KIDNEY RESEARCH AND CLINICAL PRACTICE

In silico medicine and *-omics* strategies in nephrology: contributions and relevance to the diagnosis and prevention of chronic kidney disease

Ana Checa-Ros[1,2,*], Antonella Locascio[1,*], Nelia Steib[1], Owahabanun-Joshua Okojie[1], Totte Malte-Weier[1], Valmore Bermúdez[3], Luis D'Marco[1]

[1]*Grupo de Investigación en Enfermedades Cardiorrenales y Metabólicas, Departamento de Medicina y Cirugía, Facultad de Ciencias de la Salud, Universidad Cardenal Herrera-CEU, CEU Universities, Valencia, Spain*
[2]*School of Life & Health Sciences, Aston University, Birmingham, United Kingdom*
[3]*Facultad de Ciencias de la Salud, Universidad Simón Bolívar, Barranquilla, Colombia*

Chronic kidney disease (CKD) has been increasing over the last years, with a rate between 0.49% to 0.87% new cases per year. Currently, the number of affected people is around 850 million worldwide. CKD is a slowly progressive disease that leads to irreversible loss of kidney function, end-stage kidney disease, and premature death. Therefore, CKD is considered a global health problem, and this sets the alarm for necessary efficient prediction, management, and disease prevention. At present, modern computer analysis, such as *in silico* medicine (ISM), denotes an emergent data science that offers interesting promise in the nephrology field. ISM offers reliable computer predictions to suggest optimal treatments in a case-specific manner. In addition, ISM offers the potential to gain a better understanding of the kidney physiology and/or pathophysiology of many complex diseases, together with a multiscale disease modeling. Similarly, *-omics* platforms (including genomics, transcriptomics, metabolomics, and proteomics), can generate biological data to obtain information on gene expression and regulation, protein turnover, and biological pathway connections in renal diseases. In this sense, the novel patient-centered approach in CKD research is built upon the combination of ISM analysis of human data, the use of *in vitro* models, and *in vivo* validation. Thus, one of the main objectives of CKD research is to manage the disease by the identification of new disease drivers, which could be prevented and monitored. This review explores the wide-ranging application of computational medicine and the application of *-omics* strategies in evaluating and managing kidney diseases.

Keywords: Big data, Chronic kidney diseases, Machine learning, Nephrology

Introduction

Biomedical studies are frequently conducted including *in vivo* (Latin for '*within the living*') observations; this means data collection based on the direct study of living organisms. When the usage of living organisms is not possible or available, a good alternative is the *in vitro* strategy, performed inside artificial confinements, such as test tubes,

culture dishes, or incubators. In this last case, the samples to study derive from living samples (i.e., cultivated cells, synthetic tissues or organs, heterologous systems, etc.). Both strategies are expensive, time-consuming, and require large spaces and high-technology equipment [1]. In addition, ethical issues occasionally make the performance of certain studies even more difficult or impossible.

In the past decade, an extraordinary technological advance characterized the field of biomedical research: the birth of *big data science*. This new science responded to the demands of several research groups to manage multifactorial analyses and to the increasing use of the *-omics* platforms (including genomics, transcriptomic, metabolomics, and proteomics) for the generation of biological data. The volume of data generated in *-omics* includes, among others, information on gene expression and regulation, protein turnover, and biological pathway connections. All those data require fine organization, cleaning, storage, and integration to be efficiently interpreted. The great value of these studies resides in that they can be easily shared between research groups and the information be integrated; in this way, the data collection evolves and can potentially respond to different scientific questions, just changing some of the variables or the way to analyze them. The field of data science has evolved rapidly ever since, and still, it is in constant acceleration, in parallel with the introduction of new informatics languages, like Python and R. The advancement of new informatics tools allows improving the organization of the huge volume of generated data, the organization in structures and levels, until the progression into models and predictions [2]. The big data science set is the base for the introduction and evolution of *in silico* science [3,4].

The term "*in silico*" derives from '*within the silicon*,' and it refers to the computer chips that are made on this material. In a typical *in silico* analysis, the machine creates a virtual environment where it is possible to organize, manage, visualize, reconfigure, and interpret the different variables, as well as derive predictions [5,6]. The great value of the *in silico* application in the medical field resides in enabling a multifactorial analysis, simultaneously involving physiological, molecular, cellular, environmental, social, and physical processes, and providing subject-specific predictions that would be substantially difficult or impossible to measure *in vivo*. The application of the *in silico* approach

to medicine studies is also known as 'computational medicine.' Computational medicine enables strategies for the diagnosis, prognosis, management, prevention, and treatment of diseases [7]. Noteworthy, computational analyses can bypass ethical permissions on many occasions since the data can be anonymous or even simulations (so, not collected in real human patients). In addition, data sharing between laboratory researchers and uploading to international databases is frequent and beneficial. All the predictions and modeling obtained by an *in silico* study can be scaled up to populations, recycled, and interrogated from different points of view and scenarios, allowing a dynamic adaptation of models and simulations in a relatively short time.

