Artificial intelligence and perspective for rare genetic kidney diseases

OPEN

Xiaoyi Chen[1], Anita Burgun[1,2], Olivia Boyer[3,4], Bertrand Knebelmann[5] and Nicolas Garcelon[1]

[1]Université Paris Cité, Imagine Institute, Data Science Platform, INSERM UMR 1163, Paris, France; [2]Département d'Informatique Médicale, Hôpital Necker-Enfants Malades, AP-HP, Paris, France; [3]Department of Pediatric Nephrology, APHP-Centre, Reference Center for Inherited Renal Diseases (MARHEA), Imagine Institute, Hôpital Necker-Enfants Malades, Université Paris Cité, Paris, France; [4]Laboratory of Renal Hereditary Diseases, INSERM UMR 1163, Imagine Institute, Université Paris Cité, Paris, France; and [5]Nephrology and Transplantation Department, MARHEA, Hôpital Necker-Enfants Malades, AP-HP, Université Paris Cité, Paris, France

The integration of big data and artificial intelligence (AI) has revolutionized biomedicine, enhancing our understanding of diseases and health care practices. Although AI has shown remarkable success in some medical fields, its application in nephrology faces challenges because of the complex disease mechanisms and intricate physiology. These obstacles are further compounded in rare diseases, affecting <1 in 2000 people, where data scarcity and clinical complexities create additional challenges for AI in accurate disease characterization and prediction. Rare kidney diseases encompass >150 different conditions, with significant clinical and genetic heterogeneity, posing unique challenges for AI applications. Embracing AI for rare kidney diseases is essential, not only for driving the discovery of novel genes, pathways, and mechanisms relevant to both rare and common diseases, but also for shortening the diagnostic odyssey faced by patients with rare conditions, a goal regarded as the most urgent and transformative need in rare disease care. Recent reviews highlight AI applications in nephrology, focusing on big data sources, decision support systems, imaging data, multi-omics integration, and genotype-phenotype analysis. This review explores the current landscape of AI in rare genetic kidney diseases, examining key challenges and advancements in disease characterization and clinical decision support, with an emphasis on hypothesis generation using unsupervised methods and generative AI. It shows how AI can empower physicians to interpret complex data sets, identify patterns, and generate insights that can lead to improved patient outcomes and innovative medical research for rare genetic kidney conditions.

Kidney International (2025) **108,** 388–393; https://doi.org/10.1016/j.kint.2025.03.033

KEYWORDS: artificial intelligence; clinical decision support; disease characterization; hypothesis generation; rare genetic kidney disease

Correspondence: *Nicolas Garcelon, Université Paris Cité, Imagine Institute, Data Science Platform, INSERM UMR 1163, 75015 Paris, France. E-mail: nicolas.garcelon@institutimagine.org*

Received 26 July 2024; revised 14 February 2025; accepted 25 March 2025; published online 20 June 2025

Editor's Note

Rare kidney diseases, which are often caused by genetic defects, have major clinical, social, and economic consequences, with challenges in terms of phenotype variability, diagnostic odyssey, and clinical monitoring. Despite advances in gene technologies, which have yielded unprecedented progress in defining the molecular basis of rare diseases and developing model systems, targeted therapies are lacking. The capacity of artificial intelligence (AI) technologies to integrate and analyze different types of large data sets—electronic health records, multi-omics, registries, etc.—is increasingly used to accelerate diagnosis, improve clinical care, and promote drug discovery and clinical trials in rare diseases. Yet, using AI is also challenged by the rarity and diversity of the conditions, the potentially overlapping phenotypes, the lack of training, and societal, ethical, and legal issues. In this issue of *Kidney International*, Chen *et al.* review key applications of AI for rare genetic kidney diseases. They address the importance of disease characterization, including data acquisition and deep phenotyping, and illustrate the power of AI tools for clinical decision support, prognostic modeling, and precision care. Taking a broader perspective, the authors also discuss the implications of implementing AI in clinical care and translational research. See the Kidney Genomic Medicine series at https://www.kidney-international.org/content/kidney-genomic-medicine.

