Briefings in Bioinformatics, 2024, **25(5)**, bbae364
https://doi.org/10.1093/bib/bbae364
Review

Integrated multi-omics with machine learning to uncover the intricacies of kidney disease

Xinze Liu, Jingxuan Shi, Yuanyuan Jiao, Jiaqi An, Jingwei Tian, Yue Yang, Li Zhuo*

*Corresponding author. Department of Nephrology, China-Japan Friendship Hospital, Beijing 100029, China; China-Japan Friendship Clinic Medical College, Beijing University of Chinese Medicine, 100029 Beijing, China. E-mail: zhuoli@zryhyy.com.cn

Abstract

The development of omics technologies has driven a profound expansion in the scale of biological data and the increased complexity in internal dimensions, prompting the utilization of machine learning (ML) as a powerful toolkit for extracting knowledge and understanding underlying biological patterns. Kidney disease represents one of the major growing global health threats with intricate pathogenic mechanisms and a lack of precise molecular pathology-based therapeutic modalities. Accordingly, there is a need for advanced high-throughput approaches to capture implicit molecular features and complement current experiments and statistics. This review aims to delineate strategies for integrating multi-omics data with appropriate ML methods, highlighting key clinical translational scenarios, including predicting disease progression risks to improve medical decision-making, comprehensively understanding disease molecular mechanisms, and practical applications of image recognition in renal digital pathology. Examining the benefits and challenges of current integration efforts is expected to shed light on the complexity of kidney disease and advance clinical practice.

Keywords: kidney; nephrology; multi-omics; machine learning

Introduction

Kidney disease is a major global health issue and has experienced one of the largest increases in mortality among all types of diseases over the past decade [1]. However, chronic kidney disease (CKD) remains under-recognized by both patients and healthcare providers [2]. In 2018–2019, approximately 82 million adults in China suffered from CKD with an awareness rate of merely 10% [3]. Globally, over 5 million deaths occur annually due to the unavailability of effective treatments for kidney diseases [4]. Indeed, the field of nephrology lacks targeted diagnostics and treatments tailored to the specific pathophysiological processes of individual kidney diseases [5], thus hindering the implementation of targeted therapies and precision medicine.

Omics research forms the cornerstone of precision medicine, enabling individualized therapeutic approaches [6]. The field of oncology exemplifies the progress and application of precision medicine [7], but its clinical application in nephrology falls short [8]. In current clinical practice, gathering blood, urine (a unique noninvasive method known as 'liquid biopsy' for kidney diseases), and biopsy tissues as biological samples can provide detailed molecular omics data [9], leading to a substantial increase in kidney disease studies over the last 10 years and the accumulation of extensive and intricate datasets [10]. With ongoing technological advancements, the integration of multi-omics research, emerging single-cell and spatial omics [11], radiomics [12], digital pathology, and computational image analysis [13] has become one of the primary approaches for current kidney research. The integrated analysis of different types of data has challenged traditional analytical methods, accelerating the utilization of artificial intelligence (AI) techniques and machine learning (ML) to enhance the comprehension of intrinsic and crucial information [14], often acquiring results beyond the scope of traditional statistical approaches.

This review provides an overview of the ways in which multi-omics data and ML can be integrated to improve clinical practice. We describe the technical practices with examples of clinical applicability for the precise prediction of disease onset and progression, the further understanding of kidney molecular mechanisms, and the strategies for renal digital pathology image analysis.

Integrating and elucidating multi-omics data

As a vital organ in the preservation of body fluid homeostasis, the removal of metabolic waste products, and the maintenance of blood pressure, the kidney is unique due to its extremely complex anatomy, diverse array of cell types, and intricate molecular mechanisms associated with diseases across multiple systems. This complexity makes it well-suited for integrating big data [15] in data-driven biomedical multi-omics research [16]. Specifically, the term multi-omics typically encompasses a wide spectrum of biological data, including genes (genomics), broad changes in gene expression (epigenomics), ribonucleic acid (RNA, transcriptomics) [17, 18], proteins (proteomics) [19, 20], and downstream small-molecule metabolites (metabolomics), which are generated during the processes of deoxyribonucleic acid (DNA) replication, transcription, translation, and post-translational modification. Unlike traditional experiments that measure individual biomolecules, omics technologies can comprehensively reveal all genes,

Received: March 22, 2024. **Revised:** June 20, 2024. **Accepted:** July 17, 2024

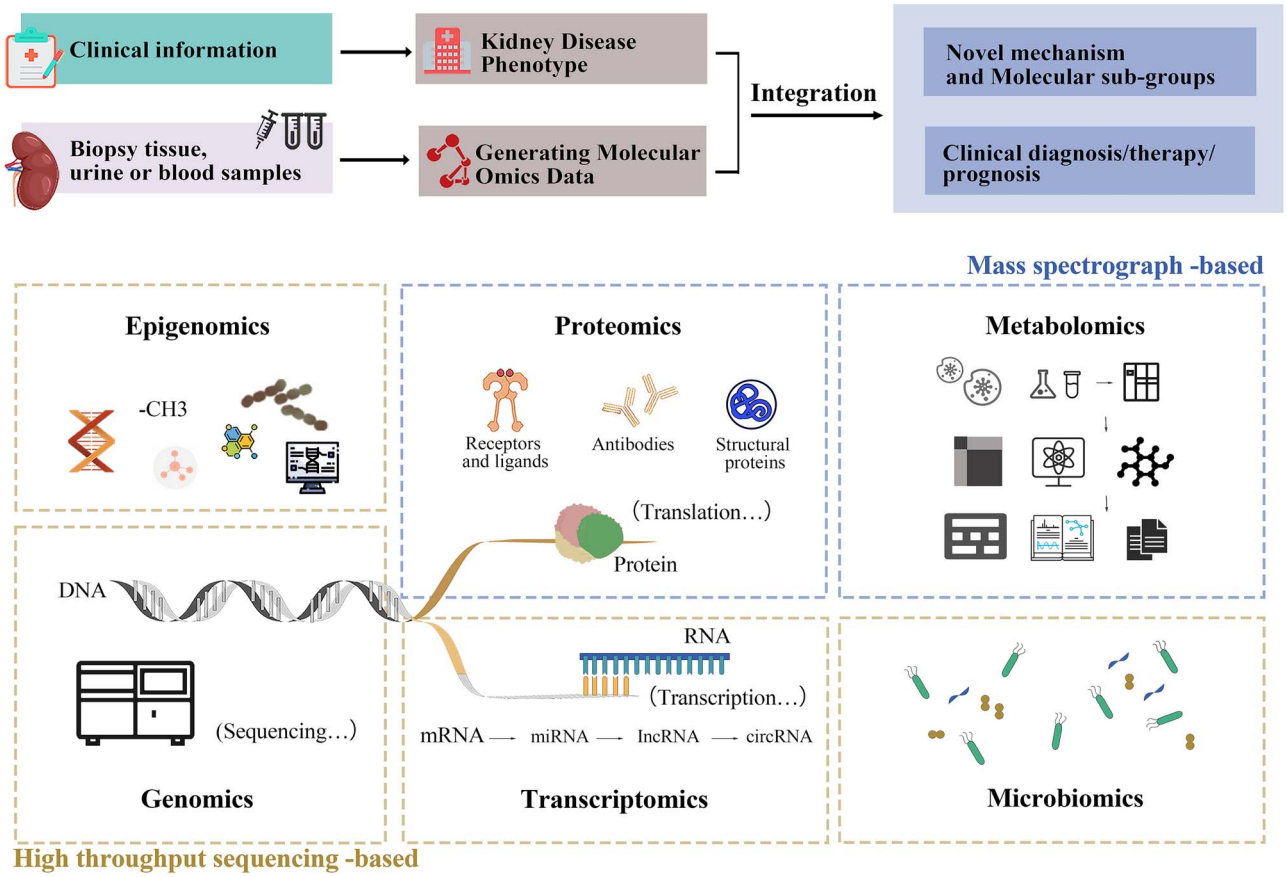

Figure 1. Overview of generating and utilizing multiple omics layers from clinical bio-samples, leading to the discovery of novel mechanisms and molecular sub-groups, which support clinical diagnosis, targeted therapy, and improved prognosis. Listed below are some common related methods, not exhaustive.

transcripts, proteins and metabolites within cells, tissues or organs from one biological origin, providing detailed molecular profiles, regulatory factors, cell types annotations, and spatial localizations spanning the entire kidney.

The integration of multi-omics combines various omics layers using advanced computational techniques, allowing for the reclassification of patient subgroups to better reveal the underlying molecular mechanisms in nephrology, thereby supporting clinical diagnosis and targeted therapy (Fig. 1). Each omics data type typically provides a list of differential factors potentially associated with the disease, such as differential expression genes (DEGs), differential expression proteins, and differential DNA methylation regions. For example, comparing transcript levels between healthy and diseased individuals allows the identification of DEGs [21] across two or more sample sets. The broad range of differential factors will be further narrowed down, followed by the validation using experimental methods or external patient cohorts [22], ultimately allowing for the identification of key genes and regulatory elements associated with kidney diseases [23–25]. For instance, to identify the key biomarkers, a recent study on membranous nephropathy (MN) and pan-cancer analysis [26] employed ML approaches to intersect a set of 318 senescence-related genes with 366 DEGs. This approach resulted in the identification of 13 senescence-related DEGs, leading to the discovery of six hub genes with further intersection and validation through immunohistochemical analysis of human renal biopsy tissues.

Powerful open data and online tools

Public data stands as a pivotal force in driving medical research forward. Various general molecular repositories alongside kidney disease-specific databases (Table 1) represent abundant sources of information concerning pathological mechanisms and molecular targets.

Table 1. General and nephrology-specific molecular data repositories.

Tool	Data types/features	Purpose	Website
General repositories			
Sequence Read Archive (SRA)	DNA sequencing data, especially 'short reads' (<1000 base pairs)	Archive raw reads from high-throughput sequencing	ncbi.nlm.nih.gov/sra
Gene Expression Omnibus (GEO)	Microarray, next-generation sequencing, and other forms of high-throughput functional genomics data	·Store high-throughput functional genomic data and gene expression profiles ·Offer easy submission procedures and formats for complete, well-annotated data ·Provide tools to query, review, and download studies and gene expression profiles	ncbi.nlm.nih.gov/geo
Encyclopedia of DNA elements (ENCODE)	Functional elements in the human genome, including protein and RNA levels, regulatory elements	Organize and search functional annotations	encodeproject.org
Online Mendelian Inheritance in Man (OMIM)	Mendelian disorders and over 16,000 genes	Discover the relationship between phenotype and genotype	omim.org
GeneCards	Gene-centric data including genomic, transcriptomic, proteomic, genetic, clinical and functional information	Provide information on all annotated and predicted human genes	genecards.org
The Cancer Genome Atlas (TCGA)	20 000+ primary cancer and matched normal samples, 33 cancer types, 2.5 petabytes of data	Improve cancer diagnosis, treatment, prevention	cancer.gov/tcga
ArrayExpress	Functional genomics data (both processed and raw data), metadata, sample annotations, protocols	Store data from high-throughput genomics experiments	ebi.ac.uk/arrayexpress
Expression Atlas	Gene and protein expression data	Provide RNA/protein abundance across species and conditions	ebi.ac.uk/gxa/home
Human Protein Atlas (HPA)	Protein expression data, high-resolution immunohistochemistry images	Map all human proteins in cells, tissues, and organs	proteinatlas.org
Human Metabolome Database (HMDB)	114 100 metabolite entries, water-soluble and lipid-soluble metabolites, protein sequences	Metabolomics, clinical chemistry, biomarker discovery	hmdb.ca
UK Biobank	Data from 500 000 participants, blood, urine, saliva samples, lifestyle information	Large-scale biomedical database and research resource	ukbiobank.ac.uk
Nephrology-specific repositories			
Nephroseq	Transcriptomic profiles of biopsy samples from patients with kidney disease Clinical metadata from patients including age, sex, UPCR, eGFR Transcriptomic profiles of kidneys from model systems	Identifying disease-related signatures Correlation of gene expression with clinical features	nephroseq.org
NephQTL	Gene expression profiles from biopsy samples, 187 NEPTUNE cohort participants, SNP genotype frequency	Discover glomerular and tubule eQTLs	nephqtl.org
Nephrocell	scRNA-seq data from kidney biopsy samples and organoids	Cell-selective gene marker identification	nephrocell.miktmc.org
Human Kidney eQTL Atlas	Compartment-specific (glomeruli and tubulointerstitial) gene expression profiles	Compartment-specific as well as whole kidney eQTL discovery	susztaklab.com/eqtl
Kidney Interactive Transcriptomics	Single-cell and single nuclear RNA-seq datasets	Cell-selective gene marker identification	humphreyslab.com/SingleCell
Kidney-Omics(Renal Epithelial Transcriptome and Proteome Databases)	Renal Epithelial general proteomics, Specialized Proteomics, Categorized Gene Lists, Chip-Seq Data, Transcriptomic Data, Meta Analysis, Urinary Exosomes, Phospho-proteomics	Gene and protein centred queries in kidney tissues, cells and segments	esbl.nhlbi.nih.gov/Databases/KSBP2/
Rebuilding a Kidney Consortium	scRNA-seq visualizations from kidney biopsy samples	Coordinate studies and data relevant to nephron regeneration Primary data access	rebuildingakidney.org

