



REVIEW OPEN ACCESS

Current Bioinformatics Tools in Precision Oncology

Tesfaye Wolde¹ | Vipul Bhardwaj² | Vijay Pandey^{1,2}¹Institute of Biopharmaceutical and Health Engineering, Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen, China |²Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen, ChinaCorrespondence: vijay.pandey@sz.tsinghua.edu.cn

Received: 12 December 2024 | Revised: 5 May 2025 | Accepted: 6 May 2025

Funding: This work was supported by the Department of Chemical Engineering-Institute of Biopharmaceutical and Health Engineering, Special Collaboration Joint Fund Project (010201000012021), Tsinghua Shenzhen International Graduate School, Tsinghua University, China; and the National Natural Science Foundation of China (Grant Nos. 81872368 and 81641051) and the Shenzhen Development and Reform Commission Subject Construction Project ([2017] 1434).

Keywords: bioinformatics | biomarkers | multiomic | oncotherapeutics | precision oncology

ABSTRACT

Integrating bioinformatics tools has profoundly transformed precision oncology by identifying essential molecular targets for personalized treatment. The rapid development of high-throughput sequencing and multiomics technologies creates complex datasets that require robust computational methods to extract meaningful insights. Nonetheless, the clinical application of multiomics data continues to pose significant challenges. This review explores advanced bioinformatics tools utilized within multiomics, emphasizing their pivotal role in discovering cancer biomarkers. Cloud-based platforms, such as Galaxy and DNAnexus, facilitate streamlined data processing, while single-cell analysis software, including Seurat, identifies rare cellular subpopulations. Further integration of artificial intelligence with machine learning approaches improves predictive modeling and diagnostic accuracy. Spatial omics technologies correlate molecular signatures within tumor microenvironments, guiding treatment strategies. Bioinformatics integrates these technologies to establish a new standard in precision oncology, thereby enhancing therapy efficacy. Collaborative initiatives between The Cancer Genome Atlas and cBioPortal expedite advancements through the sharing open data and implementing standardized methodologies. Advancing multiomics integration techniques alongside improved computational capabilities is essential for discovering new biomarkers and refining precision medicine strategies. Future efforts should focus on merging multiomics techniques with innovative computational methods to drive novel biomarker discovery and improve precision medicine applications.

1 | Introduction

Oncology is swiftly evolving from a generic, one-size-fits-all treatment model to a personalized approach rooted in precision medicine [1, 2]. This evolution is driven by advancements in molecular biology, high-throughput sequencing, and computational tools that help integrate complex multiomics data effectively [3]. Precision oncology aims to customize treatments

for individual patients, similar to how fingerprints reflect genetic, epigenetic, and environmental identities, enabling personalized strategies [4]. This approach is central to identifying and validating biomarkers that signify measurable events associated with cancer onset, progression, and therapeutic response. Biomarkers can arise from various sources, including tumor tissues, blood, and other bodily fluids, encompassing DNA, RNA, proteins, and metabolites [5–8]. Leveraging these biomarkers can

Tesfaye Wolde and Vipul Bhardwaj contributed equally to this work.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2025 The Author(s). *MedComm* published by Sichuan International Medical Exchange & Promotion Association (SCIMEA) and John Wiley & Sons Australia, Ltd.

significantly improve patient outcomes through early diagnosis, risk assessment, treatment selection, and disease monitoring. For instance, specific mutations in the EGFR gene are used as indicators for targeted therapies in non-small cell lung cancer (NSCLC), guiding the use of EGFR inhibitors [9, 10]. Moreover, epigenetic biomarkers, which influence gene expression without altering the underlying DNA sequence, play a crucial role in cancer development [11]. These include epigenetic modifications such as noncoding RNA profiles, histone changes, and DNA methylation patterns that can act as cancer biomarkers. Notably, hypermethylation and silencing of tumor suppressor gene promoters, like *MLH1* in various cancers, may drive tumor growth and serve as potential early indicators for diagnosis or response to therapy, assessing PD-L1 expression levels is also essential for determining candidacy for immunotherapy [12, 13]. Other currently used biomarkers include breast cancer gene 1/2 (*BRCA 1/2*) mutations, which signal suitability for PARP inhibitor treatment [14]. However, the discovery and validation of reliable biomarkers pose significant challenges due to the intricate nature of cancer biology and variability within and among tumors.

Recent advancements in bioinformatics have rapidly transformed precision oncology, leading to patient-centered treatment strategies based on molecular biomarkers [15]. Bioinformatics, which intersects biology, computer science, and mathematics for biological data analysis, has gained medical relevance in cancer research with the introduction of high-throughput techniques like next-generation sequencing (NGS), microarrays, and proteomics [16, 17]. The enormous data generated by these technologies can be overwhelming without suitable analytical methods. Bioinformatics tools help uncover patterns, correlations, and anomalies that could signify potential biomarkers. For instance, Wolde et al. [18] applied an integrated bioinformatics strategy to identify a novel signature of nine *immune-related genes* as potential biomarkers and therapeutic targets in ovarian carcinoma. In another study, Zhao et al. [19] discovered and validated a seven-gene signature (*AFAP1L2*, *CAMK1D*, *LOXL2*, *PIK3CG*, *PLEKHG1*, *RARRES2*, and *SPP1*) for prognosis stratification in advanced lung adenocarcinoma patients, discussing its potential to predict survival outcomes. Additionally, Snijesh et al. [20] utilized data from The Cancer Genome Atlas (TCGA) to categorize endometrial cancer (EC) tumors by sonic hedgehog (SHH) pathway activation. They found that high SHH tumors display a less aggressive phenotype, lower mutational burden, and improved survival outcomes, underlining the prognostic importance of SHH signaling in EC [20]. The integration of bioinformatics methods, statistical modeling, and network analysis could deepen our understanding of cancer biology and lead to the development of predictive tools.

In precision oncology, biomarker discovery relies on an array of bioinformatics tools that manage and analyze complex data. For example, data pipelines and specialized software are used to identify *differentially expressed genes* (DEGs), predict patient outcomes, and simulate treatment responses [16]. Genomic analysis toolkits such as the Genome Analysis Toolkit (GATK), Spliced Transcript Alignment to a Reference (STAR), and HISAT2 work together to process sequencing data, while DESeq2 and EdgeR focus on detecting differential gene expression in RNA sequencing (RNA-seq) [21]. Additionally, proteomic analysis tools like MaxQuant and Proteome Discoverer facilitate the quantification

and identification of proteins to uncover potential molecular biomarkers [22]. Integrative platforms, including cBioPortal and OncoPrint, combine multiomic datasets, providing a comprehensive perspective on tumor biology and aiding researchers in their search for promising biomarkers across various tumors [23, 24].

