Review
Enitan *et. al. Glob. South Health Horiz.* 2025 1(1):12-35
10.63950/gshh.2025.1.1.2

Integrating Genomics and Proteomics Technologies in Biological Research: Advantages, Challenges, and Prospects

Enitan SS[1], Adejumo EN[1], Osakue OE[1], Eke SS[2], Akele RY[3], Edafetanure-Ibeh OM[4], Enitan CB[1]

1 Department of Medical Laboratory Science, School of Allied Health Sciences, Babcock University, Ilishan-Remo, 121109, Ogun State, Nigeria

2 Biology Unit, Air Force Institute of Technology, Kaduna, 800282, Kaduna State, Nigeria

3 Department of Biomedical Science, School of Applied Science, University of Brighton, Brighton, ES BN2 4AT, United Kingdom

4 Department of Environmental and Occupational Health, Texas A&M School of Public Health, 212 Adriance Lab Rd, College Station, TX 77843, Texas, United States of America

Date of Submission: 27th May 2025
Date of Acceptance: 8th June 2025
Date of Publication: 30th June 2025

Abstract

Introduction: The integration of genomics and proteomics, two cornerstone omics technologies, has transformed our understanding of biological systems, enabling deeper insights into disease mechanisms, biomarker discovery, and the development of personalised medicine. While genomics uncovers genetic blueprints, proteomics reveals functional protein expressions, making their combined use essential for a holistic view of biological processes. This narrative review explores the current state, advantages, challenges, and prospects of integrating genomics and proteomics in biological research, particularly in the context of translational medicine and global health applications.

Methods: A systematic literature search was conducted across four databases—PubMed, Scopus, Web of Science, and Google Scholar, for English-language peer-reviewed articles published between January 2015 and May 2025. A PRISMA-based approach guided article screening and selection. Inclusion criteria focused on studies addressing multi-omics integration, relevant technologies, clinical applications, and associated ethical concerns. The findings were synthesised using thematic analysis.

Synthesis: The review highlights the growing synergy between genomics and proteomics in enhancing disease diagnosis, drug discovery, and precision medicine. Advances in next-generation sequencing and mass spectrometry, coupled with computational tools such as MOFA+ and AlphaFold, have improved data integration. However, challenges persist, including high data complexity, computational demands, ethical/legal concerns, and limited access in low-resource settings.

Conclusion: Integrated omics approaches hold immense potential to revolutionise healthcare and biological discovery. Addressing current barriers and fostering global collaboration, particularly in resource-limited regions, are critical to realising the full benefits of genomics and proteomics integration.

Keywords: Genomics, Proteomics, Proteogenomics, Multi-omics Integration, Research, Technologies

Introduction

The advent of high-throughput omics technologies has dramatically reshaped the landscape of biological research and medicine. Among these, genomics and proteomics stand out as foundational pillars in understanding the complexity of life at the molecular level. Genomics provides comprehensive insights into the genetic blueprint of organisms, enabling the identification of genetic variations associated with health and disease (1, 2). Proteomics, on the other hand, focuses on the large-scale study of proteins, key functional molecules that drive nearly every biological process. Together, these disciplines have become instrumental in uncovering molecular mechanisms underlying diseases, identifying diagnostic and prognostic biomarkers, and facilitating precision medicine (3, 4, 5).

Despite their power, genomics and proteomics often offer incomplete pictures when studied in isolation (6). The expression of genes does not always correlate with protein abundance or activity due to factors such as post-transcriptional regulation, alternative splicing, and post-translational modifications (6, 7). This gap has given rise to integrative approaches, particularly proteogenomics, which seeks to merge genomic and proteomic data to better understand functional outcomes and biological systems (8, 9). Integrative omics approaches have been successfully applied in diverse fields, including oncology, neuroscience, infectious disease surveillance, and agriculture (10, 11). For example, initiatives such as the Clinical Proteomic Tumour Analysis Consortium (CPTAC) and the Human Proteome Project have demonstrated the clinical relevance of combining genomics and proteomics to uncover novel drug targets and refine therapeutic strategies (12, 13).

Nevertheless, several challenges continue to hinder the full realisation of integrated omics. These include technical barriers in data generation and analysis, lack of standardised bioinformatics pipelines, ethical concerns around data privacy, and limited accessibility in low- and middle-income countries (LMICs) (14). Moreover, while numerous studies have contributed valuable insights into individual aspects of genomics and proteomics, a consolidated and critical synthesis of recent advancements in their integration, especially from a global health perspective, remains limited (1, 6, 15).

There is a particular need to explore how emerging technologies such as artificial intelligence (AI), cloud computing, and spatial omics are enhancing the integration process and addressing issues of scalability and equity. This manuscript presents a narrative review of recent literature in the last decade (2015–2025), guided by a systematic search strategy, to critically assess the advantages, challenges, and future directions of integrating genomics and proteomics technologies in biological research. Unlike a traditional systematic review, which is hypothesis-driven and strictly evidence-based, this narrative review allows for broader thematic exploration while ensuring methodological rigour through a structured screening and inclusion process. The literature was selected from four major databases, PubMed, Scopus, Web of Science, and Google Scholar, focusing on peer-reviewed articles published in English within the specified time frame. This review discusses the benefits of combining genomics and proteomics in biological research and reveals some of the key technologies and applications of genomics and proteomics integration, in addition to the challenges and opportunities that lie ahead. By synthesising recent findings and expert perspectives, this review aims to inform future research, policy, and practice in the evolving field of multi-omics integration. Special attention is given to clinical applications, translational impact, and global equity, reflecting the growing recognition that omics research must be inclusive, reproducible, and ethically sound to drive meaningful innovation in 21st-century biology.

Methods

This review employed a structured and systematic approach to identify, screen, and analyse relevant literature on proteogenomics (integration of genomics and proteomics) technologies in biological research. The aim was to provide a comprehensive and up-to-date synthesis of the advantages, challenges, and prospects of these converging fields, with an emphasis on developments between 2015 and 2025.

Data Sources and Search Strategy

Literature searches were conducted across four major electronic databases: PubMed, Scopus, Web of Science, and Google Scholar. The search was limited to articles published in English between January 1, 2015, and May 30, 2025. Keywords and Boolean operators were used to formulate the search strings. The primary search terms included: “Genomics”, “Proteomics”, “Integrative omics”, “Multi-omics”, “Systems biology”, “Biological research”, “Post-translational modification”, “Personalised medicine”, and “Data integration”.

The search strings used were combinations such as: "Genomics AND Proteomics AND (Integration OR Multi-omics) AND Biological Research"; "Post-translational modification AND Proteogenomics"; "Genomics AND Proteomics AND Personalised and Medicine"; Filters were applied for peer-reviewed journal articles, reviews, and original research studies.

Screening and Selection Process

A total of 270 records were initially retrieved. After removing duplicates (n = 78), 192 articles remained for title and abstract screening. Two independent reviewers screened articles for relevance. Discrepancies were resolved through discussion or consultation with a third reviewer. 132 articles were excluded, leaving a final total of 60 articles for inclusion in the review (Figure 1).

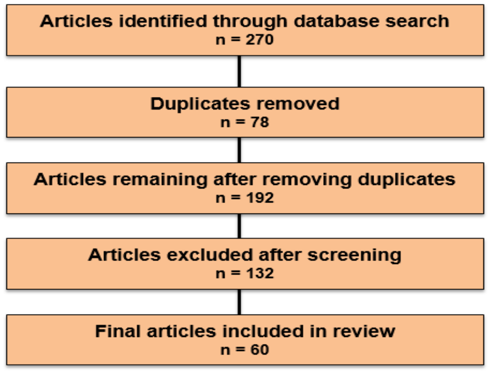

Figure 1: Simplified PRISMA Flow Diagram

Eligibility Criteria

The study included Articles published between 2015 and 2025; Peer-reviewed journal articles; Studies or reviews focusing on proteogenomics; Research relevant to biological systems, disease mechanisms, or health applications; and Articles written in English language. The review excluded non-English publications, Conference abstracts, editorials, and commentaries; Studies focusing solely on either genomics or proteomics without integrative analysis; and Articles not accessible in full text.

Data Extraction and Synthesis

From the final pool of 60 eligible articles, data were extracted, categorised and synthesised thematically under: 1) Comparison of standalone and integrated omics approaches, 2) Applications of proteogenomics, 3) Technical Approaches and Tools for Integrating Genomics and Proteomics, 4) Challenges in integrative omics and corresponding mitigation strategies, 5) Global and translational impact of integrative genomics and proteomics, and 6) Knowledge gaps, associated impact, and future research directions in integrative genomics and proteomics.

Synthesis

Results

This narrative review synthesised a decade of literature (2015–2025) on proteogenomics, with a specific focus on the emerging discipline of proteogenomics. The integration of these two omics domains reveals synergistic advantages over their standalone applications, particularly in disease research, biomarker discovery, and the advancement of precision medicine. Proteogenomics represents a key integrative strategy that combines genomic, transcriptomic, and proteomic data to better characterize the protein landscape of cells, particularly in disease contexts. As shown in Figure 2, proteogenomics uses mass spectrometry and next-generation sequencing (NGS) pipelines to generate a more accurate and comprehensive map of the proteome, including novel splice variants, isoforms, and post-translational modifications. This approach addresses key blind spots in genome-only or proteome-only analyses, such as the detection of sample-specific or disease-specific variants.

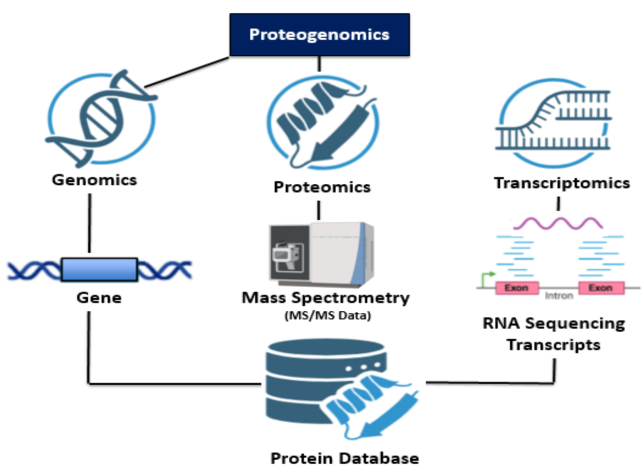

Figure 2: Schematic illustration of the proteogenomics paradigm

(*Self-created using BioRender's web-based software: https://www.biorender.com)*

The review compared the benefits and limitations of standalone versus integrated omics strategies. Table 1 presents a comparative overview, demonstrating that while standalone genomics offers insights into potential functional elements, and proteomics captures actual molecular function, it is their integration that enables functional validation of mutations, robust biomarker discovery, and deeper phenotypic characterisation. Integrated omics has led to advances in understanding cancer biology, neurodegenerative disease mechanisms, and host-pathogen interactions.

Table 1: Comparison of Standalone and Integrated Omics Approaches (16, 17, 18, 19, 20)

Feature	Standalone Omics (Genomics or Proteomics)	Integrated Omics (e.g., Proteogenomics)
Definition	Analysis of a single omic layer (e.g., genome, transcriptome, or proteome).	Concurrent analysis and interpretation of two or more omic layers.
Data Sources	DNA/RNA sequencing (genomics, transcriptomics), or mass spectrometry (proteomics).	Combined RNA-seq + MS/MS, often with epigenomics, metabolomics, or phenomics data.
Biological Insight	Offers a partial view (e.g., genetic potential or expressed proteins).	Provides systems-level understanding of biological processes and regulation.
Disease Mechanism Resolution	Identifies variants or protein levels separately.	Links genetic variation to protein function and phenotype.
Biomarker Discovery	Limited sensitivity; often false positives.	Enhanced specificity and confidence via cross-validation across omics.
Drug Target Identification	Based on sequence or expression only.	Integrates functional protein data with genomic context.
Data Interpretation Complexity	Relatively lower, single-layer analysis.	High; requires advanced statistical or ML-based methods.
Common Tools	STAR, HISAT2, MaxQuant, DESeq2.	Galaxy-P, MOFA+, OmicsIntegrator, ProteoDisco.
Clinical Translation	Often inconclusive; incomplete molecular context.	Higher translational value for diagnostics and therapeutics.
Scalability	Easier to scale; lower computational needs.	Requires high-performance computing and complex infrastructure.
Examples of Use	SNP profiling in cancer, differential protein expression in Alzheimer's.	CPTAC studies in oncology, Alzheimer's CSF proteogenomics, and COVID-19 host profiling.
Adoption in LMICs	More common due to lower cost and expertise barriers.	Emerging, with barriers related to cost, tools, and training.
Limitations	Misses regulatory interactions and post-translational changes.	High cost, technical complexity, and integration challenges.

