Big science and big data in nephrology

Check for updates

Julio Saez-Rodriguez[1,2,3], Markus M. Rinschen[4,5], Jürgen Floege[6] and Rafael Kramann[6,7]

OPEN

[1]RWTH Aachen University, Faculty of Medicine, Joint Research Centre for Computational Biomedicine (JRC-COMBINE), Aachen, Germany; [2]Institute for Computational Biomedicine, Heidelberg University, Faculty of Medicine, and Heidelberg University Hospital, Heidelberg, Germany; [3]Molecular Medicine Partnership Unit (MMPU), European Molecular Biology Laboratory and Heidelberg University, Heidelberg, Germany; [4]Department II of Internal Medicine, and Center for Molecular Medicine Cologne, University of Cologne, Cologne, Germany; [5]Center for Mass Spectrometry and Metabolomics, The Scripps Research Institute, La Jolla, California, USA; [6]RWTH Aachen, Department of Nephrology and Clinical Immunology, Aachen, Germany; and [7]Department of Internal Medicine, Nephrology and Transplantation, Erasmus Medical Center, Rotterdam, The Netherlands

There have been tremendous advances during the last decade in methods for large-scale, high-throughput data generation and in novel computational approaches to analyze these datasets. These advances have had a profound impact on biomedical research and clinical medicine. The field of genomics is rapidly developing toward single-cell analysis, and major advances in proteomics and metabolomics have been made in recent years. The developments on wearables and electronic health records are poised to change clinical trial design. This rise of 'big data' holds the promise to transform not only research progress, but also clinical decision making towards precision medicine. To have a true impact, it requires integrative and multi-disciplinary approaches that blend experimental, clinical and computational expertise across multiple institutions. Cancer research has been at the forefront of the progress in such large-scale initiatives, so-called 'big science,' with an emphasis on precision medicine, and various other areas are quickly catching up. Nephrology is arguably lagging behind, and hence these are exciting times to start (or redirect) a research career to leverage these developments in nephrology. In this review, we summarize advances in big data generation, computational analysis, and big science initiatives, with a special focus on applications to nephrology.

Kidney International (2019) **95,** 1326–1337; https://doi.org/10.1016/j.kint.2018.11.048

KEYWORDS: chronic kidney disease; gene expression; proteomic analysis

Copyright © 2019, International Society of Nephrology. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Correspondence: *Julio Saez-Rodriguez, Institute for Computational Biomedicine, University Hospital Heidelberg, BioQuant, Im Neuenheimer Feld 267, 69120 Heidelberg, Germany. E-mail: julio.saez@bioquant.uni-heidelberg.de or Rafael Kramann, Division of Nephrology and Clinical Immunology, Medical Faculty RWTH Aachen University, Pauwelsstrasse 30, 52074 Aachen, Germany. E-mail: rkramann@gmx.net*

Received 1 June 2018; revised 11 November 2018; accepted 20 November 2018; published online 5 March 2019

In the past decade, tremendous progress has been made in technological developments in the areas of large-scale molecular data generation and computational analysis. These advancements have led to an era of "big data," which in turn is fueling "precision medicine." This approach already has improved diagnosis, risk assessment, and treatment of multiple diseases, most notably in oncology, an area in which such information already is used in clinical practice.[1] This era offers the opportunity to develop novel diagnostic and therapeutic tools for kidney disease patients, as patients with the same diagnosis following biopsy often present different symptoms and tremendous variability in disease progression and response to therapy. They therefore represent a perfect patient population for application of precision medicine.[2]

Nephrology lags behind other areas in big data analyses, for various reasons. First, kidney damage is highly multifactorial, has complex and overlapping clinical phenotypes and morphologies, is often diagnosed late, and progresses chronically. Second, although biopsies are routinely taken, the amount of material and the conservation conditions (e.g., paraffin-embedding) limit the molecular profiling that can be done. Also, the kidney is an intricate organ with multiple specialized cell populations that have complex physiology.

Editor's Note

The power of big science, large consortia, and machine learning to create breakthroughs in disease management has been most prominently applied to the field of oncology. The same tools can be applied to nephrology, and the Editors believe these will be important for our readership. Thus, over the next several months, *Kidney International* will feature in-depth reviews on big science, artificial intelligence, and machine learning. This review, the first in this series, provides a broad overview of the field to introduce our readership to the concepts and frameworks of big science in nephrology.

Thus, extrapolation of bulk omics data from kidney tissue to specific (patho-)physiological processes is difficult. Third, the case of big data arguably is clearer in other areas, such as oncology, for which it has already proven valuable in the clinic and a relatively established technology (genetic sequencing) profiles the (often few) driving events of the disease. In contrast, big data efforts for kidney diseases such as proteome-driven urinary biomarkers have hardly changed clinical practice so far. Finally, and largely because of the reasons just cited, kidney disease does not get the same level of funding as other pathologies, especially in proportion to its prevalence. The level of funding is related to the relatively low awareness of the public and funding agencies, and the limited number of public–private partnerships.[3] Given that all these reasons are intertwined and affect one another, we believe that a successful big data project can break this vicious cycle.

In this review, we discuss various methods and strategies that involve generation and computational mining of big data that might transform diagnosis, prognosis, and treatment in nephrology (Figure 1). We also present the related concept of "big science," the joint effort of large consortia to generate big data to help reach a common goal, and discuss how this can have a profound impact in nephrology.

ADVANCES IN DATA GENERATION

Recent technological advances allow us to generate tremendous amounts of data, in particular "omics" data.[2,4–6] We summarize some areas that we think might be of particular interest for nephrology (Table 1[7–32]; Figure 2).

Genomics

Multiple technologies can measure the genome and its alterations. First applications in nephrology consisted of hypothesis-driven candidate gene studies, such as angiotensinogen, although results were not replicable.[33] The development of relatively inexpensive genotype arrays and the availability of samples in biobanks allowed performance of genome-wide association studies in many patients, providing important insights into risk factors and the pathogenesis of multiple kidney diseases.[33–36] Technological developments have enabled sequencing of the protein-coding regions (exons), roughly 1% of the genome (whole-exome sequencing [WES]), and even whole-genome sequencing (WGS).

WGS and WES are comprehensive technologies that inform us about substitutions, deletions, insertions, duplications, copy number changes, inversions, and translocations, providing a fairly complete view of the genome and its alterations. Currently, the information provided by WGS has outpaced our ability to interpret genetic variation, which may explain why genome sequencing is not widely used in clinical medicine in general and nephrology in particular. In addition, genetic alterations, especially those not in the coding regions, are hard to characterize functionally. Hence, genome sequencing typically provides only limited insight into functioning and disease.

Prenatal testing, diagnosis of Mendelian disorders, and cancer are the areas in which WGS and WES strategies have been introduced effectively in clinical practice.[37] Pediatric nephrology, which often confronts patients and practitioners with trying to understand the genetic causes of end-stage renal disease, probably has the most applications within nephrology for WGS and WES.[38] However, there are also compelling reasons for a thorough genetic workup in adult nephrology,[39] as inherited kidney disease accounts for approximately 10% of end-stage renal disease in adults. For

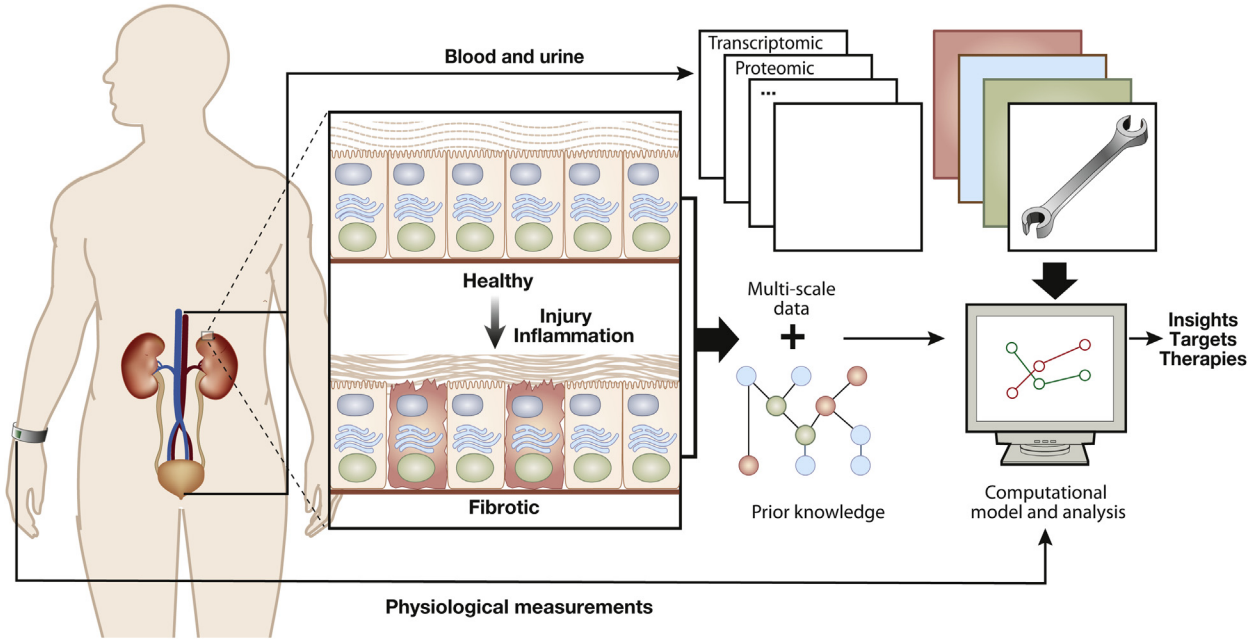

Figure 1 | Overview of data generation and analysis for nephrology.

