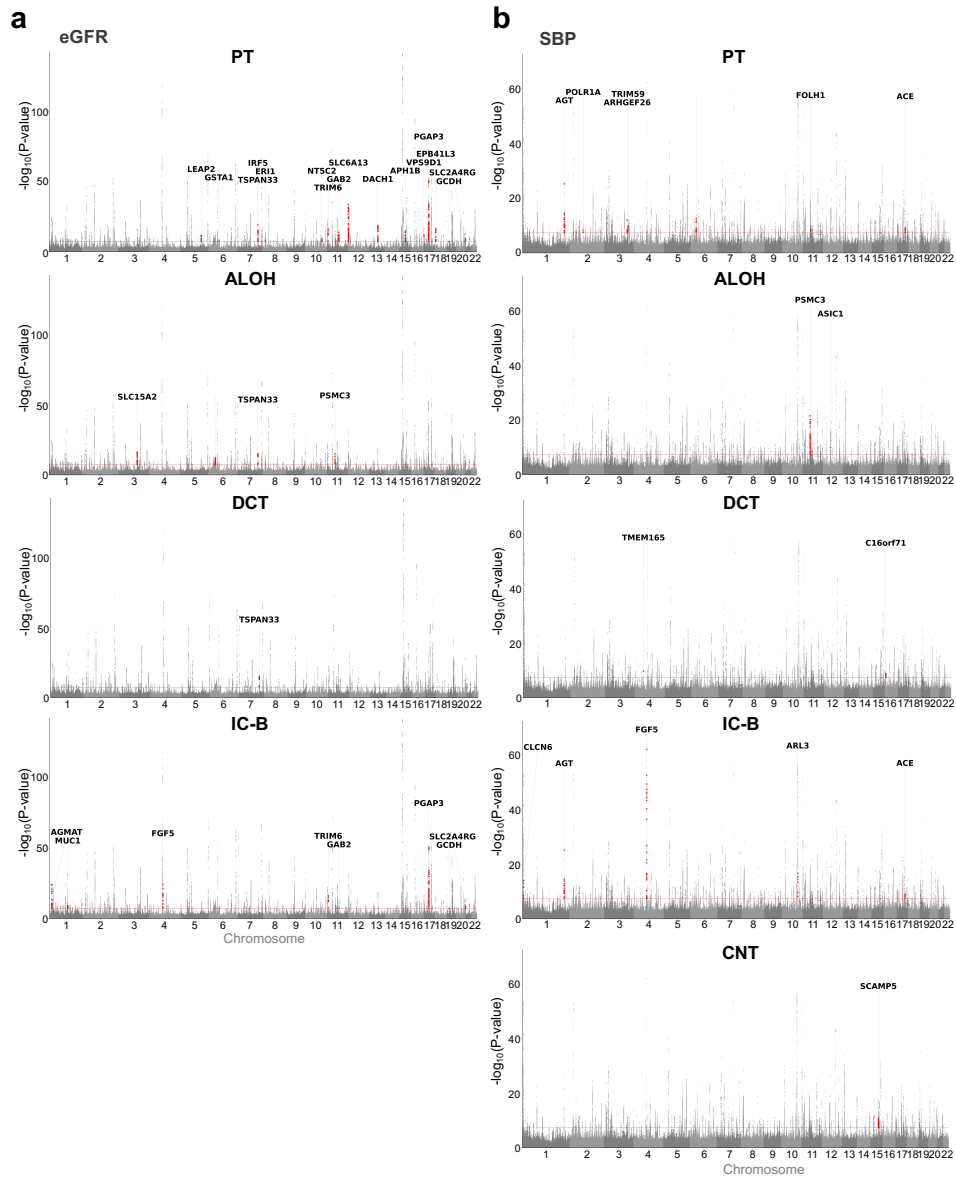
percentiles; outliers are represented by dots. N=356 human tubule samples. b) Variant effect size and direction for SBP GWAS and *FGF5* expression changes (eQTL(cf)), located within +/- 100 kb of rs3822015 significantly correlated. Each data point represents a single SNP. The data point is presented as a 'blue cross', if the two-sided P-values of both eQTL and GWAS were < 0.05. (N=356 for eQTL in tubule; N=477,054 SBP GWAS<sup>39</sup>). Two-sided P-value was calculated by linear regression. The error band indicates the 95% confidence intervals. c) The variant effect sizes of eGFR GWAS, and *FGF5* expression changes (eQTL(cf)), located within +/- 100 kb of rs3822015 are significantly correlated. Each data point represents a single SNP. The data points are shown as a 'cross', if the two-sided P-values of both eQTL and GWAS were < 0.05, the data points are shown in 'blue'. (N=356 for eQTL; N=292,258 for eGFR GWAS<sup>40</sup>). Two-sided P-value was calculated by linear regression. The error band indicates the 95% confidence intervals. d) SBP GWAS significant SNP rs3822015-overlapped PT specific open chromatin region. Single cell chromatin accessibility of human kidney snATAC-seq data. Endo: endothelial cells, Podo: podocyte, PT: proximal tubule, LOH: loop of Henle, DCT: distal convoluted tubule, PC: collecting duct principal cells, IC: collecting duct intercalated cells, Immune: immune cells. The region targeted by CRISPR-mediated gene deletion is highlighted in blue. e) Relative transcript levels of *FGF5* following genomic deletion. Data is represented as mean +/- SEM. P-value was calculated by two-sided t-test. P<0.05 is significant. N=3 independent experiments. Source data. f) Gel image of genomic PCR confirming CRISPR-mediated deletion. The experiment was repeated 3 times and a representative image is shown. g) *FGF5* expression in human scRNA-seq data<sup>15</sup>, and h) the Human Protein Atlas (<https://www.proteinatlas.org>).

## Supplementary figure 22



**Fig. S22. Miami plot showing eQTL(ci) coloc prioritized genes for eGFR (a) and SBP (b) GWAS.** X-axis represents the chromosomal location. Y-axis represents the GWAS significance  $-\log_{10}(P\text{-value})$ . EndoG: glomerular endothelial cell, Endo: endothelial cells, Podo: podocyte. a) Two-sided P-value obtained from the eGFR GWAS<sup>41</sup> and b) two-sided P-value obtained from the SBP GWAS were plotted<sup>39</sup>.

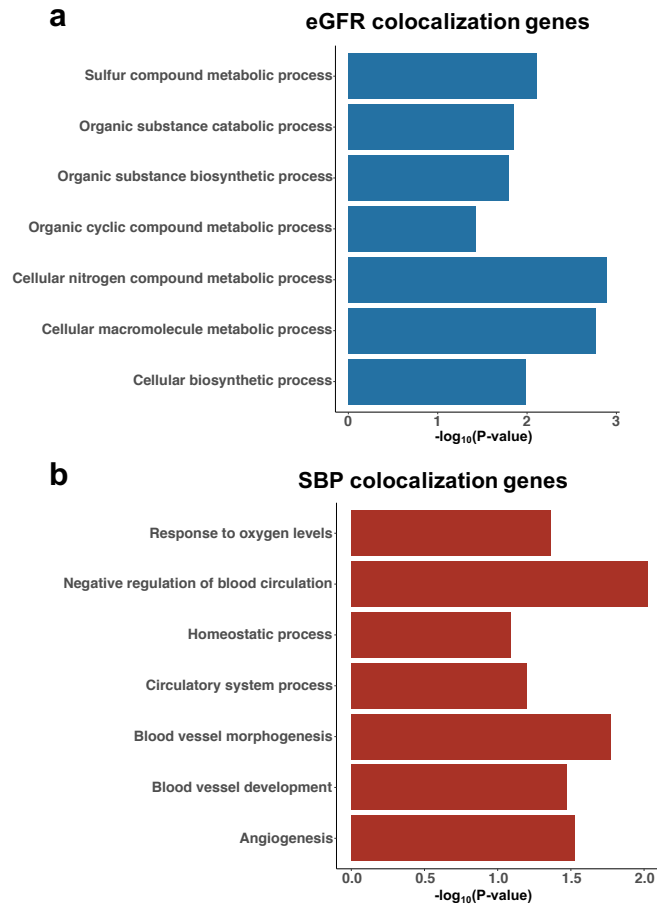
## Supplementary figure 23



**Fig. S23. Miami plot showing eQTL(ci) coloc prioritized genes for eGFR (a) and SBP (b) GWAS.** X-axis represents the chromosomal location. Y-axis represents the GWAS significance  $-\log_{10}(\text{P-value})$ . PT: proximal tubule, ALOH: ascending loop of Henle, DCT: distal convoluted tubule, IC-B: beta intercalated cells, CNT: connecting tubule. a) Two-sided P-value obtained from the eGFR GWAS<sup>41</sup> and b) two-sided P-value obtained from the SBP GWAS were plotted<sup>39</sup>.