The comparative analysis between proteogenomics and transcriptomics highlights distinct advantages and limitations in understanding gene function and disease mechanisms. Table 2 summarises the differences in focus, technologies, resolution, and applications. While transcriptomics offers broad RNA-level insights, proteogenomics provides a more accurate reflection of biological function by integrating protein data with genomic context.

Table 2: Comparison of Proteogenomics and Transcriptomics (21, 22, 23, 24, 25)

Feature	Proteogenomics	Transcriptomics
Definition	Integration of proteomic and genomic data to study the relationship between gene variants and protein expression.	Study of all RNA transcripts (mRNA, non-coding RNA) to understand gene expression.
Primary Focus	Protein expression, post-translational modifications, and variant peptides.	Gene expression profiling at the RNA level.
Data Sources	Mass spectrometry (MS/MS), next-generation sequencing (NGS), RNA-seq, reference proteomes.	RNA sequencing (RNA-seq), microarrays.
Key Technologies	MaxQuant, Galaxy-P, PeptideAtlas, Proteome Discoverer, AlphaFold.	STAR, Cufflinks, DESeq2, Salmon, EdgeR.

Strengths	Direct insight into functional proteins; identifies translational and post-translational events; detects novel peptides.	Sensitive to changes in gene expression, faster and less expensive, and scalable.
Limitations	Requires complex workflows, lower dynamic range, higher cost; sample preparation is critical.	Cannot detect post-transcriptional modifications; may not correlate with protein abundance.
Resolution	High specificity for protein isoforms and variants.	High sensitivity for low-abundance transcripts.
Integration Potential	Can integrate with genomics, transcriptomics, and metabolomics (multi-omics).	Integrates well with genomics and epigenomics.
Applications in Research	Cancer biomarker discovery, personalised medicine, and drug response analysis.	Developmental biology, disease pathway analysis, and immune profiling.
Applications in Clinical Settings	Tumour classification, therapy prediction, diagnostic protein panels.	Disease prognosis, identifying gene expression signatures, and therapy stratification.
Example Initiatives	CPTAC (Clinical Proteomic Tumour Analysis Consortium), H3Africa.	GTEx (Genotype-Tissue Expression), ENCODE, TCGA.

Furthermore, this review identified key domains where proteogenomics has significantly advanced both clinical and biological research. As illustrated in Table 3, the integration of proteomic and genomic data enhances cancer classification, enables early disease detection, and supports precision medicine. The approach has been instrumental in biomarker discovery, drug target identification, and understanding post-translational modifications. Applications also extend to infectious diseases, neurodegenerative disorders, and developmental biology. Collectively, these findings underscore the translational power of proteogenomics across multiple biological contexts and clinical settings.

Table 3: Applications of proteogenomics in clinical and biological research (26, 27, 28, 29, 30)

Application Area	Description	Impact	Example(s)
Cancer Research	Combines genomic mutations with proteomic alterations to better understand tumour biology.	Improves tumour classification, identifies driver mutations, and informs targeted therapy.	CPTAC studies in breast, lung, and colorectal cancers.
Biomarker Discovery	Identifies proteins uniquely associated with disease states by integrating transcriptomic and proteomic data.	Enhances early detection and diagnosis of complex diseases.	Discovery of CSF biomarkers in Alzheimer's disease.
Drug Target Identification	Maps altered signalling pathways by linking mutations to downstream protein effects.	Aids the development of novel therapeutics and repurposing of existing drugs.	EGFR and PI3K pathway targeting in lung and colon cancers.
Personalized Medicine	Tailor treatment strategies based on patient-specific genomic and proteomic profiles.	Enables stratified therapies with improved clinical outcomes.	Use of proteogenomic profiles to guide immunotherapy in melanoma.
Functional Annotation of Genomes	Refines gene models by confirming novel coding regions or isoforms via mass spectrometry.	Increases the accuracy of genome annotation and identification of uncharacterized genes.	Proteogenomic refinement of ENCODE annotations.
Understanding Post-translational Modifications (PTMs)	Identifies PTMs such as phosphorylation and glycosylation not evident in genomic data alone.	Reveals regulatory mechanisms of proteins and signalling dynamics.	Phosphoproteomics in signalling pathway analysis (e.g., mTOR, MAPK).
Infectious Disease Research	Maps host-pathogen interactions by studying pathogen and host proteomes alongside genomes.	Enhances vaccine development and antimicrobial target discovery.	SARS-CoV-2 proteogenomic studies identifying viral-host interactions.

Neurodegenerative Disorders	Combining omics to identify molecular changes in neurodegeneration and their progression.	Provides insight into disease aetiology and staging.	Alzheimer's and Parkinson's disease molecular profiling.
Developmental Biology	Integrates spatiotemporal gene and protein expression during organism development.	Enhances understanding of tissue-specific gene regulation.	Mouse embryogenesis proteogenomic mapping.
Population Genomics & Epidemiology	Merges genotypic variation with proteomic phenotype across diverse populations.	Supports global health genomics and disease susceptibility studies.	H3Africa and GTEx proteogenomic initiatives.

Proteogenomics offers significant advantages (Table 4) in biological research, providing a more comprehensive understanding of molecular mechanisms than either approach alone. By combining genetic information with protein expression and modifications, researchers can uncover functional impacts of genetic variants, leading to improved biomarker discovery and enhanced drug target identification. This integrative approach facilitates personalised and precision medicine by aligning genomic risk profiles with proteomic data, optimising treatment strategies. Additionally, it enables deeper insights into complex cellular processes, such as post-translational modifications and signalling pathways, thereby advancing systems biology and clinical applications across diverse diseases.

Table 4: Advantages of integrating genomics and proteomics (31, 32, 33, 34, 35)

Advantage	Description	Benefit	Example
Comprehensive Molecular Insight	Combines genetic information with protein expression and modifications for a fuller biological picture	Enables understanding of how genetic variants translate into functional proteins and phenotypes	Identifying disease mechanisms in cancer and neurodegenerative diseases
Improved Biomarker Discovery	Joint analysis enhances the detection of reliable biomarkers by correlating gene mutations with protein changes	Increases diagnostic accuracy and early disease detection	CPTAC's proteogenomic biomarkers for NSCLC
Enhanced Drug Target Identification	Integrative data pinpoint proteins and pathways influenced by genetic alterations suitable for targeted therapy	Facilitates the discovery of novel therapeutic targets and personalised treatments	Targeting HER2+ breast cancer in Sub-Saharan Africa
Personalised and Precision Medicine	Aligns genomic risk profiles with proteomic data to tailor treatment plans at individual levels	Optimises therapy selection and predicts drug response	Precision oncology and pharmacogenomics
Decoding Post-Translational Modifications (PTMs)	Proteomics reveals PTMs absent in genomics alone, essential for understanding protein function	Clarifies complex cellular signalling and regulation	Phosphorylation mapping in cancer signalling pathways
Single-Cell and Spatial Omics Integration	Combines genomics and proteomics at cellular resolution, preserving tissue context	Reveals heterogeneity and microenvironmental influences	Single-cell proteogenomics in the tumour microenvironment
Advances in Systems Biology	Integration supports building multi-layered biological networks, enabling systems-level understanding	Facilitates the discovery of novel pathways and interaction networks	Network modelling of disease pathways

The technical approaches and tools for integrating genomics and proteomics are presented in Table 5. Our synthesis highlights a diverse and evolving ecosystem of technical tools and workflows enabling proteogenomics. Key proteomics platforms such as MaxQuant and Proteome Discoverer are commonly used alongside genomic tools like STAR and GATK for variant analysis and expression profiling. Integration platforms—such as Galaxy-P, MOFA+, and OmicsIntegrator—support multi-dimensional data harmonisation. Additionally, artificial intelligence tools like AlphaFold and DeepVariant are enhancing predictive accuracy in structural and variant analyses. These tools collectively facilitate comprehensive, systems-level insights into complex biological processes and disease mechanisms.

Furthermore, integrating diverse omics data requires sophisticated workflows and analytical platforms. Figure 3 provides a schematic overview of a multi-omics data integration pipeline. This includes sample collection, extraction of DNA/RNA/proteins, sequencing, mass spectrometry, database construction, data alignment, and functional analysis. Computational tools such as MaxQuant, Galaxy-P, MOFA+, OmicsIntegrator, and AlphaFold facilitate tasks such as peptide identification, protein quantification, and molecular interaction prediction, streamlining the process of integrating and interpreting complex datasets.

Figure 3: Schematic Workflow of Multi-Omics Data Integration Pipeline

Table 5: Technical approaches and tools for integrating genomics and proteomics (35, 36, 37, 38, 39, 40)

Category	Tool/Platform	Function/Description	Application Area
Proteomics Data Analysis	MaxQuant	Quantifies peptides and proteins from MS/MS data; supports label-free and SILAC workflows	Protein identification, quantification, and PTM analysis
	Proteome Discoverer	Offers flexible workflow management for MS data, integrated with SEQUEST, Percolator	High-throughput proteomics and interactome mapping
Genomics Data Analysis	STAR	Ultrafast RNA-seq aligner; maps reads to reference genome	Transcriptome profiling, expression analysis
	GATK	Variant discovery in high-throughput sequencing data	Mutation analysis, SNP/indel detection
Proteogenomics Pipelines	Galaxy-P	Web-based platform for accessible proteogenomic data integration	MS/MS spectrum searching, variant peptide discovery
	PeptideAtlas	Repository and reanalysis tool for MS data with genomic annotation	Reference proteome generation, spectral libraries
Multi-Omics Integration	MOFA+	Matrix factorisation framework for unsupervised integration of multi-omics datasets	Pattern discovery, latent factor identification
	OmicsIntegrator	Integrates proteomic, genomic, and other omics data into biological networks	Disease module identification, pathway enrichment
Machine Learning & AI	AlphaFold	Predicts 3D protein structures from amino acid sequences using deep learning	Protein structure prediction, functional annotation
	DeepVariant	Uses deep learning for highly accurate variant calling	Variant discovery in genome and exome data

Visualisation and Workflow Tools	Cytoscape	Open-source tool for visualising molecular interaction networks	Multi-omics network integration, pathway mapping
	TIBCO Spotfire	An interactive visual analytics platform supporting omics data	Data exploration and visual trend analysis
Cloud-Based Platforms	Terra.bio (by Broad)	Cloud-native platform for scalable genomic and proteomic data analysis	Collaborative data analysis and sharing
	DNAnexus	Cloud platform for large-scale genomic and proteogenomic pipelines	Clinical genomics, population-scale studies

Despite the promise of integrative omics, several challenges persist. These range from data complexity, scale, and heterogeneity to issues of interoperability, standardisation, and computational demands. Table 6 summarises key challenges and their corresponding mitigation strategies. For example, cloud-based infrastructures and federated learning are helping overcome computational bottlenecks, while machine learning models are enabling deeper pattern recognition and predictive analytics.