The applications of *in silico* science have been explored in diverse medical areas and specialties, being nephrology and, particularly, chronic kidney disease (CKD), amongst them. CKD is the term used to describe the progressive damage and functionality loss of the kidney. In 2023, the International Society of Nephrology-Global Kidney Health Atlas (ISN-GKHA) reported that the number of affected people reached 850 million patients worldwide. CKD incidence has been increasing over the last few years, with a rate ranging between 0.49% and 0.87% new cases per year [8–10]. Age and race seem not to discriminate in this disease incidence. Due to all these considerations, CKD is considered a global health problem, and this sets the alarm for a necessarily efficient prediction, management, and prevention of the disease [11].

In this regard, the novel patient-centered approach in CKD research is built upon the combination of *in silico* analysis of human data, the extensive use of *in vitro* models, and the *in vivo* validation. One of the main objectives of CKD research is to manage the disease by the identification of new disease drivers, which may be prevented and monitored. In this review, we explore the wide-ranging application of computational medicine and the application of *-omics* strategies in evaluating and managing kidney diseases.

In silico-based medicine research

The purpose of a computer *in silico* medicine (ISM) is to collect clinical data, generally coming from different healthcare points, organize and process them, as well as to

integrate the variables to create the most accurate picture of the medical question. The final output of an *in silico* medical study is to combine mathematical and computational data science to provide a clear visualization of the medical problem. This output can be further analyzed dynamically, always combining descriptive, predictive, and prescriptive analytics.

A relevant objective for a computational model is also to create a faithful emulation of the medical case (i.e., patient/ pathology) that can be further analyzed under the view of dynamic connections between the different parts of the human body, by integrating these interactions into one physiological virtual unit (also defined as a 'digital twin'). By contrast, when studied under clinical practice (*in vivo* or *in vitro*), it is hard to solve medical questions by considering the overall complexity of the patient. The organism is frequently analyzed in parts, as if short and specific questions, that hardly can be joined to obtain the general vision of the problem.

The ISM works using different computational-based systems, including databases, artificial intelligence (AI), and machine learning (ML). In the database system, clinical data from patients and volunteers are assembled from different sources and stored in the form of electronic medical records (EMRs). The EMRs are subsequently uploaded into a server or a centralized database, to be available for future consultations by the healthcare staff. It aims to bypass the limitations of paper-based format recording where patient data are stored and put into file cabinets [12].

The core of AI is a complex intersection of algorithms and statistics methods, able to execute the functions of pattern-seeking and recognition, make predictions, and eventually, solve problems. AI can learn, recognize, and interpret human behavior, but unlike humankind, it lacks the creativity to imagine something that has never existed before. For this reason, the bottleneck of *big data science*, and in particular the ISM, is the accurate collection of clinical data and the proper organization of those before any computational analysis.

ML is a subset of AI that includes several kinds of learning machinery, as shown in Table 1 [13]. Thus, ML technology can be broadly divided into supervised learning, unsupervised learning, reinforcement learning, and *deep learning* according to different modeling needs. There are many areas of medicine to which ISM can be applied [14]. In the following section, we describe some of these applications, specifically in clinical trials.

Current evidence of *in silico* models in kidney diseases

Even though *in silico* models are well established in medicine, their use in nephrology is still underway [15]. Nevertheless, there are some interesting examples of ML applications for the prediction of acute kidney injury (AKI) and CKD, as well as to identify different structures from a kidney biopsy and drug-induced nephrotoxicity.

