

Review

# Artificial Intelligence-Driven Precision Medicine: Multi-Omics and Spatial Multi-Omics Approaches in Diffuse Large B-Cell Lymphoma (DLBCL)

Yanping Shao<sup>1</sup>, Xiuyan Lv<sup>2</sup>, Shuangwei Ying<sup>2</sup>, Qunyi Guo<sup>2,\*</sup>

<sup>1</sup>Department of Hematology, The Second Affiliated Hospital, Zhejiang University School of Medicine, 310009 Hangzhou, Zhejiang, China

<sup>2</sup>Department of Hematology, Taizhou Hospital of Zhejiang Province Affiliated to Wenzhou Medical University, 317000 Taizhou, Zhejiang, China

\*Correspondence: [guoqy@enzemed.com](mailto:guoqy@enzemed.com) (Qunyi Guo)

Academic Editor: Xudong Huang

Submitted: 16 May 2024 Revised: 17 June 2024 Accepted: 25 June 2024 Published: 28 November 2024

## Abstract

In this comprehensive review, we delve into the transformative role of artificial intelligence (AI) in refining the application of multi-omics and spatial multi-omics within the realm of diffuse large B-cell lymphoma (DLBCL) research. We scrutinized the current landscape of multi-omics and spatial multi-omics technologies, accentuating their combined potential with AI to provide unparalleled insights into the molecular intricacies and spatial heterogeneity inherent to DLBCL. Despite current progress, we acknowledge the hurdles that impede the full utilization of these technologies, such as the integration and sophisticated analysis of complex datasets, the necessity for standardized protocols, the reproducibility of findings, and the interpretation of their biological significance. We proceeded to pinpoint crucial research voids and advocated for a trajectory that incorporates the development of advanced AI-driven data integration and analytical frameworks. The evolution of these technologies is crucial for enhancing resolution and depth in multi-omics studies. We also emphasized the importance of amassing extensive, meticulously annotated multi-omics datasets and fostering translational research efforts to connect laboratory discoveries with clinical applications seamlessly. Our review concluded that the synergistic integration of multi-omics, spatial multi-omics, and AI holds immense promise for propelling precision medicine forward in DLBCL. By surmounting the present challenges and steering towards the outlined futuristic pathways, we can harness these potent investigative tools to decipher the molecular and spatial conundrums of DLBCL. This will pave the way for refined diagnostic precision, nuanced risk stratification, and individualized therapeutic regimens, ushering in a new era of patient-centric oncology care.

**Keywords:** diffuse large B-cell lymphoma; spatial multi-omics; artificial intelligence; precision medicine

## 1. Introduction

Diffuse large B-cell lymphoma (DLBCL) is the most common form of non-Hodgkin lymphoma (NHL), accounting for approximately 30% of new NHL cases worldwide, with a higher prevalence in developed countries [1]. It is identified by the uncontrolled growth of large-sized, atypical B cells widely scattered in lymph nodes [2] (Fig. 1). DLBCL usually appears as a rapidly growing lymph node mass accompanied by symptoms of systemic illness including fever, night sweats, and weight loss [3]. The disease typically strikes older people, diagnosed at a median age of approximately 70 [4] (Fig. 1).

Despite the availability of aggressive treatment regimens, including the combination of rituximab, cyclophosphamide, doxorubicin, vincristine, and prednisone (R-CHOP), the clinical outcomes for DLBCL patients vary substantially. Approximately 40% of patients do not respond to initial therapy or relapse after the initial response, leading to a poor prognosis [5]. The clinical heterogeneity of DLBCL calls for precision medicine approaches to better stratify patients and guide therapy [6] (Fig. 1).

The broad spectrum of clinical and biological heterogeneity of DLBCL complicates the task of optimizing the treatment for individual patients. The current treatment strategies, mainly the R-CHOP regimen, do not consider the intrinsic variability in tumor biology and patient characteristics, the therapeutic response rates are variable, and the clinical outcomes are diverse [7]. Potential guidance for personalized treatment selection, guided by molecular profiling of individual tumors, is one of the promises of precision medicine [6] (Fig. 1). The recent development of multi-omics and spatial multi-omics approaches offers the possibility to understand the molecular complexity of DLBCL better and design personalized treatment strategies [8].

Multi-omics involves genomics, transcriptomics, proteomics, and metabolomics data to analyze biological samples deeply. In DLBCL, multi-omics analysis has detected molecular heterogeneity, identified subtypes with distinct features, and improved pathogenesis [9]. Genomic analysis has revealed recurrent mutations and chromosomal alterations, which shed light on disease drivers and potential therapeutic targets [10–12]. Transcriptomic profiling has refined the classification of DLBCL by identi-



lying subtypes with different origins and therapeutic responses [6,13–15]. Proteomic and metabolomic studies have provided insights into protein expression changes and metabolic alterations relevant to DLBCL pathophysiology [16–18]. Spatial multi-omics is a technique that integrates molecular data with tissue architecture, which gives “insights” relating to tumor microenvironment, immune interactions, and structural organization. Tissue structure is used primarily in the study of tumor microenvironment, tumor response to therapy, and the clinical management of cancers [19–22]. These methods are avenues for precision medicine in DLBCL.

Artificial intelligence (AI) integration into multi-omics research is necessary to revolutionize our comprehension and management of complex diseases like DLBCL. Moreover, it has been noted that AI, in the form of machine learning (ML) algorithms, can be utilized to examine vast and intricate datasets beyond the human brain’s capacity for discerning patterns and associations [23]. Deep learning, a subcategory of ML, unifies different -omics data types to characterize DLBCL pathogenesis and progression [24]. Disease outcome prognoses by AI-based models are useful for patient stratification and identifying therapeutic responses [25]. Besides, natural language processing (NLP) pulls structured information from unstructured data sources that strengthen multi-omics analysis [26]. Nevertheless, challenges concerning the standardization of data, interpretability of results, and generalizability to other scenarios still hinder the use of AI in this area [26]. Combining multi-omics with artificial intelligence (AI) could be a game changer for DLBCL research.

The article examines the status of multi-omics and spatially based multi-omics techniques driven by AI in the study of DLBCL. The aim is to critically evaluate how these integrative technologies improve our understanding of molecular and spatial heterogeneity characterizing DLBCL, outline difficulties faced, and suggest possible solutions. This review also investigates the synergy between AI and multi-omics for precision medicine in DLBCL. In conclusion, this manuscript summarizes the transformative potential of AI in improving precision medicine, focusing on diagnostic improvement, specifically early-stage treatment.

## 2. Understanding DLBCL: The Conventional Approach

### 2.1 Overview of the Histopathological and Molecular Characteristics of DLBCL

DLBCL, the predominant form of non-Hodgkin lymphoma, exhibits a vast heterogeneity that manifests in its histopathological presentation and molecular underpinnings [27]. Histologically, DLBCL is marked by the diffuse proliferation of large B lymphoid cells, which surpass the size of normal macrophage nuclei within lymph nodes. This proliferation is often accompanied by a high mitotic rate and

can include necrotic areas [28]. The disease is morphologically diverse, encompassing variants such as centroblastic, immunoblastic, T-cell/histiocyte-rich, and anaplastic types, each with distinct histological features [29].

Molecularly, DLBCL is stratified into subtypes based on the cell-of-origin (COO) classification, which includes germinal center B-cell-like (GCB), activated B-cell-like (ABC), and an unclassified category. These subtypes are discernible by their unique gene expression profiles and are closely linked to clinical outcomes. Patients with ABC-DLBCL tend to have a less favorable prognosis than those with GCB-DLBCL [13,30]. Recent genomic explorations have unveiled recurrent mutations, with *MYD88* and *CD79B* mutations frequently observed in the ABC subtype [31]. Advances in multi-omics analyses, integrating genetic, epigenetic, and transcriptomic data, have allowed for delineating distinct molecular subgroups, suggesting varied therapeutic vulnerabilities within this complex disease landscape [6].

The intricate histopathological and molecular attributes of DLBCL underscore the disease’s heterogeneity, which is critical for assessing prognosis and individualizing treatment modalities. Thus, it emphasizes the need for precision approaches in therapeutic intervention and management.

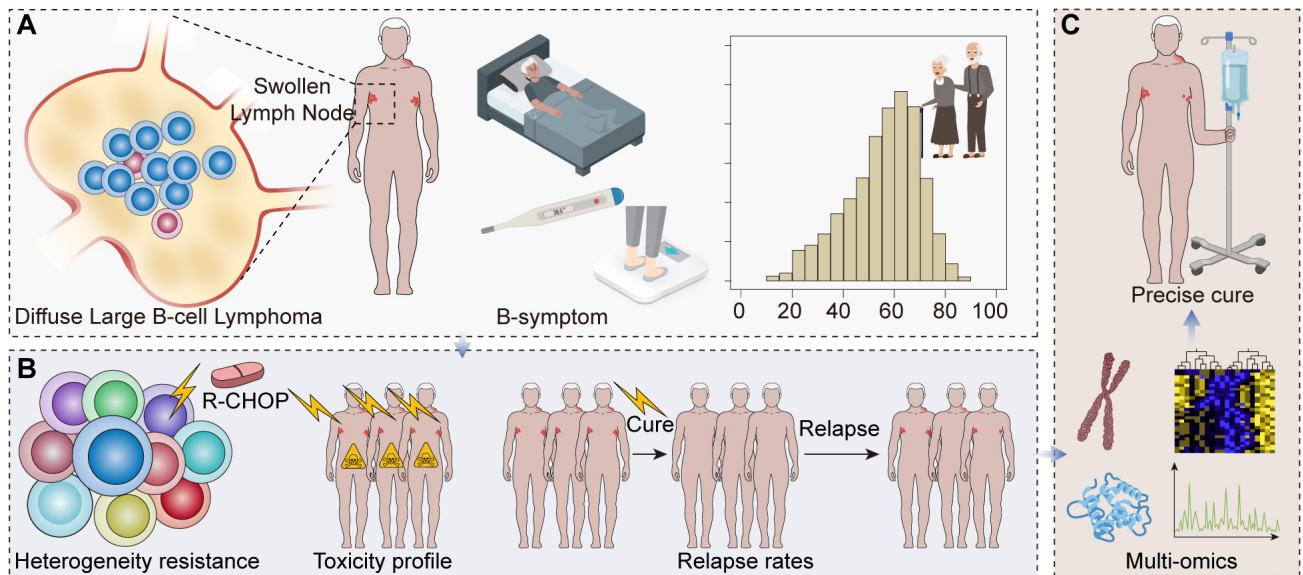
### 2.2 The Limitations of Traditional Diagnostic and Therapeutic Approaches

Traditionally, the diagnosis of DLBCL relies on histological examination of biopsy tissue, immunohistochemistry (IHC), and conventional cytogenetic analysis [29]. However, these methods have several limitations. Specifically, IHC often requires subjective interpretation and can lead to misclassification of the disease, especially in the distinction between the germinal center B-cell-like (GCB) and activated B-cell-like (ABC) subtypes [32]. Additionally, conventional cytogenetic techniques may miss cryptic or submicroscopic genetic alterations with prognostic or therapeutic implications [33].