Advancements in biomarker identification methods and the growth of extensive data repositories have increased data complexity and volume, necessitating sophisticated analytical techniques [25]. This has increased the need for machine learning (ML) and predictive algorithms. ML algorithms are designed to handle and analyze large, high-dimensional datasets, and molecular networks in network medicine, revealing patterns and relationships often overlooked by traditional methods [26]. With the expansion of artificial intelligence (AI), various ML frameworks, including Python's scikit-learn, TensorFlow, and Keras, are now utilized for predictive modeling in oncology [27]. These tools can examine historical patient data to forecast outcomes and treatment responses based on recognized biomarkers. Additionally, network and pathway analysis tools like STRING and Cytoscape investigate molecular interactions and frequently regulated biological pathways that connect and influence tumor behavior through biomarkers [28]. Clinical data integration is also experiencing growth, with software developed to merge clinical data with molecular profiles. Platforms such as REDCap and OpenClinica facilitate collecting and analyzing clinical outcomes in conjunction with genomic and proteomic data [29].

Although advancements in bioinformatics tools have accelerated biomarker discovery, several challenges remain, including data acquisition, reproducibility, quality control, interoperability among various platforms, and inconsistent reporting concerns [30]. A multidisciplinary approach that includes clinical and ethical expertise is essential for effectively applying bioinformatics tools in diagnosis, preventive medicine, and personalized therapeutic strategies. The extensive and complex datasets generated by high-throughput technologies, particularly NGS methods, necessitate the use of bioinformatics tools for their analysis [31]. The adoption of bioinformatics and related tools is transforming precision oncology by enabling swift, comprehensive genomic data analysis, evaluating disease mechanisms, and facilitating the identification of potential biomarkers for customized therapies. Moreover, the complexity and diversity of cancer present significant obstacles to identifying universally applicable biomarkers. Nonetheless, these challenges create opportunities for continued innovation in bioinformatics, which can radically enhance the analytical process. Progress in AI and ML predictive algorithms is revealing patterns and managing large datasets that exceed human analytical capabilities. Additionally, the increasing focus on multiomics approaches, integrating genomics, transcriptomics, proteomics, and metabolomics, provides a comprehensive perspective for understanding the fundamental mechanisms of cancer [32]. Collectively, these advancements hold the potential to address existing challenges and limitations in bioinformatics, transforming complex data into actionable strategies for precision-driven care.

This review provides an in-depth overview of how bioinformatics tools integrate and transform biomarker discovery and therapeutic strategies in precision oncology. We discuss their ability to identify molecular targets through the integration and

analysis of multiomics data. Additionally, we highlight several significant bioinformatics platforms, including AI and ML-driven predictive tools, as well as emerging technologies like spatial omics, and discuss their roles in cancer diagnosis, prognosis, and treatment optimization. Furthermore, we address challenges related to data interpretation and clinical translation, along with ethical implications, stressing the importance of interdisciplinary collaboration and open data-sharing initiatives. This review aims to bridge bioinformatics and clinical oncology, creating new opportunities to utilize these computational advancements for enhancing patient-specific therapeutic strategies and achieving success in precision cancer medicine.

2 | Types of Omics Data in Biomarker Discovery

The discovery and validation of biomarkers in precision oncology primarily relies on omics technologies, which facilitate the comprehensive analysis of biological molecules across diverse contexts [33]. The term “Omics” refers to various biological studies aimed at the extensive characterization of biomolecules, such as genomes, transcriptomes, proteomes, and metabolomes (Figure 1). The various types of omics, which include genomics, transcriptomics, proteomics, epigenomics, and metabolomics data, offer complementary insights that enhance our understanding of cancer biology, help identify potential biomarkers, and aid in patient stratification populations [34]. As technological advancements progress, omics technologies are set to play an increasingly vital role in precision oncology, raising innovation in biomarker discovery and personalized medicine. This section offers a comprehensive summary of the different omics data types and the bioinformatics tools utilized for their analysis.

2.1 | Genomics

Genomics constitutes a comprehensive examination of an organism’s genome, involving the sequencing, mapping, and analysis of its DNA [35]. Advancements in technology related to DNA sequencing, particularly NGS, facilitate extensive analysis of entire genomes or specific regions, thereby enabling the identification of mutations such as single-nucleotide polymorphisms (SNPs), insertions, deletions, copy number variations (CNVs), and structural modifications that are crucial to the development of cancer [36]. Techniques including whole-genome sequencing (WGS), single-cell DNA sequencing, and targeted gene panels empower researchers to precisely identify mutations that can inform therapeutic strategies. For instance, mutations in genes such as KRAS, BRAF, and TP53 correlate with specific types of cancer and therapeutic responses, underscoring the clinical significance of genomic profiling [37, 38]. The intricate nature of genomic data requires sophisticated bioinformatics support for comprehensive analysis. Specialized computational platforms are indispensable for the interpretation of variants. Bioinformatics tools, such as GATK, are recognized as leading solutions for the detection of SNPs and small insertions and deletions, while MuTect specializes in identifying somatic mutations in paired tumor-normal samples [39, 40]. ANNOVAR enhances these methodologies by offering detailed variant annotations based on functional consequences [41]. Moreover, single-cell genomics tools, such as Cell Ranger, have revolutionized gene expression

analysis at the individual cell level, essential for understanding cancer heterogeneity and evolution [42]. Tools like MORPHEUS and Monocle 3 facilitate interactive clustering of gene expression, lineage tracing, and trajectory analysis, all of which are vital for unraveling cancer progression and cellular differentiation [43, 44]. By integrating these bioinformatics instruments to convert raw genetic data from DNA sequencing techniques into meaningful clinical insights, researchers can uncover the prognostic and predictive value of specific genetic alterations, ultimately advancing personalized approaches in oncology.

2.2 | Transcriptomics

Transcriptomics provides an ongoing perspective on gene expression in cells, offering vital insights into the molecular mechanisms underlying cancer development and progression [45]. Transcriptomics focuses on analyzing RNA transcripts, such as mRNA, noncoding RNA, and miRNA, to assess gene expression variances between cancerous and healthy tissues. Advanced technologies like RNA-seq and microarrays have reinvented gene expression profiling, revealing patterns associated with different tumor types, stages, and responses to treatment [46, 47]. For example, the high and low expression of certain oncogenes and tumor suppressor genes respectively can serve as diagnostic or prognostic biomarkers, enhancing our understanding and potentially aiding in mitigating disease progression [48, 49].

