

Decoding disease and therapy through multiomics integration and systems analysis

Mano Joseph Mathew^{1,2,*}, Joyal Mathew^{3,†}, Ripsy Merrin Chacko^{1,4,5,6}, Jagadeesh Bayry^{7,8}, Jean-François Zagury^{1,*}

¹Laboratoire Génomique, Bioinformatique et Chimie Moléculaire, EA7528, Conservatoire National des Arts et Métiers, HESAM Université, 2 Rue Conté, 75003 Paris, Ile de France, France

²EFREI Research Lab, Université Paris-Panthéon-Assas, École d'Ingénieurs Généraliste du Numérique - EFREI Paris, 30–32 Avenue de la République, 94800 Villejuif, Ile de France, France

³Anhalt University of Applied Sciences, Bernburger Str. 55, 06366 Köthen (Anhalt), Germany

⁴Integrative Medicine Center, Rafael Institute, 3 Bd Bineau, 92300 Levallois-Perret, Ile de France, France

⁵Hartmann Oncology Radiotherapy Group, Hartmann Radiotherapy Institute, 4 Rue Kléber, 92300 Levallois-Perret, Ile de France, France

⁶Department of Integrative Medicine, Conservatoire National des Arts et Metiers, 292 Rue Saint-Martin, 75003 Paris, Ile de France, France

⁷Institut National de la Santé et de la Recherche Médicale, Centre de Recherche des Cordeliers, Sorbonne Université, Université Paris Cité, 15 Rue de l'École de Médecine, 75006 Paris, Ile de France, France

⁸Department of Biological Sciences and Engineering, Indian Institute of Technology Palakkad, Kanjikode, Palakkad 678623 Kerala, India

*Corresponding author. Mano Joseph Mathew, EFREI Research Lab, Université Paris-Panthéon-Assas, École d'Ingénieurs Généraliste du Numérique - EFREI Paris, 30–32 Avenue de la République, 94800 Villejuif, Ile de France, France. E-mail: mano.mathew@efrei.fr; Jean-François Zagury, Laboratoire Génomique, Bioinformatique et Chimie Moléculaire, EA7528, Conservatoire National des Arts et Métiers, HESAM Université, 2 Rue Conté, 75003 Paris, Ile de France, France. E-mail: zagury@cnam.fr

[†]Mano Joseph Mathew and Joyal Mathew contributed equally to this work.

Abstract

Computational multiomics methods are based on machine learning methods, and are primarily used for classifying patients into subtypes, discovering novel biomarkers, drug repurposing, and advancing precision medicine. Advances in high-throughput technologies have enabled comprehensive profiling of multiple molecular layers, resulting in the emergence of multiomics approaches for a more accurate understanding of disease mechanisms, therapeutic targets, and biological heterogeneity. This review examines current applications of multiomics in oncology, ageing, and immune-mediated diseases, highlighting the strengths and challenges of integrative models in understanding disease mechanisms, identifying biomarkers, and guiding precision therapies. Integration strategies, from early to late fusion and horizontal to vertical frameworks, are also examined alongside recent advances in computational platforms and preprocessing techniques.

Keywords multiomics, data integration, biomarker discovery, fusion strategies

Introduction

Recent advances in different high-throughput technologies have transformed medical research. The vast amount of data generated by high-throughput technologies such as next-generation sequencing, mass spectrometry, and nuclear magnetic resonance spectroscopy has enabled researchers to better understand the phenomena behind various biological processes and functions [1]. Understanding complex diseases requires more than examining a single layer of biological information. This is because complex diseases such as cancer arise from complex layers of intracellular and intercellular interactions, including transcriptional regulation, gene co-expression, signalling and metabolic pathways, and protein–protein interactions [2, 3]. Apart from these, a tumour is also involved in interactions with noncancerous components such as the tumour microenvironment (TME), the immune system, and gene–environment interactions

[4, 5]. The TME plays an important role in regulating and promoting tumour survival by releasing molecules and activating pathways that affect the migratory potential and invasiveness of tumour cells [3]. All these different components have major effect on the tumour growth and possess strong clinical importance for therapy and prognosis. Analysing only one of these levels typically leads to suboptimal or misleading conclusions about disease mechanisms. It is therefore important to investigate these multiple layers of interacting components involved to understand the biological phenomenon.

Multiomics models uses one or more high-throughput techniques and combines data from multiple omics layers such as genomics, transcriptomics, epigenomics, proteomics, and metabolomics [6]. As shown in Fig. 1A and B, multiple omics layers are combined to obtain relevant insights from multiple biological layers. This provides a deeper understanding of disease biology by combining

Received: February 28, 2026. **Revised:** April 22, 2026. **Accepted:** May 14, 2026

© The Author(s) 2026. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

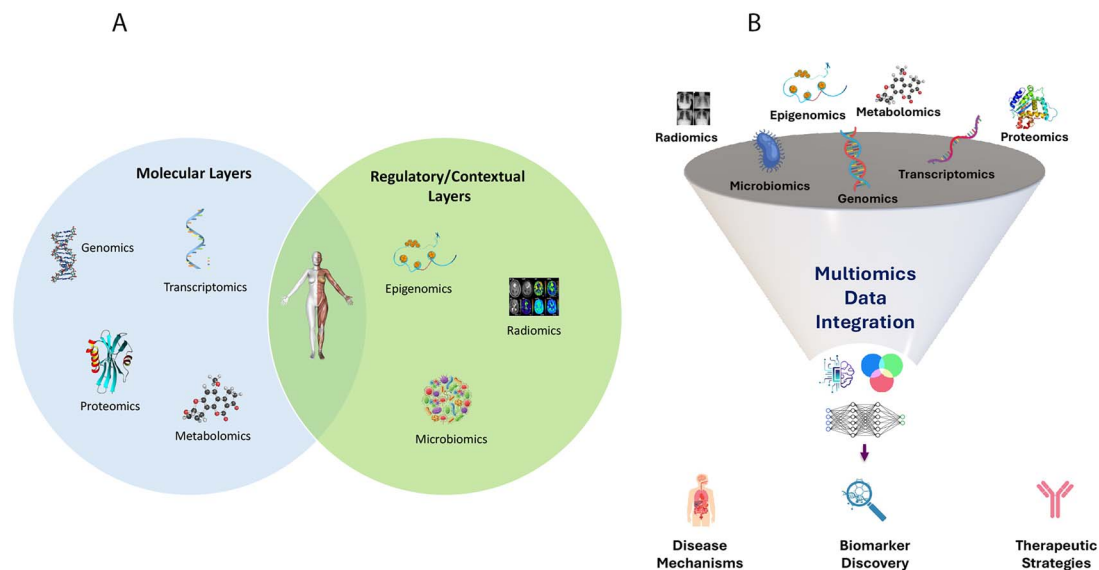


Figure 1 (a) Schematic overview of multiomics data integration. Multiple omics layers including genomics, epigenomics, transcriptomics, proteomics, metabolomics, and microbiomics are combined through integrative approaches. This unified analysis enables the extraction of clinically relevant insights, including the elucidation of disease mechanisms, the discovery of robust biomarkers, and the identification of therapeutic strategies. (b) An overview of multiomics layers. Conceptual organization of multiomics layers. Core molecular layers (genome, transcriptome, proteome, metabolome) and regulatory/contextual layers (epigenome, microbiome, radiome) together represent different dimensions of biological information contributing to phenotype.

information from RNA, DNA, proteins, and metabolites. Multiomics enables researchers to follow the molecular processes from gene alteration to phenotypic effects that are often overlooked in single omics research [7]. This systems-level strategy is particularly valuable in complex diseases like cancer, where biological variability and treatment responses can be highly different across patients [8].

While single omics studies have often provided important insights into disease mechanisms, they often fail to present a complete picture of the complex biological processes. For instance, genomic analyses provide information on DNA-level changes and copy number variations but does not capture downstream effects gene expression, protein function, or cellular phenotype [9, 10]. Similarly, transcriptomic data may show variations in gene activity but does not consider the influence of post-transcriptional regulation, protein degradation, or changes in enzymatic activity [11, 12]. Proteomic and metabolic profiles give information on the functional output of the cell, but in the absence of contextual information from upstream regulatory processes, the interpretations become incomplete [13]. Therefore, it is not possible to understand the influence of genetic mutations on cellular behaviour without considering data from multiple molecular levels. Hence, the conclusions from single-omics analyses are often limited and may miss out on crucial mechanisms that contribute to the disease progression, heterogeneity, or therapeutic resistance.

This review focuses on how multiomics helps in understanding disease mechanisms, particularly in the field of oncology. It explores the importance of combining genomic, transcriptomic, proteomic, and other omics layers in identifying therapeutic targets, improving biomarker discovery and in developing effective combination therapies.

Definitions of omics layers

Multiomics captures biological information across the central dogma from DNA to RNA, proteins, and metabolites. Genomic profiling provides critical insights into mutational burden, driver alterations, and

clinically actionable targets [14, 15]. Genomics focuses on the complete DNA sequence and its variations including insertions and deletions (INDELs), single-nucleotide variants, and structural variations such as copy number variations (CNVs) [2]. Transcriptomic data provide insights on gene regulation, phenotype, and cellular type. It records the entire set of RNA molecules expressed in a cell or tissue at a given time [16]. RNA-seq also captures noncoding RNAs such as long-noncoding RNAs (lncRNAs), microRNAs (miRNAs), and small nucleolar RNAs (snoRNAs) [17]. Transcriptomic data also enable researchers to identify dysregulated pathways, aberrant splicing, and immune signatures associated with the disease [18, 19]. Proteomics profiles the complete set of proteins expressed within a biological system giving information about the functional state of cells. Proteomics data also help in identifying dysregulated signalling pathways, neoantigens, and resistance-associated proteins aiding in biomarker discovery [20, 21]. Moreover, it can be used to map dynamic protein changes in response to therapies, which makes proteomics an important tool in drug development [22]. Metabolomics focuses on the comprehensive characterization of metabolites that contributes to or results from cellular metabolism providing information of the biochemical activity and the physiology of the cell [23]. Metabolomics plays a crucial role in understanding metabolic rewiring, a hallmark of cancer, which reflects alterations in glycolysis, amino acid metabolism, and redox state [24, 25].

Apart from core molecular layers, multiomics also capture regulatory and environmental factors that influences cellular states. Epigenomics studies the complete set of epigenetic modifications on the genetic material of a cell. It focuses on the reversible, heritable chemical changes of DNA and histone proteins that control gene expression without changing the nucleotide sequence [26]. The key modifications it focuses include DNA methylation and histone modifications that affect the gene expression without changing the underlying DNA sequence [26]. Microbiomics determines the function and composition of microbial communities using metagenomic techniques such as 16S rRNA sequencing, Metatranscriptomics (MTx), metaproteomics, or

shotgun metagenomics (MGx) [27]. Radiomics involves the extraction of high-dimensional quantitative imaging characteristics from clinical imaging modalities such as Computed Tomography (CT), Magnetic resonance imaging (MRI), and Positron emission tomography (PET) to characterize tumour heterogeneity, shape, texture, and intensity patterns. Radiomics can also be used to monitor tumour changes over time and guide adaptive treatment strategies. Radiomics is also being increasingly used in cardiovascular disease, pulmonary fibrosis, and neurological disorders, where imaging biomarkers can aid in early diagnosis and risk stratification [28–30].

Concepts and strategies of multiomics integration

Horizontal integration

Multiomics integration strategies are generally classified based on how different types of data are structured and related across biological samples. Horizontal and vertical integration strategies are the two main types of integration strategies used and require different computing approaches.

Horizontal integration involves the integration of the same type of omics data (e.g. transcriptomics or proteomics) from independent datasets, studies, or sets of experiments. Such strategies are necessary in population-level research and large meta-analyses where the aim is to increase statistical power, cross-validating results between cohorts, or improve model generalizability. However, horizontal integration faces significant challenges due to technical heterogeneity, batch effects, variable sequencing depth, and sample quality. Several statistical and machine learning models have been developed to mitigate these issues. ComBat algorithm, which uses an empirical Bayes framework, remains a foundational tool for batch correction [31]. Advanced versions of ComBat algorithms that offers better scalability and performance on complex datasets such as Nested ComBat, OPNested ComBat, ComBat-seq have been developed [32–34].

