Published in final edited form as:

Curr Opin Nephrol Hypertens. 2019 May ; 28(3): 219–226. doi:10.1097/MNH.0000000000000495.

The kidney transcriptome, from single cells to whole organs and back

Shizheng Huang^{1,*}, Xin Sheng^{1,*}, Katalin Susztak¹

¹Department of Medicine, Renal Electrolyte and Hypertension Division, University of Pennsylvania, Philadelphia, PA, USA

Abstract

Purpose of review—Transcriptome analysis of human kidney samples provides an integrated output of genetic, physiological or environmental inputs. This review will summarize recent finding including gene expression and genetic variation integration and bulk and single cell gene expression analysis and describes how such studies have improved our understanding of kidney disease development.

Recent findings—Bulk or whole tissue analysis of patient kidney samples identified a large number of genes those levels correlate with kidney function and or structural damage. These genes were enriched for metabolic and immune functions. Using expression quantitative trait analysis, genetic-variations-driven gene expression can be identified. Recent developments in single cell sequencing defined cell-type specific gene expression changes and highlighted specific cell types for disease development.

Summary—Recent advancement in whole tissue transcriptomics, specifically incorporating genotype information and single cell data have been powerful to identify kidney disease associated genes, pathways and cell types.

Keywords

Transcriptome; RNA-seq; eQTL studies; single cell RNA-seq; deconvolution; CKD

Introduction

Kidneys play a key role in maintaining homeostasis by filtering the blood and selectively reabsorbing and secreting molecules that are smaller than 60kD. The complex process of discriminatory reabsorption and secretion is achieved by lining the tubule segment with specialized cell types that contain channels and transporters to selectively process electrolytes, metabolites or other small molecules [1–7]. In clinical practice, the function of the kidney is measured by the filtering ability of the glomerulus (GFR; glomerular filtration rate) [8]. Chronic kidney disease (CKD) is defined as greater than 40% decline in GFR

Correspondence: Katalin Susztak MD, PhD, Perelman School of Medicine, University of Pennsylvania, 3400 Civic Center Blvd, 12-123 Smilow Translational Building, Philadelphia, PA 19104, 215 898 2009, ksusztak@pennmedicine.upenn.edu.

*These authors contributed equally.

Competing interests

The authors declare no competing interests.

[9,10]. Kidney fibrosis and glomerulosclerosis are uniformly observed in patients with CKD and considered common structural manifestations of CKD [11]. Other structural parameters, such as specific changes in the glomerular compartments or immune cells, can help to differentiate specific disease subgroups [12,13].

Several experimental approaches have been taken to define the mechanism of CKD [14], including genetic studies. Kidney function is a heritable trait [15,16]. Kidney function heritability follows a complex trait pattern determined by a large number of common non-coding genetic variants [17]. Genome wide association studies of large patients' cohorts have recently been performed to understand nucleotide variations associated with kidney disease [18–23]. An important limitation of GWAS studies is that they are unable to identify causal genes, because most disease associated variants are non-coding [24,25]. The lack of a unifying picture emerging from genetic analyses, in the quest for understanding common mechanisms underlying CKD, calls for alternative approaches to better describe the pathogenesis of these disorders, such as transcriptome analysis [26]. While animal models have been employed successfully to understand kidney disease development, in this review we are almost exclusively focusing on studies performed on human patient samples.

Gene expression analysis of kidneys of patients with CKD

Transcriptome analysis provides the opportunity to survey changes resulting from genetic, physiological, or environmental interactions that may be involved in CKD pathogenesis [27–29]. While proteins show a closer relationship to function, methods analyzing protein quantity and modifications are relatively cumbersome. On the other hand, gene expression microarrays and later RNA-sequencing allows accurate and affordable transcript quantification [30,31]. Several recent publications analyzed transcriptomics data obtained from kidneys of healthy and disease individuals [12,32]. Beckerman et al., generated transcriptome data from 95 micro-dissected human kidney tubule samples with a range of demographics, and functional and structural changes [12]. To define genes whose expression correlates with kidney disease manifestation, such as kidney function (GFR) or kidney fibrosis, a linear regression model adjusting for common co-variants was used [12,33]. Large numbers of genes can be identified by correlating gene expression with disease status. Gene ontology and pathway analysis indicated enrichment for genes with immune and metabolic functions (Figure 1).

Recently, unbiased analytical methods, such as Weighted Gene Co-Expression Network Analysis (WGCNA), have gained popularity for identifying gene groups for clinical phenotypes [34]. First, WGCNA identifies a group of genes in a dataset whose expression variation strongly correlated. These gene groups are then used as modules. Gene expression data from 91 kidney samples identified 16 co-expressed gene modules. Since the modules are identified in an unbiased manner, independent of the phenotype, the next step is to understand the correlation between gene modules and clinical phenotypes. In our analysis, we found that modules that strongly correlated with eGFR primarily encoded genes with metabolic functions. Gene groups that mainly encoded T-cell receptor and collagen pathways, showed the strongest correlation with fibrosis levels, suggesting that these two phenotypic manifestations might have different underlying mechanisms. Genes identified

using the unbiased (WGCNA) and targeted (Linear regression) analysis significantly overlapped. In general, the correlation between gene expression and structural changes (fibrosis, or glomerulosclerosis) was stronger than functional changes (eGFR). Therefore, it was concluded that gene expression is a very sensitive sensor of fibrosis, as the expression of 1,654 genes correlated with fibrosis even after adjusting to eGFR and other clinical parameters. The association between GFR and gene expression was mostly mediated by fibrosis. In conclusion, the transcriptome-based CKD trait dissection analysis suggests that the association between gene expression and renal function is mediated by structural changes. Differences in pathways that lead to the decline in kidney function and the development of fibrosis can be identified.

