

# Omics Strategies and Big Data: Transforming Precision Medicine and Systems Biology

Ana Checa-Ros , Owahabanun-Joshua Okojie  and Luis D'Marco \* 

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, C/Santiago Ramón y Cajal s/n, 46115 Valencia, Spain; ana.checaros@uchceu.es (A.C.-R.); joshua.okojie1@alumnos.uchceu.es (O.-J.O.)

\* Correspondence: luis.dmarcogascon@uchceu.es

## Abstract

The convergence of multi-omics strategies and big data analytics is transforming modern healthcare by shifting medicine from a reactive discipline to a proactive, personalized system. This review explores how integrating genomics, transcriptomics, proteomics, and metabolomics provides a holistic view of complex biological systems. While current literature heavily documents theoretical models, a persistent gap remains in translating these high-dimensional architectures into validated healthcare workflows. We highlight the critical role of big data infrastructure, specifically cloud computing and artificial intelligence (AI), in processing massive datasets and we benchmark our approach against existing reviews to emphasize the path toward routine clinical deployment. Through concrete case studies in oncology and metabolic disorders, we illustrate the potential clinical utility of these technologies in advancing precision medicine. Finally, we address persistent challenges including data heterogeneity, privacy concerns, and computational bottlenecks, and discuss future directions required to translate multi-omics insights into routine clinical practice.

**Keywords:** big data; precision medicine; multi-omics; computational biology; machine learning

## 1. Introduction

Medicine is undergoing a fundamental paradigm shift from a one-size-fits-all model to an individualized, data-driven discipline known as precision medicine [1]. Historically, clinical decisions relied on generalized symptoms and population averages, which often failed to account for individual biological variability [2,3]. Today, the rise of systems biology has redefined our understanding of health and disease, viewing human physiology as a dynamic network of complex molecular interactions rather than isolated pathways [4,5]. Driving this revolution is an unprecedented explosion of biological data. The rapid drop in high-throughput sequencing costs has democratized access to genomic, transcriptomic, proteomic, and metabolomic profiling. However, capturing this vast molecular landscape generates massive datasets that present significant computational challenges [6]. True precision medicine should generate data while also extracting useful clinical insights from it.

In our review, we examine the current state of multi-omics integration, evaluate the computational infrastructure required to manage big data, and discuss the clinical applications and challenges of translating these technologies into routine patient care. The



Academic Editor: Rüdiger Pryss

Received: 25 May 2026

Revised: 7 July 2026

Accepted: 24 July 2026

Published: 1 August 2026

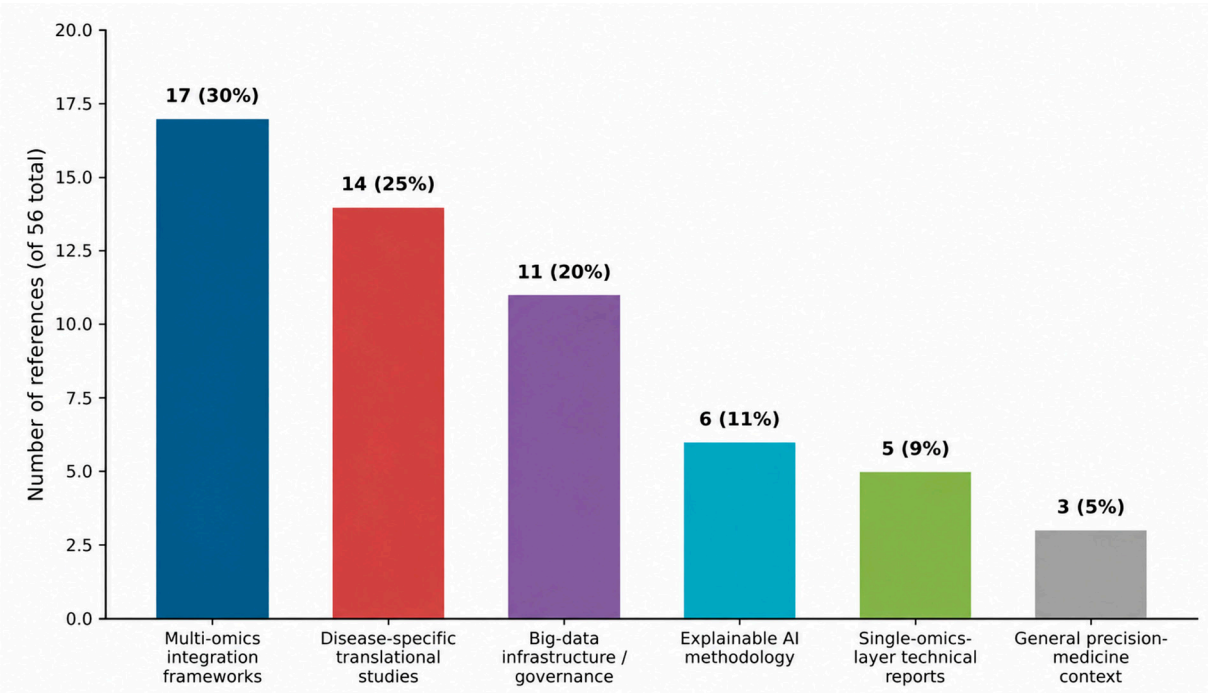
**Copyright:** © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

current biomedical literature highlights a distinct gap between computational development and clinical implementation. Single-layer omics studies dominate research, with genomics (~55%) and transcriptomics (~30%) accounting for most published datasets, and proteomics/metabolomics (~15%) accounting for a smaller volume [7,8]. Integrated multi-omics studies have grown steadily over the last five years. However, a major bottleneck remains; despite numerous papers describing theoretical multi-omics machine learning models, fewer than 5% evaluate real-world clinical cohorts (such as TCGA or iPOP), and very few systems have achieved regulatory approval (such as the FoundationOne Liquid CDx assay) [9–11]. This review focuses on this limited translational space to connect data architectures directly to validated clinical outcomes, as compared with other reviews in Table 1.