Table 6: Challenges in Integrative Omics and Corresponding Mitigation Strategies (41, 42, 43, 44, 45)

Category	Key Challenges	Mitigation Strategies
Data Complexity & Volume	High-dimensional datasets Heterogeneous data types Missing values	Use of dimensionality reduction (e.g., PCA, t-SNE, MOFA+) Data preprocessing pipelines Imputation tools
Standardization Issues	Non-uniform data formats Lack of interoperable metadata standards	Adoption of FAIR data principles Use of standard ontologies (e.g., MIAME, HUPO PSI) GA4GH frameworks
Computational Bottlenecks	Intensive processing demands Limited access to HPC or cloud infrastructure	Use of cloud platforms (e.g., DNAnexus, Terra.bio) National HPC grants Optimised workflows
Analytical Integration	Inconsistent correlation between omics layers Tool fragmentation	Ensemble modelling Integrated platforms (e.g., Galaxy-P, OmicsPipe) Network-based methods
Interpretation Challenges	Discordance between gene and protein levels Limited functional validation	Functional enrichment (GO, KEGG) Experimental validation (e.g., Western blot, CRISPR screens)
Cost & Resource Burden	Expensive equipment & reagents Maintenance & training costs	Collaborative consortia (e.g., CPTAC, H3Africa) Open-access tools Local infrastructure investment
Ethical & Legal Constraints	Data privacy concerns GDPR, HIPAA compliance Consent limitations	Dynamic consent models Blockchain for data traceability Institutional ethical training
Equity & Access (LMICs)	Underrepresentation in datasets Infrastructure & skill gaps	Equitable data sharing policies Capacity-building programs Local PI involvement in global projects
Reproducibility	Irreproducible computational results Lack of method transparency	Use of version-controlled workflows (e.g., Nextflow, Snakemake) Public code/data repositories (e.g., GitHub)
Policy & Regulation	Jurisdictional differences in data use Undefined IP rights	Harmonised international guidelines Legal support in data-sharing agreements

The global and translational relevance of integrative omics cannot be overstated. As outlined in Table 7, integrative genomics and proteomics have impacted various domains, including global health surveillance (e.g., COVID-19), cancer precision therapy, and biomarker-driven drug development. In low- and middle-income countries (LMICs), initiatives such as H3Africa have begun to bridge the omics divide, promoting genomic equity and fostering regional biobanking and computational infrastructure.

Table 7: Global and Translational Impact of Integrative Genomics and Proteomics (46, 47, 48, 49, 50)

Area of Impact	Description	Examples/Case Studies	Translational Relevance
Infectious Disease Surveillance	Integration helps decode host-pathogen interactions, identify virulence factors, and track pathogen evolution.	SARS-CoV-2 genomic surveillance, proteomic mapping of immune responses.	Enables rapid diagnostics, vaccine development, and epidemic control.
Cancer Precision Medicine	Proteogenomics identifies driver mutations, altered pathways, and therapeutic targets.	CPTAC's work in breast and lung cancer, identification of proteomic subtypes in NSCLC.	Informs personalised treatment and biomarker-guided therapies.
Neurodegenerative Diseases	Multi-omics elucidates disease heterogeneity, protein aggregation, and genetic predispositions.	Alzheimer's proteogenomic CSF profiling; APOE-related pathway studies.	Aids in early diagnosis and therapeutic stratification.
Global Health and LMIC Contexts	Omics projects tailored to local disease burden, genetic diversity, and infrastructure constraints.	H3Africa, South African Human Genome Programme, Nigeria's Sickle Cell Genomics initiative.	Supports evidence-based interventions and regional biobank development.
Antimicrobial Resistance (AMR)	Tracks resistance genes and proteomic expression in pathogens.	Integrative surveillance in *Mycobacterium tuberculosis* and *Salmonella*.	Guides stewardship and novel antimicrobial design.
Maternal and Child Health	Identifies biomarkers for preeclampsia, preterm birth, and genetic birth disorders.	Placental omics studies: integration of maternal proteomics with foetal genomics.	Enables early interventions and risk assessment models.
AI-Driven Diagnostics	Machine learning models trained on omics datasets improve pattern recognition and disease classification.	Use of AlphaFold for protein structure prediction; AI-driven cancer classifiers.	Accelerates clinical decision-making and improves diagnostic accuracy.
Nutrition and Metabolic Disorders	Connects genetic predispositions with proteomic metabolic signatures.	Studies on obesity, diabetes, and micronutrient deficiencies in LMICs.	Promotes targeted dietary guidelines and therapeutic monitoring.
Public Health Genomics	Population-scale integration of genomic and proteomic data for epidemiological modelling.	GTEx, UK Biobank, ChinaMAP, and cross-continental consortia including LMICs.	Enables stratified healthcare planning and resource allocation.
Translational Capacity Building	Involves training, open-access databases, and collaborative platforms for data sharing and innovation.	NIH-Fogarty training programs, H3ABioNet, African CDC genomic fellowship.	Fosters local expertise, sustainable research ecosystems, and equitable technology access.

Real-world case studies from CPTAC (Clinical Proteomic Tumour Analysis Consortium) demonstrate how proteogenomics has transformed cancer subtype classification and therapeutic targeting. Similarly, applications in neurodegenerative diseases like Alzheimer's disease have benefited from proteomic resolution of cerebrospinal fluid biomarkers, deepening diagnostic capabilities and research precision. Despite recent advancements, important knowledge gaps remain. These include limited data from LMIC populations, insufficient integration of single-cell and spatial omics, and underexplored ethical and legal frameworks in multi-omics research. Table 8 categorises these gaps, outlines their impacts, and proposes future research directions. Recommendations include advancing data harmonisation, enhancing representation of diverse populations, developing ethical AI frameworks, and strengthening capacity building in under-resourced settings.

Table 8: Knowledge Gaps, Associated Impact, and Future Research Directions in Integrative Genomics and Proteomics (50, 51, 52, 53, 54, 55)

Knowledge Gap	Associated Impact	Future Research Direction
Limited multi-omics integration across layers (e.g., genomics, epigenomics, proteomics, metabolomics)	Incomplete biological understanding; fragmented insights into complex diseases	Develop standardised pipelines for multi-omics integration; support federated and cross-platform models
High data volume and complexity	Difficulty in analysis, interpretation, and reproducibility	Invest in scalable computational infrastructure, cloud computing, and AI-based analytic frameworks
Underrepresentation of LMIC populations in datasets	Limited generalizability of findings; inequitable health outcomes	Expand omics consortia in LMICs (e.g., H3Africa); include diverse populations in global datasets
Limited access to advanced tools and workflows	Disparities in research capacity and innovation	Promote open-source software (e.g., Galaxy-P, MOFA+); fund bioinformatics training programs in resource-limited settings
Lack of robust data integration strategies	Hinders the discovery of functional relationships and pathways	Employ network modelling, machine learning, and dimensionality reduction techniques
Inconsistent technical standards across studies	Poor comparability and cross-study validation	Develop global quality standards for sample processing, data acquisition, and annotation
Insufficient ethical and legal frameworks in LMICs	Risk of privacy breaches, data misuse, and community mistrust	Create context-specific ELSI frameworks; promote informed consent, data sovereignty, and benefit sharing
Low genomic and proteomic literacy among researchers and the public	Mistrust in research, poor uptake of omics-driven solutions	Support educational outreach, curriculum development, and public engagement initiatives
Cost-intensive technologies and reagents	Limits adoption in LMICs; widens the innovation gap	Encourage frugal innovation; leverage public-private partnerships for subsidised technology access
Limited application of emerging tech (e.g., single-cell/spatial omics)	Missed opportunities in understanding microenvironment and cell heterogeneity	Expand affordable single-cell and spatial omics platforms; apply to infectious and chronic disease models
Sparse translational research and clinical pipelines	Slowed implementation of biomarkers and therapeutics	Develop clinical-grade proteogenomic pipelines; fund translational science centres in underserved regions

Discussion

Proteogenomics has emerged as a transformative approach in biological research, offering deeper insights into the complex mechanisms that govern cellular function, disease progression, and therapeutic response (1, 4). While genomics provides a blueprint of an organism's genetic potential, proteomics reveals the dynamic expression and functional states of proteins, often the actual effectors of biological activity. Together, these complementary technologies enable a more comprehensive understanding of biological systems than either could achieve alone. In recent years, advances in sequencing technologies, mass spectrometry, and computational biology have facilitated unprecedented opportunities for multi-omics integration (9, 10). However, the practical application of these technologies remains fraught with challenges, including data complexity, technological limitations, and interpretational gaps. This discussion synthesises current findings on the advantages and challenges of integrating genomics and proteomics, while also identifying key knowledge gaps and projecting future directions in this rapidly evolving field.

The Proteogenomics Paradigm

Proteogenomics has given rise to an evolving field known as proteogenomics, a hybrid approach that combines genomic and proteomic data to enhance the identification, annotation, and functional understanding of proteins within biological systems (3). This integration leverages the strengths of both genomics (which provides the blueprint of an organism) and proteomics (which reflects the dynamic biological processes at the protein level), bridging the gap between genetic potential and phenotypic reality.

Proteogenomics refers to the interdisciplinary approach that utilises high-throughput mass spectrometry-based proteomics data in conjunction with DNA and RNA sequencing data to improve genome annotation and uncover novel protein-coding regions, splice variants, and post-translational modifications (PTMs). At its core, proteogenomics involves generating customised protein databases derived from genomic and transcriptomic data (e.g., whole-genome sequencing or RNA-seq), which are then used to interpret mass spectrometry data for more accurate peptide identification. This approach addresses one of the longstanding limitations of proteomics: the dependence on pre-existing protein databases that may not reflect individual-specific or disease-specific protein variants. By incorporating sample-specific genomic and transcriptomic data, proteogenomics enhances the depth and precision of protein identification, particularly in complex diseases like cancer, where mutations, alternative splicing, and PTMs drive pathological outcomes (10, 12).

Comparison of Standalone and Integrated Omics Approaches

Standalone omics approaches, such as genomics, transcriptomics, and proteomics, have each contributed uniquely to understanding biological systems (16). Genomics identifies genetic variants, transcriptomics reveals gene expression dynamics, and proteomics provides insights into actual protein abundance and modifications (17). However, when used in isolation, these approaches offer a fragmented view of biological complexity. Integrated omics, especially proteogenomics, on the other hand, bridges this gap by combining datasets to provide a holistic view of molecular processes (18). For example, genomics can predict protein-coding potential, but only proteomics confirms actual protein expression. Integration improves biomarker discovery, disease classification, and mechanistic understanding, especially in complex diseases like cancer and neurodegeneration. Computational tools such as MOFA+, MaxQuant, and OmicsIntegrator enable seamless multi-omics data integration using machine learning and network modelling (19). This synergy has practical applications in personalised medicine, aligning genomic risk factors with protein-level expression and therapeutic response. While integration increases data complexity and cost, it offers more robust and actionable insights than standalone methods. Its growing use in projects like CPTAC and H3Africa underscores its translational potential. Despite resource challenges in low-income settings, integrated omics holds the key to advancing precision medicine, diagnostics, and targeted therapies globally (20).

Comparison of Proteogenomics and transcriptomics

Proteogenomics and transcriptomics serve complementary roles in modern biomedical research. Transcriptomics enables rapid and scalable profiling of gene expression, aiding in pathway analysis and biomarker discovery (21, 22). However, it does not always correlate with protein abundance or capture post-transcriptional regulation. In contrast, proteogenomics integrates proteomic and genomic data to uncover functional protein products and novel variant peptides, offering deeper insights into disease mechanisms. (23, 24). Though technically complex and resource-intensive, proteogenomics is more predictive of phenotypic outcomes. Its applications in cancer research, such as through CPTAC, have advanced tumour classification and therapeutic targeting. Integrating both approaches can overcome individual limitations, enabling a systems-level understanding of biology (25).

Applications of Proteogenomics

Proteogenomics represents a transformative advance in modern biomedical research. By combining genetic information with protein-level expression data, researchers gain a more holistic view of cellular function, disease mechanisms, and therapeutic opportunities (26). This multidimensional perspective is particularly impactful in complex disease states like cancer, where genomic mutations alone often fail to fully explain phenotypic outcomes. The applications of proteogenomics include, and not limited to:

1. Cancer Research and Precision Oncology

One of the most prominent applications of proteogenomics is in oncology, where it has been used to identify novel tumour-specific peptides, neoantigens, and biomarkers. Projects such as the Clinical Proteomic Tumour Analysis Consortium (CPTAC) have successfully applied proteogenomics to various cancer types (e.g., breast, lung, colorectal), revealing actionable molecular targets and improving cancer subtype classification. By integrating somatic mutations with downstream protein expression profiles, proteogenomics enables the identification of driver mutations and their phenotypic consequences, facilitating the development of targeted therapies and immunotherapies (26).

