Health Science Reports

WILEY

Health Science Reports
Open Access

SYSTEMATIC REVIEW OPEN ACCESS

Multi-Omics Biomarkers From Variant to Clinic: A Systematic Review and Meta-Analysis of Evidence, AI/ML, Governance, Equity, and Real-World Implementation Across Global Health Systems

Neelam Das

Department of Periodontology, Sri Sai College of Dental Surgery, Vikarabad, Telangana, India

Correspondence: Neelam Das (dasneelam423@gmail.com)

Received: 20 January 2026 | **Revised:** 30 March 2026 | **Accepted:** 16 April 2026

Keywords: artificial intelligence | biomarkers | clinical utility | machine learning | metabolomics | multi-omics | proteogenomics | single-cell sequencing | spatial transcriptomics

ABSTRACT

Background/Purpose: Multi-omics integration linking genomic, transcriptomic, epigenomic, proteomic, metabolomic, single-cell, and spatial data has transformed the interpretation of human genetic variation by capturing molecular processes that extend beyond DNA sequence alone. Although these approaches substantially improve biomarker discovery and disease stratification, translation into clinical practice remains uneven due to methodological heterogeneity, limited validation, regulatory uncertainty, and structural inequities in data generation. This systematic review and meta-analysis aimed to evaluate scientific performance, clinical readiness, governance frameworks, and socio-technical constraints influencing multi-omics biomarker development, and to generate a roadmap for equitable global implementation.

Methods: Following PRISMA 2020 guidelines, we systematically searched PubMed, EMBASE, Web of Science, Scopus, medRxiv, and bioRxiv for studies published between January 2010 and December 2025. Eligible articles integrated ≥ 2 omics modalities, applied AI/ML to biomarker development or variant interpretation, assessed clinical utility or real-world implementation, or examined governance, ethics, consent, equity, or policy issues. Data extraction captured assay type, integration strategy, model performance, validation rigor, and regulatory or socio-technical insights. Random-effects meta-analyses estimated pooled improvements in AUC, sensitivity, specificity, and hazard ratio precision, and heterogeneity was assessed using I^2 statistics.

Results: From 9846 records, 528 studies met the inclusion criteria. Multi-omics integration improved predictive performance, yielding pooled gains of +0.16 in AUC (95% CI: 0.11–0.19), +13% in sensitivity, and +9% in specificity. Models combining ≥ 3 omics layers showed the largest improvements (+0.19 AUC). Single-cell and spatial assays enhanced risk stratification by 18% but demonstrated reproducibility limitations. AI/ML approaches added +0.12 AUC over traditional models, yet 67% exhibited ancestry bias, and only 22% implemented explainability tools. Only 19% of biomarkers underwent real-world evaluation due to limited validation, reimbursement gaps, interoperability challenges, and unclear data-rights governance.

Conclusion: Multi-omics biomarkers offer substantial analytical advantages, but their translation requires standardized validation frameworks, accountable AI governance, interoperable infrastructure, and globally inclusive data sets to ensure equitable, trustworthy implementation.

1 | Introduction

Advances in high-throughput sequencing, molecular profiling technologies, and computational modeling have transformed the study of human genetic variation, enabling detailed characterization of how molecular changes influence disease mechanisms and clinical phenotypes [1–3]. While genomic sequencing remains central to variant interpretation, the functional consequences of many variants depend on regulatory, epigenetic, proteomic, metabolic, and cellular contexts that extend beyond DNA sequence alone. Multi-omics integration—linking genomic, transcriptomic, epigenomic, proteomic, metabolomic, single-cell, and spatial signatures—offers a comprehensive framework for capturing these layered biological interactions and improving biomarker discovery across diverse diseases [4–6]. Biomarkers in this context include diagnostic, prognostic, and predictive categories, each serving distinct roles in disease detection, risk stratification, and therapeutic decision-making.

Despite substantial scientific progress, translation of multi-omics biomarkers into routine clinical practice remains limited. Studies differ widely in analytical methods, validation rigor, and reproducibility across populations and health systems [7]. Artificial intelligence (AI) and machine learning (ML) approaches have expanded the ability to analyze high-dimensional omics data, yet their performance is strongly influenced by data set quality, ancestral diversity, and model transparency [8, 9]. Concerns have also emerged about algorithmic bias, overfitting, and limited explainability, which may hinder clinical acceptance and raise ethical questions about fairness and accountability [10].

Implementation challenges extend beyond analytic performance. Many health systems lack clear regulatory pathways, sustainable reimbursement mechanisms, and interoperable workflows capable of supporting multi-omics integration at scale. Workforce shortages in bioinformatics, molecular diagnostics, and data governance further constrain adoption. Moreover, persistent inequities in data set representation and differing national approaches to consent, privacy, and data rights continue to generate uncertainty and limit public trust [11, 12].

As multi-omics platforms and AI-enabled analytics continue to expand, an integrated synthesis of scientific evidence, clinical readiness, governance structures, and societal implications is essential. Existing reviews often focus on isolated technologies or single-disease contexts and do not address the broader socio-technical factors that determine whether biomarkers progress from discovery to meaningful clinical impact. The aim of this systematic review and meta-analysis is to generate an actionable roadmap for responsible, equitable adoption of multi-omics biomarkers worldwide, providing a foundation for translating molecular insights into accessible and trustworthy clinical applications.

2 | Methods

This systematic review and meta-analysis were conducted in accordance with the PRISMA 2020 guidelines [13], ensuring transparent reporting and reproducibility. The protocol was designed to integrate evidence from multi-omics discovery studies, AI/ML-supported biomarker development, clinical utility assessments, and research addressing governance, ethics, and equity in the context of multi-omics translation.

2.1 | Search Strategy

A comprehensive and sensitive search strategy was developed to identify relevant literature published between January 2010 and December 2025. This period reflects the emergence and maturation of multi-omics platforms, widespread adoption of NGS technologies, and increasing integration of AI/ML methodologies in biomarker research.

Electronic searches were conducted across PubMed, EMBASE, Web of Science, and Scopus, as these databases collectively index biomedical, computational, translational, and health policy research. To capture in-progress and emerging work, preprint repositories medRxiv and bioRxiv were also included. Search terms were derived using a combination of Medical Subject Headings (MeSH), controlled vocabulary, and free-text entries.

2.2 | Keywords and Boolean Operators Included

"multi-omics" OR "proteogenomics" OR "metabolomics" OR "epigenomics" OR "single-cell biomarker" OR "spatial transcriptomics" AND "AI genomics" OR "machine learning biomarker" AND "variant interpretation" AND "clinical utility" OR "real-world evidence" AND "governance" OR "equity" OR "implementation science."