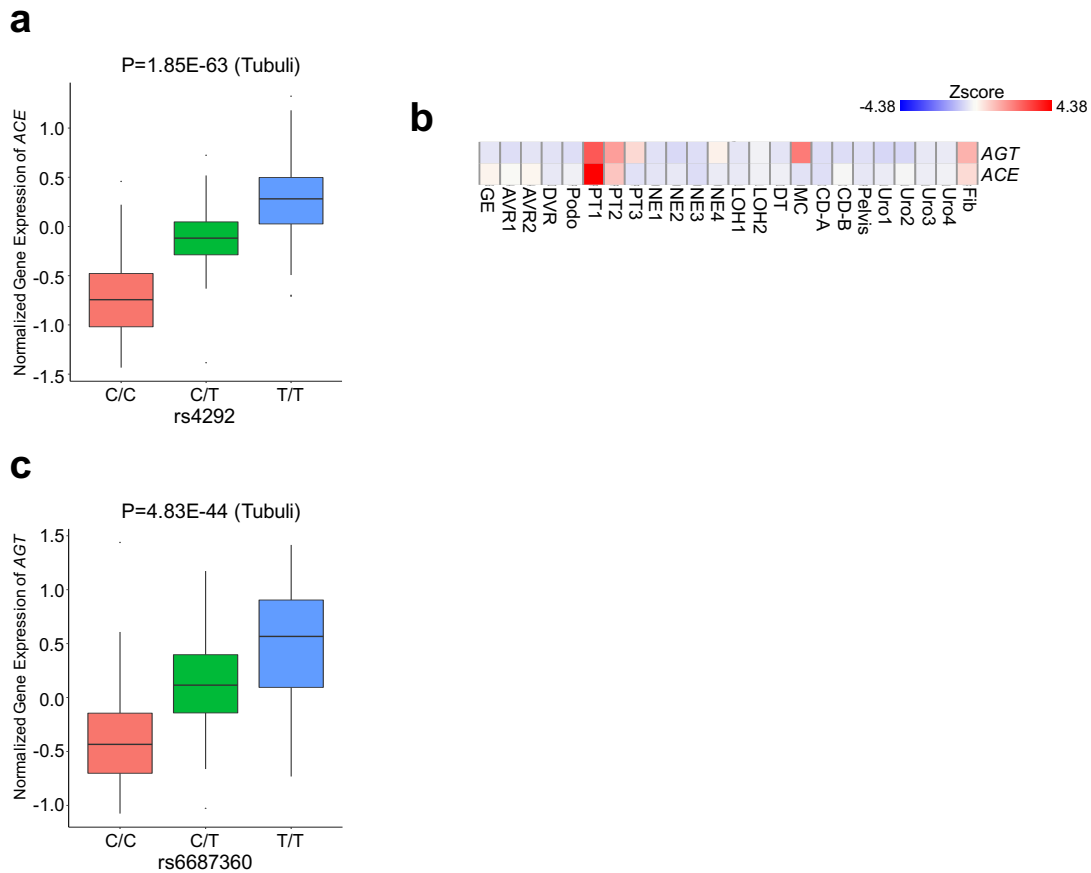


## Supplementary figure 24



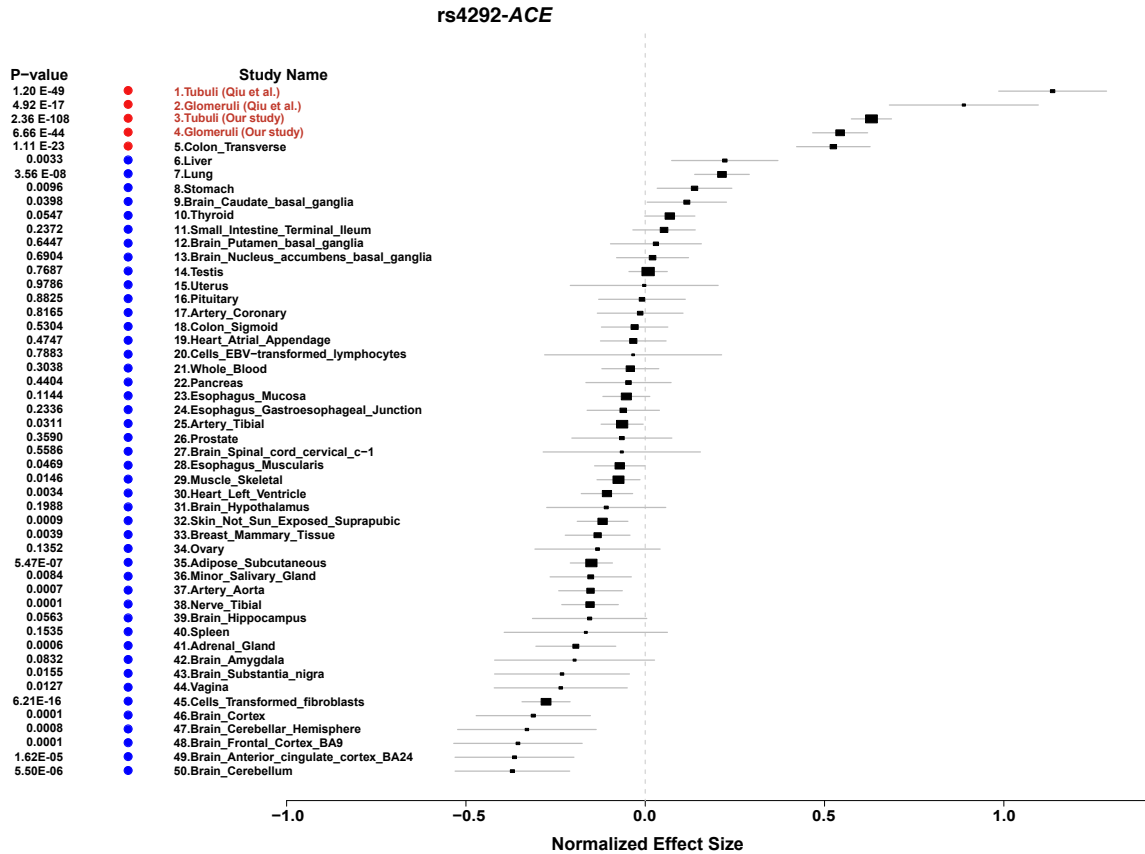
**Fig. S24. Functional annotation using gene ontology (DAVID) of eGFR<sup>41</sup> (a) and SBP<sup>39</sup> (b) GWAS prioritized genes by eQTL colocalization. a), b) Two-sided P-value was calculated by a modified Fisher's exact test (EASE) using DAVID<sup>42</sup>.**

