

Multi-Omics Integration in Nephrology: Advances, Challenges, and Future Directions

Afaf Saliba, PhD^{*}, Yuheng Du, MS[†], Tianqing Feng, BS^{*} and Lana Garmire, PhD[†]

Summary

Omics technologies have transformed nephrology, providing deep insights into molecular mechanisms of kidney disease and enabling more precise diagnostic tools, therapeutic strategies, and prognostic markers. Multi-omics data integration, spanning bulk, single-cell, and spatial omics, offers a comprehensive view of kidney biology in health and disease. In this review, we explore methods and challenges for integrating transcriptomic, epigenomic, and spatial data. By combining omics layers, researchers can uncover novel molecular interactions and spatial tissue organization, advancing our understanding of diseases like diabetic kidney disease and autosomal polycystic kidney disease. This integrated approach is reshaping diagnostic and therapeutic strategies in nephrology and is critical for optimizing insights available from spatial and multi-omics analysis.

Semin Nephrol 36:x-xx © 20xx Elsevier Inc. All rights reserved.

Semin Nephrol 000:151584 © 2025 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

Keywords: Biomarker discovery, computational methods, data challenges, kidney disease, multi-omics integration

INTRODUCTION

Recent advances in large-scale data generation and computational analysis have significantly impacted biomedical research and clinical medicine. While fields like cancer research have led precision medicine in these “big science” initiatives, other fields have lagged behind. However, recent advances hold exciting potential for nephrology to take the lead.¹ In the era of big data, the advent of high-throughput omics platforms has markedly increased the volume of molecular data, encompassing the genome, epigenome, transcriptome, proteome, and metabolome, allowing a comprehensive understanding of human health and disease development through multimodal representation.² Adopting an integrative approach to analyze multiple datasets is essential to address biological problems holistically.

Multi-omics integrative analysis can be applied to different data resolutions for various downstream tasks. For bulk omics, such integration typically aids in patient

classification, disease subtyping, and the derivation of biological insights.³ Single-cell data integration offers a finer resolution, enabling the identification of cell subpopulations, marker genes, and comparative analyses that account for tissue heterogeneity.⁴ Furthermore, spatial data integrates the spatial context on top of single-cell (or spot-level) data, allowing the identification of spatial functional domains, capturing cell–cell interplay, and developing deeper insights into cellular and tissue constitution.⁵ Accompanying such complex data, multi-omics integration is fraught with challenges ranging from data heterogeneity and dimensionality to interpretability and validation.⁶ This review summarizes key recent advancements in omics data integration applied in nephrology while highlighting the major challenges, possible solutions, and future perspectives.

OVERVIEW OF MULTI-OMICS TECHNOLOGIES IN NEPHROLOGY

Genomics, the study of an organism’s set of genes, has been instrumental in identifying genetic variants linked to kidney diseases through techniques like whole genome sequencing (WGS) and genome-wide association studies (GWAS). The WGS method identifies DNA variations beyond the exons, particularly in introns and noncoding regions, which can cause genetic disorders. These variations are missed by whole exome sequencing, which examines exons. Additionally, WGS detects complex trait variants, such as copy number variants and those in genes located within intronic and other noncoding areas of the genome. The readers interested in more details are referred to the reviews of Zoccali et al. and Devarajan et al.^{7,8} GWAS identifies single nucleotide polymorphisms among participants and is used to examine the DNA of large groups of individuals in order to

^{*}Center for Precision Medicine, Department of Medicine, University of Texas Health Science Center at San Antonio

[†]Department of Computational Medicine and Bioinformatics, University of Michigan Medical School

Financial disclosure and conflict of interest statements: AS is supported by NIH/NHLBI, T32 HL007446. YD is supported by training funding provided by the Advanced Proteogenomics of Cancer (T32 CA140044). LXG is supported by NIH/NLM R01 LM012373. The authors claim no conflict of interest to the work.

Address reprint requests to: Lana Garmire, PhD, 1600 Huron Parkway, University of Michigan, Ann Arbor, Michigan, USA 48105. E-mail: lgarmire@umich.edu

0270-9295/ - see front matter

© 2025 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

<https://doi.org/10.1016/j.semnephrol.2025.151584>

¹Contributed equally.

detect genetic variations linked to specific diseases. GWAS uncovered specific genetic variants linked to chronic kidney disease (CKD), such as those in the shroom family member 3 (*SHROOM3*) gene, which have been shown to contribute to kidney fibrosis.^{9,10} A study by Li et al. combined GWAS with epigenomic and transcriptomic resources and identified 41 genome-wide significant loci related to estimated glomerular filtration rate in individuals of African ancestry and those with ancestry in the Americas. This approach not only revealed two novel loci but also highlighted potential effector genes and key regulatory elements involved in renal function and disease, emphasizing the value of diverse, omics data for advancing understanding and improving clinical outcomes in CKD.¹¹ A limitation of GWAS is that single nucleotide polymorphisms may not directly explain changes in gene expression.

High-throughput gene expression measurements, including RNA sequencing (RNA-seq) and single-cell transcriptomics, have revolutionized nephrology by providing unprecedented insights into renal cell types and gene expression patterns. Kidney biopsy samples, fundamental to nephrology research, are a rich data source for deciphering disease etiologies and can now be used to reveal potential mechanisms with multi-omics data. These samples can be analyzed through bulk RNA-seq at various levels, including compartment specific and nephron segment specific, or via single-cell RNA-seq (scRNA-seq) to produce detailed transcriptomic profiles (reviewed in detail by Hill et al.¹²). RNA-seq has deepened our understanding of kidney fibrosis,¹³ identifying key cell populations such as myofibroblasts, which originate from pericytes and fibroblasts, contributing to extracellular matrix deposition in CKD. Moreover, scRNA-seq has further clarified distinct fibroblast subtypes and injury-associated proximal tubule cells, underscoring their critical roles in fibrosis progression and revealing new therapeutic targets such as targeting profibrotic signaling pathways and ferroptosis.¹⁴ Additionally, single-cell transcriptomics has greatly expanded our knowledge of kidney cellular diversity, including the discovery of unique podocyte subpopulations in focal segmental glomerulosclerosis, which enhances our understanding of disease mechanisms and supports personalized treatment approaches.¹⁵ A limitation of RNA-seq data is in RNA levels not always correlating with protein abundance.

Beyond transcriptomics, epigenetics and proteomics data can also enrich our comprehension of kidney biology. Epigenomics, which is the study of heritable changes in gene expression that do not involve alterations to the underlying DNA sequence, advanced our understanding of epigenetic modifications roles in kidney development, function, and disease. For example, increased DNA methylation in the Wnt/beta-catenin pathway was linked to age-related kidney dysfunction, and DNA methyltransferase 1-mediated methylation was

critical for nephron progenitor maturation, preventing developmental defects by silencing transposable elements in the kidney.¹⁶ Another study examined cytosine methylation and open chromatin states in kidney samples from 399 individuals, including both CKD patients and controls, and discovered that significant methylation changes associated with kidney disease were influenced by genetic variants linked to the condition. Additionally, the study identified regions with both methylation and chromatin alterations that affected gene expression, especially in metabolism-related genes. Methylation risk scores were found to enhance disease classification and prediction, indicating a potential causal relationship between epigenetic changes and kidney disease, suggesting new possibilities for risk stratification.¹⁷

Moreover, the use of advanced mass spectrometry techniques, often coupled with liquid chromatography, has allowed for a detailed analysis of the kidney proteome, leading to the identification of disease-specific proteins such as the M-type phospholipase A2 receptor in membranous nephropathy, which has profoundly influenced diagnostic and therapeutic approaches in kidney disease.¹⁸