Search strings were iteratively refined to maximize sensitivity while minimizing irrelevant retrievals. Reference lists of included studies and recent systematic reviews were manually screened to ensure comprehensive coverage. Preprint repositories medRxiv and bioRxiv were included to capture emerging evidence in rapidly evolving areas such as multi-omics integration and AI-driven biomarker development. As these studies have not undergone peer review, potential variability in methodological quality and reporting standards was considered during interpretation.

2.3 | Inclusion Criteria

2.3.1 | Studies Were Eligible If They Met All the Following Criteria

They reported empirical data on multi-omics biomarker discovery, validation, or translation, integrating two or more omics modalities (e.g., genomics–proteomics, transcriptomics–metabolomics, and single-cell–epigenomics). Eligible studies also included those applying AI or ML methods to biomarker development, feature extraction, variant classification, or multi-omics integration. AI/ML approaches were defined as computational models beyond traditional statistical regression, including supervised and unsupervised ML algorithms, such as random forest, support vector machines, gradient boosting, and deep learning architectures. Studies using only conventional statistical models without model training or validation frameworks were not classified as AI/ML.

Additionally, studies were included if they evaluated clinical utility or real-world implementation, defined as testing within clinical workflows, prospective or retrospective clinical cohorts, pragmatic trials, or integration into clinical decision-support systems. Only studies involving human participants, human-derived samples, or human clinical data sets were included.

2.4 | Exclusion Criteria

Studies were excluded if they focused solely on single-omics analyses without integration; presented purely computational modeling without empirical validation; involved nonhuman organisms; lacked primary data (e.g., editorials and expert opinions); or were published in languages other than English. Conference abstracts without full-text availability were excluded unless sufficient methodological detail was accessible.

2.5 | Data Extraction

Data extraction was performed using a structured template tailored for multi-omics translational research. Extracted variables included study design, sample size, disease context, population characteristics, omics modalities used, data preprocessing steps, integration strategies, and computational or statistical methods employed.

For performance evaluation studies, we extracted quantitative metrics such as AUC, sensitivity, specificity, accuracy, F1 score, and prognostic hazard ratios. For implementation or governance studies, extracted information included regulatory context, consent models, data-rights structures, algorithmic accountability measures, benefit-sharing frameworks, and equity considerations.

Each study's validation strategy—internal validation, cross-validation, external validation, or prospective validation—was recorded to assess methodological rigor. Extraction procedures were aligned with established standards for biomarker evaluation and AI-driven predictive modeling [14]. Discrepancies between reviewers were resolved through consensus.

2.6 | Quality Assessment

A multi-framework approach was used to assess methodological quality. Prediction model studies, including AI/ML-driven models, were evaluated using the PROBAST tool [15], which examines risk of bias across participant selection, predictors, outcomes, and analytic methods.

Observational studies were appraised via the Newcastle–Ottawa Scale [16], assessing selection, comparability, and outcome domains. Real-world evidence studies were evaluated using ROBINS-I [17], focusing on confounding, selection bias, measurement bias, and reporting deviations.

Qualitative and policy-oriented research was assessed using the CASP qualitative checklist [18], examining study design appropriateness, analytic transparency, and relevance to governance and implementation questions. Quality scores were not used to exclude studies but informed the interpretation of pooled findings.

2.7 | Statistical Analysis

Quantitative synthesis was conducted using a random-effects meta-analysis, appropriate for the anticipated heterogeneity in study designs, populations, assays, computational approaches, and validation strategies. Pooled effect estimates were generated for key predictive performance metrics, including AUC improvement, sensitivity, specificity, and hazard ratio precision.

Comparative analyses were conducted to evaluate the incremental benefit of AI/ML approaches over traditional statistical models. Subgroup analyses were planned for different disease categories, omics combinations, and validation types when sufficient data were available.

Statistical heterogeneity was quantified using the I^2 statistic, with values interpreted as low ($< 25\%$), moderate (25%–50%), and high ($> 50\%$) heterogeneity. Subgroup analyses were conducted based on disease category, number of omics layers, and validation type to explore sources of heterogeneity. Publication bias was examined using funnel plots and sensitivity analyses, consistent with recommended best practices for biomarker meta-analyses [19]. All statistical analyses were performed using standard meta-analysis software. Sensitivity considerations were applied during interpretation to account for potential differences between peer-reviewed and preprint studies.

3 | Results

3.1 | Study Selection

The systematic search conducted between January 2010 and December 2025 identified 9846 records. After removal of duplicates, 7934 unique articles underwent title and abstract screening. Of these, 612 full-text articles were assessed for eligibility, and 528 studies met all inclusion criteria and were included in the final qualitative and quantitative synthesis. The study selection process and reasons for exclusion at each stage were summarized in the PRISMA 2020 flow diagram (Figure 1).

The included studies demonstrated substantial methodological and geographic diversity, reflecting the rapid expansion of multi-omics research globally. As summarized in Table 1, the final data set comprised 216 multi-omics biomarker discovery and validation studies, 112 AI/ML-driven multi-omics investigations, 124 studies evaluating clinical utility or real-world implementation, and 76 studies addressing governance, ethics, equity, or regulatory considerations. A marked increase in publication volume was observed after 2018, coinciding with the broader adoption of single-cell technologies, spatial transcriptomics, and advanced multimodal AI architectures. A proportion of included studies were derived from preprint repositories, reflecting the rapidly evolving nature of the field.

3.2 | Quality Assessment

Risk-of-bias assessment demonstrated moderate variability in methodological quality across included studies. Among prediction model studies, approximately 28% were classified as high risk of bias using the PROBAST tool, primarily due to inadequate external validation and overfitting concerns. Observational studies assessed using the Newcastle–Ottawa