Table 1 | Selected technologies to generate large datasets with relevance for nephrology

Name	Definition	Examples in nephrology (reference)]
Genome-wide association studies (GWAS)	Observational study of genome-wide set of genetic variants in different individuals; aims to see if any variant is associated with a trait, such as disease	7,8
Whole-exome sequencing (WES)	Sequencing of the protein coding DNA (exons) (roughly 1% of the human genome)	9
Whole-genome sequencing (WGS)	Sequencing of the entire DNA	—
Genetic perturbation screening	Forward screening: maps specific genetic perturbations to a phenotype of interest. Genetic perturbations are mostly performed by large small, interfering RNA or short hairpin RNA libraries or by CRISPR/Cas9 combined with gRNA libraries.	10
Microarray	Multiplex lab—on a chip, can be used for DNA or protein, for example. DNA-microarrays are a collection of microscopic DNA spots (probes) attached to a solid surface that hybridize a cDNA or cRNA. Hybridization is usually detected with a fluorophore or chemiluminescent substrate.	11 12 13
Bulk RNA-seq	Pooled RNA-seq from a bulk of cells; is obtained from synthesis of DNA from RNA, and subsequent DNA sequencing	14
Single cell-RNA-seq	Sequencing of RNA from single cells, mostly performed from sorted single cells or with Microfluidics and barcoding of single cells (e.g., DropSeq, 10x chromium).	15,16,17
Targeted proteomics	A defined set of proteins (and/or their modification) is quantified in kidney tissue or body fluids using mass spectrometry, antibodies, or aptamers. It can be applied to tissues and biofluids, and to very small sample amounts.	18–21
Untargeted proteomics	In a discovery approach, proteins (and/or their modification) are identified and quantified using mass spectrometry based on stochastic frequency. It can be applied to tissues and biofluids but requires relatively large amount of material.	18,22–24
MALDI-IMS	Matrix-assisted laser desorption ionization (MALDI) - imaging mass spectrometry (IMS) allows us to obtain from a sample, typically a tissue, spatial information on the distribution of multiple molecules.	25
Untargeted metabolomics	Identification and quantification of metabolites using nuclear magnetic resonance or mass spectrometry	26
Targeted metabolomics	Targeted, mass spectrometry based quantification of metabolites	27,28
Imaging	Process of forming images, such as electron, light, and immunofluorescence microscopy, ultrasound, computer tomography, magnetic resonance imaging	29
High-throughput screening	Test compounds at large scale and study their effect on phenotype to screen potential treatment candidates.	30
Wearables	An item that can be worn and tracks information such as heart rate, physical activity, blood pressure	Reviewed in[31]
Electronic health records	Longitudinal electronic collection of health information of patients or cohorts	32

CRISPR, clustered regularly interspaced short palindromic repeats.

example, 2 common genetic variants of the *APOL1* gene have been attributed to the highly increased end-stage renal disease risk in individuals of sub-Saharan African descent.[40,41]

Transcriptomics

The development of microarrays, and later RNA sequencing (RNA-seq) has made transcriptomics (the measurement of all RNA transcripts[42]) widely accessible.[43,44] RNA-seq is currently the leading technology that allows, in contrast to microarrays, coverage, in principle, of the whole genome of any organism. This rapid development, along with decreasing costs of sequencing, has led to an immense growth of data and paved the way for profound discoveries in biomedicine.[45]

The high level of coverage (transcriptome-wide), at relatively low cost and moderate complexity (with well-established workflows for data generation and analysis), is the major advantage of transcriptomics technologies. One major weakness is that not every mRNA leads to expression of the corresponding protein; thus, measured mRNA expression might not correspond to functional effect. However, measurement of non-coding RNA can also be interpreted as a strength, as various lines of evidence suggest that non-coding RNA is important in homeostasis and disease.

Many studies using transcriptomics in nephrology have improved our knowledge of disease initiation, progression, and potential novel biomarkers and treatments too numerous to mention in the limited space here. As a hallmark study, Tuttle *et al.* and Woroniecka *et al.* analyzed microarrays of 95 microdissected (tubular and glomerular fraction) human kidney samples.[12,46] They observed lower expression of key enzymes and regulators of fatty acid oxidation in the tubular fraction and demonstrated that restoring it genetically or pharmacologically protected mice from interstitial kidney fibrosis.[11] Although clinical proof is still missing, this study suggests that correcting fatty acid oxidation might be a

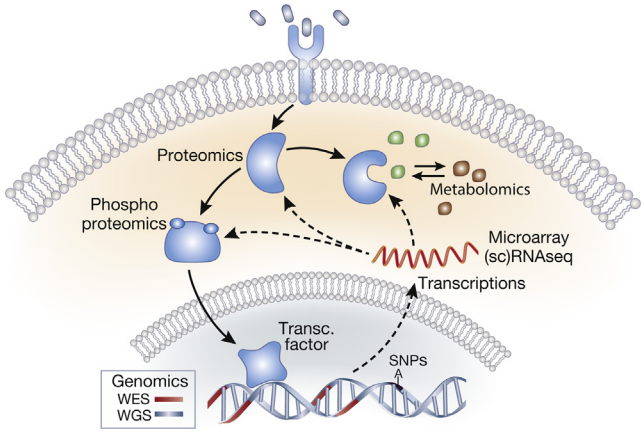

Figure 2 | The most common omics technologies and the information they provide regarding molecular processes. scRNA, single-cell RNA; SNP, single-nucleotide polymorphism; Transc., transcription; WES, whole-exome sequencing; WGS, whole-genome sequencing.

strategy to prevent kidney fibrosis and chronic kidney disease (CKD).

Transcriptomics can be linked to genetic data in so-called expression-quantitative trait loci, which uncover how alterations on the genome affect expression of genes and thereby provide insight on the effect of genetic variation. Recent studies identified lysosomal beta A mannosidase (MANBA) as a potential target in CKD[47] and identified many genes involved in nephrotic syndrome.[48] Another expression-quantitative trait loci study separating glomeruli and tubules identified disabled-2 (DAB2), an adaptor in the transforming growth factor (TGF)-β pathway, as a key protein in CKD.[49]

Single-cell transcriptomics

Techniques for high-throughput single-cell RNA sequencing (scRNA-seq) are under rapid development (summarized in Table 2[50,51]), allowing us to interrogate individual cells with unprecedented resolution.[5,50,52] In addition, the emerging field of spatial transcriptomics[53,54] enables measurement of RNA directly in tissues with spatial resolution. Plate-based methods can offer full-length coverage.[55] Other protocols sacrifice full-length coverage to increase throughput by barcoding of libraries and thus multiplexing of the amplification.[56] Recently, droplet-based microfluidic platforms have been developed that allow the encapsulation of cells together with barcoded DNA oligonucleotides and cell lysis buffer. These technologies enable a tremendously high throughput at reduced cost per cell. Every scRNA-seq technology has advantages and disadvantages, including throughput, costs, and coverage of the transcriptome.[50,51]

The strength of scRNA seq is clearly that it enables measurement of mRNA expression in individual cells. The current level of knowledge regarding the transcriptional landscape in kidney homeostasis and disease is primarily derived from profiling of whole or microdissected tissue using microarrays or RNA-seq. These studies are important but limited, as they describe only an average gene expression across the tremendous heterogeneity of kidney cell types.

One major limitation of scRNA-seq is that no current method allows measurement of the entire transcriptome in individual cells. Further, the quality of the scRNA-seq data relies heavily on various factors, including the tissue dissociation protocol. Validation is hence critical and can be achieved by such methods as immunostaining and/or spatial transcriptomics.

