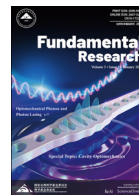
Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Fundamental Research

journal homepage: <http://www.keaipublishing.com/en/journals/fundamental-research/>

Review

Integration of artificial intelligence and multi-omics in kidney diseases

Xu-Jie Zhou^{a,b,c,d,*}, Xu-Hui Zhong^e, Li-Xin Duan^f^a Renal Division, Peking University First Hospital, Beijing 100034, China^b Kidney Genetics Center, Peking University Institute of Nephrology, Beijing 100034, China^c Key Laboratory of Renal Disease, Ministry of Health of China, Beijing 100034, China^d Key Laboratory of Chronic Kidney Disease Prevention and Treatment (Peking University), Ministry of Education, Beijing 100034, China^e Department of Pediatrics, Peking University First Hospital, Beijing, China^f The Big Data Research Center, University of Electronic Science and Technology of China, No.2006, Xiyuan Ave, West Hi-Tech Zone, Chengdu 611731, China

ARTICLE INFO

Article history:

Received 14 September 2021

Received in revised form 14 December 2021

Accepted 24 January 2022

Available online 16 March 2022

Keywords:

Artificial intelligence

Kidney diseases

Multi-omics

Network

Pathway

Precision medicine

ABSTRACT

Kidney disease is a leading cause of death worldwide. Currently, the diagnosis of kidney diseases and the grading of their severity are mainly based on clinical features, which do not reveal the underlying molecular pathways. More recent surge of ~omics studies has greatly catalyzed disease research. The advent of artificial intelligence (AI) has opened the avenue for the efficient integration and interpretation of big datasets for discovering clinically actionable knowledge. This review discusses how AI and multi-omics can be applied and integrated, to offer opportunities to develop novel diagnostic and therapeutic means in kidney diseases. The combination of new technology and novel analysis pipelines can lead to breakthroughs in expanding our understanding of disease pathogenesis, shedding new light on biomarkers and disease classification, as well as providing possibilities of precise treatment.

1. Introduction

Kidney diseases (KDs), including acute kidney injury (AKI), acute kidney disease (AKD), and chronic kidney disease (CKD), are major health problems globally. KDs have high rates of death and morbidity, and induce significant economic losses. It has been estimated that globally, around 850 million people are suffering from KDs of different causes. By 2040, CKD will increase from the 16th to the 5th ranked leading cause of death around the world, which is one of the largest projected increases of any major cause of death [1–3]. KDs are especially challenging to diagnosis and treatment, because they have diverse pathophysiologies and are always asymptomatic. When KD is not detected and treated promptly, medical care requires specialized resources and imposes high costs for both patients and medical systems [4]. In the USA, Medicare costs for CKD and end stage kidney disease (ESKD) were over \$114 billion in 2016. In China, from the 2016 Annual Data Report of China Kidney Disease Network (CK-NET), the total medical expenditure on patients with CKD was about \$4 billion USD, accounting for 6.50% of the overall Chinese medical expenditure [5]. Worldwide, it was reported that the median cost to treat a patient with CKD was significantly higher than that to treat a patient without CKD.

Currently, the diagnosis of KDs and the grading of their severity are based on clinical features, such as measurement of markers (e.g., creatinine), proteinuria, imaging examination, such as ultrasound, and biopsy findings. However, this approach does not always reveal the underlying molecular pathways, because different diseases might have similar phenotypes, symptoms, clinical parameters, or renal biopsy findings. In addition, patients with the same diagnosis might show high heterogeneity in terms of disease progression and response to the same regimen. Thus, the current classification systems remain unsatisfactory, which hampers our ability to predict long-term prognosis and apply targeted therapies. Unsurprisingly, there is a dearth of positive findings in clinical trials regarding kidney diseases.

However, there has been a surge of ~omics studies in the past decade, generating large amounts of genetic, transcriptomic, proteomic, and metabolomic data. Such studies have greatly expanded our understanding of disease pathogenesis, shedding new light on biomarkers and disease classification, as well as providing possibilities for precise treatment (Fig. 1) [6–9]. In the age of ~omics, blood, urine, and kidney biopsy samples can be taken to generate transcriptomic, proteomic, and metabolomic datasets; and digital images can be taken to produce high-dimensional, descriptive, and quantitative data for computer analysis.

* Corresponding author.

E-mail address: zhouxujie@bjmu.edu.cn (X.-J. Zhou).<https://doi.org/10.1016/j.fmre.2022.01.037>2667-3258/© 2022 The Authors. Publishing Services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

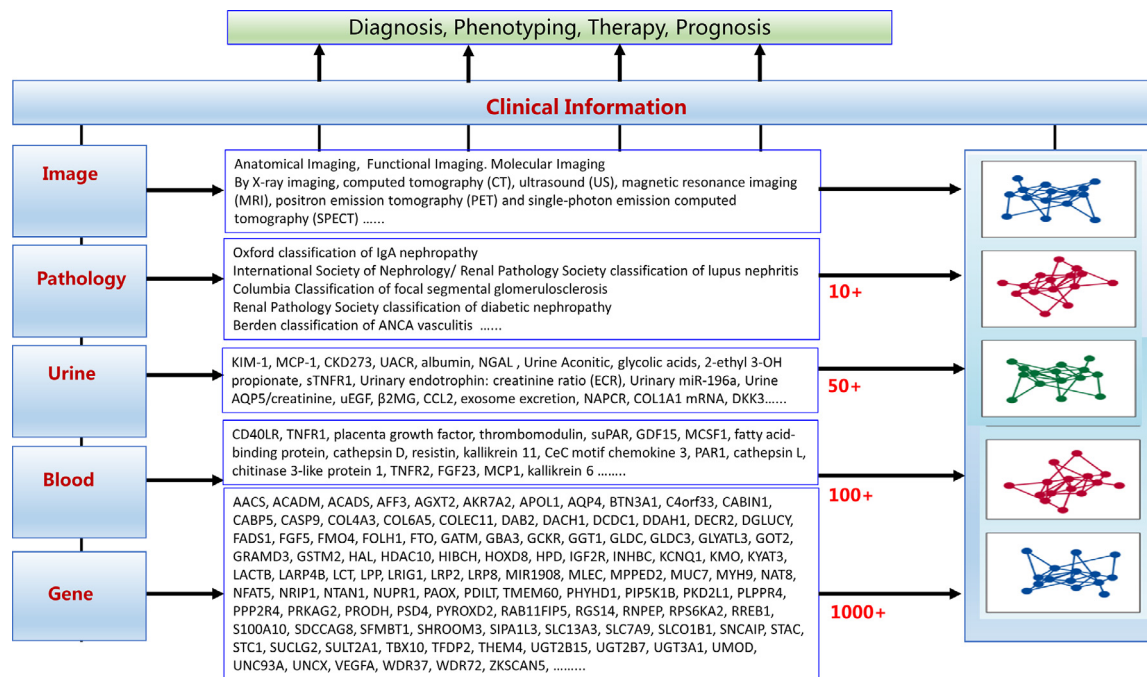


Fig. 1. Omics studies have contributed greatly to the diagnosis, phenotyping, therapy, and prognosis of kidney disease in the past ten years. Multiple biomarkers, e.g., genes, in blood and urine have been newly identified, and updated grading or classification systems in pathology and imaging have been proposed.

As biomedical research transitions into data-rich science, the era of “big data” has emerged [10]. The integration of such multi-layered datasets with longitudinal assessments of patient outcomes has the capacity to reveal different aspects of disease pathogenesis, progression, and cell-specific responses, which will be useful to guiding the design of targeted therapies.

Currently, it is imperative that kidney -omics compendia (genome sequencing, transcriptomics, proteomics, metabolomics, and microbiomics) are integrated with clinical datasets, such as electronic health records (EHRs), digital pathology repositories, and medical images (Fig. 2). Multi-source big data are thus suggested to be the major driver of precision medicine. However, without proper analysis methods, data alone cannot be transformed into clinically actionable knowledge. Thanks to the advances in computing science, artificial intelligence (AI) has been developed for robust data analysis [11–17].

This review discusses how AI and multi-omics techniques can be applied and integrated to develop novel diagnostic and therapeutic strategies for KDs. In general, the use of AI in nephrology is still in a research stage. Several other recent reviews with related foci might also be of interest [10–12,14–16].

2. Omics study

As introduced, recent -omics approaches, including genomics, transcriptomics, epigenomics, microbiomics, metabolomics, and proteomics, offer the potential to refine current disease classification paradigms and identify novel sub-phenotypes. The addition of “omics” (<http://omics.org/>) to a molecular item indicates an unbiased systemic assessment of a set of molecules. Omics studies systematically capture and assess a molecular dimension at the genome, epigenome, transcriptome, proteome, metabolome, or microbiome level. For example, transcriptomics measure all transcripts at the same time in a sample (databases for kidney such as Renal gene Expression Database, <http://rged.wall-eva.net>; Nephroseq, <http://nephroseq.org>; Kidney interactive Transcriptomics, <http://humphreyslab.com/SingleCell>), and proteomics measure all proteins (i.e. Humana Kidney and Urine Proteome Project, <http://www.hkupp.org>). All of these “omics” approaches inquire into human health and disease in an unbiased manner. The

prosperity of this field is attributed to technological advances, which have made the analysis of biological molecules cost-efficient and high-throughput. Although comprehensive at each layer, each single “-ome” can be regarded as checking only one dimension at the molecular level. Some -omics studies, such as transcriptomics, metabolomics, and proteomics, might be different at the bulk or single cell level, and offer limited correlations, reflecting reactive associations rather than causative ones. Combined and/or integrated analyses of multiple “-ome”-wide profiles were designated as “multi-omics”. This strategy is used to determine causative changes or treatment targets, because incidental association is reduced and the cause and effect can be tested.

As suggested, the concept of -omics integration can be divided into two kinds: post-analysis and the combination approach [6]. The first protocol integrates -omics data by analyzing each dataset separately and then validating the results with another omics dataset independently. There are two methods to achieve this: top-down and bottom-up. The top-down data reduction method uses genomic and transcriptomic data to suggest phenotypic responses, and to find related molecular pathways. This can be validated by further metabolomics and proteomics. However, changes in genes, proteins, and metabolites might not correlate with each other directly. In the bottom-up approach, metabolomics and proteomics are used to assess the upstream events responsible for their changes. However, the low molecule coverage limits comprehensive interpretation. In the combination method, -omics data are combined together before data interoperation and visualization. The idea is to find similarities between different -omics data using mathematical methods, such as canonical correlation analysis (CCA), knowledge-based approaches, and orthogonal two-way projection on latent structures (O2PLS) [18,19].

3. Artificial intelligence

Artificial intelligence (AI) is the science of studying computations, making it possible to perceive, reason, and act. Machine learning (ML) is a branch of AI, representing a collection of computationally intensive statistical learning techniques, to learn and improve from learning. Cur-

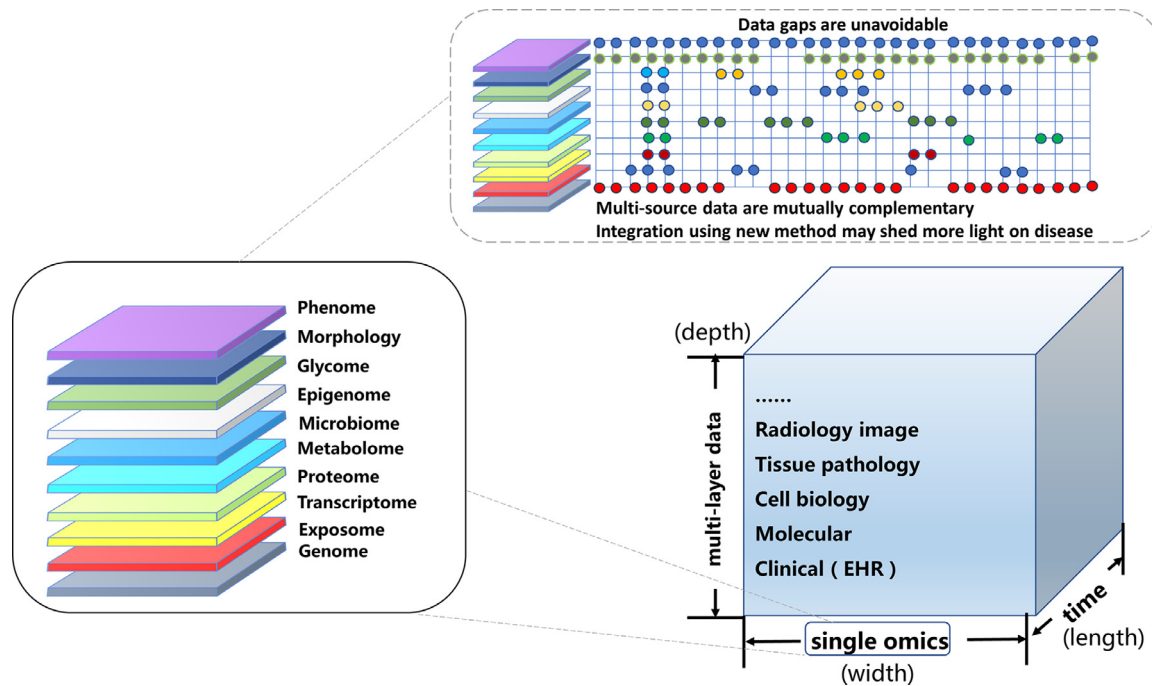


Fig. 2. Multi-layered -omics should be integrated into a whole “data ocean”. Single omics such as environmental data, epigenomics, genomics, transcriptomics, proteomics, metabolomics, microbiomics, and phenomics, represent just one dimension of data: The “width of data”. Big data are multi-layered, representing the “depth of data”. However, data may change or correlate with disease progression and time: The “length of data”. Currently, omics data frequently have inherent defects and do not fully match, thus we should incorporate robust data analysis using computing science, such as artificial intelligence.

rent AI is mostly based on ML. ML has two major branches: supervised and unsupervised. Unsupervised ML is adopted in conditions where no available outcome and ground truth annotations are present. For example, using EHR data, an unsupervised method could predict patients subgroups with an increased risk of developing outcomes such as CKD and ESKD. In contrast, currently, we more rely on supervised ML. In this scenario, a computer learns to predict the label only after being trained on a large number of training samples with ‘ground truth’ labels. For example, we can recognize specific histological features in a biopsy using supervised ML, thanks to the rapid expansion of whole slide imaging by digital slide scanners (generating whole-slide images, WSIs). The most important breakthrough in the last decade in the field of ML, may be ‘deep learning’ (DL) to train multilayered (‘deep’) neural networks (DNNs) (Fig. 3). It is anticipated that convolutional neural networks (CNNs), a specific subtype of DNNs, will create new opportunities to analyze entire histopathological slides at high resolutions. In a typical development, available WSIs are divided into training, validation, and test sets. The CNN is trained with the training set, while the process is monitored by its performance on the validation set. The test set is used to produce unbiased performance data after an optimal CNN has been achieved.