In the past decade, the integration of big data and artificial intelligence (AI) has revolutionized biomedicine, enhancing disease understanding and health care through vast biomedical data, growing computational capacity, and

broader acceptance by health care professionals. Nephrology is often considered as a prime example of big data usage, with a long history of systematically collecting diverse, multimodal, and multiscale data.[1] However, fully leveraging this vast amount of information with AI remains challenging because of the multifactorial nature, overlapping phenotypes, and morphologies of kidney damage, and the complexity of kidney function.[2]

Embracing AI applications is particularly challenging for rare kidney diseases because of the broad spectrum of >150 conditions, most of which are inherited,[3] involving >625 nephropathy-associated genes.[4] These diseases exhibit significant clinical variability, genetic heterogeneity, and overlapping features, often leading to phenocopies–phenotypic traits or diseases resembling those of a particular genotype without the individual carrying that genotype. For example, pathogenic variants in type IV collagen genes, typically associated with Alport syndrome, have also been found in patients clinically diagnosed with steroid-resistant nephrotic syndrome (focal segmental glomerulosclerosis),[5] and even in patients with a cystic kidney phenotype, and healthy individuals.[6] This exemplifies genotypic overlap with phenotypic variability. In an instance of phenotypic overlap, patients with nephronophthisis-related ciliopathies and those with monogenic kidney tubulopathies can both present with enhanced kidney echogenicity,[7] a finding also seen in many other genetic and nongenetic conditions. In summary, pathologic and imaging findings in genetic kidney diseases are often nonspecific. These complexities pose challenges for AI in accurately capturing disease phenotypes, identifying genotype-phenotype correlations, and predicting disease presence and outcomes.

Building on these specific challenges, rare genetic kidney diseases also share broader obstacles faced by AI in health care, especially in rare diseases. These include training models on a limited number of cases, ensuring data quality, achieving model transparency and interpretability, fostering trust in AI models through explainability, validating model performance in clinical settings, navigating data privacy regulations, addressing ethical considerations, and mitigating societal biases.

Despite these challenges, developing AI for rare genetic conditions remains crucial. AI offers tremendous potential to accelerate research by revealing novel genes, pathways, and mechanisms relevant to both rare and common diseases,[8] and improving patient care, especially by shortening the diagnostic odyssey faced by patients with rare conditions, a goal regarded as the most urgent and transformative need in rare disease care.[9] Moreover, progress in this field can also advance translational research and drug discovery with broader implications.

Recent reviews have covered various applications of AI in nephrology, focusing on topics like big data sources,[1,2] decision support systems,[10,11] imaging data for kidney pathology,[12,13] integration of multi-omics data,[14] and genotype-phenotype analysis.[15] Although these areas, particularly imaging and multi-omics, are also relevant to rare genetic kidney diseases, they are more established fields with substantial literature already available. Therefore, this review will not revisit those topics but instead focus on the emerging applications and unique challenges of AI in rare genetic kidney diseases.

This review is organized into 3 main sections. First, we discuss key challenges and advancements in disease characterization, because AI applications in rare genetic kidney diseases are still in their infancy and main efforts are focused on generating knowledge and improving understanding from multimodal patient data. Next, the clinical decision section examines how AI addresses diagnostic and prognostic challenges and explores the potential of digital twin models for precision medicine in rare kidney diseases. Finally, the discussion explores broader considerations, including regulatory impacts, explainability, and trust in AI models, ethical and security issues, societal biases, and the importance of AI literacy and training for clinicians. These sections collectively highlight both the current state of AI applications and the path forward in advancing care for rare genetic kidney diseases.