Various groups have already started to use scRNA-seq with mouse and human kidneys to identify novel cell populations, critical cell–cell interaction pathways, and potential therapeutic targets. Analysis of scRNA-seq data from 57,979 murine kidney cells demonstrated that Mendelian disease genes show cell-specificity, for example, podocyte-specific expression of various homologs of genes associated with monogenic inheritance of proteinuria in humans.[15] In addition, the authors reported a new transitional cell type in the collecting duct of adult mice that generates a spectrum of cell types, and revealed that Notch signaling might be critically involved in the collecting duct cell plasticity that drives metabolic acidosis in CKD.[15] Another recent study used scRNA-seq to characterize the mouse glomerulus, identifying novel marker genes for glomerular cell types, and a new subset of endothelial cells.[16] In humans, scRNA-seq data have been used to identify cell type–specific markers,[57] find novel segment-specific proinflammatory responses in kidney allograft rejection[58] and lupus,[59] and compare kidney organoids and kidney.[60]

As these pioneering studies illustrate, scRNA-seq from kidney tissue in homeostasis and disease will help illuminate the complex renal (patho-)physiology, identify novel cell populations, and develop novel therapeutics.[5,50,52] In

Table 2 | Features of major single-cell RNA methods

Method	Smart-seq (v1/v2)	Cell-seq (v1/v2)	STRT	MARS-seq	SCRB-seq	DropSeq	10x genomics	In drops
Cell isolation	Plate-sort or C1 fluidigm (v1)	Plate-sort or C1 fluidigm	C1 (fluidigm)	Plate sort			Microfluidics™	
UMI	No	Yes	Yes (v2)	Yes	Yes	Yes	Yes	Yes
Full length	Yes	No	No	No	No	No	No	No
cDNA amplification	PCR	IVT	PCR	IVT	PCR	PCR	PCR	IVT
Profiling capacity (number of cells)	Low	Low	Low	Medium	High	High	High	High

Table is based on references.[50,51]
IVT, *in vitro* transcription; MARS-seq, massively parallel single-cell RNA sequencing; PCR, polymerase chain reaction; SCRB-seq, single-cell RNA barcoding and sequencing; STRT, single-cell tagged reverse transcription; UMI, unique molecular identifier (short nucleotide sequence that tags individual mRNA molecules to distinguish original molecules from amplified ones).

Table 3 | Selected computational concepts and nephrology-specific resources

Methods	Definition	Examples in nephrology (reference)
Dimensionality reduction	Reduction of variables (e.g., genes) to be considered, using methods such as PCA or t-SNE	Ubiquitous when analyzing omics data (e.g.[15])
Pathway analysis	Estimation of activity of pathways from omics data; can be seen as a "biology-driven" dimensionality reduction	Also broadly used (e.g.[11,91,92])
Deep learning	Subset of machine learning methods built in a hierarchical manner to automatically select features	[93]

Resources	Definition
Nephroseq[94] (www.nephroseq.org)	Integrative platform with genetic, gene expression, and clinical data in nephrology
NephQTL[48] (http://nephqtl.org/)	Database of *cis*-eQTLs of the glomerular and tubulointerstitial tissues of the kidney found in 187 participants in the NEPTUNE cohort
Human kidney eQTL (http://18.217.22.69/eqtl)	eQTL database of microdissected healthy human kidney glomeruli and tubuli
Mouse kidney single-cell atlas[15] (http://18.217.22.69/sc)	Gene expression in ~70,000 individually sequenced cells from healthy mouse kidneys
A single-cell transcriptome atlas of the mouse glomerulus[16] (https://shiny.mdc-berlin.de/mgsca/)	Gene expression from ~3000 single-cell transcriptomes from wild-type mouse glomeruli

For a detailed lists of tools and resources, see, for example, reference.[6]
eQTL, expression-quantitative trait loci; NEPTUNE, Nephrotic Syndrome STudy Network; PCA, principal component analysis; t-SNE, t-distributed stochastic neighbor embedding.

addition, scRNA-seq provides information that can be used to reanalyze bulk transcriptomic data. For example, the signatures of specific cell types from scRNA-seq can be used to estimate the contribution of those cell types from bulk RNA samples.[61] Given the currently high cost of scRNA-seq, a hybrid approach, with a few samples profiled with scRNA-seq and many more with bulk RNA-seq, can be a practical strategy.[5]

Genome-wide genetic perturbation screens

Systematic high-throughput genetic perturbation technologies can help tremendously in illuminating gene function and epigenetic regulation and identifying novel therapeutic targets. They allow us to determine whether particular genes are responsible for a certain cellular phenotype. The most common methods are RNA interference (RNAi) and clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9. In particular, RNAi using short hairpin RNAs allows high-throughput repression at the transcriptional level, thereby reducing gene expression. In contrast, CRISPR/Cas9 edits the genome to truly knockout or modify genes. Of note, novel techniques also allow utilization of CRISPR/Cas9 outside of gene editing to activate or inhibit expression of genes.[62]

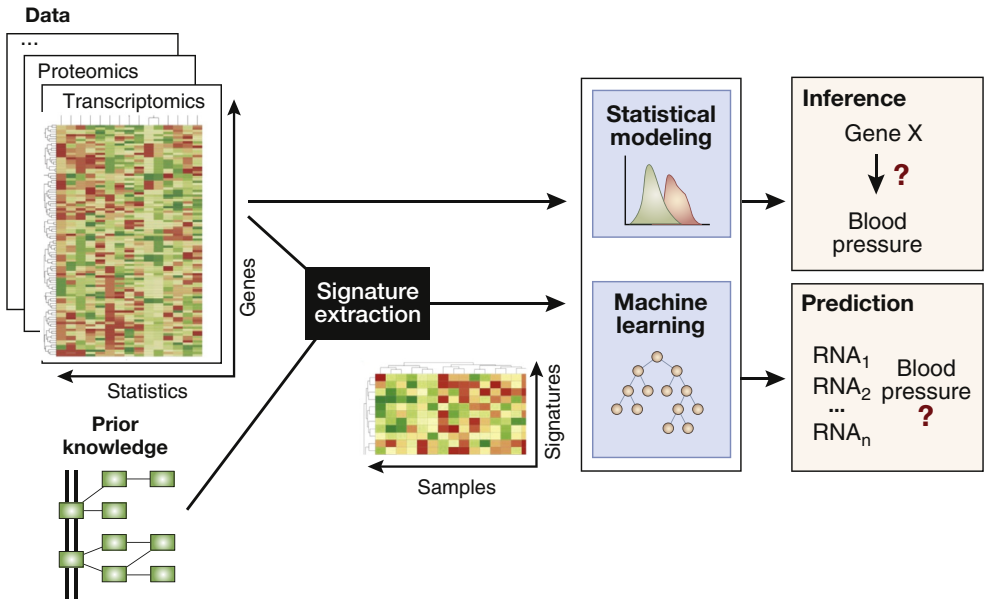

Figure 3 | Computational strategies to apply statistics and machine learning to big data. Use of prior knowledge to extract molecular signatures can facilitate subsequent analyses.

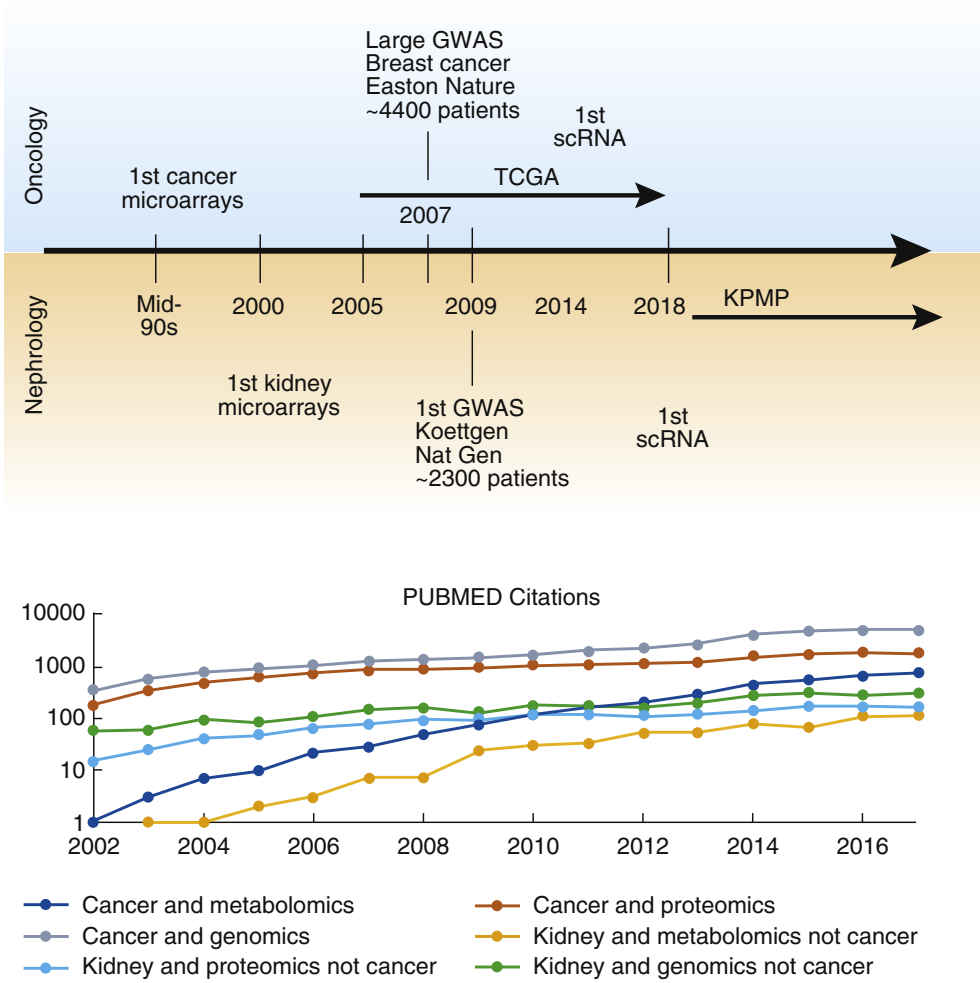

Figure 4 | Timeline of technological developments and their incorporation into basic and translational research in the renal and oncology ("cancer") fields. GWAS, Genome-Wide Association Study; KPMP, Kidney Precision Medicine Project; Nat Gen, Nature Genetics; scRNA, single-cell RNA; TCGA, The Cancer Genome Atlas.