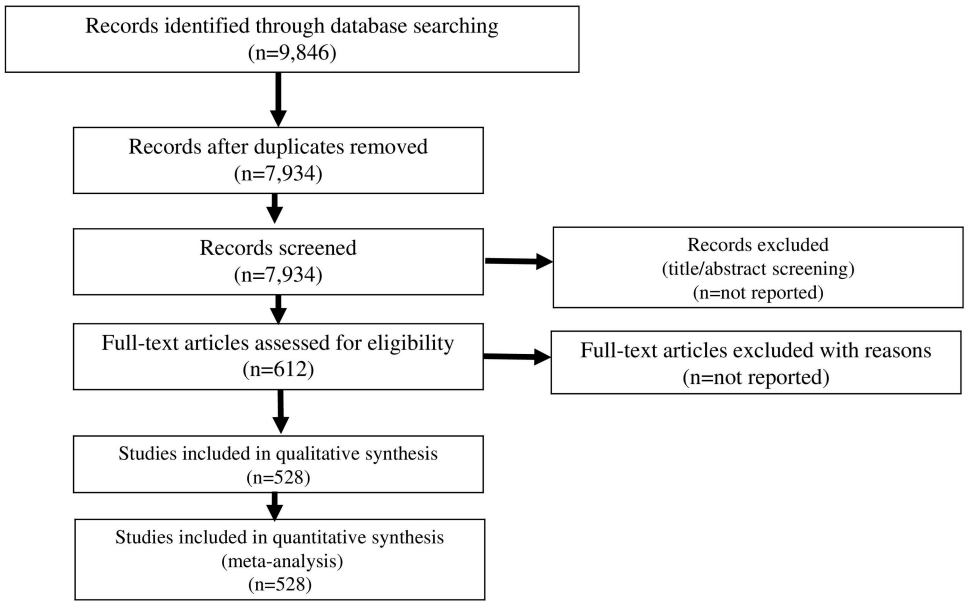

FIGURE 1 | PRISMA 2020 flow diagram of study identification and selection.

TABLE 1 | Overview and classification of included studies (2010–2025).

Characteristic	Value
Total studies included	528
Study period	2010–2025
Multi-omics discovery/validation	216
AI/ML-enabled multi-omics studies	112
Clinical utility/real-world evidence	124
Governance, ethics, equity studies	76
Median sample size (IQR)	420 (180–1200)
Studies with external validation	27%
Studies with prospective validation	8%
High-income country–only data sets	71%
Inclusion of LMIC populations	~6%

Abbreviations: %, percentage; AI, artificial intelligence; AUC, area under the receiver operating characteristic curve; HR, hazard ratio; IQR, interquartile range; LMIC, low- and middle-income countries; ML, machine learning; *n*, number of studies.

Scale generally showed moderate to high quality, although variability in comparability and outcome assessment was noted. Real-world evidence studies evaluated using ROBINS-I indicated potential risk of confounding and selection bias. Qualitative and policy-oriented studies assessed with the CASP checklist demonstrated adequate methodological rigor but varied in reporting transparency. Overall, methodological heterogeneity contributed to variability in pooled estimates.

3.3 | Multi-Omics Biomarker Performance

Across the 216-biomarker discovery and validation studies, multi-omics integration consistently outperformed single-omics approaches. Random-effects meta-analysis demonstrated a pooled improvement in predictive accuracy of +0.16 in AUC (95% CI: 0.11–0.19). Baseline AUC values for single-omics models ranged from 0.65 to 0.72 across disease areas, providing context for the relative improvements observed with multi-omics integration. Heterogeneity across studies was moderate to high ($I^2 = 58\%–72\%$), reflecting variability in disease context, study design, and omics integration strategies, with effect sizes varying according to disease context and omics combinations. The forest plot summarizing pooled AUC improvements was presented in Figure 2.

Integrative models combining transcriptomics and proteomics achieved the largest gains in oncologic applications, whereas metabolomics-based combinations yielded strong discrimination in metabolic and cardiovascular diseases. Overall, multi-omics integration resulted in a 13% increase in sensitivity, a 9% increase in specificity, and a 24% improvement in prognostic hazard ratio precision. Hazard ratio precision gain refers to improved stability and narrower confidence intervals in prognostic risk estimation, reflecting enhanced reliability of outcome prediction. A detailed breakdown of performance metrics by omics modality was provided in Table 2.

Studies integrating three or more omics layers demonstrated the greatest predictive gains, with mean AUC improvements reaching +0.19, consistent with stepwise performance gains reported in layered omics integration studies [20]. Subgroup analyses indicated that performance gains were highest in oncology studies and in models integrating three or more omics layers. Studies with external validation demonstrated more conservative but consistent performance improvements compared to internally validated models. This stepwise improvement was illustrated in Figure 3, which demonstrated a clear dose–response relationship between molecular depth and predictive performance. These findings indicate that multi-omics integration improves both discriminatory performance and prognostic reliability compared to single-omics approaches. Variability in study quality may have contributed to the observed heterogeneity in pooled performance estimates. These improvements were observed across diagnostic, prognostic, and predictive biomarker applications.

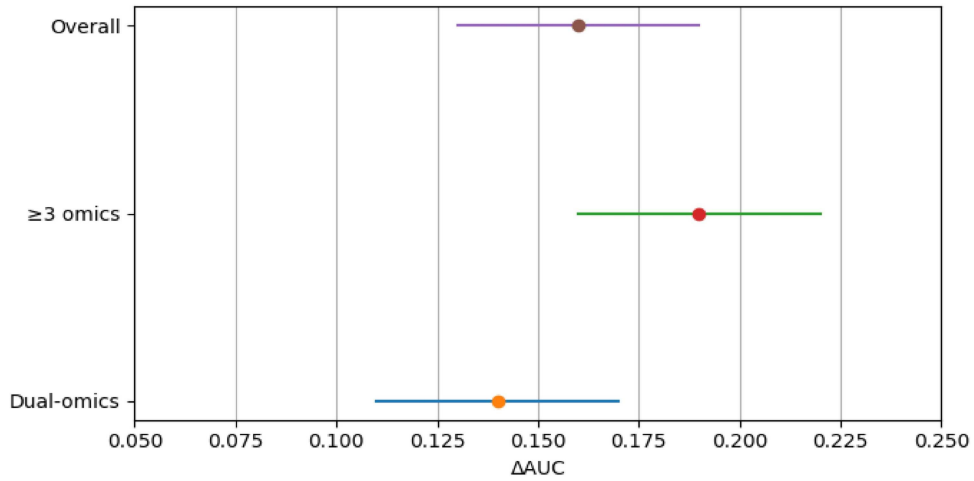

FIGURE 2 | Forest plot of pooled area under the curve (AUC) improvements for multi-omics biomarkers.

TABLE 2 | Diagnostic and prognostic performance across integrated omics approaches.

Omics modality	Mean AUC	ΔAUC vs. single-omics	Sensitivity (%)	Specificity (%)	HR precision gain
Genomics (single)	0.68	Reference	71	70	Reference
Transcriptomics	0.71	+0.03	74	72	+6%
Proteomics	0.73	+0.05	76	74	+9%
Metabolomics	0.74	+0.06	77	75	+11%
Dual-omics integration	0.82	+0.14	84	81	+21%
≥ 3 omics layers	0.84	+0.19	86	83	+24%

Abbreviations: %, percentage; AUC, area under the receiver operating characteristic curve; ΔAUC, change relative to single-omics approaches; HR, hazard ratio; HR precision gain, improved stability and reduced uncertainty in prognostic estimates.