DISEASE CHARACTERIZATION

Data acquisition and preparation

The success of AI relies heavily on the availability of high-quality and comprehensive data. Nephrology is particularly well suited for some AI applications because of availability of tissue and biological samples, easy measurable function parameters, and other well-characterized clinical and laboratory data. A recent review discussed major sources of big data in nephrology.[1] Although not exclusively focused on rare conditions, this review highlighted several registries dedicated to rare kidney diseases and a dozen of general or nephrology-specific molecular data repositories, many of which are relevant to rare genetic kidney diseases. Although AI in nephrology benefits from the availability of diverse data sources, rare kidney diseases require an additional layer of collaboration and specialized data collection to overcome their inherent challenges related to the rarity of these diseases. Adhering to the FAIR (findable, accessible, interoperable, and reusable) data principles ensures that data can be effectively shared and used. Specialized centers and networks also play key roles in this effort, such as the Polycystic Kidney Disease Foundation Registry in the United States and the European Rare Kidney Disease Reference Network,[16] which both facilitate collaborative research and data sharing.

Federated learning is an alternative to centralized registries, allowing data to remain at local centers while distributing a model that learns from each center's data. An example of such federated infrastructure is given by the Medical Informatics in Research and Care in University Medicine (MIRACUM) consortium in the context of rare disease.[17] As part of the German Medical Informatics Initiative, the consortium exemplifies this approach by integrating data across 10 German university hospitals to develop a clinical decision system that

identifies similar patient cases at each MIRACUM site to offer the physician a hint to a possible diagnosis.[18]

Deep phenotyping

Phenotypic data in patient registries, biobanks, and genomic databases for rare genetic kidney disease is usually limited to a preestablished set of structured terms. Such structured entries typically provide standardized, easy-to-process data, but they may lack the granularity needed for comprehensive phenotyping, especially for rare diseases characterized by limited clinical knowledge, where many phenotypes are not yet fully documented. In contrast, electronic health records (EHRs) contain a huge amount of unstructured data collected during routine clinical care, including narrative reports that may capture early or mild signs not documented in structured data, along with negated symptoms, genetic test results with interpretations, hypotheses, and differential diagnoses. These unstructured data are invaluable for unsupervised deep phenotyping and patient representation for downstream clinical applications, particularly when enhanced by natural language processing techniques, which enable extraction of comprehensive information from narrative entries. For example, natural language processing has been shown effective in identifying symptoms in hemodialysis patients,[19] and detecting undiagnosed Fabry disease.[20] However, the accuracy and completeness of this information depend heavily on clinicians' engagement and careful, consistent data entry, making clinician involvement and training crucial for ensuring high-quality data capture.

Since the GPT-3 release in 2020, large language models (LLMs) have gained widespread attention and driven significant progress in various natural language processing tasks. A recent study showed that prompt learning with ChatGPT could match or outperform fine-tuning classic models, like BioClinicalBERT, in identifying rare diseases and signs with only 1 annotated sample.[21] LLMs have also demonstrated better performance in extracting human phenotype ontology concepts from a corpus predominantly related to neuromuscular and digestive disorders.[22] However, although these studies highly recognize the potential of LLMs, they also emphasized the need for critical evaluation by researchers and clinicians. Challenges, such as biases due to limited high-quality data access and the use of sampling techniques that may introduce variability in results across different runs, underscore the importance of carefully considering these models' outputs. Therefore, to ensure their safe and effective use, developing and integrating LLMs into rare genetic disease research and clinical practice should involve collaboration with domain experts.