## Supplementary figure 25



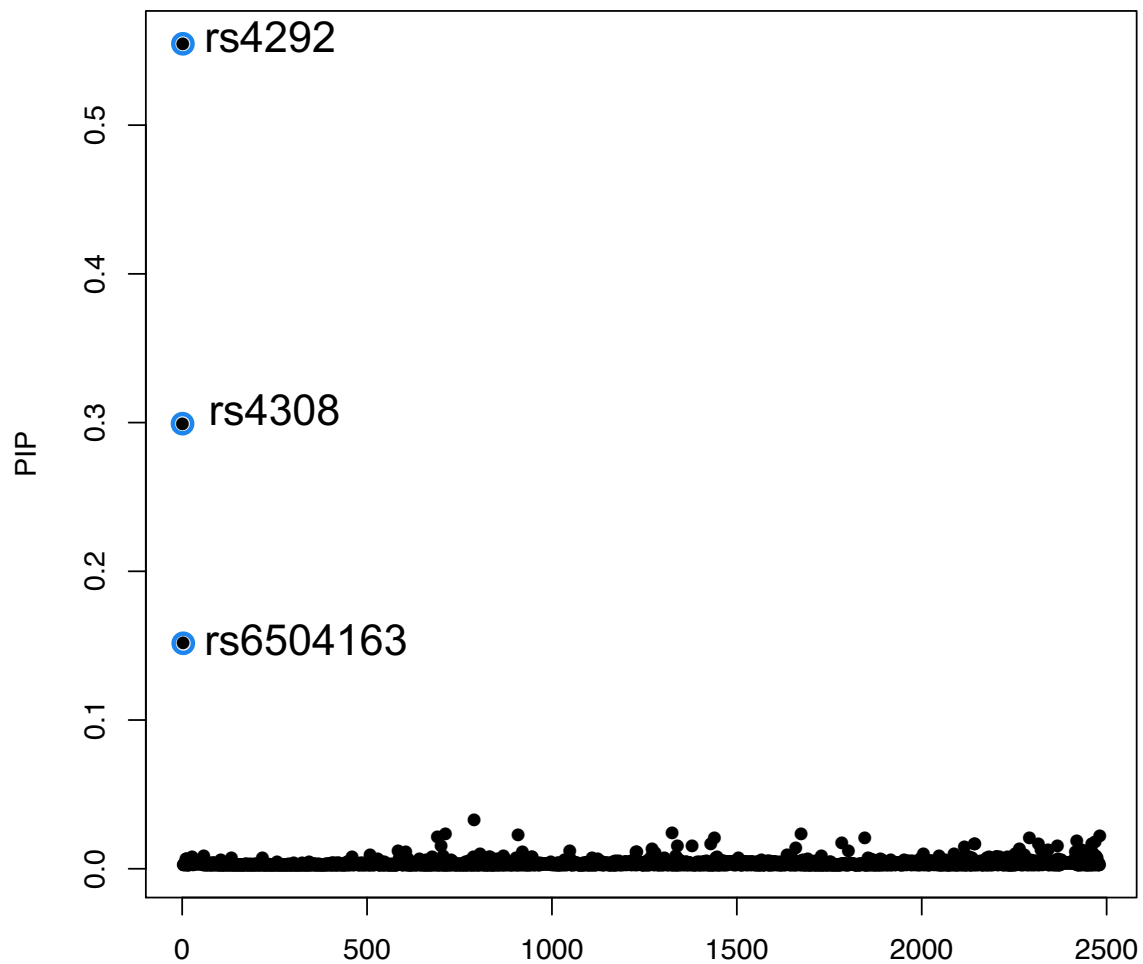
**Fig. S25. Human kidney tubule eQTL(cf)s of *ACE* and *AGT*.** a) The association between rs4292 genotype and expression of *ACE* in 356 human tubule samples. Two-sided P-value was calculated by linear regression eQTL(cf) model. Center lines show the medians; box limits indicate the 25<sup>th</sup> and 75<sup>th</sup> percentiles; whiskers extend to the 5<sup>th</sup> and 95<sup>th</sup> percentiles; outliers are represented by dots. N=356 human tubule samples. b) *ACE* and *AGT* expression in human scRNA-seq data<sup>15</sup>. c) The association between genetic variant rs6687360 and expression of *AGT* in 356 human tubule samples. Two-sided P-value was calculated by linear regression eQTL(cf) model. Center lines show the medians; box limits indicate the 25<sup>th</sup> and 75<sup>th</sup> percentiles; whiskers extend to the 5<sup>th</sup> and 95<sup>th</sup> percentiles; outliers are represented by dots. N=356 human tubule samples.

## Supplementary figure 26



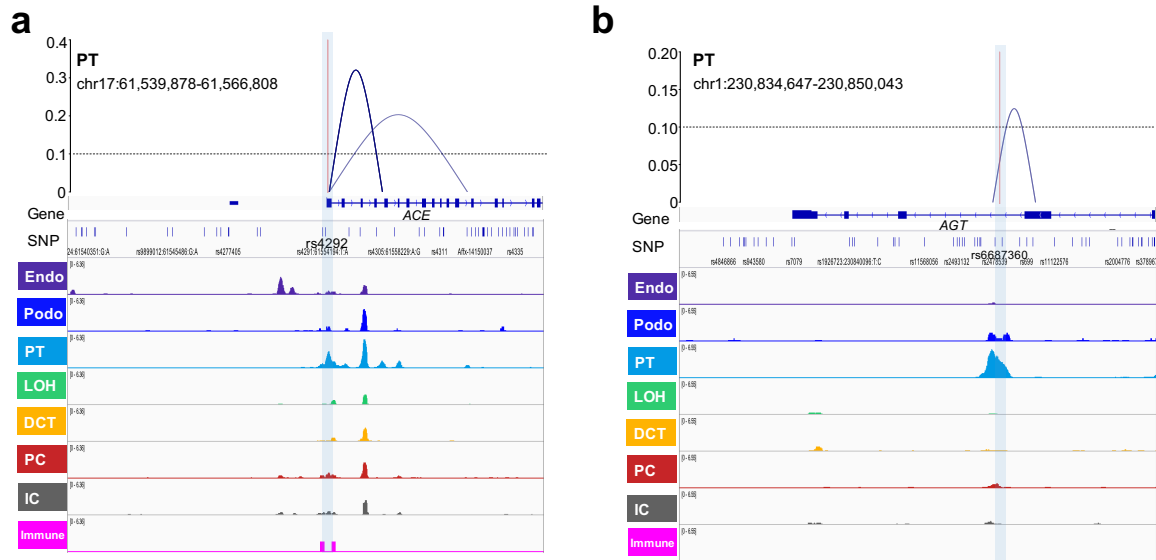
**Fig. S26. Meta-analysis of the rs4292-ACE eQTL in the kidney and 46 GTEx tissues.** The normalized effect size ( $\beta$ ) of eQTL association between rs4292 and ACE expression in kidney tubules (n=356 for this study and n=121 for study of Qiu et al.<sup>36</sup>) and glomeruli (n=303 and n=121 for glomerular of Qiu et al.) and in 46 human tissues from GTEx (v7). The x-axis shows  $\beta$  with the 95% confidence intervals (CI). EBV, Epstein-Barr virus. Data are presented as mean values +/- SE. Two-sided P-value was calculated by linear regression from GTEx (v7)<sup>43</sup>.