4. Application of multi-omics in kidney disease

It is important to bear in mind that the “central dogma”, representing the basic flow of information in bio-systems, from DNA (genome) to RNA (transcriptome) to proteins (proteome) and metabolites (metabolome) (Fig. 4). Significant technological advances have enabled us to analyze 3 billion DNA base pairs of the human genome at reduced cost. The advent of genome-wide associations studies (GWASs) and whole exome sequencing (WES) have greatly benefitted our understanding of pathophysiology, risk prediction, and treatment of KDs. A GWAS has the ability show that a disease is associated with some single-nucleotide polymorphisms (SNPs). The identification of genetic factors associated with KDs has the potential to shed critical insights into disease mecha-

nisms. For example, several loci are associated with CKD across different ethnicities, including *UMOD*, *SHROOM3*, *MPPED2*, *BCAS3*, and *UNCX*. *UMOD* was the first locus associated with CKD, with an OR of about 1.3 per copy of the risk allele, which remained unchanged after baseline estimated glomerular filtration rate (eGFR) adjustment [20–25]. And this risk variant directly increases the expression of uromodulin in a dose-dependent manner, causing salt-sensitive hypertension and kidney damage in mice and humans. Observational studies in the general population showed that urinary uromodulin levels are positively associated with eGFR, markers of tubular transport, and kidney length and volume. It thus suggested that the levels of uromodulin can be considered as biomarkers of kidney tubule function. Low urinary uromodulin is a marker of poorer tubular health in at-risk individuals (aged and early CKD), with decreased levels of uromodulin reflecting decreased production (i.e., decreased renal functional reserve). As different KDs may show differences in prevalence and severity among ancestral groups, for population specific genes, an excess of African Americans with focal segmental glomerulosclerosis (FSGS) and pathogen-triggered HIV-associated nephropathy (HIVAN) prompted the discovery of genetic factor of *APOL1* by GWASs [26–43]. It was observed that *APOL1*-associated kidney disease followed a recessive pattern of inheritance. *APOL1* genotype also influences the clinical course in kidney transplant recipients, hypertension-attributed ESKD, severe forms of lupus-associated kidney disease and subtypes of membranous nephropathy. Despite many reports of different cellular phenotypes, such as cell death, actin cytoskeletal structure, or mitochondrial dysfunction, the fundamental differences in behavior between the *APOL1* alleles at a molecular level have remained elusive. Anyhow, environmental triggers may be important. The overwhelming interaction between HIV infection and the *APOL1* risk variants may be the most powerful evidence, with an odds ratio (OR) of 29–89. And it was estimated that, in the pre-HAART (Highly Active Anti-Retroviral Therapy) era, 50% of individuals with the high-risk genotype developed HIVAN. One potential explanation is that both the high-risk *APOL1* genotype and high *APOL1* expression (driven in some cases by virus or the innate immune response to virus) may be

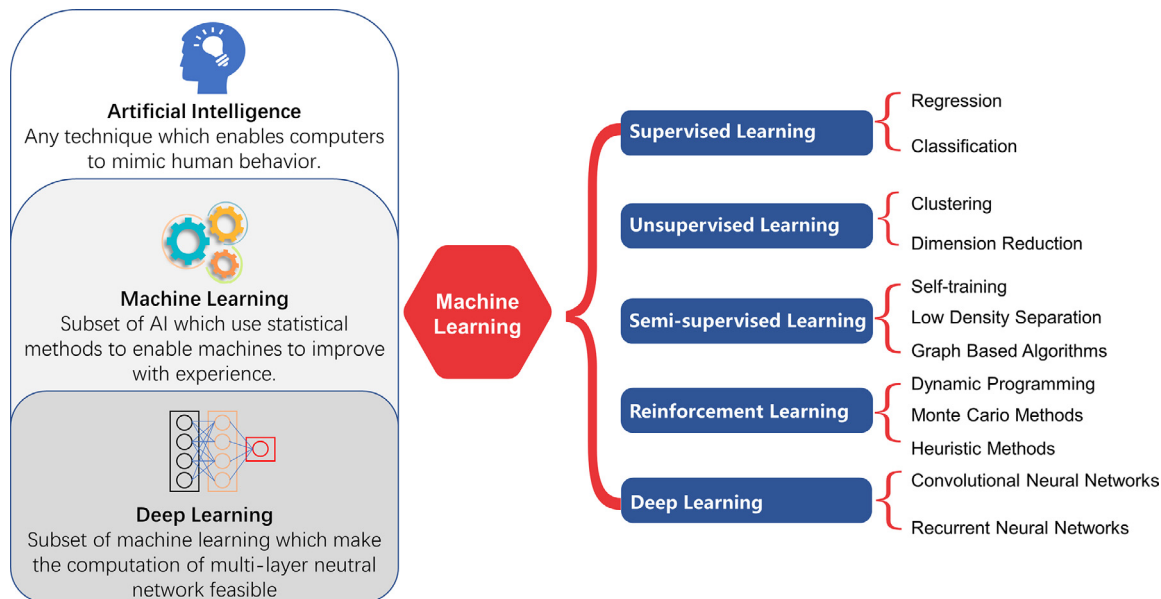


Fig. 3. Classification of artificial intelligence algorithms and relationships between artificial intelligence, machine learning, and deep learning. Artificial intelligence (AI) is a set of algorithms that enable computations making it possible to perceive, reason, and act. Machine learning (ML) is a branch of artificial intelligence in which algorithms have the ability to learn and improve from experience, without being explicitly programmed for a specific task. A popular way of classifying machine learning algorithms is by how much human supervision they require when training. Deep learning is a supervised or unsupervised machine learning algorithm based on neural networks, often specialized in image recognition, which has multiple layers of nonlinear processing units for feature extraction and transformation.

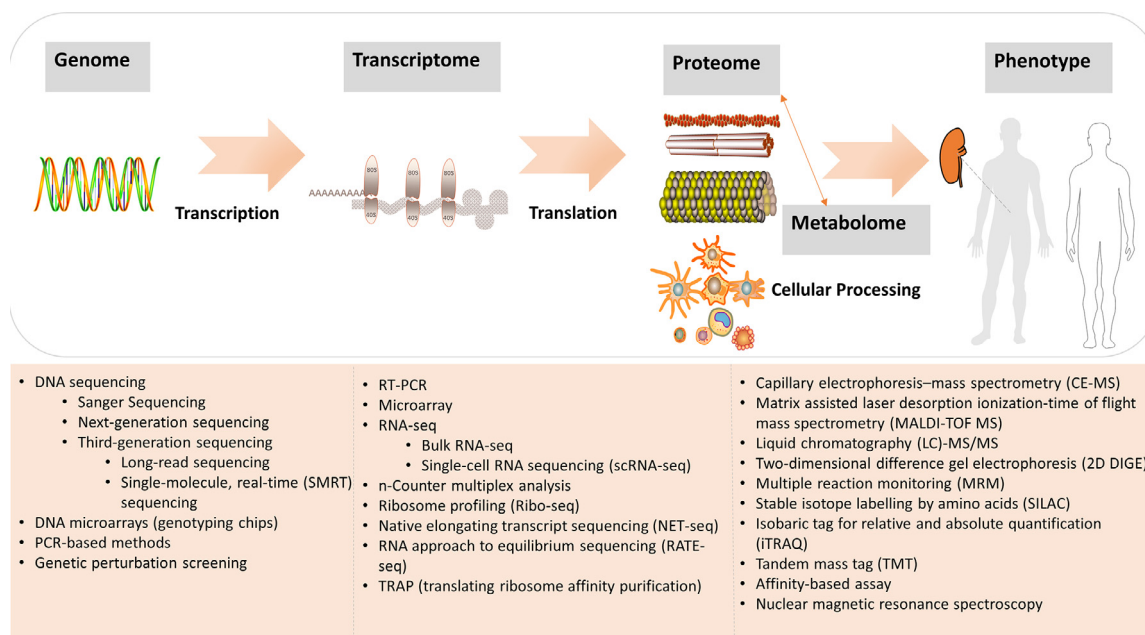


Fig. 4. Overview of the "central dogma" and related methods in -omics data generation. The upper panel depicts the central dogma and regulations of different omics layers. The lower panel presents multiple analytical platforms which can be used to interrogate samples and systems at a molecular level. In general, nucleotide-based experiments, such as genomics (genome-wide association studies), and transcriptomics studies, use PCR, microarrays and sequencing technologies. Proteomics, and metabolomics approaches share common technologies, such as mass spectrometry and nuclear magnetic resonance (NMR) spectroscopy.

required for disease to occur, as APOL1 can play as a multipurpose viral restriction factor. Several GWASs in European and East Asian populations also have identified almost 20 risk variants for IgA nephropathy (IgAN), suggesting it is shaped by selective forces from host-pathogen interactions [44,45]. However, it seems that the genetic burden of membranous nephropathy and steroid-sensitive nephrotic syndrome is determined by significantly fewer variants with much larger effects [46].

Apart from case-control GWAS, the Chronic Kidney Disease Genetics (CKDGen) Consortium has focused on identifying genetic loci associated with quantitative traits, such as the glomerular filtration rate (GFR), albuminuria, blood urea nitrogen (BUN), and serum urate levels [47]. In a more recent meta-GWAS in 42 longitudinal studies from the CKDGen Consortium and UK Biobank, Gorski *et al.* identified that seven variants located in the *UMOD*, *PRKAG2*, *WDR72*, *OR2S2*, *GATM*, and

LARP4B genes were associated with rapid kidney function decline [48]. Although GWAS has provided significant insights into pathogenic mechanisms, most loci identified still need to be followed-up to find the real causal variants and specific mechanisms. The majority of SNPs identified in GWASs are noncoding, nevertheless they implicate candidate genes or gene regions tagged by the SNPs. Further studies are needed to reveal the exact genetic changes driving the association and to prove causal inferences. However, in genomics, whether for single-gene Mendelian disorders or multi-gene complex diseases, in principle, the germ-line changes in the DNA sequence precede disease onset, indicating that the temporal relationship between genetic variation and disease onset is unidirectional.

One way to address the impact of variants on disease is by assessing the expression quantitative trait loci (eQTLs) [49,50]. eQTLs are chromosomal loci where variants explain some of the variations in RNA or in protein expressions for the candidate gene. The representative available databases for kidney include NephQTL (<https://nephqtl.org>) and the Human Kidney eQTL Atlas (<https://susztaklab.com/eqtl>). By this strategy, lysosomal beta A mannosidase (*MANBA*), disabled-2 (*DAB2*), and Dachshund homolog 1 (*DACH1*) have been suggested as potential targets in CKD [51,52]. In one early study integrating eQTL in renal transcriptomics, Martini *S et al.* selected 18 SNPs associated with eGFR in a meta-analysis of genome-wide association data by the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) and CKDGen consortia as candidates, and found that renal transcript levels for 18 genes in proximity to the SNPs significantly correlated with GFR (in *-cis*). They further constructed a co-expression network of 97 pathways linked by those genes [53]. Of these pathways, 56 pathways have been previously reported to be associated with CKD and 41 novel pathways. 78 pathways and 95% of the connections among those pathways were verified in an independent North American biopsy cohort. Of importance, the study revealed a signaling dichotomy in CKD, between activated inflammatory and immune processes and suppressed metabolic processes [54]. Several of the target factors and networks identified by this process, including the transcription factor *NRF2*, and the *JAK-STAT* and retinoid signaling pathways, have been clinically investigated or are currently underway in clinical trials. And most pathways are shared among different glomerular diseases, highlighting common pathways that mediate kidney injury. In a more recent report, integrating large-scale GWAS on kidney function with mRNA expression identified the kidney and liver as the primary organs for some traits of kidney function [55]. The proximal tubule is the critical cell type for eGFR, urate, and monogenic electrolyte or metabolic disease-related genes. Podocytes show enrichment of genes implicated in glomerular disease [55]. Notably, rare variants of some of these genes also cause Mendelian kidney diseases, providing important clues for further mechanistic studies. In the post-GWAS era, to take genetic data into clinical application, it is suggested that polygenic risk scores (PRSs) can bridge this gap by aggregating results from risk allele numbers. For example, Gorski *et al.* computed a PRS and found that individuals with a high-risk PRS (8–14 risk alleles) showed a 1.2 to 1.3-fold increased risk of AKI compared with those with a low-risk score (0–5 risk alleles) [48].

The transcriptome comprises all RNA molecules in cells. Structural similarities between DNA and RNA mean that analysis methods for genomics and transcriptomics are similar. Transcriptome analysis checks for biologically dynamic alterations, which are unstable. Thus, appropriate sampling strategies are crucial. An analysis of longitudinal blood transcriptomes showed that type 1 diabetes (T1D) was characterized by early changes in gene expression before T1D and islet autoimmunity [56]. By a genome-wide transcriptome study of tubules from biopsy samples [57], it was suggested that inflammation and metabolism as the top dysregulated pathways. Those involved genes encode key enzymes of fatty acid oxidation, including *CPT1A*, *CPT2*, *ACOX1*, *ACOX2* and their transcriptional regulators, *PPARA* and *PPARGC1A*. Further assays from mouse models with fibrosis showed a reduction in transcript and protein levels of fatty acid oxidation enzymes associated with ele-

vated lipids and triglyceride levels in tubular epithelial cells. The phenomenon can be observed prior to the occurrence of fibrosis, indicating likely pathologically causal and novel interventional targets [58]. Urinary epidermal growth factor (EGF) and monocyte chemoattractant protein-1 (MCP-1) were independently associated with disease progression as noninvasive indicators, which was supported by tissue transcriptomics and proteomics approach [59,60]. Tubulointerstitial injury was correlated with increased TGF- β and TNF- α mRNA and decreased EGF mRNA. It was reported that for a 10U decrease in the glomerular filtration rate, there was an estimated increase of 5% and 10% in TGF- β and TNF- α mRNA, respectively, whereas EGF mRNA decreased by an estimated 15% [61]. And angiotensin-converting enzyme inhibitor enalapril was effective in decreasing albumin and increasing EGF excretion [62]. Thus measurement of urinary EGF may provide a new valuable index of renal function. A “bulk” transcriptome measurement cannot remove or solve the problem of cell heterogeneity, and might be influenced by cell composition or time changes. Such data describe only averaged gene expression across the great heterogeneity of kidney cell types. Single-cell RNA sequencing (scRNA-seq) was developed to reveal transcriptome differences in heterogeneous samples, defining the transcriptome at cell resolution [63]. Using single-cell RNA sequencing, mesenchymal cells were identified as the major contributor and NKD inhibitor of WNT signaling pathway 2 (NKD2) as a myofibroblast-specific target in renal fibrosis [64]. Cadherin 11 (CDH11), SPARC related modular calcium binding 2 (SMOC2), and pigment epithelium-derived factor (PEDF) were identified as promising non-invasive biomarkers of kidney fibrosis [65]. In humans, scRNA-seq data have been used to find novel segment-specific proinflammatory responses in kidney allograft rejection [66,67], lupus nephritis [68], diabetic nephropathy [69], and IgA nephropathy [10,70,71].

Proteins are the major transcriptional products. They conduct most of the functional work in cells, playing roles as structural proteins, transcription factors, receptors, antibodies, hormones, transporters, and enzymes. The relative abundance of proteins, distinct proteins from alternative splice variants, posttranslational modifications, specific protein-protein interactions (PPIs), or expression in specific cellular components, all contribute to the complexity of the proteome. Immunoglobulins and cytokines can be detected or quantified using assays such as enzyme-linked immunosorbent assay (ELISA), immunofluorescent staining, enzyme multiplied immunoassay technique (EMIT), or mass spectrometry (MS). The identification of the M-type phospholipase A(2) receptor (PLA2R) in membranous nephropathy illustrated the potential of proteomic investigation to inform pathobiology, promoting disease diagnosis, and guiding treatment [72–74]. Importantly, by adding genomic findings in an unbiased manner, *PLA2R1* gene variants associated with membranous nephropathy in diverse populations were discovered [75–79], supporting a causal effect. In a trans-ethnic meta-analysis of 993 plasma proteins among 2,882 participants, it was observed that testican-2 might be a clinically relevant physiological marker in disease progression [80]. Urinary glycine might be a protective biomarker for IgAN [81] and higher levels of C-glycosyltryptophan might predict adverse kidney outcomes and overall mortality [82]. The gut microbiota are also important determinants of the serum metabolome, i.e. uremic toxins and secondary bile acids (SBAs) [83]. Wang *et al.* suggested that ESKD-enriched *Eggerthella lenta* and *Fusobacterium nucleatum* increased uremic toxins production and promoted the development of CKD, and a probiotic *Bifidobacterium animalis* reduced the levels of toxins and the disease severity in rats [84]. To date, however, there is still a lack of data from large cohorts with serial measurements of the proteome and metabolome over time, together with longitudinal outcomes. Furthermore, the spatial heterogeneity of the proteome and metabolome only provided limited information about intra-organ biology.