In terms of therapy, R-CHOP (rituximab, cyclophosphamide, doxorubicin, vincristine, and prednisone) remains the standard first-line treatment for DLBCL [5]. However, approximately one-third of patients are refractory to R-CHOP or relapse after an initial response [34]. This reflects the heterogeneity of DLBCL, as different subtypes and genetic variants may respond differently to R-CHOP [35].

Furthermore, R-CHOP can have serious side effects, including myelosuppression, infection, cardiotoxicity, and secondary malignancies [36]. The treatment is particularly challenging in elderly patients, who often have comorbidities and may not tolerate aggressive chemotherapy [37].

In summary, traditional diagnostic and therapeutic approaches for DLBCL face significant limitations due to their inability to fully capture the heterogeneity of the disease and lack of personalization in treatment.



**Fig. 1. The role of precision medicine and multi-omics in understanding and treating diffuse large B-cell lymphoma (DLBCL).** (A) A graphical representation of DLBCL pathology with an illustration of diffuse large B-cell proliferation in a lymph node, coupled with a demographic chart highlighting the prevalence in older populations. (B) A chart showing the efficacy of the standard rituximab, cyclophosphamide, doxorubicin, vincristine, and prednisone (R-CHOP) treatment, displaying the split between responsive and non-responsive or relapsed patients. (C) A diagram emphasizing the heterogeneity of DLBCL and the shift towards personalized medicine illustrates the diversity of treatment responses and the move away from ‘one-size-fits-all’ therapies.

### 3. The Rise of Multi-Omics in DLBCL Research

#### 3.1 Definition and Importance of Multi-Omics

Multi-omics is an integrated research approach that combines data from multiple ‘omics’ fields, such as genomics, transcriptomics, proteomics, metabolomics, and epigenomics, to comprehensively understand complex biological systems [38].

The importance of multi-omics lies in its capacity to provide a holistic view of the molecular mechanisms underlying biological processes and diseases. By examining multiple layers of biological information, multi-omics allows for a more complete understanding of disease pathogenesis than single ‘omics’ studies [39].

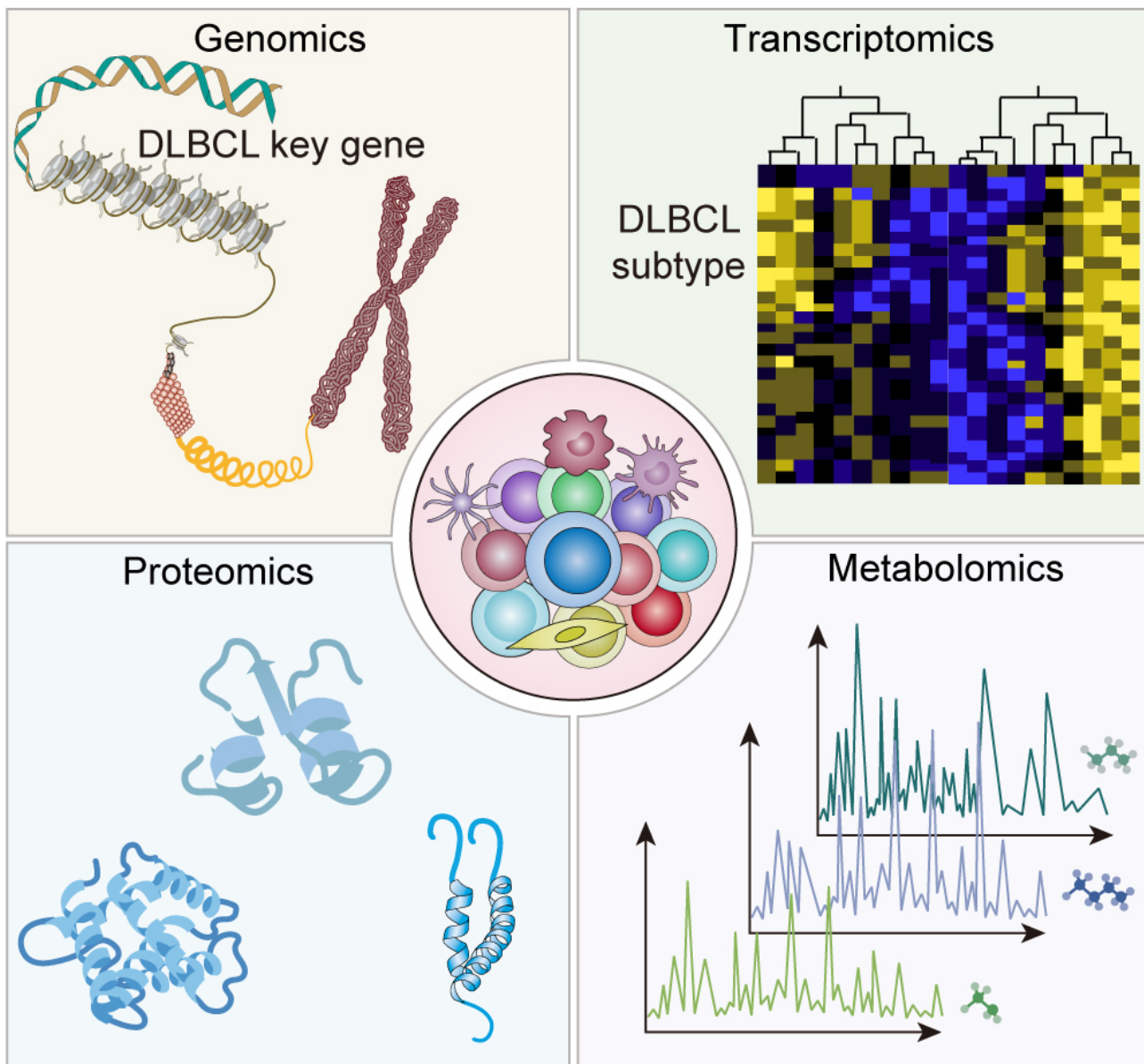
In the context of cancer research, including DLBCL, multi-omics approaches have the potential to reveal the complex interplay between genetics, epigenetics, and environmental factors that drive tumorigenesis and influence treatment response [40]. Integrating genomics and transcriptomics data can uncover the functional consequences of genetic alterations while combining proteomics and metabolomics data can shed light on the metabolic reprogramming in cancer cells [41,42].

Multi-omics profiling can inform the development of personalized medicine strategies by identifying molecular subtypes of diseases, novel therapeutic targets, and biomarkers for predicting treatment response and prognosis [38].

In summary, multi-omics represents a powerful tool for unraveling the complexity of biological systems and diseases, with promising applications in precision medicine.

#### 3.2 Multi-Omics Contribution to the Understanding of DLBCL

Genomics has been instrumental in unraveling recurrent genetic alterations that contribute to the pathogenesis of DLBCL and consequently improve our understanding. Genomic studies have identified multiple mutations within genes involved in crucial cellular pathways. Mutations affecting B-cell receptor (BCR) signaling pathways [43–45] have been reported in DLBCL, which is essential for B-cell development and activation. Further, genome-wide studies have shown various modifications of genes associated with chromatin remodeling as a process of modifying chromatin structure determining gene expression. In DLBCL, however, such are deregulated leading to abnormal gene expression patterns that can affect the disease progression [46,47]. Genomic studies have revealed numerous mutations within genes, such as those linked with programmed cell death protein 1 (PD-1) and programmed cell death ligand 1 (PD-L1). These mutations may hinder interaction between immune system cells and DLBCL cells, thus modulating response to immunotherapies [48]. Furthermore, genomic analysis has demonstrated that diverse molecular subtypes of DLBCL have varied genetic profiles and clinical outcomes [14]. The two major subtypes are the germinal center B-cell-like (GCB) and activated B-cell-like (ABC) types [49]. These



**Fig. 2. The power of multi-omics integration: unraveling the complexity of DLBCL.**

subtypes show differences in their gene expression profiles and molecular attributes, reflecting variations in their intrinsic biology or possible treatment responses. Consequently, comprehending the two major subtypes is crucial for improving the outcomes of patients as well as optimizing treatment choices. Genomics studies have identified such recurrent genetic alterations and molecular subtypes, which provides crucial insights into DLBCL pathogenesis and its underlying heterogeneity (Fig. 2). This information may help direct the development of personalized treatment strategies and targeted therapies for DLBCL patients.

**Transcriptomics:** Transcriptomics has played a role in understanding DLBCL by studying gene expression patterns using RNA sequencing. One key aspect of transcriptomics is the discovery of subtypes within DLBCL, which complements the classification based on features.

Through analyzing gene expression, a subset with a host response signature has been found, showing immune infiltration. This subset is linked to a prognosis, highlighting the significance of the response in determining DLBCL outcomes [15]. Transcriptomic investigations have also given a glimpse of the functional implications behind the recurrent genetic changes seen in DLBCL. Examining gene expression changes associated with a given mutation or genetic abnormality could help determine how the mutation alters cellular functions and signaling pathways [50]. In addition, transcriptomics has contributed to uncovering the dysregulated signaling pathways in DLBCL. This gives critical information about the molecular drivers of the disease and allows for the development of targeted therapeutics that can alter only those known pathways [31] (Fig. 2).

**Proteomics:** Several groups have used proteomic technologies, including mass spectrometry, to identify DLBCL protein changes, resulting in an understanding of the perturbations that can alter protein functions and cellular response post-transcriptionally and post-translationally. The application of mass spectrometry-based analysis has resulted in a variety of different identifications of protein alterations in DLBCL [51,52]. These alterations include differential protein expressions, post-translational modifications (e.g., phosphorylation, acetylation, and methylation), and protein-protein interactions [53]. These discoveries have brought the identification of dysregulated cellular pathways and signaling networks in DLBCL to light. Proteomic studies have captured abnormal phosphorylation events in key signaling pathways relevant to DLBCL pathobiology, including the B-cell receptor (BCR) signaling pathway [54], the phosphatidylinositol 3'-kinase (PI3K)-Akt (PI3K-AKT) pathway [55], and the nuclear factor-kappa B (NF- $\kappa$ B) pathway [56]. Together, these results provide several potential ways in which therapeutic targets may be manipulated to create precision medicines for the treatment of DLBCL (Fig. 2).