### Supplementary figure 27



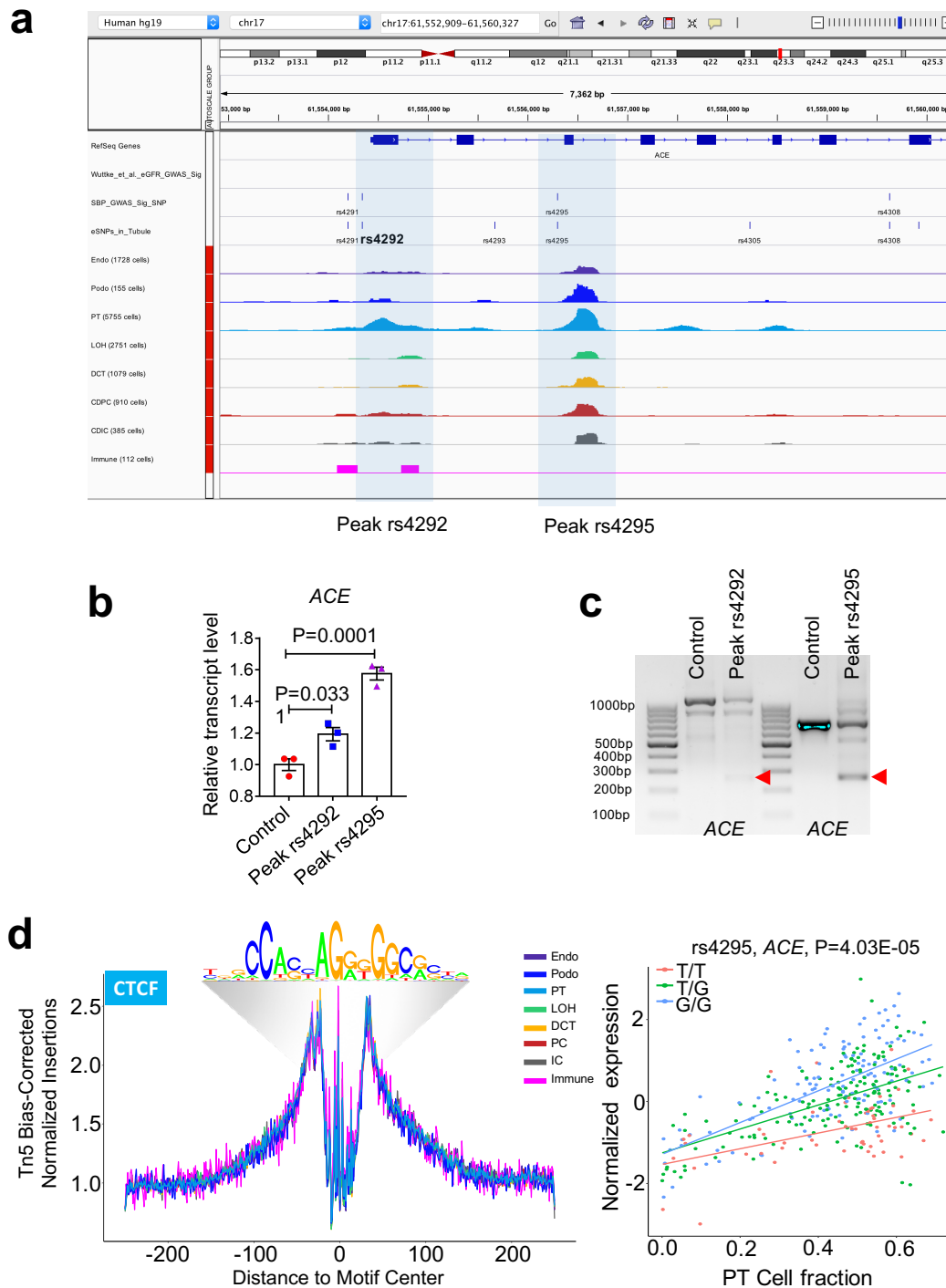
**Fig. S27. Fine mapping results of causal variants affecting *ACE* expression in human kidney tubule samples.** There are about 2,500 variants (x-axis) located within +/- 1 Mb region around the transcription start site (TSS) of *ACE*. Only three SNPs were identified as the potential causal variants for driving *ACE* expression changes in the tubule compartment. rs4292 is the top variant with the highest posterior probability (y-axis) (PIP=55.5%).

## Supplementary figure 28



**Fig. S28. Cicero-inferred co-accessibility between pairs of accessible chromatin sites at the a) *ACE* and b) *AGT* loci.** The genomic regions shown are a) chr17:61,539,878-61,566,808 and b) chr1:230,834,647-230,850,043. Single cell chromatin accessibility of human kidney snATAC-seq data. Endo: endothelial cells, Podo: podocyte, PT: proximal tubule, LOH: loop of Henle, DCT: distal convoluted tubule, PC: collecting duct principal cells, IC: collecting duct intercalated cells, Immune: immune cells.

## Supplementary figure 29

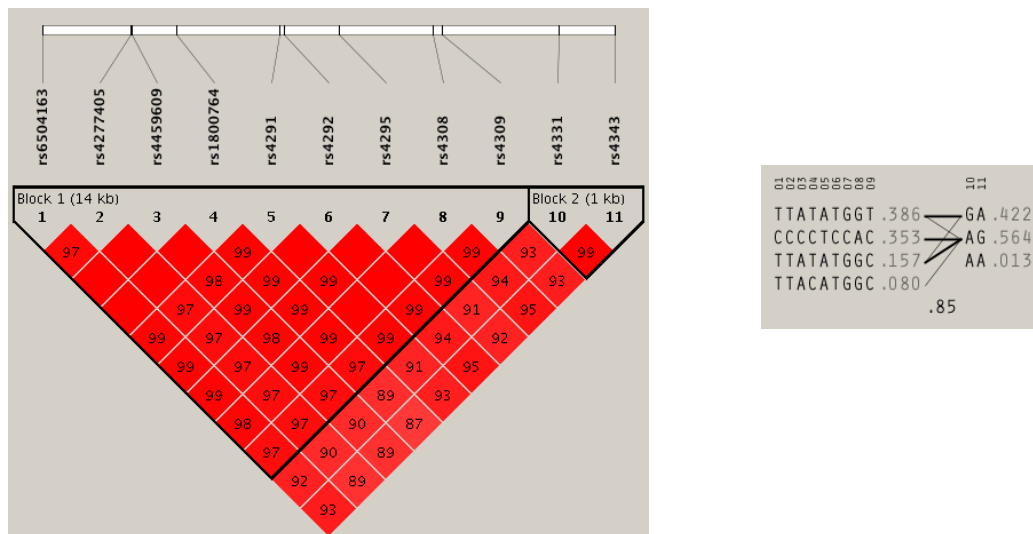


**Fig. S29. SBP GWAS significant SNPs and *ACE* expression.** a) SBP GWAS significant SNP rs4292- and rs4295- overlapped PT specific open chromatin regions. Single cell chromatin accessibility of human kidney snATAC-seq data.

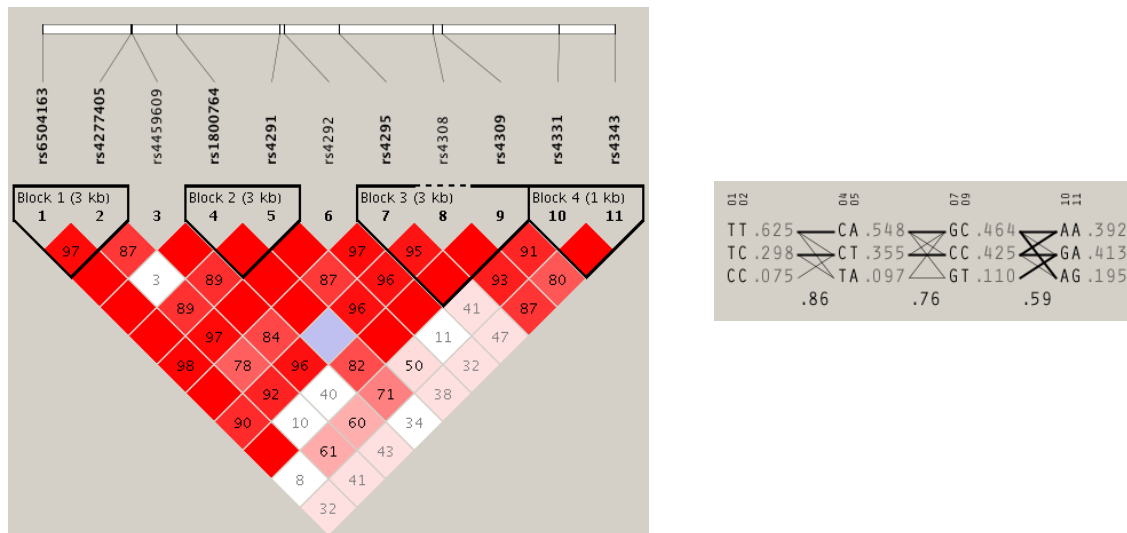
Endo: endothelial cells, Podo: podocyte, PT: proximal tubule, LOH: loop of Henle, DCT: distal convoluted tubule, PC: collecting duct principal cells, IC: collecting duct intercalated cells, Immune: immune cells. b) Relative transcript levels of *ACE* following CRISPR-Cas9 mediated deletion of genomic region containing GWAS variants (rs4292 and rs4295) and PT open chromatin region deletion. Data is represented as mean  $\pm$  SEM (standard error of the mean). P-value was calculated by two-sided t-test.  $P < 0.05$  is significant. N=3 independent experiments. Source data. c) Representative genomic PCR image showing the successful deletion of the genomic region. d) Left panel: footprinting analysis of the CTCF transcription factor across the 8 major cell types. The motif logos are shown above. Endo: endothelial cells, Podo: podocyte, PT: proximal tubule, LOH: loop of Henle, DCT: distal convoluted tubule, PC: collecting duct principal cells, IC: collecting duct intercalated cells, Immune: immune cells. Right panel: the y-axis is the relative expression of *ACE* in human kidney tubules (N=356 samples), the x-axis is the PT cell fraction colored by subjects with genotype T/T (red), T/G (green) and G/G (blue), respectively. Two-sided P-value was calculated for the eQTL(ci) model. Each data point represents a single subject. The y-axis is the normalized expression of *ACE* in human tubules (N=356 samples), the x-axis is the cell fraction colored by subjects with genotype T/T (red), T/G (green) and G/G (blue) at the rs4295 locus.