Drug-induced nephrotoxicity is one of the challenges in clinical trials, not only because of the effect it has on patients, but also because it is detected in the final stages of drug development. In a recent study, using algorithms of ML, 87 out of 565 chemical structures with potential nephrotoxicity were identified from an online chemical database [16]. Half of the 565 structures had a related nephrotoxic effect of 0.1%, as extracted from the SIDER (Side Effect Resource) database. The 87 structures, defined as structural alerts, were used as criteria to classify a drug as

Table 1. Different types of machine learning (ML) approaches

Type	Description
Supervised learning	Using logistic regression, random forest, and support vector machine, it receives input and gives a desired output value, a.k.a. supervised signal. It is commonly used in medical research.
Unsupervised learning	Receiving an uncategorized input, the ML by itself categorizes the samples into different categories based on their characteristics. For instance, it can recognize and categorize renal histological samples into different categories based on their morphological differences.
Reinforcement learning	The ML learns and applies its knowledge to the end when it will receive a reward. Reward learning strategies are used to teach ML and help it find the best outcomes in a pool of possibilities.
Deep learning	Which involves finding complex patterns in big data like the pathologic, genetic, environmental, and physiologic connections between diseases in a patient with multiple pathological conditions.

Modified from Li et al. (*Chin Med J [Engl]* 2020;133:687–698) [13] according to the Creative Commons License.

nephrotoxic.

One of the major issues in nephrology is determining the progression of CKD and predicting AKI. A generalized prediction model for CKD would enable early detection of high-risk CKD patients, as well as help to identify suitable treatment targets. Regarding the progression of CKD, a recent review revealed that in all models, the predicted outcome was calculated based on the progression towards end-stage kidney disease (ESKD), over a given interval of time [17]. To provide a consensus about the different CKD progression models, Lim et al. [17] reviewed 33 articles, which mainly used Cox proportional hazards regression models, with only a small amount employing ML or unsupervised models. The general conclusion was that the research in CKD progression based on models was very inconsistent and most of the predictions presented various limitations, concerning both data availability and methodology. The retrospective study by Jaimes Campos et al. [18] investigated the prediction of renal or cardiovascular events using previously defined urinary peptidomic classifiers CKD273, HF2, and CAD160 in a cohort of 5,585 individuals. They found a highly significant prediction of events, with an hazard ratios of 2.59, 1.71, and 4.12 for heart failure, coronary artery disease, and CKD, respectively. Therefore, they applied *in silico* treatment, implementing on each patient's urinary profile changes to the classifiers corresponding to exactly defined peptide abundance changes, following commonly used interventions.

Comparing the work by Tomašev et al. [19], who used recurrent neural networks (RNN) as multiple risk-assessment algorithms for the prediction of AKI at 48 hours, with the one by Koyner et al. [20], who used a Gradient Boosting Machine (GBM) model, is concluded that the RNN offer better predictions (area under the receiver operator characteristic curve of 0.92 for RNN vs. 0.87 for GBM). However, both studies missed an external validation, as they were carried out within the same healthcare system.

In silico models have also been tested to evaluate kidney biopsies and identify pathological tissues, as in glomerulonephritis. New studies have shown that artificial neural networks (ANN) are the method of choice for computational pathology. ANNs can not only classify kidney morphology but also predict the disease status [21]. Newer ANN models, such as convolutional neural networks, can both identify glomerular structures, as well as discriminate between non-glomerulus tissue and sclerosed tissue [13]. Although these AI techniques are not yet close to replacing pathologists, these advances will be able to assist clinicians in their decisions and, in the future, make a breakthrough in precision medicine.

The five possible data sources in nephrology are patient registries and epidemiological studies, EMRs and administrative data, clinical trials, mobile health devices, and molecular data. Molecular data on genetics and biology find their applications in demonstrating the molecular complexity of multifactorial CKD and enabling individualized detection of underlying causal variables and possible treatments of both common and rare kidney diseases [12].

One representative work, an example of this analysis in the field of nephrology, is the one presented by Li et al. [22] about the prediction of diabetic nephropathy (DN). DN is one of the most important causes of both CKD and ESKD based on system biology (the holistic field of study that assesses the molecular mechanisms in organisms on a molecular and system level by integrating computational and mathematical analysis with biomolecular disciplines) [23]. In the study [22], the Gene Expression Omnibus (GEO) dataset was used to make a transcriptomic analysis on available microarray datasets including 19 DN samples and 50 controls as input data. These data were analyzed using a variety of bioinformatics tools to make predictions, such as protein-protein interaction, network creation, analyses of differentially expressed genes, gene ontology, and gene set enrichment analysis. A multitude of new target genes for DN was identified; some of them being potential markers to be used in prognostic or therapeutic options for DN.

Cardiorenal syndrome (CRS) is another promising area in which ML techniques and system biology are integrated to provide good predictive results. CRS is a condition in which there is a primary dysfunction of either the kidney or the heart, leading to the deterioration of the other. The informatics models used by Ishrat et al. [24] were similar to those used by Li et al. [22]. The microRNA (miRNA) dataset was obtained from an external miRNA database and the literature was analyzed by R/GEO2R/miRNet/Cytoscape software. Five miRNAs were identified as significant key genes for CRS progression and as possible biomarkers for diagnosing this syndrome.