Similar studies have been performed to understand specific disease etiologies, such as diabetic kidney disease (DKD). Structural changes affecting the renal glomerulus are the most specific and the gold standard criteria to diagnose diabetic kidney disease [13,35]. One of the earliest studies, performed by Woroniecka et al, examined isolated glomerular samples from control and DKD glomeruli [36]. By analyzing glomeruli isolated from 44 human kidneys (controls and DKD) and using stringent statistical criteria, a large number (1,700) of differentially expressed probesets were identified. The majority of the probesets (1,214 [~70%]) showed decreased expression level in DKD glomeruli. Almost all podocyte-specific transcripts, such as PLCE1, NPHS1, NPHS2, SYNPO, PLA2R1 and WT1, showed severely decreased expression. These studies are consistent with the key role of podocytes and podocyte loss in DKD development. Transcripts that showed the highest increase in their expression included IGH, C3, COL1A2, CXCL6, and COL6A3. These transcripts are known to be associated with fibrosis and inflammation. The top differentially regulated pathways with the highest degree of statistical significance were the complement system, the RhoA, Cdc42, integrin, integrin-linked kinase (ILK), tight junction, and macropinocytosis pathways. In addition, semaphorins and vascular endothelial growth factor (VEGF) pathways also showed significant enrichment. Some of these pathways have been proposed to play a role in DKD development using animal model studies.

Future approaches surveying the transcriptome may reveal that changes in additional processes such as RNA splicing, editing, transport, localization and stability, as well as functional interactions between RNA transcripts, DNA or proteins, could contribute to CKD [37–40].

Genetically driven gene expression changes

A critical limitation of transcriptomic analysis is that it cannot establish a causal relationship between expression of genes and disease development [16]. While some genes and pathways are involved in disease development, others are mostly just a consequence of the disease state [41,42]. As genotype assortment always occurs prior to disease development, genotype information can help to identify causal genetic pathways [43]. Genome-wide association studies (GWASs) have identified multiple loci associated with the risk of CKD [44–47]. As almost all risk variants are localized to the noncoding region of the genome [48] molecular pathways influenced by genetic variations remains unknown. Some of the first approaches to elucidate the role of genetic variants relied on describing genes in the vicinity of these

variants that are associated with kidney disease heritability [47]. The regulation of transcripts in the vicinity of CKD-associated polymorphisms in control and diseased human kidney samples were examined to identify potentially causal genes for prioritization [49]. The expression and regulation of 226 transcripts in the vicinity of 44 single nucleotide polymorphisms using RNA sequencing and gene expression arrays from 95 micro-dissected control and diseased tubule samples and 51 glomerular samples were examined. 92 transcripts in the tubule compartment and 34 transcripts in glomeruli showed statistically significant correlation with eGFR. Many novel genes, including ACSM2A/2B, FAM47E, and PLXDC1, which previously had not been implicated in CKD development, have been identified. Interestingly, the expression of multiple genes in the vicinity of any single CKD risk allele correlated with renal function, potentially indicating that genetic variants influence multiple transcripts. Network analysis of GFR-correlating transcripts highlighted two major clusters: a positive correlation with epithelial and vascular functions and an inverse correlation with the inflammatory gene cluster.

Expression quantitative trait loci (eQTL) analysis is a method that identifies genetic variation associated with gene expression changes in specific tissue types [50]. It has been hypothesized that an integrative analysis combining CKD GWAS and kidney eQTL results can identify candidate genes for CKD (Figure 2). The first kidney eQTL analysis by correlating genotype with RNA-seq-based gene expression levels was performed in 96 human kidney samples [51]. Using stringent statistical criteria, 1,886 genes whose expression correlated with sequence variations were identified. Using direct overlap and Bayesian methods, new potential target genes for CKD were identified. With respect to one of the target genes, lysosomal beta A mannosidase (MANBA), Ko et al found that the genetic variants associated with MANBA expression in the kidney showed statistically significant colocalization with variants identified in CKD GWASs, indicating that MANBA is a potential target gene for CKD. The expression of MANBA was significantly lower in kidneys of subjects with risk alleles. Suppressing manba expression in zebrafish resulted in renal tubule defects and pericardial edema, phenotypes typically induced by kidney dysfunction. These studies emphasized that gene-expression changes driven by genetic variation in the kidney can highlight potential new target genes for CKD development.