Harmony and scanorama are two popular tools used in single-cell data integration that projects datasets into a shared low-dimensional space, offering scalable solutions. Harmony has become a popular approach in single-cell transcriptomics for integrating batches of datasets by projecting them into a shared low-dimensional embedding [35]. It has demonstrated good performance in aligning scRNA-seq datasets from various patients and conditions without the over-correction of biologically interesting signals. Similarly, scVI (single-cell variational inference) employs deep generative modelling with variational autoencoders to represent gene expression distributions, allowing for probabilistic batch correction and imputation [36]. This approach has been effective in creating harmonized atlases across a variety of tissue types and TMEs. Scanorama is another recently developed unsupervised method that performs manifold alignment of single-cell data by matching mutual nearest neighbours across batches [37]. It is particularly suited for the integration of large-scale single-cell RNA-seq datasets with varying levels of sparsity. BBKNN (Batch Balanced *K*-Nearest Neighbours) extends the *k*NN graph construction by balancing batch contributions, preserving biological variation during clustering and visualization steps in tools like Manifold Approximation and Projection (UMAP) or *t*-distributed stochastic neighbour embedding (*t*-SNE) [38].

MSstats and Triqler has been used to integrate label-free quantification datasets, correcting for missingness and technical variance in proteomics data [39, 40]. In metabolomics, tools such as Mzmine 3 and XCMS-MELTIN allow alignment and retention time correction across LC-MS batches [41, 42]. QIIME 2 is a popular tool used in

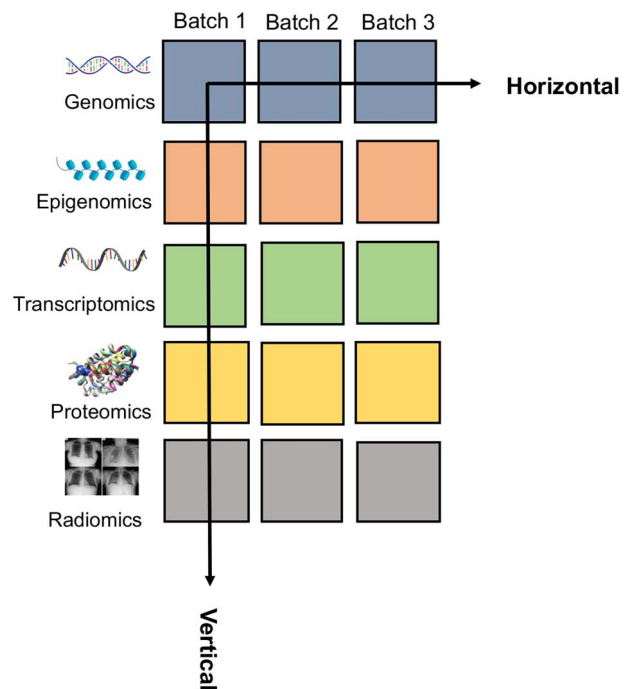


Figure 2 Horizontal versus vertical integration of multiomics data. In horizontal integration, datasets from the same omics type but different batches or cohorts are combined to increase statistical power and reduce batch-specific bias. In vertical integration, multiple omics layers from the same samples (e.g. genomics, epigenomics, transcriptomics, proteomics, radiomics) are integrated to provide cross-modal insights into biological systems and disease mechanisms.

microbiomics that supports meta-analyses by harmonizing 16S rRNA gene sequencing data across cohorts with different primers and sequencing depths [43]. These tools broaden the scope of horizontal integration by enabling reproducible analysis across diverse omics modalities. Despite its strengths in increasing statistical power and reproducibility, horizontal integration may be limited when biological variability across cohorts is affected by technical differences, making it difficult to distinguish true biological signals from batch effects. Its effectiveness also depends on the availability of similar datasets with consistent experimental design and metadata annotation. Figure 2 illustrates vertical and horizontal integration methods of multiomics data.

Vertical integration

Vertical integration enables simultaneous analysis different types of omics data measured from same samples or subjects. The aim of vertical integration is to understand multi-layered biological regulation and explore cross modal interactions such as how transcription is influenced by genomic alterations, or how metabolic products are influenced by post-translational modification. For instance, vertical integration of phosphoproteomic data, gene expression, and somatic mutation data has identified distinct oncogenic pathways to melanoma subtypes [44].

Vertical integration requires specialized tools that can address the heterogeneous nature of multiomics data. MOFA+ is one such tool that uses a factor analysis formulation to identify common and modality-specific factors across omics types, and imputes missing values [45]. iClusterBayes integrates multiomics by latent variable selection modelling, and its more recent versions available on the iClusterPlus

R package are still used for unsupervised subtype discovery of cancer [46, 47]. Vertical integration methods also support a wide range of omics data. DIABLO has gained popularity for integrating transcriptomic, proteomic, and metabolomic layers with built-in visualization and feature selection [48]. Such integrative frameworks are important when analysing complex diseases such as cancer, where regulatory interactions span across the genome, transcriptome, and proteome. While vertical integration provides deeper insight into biological regulation across multiple molecular layers and cross-modal interactions, it is often constrained by the requirement for matched multiomics datasets from the same samples, which can limit sample size and increase data sparsity. Moreover, differences in data scale, noise levels, and missing values across omics layers can complicate integration and interpretation.

Fusion strategies

Fusion methods determine at what stage processing does the integration of multiomics data takes place. Fusion strategies are typically classified into early, intermediate, and late fusion techniques. These approaches differ based on when the data are combined, whether before modelling, during feature transformation, or during decision-making.

Early-stage integration

Early-stage integration, or data-level fusion, involves the concatenation of raw or preprocessed feature values across the various omics layers into a composite matrix before downstream analysis. This approach assumes feature comparability across different omics layers either in scale, distribution, or biological meaning, so that they can be concatenated into a single matrix and analysed together. But given their intrinsic differences, it is important to perform careful normalization and scaling to make this assumption reasonable for analysis. This method is relatively easier to implement, often employing PCA, *k*-means cluster, or elastic net regression on the combined dataset. Recent developments such as OmiEmbed have introduced deep learning-based frameworks and utilize unified embedding networks to capture complex nonlinear relationships among concatenated omics features [49]. These models effectively integrate heterogeneous data types and enable better accuracy during phenotype classification, biomarker discovery, and survival prediction by uncovering complex patterns across heterogeneous data types. However, early-stage fusion methods may not be effective when omics features are inherently incompatible in scale, distribution, or biological meaning, leading to distorted representation and reduced model performance. These models also require careful optimization to overcome challenges of high dimensionality, overfitting, and noise inherent to multiomics data.

Intermediate-stage integration

Intermediate-stage integration, also known as transformation-level fusion, processes each omics dataset separately to extract meaningful latent representations and concatenates different layers together before final model building or clustering. Intermediate-level integration balances between the simplicity of early fusion and flexibility to accommodate the heterogeneous nature of multiomics data by not performing direct concatenation of raw features, thereby having the potential to solve problems regarding modality-specific noise and scale differences. Dimensionality reduction techniques such as matrix factorization, canonical correlation analysis (CCA), and manifold learning are commonly used at this stage to project omics data

into a common latent space that captures common biological signals without losing modality specific information. Neural network-based methods such as variational autoencoders (VAEs) and graph-based models are now being used to derive reliable low-dimensional representations from each data type prior to integration. DeepMO and GLUER examples of such methods that use deep learning methods to derive meaningful embeddings from each data type prior to integration [49, 50]. Intermediate integration methods can handle missing data and batch effects better than early-stage fusion and provides clear insights by identifying common and unique components in different omics data. However, intermediate integration can be limited when the learned latent representations fail to capture true biological signals or when important modality specific information is lost during dimensionality reduction. Careful design and testing methods should be used for such models to ensure that the latent features accurately represent the underlying biology rather than the technical artifacts, highlighting the need for reliable computational frameworks and reproducibility in multiomics studies. A schematic representation of early-, intermediate-, and late-stage fusion is illustrated in Fig. 3.

Late-stage integration

Late-stage fusion or decision-level fusion involves each modality being analysed separately to produce predictions or intermediate outputs such as risk scores, or clustering labels, which are then fused together using ensemble learning or meta-analysis techniques. When compared to early and intermediate methods that fuse data or latent representations, late fusion emphasizes more on modality-specific interpretability, enabling researchers to maintain the integrity of each omics level while fusing interpretations at a higher decision level. This makes it particularly suitable when the datasets are incomplete, highly heterogeneous, or come from different cohorts. Commonly used methods in late-stage fusion include majority voting, weighted averaging, and Bayesian model averaging [51, 52]. More recently developed models use ensemble learning techniques such as stacked generalization, where a meta model is trained to combine predictions from individual models built on separate omics layers [53]. MOLI is an example of one such model that applies late fusion by combining deep neural networks for each omics layer, followed by a logistic regression meta-learner for achieving high performance in predicting drug response using genomic, transcriptomic, and epigenomic data as input [54].

Although late fusion is highly resistant to modality-specific bias and data sparsity, it might lose some of the combined information captured by early and intermediate integration methods. Its accuracy will largely depend on the quality of individual models and the effectiveness of the ensemble model. Yet in cases where datasets are sparse, noisy, or asynchronously obtained, late fusion remains an adaptable and interpretable method, complementary to other integration techniques in the multiomics. However, late fusion may fail to capture cross-modal interactions between different omics layers, as integration occurs only at the decision level rather than at the feature level.

Overall, the choice of integration strategy depends on the characteristics of the data and the analytical objective. Early-stage integration is effective when omics features are comparable and normalized but may fail when feature distributions and biological meanings differ significantly across modalities and is also sensitive to missing data because it requires complete feature matrices. Intermediate integration provides a balance by learning latent representations, making it a more suitable choice for handling heterogeneity and missing data,

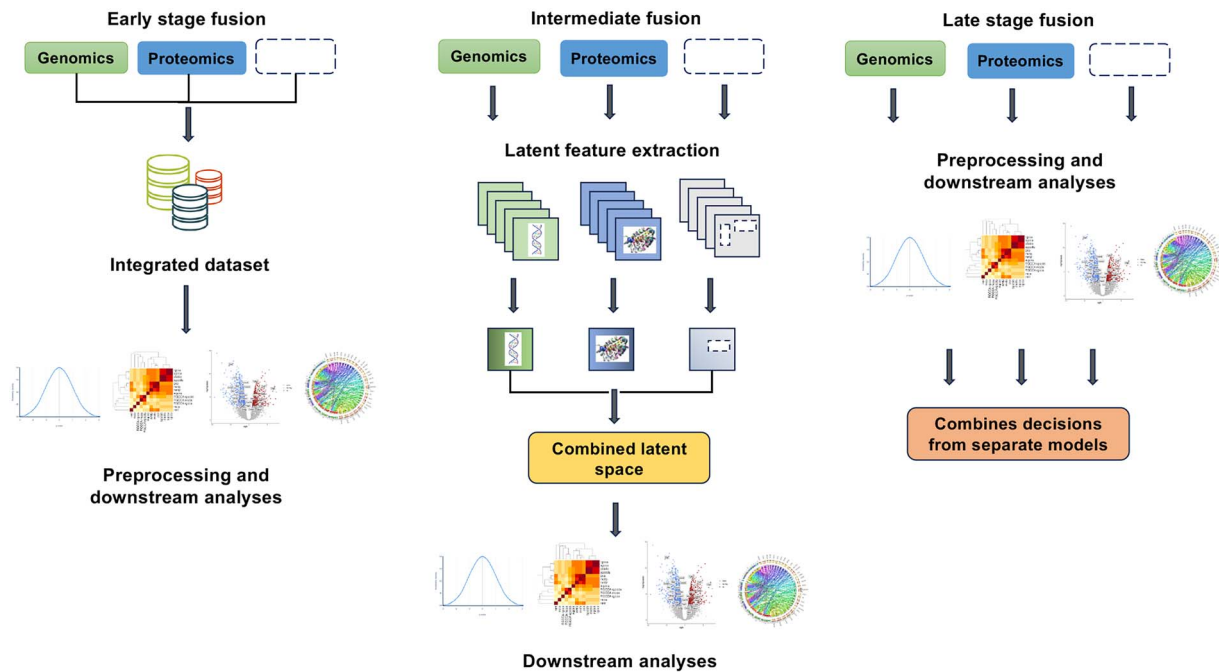


Figure 3 Fusion strategies for multiomics data integration. In early fusion, raw features from multiple omics datasets are combined into a single dataset prior to modelling. In intermediate fusion, each dataset is transformed into latent feature representations, which are subsequently merged for analysis. In late fusion, each dataset is modelled independently, and the resulting predictions are integrated at the decision level. These strategies illustrate how integration can occur at different stages of the analytical pipeline, each with distinct advantages and limitations.

although its performance depends on the quality and interpretability of the learned embeddings. Late-stage integration is particularly suitable for heterogeneous or incomplete datasets, as it preserves modality-specific information, but may fail to capture cross-modal interactions due to the lack of joint feature level modelling. Therefore, the selection of an appropriate integration strategy should consider factors such as data compatibility, sample size, missing data, and the need for interpretability or cross-modal interaction modelling. A summary of different integration strategies is presented in Table 1.