Big data science and *omics* strategy

Molecular data consists mostly of -*omics* data, including genomics, transcriptomics, proteomics, and metabolomics. The *in silico* multivariable analysis of big data obtained from -*omics* integration requires data normalization. This step is extremely relevant to ensure an unbiased hypothesis. Also, in this complementary integration of computational and experimental (clinic) techniques, both bottom-up and top-down approaches are crucial to assembling the data obtained from all levels of biological pathways and coordinating these with the physiological processes. Top-down strategies extract information from system-wide methods (metabolomics, proteomics, transcriptomics) and analyze it using diverse network models to decipher molecular mechanisms and functional patterns, allowing the prediction of phenotypic responses and, eventually, the identification of new signaling pathways [25]. On the contrary, bottom-up strategies do not start with data but with the details of a particular molecular network (from silicon-cell models) that is quantitatively studied to reconstruct predictive models applicable to drug design. In this regard, bottom-up strategies serve to the gap between molecules and physiology [23].

As previously described, molecular applications in CKD allow for the identification of risk factors and potential treatments [26]. Moreover, -*omics* data enable visualization of the impact of genetic mutations, individual classification of kidney biopsy, and uremic molecules [27]. The biological interpretation of -*omics* patterns is possible and depends on data repositories, which must be exhaustively annotated in all the main characteristics of renal clinical diseases, such as renal function over time, linked with the progression rate of a CKD [28].

The output information obtained from genomics and transcriptomics ranks these two strategies in different approaches. As an example, transcriptome-based signature matching has been successfully used in the discovery of lysine deacetylase inhibition as a potential treatment option for the progression of CKD [29]. Additionally, it was discovered that even if immunosuppression treatment of nephrotic syndrome is unsuccessful, some mutations will permit treatment with a Q10 supplement [30]. Through GWAS (genome-wide association studies), it was possible to identify alterations associated with a higher risk of CKD.

For instance, it was shown that 10% to 15% of the African American population have alterations in the gene *APOL1*. This gene is associated with more than a 10-fold risk for focal segmental glomerulosclerosis or hypertensive ESKD, in combination with other genetic or environmental factors, when two risk alleles are given [31]. The proteins TNFR1 and 2, after this study, were associated with ESKD and are now considered biomarkers for the outcome of diabetes types 1 and 2.

Again, the use of -*omics*, combined with systems biology, has been useful in the identification of JAK-STAT as a drug target in diabetic kidney disease (DKD). In this case, transcriptomics of tissues from patients with DKD revealed induction of JAK-STAT signaling [32]. A clinical trial was conducted based on a cross-species transcriptome analysis of the JAK-STAT pathway with podocyte-specific overexpression of JAK2 in DKD. This trial showed a dose-dependent decrease in albuminuria after treatment with baricitinib (JAK1 and JAK2 inhibitor), indicating the potential benefit of this therapy in DKD patients [33].

The identification of prognostic markers for the differentiation of disease courses in CKD patients was possible thanks to the integration of -*omics* data. In particular, the urinary epidermal growth factor (EGF) was identified as a predictive marker for CKD by the integration of kidney biopsy transcriptome data, urinary proteome data, and clinical data [34]. The urinary EGF protein correlates with the intrarenal EGF messenger RNA and with the eGFR slope as a measure of renal function, which is a criterion for disease progression. Furthermore, urinary EGF as a potential biomarker for the loss of renal function has been validated for different subtypes of CKD by other studies [35,36].