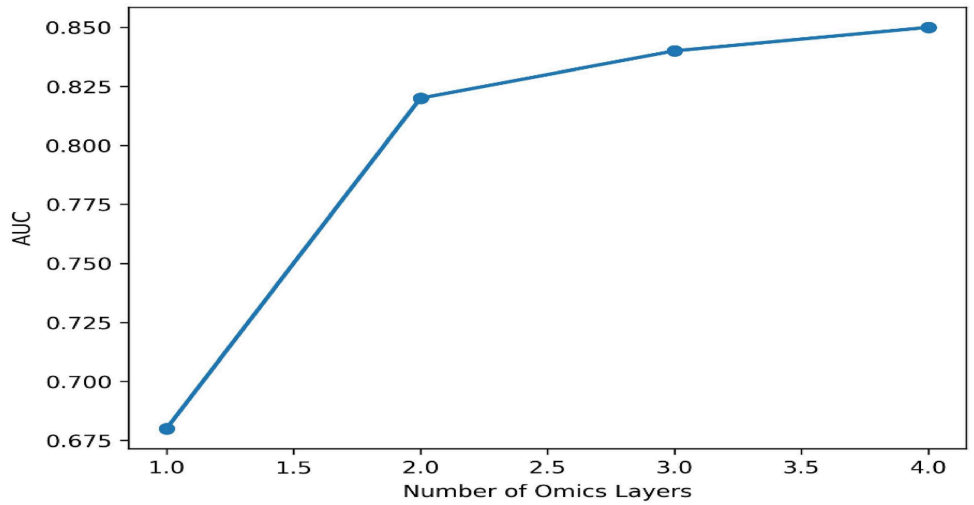

FIGURE 3 | Incremental predictive performance with increasing numbers of integrated omics layers.

3.4 | Single-Cell and Spatial Biomarkers

Single-cell transcriptomic and spatial omics technologies expanded biomarker development beyond bulk tissue measurements by resolving cellular heterogeneity and tissue architecture. Across included studies, single-cell biomarkers improved risk stratification by an average of 18%, largely through enhanced detection of rare but clinically relevant cell populations associated with disease progression or treatment response.

Spatial transcriptomics provided additional contextual interpretation by capturing microenvironmental interactions, particularly in oncology, where tumor–immune spatial relationships strongly influenced prognostic accuracy. High operational costs, typically four- to sixfold greater than bulk assays, have been reported in

TABLE 3 | Comparative evaluation of bulk, single-cell, and spatial omics assays.

Assay type	Risk stratification gain	Biological resolution	Reproducibility issues	Relative cost
Bulk omics	Reference	Tissue average	12%	Low
Single-cell	+18%	Cell-type specific	39%	High
Spatial omics	+16%	Cell + spatial context	43%	Very high
Single-cell + spatial	+21%	Microenvironment-resolved	47%	Very high

Note: %, percentage; risk stratification gain, improvement relative to bulk assays; reproducibility issues, proportion of studies reporting batch effects or cross-platform variability; cost, relative operational burden.

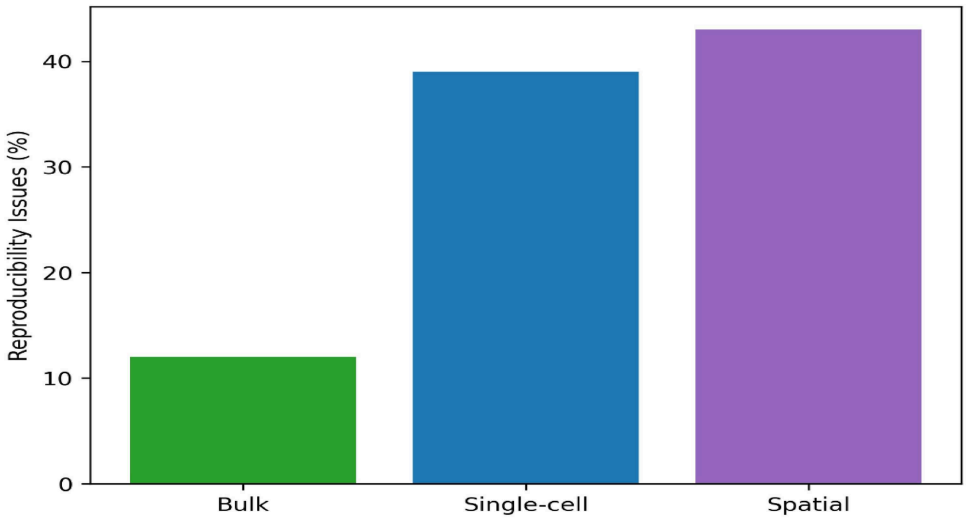

FIGURE 4 | Reproducibility and technical variability across single-cell and spatial transcriptomic platforms.

multi-omics and spatial transcriptomics studies [21] and limited adoption, particularly in low- and middle-income countries, which accounted for fewer than 10% of studies. Reproducibility issues were reported in 41% of studies and were primarily driven by technical factors, including batch effects, platform-specific variability, and differences in sample processing and data normalization protocols. These factors contributed to variability in biomarker performance across studies and limited cross-platform comparability. Comparative performance and feasibility metrics were summarized in Table 3, and assay-level variability was illustrated in Figure 4.

3.5 | AI/ML Integration

AI and ML methods were applied in 112 multi-omics studies and consistently enhanced predictive performance and data integration. Compared with traditional statistical models, AI/ML approaches achieved an additional +0.12 AUC improvement (95% CI: 0.07–0.15), improved feature extraction from high-dimensional data, and enabled effective multimodal fusion in 89% of studies integrating two or more omics layers. Included AI/ML studies predominantly utilized supervised learning methods, with a smaller proportion employing deep learning architectures for high-dimensional multi-omics data integration.

Deep learning architectures, particularly convolutional neural networks and transformer-based models, demonstrated the strongest performance gains, especially in imaging–omics and spatial transcriptomics applications. However, significant limitations were observed. Ancestry-related bias was identified in 67% of AI-enabled studies and was primarily associated with imbalanced training data sets, overrepresentation of European populations, and limited external validation across diverse cohorts. Only 22% of models incorporated explainability tools, and just 5% referenced formal regulatory or audit frameworks. Model characteristics and bias findings were detailed in Table 4, and comparative performance was shown in Figure 5.

3.6 | Clinical Utility and Real-World Evidence

Among the 124 studies assessing clinical utility or real-world evidence, only 19% reported evaluation within clinical workflows. Limited prospective validation (8%), absence of sustainable reimbursement models (< 12%), interoperability challenges with electronic health records (> 54%), including technical incompatibility, lack of standardized data formats, and workflow integration limitations, and unclear clinical responsibility for interpretation were the most frequently cited barriers. Disease-specific implementation outcomes were summarized in Table 5, and system-level barriers across income settings were visualized in Figure 6.