Genotype-phenotype correlation

Establishing genotype-phenotype correlation is another major challenge in understanding rare genetic diseases. AI approaches can integrate diverse data sets and perform feature selection and dimension reduction techniques to identify relevant genetic variants, biomarkers, pathogenic mechanisms, and clinical parameters that contribute to disease phenotypes. For example, unsupervised machine learning techniques applied to 1 of the largest cohorts of patients with nephronophthisis revealed genetic alterations in different components of the same ciliary module shared phenotypic similarities, suggesting a ciliary module-based genotype-phenotype correlation.[23] In a study of Fabry disease, both unsupervised exploratory analysis and supervised predictive models were considered in an extensive targeted plasma proteomics study to identify its molecular profiles.[24] For a deeper exploration of bioinformatics techniques, readers can refer to a recent review focusing on the role of machine learning in genotype-phenotype analysis in kidney diseases,[15] and another review addressing the integration of AI with multiomics data in nephrology.[14]

CLINICAL DECISION SUPPORT

Diagnosis challenges

Patient-centered diagnosis support. Accurate diagnosis of an individual often requires genetic testing, which is being advanced by high-throughput sequencing techniques and AI-powered variant prioritization algorithms.[25] Tools such as GDDP[26] and PubCasefinder[27] use known associations between patient phenotype and disease phenotypes embedded in knowledge bases like human phenotype ontology, Online Mendelian Inheritance in Man, and Orphanet. Additionally, tools like Exomiser[28] incorporate a patient's human phenotype ontology–encoded phenotypes into sequencing data interpretation. However, few tools specifically address rare kidney diseases. A recent study evaluated several existing tools using real-world EHR data from patients with nephronophthisis-related ciliopathies and controls, highlighting the complexities of diagnosing these heterogeneous diseases.[29] These tools can help end diagnostic wandering but require primary physicians to maintain high suspicion for rare diseases and direct integration with the patient's records in the EHR system.

Disease-centered screening. In clinical practice, many patients experience nonspecific symptoms that do not immediately suggest a rare disorder, and the primary clinician may lack the expertise to recognize subtle manifestations of these rare conditions.[30] Meanwhile, many rare genetic kidney diseases are underdiagnosed. For example, Gitelman syndrome is speculated significantly underdiagnosed given the high numbers of incidental diagnoses.[31] Its atypical representation with chronic fatigue, low blood pressure, and electrolyte abnormalities may lead to diagnostic delays. Other underdiagnosis evidence include high rates of *NPHP1* pathogenic variants found in adults with chronic kidney disease[32] and the underestimated frequencies of pathogenic variants for Alport syndrome even in individuals not known to have kidney disease.[33] Therefore, a disease-centric phenotypic screening strategy can help identify patients with potential rare genetic kidney disease from large-scale clinical data warehouses. Several efforts have been made to identify patients at risk of renal ciliopathies using phenotypes extracted from clinical

narratives, based on either a patient similarity model[34] or supervised classification.[35] Similarity models are particularly suited for scenarios with limited data, as they do not require extensive labeled data for training. These models can help identify phenotypically similar patients from large clinical data warehouses even when only 1 diagnosed patient or an archetypal patient representing the disease is available. By focusing on key shared characteristics, similarity models provided a targeted approach. On the other hand, supervised classification methods can leverage broader patterns in the data to improve the performance by considering evidence both for and against a diagnosis. Both approaches have been evaluated in imbalanced data sets, where controls significantly outnumber cases, demonstrating their complementary roles in addressing diagnostic challenges for rare diseases.

Prognostic modeling and chronic kidney disease management

Chronic kidney disease is a common feature of many rare genetic kidney diseases. The applicability of machine learning models and regression equations for estimating glomerular filtration rate and predicting chronic kidney disease progression[36] in rare kidney disease remains uncertain because of the heterogeneity in disease progression and severity. Indeed, some diseases lead to end-stage kidney disease in childhood or early adulthood (e.g., autosomal recessive polycystic kidney disease,[37] X-linked Alport syndrome[38]), whereas others have a slower progression or milder kidney involvement (e.g., autosomal dominant Alport syndrome,[38] classic Fabry disease,[39] or Dent disease[40]). AI-driven predictive models specifically tailored for rare genetic kidney diseases are scarce. A recent retrospective analysis of the UK National Registry of Rare Kidney Diseases highlighted significant variability in kidney function decline, failure, and death across 28 rare kidney diseases,[41] emphasizing the need for predictive models incorporating disease-specific data, such as extrarenal complications and molecular data.[42] Rigorous model evaluation and clinical validation is also required to be accompanied by human supervision from disease experts.