The power of these technologies is that they allow a systematic, genome-wide investigation of the function of genes. In particular, CRISPR/Cas9-based tools have rapidly developed recently to perform genetic editing at high throughput.[63] A limitation of RNAi is that it results in only incomplete knockdown of transcription and indicates high off-target activity, resulting in a low signal-to-noise ratio.[64] A potential limitation of CRISPR-based screenings is that double-strand breaks generated by the Cas9 nuclease produce gene-independent DNA damage phenotypes and thus false positives.[65,66] Further, set up of CRIPR/Cas9 screenings is challenging, including, for example, establishment of a reporter for the phenotype of interest and the culturing and sorting of several hundred million cells.

As an example of a perturbation screen in nephrology, a library of 80,000 short hairpin RNAs targeting roughly 16,000 human genes was used to identify genes whose suppression improves survival of kidney epithelium in *in vitro* models of oxygen and glucose deprivation.[10] Pharmacologic inhibition of NK1R, the product of the *TACR1* gene, one of the hits from the screen, was protective in a mouse model of renal ischemia.[10] This example illustrates that unbiased screenings can provide novel therapeutic candidates.

Proteomics and metabolomics

Proteomics. Mass-spectrometry (MS)-based, antibody-based, or aptamer-based technologies are used to quantify protein expression. Protein modifications, protein–protein, or protein–small molecule interactions can be quantified. Data acquired using MS are usually untargeted, meaning that signals are picked up stochastically and typically do not cover the entire proteome. Targeted proteomics, in contrast, provide acquisition and quantification of sets of predefined peptides with high reliability.

Recent ultrasensitive methods can acquire proteomics from very small, sub-biopsy samples and are useful for investigating tissue heterogeneity and the mechanisms driving disease. Coupling antibodies and MS via mass cytometry,[67] even a single-cell resolution of dozens of proteomic markers can be measured. Complex posttranslational modification

analyses, including that of the phosphoproteome and the degradome, have been applied only recently to kidney[68] and will likely generate major insights.

Matrix-assisted laser desorption/ionization imaging MS allows us to obtain spatially resolved information on proteins and other analytes directly from tissues.[25,69] Similarly, mass cytometry can be applied to tissues.[70,71] These technologies provide a much richer profiling than standard immunohistochemistry methods and can have a profound impact on understanding pathology.

Compared with transcriptomics, proteomics has 3 major advantages. First, transcript numbers only partially explain the abundance of proteins, particularly in very dynamic systems. Second, posttranslational modifications of proteins can be analyzed that indicate their activation, and interactions among proteins, or of proteins with small molecules, DNA, and RNA, can only be measured using proteomics. Third, the proteome also captures microenvironmental alterations that can occur independently of the transcriptome.

Proteomics has several limitations. Despite great advances, proteomics does not provide genome-wide coverage. Large-scale proteomics has a bias toward high-abundance proteins, and targeted strategies should be used to quantify low-abundance proteins in small samples. The role of noncoding RNAs can be assessed only using transcriptomics. The incompleteness of acquisition, even when peptides from every protein are observed, poses a challenge for statistical analysis.

Applications of proteomics in nephrology include analyses of biopsies to better characterize and reclassify amyloidosis and fibrillar glomerulopathy and obtain novel biomarkers.[72–74] Proteomic methods also unraveled the identity of the phospholipase A2 receptor (PLA2R) antigen in membranous nephropathy.[75] Proteomics has also been used for mechanistic studies. For instance, proteomics of renal biopsies from patients who had diabetic nephropathy revealed a change of pyruvate kinase M2 that may lead to a mechanistic understanding of potential protective signatures in diabetes.[76] First applications of deep proteomics include detailed characterization of human and mouse podocytes[77] that led to the discovery of FARP1, a previously unrecognized podocyte-specific protein with a potential role in proteinuric kidney disease.[77] An ongoing area of research focuses on posttranslational protein modifications in glomerular tissue. In particular, large-scale phosphoproteomes,[78] integrated with genomic information, can be used to predict clinically relevant phosphorylation sites. Although very informative, phosphoproteome experiments typically require significant amounts of material and are more labor intensive.

Analyses of biofluids, such as urine or serum, require specialized high-throughput setups, owing to the challenging nature of the samples and the high dynamic range of the molecules, but already have yielded first signatures that might be able to predict CKD progression in individual patients.[2] For instance, delayed graft function was successfully predicted using a targeted urinary protein assay for 167 proteins.[79,80] Exosome compartments are an alternative source of rich information.[81] Urinary epidermal growth factor protein was associated with progression of CKD as an independent predictor.[13] Predictive protein signatures in the plasma are emerging as biomarkers as well.[82,83]

Metabolomics. The metabolome, that is, the entity of metabolites and small molecules, is particularly informative because it is commonly considered to be a readout for protein function and thereby a rather suitable source for biomarkers. Compared with proteomic information, metabolomic data from body fluids and tissues are much easier to acquire. Common tools include MS, and with lower resolution, nuclear magnetic resonance.

A general strength of metabolomics data is their association with the phenotype. Generally, metabolomics enables robust acquisition of small molecules that enable rather high throughput. Another advantage is the transferability of molecules between species, as many metabolites are conserved. In addition, both technical approaches have their own strengths: MS-based approaches are generally more sensitive (detection of femto- or attomolar concentrations), and nuclear magnetic resonance–based approaches can be more robustly quantified.

Metabolomics analysis also has limitations. In general, it is not as advanced as the other omics methods. Metabolite extraction for MS is *per se* incomplete—sample preparation needs to be adjusted to the needs of the analysis, and analytical aspects need to be considered. Untargeted MS-based metabolomics approaches rely on high-resolution MS, advanced algorithms, and comprehensive standards and spectral databases that are still evolving. In contrast, targeted metabolomics, although accurate, usually covers only a limited number of analytes. Nuclear magnetic resonance–based metabolomics has less sensitivity compared with MS-based methods. Overall, metabolomics datasets are highly influenced by many factors, including genotype, lifestyle, circadian rhythm, the small molecules the patient is exposed to (exposome), drugs, diet, and most notably, the microbiome. All these potential confounding factors make the analysis challenging, and the role of these cues in renal disease and physiology are just starting to be elucidated.[84,85]

Given the central function of the kidney in human metabolism, including filtering toxins and reabsorbing nutrients, it is not surprising that several metabolites are associated with a decline in renal function; these metabolites provide an ever-increasing arena in which to find novel predictors of renal decline and disease progression.[86,87] In addition, very recent integrative studies have linked metabolomic profiles to variants in genomic information, suggesting tight associations (m genome-wide association studies), for example, between lysine abundance in urine and variants in solute carrier transporters.[27] Urinary glycine and histidine concentrations, for instance, were also found to correspond to outcomes in the Framingham offspring cohort.[28] Whereas metabolites are commonly regarded as the output of protein activity, they can also modulate biological systems and thereby drive phenotypes.[88] This activity is particularly important in the kidney

where, for example, metabolite receptors for succinate and alpha-ketoglutarate are chief modulators of nephron function.[89,90]

COMPUTATIONAL TOOLS AND METHODS

The scope and complexity of these data require adequate infrastructure to handle and process them, and analytical tools to mine them. We briefly summarize some concepts in this area (Table 3[6,11,15,48,91–94]).

Extracting information from omics data

To extract the information, each data type requires specific analytic methods to process and normalize it. Once normalized, various methods, such as principal component analysis or t-distributed stochastic neighbor embedding, help make visualization of large datasets easier and identify underlying patterns.

A common aim of computational methodologies is to compress the data into fewer variables to filter noise and increase statistical power, which are both important because omics datasets often have a larger number of measured variables than samples (Figure 3). This compression can be done in a purely data-driven manner,[95] although the resulting features might not be easily interpretable. Therefore, the data are often subject to a functional analysis, with the aim of grouping the observed values into biological processes or areas, such as cellular pathways and networks.[96–98] Many methods have been developed, particularly within oncology, but in many cases, they are transferable to nephrology. As with cancer, kidney pathologies are often characterized by the deregulation of signaling, gene regulatory processes, and metabolic processes, and largely involve the same pathways (transforming growth factor, epidermal growth factor, Wnt/Notch, Hedgehog, etc.). An important difference is that somatic mutations do not drive kidney diseases (with some exceptions, such as cystic kidney disease[99]).