Bioinformatics tools play a vital role in analyzing transcriptomic data [50]. Differential gene expression analysis is key for identifying genes that are either upregulated or downregulated in cancer, supported by statistical model-based tools such as DESeq2 and EdgeR that work with count-based RNA-seq data [21, 51]. While limma was originally created for microarray research, it is now commonly used in RNA-seq to uncover significant expression patterns [52]. Multiomics integration platforms like Sangerbox 3, MOFA+, and MixOmics improve the visualization of relationships between datasets [53–55]. For single-cell RNA-seq analysis, tools such as Scanpy and Seurat aid in clustering and in-depth examination of expression variations, providing insights on the cellular heterogeneity of cancer [56–58]. Additionally, platforms like pyBioPortal and cBioPortal enable the integration of transcriptomic data with clinical and genomic information, enhancing the overall understanding of cancer biology [59, 60].

For RNA-seq quantification, Salmon and Kallisto utilize ultra-fast, alignment-free techniques to estimate precise transcript numbers, while STAR and HTSeq offer alignment and read counting for gene expression studies [61–64]. In addition, single-cell-specific tools like CellBender improve data quality by filtering out ambient RNA, and deep learning models such as scVI support imputation and clustering analyses [65, 66]. Furthermore, Dyngen and scRNASeqDB have advanced single-cell research methods which facilitate gene expression data simulation and provide dynamic platforms for cross-cancer exploration [67–69]. Tools like scissor link single-cell RNA-seq data to clinical outcomes, integrating cellular insights with survival rates and treatment responses, thus paving the way for translational oncology applications [70]. As technological advancements progress, transcriptomics leads the way in precision oncology, presenting

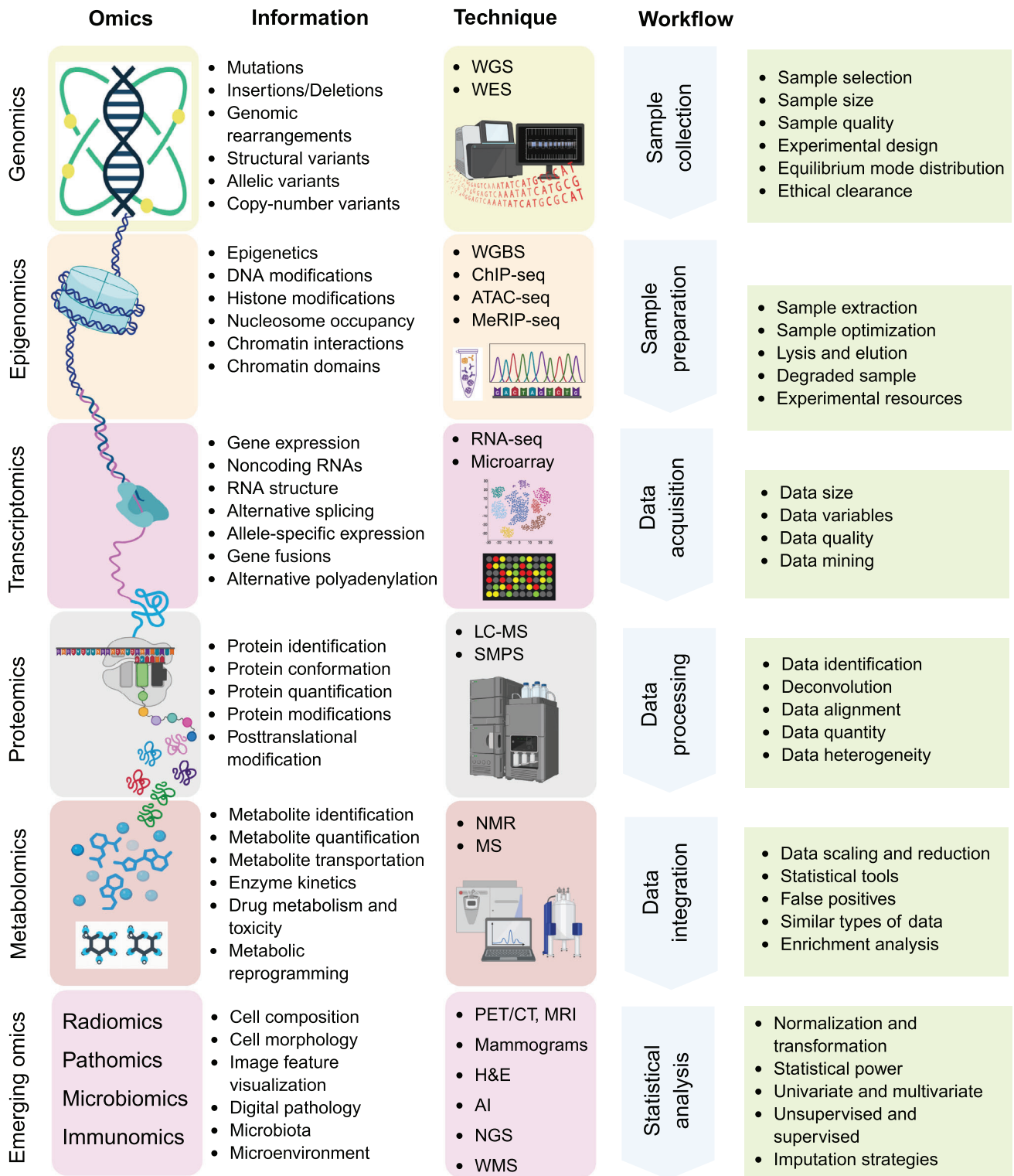


FIGURE 1 | This comprehensive figure displays the complex landscape of omics technologies and their roles in biological research. Multiomics technologies have emerged to profile genome sequences, epigenetic features, gene expression, protein levels, metabolite abundances, and more. The illustration clarifies the types of insights offered by omics, such as genetic variations and epigenetic changes. It also details essential techniques like whole-genome sequencing (WGS), whole-exome sequencing (WES), and advanced methods including whole genome bisulfite sequencing (WGBS), chromatin immunoprecipitation sequencing (ChIP-seq), and assay for transposase-accessible chromatin using sequencing (ATAC-seq) for examining chromatin dynamics. Additionally, it emphasizes the workflow from sample collection and preparation to data acquisition, pointing out challenges related to sample quality, experimental design, and ethical issues. The figure also addresses difficulties faced during the workflow. Last, it presents the emerging omics and the involvement of AI in deciphering complex datasets, highlighting the significance of multiomics integration in promoting cancer research and personalized medicine. The figure is generated using BioRender.com. NGS: next-generation sequencing, ChIP-seq: chromatin immunoprecipitation sequencing, ATAC-seq: assay for transposase-accessible chromatin sequencing, RNA-seq: RNA sequencing, LC-MS: liquid chromatography-mass spectrometry, SIMP: stable isotope mass profiling, 2D-DIGE: two-dimensional difference gel electrophoresis, SRM: selected

exceptional opportunities to decode cancer's molecular complexities and create targeted therapeutic interventions.