Computational pipelines and frameworks

Increased complexity and heterogeneity of multiomics datasets have prompted the development of end-to-end computational pipelines and frameworks that bring together the integration, analysis, and interpretation of various data modalities under a single framework. These platforms not only make it easy to harmonize and preprocess multi-layered omics data but also provide user-friendly environments to implement sophisticated statistical models and machine learning algorithms, enabling researchers to derive biological insights with greater efficiency and reproducibility. Such frameworks typically integrate a combination of statistical methods, machine learning approaches, and deep learning models for processing, integrating, and analysing multiomics data.

Integrative analysis platforms offer support for a broad range of functionalities, including data normalization, batch effect removal, dimension reduction, clustering, network inference, and biomarker discovery. Galaxy, for instance, is an open-source, web-based platform that integrates many bioinformatics software tools into modular workflows, enabling users without extensive programming expertise to perform multi-step analysis on multiple omics levels [55].

Similarly, popular workflow management tools such as Nextflow, Snakemake, and nf-core pipeline collections provide scalable and

reproducible solutions for large-scale multiomics data processing [56–58]. These tools enable automation of complex analyses involving multiple steps and support integration of different computational methods with standardized pipelines.

In addition, resources such as LinkedOmics, a curated database plus analysis tool that focuses on integrative analysis of cancer multiomics datasets by providing access to data from The Cancer Genome Atlas (TCGA) alongside visualization and statistical modules [59].

Together, these computational pipelines and frameworks form the backbone of modern multiomics analysis, facilitating high-quality data integration, enhancing interpretability, and accelerating decision-making in a variety of biomedical applications.

Together, these computational pipelines and frameworks form the backbone of modern multiomics analysis, facilitating high-quality data integration, enhancing interpretability, and accelerating decision-making in a variety of biomedical applications. Some of the popular tools and platforms commonly used are summarized in Table 2.

Data resources for multiomics integration

Publicly available datasets play an important role in enabling integrative analysis by providing access to large-scale datasets across different biological conditions, populations, and disease types. The increasing availability of such datasets has helped improve multiomics research by supporting reproducibility, cross-study comparison, and development of advanced computational models. These datasets allow researchers to combine datasets across different cohorts, validate findings, and train predictive models with improved generalizability. Moreover, standardized repositories help address challenges related to data sharing, annotation, and long-term accessibility, which are essential for advancing collaborative research in systems biology and precision medicine.

Table 1 Comparison of multiomics integration strategies.

Integration strategy	Description	Pros	Cons	Handling missing data	Best use case
Horizontal	Combine same-omics datasets across cohorts	Increases statistical power	Does not leverage multi-layer interactions	Moderate (depends on preprocessing and batch correction)	Large cohort studies
Vertical	Combine multiple omics from same samples	Captures interlayer biology	Requires matched multiomics data from same subjects	Limited (requires complete matched samples)	Mechanistic studies, pathway and regulatory analysis
Early Fusion	Concatenation before analysis	Simple, captures feature-level synergy	High dimensionality, normalization challenges	Poor (requires complete data matrix)	Small, well-aligned datasets with comparable features
Intermediate Fusion	Dimension reduction then joint modelling	Balance between structure and integration	Model complexity, sensitive to transformations	Moderate (can handle partial missing data)	Heterogeneous datasets with complex relationships
Late Fusion	Independent modelling then result-level integration	Modular; robust per modality	No interaction modelling, less synergy	Good (handles incomplete modalities effectively)	Heterogeneous or partially missing datasets

Several well-established databases support multiomics research across different domains. Large-scale cancer focused initiatives such as TCGA and the International Cancer Genome Consortium (ICGC) provide comprehensive multiomics datasets, including, genomic, transcriptomic, epigenomic, and clinical data [67, 68]. Public repositories such as the Gene Expression Omnibus (GEO) and ArrayExpress host a wide range of transcriptomic and other omics datasets from various experimental studies [69, 70]. Domain-specific repositories such as the PRIDE and Metabolights provide proteomics and metabolomics data, respectively, offering standardized formats and metadata for large-scale analysis [71, 72]. In addition, single-cell and spatial omics resources, including the Human Cell Atlas and Single Cell Expression Atlas, enable a more detailed analysis of cellular heterogeneity and tissue organization [73, 74]. Platforms such as LinkedOmics further extend these capabilities by integrating multiomics datasets with integrated analysis and visualization tools [59]. Together, these data resources form the foundation for multiomics integration, supporting both methodological development and biological discovery. Commonly used data resources for multiomics data are summarized in Table 3.

Preprocessing in multiomics

Preprocessing and harmonization are important steps in multiomics integration as they enable meaningful comparisons of biological signals across diverse datasets by effectively minimizing the technical variability. Standard preprocessing steps generally involve normalization, checking for batch effects, missing values imputation, dimensionality reduction, and scaling.

Normalization is performed to rescale the data into a common scale by reducing technical variability and ensuring comparability across samples and different data types. Quantile normalization, median normalization, log transformation, variance-stabilizing transformation (VST) of RNA-seq, and probabilistic quotient normalization (PQN) in metabolomics are some of the commonly used techniques for normalization [75]. Cross-omics normalization methods like feature-specific quantile normalization and ratio-based scaling have been

developed to enable joint analysis across heterogenous data [76, 77]. Moreover, newer methods such as Combat-Seq that uses normalization frameworks that incorporate biological covariates and sample-specific variables have shown better preservation of true biological signals [34].

Empirical Bayes methods such as ComBat and its extensions are popular approaches and can remove batch effects while preserving biological variation [78]. For single-cell multiomics, Harmony and scVI have been widely used, employing embedding alignment and probabilistic modeling, respectively [35, 79]. Tools such as MOFA+ integrate batch correction with latent factor modeling to denoise heterogeneous data sets simultaneously [45]. Emerging methods now integrate batch correction with imputation and dimensionality reduction within unified workflows. DeepBID and MIDAS are examples of such integrated methods that uses a deep learning framework for simultaneous denoising, batch handling, and dimensionality reduction of single-cell RNA-seq data [80, 81]. The choice of batch correction method can significantly influence downstream biological conclusions, so careful benchmarking and validation are important.

Missing values in multiomics data occur due to technical detection limits, sample dropout, or incomplete experimental designs or poor tissue quality [82]. Missing data imputation is a crucial step to reduce bias in downstream analysis and enable accurate interpretation of complex biological systems. Imputation methods vary depending on the omics type. In transcriptomics, methods like MAGIC and ALRA use manifold learning and matrix completion to infer missing counts. Deep learning methods are also now being used for missing value imputation in multiomics [83, 84].

Dimensionality reduction reduces the complexity and high dimensionality of multiomics data. Classical techniques such as principal component analysis (PCA) and t-SNE are standard, while more recent techniques such as Uniform UMAP are better at preserving local and global structure. Scaling techniques (e.g. z-score scaling, min-max scaling) are applied prior to dimensionality reduction to avoid features with variations in data ranges dominating analyses. Scaling normalizes these differences and ensures comparison across data types [77].

Table 2 Representative computational tools and pipelines for multiomics integration.

Tool/platform	Type	Supported omics	Functionality	Key features	Reference
MOFA+	Integration	Transcriptomics, epigenomics, proteomics, metabolomics	Factor analysis for heterogeneous multiomics	Learns shared & modality-specific latent factors; handles missing data	[45]
DIABLO (mixOmics)	Integration	Transcriptomics, proteomics, metabolomics	Sparse multivariate integration & feature selection	sparse Partial Least Squares (sPLS)-based integrative modelling; biomarker discovery	[48]
iClusterPlus	Integration	Genomics, transcriptomics, epigenomics, proteomics	Joint latent variable modelling	Integrated clustering; supports mixed data types	[47]
SNF (Similarity Network Fusion)	Integration	Genomics, transcriptomics, epigenomics, proteomics	Network-based data fusion & clustering	Fuses patient similarity graphs from multiple omics	[60]
Harmony	Integration	Single-cell RNA, ATAC, multiomics	Batch correction & integration (mainly single-cell)	Fast shared-embedding alignment across datasets	[35]
Scanorama	Integration	Single-cell RNA-seq	Large-scale single-cell alignment	Manifold alignment; robust to limited cell-type overlap	[61]
TotalVI	Integration	Single-cell RNA + protein (CITE-seq)	Joint modelling of scRNA + protein data	Variational inference; uncertainty-aware latent space	[62]
MultiVI	Integration	Single-cell RNA + ATAC	Multi-modal variational inference	Joint latent representation for discrete & continuous modalities	[63]
OmiEmbed	Integration	Genomics, transcriptomics, epigenomics, proteomics	Deep learning unified embedding	Learns nonlinear joint representations; multi-task (classification/survival)	[49]
CustOmics	Integration	Genomics, transcriptomics, proteomics	Autoencoder-based customizable integration	Flexible architectures; improved prediction & survival modelling	[64]
Galaxy	Pipeline	All major omics (genomics, transcriptomics, proteomics, metabolomics)	Modular workflow management	User-friendly web platform for reproducible multiomics pipelines	[55]
LinkedOmics	Pipeline	Genomics, transcriptomics, proteomics (TCGA/CPTAC)	Cancer-focused database and analysis	Integrates TCGA & CPTAC datasets; enrichment, correlation, and survival modules	[59]
OpenOmics	Pipeline	Genomics, transcriptomics, epigenomics	Cloud-based multiomics analysis	Scalable, high-throughput workflows; collaborative cloud environment	[65]
Seurat v5	Pipeline	Single-cell RNA, ATAC, protein, spatial	Multi-modal single-cell & spatial integration	Weighted nearest neighbour-based framework for RNA, ATAC, protein, and spatial data	[66]

Multiomics in disease mechanisms

From correlation to causation

Unravelling the complexity of disease mechanism requires causal inference across multiple molecular layers beyond simple associations. According to the central dogma of molecular biology, genetic information passes from DNA to RNA to protein and ultimately to phenotype. Evidence suggests that molecular layers closer to the phenotype, such as the proteome, correlate more with traits than transcripts or genetic variants, highlighting the importance of multiple omics layers in causal frameworks [85]. Multiomics facilitates this by integrating from DNA, RNA, protein, and metabolites, enabling reasoning that approximates the flow of causality within biological systems. This shift towards causality is also evident in network-based computational models like CausalPath, integrating proteomic changes and prior knowledge from pathway databases to infer directed causal paths between perturbations and downstream responses [86].

Similarly, probabilistic graphical models like Bayesian networks and structural equation models are being increasingly used with multiomics data to model probabilistic dependencies and infer directional effects between molecular variables, even under high dimensionality [87, 88].

Recent methodological advances have also introduced graph-based deep learning architectures such as GraphSurv, LAGProg, MoGCN, MOGAT, and MOGONet that model multiomics features on gene or patient similarity graphs to learn causal relationships for predicting clinical outcomes and disease subtypes [89–93]. In addition, recent advancements have explored multi-task graph learning frameworks to jointly model relationships between different molecular entities, such as lncRNAs, miRNAs, and diseases, improving prediction performance through shared representations across multiple related predictions [94]. Graph-based deep learning models use Graph Convolutional Networks (GCNs) to improve classification accuracy for disease

Table 3 Commonly used data resources for multiomics integration.