## Supplementary figure 30

### EUR Population (1000 Genomes Project)



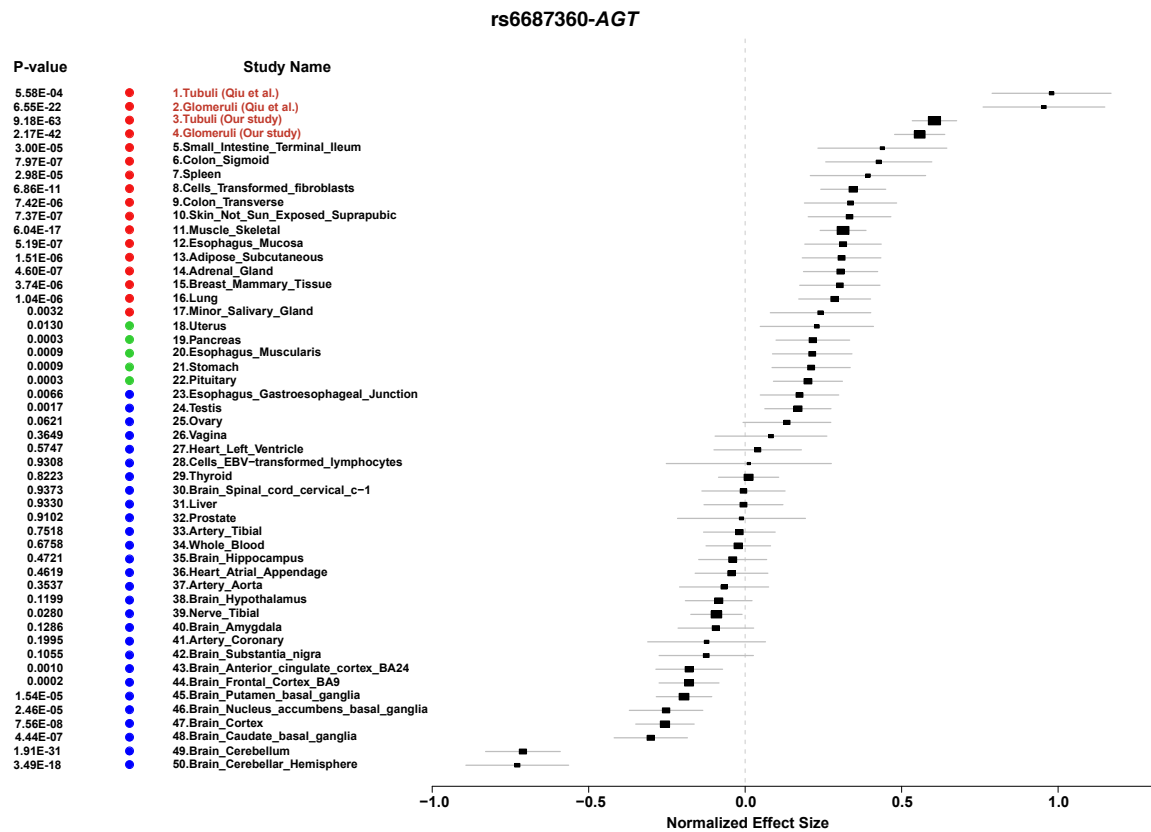
### AFR Population (1000 Genomes Project)



**Fig. S30.** LD block structure and haplotype frequencies at the *ACE* locus are different in European (EUR) and African American (AFR) population in the 1,000 Genomes Project<sup>3</sup>. Plots were generated using Haploview<sup>44</sup>.