2.3 | Proteomics

Proteomics is the comprehensive study of all proteins generated by a genome and plays a crucial role in biomedical research, notably in precision oncology [71]. Unlike genomics and transcriptomics, which infer gene expression mainly from genetic sequences, proteomics seeks to reveal the functional consequences of genes: specifically, at the protein level that affect cellular operations and define phenotypes [72]. Additionally, proteomics facilitates the direct quantification of protein levels and critical posttranslational modifications (PTMs) such as phosphorylation, glycosylation, and ubiquitination, offering valuable insights into the functional status of proteins linked to cancerous changes. Proteins discovered through proteomic analyses can provide important information about tumor characteristics, patient outlook, and possible treatment responses. For example, well-known biomarkers such as CA-125 and PSA are used in ovarian and prostate cancers, respectively [73, 74]. By integrating proteomic assessments into clinical workflows, healthcare providers can modify treatment plans according to a patient's molecular profile, ultimately improving treatment effectiveness and patient outcomes.

Recent advancements in mass spectrometry (MS) and protein microarray technologies have positioned proteomics at the forefront of scientific inquiry [75]. Techniques such as liquid chromatography–tandem mass spectrometry (LC–MS/MS) are widely utilized for high-throughput protein analysis [76]. These sophisticated methods possess the capability to detect thousands of proteins within complex biological samples and quantify their abundance, thereby facilitating comprehensive tumor proteome profiling. Protein microarrays, an additional critical tool in the field of proteomics, allow for the simultaneous investigation of multiple proteins and their interactions with various molecules [77]. By presenting arrays of known proteins on a solid substrate, this technology enables high-throughput screening of protein expression, interactions, and functions in an exceptionally efficient and informative manner.

To effectively utilize proteomics in oncology, it is essential to have robust bioinformatics tools for managing and analyzing the vast data generated by these studies. As the field advances, computational platforms that support biomarker discovery have become increasingly important. Software like MaxQuant and Skyline enables precise quantification of protein levels and PTMs, allowing for comparative analyses between tumor and normal tissues or between sensitive and resistant tumors [78, 79]. Moreover, tools like Proteome Discoverer and PeptideShaker streamline

workflows for protein identification, while OpenMS provides an open-source solution for LC–MS data analysis [80–83]. For exploring protein interactions, tools such as STRING, Cytoscape, and BioGRID aid in mapping complex protein networks, and PhosphoSitePlus offers a comprehensive catalog of verified PTMs [84–87]. Structural prediction software, including AlphaFold and I-TASSER, reveals insights into protein function, supported by visualization tools like Pymol [88–90]. High-throughput tools such as MS-DIAL and DIA-NN, alongside ML applications like DeepNovo, improve peptide identification and sequencing precision [91–93]. Additionally, statistical analysis platforms like Perseus and limma/EdgeR in R allow for detailed interpretation of proteomic data, ensuring its biological relevance while minimizing false discoveries [94, 95].

The integration of these advanced tools facilitates a substantial enhancement in our comprehension of the molecular foundations of cancer, concurrently contributing to the development of molecular biomarkers, targeted therapies, and personalized treatment strategies.

2.4 | Epigenomics

Cancer develops not solely from genetic mutations but also from significant epigenetic alterations that impact gene expression and cellular behaviors, including uncontrolled growth, invasion, and metastasis [96]. For instance, the hypermethylation of tumor suppressor genes may reduce their expression, whereas hypomethylation can potentially activate oncogenes, thereby contributing to malignancy. These epigenetic processes are essential for the identification of biomarkers that can inform personalized treatment strategies. Epigenomics, which investigates epigenetic modifications of genetic material, represents an emerging field in precision oncology [97]. In contrast to genomics, which focuses on DNA sequences, epigenomics examines how alterations such as DNA methylation, histone modifications, and noncoding RNA interactions influence gene expression. Importantly, epigenetic changes are often reversible, rendering them promising targets for therapeutic intervention. Currently, a variety of epigenetic modulators capable of altering these changes are undergoing evaluation in clinical trials, emphasizing the importance of epigenomics in cancer care [98].

Recent advances in high-throughput sequencing technologies have revolutionized epigenomic research [99]. Methods like whole-genome bisulfite sequencing (WGBS), chromatin immunoprecipitation sequencing (ChIP-seq), and RNA-seq are frequently used to map DNA methylation patterns, histone modifications, and interactions with noncoding RNAs [100–103]. For example, WGBS provides a comprehensive perspective on DNA methylation across the genome by distinguishing between

reaction monitoring, NMR: nuclear magnetic resonance, AI: artificial intelligence, ML: machine learning, WGCNA: weighted gene coexpression network analysis, GSEA: gene set enrichment analysis, TCGA: the cancer genome atlas, GEO: Gene Expression Omnibus, DDBJ: DNA Data Bank of Japan, GEPIA: gene expression profiling interactive analysis, GenBank: genetic sequence database, DNAnexus: cloud-based platform for data analysis, COSMIC: catalogue of somatic mutations in cancer, CA-125: cancer antigen 125, KRAS G12V: mutation of the *KRAS* gene at codon 12 (glycine to valine), FGFR3: fibroblast growth factor receptor 3, MGMT: O-6-methylguanine-DNA methyltransferase, TP53: tumor protein 53, microRNA: small, noncoding RNA molecules (e.g., *miR-21*, *miR-155*), *lncRNA*: long noncoding RNA, PVT1: plasmacytoma variant translocation 1.

methyated and unmethyated cytosines, thereby revealing methylation patterns associated with various cancers. This technique is crucial for understanding the epigenomic landscape of tumors and for identifying methylation signatures that may serve as biomarkers. ChIP-seq enables researchers to investigate histone modifications and transcription factor binding sites throughout the genome, offering vital insights into how these changes affect gene expression in cancer. This approach can uncover regulatory elements involved in cancer progression, helping to link epigenetic changes with clinical outcomes. Moreover, the role of noncoding RNAs, including microRNAs and long noncoding RNAs (lncRNAs), has gained significant attention in epigenomics [104]. These molecules regulate gene expression and chromatin structure, which significantly impact tumor biology [104–109]. Moreover, RNA-seq has furthered the profiling of noncoding RNAs in various cancers, emphasizing their potential as biomarkers or therapeutic targets [110].