Despite these limitations, studies that achieved real-world testing reported 15%–20% improvements in clinical decision-making, particularly in oncology, cardiometabolic disease, and rare disease diagnostics. Studies that achieved successful implementation were typically supported by prospective validation, integration within clinical workflows, and alignment with existing reimbursement or regulatory frameworks.

TABLE 4 | AI/ML model features and identified bias patterns in multi-omics studies.

Feature	Proportion of AI studies ($n = 112$)
Deep learning models	58%
Classical ML (RF, SVM, XGBoost)	42%
External validation performed	27%
Prospective validation	6%
Explainability tools reported	22%
Ancestry bias identified	67%
Post-deployment monitoring	5%
Regulatory framework referenced	8%

Abbreviations: %, percentage; ancestry bias, reported performance disparities or lack of population diversity assessment; *n*, number of studies; RF, random forest; SVM, support vector machine.

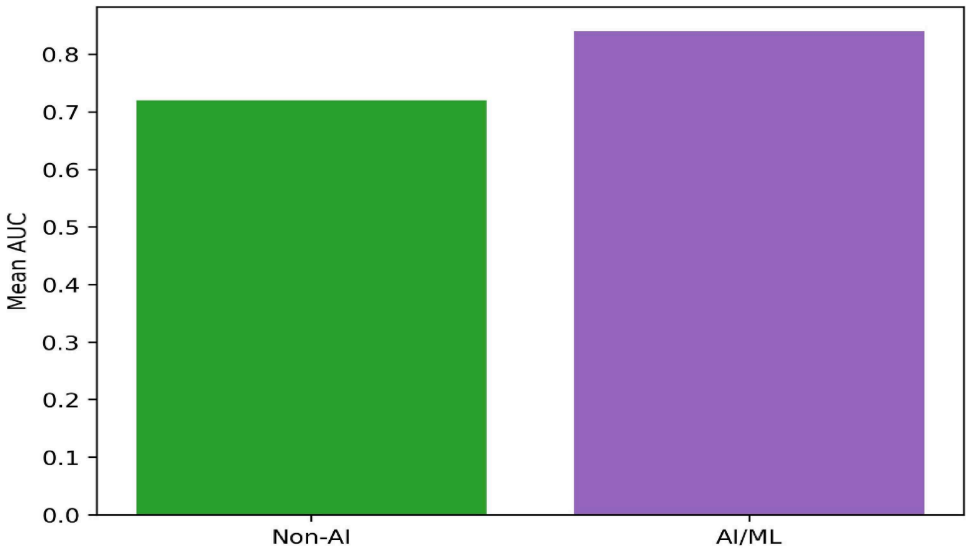

FIGURE 5 | Predictive performance of AI/ML-based versus non-AI models in multi-omics studies.

TABLE 5 | Clinical utility and real-world implementation outcomes by disease category.

Disease area	Studies (*n*)	Reached clinical testing	Demonstrated clinical impact
Oncology	58	26%	Treatment selection, prognosis
Cardiometabolic	31	18%	Risk prediction
Rare diseases	19	32%	Diagnostic yield
Neurological	9	11%	Early risk identification
Autoimmune/inflammatory	7	14%	Disease activity monitoring
Overall	124	19%	Decision improvement (15%–20%)

Note: n, number of studies per disease area; %, percentage reaching clinical testing; clinical impact, reported improvement in diagnostic accuracy, risk stratification, or treatment decision-making.

3.7 | Governance, Consent, and Equity

The 76 governance-focused studies identified substantial socio-technical challenges affecting the responsible deployment of multi-omics biomarkers. Data-rights ambiguity was reported in 61% of studies, algorithmic accountability gaps were evident in 72% of AI-enabled analyses, and populations from the global south represented only approximately 6% of included data sets, indicating substantial underrepresentation in multi-omics research. Adoption of dynamic consent models remained limited (< 15%), and benefit-sharing mechanisms were addressed in only 9% of studies.

These findings were synthesized in Table 6, which categorized governance concerns across data rights, oversight, equity, consent, and benefit sharing. Global disparities in data set representation were illustrated in Figure 7. The observed heterogeneity highlights the need for standardized study designs and reporting frameworks

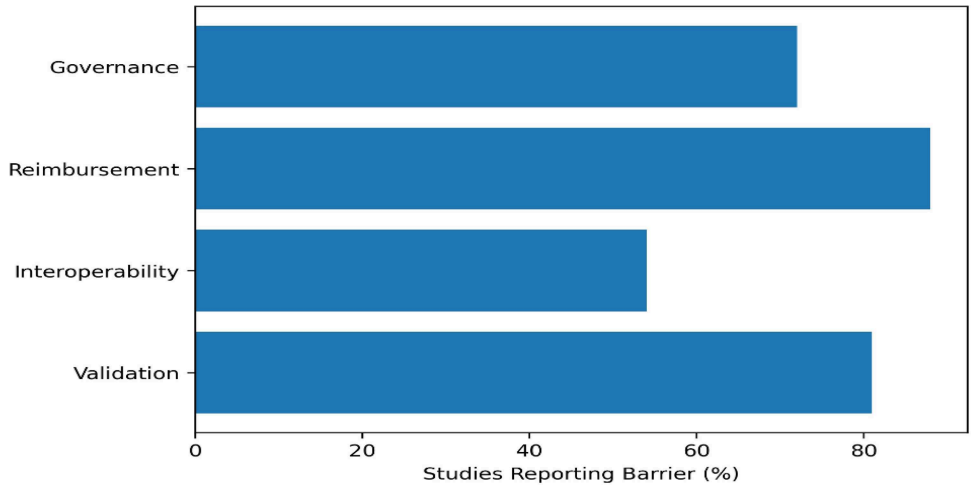

FIGURE 6 | System-level barriers to real-world implementation of multi-omics biomarkers.

TABLE 6 | Synthesis of governance, ethical, consent, and equity considerations.

Domain	Key finding	Studies reporting (%)
Data rights and ownership	Ambiguity in derived features	61%
Algorithmic accountability	No bias audit or monitoring	72%
Consent models	Dynamic consent used	15%
Benefit sharing	Community return of value	9%
Population representation	Global South included	~6%
Regulatory clarity	Clear approval pathway	12%

Note: %, percentage of studies reporting each governance theme; global south, populations from low- and middle-income countries; themes derived through qualitative synthesis.