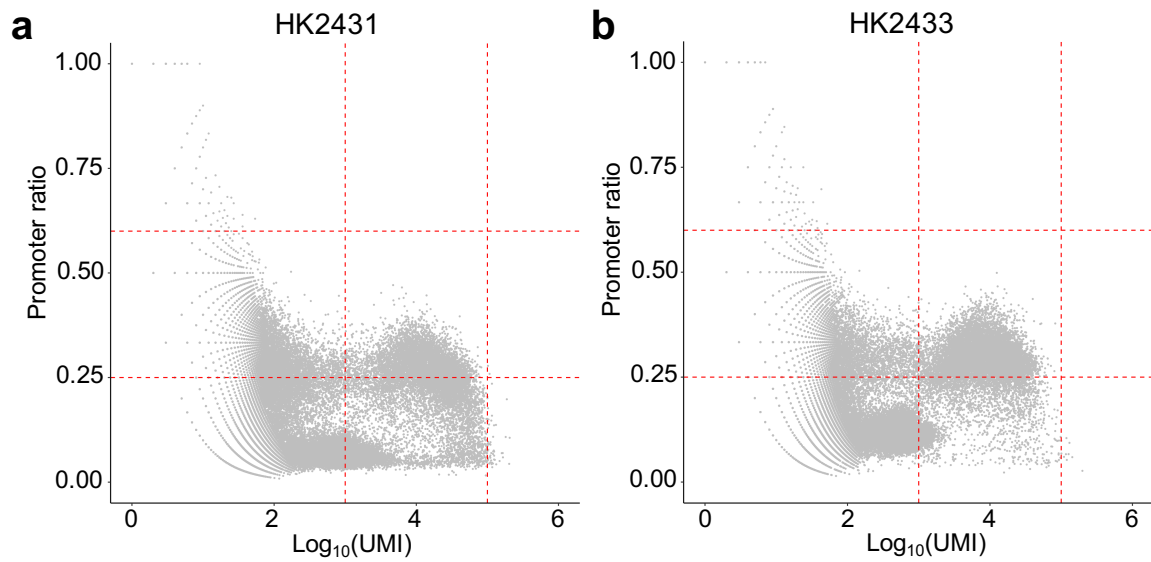


## Supplementary figure 31



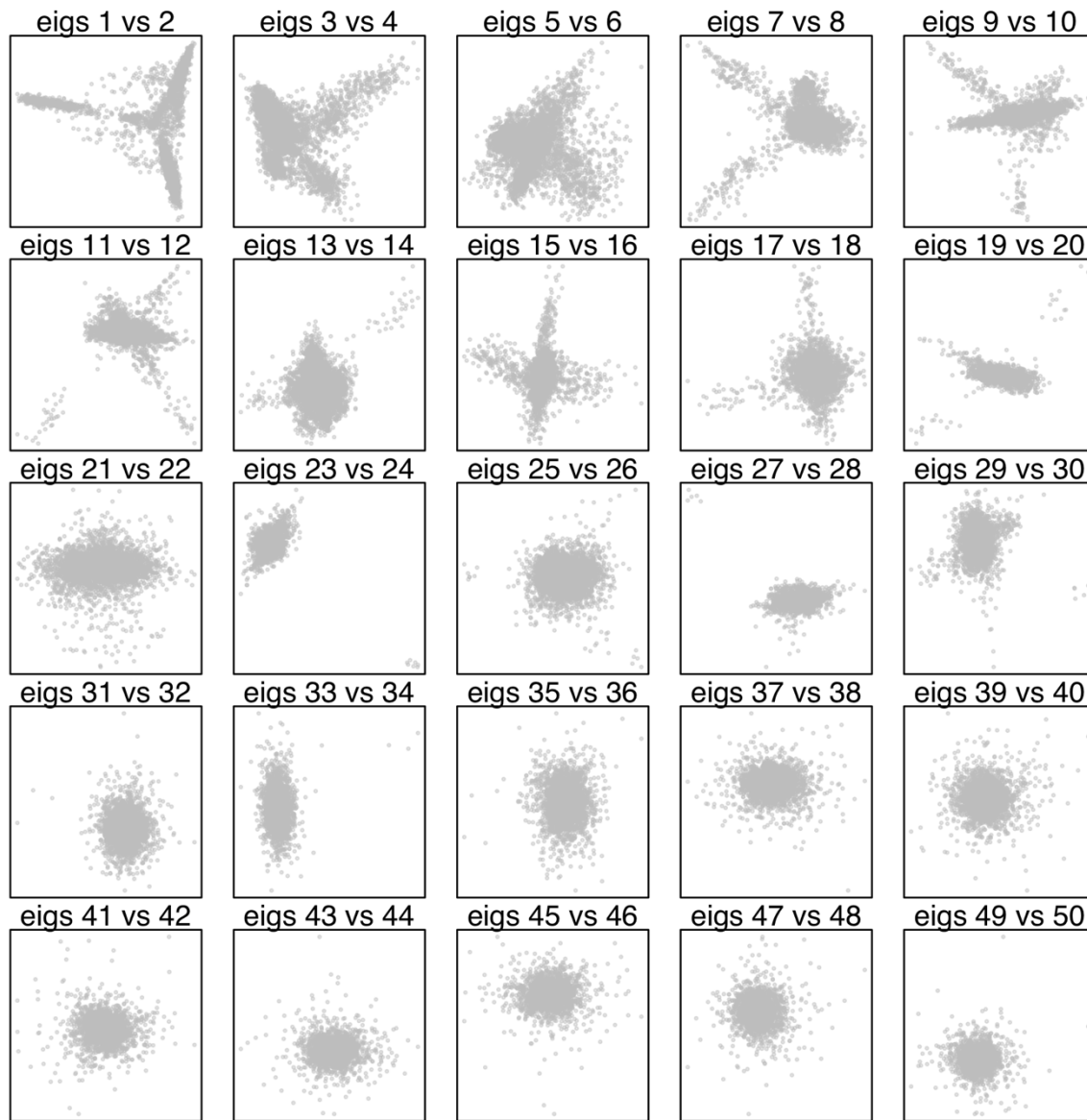
**Fig. S31. Meta-analysis of the rs6687360-AGT eQTL in human kidney compartments and 46 GTEx tissues.** The normalized effect size ( $\beta$ ) of eQTL association between rs6687360 and AGT in tubular (n=356 for this study and n=121 for study of Qiu et al.<sup>36</sup>) and glomerular (n=303 for this study and n=121 for glomerular of Qiu et al.) compartments and in 46 human tissues from GTEx (v7). The x-axis shows  $\beta$  with the 95% confidence intervals (CI). EBV, Epstein-Barr virus. Data are presented as mean values  $\pm$  SE. Two-sided P-value was calculated by linear regression from GTEx (v7)<sup>43</sup>.

### Supplementary figure 32



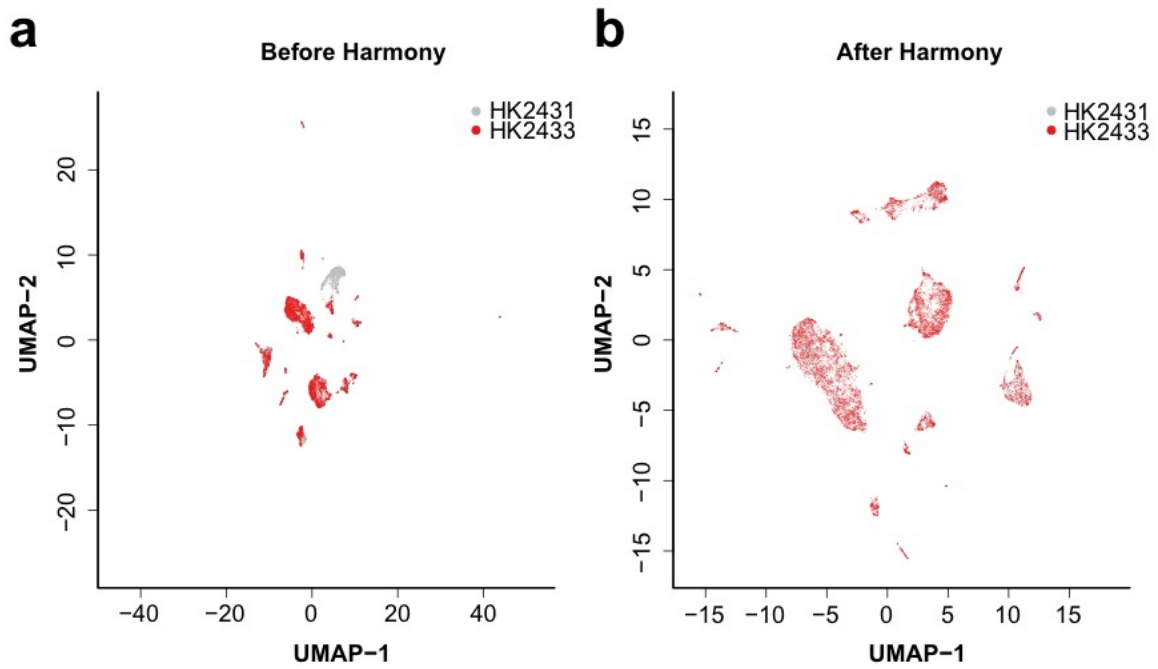
**Fig. S32. Quality Control of human kidney snATAC-seq data.** Each data point represents a cell, x-axis:  $\log_{10}(\text{UMI})$ , UMI: unique molecular identifier; y-axis: promoter ratio of sample a). HK2431 b). HK2433. Only cells with  $\log_{10}(\text{UMI})$  of 3-5 (red dash lines) and promoter ratio of 0.25-0.6 (red dash lines) were retained.

### Supplementary figure 33



**Fig. S33. Human kidney snATAC-seq dimensionality reduction.** 50 eigenvectors generated by function *runDiffusionMaps* of the SnapATAC<sup>45</sup> package.