Metabolomics, unlike the other omics mentioned earlier, uses mass spectrometry combined with liquid chromatography to profile metabolites. A key distinction is that metabolites reflect the functional output of enzymatic pathways, indicating which pathways are active or suppressed under specific conditions compared to control conditions. This technique has uncovered the significant role of gut-derived compounds like indoxyl sulfate and p-cresol sulfate in the progression of CKD, thereby providing crucial insights into the metabolic disturbances associated with renal dysfunction.¹⁹ Another urine metabolomics study revealed 13 metabolites correlating with mitochondrial metabolism in diabetic kidney disease (DKD), with further studies assessing their roles in predicting future kidney function decline and understanding kidney disease progression.^{20,21} Also, acute kidney injury (AKI) metabolomics has revealed that niacinamide, a precursor in the salvage pathway of nicotinamide adenine dinucleotide (an essential cofactor for energy metabolism and cellular health), plays a significant role in protecting kidney function.²²

While these methods have advanced our knowledge, they also have limitations, such as lack of single-cell resolution or spatial information. To address these challenges, advancements like spatial omics combining histologic imaging with spatial profiling and single-cell genomics can reveal the intricate tissue microenvironments and their functions.^{23,24} For instance, a recent study by Sharma et al.²⁴ has identified endogenous adenine as a novel biomarker for unraveling the biological mechanisms behind DKD progression, using a combination of urine metabolomics, single-cell transcriptomics, and spatial

metabolomics. Additionally, machine learning can analyze high-dimensional pathology data from digital biopsy images to reveal significant patterns.²⁵

Beyond individual omics, multi-omics studies are crucial for identifying biomarkers and understanding disease mechanisms in nephrology. As discussed in Eddy et al.,²⁶ the study of kidney diseases presents a unique opportunity to leverage multi-omics data integration, as the necessary biosamples (i.e., kidney biopsies), which are valuable resources for generating genetic, transcriptomic, proteomic, and metabolomic data, are often collected in clinical practice. While single-cell and spatial omics have advanced tissue analysis, they often capture only a single layer of complexity. Integrative multi-omics, which combines various data types, offers a more comprehensive view to enhance our understanding of diseases like DKD and AKI. This integration allows for the identification of noninvasive surrogate markers that reflect cellular processes occurring in the kidney, offering a less invasive method to monitor disease progression and treatment response.

CHALLENGES FOR MULTI-OMICS DATA INTEGRATION

Integrating multi-omics datasets poses significant challenges due to their complexity and the inherent variability in biological data (Fig. 1). In the following sections we provide an overview of these challenges.

Missing Data

Missing data are a common issue in omics studies, whether resulting from technical failures such as poor tissue quality, insufficient sample volume, or measurement system limitations or other factors such as budget restrictions or subject dropout. Missing values in large-scale 'omics data can significantly hinder downstream

analyses,²⁷ making the handling of missing data a regular challenge in multi-omics integration and analysis. The variability in the set of observations with missing data and the proportion of missingness across different omics datasets add to this complexity. Readers are referred to more details in the review of Flores et al.²⁸

Data Availability

In the context of nephrology research, a significant challenge is the scarcity of large, diverse datasets, especially in medical imaging-based deep learning.^{1, 29} Moreover, both the initial and continuous model training depend on ongoing data supplementation, validation, and improvement. Therefore, global, secure, and real-time updated resources for the research and clinical communities help improve the availability of larger datasets, which facilitates the development of highly parameterized models, thus offering significant transformative potential. Initiatives like the Kidney Precision Medicine Project,³⁰⁻³² Nephrotic Syndrome Study Network,³³ Transformative Research in Diabetic Nephropathy,³⁴ and Cure Glomerulonephropathy³⁵ aim to achieve comprehensive characterization of kidney biopsies across various kidney disease subtypes.

High Dimensionality and Data Heterogeneity

Omics datasets are typically high in dimensionality, with thousands of variables (such as gene expression levels) but relatively few samples. This “curse of dimensionality” can result in overfitting, where models excel on training data but fail on new, unseen data. Different omics layers (e.g., genomics, transcriptomics, proteomics) present data in various formats and scales. Gene expression datasets may have tens of thousands of features, while metabolomics datasets might contain only a few hundred. This heterogeneity further complicates data integration and analysis.

In nephrology, the presence of comorbidities often makes patient cohorts highly heterogeneous, complicating analysis. Therefore, data standardization and harmonization, along with the ability to integrate multimodal datasets, are critical.³⁶ Identifying biomarkers for kidney disease involves analyzing subtle patterns across different omics layers, which is complicated by the high dimensionality and variability of the data. Effectively reducing dimensionality and selecting the most influential features are essential for improving model performance and ensuring robust results.

COMPUTATIONAL METHODS FOR DATA INTEGRATION

Addressing the challenges associated with integrative multi-omics analyses requires a combination of careful

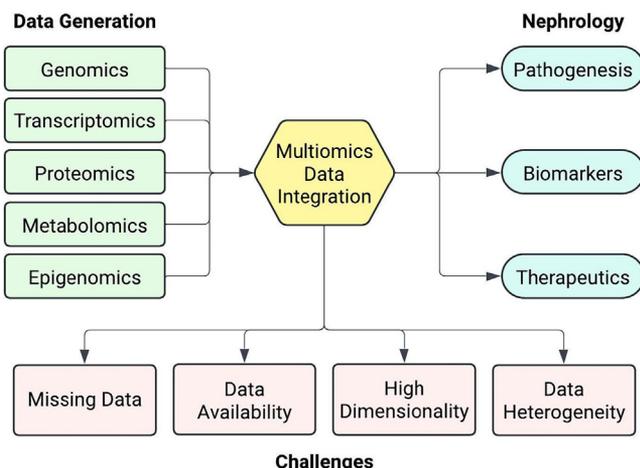


Figure 1. Multi-omics data integration in nephrology research.

data handling, advanced statistical methods, and thoughtful integration strategies to derive meaningful insights from multi-omics data. We outline the types of integration strategies structurally and algorithmically, as well as their application in recent nephrology research to advance clinical insights and potential therapeutic interventions.

Overview of Integration Strategies

From a structural point of view, most integration tasks for multi-omics measurements can be classified into three categories.³⁷ *Horizontal* integration combines data from different samples within the same modality, such as integrating gene expression data from various cohorts for a meta-analysis to increase the statistical power. The major challenge for this type of integration is proper normalization and batch correction.³⁸ *Vertical* integration uses common samples or cells as the anchor to integrate simultaneous profiling of multiple modalities. For example, it can integrate simultaneous profiling of scRNA-seq, which measures gene expression, and single-cell assay for transposase-accessible chromatin with sequencing (scATAC-seq), which assesses chromatin accessibility, using the 10x Genomics Multiome platform. Given the mixed embedding space of multimodal features, preserving biological interpretability is challenging in this type of integration. *Diagonal* integration in multi-omics refers to organizing multiple omics datasets into a block-diagonal matrix, where each omics dataset occupies a separate block along the diagonal, allowing for independent but aligned analysis of each omics layer within a unified framework. This diagonal approach presents the greatest challenge, as shared cells or omics features to serve as anchor points for integration are lacking. This approach typically depends on identifying strong correlations among low-dimensional manifolds across different data modalities.³⁹ From an algorithmic perspective, integration techniques are traditionally classified into three major levels: early, middle, and late integration, also referred to as concatenation-based, transformation-based, and model-based approaches, respectively.⁴⁰ Furthermore, Picard et al. expanded on this framework by introducing additional categories, specifically early, intermediate, mixed, late, and hierarchical integration.² We illustrate these techniques in [Figure 2](#).

Early integration directly concatenates multiple omics modalities, enabling simultaneous analysis while preserving interactions across different layers.⁴¹ However, this method is challenged by the inherent heterogeneity of input data, which can vary in distribution, noise levels, and missingness, potentially leading to biased results. Moreover, the high dimensionality relative to the sample size increases the risk of overfitting. *Mixed* integration

addresses some of these challenges by first extracting latent representations from each high-dimensional omics layer, before combining them for downstream analysis.^{2,42} This approach reduces dimensionality and noise. However, since each omics layer is transformed independently, there is a risk of losing valuable interplay and correlations between modalities. *Intermediate* integration differs from mixed integration in that it does not require prior concatenation or independent transformation of datasets.² Instead, it seeks to identify a common latent space that all omics data share, such as through nonnegative matrix factorization (NMF).⁴³ This approach assumes that multi-omics data converge in a shared latent representation, thereby considering interactions and correlations. However, it often relies on unsupervised methods, which makes it difficult to incorporate prior biological knowledge into the analysis.