To reduce the risk of false positives in -omics studies, multiple correction testing is required. Moreover, AI and pathway analyses might be helpful in organizing and decoding -omics datasets. In these datasets, the amount of variables might be collinear, which can be

accounted for by random forest regression. The least absolute shrinkage and selection operator or elastic net are suitable methods to create a parsimonious risk model. Support-vector machines or clustering can be used to categorize variables in supervised or unsupervised ML. These approaches can be adopted to select the “top hits” with an appropriate number to meet the research goal. Pathway analysis includes several methods, aiming to organize many datasets into functional biological links. Many programs can be used for pathway analysis, such as Gene Set Enrichment Analysis (GSEA); Database for Annotation, Visualization, and Integrated Discovery (DAVID); Search Tool for the Retrieval of Interacting Genes/Proteins (STRING); and Ingenuity Pathway Analysis. Biological networks aim to analyze different biological substances such as metabolites, genes, and proteins as an interacting system. There are several types of biological networks including protein–protein interaction, signaling networks, gene regulatory networks (GRN), neuronal networks, and metabolomics networks. Many bio-informatic tools have been developed to create, visualize, and analyze these networks. Recently WGCNA [85], webCEMiTool (<https://cemitoool.sysbio.tools/>) [86] and DiffCorr [87] have been applied to compare networks across health and disease stages. Other user friendly tools are also available, such as NetworkAnalyst (<https://www.networkanalyst.ca/>) [88], RegNetwork (<http://www.regnetworkweb.org/home.jsp>) [89], OmicsNet (<http://www.omicsnet.ca>) [90], Mibiomics (<https://shiny-bird.univ-nantes.fr/app/Mibiomics>) [91], Paintomics (<http://www.paintomics.org/>) [92], or Metaboanalyst (<https://www.metaboanalyst.ca/>) [93]. One of most popular visualization and integration tool is Cytoscape (<https://cytoscape.org/>) [94].

These integration protocols can be adopted in Mendelian randomization (MR) studies to test whether candidate markers lie in causal pathways. For example, in a “multi-omics” study using integrated analysis of the genome, transcriptome, and DNA methylome from 430 human kidneys, Eales *et al* observed associations between 1,038 genes and 479 GWAS association signals related to blood pressure [95]. By utilizing kidney SNPs as genetic instruments in MR analyses, causal effects of blood pressure on clinical kidney outcomes (urinary albumin-to-creatinine ratio and CKD) were identified. It provided evidence indicating the kidney as the tissue mediator of the genetic effects on blood pressure.

What’s more, innovative integration approaches have also been adopted to identify optimal CKD model systems with which to test potential therapeutic targets. For example, a cross-species comparison of glomerular transcriptional networks was previously checked between human and murine diabetic kidney disease (DKD). By a network-matching algorithm, it was observed that *STAT3* was one of the central nodes [96], suggesting the involvement of Janus kinase (JAK)/signal transducer and activator of transcription (STAT) signaling pathway in DKD. It was further reinforced by functional assay, i.e., podocyte-specific JAK2 overexpression worsened DKD in mice [97], and treatment with a specific JAK1/2 inhibitor for 2 weeks could partially reverse the major phenotypic changes of DKD. This effect was not DKD specific, which can be also observed in focal segmental glomerular sclerosis (FSGS) and IgA nephropathy [98,99].

5. Exemplars of novel approaches to combine AI with ~omics in kidney disease

Computers have been used for different tasks for several years, especially for repetitive and cumbersome work, as well as to enhance reproducibility. Technological AI-based approaches could provide a good solution for precision medicine by delivering perfectly reproducible results. AI is also suited to precise and exhaustive extraction of the multi-dimensional data, including -omics (mRNA, microRNA, protein, and others). In clinical medicine, AI has shown promising roles across a wide spectrum of applications, including diagnostics, therapeutics, public health management, administration, and regulation. The vast quantity and accessibility of EHRs or wearable devices will influence decision-

making for physicians, patients, healthcare providers, and healthcare systems (Fig. 5) [100–116].

AI has been used in nephrology, for example, to offer solutions for repetitive work requiring significant attention [16], to improve estimation of the glomerular filtration rate [117], to facilitate the early diagnosis of AKI [118] or CKD [119], to personalize anemia management in ESKD [120], and to refine drug dosing [121].

Thus, in the field of nephrology, AI may be leveraged to analyze big data for: (I) clinicians: to make more accurate diagnoses and interpret pipelines; (II) for health systems: to improve workflow and reduce medical errors; (III) for patients: to promote health education and help disease prevention. In this review, we discuss some of the spotlights in the field related to aiding clinical decision-making. Good overviews of AI in other applications can be found in previous expert perspectives [6,7,10,15,16,122–124].

5.1. Digital pathology

The significance of a kidney biopsy is to make a diagnosis, determine the pathology classification and disease stage, assist treatment selection, and predict prognosis. Using light microscopy, immunofluorescence microscopy, and electron microscopy, as well as various staining techniques, recognition of distinctive patterns are reported and interpreted by a pathologist. Meanwhile, classification and grading systems are always needed to assess the severity of pathological changes. Although pathologists are skilled in qualitative pattern recognition and quantification, the pathological scores remain semi-quantitative, which might contribute to a low diagnostic consensus rate among different pathologists. In addition, the scoring task in research with large samples can be tedious. Thus, it is necessary to develop tools to facilitate fast, objective, and consistent assessment of renal pathology, which is a prerequisite for fruitful clinical trials, better prediction, or response evaluation. The development of whole-slide scanners to generate WSIs has made the digitization of histopathology fast and high-throughput. The resulting progress in digital pathology has facilitated the implementation of AI in computer-aided diagnostics [125].

In general, the use of AI in kidney pathology is still in its infancy and mostly at the research stage [124]. Most reported studies have focused on segmentation tasks [126,127]. This can efficiently segment kidney histopathology in periodic acid–Schiff (PAS) stained sections, and it was reported that some outputs correlate well with specific scoring systems [128,129]. Several automatic segmentation [129] and classification [130] methods have been developed using AI, mostly using a supervised deep-learning approach. High accuracy of the computational method can be obtained compared with the results produced by expert nephropathologists. To provide a continuous risk model instead of a discrete category, AI has shown supportive value for clinical diagnostics. Although PAS staining is always readily available in different kidney pathology laboratories, the full workup of kidney pathology includes further staining, such as methenamine silver and periodic acid–methenamine silver staining (Jones), trichrome staining, Masson trichrome staining, and hematoxylin-eosin (HE) staining. For a definitive diagnosis, in addition to bright field microscopy, immunofluorescence (IF) and electron microscopy (EM) are essential. For both clinical use and research, AI might increase efficiency by replacing rigid tasks. Developed AI algorithms would yield objective and reproducible data from WSIs [131]. Applied in a clinical setting, this might increase accuracy and reproducibility. In addition, it might have significance in pathology education, or in areas where nephrology pathologists are unavailable. This might aid tailored treatment for individual patients if accurate quantitative descriptors are available. In research, the potential of AI to analyze large numbers of WSIs in a standardized manner could help to develop novel biomarkers. To harness these opportunities, the establishment of large, well-curated multi-center data sets is required, as well as the training and involvement of clinicians. However, currently, large cohorts are not available in kidney pathology, and all the reported

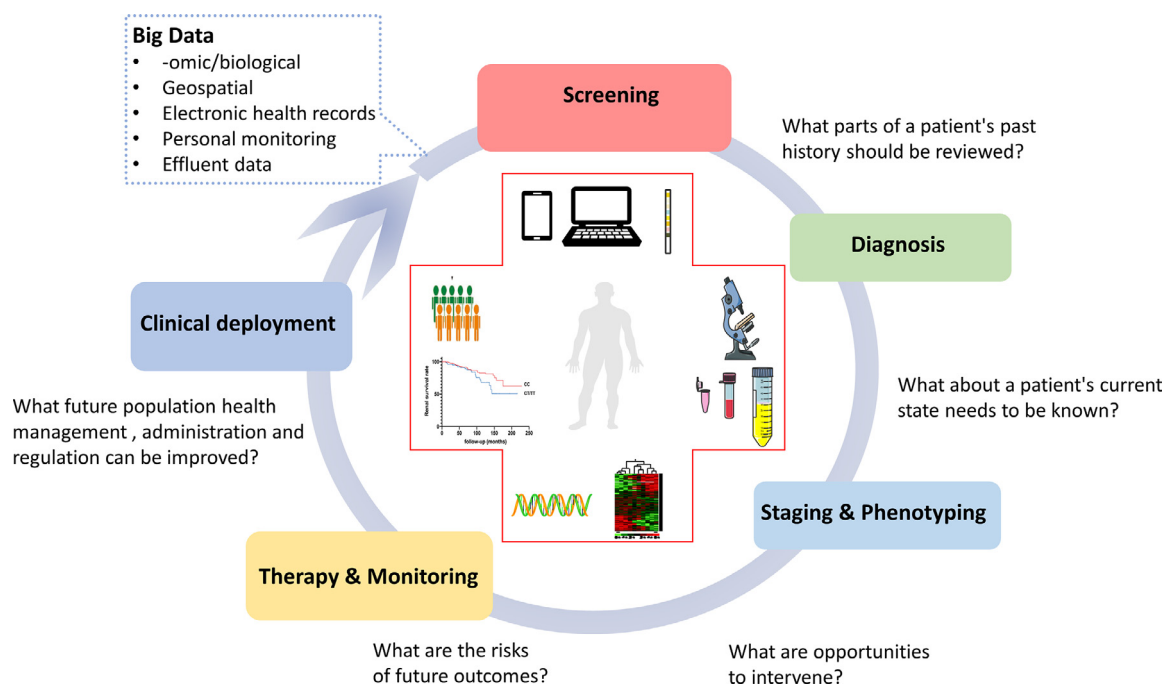


Fig. 5. Use of artificial intelligence can help to address clinically relevant questions. In the medicine, AI is poised to play major roles across a spectrum of application domains, including diagnostics, therapeutics, population health management, administration, and regulation. The vast quantity and accessibility of EHRs and ~omics will influence future decision-making on multiple fronts, including for patients, physicians, healthcare systems, healthcare providers, and regulatory bodies.

studies just analyzed tens of WSIs. Even in these small cohorts, annotation takes immense effort by pathologists. Although the renal pathology field has emphasized the precise definition of pathological findings and universally accepted definitions, gold standards for most annotations are lacking. The involvement of deep learning in pathology has emerged very recently, and compared with established diagnostic algorithms, the lack of multi-center validation and efficacy studies is not surprising. As introduced above, because molecular diagnoses are derived from objective data from measurements, they are likely to be more accurate than histology. It was suggested that a combination of molecular algorithms could improve the estimate stability [132] either by human experts or using an automated system.

5.2. Prediction and sub-phenotyping

Different approaches have been developed for sub-phenotyping, which usually take clinical characteristics as the first step. In past years, several clinical risk scores have been developed to identify patients with high risk [133]. Access to large datasets in EHRs (Table 1), and multi-omics data have contributed to recent ML approaches to identify sub-phenotypes in patients at risk of AKI [134], CKD, and ESKD.

For AKI, Tomasev *et al.* applied DL techniques to a large US Veterans Affairs (VA) longitudinal dataset of 703,782 adult patients across 172 inpatient and 1,062 outpatient sites to develop a predictive model for AKI [123,135,136]. With a lead time of up to 48 h, it was reported that the model could predict 55.8% of all inpatient episodes and 90.2% of all dialysis cases within 90 days of the initial onset. However, for every true alert, their algorithm generated two false positives. Churpek *et al.* described a predictive model for AKI developed from 495,971 adult patients. The results showed an area under the curve (AUC) as high as 0.92 for AKI and 0.97 for dialysis. However, considering the low prevalence of outcome events (3.4% stage 2 AKI), the positive-predictive value was low [137]. In an assessment of the MIMIC 3 dataset, Xu *et al.* identified three phenotypes in patients with a high risk of AKI

and associated with the subsequent severity stage of AKI [138]. Sandokji *et al.* reviewed the EHRs of 6328 hospitalized children retrospectively, and identified a time-updated prediction model with ten readily available variables to predict AKI [139]. The retrospective nature of the study limited its ability to make clear conclusions on the causality between predictor variables and the outcome [140]. Future AI algorithms developed to support real-time decisions in AKI, should be based on complete medical information, including EHRs, medical history, physical signs, laboratory data, and medications. AKI risk scores based on such an algorithm would stratify patients and inform clinical decisions to enable personalized therapy [141]. Utilizing an AI approach through perioperative data-driven learning to predict cardiac surgery-associated acute kidney injury, suggested that the top three most influential features were the intraoperative urine output, the pre-operative hemoglobin level, and transfused blood units during surgery [142]. This study represented a commendable attempt to optimize postoperative strategies to minimize complications following cardiac surgery.

CKD screening is a major challenge for both community and primary care settings, even in high-income countries. Sabanayagam *et al.* developed a DL algorithm to detect CKD using retinal images, with an AUC of 0.81–0.94 [143]. Kuo *et al.* predicted kidney function based on ultrasound images with an overall CKD classification accuracy of 85.6%, offering a possibility of noninvasive assessment of kidney function [144]. It was reported that a deep-learning model based on electrocardiogram (ECG) data might improve the detection of life-threatening hyperkalemia in CKD, enabling noninvasive screening [145]. Shang *et al.* designed an electronic CKD phenotype [146] using a combination of rule-based and ML methods to identify patients on the staging grid of albuminuria by the glomerular filtration rate (“A-by-G” grid). Applying this phenotype to 1.3 million patients indicated that over 80% of CKD are undetected. This is might significant to improve phenotype definitions in studies of CKD.

Accurate ESKD prediction is vitally important in improving clinical outcomes. To predict rapid kidney function decline, using EHR data

Table 1
Examples of current application of machine learning in CKD (taking IgA nephropathy as example) and AKI.