Stated explicitly, the review question addressed here is: To what extent has the multi-omics big-data literature translated into validated, bedside-deployed clinical tools, and what specific data-infrastructure and governance gaps remain? Unlike prior reviews, which either treat the computational methodology and the clinical application as separate studies (Table 1) or discuss translational gaps qualitatively without quantifying them, this review’s unique focus is to (i) quantify, across the reviewed literature, how much of the field has reached real-world clinical validation (Figure 1); and (ii) connect that quantified gap directly to the specific infrastructure, governance, and explainability requirements needed to close it.



**Figure 1.** Compositional analysis of the 56 references cited in this review, classified by primary scope. Only the 11% subset annotated in red text on the chart combines multi-omics integration with validation in real patient cohorts, illustrating the translational bottleneck addressed by the current review. [12–17].

Table 1 positions this review qualitatively against nine recent surveys. To make the added value of the present review explicit, Table 2 provides a detailed, study-by-study statement of the reported scope, remaining gap, and specific contribution of the current work relative to the six most closely related reviews.

**Table 1.** Comparison of recent literature reviews on multi-omics data integration.

| Study                            | Scope & Core Focus                                               | Analysis of Computational Data Stages | Incorporation of Data Governance Standards | Explanation of Bedside Clinical Interventions |
|----------------------------------|------------------------------------------------------------------|---------------------------------------|--------------------------------------------|-----------------------------------------------|
| Yetgin et al. [6]                | Broad overview of AI and data processing in Omics.               | Yes                                   | No                                         | No                                            |
| Martinez et al. [8]              | Computational challenges and hurdles in ML precision medicine.   | No                                    | No                                         | No                                            |
| Abbas et al. [12]                | Multi-omics computational framework for cancer subtyping.        | Yes                                   | No                                         | No                                            |
| Picard et al. [18]               | Strict mathematical classification of ML integration strategies. | Yes                                   | No                                         | No                                            |
| Yagin et al. [19]                | Medical applications of ML data integration.                     | Yes                                   | No                                         | Yes                                           |
| Jiang et al. [20]                | Network-based Omics in early drug discovery.                     | No                                    | No                                         | No                                            |
| Pallikkavaliyaveetil et al. [21] | Data-limited ML for drug discovery                               | Yes                                   | No                                         | No                                            |
| Alemu et al. [22]                | Gene-environment interactions in non-communicable diseases.      | No                                    | No                                         | Yes                                           |
| Lac et al. [23]                  | Deep learning and statistical models for multimodal omics.       | Yes                                   | No                                         | No                                            |
| Baysoy et al. [24]               | Technical landscape and applications of single-cell multi-omics. | No                                    | No                                         | No                                            |
| Current review                   | End-to-end translation from data stages to clinical application. | Yes                                   | Yes                                        | Yes                                           |

To move beyond a purely qualitative positioning statement, we quantified the composition of the 56 references cited in this review by primary scope. Approximately 30% (17/56) describe methodological or conceptual frameworks for multi-omics integration without a disease-specific application; 25% (14/56) report a disease-specific translational study; 20% (11/56) address big-data infrastructure, governance, or privacy; 11% (6/56) focus on explainable AI methodology; 9% (5/56) are single-omics-layer technical reports; and the remaining 5% (3/56) provide general precision-medicine context (Figure 1).

Critically, only 6 of these 56 references (11%) report a multi-omics analysis that was both integrated across two or more molecular layers *and* was validated against real patient cohorts or regulatory-grade evidence [12–17]; the remainder either model a single omics layer, discuss infrastructure in the abstract, or address multi-omics integration purely at the methodological level. This quantified translational gap, rather than the integration mathematics itself, is the specific bottleneck that the present review targets, and it is not explicitly measured in any of the six most closely related reviews summarized in Table 2.