Expression data generated from whole kidney cortical samples identified a remarkably small number of CKD GWAS target genes. It was hypothesized that this was mostly attributed to the cellular and phenotypic heterogeneity observed in whole kidney lysates. To reduce cell heterogeneity, a new eQTL dataset was generated from kidney samples microdissected into glomeruli and tubuli, restricting samples only for single ethnicity and samples that are obtained from kidneys with normal kidney function and absence of any structural alterations (no fibrosis) [52]. This new eQTL dataset significantly improved the eGene identification and identified close to 4,000 genes whose levels were influenced by genetic variation in the tubule or glomerular compartments. By integrating this new set of eGenes with prior genetic studies (GWAS), Qiu et al was able to identify 27 likely causal genes for kidney disease. Furthermore, the team demonstrated that target genes are enriched for proximal tubule expression. Functional analysis indicated that putative causal genes were enriched for endolysosomal function, where DAB2, an adaptor protein in the TGF- β pathway, formed a central node. Functional experiments confirmed that reducing Dab2 expression in renal

tubules protected mice from CKD. These studies indicate that compartment-specific eQTL analysis is an important avenue for the identification of novel genes and cellular pathways involved in CKD development. In the future it will be important to compare eQTL studies performed in healthy kidney tissue samples to those obtained from patients with kidney disease. The Sampson group has generated a compartment specific eQTL dataset from subjects with nephrotic syndrome [53]. The number of eGenes identified in this dataset was significantly lower when compared to healthy tissue studies (by Qiu et al), indicating that cell heterogeneity might be an important confounder in disease tissue analysis. On the other hand, there was a large number of eGenes that showed consistent regulation in healthy and disease samples. Future studies will be essential to explore the role of disease state and cell heterogeneity in eQTL datasets.

Gene expression analysis at a single cell level

Tissue (bulk) RNA sequencing (RNA-seq) is a widely adopted method used to understand genome wide transcriptomic variations in different conditions such as disease states [54]. Bulk RNA-seq generates an output for the average expression of genes, which is the sum of cell specific gene expression differences and cell composition changes. Knowledge of cell type composition and their proportions in intact tissues is important, because certain cell types could be more vulnerable to diseases than others. Characterizing the variation in cell type composition across subjects can identify cellular targets of diseases, and adjusting for these variations can clarify downstream analysis (Figure 3).

The rapid development of single-cell RNA-seq (scRNA-seq) technologies has enabled cell-type specific transcriptome profiling [55]. Although cell type composition and proportions are obtainable from scRNA-seq, scRNA-seq is still costly, prohibiting its application in clinical studies that involve a large number of subjects [56]. Furthermore, scRNA-seq is not well suited to characterizing cell type proportions in a solid tissue, due to the cell dissociation induced bias [56].

Young et al. was the first to publish single cell resolution gene expression maps for the human kidney [57]. They had used normal tissue biopsies taken from macroscopically normal regions of kidneys resected for cancer or transplantation. They performed technical replicates of each biopsy and prepared biological replicates. They processed kidneys immediately after resection, generating single-cell solutions enriched for viable cells. Total of 72,501 fetal, normal, and tumor cells were sequenced. Using a community detection algorithm, they further segregated transcriptomes into distinct clusters of cells. They next generated a reference map of normal mature and fetal cells, assigning an identity to each cluster by cross-referencing cluster-defining transcripts with canonical markers curated from the literature. Ambiguous clusters were not included in the reference map. Among 42,809 nonmalignant cells, 37,951 mature kidney cells represented epithelial cells from distinct micro-anatomical regions of the nephron, with a large proportion of proximal tubular cells. They identified fibroblasts, myofibroblasts, and different vascular endothelial cells such as glomerular endothelium and ascending and descending vasa recta cells. Cell type specific markers showed very strong correlation with cell type specific gene expression signatures previously generated for mouse kidney samples by Park et al [58].

The Humphreys group performed single nuclear RNA-seq (snRNAseq) using the 10X Chromium platform by sequencing 4,524 nuclei and identified 12 distinct epithelial cell clusters, including podocytes, proximal tubule cells, loop of Henle cells, distal tubule cells, connecting segment cells, principal cells, and intercalated cells [59]. They speculated that the absence of stromal or leukocyte populations reflected either a dissociation bias and/or a cell frequency below our limit of detection. Given the significant heterogeneity in the human population, much larger studies will be needed to generate reference kidney single cell atlases and expression data from different disease conditions.

Connecting bulk and single cell RNAseq analysis to define cell composition changes

Changes in cell composition underlie diverse physiological and pathological states [60,61]. For example, in malignant tumors, levels of infiltrating immune cells are associated with tumor growth and patient outcome [61]. Common methods for studying cell heterogeneity, such as immunohistochemistry and flow cytometry, rely on a limited repertoire of phenotypic markers and tissue disaggregation, and therefore are prone to bias [62]. Most single cell sequencing methods suffer from cell capture bias and as a result they likely cannot accurately map cell proportion changes in different conditions.