This table presents some, but not all, of the commonly used database and online website tools. eGFR: glomerular filtration rate; UPCR: urine protein-creatinine ratio

Identified genetic variants associated with kidney disease

Many factors influence kidney function and disease status, with genetic background standing out as a key determining factor among them [27]. Previous studies have identified many monogenic mutations leading to kidney diseases [28] such as Alport syndrome [29] and Fabry disease [30]. The term genetic variation encompasses three scenarios: (1) single nucleotide substitutions, including rare mutations, common polymorphisms, or single nucleotide polymorphisms (SNPs); (2) insertions/deletions (indels); and (3) structural variations. For instance, numerous studies have found that genetic variations in APOL1 significantly increase the risk of various severe kidney diseases among individuals of African descent [31], with the APOL1 G1 variant consisting of two amino acid-changing SNPs (mutations), and the APOL1 G2 variant involving a six-nucleotide deletion. Single nucleotide substitutions represent the most studied type of genetic variation [32]. The term SNP typically refers to single nucleotide changes at specific positions in the genome. While some single nucleotide substitutions may have no apparent effect on phenotype, others may be lethal.

Identifying expression quantitative trait loci analysis (eQTLs) is a critical analytical method for studying the impact of genetic variation on disease. Analyzing the polymorphisms manifested at these loci can demonstrate partial variations in RNA or protein expression of specific gene products. Integration of genomic sequencing data with transcriptomic or proteomic expression data enables the determination of these loci. Importantly, these studies shed light on the functional consequences of gene variants in regulating kidney disease transcriptional mechanisms. For example, using microscopic anatomical samples from 240 glomeruli and 311 tubulointerstitial compartments obtained from human kidney biopsies, genomic regulatory maps of kidney diseases and traits can be constructed [33], and the target loci can be finely mapped via genome-wide association studies (GWAS).

Currently, kidney-specific eQTLs have been employed to identify potential novel disease modifiers and targets, such as the expression of lysosomal β-glucosidase [34] and disease severity. Moreover, compartment-specific eQTLs [35] contribute to the identification of novel gene targets and cellular pathways involved in the progression of CKD, such as TGF-β and DAB2. The increasing number of genetic and transcriptomic studies will further deepen our understanding of the genetic determinants of kidney diseases and help identify the initial insults and transcriptional pathways leading to disease progression in genetically susceptible populations.

Epigenomics mediates crosstalk between genes and environmental factors in the kidney

Emerging evidence suggests that epigenetic regulation contributes to various kidney diseases [36] by playing remarkable roles in mediating crosstalk between genes and the environment, and inducing phenotypic changes [37]. Without changing the primary nucleotide sequence, epigenomics explores heritable mechanisms that control gene expression [37], which are considered to be stable, heritable, and reversible during cell divisions [38]. The most well-studied epigenetic marks include DNA methylation of cytosines [39, 40], histone post-translational modifications (PTMs) [41, 42], and non-coding RNAs [43, 44]. Classically, dense promoter DNA methylation is associated with transcriptional repression [45]. For instance, hypermethylation leads to the loss of HOXA5, resulting in JAG1 expression and NOTCH signaling contributing to kidney fibrosis [46]. However, growing evidence suggests that promoter hypermethylation also appears to be associated with high transcriptional activity [47]. Collectively, targeting DNA methylation and other epigenetic mechanisms has been believed to effectively affect the progression of nephrology [48]. A previous epigenome-wide association study (EWAS) found 19 DNA methylation sites that were significantly and reproducibly associated with eGFR or CKD [49]. And a recent study further demonstrated that methylation risk scores can improve disease state annotation and prediction of kidney disease development [50], providing potential pathways for the development of novel risk stratification methods, suggesting that EWAS can complement genotype variations uncovered by GWAS and provide powerful information about disease susceptibility and causality.

The study of epigenetics, epigenomics, and metabolic memory may fill a critical gap in our understanding of kidney disease development, notably in diabetes, hypertension, and obesity-attributed kidney disease areas. Genetic predisposition, as well as aging, contributes to epigenetic variability, and several environmental factors, including exercise and diet, further interact with the human epigenome [51]. The persistent effects of high glucose through metabolic memory remain a major hurdle in the effective management of diabetic kidney disease (DKD) [52]. The senescence-associated cyclin-dependent kinase inhibitor p21 (Cdkn1a) was the top hit among genes persistently induced by hyperglycemia and was associated with induction of the p53-p21 pathway. Recent research indicates that prolonged expression of tubular p21 in DKD correlates with the demethylation of its promoter and a decrease in DNA methyltransferase 1 (DNMT1) expression, while tubular and urinary p21 levels are linked to the severity of DKD and stay high even with better human blood glucose levels [53]. These studies support not only a role for epigenetics in kidney disease development but also epigenetic alterations as a response to disease, which hold promise for future therapeutic strategies.

Proteomics and metabolomics relate directly to the pathological symptoms and clinical parameters

As downstream molecules of the genome, the proteome and metabolome represent the integrated effects of gene function, also known as the 'functional genome'. The purpose is to understand the genotype–phenotype relationships on a genome-wide scale and to reflects the influence of environmental exposures beyond gene coding [54]. The proteome and metabolome offer distinct advantages in kidney disease [55]: the core specimens for clinical testing of kidney disease, such as blood and urine, contain metabolites (such as urea, creatinine, glucose, and uric acid) and proteins (such as albumin, cystatin C, complement, and parathyroid hormone), relate more directly to the pathological symptoms and clinical parameters observed in patients, which can also serve as dynamic therapeutic targets in response to disease and treatment changes, as well as specialized tools for metabolic biomarker and pathway analysis [56]. Furthermore, compared to the genome, the proteome and metabolome provide biological information at distinct times and locations: as functional products of gene expression, they exhibit considerable dynamism and variability, yielding different results in different locations such as the liver, muscles, kidneys, blood, and urine, and showing significant heterogeneity among tissues like glomerular cells, endothelial cells, and tubular cells [57]. Therefore, targeted proteomics is advantageous for identifying the heterogeneous

disease mechanisms underlying clinical manifestations and identifying drug targets for targeted therapy.

Significantly, in contrast to genomic studies, the proteome and metabolome do not deduce causality. Proteins found in urine could indicate distinct biological activities in the kidney, yet they might also suffer general damage due to the glomerular filtration barrier. Nonetheless, the proteome and metabolome are crucial in comprehending the disease's developmental phase and directing both diagnosis and treatment. An illustrative milestone in kidney proteomic research involves the discovery and precise identification of anti-PLA2R in the serum of MN patients. Serum levels of these autoantibodies correlate with MN disease activity and response to immunosuppression, establishing it as a widely used non-invasive marker for MN detection in clinical settings [58]. Similar approaches have identified additional markers such as THSD7A and amyloid A1 [59], which offer additional prognostic insights based on PLA2R antibody levels.

Single cell and spatial multi-omics: defining the atlas of cell states and niches in kidney

Understanding kidney disease relies on recognizing the complexity of different renal cell types and states, their associated molecular profiles, and interactions within tissue neighborhoods. When kidney function progressively declines after injury, dynamic acute and chronic changes occur in the renal tubules and surrounding interstitial niche, leading to molecular diversity at the single-cell level [60]. The heterogeneity among cells is constituted by multiple complex intracellular and intercellular interactions, hierarchical structures, and environmental variables, as well as temporal and spatial informational regulation [61]. Therefore, it is imperative to employ finely-grained single-cell and spatially-resolved multi-omics approaches to understand the molecular hierarchy of a single cell from genome to phenome. Especially for RNA sequencing, This most widely used technology in genomics tool box has evolved from classic bulk RNA sequencing to popular single cell RNA sequencing and newly emerged spatial RNA sequencing [62].

In recent years, the explosive growth of single-cell technologies has unveiled previously underappreciated cellular heterogeneity and new cell state associations with gender, diseases, development, and other processes [63]. Single-cell transcriptomics, currently the most mature single-cell omics method, initially redefined cell types and subtypes in the kidney through the transcriptional fingerprints of individual cells, generating comprehensive cellular atlases and identifying cell type-specific markers [64]. Recent research developments have extensively utilized these cell-specific gene maps to delineate pathways of disease progression and identify new molecular targets. For instance, a comprehensive analysis of macrophage transcriptomes in early diabetic nephropathy revealed dynamic changes in cellular phenotypes during disease progression and enhanced expression of pro-inflammatory or anti-inflammatory genes in a subset-specific manner [65].

Spatial omics is widely acclaimed as the emerging frontier of life sciences [66]. Since spatial information in tissue context remains elusive despite the findings provided by scRNA-seq technologies regarding cellular heterogeneity within tissue types, it has given rise to the development of spatial omics [67]. Methods combining single-cell and spatial omics facilitate a deeper understanding of cell type-specific metabolism in complex tissues and greatly illustrate spatial characteristics and patterns of cells and genes. For example, single-cell spatial genomics studies of the human kidney can identify cell types as well as complex states associated with molecular signatures, and interactions within tissue neighborhoods in renal disease by establishing a multidimensional single-cell-referenced map of healthy and damaged cell states and ecological niches [68]. Thus, the rise of 'spatial multi-omics' builds upon spatial single-omics (spatial genomics [69], spatial proteomics [70], spatial metabolomics [70], etc.) and encompasses a range of emerging technologies including array-based spatial transcriptomics, microfluidic deterministic barcoding strategies [71–73], DNA antibody labeling [74–77], and multiplex single-molecule fluorescence in situ hybridization [78, 79], offering a deeper understanding of molecular patterns of complex kidney tissues at multiple hierarchical dimensions.

How to select proper machine leaning strategies

Past difficulties in conventional analysis methods underscore the necessity for computers to possess the ability to acquire knowledge autonomously. ML arises at the intersection of statistics and computer science, where the former learns relationships from data while the latter emphasizes efficient computational algorithms [80]. Moreover, ML holds a crucial position for datasets that are too vast (comprising numerous independent data points) and intricate (involving numerous diverse features) for manual examination, or for the requirements to develop an automated, replicable, and efficient research route [81]. For instance, computer-based methods can identify drug–target interactions (DTI), reducing traditional experimental costs [82], especially playing a significant role in new drug development processes. Utilizing omics data with ML approaches can establish classification models for various types of renal diseases [83], and even engage in numerous steps of patient disease management, such as predicting clinical risks, improving clinical care, assisting clinicians in diagnosis and treatment [84]. In practical clinical applications, the Food and Drug Administration has already permitted clinicians to utilize AI in various domains, such as diabetic retinopathy [85], where AI can perform routine diagnoses without the need for ophthalmologists to confirm them [86].

ML is becoming an indispensable tool in the analysis of biological data workflows. As its application proliferates explosively, understanding ML theory, appropriately selecting ML strategies based on biological theories [81], and evaluating the suitability of these methods are becoming increasingly critical (Fig. 2).