**Table 2.** Added value of the current review relative to the most closely related prior reviews identified in Table 1.

| Study                                        | Primary Contribution                                                        | Reported Analytical Scope                                                                                                          | Gap Relative to Current Review                                                                                         | Added Value of Current Review                                                                                                 |
|----------------------------------------------|-----------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------|
| Checa-Ros et al. (2025) [5]                  | In silico medicine and -omics strategies in nephrology                      | Narrative, single-disease (CKD), non-quantitative                                                                                  | Concept-level; no cross-study literature quantification or infrastructure/governance comparison                        | Extends the same clinical domain into a quantified, end-to-end computational-to-bedside pipeline                              |
| Yetgin (2025) [6]                            | AI and data processing across omics layers                                  | Broad, technology-first narrative                                                                                                  | No governance or bedside-translation analysis (Table 1)                                                                | Adds the governance and bedside-translation layers absent from                                                                |
| Martinez-Garcia & Hernandez-Lemus (2022) [8] | Data-integration challenges for machine learning in precision medicine      | Computational-hurdles focus; no disease case studies                                                                               | No disease-specific case studies or quantitative synthesis of the literature                                           | Anchors the same computational hurdles to concrete oncology and CKD/T2DM case studies with a quantified evidence-gap analysis |
| Li-Pook-Than & Snyder (2013) [10]            | iPOP integrated personal omics profiling cohort                             | Single landmark cohort, perspective article, predates current cloud/AI tooling                                                     | Predates cloud-native infrastructure (GA4GH, containerized pipelines) and explainable-AI standards                     | Updates the iPOP-era vision with current cloud, governance and explainable-AI frameworks                                      |
| Li et al. (2026) [13]                        | Multi-omics and AI in diabetic kidney disease                               | Narrative, single-disease (DKD)                                                                                                    | Confined to DKD; no cross-disease comparison, infrastructure/governance analysis, or quantitative literature synthesis | Generalizes the DKD example within a broader, quantified, cross-disease (oncology plus CKD/T2DM) framework                    |
| Picard et al. (2021) [18]                    | Mathematical taxonomy of multi-omics integration strategies                 | Rigorous but purely methodological                                                                                                 | Does not connect integration mathematics to regulatory status or bedside deployment                                    | Applies the same taxonomy to regulatory-approved assays and an active clinical trial                                          |
| Jiang et al. (2025) [20]                     | Network-based multi-omics integration in drug discovery                     | Domain-restricted to drug discovery                                                                                                | Does not cover diagnostic/prognostic clinical workflows, infrastructure, or governance                                 | Broadens scope to diagnostic and prognostic bedside applications beyond drug discovery                                        |
| Current review                               | End-to-end translation from multi-omics data stages to clinical application | Qualitative comparison (Table 1) plus quantitative bibliometric synthesis (Figure 1) spanning governance and explainable-AI layers | —                                                                                                                      | —                                                                                                                             |

## 2. Multi-Omics Data Integration

While individual “omics” fields offer valuable insights, single-layer analyses inherently fail to capture the holistic complexity of human biology [25]. Genomics reveals inherited predispositions, but it cannot capture the dynamic environmental responses reflected in the transcriptome, proteome, or metabolome [16]. Multi-omics data integration bridges these layers, tracing the flow of biological information from genetic code to functional phenotype. By combining these distinct data types, researchers can move past simple correlations to map out complex associative and potential regulatory biological networks, shifting how we understand disease mechanisms.

Computationally, integrating these massive, disparate datasets requires robust mathematical frameworks capable of handling high-dimensional data. Methods generally fall into five core strategies: early, late, intermediate, model-based and multi-staged integration [18]. For the scope of this review, we focus primarily on the three main data-stage frameworks: early, late and intermediate methods. Early integration combines all data types into a single matrix before analysis, which is simple but often loses layer-specific nuances. The model can become overwhelmed by the largest data type (usually genomics or transcriptomics), while missing signals from smaller layers (like metabolomics) [18]. Late integration isolates the analysis of individual omics layers, relying on ensemble methods (e.g., averaging or voting) to combine final predictions, though it risks missing cross-layer interactions [18,19]. Intermediate integration resolves these limitations mathematically; instead of flattening raw matrices, it deploys joint matrix factorization and deep autoencoders to project distinct layers into a shared, lower-dimensional feature space [18,20]. These methods preserve the unique statistical properties of each biological tier while uncovering shared underlying features. The remaining two strategies; model-based integration (which incorporates prior biological pathways directly into machine learning models) and multi-staged integration (which performs step-wise hierarchical filtering across layers) are excluded from deep analysis here, as they represent specialized hybrid frameworks rather than pure data-stage merging strategies [18,20].

Large-scale public repositories have been instrumental in benchmarking and validating these integration frameworks. The Cancer Genome Atlas (TCGA) stands as a prime example, offering matched genomic, transcriptomic, and epigenetic profiles across thousands of tumor samples [11]. By applying multi-omics integration to TCGA data, computational biologists have successfully identified novel molecular subtypes of cancer that appear identical under a microscope but behave differently clinically [14]. This integration of diverse biological layers ultimately fuels the discovery of robust, multi-dimensional biomarkers and targeted clinical therapies.