Late integration is a straightforward approach involving the combination of outputs from individual analyses of each omics layer. For example, Chen et al.⁴⁴ used late integration to combine data from various sources, including epigenome-wide DNA methylation from blood cells, genetic variants, circulating proteins, and microRNAs (miRNAs). They identified specific DNA sites (CpG loci) associated with kidney failure risk in individuals with type 1 diabetes and DKD. By analyzing these sites along plasma proteins and miRNAs, they performed mediation analyses to determine how proteins and miRNAs affect the link between DNA methylation and kidney failure risk. While simple, late integration is limited by its inability to capture shared knowledge across omics layers, thereby undermining the holistic nature of multi-omics analysis.²

Hierarchical integration adopts a sequential process that incorporates prior biological knowledge, such as pathway and regulatory relationships among omics layers.⁴⁵ Although this method can be highly specific and informative for certain types of data, it often requires extensive external biological knowledge, making it less robust and versatile in handling diverse data types. Methods like Mergeomics⁴⁶ exemplify hierarchical integration. Mergeomics leverages multiple biological databases, including over 25 different pathways and network databases, to identify regulatory pathways using prior biological knowledge. In a study on cisplatin-induced AKI by Deng et al.,⁴⁷ Mergeomics employed marker set enrichment analysis and weighted key driver analysis to integrate multi-omics data and pinpoint key molecular targets. This approach identified tumor-associated calcium signal transducer 2 (*Tacstd2*) as a key driver in the ferroptosis network associated with cisplatin-induced AKI. Functional analyses, including gene ontology and Kyoto encyclopedia of genes and genomes (KEGG), along with network visualization using Cytoscape,

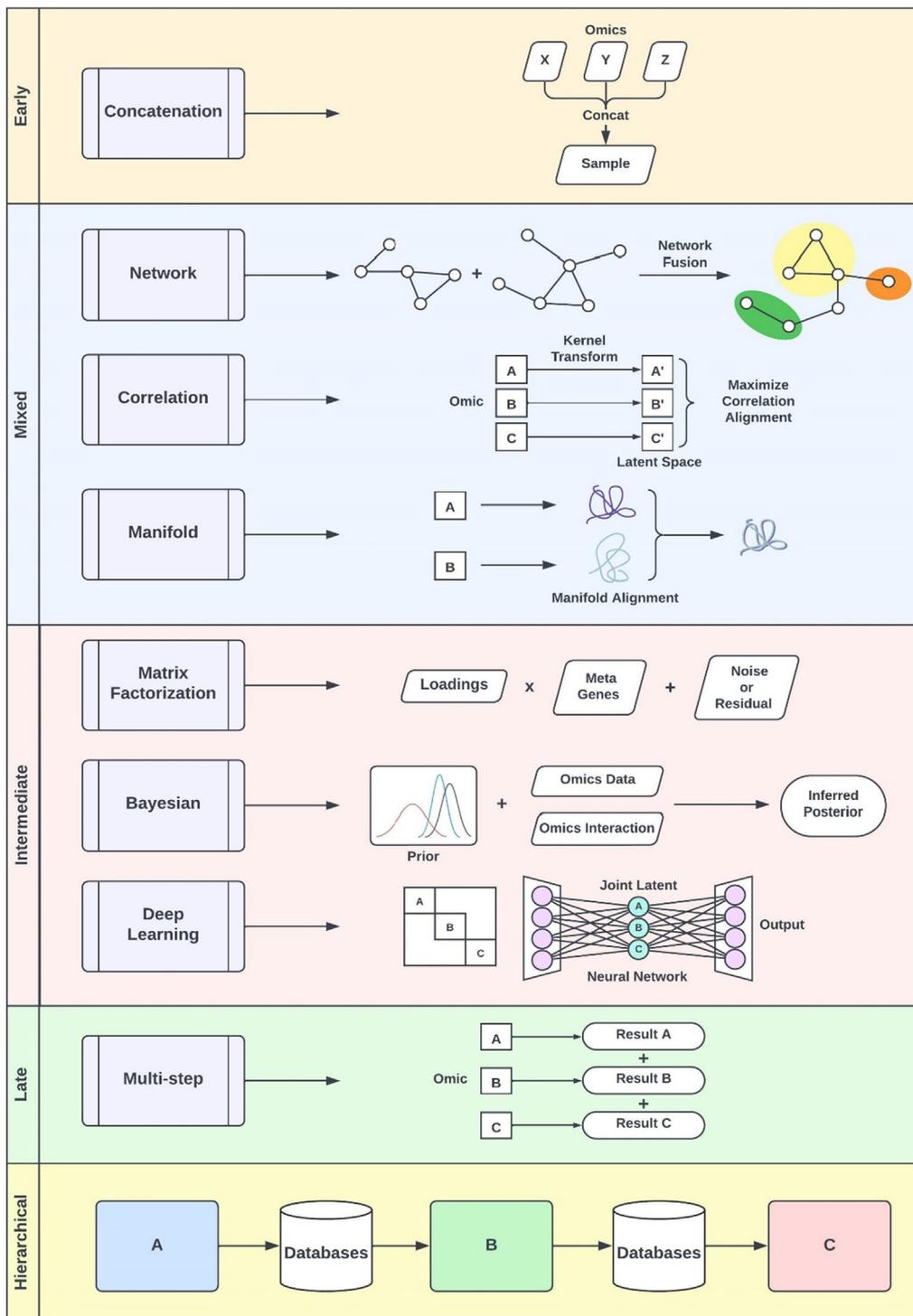


Figure 2. Overview of integration techniques in multi-omics data analysis.

provided further insights into *Tacstd2*'s role and its interactions with other genes.

spatial data, which have been developed to address the aforementioned challenges (Tables 1-3).

Description of Data Integration Methods

Herein, we organized the state-of-the-art data integration methods that can be applied to bulk, single-cell, and

Integration Methods for Bulk Omics Data

Bulk omics integration methods are developed to address problems, mainly focusing on disease subtyping,

Table 1. Integration Methods for Bulk Omics Data

Method	Field of Application	Algorithm	Sample Requirement	Supported Input Data	Reference
BCC	Subtyping	Bayesian consensus clustering	Paired	Omics data	98
iClusterPlus	Subtyping, Biomarker	Gaussian latent variable model	Paired	CNV, Omics data	57
intNMF	Subtyping	Non-negative matrix factorization	Paired	CNV, Omics data	56
JIVE	Subtyping, Insights	Joint and individual variation	Partially paired	Omics data	58
mixOmics	Subtyping, Biomarker	Canonical correlation, Partial least squares	Paired	Omics data	60
MOFA	Subtyping, Biomarker	Factor analysis	Partially paired	Mutation, Omics data, Drug Response	51
NEMO	Subtyping	Similarity	Partially paired	Omics data	54
PSDF	Subtyping	Bayesian	Paired	Omics data	50
SNF	Subtyping	Similarity network fusion	Paired	Omics data, Discrete data	53

survival prediction of subtypes, biomarker discovery, and biological insights using large-scale datasets such as gene expression, DNA methylation, CNVs, and mutations.⁴⁸ Bulk omics inputs are usually composed of matrices with rows representing features (gene, CpG, protein, drug response, etc.) and columns representing bulk tissue samples or patients. Huang et al. classified joint analysis of bulk multi-omics data into various categories, including Bayesian, network-based, matrix factorization-based, and multi-step correlation-based methods, involving mix and intermediate integration approaches as mentioned above.⁴⁹

Bayesian methods such as Bayesian consensus clustering (BCC)⁹⁸ and patient-specific data fusion (PSDF)⁵⁰ use the Dirichlet process foundation for clustering and feature selection. MOFA uses a probabilistic Bayesian model supporting a combination of different noise models.⁵¹ For instance, Clos-Garcia et al. found distinct gut microbiome and blood metabolome signatures in individuals with type 1 diabetes and progressive kidney disease, stratified by albuminuria levels, by integrating

multi-omics data using methods like shotgun sequencing, metabolomics profiling, and the MOFA tool for multi-omics analysis.⁵²

Network-based methods such as Similarity Network Fusion (SNF)⁵³ and Neighborhood-based multi-omics clustering (NEMO)⁵⁴ handle data heterogeneity by constructing and fusing omic-specific similarity matrices, enabling robust subgroup identification across samples.