As an additional example, the metabolomic approach has improved our understanding of uremia [37]. Specifically, all retained molecules and their toxicity when the kidney fails have been identified. The correlation and quantification of symptoms linked to retained metabolite toxicity was a challenge until then. For instance, kidney failure caused by uremic illness is affected either by environmental or genetic traits. In this disease, there is a simultaneous impact of filtration, reabsorption, secretion, and metabolism, of toxic solutes, in both blood and urine [38,39]. The identification of which of those solutes were toxic was extremely difficult to predict [40–43]. The Fig. 1 summarizes the current ISM workflow and applications in nephrology.
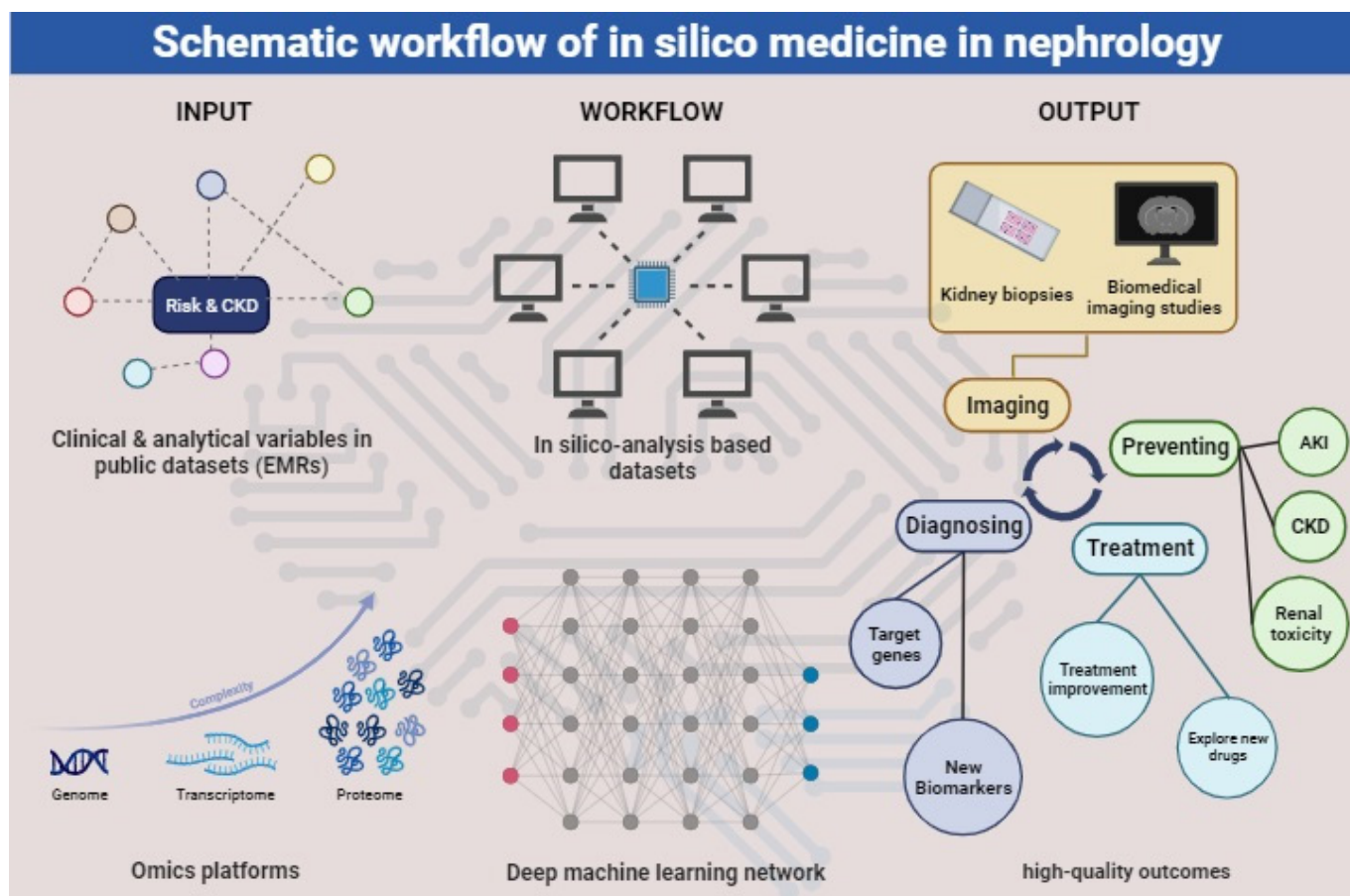
Figure 1. Schematic workflow of *in silico* medicine in nephrology. At the initial stages, the input from EMRs and other healthcare sources of public datasets can provide much clinical information related to a specific medical problem or disease; in this particular case, information related to CKD/AKI or renal-related risk factors. Also, data from the -*omics* platform, once validated, serve to evaluate patients at risk. Then, the *in silico* medicine and deep machine learning workflow collects all clinical and analytical data, to integrate the variables to create the most accurate analysis of a specific kidney disease. The final output is to combine mathematical and computational data science to provide a clear visualization of the medical problem. This output can be further analyzed dynamically, always combining descriptive, predictive, and prescriptive analytics to obtain an improvement for those patients affected.
AKI, acute kidney disease; CKD, chronic kidney disease; EMRs, electronic medical records.