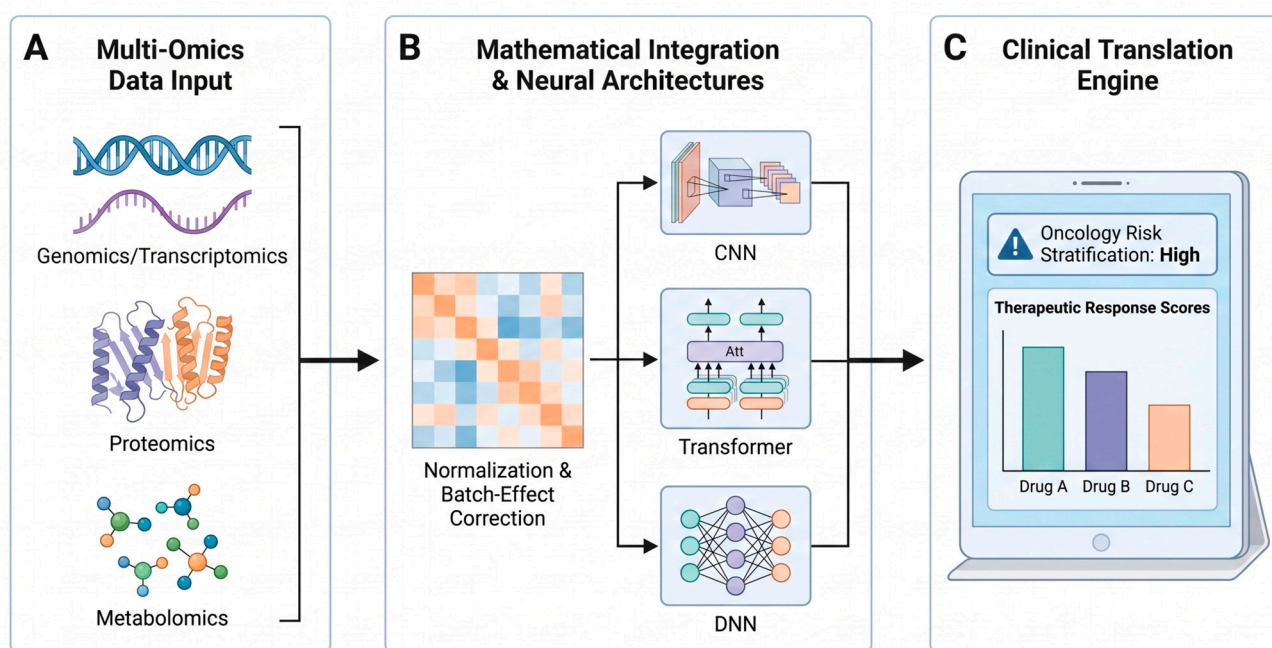
### 3. Big Data Infrastructure

Managing the sheer volume and velocity of multi-omics data requires a comprehensive overhaul of traditional computational infrastructure. A single human genome generates dozens of gigabytes of raw data [26]; when scaled across clinical cohorts and layered with transcriptomic and proteomic profiles, data storage and processing needs quickly exceed the capacity of local servers [8]. Cloud computing platforms, such as Amazon Web Services (AWS), Google Cloud Platform (GCP), and Microsoft Azure, have become indispensable in solving these bottlenecks. Crucially, the hosting landscape has begun to evolve from general purpose hosting towards purpose-built life science verticals, utilizing platforms like AWS HealthOmics for automated bioinformatics pipeline orchestration and Google Cloud's Healthcare API to ingest multimodal datasets at scale [27,28]. These platforms offer scalable storage, massive parallel processing capabilities, and secure environments that allow global research consortia to collaborate and analyze massive datasets without the financial burden of maintaining physical supercomputers [29].

Once this data is securely hosted in the cloud, artificial intelligence (AI) serves as the primary engine for pattern extraction. However, as clinical translation relies on practical data governance frameworks, modern architectures require Global Alliance for Genomics and Health (GA4GH) standards, containerized workflow languages such as Nextflow, Workflow Description Language (WDL) and git-based version control, in order to guarantee reproducible and interoperable pipelines across separate institutions [30,31]. Furthermore, standard metadata schemas preserve biological provenance, while federated learning systems enable secure, decentralized clinical data sharing without exposing raw se-

quence files [32,33]. Under this standardized framework, researchers bypass the traditional mathematical “curse of dimensionality,” where the number of biological variables vastly outnumbers the number of patients [34]. To navigate this, researchers deploy classical ML architectures like Random Forests and Support Vector Machines for robust feature selection and sample classification [5,21]. For deeper, non-linear biological networks, distinct deep learning frameworks are strictly tailored to the underlying biological data types rather than applied universally. For instance, Convolutional Neural Networks (CNNs) like *DeepVariant* process sequencing alignments reframed as image-like pileup matrices to accurately detect single nucleotide variants [35]. In structural biochemistry, Transformer-based models like AlphaFold3 process tokenized amino acid matrices and molecular graphs to predict 3D protein–ligand binding configurations [29]. Finally, for clinical prognosis, specialized Deep Neural Networks (DNNs) like *DeepSurv* integrate high-dimensional clinical omics streams with non-linear Cox proportional hazards models to forecast patient survival trajectories [36]. Figure 2 outlines the steps for using multi-omics data in clinical practice. The diagram shows raw molecular layers (A), the computational methods used for integration (B), and the resulting patient risk report (C).

### A Multi-Omics Integrative Framework for Clinical Risk Stratification



**Figure 2.** A Multi-Omics Integrative Framework for Clinical Risk Stratification: The figure outlines the sequential workflows that translate multi-omics information into clinically applicable data.

#### 4. Clinical Translation and Disease Case Studies

The clinical utility of multi-omics big data is moving past broad theoretical models to drive major advancements in clinical oncology, shifting patient care from a reactive to a proactive and precise management. A primary case study lies in the molecular stratification of Breast Invasive Carcinoma using public data repositories like The Cancer Genome Atlas (TCGA) [15]. In these workflows, computational frameworks combine matched genomic copy number variations (CNVs), transcriptomic RNA-Seq profiles, and proteomic mass spectrometry arrays derived from identical tumor tissue samples [12,15]. By applying computational methods such as Joint Matrix Factorization (JMF) and deep autoencoders,

algorithms can simultaneously cluster these distinct data layers without losing cross layer statistical relationships [23]. This multi-omics approach would improve patient stratification. While traditional microscopy classifies breast tumors based on surface receptors (PR, ER and HER2), multi-omics clustering reveals distinct molecular subtypes (such as Basal-like and Luminal A & B) within identical receptor-positive cohorts [16,37]. This allows oncologists to better predict therapeutic resistance, as Basal-like tumors (which present clinically as Triple-Negative Breast Cancer [TNBC]) commonly fail to respond to standard hormone therapies but show high sensitivity to platinum-based chemotherapy combinations, thereby improving patient survival and long-term remission outcomes [16,38]. This clinical translation is supported by real-world, high-dimensional frameworks like the FDA-approved FoundationOne Liquid CDx assay and the landmark I-PREDICT precision oncology trial [9,39].