Recently, computational methods were reported for predicting fractions of multiple cell types in gene expression profiles of bulk (or whole tissue) studies [62]. One commonly used in silico method for cell fraction estimation is CIBERSORT, a computational approach that accurately resolves relative fractions of diverse cell subsets from complex tissues [63]. To deconvolve the cell mixture data, they employ a novel application of linear support vector regression (SVR), which is a robust machine learning approach. SVR performs a feature selection where genes from the signature matrix are adaptively selected to deconvolve a given mixture. Multiple additional deconvolution methods have been published, such as linear least-squares regression (LLSR) [64], quadratic programming (QP) [65], perturbation model for gene expression deconvolution (PERT) [66], robust linear regression (RLR) [67], microarray microdissection with analysis of differences (MMAD) [68] and digital sorting algorithm (DSA) [69]. A more recent method is BSEQ-sc [70] is an extension from CIBERSORT and allows the use of scRNA-seq gene expression as a reference. TIMER [71], developed for cancer data, focuses on immune cell infiltration quantification. These methods rely on pre-selected cell type-specific marker genes, and thus are sensitive to the choice of significance threshold. More importantly, these methods ignore cross-subject heterogeneity in cell type-specific gene expression as well as within-cell type stochasticity of single cell expression, both of which are important when analyzing multiple scRNA-seq datasets.

The MUlti-Subject Single Cell deconvolution (MuSiC) method utilizes cross-subject scRNA-seq to estimate cell type proportions in bulk RNA-seq data [72]. A key concept in MuSiC is “marker gene stability”. Another method is called DESCEND, a statistical method that deconvolves the true cross-cell gene expression distribution from observed scRNA-seq counts and quantifies how this distribution depends on cell-level [73].

In a recent paper Park et al have used CIBERSORT to understand cell proportion changes in disease mouse and human kidney samples [58]. By examining cell fraction changes in the collecting duct they observed differences in principal and intercalated cell fractions in CKD samples. Combining single cell sequencing with bulk RNAsequencing and deconvolution analysis appears to be an important path for understanding cell heterogeneity and disease development in various organs.

Conclusion

A revolution in global gene expression analysis started 15–20 years ago with the introduction of spotted microarrays and continued with the development of RNAsequencing. These methods can now very precisely determine the expression and regulation of transcripts in small sample quantities. Single cell sequencing, which is less than five years old, has developed rapidly and is able to provide unprecedented insight into gene regulation at cellular resolution in thousands and potentially millions of cells [74]. The current limitation of these methods lies in translating genome wide observations into understanding kidney health and disease states. Combining gene expression analysis with genetic information such as eQTL and GWAS studies has been highly informative for identifying putative causal genes for kidney disease development. Combining single cell and bulk (whole tissue) RNA-sequencing methods can be promising for understanding cell type specific and cell heterogeneity changes in disease states.

Acknowledgements

The authors thank Chang (Linda) Liu for her assistance with proofreading. This work in the Susztak lab has been supported by the National Institute of Health NIH R01 DK076077, DK087635 and DP3108220, and the Juvenile Diabetes Research Foundation.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- of special interest
- of outstanding interest

1. Roy A, Al-bataineh MM, Pastor-Soler NM: Collecting duct intercalated cell function and regulation. *Clin J Am Soc Nephrol* (2015) 10(2):305–324. [PubMed: 25632105]
2. Nigam SK, Wu W, Bush KT, Hoenig MP, Blantz RC, Bhatnagar V: Handling of drugs, metabolites, and uremic toxins by kidney proximal tubule drug transporters. *Clin J Am Soc Nephrol* (2015) 10(11):2039–2049. [PubMed: 26490509]
3. Mount DB: Thick ascending limb of the loop of henle. *Clin J Am Soc Nephrol* (2014) 9(11):1974–1986. [PubMed: 25318757]
4. Pearce D, Soundararajan R, Trimpert C, Kashlan OB, Deen PM, Kohan DE: Collecting duct principal cell transport processes and their regulation. *Clin J Am Soc Nephrol* (2015) 10(1):135–146. [PubMed: 24875192]
5. Subramanya AR, Ellison DH: Distal convoluted tubule. *Clin J Am Soc Nephrol* (2014) 9(12):2147–2163. [PubMed: 24855283]