Future opportunities in the nephrology field

Due to the multitude of underlying pathophysiological mechanisms in kidney diseases, there are great challenges in their staging, treatment, and prognosis. Systems biology offers the potential to complement medical research by integrating multi-*omics* data, clinical phenotype data, and pathological data mapping, to identify CKD subtypes and disease determinants. As aforementioned, linking cellular function, alterations, and interaction may improve future classification, prognostic, and more specific novel molecular therapies [29]. As CKD patients are at high risk for car-diovascular disease and ESKD and have no effective treatment to halt or reverse progression, future efforts need to focus on preventing and reversing the situation (treatment), and not merely on disease detection. The purpose of this final section is to describe and summarize those aspects in need of improvement.

First, the integration of imaging and -*omics* data is required to correlate them with biomarkers. Equally important is the establishment of standardized sampling and imaging protocols to ensure the quality and improvement of the translation of two-dimensional sections to three-dimensional constructions [21]. In this case, AI can visualize

subtle differences not distinguishable by experts and assist in the identification and prediction of results. The noninvasive diagnosis offered by AI might substitute renal biopsy, in the future [13]. ML is well suited for diagnosis and treatment using EMRs genomics and biomedical image analysis of renal pathology. The future key potential of ML lies in decision-making, as it has already been applied to the analysis of renal pathological images and diagnosis, including the prognosis of CKD and AKI. Concerning this, temporal medical history improves the prediction and identification of high-risk patients. Yet, CKD prediction models require more calibration and external validation to be incorporated into the guidelines [15,28]. ML has the potential to be widely used in the prescription, monitoring, diagnosis, and prediction of dialytic treatment in the future. However, at present, the use of ML in kidney diseases resides in its application at the initial stages of renal disease prediction, as fulfilling criteria for enough patients, eliminating demographic bias, or success in external validation [15].

Second, qualitative big data requires data standardization through consensus, data curation to maintain quality, competition among the entities, patient privacy, comparability of results, and independence from industrial influence. Integration and interoperation of consolidated and exchangeable big data is a key feature for a future comprehensive approach [4]. Since the accuracy of a model depends on reliable data reflection, improvement of funding for data collection and unified standardization for higher data quality is crucial. So far, data available are scattered, non-shared, and need mapping and pre-processing, to be used in a model. Furthermore, external validation of the model is rarely possible without enough data [13].

The third aspect refers to the identification of therapeutic targets. *Omics* has successfully discovered CKD disease biomarkers, but not clinically relevant therapeutic targets, indicating that systems biology predictive modeling requires new bioinformatics solutions. In addition, differentiation of tubular and glomerular physiology by -*omics* poses a challenge, since an abundance of molecules in the urine or blood can be caused by increased glomerular filtration barrier passage, lower tubular reabsorption, or damaged renal cells [27]. Future -*omics* approaches have to advance to combine single cell/ single nuclear profiles with spatial resolution, and small conditional RNA profile [29]. Moreover, mapping cellular interactions to disease phenotypes

in molecular models is crucial in the future to seek targets with structural information [26].

Conclusions

The ISM is an emergent data science that poses interesting promise in the field of nephrology and offers reliable computer predictions to suggest optimal treatments, in a case-specific manner. In addition, ISM has the potential to acquire a better understanding of the kidney physiology and/or pathophysiology of many complex diseases, together with a multiscale disease modeling. However, the use of this approach is limited by the low coverage of metabolomics, as there are much fewer metabolites measured than genes and proteins. This ultimately hinders comprehensive mapping and the discovery of clear dysfunction correlations.

The prediction of renal function loss is complex since there are no suitable biomarkers available in routine diagnostics to enhance the predictive power of the already-established markers, like proteinuria, albuminuria, and estimated glomerular filtration rate. Systems biology, however, has been used to identify prognostic markers for the differentiation of disease courses in CKD patients.

Finally, a current and future issue that arises from *in silico* models is that they present some impairment in the translation of molecular simulations (sub-organelle level) to physiology-based models (organ-system level). It remains hard to adapt mathematical theories to physiology and biology starting from such models. The challenge is to describe the dynamic behavior of a high-dimensional model through a lower level using a simplified dimensional model, even with the use of the most advanced technology. Multiscale modeling can help to fix this issue, by fusing models at different levels.

Conflicts of interest

All authors have no conflicts of interest to declare.

Funding

the study design, data collection, analysis, and interpretation of data, or the writing of this manuscript.

Data sharing statement

The data presented in this study are available from the corresponding author upon reasonable request.