2. Biomarker Discovery and Disease Diagnosis

Proteogenomics enhances the discovery of disease-specific biomarkers by detecting aberrant protein isoforms and fusion proteins that may not be identifiable through genomics alone. This is particularly relevant in diseases characterised by complex splicing events or RNA editing. For example, in neurodegenerative diseases like Alzheimer's, proteogenomics has uncovered novel proteoforms linked to pathogenesis, opening new diagnostic and therapeutic avenues (27).

3. Personalised and Precision Medicine
The field is increasingly contributing to the development of personalised medicine, where individual genetic and proteomic profiles are used to tailor treatment strategies. In this context, proteogenomics provides a more comprehensive view of disease mechanisms by linking genetic variation to proteomic expression, aiding clinicians in making informed decisions about drug selection, dosing, and monitoring therapeutic response (28).

4. Functional Annotation of Genomes
Proteogenomics plays a crucial role in refining genome annotations by validating predicted open reading frames (ORFs), identifying non-canonical translation initiation sites, and confirming novel gene models. This is particularly important for under-characterised genomes, such as those from emerging pathogens, non-model organisms, or individuals from diverse ancestries (29).

5. Vaccine Development and Infectious Disease Research
In infectious disease settings, proteogenomics supports epitope mapping and vaccine design by identifying pathogen-specific proteins that are expressed during infection. During the COVID-19 pandemic, integrative omics approaches, including proteogenomics, were pivotal in mapping viral-host interactions and informing vaccine and antiviral drug development.
No doubt, proteogenomics represents a transformative advance in life science research, offering a more nuanced understanding of how genetic information is translated into functional biology. By capturing the interplay between the genome and proteome, this integrative approach enhances our ability to characterise disease mechanisms, improve diagnostic accuracy, and develop personalised therapeutic interventions. As technologies continue to evolve, particularly in areas like single-cell analysis, spatial omics, and AI-driven analytics, proteogenomics is poised to play a central role in shaping the future of biomedical research and precision medicine (30).

Advantages of Genomes and Proteomics Integration
Proteogenomics has become a cornerstone of modern biological research and translational medicine (31). This synergy offers a holistic understanding of cellular function by combining DNA-level insights with real-time protein activity, thus overcoming the limitations of standalone omics approaches. While genomics reveals what *could* happen by decoding the static blueprint of life, proteomics discloses what *is happening*, capturing the dynamic state of the cell through post-translational modifications, interactions, and abundance profiles. When integrated, these technologies unlock multidimensional insights with significant scientific and clinical implications.

1. Functional Synergy between Genomics and Proteomics
At the core of integrative omics lies the functional synergy between gene and protein expression. Genomics identifies mutations, copy number variations, and gene expression patterns, yet it cannot predict how those changes translate into phenotypic outcomes. Proteomics, conversely, provides insight into protein isoforms, post-translational modifications (PTMs), and interaction networks, many of which are regulated independently of mRNA levels. For instance, a gene may be upregulated without corresponding increases in protein levels due to miRNA interference or protein degradation pathways. Proteogenomics bridges this discordance, allowing researchers to correlate genetic alterations with actual protein products and functions. This integration is particularly vital in oncology, where oncogenic mutations often drive aberrant protein networks that genomics alone cannot capture (32).

2. Improved Biomarker Discovery
Biomarkers play an essential role in early diagnosis, prognosis, and therapeutic monitoring. However, biomarkers identified solely through genomics often lack specificity due to transcriptional noise and the gap between mRNA and protein levels. The integration of proteomics adds an essential layer of validation, increasing confidence in candidate biomarkers by confirming their expression at the protein level. A landmark example is the Clinical Proteomic Tumour Analysis Consortium (CPTAC), which demonstrated that integrating proteomic data with genomic profiles in breast, ovarian, and colon cancers could uncover novel proteogenomic subtypes (1). In colon cancer, CPTAC identified unique proteomic patterns corresponding to microsatellite instability not evident in the genomic landscape alone (2).
In Alzheimer's disease, for instance, integrative analyses combining genome-wide association studies (GWAS) and cerebrospinal fluid (CSF) proteomics have identified risk-associated genes (e.g., *APOE*, *BIN1*) and corresponding dysregulated proteins, improving early detection and mechanistic understanding (3, 31).

3. Drug Target Identification and Validation
Proteogenomics expedites drug target discovery by linking genotypic variations with functional protein outputs. Genomics can pinpoint mutations or expression patterns associated with disease, but proteomics confirms whether the resulting proteins are druggable, e.g., their structure, abundance, location, or PTMs. For instance, in non-small cell lung cancer (NSCLC), proteogenomic profiling has uncovered unique mutations in the *EGFR* gene and correlated these with phosphoproteomic data, guiding the use of tyrosine kinase inhibitors (4). Moreover, proteomics can help predict off-target effects and resistance mechanisms by examining protein interaction networks. In low- and middle-income countries (LMICs), where drug repurposing is vital due to limited resources, integrative

omics has enabled identification of shared molecular targets across diseases such as malaria, tuberculosis, and cancers, optimising existing therapies (5, 32).

4. Personalised and Precision Medicine
A major promise of integrative omics lies in its potential to personalise diagnosis and therapy. Precision medicine aims to tailor treatment strategies based on an individual's molecular profile. Genomics alone often lacks the resolution to account for disease heterogeneity, especially in multifactorial conditions like cancer, diabetes, and neurodegenerative diseases. By combining patient-specific genomic alterations with proteomic phenotypes, clinicians can stratify patients more effectively, predict therapeutic responses, and minimise adverse effects. A notable clinical implementation is in breast cancer subtyping: while genomic assays (e.g., Oncotype DX) assess recurrence risk, adding proteomic data refines therapeutic decisions, especially for borderline cases (6). Further, integrated profiles can reveal mechanisms of drug resistance, a major challenge in oncology. For example, proteogenomic studies in ovarian cancer patients receiving PARP inhibitors have identified compensatory signalling pathways activated at the protein level, guiding combination therapies (7, 32).

5. Single-Cell and Spatial Omics Integration
Recent advancements in single-cell and spatial omics have further enhanced integrative strategies. Traditional bulk analyses average signals across heterogeneous cell populations, masking crucial variability. Single-cell genomics and proteomics capture variability at cellular resolution, while spatial techniques preserve tissue context, enabling the mapping of molecular signatures *in situ*. Proteogenomic integration at the single-cell level is particularly transformative in immuno-oncology. For instance, mapping tumour microenvironment heterogeneity in melanoma has revealed distinct immune cell subtypes with unique transcriptomic and proteomic features, informing immune checkpoint inhibitor strategies [8]. Moreover, in brain research, combining spatial transcriptomics with region-specific proteomics has illuminated neurodegenerative trajectories in diseases like Parkinson's and Alzheimer's, offering region- and cell-type-specific targets (9, 33).

6. Advantages in LMIC Contexts and Global Health
In resource-constrained settings, integrative omics offers a cost-effective pathway to maximise data utility. While full-scale proteogenomic platforms may be challenging, tiered approaches, e.g., low-pass sequencing coupled with targeted proteomics, can yield valuable insights into local disease patterns. Collaborations like H3Africa and Africa CDC's Pathogen Genomics Initiative are beginning to incorporate proteomics into their workflows to monitor disease outbreaks, enhance pharmacogenomic surveillance, and understand genetic diversity in African populations (10). Integration of these data layers will be crucial for addressing global disparities in healthcare research and implementation.

7. Systems Biology and Network Modelling
Integrative omics underpins systems biology approaches, enabling the construction of comprehensive molecular interaction networks. These models go beyond linear cause-effect relationships to identify regulatory hubs, pathway crosstalk, and emergent behaviours of biological systems. For example, MOFA+ (Multi-Omics Factor Analysis) and OmicsIntegrator are computational frameworks that combine genomic, transcriptomic, and proteomic data to infer latent factors driving disease (11, 12). These tools have already been applied in complex diseases like rheumatoid arthritis and cardiomyopathy to reveal therapeutic opportunities that were previously obscured (34).
Proteogenomics represents a paradigm shift in biological research and clinical practice. It offers unparalleled depth in characterising disease mechanisms, refining biomarker discovery, validating therapeutic targets, and advancing personalised medicine (e.g., genomics-based cancer therapies). While standalone omics have provided foundational knowledge, only through their synergy can we capture the complexity and dynamism of biological systems. Continued investment in integrative methodologies, data infrastructure, and equitable access, especially in LMICs, will be vital for translating these scientific insights into global health solutions (35).

Multi-Omics Integration Frameworks
Multi-omics integration frameworks, particularly those combining genomics, proteomics, and transcriptomics- offer transformative potential in advancing precision medicine and public health, especially in low- and middle-income countries (LMICs). In Sub-Saharan Africa, initiatives like the African Genomics Initiative and H3Africa have laid critical groundwork by generating population-specific genomic data. These efforts are increasingly embracing proteogenomics, especially in oncology. For example, integrating proteomic and genomic profiles of HER2+ breast cancer in African women has revealed unique molecular subtypes, informing more context-specific diagnostic and therapeutic strategies (7).
During the COVID-19 pandemic, multi-omics frameworks were pivotal for real-time pathogen surveillance. Genomics tracked viral evolution and spread, while proteomics helped decode host-pathogen interactions and immune responses. In LMICs, partnerships with global consortia enabled the deployment of integrative tools for early detection, vaccine target identification, and epidemiological modelling, demonstrating the value of accessible, cross-omic infrastructures in crisis response. These examples underscore the growing relevance of multi-omics in global health equity and precision public health, while also highlighting the need for sustainable investment in data

infrastructure, training, and ethical frameworks (10, 11, 12, 13).

Case Studies in Multi-Omics Integration

1. CPTAC and Biomarker Discovery in Non-Small Cell Lung Cancer (NSCLC)

The Clinical Proteomic Tumour Analysis Consortium (CPTAC) has been instrumental in advancing multi-omics research, particularly in oncology. One of its landmark achievements includes the integrative analysis of non-small cell lung cancer (NSCLC) using proteogenomics. By combining genomic sequencing, transcriptomic data, and mass spectrometry-based proteomics, CPTAC identified novel protein biomarkers and mutation-driven signalling pathways that were not apparent through genomic data alone. This approach improved tumour classification, revealed therapeutic targets, and provided a clearer understanding of tumour heterogeneity, paving the way for precision medicine in lung cancer (7).

2. Proteogenomic Profiling in Alzheimer's Disease

In neurodegenerative disease research, particularly Alzheimer's disease (AD), proteogenomic approaches have been used to analyse cerebrospinal fluid (CSF) samples. By integrating RNA-seq data from brain tissue with proteomic signatures from CSF, researchers have uncovered dysregulated pathways involved in synaptic dysfunction, inflammation, and protein misfolding. These findings are refining biomarker panels for early detection and providing insights into disease progression at both molecular and systemic levels—advancing therapeutic strategies for AD (2, 8).

3. Integrated Omics in Breast Cancer Therapy in Sub-Saharan Africa

In Sub-Saharan Africa, the application of multi-omics, particularly in HER2-positive breast cancer, is gaining traction through regional initiatives and collaborations. Integrative analyses combining whole-exome sequencing, transcriptomics, and proteomics have revealed ethnically distinct molecular signatures, leading to the development of tailored treatment approaches. This precision strategy is crucial in regions with limited access to standardised care and emphasises the importance of population-specific multi-omics data to address disparities in cancer outcomes (9, 13).

Technical Approaches and Tools for Integrating Genomics and Proteomics

Successful proteogenomics relies not only on conceptual synergy but also on robust technical workflows and computational tools. This section outlines the primary methodologies that enable such integration, covering experimental workflows (e.g., mass spectrometry and RNA-sequencing), software tools (e.g., MaxQuant, Galaxy-P), and advanced data integration strategies including machine learning and network modelling (35).