The deregulation typically has a subtler origin and is less strongly manifested, resulting in less-distorted phenotypes. This subtlety makes analysis of the phenotypes more challenging, and combining multiple omics in particular can provide an integrated view on signaling, regulatory, and metabolic mechanisms within the cell, all of which can be deregulated in CKD. The integration of multiple omics is challenging and under active development.[100–102]

Statistics and machine learning

Once big data have been adequately processed, and potentially compressed into fewer and more easily interpretable features such as pathways, they are typically analyzed to identify differences across samples (e.g., comparing CKD patients with healthy patients), or to predict a phenotype of interest (e.g., estimated glomerular filtration rate). Statistical models such as analysis of variance or linear regression models are standard yet suitable tools for these tasks. Alternatively, more complex computational methods can be applied, often from machine learning. Although the line between statistical models and machine learning can be blurry, typically the former are more suitable to learn about underlying processes (e.g., to determine if a given gene affects blood pressure), whereas the latter focus on prediction (e.g., provide an expected blood pressure level as accurately as possible[103]; Figure 3).

Exciting advances have been made recently in machine learning thanks to so-called deep learning techniques. These methods can automatically derive informative features from many types of raw data, a task that otherwise requires domain expertise, leading to major performance improvements across many fields, from speech recognition to reconstruction of brain circuits and sentiment analysis.[104] Deep learning techniques are increasingly applied to biomedical data, from image processing to genomic data analysis.[105] For example, when trained with 120,000 images, benign nevi can be distinguished from malignant melanoma with the accuracy of an experienced dermatologist.[106] Similarly, such methods might outperform pathologists' fibrosis scores from histological renal biopsy images.[93] Deep learning has also been applied to EHRs for solving problems such as extraction of information from EHRs, prediction of disease outcome, and de-identification.[107] In nephrology, the analysis of images and EHRs will probably benefit most from application of these approaches. Omics data analysis will likely benefit as well, once enough high-quality data are generated. Many deep learning methods exist already or are under active development. Most of them are based on artificial neural networks—sets of interconnected nodes (neurons), arranged in layers. A prominent technique is convolutional neural networks,[105] built with multiple layers of neurons that share parameters and are connected to only a few neurons in the previous layer. Recurrent neural networks are designed to model sequentially ordered data such as time series.[107] Although deep learning is a very powerful technology, no single method is universally applicable, and conventional approaches are relevant and have advantages, particularly when data are scarce.[105]

The results of computational analysis can be only as good as the data used. Big data, in particular clinical data, often have biases and can be misinterpreted.[108] In addition, large studies and subsequent computational analyses typically find the most prominent alterations in the data. However, they are less suitable for finding infrequent yet potentially critical events. For this, more-focused and knowledge-driven strategies, in particular mechanistic models, should be used to complement statistical and machine learning approaches.[6]

Technical infrastructures for big science

In addition to the challenge of finding adequate analytical methods for large and detailed datasets, they also pose challenges in terms of infrastructure. The huge amount of data imposes considerable cost and logistic requirements. Multiple databases have been developed for omics data, typically focused on one type of data (transcriptomics, proteomics, etc.). Most are general purpose, whereas a few are specific for nephrology,[109] such as Nephroseq.[94]

A fundamental need in storing clinical data is to make them accessible to those performing the analysis and, at the same time, guarantee patients' privacy. A particularly challenging issue in this regard is the identification of patients based on genomic or other molecular data.[110,111] Initiatives such as the Global Alliance for Genomics and Health (GA4GH) are actively working to solve these issues. For example, they are establishing services that allow data to be queried only for specific information, such as the presence of a particular allele.[112] For larger systematic analysis, cryptographic techniques are being developed.[113] Alternatively, so-called virtualization technologies allow scientists to submit their analytical tools to be run remotely on a server, enabling analysis to be performed without having to share the actual data ("the algorithm goes to the data, instead of the data to the algorithm").[114]

BIG SCIENCE PROJECTS

Large consortia in biomedicine

The amount of data required to perform analyses such as those described earlier can rarely be generated by one center

alone. This is due to not only high cost, but also accessibility to enough patients—the richer the characterization desired, the larger the number of patients needed to have enough statistical power. To overcome this issue, consortia are built that share standard operational procedures. In oncology, such initiatives have already been running for a number of years, demonstrating the possibility of generating invaluable resources for the research and clinical community, largely coordinated under the umbrella of the International Cancer Genomic Consortium (ICGC).[115] For example, The Cancer Genome Atlas (TCGA) has recently concluded the characterization of 11,000 tumors of 33 types,[116] including copy number alterations, mutations, transcriptomics, and for a subset, proteomics through the Clinical Proteomic Tumor Analysis Consortium (CPTAC). This initiative has yielded a large number of insights and a formidable resource for oncologists.

Even larger national initiatives are starting to generate molecular and clinical data for massive cohorts. The UK Biobank recruited half a million people from all over the UK, from whom various measurements are taken, including genotyping, and biofluids are collected for future analyses.[117] The characterization will continue, partially in subgroups of patients; for example, 100,000 of the participants are undergoing imaging of major organs. This colossal resource can be used by researchers from around the globe and should help to shed light on multiple health issues. Similar efforts are ramping up in various places, from relatively small countries with homogenous populations, such as Finland (Finngen) and Denmark (Danish National Biobank), to the United States, with the "All of Us" initiative that aims to sign up one million participants[118] by 2022.

Towards big science in nephrology

In nephrology, several consortia gathering human kidney tissue biopsy biobanks have been initiated to perform this collaborative research.[119] Various initiatives that aim for a comprehensive characterization of kidney biopsies for different CKD subtypes have been launched, including NEPTUNE (Nephrotic Syndrome STudy Network), ERCB (European Renal cDNA Bank), EURenOmics, C-PROBE (Clinical Phenotyping and Resource Biobank), PKU-IgAN, and more recently, TRIDENT (for diabetic nephropathy), CureGN (for glomerulopathies), and the NIDDK (National Institute of Diabetes and Digestive and Kidney Diseases) Kidney Precision Medicine Project (KPMP).

These large biobanks can be profiled to obtain novel insights. For example, analysis of human and mouse glomerular transcriptomics revealed activation of the Jak-STAT pathway in diabetic nephropathy,[120,121] leading to a phase 2 clinical trial testing the Jak1/Jak2 inhibitor baricitinib in type 2 diabetics with diabetic kidney disease, with promising preliminary results.[46] In another study, leveraging biopsies of the aforementioned ERCB, C-PROBE, NEPTUNE, and PKU-IgAN consortia, analysis of transcriptomics data led to the identification of urinary epidermal growth factor as an independent risk predictor of CKD progression.[13] Many studies have been performed on *ad hoc* cohorts, including an omics characterization, in particular transcriptomics. Schena and colleagues[91] recently assembled 19 studies with microarrays on microdissections from biopsies from patients with various kidney diseases. Important limitations to integrate *post hoc* in these studies are batch effects and other confounding factors. When we analyzed 5 such studies, we had to perform a very stringent normalization, discarding many genes, to be able to robustly combine them.[92]

Coordinated efforts across multiple centers can overcome these problems, such as the aforementioned KPMP initiative, which has just started to try to develop a rich molecular characterization using cutting-edge technologies of kidney biopsies. This initiative is similar to the TCGA consortium in oncology that was started over a decade earlier. When completed, KPMP will be a major resource in developing precision nephrology (Figure 2).

Crowdsourcing and citizen science

The analysis of such complex data requires a broad set of expertise unlikely to be present in a single group. Although the aforementioned initiatives involve multiple analytical groups, they are not able to involve the whole relevant community, and are even less able to include contributors from outside the community. To maximize the number of scientists that can contribute to solving a complex project, crowdsourcing is becoming increasingly popular. Here, the help of large communities is used to solve problems posed by an organization.[122,123]

Contributions can range from sharing of medical information by participants to supporting development of a resource[124] (such as Wikipedia), to involvement in collaborative competitions called challenges. The latter is an effective strategy to identify the most effective algorithms and best practices to solve computational problems, ranging from determination of a protein structure to prognosis for disease development.[122]

CONCLUSION

The advent of new technologies creates exciting opportunities for nephrology. In this review, we have focused on molecular-based technologies. Imaging technologies are also rapidly evolving, including novel high-resolution ultrasound contrast methods to assess the renal vasculature, and high-performance computer tomography and high-resolution magnetic resonance imaging.[125,126] Although they cannot replace biopsies, they give novel insights into what is occurring in the kidney during disease progression.