Beyond oncology, this integrated systems approach is fundamentally reshaping the prevention and management of complex metabolic diseases such as Type 2 Diabetes Mellitus (T2DM) and Diabetic Kidney Disease (DKD). Traditional clinical tools, such as HbA1c tracking or estimated Glomerular Filtration Rate (eGFR) calculation, sometimes fail to capture early, sub-clinical metabolic and microvascular dysregulation [40]. To resolve this limitation, researchers leverage large-scale datasets such as the ‘Integrative Personal Omics Profile (iPOP)’ cohort, among others, to merge genomic polygenic risk scores with longitudinal plasma proteomics, untargeted lipidomic arrays, and urinary proteomics [10,17,41]. Computational pipelines process these high-dimensional streams using supervised machine learning classifiers, including Random Forests and Support Vector Machines, to map early pathological fingerprints [21]. The clinical utility of this multi-omics methodology lies in the ability to detect diabetic nephropathy long before symptoms such as microalbuminuria or drops in eGFR appear [13]. By flagging early changes in specific plasma and urine markers, clinicians can identify which patients face the highest risk for rapid renal decline [13]. This would enable the timely and precise administration of nephroprotective and anti-diabetic agents like sodium-glucose cotransporter 2 inhibitors (SGLT2is) and glucagon-like peptide-1 receptor agonists (GLP-1RAs), successfully altering the patient’s long-term disease course before irreversible damage occurs [42,43]. This multi-layered architecture is actively validated at scale by the ongoing NIH Nutrition for Precision Medicine landmark multi-omics cohort trial [44].

## 5. Challenges and Future Directions

Despite the immense promise of multi-omics big data, several critical limitations must be resolved before its full potential can be realized in routine clinical care. From a computational perspective, data heterogeneity and standardization remain massive hurdles [8]. Different laboratories use varying experimental protocols, sequencing depths, and extraction techniques, introducing significant “batch effects” that can skew mathematical models and lead to false-positive discoveries [45]. To resolve this, standard multi-omics pipelines must mandate uniform preprocessing and computational normalization protocols, leveraging empirical Bayes frameworks like ComBat or Surrogate Variable Analysis (SVA) to systematically eliminate non-biological batch variance [46,47]. Furthermore, integrating datasets across completely different biological modalities involves resolving massive discrepancies in data scale and noise levels [21,48]. Translating raw data files into standardized, clinically actionable reports requires universally accepted data formats, reproducible pipelines, and rigorous benchmarking standards that do not yet exist at scale.

Parallel to these technological hurdles, ethical, legal, and institutional challenges persist. High-throughput sequencing exposes deeply personal, immutable genetic information, raising substantial concerns regarding patient data privacy, informed consent, and security

against unauthorized access or genetic discrimination [49]. Facilitating the open data-sharing necessary to train robust AI models, while strictly protecting patient anonymity across international borders, requires sophisticated cryptographic frameworks. Clinically, the field faces an interpretation bottleneck. Most practicing physicians struggle to trust complex, uninterpretable “black box” multi-omics network models, as making blind high-stakes clinical interventions based on hidden node computations poses massive ethical and safety risks [50]. To make these models transparent for bedside decision-making, current developer architectures deploy Explainable AI (XAI) local feature attribution tools (such as SHAP [Shapley Additive exPlanations] or LIME [Local interpretable Model-agnostic Explanations] [51–53]). These tools convert black boxes into intuitive feature weight maps that reveal exactly which molecular layer or biomarker drove a patient’s risk calculation [51,54]. Ultimately, these XAI summaries must be translated into standardized clinical reports and embedded as automated Clinical Decision Support (CDS) alerts within Electronic Health Records (EHRs), providing a practical, transparent roadmap for clinicians in real-world practice [51,52,55].

Looking ahead, the future of precision medicine lies in capturing biological systems with higher resolution and spatial context. Emerging single-cell multi-omics technologies are revolutionizing the field by allowing researchers to analyze genomic, transcriptomic, and proteomic layers simultaneously within individual cells, exposing the hidden diversity within complex tissue microenvironments [24]. Furthermore, spatial transcriptomics adds a crucial structural dimension, mapping exactly where these molecular activities occur within the physical architecture of a tissue [56]. By pairing these high-resolution single-cell techniques with advanced cloud computing and explainable AI models, the next generation of systems biology will move past broad tissue averages to deliver accurate, localized, and dynamically responsive patient care.

## 6. Conclusions

The convergence of multi-omics data integration, cloud-scalable infrastructure, and advanced artificial intelligence represents a monumental shift toward the realization of true precision medicine. By moving beyond single-variable methodologies, these computational frameworks and deep learning models support the ongoing translation of massive biological datasets into tangible clinical applications, such as multidimensional biomarker discovery and proactive metabolic risk stratification. However, a cautious approach remains necessary as these technologies transition from the research laboratory to the patient’s bedside. The path forward is still limited by systemic challenges in data standardization, batch effects, clinical interpretation, and patient data privacy. For multi-omics to successfully guide routine diagnostic and therapeutic decisions, the medical and computational communities must collaboratively develop standardized, reproducible pipelines, secure ethical sharing frameworks, and explainable AI tools that clinicians can easily interpret and act upon.

**Author Contributions:** Conceptualization, L.D. and A.C.-R.; methodology, L.D., A.C.-R. and O.-J.O.; writing—original draft, L.D. and A.C.-R.; writing—review and editing, A.C.-R. and O.-J.O. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** No new data were created or analyzed in this study.