Supervised learning versus unsupervised learning

Defined by the presence/absence of labels in the datasets, ML can be classified into supervised learning and unsupervised learning (Fig. 2, Step 3).

Supervised learning harnesses the power of labeled data to train models. Through training, the machine learns the relationship between features and labels, enabling it to predict labels for new unlabeled feature data. For instance, gene expression prediction for genomic genes using classical labeled genes [87] or protein secondary structure prediction based on existing protein databases [88]. Supervised learning can further be categorized into classification and regression tasks. Common algorithms include Support Vector Machine (SVM, a powerful regression and classification model that uses kernel functions to transform a non-separable problem into an easily solvable separable one), K-Nearest Neighbors (one of the simplest classification methods), and Naive Bayesian Model (stable classification efficiency with few parameters to estimate) [89]. Additionally, widely used tree-based models use a series of if-then rules to generate predictions from one or more decision trees. Examples include Random Forest

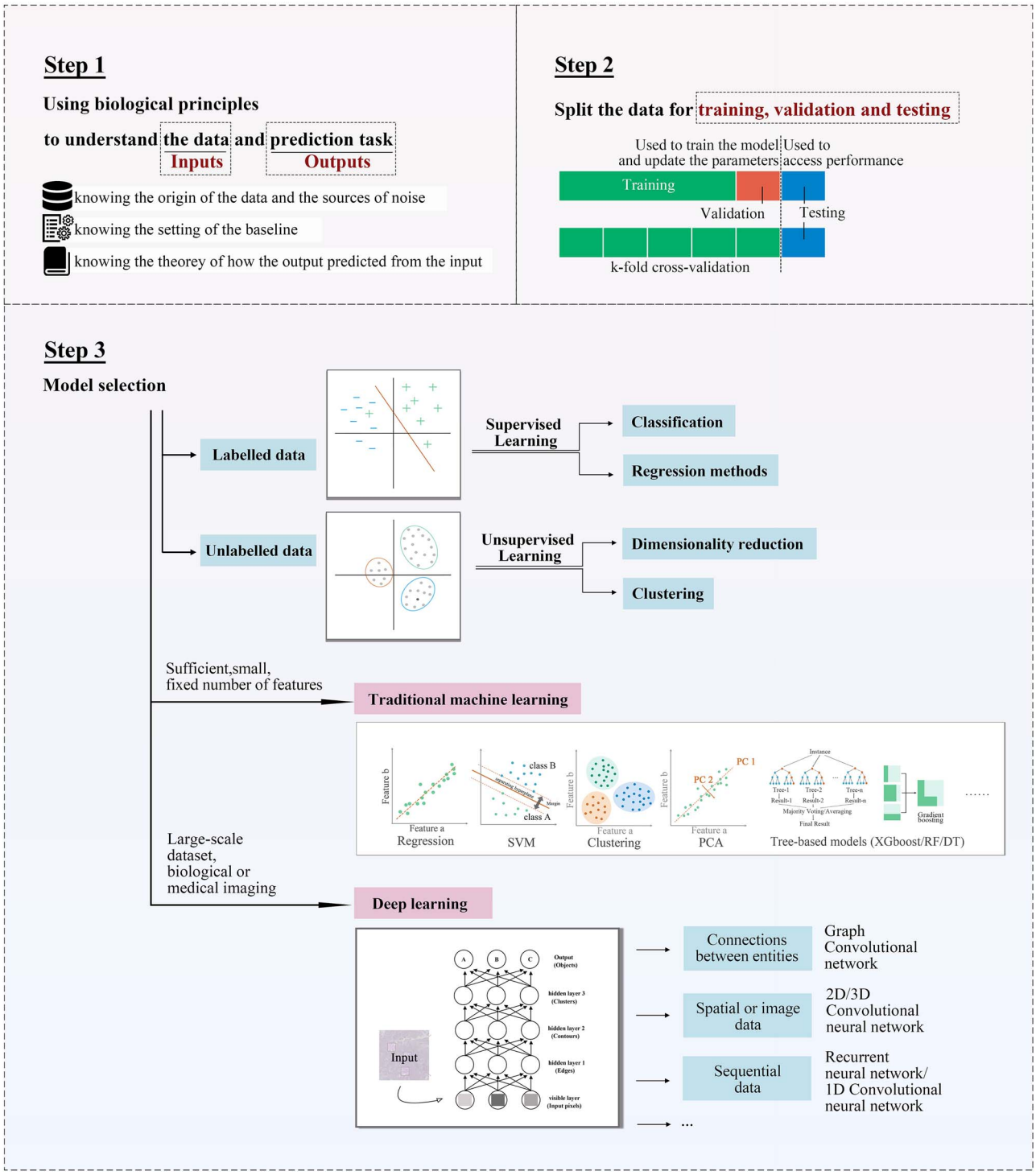

Figure 2. Steps of training an ML model: in general, the process of training ML models using biomedical data involves three primary steps. The first step entails comprehensively understanding the input data and the tasks to be performed, thereby grasping the problem and significance from a biomedical perspective. The second step involves data partitioning for training, validation, and testing purposes. The training set is directly employed to train the model, the validation set is used to monitor training progress, and the testing set is utilized to evaluate model performance. Meanwhile, k-fold cross-validation with a separate testing set can also be employed. The third step involves model selection, contingent upon the nature of the data and prediction tasks, such as the number of features available per data point and the presence of labels. Subsequently, the accuracy of the selected model on the testing set is assessed and validated. Note: this schematic shows a fundamental process, not all scenarios. Additional issues like overfitting and hyperparameter tuning also need consideration.

(RF, an ensemble method that builds many decision trees in parallel), and eXtreme Gradient Boosting (XGBoost [90], an ensemble method that builds many decision trees sequentially and is known for its exceptional performance in both speed and accuracy).

In contrast, unsupervised learning focuses on uncovering hidden structures and patterns within unlabeled data. The unsupervised learning models are used for three main tasks: clustering, association, and dimensionality reduction. For instance,

predicting drug responsiveness based on gene expression profiles of new patients where different patient subgroups are identified solely based on expression profiles without any information regarding drug responsiveness [91]. These identified subgroups can then be further studied for differential drug responsiveness, and new patients can be assigned to the most similar cluster based on their own expression profiles.

Traditional ML versus deep learning

Previously, some fundamental ML algorithms were mentioned and illustrated in Fig. 2, Step3, which are often referred to as 'traditional machine learning'. When developing ML methods for biological data, traditional ML is still regarded as the primary exploratory domain for finding the most suitable approaches for a given task. Many packages can be utilized to train such models, including scikit-learn [92] in Python, caret [93] in R, and MLJ [94] in Julia.

In recent years, deep learning (DL) has emerged as the most effective approach for many tasks and a leading trend. Due to the large volume, diversity, heterogeneity, complexity, and often ill-understood nature of data in biology and medicine, DL techniques may be particularly well-suited to solving problems in these data-rich disciplines [95]. As a specific type of ML, DL conceptualizes the vast world as nested hierarchical systems of concepts, defining complex concepts in terms of simpler ones. The specific operation involves presenting inputs in the visible layer, then extracting a series of increasingly abstract features in hidden layers, and finally establishing an output layer. Artificial Neural Networks (ANNs) are a method of DL and the primary mode adopted. Of which Convolutional Neural Networks (CNNs) are specifically designed for processing data with grid-like structures, making them well-suited for image-like data and widely applied in various medical images, including radiology, ultrasound, endoscopy, ophthalmology, and pathology. Currently popular algorithms include R-CNN, Fast R-CNN, Faster R-CNN, PFN, PSPNet, SSD, YOLO, CenterNet, and EfficientNet [96].

However, despite its numerous advantages, the application of DL remains restricted to specific domains characterized by large datasets (e.g. millions of data points), numerous features per data point, and highly structured features (e.g. adjacent pixels in images). Biological data, such as DNA, RNA, protein sequences [97], and microscopic images [98], fulfills these criteria and has seen successful implementation. Nevertheless, the demand for substantial datasets can also render DL suboptimal, even when the other conditions are met. Thus, developing architectures for deep neural networks and training them remains a time-consuming and computationally expensive endeavor. In contrast, traditional models such as SVM and RF offer faster development and testing cycles for specific problems. Therefore, when exploring and selecting ANNs, it is advisable to concurrently train a traditional ML model and conduct a systematic comparison with ANN-based models [99].

Data augmentation [100] significantly expands the amount and variety of data available for training without actually collecting new samples. This is particularly valuable for biological and medical data, where collecting large datasets is challenging due to privacy concerns and labeling costs. Data augmentation techniques range from basic yet highly effective transformations such as cropping, padding, and flipping, to advanced generative models [101]. These data augmentation techniques can be divided into two broad categories: transformation of original data (including affine, erasing, elastic and pixel-level) and generation of artificial data (including generative models, feature mixing, model based and reconstruction-based method). Depending on the nature of the input and the visual task, different data augmentation strategies may perform differently. For this reason, it is conceivable that each biological task requires specific augmentation strategies that generate plausible data samples and effectively regularize deep neural networks. For example, automatically segmenting kidneys in different clinical imaging modalities remains a significant challenge due to the kidneys' varied shapes and image intensity distributions. To build a robust kidney segmentation model, several studies have been proposed in the literature of computed tomography [102, 103], magnetic resonance [104], and ultrasound [105], A recent systematic literature review found consistent benefits across all organs, modalities, and tasks, with the use of data augmentation, from the simplest affine transformations to the most complex generative models [106].

Current applications and clinical insights in kidney research

In summary, there are three key aspects of kidney disease applications (Fig. 3): (1) accurate prediction: predicting the risk of disease progression and improving medical decisions; (2) mechanism elucidation: emphasizing the extraction of regularities from the biological internal mechanisms to further understand the molecular mechanisms of diseases; and (3) digital pathological image analysis of kidneys.

Making accurate prediction

Predicting the risk of disease progression

The risk prediction models not only aid clinicians in diagnosis and treatment but also identify new risk factors for timely intervention in disease management. Acute kidney injury (AKI) is a common life-threatening condition in kidney disease [107], responsible for 11% of inpatient deaths due to failure to recognize and treat it promptly. Hence, early identification, timely detection of risk factors and early intervention are vital for their survival and prognosis [108–111]. A common framework involves inputting features at each time point into the statistical model and outputting the probability of any severity stage of AKI occurring in a future time, which exceeds a selected operational threshold to produce a positive prediction. As a case in point, a DL-based continuous AKI risk prediction model can predict AKI events of any severity occurring 48 hours in advance with an accuracy of 55.8% and predict 90.2% of AKI cases requiring dialysis [108], demonstrating its universality and potential application as a clinical decision support tool for improving AKI detection and outcomes [112].