Year	Disease	Type of data	Sample size	Purpose of AI	Output Data (Label)	Results	Algorithms	Notes	PMID
1998	IgAN	Phenomics (6 input variables)	54 patients with IgAN	Method investigation	Outcome was assigned as 'stable' if serum creatinine was < 150 $\mu\text{mol/l}$ after 7 years and 'non-stable' if serum creatinine was $\geq 150 \mu\text{mol/l}$.	The ANN assigned the correct outcome to 47/54 (87.0%) patients: sensitivity 19/22(86.4%), specificity 28/32(87.5%). The mean score for nephrologists was 37.5/54 (69.4%, range 35–40), mean sensitivity 72% and mean specificity 66%.	artificial neural network (ANN)	One of the main purposes of the study was to demonstrate the potential application of artificial neural networks to clinical nephrology. Studies of ANN to predict 10-, 15- and 20-year outcome would be of value and would require data from more than one centre to provide sufficient training data. The first Clinical decision support system for end-stage kidney disease risk estimation in IgAN patients	9481717
2016	IgAN	Phenomics (6 input variables)	1040 patients with IgAN	Risk prediction	predict ESKD status and the time to ESKD (defined as three categories: ≤ 3 years, between > 3 and 8 years and over 8 years)	The ANNs demonstrated high performance for both the prediction of ESKD (with an AUC of 89.9, 93.3 and 100% in the Italian, Norwegian and Japanese IgAN population, respectively) and its timing (f-measure of 90.7% in the cohort from Italy and 70.8% in the one from Norway).	artificial neural network (ANN)	In addition to the predictors in the CDSS, Oxford-MEST scores, C3 staining and eGFR conveyed additional information for ESRD prediction in Chinese IgAN patients using a RF model.	26047632
2018	IgAN	Phenomics (35 input variables)	262 patients with IgAN	Risk prediction	predict the ESKD status	This RF model with above CDSS (http://www.igan.net/) 6 predictors (gender, age, hypertension, 24-h urine protein levels and histological grading) achieved an F-measure of 0.8 and the an AUC of 92.57%.	random forest model	a simplified scoring scale model (SSM) included 3 variables: urine protein excretion, global sclerosis, and tubular atrophy/interstitial fibrosis was derived	30537719
2019	IgAN	Phenomics (36 input variables)	2,047 patients with IgAN	Risk prediction	combined event of end-stage kidney disease (ESKD) or 50% reduction in estimated glomerular filtration rate within 5 years after diagnostic kidney biopsy	a C statistic of 0.89 for the derivation cohort and 0.84 for the validation cohort while using the 10 most important variables	eXtreme Gradient Boosting (XGBoost)		31031086

(continued on next page)

Table 1 (continued)

Year	Disease	Type of data	Sample size	Purpose of AI	Output Data (Label)	Results	Algorithms	Notes	PMID
2019	IgAN	Phenomics (14 input variables)	1,370 patients with IgAN	Risk prediction	predict the status of ESRD	Logistic regression [area under the curve (AUC) =96.1%] outperformed the other models [random forest (AUC =95.5%), SVM (AUC =95.8%), decision tree (AUC =84.3%), ANN (AUC =92.2%) and KNN (AUC =94.6%)].	random forest model	The formula was employed: $P = eX/(1+eX)$ where $X = -1.354 - (1.887 * M) - (2.970 * T) + (0.22 * 24 \text{ h urine protein}) + (1.792 * \text{treatment})$, e is the base of the natural logarithm. For M and T score, 0 and 1 were defined as absent and present. For treatment, 1 and 2 were defined as no glucocorticoid use and glucocorticoid use. A P value of 0.645 was selected as a cut-off point and P value >0.645 should be considered positive ESRD status and P < 0.645 should be considered negative.	31317004
2021	IgAN	Phenomics (7 input variables)	948 patients with IgAN	Risk prediction	a two-step procedure of a classifier model that predicts ESKD, and a regression model that predicts development of ESKD over time	The classifier model showed a performance value of 0.82 (area under the receiver operating characteristic curve) in patients with a follow-up of five years, which improved to 0.89 at the ten year follow-up. The regression model showed a mean absolute error of 1.78 years and a root mean square error of 2.15 years.	artificial neural network	A clinical decision support system (CDSS) (https://igan.poliba.it) may predict ESKD in patients with IgAN with a median follow-up of 5 and 10 years.	32889014
2021	IgAN	Phenomics (7 input variables)	80 patients with IgAN	Risk prediction	analyze regression techniques, estimating the deterioration of renal function, as expressed by the difference in creatinine concentration, compared to the baseline	The performance of the tested models was obtained: Random Forest Classifier showed an accuracy of 0.8–1.0, Multi-Layer Perceptron an Area Under Curve of 0.8842–0.9035 and an accuracy of 0.7527–1.0) and regressors with a low estimation error (Decision Tree Regressor showed MAE 0.2059, RMSE 0.2645).	Gaussian Naive Bayes Classifier; Support Vector Machine; Random Forest Classifier; K-nearest Neighbor Classifier; and Multi-Layer Perceptron (an artificial neural network, ANN)	The choice of input data and the choice of the right model have a direct impact on the performance. The limitation of our study is that it uses a retrospective design and a small number of samples.	33920611

(continued on next page)

Table 1 (continued)

Year	Disease	Type of data	Sample size	Purpose of AI	Output Data (Label)	Results	Algorithms	Notes	PMID
2015	IgAN	Phenomics (8 input variables)	1,174 patients with IgAN	Risk prediction	identify significant predictors of ESKD and time-to-ESKD	ANNs was superior performance compared to the other models. The ANN for ESKD prediction has accuracy greater than 90% as well as precision, recall, and f-measure for the class of patients not reaching ESKD, while precision, recall, and f-measure for the class of patients reaching ESKD are slightly lower.	artificial neural networks (ANNs), neuro fuzzy systems (NFSs), support vector machines (SVMs), and decision trees (DTs)	Four different data-driven models (artificial neural networks, neuro fuzzy systems, support vector machines, and decision trees) have been trained. Difference is due to the peculiarities of the dataset which includes a lower number of clinical records.	26453758
2021	IgAN	Phenomics (36 input variables)	2,047 patients with IgAN	Risk prediction	predict the prognosis of IgAN patients by taking the time-to-event information	The XGBoost-Surv achieved overwhelmed performance over the cox regression (0.655), lasso-cox (0.795), RSF (0.773) and SVM-Surv (0.781) on the external validation set.	EXtreme Gradient Boosting for survival (XSBoost-Surv)	Shapley Additive exPlanations (SHAP) was utilized to interpret the prediction result of our model. The Tubular atrophy/Interstitial fibrosis (%) and global sclerosis (%) were the strongest contributors to the model's final decision.	33936448
2020	IgAN	Phenomics (40 input variables)	4,047 patients with IgAN	Subphenotyping	different benefit from immunosuppression (IS) therapy measured by ESKD or 30%-50% reduction in estimated glomerular filtration rate	Three identified subgroups obtained a significant IS benefits. In patients with serum creatinine ≤ 1.437 mg/dl, the benefits of IS were observed in those with proteinuria > 1.525 g/24h, especially in those with proteinuria > 2.480 g/24h. In patients with serum creatinine > 1.437 mg/dl, those with high proteinuria and crescents benefitted from IS.	model-based recursive partitioning	This study shed some light on individualised therapy in IgAN.	32062356
2018	AKI	Phenomics (EHR data from the the University of Chicago)	121,158 patients	Risk prediction	develop an acute kidney injury risk prediction model using electronic health record data for longitudinal use in hospitalized patients	The AUC (95% CI) was 0.90 for predicting stage 2 acute kidney injury within 24 hours and 0.87 within 48 hours. The AUC was 0.96 for receipt of renal replacement therapy (n = 821) in the next 48 hours. Accuracy was similar across hospital settings and admitting serum	Gradient Boosting Machine algorithm	The algorithm provides a nearly 2-day lead time in advance of more traditional SCr-based AKI definitions. This window of time between evidence of increased AKI risk and clinical AKI being present is the ideal period for a clinical intervention.	29596073

(continued on next page)

Table 1 (continued)

Year	Disease	Type of data	Sample size	Purpose of AI	Output Data (Label)	Results	Algorithms	Notes	PMID
	AKI					creatinine groupings. At a probability threshold of greater than or equal to 0.022, the algorithm had a sensitivity of 84% and a specificity of 85% for stage 2 acute kidney injury and predicted the development of stage 2 a median of 41 hours prior to the development of stage 2 acute kidney injury.			
2018	AKI	Phenomics (institutional EHR data)	1211 patients who underwent living donor liver transplantation (LDLT) or deceased donor liver transplantation (DDLTL)	Risk prediction	predict AKI after liver transplantation	Gradient boosting machine showed the largest test AUROC (0.90) and highest accuracy (84%). AUROC of the multivariable logistic prediction model for the test dataset was 0.61. Decision tree and random forest techniques showed moderate performance (AUROC 0.86 and 0.85, respectively).	Scikit-learn, XGboost, and Keras.	Variants important in the gradient boosting machine were SvO2 (mixed venous oxygen saturation); Hb(hemoglobin); MEDL (model for end-stage liver disease); EBL(estimated blood loss); BMI(body-mass index); and ABP(arterial blood pressure) .	30413107
2018	AKI	Phenomics (42 input variables)	2010 patients who underwent open heart surgery and thoracic aortic surgery	Risk prediction	predict acute kidney injury after cardiac surgery	During the first postoperative week, AKI occurred in 770 patients (38.3%). The best performance regarding AUC was achieved by the gradient boosting machine to predict the AKI of all stages (0.78) or stage 2 or 3 AKI. The AUC of logistic regression analysis was 0.69. Decision tree, random forest, and support vector machine showed similar performance to logistic regression.	decision tree, random forest, extreme gradient boosting, support vector machine, neural network classifier, and deep learning	Important covariates in the final model included angiotensin receptor blocker,body-mass index, coronary artery bypass graft, calcium channel blocker, chronic kidney disease, creatinine, history of cerebrovascular accident, ejection fraction, preoperative ratio of early transmitral flow velocity to early diastolic velocity of the mitral annulus, fresh frozen plasma, hematocrit, hypertension, intraoperative mean mixed venous oxygen saturation, three vessel coronary disease, preoperative packed red blood cells.	

(continued on next page)

Table 1 (continued)

Year	Disease	Type of data	Sample size	Purpose of AI	Output Data (Label)	Results	Algorithms	Notes	PMID
2018	AKI	Phenomics (45 input variables)	970,869 patients undergoing PCIs	Risk prediction	AKI risk prediction after percutaneous coronary intervention (PCI) from the National Cardiovascular Data Registry (NCDR) CathPCI registry	A total of 69,826 (7.4%) patients developed AKI. The XGBoost model using permutation selection, had the highest AUC (0.725), best Brier score (0.0630), highest resolution (0.0043), and best reliability (0.0004×10^{-2}).	XGBoost	The improvement in performance of the proposed machine learning model over the current AKI model was a result of employing all available variables to modeling, using permutation test for variable selection and implementing XGBoost to model the relationship between variables and outcome. The best machine learning model only entailed the use of 2 more variables than the current AKI model, posing minimal additional burden on data extraction and processing.	30481186
2019	AKI	Phenomics (EHR data from the Medical Information Mart for Intensive Care III (MIMIC-III) database)	23,950 patients	Risk prediction	develop and validate a data driven multivariable clinical predictive model for early detection of AKI among a large cohort of adult critical care patients	It demonstrated that using machine learning models (multivariate logistic regression, random forest and artificial neural networks) with demographics and physiologic features can predict AKI onset as defined by the current clinical guideline with a competitive AUC (mean AUC 0.783 by our all-feature, logistic-regression model).	multivariate logistic regression, random forest and artificial neural networks	The comprehensive demographics and physiologic features can accurately predict max serum creatinine level during Day 2 and Day 3 with a root mean square error of 0.224 mg/dL.	30700291
2019	AKI	Phenomics (EHR data from the Medical Information Mart for Intensive Care III (MIMIC-III) database)	6,682 patients	Risk prediction	prediction of volume responsiveness in patients with oliguric acute kidney injury in critical care	The machine learning XGBoost model outperformed the traditional logistic regression model in differentiating between the volume-responsive (VR) and volume-unresponsive (VU) groups (AU-ROC, 0.860 vs. 0.728).	extreme gradient boosting (XGBoost), logistic regression	Using advanced machine learning techniques, it was observed that some important clinical factors associated with VR-AKI such as age, urinary creatinine concentration, maximum BUN concentration, and albumin. These results have some implications and require further consideration.	30961662

(continued on next page)

Table 1 (continued)

Year	Disease	Type of data	Sample size	Purpose of AI	Output Data (Label)	Results	Algorithms	Notes	PMID
2019	AKI	Phenomics (EHR data from the Medical Information Mart for Intensive Care III (MIMIC-III) database)	58,976 patients	Risk prediction	prediction of mortality risk of AKI patients who are stratified according to their AKI stages	Comparing the performance of four predictors, GBDT acquires better results. Comparing the Stages, Stage-III shows a higher mortality predictive performance.	Logistic Regression (LR), L2 norm regularized Logistic Regression (Ridge), Random Forest (RF), and Gradient Boosting Decision Tree (GBDT)	The technique of stratifying patients with AKI into different stages obtained a better performance compared to the mixed cohort. The technique of constructing a mortality predictor based on different cohorts has the potential to consider different AKI causes. It might be useful to incorporate more information (e.g., imaging and genomic biomarkers) to further improve the performance of dynamic clinical mortality risk prediction models.	31437966
2019	AKI	Phenomics (EHR data with 3,599 clinically relevant features)	703,782 adult patients across 172 inpatient and 1,062 outpatient sites	Risk prediction	develop a deep learning approach for the continuous risk prediction of future deterioration in patients from diverse clinical environments	It predicts 55.8% of all inpatient episodes of acute kidney injury, and 90.2% of all acute kidney injuries that required subsequent administration of dialysis, with a lead time of up to 48 h and a ratio of 2 false alerts for every true alert. This corresponds to an area under the receiver operating characteristic curve of 92.1%, and an area under the precision–recall curve of 29.7%.	Recurrent neural networks	Recurrent neural networks achieve the highest performance for both PR AUC and ROC AUC. Feed-forward models (deep MLP, shallow MLP, Logistic Regression, Random Forest, Gradient Boosted Trees) do not have the capacity to aggregate the information about a patient over time, which necessitates manual collection and engineering of patient historical features. Gradient Boosted Trees (GBTs) benefited from heavy overweighting of observations with positive-labels while equivalent oversampling for random forest and neural-networkbased models did not bring a similar improvement.	31367026

(continued on next page)

Table 1 (continued)

Year	Disease	Type of data	Sample size	Purpose of AI	Output Data (Label)	Results	Algorithms	Notes	PMID
2019	AKI	Phenomics (EHR data from the Medical Information Mart for Intensive Care III (MIMIC-III) database)	19044 patients with AKI in the ICU	Risk prediction	construct a mortality prediction model using the random forest (RF) algorithm for acute kidney injury (AKI) patients in the intensive care unit (ICU)	The observed in-hospital mortality of AKI patients is 13.6% in the final cohort. The results of model performance comparison show that the Brier score of the RF model is 0.085 and AUROC of the RF model is 0.866. The accuracy of the RF model is 0.728. The F1 score of the RF model is 0.459. The calibration plots show that the RF model slightly overestimates mortality in patients with low risk of death and underestimates mortality in patients with high risk of death.	Random forest model, Artificial neural network model, Support vector machine model, Customized SAPS II model	The top five important predictor variables that have influence on the in-hospital mortality of AKI patients in the ICU produced by the RF model are urine output, systolic blood pressure, age, serum bicarbonate level and heart rate.	30914181
2020	AKI	Phenomics (13 input variables)	212 patients received thoracoabdominal aortic aneurysm repair (TAAAR)	Risk prediction	acute renal failure (ARF)	Five-fold cross-validation showed that among the four classification models, random forest was the most precise model for predicting ARF, with an average area under the curve (AUC) of 0.89 ± 0.08 .	logistic regression, linear kernel SVM, Gaussian kernel SVM, and RF	Machine learning models can precisely predict ARF and paraplegia during early stages after surgery. This program allows cardiac surgeons to address complications earlier and may help improve the clinical outcomes of TAAAR.	31765025
2020	AKI	Phenomics (EHR data from the Medical Information Mart for Intensive Care III (MIMIC-III) database)	a total of 37,486 ICU stays including 7,657 AKI cases from 6,933 unique patients, and 29,829 controls from 24,366 unique patients	Subphenotyping	discover AKI sub-phenotypes using structured and unstructured electronic health record (EHR) data of patients before AKI diagnosis	Three distinct sub-phenotypes are obtained by using cluster method based on patients representations.	memory networks (MN), Logistic Regression (LR), Random Forest (RF), and Gradient Boosting Decision Tree (GBDT)	After Age adjustment, the following features are significantly different across the three subtypes: Glucose, Albumin, Creatinine, White Blood Count (WBC), Urine, and estimated Glomerular Filtration Rate (eGFR).	31911172
2020	AKI	Phenomics and some biomarkers (9 input variables)	101 adult patients with $\geq 20\%$ total body surface area (TBSA) burns or non-burn trauma-related injuries requiring surgery	Risk prediction	early recognition of burn- and trauma-related AKI	The AI/ML algorithm helped predict AKI 61.8 (32.5) hours faster than the Kidney Disease and Improving Global Disease Outcomes (KDIGO) criteria for burn and non-burned trauma patients.	logistic regression (LR), k-nearest neighbor (k-NN), random forest (RF), support vector machine (SVM), and our multi-layer perceptron (MLP) deep neural network (DNN)	NGAL was analytically superior to traditional AKI biomarkers such as creatinine and UOP. With ML, the AKI predictive capability of NGAL was further enhanced when combined with NT-proBNP or creatinine.	31937795