**Acknowledgments:** During the preparation of this manuscript, the authors used ChatGPT (v5.5 OpenAI) to improve the academic quality and fluency of the English language. After using this AI tool, the authors carefully reviewed and edited the content as needed and assume full responsibility for the final version of the manuscript.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Stefanicka-Wojtas, D.; Kurpas, D. Personalised Medicine—Implementation to the Healthcare System in Europe (Focus Group Discussions). *J. Pers. Med.* **2023**, *13*, 380. [\[CrossRef\]](#) [\[PubMed\]](#)
2. Coskun, A. Diagnosis Based on Population Data versus Personalized Data: The Evolving Paradigm in Laboratory Medicine. *Diagnostics* **2024**, *14*, 2135. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Prosperi, M.; Min, J.S.; Bian, J.; Modave, F. Big data hurdles in precision medicine and precision public health. *BMC Med. Inform. Decis. Mak.* **2018**, *18*, 139. [\[CrossRef\]](#) [\[PubMed\]](#)
4. David, L.; Gourraud, P.-A.; Lamirault, G.; Lemarchand, P. Rethinking medical education through systems biology to address complexity. *npj Syst. Biol. Appl.* **2026**, *12*, 12. [\[CrossRef\]](#) [\[PubMed\]](#)
5. Checa-Ros, A.; Locascio, A.; Steib, N.; Okojie, O.-J.; Malte-Weier, T.; Bermúdez, V.; D’Marco, L. In silico medicine and -omics strategies in nephrology: Contributions and relevance to the diagnosis and prevention of chronic kidney disease. *Kidney Res. Clin. Pract.* **2025**, *44*, 49–57. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Yetgin, A. Revolutionizing multi-omics analysis with artificial intelligence and data processing. *Quant. Biol.* **2025**, *13*, e70002. [\[CrossRef\]](#) [\[PubMed\]](#)
7. Hasin, Y.; Seldin, M.; Lusi, A. Multi-omics approaches to disease. *Genome Biol.* **2017**, *18*, 83. [\[CrossRef\]](#) [\[PubMed\]](#)
8. Martínez-García, M.; Hernández-Lemus, E. Data Integration Challenges for Machine Learning in Precision Medicine. *Front. Med.* **2022**, *8*, 784455. [\[CrossRef\]](#) [\[PubMed\]](#)
9. Woodhouse, R.; Li, M.; Hughes, J.; Delfosse, D.; Skoletsky, J.; Ma, P.; Meng, W.; Dewal, N.; Milbury, C.; Clark, T.; et al. Clinical and analytical validation of FoundationOne Liquid CDx, a novel 324-Gene cfDNA-based comprehensive genomic profiling assay for cancers of solid tumor origin. *PLoS ONE* **2020**, *15*, e0237802. [\[CrossRef\]](#) [\[PubMed\]](#)
10. Li-Pook-Than, J.; Snyder, M. iPOP goes the world: Integrated personalized Omics profiling and the road toward improved health care. *Chem. Biol.* **2013**, *20*, 660–666. [\[CrossRef\]](#) [\[PubMed\]](#)
11. Tomczak, K.; Czerwińska, P.; Wiznerowicz, M. The Cancer Genome Atlas (TCGA): An immeasurable source of knowledge. *Contemp. Oncol.* **2015**, *19*, A68–A77. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Abbasi, A.F.; Sajjad, M.; Asim, M.N.; Vollmer, S.; Dengel, A. Multi-omics driven computational framework for cancer molecular subtype classification. *Sci. Rep.* **2025**, *15*, 44141. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Li, T.; Chen, K.; Sun, Y.; Zhang, L. Diabetic kidney disease: Integrating multi-omics insights, artificial intelligence, and novel therapeutics for precision medicine. *Front. Genet.* **2026**, *17*, 1760654. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Ge, J.; Cai, J.; Zhang, G.; Li, D.; Tao, L. Multi-omics integration and machine learning uncover molecular basal-like subtype of pancreatic cancer and implicate A2ML1 in promoting tumor epithelial-mesenchymal transition. *J. Transl. Med.* **2025**, *23*, 741. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Thennavan, A.; Beca, F.; Xia, Y.; Garcia-Recio, S.; Allison, K.; Collins, L.C.; Tse, G.M.; Chen, Y.Y.; Schnitt, S.J.; Hoadley, K.A.; et al. Molecular analysis of TCGA breast cancer histologic types. *Cell Genom.* **2021**, *1*, 100067. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Sharma, A.; Debik, J.; Naume, B.; Ohnstad, H.O.; Oslo Breast Cancer Consortium (OSBREAC); Bathen, T.F.; Giskeødegård, G.F. Comprehensive multi-omics analysis of breast cancer reveals distinct long-term prognostic subtypes. *Oncogenesis* **2024**, *13*, 22. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Woerner, J.; Westbrook, T.M.; Joo, J.; Shivakumar, M.; Venkatesh, R.; Cherlin, T.; Jung, S.-H.; Jeong, S.; Maseda, D.; McKeague, M.; et al. Large-scale evaluation of proteomic and polygenic risk scores reveals complementary contributions to incident disease prediction. *bioRxiv* **2025**. [\[CrossRef\]](#) [\[PubMed\]](#)
18. Picard, M.; Scott-Boyer, M.-P.; Bodein, A.; Périn, O.; Droit, A. Integration strategies of multi-omics data for machine learning analysis. *Comput. Struct. Biotechnol. J.* **2021**, *19*, 3735–3746. [\[CrossRef\]](#) [\[PubMed\]](#)
19. Yagin, F.H. Machine Learning Approaches for Multi-omics Data Integration in Medicine. In *Machine Learning Methods for Multi-Omics Data Integration*; Springer International Publishing: Cham, Switzerland, 2024; pp. 23–38.
20. Jiang, W.; Ye, W.; Tan, X.; Bao, Y.-J. Network-based multi-omics integrative analysis methods in drug discovery: A systematic review. *BioData Min.* **2025**, *18*, 27. [\[CrossRef\]](#) [\[PubMed\]](#)
21. Pallikkavaliyaveetil, N.; Chandrasekaran, S. Small data, big challenges: Machine- and deep-learning strategies for data-limited drug discovery. *Adv. Drug Deliv. Rev.* **2026**, *229*, 115762. [\[CrossRef\]](#) [\[PubMed\]](#)