Authors' contributions

Conceptualization, Data curation, Visualization: LDM
Funding acquisition: ACR, LDM
Investigation: NS, OJO, TMW
Supervision: VB
Writing–original draft: ACR, AL
Writing–review & editing: VB
All authors read and approved the final manuscript.

ORCID

Ana Checa-Ros, https://orcid.org/0000-0001-8958-6711
Antonella Locascio, https://orcid.org/0000-0001-8026-1937
Nelia Steib, https://orcid.org/0009-0008-0935-8208
Owahabanun-Joshua Okojie, https://orcid.org/0009-0001-6805-7738
Totte Malte-Weier, https://orcid.org/0009-0006-9105-6756
Valmore Bermúdez, https://orcid.org/0000-0003-1880-8887
Luis D'Marco, https://orcid.org/0000-0003-0148-891X

References

1. Westreich D, Pence BW, Turner AN. In populo. *Epidemiology* 2010;21:152–153.
2. Subrahmanya SV, Shetty DK, Patil V, et al. The role of data science in healthcare advancements: applications, benefits, and future prospects. *Ir J Med Sci* 2022;191:1473–1483.
3. Khanna D, Jindal N, Singh H, Rana PS. Applications and challenges in healthcare big data: a strategic review. *Curr Med Imaging* 2022;19:27–36.
4. Kaur N, Bhattacharya S, Butte AJ. Big data in nephrology. *Nat Rev Nephrol* 2021;17:676–687.
5. Gao J, Liu YY, D'Souza RM, Barabási AL. Target control of complex networks. *Nat Commun* 2014;5:5415.
6. Goh WW, Wong L. The birth of bio-data science: trends, expectations, and applications. *Genomics Proteomics Bioinformatics* 2020;18:5–15.
7. Lyu LQ, Cui HY, Shao MY, Fu Y, Zhao RX, Chen QP. Computational medicine: past, present and future. *Chin J Integr Med* 2022;28:453–462.
8. Kampmann JD, Heaf JG, Mogensen CB, Mickley H, Wolff DL, Brandt F. Prevalence and incidence of chronic kidney disease stage 3-5: results from KidDiCo. *BMC Nephrol* 2023;24:17.
9. Eckardt KU, Coresh J, Devuyst O, et al. Evolving importance of kidney disease: from subspecialty to global health burden. *Lancet* 2013;382:158–169.
10. Tuttle KR, Jones CR, Daratha KB, et al. Incidence of chronic kidney disease among adults with diabetes, 2015-2020. *N Engl J Med* 2022;387:1430–1431.
11. Bello AK, Levin A, Tonelli M, et al. Assessment of global kidney health care status. *JAMA* 2017;317:1864–1881.
12. Singh SK, Malik A, Firoz A, Jha V. CDKD: a clinical database of kidney diseases. *BMC Nephrol* 2012;13:23.
13. Li Q, Fan QL, Han QX, et al. Machine learning in nephrology: scratching the surface. *Chin Med J (Engl)* 2020;133:687–698.
14. Pappalardo F, Russo G, Tshinanu FM, Viceconti M. In silico clinical trials: concepts and early adoptions. *Brief Bioinform* 2019;20:1699–1708.
15. Chan L, Vaid A, Nadkarni GN. Applications of machine learning methods in kidney disease: hope or hype? Curr Opin Nephrol Hypertens 2020;29:319–326.
16. Shi Y, Hua Y, Wang B, Zhang R, Li X. In silico prediction and insights into the structural basis of drug induced nephrotoxicity. *Front Pharmacol* 2021;12:793332.
17. Lim DK, Boyd JH, Thomas E, et al. Prediction models used in the progression of chronic kidney disease: a scoping review. *PLoS One* 2022;17:e0271619.
18. Jaimes Campos MA, Andújar I, Keller F, et al. Prognosis and personalized in silico prediction of treatment efficacy in cardiovascular and chronic kidney disease: a proof-of-concept study. *Pharmaceuticals (Basel)* 2023;16:1298.
19. Tomašev N, Glorot X, Rae JW, et al. A clinically applicable approach to continuous prediction of future acute kidney injury. *Nature* 2019;572:116–119.

20. Koyner JL, Carey KA, Edelson DP, Churpek MM. The development of a machine learning inpatient acute kidney injury prediction model. *Crit Care Med* 2018;46:1070–1077.

21. Santo BA, Rosenberg AZ, Sarder P. Artificial intelligence driven next-generation renal histomorphometry. *Curr Opin Nephrol Hypertens* 2020;29:265–272.