**Metabolomics:** Metabolomics has been used to provide new insight into DLBCL, unveiling the metabolic reprogramming that supports the distinguishing hallmark of this disease. Metabolic reprogramming in DLBCL: One of the key metabolic alterations identified in DLBCL is metabolic reprogramming. To meet the increased energy and biosynthesis demands, DLBCL cells undergo a metabolic reprogramming to rely on nuclear-pentose phosphate pathway (n-PPP), which potentially favors rapid proliferation and survival. Two pivotal changes in metabolism, identified in metabolomic studies, include potentiated glycolysis and glutaminolysis (Fig. 2). In addition, the metabolomic studies also showed dysregulation of lipid metabolism, nucleotide metabolism, amino acid metabolism, and others in DLBCL. These changes all feed into the widespread metabolic reprogramming in DLBCL to enable the aggressive growth characteristic of DLBCL cells [57–59].

In conclusion, multi-omics data integration has significantly enriched our understanding of DLBCL, revealing the complex interplay between genetic, transcriptional, translational, and metabolic changes underlying this disease.

## 4. Spatial Multi-Omics: Adding a New Dimension to DLBCL Research

### 4.1 Explanation of the Concept and Significance of Spatial Multi-Omics

Spatial multi-omics is an emerging field that combines traditional multi-omics analysis with spatial context within biological tissues. Traditional multi-omics approaches, which include genomics, transcriptomics, proteomics, and

metabolomics, provide a comprehensive view of the molecular mechanisms within a cell. However, these methods often neglect the spatial context of these molecular events. Spatial multi-omics fills this gap by preserving the spatial information during the analysis, thereby providing a comprehensive, contextual view of the molecular landscape within tissues [22].

The significance of spatial multi-omics lies in its ability to provide a more holistic view of biological tissues. By preserving the spatial information, researchers can correlate molecular events with their precise location within the tissue, leading to a more thorough understanding of the complex interplay between cells and their environment. This is particularly important in diseases such as cancer, where the tumor microenvironment plays a critical role in disease progression and response to treatment [60].

Spatial multi-omics has wide-ranging applications in biology and medicine. It has been utilized to study the complex tissue architecture of the brain, the heterogeneous tumor microenvironment in cancer, and the intricate cellular interplay in the immune system. One study has provided novel insights into disease mechanisms, leading to the identification of new therapeutic targets [61].

The future of spatial multi-omics is promising. As technology advances, we can expect more accurate, high-throughput methods to capture a broader range of molecular events. This will undoubtedly deepen our understanding of complex diseases and pave the way for personalized medicine [62].

### 4.2 Insights into DLBCL from Spatial Multi-Omics Investigations

DLBCL is the most common type of non-Hodgkin lymphoma, characterized by significant heterogeneity at the clinical, pathological, and molecular levels [13]. Spatial multi-omics has recently been applied to DLBCL research, providing vital insights into the tumor microenvironment and the spatially distinct molecular patterns within the tumor [22].

Spatial multi-omics has revealed intricate heterogeneity within DLBCL tumors. By capturing the spatial context of genomic, transcriptomic, proteomic, and metabolomic data, researchers have identified distinct subpopulations of cancer cells within a single DLBCL tumor, each characterized by unique molecular profiles [61].

The application of spatial multi-omics in DLBCL has also significantly advanced our understanding of the tumor microenvironment. The spatial distribution of immune cells and their interactions with cancer cells have been mapped in detail, shedding light on the complex interplay between the immune system and cancer cells and its impact on disease progression and treatment response [60].

Spatial multi-omics has not only deepened our understanding of DLBCL but has also opened new avenues for treatment. More targeted and personalized treatment ap-

proaches can be devised by identifying molecularly distinct subpopulations of cells and understanding their spatial distribution within the tumor [63].

## 5. Successful Applications of Multi-Omics and Spatial Multi-Omics in DLBCL

### 5.1 In-depth Analysis of Multi-Omics Approaches

The seminal work of Chapuy *et al.* [15] on multi-omics highlighted the utility of integrating genomic, transcriptomic, and epigenomic data to stratify DLBCL into five molecular subtypes with distinct characteristics [6]. Their genomic data, derived from whole-exome sequencing, provides insights into somatic mutations and copy number variations. Transcriptomic data from RNA sequencing offers a landscape of gene expression, while epigenomic data, acquired through DNA methylation profiling, reveals patterns of epigenetic modifications across the subtypes.

The analytical tools employed in this study included bioinformatic pipelines for variant calling, differential gene expression analysis, and methylation array analysis. Unsupervised clustering algorithms, such as hierarchical clustering, delineated molecular subtypes based on the multi-dimensional data. The biological insights gained included the identification of unique driver mutations, dysregulated pathways, and epigenetic markers that define each subtype, offering new avenues for targeted therapies and prognostic biomarkers.

Schmitz *et al.* [6] focused on the genetic underpinnings of DLBCL by integrating whole-exome sequencing with transcriptomic data to unearth four genetic subtypes. Advanced computational methods, such as mutual exclusivity analysis and gene set enrichment analysis, were applied to dissect the intricate network of genetic interactions and pathway involvements.

The study by Schmitz *et al.* [6] provided a nuanced genetic classification system that holds prognostic value and identifies actionable targets. For instance, one subtype, characterized by B-cell receptor pathway mutations, might be amenable to B-cell receptor signaling inhibitors.

### 5.2 Advancements in Spatial Multi-Omics

Vickovic *et al.* [64] took advantage of the spatial transcriptomics platform to map the gene expression landscape of DLBCL in relation to the tumor's histological structure. This technique enabled the generation of spatially resolved gene expression data by capturing RNA from histological tissue sections and associating each transcript with its original location.

The analytical methodology involved spatial barcoding, image registration, and computational deconvolution to integrate the spatial information with gene expression profiles. Vickovic *et al.*'s study's [64] insights provided a molecular counterpart to the morphological features ob-

served in DLBCL, which has implications for understanding the tumor microenvironment and its role in lymphomagenesis.

Moncada *et al.* [65] combined spatial transcriptomics with single-cell RNA sequencing to elucidate the cellular heterogeneity and spatial organization within DLBCL tumors. Single-cell RNA-seq provided a high-resolution view of the transcriptomic diversity at the individual cell level. At the same time, spatial transcriptomics offered a global view of the spatial distribution of these heterogeneous populations.

These datasets were integrated using computational frameworks that allow for the alignment of single-cell transcriptomic data with spatial gene expression patterns. This multi-faceted approach uncovered spatially distinct cell subpopulations, which could have profound implications for understanding the microenvironmental niches within DLBCL, identifying potential therapeutic strategies, and predicting treatment responses.

### 5.3 Advancing toward Precision Medicine through Multi-Omics Studies

The multi-omics and spatial multi-omics studies by Chapuy *et al.* [15], Schmitz *et al.* [6], and Davidson-Moncada *et al.* [66] have collectively advanced our understanding of DLBCL's molecular and spatial heterogeneity. These studies have laid the groundwork for precision medicine by uncovering distinct molecular subtypes and spatial heterogeneity.

The identification of genetic and epigenetic alterations, as well as the spatial organization of tumor cells, provides critical insights for the development of personalized therapies. For instance, targeting subtype-specific driver mutations or dysregulated pathways can lead to more effective and less toxic treatments.

Furthermore, understanding the spatial heterogeneity within DLBCL tumors can lead to strategies that account for the tumor's microenvironmental complexity, potentially improving the efficacy of immunotherapies or other targeted treatments. This could involve spatially targeted drug delivery or the modulation of the tumor microenvironment to enhance treatment response.

In conclusion, these case studies underscore the transformative potential of multi-omics and spatial multi-omics in pursuing precision medicine for DLBCL. They exemplify how integrating diverse data types can yield a more comprehensive understanding of cancer biology, ultimately guiding the development of more personalized and effective therapeutic strategies. Continued research in this domain is expected to refine our understanding of DLBCL further and improve patient outcomes in the clinical setting.

## 6. Artificial Intelligence: A Catalyst for Multi-Omics Data Interpretation

### 6.1 Introduction to AI and Machine Learning in Biomedical Research

AI has emerged as a transformative force in biomedical research, particularly in analyzing complex datasets generated by multi-omics studies. AI encompasses a suite of computational techniques that emulate aspects of human cognition, such as learning and problem-solving, to process and interpret data. ML, a subset of AI, involves algorithms that enable computers to learn from and make predictions or decisions based on data [67]. This capability is particularly pertinent to biomedical research, where the volume and complexity of data have outstripped the capacity of conventional statistical approaches.

Supervised, unsupervised, and reinforcement learning are three broad categories of ML algorithms. Support vector machines (SVMs) are a group of powerful supervised learning algorithms used for classification in a wide range of biomedical research. SVMs find a hyperplane that maximizes the margin between classes—so it classifies more effectively [68]. On the other hand, random forests are ensemble learning methods that combine multiple decision trees to make predictions [69]. They can deal with high-dimensional data and encode complex relationships. Applications of unsupervised learning algorithms are used to extract the patterns or clusters in the data that have no previous labels. This feature makes them ideal for conducting exploratory analysis and discovering hidden structures within datasets. For example, K-means clustering is a method of clustering data into a set number of clusters using data point similarity. Data points are associated with the nearest mean of a cluster [70]. Principal component analysis (PCA) is a dimensionality reduction method that can convert high-dimensional data into a set of lower-dimension data reflecting most of the information in the data [71]. Reinforcement learning is a ML technique that learns to act in a dynamic environment to obtain a reward [72].

The advent of transformer models, initially conceived for complex sequence-based tasks in natural language processing (NLP), presents a novel paradigm for analyzing multi-omics data in DLBCL research [73]. These models, particularly noteworthy for their self-attention mechanisms, can recognize intricate patterns within vast datasets, making them well-suited for identifying biomarkers and therapeutic targets [74]. Transformers can be applied to multi-omics data by treating molecular profiles as sequences, where each omics feature—be it a gene expression level, a mutational status, or an epigenetic mark—can be considered analogous to a word in a sentence [46]. This allows for capturing long-range dependencies between disparate omics features, potentially unraveling complex biological interactions that underpin DLBCL pathogenesis and patient heterogeneity [75].

In DLBCL research, transformers can be trained to predict clinical outcomes by learning the high-dimensional relationships between genetic alterations, transcriptomic profiles, and patient responses to therapy. Such models could also be fine-tuned to stratify patients into molecular subtypes with greater precision, thereby facilitating the development of subtype-specific treatments [76]. Moreover, transformer models can facilitate the integration of spatial multi-omics data, which combines molecular profiling with spatial information, providing insights into the tumor microenvironment and its role in lymphomagenesis. By doing so, researchers can gain a more comprehensive understanding of DLBCL, leading to the identification of novel biomarkers and therapeutic targets [77].