Database	Data type	Key features	Reference
TCGA (The Cancer Genome Atlas)	Genomics, transcriptomics, epigenomics, clinical	Large-scale cancer multiomics datasets	[67]
ICGC (International Cancer Genome Consortium)	Genomics, transcriptomics, clinical	International cancer cohort data; cross-population studies	[68]
GEO (Gene Expression Omnibus)	Transcriptomics, epigenomics	Public repository for gene expression data; diverse datasets	[69]
ArrayExpress	Transcriptomics, multiomics	Curated functional genomics data; standardized metadata	[70]
PRIDE	Proteomics	Mass spectrometry-based proteomics datasets	[72]
MetaboLights	Metabolomics	Metabolite profiling data	[71]
Human Cell Atlas	Single-cell omics	High-resolution cellular data across tissues	[73]
Single Cell Expression Atlas	Single-cell transcriptomics	Processed single-cell datasets with visualization tools	[74]
LinkedOmics	Multiomics (cancer-focused)	Integrated TCGA data with analysis and visualization tools	[59]

subtyping, diagnosis, and identification of new targets or driver genes [95]. These models typically present multiomics data as interconnected graphs, where nodes represent genes or samples and edges capture the biological relationships or similarity. For instance, models such as MoGCN and MOGONet use GCNs to learn modality-specific representations from each omics layer and integrate them using a supervised classification model for patient classification and biomarker prediction [89, 92]. Other models such as MOGAT extend this approach by using attention mechanisms to weight the relative importance of different omics modalities during integration [91]. Frameworks like GraphSurv use graphical representations to identify patient-level relationships for predicting a patient’s survival rate [90]. These supervised models generally outperform purely correlational ones by taking advantage of structured dependencies due to known pathways or network topology inference.

In addition to correlation-based methods, causal inference methods such as Mendelian randomization are also used in multiomics research to infer effects from genetic variation to downstream molecular and clinical phenotypes [96].

Multiomics in cancer

Recent multiomics studies have transformed our understanding of cancer biology by enabling integration and analyses of diverse molecular layers to yield actionable insights.

The Pan-Cancer Analysis of Whole Genomes (PCAWG) is considered as a landmark study, integrating whole genome and transcriptomic profiles of metastatic and primary tumours in 38 cancer types [67]. The study revealed that metastatic tumours exhibit a distinct structural variant and clonal mutation compared to primary tumours, offering insights into disease progression. These findings highlight the ability of pan-cancer multiomics approaches to distinguish tissue-specific and shared molecular mechanisms. The Clinical Proteomic Tumour Analysis Consortium (CPTAC) study on colon cancer, using whole-exome sequencing, transcriptomics, global proteomics, and phospho-proteomics data, identified the correlation between Rb phosphorylation and increased proliferation, while microsatellite instability-high (MSI-H) tumours exhibit glycolytically driven immune evasion by suppressing CD8+ T-cell activity [97]. These observations highlight

the potential of targeting metabolic pathways and kinase inhibitors to improve outcomes in specific patient subsets.

The emergence of spatial transcriptomics has further enriched the context by introducing a spatial dimension to molecular profiling. In a recent study on hepatocellular carcinoma (HCC), combining scRNA-seq and spatial transcriptomics showed overexpression of PRKDC in immune-excluded tumour regions [98]. Functional assay revealed PRKDC’s role in cell proliferation, making it a potential prognostic biomarker and therapeutic target. Another study using spatial single-cell multiomics on HCC identified a specific stromal fibroblast subpopulation marked by COL1A2, COL4A2, CTGF, and FSTL1 [99]. These cancer-associated fibroblasts were found to promote aggressive growth and were associated with poor prognosis.

Melanoma, characterized by its high immunogenicity and therapeutic heterogeneity has been widely studied using multiomics approaches. An integrated single cell and spatial transcriptomic analysis of acral melanoma found an association of MYC+ tumour cells and fatty acid oxidation-driven metabolism in localized lymph node metastasis, identifying cell state and metabolic dependencies that lead to dissemination [100]. Another multiomics analysis of 68 melanoma cell lines integrating genomics, methylation, RNA, miRNA, lncRNA, and proteomic data revealed how intrinsic cell-state phenotypes are regulated through complex multi-layer regulatory networks, linking phenotypic plasticity to variable drug resistance and treatment outcomes [101]. Furthermore, a multiomics study comparing melanoma brain metastases to extracranial lesions showed an enrichment of oxidative phosphorylation pathways and lower immunogenicity in the brain microenvironment, revealing the site-specific adaptations in metastatic melanoma [102].

Similar studies were also found to be effective in many types of cancers, including breast and lung cancers, as well as rare malignancies such as skull-base chordoma, and is summarized in Table 4. These studies highlight the effectiveness of multiomics studies in understanding disease mechanisms in oncology.

Multiomics in aging research

Multiomics approaches are now also increasingly used in ageing research. Ageing is a time-dependent physiological process that is

Table 4 Applications of multiomics in cancer: insights from recent studies.

Disease	Omics layers	Key findings	Reference
Pan-cancer (PCAWG, 38 tumour types)	WGS, RNA-seq	Metastatic tumours show distinct structural variants and clonal evolution pathways	[67]
Colon cancer (CPTAC proteogenomics)	WES, RNA, proteomics, phosphoproteomics	Rb phosphorylation and metabolic rewiring linked to immune exclusion and aggressiveness	[97]
Hepatocellular carcinoma (HCC)	scRNA-seq + spatial transcriptomics	PRKDC overexpressed in immune-excluded tumour niches	[98]
Liver (HCC)	Single-cell & spatial transcriptomics	F5-CAF fibroblast subset promotes stemness and adversity via niche interactions	[99]
Breast cancer—difficult-to-treat basal-like (DTBC)	WGS, RNA-seq, global proteomics, phosphoproteomics	High-relapse DTBC shows 1q21 amplification and phosphoproteomic signatures linked to proliferation and signaling rewiring.	[103]
Melanoma	scRNA + spatial transcriptomics	MYC+ melanoma cells and FAO metabolism linked to lymph node metastasis niches	[100]
Melanoma	WES, transcriptome, miRNA, lncRNA, methylation, RPPA	Distinct intrinsic phenotypic states controlled by multi-layer regulation, tied to therapy response patterns	[101]
Melanoma brain metastases	Genomics, transcriptomics, proteomics	Brain metastases enriched for oxidative phosphorylation and display reduced immunogenicity	[102]

characterized by multiple biological layers. Multiomics studies provide a comprehensive view of how molecular changes across multiple omics layers affect ageing, uncovering tissue- and cell-type-specific patterns that are often missed by single omics analyses.

A recent study by analyzing proteomic, metabolomic, and methylomic profiles from multiple cohorts over time reported that ageing progresses in a nonlinear manner, with each omics layer exhibiting distinct transition points that single-layer ageing clocks fail to detect [104]. Multiomics ageing clocks that integrate transcriptomic, epigenetic, proteomic, and metabolic data outperform single-omics models by capturing complementary biological hallmarks, resulting in improved accuracy and generalization across diverse populations and tissues [104, 105].

Single cell multiomics has transformed ageing research by providing a granular-level view of cellular changes that occur over time. For instance, a recent study combining single-nucleus RNA seq and ATAC-seq data mapped how ageing affects human skeletal muscles and created the first atlas of cellular senescence in ageing human skeletal muscle [106]. Multiomics has also enabled mapping of molecular ageing across multiple tissues. A cross-tissue multiomics study combining proteomic and metabolomic profiling across 21 organs in mice have identified both tissue-specific and shared molecular remodelling during ageing [107]. The findings reveal disruption of coordinated molecular pathways, with alterations in complement activation, lipid metabolism, and immune regulation, providing a systems-level view of ageing in different organs.

Together, these studies highlight the contribution of multiomics studies in unravelling the complexity of ageing. A selection of recent high-impact studies including those already mentioned is summarized in Table 5.

Multiomics in immunology

The immune system is highly complex and exhibits high cellular heterogeneity, dynamic signaling networks, and metabolic plasticity, making it difficult for single omics approaches to capture [111, 112]. Multiomics approaches integrate multiple omics layers at both bulk

and single-cell resolution, enabling detailed insights into immune regulation mechanisms.

A multi layered analysis using scRNA seq, ATAC-seq, spatial transcriptomics, and imaging data helped reveal the microenvironment of the inflamed synovium at an unprecedented level [113]. While transcriptomic profiling alone can identify types of immune cells, this multiomics approach revealed the distinct epigenetic landscapes of fibroblast-like synoviocytes (FLSs) that shape their inflammatory behaviour. The study identified key FLS subsets that drive pathogenic immune and stromal interactions by connecting chromatin accessibility patterns with spatial transcriptional profiles, providing insights that would otherwise remain elusive with single omics analyses.

A study by Mennillo *et al.* [114] combined single-cell and spatial multiomics to investigate the effects of anti-integrin therapy in inflammatory bowel disease (IBD). The study revealed that response to the therapy correlated with shifts in mononuclear phagocyte and fibroblast populations and their spatial organization, demonstrating how tissue microenvironments also play a role in therapeutic resistance. Such findings highlight the ability of multiomics studies to capture intercompartment interactions that single-cell transcriptomics alone could overlook. Similar studies were also performed on Covid-19. An integrated study on Covid-19 using proteomics, metabolomics, and transcriptomics data revealed disruptions of central metabolism and mitochondrial pathways correlated with disease severity [115].

The strength of multiomics lies in its capability to integrate multiple molecular levels while simultaneously capturing the diversity of cells and their interaction within the microenvironment. This dual perspective offers a more comprehensive framework to understand immunological diseases. Some of the recent studies that effectively used multiomics to understand the disease mechanisms in immune mediated diseases are summarized in Table 6.

Multiomics in systems vaccinology

The advent of multiomics systems vaccinology integrating transcriptomics, proteomics, metabolomics, cytometry, and other high-dimensional modalities has enabled the delineation of early molecular signatures that correlate with vaccine-induced immune outcomes (e.g.

Table 5 Multiomics studies in aging: models, modalities, and molecular insights from recent studies.

Model/cohort	Omics layers	Key findings	Reference
Human longitudinal cohorts (multi-cohort)	Proteome, metabolome, methylome, clinical	Multiomics revealed nonlinear ageing patterns with each omics layer showing distinct transition points.	[104]
Human immune landscape across lifespan	scRNA-seq, TCR/BCR sequencing, mass cytometry proteomics	Elucidates distinct ageing patterns across T- and B-cell subsets, trajectories in clonality, and protein markers, enabling an immune-age prediction model based on integrated omics	[108]
Cross-tissue mouse atlas of ageing (21 organs)	Proteome, metabolome, metagenomic (cross-tissue)	Tissue-specific metabolic and protein remodelling disrupt coordinated pathways across organs	[107]
Population cohorts (UK Biobank-scale multi-omics on ageing & longevity)	Genomics, transcriptomics, metabolomics	Ageing and longevity driven by distinct biological pathways beyond single-omics signals	[109]
Longitudinal human proteome (3 time points, 9 years)	Proteome + clinical traits	Longitudinal changes in serum proteins identify biomarkers linked to age-related chronic disease risk	[110]
Human skeletal muscle ageing (single-cell multiome atlas)	snRNA-seq + snATAC-seq	Defined a cellular senescence atlas in ageing skeletal muscle, revealing heterogeneity in SASP regulation.	[106]

Table 6 Multiomics in immune-mediated diseases: disease-specific insights from integrated analyses.