6. Dantzler WH, Layton AT, Layton HE, Pannabecker TL: Urine-concentrating mechanism in the inner medulla: Function of the thin limbs of the loops of henle. *Clin J Am Soc Nephrol* (2014) 9(10): 1781–1789. [PubMed: 23908457]
7. Curthoys NP, Moe OW: Proximal tubule function and response to acidosis. *Clin J Am Soc Nephrol* (2014) 9(9):1627–1638. [PubMed: 23908456]
8. Porrini E, Ruggerenti P, Luis-Lima S, Carrara F, Jimenez A, de Vries APJ, Torres A, Gaspari F, Remuzzi G: Estimated gfr: Time for a critical appraisal. *Nat Rev Nephrol* (2018).
9. Levey AS, Inker LA, Matsushita K, Greene T, Willis K, Lewis E, de Zeeuw D, Cheung AK, Coresh J: Gfr decline as an end point for clinical trials in ckd: A scientific workshop sponsored by the national kidney foundation and the us food and drug administration. *Am J Kidney Dis* (2014) 64(6): 821–835. [PubMed: 25441437]
10. Thompson A, Lawrence J, Stockbridge N: Gfr decline as an end point in trials of ckd: A viewpoint from the fda. *Am J Kidney Dis* (2014) 64(6):836–837. [PubMed: 25446026]
11. An Y, Xu F, Le W, Ge Y, Zhou M, Chen H, Zeng C, Zhang H, Liu Z: Renal histologic changes and the outcome in patients with diabetic nephropathy. *Nephrol Dial Transplant* (2015) 30(2):257–266. [PubMed: 25063425]
12. Beckerman P, Qiu C, Park J, Ledo N, Ko YA, Park AD, Han SY, Choi P, Palmer M, Susztak K: Human kidney tubule-specific gene expression based dissection of chronic kidney disease traits. *EBioMedicine* (2017) 24(267–276).
13. Reidy K, Kang HM, Hostetter T, Susztak K: Molecular mechanisms of diabetic kidney disease. *J Clin Invest* (2014) 124(6):2333–2340. [PubMed: 24892707]
14. Breyer MD, Susztak K: The next generation of therapeutics for chronic kidney disease. *Nat Rev Drug Discov* (2016).
15. Fox CS, Yang Q, Cupples LA, Guo CY, Larson MG, Leip EP, Wilson PW, Levy D: Genomewide linkage analysis to serum creatinine, gfr, and creatinine clearance in a community-based population: The framingham heart study. *J Am Soc Nephrol* (2004) 15(9):2457–2461. [PubMed: 15339995]
16. Emilsson V, Thorleifsson G, Zhang B, Leonardson AS, Zink F, Zhu J, Carlson S, Helgason A, Walters GB, Gunnarsdottir S, Mouy M et al.: Genetics of gene expression and its effect on disease. *Nature* (2008) 452(7186):423–428. [PubMed: 18344981]
17. Price PM, Hirschhorn K, Safirstein RL: Chronic kidney disease and gwas: “The proper study of mankind is man”. *Cell Metab* (2010) 11(6):451–452. [PubMed: 20519117]
18. Kottgen A, Pattaro C, Boger CA, Fuchsberger C, Olden M, Glazer NL, Parsa A, Gao X, Yang Q, Smith AV, O’Connell JR et al.: New loci associated with kidney function and chronic kidney disease. *Nat Genet* (2010) 42(5):376–384. [PubMed: 20383146]
19. Chambers JC, Zhang W, Lord GM, van der Harst P, Lawlor DA, Sehmi JS, Gale DP, Wass MN, Ahmadi KR, Bakker SJ, Beckmann J et al.: Genetic loci influencing kidney function and chronic kidney disease. *Nat Genet* (2010) 42(5):373–375. [PubMed: 20383145]
20. Kottgen A, Glazer NL, Dehghan A, Hwang SJ, Katz R, Li M, Yang Q, Gudnason V, Launer LJ, Harris TB, Smith AV et al.: Multiple loci associated with indices of renal function and chronic kidney disease. *Nat Genet* (2009) 41(6):712–717. [PubMed: 19430482]
21. Morris AP, Le TH, Wu H, Akbarov A, van der Most PJ, Hemani G, Smith GD, Mahajan A, Gaulton KJ, Nadkarni GN, Valladares-Salgado A et al.: Trans-ethnic kidney function association study reveals putative causal genes and effects on kidney-specific disease aetiologies. *Nat Commun* (2019) 10(1):29. [PubMed: 30604766]
22. Salem R, Todd JN, Sandholm N, Cole JB, Chen W-M, Andrews D, Pezzolesi MG, MCKEIGUE PM, Hiraki LT, Qiu C, Nair V et al.: Genome-wide association study of diabetic kidney disease highlights biology involved in renal basement membrane collagen. *bioRxiv* (2018) 499616.
23. Pattaro C, Teumer A, Gorski M, Chu AY, Li M, Mijatovic V, Garnaas M, Tin A, Sorice R, Li Y, Taliun D et al.: Genetic associations at 53 loci highlight cell types and biological pathways relevant for kidney function. *Nat Commun* (2016) 7(10023).
24. Zhang F, Lupski JR: Non-coding genetic variants in human disease. *Hum Mol Genet* (2015) 24(R1):R102–110. [PubMed: 26152199]