Many statistical considerations and practices must be addressed when applying AI techniques, including ML and transformer models, to biomedical research and analysis of multi-omics data in DLBCL. It is important to appreciate that the specific implementation of statistical tests, the validated assumptions, the  $p$ -values, and the correction for multiple testing may differ depending on the ML algorithms, programming libraries, and software tools used in DLBCL research. Therefore, DLBCL research must include proper statistical approaches and guidelines for robust and reproducible multi-omics data analysis.

### 6.2 Specific AI Algorithms and Their Applications in Multi-Omics

In multi-omics, AI algorithms are applied to synthesize and interpret data across genomic, transcriptomic, proteomic, and metabolomic layers. Deep learning, an advanced ML technique characterized by the use of neural networks with multiple layers, has proven particularly adept at handling the non-linear and high-dimensional nature of multi-omics data. Convolutional neural networks (CNNs) are a particular type of feedforward neural network and have been shown to be quite successful in interpreting DNA sequences [78]. Convolutional layers in CNNs process input data to extract local features of input data and identify patterns. CNNs can also be used to analyze genomic data (e.g., DNA sequences) to discover significant genomic regions, regulatory elements, or sequence motifs associated with the progression of the disease or specific molecular subtypes in DLBCL research [79]. In contrast, recurrent neural networks (RNNs) appear well-suited for capturing the inherent sequential information present in transcriptomic data [80]. RNNs are meant to deal with sequence data introduction of feedback connections that store information from previous inputs. Thus, they are ideal for uncovering patterns in time series or longitudinal data (e.g., gene expression profiles at different stages of DLBCL progression or treatment response) [81]. The model can learn dynamic gene expression patterns, regulatory networks, or disease outcome prediction, as RNNs can capture temporal dependencies within the data [82].

The utilization of AI in oncology currently faces a pivotal challenge, as the prevalent single-modality approach needs to harness the comprehensive clinical landscape richly detailed by the array of patient data available. To elevate the precision and dependability of AI-driven diagnostic and prognostic models—and to facilitate their integration into clinical practice—it is imperative that we amalgamate disparate data modalities. These include, but are not limited to, radiology, histology, genomics, and electronic health records. The strategic integration of such multimodal data enhances model performance and positions AI to reveal intricate patterns that traverse these modalities. These patterns may prove crucial in elucidating the heterogeneity in patient outcomes and treatment resistance. Discoveries precipitated by such an integrated AI approach hold the promise of propelling investigative studies and fostering the emergence of novel biomarkers and therapeutic targets. Moreover, the application of AI methods in multimodal data fusion—and the pursuit of strategies to render AI models interpretable and exploratory—will be instrumental in addressing the existing challenges and catalyzing the clinical adoption of AI in oncology [83].

### 6.3 AI in Enhancing Data Integration, Analysis, and Visualization

AI facilitates multi-omics data integration by aligning disparate data types onto a unified analytical framework. This process is critical, given that different -omics data can have varying scales, distributions, and levels of noise. AI-powered integration strategies enhance the ability to uncover meaningful biological insights that would be obscured if each data type were analyzed in isolation [84].

Regarding analysis, AI algorithms are adept at sifting through the noise to identify signals of biological relevance. For instance, feature selection techniques reduce dimensionality and improve model interpretability by isolating the most informative attributes from complex datasets.

Visualization is another area where AI contributes significantly. Techniques such as t-distributed stochastic neighbor embedding (t-SNE) and uniform manifold approximation and projection (UMAP) employ ML to reduce the dimensionality of data for visualization, enabling researchers to discern patterns and clusters within multi-omics datasets [85].

### 6.4 Predictive Modeling and Its Implications for DLBCL

The advent of predictive modeling through AI has opened new horizons in managing and treating DLBCL. AI's capacity to integrate and analyze complex multi-omics data allows for the generation of prediction models that can forecast disease trajectory, evaluate patient outcomes, and predict individual responses to various therapeutic interventions [86].

The power of AI in this context lies in its ability to sift through extensive datasets, identifying patterns and

biomarkers that might otherwise elude human analysis. Such biomarkers are essential tools for patient stratification, enabling clinicians to categorize individuals into distinct groups based on their risk profiles or anticipated responsiveness to specific treatments. This stratification is paramount, as it facilitates tailoring treatment regimens to individual patient needs, thereby enhancing the efficacy of therapeutic interventions [87].

Furthermore, AI-driven predictive models can potentially revolutionize the clinical approach to DLBCL by providing a more nuanced understanding of the disease's heterogeneous nature. By elucidating the underlying biological processes that drive DLBCL progression and response to treatment, AI can assist in designing personalized medicine strategies [88]. These strategies aim not only to improve survival rates but also to optimize the quality of life for patients by minimizing exposure to ineffective treatments and their associated toxicities.

Moreover, the prognostic precision offered by AI models is invaluable. Accurate prognostic assessments can inform clinical decision-making, leading to better resource allocation and patient counseling. As such, integrating AI into the clinical workflow for DLBCL can significantly enhance patient management, from diagnosis to survivorship [89].

To harness the full potential of predictive modeling in DLBCL, ongoing research and collaboration between computational scientists and clinical experts are essential. The continuous refinement of AI algorithms and the expansion of high-quality, annotated omics datasets will undoubtedly lead to improved model accuracy and reliability. As AI models become increasingly sophisticated and validated in clinical settings, they promise to become indispensable tools in the fight against DLBCL, ultimately leading to more informed, effective, and personalized patient care.

## 7. AI-Based Achievements in DLBCL

The recent progress in DLBCL driven by AI has been awe-inspiring. These accomplishments have changed research and medical practices, such as predicting prognosis, diagnosing, evaluating treatment responses, and classifying subtypes.

A notable advancement is the creation of the “contained model” (McPM) by Peng *et al.* [81]. This innovative model combines AI analysis of markers with information to predict overall survival (OS) and progression-free survival (PFS) in DLBCL patients. By surpassing models like the international prognostic index (IPI), the McPM has opened doors for personalized treatment approaches by integrating data into prognostic models.

AI-driven image analysis has made strides in assessing programmed death-ligand 1 (PD-L1) expression in patients [90]. These developments have enhanced the objectivity and clarity of PD-L1 measurement, thereby advancing targeted immunotherapy research for DLBCL. The in-

roduction of a PD-L1 scoring system highlights how AI can precisely measure PD-L1 expressions, supporting immunotherapy advancements.

On the diagnosis front, deep learning-based diagnostic systems have achieved unprecedented diagnostic accuracy in DLBCL cases [91] by using convolutional neural networks and addressing image-to-image variation and data rarity challenges with pre-trained networks, data augmentation, regularization, and cross-validation. A consequent result is diagnostic accuracy of nearly 100%, which is instrumental in improving DLBCL diagnosis and patient management.

Moreover, deep learning and digital pathology have enabled the prediction of DLBCL treatment response [92]. Combining clinical data with pathology image features, multimodal prediction models have been able to predict relapse-free survival as well as identify histological features related to treatment response. This innovation supports the potential of AI-based diagnosis as a robust diagnostic and prognostic tool in DLBCL management.

It is possible to predict treatment responses using meticulous algorithms of convolutional neural networks and maximum intensity projection (MIP) images [93]. In less than two years, these AI networks have exceeded accuracy in predicting time-to-progression (TTP) over traditional models, including the established IPI. It allows clinicians to adapt treatment strategies at an early stage for these more at-risk patients in order to improve care.

AI networks have also helped categorize lymphomas. Specifically, interpretable AI methods such as LymphoML reach a diagnostic accuracy similar to pathologists on H&E-stained tissue [94]. Three dimensional (3D)—and AI-based approaches have had a major impact on morphologic boundary-breaking, texture, and architectural boundary-breaking. By integrating these features with immunostains, AI tools can help pathologists improve diagnostic accuracy and subtyping of lymphomas.

ML approaches to predict DLBCL molecular subtypes using gene expression data have significantly advanced in recent years [95]. Different models from ML methods have exhibited substantial accuracy for extracting subtypes in subtype classification, critically highlighting the importance of AI in this domain.

In summary, AI has revolutionized the way that DLBCL is currently challenged. These advances have been major in the context of prognosis prediction, diagnosis, treatment response evaluation, and subtyping. They have provided opportunities for improving patient outcomes in DLBCL through better molecular integration, more understandable prediction, and wider applicability given currently advanced AI techniques.

## 8. Future Directions for AI-Enhanced Multi-Omics in DLBCL

### 8.1 Advanced Computational Methods for Data Integration

The evolution of AI in multi-omics necessitates the development of more sophisticated computational methods to integrate diverse -omics data types effectively. Novel integration approaches, such as deep integration, that leverage the representational learning abilities of deep learning architectures promise to capture the complex, non-linear relationships inherent to multi-omics data [96]. One emerging method uses variational autoencoders (VAEs) that can learn latent representations of data, potentially unveiling new biological insights that are not readily apparent from individual-omics layers [97].

### 8.2 High-Resolution Technologies and Their Expected Advancements

High-resolution technologies, such as single-cell sequencing and spatial transcriptomics, are expected to advance, enabling the understanding of DLBCL in unprecedented detail. Integrating AI with these technologies will facilitate the identification of cellular heterogeneity and microenvironmental influences on DLBCL pathogenesis [97]. Advances in imaging technologies, combined with AI-like convolutional neural networks, will enhance our ability to visualize and quantify the spatial arrangement of tumor cells and their microenvironments.

### 8.3 Large-Scale, Well-Annotated Multi-Omics Datasets for DLBCL

The prognostic and therapeutic landscapes of DLBCL are poised to benefit from the compilation of large-scale, well-annotated multi-omics datasets. Such datasets should encompass comprehensive genetic, epigenetic, transcriptomic, proteomic, and metabolomic profiles and clinical data such as treatment history and patient outcomes. Efforts like The Cancer Genome Atlas (TCGA) have laid the groundwork for multi-omics data repositories, but DLBCL-specific initiatives are needed to address the heterogeneity of the disease [98]. These datasets will be invaluable for training AI models to uncover novel biomarkers and therapeutic targets.

### 8.4 Translational Research: From Bench to Bedside

The path from bench to bedside for AI-enhanced multi-omics in DLBCL involves the translation of computational discoveries into clinical applications. AI algorithms must be validated in preclinical studies and clinical trials to assess their predictive power and clinical utility [99]. Additionally, integrating AI with electronic health records (EHRs) will facilitate real-time predictive analytics, potentially improving the management of DLBCL.

In the era of personalized medicine, treatments for DLBCL will increasingly be tailored to the molecular and spatial profiles of individual tumors. AI will play a critical role in interpreting these complex datasets, aiding in selecting targeted therapies and predicting treatment responses [100]. Furthermore, integrating spatial multi-omics will provide a more comprehensive understanding of the tumor microenvironment, thus enabling the identification of novel therapeutic windows and the development of more efficacious combination therapies [101].