1. Proteogenomics Workflows

Proteogenomics pipelines combine data from RNA sequencing (RNA-seq) and mass spectrometry-based proteomics (MS/MS) to construct customised protein databases and enhance peptide identification. The typical proteogenomics workflow includes the following steps:

a. RNA-Seq and Custom Protein Databases

High-throughput RNA-seq is employed to capture the transcriptomic landscape of a biological sample, providing evidence for canonical transcripts, alternative splicing events, single-nucleotide variants (SNVs), and gene fusions. This data is then translated *in silico* into protein sequences to generate sample-specific or disease-specific proteome databases (1). Tools such as RSEM, STAR, and HISAT2 assist in mapping and quantifying transcripts, which are subsequently converted to protein sequences using open-source software like TransDecoder (36).

b. Mass Spectrometry (MS/MS)

Mass spectrometry, particularly tandem MS/MS, enables high-throughput and high-accuracy identification and quantification of peptides and proteins. Shotgun proteomics, using data-dependent acquisition (DDA) or data-independent acquisition (DIA), produces spectra that are matched against the custom protein databases derived from RNA-seq, enhancing the detection of non-canonical peptides, splice variants, and PTMs (2, 32).

c. Reference Proteomes and Variant Detection

Standard proteomic analyses often rely on reference databases such as UniProt or RefSeq, which may not account for patient- or disease-specific variants. Integrating RNA-seq allows for the identification of novel peptides, offering a more comprehensive and accurate view of the proteome. This is especially crucial in cancer research, where tumour-specific mutations result in unique neoantigens with diagnostic or therapeutic relevance (3).

2. Computational Tools for Proteogenomic Integration

Several computational platforms have been developed to support the complex tasks of aligning, filtering, identifying, and integrating multi-omic datasets. The most widely used tools include:

a. MaxQuant

MaxQuant is a powerful open-source software for analysing large-scale mass spectrometry data. It performs peak detection, peptide identification, protein quantification, and PTM mapping. In proteogenomics, MaxQuant can be used with custom FASTA files from RNA-seq data to identify novel peptides and splice isoforms. It also integrates well with Perseus for downstream statistical and bioinformatic analyses (4).

b. Galaxy-P
Galaxy-P is a user-friendly, web-based platform tailored for proteogenomics workflows. Built on the Galaxy Project infrastructure, it supports reproducible and scalable analysis of both genomics and proteomics data. Galaxy-P features modular workflows for converting RNA-seq data into protein FASTA files, running peptide searches against custom databases, and visualising peptide-spectrum matches (PSMs) (5). Its accessibility and graphical interface make it particularly attractive for users in low-resource environments.

c. AlphaFold and Structural Insights
AlphaFold, developed by DeepMind, uses deep learning to predict 3D protein structures from amino acid sequences with remarkable accuracy. In proteogenomics, AlphaFold facilitates the structural validation of novel proteins and variants discovered via MS/RNA-seq. It allows researchers to model protein function, drug-binding potential, and structural consequences of disease mutations (6, 35).

d. MOFA+ (Multi-Omics Factor Analysis)
MOFA+ is a statistical framework for discovering latent factors that explain variability across different omic layers. It performs unsupervised integration of genomics, transcriptomics, and proteomics to uncover shared and unique patterns, making it ideal for discovering patient subtypes, molecular mechanisms, and regulatory interactions. MOFA+ has been successfully applied in diseases like chronic lymphocytic leukaemia and Alzheimer's (7).

e. OmicsIntegrator
OmicsIntegrator employs network-based modelling to combine diverse omic datasets with prior biological knowledge (e.g., protein-protein interaction databases). It reconstructs probable biological pathways connecting observed mutations to downstream effects, helping prioritise therapeutic targets and interpret molecular mechanisms (36).

3. Data Integration Strategies
Integrating multi-omic datasets requires specialised computational approaches to handle the volume, heterogeneity, and complexity of the data. The major strategies include:

a. Machine Learning and Artificial Intelligence
Machine learning (ML) approaches are increasingly used for pattern recognition, classification, and prediction in multi-omic data integration. Algorithms such as random forests, support vector machines (SVMs), and deep neural networks can identify biomarkers, stratify patients, and predict therapeutic outcomes. For instance, in lung adenocarcinoma, ML models trained on integrated proteogenomic data outperformed genomics-alone models in predicting immunotherapy response [9]. Further, AutoML platforms like Google's AutoML Tables are now being used to automate model selection and tuning, enhancing accessibility for non-experts (37).

b. Network-Based Modelling
Network modelling methods conceptualise biological systems as graphs of interacting molecules. Tools such as Cytoscape, STRING, and GeneMANIA allow visualisation and exploration of complex interaction networks derived from integrated data. Network-based integration not only helps identify central hubs and regulatory bottlenecks but also allows contextual interpretation of variant and expression data within pathways. This is especially helpful in precision oncology for identifying druggable pathways and resistance mechanisms (38).

c. Dimensionality Reduction and Multi-Omic Fusion
Techniques such as principal component analysis (PCA), canonical correlation analysis (CCA), and non-negative matrix factorisation (NMF) reduce data dimensionality and identify latent variables shared across omic layers. These are often used in conjunction with visualisation tools like t-SNE and UMAP to explore molecular landscapes and clusters. Some platforms also support "data fusion," where omic datasets are concatenated or hierarchically merged to form a single analytical matrix. While powerful, this approach requires careful normalisation and bias correction (11, 39).

4. Data Standards and FAIR Principles
A critical enabler of integration is the adoption of data standards and FAIR (Findable, Accessible, Interoperable, and Reusable) principles. Repositories like PRIDE (for proteomics) and GEO/SRA (for genomics) support data sharing and validation. Ontologies such as HUPO-PSI and OBI ensure consistent annotation, facilitating cross-platform analyses (12). Proteogenomics is made possible by a rich array of experimental and computational tools. From RNA-seq and MS workflows to sophisticated software like MaxQuant, Galaxy-P, and MOFA+, these tools have enabled a new era of systems biology and precision medicine. Machine learning and network modelling offer advanced strategies for data synthesis, ensuring that biological insights are both accurate and actionable. As technologies continue to evolve, the need for standardisation, scalability, and inclusivity, especially for LMIC researchers, will become ever more pressing. Future efforts should focus on improving accessibility to these platforms, training interdisciplinary scientists, and expanding integrative pipelines into clinical and global health contexts (40).

Challenges and Limitations of Integrating Genomics and Proteomics Technologies
While proteogenomics has heralded unprecedented insights into biological systems and human health, the

implementation of these technologies at scale is fraught with several challenges. These barriers span data complexity, computational and analytical limitations, economic constraints, and deeply rooted ethical and legal considerations—especially in the context of low- and middle-income countries (LMICs) (41).

1. Data Complexity and Volume

The volume of data generated from genomics and proteomics platforms is staggering. Whole genome sequencing (WGS), RNA sequencing (RNA-seq), and mass spectrometry (MS)-based proteomics generate terabytes of data per study. This complexity multiplies exponentially when multi-omics datasets are integrated, including transcriptomics, epigenomics, and metabolomics. Heterogeneity of data is a primary concern. Genomic data is mostly discrete and static, reflecting DNA sequences that remain constant across time and tissues. In contrast, proteomic data is highly dynamic, context-dependent, and influenced by post-translational modifications (PTMs), environmental cues, and tissue-specific expression. Reconciling these differences requires advanced computational methods and domain-specific expertise (14). Moreover, data quality and standardisation are persistent challenges. Inconsistent sample collection protocols, sequencing depths, and proteomic platform sensitivity contribute to variability that complicates cross-study integration. Batch effects, missing values, and false positives can undermine the reliability of downstream analyses. There remains a lack of universal data standards and formats that enable seamless cross-platform and cross-institutional collaboration. Although initiatives like the Global Alliance for Genomics and Health (GA4GH) have made progress, many datasets are still siloed due to format incompatibility or institutional restrictions (9, 40).

2. Integration and Interpretation Challenges

One of the central promises of integrative omics is the ability to contextualise genes within functional protein networks and pathways. However, bioinformatics integration pipelines remain a major bottleneck. Traditional statistical methods often fall short in handling high-dimensional omics datasets, especially when sample sizes are small relative to feature numbers. Advanced approaches such as machine learning, deep learning, and network-based modelling are being increasingly adopted. Tools like MOFA+ (Multi-Omics Factor Analysis), OmicsIntegrator, and iCluster+ are examples of integrative frameworks. Yet these tools often require sophisticated expertise in programming, statistics, and domain biology, limiting accessibility (5, 9).

Interpretation of integrated datasets also presents a significant challenge. For instance, a gene with upregulated expression might not correspond to increased protein abundance due to post-transcriptional regulation. Such discrepancies necessitate careful contextualisation, often requiring prior biological knowledge and experimental validation, a time- and resource-intensive endeavour. Furthermore, integrating single-cell and spatial omics adds layers of complexity, including spatial mapping, temporal resolution, and tissue heterogeneity. There is a dearth of standardised workflows for processing, integrating, and visualising these complex data types across modalities (40).

3. Cost and Resource Demands

Despite decreasing costs for sequencing and proteomic analysis, integrative omics remains **cost-**prohibitive for many institutions, particularly in LMICs. Costs arise not only from laboratory reagents and consumables but also from the need for:

a. High-throughput sequencing and mass spectrometry instruments.

b. Large-scale data storage infrastructure.

c. Cloud computing or high-performance computing clusters.

d. Licences for bioinformatics tools.

e. Training and capacity building for personnel.

In LMICs, many research institutions lack access to sustained funding mechanisms to support multi-omics research. Moreover, technical maintenance and vendor support are often limited or unavailable, leading to equipment downtime or underutilisation. Cost-related limitations also hinder population-level studies, which are critical for identifying rare variants or disease-specific biomarkers. For instance, the African genomics landscape remains vastly underrepresented, despite the continent's high genetic diversity and disease burden. Additionally, intellectual property issues and licensing restrictions around proprietary algorithms and databases can hinder collaboration and innovation in low-resource settings (41).

4. Ethical, Legal, and Social Implications (ELSI)

The ethical dimension of genomics and proteomics integration has gained increasing attention, particularly as datasets grow in granularity and global reach. Ethical challenges are not merely abstract; they have direct implications for research participation, data sharing, and translational impact.

a. Data Privacy and Governance: One of the core concerns is ensuring data privacy and security. Omics data are inherently identifiable; even de-identified genomic sequences can be traced back to individuals using reference databases. As such, compliance with regulations like the General Data Protection Regulation (GDPR) in Europe is critical for cross-border research. Robust frameworks for informed consent are essential, particularly for longitudinal and re-usable data studies. However, consent frameworks are often vague or do not cover secondary uses, especially in biobanking initiatives. Many LMICs lack formalised Institutional Review Board (IRB) frameworks for evaluating multi-omics studies, creating grey zones for ethical conduct and oversight (42).

b. Risk of Discrimination and Stigmatisation: There is increasing concern that omics research, if misused, could exacerbate health disparities or lead to genetic discrimination. This includes the potential misuse of genomic data by insurers, employers, or governments, particularly where legal protections are absent or underdeveloped. Research focusing on population-specific biomarkers or disease susceptibilities must navigate sensitive issues related to ethnicity, ancestry, and health inequity. Without community engagement and transparent communication, such research risks misinterpretation and mistrust, particularly in Indigenous or historically marginalised populations (43).

c. Equity in Low-Resource Settings: Global omics initiatives often leave out LMICs, exacerbating existing scientific inequities. Even when samples from LMICs are used, data ownership and access are often retained by institutions in high-income countries (HICs), a practice known as "helicopter science." To address this, funding agencies and consortia must enforce principles of data sovereignty, equitable authorship, and capacity building. Programs like H3Africa provide a model for fostering genomics research leadership within Africa. However, similar efforts are needed in proteomics and integrative omics. There is also a need to increase genetic and proteomic literacy among communities and clinicians in LMICs to enable informed decision-making and translational uptake (44).

d. Legal and Policy Gaps: There is currently no universal legal framework governing the integration, use, and sharing of multi-omics data. Policies differ across jurisdictions, complicating international collaborations. The key legal issues include: Data ownership and custodianship; Liability for erroneous interpretation of clinical omics data; Re-use of public omics datasets by third parties; and Benefit sharing with communities and nations contributing biological samples.

Without clear guidelines, research may proceed in ethically murky waters, undermining trust and long-term sustainability. Despite its revolutionary potential, proteogenomics is not without substantial barriers. Data complexity, analytical challenges, and resource limitations continue to hinder widespread adoption. Moreover, ethical, legal, and social considerations—especially in the context of LMICs- must be proactively addressed to ensure that the benefits of omics integration are equitably distributed. A multifaceted response is required: investment in infrastructure, development of open-access tools, training of local researchers, and creation of inclusive ethical frameworks. Only through such concerted efforts can integrative omics fulfil its promise of transforming biology and medicine for all populations, not just a privileged few (45).