22. Alemu, R.; Sharew, N.T.; Arsano, Y.Y.; Ahmed, M.; Tekola-Ayele, F.; Mersha, T.B.; Amare, A.T. Multi-omics approaches for understanding gene-environment interactions in noncommunicable diseases: Techniques, translation, and equity issues. *Hum. Genom.* **2025**, *19*, 8. [CrossRef] [PubMed]
23. Lac, L.; Leung, C.K.; Hu, P. Computational frameworks integrating deep learning and statistical models in mining multimodal omics data. *J. Biomed. Inform.* **2024**, *152*, 104629. [CrossRef] [PubMed]
24. Baysoy, A.; Bai, Z.; Satija, R.; Fan, R. The technological landscape and applications of single-cell multi-omics. *Nat. Rev. Mol. Cell Biol.* **2023**, *24*, 695–713. [CrossRef] [PubMed]
25. Fan, B.-L.; Chen, L.-H.; Chen, L.-L.; Guo, H. Integrative Multi-Omics Approaches for Identifying and Characterizing Biological Elements in Crop Traits: Current Progress and Future Prospects. *Int. J. Mol. Sci.* **2025**, *26*, 1466. [CrossRef] [PubMed]
26. Logsdon, G.A.; Vollger, M.R.; Eichler, E.E. Long-read human genome sequencing and its applications. *Nat. Rev. Genet.* **2020**, *21*, 597–614. [CrossRef] [PubMed]
27. Greene, R. Exact Sciences Transforms Bioinformatics Infrastructure with AWS HealthOmics. AWS for Industries. Available online: <https://aws.amazon.com/blogs/industries/exact-sciences-transforms-bioinformatics-infrastructure-with-aws-healthomics/> (accessed on 17 June 2026).
28. Von See, C. Getting to Know the Google Cloud Healthcare API: Part 1. Google Cloud Blog. Available online: <https://cloud.google.com/blog/topics/healthcare-life-sciences/getting-to-know-the-google-cloud-healthcare-api-part-1> (accessed on 17 June 2026).
29. Alkhatib, A.; Shaheen, A.; Albustanji, R.N. A Comparative Analysis of Cloud Computing Services: AWS, Azure, and GCP. *Int. J. Comput. Digit. Syst.* **2025**, *18*, 1–15. [CrossRef]
30. Rehm, H.L.; Page, A.J.H.; Smith, L.; Adams, J.B.; Alterovitz, G.; Babb, L.J.; Barkley, M.P.; Baudis, M.; Beauvais, M.J.S.; Beck, T.; et al. GA4GH: International policies and standards for data sharing across genomic research and healthcare. *Cell Genom.* **2021**, *1*, 100029. [CrossRef] [PubMed]
31. Berke, S.R.; Kanchan, K.; Marazita, M.L.; Tobin, E.; Ruczinski, I. Fundamentals of FAIR biomedical data analyses in the cloud using custom pipelines. *PLoS Comput. Biol.* **2025**, *21*, e1013215. [CrossRef] [PubMed]
32. Harshith, M.; Ansari, Z.A.; Fatima, S.; Siddiqui, S.; Swarna, S.; Reddy, D.R.N.; Mohsin, S.W. Federated microservices architecture with blockchain for privacy-preserving and scalable healthcare analytics. *Sci. Rep.* **2026**, *16*, 13928. [CrossRef] [PubMed]
33. Li, R.; Romano, J.D.; Chen, Y.; Moore, J.H. Centralized and Federated Models for the Analysis of Clinical Data. *Annu. Rev. Biomed. Data Sci.* **2024**, *7*, 179–199. [CrossRef] [PubMed]
34. Arabshahi, S.; Davoodi Karsalari, P.; Baghsheikhi, H.; Mogharrabi, F.; Samieefar, N.; Rezaei, N. The Use of Machine Learning Algorithms in the Identification of Novel Biomarkers for Disorders of the Immune System. *Adv. Biomark. Sci. Technol.* **2026**, *8*, 509–524. [CrossRef]
35. Cai, L.; Wu, Y.; Gao, J. DeepSV: Accurate calling of genomic deletions from high-throughput sequencing data using deep convolutional neural network. *BMC Bioinform.* **2019**, *20*, 665. [CrossRef] [PubMed]
36. Katzman, J.L.; Shaham, U.; Cloninger, A.; Bates, J.; Jiang, T.; Kluger, Y. DeepSurv: Personalized treatment recommender system using a Cox proportional hazards deep neural network. *BMC Med. Res. Methodol.* **2018**, *18*, 24. [CrossRef] [PubMed]
37. Cho, N. Imaging features of breast cancer molecular subtypes: State of the art. *J. Pathol. Transl. Med.* **2021**, *55*, 16–25. [CrossRef] [PubMed]
38. Greco, S.; Fabbri, N.; Spaggiari, R.; De Giorgi, A.; Fabbian, F.; Giovine, A. Update on Classic and Novel Approaches in Metastatic Triple-Negative Breast Cancer Treatment: A Comprehensive Review. *Biomedicines* **2023**, *11*, 1772. [CrossRef] [PubMed]
39. Sicklick, J.K.; Nishizaki, D.; Miyashita, H.; Okamura, R.; Hahn, M.E.; Nikanjam, M.; Fanta, P.T.; Piccioni, D.E.; Patel, H.; Eskander, R.N.; et al. Investigation of Profile-Related Evidence Determining Individualized Cancer Therapy (I-PREDICT) N-of-1 Precision Oncology Study: Molecular Profiling to Match Individually Dosed, Personalized Drug Combinations. *J. Clin. Oncol.* **2026**, *44*, 540–552. [CrossRef] [PubMed]
40. Lee, C.-L.; Chen, C.-H.; Wu, M.-J.; Tsai, S.-F. The variability of glycated hemoglobin is associated with renal function decline in patients with type 2 diabetes. *Ther. Adv. Chronic Dis.* **2020**, *11*, 2040622319898370. [CrossRef] [PubMed]
41. Mardinoglu, A.; Turkez, H.; Shong, M.; Srinivasulu, V.P.; Nielsen, J.; Palsson, B.O.; Hood, L.; Uhlen, M. Longitudinal big biological data in the AI era. *Mol. Syst. Biol.* **2025**, *21*, 1147–1165. [CrossRef] [PubMed]
42. Checa-Ros, A.; Okojie, O.-J.; D'Marco, L. SGLT2 Inhibitors: Multifaceted Therapeutic Agents in Cardiometabolic and Renal Diseases. *Metabolites* **2025**, *15*, 536. [CrossRef] [PubMed]
43. Checa-Ros, A.; Okojie, O.J.; Wassouf, J.G.; Yedean, A.; Hsueh, W.-C.; Hebda, P.; Llobell, E.R.; Muhmenthaler, G.B.; Tran, M.D.-D.; D'Marco, L. Glucagon-like Peptide Receptor Agonists and Kidney Outcomes in the Era of Personalized Medicine: Focus on Albuminuria. *J. Pers. Med.* **2026**, *16*, 97. [CrossRef] [PubMed]
44. Nourazarain, A.; Vaziri, Y. Nutrigenomics meets multi-omics: Integrating genetic, metabolic, and microbiome data for personalized nutrition strategies. *Genes Nutr.* **2025**, *20*, 30. [CrossRef] [PubMed]