Digital twins for precision medicine

The digital twin concept has begun to emerge in medicine, particularly for precision and individualized care.[43] The aim is to integrate physiopathologic "mechanistic" models, AI, and data-driven models with patient data to simulate unique patient-specific conditions and treatment responses. Although currently focused on digital organs, like the heart or brain, digital twins offer potential for understanding the complex pathophysiology of rare genetic kidney diseases, by continuously learning and adapting. Furthermore, digital twins could play a key role in simulating clinical trials by virtually testing treatment responses in diverse patient models, thereby informing trial design, predicting outcomes, and optimizing patient selection before real-world implementation, especially for rare disease with a limited patient population.

BROADER IMPLICATIONS AND PERSPECTIVES IN SPECIALIZED MEDICINE

Regulatory challenges

The implementation of AI in rare genetic disease must also navigate a complex regulatory environment. The European General Data Protection Regulation enforces strict data protection rules, ensuring that sensitive patient data are handled with the utmost care. Additionally, the European Union AI Act, which classifies AI systems based on risk, introduces further complexity for health care applications, as high-risk AI systems must demonstrate transparency, safety, and robustness. These regulations present both challenges and opportunities for the integration of AI in clinical practice for rare diseases.

Model explainability and trust

Explainability and trust in AI models are critical in medicine, where clinical decisions often carry significant consequences. Despite their potential, AI models can sometimes rely on unexpected or irrelevant features, leading to erroneous prediction and undermining clinician trust. For example, a model trained to identify melanoma used the presence of a ruler in the image as a deciding feature.[44] Similarly, in a methodological study evaluating innovative AI models for kidney failure prediction,[45] features such as human phenotype ontology term "stage 5 chronic kidney disease" and immunosuppressive drugs for postkidney transplantation, which are direct consequences of kidney failure, were included as predictors, artificially inflating performance and reducing reliability. These examples highlight the importance of involving domain experts in model development and validation to ensure rigorous feature selection, detailed population characterization, and comparison with traditional approaches.

Enhancing clinician engagement

AI models with visual interfaces can play a crucial role in facilitating clinician engagement. By providing interactive and interpretable insights, they help clinicians better understand and engage with the AI application. A recent study introduced an integrative tool with a visual interface that allows physicians to create cohorts of interest from EHRs, visualize patient clustering, and interpret these clusters by identifying the most discriminative features.[46] This system also enables users to test hypotheses interactively through exploring association between clusters and user-defined queries, aiding the discovery of novel insights within the data and enhancing research. Such visual tools interoperable with the EHR system can represent a critical step in creating a new dialogue between clinicians, researchers, and outside people (e.g., engineers, developers) using health data, particularly in areas like rare genetic diseases, where complex data required careful interpretation.

Ethical and equity considerations

Security and ethical considerations are paramount in health care, and even more so for rare genetic diseases, where the

risk of reidentification is more significant than in common diseases. Promoting equity in AI model development is also crucial. In nephrology, efforts have been made to remove race from glomerular filtration rate estimation to avoid underestimating risk for individuals of African descent.[47] However, more efforts are needed in rare genetic diseases, where existing knowledge is often based on studies conducted with limited diversity in population representation. It is similar for AI models where data acquisition usually focuses on specific demographics, often overlooking data from developing countries. As a result, the transferability of AI models across different demographics is limited, potentially introducing biases that affect other racial and ethnic groups. Recent studies have also showed that even LLMs can perpetuate racial and gender biases,[48,49] highlighting the need for greater ethical considerations in health care application development.