(continued on next page)

Table 1 (continued)

Year	Disease	Type of data	Sample size	Purpose of AI	Output Data (Label)	Results	Algorithms	Notes	PMID
2020	AKI	Phenomics (17 input variables)	22,542 patients in the Medical Information Mart for Intensive Care dataset	Risk prediction	prediction of acute kidney injury	Random forest was the best classifier using structural temporal pattern detection. The accuracy of the classifier with local and global trend features was significantly higher than that while using symbolic temporal pattern detection and the last recorded value (81.3% vs 70.6% vs 58.1%; $P < .001$).	Random Forest, Extreme Gradient Boosting Tree, Kernel-based Bayesian Network, Support Vector Machine (SVM), Logistic Regression, Naïve Bayes, K-Nearest Neighbor, and Artificial Neural Network (ANN)	It highlights the importance of using all information in time series data rather than using a single value.	32181753
2020	AKI	Phenomics (13 input variables)	1,173 hepatectomy patients, 77 (6.6%) of whom had AKI and 1,096 (93.4%) who did not	Risk prediction	predict the likelihood of acute kidney injury after liver cancer resection	The AUC values for the four algorithms were: Gbdt (0.772), gbm (0.725), forest (0.662) and DecisionTree (0.628).	random forest (forest), LightGBM (gbm), decision tree (tr), gradient boosting (Gbdt)	Age, cholesterol, tumor size, surgery duration and PLT influence the likelihood and development of postoperative acute kidney injury.	32140301
2020	AKI	Phenomics (the EHRs of 15 hospitals that are part of Aurora Health Care system)	36,614 patients with 44,691 total hospital stays	Risk prediction	continually predicts AKI to occur any time during the rest of the hospital stay	The continual prediction model obtained statistically significantly better AUC than the one-time prediction model (0.724 vs. 0.653; $p < 0.05$; two-tailed paired t-test).	Weka machine learning software	Top ten medication and comorbidity features whose prescription or diagnosis during a hospital stay were associated with a change in the model's prediction from non-AKI to AKI. Those include Cisplatin, Aminoglycosides, Hypercalcemia, Diuretics, Acyclovir, ACE inhibitors or NSAIDS or Diuretics, ARB or ACE inhibitors or NSAIDS or Diuretics, Respiratory failure, Rhabdomyolysis, K sparing.	32001013

(continued on next page)

Table 1 (continued)

Year	Disease	Type of data	Sample size	Purpose of AI	Output Data (Label)	Results	Algorithms	Notes	PMID
2020	AKI	Phenomics (32 input variables)	1571 adult patients who started CRRT for acute kidney injury	Risk prediction	predict mortality in patients undergoing continuous renal replacement therapy	For the ICU mortality, the random forest model showed the highest AUC (0.784), and the artificial neural network and extreme gradient boost models demonstrated the next best results (0.776). The AUC of the random forest model was higher than 0.611, 0.677, and 0.722, as achieved by APACHE II, SOFA, and MOSAIC, respectively.	κ -nearest neighbor (KNN), support vector machine (SVM), multivariate adaptive regression splines (MARS), random forest (RF), extreme gradient boost (XGB), and artificial neural network (ANN)	The models developed using machine learning algorithms better predicted ICU and in-hospital mortalities than conventional scoring systems such as APACHE II and SOFA, and MOSAIC.	32028984
2021	AKI	Phenomics (64 input variables)	200,859 admissions between 2014 and 2015 at 208 hospitals of the United States	Risk prediction	use XGBoost to construct a predictive mortality model for AKI patients in the ICU	The overall in-hospital mortality of AKI patients was 16.35%. The best performing algorithm in this study was XGBoost with the highest AUROC (0.796, $p < 0.01$), F1 (0.922, $p < 0.01$) and accuracy (0.860). The precision (0.860) and recall (0.994) of the XGBoost model rank second among the four models.	the XGBoost (eXtreme Gradient Boosting) decision tree model, LR (logistic regression), SVM (support vector machines), and RF (random forest)	It indicates that the minimum creatinine was more useful in predicting AKI mortality than any of the other laboratory measurements or vital signs	33539390
2021	AKI	Phenomics (64 input variables)	2,009 patients	Risk prediction	develop a predictive model for contrast induced nephropathy (CIN) in patients with coronary artery disease (CAD) with relatively normal renal function (NRF)	With baseline UA ≥ 450 $\mu\text{mol/L}$, CK-MB ≥ 48 U/L, and NT-proBNP ≥ 850 pg/mL as the cut-off values, patients were divided into two risk groups (low risk and high risk, corresponding to a total score of < 10 and ≥ 10 , respectively). The CIN incidence of patients in the low-risk group was 1.0%, while the incidence increased to 14.8% when the total score was ≥ 10 .	elastic net	Three independent risk factors in patients with CAD with relatively NRF for CI-AKI were: baseline UA level, CK-MB level, and NT-proBNP level of CIN.	34778413

(continued on next page)

Table 1 (continued)

Year	Disease	Type of data	Sample size	Purpose of AI	Output Data (Label)	Results	Algorithms	Notes	PMID
2021	AKI	Metabolomics	214 individuals (122 patients with AKI, 92 patients without AKI as controls)	Risk prediction	develop and evaluate the diagnostic use of metabolomics-based biomarkers in patients with Cardiac surgery-associated acute kidney injury (CSA-AKI)	Gluconic acid, fumaric acid, and pseudouridine were significantly upregulated in patients with AKI. A random forest model constructed with selected clinical parameters and metabolites exhibited excellent discriminative ability (area under curve, 0.939). In the AKI swine model, plasma levels of the 3 discriminating metabolites increased in a time-dependent manner. The predictive model remained robust when tested in a subset of patients with early-stage AKI in the validation cohort (area under curve, 0.943).	random forest (RF), support vector machine (SVM), and logistic regression (LR)	High-resolution metabolomics is sufficiently powerful for developing novel biomarkers. Plasma levels of 3 metabolites were useful for the early identification of CSA-AKI.	34719239
2021	AKI	Phenomics (NA)	4,289 hospitalized adult patients with acute kidney injury at admission	Subphenotyping	cluster patients with acute kidney injury at hospital admission into clinically distinct subtypes using an unsupervised machine learning approach and assess the mortality risk among the distinct clusters	Consensus clustering analysis identified four distinct clusters. Cluster 2 patients had lower serum bicarbonate, strong ion difference, and hemoglobin, but higher serum chloride, whereas cluster 3 patients had lower serum chloride but higher serum bicarbonate and strong ion difference. Cluster 4 patients were younger and more likely to be admitted for genitourinary disease and infectious disease but less likely to be admitted for cardiovascular disease. Cluster 4 patients also had more severe acute kidney injury, lower serum sodium, serum chloride, and serum bicarbonate, but higher serum potassium and anion gap. Cluster 2, 3, and 4 patients had significantly higher hospital and one-year mortality than cluster 1 patients ($p < 0.001$).	unsupervised consensus clustering analysis	It demonstrated using machine learning consensus clustering analysis to characterize a heterogeneous cohort of patients with acute kidney injury on hospital admission into four clinically distinct clusters with different associated mortality risks.	34698185

(continued on next page)

Table 1 (continued)

Year	Disease	Type of data	Sample size	Purpose of AI	Output Data (Label)	Results	Algorithms	Notes	PMID
2021	AKI	Metabolomics	1st group included 61 patients (non-AKI: 23, mild AKI: 24, severe AKI: 14) and 2nd group included 60 additional patients (non-AKI: 40, mild AKI: 20).	Risk prediction	identify metabolites potentially associated with the progression of AKI	Uinre glycine and ethanolamine were decreased in patients with AKI compared with non-AKI patients at 6-24 h in the two groups. The linear statistical model constructed at each time point by machine learning achieved the best performance at 24 h (median AUC 89%) for the 1st group. When cross-validated between the two groups, the AUC showed the best value of 70% at 12 h.	the least absolute shrinkage and selection operator (LASSO)	Urinary metabolomic profile analyses identified metabolites and time points that show patterns specific to patients who develop AKI.	34677386
2021	AKI	Phenomics (52input variables)	6093 patients (2442 in training and 3651 in external validation)	Risk prediction	predict acute dialysis initiation in AKI of hospitalized patients with coronavirus disease 2019 (COVID-19)	Of the five machine learning models tested, the nonimputed XGBoost had the best performance, with an AUROC of 0.98 at 1 day, 0.96 at 3 days, 0.94 at 5 days, and 0.93 at 7 days on internal validation. The AUPRC was 0.82 at 1 day, 0.80 at 3 days, 0.79 at 5 days, and 0.78 at 7 days. Using the optimal threshold derived using the Youden J statistic, nonimputed XGBoost had the highest sensitivity (ranging between 0.84 and 0.95) and specificity (0.9–0.96) across time points in the internal validation cohort.	logistic regression, Least Absolute Shrinkage and Selection Operator (LASSO), random forest, and eXtreme GradientBoosting (XGBoost0)	During all time horizons, serum creatinine at admission was one of the major features driving model predictions. Other clinically relevant features included BUN, systolic BP, age, and oxygen saturation.	34031183
2021	AKI	Phenomics (EHR data from the Medical Information Mart for Intensive Care III (MIMIC-III) database)	12,347 patient	Risk prediction	evaluate the ability of a machine learning algorithm to predict for AKI as defined by KDIGO stage 2 or 3 up to 48 hours	On a hold-out test set, the CNN algorithm attained an AUROC of 0.86 and PPV of 0.24, relative to a cohort AKI prevalence of 7.62%, for long-horizon AKI prediction at a 48-hour window before onset.	XGBoost, and convolutional neural networks (CNNs)	A CNN machine learning-based AKI prediction model outperforms XGBoost and the SOFA scoring system, revealing superior performance in predicting AKI 48 hours before onset, without reliance on serum creatinine (SCr) measurements.	34013107

An increasing amount of studies taking artificial intelligence for predicting, diagnosing and sub-phenotyping. And there are many promising models, but exhibit variable performance. The variability in ML prediction can be attributed, in part, to the specific ML model utilized, variable selection and processing, study and subject characteristics, and the steps associated with model training, validation, testing, and calibration. Apart from traditional models, there are some studies investigating ~omics in AKI prediction. IgAN: IgA nephropathy; AKI: Acute kidney injury; EHRs: Electronic health records; PCI: Percutaneous coronary intervention; GBM: Gradient boosting machine; DT: Decision tree; RF: Random forest; SVM: Support vector machine; LR: Logistic regression; RNN: Recurrent neural network; ANN: Artificial neural network; ROC: Receiver operating characteristic curve; AUC: Area under curve.

from 118,584 patients, Inaguma *et al.* observed that the amount of urine protein, especially an increasing tendency, could serve as a prominent risk factor [147]. Using an unsupervised latent class mixed model [148], Raynaud *et al.* identified that donor age, eGFR, proteinuria, and histological features including graft scarring, interstitial inflammation and tubulitis, microcirculation inflammation, and circulating anti-HLA donor specific antibodies, could predict long-term kidney allograft outcomes. Using a retrospective cohort of 948 patients, Schena *et al.* developed an ANN ESKD prediction model in IgA nephropathy [149]. Chen *et al.* constructed a scoring scale model (SSM) with three variables (tubular atrophy/interstitial fibrosis, global sclerosis, and urine protein excretion). They reported that the 5-year risks for the combined event using the SSM risk scores for 0–4 points were 2.7%, 4.8%, 16.4%, 30.8%, and 72.4%, respectively. To date, several different models have been developed and tested to predict ESKD in IgAN [150–154] (Table 1); however, results were conflicting, partially because different predictors were involved and the selected predictors were limited. If the algorithm captures biases in the training data, the conclusions would not be strongly supported. In cases of erroneous discoveries, it will be important to have a better understanding of the disease and relationships among predictors, as well as robust validation and mechanistic follow-up studies. Interestingly, Gonzalo-Calvo *et al.* suggested that adding -omics data, including circulating microribonucleic acids, to machine learning modeling could improve cardiovascular risk prediction in patients with ESKD on hemodialysis [155]. In addition, the inclusion of metabolites might also improve risk prediction in people with diabetes [156]. However, sample sizes in -omics studies need to be increased for reliable readouts.

A beneficial impact of AI on life quality and survival in dialysis and kidney transplant patients has been proposed. It was suggested that AI-based algorithms were better than nephrologists to predict volumes, Kt/V, hypotension, and cardiovascular events [157]. The applications of AI in dialysis also include: service process, dialysis procedure, anemia management, diet, arteriovenous fistula assessment, peritoneal techniques, infections, and outcome prediction. Applications of AI in transplantation studies include pretransplant organ-matching, graft rejection prediction, and therapy dose modulation.

5.3. AI based system for rare kidney diseases

There are over 8000 kinds of rare diseases (RDs) worldwide. About 80% of RDs are genetic and about 75% are childhood onset. Although individually rare, it is estimated that they could affect 350 million people in total, representing a major challenge for healthcare. Patients with RDs frequently have several obstacles: multi-organ involvement without specific symptoms leading to late diagnosis and misdiagnosis, lack of targeted therapies, and an absence of monitoring systems. Thus, these patients could derive a particular benefit from AI technologies [158–160]. To integrate and analyze data from different sources (e.g., multi-omics and patient registries), the use of AI can overcome the above challenges. Opportunities will include increased access to various expertise, improved patient participation and monitoring, and better patient empowerment. AI technologies can be used to identify patients from EHR data [161], to provide real-world information about their medical history, and to determine effective drug selection and therapeutic regimens [162,163]. In genetic counseling, AI has also shown its superiority to improve the accuracy for variant calling, prediction, and classification [164–166].

5.4. Drug discovery

A variety of AI methods, such as Bayesian, support vector machines, and DNNs have demonstrated their merits in drug development. These methods leverage big datasets obtained from high-throughput screening and allow the prediction of target properties with enhanced accuracy [167–169]. AI has also been used to text-mine literature using natural

language processing (NLP), to search druggable targets using genome wide data [170–172], and to perform ‘virtual screens’ on vast numbers of compounds. In the field of structure-based drug development or targeted mutagenesis, knowledge about protein structure is imperative. Despite of decades of effort, only about 100,000 proteins revealed structures, representing 17% of human protein sequences. AlphaFold is a newly developed ML approach that incorporates physical and biological knowledge, leveraging multi-sequence alignments into the algorithm design [173–177]. By applying AlphaFold2 (<https://alphafold.ebi.ac.uk/>) to 98.5% of human proteins at scale, it disclosed 58% of the residues with a confident prediction. AlphaFold is a co-evolution-dependent method. By contrast, physics-based approaches, such as end-to-end differentiable models, semi-supervised approaches, and generative approaches, are also gradually maturing, and might provide a complementary path to tackle the prediction of protein structure [178]. These AI tools are predicted to accelerate future drug design. Unfortunately, drug development is a lengthy process.