For example, gene expression and DNA methylation data were integrated using SNF in a study⁵⁵ to classify kidney renal clear cell carcinoma (KIRC) into three stages. By combining these data types, a fused network was created, which improved cancer stage prediction compared to using either data type alone. The fused network's structure captured patient relationships more effectively, allowing for more accurate classification. This approach demonstrated the power of multi-omics data integration for better disease diagnosis and classification of KIRC.

Matrix factorization-based methods such as Integrative NMF (intNMF),⁵⁶ iCluster+,⁵⁷ and JIVE⁵⁸ project

Table 2. Integration Methods for Single-Cell Omics Data

Method	Field of Application	Algorithm	Sample Requirement	Supported Input Data	Reference
Cobolt	Joint latent	Hierarchical generative model	Paired	scRNA-seq, scATAC-seq	68
GLUE	Joint latent	Graph-linked unified embedding	Unpaired	scRNA-seq, scATAC-seq, scmC-seq	75
Signac	Joint latent	Reciprocal LSI projection	Paired	scRNA-seq, scATAC-seq	69
UnionCom	Joint latent	Manifold alignment	Unpaired	scRNA-seq, scATAC-seq, scDNAm	99
MOFA+	Joint latent	Stochastic variational inference	Paired	scRNA-seq, scATAC-seq, Methyl	66
MultiVI	Joint latent	Deep generative model	Paired	scRNA-seq, scATAC-seq, Protein	67
LIGER	Joint latent	Matrix factorization	Unpaired	scRNA-seq, scATAC-seq, scDNAm	74
Seurat	Joint latent	Weighted nearest neighbor	Unpaired	scRNA-seq, scATAC-seq, Protein	76

Table 3. Integration Methods for Spatial Omics Data

Method	Field of Application	Algorithm	Sample Requirement	Supported Input Data	Reference
LIGER	Joint latent	Matrix factorization	Unpaired	Single-cell data, Spatial data	74
Seurat	Joint latent	Weighted nearest neighbor	Unpaired	Single-cell data, Spatial data	76
ENVI	Mapping	Variational inference	Unpaired	scRNA-seq, Spatial transcriptomics	82
Tangram	Mapping	Nonconvex optimization	Paired	scRNA-seq, Spatial transcriptomics	83
stPlus	Mapping	Autoencoder, Weighted k-NN	Unpaired	scRNA-seq, Spatial transcriptomics	84
CellTrek	Mapping	Random forest, Mutual nearest neighbors	Unpaired	scRNA-seq, Spatial transcriptomics	85
SpaGCN	Joint latent, Spatial domain	Graph convolutional network	Paired	Spatial transcriptomics with histology	88
GraphST	Alignment, Spatial domain	Graph self-supervised contrastive learning	Paired	Spatial transcriptomics with histology	89
SpatialGlue	Joint latent, Spatial domain	Graph neural network with dual attention	Partially paired	Spatial epigenome, transcriptome, proteome	90

the data into lower-rank matrices including metagenes (combination of original genes) and loadings (coefficient matrix), facilitating the identification of shared biological patterns while reducing noise. A recent study⁵⁹ integrated data on aneuploidy, DNA hypermethylation, mRNA, and miRNA expression, identifying 28 distinct molecular subtypes, with clustering patterns reflecting histology and tissue type. Using the iCluster algorithm, three kidney cancer subtypes were identified: a pan-kidney cluster, an epithelial-mesenchymal transition cluster, and a heterogeneous group consisting of clear cell, papillary, and chromophobe renal cell carcinomas (RCC). The pan-kidney cluster showed high hypoxia signaling and active immune-related pathways, while chromophobe RCC co-clustered with adrenal cortical carcinoma, lacking hypoxic and immune signals. This analysis enhances the understanding of kidney cancer's molecular landscape.

On the other hand, correlation-based methods such as mixOmics⁶⁰ implement both canonical correlation analysis and partial least squares to maximize the covariance between different data modalities and mitigate sensitivity to outliers. This analysis is exemplified in a study on kidney transplant rejection⁶¹ applying various techniques, including principal component analysis (PCA), integrative PCA, sparse partial least squares discriminant analysis, and sparse generalized canonical correlation analysis (currently accessible in mixOmics). The study integrated genomics and proteomics data from patients with acute rejection and nonrejecting controls, identifying key gene and protein signatures associated with immune responses. Although most integration methods require paired samples across multiple modalities to proceed, methods like JIVE,⁵⁸ multi-omics factor analysis (MOFA),⁵¹ and NEMO⁵⁴ can be applied to partially

paired samples to solve the missing data dilemma described earlier.

Last but not least, the multistep integration combines data from various omics layers to extract meaningful insights in a structured manner and then integrates them to capture shared patterns and interactions, focusing on the consensus and association between omics.⁴⁰ Haug et al.⁶² used the multistep approach to explore the renal medulla, creating a detailed genomic reference for adult kidneys. By integrating individual results from bulk RNA-seq, ATAC-seq, and chromatin conformation with spatial transcriptomics and immunohistochemistry, they identified 31 high-confidence marker genes and provided clear distinctions between medullary and cortical tissues. Additionally, the study reclassified genotype-tissue expression project (GTEx) kidney samples into medullary or cortical categories using PCA. This combined GTEx data with the study's RNA-seq data, allowing for differential analysis across a broader dataset. It also highlighted medullary-specific enhancer regions and potential regulatory roles of genes like claudin 14 (*CLDN14*) and Wnt family member 7B (*WNT7B*).

Integration Methods for Single-Cell Omics Data

Unlike bulk omics data, single-cell omics datasets provide information at a single-cell resolution, enabling more precise understanding of various aspects, including cell type classification, detailed modeling of gene regulatory networks, and the biological process under study.⁶³ In single-cell multi-omics analysis, the objective is to integrate diverse modalities into a shared latent space where the data can be jointly analyzed to foster downstream analyses (cell lineage tracing, tissue atlas, assessing microenvironment, etc.).⁶⁴ These integration

methods can be broadly categorized into paired and unpaired strategies, the latter of which, unpaired methods, is also called *alignment*.⁶⁵

In the case of paired data where multiple omics layers are collected simultaneously from the same cell, Stanovjevic et al. summarized the common integration method as matrix factorization–based, neural network–based, and network-based approaches.⁶⁵ Matrix factorization–based methods like MOFA+⁶⁶ employ hierarchical generative modeling and variational inference to produce unified latent representations. Deep learning–based methods like MultiVI⁶⁷ and Cobolt⁶⁸ similarly use variational autoencoders that model each omic layer with its own distinct distribution. Network-based methods like Signac⁶⁹ focus specifically on two omics, such as scRNA-seq and scATAC-seq, leveraging latent semantic indexing (LSI) to treat chromatin accessibility data in a way analogous to natural language processing.⁷⁰

In a study of DKD,⁷¹ researchers used Signac to analyze single-nucleus RNA-seq (snRNA-seq) and single-nucleus ATAC-seq (snATAC-seq) data from kidney cortex samples from patients with and without type 2 diabetes, to integrate chromatin accessibility with gene expression profiles. They identified accessible chromatin regions across various cells, with a focus on proximal convoluted tubules cells. Signac enabled batch effect correction and doublet removal, facilitating the identification of chromatin accessibility linked to cell type–specific markers. This approach uncovered differential chromatin features and transcription factor binding sites associated with glucose metabolism and glucocorticoid signaling.

A study investigating the role of mineralocorticoid excess in hypertension and kidney disease⁷² employed Signac for scATAC-seq analysis identifying chromatin accessibility changes in principal and connecting tubule cells and highlighting the protective effects of antihypertensive therapies in a rat model. Signac was also used to explore how a Wilms tumor–associated mutation in the histone acetylation reader eleven-nineteen-leukemia (ENL) disrupts kidney differentiation in mice,⁷³ revealing alterations in the gene regulatory landscape that promote nephron progenitor commitment while hindering their maturation, ultimately leading to severe developmental defects, which could be reversed by a small molecule inhibitor targeting mutant ENL.