Disease	Omics layers	Key findings	Reference
Rheumatoid arthritis (RA) synovium	scRNA-seq + scATAC-seq + spatial transcriptomics + multiplexed imaging	Reveal pathogenic fibroblast-like synoviocyte (FLS) subsets driving inflammatory circuits.	[113]
Inflammatory bowel disease (IBD)	Single-cell & spatial multi-omics	Shows how stromal, epithelial, and immune compartments reorganize under anti-integrin therapy, explaining response versus nonresponse	[114]
COVID-19 (severity stratification)	Proteomics + metabolomics + transcriptomics (multi-omics networks)	Progressive disruption of central metabolism mechanistically tracks with disease severity	[115]
Multiple sclerosis (MS)	Genetics + plasma & brain proteomics (multi-omics causal integration)	Prioritizes 18 proteins with potential causal roles in MS by integrating multi-omics and causal inference	[116]
Systemic lupus erythematosus (SLE)	Transcriptomics + metabolomics + (immune) proteomics/lipidomics (integrative)	Integrative B-cell and systemic profiling reveal metabolic changes linked to SLE activity	[108]

magnitude of antibody- or T-cell response) across diverse platforms [117, 118]. In a systems vaccinology study of H5N1 + AS03 vaccination in humans, a platelet- and adhesion-related transcriptional signature at Day 7 postvaccination was predictive of antibody durability (i.e. longer-term maintenance) across individuals [119]. Furthermore, a recent ‘universal signature’ across 13 different vaccines was shown to forecast the magnitude of antibody induction, setting the stage for pan-vaccine predictive biomarkers [120]. Multiomics resources such as the CMI-PB dataset (Tdap booster context) facilitate benchmarking predictive modelling approaches by providing matched baseline and postvaccine omics data across modalities [121]. Finally, methodological advances like RISE (Rank-based Identification of high-dimensional Surrogate Markers) address the challenge of inferring longitudinal immunity (e.g. antibody titres at late timepoints) from early omics measurements and may help flag nonresponders or subjects with waning responses [122]. These findings offered a systems-level perspective, enabling more accurate patient stratification.

Multiomics in identification of biomarkers and design of novel therapies

The development of effective therapies depends on the detailed understanding of molecular mechanisms driving the disease. Traditional single-omics approaches only provide a limited knowledge of the complex biological systems involved. Multiomics integration not only helps in understanding the complex disease mechanisms involved but also provides novel insights into drug response mechanisms and aids in revealing new therapeutic targets and pathways. Compared to single-omics approaches, multiomics approaches enables researchers to identify the molecular changes and the drivers of those changes, paving the way for precision diagnostics and targeted therapies in various diseases.

In oncology, disease heterogeneity and therapy resistance often undermine the predictive value of individual biomarkers. Predictive models integrating multiple omics layers perform better than those solely based on a single layer, enabling accurate patient

categorization and guiding the selection of drug combinations targeted to specific disease subtypes. A recent multiomics study by Chen *et al.* on metastatic melanoma patients characterize pyroptosis-related signalling modules that classify immunotherapy responder and nonresponder groups [123]. The study predicted response beyond T-cell infiltration markers alone, pointing to pyroptosis and metabolic inflammation as a target for therapeutic intervention. Similarly, another integrated study on melanoma defined adenosine-signalling program predictive of immune checkpoint inhibitor responsiveness [124]. The adenosine pathway was found to represent a checkpoint responsive to therapeutic modulation, as a potential biomarker and providing insights into resistance pathways. In head and neck squamous cell carcinoma, integrated analysis of genomics, copy number variation, DNA methylation, and transcriptomics found three different molecular subtypes with distinct genomic features and TME compositions. Notably, one of the subtypes was characterized by activated epidermal growth factor receptor (EGFR) signalling and an inflamed microenvironment with cancer-associated fibroblast and vascular infiltration, demonstrating sensitivity to EGFR inhibitors. Such integrative studies enable efficient patient subtype classification and targeted therapeutic intervention.

Multiomics approaches are also being increasingly applied in ageing research and immunology for biomarker identification and drug design. In a longitudinal study of ageing individuals by combining proteomic, metabolic, and methylation profiles suggested that molecular changes occur in distinct phases, with major changes happening around midlife and in later stages [104]. These changes were linked with serious health issues such as heart and kidney disorders, and the findings offered insights into biomarkers that correlates with ageing and chronic disease. In a randomized placebo-controlled trial, multiomics analysis found reduced biological age in older adults when treated with therapeutic plasma exchange combined with intravenous immunoglobulin (IVIG) [125, 126]. Similarly, in immune mediated diseases, multiomics has proven effective in identifying targets and predicting therapeutic responses. In Crohn's disease, baseline gut microbiome and bile acid metabolism profiles predicted responses to anti- $\alpha 4\beta 7$ -integrin therapy. Increased abundance of secondary bile acid-producing microbes (*Clostridium*, *Lactobacillus*) and elevated levels of secondary bile acids were predictive of good outcomes [127]. Such multi-omics biomarkers not only outperform conventional inflammatory markers but also implicate the microbial-metabolite-receptor axis (FXR/TGR5) as a mechanistic target for probiotic or metabolic therapies.

Multiomics approaches are redefining biomarker discovery and drug design across various disease domains. These approaches move beyond mere correlation by connecting the markers with causal pathways or regulatory mechanisms, providing deeper understandings that can guide targeted interventions. Table 7 summarizes some of the recent studies that used multiomics approaches to identify novel markers and develop new therapeutic combinations. These examples highlight the impact of multiomics frameworks identifying targets and in providing personalized care. As technologies advance and datasets grow, multiomics is poised to become a central pillar of precision medicine.

Discussion

Limitations and current challenges in multiomics integration

Despite the growing success of multiomics in understanding disease mechanisms and guiding precision medicine, several challenges

still hinder their broader application. These limitations arise from both computational and biological aspects. One of the major challenges is the technical and analytical complexity in integrating heterogeneous data types from various platforms. Each omics layer has unique characteristics in terms of dimensionality and scale. Integrating these different layers without losing biological signal is highly important. While early-stage integration can oversimplify relationships, intermediate- and late-stage integration machine learning or graphical models require considerable computational knowledge and resources.

Multiomics data are often generated from various experimental platforms, laboratories, and protocols, resulting in batch effects, scale discrepancies, and missing data that cause challenges in downstream analyses. Addressing these challenges are essential for ensuring the integrity and comparability of multiomics data prior to integration. Various preprocessing methods required increases the computational complexity of integrating multiomics data. Each omics modality presents its own preprocessing challenges. For instance, transcriptomics count data require normalization for sequencing depth and gene length, proteomics data must be adjusted for batch-specific biases, and metabolomics data often require scaling to account for large variations in metabolite concentrations.

Sample heterogeneity and small cohort sizes also pose significant challenges. Generating high-quality multiomics datasets from the same sample is technically demanding and expensive, often requiring compromises in resolution or coverage. A smaller cohort size reduces the statistical power of the study and increases the likelihood of false positives or overfitting. Smaller sample size also makes it difficult to capture the biological heterogeneity of the disease.

Moreover, batch effects and platform biases might be mistaken as biological variation, requiring normalization and corrections. Batch effects are caused by nonbiological variation due to differences in sample processing times, reagents, or equipment and can hide true biological signals leading to inaccurate conclusions. Correcting batch effects is particularly challenging in multiomics studies where each omics type may have distinct batch structures. High dimensionality and presence of missing data also increase the computational complexity involved in integrating multiomics data. Missing value imputation has shown to improve accuracy in downstream analyses. Despite advances, imputing missing values remains a challenge when the proportion of missing data is high and is varied across different omics datasets [82].

Lack of end-to-end computational pipelines and standardized workflows is another major bottleneck in multiomics integration. Most analysis are performed on custom-built pipelines that leads to poor reproducibility and comparability across studies. Absence of automated pipelines makes it difficult to handle high-dimensional datasets, slowing down analysis and limiting accessibility for noncomputational researchers.

Biologically, another concern is sparse functional annotation of multiomics correlations, highlighting the need for approaches that go beyond simple correlations and focus on uncovering true biological mechanisms, such as casual inference models. Current integration methods continue to face persistent technical challenges, including limited scalability, reduced interpretability, and poor generalizability across independent datasets. Furthermore, the absence of standardized benchmarking frameworks hinders objective and reproducible comparison of model performance. Although existing computational platforms can effectively identify patterns across multiple data layers, the mechanistic and causal interpretation of detected associations

Table 7 Recent applications of multiomics in biomarker discovery and therapy design.

Disease/focus	Omics layers	Key findings	Type	Reference
Head & neck SCC (cancer therapy stratification)	Genomics, CNVs, methylome, transcriptome	Defines molecular subtypes with distinct therapeutic vulnerabilities (e.g. EGFRi sensitivity)	Therapy Design	[128]
Metastatic melanoma (pyroptosis & immune microenvironment)	Transcriptomics, proteomics, metabolomics	Identifies pyroptosis-related signalling modules predictive of immunotherapy response	Biomarker Discovery	[123]
Melanoma tumour ecosystem (adenosine signalling)	Transcriptomics, proteomics, metabolomics	Adenosine pathway signalling identified as predictive of immunotherapy benefit	Biomarker Discovery	[124]
Difficult-to-treat breast cancer (DTBC)	Whole-genome sequencing, transcriptomics, global proteomics, phosphoproteomics	Integrated proteogenomic profiling identifies relapse-risk subtypes and proliferation-associated vulnerabilities in difficult-to-treat breast cancer.	Biomarker Discovery/Therapy	[103]
Ageing cohort (longitudinal proteome + clinical)	Proteome, clinical	Identifies serum proteome ageing score predictive of cardiometabolic disease risk	Biomarker Discovery	[110]
Lung adenocarcinoma and squamous NSCLC	Genomics, transcriptomics, proteomics, phosphoproteomics	Five subtypes identified; proliferative subtype with PI3K–Akt activity and Whole-Genome Doubling linked to poor survival; phosphoproteomics reveals signalling and metastatic drivers.	Biomarker Discovery/Therapy	[129]
Cervical cancer (plasma proteomics)	Targeted proteomics (Olink panel)	Five-protein signature predictive of immunotherapy response and prognosis	Biomarker Discovery	[130]
COVID-19 immunopathology	Plasma proteomics + metabolomics, transcriptomics	Multiomics networks decouple metabolic versus inflammatory drivers of severity, guiding therapy	Biomarker Discovery/Therapy	[131]
IBD integrin therapy response	Whole-genome metagenomics, metabolomics	Microbiome-mediated bile acid metabolism predicts response to anti- $\alpha 4\beta 7$ therapy	Biomarker Discovery	[127]
Ageing rejuvenation interventions	Transcriptome + methylome + proteome	Multiomics clocks detect coordinated biological age reversal postplasma exchange + IVIG	Biomarker Discovery/Therapy	[125]

remains uncertain without complementary experimental validation. This uncertainty substantially restricts the translational relevance of multiomics findings in clinical contexts. To address these limitations, future research should prioritize the incorporation of explainable artificial intelligence, network-based modelling approaches, and rigorous experimental validation strategies. Specifically, adopting causal inference frameworks will enable researchers to transition from purely correlative analyses towards more biologically meaningful and clinically actionable interpretations. Collectively, resolving these methodological and biological challenges is critical for enhancing the reliability, reproducibility, and translational applicability of multiomics research while simultaneously defining key priorities for advancing the field.

Future directions

The future of multiomics is likely to be determined by several converging developments. The application of deep learning techniques is expected to improve pattern recognition in complex multilayered datasets. Emerging techniques such as digital twins, causal inference models aimed at improving mechanistic interpretation, and temporal

multiomics are showing promise in dynamic disease modelling and optimizing therapy. Experimentally, the advent of ultra-sensitive multiomics profiling of low-input samples such as single-cell multiomics or spatial-omics will enable us to improve tissue microenvironment understanding and cellular states.

Importantly, multiomics also holds immediate translational relevance. Integrative and personalized medicine could be enhanced by multiomics via tailoring interventions to individual molecular profiles [132]. This approach enables practitioners to select conventional or complementary therapies based on a patient's unique genetic, proteomic, and metabolic signatures, thereby optimizing outcomes and minimizing adverse effects. In oncology, for instance, multiomics-driven biomarker panels already guide drug selection and monitor therapeutic efficacy [133, 134].

Besides, initiatives such as the Human Cell Atlas and TCGA Pan-Cancer Atlas are laying the foundation for open-access, harmonized multiomics datasets for cross-cohort meta-analyses and accelerating discovery. Eventually, the full potential of multiomics will be achieved through not only technological development but also deliberate effort towards standardization, multidisciplinary collaboration, and equitable data sharing. More recently, large language models (LLMs) have

emerged as a promising tool in multiomics research. These models can integrate heterogeneous biomedical data with textual knowledge from scientific literature, enabling improved interpretation of complex biological relationships [135]. LLMs also facilitate automated hypothesis generation, data annotation, and knowledge extraction, offering a new paradigm for integrating omics data with clinical and biological insights [135]. However, challenges such as interpretability, data bias, and validation remain to be addressed before widespread adoption.