Recommendation for advancing AI for rare genetic diseases

Despite the challenges inherent in rare genetic diseases, AI offers significant potential to advance translational research and improve patient care. To fully realize this potential, the key recommendation is to integrate comprehensive knowledge bases and leverage all available data resources. Multimodal approaches, combining diverse data types, should be adopted to enhance the accuracy and robustness of AI models. Additionally, involving domain experts throughout the development process is crucial to ensure the relevance and reliability of models. Finally, rigorous evaluation and clinical validation are necessary to ensure that AI applications are both effective and applicable to real-world health care settings, particularly for rare diseases where clinical data are limited.

Empower physician with AI integration

Physicians often struggle to fully leverage the data they generate through patient interactions, diagnostics, and treatments. Despite being the primary producers of vast amounts of medical data, they are usually the last to access and leverage this information, hindering their ability to make data-driven decisions and generate hypotheses. This issue is further complicated by the challenges physicians face in integrating AI into their workflows, as many already underuse the capabilities of the existing EHR system.[50] Additionally, the AI literacy of health care providers plays a critical role in the use of information technology pipelines, ranging from the quality of information entered into EHRs, to the responsible integration of AI tools into their workflows. AI should be designed to empower physicians to enable easier access to and better interpretation of their data, which can lead to improved patient outcomes and innovative medical research.

Conclusion

AI has the potential to transform rare genetic disease research and care, but its implementation faces regulatory, technical, and ethical challenges. Collaboration among clinicians, researchers, and developers is key to building transparent, equitable, and effective AI tools.

DISCLOSURES

NG is a cofounder of codoc, a French company on data integration. All the other authors declared no competing interests.

ACKNOWLEDGMENTS

This work was supported by state funding from the Agence Nationale de la Recherche under "Investissements d'Avenir" program (ANR-10-IAHU-01), and funding from the European Union's Horizon Europe research and innovation program under grant agreement number 101080717.

REFERENCES

1. Kaur N, Bhattacharya S, Butte AJ. Big data in nephrology. *Nat Rev Nephrol*. 2021;17:676–687.
2. Saez-Rodriguez J, Rinschen MM, Floege J, Kramann R. Big science and big data in nephrology. *Kidney Int*. 2019;95:1326–1337.
3. Aymé S, Bockenhauer D, Day S, et al. Common elements in rare kidney diseases: conclusions from a Kidney Disease: Improving Global Outcomes (KDIGO) Controversies Conference. *Kidney Int*. 2017;92:796–808.
4. Groopman EE, Marasa M, Cameron-Christie S, et al. Diagnostic utility of exome sequencing for kidney disease. *N Engl J Med*. 2019;380:142–151.
5. Malone AF, Phelan PJ, Hall G, et al. Rare hereditary COL4A3/COL4A4 variants may be mistaken for familial focal segmental glomerulosclerosis. *Kidney Int*. 2014;86:1253–1259.
6. Zeni L, Mescia F, Toso D, et al. Clinical significance of the cystic phenotype in Alport syndrome. *Am J Kidney Dis*. 2024;84:320–328.e1.
7. Groopman EE, Povysil G, Goldstein DB, Gharavi AG. Rare genetic causes of complex kidney and urological diseases. *Nat Rev Nephrol*. 2020;16:641–656.
8. Devuyst O, Knoers NVAM, Remuzzi G, Schaefer F. Board of the Working Group for Inherited Kidney Diseases of the European Renal Association and European Dialysis and Transplant Association. Rare inherited kidney diseases: challenges, opportunities, and perspectives. *Lancet Lond Engl*. 2014;383:1844–1859.
9. Iskrov G, Raycheva R, Kostadinov K, et al. Are the European reference networks for rare diseases ready to embrace machine learning? a mixed-methods study. *Orphanet J Rare Dis*. 2024;19:25.
10. Loftus TJ, Shickel B, Ozrazgat-Baslanti T, et al. Artificial intelligence-enabled decision support in nephrology. *Nat Rev Nephrol*. 2022;18:452–465.
11. Garcelon N, Burgun A, Salomon R, Neuraz A. Electronic health records for the diagnosis of rare diseases. *Kidney Int*. 2020;97:676–686.
12. Barisoni L, Lafata KJ, Hewitt SM, et al. Digital pathology and computational image analysis in nephropathology. *Nat Rev Nephrol*. 2020;16:669–685.
13. Huo Y, Deng R, Liu Q, et al. AI applications in renal pathology. *Kidney Int*. 2021;99:1309–1320.
14. Zhou XJ, Zhong XH, Duan LX. Integration of artificial intelligence and multi-omics in kidney diseases. *Fundam Res*. 2023;3:126–148.
15. Sealfon RSG, Mariani LH, Kretzler M, Troyanskaya OG. Machine learning, the kidney, and genotype-phenotype analysis. *Kidney Int*. 2020;97:1141–1149.
16. Bassanese G, Wlodkowski T, Servais A, et al. The European Rare Kidney Disease Registry (ERKReg): objectives, design and initial results. *Orphanet J Rare Dis*. 2021;16:251.
17. Kapsner LA, Mang JM, Mate S, et al. Linking a consortium-wide data quality assessment tool with the MIRACUM metadata repository. *Appl Clin Inform*. 2021;12:826–835.
18. Schaaf J, Boeker M, Ganslandt T, et al. Finding the needle in the hay stack: an open architecture to support diagnosis of undiagnosed patients. *Stud Health Technol Inform*. 2019;264:1580–1581.
19. Chan L, Beers K, Yau AA, et al. Natural language processing of electronic health records is superior to billing codes to identify symptom burden in hemodialysis patients. *Kidney Int*. 2020;97:383–392.