In contrast, unpaired methods are designed to integrate omics data obtained from different experiments and cell types without requiring direct correspondence between datasets. A common approach in these methods is to map the data into a shared co-embedded space or a nonlinear manifold to identify common patterns.⁶⁵ LIGER⁷⁴ uses iNMF to align and analyze heterogeneous datasets. GLUE⁷⁵ applies graph-linked variational autoencoders and adversarial alignment to learn unified feature representations while simultaneously

reconstructing omics layers from unaligned datasets. Seurat⁷⁶ integrates data using CCA combined with mutual nearest neighbor identification to anchor features. One study on autosomal dominant polycystic kidney disease (ADPKD)⁷⁷ analyzed single-nucleus multi-omics data from a mouse model, integrating transcriptomic and epigenetic datasets while using Harmony for batch effect correction. Uniform manifold approximation and projection (UMAP) and a weighted nearest neighbor graph were employed for dimensional reduction and integration of snRNA-seq and snATAC-seq data. Seurat’s FindMarkers function identified differentially expressed genes and accessible chromatin regions, revealing cell type–specific responses to polycystic kidney disease (*Pkd1*) deletion. In another related study,⁷⁸ snRNA-seq data from ADPKD and healthy human kidney samples were analyzed. Harmony was again used for batch effect correction, with UMAP for dimensionality reduction and the Louvain algorithm for clustering. This study identified key signaling pathways, particularly in proximal tubular cells, and highlighted G protein–coupled receptor class C group 5 member A (*GPRC5A*) as a marker for cyst-lining collecting duct cells. Both studies by Humphreys et al.^{77,78} utilized Seurat for single-cell-level integration and analysis, providing insight into the early molecular deregulation in a mouse model of PKD and human ADPKD kidney pathology.

Spatial Omics Integration Methods

Spatial omics integration is crucial for capturing tissue architecture and deciphering intricate cell–cell interactions with spatial and temporal precision. However, spatial assays often face inherent trade-offs between resolution and the breadth of molecular features they can access. For instance, platforms like 10x Genomics’ Visium offer spot-level resolution, typically capturing around 10 cells per spot while offering comprehensive transcriptomic profiling. On the other hand, matrix-assisted laser desorption/ionization mass spectrometry imaging (MALDI-MSI) provides high spatial resolution, enabling visualization of spatial metabolites within tissues.

This integration of datasets is further enhanced when combining transcriptomic, chromatin accessibility, and metabolic features that overlap with single-cell data. Computational tools like LIGER and Seurat can leverage these shared features for robust alignment and joint analysis. For example, in a recent study, Humphreys and colleagues used the 10X Visium platform to spatially map transcriptomic data in human kidney tissue. They integrated these data with histopathological analysis by performing morphology-based clustering. Seurat was then used to normalize the transcriptomic data, perform clustering, and reduce dimensionality, allowing morphology-based clusters integration with gene expression

profiles. This integration helped align molecular signatures with specific tissue morphologies, enhancing the interpretation of kidney lesions associated with CKD.⁷⁹

Rabelink and colleagues employed MALDI-MSI to map lipidome and metabolome across renal structures in the developing human fetal kidney. They performed post-MALDI immunofluorescence staining for coregistration with the MALDI-MSI data, which enhanced the identification of distinct renal cell types. Using Seurat, they integrated scRNA-seq data with spatial metabolomics and isotopic tracing. This multi-omics approach linked specific metabolic signatures to distinct renal cell types, revealing a shift in substrate utilization from glycolysis to fatty acid oxidation during proximal tubules differentiation.⁸⁰ Similarly, in a related study investigating metabolic changes in kidney tissue after ischemia-reperfusion injury in mice, the team also integrated MALDI-MSI with immunofluorescence data. They conducted MSI measurements on kidney tissue sections at various time points and with different isotope-enriched nutrients. To capture dynamic metabolic changes at pixel level while minimizing batch effects, they employed a two-step process from the Seurat package. First, anchors were identified between datasets using single-pixel lipid profiles; then, they transferred the abundance of isotope-enriched metabolites into the control dataset via k-nearest neighbors (k-NN) analysis. This process established comprehensive isotope-labeling information, enabling calculation of dynamic metabolic changes and the generation of pseudoimages that visualized the *in situ* heterogeneity of metabolic dynamics in ischemic injury.⁸¹

Despite the advantages of spatial assays, challenges remain in balancing resolution with the breadth of omics features. Technologies such as 10x Xenium, NanoString, CosMx, and MERFISH offer single-cell resolution but are limited in transcriptomic coverage, typically measuring a few hundred to a few thousand genes. This reduced depth presents challenges for detailed biological analysis. To address this, methods like ENVI,⁸² Tangram,⁸³ stPlus,⁸⁴ and CellTrek⁸⁵ have been developed to impute missing gene expression in spatial data by leveraging scRNA-seq information. These tools integrate reconstructed expression data with spatial datasets, facilitating the alignment and joint analysis of single-cell and spatial omics data. Instead of imputing gene expression, methods like CellTrek use a spatial mapping algorithm to project cells onto spatial transcriptomic coordinates and employ random forest and mutual nearest neighbor to impute the spatial coordinates of single-cell data. For example, in a recent study addressing the kidney's complex three-dimensional cellular structure, Susztak et al.⁸⁶ generated high-quality datasets from numerous samples of healthy, diabetic, and hypertensive kidney tissues. They utilized single-cell, single-nucleus, and spatial transcriptomics data. Using Signac, they analyzed snRNA-seq and snATAC-seq data to identify cell types

and chromatin accessibility. These cell types were mapped to spatial transcriptomic data using Cell2location, a Bayesian model that enables the integration of single-cell and spatial transcriptomics with higher sensitivity and resolution.⁸⁷ CellTrek was employed to impute precise cell locations at near single-cell resolution. This integration revealed distinct kidney microenvironments, such as glomerular, immune, tubular, and fibrotic regions, providing new insights into the spatial organization of healthy and diseased kidney tissues.

Other spatial omics tools, including SpaGCN⁸⁸ and GraphST,⁸⁹ integrate spatial information and single-omics modalities by applying graph neural networks (GNNs) to capture spatial domains, enhancing the interpretation of tissue architecture. They can identify spatial domains that preserve temporal and functional information. However, these methods only support unimodal input, and feature concatenation is needed if applied on the multimodality scale assuming the same weight.

Unlike the previously mentioned methods, which focus primarily on aligning spatial transcriptomics with single-omics or single-modality data, SpatialGlue⁹⁰ is currently the only method designed for spatial multi-omics data integration using a GNN framework with the dual-attention mechanism. SpatialGlue first uses k-NN to construct a spatial neighborhood graph and then applies a GNN with an attention mechanism to learn modality-specific representations. This allows for the detailed analysis of complex tissue structures and cell interactions across different molecular modalities.

FUTURE PERSPECTIVE AND PREDICTIONS IN THE FIELD

The field of nephrology stands on the cusp of a transformative era driven by the integration of high-dimensional omics data across different measurement scales. The advancements in multi-omics technologies, as discussed in this review, have already begun to unravel the complex molecular underpinnings of kidney diseases. We are optimistic that the future holds even greater promise as these technologies continue to evolve and become more sophisticated. These developments will be important to pursue to implement precision nephrology.

Enhanced Data Integration and Interpretation

Future efforts in nephrology will likely focus on enhancing the integration of multi-omics data across different resolutions and modalities. The development of more advanced computational algorithms that can handle the increasing complexity and heterogeneity of these datasets will be crucial for this task. These algorithms will need to seamlessly integrate genomics, transcriptomics, proteomics, metabolomics, and spatial data to provide a holistic understanding of kidney disease mechanisms.