Global and Translational Impact of Integrative Genomics and Proteomics

Proteogenomics technologies has transcended academic research and is now reshaping clinical and public health paradigms globally [4, 10]. With the convergence of these "omics" approaches, healthcare systems are beginning to address diseases with unprecedented precision, particularly in infectious disease management, non-communicable diseases (NCDs), and rare genetic disorders. The real-world impact of these advancements is especially significant in low-resource and high-disease-burden regions, where omics-informed interventions are driving innovation, accessibility, and equity (46).

1. Infectious Disease and Pandemic Preparedness

One of the most striking demonstrations of integrative omics in global health was observed during the COVID-19 pandemic. Genomic surveillance enabled rapid tracking of SARS-CoV-2 variants through whole genome sequencing (WGS), while proteomic profiling of patient samples contributed to understanding host responses and cytokine storms. This integration supported the stratification of patients by severity, guided treatment interventions, and accelerated the development of diagnostic assays and vaccines. For instance, studies combining RNA-seq and mass spectrometry in COVID-19 patients identified key biomarkers such as IL-6, ferritin, and D-dimer, enhancing early triage and outcome prediction. Proteogenomic tools also supported serological assay development and mRNA vaccine design by identifying immunogenic viral peptides. These rapid-response strategies highlight how integrative omics can be repurposed for pandemic preparedness and emerging infectious disease threats, including antimicrobial resistance (AMR) and neglected tropical diseases (45).

2. Precision Medicine in Low- and Middle-Income Countries (LMICs)

Historically, the Global South has been underrepresented in genomic datasets, limiting the applicability of precision medicine. However, initiatives such as H3Africa (Human Heredity and Health in Africa) and the Nigerian 100K Genomes Project have significantly expanded African genomic data infrastructure. These efforts are enabling a better understanding of population-specific genetic variants, pharmacogenomics, and disease susceptibility patterns, especially in cancer, sickle cell disease, and diabetes (7). The integration of proteomic data, via low-cost technologies such as label-free quantification (LFQ) and selected reaction monitoring (SRM), into these programs is facilitating protein biomarker discovery for diseases like cervical cancer and malaria. By aligning proteomic findings with genomic variation, researchers in Africa are beginning to develop tailored therapeutic and diagnostic models, increasing the relevance and uptake of precision healthcare. For example, in Ghana and South Africa, multi-omics research has helped in characterising tuberculosis (TB) pathophysiology and host immune signatures. Proteogenomic profiling in HIV-positive TB patients revealed novel biomarkers for latent TB reactivation risk, guiding more targeted prophylaxis in these settings. These integrative efforts bridge the gap between discovery science and public health intervention, especially when supported by governmental policies and international collaborations (46).

3. AI-Enhanced Diagnostics and Predictive Modelling
Artificial intelligence (AI) is further amplifying the translational utility of omics data. Machine learning (ML) algorithms trained on integrative omics datasets are increasingly being deployed for disease classification, patient stratification, and treatment response prediction. In cancer, for example, tools like DeepOmix and AutoML frameworks have demonstrated the ability to distinguish subtypes of breast cancer by combining gene expression and proteomic data. In LMIC contexts, AI-enabled omics solutions are emerging to bypass limitations in clinical infrastructure. Mobile-linked AI diagnostics powered by multi-omics data, such as smartphone-integrated TB and malaria tests, are being piloted in East Africa, offering decentralised, rapid diagnostics even in rural settings. Moreover, platforms like PathAI and BioMind are using omics-guided models to interpret pathology slides and imaging in real time, dramatically reducing diagnosis turnaround times where expert personnel are scarce. Crucially, AI tools also enable integrative analysis of diverse datasets, including social determinants of health, electronic health records (EHRs), and wearable sensor outputs, broadening the clinical applicability of omics technologies from the lab bench to bedside decision-making in a globally inclusive manner (47).

4. Expanding Accessibility of Omics Technologies in the Global South
Despite the transformative potential, cost and infrastructure have historically been barriers to the adoption of omics platforms in LMICs [5]. However, several developments are closing this accessibility gap:
Miniaturised Sequencing Devices: Technologies such as Oxford Nanopore's MinION are portable, low-cost, and have been successfully deployed for Ebola and COVID-19 sequencing in field conditions.
a. Open Access Platforms: Initiatives like the Pan-African Bioinformatics Network (H3ABioNet) provide freely available pipelines and computational resources for omics data analysis tailored to resource-limited contexts.
b. Training and Capacity Building: Programs supported by WHO, Wellcome Trust, and NIH are building local expertise through fellowships and bioinformatics hubs, reducing dependence on external collaborators.
c. Cloud-Based Data Sharing: Platforms like DNAnexus, Terra.bio, and Galaxy allow global access to omics datasets and collaborative analysis, minimising the need for local data servers or high-performance computing.
Moreover, low-cost proteomics tools such as mass spectrometry-based kits and open-source platforms (e.g., Skyline, OpenMS) are enabling laboratories in LMICs to participate in biomarker discovery and drug target validation efforts without needing high-end commercial systems (48).

5. Ethical, Legal, and Policy Implications
To ensure equitable global benefit, integrative omics in translational research must address ethical and legal frameworks [4]. Challenges such as informed consent, data ownership, community engagement, and genetic literacy are magnified in resource-limited and ethnically diverse settings. Initiatives like the Global Alliance for Genomics and Health (GA4GH) and EGA Africa are promoting culturally sensitive consent models and equitable data sharing practices. In LMICs, community-based participatory research and ethical review boards are playing an increasing role in guiding omics implementation. For example, genetic data from indigenous populations is now being co-managed with community input in several African genomics projects, fostering trust and ensuring social validity of biomedical research (49).
The global and translational impact of integrated genomics and proteomics is accelerating scientific discovery, clinical translation, and public health innovation across the world. From pandemic preparedness and AI-enabled diagnostics to LMIC-led biomarker discovery and indigenous data stewardship, integrative omics is reshaping biomedical research with a clear commitment to inclusion and equity. As accessibility grows and cross-disciplinary collaborations expand, the promise of omics technologies will become increasingly realised at the global scale (50).

Knowledge Gaps and Associated Impact
Proteogenomics in biological research has led to significant advancements in disease characterisation, drug discovery, and personalised medicine (2, 51). However, the full potential of these technologies remains constrained by persistent knowledge gaps, especially in the context of scalability, standardisation, interpretation, and equitable access, particularly in low- and middle-income countries (LMICs). Addressing these gaps is critical to enhancing the translational utility and global relevance of multi-omics research.

1. Limited Integration Across Omics Layers
Despite growing interest in multi-omics integration, most research remains focused on isolated omics layers (e.g., genomics, transcriptomics, or proteomics independently). The lack of standardised pipelines for integrating diverse datasets (e.g., epigenomics, metabolomics, microbiomics) impedes comprehensive systems biology insights.
Impact: This siloed approach leads to partial interpretations of biological phenomena, limiting diagnostic precision and therapeutic innovation. Complex diseases such as cancer,

diabetes, and neurodegeneration cannot be fully understood without integrative models (52).

2. Data Volume, Complexity, and Interpretation
High-throughput technologies generate vast, heterogeneous datasets that are difficult to harmonise. Variations in experimental design, sample preparation, and platform sensitivity add another layer of complexity.
Impact: Researchers often struggle to derive biologically meaningful conclusions, slowing down biomarker validation and translation to clinical practice. Poor interoperability also hinders reproducibility and meta-analysis across studies (53).

3. Underrepresentation of LMIC Populations
Current omics databases are heavily skewed toward European and North American populations. Genetic, proteomic, and environmental diversity in LMICs is largely uncharted, leading to biases in disease association studies and therapeutic development.
Impact: This gap exacerbates global health inequities and limits the applicability of precision medicine to diverse populations. It also undermines the identification of population-specific biomarkers and drug responses (54).

4. Ethical, Legal, and Social Implications (ELSI)
Multi-omics research presents complex ethical challenges, particularly around data privacy, informed consent, and the ownership of genetic data. Regulatory frameworks such as the General Data Protection Regulation (GDPR) are often lacking or poorly enforced in LMICs (10, 14).
Impact: Concerns over data misuse and discrimination may discourage participation in research, especially in communities with histories of exploitation or limited genomic literacy.

5. Technical Standardisation and Scalability
There is no consensus on the best practices for proteogenomics workflows—whether in sample processing, MS/MS analysis, or downstream computational modelling. Tools like MaxQuant, Galaxy-P, and MOFA+ are powerful but often inaccessible due to technical or financial constraints.
Impact: Inconsistent workflows affect data quality, limit reproducibility, and delay cross-study validation, especially in resource-limited settings (55).

Future Research Directions

The future of integrative omics is poised to transform biomedical research and clinical practice through the convergence of cutting-edge technologies, inclusive data-sharing frameworks, and equitable global collaborations. While genomics and proteomics have independently contributed significantly to disease understanding, the next frontier lies in integrating these, and other "omics" layers, at high resolution, across diverse populations, and in real-time clinical settings (56).

1. Emerging Technologies: Spatial and Single-Cell Multi-Omics
Spatial omics and single-cell multi-omics are two of the most groundbreaking technological advances that will redefine systems biology. Traditional bulk sequencing approaches mask cell-to-cell variability and spatial context within tissues. In contrast, single-cell RNA sequencing (scRNA-seq), single-cell proteomics (e.g., CyTOF), and ATAC-seq provide a granular view of gene and protein activity within individual cells. Moreover, spatial transcriptomics and proteomics technologies such as 10x Genomics Visium, NanoString GeoMx, and imaging mass cytometry offer spatial resolution of gene and protein expression within tissue architecture. These tools are particularly useful in oncology, neurology, and immunology, where understanding microenvironmental interactions (e.g., tumour-immune dynamics) is crucial for therapy development. Integrating spatial and single-cell multi-omics allows for unprecedented insights into cellular heterogeneity, developmental biology, and drug resistance mechanisms. As these platforms become more scalable and cost-efficient, their integration into routine diagnostics and translational research, especially in resource-constrained settings, will become increasingly feasible (4, 57).

2. Integrative Pipelines: AI, Federated Learning, and Cloud Computing
High-throughput omics technologies generate massive, multidimensional datasets that demand advanced analytical infrastructure. Future pipelines will increasingly rely on artificial intelligence (AI) and machine learning (ML) to decipher complex biological patterns. Deep learning models are already being used for protein structure prediction (e.g., AlphaFold2), regulatory network inference, and disease classification based on multi-omics signatures. A particularly promising avenue is federated learning, where decentralised ML models can be trained on distributed datasets without transferring sensitive patient data. This approach protects data privacy while enabling collaborative analysis across institutions and geographies. Federated omics analysis holds transformative potential for rare disease research, where data aggregation across multiple centres is essential but often hampered by legal or ethical constraints. Additionally, cloud computing platforms such as Terra (Broad Institute), DNAnexus, and AWS Omics provide scalable, secure environments for data storage and computation. These platforms host standardised pipelines for multi-omics integration, variant calling, proteomic analysis, and data visualisation, making sophisticated analytics accessible even to teams lacking in-house infrastructure (9, 58).

3. Open Science and Data-Sharing Initiatives

To realise the full power of integrative omics, open-access and FAIR (Findable, Accessible, Interoperable, Reusable) data practices are indispensable. Several global initiatives are setting the standard for collaborative science:

a. HuBMAP (Human BioMolecular Atlas Program) aims to map cellular and molecular architecture across the human body using integrated omics and spatial profiling.

b. GTEx (Genotype-Tissue Expression Project) provides high-quality transcriptomic and proteomic data across various human tissues, supporting cross-omics correlation studies.

c. H3Africa (Human Heredity and Health in Africa) promotes African leadership in genomics and bioinformatics, enabling population-specific insights relevant to the African continent.

These initiatives foster reproducibility, equity, and innovation by making high-quality datasets and tools openly available to researchers worldwide. Future directions should include greater representation from LMICs in data generation, analysis, and authorship, ensuring that global benefits reflect global participation (14, 57).

4. Capacity Building and Infrastructure Development in LMICs

While omics technologies are rapidly advancing, their impact remains unevenly distributed. For integrative omics to achieve global relevance, strategic investments must be made in LMICs to enhance infrastructure, human capacity, and institutional resilience.