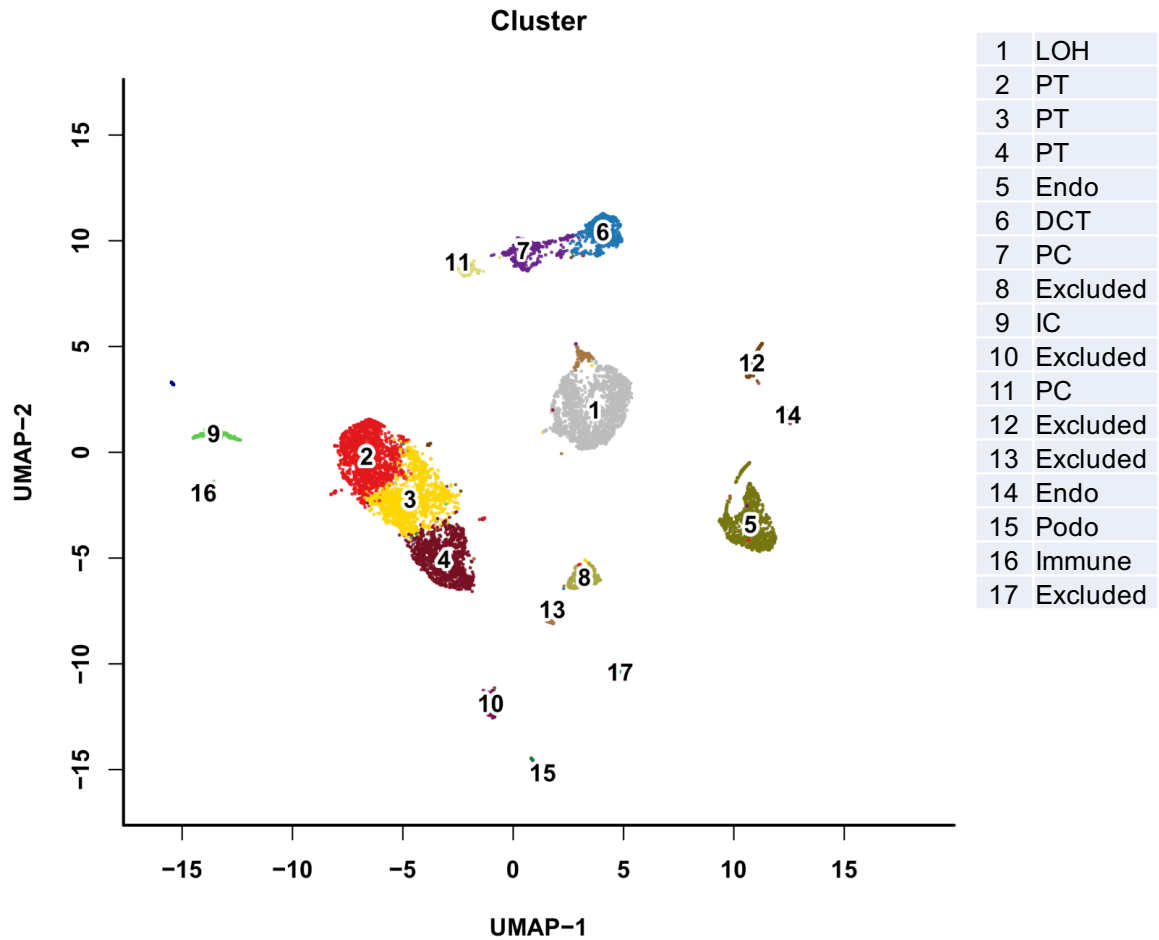
### Supplementary figure 34



**Fig. S34. Batch effect correction of human kidney snATAC-seq by Harmony<sup>46</sup>.**

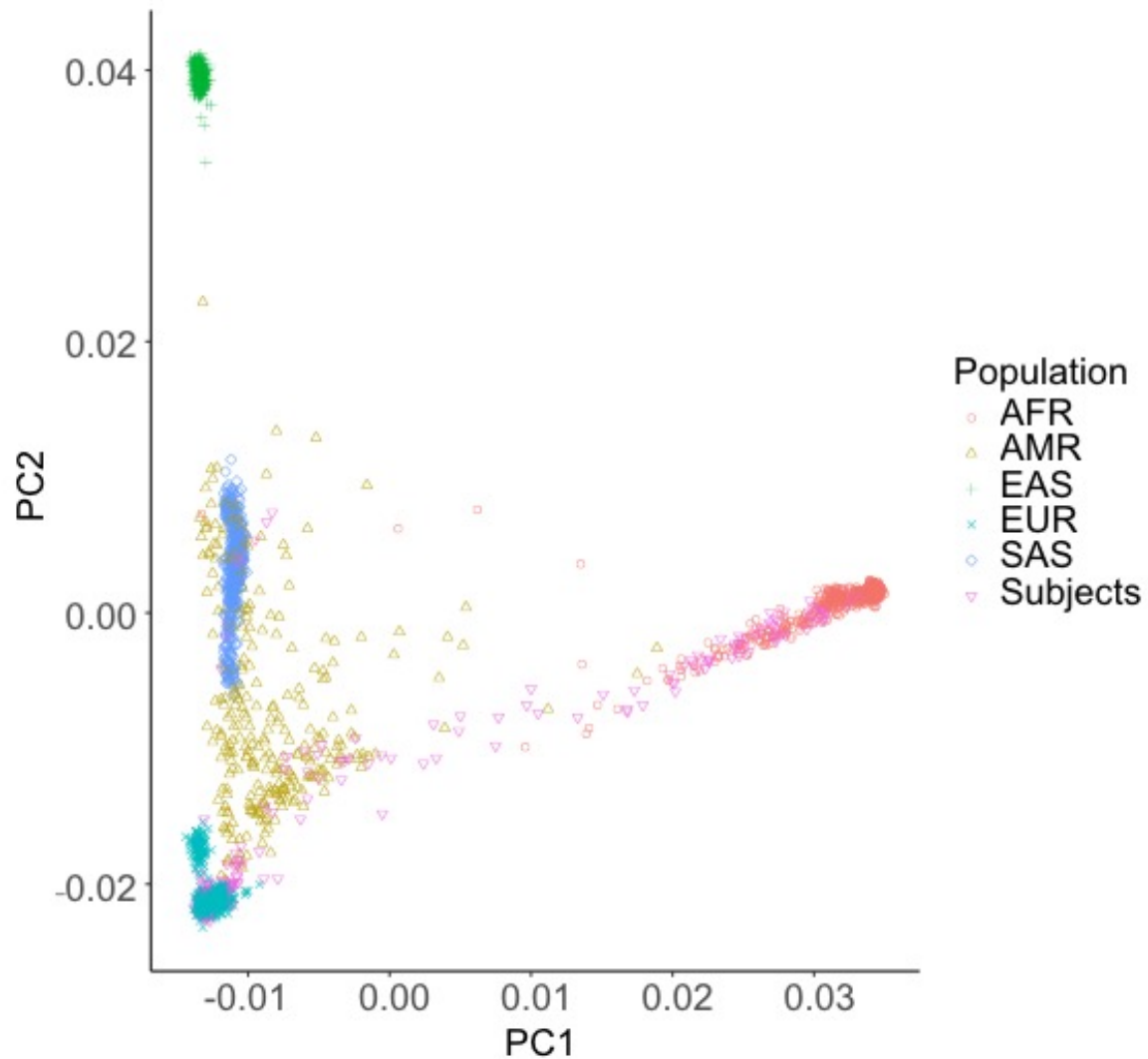
a) The UMAP before and b) after batch effect correction.

# Supplementary figure 35



**Fig. S35. The initial UMAP of the human kidney snATAC-seq.** Endo: endothelial cells, Podo: podocyte, PT: proximal tubule, LOH: loop of Henle, DCT: distal convoluted tubule, PC: collecting duct principal cells, IC: collecting duct intercalated cells, Immune: immune cells.

**Supplementary figure 36**



**Fig. S36. Genotype data distribution.** Principal component analysis of genotype data of 367 participants (individuals with high quality genotype data available) and 1,000 Genomes Project Phase 3 reference genome data. Magenta triangles indicate 367 kidney samples (of the current study).