5.5. Precision medicine

Precision medicine is generalized as the delivery of tailored interventions to specific patients, i.e., “the right drug for the right patient at the right time”. The practice largely depends on detailed molecular characteristics generated from -omics investigations. Omics data must be integrated with accurate clinical measurements and detailed pathological assessments. While there are intrinsic limitations to the use of traditional statistical methods to deal with multi-layered big data, AI is a promising technique to enable the use and integration of -omics data. For instance, in inflammatory bowel disease, ML techniques are able to stratify patients into sub-phenotypes based on the integration of immunological findings, endoscopic observation, and histological data [179,180]. Decision support systems based on AI have been used in patients receiving hemodialysis to guide drug and dialysis dose prescription [181,182]. In addition, ML models can help to cluster clinically similar patients into molecularly different phenotypes, based on ‘-omics’ data, to select more effective treatment options [183]. These perspectives can be used as a reference to stratify benefits and risks, as well as to develop individualize treatment, by introducing non-invasive and more specific diagnostic classifications.

Several collaborative projects have been launched to achieve the goal of precision medicine in KDs. The Kidney Precision Medicine Project (KPMP) (<https://www.kpmp.org/>) is an NIH-funded plan that is actively generating molecular and 3D imaging maps of reference kidneys from patients with AKI and CKD. The goal is to promote data sharing and accelerate molecular diagnosis of common KDs, finally aiming to develop more precise advanced therapy [184]. In this context, a repository of clinical and biospecimen data named the Kidney Tissue Atlas has been created. It will be helpful to use molecular data to define novel subtypes, which currently suffers from insufficient classification. Then, the anticipated targeted and effective therapies could be discovered. Other cohorts/consortia initiated include the ERCB (European Renal cDNA Bank), the CKDGen Consortium (<https://ckdgen.imbi.uni-freiburg.de>), NEPTUNE (Nephrotic Syndrome Study Network) (<https://www.neptune-study.org/>), Clinical Phenotyping Resource and Biobank Core (C-Probe) (<https://kidneycenter.med.umich.edu>), and the CRIC cohort (Chronic Renal Insufficiency Cohort) (<http://www.cristudy.org>).

6. Current limitations and future directions

As discussed in this review, new opportunities are opening up for nephrologists because of the development of -omics technologies and AI methods. These new technologies will contribute to a better understanding of the molecular pathophysiology of kidney diseases, their mechanistic classification, and their targeted therapy (Fig. 6). In this

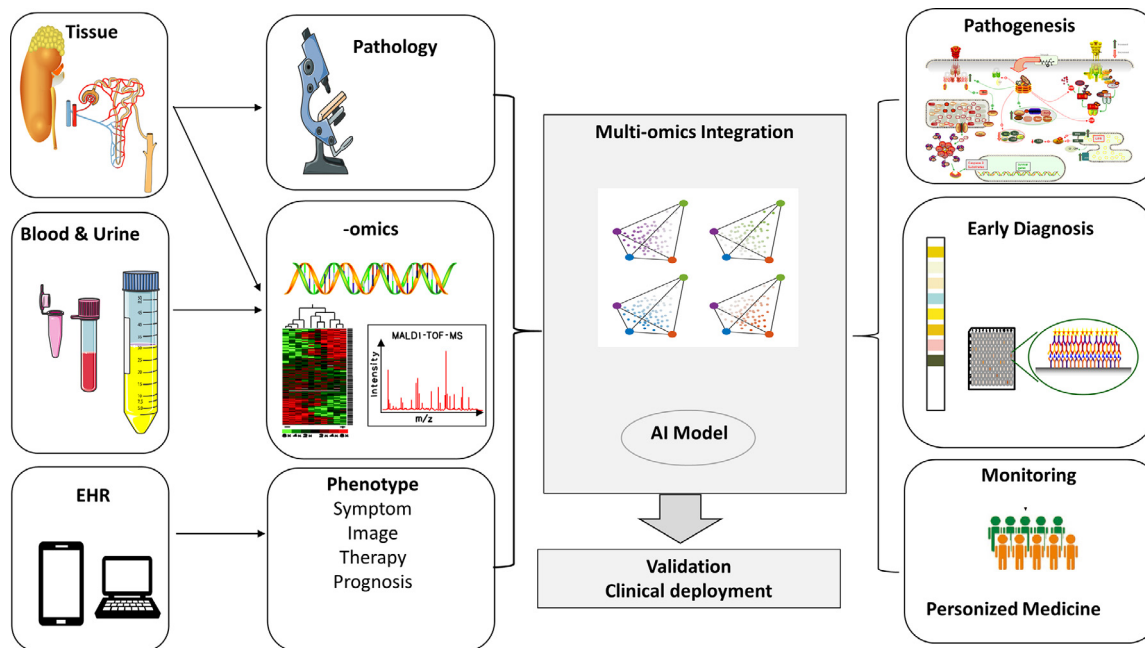


Fig. 6. Interrogation of multi-layered omics data using by artificial intelligence to address disease molecular subtypes, to diagnosis and predict diseases early, and to develop precise targeted therapy are the current challenges in kidney diseases. Information from an individual can be de-identified and coded at the individual level, and then captured across multiple data domains, including omics and/or pathology data from tissue, urine or blood samples as well as electronic health records (EHR). Such approaches have the potential to aid risk stratification, provide insights into novel mechanisms associated with kidney disease, and aid the identification of clinically relevant subpopulations or subgroups who may benefit from targeted therapy.

scenario, multi-source data integration plays an important role. However, we need to recognize that the meaningful results of computational analysis can be only derived from valid data. Big data, particularly data from real-world clinical practice, are usually biased and might be misinterpreted. Most reported studies also suffer from the biases resulting from the use of retrospective cohorts. The structures of clinical data are complex, and need to be linked with different data resources. Data storage might result in errors and missing information is almost inevitable. In addition, there is a lack of standardized reporting systems that are easy to process and identify. For -omics studies, the quality of research, the transparency and reproducibility of assay data, generalization issues resulting from results being produced on different platforms or laboratories, are all critical points. Using standardized references, assay protocols, and procedures for quality control will improve reproducibility among studies. In addition, large -omics studies have the potential to find the most noteworthy changes, but are less useful to observe uncommon, yet likely defining, events. Thus, it is necessary to complement strategies based on *a priori* knowledge and mechanistic models. Considerations of ethical issues are also required when merging data access and sharing. Data disparities among different populations, inappropriate use of datasets, inaccurate or inappropriate disclosures, limitations in deidentification techniques, and little incentive for data sharing in the current healthcare systems, may lead to privacy breaches, minority discriminatory and biased treatment. Technologies of AI will undoubtedly alter the traditional physician-patient relationship, but it may also introduce safety challenges once system performance fluctuates. In addition, for all medical devices and regimens, there is tension between providing ethical medical care and generating profit. One key element to implementing AI in healthcare is development of regulatory standards for assessment of safety and efficacy. One of the organizations leading guidance in this area is the International Medical Device Regulators Forum (IMDRF), whose members currently include Australia, Brazil, Canada, China, Europe, Japan, Russia, Singapore, South Korea, and the United States [100–116]. The application of multi-omics, and artificial intelligence in nephrology is still at an early

discovery stage. We need to create and execute a roadmap to validate those basic discoveries and further translate into clinical care, creating a new era of personalized medicine to improve health and kidney outcomes.

Author contribution statement

ZXJ: Collected data, wrote the manuscript, revised the manuscript, and gave final approval of the version to be published. ZXH: Edited the manuscript. ZXH: Edited the manuscript.

Declaration of competing interest

The authors declare that they have no conflicts of interest in this work.

Acknowledgments

Support was provided by the National Science Foundation of China (82022010, 82131430172, 81970613); the Beijing Natural Science Foundation (Z190023); Academy of Medical Sciences –Newton Advanced Fellowship (NAFR13\1033); the Fok Ying Tung Education Foundation (171030); the Beijing Nova Program Interdisciplinary Cooperation Project (Z191100001119004); and the Chinese Academy of Medical Sciences Research Unit (2019RU023). The funders had no role in the study design, data collection and analysis, the decision to publish, or the preparation of the manuscript. We apologize to those whose work or original publication could not be cited because of space limitations.

References

- [1] K.J. Foreman, N. Marquez, A. Dolgert, et al., Forecasting life expectancy, years of life lost, and all-cause and cause-specific mortality for 250 causes of death: Reference and alternative scenarios for 2016–40 for 195 countries and territories, *Lancet* 392 (10159) (2018) 2052–2090.
- [2] V.A. Luyckx, Z. Al-Aly, A.K. Bello, et al., Sustainable Development Goals relevant to kidney health: An update on progress, *Nat. Rev. Nephrol.* 17 (1) (2021) 15–32.