25. Leiserson MD, Eldridge JV, Ramachandran S, Raphael BJ: Network analysis of gwas data. *Curr Opin Genet Dev* (2013) 23(6):602–610. [PubMed: 24287332]
26. Gluck C, Ko YA, Susztak K: Precision medicine approaches to diabetic kidney disease: Tissue as an issue. *Curr Diab Rep* (2017) 17(5):30. [PubMed: 28364354]
27. Tajti F, Antoranz A, Ibrahim MM, Kim H, Ceccarelli F, Kuppe C, Alexopoulos LG, Kramann R, Saez-Rodriguez J: A functional landscape of chronic kidney disease entities from public transcriptomic data. *bioRxiv* (2018) 265447.
28. Harlan SM, Heinz-Taheny KM, Overstreet JM, Breyer MD, Harris RC, Heuer JG: Pathological and transcriptome changes in the reninangiotensin model of advanced diabetic kidney disease exhibit features of human disease. *Toxicologic pathology* (2018) 46(8):991–998. [PubMed: 30392455]
29. Xu X, Eales JM, Akbarov A, Guo H, Becker L, Talavera D, Ashraf F, Nawaz J, Pramanik S, Bowes J: Molecular insights into genome-wide association studies of chronic kidney disease-defining traits. *Nature communications* (2018) 9(1):4800.
30. Wang Z, Gerstein M, Snyder M: Rna-seq: A revolutionary tool for transcriptomics. *Nature reviews genetics* (2009) 10(1):57.
31. Schena M, Shalon D, Davis RW, Brown PO: Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* (1995) 270(5235):467–470. [PubMed: 7569999]
32. Kang HM, Ahn SH, Choi P, Ko YA, Han SH, Chinga F, Park AS, Tao J, Sharma K, Pullman J, Bottinger EP et al.: Defective fatty acid oxidation in renal tubular epithelial cells has a key role in kidney fibrosis development. *Nat Med* (2015) 21(1):37–46. [PubMed: 25419705]
33. Lemos DR, McMurdo M, Karaca G, Wilflingseder J, Leaf IA, Gupta N, Miyoshi T, Susa K, Johnson BG, Soliman K: Interleukin-1 β activates a myc-dependent metabolic switch in kidney stromal cells necessary for progressive tubulointerstitial fibrosis. *Journal of the American Society of Nephrology* (2018) 29(6):1690–1705. [PubMed: 29739813]
34. Langfelder P, Horvath S: Wgcna: An R package for weighted correlation network analysis. *BMC Bioinformatics* (2008) 9(559).
35. Lin JS, Susztak K: Podocytes: The weakest link in diabetic kidney disease? *Curr Diab Rep* (2016) 16(5):45. [PubMed: 27053072]
36. Woroniecka KI, Park AS, Mohtat D, Thomas DB, Pullman JM, Susztak K: Transcriptome analysis of human diabetic kidney disease. *Diabetes* (2011) 60(9):2354–2369. [PubMed: 21752957]
37. Wang D, Liu S, Warrell J, Won H, Shi X, Navarro FCP, Clarke D, Gu M, Emani P, Yang YT, Xu M et al.: Comprehensive functional genomic resource and integrative model for the human brain. *Science* (2018) 362(6420).
38. Gandal MJ, Zhang P, Hadjimichael E, Walker RL, Chen C, Liu S, Won H, van Bakel H, Varghese M, Wang Y, Shieh AW et al.: Transcriptome-wide isoform-level dysregulation in asd, schizophrenia, and bipolar disorder. *Science* (2018) 362(6420).
39. An JY, Lin K, Zhu L, Werling DM, Dong S, Brand H, Wang HZ, Zhao X, Schwartz GB, Collins RL, Currall BB et al.: Genome-wide de novo risk score implicates promoter variation in autism spectrum disorder. *Science* (2018) 362(6420).
40. Rajarajan P, Borrmann T, Liao W, Schrode N, Flaherty E, Casino C, Powell S, Yashaswini C, LaMarca EA, Kassim B, Javidfar B et al.: Neuron-specific signatures in the chromosomal connectome associated with schizophrenia risk. *Science* (2018) 362(6420).
41. Yu W, Li Y, Wang Z, Liu L, Liu J, Ding F, Zhang X, Cheng Z, Chen P, Dou J: Transcriptomic changes in human renal proximal tubular cells revealed under hypoxic conditions by rna sequencing. *International journal of molecular medicine* (2016) 38(3):894–902. [PubMed: 27432315]
42. Chen D-Q, Cao G, Chen H, Liu D, Su W, Yu X-Y, Vaziri ND, Liu X-H, Bai X, Zhang L: Gene and protein expressions and metabolomics exhibit activated redox signaling and wnt/ β -catenin pathway are associated with metabolite dysfunction in patients with chronic kidney disease. *Redox biology* (2017) 12(505–521).
43. Kircher M, Witten DM, Jain P, O’Roak BJ, Cooper GM, Shendure J: A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* (2014) 46(3):310–315. [PubMed: 24487276]