Bioinformatics tools are essential for analyzing the large datasets generated by epigenomic research. Bismark is often used for DNA methylation analysis, aligning sequencing data and identifying differential methylation regions at a single-base level [111]. MethylKit and DSS are popular for detecting methylation changes under different conditions in differential analysis [112, 113]. For ChIP-seq and ATAC-seq data, MACS is a favored tool for detecting peaks in chromatin accessibility, while SICER is particularly effective at revealing larger enrichment areas in histone modification studies [114, 115]. Tools like DiffReps and ChromHMM provide comprehensive analyses of chromatin states and accessibility across genomic regions, emphasizing the epigenetic patterns critical for understanding cancer biology [116, 117]. To separate meaningful biological signals from background noise, researchers frequently employ limma and EdgeR, R-based packages that assist in identifying significant epigenetic changes between cancerous and normal tissues [118, 119].

The integration of epigenomic data with bioinformatics creates numerous opportunities for identifying biomarkers in precision oncology. Epigenetic profiling has already revealed several promising biomarkers, including specific methylation patterns linked to certain cancer types and clinical outcomes [120]. For example, the methylation status of genes such as RASSF1A and GSTP1 has been investigated as a potential biomarker for diagnosing prostate cancer [121]. Additionally, epigenomic biomarkers can help in making therapeutic choices [122]. Genes exhibiting abnormal methylation can support the selection of patients for epigenetic treatments, while biomarkers related to drug response and resistance can inform personalized treatment plans based on an individual's epigenetic profile.

2.5 | Metabolomics

Cancer is frequently characterized by altered metabolic pathways, particularly exemplified by the Warburg effect, wherein tumor cells predominantly favor glycolysis followed by lactic acid fermentation for energy production, even in the presence of oxygen [123]. This modification in metabolic processes facilitates enhanced cellular proliferation and contributes to tumor progression and metastasis. Metabolomics, defined as the comprehensive analysis of metabolites within biological systems,

yields significant insights into the biochemical processes driving cancer biology [124]. Given that metabolites represent the end products of cellular activities, they reflect the physiological conditions of cells and tissues, thereby revealing critical information regarding metabolic dysregulation in cancer. Within the field of precision oncology, where treatments are tailored to individual patient profiles, metabolomics proves indispensable for identifying biomarkers [125]. Advancements in metabolomics and bioinformatics have already discovered several promising cancer biomarkers applicable for detection and prognosis. For example, elevated levels of metabolites such as 2-hydroxyglutarate have been correlated with specific cancers, including glioma and acute myeloid leukemia, thereby presenting new opportunities for early detection and personalized treatment [126, 127]. Furthermore, metabolomic profiling can enhance therapeutic strategies by identifying metabolites associated with drug resistance or sensitivity [128]. For instance, alterations in lipid metabolism have been linked to chemotherapy responses in breast cancer, suggesting that modifying treatments based on metabolic profiles may enhance efficacy [129].

Metabolomics has wide-ranging clinical applications in oncology. Metabolites can serve as early diagnostic markers, enabling the identification of cancer even in asymptomatic stages [130]. Additionally, metabolic profiles provide prognostic information, assisting in identifying patients at an elevated risk of recurrence or treatment resistance. Furthermore, metabolites associated with therapeutic response can inform personalized treatment strategies, enhancing targeted therapies' effectiveness [128]. MS and nuclear magnetic resonance (NMR) spectroscopy are the principal techniques utilized for metabolomic profiling [131, 132]. MS is extensively used due to its sensitivity and capability to analyze complex mixtures. When integrated with chromatographic techniques such as gas chromatography (GC–MS) or liquid chromatography (LC–MS), MS facilitates both qualitative and quantitative analysis of a wide array of metabolites [133, 134]. This versatility makes MS an invaluable tool for investigating tumor metabolism. NMR, while less sensitive, offers distinct advantages, including nondestructive testing and the ability to ascertain metabolite structures. It is particularly beneficial for profiling metabolites within intact biological matrices, such as tissues or biofluids, and can corroborate findings from MS-based investigations due to its reproducibility and quantitative attributes.

The rapid expansion of metabolomics has led to vast datasets that require advanced computational and bioinformatics tools for effective analysis and interpretation. For instance, XCMS is commonly used for processing MS data, including peak detection, retention time adjustment, and metabolite quantification to uncover cancer-specific metabolic alterations [135]. Other popular tools such as MZmine and OpenMS facilitate the analysis of LC–MS and GC–MS liquid data, along with Sangerbox 3, which supports multiomics integration [136]. Furthermore, tools like MetaboAnalyst and MZmine play a crucial role in managing raw metabolomic data, performing peak identification, retention time adjustment, and normalization, thus ensuring high-quality data for further analyses [137, 138].

In metabolite identification, MetFrag is a valuable tool for in silico fragmentation and matching MS/MS data [139]. On the

other hand, Global Natural Products Social (GNPS) provides web-based resources tailored for sharing and analyzing MS/MS data, particularly focused on natural products research [140]. Moreover, databases like the Kyoto Encyclopedia of Genes and Genomes (KEGG), the Human Metabolome Database (HMDB), and Reactome provide crucial information about biological pathways related to metabolites, aiding in pathway enrichment analysis [141, 142]. KEGG organizes pathways by merging genomic, chemical, and functional data, in cancer biology. Additionally, specialized software tools such as LipidSearch, LIMS, and Lipid-Blast significantly improve the identification and quantification of lipids, thereby expanding the focus of metabolomic studies to include lipidomics, a vital component of cancer metabolism [143].

In metabolomics, statistical tools like R packages *limma* and *glmnet*, alongside software such as SPSS and SAS, help researchers conduct multivariate analyses, including principal component analysis (PCA) and partial least squares discriminant analysis [144, 145]. These methods aid in distinguishing metabolic patterns among various cancer types or clinical subgroups, revealing potential biomarkers linked to particular tumor traits or treatment responses.

Researchers are enhancing the identification of biomarkers that can substantially improve cancer diagnosis, prognosis, and therapeutic monitoring by integrating advanced bioinformatics tools with metabolomic analyses. Consequently, metabolomics presents significant potential for the future of precision oncology, paving the way for personalized treatment strategies (Table 1).