Conclusion

Multiomics has evolved from a conceptual framework into a practical approach, improving our ability to understand the molecular architecture of complex diseases. By integrating heterogeneous layers of genomics, transcriptomics, proteomics, metabolomics, epigenomics, and medical imaging, it overcomes the limitations of single-omics approaches and provides a better understanding of disease heterogeneity, progression, and treatment response. However, current integration strategies have many limitations. Early integration captures interactions between various data types but is limited by dimensionality and noisy data. Intermediate approaches, including network-based and latent factor models such as MOFA+ and similarity network fusion, preserve the modality-specific structure and are easier to interpret, yet remain sensitive to batch effects. Late-stage integration is more flexible and handles missing data better but may overlook interactions between different omics layers.

Deep learning methods, particularly graph neural networks, are now more widely used and can capture complex patterns in multiomics data. However, their limited interpretability remains a major challenge for clinical adoption, leading to growing interest in explainable models that incorporate biological knowledge.

Several challenges still remain, including differences between datasets from different platforms, lack of standardized preprocessing methods, and the computational demands of large and sparse datasets such as single cell and spatial data. However, new developments in spatial and single-cell multiomics are helping to address these challenges by providing better understanding of tissue organization and cellular features. In addition, single-cell multimodal assays such as CITE seq and 10× Multiome enable the simultaneous profiling of regulatory and phenotypic states within the same cell, while approaches such as federated learning support data analysis across institutions without compromising on data privacy. Moreover, the integration of multiomics with artificial intelligence, digital pathology, and clinical data is enabling the development of patient-specific digital twins and predictive disease modelling.

Looking ahead, progress in multiomics will depend on better data standardization, scalable computational frameworks, and improved model interpretability. With these developments, multiomics is likely to play an important role in biomarker discovery, therapy development, and precision medicine.

Key Points

- We review current multiomics technologies and integration strategies across genomics, transcriptomics, proteomics, metabolomics, and other regulatory layers, providing a systems-level perspective of disease biology.

- The article covers major computational frameworks for multiomics integration, including statistical models, network-based approaches, and machine learning methods for extracting mechanistic and clinically relevant insights.
- Representative applications demonstrating how multiomics has been used in decoding disease mechanisms, refining molecular subtypes, identifying biomarkers, and enabling therapeutic discoveries across cancer, ageing, and immunological disorders is discussed.
- Finally, we highlight the current challenges, including data heterogeneity, computational complexity, and causal inference limitations, and discuss future directions involving single-cell, spatial multiomics, and advanced integrative modeling approaches.

Acknowledgements

The authors wish to thank their colleagues for insightful discussions and critical reading of the manuscript. We are also grateful to the anonymous reviewers for their constructive comments, which helped improve this work.

Author contributions

Joyal Mathew (Writing—original draft, Methodology, Investigation, Formal analysis), Ripsy Merrin Chacko (Writing—review & editing, Methodology), Jagadeesh Bayry (Writing—review & editing), Jean-François Zagury (Writing—review & editing), and Mano Joseph Mathew (Writing—review & editing, Supervision, Project administration, Investigation, Formal analysis, Conceptualization)

Funding

None declared.

Data availability

No new data were generated or analysed in support of this research.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this manuscript, the authors used ChatGPT (GPT-4o, OpenAI, May 2025) solely to improve grammar, syntax, and stylistic clarity. The tool was not employed for conceptualization, study design, data analysis, interpretation, or drawing scientific conclusions.

References

1. Juan H, Huang H. Quantitative analysis of high-throughput biological data. *Wiley Interdiscip Rev Comput Mol Sci* 2023;**13**. <https://doi.org/10.1002/wcms.1658>
2. de Anda-Jáuregui G, Hernández-Lemus E. Computational oncology in the multi-omics era: state of the art. *Front Oncol* 2020;**10**. <https://doi.org/10.3389/fonc.2020.00423>
3. Neophytou CM, Panagi M, Stylianopoulos T *et al*. The role of tumor microenvironment in cancer metastasis: molecular mechanisms

- and therapeutic opportunities. *Cancers* 2021;**13**:2053. <https://doi.org/10.3390/cancers13092053>
4. Erasha AM, EL-Gendy H, Aly AS *et al.* The role of the tumor microenvironment (TME) in advancing cancer therapies: immune system interactions, tumor-infiltrating lymphocytes (TILs), and the role of exosomes and Inflammasomes. *Int J Mol Sci* 2025;**26**:2716. <https://doi.org/10.3390/ijms26062716>
 5. Entenberg D, Oktay MH, Condeelis JS. Intravital imaging to study cancer progression and metastasis. *Nat Rev Cancer* 2023;**23**:25–42. <https://doi.org/10.1038/s41568-022-00527-5>
 6. Chu X, Zhang B, Koeken VACM *et al.* Multi-omics approaches in immunological research. *Front Immunol* 2021;**12**. <https://doi.org/10.3389/fimmu.2021.668045>
 7. Pettini F, Visibelli A, Cicaloni V *et al.* Multi-omics model applied to cancer genetics. *Int J Mol Sci* 2021;**22**. <https://doi.org/10.3390/ijms22115751>
 8. Campanharo CV, Dos Santos Silveira LV, Meira DD *et al.* Pan-cancer and multiomics: advanced strategies for diagnosis, prognosis, and therapy in the complex genetic and molecular universe of cancer. *Clin Transl Oncol* 2025;**27**:2936–54. <https://doi.org/10.1007/s12094-024-03819-4>
 9. Engreitz JM, Lawson HA, Singh H *et al.* Deciphering the impact of genomic variation on function. *Nature* 2024;**633**:47–57. <https://doi.org/10.1038/s41586-024-07510-0>
 10. Titlow JS, Kiourlappou M, Palanca A *et al.* Systematic analysis of YFP traps reveals common mRNA/protein discordance in neural tissues. *J Cell Biol* 2023;**222**. <https://doi.org/10.1083/jcb.202205129>
 11. Zeng Z, Li Y, Li Y *et al.* Statistical and machine learning methods for spatially resolved transcriptomics data analysis. *Genome Biol* 2022;**23**:83. <https://doi.org/10.1186/s13059-022-02653-7>
 12. Dar MA, Arafah A, Bhat KA *et al.* Multiomics technologies: role in disease biomarker discoveries and therapeutics. *Brief Funct Genomics* 2023;**22**:76–96. <https://doi.org/10.1093/bfpg/elac017>
 13. Kwon YW, Jo H-S, Bae S *et al.* Application of proteomics in cancer: recent trends and approaches for biomarkers discovery. *Front Med* 2021;**8**. <https://doi.org/10.3389/fmed.2021.747333>
 14. Jan Y-H, Tan KT, Chen S-J *et al.* Comprehensive assessment of actionable genomic alterations in primary colorectal carcinoma using targeted next-generation sequencing. *Br J Cancer* 2022;**127**:1304–11. <https://doi.org/10.1038/s41416-022-01913-4>
 15. Teuwen L-A, Roets E, D'Hoore P *et al.* Comprehensive genomic profiling and therapeutic implications for patients with advanced cancers: The experience of an academic hospital. *Diagnostics* 2023;**13**:1619. <https://doi.org/10.3390/diagnostics13091619>
 16. Mubarak G, Zahir FR. Recent major transcriptomics and epitranscriptomics contributions toward personalized and precision medicine. *J Pers Med* 2022;**12**:199. <https://doi.org/10.3390/jpm12020199>
 17. Dindhoria K, Monga I, Thind AS. Computational approaches and challenges for identification and annotation of non-coding RNAs using RNA-Seq. *Funct Integr Genomics* 2022;**22**:1105–12. <https://doi.org/10.1007/s10142-022-00915-y>
 18. Han N, Liu Z. Targeting alternative splicing in cancer immunotherapy. *Front Cell Dev Biol* 2023;**11**. <https://doi.org/10.3389/fcell.2023.1232146>
 19. Stanley RF, Abdel-Wahab O. Dysregulation and therapeutic targeting of RNA splicing in cancer. *Nat Cancer* 2022;**3**:536–46. <https://doi.org/10.1038/s43018-022-00384-z>
 20. Rivero-Hinojosa S, Grant M, Panigrahi A *et al.* Proteogenomic discovery of neoantigens facilitates personalized multi-antigen targeted T cell immunotherapy for brain tumors. *Nat Commun* 2021;**12**:6689. <https://doi.org/10.1038/s41467-021-26936-y>
 21. Wenk D, Zuo C, Kislinger T *et al.* Recent developments in mass-spectrometry-based targeted proteomics of clinical cancer biomarkers. *Clin Proteomics* 2024;**21**:6. <https://doi.org/10.1186/s12014-024-09452-1>
 22. Meissner F, Geddes-McAlister J, Mann M *et al.* The emerging role of mass spectrometry-based proteomics in drug discovery. *Nat Rev Drug Discov* 2022;**21**:637–54. <https://doi.org/10.1038/s41573-022-00409-3>
 23. Galal A, Talal M, Moustafa A. Applications of machine learning in metabolomics: disease modeling and classification. *Front Genet* 2022;**13**. <https://doi.org/10.3389/fgene.2022.1017340>
 24. Bonanomi M, Salmistraro N, Fisco G *et al.* Transcriptomics and metabolomics integration reveals redox-dependent metabolic rewiring in breast cancer cells. *Cancer* 2021;**13**:5058. <https://doi.org/10.3390/cancers13205058>
 25. Wei Q, Qian Y, Yu J *et al.* Metabolic rewiring in the promotion of cancer metastasis: mechanisms and therapeutic implications. *Oncogene* 2020;**39**:6139–56. <https://doi.org/10.1038/s41388-020-01432-7>
 26. Davalos V, Esteller M. Cancer epigenetics in clinical practice. *CA Cancer J Clin* 2023;**73**:376–424. <https://doi.org/10.3322/caac.21765>
 27. Aitmanaitė L, Širmonaitis K, Russo G. Microbiomes, their function, and cancer: how Metatranscriptomics can close the knowledge gap. *Int J Mol Sci* 2023;**24**:13786. <https://doi.org/10.3390/ijms241813786>
 28. Yang C-C, Chen C-Y, Kuo Y-T *et al.* Radiomics for the prediction of response to Antifibrotic treatment in patients with idiopathic pulmonary fibrosis: a pilot study. *Diagnostics* 2022;**12**:1002. <https://doi.org/10.3390/diagnostics12041002>
 29. Cetin I, Raisi-Estabragh Z, Petersen SE *et al.* Radiomics signatures of cardiovascular risk factors in cardiac MRI: results from the UK biobank. *Front Cardiovasc Med* 2020;**7**. <https://doi.org/10.3389/fcvm.2020.591368>
 30. Xiao H, He X, Zhou W *et al.* The application of radiomics in the diagnosis and evaluation of cognitive impairment related to neurological diseases. *Front Neurosci* 2025;**19**:19. <https://doi.org/10.3389/fnins.2025.1591605>
 31. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* 2007;**8**:118–27. <https://doi.org/10.1093/biostatistics/kxj037>
 32. Horng H, Singh A, Yousefi B *et al.* Improved generalized ComBat methods for harmonization of radiomic features. *Sci Rep* 2022;**12**:19009. <https://doi.org/10.1038/s41598-022-23328-0>
 33. Horng H, Singh A, Yousefi B *et al.* Generalized ComBat harmonization methods for radiomic features with multi-modal distributions and multiple batch effects. *Sci Rep* 2022;**12**:4493. <https://doi.org/10.1038/s41598-022-08412-9>
 34. Zhang Y, Parmigiani G, Johnson WE. ComBat-seq: batch effect adjustment for RNA-seq count data. *NAR Genom Bioinform* 2020;**2**. <https://doi.org/10.1093/nargab/lqaa078>
 35. Korsunsky I, Millard N, Fan J *et al.* Fast, sensitive and accurate integration of single-cell data with harmony. *Nat Methods* 2019;**16**:1289–96. <https://doi.org/10.1038/s41592-019-0619-0>
 36. Gayoso A, Lopez R, Xing G *et al.* A python library for probabilistic analysis of single-cell omics data. *Nat Biotechnol* 2022;**40**:163–6. <https://doi.org/10.1038/s41587-021-01206-w>
 37. Hie BL, Kim S, Rando TA *et al.* Scanorama: integrating large and diverse single-cell transcriptomic datasets. *Nat Protoc* 2024;**19**:2283–97. <https://doi.org/10.1038/s41596-024-00991-3>