44. Olden M, Teumer A, Bochud M, Pattaro C, Köttgen A, Turner ST, Rettig R, Chen M-H, Dehghan A, Bastardot F: Overlap between common genetic polymorphisms underpinning kidney traits and cardiovascular disease phenotypes: The ckdgen consortium. *American Journal of Kidney Diseases* (2013) 61(6):889–898. [PubMed: 23474010]
45. Köttgen A, Pattaro C, Böger CA, Fuchsberger C, Olden M, Glazer NL, Parsa A, Gao X, Yang Q, Smith AV: New loci associated with kidney function and chronic kidney disease. *Nature genetics* (2010) 42(5):376. [PubMed: 20383146]
46. Pattaro C, Teumer A, Gorski M, Chu AY, Li M, Mijatovic V, Garnaas M, Tin A, Sorice R, Li Y: Genetic associations at 53 loci highlight cell types and biological pathways relevant for kidney function. *Nature communications* (2016) 7(10023).
47. Köttgen A, Glazer NL, Dehghan A, Hwang S-J, Katz R, Li M, Yang Q, Gudnason V, Launer LJ, Harris TB: Multiple loci associated with indices of renal function and chronic kidney disease. *Nature genetics* (2009) 41(6):712. [PubMed: 19430482]
48. Ward LD, Kellis M: Interpreting noncoding genetic variation in complex traits and human disease. *Nat Biotechnol* (2012) 30(11):1095–1106. [PubMed: 23138309]
49. Ledo N, Ko YA, Park AS, Kang HM, Han SY, Choi P, Susztak K: Functional genomic annotation of genetic risk loci highlights inflammation and epithelial biology networks in ckd. *J Am Soc Nephrol* (2015) 26(3):692–714. [PubMed: 25231882]
50. Consortium GT, Laboratory DA, Coordinating Center -Analysis Working G, Statistical Methods groups-Analysis Working G, Enhancing Gg, Fund NIHC, Nih/Nci, Nih/Nhgri, Nih/Nimh, Nih/Nida, Biospecimen Collection Source Site N et al.: Genetic effects on gene expression across human tissues. *Nature* (2017) 550(7675):204–213. [PubMed: 29022597]
51. Ko YA, Yi H, Qiu C, Huang S, Park J, Ledo N, Kottgen A, Li H, Rader DJ, Pack MA, Brown CD et al.: Genetic-variation-driven gene-expression changes highlight genes with important functions for kidney disease. *Am J Hum Genet* (2017) 100(6):940–953. [PubMed: 28575649]
52. Qiu C, Huang S, Park J, Park Y, Ko YA, Seasock MJ, Bryer JS, Xu XX, Song WC, Palmer M, Hill J et al.: Renal compartment-specific genetic variation analyses identify new pathways in chronic kidney disease. *Nat Med* (2018) 24(11):1721–1731. [PubMed: 30275566] • An expression quantitative trait (eQTL) atlas was created for the tubular and glomerular compartments of the human kidney. By integrating CKD GWAS information with eQTL, single-cell RNA sequencing and regulatory region map, putative causal genes for CKD were identified. An adaptor protein in the TGF- β pathway, DAB2, was identified and confirmed by experiments.
53. Gillies CE, Putler R, Menon R, Otto E, Yasutake K, Nair V, Hoover P, Lieb D, Li S, Eddy S, Fermin D et al.: An eqtl landscape of kidney tissue in human nephrotic syndrome. *Am J Hum Genet* (2018) 103(2):232–244. [PubMed: 30057032] • eQTL landscape was established based on Whole-genome sequencing (WGS) and microdissected glomerular (n=136) and tubulointerstitial (n=166) transcriptomes from 187 individuals with nephrotic syndrome. Cell type specific eQTLs were identified by combining single-cell RNA-seq data. Deconvoluted cell-specific signals were used to characterize known GWAS alleles.
54. Oszolak F, Milos PM: Rna sequencing: Advances, challenges and opportunities. *Nat Rev Genet* (2011) 12(2):87–98. [PubMed: 21191423]
55. Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, Tirosh I, Bialas AR, Kamitaki N, Martersteck EM, Trombetta JJ et al.: Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* (2015) 161(5):1202–1214. [PubMed: 26000488]
56. Shalek AK, Benson M: Single-cell analyses to tailor treatments. *Sci Transl Med* (2017) 9(408).
57. Young MD, Mitchell TJ, Vieira Braga FA, Tran MGB, Stewart BJ, Ferdinand JR, Collord G, Botting RA, Popescu DM, Loudon KW, Vento-Tormo R et al.: Single-cell transcriptomes from human kidneys reveal the cellular identity of renal tumors. *Science* (2018) 361(6402):594–599. [PubMed: 30093597] •• This study was based on 72,501 single-cell transcriptomes of human renal tumors and normal tissue from fetal, pediatric and adult kidneys. Evidence of aberrant Wilms tumor cells were found by matching childhood Wilms tumors with specific fetal cell types. Cancer-associated normal cells were defined and a complex vascular endothelial growth factor (VEGF) signaling circuit was delineated by analyzing the tumor composition.
58. Park J, Shrestha R, Qiu C, Kondo A, Huang S, Werth M, Li M, Barasch J, Susztak K: Single-cell transcriptomics of the mouse kidney reveals potential cellular targets of kidney disease. *Science*

(2018) 360(6390):758–763. [PubMed: 29622724] • 57,979 cells from healthy mouse kidneys were characterized by unbiased single-cell RNA sequencing. In adult mice kidneys, the collecting duct generates a spectrum of cell types through a newly identified transitional cell. Intercalated cells and principal cells undergo transitions mediated by the Notch signaling pathway suggested by computational cell trajectory analysis and in vivo lineage tracing.