Integration With Artificial Intelligence and Machine Learning

The future of multi-omics integration can be closely tied to and benefit from the most recent advancements in AI and ML. These technologies will play a pivotal role in identifying patterns and correlations within complex omics datasets that may not be apparent through traditional analysis methods. AI-driven models could predict disease progression, identify novel therapeutic targets, and even suggest personalized treatment plans based on a patient's unique omics profile. Importantly, some AI and ML models can be adapted to increasingly accessible large language models⁹¹ and foundation models,⁹² making model retraining more efficient.

Broader Data Accessibility and Global Collaboration

The accuracy of the multi-omics integration and AI models in nephrology will undoubtedly depend on broader accessibility to these technologies and global collaboration. Developing open-access data repositories and open-source code/packages will ensure that researchers and clinicians worldwide have the necessary tools and datasets to advance the field. International consortiums and cross-disciplinary collaborations will foster the sharing of knowledge and resources, accelerating the pace of discovery and translating research findings into clinical practice more rapidly.

Real-Time Personalized Medicine

The ultimate goal of multi-omics integration in nephrology is the realization of real-time, personalized medicine. As these technologies mature, we anticipate the emergence of clinical tools that can integrate patient-specific omics data in real time, enabling personalized treatment strategies. This will be particularly transformative for diseases like CKD and AKI, where early detection and individualized treatment can significantly improve outcomes.

Crosstalk Between Nephrology and Other Diseases

The kidneys play a critical role in maintaining overall health by filtering waste, balancing fluids, and regulating blood pressure, among other functions. Disruptions in kidney function can have widespread effects on other systems, and vice versa. There have been known connections between kidney diseases and other diseases including cardiovascular diseases, lung diseases, cancers, and neurodegenerative diseases.^{93–96,100} Kidney diseases could also be complications of other diseases, such as preeclampsia during pregnancy.⁹⁷ It is expected that multi-organ or system-level research using tools such as

multi-omics will reveal new holistic insights for the long-term management of patients to improve their life span and quality of life.

REFERENCES

1. 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>.
2. Picard M, Scott-Boyer MP, Bodein A, et al. Integration strategies of multi-omics data for machine learning analysis. *Comput Struct Biotechnol J.* 2021;19:3735-46. <https://doi.org/10.1016/j.csbj.2021.06.030>.
3. Wang Y, Tang S, Ma R, et al. Multi-modal intermediate integrative methods in neuropsychiatric disorders: a review. *Comput Struct Biotechnol J.* 2022;20:6149-62. <https://doi.org/10.1016/j.csbj.2022.11.008>.
4. Ryu Y, Han GH, Jung E, et al. Integration of single-cell RNA-Seq datasets: a review of computational methods. *Mol Cells.* 2023;46:106-19. <https://doi.org/10.14348/molcells.2023.0009>.
5. Kiessling P, Kuppe C. Spatial multi-omics: novel tools to study the complexity of cardiovascular diseases. *Genome Med.* 2024;16:14. <https://doi.org/10.1186/s13073-024-01282-y>.
6. Odenkirk MT, Reif DM, Baker ES. Multiomic big data analysis challenges: increasing confidence in the interpretation of artificial intelligence assessments. *Anal Chem.* 2021;93:7763-73. <https://doi.org/10.1021/acs.analchem.0c04850>.
7. Devarajan P, Chertow GM, Susztak K, et al. Emerging role of clinical genetics in CKD. *Kidney Med.* 2022;4:100435. <https://doi.org/10.1016/j.xkme.2022.100435>.
8. Zoccali C, Mallamaci F, Lightstone L, et al. A new era in the science and care of kidney diseases. *Nat Rev Nephrol.* 2024;20:460-72. <https://doi.org/10.1038/s41581-024-00828-y>.
9. Boger CA, Gorski M, Li M, et al. Association of eGFR-related loci identified by GWAS with Incident CKD and ESRD. *PLoS Genet.* 2011;7:e1002292. <https://doi.org/10.1371/journal.pgen.1002292>.
10. Menon MC, Chuang PY, Li Z, et al. Intronic locus determines SHROOM3 expression and potentiates renal allograft fibrosis. *J Clin Invest.* 2015;125:208-21. <https://doi.org/10.1172/JCI76902>.
11. Li D, Hsu FC, Palmer ND, et al. Multiomics analyses identify AKR1A1 as a biomarker for diabetic kidney disease. *Diabetes.* 2024;73:1188-95. <https://doi.org/10.2337/db23-0540>.
12. Hill C, Avila-Palencia I, Maxwell AP, et al. Harnessing the full potential of multi-omic analyses to advance the study and treatment of chronic kidney disease. *Front Nephrol.* 2022;2:923068. <https://doi.org/10.3389/fneph.2022.923068>.
13. Craciun FL, Bijol V, Ajay AK, et al. RNA sequencing identifies novel translational biomarkers of kidney fibrosis. *J Am Soc Nephrol.* 2016;27:1702-13. <https://doi.org/10.1681/ASN.2015020225>.
14. Schreiber F, Anslinger TM, Kramann R. Fibrosis in pathology of heart and kidney: from deep RNA-sequencing to novel molecular targets. *Circ Res.* 2023;132:1013-33. <https://doi.org/10.1161/CIRCRESAHA.122.321761>.
15. Wu H, Malone AF, Donnelly EL, et al. Single-cell transcriptomics of a human kidney allograft biopsy specimen defines a diverse inflammatory response. *J Am Soc Nephrol.* 2018;29:2069-80. <https://doi.org/10.1681/ASN.2018020125>.
16. Wanner N, Vornweg J, Combes A, et al. DNA methyltransferase 1 controls nephron progenitor cell renewal and differentiation. *J Am Soc Nephrol.* 2019;30:63-78. <https://doi.org/10.1681/ASN.2018070736>.
17. 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>.
18. 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>.
 19. Aronov PA, Luo FJ, Plummer NS, et al. Colonic contribution to uremic solutes. *J Am Soc Nephrol.* 2011;22:1769-76. <https://doi.org/10.1681/ASN.2010121220>.
 20. Sharma K, Karl B, Mathew AV, et al. Metabolomics reveals signature of mitochondrial dysfunction in diabetic kidney disease. *J Am Soc Nephrol.* 2013;24:1901-12. <https://doi.org/10.1681/ASN.2013020126>.
 21. Zhang J, Fuhrer T, Ye H, et al. High-throughput metabolomics and diabetic kidney disease progression: evidence from the chronic renal insufficiency (CRIC) study. *Am J Nephrol.* 2022;53:215-25. <https://doi.org/10.1159/000521940>.
 22. Poyan Mehr A, Tran MT, Ralto KM, et al. De novo NAD(+) biosynthetic impairment in acute kidney injury in humans. *Nat Med.* 2018;24:1351-9. <https://doi.org/10.1038/s41591-018-0138-z>.
 23. Hu J, Wang SG, Hou Y, et al. Multi-omic profiling of clear cell renal cell carcinoma identifies metabolic reprogramming associated with disease progression. *Nat Genet.* 2024;56:442-57. <https://doi.org/10.1038/s41588-024-01662-5>.
 24. Sharma K, Zhang G, Hansen J, et al. Endogenous adenine mediates kidney injury in diabetic models and predicts diabetic kidney disease in patients. *J Clin Invest.* 2023;133:e170341. <https://doi.org/10.1172/JCI170341>.
 25. 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>.
 26. 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>.
 27. Ouyang M, Welsh WJ, Georgopoulos P. Gaussian mixture clustering and imputation of microarray data. *Bioinformatics.* 2004;20:917-23. <https://doi.org/10.1093/bioinformatics/bth007>.
 28. Flores JE, Claborn DM, Weller ZD, et al. Missing data in multi-omics integration: recent advances through artificial intelligence. *Front Artif Intell.* 2023;6:1098308. <https://doi.org/10.3389/frai.2023.1098308>.
 29. Zhang M, Ye Z, Yuan E, et al. Imaging-based deep learning in kidney diseases: recent progress and future prospects. *Insights Imag.* 2024;15:50. <https://doi.org/10.1186/s13244-024-01636-5>.
 30. Menon R, Otto EA, Hoover P, et al. Single cell transcriptomics identifies focal segmental glomerulosclerosis remission endothelial biomarker. *JCI Insight.* 2020;5:e133267. <https://doi.org/10.1172/jci.insight.133267>.
 31. 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>.
 32. El-Achkar TM, Eadon MT, Kretzler M, et al. Precision medicine in nephrology: an integrative framework of multidimensional data in the kidney precision medicine project. *Am J Kidney Dis.* 2024;83:402-10. <https://doi.org/10.1053/j.ajkd.2023.08.015>.
 33. 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>.
 34. 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>.
 35. 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>.
 36. 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>.
 37. Argelaguet R, Cuomo ASE, Stegle O, et al. Computational principles and challenges in single-cell data integration. *Nat Biotechnol.* 2021;39:1202-15. <https://doi.org/10.1038/s41587-021-00895-7>.
 38. Rautenstrauch P, Ohler U. Liam tackles complex multimodal single-cell data integration challenges. *Nucleic Acids Res.* 2024;52:e52. <https://doi.org/10.1093/nar/gkae409>.
 39. Xu Y, McCord RP. Diagonal integration of multimodal single-cell data: potential pitfalls and paths forward. *Nat Commun.* 2022;13:3505. <https://doi.org/10.1038/s41467-022-31104-x>.
 40. Ritchie MD, Holzinger ER, Li R, et al. Methods of integrating data to uncover genotype-phenotype interactions. *Nat Rev Genet.* 2015;16:85-97. <https://doi.org/10.1038/nrg3868>.
 41. Rappoport N, Shamir R. Multi-omic and multi-view clustering algorithms: review and cancer benchmark. *Nucleic Acids Res.* 2018;46:10546-62. <https://doi.org/10.1093/nar/gky889>.
 42. Reel PS, Reel S, Pearson E, et al. Using machine learning approaches for multi-omics data analysis: a review. *Biotechnol Adv.* 2021;49:107739. <https://doi.org/10.1016/j.biotechadv.2021.107739>.
 43. Lee DD, Seung HS. Learning the parts of objects by non-negative matrix factorization. *Nature.* 1999;401:788-91. <https://doi.org/10.1038/44565>.
 44. Chen Z, Satake E, Pezzolesi MG, et al. Integrated analysis of blood DNA methylation, genetic variants, circulating proteins, microRNAs, and kidney failure in type 1 diabetes. *Sci Transl Med.* 2024;16:eadj3385. <https://doi.org/10.1126/scitranslmed.adj3385>.
 45. Zarayeneh N, Ko E, Oh JH, et al. Integration of multi-omics data for integrative gene regulatory network inference. *Int J Data Min Bioinform.* 2017;18:223-39. <https://doi.org/10.1504/IJDMB.2017.10008266>.
 46. Ding J, Blencowe M, Nghiem T, et al. Mergeomics 2.0: a web server for multi-omics data integration to elucidate disease networks and predict therapeutics. *Nucleic Acids Res.* 2021;49:W375-87. <https://doi.org/10.1093/nar/gkab405>.
 47. Deng Z, Dong Z, Wang Y, et al. Identification of TACSTD2 as novel therapeutic targets for cisplatin-induced acute kidney injury by multi-omics data integration. *Hum Genet.* 2024;143:1061-80. <https://doi.org/10.1007/s00439-024-02641-w>.
 48. Subramanian I, Verma S, Kumar S, et al. Multi-omics data integration, interpretation, and its application. *Bioinform Biol Insights.* 2020;14:1177932219899051. <https://doi.org/10.1177/1177932219899051>.
 49. Huang S, Chaudhary K, Garmire LX. More is better: recent progress in multi-omics data integration methods. *Front Genet.* 2017;8:84. <https://doi.org/10.3389/fgene.2017.00084>.
 50. Yuan Y, Savage RS, Markowitz F. Patient-specific data fusion defines prognostic cancer subtypes. *PLoS Comput Biol.* 2011;7:e1002227. <https://doi.org/10.1371/journal.pcbi.1002227>.
 51. Argelaguet R, Velten B, Arnol D, et al. Multi-omics factor analysis: a framework for unsupervised integration of multi-omics data sets. *Mol Syst Biol.* 2018;14:e8124. <https://doi.org/10.15252/msb.20178124>.
 52. Clos-Garcia M, Ahluwalia TS, Winther SA, et al. Multiomics signatures of type 1 diabetes with and without albuminuria. *Front*