20. Michalski AA, Lis K, Stankiewicz J, et al. Supporting the diagnosis of Fabry disease using a natural language processing-based approach. *J Clin Med*. 2023;12:3599.
21. Shyr C, Hu Y, Bastarache L, et al. Identifying and extracting rare diseases and their phenotypes with large language models. *J Healthc Inform Res*. 2024;8:438–461.
22. Groza T, Caufield H, Gration D, et al. An evaluation of GPT models for phenotype concept recognition. *BMC Med Inform Decis Mak*. 2024;24:30.
23. Petzold F, Billot K, Chen X, et al. The genetic landscape and clinical spectrum of nephronophthisis and related ciliopathies. *Kidney Int*. 2023;104:378–387.
24. Tebani A, Barbey F, Dormond O, et al. Deep next-generation proteomics and network analysis reveal systemic and tissue-specific patterns in Fabry disease. *Transl Res J Lab Clin Med*. 2023;258:47–59.
25. Stenton SL, O'Leary MC, Lemire G, et al. Critical assessment of variant prioritization methods for rare disease diagnosis within the rare genomes project. *Hum Genomics*. 2024;18:44.
26. Chen J, Xu H, Jegga A, et al. Novel phenotype–disease matching tool for rare genetic diseases. *Genet Med*. 2019;21:339–346.
27. Fujiwara T, Shin J, Yamaguchi A. Advances in the development of PubCaseFinder, including the new application programming interface and matching algorithm. *Hum Mutat*. 2022;43:734–742.
28. Smedley D, Jacobsen JOB, Jäger M, et al. Next-generation diagnostics and disease-gene discovery with the Exomiser. *Nat Protoc*. 2015;10:2004–2015.
29. Faviez C, Chen X, Garcelon N, et al. Objectivizing issues in the diagnosis of complex rare diseases: lessons learned from testing existing diagnosis support systems on ciliopathies. *BMC Med Inform Decis Mak*. 2024;24:134.
30. Faviez C, Chen X, Garcelon N, et al. Diagnosis support systems for rare diseases: a scoping review. *Orphanet J Rare Dis*. 2020;15:94.
31. Blanchard A, Bockenhauer D, Bolignano D, et al. Gitelman syndrome: consensus and guidance from a Kidney Disease: Improving Global Outcomes (KDIGO) Controversies Conference. *Kidney Int*. 2017;91:24–33.
32. Snoek R, van Setten J, Keating BJ, et al. NPHP1 (nephrocystin-1) gene deletions cause adult-onset ESRD. *J Am Soc Nephrol*. 2018;29:1772–1779.
33. Gibson J, Fieldhouse R, Chan MMY, et al. Prevalence estimates of predicted pathogenic COL4A3–COL4A5 variants in a population sequencing database and their implications for Alport syndrome. *J Am Soc Nephrol*. 2021;32:2273–2290.
34. Chen X, Faviez C, Vincent M, et al. Patient-patient similarity-based screening of a clinical data warehouse to support ciliopathy diagnosis. *Front Pharmacol*. 2022;13:786710.