59. Wu H, Malone AF, Donnelly EL, Kirita Y, Uchimura K, Ramakrishnan SM, Gaut JP, Humphreys BD: Single-cell transcriptomics of a human kidney allograft biopsy specimen defines a diverse inflammatory response. *J Am Soc Nephrol* (2018) 29(8):2069–2080. [PubMed: 29980650]
60. Meacham CE, Morrison SJ: Tumour heterogeneity and cancer cell plasticity. *Nature* (2013) 501(7467):328–337. [PubMed: 24048065]
61. Hanahan D, Weinberg RA: Hallmarks of cancer: The next generation. *Cell* (2011) 144(5):646–674. [PubMed: 21376230]
62. Bendall SC, Nolan GP: From single cells to deep phenotypes in cancer. *Nat Biotechnol* (2012) 30(7):639–647. [PubMed: 22781693]
63. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, Alizadeh AA: Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods* (2015) 12(5):453–457. [PubMed: 25822800]
64. Abbas AR, Wolslegel K, Seshasayee D, Modrusan Z, Clark HF: Deconvolution of blood microarray data identifies cellular activation patterns in systemic lupus erythematosus. *PloS one* (2009) 4(7):e6098. [PubMed: 19568420]
65. Bertsekas DP: Nonlinear programming Athena scientific Belmont, (1999).
66. Qiao W, Quon G, Csaszar E, Yu M, Morris Q, Zandstra PW: Pert: A method for expression deconvolution of human blood samples from varied microenvironmental and developmental conditions. *PLoS computational biology* (2012) 8(12):e1002838. [PubMed: 23284283]
67. Holland PW, Welsch RE: Robust regression using iteratively reweighted least-squares. *Communications in Statistics-theory and Methods* (1977) 6(9):813–827.
68. Liebner DA, Huang K, Parvin JD: Mmad: Microarray microdissection with analysis of differences is a computational tool for deconvoluting cell type-specific contributions from tissue samples. *Bioinformatics* (2013) 30(5):682–689. [PubMed: 24085566]
69. Zhong Y, Wan Y-W, Pang K, Chow LM, Liu Z: Digital sorting of complex tissues for cell type-specific gene expression profiles. *BMC bioinformatics* (2013) 14(1):89. [PubMed: 23497278]
70. Baron M, Veres A, Wolock SL, Faust AL, Gaujoux R, Vetere A, Ryu JH, Wagner BK, Shen-Orr SS, Klein AM, Melton DA et al.: A single-cell transcriptomic map of the human and mouse pancreas reveals inter- and intra-cell population structure. *Cell Syst* (2016) 3(4):346–360 e344. [PubMed: 27667365]
71. Li B, Severson E, Pignon JC, Zhao H, Li T, Novak J, Jiang P, Shen H, Aster JC, Rodig S, Signoretti S et al.: Comprehensive analyses of tumor immunity: Implications for cancer immunotherapy. *Genome biology* (2016) 17(1):174. [PubMed: 27549193]
72. Wang X, Park J, Susztak K, Zhang N, Li M: Bulk tissue cell type deconvolution with multi-subject single-cell expression reference. *bioRxiv* (2018) 354944.
73. Wang J, Huang M, Torre E, Dueck H, Shaffer S, Murray J, Raj A, Li M, Zhang NR: Gene expression distribution deconvolution in single-cell rna sequencing. *Proc Natl Acad Sci U S A* (2018) 115(28):E6437–E6446. [PubMed: 29946020] • The MUlti-Subject Single Cell deconvolution (MuSiC) method utilizes cross-subject scRNA-seq to estimate cell type proportions in bulk RNA-seq data [72]. A key concept in MuSiC is “marker gene stability”.
74. Wu H, Humphreys BD: The promise of single-cell rna sequencing for kidney disease investigation. *Kidney Int* (2017) 92(6):1334–1342. [PubMed: 28893418]

Key points

- Transcriptome survey studies of human kidney samples have facilitated the characterization of global changes in gene expression in kidney disease states and phenotypes.
- Expression of quantitative trait analysis (eQTL) is a powerful method to identify genetic variation driven gene expression changes. Integration of eQTL studies with GWAS studies can also identify causal genes for disease development.
- New single cell RNAsequencing methods can quantify transcript levels in thousands or millions of cells. Integrating single cell and bulk RNAsequencing methods can identify cell proportion and cell type specific changes in disease states.

Figure 1: Gene expression and phenotype analysis of human kidney samples.

Samples can be obtained from a broad range of individuals with clinical and histological description. The analysis can include whole kidney, micro-dissected tubule and glomerulus compartment or single cell samples. Association between gene expression and phenotype can be studied.

Figure 2: Integration of genotype and gene expression information allows the identification of causal genes for kidney disease.

A. Genotype and phenotype association is established in GWAS studies and genotype and gene expression analysis is correlated in eQTL studies. Loci associated with phenotype and gene level changes can highlight causal genes for kidney disease B. Genotype (CKD-associated GWAS SNP rs170563) and phenotype (eGFR) association; C. Genotype (rs170563) and gene expression (MANBA) association in human kidney.

Figure 3: Cell fraction changes can be analyzed in whole kidney samples
Multiple different computational methods can be used to define cell composition changes in transcriptome data obtained from whole tissue samples. These methods rely on pre-selected cell type-specific marker genes. CIBERSORT adopted a linear support vector regression. BSEQ-sc extended CIBERSORT by using scRNA-seq gene expression as a reference.