- Endocrinol (Lausanne). 2022;13:1015557. <https://doi.org/10.3389/fendo.2022.1015557>.
53. Wang B, Mezlini AM, Demir F, et al. Similarity network fusion for aggregating data types on a genomic scale. *Nat Methods*. 2014;11:333-7. <https://doi.org/10.1038/nmeth.2810>.
 54. Rappoport N, Shamir R. NEMO: cancer subtyping by integration of partial multi-omic data. *Bioinformatics*. 2019;35:3348-56. <https://doi.org/10.1093/bioinformatics/btz058>.
 55. Deng SP, Cao S, Huang DS, et al. Identifying stages of kidney renal cell carcinoma by combining gene expression and DNA methylation data. *IEEE/ACM Trans Comput Biol Bioinform*. 2017;14:1147-53. <https://doi.org/10.1109/TCBB.2016.2607717>.
 56. Chalise P, Fridley BL. Integrative clustering of multi-level 'omic data based on non-negative matrix factorization algorithm. *PLoS One*. 2017;12:e0176278. <https://doi.org/10.1371/journal.pone.0176278>.
 57. Shen R, Olshen AB, Ladanyi M. Integrative clustering of multiple genomic data types using a joint latent variable model with application to breast and lung cancer subtype analysis. *Bioinformatics*. 2009;25:2906-12. <https://doi.org/10.1093/bioinformatics/btp543>.
 58. Lock EF, Hoadley KA, Marron JS, et al. Joint and individual variation explained (Jive) for integrated analysis of multiple data types. *Ann Appl Stat*. 2013;7:523-42. <https://doi.org/10.1214/12-AOAS597>.
 59. Hoadley KA, Yau C, Hinoue T, et al. Cell-of-origin patterns dominate the molecular classification of 10,000 tumors from 33 types of cancer. *Cell*. 2018;173:291-304. <https://doi.org/10.1016/j.cell.2018.03.022>.
 60. Rohart F, Gautier B, Singh A, et al. mixOmics: an R package for 'omics feature selection and multiple data integration. *PLoS Comput Biol*. 2017;13:e1005752. <https://doi.org/10.1371/journal.pcbi.1005752>.
 61. Gunther OP, Shin H, Ng RT, et al. Novel multivariate methods for integration of genomics and proteomics data: applications in a kidney transplant rejection study. *Omic*. 2014;18:682-95. <https://doi.org/10.1089/omi.2014.0062>.
 62. Haug S, Muthusamy S, Li Y, et al. Multi-omic analysis of human kidney tissue identified medulla-specific gene expression patterns. *Kidney Int*. 2024;105:293-311. <https://doi.org/10.1016/j.kint.2023.10.024>.
 63. Miao Z, Humphreys BD, McMahon AP, et al. Multi-omics integration in the age of million single-cell data. *Nat Rev Nephrol*. 2021;17:710-24. <https://doi.org/10.1038/s41581-021-00463-x>.
 64. Baysoy A, Bai Z, Satija R, et al. The technological landscape and applications of single-cell multi-omics. *Nat Rev Mol Cell Biol*. 2023;24:695-713. <https://doi.org/10.1038/s41580-023-00615-w>.
 65. Stanojevic S, Li Y, Ristivojevic A, et al. Computational methods for single-cell multi-omics integration and alignment. *Genom Proteom Bioinform*. 2022;20:836-49. <https://doi.org/10.1016/j.gpb.2022.11.013>.
 66. Argelaguet R, Arnol D, Bredikhin D, et al. MOFA+: a statistical framework for comprehensive integration of multi-modal single-cell data. *Genome Biol*. 2020;21:111. <https://doi.org/10.1186/s13059-020-02015-1>.
 67. Ashuaq T, Gabitto MI, Koodli RV, et al. MultiVI: deep generative model for the integration of multimodal data. *Nat Methods*. 2023;20:1222-31. <https://doi.org/10.1038/s41592-023-01909-9>.
 68. Gong B, Zhou Y, Purdom E. Cobolt: integrative analysis of multimodal single-cell sequencing data. *Genome Biol*. 2021;22:351. <https://doi.org/10.1186/s13059-021-02556-z>.
 69. Stuart T, Srivastava A, Madad S, et al. Single-cell chromatin state analysis with Signac. *Nat Methods*. 2021;18:1333-41. <https://doi.org/10.1038/s41592-021-01282-5>.
 70. Cusanovich DA, Daza R, Adey A, et al. Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science*. 2015;348:910-4. <https://doi.org/10.1126/science.aab1601>.
 71. Wilson PC, Muto Y, Wu H, et al. Multimodal single cell sequencing implicates chromatin accessibility and genetic background in diabetic kidney disease progression. *Nat Commun*. 2022;13:5253. <https://doi.org/10.1038/s41467-022-32972-z>.
 72. Abedini A, Sanchez-Navarro A, Wu J, et al. Single-cell transcriptomics and chromatin accessibility profiling elucidate the kidney-protective mechanism of mineralocorticoid receptor antagonists. *J Clin Invest*. 2024;134:e157165. <https://doi.org/10.1172/JCI157165>.
 73. Song L, Li Q, Xia L, et al. Single-cell multiomics reveals ENL mutation perturbs kidney developmental trajectory by rewiring gene regulatory landscape. *Nat Commun*. 2024;15:5937. <https://doi.org/10.1038/s41467-024-50171-w>.
 74. Liu J, Gao C, Sodicoff J, et al. Jointly defining cell types from multiple single-cell datasets using LIGER. *Nat Protoc*. 2020;15:3632-62. <https://doi.org/10.1038/s41596-020-0391-8>.
 75. Cao ZJ, Gao G. Multi-omics single-cell data integration and regulatory inference with graph-linked embedding. *Nat Biotechnol*. 2022;40:1458-66. <https://doi.org/10.1038/s41587-022-01284-4>.
 76. Hao Y, Hao S, Andersen-Nissen E, et al. Integrated analysis of multimodal single-cell data. *Cell*. 2021;184:3573-3587. <https://doi.org/10.1016/j.cell.2021.04.048>.
 77. Muto Y, Yoshimura Y, Wu H, et al. Multi-omics profiling of mouse polycystic kidney disease progression at a single cell resolution. *bioRxiv*. 2024. <https://doi.org/10.1101/2024.05.27.595830>.
 78. Muto Y, Dixon EE, Yoshimura Y, et al. Defining cellular complexity in human autosomal dominant polycystic kidney disease by multimodal single cell analysis. *Nat Commun*. 2022;13:6497. <https://doi.org/10.1038/s41467-022-34255-z>.
 79. Isnard P, Li D, Xuanyuan Q, et al. Histopathological-based analysis of human kidney spatial transcriptomics data: toward precision pathology. *Am J Pathol*. 2025;195:69-88. <https://doi.org/10.1016/j.ajpath.2024.06.011>.
 80. Wang G, Heijs B, Kostidis S, et al. Spatial dynamic metabolomics identifies metabolic cell fate trajectories in human kidney differentiation. *Cell Stem Cell*. 2022;29:1580-93. <https://doi.org/10.1016/j.stem.2022.10.008>.
 81. Wang G, Heijs B, Kostidis S, et al. Analyzing cell-type-specific dynamics of metabolism in kidney repair. *Nat Metab*. 2022;4:1109-18. <https://doi.org/10.1038/s42255-022-00615-8>.
 82. Haviv D, Remsik J, Gatie M, et al. The covariance environment defines cellular niches for spatial inference. *Nat Biotechnol*. 2025;43:269-80. <https://doi.org/10.1038/s41587-024-02193-4>.
 83. Biancalani T, Scalia G, Buffoni L, et al. Deep learning and alignment of spatially resolved single-cell transcriptomes with Tangram. *Nat Methods*. 2021;18:1352-62. <https://doi.org/10.1038/s41592-021-01264-7>.
 84. Shengquan C, Boheng Z, Xiaoyang C, et al. stPlus: a reference-based method for the accurate enhancement of spatial transcriptomics. *Bioinformatics*. 2021;37:i299-307. <https://doi.org/10.1093/bioinformatics/btab298>.
 85. Wei R, He S, Bai S, et al. Spatial charting of single-cell transcriptomes in tissues. *Nat Biotechnol*. 2022;40:1190-9. <https://doi.org/10.1038/s41587-022-01233-1>.
 86. Abedini A, Levinsohn J, Klötzer KA, et al. Spatially resolved human kidney multi-omics single cell atlas highlights the key role of the fibrotic microenvironment in kidney disease progression. *bioRxiv*. 2024. <https://doi.org/10.1101/2022.10.24.513598>. 2022.2010.2024.513598.
 87. Kleshchevnikov V, Shmatko A, Dann E, et al. Cell2location maps fine-grained cell types in spatial transcriptomics. *Nat Biotechnol*. 2022;40:661-71. <https://doi.org/10.1038/s41587-021-01139-4>.