3 | Key Bioinformatics Tools for Biomarker Discovery

The development of bioinformatics tools has revolutionized biomarker discovery in precision oncology, enabling researchers to conduct a comprehensive analysis of complex datasets. Identifying biomarkers in precision oncology requires employing specialized bioinformatics tools throughout various phases of data analysis, from preprocessing to validation [167] (Figure 2). Each stage is essential to confirm that the identified biomarkers are reliable, relevant, and clinically significant. Below is an in-depth discussion of the primary bioinformatics tools utilized for biomarker discovery.

3.1 | Data Preprocessing and Quality Control Tools

Data preprocessing and quality control are essential steps in analyzing high-throughput sequencing data, enhancing the integrity and reliability of subsequent analyses. The process typically includes several steps, such as trimming low-quality reads, removing sequencing adapters, and filtering out contaminants. This initial step is crucial for improving the integrity and reliability of subsequent analyses [168].

The quality control of sequencing data performed during the preprocessing phase. FastQC is a widely utilized tool for assessing the quality of raw sequencing data, generating comprehensive reports on base quality scores, GC content, and sequencing depth. Researchers utilize these reports to identify potential data issues,

such as low-quality reads or biases, which may adversely affect the analysis [169].

Upon assessing the quality, Trimmomatic is used to trim low-quality bases and eliminate sequencing adapters, thereby ensuring the retention of only high-quality reads for subsequent analyses [170]. This procedure significantly enhanced the accuracy of variant calling and gene expression quantification, establishing a robust foundation for data interpretation. Following preprocessing, normalization is often essential to address variations attributable to discrepancies in sequencing depth or batch effects. ComBat and ComBat-ref are frequently employed to adjust for batch effects arising from technical variations in sample preparation or sequencing, thereby preventing false biomarker identification [171]. Furthermore, surrogate variable analysis (SVA) serves to detect and remove batch effects and other unwanted variations that are not directly measurable [172]. SVA is widely applied in RNA-seq and microarray studies to uphold data comparability across samples and experiments, ultimately improving the reliability of analytical results [173].

3.2 | Biomarker Discovery Algorithms

Once the quality of data is assured, the subsequent step involves identifying potential biomarkers through computational algorithms that differentiate between relevant features (such as genes or proteins) and background noise. Feature selection methods, including Lasso and Elastic Net, are employed to mitigate the complexities of large datasets and to select the most predictive biomarkers [174]. Lasso is particularly advantageous for high-dimensional genomic data, while Elastic Net is adept at handling correlated datasets with interdependent variables [175, 176]. ML methods play a crucial role in biomarker discovery by uncovering intricate relationships within extensive datasets [177]. Random Forests amalgamate decision trees to facilitate robust feature selection, whereas support vector machines (SVMs) are utilized to differentiate between cancerous and noncancerous samples [178, 179]. Furthermore, deep learning techniques, especially in the context of multiomics and image data, unveil complex patterns through the application of neural networks [180].

The primary instruments used in ML include Random Forest and scikit-learn, with a focus on model training in R, along with TensorFlow and PyTorch for deep learning applications [181–185]. Furthermore, additional tools such as Weka, Bioinformatics Toolbox, and R/Bioconductor offer methodologies for statistical analysis [186, 187]. In addition, Cytoscape and GenePattern facilitate the visualization of biomarker interactions within pathways, thereby assisting in the interpretation of biomarker networks in the context of precision oncology [188, 189].

3.3 | Multiomics Integration Platforms

Biomarkers are rarely identified through the utilization of a singular data type, as cancer encompasses intricate interactions across molecular strata. Multiomics integration platforms connect genomic, transcriptomic, proteomic, epigenomic, and metabolomic data, offering a holistic perspective on cancer biology and robust biomarkers supported by cross-omics evidence.

TABLE 1 | Current bioinformatics tools for biomarker discovery in precision oncology, sub-categorized to omics level, function, and application limitations.

Omics	Category	Tools	Description	Limitations	References
Genomics	Gene expression analysis	DESeq2, EdgeR, limma, Cufflinks, Ballgown	Differential expression analysis of RNA-seq and microarray data	Limited performance with extreme gene expression levels; normalization and quality control are essential	[21]
	Variant calling	GATK, MuTect2, Haplotype Caller, VarScan, FreeBayes	Tools for variant discovery, SNPs, and somatic mutation calling	GATK requires large computational resources; performance can vary depending on reference genome quality.	[146]
		Strelka, Lofreq, Platypus	High-sensitivity variant callers	May require extensive filtering; complex workflows	[147]
	Functional genomics	GSEA, DAVID, KEGG, Reactome, Sangerbox 3	Gene set enrichment, pathway analysis, and functional annotation	May overlook context-specific pathways or interactions, relying heavily on gene annotations that can differ across databases	[148]
	Copy number alterations	CNVkit, Control-FREEC, ExomeCNV	Detection of copy number variations from exome or whole genome data	Sensitivity depends on sequencing depth and genome complexity, with noisy data increasing the risk of false positives.	[149]
Epigenomics	DNA methylation	Bismark, MethylKit, Bis-SNP, DSS,	Tools for methylation calling and differential methylation analysis from bisulfite sequencing	Sensitive to sequencing errors in bisulfite-treated reads, with difficulty accurately calling methylation status in repetitive regions	[150]
	Chromatin accessibility	MACS, SICER, ChIPseeker, ATAC-Seq Tools (HMMRATAC)	Tools for peak calling, analysis of chromatin accessibility (e.g., ChIP-seq, ATAC-seq)	Difficult to distinguish between biologically relevant peaks and noise in low-signal data; requires deep sequencing	[151]
	Epigenetic regulation	EpiDISH, MethylSig, BSmooth	Tools for deconvolution of methylation data and analysis of epigenetic regulation	Interpretation of epigenetic regulation is context dependent; there are limited databases for noncoding regions.	[152]
	Histone modification analysis	diffReps, RSEG, ChromHMM	Analysis of histone modification data from ChIP-seq	False positive peaks and misalignment of histone marks can lead to inaccurate functional predictions.	[153]
Proteomics	Protein identification	MaxQuant, Proteome Skyline, Discoverer, MSFragger	Tools for mass spectrometry-based protein identification and quantification	MS often misses low-abundance proteins; identification relies on high-quality spectral libraries.	[78]
		Trans-proteomic pipeline, Open MS, Peptide Shaker,	Open-source tools for peptide and protein identification	High data preprocessing burden, especially for large datasets; lower sensitivity compared with commercial tools.	[154]
	Protein–protein interaction	STRING, Cytoscape, BioGRID, IntAct	Visualization and analysis of protein interaction networks	Often relies on predicted interactions, which can result in false positives; limited experimental validation	[155]