Collectively, these future directions underscore the potential of AI-enhanced multi-omics to revolutionize our understanding and treatment of DLBCL, ultimately leading to more precise and effective clinical interventions. Continuous advancements in computational methods, high-resolution technologies, and the acquisition of large-scale multi-omics datasets will be the cornerstone of this transformative journey in cancer immunology.

## 9. Challenges and Limitations

### 9.1 Current Limitations of Multi-Omics

One of the key challenges in multi-omics studies is the integration of data from different omics platforms. When we provide data for each omics layer (genomics, transcriptomics, proteomics), each comes with a large amount of data, and individual properties and difficulties [38]. Meeting the challenge: Computational methods and algorithms are being developed to provide robust data integration for analyzing multi-omics data [102–105]. These methods intend to capture the inherent relationships and dependencies that exist among diverse omics layers and, in doing so, provide a more systems-level interpretation of the biological system being interrogated [102]. However, accurate and robust integration across this diversity of omics layers into the network view of the cell is a formidable challenge, as the data formats, measurement technologies, and noise levels across those omics layers are inevitably different [106]. In addition, the interpretation and validation of integrated multi-omics results pose additional hurdles [103]. Conclusion-Comprehensive biological validation: Studies that perform biological validation of findings across multiple omics layers are essential but must be carefully designed and executed. Moreover, the integration of multi-omics data is typically reductionist in practice, requiring considerable approximations and assumptions that could compromise the accuracy and robustness of the outcome.

Batch effects, the non-biological variations that occur due to differences in experimental conditions, pose a significant challenge to multi-omics studies [107]. These effects can confound the interpretation of results and affect the re-

producibility of the studies. Therefore, proper experimental design and normalization techniques are necessary to mitigate these effects.

For spatial multi-omics, one of the primary technical challenges is maintaining the spatial resolution while obtaining omics data [22]. While methods like spatial transcriptomics provide spatial context, they are often limited by their resolution. Additionally, linking the spatial distribution of different omics layers (like mRNA and proteins) is currently a challenge due to limitations in technological capabilities.

Interpreting the results from spatial multi-omics studies can also be challenging. Understanding how the spatial organization of different molecular features correlates with biological function and disease progression requires sophisticated computational tools and a deep understanding of the biology [65]. Developing new analytical methods and visualization tools that can handle spatially resolved multi-omics data is an active area of research.

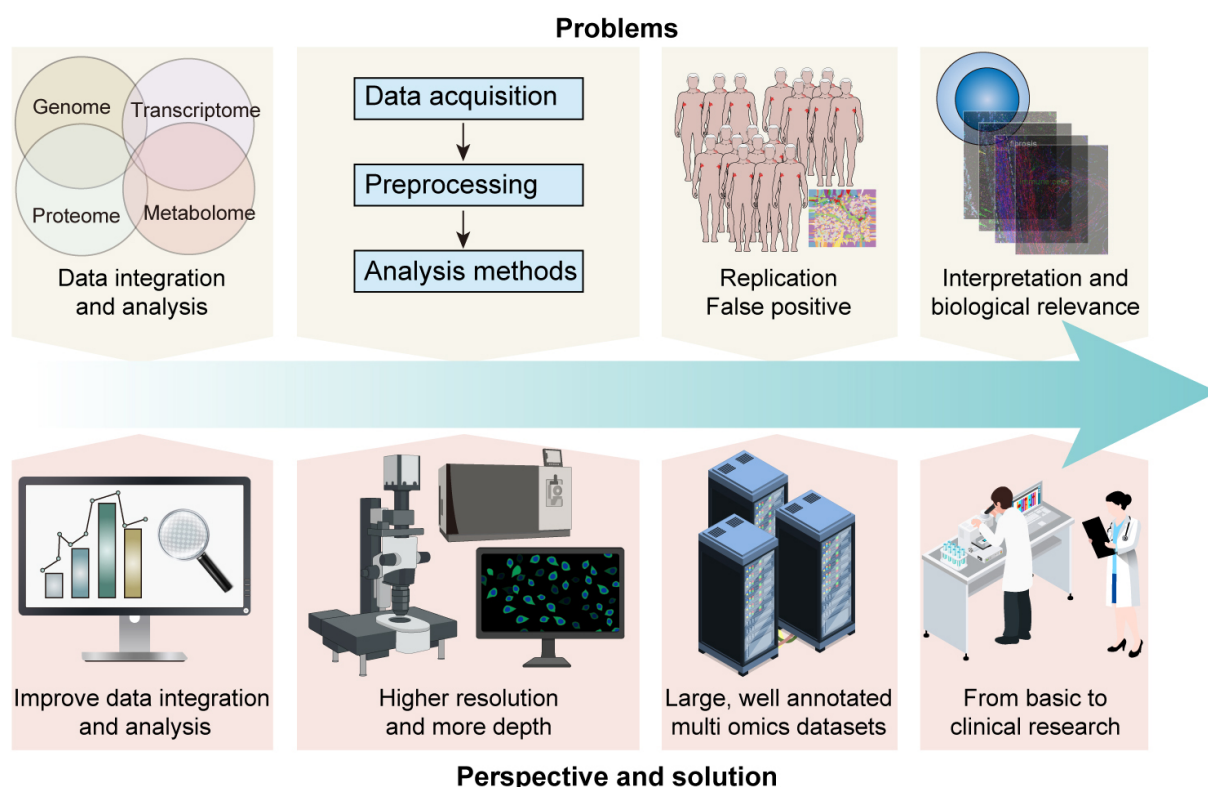
### 9.2 Current Limitations of Spatial Multi-Omics

One of the major challenges in multi-omics and spatial multi-omics is the integration and analysis of high-dimensional data from different omics layers [38]. Each layer (genomic, transcriptomic, proteomic) generates a massive amount of data with unique characteristics, and integrating these data in a biologically meaningful way poses a significant computational challenge.

Multi-omics data is high dimensional data, which results in an increased complexity of analysis and interpretation. In addition, interoperability of multi-omics and spatial omics data relies on specific computational tools [108]. On the other hand, spatial omics, e.g., spatial transcriptomics or imaging mass cytometry, deliver information on the spatial distributions of molecularly distinct profiles within tissues or cells. However, with the integration of these spatially-resolved datasets with multi-omics data, the computational analysis composes a more complicated problem [108]. Identifying crucial patterns and correlations within the data of high-dimensional multi-omics is another challenge. As many factors are involved, it becomes more difficult to separate actual biological signals from noise [109].

Various computational methods to address the computational challenges in multi-omic and spatial multi-omics analyses — dimensionality reduction approaches, network-based analysis methods, and state-of-the-art ML algorithms — are being actively investigated by researchers [108,110]. Such approaches are designed to reduce the dimensionality, select relevant points, and extract biologically informative data from high-dimensional data.

Another area for improvement in the field is the need for more standardization in data acquisition, preprocessing, and analysis methods. Different labs may use different protocols, which could result in inconsistencies when compar-



**Fig. 3. Overcoming challenges and shaping future directions in DLBCL: Leveraging multi-omics and spatial multi-omics for precision medicine.** This figure illustrates the challenges of data integration, method standardization, reproducibility, and biological interpretation in DLBCL research. It further outlines future directions, including enhanced computational tools, advanced spatial multi-omics technologies, the creation of comprehensive omics databases, and translational applications. The central theme of DLBCL is visually anchored, with arrows indicating the flow from current challenges to future innovations.

ing results from different studies [107]. Efforts are ongoing to establish standardized protocols and best practices, but this remains a significant hurdle for the field.

Another important issue is the reproducibility of findings from multi-omics and spatial multi-omics studies. Given the complexity and high dimensionality of the data, false positives can be a problem [111]. Therefore, rigorous validation of findings, both computationally and experimentally, is necessary to ensure the robustness of the results.

Finally, interpreting the results from multi-omics and spatial multi-omics studies can be challenging. It is not always clear how the observed molecular and spatial patterns correlate with biological function and disease phenotypes [103]. Further biological experiments are often required to validate and explore the biological relevance of these findings (Fig. 3).

### 9.3 Challenges in AI Application and Potential Solutions

Despite the potential of AI in multi-omics, several challenges persist. One significant hurdle is the “black-box” nature of some AI algorithms, particularly deep learning, which can make the interpretation of results difficult for clinicians. Efforts are being made to improve the trans-

parency and interpretability of AI models. Techniques such as model explainability and interpretability allow researchers to gain insights into the decision-making process of AI models. This can help identify and mitigate biases and ensure that AI models make fair and reliable predictions [112].

Another challenge is the need for large, well-annotated datasets to train AI models. Biases can arise from various sources, such as imbalances in the training data, human annotator biases, or systemic biases in the data collection process. If the training data does not represent the target population or contains inherent biases, the AI model may perpetuate or amplify these biases in its predictions and decisions. Techniques such as transfer learning, where a model trained on one task is repurposed for another related task, and data augmentation, which artificially expands the training dataset, are potential solutions to this challenge [113].

Lastly, ethical considerations must be addressed, particularly regarding patient privacy and data security. Ensuring the ethical use of AI in biomedical research requires adherence to established guidelines and ongoing dialogue between researchers, ethicists, and policymakers [114].

#### 9.4 Limitations of AI-driven Diagnosis and Treatment of DLBCL

AI technologies can achieve important insights and assist in risk stratification, prognosis prediction, and subtype classification, but they have not transformed the DLBCL diagnosis and treatment landscape.

One of the primary limitations is the complexity and heterogeneity of DLBCL [115,116]. This hematological malignancy shows high inter- and intra-tumor heterogeneity, thereby hampering an accurate diagnosis and therapy. Significant heterogeneity is found within DLBCL, describing the disease as several molecular subtypes with apparent differences in clinical behavior and requiring different treatment modalities [117]. Even though AI models may help find subtypes and predict outcomes, translating these findings into personalized treatment decisions is difficult.

DLBCL diagnosis requires multiple time-consuming diagnostic modalities, such as histopathology, immunohistochemistry, and genetic testing [118]. It is notoriously difficult to integrate those diverse data sources and interpret the results cohesively. Although AI algorithms can help in data integration and exploration analysis, they have not yet had a large impact on enhancing the accuracy and efficiency of DLBCL diagnosis.

Treatment decision-making in DLBCL is mainly driven by established protocols and clinical guidelines derived from large-scale clinical studies [118]. Although precision medicine has progressed tremendously, AI-based models have only recently been integrated into treatment planning. Treatment of DLBCL has been adapted to patient characteristics, disease stage, genetic markers and a vast number of predictors of response for different interventions [118]. Whilst AI models may provide support for treatment stratification and prediction, their implementation within routine clinical care and validation of their effectiveness is a recurring challenge.