End-stage renal disease (ESRD) marks the final stage of renal failure. Early prevention and intervention can significantly postpone the initiation of renal replacement therapy, improving patient quality of life. Recent studies have utilized ANNs to develop neural network classifiers [113], also known as clinical decision support systems, to predict ESRD based on clinical data and omics data from kidney biopsies, enabling the identification of high-risk individuals, forecasting time-to-event endpoints, and conducting external validation through follow-up. For example, for type 1 diabetes patients, currently developed ESRD risk prediction models can predict the risk of ESKD for 5 years based on routine clinical data (age, gender, duration of diabetes, estimated glomerular filtration rate, micro and macroalbuminuria, glycated hemoglobin, smoking, and history of cardiovascular disease), providing a basis for clinical decision-making [114]. However, the 5-year prediction period is relatively short for type 1 diabetes patients (most of whom are young, yet ESRD progression is very long), posing a common challenge for such prediction

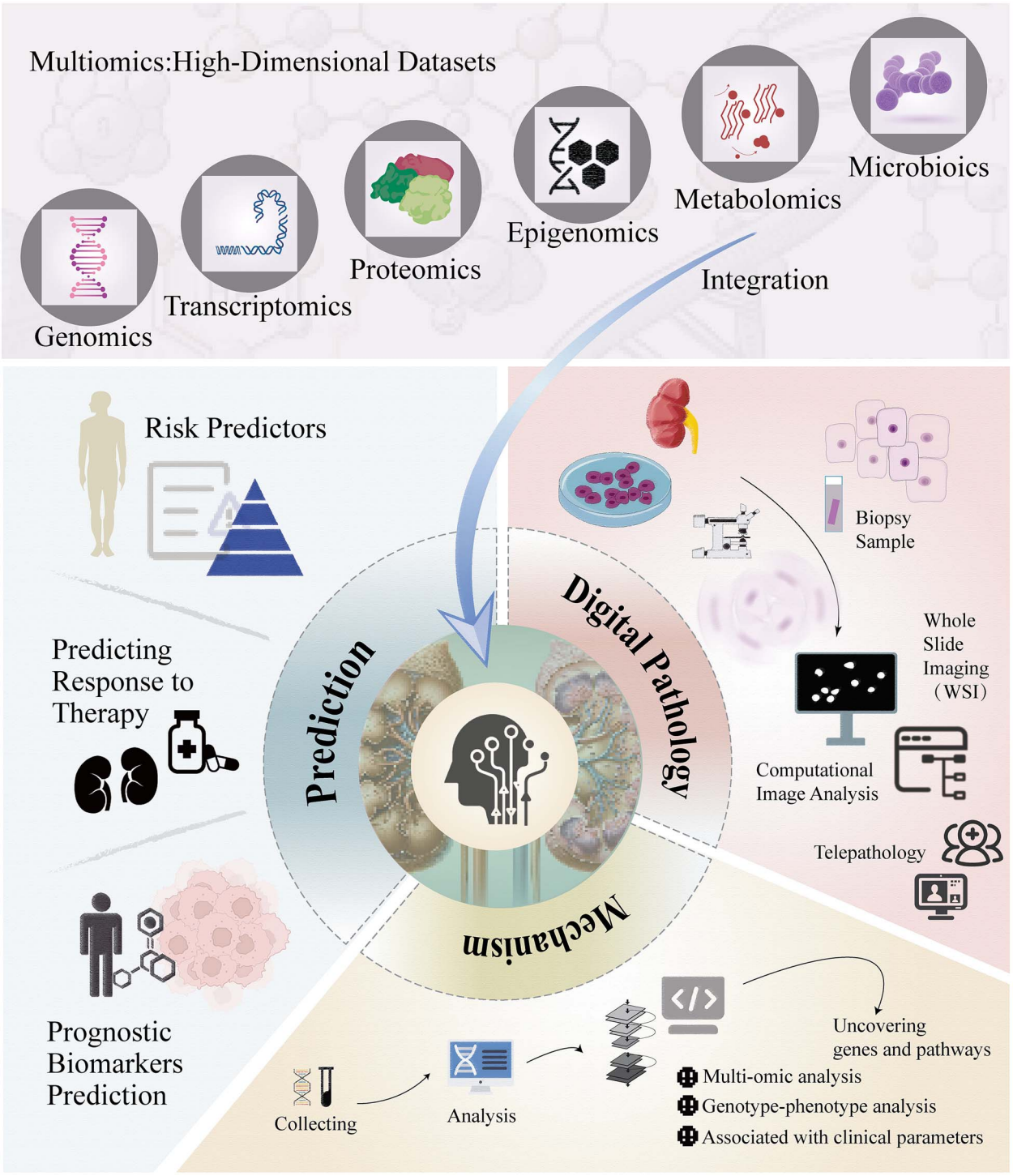

Figure 3. Employing ML to integrate multi-omics molecular data and clinical data for kidney disease research.

models [115]. One solution is to consider establishing lifetime prediction models to cover longer time spans, which can not only improve the accuracy but also estimate the effects of lifestyle changes and preventive drug use (such as reducing blood pressure, HbA1c, etc.).

Predicting response to treatment

As chronic diseases, kidney diseases critically require novel methods to elucidate intrinsic therapeutic effects and evaluate treatment outcomes. After a certain treatment, transcriptomic and metabolomic data can quantitatively compare the patient's activation levels in a certain pathway at different time points to predict the response to a specific therapy. To elaborate, by connecting genes, drugs, and disease states through common gene expression features [116], the mechanism of small molecules can be inferred from transcript expression levels, allowing functional annotation of genetic variation in disease genes, and informing clinical trials for drug development. The quantitative scoring of transcriptional features has been used to identify diverse features related to kidney disease, including features of podocyte

development reactivated [117] in patients with glomerular disease and endothelial cell characteristics indicating the response to steroids in patients with focal segmental glomerulosclerosis [118]. These features would be essential for identifying specific pathway activations and evaluating drug efficacy in disease settings.

The predictive model for renal replacement therapy is another important research area. For example, transplant renal dysfunction is a common adverse outcome observed after kidney transplantation. A computer-aided diagnostic system based on DL can early detect acute renal transplant rejection [119]. An unsupervised archetype analysis learning method integrating clinical, functional, immunological, and histological parameters can stratify the heterogeneity of transplant renal dysfunction based on different long-term allograft survival rates and establish an online application for clinical practice based on real patients [120].

Prognostic biomarkers prediction

CKD typically evolves over many years, often with a long latent period where the disease remains clinically silent. Diagnosis, evaluation, and treatment rely primarily on biomarkers, which serve as vital indicators marking structural and functional changes in organisms, crucial for disease staging, drug development, and treatment assessment. Studies have been conducted to predict potential targets and new molecular markers among a variety of kidney-related diseases such as FOSL1/2 in IgA Nephropathy (IgAN) [121], IFI27 in lupus nephritis [122, 123], DUSP1 in hypertensive nephropathy [124], and RPTOR in diabetic nephropathy [125]. Nonetheless, despite the theoretical significance of these biomarkers, they still need high-quality prospective cohort to validate their clinical utility and mechanistic implications.

The development of new biomarkers contributes to the advancement of existing clinical diagnostics. Currently, the diagnostic type of CKD and its severity are based on clinical features such as eGFR, proteinuria [126], and pathologic features from renal biopsy samples. However, this categorization fails to capture the diversity of molecular pathways that may lead to phenotypically similar renal diseases, which in turn hampers our ability to predict long-term prognosis or to test and apply targeted therapies. Therefore, an increasing number of studies are focusing on developing new biomarkers to identify CKD progression, improving the diagnostic classification of CKD [127]. Algorithms based on differential network enrichment analysis can partition lipidomic profiles associated with CKD progression severity, suggesting that alterations in triacylglycerol and cardiolipin-phosphatidylethanolamine precede the clinical outcome of ESRD by several years [128]. In addition, identifying injury features of the kidney in urine proteomics is also a significant research issue [129–132]. Integrating urine proteomic datasets with kidney biopsy tissue transcriptomic data and other clinical information can develop risk prediction models for CKD progression. Urinary epidermal growth factor (uEGF) may be an effective biomarker for predicting pediatric CKD progression [129], where low levels of uEGF can predict CKD progression, and reflecting the degree of tubulointerstitial damage.

Identify novel disease mechanisms

For complicated diseases like nephrology, distinguishing causative factors is critical to clarifying diagnosis and guiding treatment selection. Nevertheless, substantial variability in disease progression risk and treatment response within identical diagnostic conditions underscores the heterogeneity of underlying molecular mechanisms. Thus, identifying pivotal therapeutic pathways for complex, multifactorial diseases and elucidating their intrinsic mechanisms remain formidable challenges [133]. High-throughput analysis offers new opportunities for understanding the intrinsic molecular mechanisms corresponding to these complex pathophysiological processes. Integrated multi-omics approaches can be used for novel disease classification [127], reclassifying patients into molecularly defined subgroups, thereby revealing the intrinsic molecular mechanisms and biological pathways of various diseases. For instance, integrating IgAN gene expression datasets into blood cells and systematically validating them through experimental verification to identify aberrantly expressed genes and biological pathways [134]. It was found that these aberrantly expressed genes and pathways are mainly enriched in the intestinal immune network and are involved in IgA production and autophagy processes. Additionally, PTEN in B cells may be involved in the mechanism of Gd-IgA1 production. Another transcriptomic analysis found expression characteristics and possible regulatory mechanisms of interferon-stimulated genes in lupus patients [135]: monocytes, B cells, dendritic cells, and granulocytes significantly increased, while subsets of T cells significantly decreased. Genomic and epigenomic omics research has also identified kidney mechanisms mediated by genes associated with hypertension susceptibility, revealing 179 unique renal genes involved in blood pressure control [136].

Radiomics and image analysis: digital pathology

With collaborative efforts in collecting, analyzing, and integrating pathological data, renal pathology is entering the digital era [13]. Conventional stained images on slides are being transformed into digital format images, known as whole slide images (WSI), which involve four consecutive processes [137]: image acquisition, storage, processing, and visualization. WSI contains rich information from traditional staining, single-channel, or multi-channel immunohistochemistry staining, as well as multi-omics data [138]. Continuous technological advancements in digital scanners, image visualization methods, and their integration with algorithms provide opportunities for the application and development of WSI. WSI has been widely used in various aspects such as digital diagnosis, remote consultations, and research assistance, with studies confirming its high consistency with traditional light microscopy (CLM) for diagnosis [139].

The main uses of digital imaging in renal pathology can be divided into three main operational modes: telepathology, digital pathology, and computational image analysis [13]. Digital pathology includes digital workflows and imaging solutions aiming to create an application environment for accessing, managing, interpreting, and searching WSI or other digital content. Telepathology, one of the earliest applications of WSI, involves transmitting digital images to another remote site for analysis. It has now become a common tool for real-time assessment of biopsy tissue adequacy and diagnosis with widespread validation [140]. Especially for kidney transplantation, assessment models, evaluating the proportion of glomerulosclerosis can rapidly and accurately assess whether living donor kidney tissue is suitable for transplantation [141], potentially becoming an important part of clinical assessment of living donor kidney biopsies. Telepathology can significantly optimize the workflow of nephrologists in the process of kidney transplantation procurement and evaluation. Computational image analysis, which generates extensive data, relies heavily on advanced ML techniques to comprehensively extract features, patterns, and information in tissue pathology.