35. Faviez C, Vincent M, Garcelon N, et al. Performance and clinical utility of a new supervised machine-learning pipeline in detecting rare ciliopathy patients based on deep phenotyping from electronic health records and semantic similarity. *Orphanet J Rare Dis*. 2024;19:55.
36. Ramspek CL, Evans M, Wanner C, et al. Kidney failure prediction models: a comprehensive external validation study in patients with advanced CKD. *J Am Soc Nephrol*. 2021;32:1174–1186.
37. Guay-Woodford LM, Desmond RA. Autosomal recessive polycystic kidney disease: the clinical experience in North America. *Pediatrics*. 2003;111(pt 1):1072–1080.
38. Solanki KV, Hu Y, Moore BS, et al. The phenotypic spectrum of COL4A3 heterozygotes. *Kidney Int Rep*. 2023;8:2088–2099.
39. Ortiz A, Germain DP, Desnick RJ, et al. Fabry disease revisited: management and treatment recommendations for adult patients. *Mol Genet Metab*. 2018;123:416–427.
40. Blanchard A, Curis E, Guyon-Roger T, et al. Observations of a large Dent disease cohort. *Kidney Int*. 2016;90:430–439.
41. Wong K, Pitcher D, Braddon F, et al. Effects of rare kidney diseases on kidney failure: a longitudinal analysis of the UK National Registry of Rare Kidney Diseases (RaDaR) cohort. *Lancet Lond Engl*. 2024;403:1279–1289.
42. Eddy S, Mariani LH, Kretzler M. Integrated multi-omics approaches to improve classification of chronic kidney disease. *Nat Rev Nephrol*. 2020;16:657–668.
43. Katsoulakis E, Wang Q, Wu H, et al. Digital twins for health: a scoping review. *NPJ Digit Med*. 2024;7:77.
44. Narla A, Kuprel B, Sarin K, et al. Automated classification of skin lesions: from pixels to practice. *J Invest Dermatol*. 2018;138:2108–2110.
45. Guo J, Yuan C, Shang N, et al. Similarity-based health risk prediction using domain fusion and electronic health records data. *J Biomed Inform*. 2021;116:103711.
46. Chen X, Wang J, Faviez C, et al. An integrated pipeline for phenotypic characterization, clustering and visualization of patient cohorts in a rare disease-oriented clinical data warehouse. *Stud Health Technol Inform*. 2024;316:1785–1789.
47. Delgado C, Baweja M, Ríos Burrows N, et al. Reassessing the inclusion of race in diagnosing kidney diseases: an interim report from the NKF-ASN task force. *Am J Kidney Dis*. 2021;78:103–115.
48. Zack T, Lehman E, Suzgun M, et al. Assessing the potential of GPT-4 to perpetuate racial and gender biases in health care: a model evaluation study. *Lancet Digit Health*. 2024;6:e12–e22.
49. Omiye JA, Lester JC, Spichak S, et al. Large language models propagate race-based medicine. *Npj Digit Med*. 2023;6:1–4.
50. Rumball-Smith J, Shekelle P, Damberg CL. Electronic health record "super-users" and "under-users" in ambulatory care practices. *Am J Manag Care*. 2018;24:26–31.