38. Polański K, Young MD, Miao Z *et al.* BBKNN: fast batch alignment of single cell transcriptomes. *Bioinformatics* 2020;**36**:964–5. <https://doi.org/10.1093/bioinformatics/btz625>
39. The M, Käll L. Integrated identification and quantification error probabilities for shotgun proteomics. *Mol Cell Proteomics* 2019;**18**:561–70. <https://doi.org/10.1074/mcp.RA118.001018>
40. Kohler D, Staniak M, Yu F *et al.* An MSstats workflow for detecting differentially abundant proteins in large-scale data-independent acquisition mass spectrometry experiments with FragPipe processing. *Nat Protoc* 2024;**19**:2915–38. <https://doi.org/10.1038/s41596-024-01000-3>
41. Giera M, Aisporna A, Uritboonthai W *et al.* XCMS-METLIN: data-driven metabolite, lipid, and chemical analysis. *Mol Syst Biol* 2024;**20**:1153–5. <https://doi.org/10.1038/s44320-024-00063-4>
42. Schmid R, Heuckeroth S, Korf A *et al.* Integrative analysis of multimodal mass spectrometry data in MZmine 3. *Nat Biotechnol* 2023;**41**:447–9. <https://doi.org/10.1038/s41587-023-01690-2>
43. Bolyen E, Rideout JR, Dillon MR *et al.* Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol* 2019;**37**:852–7. <https://doi.org/10.1038/s41587-019-0209-9>
44. Kuras M, Betancourt LH, Hong R *et al.* Proteogenomic profiling of treatment-naïve metastatic malignant melanoma. *Cancer* 2025;**17**:832. <https://doi.org/10.3390/cancers17050832>
45. Argelaguet R, Arnol D, Bredikhin D *et al.* MOFA+: a statistical framework for comprehensive integration of multi-modal single-cell data. *Genome Biol* 2020;**21**:111. <https://doi.org/10.1186/s13059-020-02015-1>
46. Jiangzhou H, Zhang H, Sun R *et al.* Integrative omics analysis reveals effective stratification and potential prognosis markers of pan-gastrointestinal cancers. *iScience* 2021;**24**:102824. <https://doi.org/10.1016/j.isci.2021.102824>
47. Mo Q, Shen R, Guo C *et al.* A fully Bayesian latent variable model for integrative clustering analysis of multi-type omics data. *Biostatistics* 2018;**19**:71–86. <https://doi.org/10.1093/biostatistics/kxx017>
48. Singh A, Shannon CP, Gautier B *et al.* DIABLO: an integrative approach for identifying key molecular drivers from multi-omics assays. *Bioinformatics* 2019;**35**:3055–62. <https://doi.org/10.1093/bioinformatics/bty1054>
49. Zhang X, Xing Y, Sun K *et al.* OmiEmbed: a unified multi-task deep learning framework for multi-omics data. *Cancer* 2021;**13**. <https://doi.org/10.3390/cancers13123047>
50. Peng T, Pourfarhangi KE, Tan K. Abstract PO-026: GLUER: integrative analysis of multi-omics data at single-cell resolution. *Cancer Res* 2020;**80**:PO-026. <https://doi.org/10.1158/1538-7445.TUMHET2020-PO-026>
51. Hai Y, Ma J, Yang K *et al.* Bayesian linear mixed model with multiple random effects for prediction analysis on high-dimensional multi-omics data. *Bioinformatics* 2023;**39**. <https://doi.org/10.1093/bioinformatics/btad647>
52. Spooner A, Moridani MK, Toplis B *et al.* Benchmarking ensemble machine learning algorithms for multi-class, multi-omics data integration in clinical outcome prediction. *Brief Bioinform* 2025;**26**. <https://doi.org/10.1093/bib/bbaf116>
53. Zhang G, Ma C, Yan C *et al.* MSFN: a multi-omics stacked fusion network for breast cancer survival prediction. *Front Genet* 2024;**15**. <https://doi.org/10.3389/fgene.2024.1378809>
54. Sharifi-Noghabi H, Zolotareva O, Collins CC *et al.* MOLI: multi-omics late integration with deep neural networks for drug response prediction. *Bioinformatics* 2019;**35**:i501–9. <https://doi.org/10.1093/bioinformatics/btz318>
55. Afgan E, Baker D, Batut B *et al.* The galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res* 2018;**46**:W537–44. <https://doi.org/10.1093/nar/gky379>
56. Ewels PA, Peltzer A, Fillinger S *et al.* The nf-core framework for community-curated bioinformatics pipelines. *Nat Biotechnol* 2020;**38**:276–8. <https://doi.org/10.1038/s41587-020-0439-x>
57. Köster J, Rahmann S. Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics* 2012;**28**:2520–2. <https://doi.org/10.1093/bioinformatics/bts480>
58. Di Tommaso P, Chatzou M, Floden EW *et al.* Nextflow enables reproducible computational workflows. *Nat Biotechnol* 2017;**35**:316–9. <https://doi.org/10.1038/nbt.3820>
59. Vasaikar SV, Straub P, Wang J *et al.* LinkedOmics: Analyzing multi-omics data within and across 32 cancer types. *Nucleic Acids Res* 2018;**46**:D956–63. <https://doi.org/10.1093/nar/gkx1090>
60. Wang B, Mezlini AM, Demir F *et al.* Similarity network fusion for aggregating data types on a genomic scale. *Nat Methods* 2014;**11**:333–7. <https://doi.org/10.1038/nmeth.2810>
61. Hie B, Bryson B, Berger B. Efficient integration of heterogeneous single-cell transcriptomes using Scanorama. *Nat Biotechnol* 2019;**37**:685–91. <https://doi.org/10.1038/s41587-019-0113-3>
62. Gayoso A, Steier Z, Lopez R *et al.* Joint probabilistic modeling of single-cell multi-omic data with totalVI. *Nat Methods* 2021;**18**:272–82. <https://doi.org/10.1038/s41592-020-01050-x>
63. Ashuach T, Gabitto MI, Koodli RV *et al.* MultiVI: deep generative model for the integration of multimodal data. *Nat Methods* 2023;**20**:1222–31. <https://doi.org/10.1038/s41592-023-01909-9>
64. Benkirane H, Pradat Y, Michiels S *et al.* CustOmics: a versatile deep-learning based strategy for multi-omics integration. *PLoS Comput Biol* 2023;**19**:e1010921. <https://doi.org/10.1371/journal.pcbi.1010921>
65. Tran N, Gao J. OpenOmics: a bioinformatics API to integrate multi-omics datasets and interface with public databases. *J Open Source Softw* 2021;**6**:3249. <https://doi.org/10.21105/joss.03249>
66. Hao Y, Stuart T, Kowalski MH *et al.* Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat Biotechnol* 2024;**42**:293–304. <https://doi.org/10.1038/s41587-023-01767-y>
67. Aaltonen LA, Abascal F, Abeshouse A *et al.* Pan-cancer analysis of whole genomes. *Nature* 2020;**578**:82–93. <https://doi.org/10.1038/s41586-020-1969-6>
68. International Cancer Genome Consortium, Hudson TJ, Anderson W *et al.* International network of cancer genome projects. *Nature* 2010;**464**:993–8. <https://doi.org/10.1038/nature08987>
69. Barrett T, Wilhite SE, Ledoux P *et al.* NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res* 2012;**41**:D991–5. <https://doi.org/10.1093/nar/gks1193>
70. Brazma A, Parkinson H, Sarkans U *et al.* ArrayExpress—a public repository for microarray gene expression data at the EBI. *Nucleic Acids Res* 2003;**31**:68–71. <https://doi.org/10.1093/nar/gkg091>
71. Yurekten O, Payne T, Tejera N *et al.* MetaboLights: open data repository for metabolomics. *Nucleic Acids Res* 2024;**52**:D640–6. <https://doi.org/10.1093/nar/gkad1045>
72. Perez-Riverol Y, Bandla C, Kundu DJ *et al.* The PRIDE database at 20 years: 2025 update. *Nucleic Acids Res* 2025;**53**:D543–53. <https://doi.org/10.1093/nar/gkae1011>
73. Regev A, Teichmann SA, Lander ES *et al.* The human cell atlas. *elife* 2017;**6**. <https://doi.org/10.7554/eLife.27041>
74. Madrigal P, Thanki AS, Fexova S *et al.* Expression atlas in 2026: enabling FAIR and open expression data through community