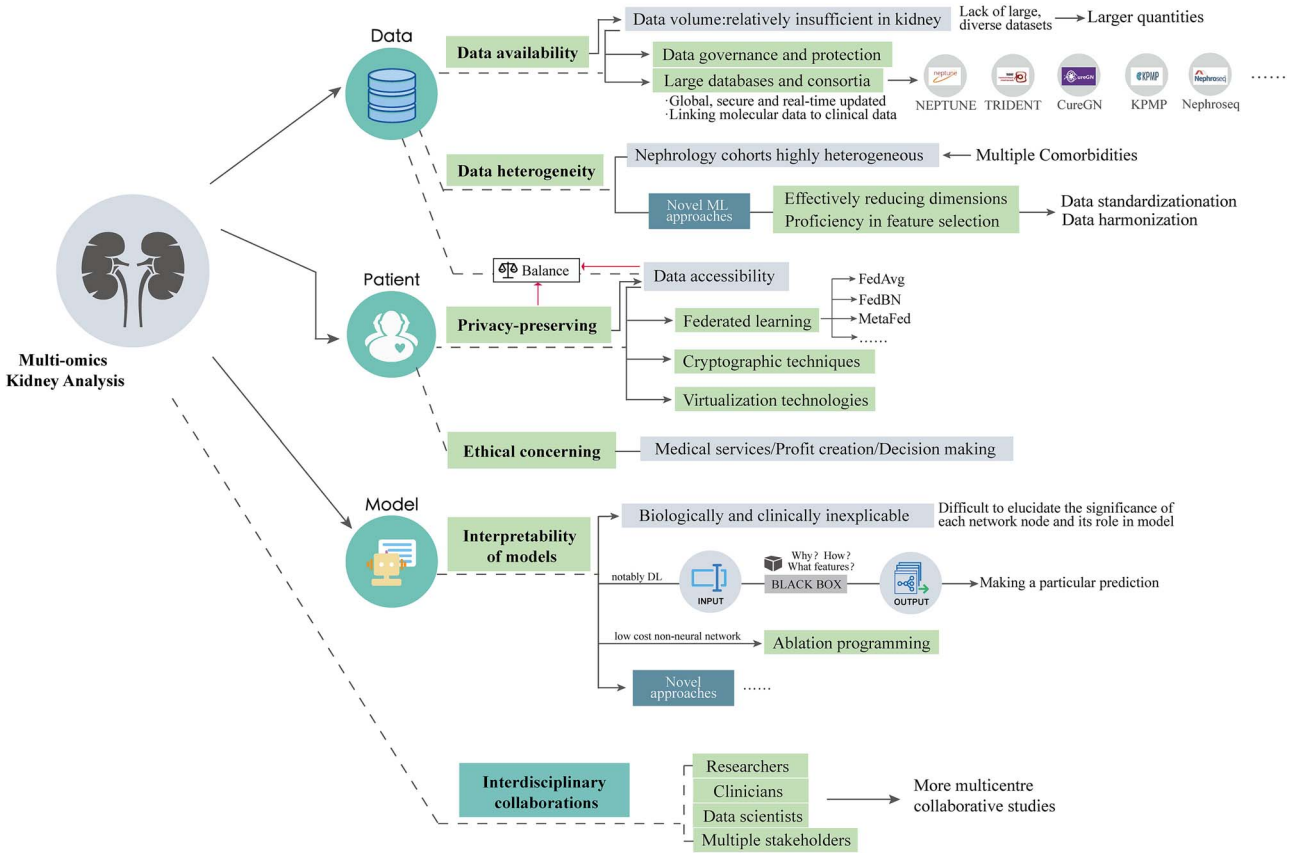

Figure 4. The framework of current challenges and related methods in multi-omics kidney analysis. Overall, due to the structural and mechanistic complexity of kidney disease and the relative scarcity of research and data, there are challenges related to data availability, data heterogeneity, and model interpretability. Additionally, privacy protection concerns are more pronounced given the long-term chronic nature of the disease. Addressing these challenges requires substantial collaboration across various fields and global cooperation.

In the past, ML was commonly used for quantitative analysis to assist in identifying pathological features, such as histological features of diabetic nephropathy in rats [142], identifying glomerular lesions and intrinsic glomerular cell types [143]. However, with the explosive development of algorithms, the use of ML has the potential to elevate digital images from their basic role as visual assessments of disease status to more complex and comprehensive roles, such as facilitating disease trajectory prediction and risk scoring for IgAN [144]. The implementation of these novel tools is positioning nephropathology at the forefront of defining new, integrated, biologically and clinically homogeneous disease categories, identifying patients at risk of progression, and transforming current paradigms for the treatment and prevention of kidney diseases.

Challenges and perspectives

It is worth noting that a meta-analysis showed that ML models did not outperform traditional statistical prediction models like logistic regression (LR) in predicting AKI [145]. We must recognize that various AI technologies are still in development, and there remains a gap in achieving the ideal form of AI. Although DL is capable of tackling singular issues, it falls short as a comprehensive remedy for *a range* of different *problems* [146]. With approximately 33% of research being irreproducible in the stem cell field [147] and a significant lag in the field of big science and big data in nephrology (as mentioned before) [148], there are some common and specific issues that need to be considered (Fig. 4).

Data availability

One of the main challenges in kidney diseases is the relative scarcity of large, diverse datasets, particularly in the context of medical imaging-based DL [149]. Additionally, due to the involvement of multiple technical domains, the quality and reliability of data often face difficulties such as batch effects [150], missing values [151], and measurement errors. Moreover, not only initial model training requires data, but continuous model training also relies on ongoing data supplement, validation, and improvement. Therefore, generating more global, secure and real-time updated invaluable resources for the research and clinical community is imperative. Various initiatives have been undertaken to achieve a comprehensive characterization of kidney biopsies across various CKD subtypes, including the Nephrotic Syndrome Study Network [152], *Transformative Research in Diabetic Nephropathy* [153], Cure Glomerulonephropathy [154], and Kidney Precision Medicine Project [155]. When larger quantities are available, it becomes feasible to consider using more highly parameterized models, which hold great transformative potential. For instance, linking molecular data to EHRs could uncover molecular phenotypes of kidney diseases, enabling targeted monitoring, personalized treatment, and improved family counseling [156].

Data heterogeneity

Specifically, patients with kidney disease often have comorbidities that make nephrology cohorts highly heterogeneous. Therefore, data standardization and data harmonization with the capability to arbitrarily integrate multi-modal datasets stand out to be

concerned [157]. Additionally, since the model training processing also suffered from the classic 'curse of dimensionality', effectively reducing dimensions and selecting the most influential features and variables are crucial. To address these challenges, multiple new ML approaches have been proposed and employed such as a new deep neuro-fuzzy system consists of a deep structure in the rule layer and novel architecture in the fuzzifier layer to classify kidney cancer subgroups [158], as well as algorithms like RECODE for reducing noise in scRNA-seq data [159] and multifactor dimensionality reduction for analyzing exponentially growing SNPs [160].

Model interpretability

While pathological imaging radiomics research holds significant importance in nephropathology studies, a major limitation of current DL models is their lack of interpretability compared to basic statistical regression models. This makes it challenging to understand the significance of each network node and its role in model efficacy. In contrast, the low cost of training non-neural networks supports ablation programming [161], which helps identify useful features, leading to more robust, efficient, and interpretable models by revealing the significance of different model components and making the decision-making process more transparent.

Recognizing this challenge, the ML community has also focused on developing new techniques to elucidate 'black-box' DL models. For example, activation maximization encompasses algorithms that use gradient descent to find inputs maximizing the model's response, aiming to generate inputs that best represent a desired outcome [146].

Privacy preserving and data accessibility

As data dissemination for training purposes increases, the standardization of secure data storage, retrieval, and access becomes crucial [162]. Sensitive medical information, such as CKD data containing long-term private information, cannot be shared without ensuring patient confidentiality and data security. Thus, achieving an appropriate balance between data accessibility and privacy preservation is essential and presents significant challenges. Algorithms have been developed for efficient federated learning [163], where many clients collaboratively train a model under the orchestration of a central server while keeping the training data decentralized, including FedAvg, FedBN, and the recent MetaFed [164]. Additionally, cryptographic techniques [165] and other alternative models [166] such as virtualization technologies have been introduced, enabling analysis without sharing the actual data.

Interdisciplinary collaborations

Combined efforts from researchers, clinicians, and data scientists, along with engagement from multiple stakeholders including healthcare organizations, government bodies, and the pharmaceutical and biotech industries, are necessary to better understand the pathogenesis and prognosis of kidney disease, which is pivotal for final clinical deployment. The kidney community must mobilize to conduct more multi-center collaborative studies and to collect more data on metrics for monitoring diseases such as AKI and CKD.

Conclusions

Understanding and optimizing the advantages, strategies, implementation, and limitations of these ML approaches and multi-omics techniques are essential for translating research findings into clinical practice. Overall, this integration has emerged as a revolutionary tool in the era of high-throughput kidney research. The success of this new integrated scientific paradigm undoubtedly requires active collaboration and communication across various disciplines. We believe that these specific measures will significantly contribute to the clinical prevention, early diagnosis, disease management, and monitoring of kidney diseases, thereby facilitating accurate disease diagnosis and personalized treatment approaches.

Key Points

- This paper provides a comprehensive review of current integration of multi-omics and machine learning in nephrology.
- We review the multi-omics data generated and utilized in kidney research, especially for genetic variants as a key determining factor of disease and proteomics, epigenomics mediate crosstalk between genes and environmental factors, proteomics and metabolomics that relate directly to the pathological symptoms and clinical parameters, as well as insights on single cell and spatial multi-omics defining the atlas of cell states and niches.
- We demonstrate the general workflow for appropriately selecting ML strategies based on biological theories.
- The key purpose of the integration is summarized into three aspects: making accurate prediction, including risk predictors, predicting response to therapy and prognostic biomarkers prediction; uncovering further mechanism; and digital pathological image analysis of kidneys.
- We discuss major kidney-specific challenges and possible methods as a general framework about data availability, data heterogeneity, model interpretability, data accessibility, privacy-preserving issues, and our expectations of active interdisciplinary collaborations.

Supplementary data

Supplementary data are available at *Briefings in Bioformatics* online.

Funding

This work was supported by the Elite Medical Professionals project of China-Japan Friendship Hospital(NO.ZRJY2021-BJ07).

Conflict of interest

The authors declare that they have no conflicts of interest.

Data availability

The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

References

1. Foreman KJ, Marquez N, Dolgert A. *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* 2018;**392**:2052–90. https://doi.org/10.1016/S0140-6736(18)31694-5.