To overcome these limitations, concerted efforts involving multidisciplinary collaborations between researchers, clinicians, and regulatory bodies are required. Longitudinal studies, prospective clinical trials, and the establishment of comprehensive databases are essential for generating high-quality data, validating AI models, and integrating them into clinical workflows.

In conclusion, AI is proving to be a catalyst for interpreting and applying multi-omics data in DLBCL research. By leveraging AI's capabilities in data integration, analysis, and predictive modeling, researchers can advance the frontiers of personalized medicine for DLBCL. However, the challenges associated with AI, such as interpretability, data requirements, ethical considerations, and the complex nature of DLBCL, must be diligently navigated to realize AI's transformative potential fully.

## 10. Future Directions: Moving toward Precision Medicine in DLBCL

Despite the advent of multi-omics and spatial multi-omics, downstream challenges in integrating and analyzing high-dimensional data remain [38]. The complexity and interdependence of multi-omics data represent a significant clinical challenge critical in developing computational tools and statistical models. In light of this, it is necessary to create more advanced data integration tools, and the focus should be on creating easily usable tools that can deal with all the complexities of multi-omics datasets.

Existing spatial multi-omics technologies are restricted by resolution and the number of features that can be simultaneously measured [22]. Thus, a critical avenue for future research is to develop the technology to increase the resolution and the multi-omics depth. Improving these parameters allows researchers to get more accurate and in-depth information about the spatial structure and molecular interactions in the biological states.

Most diseases need more large, well-annotated multi-omics datasets that are publicly available [103]. These datasets are precious resources used to develop and test new computational tools and better understand disease. Therefore, there is a need to focus on creating and disseminating complete multi-omics data to promote discovery and help make considerable progress toward a better comprehension of complex diseases.

Although multi-omics and spatial multi-omics have considerable benefits in discovering new biology and diseases, using these data to transform clinical care is a significant challenge [119]. Future research needs to concentrate on how to apply these techniques to the prevention, diagnosis, prognosis, and treatment of a myriad of disease forms (Fig. 3). In translating research findings to the clinic, the integration of multi-omics and spatial multi-omics can bring new knowledge and insights to personalized medicine, providing effective healthcare.

## 11. Conclusion

In conclusion, the review underscores the transformative potential of integrating artificial intelligence (AI) with multi-omics and spatial multi-omics technologies in the realm of Diffuse Large B-cell Lymphoma (DLBCL) research. Such sophisticated methodologies are being used by researchers in an attempt to make sense of the molecular complexities and spatial variability characteristic of DLBCL, thereby leading to improvements in diagnostic accuracy, individualized treatment strategies, and improved risk classification. While considering the challenges in data standardization, result interpretability, and generalizability, the synergy between AI and multi-omics can hopefully propel precision medicine in DLBCL. More sophisticated AI-driven frameworks are needed, and large multi-omics

datasets should accumulate so that the laboratory findings can be translated into bedside practice in the new era of patient-centric oncology care for DLBCL treatment.

## Author Contributions

YS wrote the manuscript, edited the drafts, and conceptualized the figures. XL and SY searched the literature and created the charts, and wrote the original draft. QG designed the study and reviewed and edited the manuscript. All authors read and approved the final manuscript. All authors contributed to editorial changes in the manuscript. All authors have participated sufficiently in the work to take public responsibility for appropriate portions of the content and agreed to be accountable for all aspects of the work in ensuring that questions related to its accuracy or integrity.

## Ethics Approval and Consent to Participate

Not applicable.

## Acknowledgment

Not applicable.

## Funding

This research received no external funding.

## Conflict of Interest

The authors declare no conflict of interest.

## References

- [1] Armitage JO, Weisenburger DD. New approach to classifying non-Hodgkin's lymphomas: clinical features of the major histologic subtypes. Non-Hodgkin's Lymphoma Classification Project. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*. 1998; 16: 2780–2795.
- [2] Sehn LH, Gascoyne RD. Diffuse large B-cell lymphoma: optimizing outcome in the context of clinical and biologic heterogeneity. *Blood*. 2015; 125: 22–32.
- [3] Friedberg JW. Relapsed/refractory diffuse large B-cell lymphoma. *Hematology. American Society of Hematology. Education Program*. 2011; 2011: 498–505.
- [4] Teras LR, DeSantis CE, Cerhan JR, Morton LM, Jemal A, Flowers CR. 2016 US lymphoid malignancy statistics by World Health Organization subtypes. *CA: a Cancer Journal for Clinicians*. 2016; 66: 443–459.
- [5] Coiffier B, Thieblemont C, Van Den Neste E, Lepeu G, Plantier I, Castaigne S, *et al.* Long-term outcome of patients in the LNH-98.5 trial, the first randomized study comparing rituximab-CHOP to standard CHOP chemotherapy in DLBCL patients: a study by the Groupe d'Etudes des Lymphomes de l'Adulte. *Blood*. 2010; 116: 2040–2045.
- [6] Schmitz R, Wright GW, Huang DW, Johnson CA, Phelan JD, Wang JQ, *et al.* Genetics and Pathogenesis of Diffuse Large B-Cell Lymphoma. *The New England Journal of Medicine*. 2018; 378: 1396–1407.
- [7] Collins FS, Varmus H. A new initiative on precision medicine. *The New England Journal of Medicine*. 2015; 372: 793–795.
- [8] Moffitt AB, Dave SS. Clinical Applications of the Genomic Landscape of Aggressive Non-Hodgkin Lymphoma. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*. 2017; 35: 955–962.
- [9] García-Sanz R, Jiménez C. Time to Move to the Single-Cell Level: Applications of Single-Cell Multi-Omics to Hematological Malignancies and Waldenström's Macroglobulinemia—A Particularly Heterogeneous Lymphoma. *Cancers*. 2021; 13: 1541.
- [10] Morin RD, Mendez-Lago M, Mungall AJ, Goya R, Mungall KL, Corbett RD, *et al.* Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. *Nature*. 2011; 476: 298–303.
- [11] Lohr JG, Stojanov P, Lawrence MS, Auclair D, Chapuy B, Sougnez C, *et al.* Discovery and prioritization of somatic mutations in diffuse large B-cell lymphoma (DLBCL) by whole-exome sequencing. *Proceedings of the National Academy of Sciences of the United States of America*. 2012; 109: 3879–3884.
- [12] Davis RE, Ngo VN, Lenz G, Tolar P, Young RM, Romesser PB, *et al.* Chronic active B-cell-receptor signalling in diffuse large B-cell lymphoma. *Nature*. 2010; 463: 88–92.
- [13] Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, *et al.* Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature*. 2000; 403: 503–511.
- [14] Lenz G, Wright GW, Emre NCT, Kohlhammer H, Dave SS, Davis RE, *et al.* Molecular subtypes of diffuse large B-cell lymphoma arise by distinct genetic pathways. *Proceedings of the National Academy of Sciences of the United States of America*. 2008; 105: 13520–13525.
- [15] Chapuy B, Stewart C, Dunford AJ, Kim J, Kamburov A, Redd RA, *et al.* Molecular subtypes of diffuse large B cell lymphoma are associated with distinct pathogenic mechanisms and outcomes. *Nature Medicine*. 2018; 24: 679–690.
- [16] Green MR, Kihira S, Liu CL, Nair RV, Salari R, Gentles AJ, *et al.* Mutations in early follicular lymphoma progenitors are associated with suppressed antigen presentation. *Proceedings of the National Academy of Sciences of the United States of America*. 2015; 112: E1116–E1125.
- [17] Ludvigsen M, Campbell AJ, Enemark MB, Hybel TE, Karjalainen-Lindsberg ML, Beiske K, *et al.* Proteomics uncovers molecular features for relapse risk stratification in patients with diffuse large B-cell lymphoma. *Blood Cancer Journal*. 2023; 13: 161.
- [18] Kluckova K, D'Avola A, Riches JC. Advances in Understanding of Metabolism of B-Cell Lymphoma: Implications for Therapy. *Cancers*. 2022; 14: 5552.
- [19] Magnes T, Wagner S, Thorner AR, Neureiter D, Klieser E, Rinnerthaler G, *et al.* Spatial Heterogeneity in Large Resected Diffuse Large B-Cell Lymphoma Bulks Analysed by Massively Parallel Sequencing of Multiple Synchronous Biopsies. *Cancers*. 2021; 13: 650.
- [20] Liu M, Bertolazzi G, Sridhar S, Lee RX, Jaynes P, Mulder K, *et al.* Spatially-resolved transcriptomics reveal macrophage heterogeneity and prognostic significance in diffuse large B-cell lymphoma. *Nature Communications*. 2024; 15: 2113.
- [21] Angelo M, Bendall SC, Finck R, Hale MB, Hitzman C, Borowsky AD, *et al.* Multiplexed ion beam imaging of human breast tumors. *Nature Medicine*. 2014; 20: 436–442.
- [22] Ståhl PL, Salmén F, Vickovic S, Lundmark A, Navarro JF, Magnusson J, *et al.* Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science (New York, N.Y.)*. 2016; 353: 78–82.
- [23] Libbrecht MW, Noble WS. Machine learning applications in genetics and genomics. *Nature Reviews. Genetics*. 2015; 16: 321–332.
- [24] Min S, Lee B, Yoon S. Deep learning in bioinformatics. *Briefings in Bioinformatics*. 2017; 18: 851–869.
- [25] Ballester PJ, Mitchell JBO. A machine learning approach to