Key strategies include:

a. Bioinformatics training programs in universities and research institutions to cultivate a new generation of omics scientists.

b. South-South and North-South partnerships for mentorship, joint research, and shared infrastructure (e.g., regional sequencing centres).

c. Policy frameworks and funding mechanisms that prioritise context-relevant research in infectious diseases, nutrition, and genetic disorders prevalent in LMICs.

Several promising examples already exist: the African Centre of Excellence for Genomics of Infectious Diseases (ACEGID) in Nigeria has been instrumental in SARS-CoV-2 variant surveillance. Similarly, H3ABioNet provides training and pipeline development for over 30 African institutions. Future success will depend on democratizing access to technologies like mass spectrometry, next-generation sequencing (NGS), and cloud computing, paired with local ownership of data and research priorities (13, 58).

5. Ethical, Legal, and Social Implications (ELSI): Building a Responsible Future

As integrative omics becomes more embedded in healthcare and research, ethical governance frameworks must evolve in parallel. Emerging concerns around data privacy, ownership, and equity demand proactive measures to ensure responsible implementation (11, 23).

Recommended directions include:

a. Community-based participatory research models that involve local populations in consent design, priority setting, and data governance.

b. Stronger enforcement of data protection laws, such as GDPR in Europe and evolving local laws in Africa, Asia, and Latin America.

c. Equity-focused metrics to monitor diversity in genomic and proteomic datasets, publication authorship, and clinical trial participation.

d. Cross-border legal harmonisation to support federated data sharing while respecting cultural and regulatory differences.

The global research community must move toward reciprocal benefit models, where populations contributing data also share in the resulting scientific, therapeutic, and economic gains. The next decade of integrative omics will be defined by convergence: of disciplines (e.g., biology, computer science, ethics), technologies (e.g., spatial profiling, AI), and stakeholders (e.g., patients, scientists, policymakers). By aligning innovation with inclusion, openness with responsibility, and data with human impact, we can create a future where integrative genomics and proteomics catalyse not just scientific progress but global health equity. Embracing these future directions will not only enhance our molecular understanding of life but also ensure that the benefits are broadly and ethically distributed (59, 60).

Conclusion

Integrating genomics and proteomics represents a transformative approach in modern biological and clinical research, enabling a more comprehensive understanding of cellular function, disease mechanisms, and therapeutic opportunities. While genomics provides insights into the genetic blueprint, proteomics offers dynamic context through the analysis of protein expression, modifications, and interactions. The synergy between these fields enhances our ability to discover robust biomarkers, identify drug targets, and advance precision medicine strategies. Real-world applications, such as those seen in cancer proteogenomics and neurodegenerative disease research, underscore the translational value of this integrative approach. Despite these advances, challenges persist—including data complexity, integration barriers, cost constraints, and ethical concerns, especially in low- and middle-income countries (LMICs). Addressing these issues through standardisation, equitable access, and global collaboration will be crucial for realising the full potential of integrative omics. As technologies such as single-cell and spatial omics, AI, and cloud computing mature, the future holds promise for scalable, reproducible, and ethically grounded multi-omics research. This narrative review underscores the urgent need for inclusive, interdisciplinary efforts to overcome current limitations and drive innovation across biomedical domains. Ultimately, integrative omics is not just a scientific imperative but a global health opportunity.

Research Highlights
1. Integrating genomics and proteomics enables deeper insights into cellular processes, disease mechanisms, and biomarker discovery for improved diagnostics and therapeutics.
2. Recent advances in next-generation sequencing and high-throughput proteomics have enhanced multi-omics data accuracy, resolution, and clinical applicability.
3. Despite progress, challenges such as data integration complexity, cost, and limited infrastructure in low-resource settings hinder widespread adoption.

Study limitations
While this narrative review provides a comprehensive and timely synthesis of recent advances in proteogenomics technologies, several limitations must be acknowledged. First, the literature included in this review was restricted to publications written in English. This language limitation may have excluded valuable contributions published in other languages, particularly from non-English-speaking regions that are increasingly active in omics research. Second, the time frame for article inclusion was intentionally set between 2015 and 2025 to ensure the review focused on the most recent developments and technological advances. However, this may have excluded earlier foundational studies that could offer historical context or longitudinal insights. Third, the methodology employed was a narrative review, rather than a systematic review or meta-analysis. While this approach allows for thematic exploration and conceptual synthesis, it does not provide statistical rigour or quantify outcomes across studies. Future work may benefit from systematic meta-analytical techniques to assess the efficacy of outcomes of integrative omics strategies.

Abbreviations

AI: Artificial Intelligence
CPTAC: Clinical Proteomic Tumour Analysis Consortium
CSF: Cerebrospinal Fluid
DDA: Data-Dependent Acquisition
DIA: Data-Independent Acquisition
DNA: Deoxyribonucleic Acid
DNN: Deep Neural Network
EGFR: Epidermal Growth Factor Receptor
ENCODE: Findable, Accessible, Interoperable, and Reusable (principles for data management)
GA4GH: Global Alliance for Genomics and Health
GATK: Genome Analysis Toolkit
GTEx: Genotype-Tissue Expression project
GWAS: Genome-Wide Association Studies
HIPAA: Health Insurance Portability and Accountability Act
HISAT2: High-Performance Computing
H3Africa: Human Heredity and Health in Africa
IRB: Institutional Review Board
KEGG: Kyoto Encyclopedia of Genes and Genomes
LMICs: Low- and Middle-Income Countries
MAPK: Mitogen-Activated Protein Kinase
MIAME: Minimum Information About a Microarray Experiment
mRNA: Messenger RNA
MS/MS: Tandem Mass Spectrometry
NGS: Next-Generation Sequencing
NSCLC: Non-Small Cell Lung Cancer
ORF: Open Reading Frame
PCA: Principal Component Analysis
PI3K: Phosphoinositide 3-Kinase
PRIDE: Proteomics Identifications Database
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PTM: Post-Translational Modification
RNA-seq: RNA Sequencing
RSEM: RNA-Seq by Expectation-Maximisation
SNP: Single Nucleotide Polymorphism
STAR: Spliced Transcripts Alignment to a Reference
TCGA: The Cancer Genome Atlas
UMAP: Uniform Manifold Approximation and Projection
WGS: Whole Genome Sequencing

Declarations

Ethics Approval and Consent to Participate: Not applicable
Data Availability: All data used for this review are publicly available
Competing Interests: The authors declared no competing interests
Funding sources: This review did not receive any funding from an external source
Author contributions: Conceptualisation: ESS and AEN; Visualisation: SSE; Writing – original draft: ESS, AEN and OEOO; Writing – review & editing: ESS, ES, ARY, EOM and ECB
Acknowledgements: None

References

1. Kang MG, Byun K, Kim JH, Park NH, Heinsen H, Ravid R, Steinbusch HW, Lee B, Park YM. Proteogenomics of the human hippocampus: The road ahead Biochim Biophys Acta. 2015 Jul;1854(7):788-97. https://doi.org/10.1016/j.bbapap.2015.02.010
2. Shukla HD, Mahmood J, Vujaskovic Z. Integrated proteo-genomic approach for early diagnosis and prognosis of cancer Cancer Lett. 2015 Dec 1;369(1):28-36. https://doi.org/10.1016/j.canlet.2015.08.003
3. Gangfuß A, Schara-Schmidt U, Roos A. [Genomics and proteomics in the research of neuromuscular diseases]

Nervenarzt. 2022 Feb;93(2):114-121. https://doi.org/10.1007/s00115-021-01201-1
4. Fu S, Liu X, Luo M, Xie K, Nice EC, Zhang H, Huang C. Proteogenomic studies on cancer drug resistance: towards biomarker discovery and target identification Expert Rev Proteomics. 2017 Apr;14(4):351-362. https://doi.org/10.1080/14789450.2017.1299006.
5. Rodriguez H, Zenklusen JC, Staudt LM, Doroshow JH, Lowy DR. The next horizon in precision oncology: Proteogenomics to inform cancer diagnosis and treatment Cell. 2021 Apr 1;184(7):1661-1670. https://doi.org/10.1016/j.cell.2021.02.055
6. Bischoff R, Permentier H, Guryev V, Horvatovich P. Genomic variability and protein species - Improving sequence coverage for proteogenomics J Proteomics. 2016 Feb 16;134:25-36. https://doi.org/10.1016/j.jprot.2015.09.021
7. Heo YJ, Hwa C, Lee GH, Park JM, An JY. Integrative Multi-Omics Approaches in Cancer Research: From Biological Networks to Clinical Subtypes Mol Cells. 2021 Jul 31;44(7):433-443. https://doi.org/10.14348/molcells.2021.0042
8. Menschaert G, Fenyö D. Proteogenomics from a bioinformatics angle: A growing field Mass Spectrom Rev. 2017 Sep;36(5):584-599. https://doi.org/10.1002/mas.21483
9. Husain H, Ahmad R, Khan A, Inamuddin, Asiri AM. Proteomic-genomic adjustments and their confluence for elucidation of pathways and networks during liver fibrosis Int J Biol Macromol. 2018 May;111:379-392. https://doi.org/10.1016/j.ijbiomac.2017.12.168
10. Dimitrakopoulos L, Prassas I, Diamandis EP, Charames GS. Onco-proteogenomics: Multi-omics level data integration for accurate phenotype prediction Crit Rev Clin Lab Sci. 2017 Sep;54(6):414-432. https://doi.org/10.1080/10408363.2017.1384446.
11. Kumar D, Dash D. Proteogenomic Tools and Approaches to Explore Protein Coding Landscapes of Eukaryotic Genomes Adv Exp Med Biol. 2016;926:1-10. https://doi.org/10.1007/978-3-319-42316-6_1
12. Eicher T, Patt A, Kautto E, Machiraju R, Mathé E, Zhang Y. Challenges in proteogenomics: a comparison of analysis methods with the case study of the DREAM proteogenomics sub-challenge BMC Bioinformatics. 2019 Dec 20;20(Suppl 24):669. https://doi.org/10.1186/s12859-019-3253-z
13. Staal JA, Pei Y, Rood BR. A Proteogenomic Approach to Understanding MYC Function in Metastatic Medulloblastoma Tumors Int J Mol Sci. 2016 Oct 19;17(10):1744. https://doi.org/10.3390/ijms17101744
14. Vaudel M, Barsnes H, Ræder H, Berven FS. Using Proteomics Bioinformatics Tools and Resources in Proteogenomic Studies Adv Exp Med Biol. 2016;926:65-75. https://doi.org/10.1007/978-3-319-42316-6_5
15. Kumar D, Bansal G, Narang A, Basak T, Abbas T, Dash D. Integrating transcriptome and proteome profiling: Strategies and applications Proteomics. 2016 Oct;16(19):2533-2544. https://doi.org/10.1002/pmic.201600140
16. Barbieri R, Guryev V, Brandsma CA, Suits F, Bischoff R, Horvatovich P. Proteogenomics: Key Driver for Clinical Discovery and Personalized Medicine Adv Exp Med Biol. 2016;926:21-47. https://doi.org/10.1007/978-3-319-42316-6_3
17. Ansari D, Torén W, Zhou Q, Hu D, Andersson R. Proteomic and genomic profiling of pancreatic cancer Cell Biol Toxicol. 2019 Aug;35(4):333-343. https://doi.org/10.1007/s10565-019-09465-9
18. Verma A, Halder A, Marathe S, Purwar R, Srivastava S. A proteogenomic approach to target neoantigens in solid tumors Expert Rev Proteomics. 2020 Nov-Dec;17(11-12):797-812. https://doi.org/10.1080/14789450.2020.1881889
19. Das T, Andrieux G, Ahmed M, Chakraborty S. Integration of Online Omics-Data Resources for Cancer Research Front Genet. 2020 Oct 23;11:578345. https://doi.org/10.3389/fgene.2020.578345
20. Chiou SH, Lee KT. Proteomic analysis and translational perspective of hepatocellular carcinoma: Identification of diagnostic protein biomarkers by an onco-proteogenomics approach Kaohsiung J Med Sci. 2016 Nov;32(11):535-544. https://doi.org/10.1016/j.kjms.2016.09.002
21. Fernández-Lázaro D, Garrosa E, Seco-Calvo J, Garrosa M. Potential Satellite Cell-Linked Biomarkers in Aging Skeletal Muscle Tissue: Proteomics and Proteogenomics to Monitor Sarcopenia Proteomes. 2022 Aug 19;10(3):29. https://doi.org/10.3390/proteomes10030029
22. Verstappe B, Scott CL. Implementing distinct spatial proteogenomic technologies: opportunities, challenges, and key considerations Clin Exp Immunol. 2024 Oct 16;218(2):151-162. https://doi.org/10.1093/cei/uxae077
23. Shi L, Ng JK, Xiong Q, Ao KF, Shin SK, Law CT, Mu W, Liu GM, Rao S, Tsui SK. Comparative genomic analysis of immune-related genes and chemosensory receptors provides insights into the evolution and adaptation of four major domesticated Asian carps. BMC Genomics. 2025 May 26;26(1):529. https://doi.org/10.1186/s12864-025-11719-2
24. Lin MR, Tsai CL, Liao CS, Wei CY, Chou WH, Hsiao TH, Chang WC. Exploring the genomic and transcriptomic profiles of glycemic traits and drug repurposing J Biomed Sci. 2025 May 21;32(1):50. https://doi.org/10.1186/s12929-025-01137-7
25. Lee Y, Lee M, Shin Y, Kim K, Kim T. Spatial Omics in Clinical Research: A Comprehensive Review of Technologies and Guidelines for Applications Int J Mol Sci. 2025 Apr 22;26(9):3949. https://doi.org/10.3390/ijms26093949