2. Chen TK, Hoenig MP, Nitsch D. *et al.* Advances in the management of chronic kidney disease. *BMJ* 2023;**383**:e074216. https://doi.org/10.1136/bmj-2022-074216.
3. Wang L, Xu X, Zhang M. *et al.* Prevalence of chronic kidney disease in China: results from the Sixth China Chronic Disease and Risk Factor surveillance. *JAMA Intern Med* 2023;**183**:298–310. https://doi.org/10.1001/jamainternmed.2022.6817.
4. Luyckx VA, Al-Aly Z, Bello AK. *et al.* Sustainable development goals relevant to kidney health: an update on progress. *Nat Rev Nephrol* 2021;**17**:15–32. https://doi.org/10.1038/s41581-020-00363-6.
5. Precision medicine in nephrology. Precision medicine in nephrology. *Nat Rev Nephrol* 2020;**16**:615–5. https://doi.org/10.1038/s41581-020-00360-9.
6. Collins FS, Varmus H. A new initiative on precision medicine. *N Engl J Med* 2015;**372**:793–5. https://doi.org/10.1056/NEJMp1500523.
7. Jang Y, Choi T, Kim J. *et al.* An integrated clinical and genomic information system for cancer precision medicine. *BMC Med Genomics* 2018;**11**:34. https://doi.org/10.1186/s12920-018-0347-9.
8. Verma A, Chitalia VC, Waikar SS. *et al.* Machine learning applications in nephrology: a bibliometric analysis comparing kidney studies to other medicine subspecialities. *Kidney Med* 2021;**3**:762–7. https://doi.org/10.1016/j.xkme.2021.04.012.
9. Kang M, Ko E, Mersha TB. A roadmap for multi-omics data integration using deep learning. *Brief Bioinform* 2022;**23**:bbab454. https://doi.org/10.1093/bib/bbab454.
10. Rhee EP. How omics data can Be used in nephrology. *Am J Kidney Dis* 2018;**72**:129–35. https://doi.org/10.1053/j.ajkd.2017.12.008.
11. Vandereyken K, Sifrim A, Thienpont B. *et al.* Methods and applications for single-cell and spatial multi-omics. *Nat Rev Genet* 2023;**24**:494–515. https://doi.org/10.1038/s41576-023-00580-2.
12. Mayerhoefer ME, Materka A, Langs G. *et al.* Introduction to radiomics. *J Nucl Med* 2020;**61**:488–95. https://doi.org/10.2967/jnumed.118.222893.
13. Barisoni L, Lafata KJ, Hewitt SM. *et al.* Digital pathology and computational image analysis in nephropathology. *Nat Rev Nephrol* 2020;**16**:669–85. https://doi.org/10.1038/s41581-020-0321-6.
14. Li R, Li L, Xu Y. *et al.* Machine learning meets omics: applications and perspectives. *Brief Bioinform* 2022;**23**:bbab460. https://doi.org/10.1093/bib/bbab460.
15. Kaur N, Bhattacharya S, Butte AJ. Big data in nephrology. *Nat Rev Nephrol* 2021;**17**:676–87. https://doi.org/10.1038/s41581-021-00439-x.
16. Cirillo D, Valencia A. Big data analytics for personalized medicine. *Curr Opin Biotechnol* 2019;**58**:161–7. https://doi.org/10.1016/j.copbio.2019.03.004.
17. Levy SE, Boone BE. Next-generation sequencing strategies. *Cold Spring Harb Perspect Med* 2019;**9**:a025791. https://doi.org/10.1101/cshperspect.a025791.
18. Koboldt DC, Steinberg KM, Larson DE. *et al.* The next-generation sequencing revolution and its impact on genomics. *Cell* 2013;**155**:27–38. https://doi.org/10.1016/j.cell.2013.09.006.
19. Mann M, Kumar C, Zeng W-F. *et al.* Artificial intelligence for proteomics and biomarker discovery. *Cell Syst* 2021;**12**:759–70. https://doi.org/10.1016/j.cels.2021.06.006.
20. Zhao C, Dong J, Deng L. *et al.* Molecular network strategy in multi-omics and mass spectrometry imaging. *Curr Opin Chem Biol* 2022;**70**:102199. https://doi.org/10.1016/j.cbpa.2022.102199.
21. Porcu E, Sadler MC, Lepik K. *et al.* Differentially expressed genes reflect disease-induced rather than disease-causing changes in the transcriptome. *Nat Commun* 2021;**12**:5647. https://doi.org/10.1038/s41467-021-25805-y.
22. Jiao Y, Jiang S, Wang Y. *et al.* Activation of complement C1q and C3 in glomeruli might accelerate the progression of diabetic nephropathy: evidence from transcriptomic data and renal histopathology. *J Diabetes Investig* 2022;**13**:839–49. https://doi.org/10.1111/jdi.13739.
23. Jiao Y, Liu X, Shi J. *et al.* Unraveling the interplay of ferroptosis and immune dysregulation in diabetic kidney disease: a comprehensive molecular analysis. *Diabetol Metab Syndr* 2024;**16**:86. https://doi.org/10.1186/s13098-024-01316-w.
24. Lu H-T, Jiao Y-Y, Yu T-Y. *et al.* Unraveling DDIT4 in the VDR-mTOR pathway: a novel target for drug discovery in diabetic kidney disease. *Front Pharmacol* 2024;**15**:1344113. https://doi.org/10.3389/fphar.2024.1344113.
25. Jiang S, Jiao Y, Zou G. *et al.* Activation of complement pathways in kidney tissue may mediate tubulointerstitial injury in diabetic nephropathy. *Front Med (Lausanne)* 2022;**9**:845679. https://doi.org/10.3389/fmed.2022.845679.
26. Yang Y, Zou G, Wei X. *et al.* Identification and validation of biomarkers in membranous nephropathy and pan-cancer analysis. *Front Immunol* 2024;**15**:1–13. https://doi.org/10.3389/fimmu.2024.1302909.
27. Manzoni C, Kia DA, Vandrovcova J. *et al.* Genome, transcriptome and proteome: the rise of omics data and their integration in biomedical sciences. *Brief Bioinform* 2018;**19**:286–302. https://doi.org/10.1093/bib/bbw114.
28. Groopman EE, Marasa M, Cameron-Christie S. *et al.* Diagnostic utility of exome sequencing for kidney disease. *N Engl J Med* 2019;**380**:142–51. https://doi.org/10.1056/NEJMoa1806891.
29. Kashtan CE. Alport syndrome: achieving early diagnosis and treatment. *Am J Kidney Dis* 2021;**77**:272–9. https://doi.org/10.1053/j.ajkd.2020.03.026.
30. Kim JW, Kim HW, Nam SA. *et al.* Human kidney organoids reveal the role of glutathione in Fabry disease. *Exp Mol Med* 2021;**53**:1580–91. https://doi.org/10.1038/s12276-021-00683-y.
31. Friedman DJ, Pollak MR. APOL1 nephropathy: from genetics to clinical applications. *Clin J Am Soc Nephrol* 2021;**16**:294–303. https://doi.org/10.2215/CJN.15161219.
32. Pollak MR, Friedman DJ. The genetic architecture of kidney disease. *Clin J Am Soc Nephrol* 2020;**15**:268–75. https://doi.org/10.2215/CJN.09340819.
33. Han SK, McNulty MT, Benway CJ. *et al.* Mapping genomic regulation of kidney disease and traits through high-resolution and interpretable eQTLs. *Nat Commun* 2023;**14**:2229. https://doi.org/10.1038/s41467-023-37691-7.
34. Gu X, Yang H, Sheng X. *et al.* Kidney disease genetic risk variants alter lysosomal beta-mannosidase (MANBA) expression and disease severity. *Sci Transl Med* 2021;**13**:eaaz1458. https://doi.org/10.1126/scitranslmed.aaz1458.
35. Qiu C, Huang S, Park J. *et al.* Renal compartment-specific genetic variation analyses identify new pathways in chronic kidney disease. *Nat Med* 2018;**24**:1721–31. https://doi.org/10.1038/s41591-018-0194-4.
36. Guo C, Dong G, Liang X. *et al.* Epigenetic regulation in AKI and kidney repair: mechanisms and therapeutic implications. *Nat Rev Nephrol* 2019;**15**:220–39. https://doi.org/10.1038/s41581-018-0103-6.
37. Wang KC, Chang HY. Epigenomics: technologies and applications. *Circ Res* 2018;**122**:1191–9. https://doi.org/10.1161/CIRCRESAHA.118.310998.

38. Keating ST, Van Diepen JA, Riksen NP. *et al.* Epigenetics in diabetic nephropathy, immunity and metabolism. *Diabetologia* 2018;**61**:6–20. https://doi.org/10.1007/s00125-017-4490-1.
39. Poeta L, Drongitis D, Verrillo L. *et al.* DNA Hypermethylation and unstable repeat diseases: a paradigm of transcriptional silencing to decipher the basis of pathogenic mechanisms. *Genes (Basel)* 2020;**11**:684. https://doi.org/10.3390/genes11060684.
40. Singh K, Rustagi Y, Abouhashem AS. *et al.* Genome-wide DNA hypermethylation opposes healing in patients with chronic wounds by impairing epithelial-mesenchymal transition. *J Clin Invest* 2022;**132**:e157279. https://doi.org/10.1172/JCI157279.
41. Torres-Perez JV, Irfan J, Febrianto MR. *et al.* Histone post-translational modifications as potential therapeutic targets for pain management. *Trends Pharmacol Sci* 2021;**42**:897–911. https://doi.org/10.1016/j.tips.2021.08.002.
42. Millán-Zambrano G, Burton A, Bannister AJ. *et al.* Histone post-translational modifications - cause and consequence of genome function. *Nat Rev Genet* 2022;**23**:563–80. https://doi.org/10.1038/s41576-022-00468-7.
43. Herman AB, Tsitsipatis D, Gorospe M. Integrated lncRNA function upon genomic and epigenomic regulation. *Mol Cell* 2022;**82**:2252–66. https://doi.org/10.1016/j.molcel.2022.05.027.
44. Guo Y, Zhao S, Wang GG. Polycomb gene silencing mechanisms: PRC2 chromatin targeting, H3K27me3 'Readout', and phase separation-based compaction. *Trends Genet* 2021;**37**:547–65. https://doi.org/10.1016/j.tig.2020.12.006.
45. Loaeza-Loaeza J, Beltran AS, Hernández-Sotelo D. DNMTs and impact of CpG content, transcription factors, consensus motifs, lncRNAs, and histone marks on DNA methylation. *Genes* 2020;**11**:1336. https://doi.org/10.3390/genes11111336.
46. Xiao X, Wang W, Guo C, Wu J., Zhang S., Shi H., Kwon S., Chen J., Dong Z. Hypermethylation leads to the loss of HOXA5, resulting in JAG1 expression and NOTCH signaling contributing to kidney fibrosis. *Kidney Int* 2024; **106**(24)98–114, https://doi.org/10.1016/j.kint.2024.02.023
47. Smith J, Sen S, Weeks RJ. *et al.* Promoter DNA hypermethylation and paradoxical gene activation. *Trends in Cancer* 2020;**6**:392–406. https://doi.org/10.1016/j.trecan.2020.02.007.
48. Linehan WM, Ricketts CJ. The Cancer Genome Atlas of renal cell carcinoma: findings and clinical implications. *Nat Rev Urol* 2019;**16**:539–52. https://doi.org/10.1038/s41585-019-0211-5.
49. Chu AY, Tin A, Schlosser P. *et al.* Epigenome-wide association studies identify DNA methylation associated with kidney function. *Nat Commun* 2017;**8**:1286. https://doi.org/10.1038/s41467-017-01297-7.
50. Yan Y, Liu H, Abedini A. *et al.* Unraveling the epigenetic code: human kidney DNA methylation and chromatin dynamics in renal disease development. *Nat Commun* 2024;**15**:873. https://doi.org/10.1038/s41467-024-45295-y.
51. Ling C, Rönn T. Epigenetics in human obesity and type 2 diabetes. *Cell Metab* 2019;**29**:1028–44. https://doi.org/10.1016/j.cmet.2019.03.009.
52. Kato M, Natarajan R. Epigenetics and epigenomics in diabetic kidney disease and metabolic memory. *Nat Rev Nephrol* 2019;**15**:327–45. https://doi.org/10.1038/s41581-019-0135-6.
53. Al-Dabet MM, Shahzad K, Elwakiel A. *et al.* Reversal of the renal hyperglycemic memory in diabetic kidney disease by targeting sustained tubular p21 expression. *Nat Commun* 2022;**13**:5062. https://doi.org/10.1038/s41467-022-32477-9.
54. Rinschen MM, Saez-Rodriguez J. The tissue proteome in the multi-omic landscape of kidney disease. *Nat Rev Nephrol* 2021;**17**:205–19. https://doi.org/10.1038/s41581-020-00348-5.
55. Dubin RF, Rhee EP. Proteomics and metabolomics in kidney disease, including insights into etiology, treatment, and prevention. *Clin J Am Soc Nephrol* 2020;**15**:404–11. https://doi.org/10.2215/CJN.07420619.
56. Qiu S, Cai Y, Yao H. *et al.* Small molecule metabolites: discovery of biomarkers and therapeutic targets. *Signal Transduct Target Ther* 2023;**8**:132. https://doi.org/10.1038/s41392-023-01399-3.
57. Hoyer KJR, Dittrich S, Bartram MP. *et al.* Quantification of molecular heterogeneity in kidney tissue by targeted proteomics. *J Proteomics* 2019;**193**:85–92. https://doi.org/10.1016/j.jprot.2018.03.001.
58. van de Logt A-E, Fresquet M, Wetzels JF. *et al.* The anti-PLA2R antibody in membranous nephropathy: what we know and what remains a decade after its discovery. *Kidney Int* 2019;**96**:1292–302. https://doi.org/10.1016/j.kint.2019.07.014.
59. Yu X, Cai J, Jiao X. *et al.* Response predictors to calcineurin inhibitors in patients with primary membranous nephropathy. *Am J Nephrol* 2018;**47**:266–74. https://doi.org/10.1159/000488728.
60. Kirita Y, Wu H, Uchimura K. *et al.* Cell profiling of mouse acute kidney injury reveals conserved cellular responses to injury. *Proc Natl Acad Sci U S A* 2020;**117**:15874–83. https://doi.org/10.1073/pnas.2005477117.
61. van Vliet S, Dal Co A, Winkler AR. *et al.* Spatially correlated gene expression in bacterial groups: the role of lineage history, spatial gradients, and cell-cell interactions. *Cell Syst* 2018;**6**:496–507.e6. https://doi.org/10.1016/j.cels.2018.03.009.
62. Li X, Wang C-Y. From bulk, single-cell to spatial RNA sequencing. *Int J Oral Sci* 2021;**13**:36. https://doi.org/10.1038/s41368-021-00146-0.
63. Wang G, Heijs B, Kostidis S. *et al.* Spatial dynamic metabolomics identifies metabolic cell fate trajectories in human kidney differentiation. *Cell Stem Cell* 2022;**29**:1580–1593.e7. https://doi.org/10.1016/j.stem.2022.10.008.
64. Park J, Shrestha R, Qiu C. *et al.* Single-cell transcriptomics of the mouse kidney reveals potential cellular targets of kidney disease. *Science* 2018;**360**:758–63. https://doi.org/10.1126/science.aar2131.
65. Fu J, Sun Z, Wang X. *et al.* The single-cell landscape of kidney immune cells reveals transcriptional heterogeneity in early diabetic kidney disease. *Kidney Int* 2022;**102**:1291–304. https://doi.org/10.1016/j.kint.2022.08.026.
66. Bressan D, Battistoni G, Hannon GJ. The dawn of spatial omics. *Science* 2023;**381**:eabq4964. https://doi.org/10.1126/science.abq4964.
67. Burgess DJ. Spatial transcriptomics coming of age. *Nat Rev Genet* 2019;**20**:317. https://doi.org/10.1038/s41576-019-0129-z.
68. Lake BB, Menon R, Winfree S. *et al.* An atlas of healthy and injured cell states and niches in the human kidney. *Nature* 2023;**619**:585–94. https://doi.org/10.1038/s41586-023-05769-3.
69. Bouwman BAM, Crosetto N, Bienko M. The era of 3D and spatial genomics. *Trends Genet* 2022;**38**:1062–75. https://doi.org/10.1016/j.tig.2022.05.010.
70. Mund A, Brunner A-D, Mann M. Unbiased spatial proteomics with single-cell resolution in tissues. *Mol Cell* 2022;**82**:2335–49. https://doi.org/10.1016/j.molcel.2022.05.022.
71. Deng Y, Bartosovic M, Ma S. *et al.* Spatial profiling of chromatin accessibility in mouse and human tissues. *Nature* 2022;**609**:375–83. https://doi.org/10.1038/s41586-022-05094-1.
72. Deng Y, Bartosovic M, Kukanja P. *et al.* Spatial-CUT&Tag: spatially resolved chromatin modification profiling at the cellular level. *Science* 2022;**375**:681–6. https://doi.org/10.1126/science.abg7216.