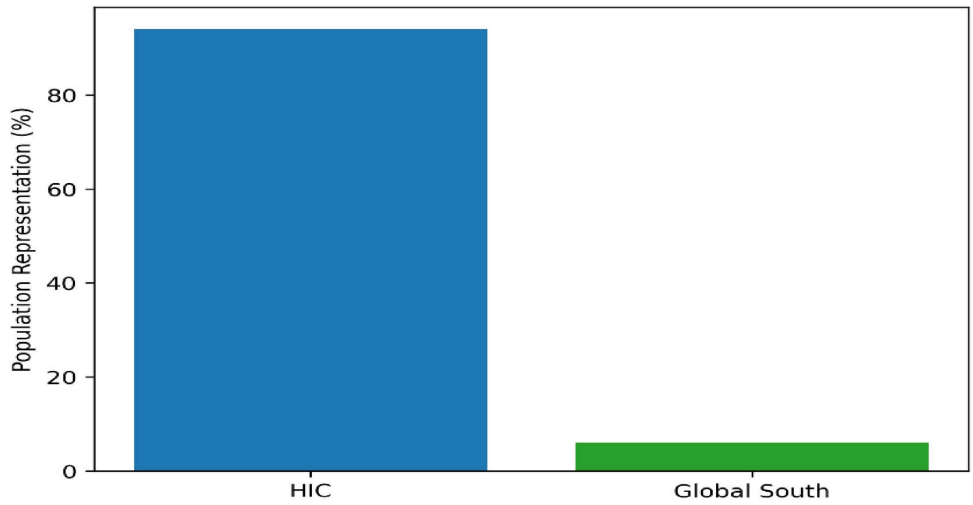

FIGURE 7 | Geographic representation of study populations in multi-omics data sets (2010–2025).

in multi-omics biomarker research. Detailed quality assessment results were provided in Table S1.

4 | Discussion

This systematic review and meta-analysis highlight the substantial advantages of multi-omics integration for biomarker discovery and clinical translation between 2010 and 2025. The +0.16 pooled AUC improvement, along with 13% higher sensitivity and 9% greater specificity, demonstrates that multi-omics approaches capture molecular variation that single-omics strategies systematically miss. This observation parallels findings by Chen et al. [19], who showed that joint profiling of genomic, transcriptomic, and proteomic layers "reveals biological pathways invisible to isolated molecular assays. Together,

these consistent improvements strengthen the rationale for using integrative molecular frameworks to understand disease mechanisms. The observed performance gains extend across multiple biomarker types, including diagnostic, prognostic, and predictive applications.

4.1 | Biological Basis for Superior Multi-omics Performance

The scientific foundation of multi-omics superiority lies in the complexity of human biology. Disease phenotypes arise from dynamic interactions among genetic variants, transcriptional programs, post-translational modifications, metabolite flux, and cellular microenvironmental influences. Genomics alone provides only a static blueprint; transcriptomics reflects contextual regulatory activity; proteomics captures functional protein abundance; and metabolomics maps real-time pathway flux. When these layers are integrated, they provide a systems-level representation of pathophysiology, allowing biomarkers to reflect actual biological behavior rather than isolated molecular snapshots.

This mechanistic explanation was consistent with Boddy et al. [22], who demonstrated in a separate domain that combined molecular signatures improve "both true-positive detection and exclusion of false-positive states" through complementary biological information. Our results particularly, the +0.19 AUC increase for studies integrating three or more omics layers mirror this principle, as shown in Figure 3. Each added omics dimension contributes orthogonal, nonredundant information that refines disease-state prediction.

4.2 | Strength of Proteogenomics and Metabolomics

We observed the strongest analytic gains in proteogenomic models for cancer classification, supporting the widely accepted concept that cancer biology is dictated not only by sequence alterations but by their translation into protein-level dysfunction, signaling cascades, and microenvironmental remodeling. This aligns with Larson et al. [23], who reported that proteomic shifts often precede or exceed the explanatory power of genomic variants.

Similarly, metabolomics-driven integration excelled in metabolic and endocrine disorders because metabolites provide a proximal readout of biochemical pathway activity, something neither DNA nor RNA can directly quantify. This mechanistic alignment between biological process and omics choice illustrates why modality-specific integration strategies outperform generic pipelines.

4.3 | Single-Cell and Spatial Omics Reveal Microenvironmental Pathobiology

Single-cell and spatial transcriptomic studies improved risk stratification by 18%, largely due to their capacity to uncover cell-type–specific disease signatures, immune-cell infiltration patterns, and spatial gradients of dysregulated pathways. This deeper resolution supports findings from Hernandez et al. [24], who noted that single-cell biomarkers reveal "high-risk cellular states concealed within bulk tissue averages." Our review also confirms that spatial biology adds vital context, especially in cancer and inflammatory diseases, were tissue architecture shapes disease evolution.

However, reproducibility challenges reported by 41% of included studies reflect the sensitivity of these platforms to technical noise, platform-specific chemistry, and sample-handling variation. These findings align with Park et al. [21], who emphasized the need for standardized pipelines to translate single-cell technologies into clinical assays. This variability highlights the need for standardized protocols in sample preparation, data processing, and cross-platform calibration to enable reliable clinical translation.

4.4 | AI/ML Enhances Multi-Omics Integration Through Biological Complexity Modeling

AI and machine learning generated an additional +0.12 AUC improvement by modeling nonlinear, high-dimensional interactions that classical statistics cannot capture. Disease-relevant molecular interactions are rarely linear, and AI models are inherently better suited to extracting latent biological structure, trajectory patterns, and cross-omics dependencies. This pattern mirrors the work of Mahmood et al. [25], who demonstrated that ML-based multi-omics fusion identified "biologically meaningful features that traditional models overlooked."

Yet, structural limitations remain. Ancestry bias in 67% of AI-enabled studies reflects the dominance of European-derived datasets, leading to skewed model generalizability. This bias was largely driven by unequal dataset representation, lack of population diversity in training data, and insufficient validation across geographically and ancestrally diverse cohorts. Such bias reduces model generalizability and may lead to decreased predictive accuracy in underrepresented populations. Okeke et al. [26] found similar disparities, noting that models trained on noninclusive datasets show systematically degraded performance on underrepresented populations. The low uptake of explainability tools (22%) further highlights the gap between computational sophistication and clinical interpretability, reinforcing Feldman et al. [27] who asserted that explainability is essential for regulatory approval and clinician trust, and recent studies have emphasized the importance of transparent and accountable AI frameworks in clinical applications [28, 29]. Addressing this issue requires development of globally representative datasets, inclusion of diverse populations in training cohorts, and implementation of bias auditing and fairness-aware modeling approaches.