- predicting protein-ligand binding affinity with applications to molecular docking. *Bioinformatics* (Oxford, England). 2010; 26: 1169–1175.
- [26] Spasić I, Livsey J, Keane JA, Nenadić G. Text mining of cancer-related information: review of current status and future directions. *International Journal of Medical Informatics*. 2014; 83: 605–623.
- [27] Shipp MA, Ross KN, Tamayo P, Weng AP, Kutok JL, Aguiar RCT, *et al.* Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. *Nature Medicine*. 2002; 8: 68–74.
- [28] Martelli M, Ferreri AJM, Agostinelli C, Di Rocco A, Pfreundschuh M, Pileri SA. Diffuse large B-cell lymphoma. *Critical Reviews in Oncology/hematology*. 2013; 87: 146–171.
- [29] Swerdlow SH, Campo E, Pileri SA, Harris NL, Stein H, Siebert R, *et al.* The 2016 revision of the World Health Organization classification of lymphoid neoplasms. *Blood*. 2016; 127: 2375–2390.
- [30] Wright G, Tan B, Rosenwald A, Hurt EH, Wiestner A, Staudt LM. A gene expression-based method to diagnose clinically distinct subgroups of diffuse large B cell lymphoma. *Proceedings of the National Academy of Sciences of the United States of America*. 2003; 100: 9991–9996.
- [31] Reddy A, Zhang J, Davis NS, Moffitt AB, Love CL, Waldrop A, *et al.* Genetic and Functional Drivers of Diffuse Large B Cell Lymphoma. *Cell*. 2017; 171: 481–494.e15.
- [32] Carey CD, Gusenleitner D, Chapuy B, Kovach AE, Kluk MJ, Sun HH, *et al.* Molecular classification of MYC-driven B-cell lymphomas by targeted gene expression profiling of fixed biopsy specimens. *The Journal of Molecular Diagnostics: JMD*. 2015; 17: 19–30.
- [33] Sesques P, Johnson NA. Approach to the diagnosis and treatment of high-grade B-cell lymphomas with MYC and BCL2 and/or BCL6 rearrangements. *Blood*. 2017; 129: 280–288.
- [34] Gisselbrecht C, Glass B, Mounier N, Singh Gill D, Linch DC, Trneny M, *et al.* Salvage regimens with autologous transplantation for relapsed large B-cell lymphoma in the rituximab era. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*. 2010; 28: 4184–4190.
- [35] Nowakowski GS, Blum KA, Kahl BS, Friedberg JW, Baizer L, Little RF, *et al.* Beyond RCHOP: A Blueprint for Diffuse Large B Cell Lymphoma Research. *Journal of the National Cancer Institute*. 2016; 108: djw257.
- [36] Zhou Z, Sehn LH, Rademaker AW, Gordon LI, Lacasce AS, Crosby-Thompson A, *et al.* An enhanced International Prognostic Index (NCCN-IPi) for patients with diffuse large B-cell lymphoma treated in the rituximab era. *Blood*. 2014; 123: 837–842.
- [37] Maurer MJ, Ghesquières H, Jais JP, Witzig TE, Haioun C, Thompson CA, *et al.* Event-free survival at 24 months is a robust end point for disease-related outcome in diffuse large B-cell lymphoma treated with immunochemotherapy. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*. 2014; 32: 1066–1073.
- [38] Hasin Y, Seldin M, Lusis A. Multi-omics approaches to disease. *Genome Biology*. 2017; 18: 83.
- [39] Chen R, Mias GI, Li-Pook-Than J, Jiang L, Lam HYK, Chen R, *et al.* Personal omics profiling reveals dynamic molecular and medical phenotypes. *Cell*. 2012; 148: 1293–1307.
- [40] Makohon-Moore A, Iacobuzio-Donahue CA. Pancreatic cancer biology and genetics from an evolutionary perspective. *Nature Reviews. Cancer*. 2016; 16: 553–565.
- [41] Mertins P, Mani DR, Ruggles KV, Gillette MA, Clauser KR, Wang P, *et al.* Proteogenomics connects somatic mutations to signalling in breast cancer. *Nature*. 2016; 534: 55–62.
- [42] Johnson CH, Ivanisevic J, Siuzdak G. Metabolomics: beyond biomarkers and towards mechanisms. *Nature Reviews. Molecular Cell Biology*. 2016; 17: 451–459.
- [43] Young RM, Shaffer AL, 3rd, Phelan JD, Staudt LM. B-cell receptor signaling in diffuse large B-cell lymphoma. *Seminars in Hematology*. 2015; 52: 77–85.
- [44] Young RM, Phelan JD, Wilson WH, Staudt LM. Pathogenic B-cell receptor signaling in lymphoid malignancies: New insights to improve treatment. *Immunological Reviews*. 2019; 291: 190–213.
- [45] Eken JA, Koning MT, Kupcova K, Sepúlveda Yáñez JH, de Groen RAL, Quinten E, *et al.* Antigen-independent, autonomous B cell receptor signaling drives activated B cell DLBCL. *The Journal of Experimental Medicine*. 2024; 221: e20230941.
- [46] Zhuang S, Yang Z, Cui Z, Zhang Y, Che F. Epigenetic alterations and advancement of lymphoma treatment. *Annals of Hematology*. 2024; 103: 1435–1454.
- [47] Bakhshi TJ, Georgel PT. Genetic and epigenetic determinants of diffuse large B-cell lymphoma. *Blood Cancer Journal*. 2020; 10: 123.
- [48] Vari F, Arpon D, Keane C, Hertzberg MS, Talaulikar D, Jain S, *et al.* Immune evasion via PD-1/PD-L1 on NK cells and monocyte/macrophages is more prominent in Hodgkin lymphoma than DLBCL. *Blood*. 2018; 131: 1809–1819.
- [49] Blenk S, Engelmann J, Weniger M, Schultz J, Dittrich M, Rosenwald A, *et al.* Germinal center B cell-like (GCB) and activated B cell-like (ABC) type of diffuse large B cell lymphoma (DLBCL): analysis of molecular predictors, signatures, cell cycle state and patient survival. *Cancer Informatics*. 2007; 3: 399–420.
- [50] Ren W, Wan H, Own SA, Berglund M, Wang X, Yang M, *et al.* Genetic and transcriptomic analyses of diffuse large B-cell lymphoma patients with poor outcomes within two years of diagnosis. *Leukemia*. 2024; 38: 610–620.
- [51] Gao HX, Nuerlan A, Abulajiang G, Cui WL, Xue J, Sang W, *et al.* Quantitative proteomics analysis of differentially expressed proteins in activated B-cell-like diffuse large B-cell lymphoma using quantitative proteomics. *Pathology, Research and Practice*. 2019; 215: 152528.
- [52] Guo Z, Wang C, Shi X, Wang Z, Tao J, Ma J, *et al.* Advances in proteomics in diffuse large B cell lymphoma (Review). *Oncology Reports*. 2024; 51: 87.
- [53] Deeb SJ, D'Souza RCJ, Cox J, Schmidt-Supprian M, Mann M. Super-SILAC allows classification of diffuse large B-cell lymphoma subtypes by their protein expression profiles. *Molecular & Cellular Proteomics: MCP*. 2012; 11: 77–89.
- [54] Carvalho AS, Baeta H, Henriques AFA, Ejtehadifar M, Tranfield EM, Sousa AL, *et al.* Proteomic Landscape of Extracellular Vesicles for Diffuse Large B-Cell Lymphoma Subtyping. *International Journal of Molecular Sciences*. 2021; 22: 11004.
- [55] Huang X, Shen Y, Liu M, Bi C, Jiang C, Iqbal J, *et al.* Quantitative proteomics reveals that miR-155 regulates the PI3K-AKT pathway in diffuse large B-cell lymphoma. *The American Journal of Pathology*. 2012; 181: 26–33.
- [56] Ejtehadifar M, Zahedi S, Gameiro P, Cabeçadas J, da Silva MG, Beck HC, *et al.* Meta-Analysis of MS-Based Proteomics Studies Indicates Interferon Regulatory Factor 4 and Nucleobindin1 as Potential Prognostic and Drug Resistance Biomarkers in Diffuse Large B Cell Lymphoma. *Cells*. 2023; 12: 196.
- [57] Mi M, Liu Z, Zheng X, Wen Q, Zhu F, Li J, *et al.* Serum metabolomic profiling based on GC/MS helped to discriminate Diffuse Large B-cell Lymphoma patients with different prognosis. *Leukemia Research*. 2021; 111: 106693.
- [58] Barberini L, Noto A, Fattuoni C, Satta G, Zucca M, Cabras MG, *et al.* The Metabolomic Profile of Lymphoma Subtypes: A Pilot Study. *Molecules* (Basel, Switzerland). 2019; 24: 2367.
- [59] Alfaifi A, Refai MY, Alsaadi M, Bahashwan S, Malhan H, Al-Kahiry W, *et al.* Metabolomics: A New Era in the Diagnosis