45. Nuanpirom, J.; Charoensawan, V. Multimodal bioinformatic analyses of genome-scale expression beyond gene-centric differential expression. *Brief. Bioinform.* **2026**, *27*, bbag152. [[CrossRef](#)] [[PubMed](#)]
46. Vo, T.H.; Le, N.Q.K. Toward trustworthy artificial intelligence in multi-omics: A review of reproducibility, stability, and interpretability. *Brief. Bioinform.* **2026**, *27*, bbag227. [[CrossRef](#)] [[PubMed](#)]
47. Wang, J. ComBat-met: Adjusting batch effects in DNA methylation data. *NAR Genom. Bioinform.* **2025**, *7*, lqaf062. [[CrossRef](#)] [[PubMed](#)]
48. Angermueller, C.; Pärnamaa, T.; Parts, L.; Stegle, O. Deep learning for computational biology. *Mol. Syst. Biol.* **2016**, *12*, MSB156651. [[CrossRef](#)] [[PubMed](#)]
49. Oliva, A.; Kaphle, A.; Reguant, R.; Sng, L.M.F.; Twine, N.A.; Malakar, Y.; Wickramarachchi, A.; Keller, M.; Ranbaduge, T.; Chan, E.K.F.; et al. Future-proofing genomic data and consent management: A comprehensive review of technology innovations. *Gigascience* **2024**, *13*, giae021. [[CrossRef](#)] [[PubMed](#)]
50. Rasheed, K.; Qayyum, A.; Ghaly, M.; Al-Fuqaha, A.; Razi, A.; Qadir, J. Explainable, trustworthy, and ethical machine learning for healthcare: A survey. *Comput. Biol. Med.* **2022**, *149*, 106043. [[CrossRef](#)] [[PubMed](#)]
51. Zhang, K.; Wang, D.; Lin, F.; Xie, J.; Zhou, W. A comprehensive review of explainable artificial intelligence in healthcare methods, evaluation, and clinical integration. *iScience* **2026**, *29*, 115026. [[CrossRef](#)] [[PubMed](#)]
52. Nohara, Y.; Matsumoto, K.; Soejima, H.; Nakashima, N. Explanation of machine learning models using shapley additive explanation and application for real data in hospital. *Comput. Methods Programs Biomed.* **2022**, *214*, 106584. [[CrossRef](#)] [[PubMed](#)]
53. Shin, J. Feasibility of local interpretable model-agnostic explanations (LIME) algorithm as an effective and interpretable feature selection method: Comparative fNIRS study. *Biomed. Eng. Lett.* **2023**, *13*, 689–703. [[CrossRef](#)] [[PubMed](#)]
54. Salih, A.; Raisi-Estabragh, Z.; Galazzo, I.B.; Radeva, P.; Petersen, S.E.; Menegaz, G.; Lekadir, K. A Perspective on Explainable Artificial Intelligence Methods: SHAP and LIME. *Adv. Intell. Syst.* **2024**, *6*, 2400304. [[CrossRef](#)]
55. Fernandes Prabhu, D.; Gurupur, V.; Stone, A.; Trader, E. Integrating Artificial Intelligence, Electronic Health Records, and Wearables for Predictive, Patient-Centered Decision Support in Healthcare. *Healthcare* **2025**, *13*, 2753. [[CrossRef](#)] [[PubMed](#)]
56. Rao, A.; Barkley, D.; França, G.S.; Yanai, I. Exploring tissue architecture using spatial transcriptomics. *Nature* **2021**, *596*, 211–220. [[CrossRef](#)] [[PubMed](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.