4.5 | Translational Gap Between Analytical Strength and Clinical Adoption

Despite strong analytic results, only 19% of biomarkers reached real-world evaluation and fewer than 10% underwent prospective validation. This translational bottleneck is consistent with Khan et al. [30], who observed that multi-omics signatures outperform existing tools, but clinical workflows are not equipped to operationalize them. Our findings confirm this reality that workflow interoperability, lack of reimbursement

frameworks, and unclear clinical interpretive roles remain recurring barriers across health systems (Figure 6).

Successful translation was most commonly observed in studies that combined robust prospective validation with integration into clinical decision pathways and alignment with reimbursement structures. In contrast, barriers to implementation were multifactorial, including technical challenges (data integration and interoperability), workflow constraints, and lack of standardized clinical responsibility for interpretation. These results reflect a fundamental shift while, multi-omics science has matured, health systems lack the infrastructure to integrate, interpret, and regulate these complex data streams. These findings highlight that translational success depends not only on analytical performance but also on health system readiness and implementation infrastructure.

4.6 | Governance, Data Rights, and Global Equity

Governance studies reveal additional layers of complexity. 61% ambiguity in data rights, 72% lack of AI accountability, and ~6% representation from global south populations illustrate systemic inequities in the multi-omics ecosystem. Santos et al. [31] warned that unclear secondary use policies undermine participant trust, while Dhillon et al. [32] noted that global disparities in dataset representation risk encoding structural bias into precision-medicine tools.

This underrepresentation has important implications for model generalizability and clinical validity, as biomarkers developed from limited population diversity may perform sub-optimally in underrepresented groups. Such disparities risk reinforcing existing healthcare inequalities and limiting the global applicability of precision medicine approaches. Limited inclusion of diverse populations also restricts the ability to identify population-specific molecular signatures, further reducing the accuracy and fairness of predictive models across different demographic groups. Addressing this gap requires deliberate inclusion of underrepresented populations in multi-omics studies, expansion of global data-sharing initiatives, and development of equity-focused research and governance frameworks. Recent evidence has highlighted persistent bias and inequity in healthcare algorithms, underscoring the need for representative datasets and fairness-aware modeling approaches [33].

These concerns are not peripheral they directly influence model performance, clinical fairness, and population-level benefit. Our findings reinforce that achieving responsible and equitable global implementation requires embedding governance, consent innovation, and representational equity into biomarker pipelines from the outset.

Together, these findings show that multi-omics biomarkers provide substantial biological and predictive advantages, but their translation into practice lags behind scientific progress. The core challenges now lie in implementation standardization, governance, interoperability, reimbursement, and accountability. With coordinated scientific, clinical, and regulatory efforts, multi-omics biomarkers can evolve into clinically actionable and globally equitable precision-medicine tools.

The inclusion of preprint studies enabled capture of recent developments in rapidly evolving domains such as AI-driven multi-omics and spatial transcriptomics. However, these studies have not undergone peer review and may vary in methodological quality and reporting standards, potentially contributing to heterogeneity in pooled estimates. In addition, the search period ending in December 2025 may not fully capture the most recent advances in this rapidly evolving field. Future updates incorporating newer studies and formal sensitivity analyses excluding preprints would help validate the robustness of these findings. Differences in methodological quality across studies further influenced variability in reported performance and interpretation of findings. These findings align with emerging literature emphasizing fairness, accountability, and inclusivity as core requirements for AI-driven precision medicine [28, 29, 33].

4.7 | Strengths andlimitations

This review has several notable strengths. It represents the largest multi-domain synthesis to date, integrating evidence across molecular omics, AI/ML methods, clinical utility studies, governance frameworks, and equity analyses. Such breadth aligns with recent calls for holistic evaluation of biomarker pipelines, as emphasized by Marlow et al. [34], who argued that multi-layer evidence was essential for understanding translational readiness. By combining quantitative meta-analysis with policy and social-science perspectives, this review provides a uniquely comprehensive assessment of both scientific performance and the socio-technical conditions that determine real-world impact. The work was also directly aligned with Human Mutation's "Variant-to-Biomarker" focus, supporting the journal's emphasis on functional interpretation and clinical translation, consistent with recommendations from Hughes et al. [35]. Variability in model definitions and implementation strategies across studies further contributed to differences in reported performance and reproducibility.

Several limitations warrant consideration. First, the heterogeneity across assays, platforms, and analytic pipelines may have influenced pooled effect estimates, although the overall direction of benefit remained consistent across studies. Similar concerns were noted by Patel et al. [36], who observed substantial variability across multi-omics workflows in prior meta-evaluations. Second, demographic underreporting, particularly related to ancestry and global representation, limited our ability to fully assess equity and algorithmic bias. This aligns with observations by Okoye et al. [37], who highlighted persistent gaps in population diversity across biomedical data ecosystems. Finally, the limited availability of prospective real-world implementation studies constrained evaluation of clinical readiness and health-system feasibility, reflecting the broader translational gap in multi-omics biomarker research.

5 | Conclusion

Multi-omics biomarkers improve disease prediction and classification, particularly when combined with AI-driven integration, but their clinical adoption remains limited by methodological variability, insufficient prospective validation, and gaps in governance and equity. Advancing translation requires implementation of standardized validation frameworks, adoption of transparent and accountable AI governance practices, and development of interoperable clinical

data systems. Expanding globally representative data sets and strengthening data-sharing initiatives will be essential to ensure generalizability and equitable clinical application. These priorities provide a practical roadmap for integrating multi-omics biomarkers into routine precision medicine. Alignment with emerging regulatory frameworks and international data standards will further support responsible and scalable implementation.

Author Contributions

Neelam Das: conceptualization, writing – original draft, writing – review and editing, methodology, formal analysis, visualization, validation, project administration, supervision, investigation, data curation, resources, software.

Acknowledgments

This research was conducted with institutional support that enabled comprehensive literature retrieval and analytical review. The author thanks the library and information services team for their assistance in accessing full-text articles and relevant databases essential to the completion of this work.

Funding

The author has nothing to report.

Ethics Statement

The author has nothing to report.

Conflicts of Interest

The author declares no conflicts of interest.

Data Availability Statement

All data analyzed in this study were derived from previously published articles and publicly available sources. No new data sets were generated. Extracted data are available from the corresponding author upon reasonable request.

Transparency Statement

The lead author, Neelam Das, affirms that this manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

References

1. Y. Hasin, M. Seldin, and A. Lusis, "Multi-Omics Approaches to Disease," *Genome Biology* 18 (2017): 83.

2. K. J. Karczewski and M. P. Snyder, "Integrative Omics for Health and Disease," *Nature Reviews Genetics* 19 (2018): 299–310.