Another stream of large-scale data is becoming rapidly available from the continuous monitoring of patients using wearables, leveraging our increased connectivity.[127] These technologies hold strong potential in nephrology to measure physical activity and parameters related to diabetes and cardiovascular status, including blood pressure, blood glucose, peripheral oxygen saturation (SpO_2), and electrocardiogram

results. Some CKD-related parameters may be available soon, such as the level of potassium in sweat.[128,129] These parameters may enable detection of patients at risk[31] and may be used to inform treatment and prognosis, and guide clinical trials.[130] Important challenges remain to be addressed, ranging from accuracy to the lack of an adequate regulatory framework.[130,131]

A vast amount of information is routinely stored for CKD patients in the form of EHRs.[132] The prevalence of CKD, the room for improvement in its detection and management, and the fact that it is largely defined by laboratory data, make CKD ideal for leveraging EHRs.[133] The recent development of devices for cloud-based monitoring facilitates remote participation and enhanced monitoring, paving the way for big data in the context of clinical trials.[134] Major hurdles need to be overcome: data structures are complex and difficult to link to other biomedical resources. Data are also hard to use (e.g., contain errors and missing information), require appropriate reporting,[135] and suffer from the biases of retrospective cohorts.[136]

The kidney field is lagging somewhat behind many other areas in terms of big data usage (Figure 4), in research and especially in diagnostic and treatment decision-making. However, this lag also represents a chance for researchers to begin to work in (or redirect to) the area of nephrology for their career and have a tremendous impact on the field.

Any larger data acquisition endeavor will face current big data problems that include not only infrastructure but also ethical issues. This challenge holds for all kinds of data, in particular EHRs, for which regulating access, sharing, and the balance of commercial value versus open data is important. However, these considerations are not specific to nephrology, and maybe our lag can be turned to the good. To quote Alexander Pope: "Be not the first by whom the new are tried, nor yet the last to lay the old aside."[137]

ACKNOWLEDGMENTS

This work was supported by grants to RK from the German Research Foundation (KR-4073/3-1, SCHN1188/5-1, SFB/TRR57, SFB/TRR219), the European Research Council (ERC-StG 677448), and the State of Northrhine Westfalia (return to NRW), and grants to JF from the German Research Foundation (SFB/TRR57, SFB/TRR219). MMR was supported by the German Research Foundation (RI-2811/1-1 and RI-2811/2-1). JS-R has received funding from the Joint Research Center for Computational Biomedicine (which is partially funded by Bayer AG). Thanks to Nicolas Palacio for help with the figures.

DISCLOSURE

All the authors declared no competing interests.

REFERENCES

1. Jang Y, Choi T, Kim J, et al. An integrated clinical and genomic information system for cancer precision medicine. *BMC Med Genomics*. 2018;l1(suppl 2):34.
2. Mariani LH, Pendergraft WF 3rd, Kretzler M. Defining glomerular disease in mechanistic terms: implementing an integrative biology approach in nephrology. *Clin J Am Soc Nephrol*. 2016;11:2054–2060.
3. Tomilo M, Ascani H, Mirel B, et al. Renal Pre-Competitive Consortium (RPC2): discovering therapeutic targets together. *Drug Discov Today*. 2018;23:1695–1699.
4. Karczewski KJ, Snyder MP. Integrative omics for health and disease. *Nat Rev Genet*. 2018;19:299–310.
5. Kiryluk K, Bomback AS, Cheng Y-L, et al. Precision medicine for acute kidney injury (AKI): redefining AKI by agnostic kidney tissue interrogation and genetics. *Semin Nephrol*. 2018;38:40–51.
6. He JC, Chuang PY, Ma'ayan A, et al. Systems biology of kidney diseases. *Kidney Int*. 2012;81:22–39.
7. Köttgen A, Glazer NL, Dehghan A, et al. Multiple loci associated with indices of renal function and chronic kidney disease. *Nat Genet*. 2009;41:712–717.
8. Kiryluk K, Li Y, Scolari F, et al. Discovery of new risk loci for IgA nephropathy implicates genes involved in immunity against intestinal pathogens. *Nat Genet*. 2014;46:1187–1196.
9. Ashraf S, Kudo H, Rao J, et al. Mutations in six nephrosis genes delineate a pathogenic pathway amenable to treatment. *Nat Commun*. 2018;9:1960.
10. Zynda ER, Schott B, Babagana M, et al. An RNA interference screen identifies new avenues for nephroprotection. *Cell Death Differ*. 2016;23:608–615.
11. Kang HM, Ahn SH, Choi P, et al. Defective fatty acid oxidation in renal tubular epithelial cells has a key role in kidney fibrosis development. *Nat Med*. 2015;21:37–46.
12. Woroniecka KI, Park ASD, Mohtat D, et al. Transcriptome analysis of human diabetic kidney disease. *Diabetes*. 2011;60:2354–2369.
13. Ju W, Nair V, Smith S, et al. Tissue transcriptome-driven identification of epidermal growth factor as a chronic kidney disease biomarker. *Sci Transl Med*. 2015;7:316ra193.
14. Long J, Badal SS, Ye Z, et al. Long noncoding RNA Tug1 regulates mitochondrial bioenergetics in diabetic nephropathy. *J Clin Invest*. 2016;126:4205–4218.
15. Park J, Shrestha R, Qiu C, et al. Single-cell transcriptomics of the mouse kidney reveals potential cellular targets of kidney disease. *Science*. 2018;360:758–763.
16. Karaiskos N, Rahmatollahi M, Boltengagen A, et al. A single-cell transcriptome atlas of the mouse glomerulus. *J Am Soc Nephrol*. 2018;29:2060–2068.
17. Kramann R, Machado F, Wu H, et al. Parabiosis and single-cell RNA sequencing reveal a limited contribution of monocytes to myofibroblasts in kidney fibrosis. *JCI Insight*. 2018;3. pii: 99561.
18. Höhne M, Frese CK, Grahammer F, et al. Single-nephron proteomes connect morphology and function in proteinuric kidney disease. *Kidney Int*. 2018;93:1308–1319.
19. Hoyer KJR, Dittrich S, Bartram MP, et al. Quantification of molecular heterogeneity in kidney tissue by targeted proteomics. *J. Proteomics*. 2019;93:85–92.
20. Carlsson AC, Ingelsson E, Sundström J, et al. Use of proteomics to investigate kidney function decline over 5 years. *Clin J Am Soc Nephrol*. 2017;12:1226–1235.
21. Konvalinka A, Batruch I, Tokar T, et al. Quantification of angiotensin II-regulated proteins in urine of patients with polycystic and other chronic kidney diseases by selected reaction monitoring. *Clin Proteomics*. 2016;13:16.
22. Rinschen MM, Gödel M, Grahammer F, et al. A multi-layered quantitative in vivo expression atlas of the podocyte unravels kidney disease candidate genes. *Cell Rep*. 2018;23:2495–2508.
23. Kota SK, Pernicone E, Leaf DE, et al. BPI fold-containing family A member 2/parotid secretory protein is an early biomarker of AKI. *J Am Soc Nephrol*. 2017;28:3473–3478.
24. Clotet S, Soler MJ, Riera M, et al. Stable isotope labeling with amino acids (SILAC)-based proteomics of primary human kidney cells reveals a novel link between male sex hormones and impaired energy metabolism in diabetic kidney disease. *Mol Cell Proteomics*. 2017;16:368–385.
25. Casadonte R, Kriegsmann M, Deininger S-O, et al. Imaging mass spectrometry analysis of renal amyloidosis biopsies reveals protein co-localization with amyloid deposits. *Anal Bioanal Chem*. 2015;407:5323–5331.
26. Yu B, Zheng Y, Nettleton JA, et al. Serum metabolomic profiling and incident CKD among African Americans. *Clin J Am Soc Nephrol*. 2014;9:1410–1417.
27. Li Y, Sekula P, Wuttke M, et al. Genome-wide association studies of metabolites in patients with CKD identify multiple loci and illuminate tubular transport mechanisms. *J Am Soc Nephrol*. 2018;29:1513–1524.