(Continues)

TABLE 1 | (Continued)

Omics	Category	Tools	Description	Limitations	References
Metabolomics	PTMs	PhosphoSitePlus, PTMScan, MODa	Tools for identifying and analyzing PTMs	PTM analysis is highly sensitive to sample preparation and detection techniques and has limited coverage of PTM types.	[87]
	Protein structural analysis	AlphaFold, Pymol, I-TASSER, MODELLER	Tools for predicting and visualizing protein structures	Prediction accuracy decreases for highly disordered regions or multidomain proteins; it is limited for protein interaction predictions.	[156]
	Metabolite identification	XCMS, Metabo Analyst, MZmine, OpenMS, Sangerbox 3	Tools for metabolite identification and quantification from MS data	Limited spectral libraries for certain metabolite classes; metabolite annotation is challenging due to overlap in mass/charge ratio.	[157]
		MetFrag, GNPS	Fragmentation-based metabolite identification	High false discovery rates for low-resolution MS; limited databases for novel compounds	[158]
	Metabolic pathway analysis	KEGG, HMDB, MSEA, Reactome, Pathway Commons	Pathway and network analysis tools for connecting metabolomics data to biological functions	Pathways are often curated from general organism models, which might not reflect species-specific or condition-specific pathways.	[159]
Transcriptomics	Lipidomics	LipidSearch, LIMSA, LipidBlast	Tools for analyzing lipidomics data from NMR or MS	Lipid identification is challenging due to the diversity of lipid structures; it requires specific libraries and standards.	[160]
	Bulk RNA-seq	Salmon, Kallisto, STAR, RSEM, HTSeq	RNA-seq quantification tools for transcript-level and gene expression analysis	Transcript quantification accuracy can drop for low-expressed genes or those with highly similar sequences.	[161]
	Single-cell RNA-seq	Seurat, Scanpy, CellRanger, Monocle, Sangerbox 3	Tools for clustering, differential expression, and trajectory analysis of single-cell RNA-seq data	Single-cell RNA-seq data often has high dropout rates; clustering can be influenced by noise, making cell type annotation complex.	[162]
	Noncoding RNA Analysis	miRDeep2, Infernal, snoSeeker, lncRNator,	Tools for identification of noncoding RNAs (<i>miRNAs</i> , <i>snoRNAs</i> , <i>lncRNAs</i>) from <i>RNA-seq</i>	Noncoding RNAs can be difficult to detect due to short lengths or sequence similarity; functional annotation is often incomplete.	[163]
Integrative multiomics	Multiomics integration	iCluster, MixOmics, MOFA+, Galaxy, Sangerbox 3	Tools for integrating genomics, proteomics, transcriptomics, and metabolomics data	The integration of multiomics datasets can be computationally intensive and requires careful interpretation of correlations between different omics layers.	[55]

(Continues)

TABLE 1 | (Continued)

Omics	Category	Tools	Description	Limitations	References
Cancer-specific tools	Cancer genomics & biomarkers	cBioPortal, GEPIA, OncoKB, TIMER, Sangerbox 3	Tools for cancer biomarker discovery and analysis of public datasets like TCGA, GTEx	Limited to public cancer datasets; may not capture rare mutations or ethnic-specific biomarker data.	[164]
	Drug sensitivity prediction	GDSC, PRISM Repurposing, CellMiner, Pharmacodb	Tools for predicting drug sensitivity and resistance based on cancer cell line data	in vitro data may not fully reflect real-world drug responses due to tumor microenvironment and patient variability.	[165]
	Immune infiltration analysis	TIMER, CIBERSORT, xCell, EPIC	Tools for estimating immune cell infiltration from bulk RNA-seq and other datasets	Immune cell quantification accuracy depends on deconvolution algorithms, which may perform poorly on low-purity or heterogeneous tumor samples.	[166]

References [61, 95, 176–197] present evidence regarding the applications and limitations of the listed tools across different domains. Genomics tools, including those for differential expression analysis and variant calling, encounter challenges such as performance variability in extreme expression values and high demands for computational resources. Functional genomics tools, especially those employed for pathway enrichment analysis, might face issues tied to pathway relevance, as their precision relies on the quality and completeness of gene annotations. Epigenomics tools, which analyze DNA methylation and chromatin accessibility, often find it difficult to accurately detect methylation in repetitive regions and to differentiate significant peaks. Tools for proteomics and metabolomics aid in the identification of proteins and metabolites; nonetheless, their sensitivity tends to be limited when it comes to detecting low-abundance analytes. Transcriptomics tools, whether used for single-cell or bulk RNA sequencing, experience challenges from noise and dropout events affecting lowly expressed genes, complicating cell type annotation. Multiomics integration tools, vital for thorough analyses, necessitate careful interpretation of data due to the complexities involved in merging diverse data types. Cancer-specific tools, frequently utilized in biomarker discovery, estimation of immune infiltration, and prediction of drug sensitivity, primarily depend on public datasets. Consequently, their accuracy may fluctuate, especially in heterogeneous tissue environments or when examining rare mutations.

Tools to integrate the multiomics data include SNF and *iCluster*. SNF constructs similarity networks for each omics category, which are then merged into a consensus network [190, 191]. This process facilitates biomarker discovery with consistent evidence across data layers. In comparison, *iCluster* groups multiomics data to identify cancer subtypes through common molecular traits, which may reveal biomarkers specific to subtypes that are relevant for prognosis or treatment response [192, 193].

Network-based methods are essential for identifying biomarkers. Tools like netboxr combine protein–protein interaction networks with genomic information, pinpointing regulatory pathways and biomarkers across various omics layers [194]. Cytoscape acts as a flexible platform to visualize molecular interaction networks, integrate multiomics data, and investigate biomarkers within biological pathways and network frameworks [188]. Moreover, Cytoscape serves as a platform for visualizing these networks and performing pathway enrichment analyses through various plugins, such as ClueGO and Bingo. These tools enhance the biomarker discovery domain and encourage personalized strategies in oncology (Figure 2).

3.4 | Pathway and Network Analysis

In cancer biomarker discovery, there is a growing emphasis on analyzing pathways and networks rather than solely examining individual genes or proteins. This transition recognizes the connections among molecular processes. Disruptions in biological pathways can occur due to altered gene expression or

protein activity, and recognizing these changes may facilitate the development of robust biomarkers.