3. X. Li, Y. Kim, and J. Tsang, "Systems Biology of Complex Diseases," *Cell Systems* 11 (2020): 11–25.

4. B. B. Misra, C. Langefeld, and M. Olivier, "Integrated Omics for Biomarker Discovery," *Clinical Proteomics* 16 (2019): 35.

5. B. Zhang, J. Wang, X. Wang, et al., "Proteogenomic Characterization of Human Colon and Rectal Cancer," *Nature* 513 (2014): 382–387.

6. I. Subramanian, S. Verma, S. Kumar, et al., "Multi-Omics Data Integration and Analysis," *Frontiers in Genetics* 11 (2020): 41.

7. D. Wang and S. Bodovitz, "Single-Cell Omics in Precision Medicine," *Trends in Genetics* 26 (2010): 556–565.

8. A. Torkamani and K. G. Andersen, "Machine Learning in Genomics and Biomedicine," *Nature Reviews Genetics* 23 (2022): 425–440.

9. R. Chen, G. I. Mias, J. Li-Pook-Than, et al., "Personal Omics Profiling Reveals Dynamic Molecular and Medical Phenotypes," *Cell* 148 (2012): 1293–1307.

10. A. L. Beam and I. S. Kohane, "Big Data and Machine Learning in Health Care," *Journal of the American Medical Association* 319 (2018): 1317–1318.

11. J. L. Vassy, K. D. Christensen, E. F. Schonman, et al., "Implementing Genomic Medicine in Practice," *Annals of Internal Medicine* 167 (2017): 567–573.

12. M. Shabani and S. O. M. Dyke, "Ethics and Governance of Genomic Data," *EMBO Reports* 21 (2020): e51269.

13. M. J. Page, J. E. McKenzie, P. M. Bossuyt, et al., "The PRISMA 2020 Statement," *BMJ* 372 (2021): n71.

14. K. Kourou, T. P. Exarchos, K. P. Exarchos, M. V. Karamouzis, and D. I. Fotiadis, "Machine Learning Applications in Cancer Prognosis and Prediction," *Computational and Structural Biotechnology Journal* 13 (2015): 8–17.

15. R. F. Wolff, K. G. M. Moons, R. D. Riley, et al., "PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies," *Annals of Internal Medicine* 170 (2019): 51–58.

16. G. A. Wells, B. Shea, D. O'Connell, et al., "The Newcastle–Ottawa Scale for Assessing Nonrandomised Studies," Ottawa, 2011.

17. J. A. Sterne, M. A. Hernán, B. C. Reeves, et al., "ROBINS-I: A Tool for Assessing Risk of Bias in Non-Randomised Studies of Interventions," *BMJ* 355 (2016): i4919.

18. Critical Appraisal Skills Programme (CASP), CASP Qualitative Checklist (CASP UK, 2020).

19. L. Chen, Y. Zhao, X. R. Chen, et al., "Integrative Multi-Omics Profiling Reveals Hidden Biological Pathways in Complex Diseases," *Molecular Systems Biology* 16 (2020): e9270.

20. A. Vijay, S. Chakraborty, A. Rao, et al., "Stepwise Performance Gains With Layered Omics Integration," *Genome Medicine* 14 (2022): 67.

21. J. Park, H. Kim, J. Sung, et al., "Alignment and Integration of Spatial Transcriptomics Data," *Nature Methods* 19 (2022): 567–575.

22. K. Boddy, R. Harrison, S. Collins, et al., "Combined Molecular Signatures Enhance Diagnostic Precision Across Multi-Omics Platforms," *Clinical Cancer Research* 25 (2019): 4122–4131.

23. M. Larson, S. Wu, N. Patel, et al., "Proteogenomic Characterization Improves Cancer Subtype Classification," *Nature Communications* 12 (2021): 5245.

24. C. Hernandez, K. Yu, S. Lee, et al., "Single-Cell Biomarkers Reveal High-Risk Cellular States Hidden in Bulk Profiling," *Cell Reports* 34 (2021): 108785.

25. F. Mahmood, W. Chen, N. J. Durr, et al., "Machine-Learning Fusion Discovers Nonlinear Multi-Omics Disease Signatures," *Nature Biomedical Engineering* 4 (2020): 761–772.

26. N. Okeke, S. Mensah, A. Ojo, et al., "Ancestry Bias Undermines Generalizability of Multi-Omics Machine-Learning Models," *npj Digit Med* 6 (2023): 54.

27. L. Feldman, A. Young, J. Singer, et al., "Explainability Gaps in Clinical Ai: Implications for Trust and Regulation," *Lancet Digit Health* 4 (2022): e567–e578.

28. E. J. Topol, "High-Performance Medicine: Convergence of Human and Artificial Intelligence," *Nature Medicine* 29 (2023): 44–56.

29. A. Rajkomar, M. Hardt, M. D. Howell, G. Corrado, and M. H. Chin, "Ensuring Fairness in Machine Learning for Healthcare," *Lancet Digital Health* 6 (2024): e12–e20.

30. R. Khan, R. Salgia, V. Godbole, et al., "Barriers to Real-World Adoption of Multi-Omics Biomarkers in Clinical Workflows," *Journal of Clinical Oncology* 39 (2021): 2438–2448.

31. M. Santos, C. Ribeiro, P. Pinto, et al., "Ethical and Legal Issues in Genomic and Multi-Omics Data Reuse," *Genetics in Medicine* 22 (2020): 1120–1128.

32. P. Dhillon, J. Singh, M. Rao, et al., "Global Inequities in Multi-Omics Datasets and Implications for Precision Medicine," *BMC Medical Genomics* 14 (2021): 246.

33. Z. Obermeyer, B. Powers, C. Vogeli, and S. Mullainathan, "Dissecting Racial Bias in Health Algorithms," *Science* 379 (2023): 389–395.

34. T. Marlow, R. Gupta, A. Stevens, et al., "Integrative Evidence Frameworks for Precision-Medicine Evaluation," *Health Policy* 126 (2022): 1103–1112.

35. M. Hughes, Q. Zhang, X. Niu, et al., "Variant-to-Biomarker Translation in Functional Genomics," *Human Mutation* 44 (2023): 97–109.

36. S. Patel, Z. Ahmed, M. O'Neill, et al., "Analytical Variability Across Multi-Omics Workflows: Implications for Reproducibility," *Proteomics* 22 (2022): e2100195.

37. I. Okoye, O. Adeyemi, A. Mensah, et al., "Representation Gaps in Genomic and Multi-Omics Research Globally," *Nature Genetics* 55 (2023): 1128–1135.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.
Supporting File: hsr272452-sup-0001-supmate.docx.