73. Liu Y, Yang M, Deng Y. *et al.* High-spatial-resolution multi-omics sequencing via deterministic barcoding in tissue. *Cell* 2020;**183**:1665–1681.e18. https://doi.org/10.1016/j.cell.2020.10.026.
74. JanP G, Van Buggenum JAG, Tanis SEJ. *et al.* Combined quantification of intracellular (phospho-)proteins and transcriptomics from fixed single cells. *Sci Rep* 2019;**9**:1469. https://doi.org/10.1038/s41598-018-37977-7.
75. Mimitou EP, Lareau CA, Chen KY. *et al.* Scalable, multimodal profiling of chromatin accessibility, gene expression and protein levels in single cells. *Nat Biotechnol* 2021;**39**:1246–58. https://doi.org/10.1038/s41587-021-00927-2.
76. Swanson E, Lord C, Reading J. *et al.* Simultaneous trimodal single-cell measurement of transcripts, epitopes, and chromatin accessibility using TEA-seq. *Elife* 2021;**10**:e63632. https://doi.org/10.7554/eLife.63632.
77. Zhang B, Srivastava A, Mimitou E. *et al.* Characterizing cellular heterogeneity in chromatin state with scCUT&Tag-pro. *Nat Biotechnol* 2022;**40**:1220–30. https://doi.org/10.1038/s41587-022-01250-0.
78. Su J-H, Zheng P, Kinrot SS. *et al.* Genome-scale imaging of the 3D organization and transcriptional activity of chromatin. *Cell* 2020;**182**:1641–1659.e26. https://doi.org/10.1016/j.cell.2020.07.032.
79. Takei Y, Yun J, Zheng S. *et al.* Integrated spatial genomics reveals global architecture of single nuclei. *Nature* 2021;**590**:344–50. https://doi.org/10.1038/s41586-020-03126-2.
80. Deo RC. Machine learning in medicine. *Circulation* 2015;**132**:1920–30. https://doi.org/10.1161/CIRCULATIONAHA.115.001593.
81. Greener JG, Kandathil SM, Moffat L. *et al.* A guide to machine learning for biologists. *Nat Rev Mol Cell Biol* 2022;**23**:40–55. https://doi.org/10.1038/s41580-021-00407-0.
82. Bagherian M, Sabeti E, Wang K. *et al.* Machine learning approaches and databases for prediction of drug-target interaction: a survey paper. *Brief Bioinform* 2021;**22**:247–69. https://doi.org/10.1093/bib/bbz157.
83. Zhao W. *et al.* KDClassifier: urinary proteomic spectra analysis based on machine learning for classification of kidney diseases. *medRxiv* 2020, 2020–12.
84. Niel O, Bastard P. Artificial intelligence in nephrology: core concepts, clinical applications, and perspectives. *Am J Kidney Dis* 2019;**74**:803–10. https://doi.org/10.1053/j.ajkd.2019.05.020.
85. Gulshan V, Peng L, Coram M. *et al.* Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* 2016;**316**:2402–10. https://doi.org/10.1001/jama.2016.17216.
86. Commissioner O of the. *FDA Permits Marketing of Artificial Intelligence-Based Device to Detect Certain Diabetes-Related Eye Problems*. U.S. Food and Drug Administration, Silver Spring, MD, 2020.
87. Sasse A, Ng B, Spiro AE. *et al.* Benchmarking of deep neural networks for predicting personal gene expression from DNA sequence highlights shortcomings. *Nat Genet* 2023;**55**:2060–4. https://doi.org/10.1038/s41588-023-01524-6.
88. Buchan DWA, Jones DT. The PSIPRED protein analysis workbench: 20 years on. *Nucleic Acids Res* 2019;**47**:W402–7. https://doi.org/10.1093/nar/gkz297.
89. Zhong Y, He J, Duan S. *et al.* Revealing the mechanism of novel nitrogen-doped biochar supported magnetite (NBM) enhancing anaerobic digestion of waste-activated sludge by sludge characteristics. *J Environ Manage* 2023;**340**:117982. https://doi.org/10.1016/j.jenvman.2023.117982.
90. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY, USA: Association for Computing Machinery, 2016, 785–94.
91. Sealfon RSG, Mariani LH, Kretzler M. *et al.* Machine learning, the kidney, and genotype-phenotype analysis. *Kidney Int* 2020;**97**:1141–9. https://doi.org/10.1016/j.kint.2020.02.028.
92. Pedregosa F, Varoquaux G, Gramfort A. *et al.* Scikit-learn: machine learning in Python. *J Mach Learn Res* 2011;**12**:2825–30.
93. Kuhn M. Building predictive models in R using the caret package. *J Stat Soft* 2008;**28**:28. https://doi.org/10.18637/jss.v028.i05.
94. Blaom A, Kiraly F, Lienart T. *et al.* MLJ: a Julia package for composable machine learning. *JOSS* 2020;**5**:2704. https://doi.org/10.21105/joss.02704.
95. Ching T, Himmelstein DS, Beaulieu-Jones BK. *et al.* Opportunities and obstacles for deep learning in biology and medicine. *J R Soc Interface* 2018;**15**:20170387. https://doi.org/10.1098/rsif.2017.0387.
96. Yang R, Yu Y. Artificial convolutional neural network in object detection and semantic segmentation for medical imaging analysis. *Front Oncol* 2021;**11**:638182. https://doi.org/10.3389/fonc.2021.638182.
97. Senior AW, Evans R, Jumper J. *et al.* Improved protein structure prediction using potentials from deep learning. *Nature* 2020;**577**:706–10. https://doi.org/10.1038/s41586-019-1923-7.
98. Tegunov D, Cramer P. Real-time cryo-electron microscopy data preprocessing with Warp. *Nat Methods* 2019;**16**:1146–52. https://doi.org/10.1038/s41592-019-0580-y.
99. Smith AM, Walsh JR, Long J. *et al.* Standard machine learning approaches outperform deep representation learning on phenotype prediction from transcriptomics data. *BMC Bioinformatics* 2020;**21**:119. https://doi.org/10.1186/s12859-020-3427-8.
100. Chaitanya K, Karani N, Baumgartner CF. *et al.* Semi-supervised task-driven data augmentation for medical image segmentation. *Med Image Anal* 2021;**68**:101934. https://doi.org/10.1016/j.media.2020.101934.
101. Tang B, Chen X, Wang S. *et al.* Generalized heterophily graph data augmentation for node classification. *Neural Netw* 2023;**168**:339–49. https://doi.org/10.1016/j.neunet.2023.09.021.
102. Sandfort V, Yan K, Pickhardt PJ. *et al.* Data augmentation using generative adversarial networks (CycleGAN) to improve generalizability in CT segmentation tasks. *Sci Rep* 2019;**9**:16884. https://doi.org/10.1038/s41598-019-52737-x.
103. Qin T, Wang Z, He K. *et al.* Automatic data augmentation via deep reinforcement learning for effective kidney tumor segmentation. In: *ICASSP 2020 – 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. New York, NY: IEEE, 2020, 1419–23.
104. Chen Y, Ruan D, Xiao J. *et al.* Fully automated multiorgan segmentation in abdominal magnetic resonance imaging with deep neural networks. *Med Phys* 2020;**47**:4971–82. https://doi.org/10.1002/mp.14429.
105. Yin S, Peng Q, Li H. *et al.* Automatic kidney segmentation in ultrasound images using subsequent boundary distance regression and pixelwise classification networks. *Med Image Anal* 2020;**60**:101602. https://doi.org/10.1016/j.media.2019.101602.
106. Garcea F, Serra A, Lamberti F. *et al.* Data augmentation for medical imaging: a systematic literature review. *Comput Biol Med* 2023;**152**:106391. https://doi.org/10.1016/j.compbiomed.2022.106391.
107. Khwaja A. KDIGO clinical practice guidelines for acute kidney injury. *Nephron Clin Pract* 2012;**120**:c179–84. https://doi.org/10.1159/000339789.