- [3] K. Kalantar-Zadeh, T.H. Jafar, D. Nitsch, et al., Chronic kidney disease, *Lancet* (2021).
- [4] S. Luo, M.E. Grams, Epidemiology research to foster improvement in chronic kidney disease care, *Kidney Int.* 97 (3) (2020) 477–486.
- [5] L. Zhang, M.H. Zhao, L. Zuo, et al., China Kidney Disease Network (CK-NET) 2016 Annual Data Report, *Kidney Int. Suppl.* 10 (2) (2020) e97–e185 (2011).
- [6] M.T. Lindenmeyer, F. Alakwaa, M. Rose, et al., Perspectives in systems nephrology, *Cell Tissue Res.* (2021).
- [7] S. Eddy, L.H. Mariani, M. Kretzler, Integrated multi-omics approaches to improve classification of chronic kidney disease, *Nat. Rev. Nephrol.* 16 (11) (2020) 657–668.
- [8] R.F. Dubin, E.P. Rhee, Proteomics and Metabolomics in Kidney Disease, including Insights into Etiology, Treatment, and Prevention, *Clin J Am Soc Nephrol* 15 (3) (2020) 404–411.
- [9] M.B. Stokes, Classification systems in renal pathology: Promises and problems, *Surg Pathol Clin* 7 (3) (2014) 427–441.
- [10] J. Saez-Rodriguez, M.M. Rinschen, J. Floege, et al., Big science and big data in nephrology, *Kidney Int.* 95 (6) (2019) 1326–1337.
- [11] O. Niel, P. Bastard, Artificial intelligence in nephrology: Core concepts, clinical applications, and perspectives, *Am. J. Kidney Dis.* 74 (6) (2019) 803–810.
- [12] L. Chan, A. Vaid, G.N. Nadkarni, Applications of machine learning methods in kidney disease: Hope or hype? *Curr. Opin. Nephrol. Hypertens.* 29 (3) (2020) 319–326.
- [13] C. Thongprayoon, W. Kaewput, K. Kovvuru, et al., Promises of big data and artificial intelligence in nephrology and transplantation, *J Clin Med* 9 (4) (2020).
- [14] M. Hermesen, B. Smeets, L. Hilbrands, et al., Artificial intelligence: Is there a potential role in nephropathology? *Nephrol. Dial. Transplant.* (2020).
- [15] R. Sealfon, L.H. Mariani, M. Kretzler, et al., Machine learning, the kidney, and genotype-phenotype analysis, *Kidney Int.* 97 (6) (2020) 1141–1149.
- [16] J.U. Becker, D. Mayerich, M. Padmanabhan, et al., Artificial intelligence and machine learning in nephropathology, *Kidney Int.* 98 (1) (2020) 65–75.
- [17] K.V. Lemley, Machine learning comes to nephrology, *J. Am. Soc. Nephrol.* 30 (10) (2019) 1780–1781.
- [18] H.K. Pedersen, S.K. Forslund, V. Gudmundsdottir, et al., A computational framework to integrate high-throughput ‘-omics’ datasets for the identification of potential mechanistic links, *Nat. Protoc.* 13 (12) (2018) 2781–2800.
- [19] J. Chong, J. Xia, Computational approaches for integrative analysis of the metabolome and microbiome, *Metabolites* 7 (4) (2017).
- [20] O. Devuyst, C. Pattaro, The UMOD locus: Insights into the pathogenesis and prognosis of kidney disease, *J. Am. Soc. Nephrol.* 29 (3) (2018) 713–726.
- [21] A. Köttgen, N.L. Glazer, A. Dehghan, et al., Multiple loci associated with indices of renal function and chronic kidney disease, *Nat. Genet.* 41 (6) (2009) 712–717.
- [22] M. Trudu, S. Janas, C. Lanzani, et al., Common noncoding UMOD gene variants induce salt-sensitive hypertension and kidney damage by increasing uromodulin expression, *Nat. Med.* 19 (12) (2013) 1655–1660.
- [23] O. Devuyst, E. Olinger, S. Weber, et al., Autosomal dominant tubulointerstitial kidney disease, *Nat. Rev. Dis. Primers* 5 (1) (2019) 60.
- [24] O. Devuyst, E. Olinger, L. Rampoldi, Uromodulin: From physiology to rare and complex kidney disorders, *Nat. Rev. Nephrol.* 13 (9) (2017) 525–544.
- [25] M.G. Shlipak, E.C. Day, Biomarkers for incident CKD: A new framework for interpreting the literature, *Nat. Rev. Nephrol.* 9 (8) (2013) 478–483.
- [26] D.J. Friedman, M.R. Pollak, APOL1 and kidney disease: From genetics to biology, *Annu. Rev. Physiol.* 82 (2020) 323–342.
- [27] X.J. Zhou, D.J. Klionsky, H. Zhang, Podocytes and autophagy: A potential therapeutic target in lupus nephritis, *Autophagy* 15 (5) (2019) 908–912.
- [28] B.I. Freedman, J.R. Sedor, Intensive blood-pressure control in hypertensive chronic kidney disease, *N. Engl. J. Med.* 363 (26) (2010) 2565–2566.
- [29] A. Parsa, W.H. Kao, D. Xie, et al., APOL1 risk variants, race, and progression of chronic kidney disease, *N. Engl. J. Med.* 369 (23) (2013) 2183–2196.
- [30] W.W. Williams, M.R. Pollak, Health disparities in kidney disease—emerging data from the human genome, *N. Engl. J. Med.* 369 (23) (2013) 2260–2261.
- [31] S.D. Cohen, J.B. Kopp, P.L. Kimmel, Kidney diseases associated with human immunodeficiency virus infection, *N. Engl. J. Med.* 377 (24) (2017) 2363–2374.
- [32] G.N. Nadkarni, C.R. Gignoux, E.P. Sorokin, et al., Worldwide frequencies of APOL1 renal risk variants, *N. Engl. J. Med.* 379 (26) (2018) 2571–2572.
- [33] P. Beckerman, J. Bi-Karchin, A.S. Park, et al., Transgenic expression of human APOL1 risk variants in podocytes induces kidney disease in mice, *Nat. Med.* 23 (4) (2017) 429–438.
- [34] S.S. Hayek, K.H. Koh, M.E. Grams, et al., A tripartite complex of suPAR, APOL1 risk variants and $\alpha(v)\beta(3)$ integrin on podocytes mediates chronic kidney disease, *Nat. Med.* 23 (8) (2017) 945–953.
- [35] S. Rosset, S. Tzur, D.M. Behar, et al., The population genetics of chronic kidney disease: Insights from the MYH9-APOL1 locus, *Nat. Rev. Nephrol.* 7 (6) (2011) 313–326.
- [36] P.E. Drawz, J.R. Sedor, The genetics of common kidney disease: A pathway toward clinical relevance, *Nat. Rev. Nephrol.* 7 (8) (2011) 458–468.
- [37] G. Genovese, D.J. Friedman, M.R. Pollak, APOL1 variants and kidney disease in people of recent African ancestry, *Nat. Rev. Nephrol.* 9 (4) (2013) 240–244.
- [38] P. Sidaway, Glomerular disease: Innate immunity-APOL1 interaction, *Nat. Rev. Nephrol.* 10 (10) (2014) 543.
- [39] A.Z. Rosenberg, S. Naicker, C.A. Winkler, et al., HIV-associated nephropathies: Epidemiology, pathology, mechanisms and treatment, *Nat. Rev. Nephrol.* 11 (3) (2015) 150–160.
- [40] E. Nobakht, S.D. Cohen, A.Z. Rosenberg, et al., HIV-associated immune complex kidney disease, *Nat. Rev. Nephrol.* 12 (5) (2016) 291–300.
- [41] K.D. Brown, C. Campbell, G.V. Roberts, Precision medicine in kidney disease: The patient’s view, *Nat. Rev. Nephrol.* 16 (11) (2020) 625–627.
- [42] A.A. Adeyemo, D. Shriner, A.R. Bentley, et al., Evolutionary genetics and acclimatization in nephrology, *Nat. Rev. Nephrol.* 17 (12) (2021) 827–839.
- [43] G. Genovese, D.J. Friedman, M.D. Ross, et al., Association of trypanolytic ApoL1 variants with kidney disease in African Americans, *Science* 329 (5993) (2010) 841–845.
- [44] X.Q. Yu, M. Li, H. Zhang, et al., A genome-wide association study in Han Chinese identifies multiple susceptibility loci for IgA nephropathy, *Nat. Genet.* 44 (2) (2011) 178–182.
- [45] K. Kiryluk, Y. Li, F. Scolori, et al., Discovery of new risk loci for IgA nephropathy implicates genes involved in immunity against intestinal pathogens, *Nat. Genet.* 46 (11) (2014) 1187–1196.
- [46] E. Sanchez-Rodriguez, C.T. Southard, K. Kiryluk, GWAS-Based discoveries in IgA nephropathy, membranous nephropathy, and Steroid-Sensitive nephrotic syndrome, *Clin J Am Soc Nephrol* 16 (3) (2021) 458–466.
- [47] A. Köttgen, C. Pattaro, The CKDGen Consortium: Ten years of insights into the genetic basis of kidney function, *Kidney Int.* 97 (2) (2020) 236–242.
- [48] M. Gorski, B. Jung, Y. Li, et al., Meta-analysis uncovers genome-wide significant variants for rapid kidney function decline, *Kidney Int.* 99 (4) (2021) 926–939.
- [49] C.E. Gillies, R. Putler, R. Menon, et al., An eQTL landscape of kidney tissue in human nephrotic syndrome, *Am. J. Hum. Genet.* 103 (2) (2018) 232–244.
- [50] C. Qiu, S. Huang, J. Park, et al., Renal compartment-specific genetic variation analyses identify new pathways in chronic kidney disease, *Nat. Med.* 24 (11) (2018) 1721–1731.
- [51] T. Doko, S. Huang, C. Qiu, et al., Transcriptome-wide association analysis identifies DACH1 as a kidney disease risk gene that contributes to fibrosis, *J. Clin. Invest.* 131 (10) (2021).
- [52] Y.A. Ko, H. Yi, C. Qiu, et al., Genetic-Variation-Driven Gene-Expression changes highlight genes with important functions for kidney disease, *Am. J. Hum. Genet.* 100 (6) (2017) 940–953.
- [53] S. Martini, V. Nair, B.J. Keller, et al., Integrative biology identifies shared transcriptional networks in CKD, *J. Am. Soc. Nephrol.* 25 (11) (2014) 2559–2572.
- [54] N. Ledo, Y.A. Ko, A.S. Park, et al., Functional genomic annotation of genetic risk loci highlights inflammation and epithelial biology networks in CKD, *J. Am. Soc. Nephrol.* 26 (3) (2015) 692–714.
- [55] Y. Li, S. Haug, P. Schlosser, et al., Integration of GWAS summary statistics and gene expression reveals target cell types underlying kidney function traits, *J. Am. Soc. Nephrol.* 31 (10) (2020) 2326–2340.
- [56] L.P. Xhonneux, O. Knight, Å. Lernmark, et al., Transcriptional networks in at-risk individuals identify signatures of type 1 diabetes progression, *Sci. Transl. Med.* 13 (587) (2021).
- [57] H.M. Kang, S.H. Ahn, P. Choi, et al., Defective fatty acid oxidation in renal tubular epithelial cells has a key role in kidney fibrosis development, *Nat. Med.* 21 (1) (2015) 37–46.
- [58] X. Zhu, L. Jiang, M. Long, et al., Metabolic reprogramming and renal fibrosis, *Front Med (Lausanne)* 8 (2021) 746920.
- [59] S. Menez, W. Ju, R. Menon, et al., Urinary EGF and MCP-1 and risk of CKD after cardiac surgery, *JCI Insight* 6 (11) (2021).
- [60] W. Ju, V. Nair, S. Smith, et al., Tissue transcriptome-driven identification of epidermal growth factor as a chronic kidney disease biomarker, *Sci. Transl. Med.* 7 (316) (2015) 193r–316r.
- [61] S.A. Teteris, S.A. Menahem, G. Perry, et al., Dysregulated growth factor gene expression is associated with tubulointerstitial apoptosis and renal dysfunction, *Kidney Int.* 71 (10) (2007) 1044–1053.
- [62] Z. Josefsberg, S.A. Ross, A. Lev-Ran, et al., Effects of enalapril and nitrendipine on the excretion of epidermal growth factor and albumin in hypertensive NIDDM patients, *Diabetes Care.* 18 (5) (1995) 690–693.
- [63] D. Barwinska, T.M. El-Achkar, F.R. Melo, et al., Molecular characterization of the human kidney interstitium in health and disease, *Sci. Adv.* 7 (7) (2021).
- [64] C. Kuppe, M.M. Ibrahim, J. Kranz, et al., Decoding myofibroblast origins in human kidney fibrosis, *Nature* 589 (7841) (2021) 281–286.
- [65] I.M. Schmidt, M.R. Colona, B.R. Kestenbaum, et al., Cadherin-11, Sparc-related modular calcium binding protein-2, and Pigment epithelium-derived factor are promising non-invasive biomarkers of kidney fibrosis, *Kidney Int.* (2021).
- [66] H. Wu, A.F. Malone, E.L. Donnelly, et al., Single-Cell transcriptomics of a human kidney allograft biopsy specimen defines a diverse inflammatory response, *J. Am. Soc. Nephrol.* 29 (8) (2018) 2069–2080.
- [67] Y. Liu, J. Hu, D. Liu, et al., Single-cell analysis reveals immune landscape in kidneys of patients with chronic transplant rejection, *Theranostics* 10 (19) (2020) 8851–8862.
- [68] E. Der, S. Ranabothu, H. Suryawanshi, et al., Single cell RNA sequencing to dissect the molecular heterogeneity in lupus nephritis, *JCI Insight* 2 (9) (2017).
- [69] P.C. Wilson, H. Wu, Y. Kirita, et al., The single-cell transcriptomic landscape of early human diabetic nephropathy, *Proc Natl Acad Sci U S A.* 116 (39) (2019) 19619–19625.
- [70] E.D. O’Sullivan, K.J. Mylonas, J. Hughes, et al., Complementary roles for Single-Nucleus and Single-Cell RNA sequencing in kidney disease research, *J. Am. Soc. Nephrol.* 30 (4) (2019) 712–713.
- [71] Y. Zheng, P. Lu, Y. Deng, et al., Single-Cell transcriptomics reveal immune mechanisms of the onset and progression of IgA nephropathy, *Cell Rep.* 33 (12) (2020) 108525.
- [72] P. Ronco, H. Debiec, Molecular pathogenesis of membranous nephropathy, *Annu Rev Pathol* 15 (2020) 287–313.

- [73] P. Ronco, H. Debiec, Pathophysiological advances in membranous nephropathy: Time for a shift in patient's care, *Lancet* 385 (9981) (2015) 1983–1992.
- [74] L.J. Beck, R.G. Bonegio, G. Lambeau, et al., M-type phospholipase A2 receptor as target antigen in idiopathic membranous nephropathy, *N. Engl. J. Med.* 361 (1) (2009) 11–21.
- [75] H.C. Stanescu, M. Arcos-Burgos, A. Medlar, et al., Risk HLA-DQA1 and PLA(2)R1 alleles in idiopathic membranous nephropathy, *N. Engl. J. Med.* 364 (7) (2011) 616–626.
- [76] Z. Cui, L.J. Xie, F.J. Chen, et al., MHC class II risk alleles and amino acid residues in idiopathic membranous nephropathy, *J. Am. Soc. Nephrol.* 28 (5) (2017) 1651–1664.
- [77] J. Lv, W. Hou, X. Zhou, et al., Interaction between PLA2R1 and HLA-DQA1 variants associates with anti-PLA2R antibodies and membranous nephropathy, *J. Am. Soc. Nephrol.* 24 (8) (2013) 1323–1329.
- [78] H.Y. Wang, Z. Cui, L.J. Xie, et al., HLA class II alleles differing by a single amino acid associate with clinical phenotype and outcome in patients with primary membranous nephropathy, *Kidney Int.* 94 (5) (2018) 974–982.
- [79] J. Xie, L. Liu, N. Mladkova, et al., The genetic architecture of membranous nephropathy and its potential to improve non-invasive diagnosis, *Nat. Commun.* 11 (1) (2020) 1600.
- [80] P. Matias-García, R. Wilson, Q. Guo, et al., Plasma proteomics of renal function: A trans-ethnic meta-analysis and mendelian randomization study, *J. Am. Soc. Nephrol.* (2021).
- [81] S. Park, J. Lee, S.H. Yang, et al., Comprehensive metabolomic profiling in early IgA nephropathy patients reveals urine glycine as a prognostic biomarker, *J. Cell. Mol. Med.* 25 (11) (2021) 5177–5190.
- [82] I. Steinbrenner, U.T. Schultheiss, F. Kotsis, et al., Urine metabolite levels, adverse kidney outcomes, and mortality in CKD patients: A metabolome-wide association study, *Am. J. Kidney Dis.* (2021).
- [83] I.W. Wu, S.S. Gao, H.C. Chou, et al., Integrative metagenomic and metabolomic analyses reveal severity-specific signatures of gut microbiota in chronic kidney disease, *Theranostics* 10 (12) (2020) 5398–5411.
- [84] X. Wang, S. Yang, S. Li, et al., Aberrant gut microbiota alters host metabolome and impacts renal failure in humans and rodents, *Gut* 69 (12) (2020) 2131–2142.
- [85] P. Langfelder, S. Horvath, WGCNA: An R package for weighted correlation network analysis, *BMC Bioinformatics* 9 (2008) 559.
- [86] L.E. Cardozo, P. Russo, B. Gomes-Correia, et al., WebCEMiTool: Co-expression modular analysis made easy, *Front Genet* 10 (2019) 146.
- [87] A. Fukushima, DiffCorr: An R package to analyze and visualize differential correlations in biological networks, *Gene* 518 (1) (2013) 209–214.
- [88] G. Zhou, O. Soufan, J. Ewald, et al., NetworkAnalyst 3.0: A visual analytics platform for comprehensive gene expression profiling and meta-analysis, *Nucleic. Acids. Res.* 47 (W1) (2019) W234–W241.
- [89] Z.P. Liu, C. Wu, H. Miao, et al., RegNetwork: An integrated database of transcriptional and post-transcriptional regulatory networks in human and mouse, *Database (Oxford)* 2015 (2015).
- [90] G. Zhou, J. Xia, OmicsNet: a web-based tool for creation and visual analysis of biological networks in 3D space, *Nucleic. Acids. Res.* 46 (W1) (2018) W514–W522.
- [91] J. Zoppi, J.F. Guillaume, M. Neunlist, et al., MiBioOmics: An interactive web application for multi-omics data exploration and integration, *BMC Bioinformatics* 22 (1) (2021) 6.
- [92] R. Hernández-de-Diego, S. Tarazona, C. Martínez-Mira, et al., PaintOmics 3: A web resource for the pathway analysis and visualization of multi-omics data, *Nucleic. Acids. Res.* 46 (W1) (2018) W503–W509.
- [93] J. Chong, O. Soufan, C. Li, et al., MetaboAnalyst 4.0: Towards more transparent and integrative metabolomics analysis, *Nucleic. Acids. Res.* 46 (W1) (2018) W486–W494.
- [94] D. Otasek, J.H. Morris, J. Bouças, et al., Cytoscape Automation: Empowering workflow-based network analysis, *Genome Biol.* 20 (1) (2019) 185.
- [95] J.M. Eales, X. Jiang, X. Xu, et al., Uncovering genetic mechanisms of hypertension through multi-omic analysis of the kidney, *Nat. Genet.* 53 (5) (2021) 630–637.
- [96] J.B. Hodgin, V. Nair, H. Zhang, et al., Identification of cross-species shared transcriptional networks of diabetic nephropathy in human and mouse glomeruli, *Diabetes* 62 (1) (2013) 299–308.
- [97] H. Zhang, V. Nair, J. Saha, et al., Podocyte-specific JAK2 overexpression worsens diabetic kidney disease in mice, *Kidney Int.* 92 (4) (2017) 909–921.
- [98] J. Tao, L. Mariani, S. Eddy, et al., JAK-STAT activity in peripheral blood cells and kidney tissue in IgA nephropathy, *Clin J Am Soc Nephrol* 15 (7) (2020) 973–982.
- [99] J. Tao, L. Mariani, S. Eddy, et al., JAK-STAT signaling is activated in the kidney and peripheral blood cells of patients with focal segmental glomerulosclerosis, *Kidney Int.* 94 (4) (2018) 795–808.
- [100] J. He, S.L. Baxter, J. Xu, et al., The practical implementation of artificial intelligence technologies in medicine, *Nat. Med.* 25 (1) (2019) 30–36.
- [101] Medicine in the digital age, *Nat. Med.* 25 (1) (2019) 1.
- [102] B. Norgeot, B.S. Glicksberg, A.J. Butte, A call for deep-learning healthcare, *Nat. Med.* 25 (1) (2019) 14–15.
- [103] E.J. Topol, High-performance medicine: The convergence of human and artificial intelligence, *Nat. Med.* 25 (1) (2019) 44–56.
- [104] A. Esteva, A. Robicquet, B. Ramsundar, et al., A guide to deep learning in healthcare, *Nat. Med.* 25 (1) (2019) 24–29.
- [105] K.Y. Ngiam, Braving the new world of artificial intelligence, *Nat. Med.* 25 (1) (2019) 13.
- [106] O. Gottesman, F. Johansson, M. Komorowski, et al., Guidelines for reinforcement learning in healthcare, *Nat. Med.* 25 (1) (2019) 16–18.
- [107] W.N. Price, I.G. Cohen, Privacy in the age of medical big data, *Nat. Med.* 25 (1) (2019) 37–43.
- [108] A. Mincholé, B. Rodriguez, Artificial intelligence for the electrocardiogram, *Nat. Med.* 25 (1) (2019) 22–23.
- [109] Y. Gurovich, Y. Hanani, O. Bar, et al., Identifying facial phenotypes of genetic disorders using deep learning, *Nat. Med.* 25 (1) (2019) 60–64.
- [110] B. Abbott, Deeper learning, *Nat. Med.* 25 (1) (2019) 9–11.
- [111] A.Y. Hannun, P. Rajpurkar, M. Haghpanahi, et al., Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network, *Nat. Med.* 25 (1) (2019) 65–69.
- [112] Z.I. Attia, S. Kapa, F. Lopez-Jimenez, et al., Screening for cardiac contractile dysfunction using an artificial intelligence-enabled electrocardiogram, *Nat. Med.* 25 (1) (2019) 70–74.
- [113] R. Arnaout, Toward a clearer picture of health, *Nat. Med.* 25 (1) (2019) 12.
- [114] A. Rajkumar, J. Dean, I. Kohane, Machine learning in medicine, *N. Engl. J. Med.* 380 (14) (2019) 1347–1358.
- [115] K.Y. Ngiam, I.W. Khor, Big data and machine learning algorithms for health-care delivery, *Lancet Oncol.* 20 (5) (2019) e262–e273.
- [116] N. Schwalbe, B. Wahl, Artificial intelligence and the future of global health, *Lancet* 395 (10236) (2020) 1579–1586.
- [117] X. Liu, N. Li, L. Lv, et al., Improving precision of glomerular filtration rate estimating model by ensemble learning, *J. Transl. Med.* 15 (1) (2017) 231.
- [118] B.A. Molitoris, Beyond biomarkers: Machine learning in diagnosing acute kidney injury, *Mayo Clin. Proc.* 94 (5) (2019) 748–750.
- [119] N.A. Almansour, H.F. Syed, N.R. Khayat, et al., Neural network and support vector machine for the prediction of chronic kidney disease: A comparative study, *Comput. Biol. Med.* 109 (2019) 101–111.
- [120] M.E. Brier, A.E. Gaweda, G.R. Aronoff, Personalized anemia management and precision medicine in ESA and iron pharmacology in End-Stage kidney disease, *Semin. Nephrol.* 38 (4) (2018) 410–417.
- [121] J. Tang, R. Liu, Y.L. Zhang, et al., Application of Machine-Learning models to predict tacrolimus stable dose in renal transplant recipients, *Sci. Rep.* 7 (2017) 42192.
- [122] P. Rashidi, A. Bihorac, Artificial intelligence approaches to improve kidney care, *Nat. Rev. Nephrol.* 16 (2) (2020) 71–72.
- [123] J.A. Kellum, A. Bihorac, Artificial intelligence to predict AKI: Is it a breakthrough? *Nat. Rev. Nephrol.* 15 (11) (2019) 663–664.
- [124] L. Barisoni, K.J. Lafata, S.M. Hewitt, et al., Digital pathology and computational image analysis in nephropathology, *Nat. Rev. Nephrol.* 16 (11) (2020) 669–685.
- [125] K. Bera, K.A. Schalper, D.L. Rimm, et al., Artificial intelligence in digital pathology - new tools for diagnosis and precision oncology, *Nat. Rev. Clin. Oncol.* 16 (11) (2019) 703–715.
- [126] I. Alnazer, P. Bourdon, T. Urruty, et al., Recent advances in medical image processing for the evaluation of chronic kidney disease, *Med. Image Anal.* 69 (2021) 101960.
- [127] E. Uchino, K. Suzuki, N. Sato, et al., Classification of glomerular pathological findings using deep learning and nephrologist-AI collective intelligence approach, *Int. J. Med. Inform.* 141 (2020) 104231.
- [128] R.D. Bülow, J. Kers, P. Boor, Multistain segmentation of renal histology: First steps toward artificial intelligence-augmented digital nephropathology, *Kidney Int.* 99 (1) (2021) 17–19.
- [129] C.P. Jayapandian, Y. Chen, A.R. Janowczyk, et al., Development and evaluation of deep learning-based segmentation of histologic structures in the kidney cortex with multiple histologic stains, *Kidney Int.* 99 (1) (2021) 86–101.
- [130] B. Ginley, B. Lutnick, K.Y. Jen, et al., Computational segmentation and classification of diabetic glomerulosclerosis, *J. Am. Soc. Nephrol.* 30 (10) (2019) 1953–1967.
- [131] B. Ginley, K.Y. Jen, S.S. Han, et al., Automated computational detection of interstitial fibrosis, tubular atrophy, and glomerulosclerosis, *J. Am. Soc. Nephrol.* 32 (4) (2021) 837–850.
- [132] J. Reeve, G.A. Böhmig, F. Eskandary, et al., Generating automated kidney transplant biopsy reports combining molecular measurements with ensembles of machine learning classifiers, *Am. J. Transplant.* 19 (10) (2019) 2719–2731.
- [133] S.J. Barbour, R. Coppo, H. Zhang, et al., Evaluating a new international Risk-Prediction tool in IgA nephropathy, *JAMA Intern Med* 179 (7) (2019) 942–952.
- [134] Z.H. Endre, R.L. Mehta, Identification of acute kidney injury subphenotypes, *Curr. Opin. Crit. Care.* 26 (6) (2020) 519–524.
- [135] N. Tomašev, N. Harris, S. Baur, et al., Use of deep learning to develop continuous-risk models for adverse event prediction from electronic health records, *Nat. Protoc.* 16 (6) (2021) 2765–2787.
- [136] N. Tomašev, X. Glorot, J.W. Rae, et al., A clinically applicable approach to continuous prediction of future acute kidney injury, *Nature* 572 (7767) (2019) 116–119.
- [137] M.M. Churpek, K.A. Carey, D.P. Edelson, et al., Internal and external validation of a machine learning risk score for acute kidney injury, *JAMA Netw Open* 3 (8) (2020) e2012892.
- [138] Z. Xu, J. Chou, X.S. Zhang, et al., Identifying sub-phenotypes of acute kidney injury using structured and unstructured electronic health record data with memory networks, *J. Biomed. Inform.* 102 (2020) 103361.