26. Sun H, Chen S, Kong J. Cerebrospinal Fluid Metabolomics and Proteomics Integration in Neurological Syndromes Methods Mol Biol. 2025;2914:303-321. https://doi.org/10.1007/978-1-0716-4462-1_21
27. Wang Z, Zhang T, Tang M. Navigating nanotoxicity: Unraveling nanomaterial-induced effects via multi-omics integration NanoImpact. 2025 May 16;38:100565. https://doi.org/10.1016/j.impact.2025.100565
28. Morabito A, De Simone G, Pastorelli R, Brunelli L, Ferrario M. Algorithms and tools for data-driven omics integration to achieve multilayer biological insights: a narrative review J Transl Med. 2025 Apr 10;23(1):425. https://doi.org/10.1186/s12967-025-06446-x
29. Kahraman Ilıkkan Ö. Lactobacillomics as a new notion in lactic acid bacteria research through omics integration World J Microbiol Biotechnol. 2025 Feb 11;41(2):68. https://doi.org/10.1007/s11274-025-04285-y
30. Zhang F, Zhu M, Chen Y, Wang G, Yang H, Lu X, Li Y, Chang HM, Wu Y, Ma Y, Yuan S, Zhu W, Dong X, Zhao Y, Yu Y, Wang J, Mu L. Harnessing omics data for drug discovery and development in ovarian aging Hum Reprod Update. 2025 May 1;31(3):240-268. https://doi.org/10.1093/humupd/dmaf002
31. Gong X, Su L, Huang J, Liu J, Wang Q, Luo X, Yang G, Chi H. An overview of multi-omics technologies in rheumatoid arthritis: applications in biomarker and pathway discovery Front Immunol. 2024 Jul 30;15:1381272. https://doi.org/10.3389/fimmu.2024.1381272
32. Luo Y, Zhao C, Chen F. Multiomics Research: Principles and Challenges in Integrated Analysis Biodes Res. 2024 Dec 5;6:0059. https://doi.org/10.34133/bdr.0059.
33. Lin M, Xu F, Sun J, Song J, Shen Y, Lu S, Ding H, Lan L, Chen C, Ma W, Wu X, Song Z, Wang W. Integrative multi-omics analysis unravels the host response landscape and reveals a serum protein panel for early prognosis prediction for ARDS Crit Care. 2024 Jul 2;28(1):213. https://doi.org/10.1186/s13054-024-05000-3
34. Zhang H, Zhou Y, Yu B, Deng Y, Wang Y, Fang S, Song X, Fan X, Zhou H. Multi-Omics Approaches to Discover Biomarkers of Thyroid Eye Disease: A Systematic Review Int J Biol Sci. 2024 Nov 11;20(15):6038-6055. https://doi.org/10.7150/ijbs.103977
35. George NP, Kwon M, Jang YE, Kim SG, Hwang JS, Lee SS, Lee G. Integrative Analysis of Metabolome and Proteome in the Cerebrospinal Fluid of Patients with Multiple System Atrophy Cells. 2025 Feb 12;14(4):265. https://doi.org/10.3390/cells14040265
36. Cominetti O, Dayon L. Unravelling disease complexity: integrative analysis of multi-omic data in clinical research Expert Rev Proteomics. 2025 Apr;22(4):149-162. https://doi.org/10.1080/14789450.2025.2491357
37. Zhang W, Huang H, Wang L, Lehmann BD, Chen SX. An Integrative Multi-Omics Random Forest Framework for Robust Biomarker Discovery bioRxiv [Preprint]. 2025 Mar 6:2025.03.05.641533. https://doi.org/10.1101/2025.03.05.641533
38. Ambikan A, Akusjärvi SS, Sperk M, Neogi U. System-level integrative omics analysis to identify the virus-host immunometabolic footprint during infection Adv Immunol. 2024;164:73-100. https://doi.org/10.1016/bs.ai.2024.08.002
39. Sathyanarayanan A, Mueller TT, Ali Moni M, Schueler K; ECNP TWG Network members; Baune BT, Lio P, Mehta D, Baune BT, Dierssen M, Ebert B, Fabbri C, Fusar-Poli P, Gennarelli M, Harmer C, Howes OD, Janzing JGE, Lio P, Maron E, Mehta D, Minelli A, Nonell L, Pisanu C, Potier MC, Rybakowski F, Serretti A, Squassina A, Stacey D, van Westrhenen R, Xicota L. Multi-omics data integration methods and their applications in psychiatric disorders Eur Neuropsychopharmacol. 2023 Apr;69:26-46. https://doi.org/10.1016/j.euroneuro.2023.01.001
40. Fang S, Chen B, Zhang Y, Sun H, Liu L, Liu S, Li Y, Xu X. Computational Approaches and Challenges in Spatial Transcriptomics Genomics Proteomics Bioinformatics. 2023 Feb;21(1):24-47. https://doi.org/10.1016/j.gpb.2022.10.001
41. Pokhriyall M, Shukla N, Singh TR, Suravajhala P. Proteogenomic Approaches for Diseasome Studies Methods Mol Biol. 2025;2859:253-264. https://doi.org/10.1007/978-1-0716-4152-1_14
42. Tsang O, Wong JWH. Proteogenomic interrogation of cancer cell lines: an overview of the field. Expert Rev Proteomics. 2021 Mar;18(3):221-232. https://doi.org/10.1080/14789450.2021.1914594
43. Binetti M, Lauro A, Vaccari S, Cervellera M, Tonini V. Proteogenomic biomarkers in colorectal cancers: clinical applications Expert Rev Proteomics. 2020 May;17(5):355-363. https://doi.org/10.1080/14789450.2020.1782202
44. Philips TJ, Erickson BK, Thomas SN. Opportunities for predictive proteogenomic biomarkers of drug treatment sensitivity in epithelial ovarian cancer. Front. Oncol. 2025 Jan 7;14:1503107. https://doi.org/10.3389/fonc.2024.1503107
45. Ruggles KV, Krug K, Wang X, Clauser KR, Wang J, Payne SH, Fenyö D, Zhang B, Mani DR. Methods, Tools and Current Perspectives in Proteogenomics Mol Cell Proteomics. 2017 Jun;16(6):959-981. https://doi.org/10.1074/mcp.MR117.000024
46. Song YC, Das D, Zhang Y, Chen MX, Fernie AR, Zhu FY, Han J. Proteogenomics-based functional genome research: approaches, applications, and perspectives in plants. Trends Biotechnol. 2023 Dec;41(12):1532-1548. https://doi.org/10.1016/j.tibtech.2023.05.010
47. Nishimura T, Nakamura H, Végvári Á, Marko-Varga G, Furuya N, Saji H. Current status of clinical proteogenomics in lung cancer. Expert Rev Proteomics.

2019 Sep;16(9):761-772. https://doi.org/10.1080/14789450.2019.1654861

48. Rao R, Gulfishan M, Kim MS, Kashyap MK. Deciphering Cancer Complexity: Integrative Proteogenomics and Proteomics Approaches for Biomarker Discovery Methods Mol Biol. 2025;2859:211-237. https://doi.org/10.1007/978-1-0716-4152-1_12
49. Ruggles KV, Fenyö D. Next Generation Sequencing Data and Proteogenomics Adv Exp Med Biol. 2016;926:11-19. https://doi.org/10.1007/978-3-319-42316-6_2
50. Roychowdhury R, Das SP, Gupta A, Parihar P, Chandrasekhar K, Sarker U, Kumar A, Ramrao DP, Sudhakar C. Multi-Omics Pipeline and Omics-Integration Approach to Decipher Plant's Abiotic Stress Tolerance Responses Genes (Basel). 2023 Jun 16;14(6):1281. https://doi.org/10.3390/genes14061281
51. Díez P, Fuentes M. Proteogenomics for the Comprehensive Analysis of Human Cellular and Serum Antibody Repertoires Adv Exp Med Biol. 2016;926:153-162. https://doi.org/10.1007/978-3-319-42316-6_10
52. Sheynkman GM, Shortreed MR, Cesnik AJ, Smith LM. Proteogenomics: Integrating Next-Generation Sequencing and Mass Spectrometry to Characterize Human Proteomic Variation Annu Rev Anal Chem (Palo Alto Calif). 2016 Jun 12;9(1):521-45. https://doi.org/10.1146/annurev-anchem-071015-041722
53. Ang MY, Low TY, Lee PY, Wan Mohamad Nazarie WF, Guryev V, Jamal R. Proteogenomics: From next-generation sequencing (NGS) and mass spectrometry-based proteomics to precision medicine Clin Chim Acta. 2019 Nov;498:38-46. https://doi.org/10.1016/j.cca.2019.08.010
54. Choi S, An JY. Multiomics in cancer biomarker discovery and cancer subtyping. Adv Clin Chem. 2025;124:161-195. https://doi.org/10.1016/bs.acc.2024.10.004
55. Roos A, Thompson R, Horvath R, Lochmüller H, Sickmann A. Intersection of Proteomics and Genomics to "Solve the Unsolved" in Rare Disorders such as Neurodegenerative and Neuromuscular Diseases Proteomics Clin Appl. 2018 Mar;12(2). https://doi.org/10.1002/prca.201700073
56. Schiebenhoefer H, Van Den Bossche T, Fuchs S, Renard BY, Muth T, Martens L. Challenges and promise at the interface of metaproteomics and genomics: an overview of recent progress in metaproteogenomic data analysis. Expert Rev Proteomics. 2019 May;16(5):375-390. https://doi.org/10.1080/14789450.2019.1609944
57. Hernandez-Valladares M, Vaudel M, Selheim F, Berven F, Bruserud Ø. Proteogenomics approaches for studying cancer biology and their potential in the identification of acute myeloid leukemia biomarkers. Expert Rev Proteomics. 2017 Aug;14(8):649-663. https://doi.org/10.1080/14789450.2017.1352474
58. Fels U, Gevaert K, Van Damme P. Proteogenomics in Aid of Host-Pathogen Interaction Studies: A Bacterial Perspective Proteomes. 2017 Oct 11;5(4):26. https://doi.org/10.3390/proteomes5040026
59. Subbannayya Y, Pinto SM, Gowda H, Prasad TS. Proteogenomics for understanding oncology: recent advances and future prospects. Expert Rev Proteomics. 2016;13(3):297-308. https://doi.org/10.1586/14789450.2016.1136217
60. Paraskar G, Bhattacharya S, Kuttiappan A. The Role of Proteomics and Genomics in the Development of Colorectal Cancer Diagnostic Tools and Potential New Treatments ACS Pharmacol Transl Sci. 2025 Apr 10;8(5):1227-1250. https://doi.org/10.1021/acsptsci.4c00686