28. McMahon GM, Hwang S-J, Clish CB, et al. Urinary metabolites along with common and rare genetic variations are associated with incident chronic kidney disease. *Kidney Int*. 2017;91:1426–1435.
29. Peti-Peterdi J, Kidokoro K, Riquier-Brison A. Novel in vivo techniques to visualize kidney anatomy and function. *Kidney Int*. 2015;88:44–51.
30. Sieber J, Wieder N, Clark A, et al. GDC-0879, a BRAFV600E inhibitor, protects kidney podocytes from death. *Cell Chem Biol*. 2018;25:175–184.e4.
31. Wieringa FP, Broers NJH, Kooman JP, et al. Wearable sensors: can they benefit patients with chronic kidney disease? *Expert Rev Med Devices*. 2017;14:505–519.
32. Kashani KB. Automated acute kidney injury alerts. *Kidney Int*. 2018;94:484–490.
33. O'Seaghdha CM, Fox CS. Genome-wide association studies of chronic kidney disease: what have we learned? *Nat Rev Nephrol*. 2011;8:89–99.
34. Wuttke M, Köttgen A. Insights into kidney diseases from genome-wide association studies. *Nat Rev Nephrol*. 2016;12:549–562.
35. Ahlqvist E, van Zuydam NR, Groop LC, et al. The genetics of diabetic complications. *Nat Rev Nephrol*. 2015;11:277–287.
36. Mohan C, Putterman C. Genetics and pathogenesis of systemic lupus erythematosus and lupus nephritis. *Nat Rev Nephrol*. 2015;11:329–341.
37. Shendure J, Balasubramanian S, Church GM, et al. DNA sequencing at 40: past, present and future. *Nature*. 2017;550:345–353.
38. Gulati A, Somlo S. Whole exome sequencing: a state-of-the-art approach for defining (and exploring!) genetic landscapes in pediatric nephrology. *Pediatr Nephrol*. 2018;33:745–761.
39. Groopman EE, Rasouly HM, Gharavi AG. Genomic medicine for kidney disease. *Nat Rev Nephrol*. 2018;14:83–104.
40. Genovese G, Friedman DJ, Ross MD, et al. Association of trypanolytic ApoL1 variants with kidney disease in African Americans. *Science*. 2010;329:841–845.
41. Tzur S, Rosset S, Shemer R, et al. Missense mutations in the APOL1 gene are highly associated with end stage kidney disease risk previously attributed to the MYH9 gene. *Hum Genet*. 2010;128:345–350.
42. Lowe R, Shirley N, Bleackley M, et al. Transcriptomics technologies. *PLoS Comput Biol*. 2017;13:e1005457.
43. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*. 2009;10:57–63.
44. Nelson NJ. Microarrays have arrived: gene expression tool matures. *J Natl Cancer Inst*. 2001;93:492–494.
45. Muir P, Li S, Lou S, et al. The real cost of sequencing: scaling computation to keep pace with data generation. *Genome Biol*. 2016;17:53.
46. Tuttle KR, Brosius FC 3rd, Adler SG, et al. JAK1/JAK2 inhibition by baricitinib in diabetic kidney disease: results from a phase 2 randomized controlled clinical trial. *Nephrol Dial Transplant*. 2018;33:1950–1959.
47. Ko Y-A, Yi H, Qiu C, et al. Genetic-variation-driven gene-expression changes highlight genes with important functions for kidney disease. *Am J Hum Genet*. 2017;100:940–953.
48. Gillies CE, Putler R, Menon R, et al. An eQTL landscape of kidney tissue in human nephrotic syndrome. *Am J Hum Genet*. 2018;103:232–244.
49. Qiu C, Huang S, Park J, et al. Renal compartment–specific genetic variation analyses identify new pathways in chronic kidney disease. *Nat Med*. 2018;24:1721–1731.
50. Wu H, Humphreys BD. The promise of single-cell RNA sequencing for kidney disease investigation. *Kidney Int*. 2017;92:1334–1342.
51. Ziegenhain C, Vieth B, Parekh S, et al. Comparative analysis of single-cell RNA sequencing methods. *Mol Cell*. 2017;65:631–643.e4.
52. Potter SS. Single-cell RNA sequencing for the study of development, physiology and disease. *Nat Rev Nephrol*. 2018;14:479–492.
53. Moor AE, Itzkovitz S. Spatial transcriptomics: paving the way for tissue-level systems biology. *Curr Opin Biotechnol*. 2017;46:126–133.
54. Lein E, Borm LE, Linnarsson S. The promise of spatial transcriptomics for neuroscience in the era of molecular cell typing. *Science*. 2017;358:64–69.
55. Picelli S, Faridani OR, Björklund AK, et al. Full-length RNA-seq from single cells using Smart-seq2. *Nat Protoc*. 2014;9:171–181.
56. Kivioja T, Vähärautio A, Karlsson K, et al. Counting absolute numbers of molecules using unique molecular identifiers. *Nat Methods*. 2011;9:72–74.
57. Sivakamasundari V, Bolisetty M, Sivajothi S. Comprehensive cell type specific transcriptomics of the human kidney. *bioRxiv*. 2017.
58. Wu H, Malone AF, Donnelly EL, et al. Single-cell transcriptomics of a human kidney allograft biopsy specimen defines a diverse inflammatory response. *J Am Soc Nephrol*. 2018;29:2069–2080.
59. Der E, Ranabothu S, Suryawanshi H, et al. Single cell RNA sequencing to dissect the molecular heterogeneity in lupus nephritis. *JCI Insight*. 2017;2:e93009.
60. Wu H, Uchimura K, Donnelly E, et al. Comparative analysis of kidney organoid and adult human kidney single cell and single nucleus transcriptomes. *bioRxiv*. 2017.
61. Schelker M, Feau S, Du J, et al. Estimation of immune cell content in tumour tissue using single-cell RNA-seq data. *Nat Commun*. 2017;8:2032.
62. Wang H, La Russa M, Qi LS. CRISPR/Cas9 in genome editing and beyond. *Annu Rev Biochem*. 2016;85:227–264.
63. Gilbert LA, Horlbeck MA, Adamson B, et al. Genome-scale CRISPR-mediated control of gene repression and activation. *Cell*. 2014;159:647–661.
64. Joung J, Konermann S, Gootenberg JS, et al. Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening. *Nat Protoc*. 2017;12:828–863.
65. Munoz DM, Cassiani PJ, Li L, et al. CRISPR screens provide a comprehensive assessment of cancer vulnerabilities but generate false-positive hits for highly amplified genomic regions. *Cancer Discov*. 2016;6:900–913.
66. Aguirre AJ, Meyers RM, Weir BA, et al. Genomic copy number dictates a gene-independent cell response to CRISPR/Cas9 targeting. *Cancer Discov*. 2016;6:914–929.
67. Spitzer MH, Nolan GP. Mass cytometry: single cells, many features. *Cell*. 2016;165:780–791.
68. Rinschen MM, Hoppe AK, Grahammer F, et al. N-degradomic analysis reveals a proteolytic network processing the podocyte cytoskeleton. *J Am Soc Nephrol*. 2017;28:2867–2878.
69. Gessel MM, Norris JL, Caprioli RM. MALDI imaging mass spectrometry: spatial molecular analysis to enable a new age of discovery. *J Proteomics*. 2014;107:71–82.
70. Bodenmiller B. Multiplexed epitope-based tissue imaging for discovery and healthcare applications. *Cell Syst*. 2016;2:225–238.
71. Sivakamasundari V, Bolisetty M, Sivajothi S. Comprehensive cell type specific transcriptomics of the human kidney. *bioRxiv*. 2017.
72. Dasari S, Alexander MP, Vrana JA, et al. DnaJ heat shock protein family B member 9 is a novel biomarker for fibrillary GN. *J Am Soc Nephrol*. 2018;29:51–56.
73. Andeen NK, Yang H-Y, Dai D-F, et al. DnaJ homolog subfamily B member 9 is a putative autoantigen in fibrillary GN. *J Am Soc Nephrol*. 2018;29:231–239.
74. Sethi S, Theis JD. Pathology and diagnosis of renal non-AL amyloidosis. *J Nephrol*. 2018;31:343–350.
75. Beck LH Jr, Bonegio RG, Lambeau G, et al. M-type phospholipase A2 receptor as target antigen in idiopathic membranous nephropathy. *N Engl J Med*. 2009;361:11–21.
76. Qi W, Keenan HA, Li Q, et al. Pyruvate kinase M2 activation may protect against the progression of diabetic glomerular pathology and mitochondrial dysfunction. *Nat Med*. 2017;23:753–762.
77. Rinschen MM, Gödel M, Grahammer F, et al. A multi-layered quantitative in vivo expression atlas of the podocyte unravels kidney disease candidate genes. *Cell Rep*. 2018;23:2495–2508.
78. Rinschen MM, Wu X, König T, et al. Phosphoproteomic analysis reveals regulatory mechanisms at the kidney filtration barrier. *J Am Soc Nephrol*. 2014;25:1509–1522.
79. Williams KR, Colangelo CM, Hou L, et al. Use of a targeted urine proteome assay (TUPA) to identify protein biomarkers of delayed recovery after kidney transplant. *Proteomics Clin Appl*. 2017;11.
80. Cantley LG, Colangelo CM, Stone KL, et al. Development of a targeted urine proteome assay for kidney diseases. *Proteomics Clin Appl*. 2016;10:58–74.
81. Salih M, Demmers JA, Bezstarosti K, et al. Proteomics of urinary vesicles links plakins and complement to polycystic kidney disease. *J Am Soc Nephrol*. 2016;27:3079–3092.
82. Heinzel A, Kammer M, Mayer G, et al. Validation of plasma biomarker candidates for the prediction of eGFR decline in patients with type 2 diabetes. *Diabetes Care*. 2018;41:1947–1954.
83. Hayek SS, Sever S, Ko Y-A, et al. Soluble urokinase receptor and chronic kidney disease. *N Engl J Med*. 2015;373:1916–1925.