- [139] I. Sandokji, Y. Yamamoto, A. Biswas, et al., A Time-Updated, parsimonious model to predict AKI in hospitalized children, *J. Am. Soc. Nephrol.* 31 (6) (2020) 1348–1357.
- [140] G. De Vlioger, K. Kashani, G. Meyfroidt, Artificial intelligence to guide management of acute kidney injury in the ICU: A narrative review, *Curr. Opin. Crit. Care.* 26 (6) (2020) 563–573.
- [141] J.L. Koynier, K.A. Carey, D.P. Edelson, et al., The development of a machine learning important acute kidney injury prediction model, *Crit. Care Med.* 46 (7) (2018) 1070–1077.
- [142] P.Y. Tseng, Y.T. Chen, C.H. Wang, et al., Prediction of the development of acute kidney injury following cardiac surgery by machine learning, *Crit. Care.* 24 (1) (2020) 478.
- [143] C. Sabanayagam, D. Xu, D. Ting, et al., A deep learning algorithm to detect chronic kidney disease from retinal photographs in community-based populations, *Lancet Digit Health* 2 (6) (2020) e295–e302.
- [144] C.C. Kuo, C.M. Chang, K.T. Liu, et al., Automation of the kidney function prediction and classification through ultrasound-based kidney imaging using deep learning, *NPJ Digit Med* 2 (2019) 29.
- [145] C.D. Galloway, A.V. Valys, J.B. Shreibati, et al., Development and validation of a Deep-Learning model to screen for hyperkalemia from the electrocardiogram, *JAMA Cardiol* 4 (5) (2019) 428–436.
- [146] N. Shang, A. Khan, F. Polubriaginof, et al., Medical records-based chronic kidney disease phenotype for clinical care and "big data" observational and genetic studies, *NPJ Digit Med* 4 (1) (2021) 70.
- [147] D. Inaguma, A. Kitagawa, R. Yanagiya, et al., Increasing tendency of urine protein is a risk factor for rapid eGFR decline in patients with CKD: A machine learning-based prediction model by using a big database, *PLoS One* 15 (9) (2020) e239262.
- [148] M. Raynaud, O. Aubert, P.P. Reese, et al., Trajectories of glomerular filtration rate and progression to end stage kidney disease after kidney transplantation, *Kidney Int.* 99 (1) (2021) 186–197.
- [149] F.P. Schena, V.W. Anelli, J. Trotta, et al., Development and testing of an artificial intelligence tool for predicting end-stage kidney disease in patients with immunoglobulin a nephropathy, *Kidney Int.* 99 (5) (2021) 1179–1188.
- [150] M. Diciolla, G. Binetti, T. Di Noia, et al., Patient classification and outcome prediction in IgA nephropathy, *Comput. Biol. Med.* 66 (2015) 278–286.
- [151] F. Pesce, M. Diciolla, G. Binetti, et al., Clinical decision support system for end-stage kidney disease risk estimation in IgA nephropathy patients, *Nephrol. Dial. Transplant.* 31 (1) (2016) 80–86.
- [152] Y. Liu, Y. Zhang, D. Liu, et al., Prediction of ESRD in IgA nephropathy patients from an asian cohort: A random forest model, *Kidney Blood Press. Res.* 43 (6) (2018) 1852–1864.
- [153] R.D. Bülow, D. Dimitrov, P. Boor, et al., How will artificial intelligence and bioinformatics change our understanding of IgA Nephropathy in the next decade? *Semin. Immunopathol.* (2021).
- [154] T. Chen, X. Li, Y. Li, et al., Prediction and risk stratification of kidney outcomes in IgA nephropathy, *Am. J. Kidney Dis.* 74 (3) (2019) 300–309.
- [155] D. de Gonzalo-Calvo, P. Martínez-Cambor, C. Bär, et al., Improved cardiovascular risk prediction in patients with end-stage renal disease on hemodialysis using machine learning modeling and circulating microribonucleic acids, *Theranostics* 10 (19) (2020) 8665–8676.
- [156] J. Huang, C. Huth, M. Covic, et al., Machine learning approaches reveal metabolic signatures of incident chronic kidney disease in individuals with prediabetes and type 2 diabetes, *Diabetes* 69 (12) (2020) 2756–2765.
- [157] A. Burlacu, A. Iftene, D. Jugrin, et al., Using artificial intelligence resources in dialysis and kidney transplant patients: A literature review, *Biomed Res. Int.* 2020 (2020) 9867872.
- [158] N. Hurvitz, H. Azmanov, A. Kesler, et al., Establishing a second-generation artificial intelligence-based system for improving diagnosis, treatment, and monitoring of patients with rare diseases, *Eur. J. Hum. Genet.* (2021).
- [159] M.C. Hirsch, S. Ronicke, M. Krusche, et al., Rare diseases 2030: How augmented AI will support diagnosis and treatment of rare diseases in the future, *Ann. Rheum. Dis.* 79 (6) (2020) 740–743.
- [160] H. Amiri, I.S. Kohane, Machine learning of patient characteristics to predict admission outcomes in the undiagnosed diseases network, *JAMA Netw Open* 4 (2) (2021) e2036220.
- [161] H. Bergier, L. Duron, C. Sordet, et al., Digital health, big data and smart technologies for the care of patients with systemic autoimmune diseases: Where do we stand? *Autoimmun. Rev.* 20 (8) (2021) 102864.
- [162] S. Brasil, C. Pascoal, R. Francisco, et al., Artificial intelligence (AI) in rare diseases: Is the future brighter? *Genes (Basel)* 10 (12) (2019).
- [163] S. Das, S. Huang, A.W. Lo, Acceleration of rare disease therapeutic development: A case study of AGIL-AADC, *Drug Discov. Today.* 24 (3) (2019) 678–684.
- [164] Z. Liu, L. Zhu, R. Roberts, et al., Toward clinical implementation of Next-Generation Sequencing-Based genetic testing in rare diseases: Where are we? *Trends Genet.* 35 (11) (2019) 852–867.
- [165] Q. Li, K. Zhao, C.D. Bustamante, et al., Xrare: A machine learning method jointly modeling phenotypes and genetic evidence for rare disease diagnosis, *Genet. Med.* 21 (9) (2019) 2126–2134.
- [166] G. McInnes, A.G. Sharo, M.L. Koleske, et al., Opportunities and challenges for the computational interpretation of rare variation in clinically important genes, *Am. J. Hum. Genet.* 108 (4) (2021) 535–548.
- [167] X. Yang, Y. Wang, R. Byrne, et al., Concepts of artificial intelligence for Computer-Assisted drug discovery, *Chem. Rev.* 119 (18) (2019) 10520–10594.
- [168] P. Schneider, W.P. Walters, A.T. Plowright, et al., Rethinking drug design in the artificial intelligence era, *Nat. Rev. Drug Discov.* 19 (5) (2020) 353–364.
- [169] S. Ekins, A.C. Puhl, K.M. Zorn, et al., Exploiting machine learning for end-to-end drug discovery and development, *Nat. Mater.* 18 (5) (2019) 435–441.
- [170] À. Bravo, J. Piñero, N. Queralt-Rosinach, et al., Extraction of relations between genes and diseases from text and large-scale data analysis: Implications for translational research, *BMC Bioinformatics* 16 (2015) 55.
- [171] J. Kim, J.J. Kim, H. Lee, An analysis of disease-gene relationship from Medline abstracts by, *DigSee. Sci Rep.* 7 (2017) 40154.
- [172] J. Jeon, S. Nim, J. Teyra, et al., A systematic approach to identify novel cancer drug targets using machine learning, inhibitor design and high-throughput screening, *Genome Med* 6 (7) (2014) 57.
- [173] K. Tunyasuvunakool, J. Adler, Z. Wu, et al., Highly accurate protein structure prediction for the human proteome, *Nature* (2021).
- [174] J. Jumper, R. Evans, A. Pritzel, et al., Highly accurate protein structure prediction with AlphaFold, *Nature* (2021).
- [175] A.W. Senior, R. Evans, J. Jumper, et al., Improved protein structure prediction using potentials from deep learning, *Nature* 577 (7792) (2020) 706–710.
- [176] T. Hey, K. Butler, S. Jackson, et al., Machine learning and big scientific data, *Philos Trans A Math Phys Eng Sci* 378 (2166) (2020) 20190054.
- [177] A.W. Senior, R. Evans, J. Jumper, et al., Protein structure prediction using multiple deep neural networks in the 13th Critical Assessment of Protein Structure Prediction (CASP13), *Proteins* 87 (12) (2019) 1141–1148.
- [178] M. AlQuraishi, AlphaFold at CASP13, *Bioinformatics* 35 (22) (2019) 4862–4865.
- [179] T. Coelho, E. Mossotto, Y. Gao, et al., Immunological profiling of paediatric inflammatory bowel disease using unsupervised machine learning, *J. Pediatr. Gastroenterol. Nutr.* 70 (6) (2020) 833–840.
- [180] E. Mossotto, J.J. Ashton, T. Coelho, et al., Classification of Paediatric Inflammatory Bowel Disease using, *Machine Learning. Sci Rep.* 7 (1) (2017) 2427.
- [181] M. Hueso, L. de Haro, J. Calabia, et al., Leveraging data science for a personalized haemodialysis, *Kidney Dis (Basel)* 6 (6) (2020) 385–394.
- [182] C. Barbieri, M. Molina, P. Ponce, et al., An international observational study suggests that artificial intelligence for clinical decision support optimizes anemia management in hemodialysis patients, *Kidney Int.* 90 (2) (2016) 422–429.
- [183] A. Bierzynska, H.J. McCarthy, K. Soderquest, et al., Genomic and clinical profiling of a national nephrotic syndrome cohort advocates a precision medicine approach to disease management, *Kidney Int.* 91 (4) (2017) 937–947.
- [184] E. Ong, L.L. Wang, J. Schaub, et al., Modelling kidney disease using ontology: Insights from the Kidney Precision Medicine Project, *Nat. Rev. Nephrol.* 16 (11) (2020) 686–696.



Xu-jie Zhou obtained his PhD degree from Peking University, China. He started his career at Peking University First Hospital since 2011, and is currently the associate professor. His major research interests focus on basic and translational research in kidney diseases. Zhou has published over 90 peer-reviewed papers.