- collaboration and integration. *Nucleic Acids Res* 2026;**54**:D147–57. <https://doi.org/10.1093/nar/gkaf1238>
75. Tseng CY, Salguero JA, Breidenbach JD *et al.* Evaluation of normalization strategies for mass spectrometry-based multi-omics datasets. *Metabolomics* 2025;**21**:98. <https://doi.org/10.1007/s11306-025-02297-1>
 76. Skubleny D, Ghosh S, Spratlin J *et al.* Feature-specific quantile normalization and feature-specific mean–variance normalization deliver robust bi-directional classification and feature selection performance between microarray and RNAseq data. *BMC Bioinformatics* 2024;**25**:136. <https://doi.org/10.1186/s12859-024-05759-w>
 77. Yu Y, Zhang N, Mai Y *et al.* Correcting batch effects in large-scale multiomics studies using a reference-material-based ratio method. *Genome Biol* 2023;**24**:201. <https://doi.org/10.1186/s13059-023-03047-z>
 78. Behdenna A, Colange M, Haziza J *et al.* pyComBat, a python tool for batch effects correction in high-throughput molecular data using empirical Bayes methods. *BMC Bioinformatics* 2023;**24**:459. <https://doi.org/10.1186/s12859-023-05578-5>
 79. Lopez R, Regier J, Cole MB *et al.* Deep generative modeling for single-cell transcriptomics. *Nat Methods* 2018;**15**:1053–8. <https://doi.org/10.1038/s41592-018-0229-2>
 80. He Z, Hu S, Chen Y *et al.* Mosaic integration and knowledge transfer of single-cell multimodal data with MIDAS. *Nat Biotechnol* 2024;**42**:1594–605. <https://doi.org/10.1038/s41587-023-02040-y>
 81. Qin L, Zhang G, Zhang S *et al.* Deep batch integration and Denoise of single-cell RNA-Seq data. *Adv Sci* 2024;**11**. <https://doi.org/10.1002/advs.202308934>
 82. Flores JE, Claborne DM, Weller ZD *et al.* Missing data in multi-omics integration: recent advances through artificial intelligence. *Front Artif Intell* 2023;**6**. <https://doi.org/10.3389/frai.2023.1098308>
 83. Xu J, Wu P, Chen Y *et al.* A hierarchical integration deep flexible neural forest framework for cancer subtype classification by integrating multi-omics data. *BMC Bioinformatics* 2019;**20**:527. <https://doi.org/10.1186/s12859-019-3116-7>
 84. Xie G, Dong C, Kong Y *et al.* Group lasso regularized deep learning for cancer prognosis from multi-omics and clinical features. *Genes* 2019;**10**:240. <https://doi.org/10.3390/genes10030240>
 85. Qin H, Niu T, Zhao J. Identifying multi-omics causers and causal pathways for complex traits. *Front Genet* 2019;**10**. <https://doi.org/10.3389/fgene.2019.00110>
 86. Babur Ö, Luna A, Korkut A *et al.* Causal interactions from proteomic profiles: molecular data meet pathway knowledge. *Patterns* 2021;**2**:100257. <https://doi.org/10.1016/j.patter.2021.100257>
 87. Ing A, Andrades A, Cosenza MR *et al.* Integrating multimodal cancer data using deep latent variable path modelling. *Nat Mach Intell* 2025;**7**:1053–75. <https://doi.org/10.1038/s42256-025-01052-4>
 88. Suter P, Dazert E, Kuipers J *et al.* Multi-omics subtyping of hepatocellular carcinoma patients using a Bayesian network mixture model. *PLoS Comput Biol* 2022;**18**:e1009767. <https://doi.org/10.1371/journal.pcbi.1009767>
 89. Li X, Ma J, Leng L *et al.* MoGCN: a multi-omics integration method based on graph convolutional network for cancer subtype analysis. *Front Genet* 2022;**13**. <https://doi.org/10.3389/fgene.2022.806842>
 90. Liu P, Ji L, Ye F *et al.* GraphLSurv: a scalable survival prediction network with adaptive and sparse structure learning for histopathological whole-slide images. *Comput Methods Prog Biomed* 2023;**231**:107433. <https://doi.org/10.1016/j.cmpb.2023.107433>
 91. Tanvir RB, Islam MM, Sobhan M *et al.* MOGAT: a multi-omics integration framework using graph attention networks for cancer subtype prediction. *Int J Mol Sci* 2024;**25**:2788. <https://doi.org/10.3390/ijms25052788>
 92. Wang T, Shao W, Huang Z *et al.* MOGONET integrates multi-omics data using graph convolutional networks allowing patient classification and biomarker identification. *Nat Commun* 2021;**12**:3445. <https://doi.org/10.1038/s41467-021-23774-w>
 93. Zhang Y, Xiong S, Wang Z *et al.* Local augmented graph neural network for multi-omics cancer prognosis prediction and analysis. *Methods* 2023;**213**:1–9. <https://doi.org/10.1016/j.ymeth.2023.02.011>
 94. Sheng N, Wang Y, Huang L *et al.* Multi-task prediction-based graph contrastive learning for inferring the relationship among lncRNAs, miRNAs and diseases. *Brief Bioinform* 2023;**24**. <https://doi.org/10.1093/bib/bbad276>
 95. Wu Y, Xu J, Li J *et al.* Deep graph convolutional network-based multi-omics integration for cancer driver gene identification. *Brief Bioinform* 2025;**26**. <https://doi.org/10.1093/bib/bbaf364>
 96. Yi E, Cui J, Wang H *et al.* Multiomics Mendelian randomization identifies serpin family G member 1 as a chronic obstructive pulmonary disease modulator. *Signal Transduct Target Ther* 2026;**11**:34. <https://doi.org/10.1038/s41392-025-02547-7>
 97. Vasaikar S, Huang C, Wang X *et al.* Proteogenomic analysis of human colon cancer reveals new therapeutic opportunities. *Cell* 2019;**177**:1035–1049.e19. <https://doi.org/10.1016/j.cell.2019.03.030>
 98. Pang W, Wang Y, Lu X *et al.* Integrated spatial and single cell transcriptomics identifies PRKDC as a dual prognostic biomarker and therapeutic target in hepatocellular carcinoma. *Sci Rep* 2025;**15**:14834. <https://doi.org/10.1038/s41598-025-98866-4>
 99. Jing S, Liu D, Feng N *et al.* Spatial multiomics reveals a subpopulation of fibroblasts associated with cancer stemness in human hepatocellular carcinoma. *Genome Med* 2024;**16**:98. <https://doi.org/10.1186/s13073-024-01367-8>
 100. Wei C, Sun W, Shen K *et al.* Delineating the early dissemination mechanisms of acral melanoma by integrating single-cell and spatial transcriptomic analyses. *Nat Commun* 2023;**14**:8119. <https://doi.org/10.1038/s41467-023-43980-y>
 101. Andrews MC, Oba J, Wu C-J *et al.* Multi-modal molecular programs regulate melanoma cell state. *Nat Commun* 2022;**13**:4000. <https://doi.org/10.1038/s41467-022-31510-1>
 102. In GK, Ribeiro JR, Yin J *et al.* Multi-omic profiling reveals discrepant immunogenic properties and a unique tumor microenvironment among melanoma brain metastases. *NPJ Precis Oncol* 2023;**7**:120. <https://doi.org/10.1038/s41698-023-00471-z>
 103. Raj-Kumar P-K, Lin X, Liu T *et al.* Proteogenomic characterization of difficult-to-treat breast cancer with tumor cells enriched through laser microdissection. *Breast Cancer Res* 2024;**26**:76. <https://doi.org/10.1186/s13058-024-01835-4>
 104. Shen X, Wang C, Zhou X *et al.* Nonlinear dynamics of multi-omics profiles during human ageing. *Nat Aging* 2024;**4**:1619–34. <https://doi.org/10.1038/s43587-024-00692-2>
 105. Liu G-H, Bao Y, Qu J *et al.* Aging atlas: a multi-omics database for aging biology. *Nucleic Acids Res* 2021;**49**:D825–30. <https://doi.org/10.1093/nar/gkaa894>
 106. Li Y, Li C, Zhou Q *et al.* Multiomics and cellular senescence profiling of aging human skeletal muscle uncovers Maraviroc as a senotherapeutic approach for sarcopenia. *Nature Communications* 2025;**16**:6207. <https://doi.org/10.1038/s41467-025-61403-y>
 107. Zhang F, Li R, Liu Y *et al.* Integrative cross-tissue analysis unveils complement-immunoglobulin augmentation and dysbiosis-related fatty acid metabolic remodeling during mammalian aging. *iMeta* 2025;**4**:e70027. <https://doi.org/10.1002/imt2.70027>

108. Wang Y, Li R, Tong R *et al.* Integrating single-cell RNA and T cell/B cell receptor sequencing with mass cytometry reveals dynamic trajectories of human peripheral immune cells from birth to old age. *Nat Immunol* 2025;**26**:308–22. <https://doi.org/10.1038/s41590-024-02059-6>
109. Mavromatis LA, Rosoff DB, Bell AS *et al.* Multi-omic underpinnings of epigenetic aging and human longevity. *Nat Commun* 2023;**14**:2236. <https://doi.org/10.1038/s41467-023-37729-w>
110. Tang J, Yue L, Xu Y *et al.* Longitudinal serum proteome mapping reveals biomarkers for healthy ageing and related cardiometabolic diseases. *Nat Metab* 2025;**7**:166–81. <https://doi.org/10.1038/s42255-024-01185-7>
111. Wang X, Fan D, Yang Y *et al.* Integrative multi-omics approaches to explore immune cell functions. *Challenges and opportunities iScience* 2023;**26**:106359. <https://doi.org/10.1016/j.isci.2023.106359>
112. Chen C, Wang J, Pan D *et al.* Applications of multi-omics analysis in human diseases. *MedComm* 2023;**4**:e315. <https://doi.org/10.1002/mco2.315>
113. Smith MH, Gao VR, Periyakoil PK *et al.* Drivers of heterogeneity in synovial fibroblasts in rheumatoid arthritis. *Nat Immunol* 2023;**24**:1200–10. <https://doi.org/10.1038/s41590-023-01527-9>
114. Mennillo E, Kim YJ, Lee G *et al.* Single-cell and spatial multi-omics highlight effects of anti-integrin therapy across cellular compartments in ulcerative colitis. *Nat Commun* 2024;**15**:1493. <https://doi.org/10.1038/s41467-024-45665-6>
115. Ambikan AT, Yang H, Krishnan S *et al.* Multi-omics personalized network analyses highlight progressive disruption of central metabolism associated with COVID-19 severity. *Cell Syst* 2022;**13**:665–681.e4. <https://doi.org/10.1016/j.cels.2022.06.006>
116. Jiang Y, Liu Q, Stridh P *et al.* Multiomics integration prioritizes potential drug targets for multiple sclerosis. *Proc Natl Acad Sci* 2025;**122**. <https://doi.org/10.1073/pnas.2425537122>
117. Wimmers F, Pulendran B. Emerging technologies for systems vaccinology — multi-omics integration and single-cell (epi)genomic profiling. *Curr Opin Immunol* 2020;**65**:57–64. <https://doi.org/10.1016/j.coi.2020.05.001>
118. Hegde NR, Gauthami S, Sampath Kumar HM *et al.* The use of databases, data mining and immunoinformatics in vaccinology: where are we? *Expert Opin Drug Discov* 2018;**13**:117–30. <https://doi.org/10.1080/17460441.2018.1413088>
119. Howard LM, Hoek KL, Goll JB *et al.* Cell-based systems biology analysis of human AS03-Adjuvanted H5N1 avian influenza vaccine responses: a phase I randomized controlled trial. *PLoS One* 2017;**12**:e0167488. <https://doi.org/10.1371/journal.pone.0167488>
120. Hagan T, Gerritsen B, Tomalin LE *et al.* Transcriptional atlas of the human immune response to 13 vaccines reveals a common predictor of vaccine-induced antibody responses. *Nat Immunol* 2022;**23**:1788–98. <https://doi.org/10.1038/s41590-022-01328-6>
121. Shinde P, Soldevila F, Reyna J *et al.* A multi-omics systems vaccinology resource to develop and test computational models of immunity. *Cell Rep Methods* 2024;**4**:100731. <https://doi.org/10.1016/j.crmeth.2024.100731>
122. Hughes A, Parast L, Thiébaud R *et al.* RISE: two-stage rank-based identification of high-dimensional surrogate markers applied to vaccinology. *Stat Med* 2025;**44**:e70241. <https://doi.org/10.1002/sim.70241>
123. Chen W, He Y, Zhou G *et al.* Multiomics characterization of pyroptosis in the tumor microenvironment and therapeutic relevance in metastatic melanoma. *BMC Med* 2024;**22**:24. <https://doi.org/10.1186/s12916-023-03175-0>
124. Xu Y, Lau P, Chen X *et al.* Integrated multiomics revealed adenosine signaling predict immunotherapy response and regulate tumor ecosystem of melanoma. *Hum Genomics* 2024;**18**:101. <https://doi.org/10.1186/s40246-024-00651-3>
125. Fuentealba M, Kiproff D, Schneider K *et al.* Multi-omics analysis reveals biomarkers that contribute to biological age rejuvenation in response to single-blinded randomized placebo-controlled therapeutic plasma exchange. *Aging Cell* 2025;**24**:e70103. <https://doi.org/10.1111/acer.70103>
126. Rambabu N, Mathew MJ, Kaveri SV *et al.* Boolean analysis of the transcriptomic data to identify novel biomarkers of IVIG response. *Autoimmun Rev* 2021;**20**:102850. <https://doi.org/10.1016/j.autrev.2021.102850>
127. Han B, Tang D, Lv X *et al.* Integrated multi-omics reveal gut microbiota-mediated bile acid metabolism alteration regulating immunotherapy responses to anti- $\alpha 4\beta 7$ -integrin in Crohn's disease. *Gut Microbes* 2024;**16**. <https://doi.org/10.1080/19490976.2024.2310894>
128. Diao P, Dai Y, Wang A *et al.* Integrative multiomics analyses identify molecular subtypes of head and neck squamous cell carcinoma with distinct therapeutic vulnerabilities. *Cancer Res* 2024;**84**:3101–17. <https://doi.org/10.1158/0008-5472.CAN-23-3594>
129. Song KJ, Choi S, Kim K *et al.* Proteogenomic analysis reveals non-small cell lung cancer subtypes predicting chromosome instability, and tumor microenvironment. *Nat Commun* 2024;**15**:10164. <https://doi.org/10.1038/s41467-024-54434-4>
130. Li J, Zhang X, Yang L *et al.* Targeted proteomics-determined multi-biomarker profiles developed classifier for prognosis and immunotherapy responses of advanced cervical cancer. *Front Immunol* 2024;**15**. <https://doi.org/10.3389/fimmu.2024.1391524>
131. Overmyer KA, Shishkova E, Miller IJ *et al.* Large-scale multi-omic analysis of COVID-19 severity. *Cell Syst* 2021;**12**:23–40.e7. <https://doi.org/10.1016/j.cels.2020.10.003>
132. Jakka AL, Chacko RM, Vasam M *et al.* From genomics to clinic: The transformative impact of AI in pharmacogenomics and personalized medicine. *Pharmacogenomics* 2025;**26**:573–85. <https://doi.org/10.1080/14622416.2025.2591596>
133. Kumar-Sinha C, Chinnaiyan AM. Precision oncology in the age of integrative genomics. *Nat Biotechnol* 2018;**36**:46–60. <https://doi.org/10.1038/nbt.4017>
134. Kang W, Qiu X, Luo Y *et al.* Application of radiomics-based multiomics combinations in the tumor microenvironment and cancer prognosis. *J Transl Med* 2023;**21**:598. <https://doi.org/10.1186/s12967-023-04437-4>
135. Lin A, Ye J, Qi C *et al.* Bridging artificial intelligence and biological sciences: a comprehensive review of large language models in bioinformatics. *Brief Bioinform* 2025;**26**. <https://doi.org/10.1093/bib/bbaf357>