## **Supplemental Tables**

**Table S1.** Baseline demographic and clinical characteristics of human kidney tubule tissue samples

**Table S2.** Baseline demographic and clinical characteristics of human kidney glomerular tissue samples

**Table S3.** Top 100 protein-coding eGenes in tubule (ranked by effect size)

**Table S4.** Top 100 protein-coding eGenes in glom (ranked by effect size)

**Table S5.** 303 eGenes in both tubule and glomeruli that overlapped with previously curated 625 nephropathy-associated genes

**Table S6.** Summary statistics of the eGenes (lead SNP-eGene) that did not overlap with previously curated 625 nephropathy-associated genes in tubules (identified by eQTL(cf) model)

**Table S7.** Summary statistics of the eGenes (lead SNP-eGene) that did not overlap with previously curated 625 nephropathy-associated genes in glom (identified by eQTL(cf) model)

**Table S8.** Functional annotation of the eGenes that did not overlap with previously curated 625 nephropathy-associated genes (identified by eQTL(cf) model)

**Table S9.** Sequencing data for each deleted genomic regions

**Table S10.** Cell type interaction eQTLs (eQTL(ci)s) in Tubule

**Table S11.** Cell type interaction eQTLs (eQTL(ci)s) in Glomeruli

**Table S12.** Two-sided P-value of replicated eQTL(ci)s using healthy mouse kidney scRNA-seq data for deconvolution

**Table S13.** The top 5 possible near/disrupting motifs of TFs for each eSNP (eQTL(ci)s)

**Table S14.** The enrichment of snATAC-seq open chromatin peaks in the adult human kidney

**Table S15.** DeltaSVM scores for each eSNP (eQTL(ci)s)

**Table S16.** Cell type eQTL(ci) SNPs enrichment in SNPs with significant deltaSVM scores estimated based on snATAC-seq data for each cell type

**Table S17.** Enrichment of the bulk eQTL(cf) in the snATAC-seq data

**Table S18.** Tubule coloc genes (eQTL(cf) model)

**Table S19.** Glom coloc genes (eQTL(cf) model)

**Table S20.** eGFR coloc genes (in both tubuli and glomeruli)

**Table S21.** Blood pressure (SBP) coloc genes (in both tubuli and glomeruli)

**Table S22.** Coloc genes of 6 kidney related function GWAS traits based on cell type interaction eQTLs

**Table S23.** eGFR coloc gene function (DAVID)

**Table S24.** SBP coloc gene function (DAVID)

**Table S25.** Gene drug combinations (coloc genes for all 6 traits)

**Table S26.** Druggable categories for genes (coloc genes for all 6 traits)

**Table S27.** Sequencing information for human kidney snATAC-seq samples

**Table S28.** Quality control of genotype data

**Table S29.** sgRNA and primer sequence for CRISPR based genome editing



## References

1. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics* **81**, 559-575 (2007).
2. Price, A.L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nature genetics* **38**, 904 (2006).
3. Siva, N. 1000 Genomes project. (Nature Publishing Group, 2008).
4. Delaneau, O., Marchini, J. & Zagury, J.-F. A linear complexity phasing method for thousands of genomes. *Nature methods* **9**, 179 (2012).
5. Delaneau, O., Zagury, J.-F. & Marchini, J. Improved whole-chromosome phasing for disease and population genetic studies. *Nature methods* **10**, 5 (2013).
6. Howie, B.N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS genetics* **5**, e1000529 (2009).
7. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nature genetics* **39**, 906 (2007).
8. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21. (2013).
9. Harrow, J. *et al.* GENCODE: the reference human genome annotation for The ENCODE Project. *Genome research* **22**, 1760-1774 (2012).
10. Li, B. & Dewey, C.N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC bioinformatics* **12**, 323 (2011).
11. Love, M.I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology* **15**, 550 (2014).
12. Robinson, M.D., McCarthy, D.J. & Smyth, G.K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140 (2010).
13. Newman, A.M. *et al.* Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nature biotechnology* **37**, 773-782 (2019).
14. Jew, B. *et al.* Accurate estimation of cell composition in bulk expression through robust integration of single-cell information. *Nature communications* **11**, 1-11 (2020).
15. Young, M.D. *et al.* Single-cell transcriptomes from human kidneys reveal the cellular identity of renal tumors. *Science* **361**, 594-599 (2018).

16. Wilson, P.C. *et al.* The single-cell transcriptomic landscape of early human diabetic nephropathy. *Proceedings of the National Academy of Sciences* **116**, 19619-19625 (2019).
17. Park, J. *et al.* Single-cell transcriptomics of the mouse kidney reveals potential cellular targets of kidney disease. *Science* **360**, 758-763 (2018).
18. Park, J. *et al.* Renal proximal tubule cell state and metabolism are coupled by nuclear receptors. *bioRxiv* (2020).
19. Park, J. *et al.* Functional methylome analysis of human diabetic kidney disease. *JCI insight* **4**(2019).
20. Han, B. & Eskin, E. Interpreting meta-analyses of genome-wide association studies. *PLoS Genet* **8**, e1002555 (2012).
21. Schep, A.N., Wu, B., Buenrostro, J.D. & Greenleaf, W.J. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nature methods* **14**, 975-978 (2017).
22. Khan, A. *et al.* JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic acids research* **46**, D260-D266 (2018).
23. Schep, A. motifmatchr: Fast Motif Matching in R. *motifmatchr: Fast Motif Matching in R* (2018).
24. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular cell* **38**, 576-589 (2010).
25. Corces, M.R. *et al.* The chromatin accessibility landscape of primary human cancers. *Science* **362**(2018).
26. Quinlan, A.R. & Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842 (2010).
27. Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nature methods* **9**, 215 (2012).
28. Ghandi, M. *et al.* gkmSVM: an R package for gapped-kmer SVM. *Bioinformatics* **32**, 2205-2207 (2016).
29. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nature biotechnology* **36**, 411-420 (2018).
30. Leek, J.T., Johnson, W.E., Parker, H.S., Jaffe, A.E. & Storey, J.D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882-883 (2012).
31. Lee, D. LS-GKM: a new gkm-SVM for large-scale datasets. *Bioinformatics* **32**, 2196-2198 (2016).

32. Storey, J.D. A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **64**, 479-498 (2002).
33. Pliner, H.A. *et al.* Cicero predicts cis-regulatory DNA interactions from single-cell chromatin accessibility data. *Molecular cell* **71**, 858-871. e8 (2018).
34. Wang, G., Sarkar, A.K., Carbonetto, P. & Stephens, M. A simple new approach to variable selection in regression, with application to genetic fine-mapping. *bioRxiv*, 501114 (2019).
35. Dhillon, P. *et al.* The Nuclear Receptor ESRRA Protects from Kidney Disease by Coupling Metabolism and Differentiation. *Cell Metabolism* (2020).
36. Qiu, C. *et al.* Renal compartment-specific genetic variation analyses identify new pathways in chronic kidney disease. *Nature medicine* **24**, 1721 (2018).
37. Urbut, S.M., Wang, G., Carbonetto, P. & Stephens, M. Flexible statistical methods for estimating and testing effects in genomic studies with multiple conditions. *Nature genetics* **51**, 187-195 (2019).
38. Zhang, X. *et al.* CellMarker: a manually curated resource of cell markers in human and mouse. *Nucleic acids research* **47**, D721-D728 (2019).
39. Giri, A. *et al.* Trans-ethnic association study of blood pressure determinants in over 750,000 individuals. *Nature genetics* **51**, 51-62 (2019).
40. Hellwege, J.N. *et al.* Mapping eGFR loci to the renal transcriptome and phenome in the VA Million Veteran Program. *Nat Commun* **10**, 3842 (2019).
41. Wuttke, M. *et al.* A catalog of genetic loci associated with kidney function from analyses of a million individuals. *Nature genetics* **51**, 957 (2019).
42. Dennis, G. *et al.* DAVID: database for annotation, visualization, and integrated discovery. *Genome biology* **4**, R60 (2003).
43. Consortium, G. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science* **348**, 648-660 (2015).
44. Barrett, J.C., Fry, B., Maller, J. & Daly, M.J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263-265 (2005).
45. Fang, R. *et al.* Fast and accurate clustering of single cell epigenomes reveals cis-regulatory elements in rare cell types. *bioRxiv*, 615179 (2019).
46. Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. *Nature methods*, 1-8 (2019).

## Source Data

### Source Data Supplementary Fig. 21. Relative transcript level (*FGF5*)

#### Supplementary figure 21e. Relative transcript level (*FGF5*)

| Control | rs3822015 |
|---------|-----------|
| 1.05    | 0.67      |
| 0.91    | 0.69      |
| 1.04    | 0.76      |

### Source Data Supplementary Fig. 29. Relative transcript level (*ACE*)

#### Supplementary figure 29b. Relative transcript level (*ACE*)

| Control | rs4292 | rs4295 |
|---------|--------|--------|
| 1.04    | 1.26   | 1.61   |
| 0.93    | 1.20   | 1.62   |
| 1.04    | 1.11   | 1.50   |