22. Li B, Zhao X, Xie W, Hong Z, Zhang Y. Integrative analyses of biomarkers and pathways for diabetic nephropathy. *Front Genet* 2023;14:1128136.

23. Bruggeman FJ, Hornberg JJ, Boogerd FC, Westerhoff HV. Introduction to systems biology. In: Baginsky S, Fernie AR, eds. Plant systems biology. Experientia supplementum. Vol. 97. Birkhäuser Basel; 2007.

24. Ishrat R, Ahmed MM, Tazyeen S, et al. In silico integrative approach revealed key microRNAs and associated target genes in cardiorenal syndrome. *Bioinform Biol Insights* 2021; 15:11779322211027396.

25. Thomas S, Hao L, Ricke WA, Li L. Biomarker discovery in mass spectrometry-based urinary proteomics. *Proteomics Clin Appl* 2016;10:358–370.

26. Cisek K, Krochmal M, Klein J, Mischak H. The application of multi-omics and systems biology to identify therapeutic targets in chronic kidney disease. *Nephrol Dial Transplant* 2016;31:2003–2011.

27. Rhee EP. How omics data can be used in nephrology. *Am J Kidney Dis* 2018;72:129–135.

28. Buvall L, Menzies RI, Williams J, et al. Selecting the right therapeutic target for kidney disease. *Front Pharmacol* 2022; 13:971065.

29. Lindenmeyer MT, Alakwaa F, Rose M, Kretzler M. Perspectives in systems nephrology. *Cell Tissue Res* 2021;385:475–488.

30. Desbats MA, Lunardi G, Doimo M, Trevisson E, Salviati L. Genetic bases and clinical manifestations of coenzyme Q10 (CoQ 10) deficiency. *J Inherit Metab Dis* 2015;38:145–156.

31. Kopp JB, Smith MW, Nelson GW, et al. MYH9 is a major-effect risk gene for focal segmental glomerulosclerosis. *Nat Genet* 2008;40:1175–1184.

32. Hodgin JB, Nair V, Zhang H, et al. Identification of cross-species shared transcriptional networks of diabetic nephropathy in human and mouse glomeruli. *Diabetes* 2013;62:299–308.

33. Tuttle KR, Brosius FC, Adler SG, et al. JAK1/JAK2 inhibition by baricitinib in diabetic kidney disease: results from a Phase 2 randomized controlled clinical trial. *Nephrol Dial Transplant* 2018;33:1950–1959.

34. Ju W, Nair V, Smith S, et al. Tissue transcriptome-driven identification of epidermal growth factor as a chronic kidney disease biomarker. *Sci Transl Med* 2015;7:316ra193.

35. 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;96:214–221.

36. Segarra-Medrano A, Carnicer-Caceres C, Valtierra-Carmeno N, et al. Value of urinary levels of interleukin-6, epidermal growth factor, monocyte chemoattractant protein type1 and transforming growth factor β1 in predicting the extent of fibrosis lesions in kidney biopsies of patients with IgA nephropathy. *Nefrologia* 2017;37:531–538.

37. Van JA, Scholey JW, Konvalinka A. Insights into diabetic kidney disease using urinary proteomics and bioinformatics. *J Am Soc Nephrol* 2017;28:1050–1061.

38. Sirich TL, Funk BA, Plummer NS, Hostetter TH, Meyer TW. Prominent accumulation in hemodialysis patients of solutes normally cleared by tubular secretion. *J Am Soc Nephrol* 2014;25:615–622.

39. Zoccali C, Bode-Böger S, Mallamaci F, et al. Plasma concentration of asymmetrical dimethylarginine and mortality in patients with end-stage renal disease: a prospective study. *Lancet* 2001;358:2113–2117.

40. Kurella Tamura M, Chertow GM, Depner TA, et al. Metabolic profiling of impaired cognitive function in patients receiving dialysis. *J Am Soc Nephrol* 2016;27:3780–3787.

41. Shafi T, Powe NR, Meyer TW, et al. Trimethylamine N-oxide and cardiovascular events in hemodialysis patients. *J Am Soc Nephrol* 2017;28:321–331.

42. Tanaka H, Sirich TL, Plummer NS, Weaver DS, Meyer TW. An enlarged profile of uremic solutes. *PLoS One* 2015;10:e0135657.

43. Sato E, Kohno M, Yamamoto M, Fujisawa T, Fujiwara K, Tanaka N. Metabolomic analysis of human plasma from haemodialysis patients. *Eur J Clin Invest* 2011;41:241–255.