- or Prognosis of B-Cell Non-Hodgkin's Lymphoma. *Diagnostics* (Basel, Switzerland). 2023; 13: 861.
- [60] Rodriques SG, Stickels RR, Goeva A, Martin CA, Murray E, Vanderburg CR, *et al.* Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science* (New York, N.Y.). 2019; 363: 1463–1467.
- [61] Thrane K, Eriksson H, Maaskola J, Hansson J, Lundeberg J. Spatially Resolved Transcriptomics Enables Dissection of Genetic Heterogeneity in Stage III Cutaneous Malignant Melanoma. *Cancer Research*. 2018; 78: 5970–5979.
- [62] Lee JH, Daugharthy ER, Scheiman J, Kalhor R, Ferrante TC, Terry R, *et al.* Fluorescent in situ sequencing (FISSEQ) of RNA for gene expression profiling in intact cells and tissues. *Nature Protocols*. 2015; 10: 442–458.
- [63] Schapiro D, Jackson HW, Raghuraman S, Fischer JR, Zanutelli VRT, Schulz D, *et al.* histoCAT: analysis of cell phenotypes and interactions in multiplex image cytometry data. *Nature Methods*. 2017; 14: 873–876.
- [64] Vickovic S, Eraslan G, Salmén F, Klughammer J, Stenbeck L, Schapiro D, *et al.* High-definition spatial transcriptomics for in situ tissue profiling. *Nature Methods*. 2019; 16: 987–990.
- [65] Moncada R, Barkley D, Wagner F, Chiodin M, Devlin JC, Baron M, *et al.* Integrating microarray-based spatial transcriptomics and single-cell RNA-seq reveals tissue architecture in pancreatic ductal adenocarcinomas. *Nature Biotechnology*. 2020; 38: 333–342.
- [66] Davidson-Moncada JK, McDuffee E, Roschewski M. CD5+ diffuse large B-cell lymphoma with hemophagocytosis. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*. 2013; 31: e76–e79.
- [67] Alzubi J, Nayyar A, Kumar A. Machine learning from theory to algorithms: an overview. *Journal of Physics: Conference Series*. IOP Publishing. 2018; 1142: 012012.
- [68] Huang S, Cai N, Pacheco PP, Narrandes S, Wang Y, Xu W. Applications of Support Vector Machine (SVM) Learning in Cancer Genomics. *Cancer Genomics & Proteomics*. 2018; 15: 41–51.
- [69] Biau G. Analysis of a random forests model. *The Journal of Machine Learning Research*. 2012; 13: 1063–1095.
- [70] Sinaga KP, Yang MS. Unsupervised K-means clustering algorithm. *IEEE Access*. 2020; 8: 80716–80727.
- [71] Shlens J. A tutorial on principal component analysis. *arXiv*. 2014. (preprint)
- [72] Lewis FL, Vrabie D, Vamvoudakis KG. Reinforcement learning and feedback control: Using natural decision methods to design optimal adaptive controllers. *IEEE Control Systems Magazine*. 2012; 32: 76–105.
- [73] Ching T, Himmelstein DS, Beaulieu-Jones BK, Kalinin AA, Do BT, Way GP, *et al.* Opportunities and obstacles for deep learning in biology and medicine. *Journal of the Royal Society, Interface*. 2018; 15: 20170387.
- [74] Lee J, Yoon W, Kim S, Kim D, Kim S, So CH, *et al.* BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics* (Oxford, England). 2020; 36: 1234–1240.
- [75] Luo R, Sun L, Xia Y, Qin T, Zhang S, Poon H, *et al.* BioGPT: generative pre-trained transformer for biomedical text generation and mining. *Briefings in Bioinformatics*. 2022; 23: bbac409.
- [76] Rafique R, Islam SMR, Kazi JU. Machine learning in the prediction of cancer therapy. *Computational and Structural Biotechnology Journal*. 2021; 19: 4003–4017.
- [77] Zhao C, Xu Z, Wang X, Tao S, MacDonald WA, He K, *et al.* Innovative super-resolution in spatial transcriptomics: a transformer model exploiting histology images and spatial gene expression. *Briefings in Bioinformatics*. 2024; 25: bbae052.
- [78] Le NQK, Ho QT, Nguyen TTD, Ou YY. A transformer architecture based on BERT and 2D convolutional neural network to identify DNA enhancers from sequence information. *Briefings in Bioinformatics*. 2021; 22: bbab005.
- [79] Blanc-Durand P, Jégou S, Kanoun S, Berriolo-Riedinger A, Bodet-Milin C, Kraeber-Bodéré F, *et al.* Fully automatic segmentation of diffuse large B cell lymphoma lesions on 3D FDG-PET/CT for total metabolic tumour volume prediction using a convolutional neural network. *European Journal of Nuclear Medicine and Molecular Imaging*. 2021; 48: 1362–1370.
- [80] Gan Y, Hu X, Zou G, Yan C, Xu G. Inferring Gene Regulatory Networks From Single-Cell Transcriptomic Data Using Bidirectional RNN. *Frontiers in Oncology*. 2022; 12: 899825.
- [81] Peng H, Su M, Guo X, Shi L, Lei T, Yu H, *et al.* Artificial intelligence-based prognostic model accurately predicts the survival of patients with diffuse large B-cell lymphomas: analysis of a large cohort in China. *BMC Cancer*. 2024; 24: 621.
- [82] Mamoshina P, Vieira A, Putin E, Zhavoronkov A. Applications of Deep Learning in Biomedicine. *Molecular Pharmaceutics*. 2016; 13: 1445–1454.
- [83] Lipkova J, Chen RJ, Chen B, Lu MY, Barbieri M, Shao D, *et al.* Artificial intelligence for multimodal data integration in oncology. *Cancer Cell*. 2022; 40: 1095–1110.
- [84] Leng D, Zheng L, Wen Y, Zhang Y, Wu L, Wang J, *et al.* A benchmark study of deep learning-based multi-omics data fusion methods for cancer. *Genome Biology*. 2022; 23: 171.
- [85] Yang Y, Sun H, Zhang Y, Zhang T, Gong J, Wei Y, *et al.* Dimensionality reduction by UMAP reinforces sample heterogeneity analysis in bulk transcriptomic data. *Cell Reports*. 2021; 36: 109442.
- [86] Johnson KB, Wei WQ, Weeraratne D, Frisse ME, Misulis K, Rhee K, *et al.* Precision Medicine, AI, and the Future of Personalized Health Care. *Clinical and Translational Science*. 2021; 14: 86–93.
- [87] Glaab E, Rauschenberger A, Banzi R, Gerardi C, Garcia P, Demotes J. Biomarker discovery studies for patient stratification using machine learning analysis of omics data: a scoping review. *BMJ Open*. 2021; 11: e053674.
- [88] Hasani N, Paravastu SS, Farhadi F, Yousefirizi F, Morris MA, Rahmim A, *et al.* Artificial Intelligence in Lymphoma PET Imaging:: A Scoping Review (Current Trends and Future Directions). *PET Clinics*. 2022; 17: 145–174.
- [89] Walter W, Pohlkamp C, Meggendorfer M, Nadarajah N, Kern W, Haferlach C, *et al.* Artificial intelligence in hematological diagnostics: Game changer or gadget? *Blood Reviews*. 2023; 58: 101019.
- [90] Yan F, Da Q, Yi H, Deng S, Zhu L, Zhou M, *et al.* Artificial intelligence-based assessment of PD-L1 expression in diffuse large B cell lymphoma. *NPJ Precision Oncology*. 2024; 8: 76.
- [91] Li D, Bledsoe JR, Zeng Y, Liu W, Hu Y, Bi K, *et al.* A deep learning diagnostic platform for diffuse large B-cell lymphoma with high accuracy across multiple hospitals. *Nature Communications*. 2020; 11: 6004.
- [92] Lee JH, Song GY, Lee J, Kang SR, Moon KM, Choi YD, *et al.* Prediction of immunochemotherapy response for diffuse large B-cell lymphoma using artificial intelligence digital pathology. *The Journal of Pathology. Clinical Research*. 2024; 10: e12370.
- [93] Ferrández MC, Golla SSV, Eertink JJ, de Vries BM, Lugtenburg PJ, Wieggers SE, *et al.* An artificial intelligence method using FDG PET to predict treatment outcome in diffuse large B cell lymphoma patients. *Scientific Reports*. 2023; 13: 13111.
- [94] Shankar V, Yang X, Krishna V, Tan B, Silva O, Rojansky R, *et al.* LymphoML: An interpretable artificial intelligence-based method identifies morphologic features that correlate with lymphoma subtype. In *Machine Learning for Health (ML4H)* (pp. 528–558). PMLR. 2023.
- [95] Carreras J, Nakamura N. Artificial Intelligence, Lymphoid Neo-

- plasms, and Prediction of MYC, BCL2, and BCL6 Gene Expression Using a Pan-Cancer Panel in Diffuse Large B-Cell Lymphoma. *Hemato*. 2024; 5: 119–143.
- [96] Rappoport N, Shamir R. Multi-omic and multi-view clustering algorithms: review and cancer benchmark. *Nucleic Acids Research*. 2018; 46: 10546–10562.
- [97] Lopez R, Regier J, Cole MB, Jordan MI, Yosef N. Deep generative modeling for single-cell transcriptomics. *Nature Methods*. 2018; 15: 1053–1058.
- [98] Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, *et al*. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discovery*. 2012; 2: 401–404.
- [99] Jiang P, Gu S, Pan D, Fu J, Sahu A, Hu X, *et al*. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nature Medicine*. 2018; 24: 1550–1558.
- [100] You Y, Lai X, Pan Y, Zheng H, Vera J, Liu S, *et al*. Artificial intelligence in cancer target identification and drug discovery. *Signal Transduction and Targeted Therapy*. 2022; 10: 156.
- [101] Bolouri H, Farrar JE, Triche T, Jr, Ries RE, Lim EL, Alonzo TA, *et al*. The molecular landscape of pediatric acute myeloid leukemia reveals recurrent structural alterations and age-specific mutational interactions. *Nature Medicine*. 2018; 24: 103–112.
- [102] Kang M, Ko E, Mersha TB. A roadmap for multi-omics data integration using deep learning. *Briefings in Bioinformatics*. 2022; 23: bbab454.
- [103] Subramanian I, Verma S, Kumar S, Jere A, Anamika K. Multi-omics Data Integration, Interpretation, and Its Application. *Bioinformatics and Biology Insights*. 2020; 14: 1177932219899051.
- [104] Canzler S, Schor J, Busch W, Schubert K, Rolle-Kampczyk UE, Seitz H, *et al*. Prospects and challenges of multi-omics data integration in toxicology. *Archives of Toxicology*. 2020; 94: 371–388.
- [105] Graw S, Chappell K, Washam CL, Gies A, Bird J, Robeson MS, 2nd, *et al*. Multi-omics data integration considerations and study design for biological systems and disease. *Molecular Omics*. 2021; 17: 170–185.
- [106] Picard M, Scott-Boyer MP, Bodein A, Périn O, Droit A. Integration strategies of multi-omics data for machine learning analysis. *Computational and Structural Biotechnology Journal*. 2021; 19: 3735–3746.
- [107] Leek JT, Scharpf RB, Bravo HC, Simcha D, Langmead B, Johnson WE, *et al*. Tackling the widespread and critical impact of batch effects in high-throughput data. *Nature Reviews. Genetics*. 2010; 11: 733–739.
- [108] Vandereyken K, Sifrim A, Thienpont B, Voet T. Methods and applications for single-cell and spatial multi-omics. *Nature Reviews. Genetics*. 2023; 24: 494–515.
- [109] Wang Z, Czerniak B, Wei P. Spatial IMIX: A Mixture Model Approach to Spatially Correlated Multi-Omics Data Integration. *bioRxiv*. 2023. (preprint)
- [110] Lin S, Zhao F, Wu Z, Yao J, Zhao Y, Yuan Z. Streamlining spatial omics data analysis with Pysodb. *Nature Protocols*. 2024; 19: 831–895.
- [111] Ioannidis JPA, Allison DB, Ball CA, Coulibaly I, Cui X, Culhane AC, *et al*. Repeatability of published microarray gene expression analyses. *Nature Genetics*. 2009; 41: 149–155.
- [112] Holzinger A, Biemann C, Pattichis CS, Kell DB. What do we need to build explainable AI systems for the medical domain?. *arXiv*. 2017. (preprint)
- [113] Pan SJ, Yang Q. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*. 2009; 22: 1345–1359.
- [114] Char DS, Shah NH, Magnus D. Implementing Machine Learning in Health Care - Addressing Ethical Challenges. *The New England Journal of Medicine*. 2018; 378: 981–983.
- [115] Cocco N, Anelli L, Zagaria A, Perrone T, Specchia G, Albano F. Molecular Complexity of Diffuse Large B-Cell Lymphoma: Can It Be a Roadmap for Precision Medicine? *Cancers*. 2020; 12: 185.
- [116] Morin RD, Arthur SE, Hodson DJ. Molecular profiling in diffuse large B-cell lymphoma: why so many types of subtypes? *British Journal of Haematology*. 2022; 196: 814–829.
- [117] Hilton LK, Scott DW, Morin RD. Biological heterogeneity in diffuse large B-cell lymphoma. *Seminars in Hematology*. 2023; 60: 267–276.
- [118] Liu Y, Barta SK. Diffuse large B-cell lymphoma: 2019 update on diagnosis, risk stratification, and treatment. *American Journal of Hematology*. 2019; 94: 604–616.
- [119] Hood L, Friend SH. Predictive, personalized, preventive, participatory (P4) cancer medicine. *Nature Reviews. Clinical Oncology*. 2011; 8: 184–187.