84. Yang T, Richards EM, Pepine CJ, et al. The gut microbiota and the brain-gut-kidney axis in hypertension and chronic kidney disease. *Nat Rev Nephrol*. 2018;14:442–456.
85. Goldfarb DS. The exposome for kidney stones. *Urolithiasis*. 2016;44:3–7.
86. Hocher B, Adamski J. Metabolomics for clinical use and research in chronic kidney disease. *Nat Rev Nephrol*. 2017;13:269–284.
87. Grams ME, Shafi T, Rhee EP. Metabolomics research in chronic kidney disease. *J Am Soc Nephrol*. 2018;29:1588–1590.
88. Guijas C, Montenegro-Burke JR, Warth B, et al. Metabolomics activity screening for identifying metabolites that modulate phenotype. *Nat Biotechnol*. 2018;36:316–320.
89. Toma I, Kang JJ, Sipos A, et al. Succinate receptor GPR91 provides a direct link between high glucose levels and renin release in murine and rabbit kidney. *J Clin Invest*. 2008;118:2526–2534.
90. Grimm PR, Lazo-Fernandez Y, Delpire E, et al. Integrated compensatory network is activated in the absence of NCC phosphorylation. *J Clin Invest*. 2015;125:2136–2150.
91. Schena FP, Nistor I, Curci C. Transcriptomics in kidney biopsy is an untapped resource for precision therapy in nephrology: a systematic review. *Nephrol Dial Transplant*. 2017;7:1094–1102.
92. Tajti F, Antoranz A, Ibrahim MM, et al. A functional landscape of chronic kidney disease entities from public transcriptomic data. *bioRxiv*. 2018: 265447.
93. Kolachalama VB, Singh P, Lin CQ, et al. Association of pathological fibrosis with renal survival using deep neural networks. *Kidney Int Rep*. 2018;3:464–475.
94. Martini S, Eichinger F, Nair V, Kretzler M. Defining human diabetic nephropathy on the molecular level: integration of transcriptomic profiles with biological knowledge. *Rev Endocr Metab Disord*. 2008;9: 267–274.
95. Meng C, Zeleznik OA, Thallinger GG, et al. Dimension reduction techniques for the integrative analysis of multi-omics data. *Brief Bioinform*. 2016;17:628–641.
96. Mitra K, Carvunis A-R, Ramesh SK, et al. Integrative approaches for finding modular structure in biological networks. *Nat Rev Genet*. 2013;14:719–732.
97. Khatri P, Sirota M, Butte AJ. Ten years of pathway analysis: current approaches and outstanding challenges. *PLoS Comput Biol*. 2012;8: e1002375.
98. Schubert M, Klinger B, Klünemann M, et al. Perturbation-response genes reveal signaling footprints in cancer gene expression. *Nat Commun*. 2018;9:20.
99. Tan AY, Zhang T, Michaeel A, et al. Somatic mutations in renal cyst epithelium in autosomal dominant polycystic kidney disease. *J Am Soc Nephrol*. 2018;29:2139–2156.
100. Hasin Y, Seldin M, Lusis A. Multi-omics approaches to disease. *Genome Biol*. 2017;18:83.
101. Huang S, Chaudhary K, Garmire LX. More is better: recent progress in multi-omics data integration methods. *Front Genet*. 2017;8:84.
102. Argelaguet R, Velten B, Arnol D, et al. Multi-omics factor analysis disentangles heterogeneity in blood cancer. *bioRxiv*. 2017:217554.
103. Bzdok D, Altman N, Krzywinski M. Points of significance: statistics versus machine learning. *Nat Methods*. 2018;15:233–234.
104. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521:436–444.
105. Angermueller C, Pärnamaa T, Parts L, et al. Deep learning for computational biology. *Mol Syst Biol*. 2016;12:878.
106. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017;542:115–118.
107. Shickel B, Tighe PJ, Bihorac A, Rashidi P. Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis. *IEEE J Biomed Health Inform*. 2018;22:1589–1604.
108. Agarwal R, Sinha AD. Big data in nephrology—a time to rethink. *Nephrol Dial Transplant*. 2018;33:1–3.
109. Papadopoulos T, Krochmal M, Cisek K, et al. Omics databases on kidney disease: where they can be found and how to benefit from them. *Clin Kidney J*. 2016;9:343–352.
110. Rodriguez LL, Brooks LD, Greenberg JH, et al. Research ethics. The complexities of genomic identifiability. *Science*. 2013;339:275–276.
111. Erlich Y, Narayanan A. Routes for breaching and protecting genetic privacy. *Nat Rev Genet*. 2014;15:409–421.
112. Global Alliance for Genomics and Health. GENOMICS. A federated ecosystem for sharing genomic, clinical data. *Science*. 2016;352:1278–1280.
113. Cho H, Wu DJ, Berger B. Secure genome-wide association analysis using multiparty computation. *Nat Biotechnol*. 2018;36:547–551.
114. Guinney J, Saez-Rodriguez J. Alternative models for sharing confidential biomedical data. *Nat Biotechnol*. 2018;36:391–392.
115. International Cancer Genome Consortium, Hudson TJ, Anderson W, et al. International network of cancer genome projects. *Nature*. 2010;464:993–998.
116. Ding L, Bailey MH, Porta-Pardo E, et al. Perspective on oncogenic processes at the end of the beginning of cancer genomics. *Cell*. 2018;173:305–320.e10.
117. Sudlow C, Gallacher J, Allen N, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med*. 2015;12:e1001779.
118. Sankar PL, Parker LS. The Precision Medicine Initiative's All of Us Research Program: an agenda for research on its ethical, legal, and social issues. *Genet Med*. 2017;19:743–750.
119. Lindenmeyer MT, Kretzler M. Renal biopsy-driven molecular target identification in glomerular disease. *Pflugers Arch*. 2017;469:1021–1028.
120. Berthier CC, Zhang H, Schin M, et al. Enhanced expression of Janus kinase-signal transducer and activator of transcription pathway members in human diabetic nephropathy. *Diabetes*. 2009;58: 469–477.
121. Hodgin JB, Nair V, Zhang H, et al. Identification of cross-species shared transcriptional networks of diabetic nephropathy in human and mouse glomeruli. *Diabetes*. 2013;62:299–308.
122. Saez-Rodriguez J, Costello JC, Friend SH, et al. Crowdsourcing biomedical research: leveraging communities as innovation engines. *Nat Rev Genet*. 2016;17:470–486.
123. Khare R, Good BM, Leaman R, et al. Crowdsourcing in biomedicine: challenges and opportunities. *Brief Bioinform*. 2016;17:23–32.
124. Gonzalez-Vicente A, Hopfer U, Garvin JL. Developing tools for analysis of renal genomic data: an invitation to participate. *J Am Soc Nephrol*. 2017;28:3438–3440.
125. Khosroshahi HT, Abedi B, Daneshvar S, et al. Future of the renal biopsy: time to change the conventional modality using nanotechnology. *Int J Biomed Imaging*. 2017;2017:6141734.
126. Xie L, Bennett KM, Liu C, et al. MRI tools for assessment of microstructure and nephron function of the kidney. *Am J Physiol Renal Physiol*. 2016;311:F1109–F1124.
127. Steinhubl SR, Muse ED, Topol EJ. The emerging field of mobile health. *Sci Transl Med*. 2015;7:283rv3.
128. Vairo D, Bruzzese L, Marlinge M, et al. Towards addressing the body electrolyte environment via sweat analysis:pilocarpine iontophoresis supports assessment of plasma potassium concentration. *Sci Rep*. 2017;7:11801.
129. Gao W, Emaminejad S, Nyein HYY, et al. Fully integrated wearable sensor arrays for multiplexed in situ perspiration analysis. *Nature*. 2016;529:509–514.
130. Izmailova ES, Wagner JA, Perakslis ED. Wearable devices in clinical trials: hype and hypothesis. *Clin Pharmacol Ther*. 2018;104:42–52.
131. Steinhubl SR, Muse ED, Topol EJ. The emerging field of mobile health. *Sci Transl Med*. 2015;7:283rv3.
132. Navaneethan SD, Jolly SE, Sharp J, et al. Electronic health records: a new tool to combat chronic kidney disease? *Clin Nephrol*. 2013;79:175–183.
133. Drawz PE, Archdeacon P, McDonald CJ, et al. CKD as a model for improving chronic disease care through electronic health records. *Clin J Am Soc Nephrol*. 2015;10:1488–1499.
134. Steinhubl SR, McGovern P, Dylan J, et al. The digitised clinical trial. *Lancet*. 2017;390:2135.
135. Nadkarni GN, Coca SG, Wyatt CM. Big data in nephrology: promises and pitfalls. *Kidney Int*. 2016;90:240–241.
136. Glicksberg BS, Johnson KW, Dudley JT. The next generation of precision medicine: observational studies, electronic health records, biobanks and continuous monitoring. *Hum Mol Genet*. 2018;27:R56–R62.
137. Pope A. *An Essay on Criticism*. London: W. Lewis; 1711.