88. Hu J, Li X, Coleman K, et al. SpaGCN: integrating gene expression, spatial location and histology to identify spatial domains and spatially variable genes by graph convolutional network. *Nat Methods*. 2021;18:1342-51. <https://doi.org/10.1038/s41592-021-01255-8>.
89. Long Y, Ang KS, Li M, et al. Spatially informed clustering, integration, and deconvolution of spatial transcriptomics with GraphST. *Nat Commun*. 2023;14:1155. <https://doi.org/10.1038/s41467-023-36796-3>.
90. Long Y, Ang KS, Sethi R, et al. Deciphering spatial domains from spatial multi-omics with SpatialGlue. *Nat Methods*. 2024;21:1658-67. <https://doi.org/10.1038/s41592-024-02316-4>.
91. Toufiq M, Rinchai D, Bettacchioli E, et al. Harnessing large language models (LLMs) for candidate gene prioritization and selection. *J Transl Med*. 2023;21:728. <https://doi.org/10.1186/s12967-023-04576-8>.
92. Cui H, Wang C, Maan H, et al. scGPT: toward building a foundation model for single-cell multi-omics using generative AI. *Nat Methods*. 2024;21:1470-80. <https://doi.org/10.1038/s41592-024-02201-0>.
93. Etgen T. Kidney disease as a determinant of cognitive decline and dementia. *Alzheimers Res Ther*. 2015;7:29. <https://doi.org/10.1186/s13195-015-0115-4>.
94. Kitchlu A, Reid J, Jeyakumar N, et al. Cancer risk and mortality in patients with kidney disease: a population-based cohort study. *Am J Kidney Dis*. 2022;80:436-48. <https://doi.org/10.1053/ajkd.2022.02.020>.
95. Parivakkam Mani A, K S, Sundar R, et al. Pulmonary manifestations at different stages in the chronic kidney disease: an observational study. *Cureus*. 2023;15:e39235. <https://doi.org/10.7759/cureus.39235>.
96. Kon V, Shelton EL, Pitzer A, et al. Inflammation, lymphatics, and cardiovascular disease: amplification by chronic kidney disease. *Curr Hypertens Rep*. 2022;24:455-63. <https://doi.org/10.1007/s11906-022-01206-4>.
97. Kattah A. Preeclampsia and kidney disease: deciphering cause and effect. *Curr Hypertens Rep*. 2020;22:91. <https://doi.org/10.1007/s11906-020-01099-1>.
98. Lock EF, Dunson DB. Bayesian consensus clustering. *Bioinformatics*. 2013;29:2610-6. <https://doi.org/10.1093/bioinformatics/btt425>.
99. Cao K, Bai X, Hong Y, et al. Unsupervised topological alignment for single-cell multi-omics integration. *Bioinformatics*. 2020;36:i48-56. <https://doi.org/10.1093/bioinformatics/btaa443>.
100. Saliba A, Debnath S, Tamayo I, et al. Quinolinic acid potentially links kidney injury to brain toxicity. *JCI Insight*. 2025;10(6):e180229. <https://doi.org/10.1172/jci.insight.180229>.