108. Tomašev N, Glorot X, Rae JW. *et al.* A clinically applicable approach to continuous prediction of future acute kidney injury. *Nature* 2019;**572**:116–9. https://doi.org/10.1038/s41586-019-1390-1.
109. Koyner JL, Carey KA, Edelson DP. *et al.* The development of a machine learning inpatient acute kidney injury prediction model. *Crit Care Med* 2018;**46**:1070–7. https://doi.org/10.1097/CCM.0000000000003123.
110. Rajkomar A, Oren E, Chen K. *et al.* Scalable and accurate deep learning with electronic health records. *NPJ Digit Med* 2018;**1**:18. https://doi.org/10.1038/s41746-018-0029-1.
111. Song X, Yu ASL, Kellum JA. *et al.* Cross-site transportability of an explainable artificial intelligence model for acute kidney injury prediction. *Nat Commun* 2020;**11**:5668. https://doi.org/10.1038/s41467-020-19551-w.
112. Churpek MM, Carey KA, Edelson DP. *et al.* Internal and external validation of a machine learning risk score for acute kidney injury. *JAMA Netw Open* 2020;**3**:e2012892. https://doi.org/10.1001/jamanetworkopen.2020.12892.
113. Schena FP, Anelli VW, Trotta J. *et al.* Development and testing of an artificial intelligence tool for predicting end-stage kidney disease in patients with immunoglobulin A nephropathy. *Kidney Int* 2021;**99**:1179–88. https://doi.org/10.1016/j.kint.2020.07.046.
114. Vistisen D, Andersen GS, Hulman A. *et al.* A validated prediction model for end-stage kidney disease in type 1 diabetes. *Diabetes Care* 2021;**44**:901–7. https://doi.org/10.2337/dc20-2586.
115. Østergaard HB, van der Leeuw J, Visseren FLJ. *et al.* Comment on Vistisen et al. a validated prediction model for end-stage kidney disease in type 1 diabetes. *Diabetes Care* 2021;**44**:901–7. https://doi.org/10.2337/dc21-0364.
116. Subramanian A, Narayan R, Corsello SM. *et al.* A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. *Cell* 2017;**171**:1437–1452.e17. https://doi.org/10.1016/j.cell.2017.10.049.
117. Harder JL, Menon R, Otto EA. *et al.* Organoid single cell profiling identifies a transcriptional signature of glomerular disease. *JCI Insight* 2019;**4**:122697. https://doi.org/10.1172/jci.insight.122697.
118. Menon R, Otto EA, Hoover P. *et al.* Single cell transcriptomics identifies focal segmental glomerulosclerosis remission endothelial biomarker. *JCI Insight* 2020;**5**:133267. https://doi.org/10.1172/jci.insight.133267.
119. Abdeltawab H, Shehata M, Shalaby A. *et al.* A novel CNN-based CAD system for early assessment of transplanted kidney dysfunction. *Sci Rep* 2019;**9**:5948. https://doi.org/10.1038/s41598-019-42431-3.
120. Aubert O, Higgins S, Bouatou Y. *et al.* Archetype analysis identifies distinct profiles in renal transplant recipients with transplant glomerulopathy associated with allograft survival. *JASN* 2019;**30**:625–39. https://doi.org/10.1681/ASN.2018070777.
121. Qian W, Xiaoyi W, Zi Y. Screening and bioinformatics analysis of IgA nephropathy gene based on GEO databases. *Biomed Res Int* 2019;**2019**:1–7. https://doi.org/10.1155/2019/8794013.
122. Luan S, Li P, Yi T. Series test of cluster and network analysis for lupus nephritis, before and after IFN-K-immunosuppressive therapy. *Nephrol Ther* 2018;**23**:997–1006. https://doi.org/10.1111/nep.13159.
123. Zhao X, Zhang L, Wang J. *et al.* Identification of key biomarkers and immune infiltration in systemic lupus erythematosus by integrated bioinformatics analysis. *J Transl Med* 2021;**19**:35. https://doi.org/10.1186/s12967-020-02698-x.
124. Chen X, Cao Y, Wang Z. *et al.* Bioinformatic analysis reveals novel hub genes and pathways associated with hypertensive nephropathy. *Nephrol Ther* 2019;**24**:1103–14. https://doi.org/10.1111/nep.13508.
125. Wu I-W, Tsai T-H, Lo C-J. *et al.* Discovering a trans-omics biomarker signature that predisposes high risk diabetic patients to diabetic kidney disease. *NPJ Digit Med* 2022;**5**:166. https://doi.org/10.1038/s41746-022-00713-7.
126. Levey AS, Grams ME, Inker LA. Uses of GFR and albuminuria level in acute and chronic kidney disease. *New Engl J Med* 2022;**386**:2120–8. https://doi.org/10.1056/NEJMra2201153.
127. Eddy S, Mariani LH, Kretzler M. Integrated multi-omics approaches to improve classification of chronic kidney disease. *Nat Rev Nephrol* 2020;**16**:657–68. https://doi.org/10.1038/s41581-020-0286-5.
128. Ma J, Karnovsky A, Afshinnia F. *et al.* Differential network enrichment analysis reveals novel lipid pathways in chronic kidney disease. *Bioinformatics* 2019;**35**:3441–52. https://doi.org/10.1093/bioinformatics/btz114.
129. Azukaitis K, Ju W, Kirchner M. *et al.* Low levels of urinary epidermal growth factor predict chronic kidney disease progression in children. *Kidney Int* 2019;**96**:214–21. https://doi.org/10.1016/j.kint.2019.01.035.
130. Satirapoj B, Pooluea P, Nata N. *et al.* Urinary biomarkers of tubular injury to predict renal progression and end stage renal disease in type 2 diabetes mellitus with advanced nephropathy: a prospective cohort study. *J Diabetes Complications* 2019;**33**:675–81. https://doi.org/10.1016/j.jdiacomp.2019.05.013.
131. Hsu C-Y, Xie D, Waikar SS. *et al.* Urine biomarkers of tubular injury do not improve on the clinical model predicting chronic kidney disease progression. *Kidney Int* 2017;**91**:196–203. https://doi.org/10.1016/j.kint.2016.09.003.
132. Yuan Q, Xie Y, Peng Z. *et al.* Urinary magnesium predicts risk of cardiovascular disease in chronic kidney disease stage 1-4 patients. *Clin Nutr* 2021;**40**:2394–400. https://doi.org/10.1016/j.clnu.2020.10.036.
133. Inrig JK, Califf RM, Tasneem A. *et al.* The landscape of clinical trials in nephrology: a systematic review of Clinicaltrials.gov. *Am J Kidney Dis* 2014;**63**:771–80. https://doi.org/10.1053/j.ajkd.2013.10.043.
134. Liu Y, Liu X, Jia J. *et al.* Comprehensive analysis of aberrantly expressed profiles of mRNA and its relationship with serum galactose-deficient IgA1 level in IgA nephropathy. *J Transl Med* 2019;**17**:320. https://doi.org/10.1186/s12967-019-2064-3.
135. Deng Y, Zheng Y, Li D. *et al.* Expression characteristics of interferon-stimulated genes and possible regulatory mechanisms in lupus patients using transcriptomics analyses. *EBioMedicine* 2021;**70**:103477. https://doi.org/10.1016/j.ebiom.2021.103477.
136. Eales JM, Jiang X, Xu X. *et al.* Uncovering genetic mechanisms of hypertension through multi-omic analysis of the kidney. *Nat Genet* 2021;**53**:630–7. https://doi.org/10.1038/s41588-021-00835-w.
137. Kumar N, Gupta R, Gupta S. Whole slide imaging (WSI) in pathology: current perspectives and future directions. *J Digit Imaging* 2020;**33**:1034–40. https://doi.org/10.1007/s10278-020-00351-z.
138. Becker JU, Mayerich D, Padmanabhan M. *et al.* Artificial intelligence and machine learning in nephropathology. *Kidney Int* 2020;**98**:65–75. https://doi.org/10.1016/j.kint.2020.02.027.
139. Araújo ALD, Arboleda LPA, Palmier NR. *et al.* The performance of digital microscopy for primary diagnosis in human

pathology: a systematic review. *Virchows Arch* 2019;**474**:269–87. https://doi.org/10.1007/s00428-018-02519-z.
140. Dietz RL, Hartman DJ, Pantanowitz L. Systematic review of the use of Telepathology during intraoperative consultation. *Am J Clin Pathol* 2020;**153**:198–209. https://doi.org/10.1093/ajcp/aqz155.
141. Marsh JN, Matlock MK, Kudose S. *et al.* Deep learning global Glomerulosclerosis in transplant kidney frozen sections. *IEEE Trans Med Imaging* 2018;**37**:2718–28. https://doi.org/10.1109/TMI.2018.2851150.
142. Bukowy JD, Dayton A, Cloutier D. *et al.* Region-based convolutional neural nets for localization of glomeruli in trichrome-stained whole kidney sections. *J Am Soc Nephrol* 2018;**29**:2081–8. https://doi.org/10.1681/ASN.2017111210.
143. Zeng C, Nan Y, Xu F. *et al.* Identification of glomerular lesions and intrinsic glomerular cell types in kidney diseases via deep learning. *J Pathol* 2020;**252**:53–64. https://doi.org/10.1002/path.5491.
144. Barbour SJ, Coppo R, Zhang H. *et al.* Evaluating a new international risk-prediction tool in IgA nephropathy. *JAMA Intern Med* 2019;**179**:942–52. https://doi.org/10.1001/jamainternmed.2019.0600.
145. Song X, Liu X, Liu F. *et al.* Comparison of machine learning and logistic regression models in predicting acute kidney injury: a systematic review and meta-analysis. *Int J Med Inform* 2021;**151**:104484. https://doi.org/10.1016/j.ijmedinf.2021.104484.
146. Sapoval N, Aghazadeh A, Nute MG. *et al.* Current progress and open challenges for applying deep learning across the biosciences. *Nat Commun* 2022;**13**:1728. https://doi.org/10.1038/s41467-022-29268-7.
147. Choi J, Pacheco CM, Mosbergen R. *et al.* Stemformatics: visualize and download curated stem cell data. *Nucleic Acids Res* 2019;**47**:D841–6. https://doi.org/10.1093/nar/gky1064.
148. Saez-Rodriguez J, Rinschen MM, Floege J. *et al.* Big science and big data in nephrology. *Kidney Int* 2019;**95**:1326–37. https://doi.org/10.1016/j.kint.2018.11.048.
149. Zhang M, Ye Z, Yuan E. *et al.* Imaging-based deep learning in kidney diseases: recent progress and future prospects. *Insights Imaging* 2024;**15**:1–13. https://doi.org/10.1186/s13244-024-01636-5.
150. Haghverdi L, Lun ATL, Morgan MD. *et al.* Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nat Biotechnol* 2018;**36**:421–7. https://doi.org/10.1038/nbt.4091.
151. Austin PC, White IR, Lee DS. *et al.* Missing data in clinical research: a tutorial on multiple imputation. *Can J Cardiol* 2021;**37**:1322–31. https://doi.org/10.1016/j.cjca.2020.11.010.
152. Gadegbeku CA, Gipson DS, Holzman LB. *et al.* Design of the Nephrotic Syndrome Study Network (NEPTUNE) to evaluate primary glomerular nephropathy by a multidisciplinary approach. *Kidney Int* 2013;**83**:749–56. https://doi.org/10.1038/ki.2012.428.
153. Townsend RR, Guarnieri P, Argyropoulos C. *et al.* Rationale and design of the Transformative Research in Diabetic Nephropathy (TRIDENT) study. *Kidney Int* 2020;**97**:10–3. https://doi.org/10.1016/j.kint.2019.09.020.
154. Mariani LH, Bomback AS, Canetta PA. *et al.* CureGN study rationale, design, and methods: establishing a large prospective observational study of glomerular disease. *Am J Kidney Dis* 2019;**73**:218–29. https://doi.org/10.1053/j.ajkd.2018.07.020.
155. Tuttle KR, Bebiak J, Brown K. *et al.* Patient perspectives and involvement in precision medicine research. *Kidney Int* 2021;**99**:511–4. https://doi.org/10.1016/j.kint.2020.10.036.
156. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. *Future Healthc J* 2019;**6**:94–8. https://doi.org/10.7861/futurehosp.6-2-94.
157. Chen T, Tyagi S. Integrative computational epigenomics to build data-driven gene regulation hypotheses. *Gigascience* 2020;**9**:giaa064. https://doi.org/10.1093/gigascience/giaa064.
158. Pirmoradi S, Teshnehlab M, Zarghami N. *et al.* A self-organizing deep neuro-fuzzy system approach for classification of kidney cancer subtypes using miRNA genomics data. *Comput Methods Programs Biomed* 2021;**206**:106132. https://doi.org/10.1016/j.cmpb.2021.106132.
159. Imoto Y, Nakamura T, Escolar EG. *et al.* Resolution of the curse of dimensionality in single-cell RNA sequencing data analysis. *Life Sci Alliance* 2022;**5**:e202201591. https://doi.org/10.26508/lsa.202201591.
160. Chattopadhyay A, Lu T-P. Gene-gene interaction: the curse of dimensionality. *Ann Transl Med* 2019;**7**:813. https://doi.org/10.21037/atm.2019.12.87.
161. Sheikholeslami S. *Ablation Programming for Machine Learning*. New York, NY: Springer, 2019.
162. He J, Baxter SL, Xu J. *et al.* The practical implementation of artificial intelligence technologies in medicine. *Nat Med* 2019;**25**:30–6. https://doi.org/10.1038/s41591-018-0307-0.
163. Kairouz P, McMahan HB, Avent B. *et al.* Advances and open problems in federated learning. *MAL* 2021;**14**:1–210. https://doi.org/10.1561/9781680837896.
164. Chen Y, Lu W, Qin X. *et al.* MetaFed: federated learning among federations with cyclic knowledge distillation for personalized healthcare. *IEEE Trans Neural Netw Learn Syst* 2023;**34**:1–12. https://doi.org/10.1109/TNNLS.2023.3297103.
165. Aziz MMA, Sadat MN, Alhadidi D. *et al.* Privacy-preserving techniques of genomic data-a survey. *Brief Bioinform* 2019;**20**:887–95. https://doi.org/10.1093/bib/bbx139.
166. Guinney J, Saez-Rodriguez J. Alternative models for sharing confidential biomedical data. *Nat Biotechnol* 2018;**36**:391–2. https://doi.org/10.1038/nbt.4128.