A range of tools enhances pathway enrichment and network development. Gene set enrichment analysis (GSEA) detects enriched gene sets or pathways crucial for understanding differential gene expression in biological contexts [195]. The database for annotation, visualization, and integrated discovery (DAVID) aids in functional annotation, highlighting enriched pathways and cellular components [196, 197]. The Reactome provides a curated collection of biological pathways and mapping tools that relate data to pathway disruption for biomarker identification [198, 199]. Ingenuity pathway analysis (IPA) is utilized for pathway analysis and network construction, widely applied in therapeutic biomarker research [200]. Network analysis helps identify biomarkers by examining genes or proteins within networks, focusing on key “hubs” that regulate vital processes. STRING creates protein–protein interaction networks and identifies interactions specific to cancer [155]. Another essential bioinformatics tool known as BioGRID used to display essential protein and genetic interactions for networks in cancer biology [201]. *ClusterProfiler*, an R package, enhances these tools by visualizing pathway enrichment and aiding in dataset interpretation [202].

Recently developed tools have significantly advanced the discovery of biomarkers using network-based approaches. For instance, NetPath meticulously curates the signaling pathways associated with cancer, facilitating the identification of biomarkers through the dysregulation of these pathways [203]. Metascape, on the

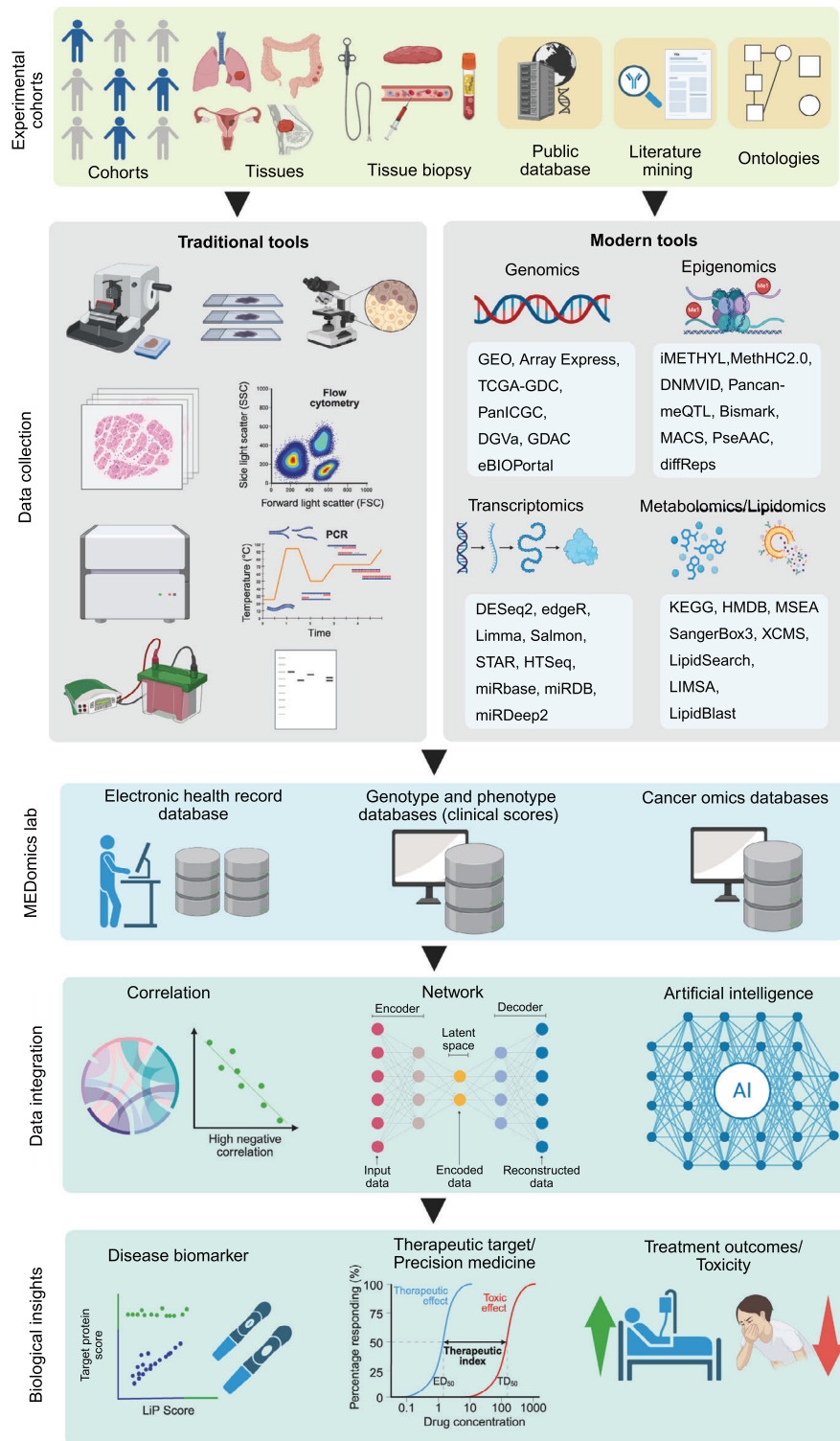


FIGURE 2 | Detailed workflow showing the combination of in silico and wet lab methods for generating biological insights and developing oncotherapeutic strategies in precision medicine: This diagram depicts the smooth integration of computational (in silico) and experimental (wet lab) techniques to extract biological insights and formulate oncotherapeutic strategies for precision medicine. The process initiates with gathering data from various sources, including clinical cohorts, patient-derived tissue samples, biopsies, publicly accessible databases, literature reviews, and ontologies that guarantee standardized knowledge representation. Standard laboratory methods like histopathology, flow cytometry, gel electrophoresis, and spectrophotometry are enhanced by high-throughput omics technologies, encompassing genomics, epigenomics, transcriptomics, and metabolomics/lipidomics. These data are meticulously organized and analyzed through specialized bioinformatics tools and repositories, such as GEO, TCGA, and HMDB. These datasets are integrated in MEDomics labs, leveraging electronic health records, genotype–phenotype connections, and cancer-specific multiomics repositories for thorough data harmonization. Subsequently, computational analyses are utilized to draw significant insights—correlation studies disclose biological relationships, network modeling portrays the complexity of molecular interactions, and AI methodologies, including machine learning and deep learning, assist in pattern recognition, predictive modeling, and simulation of biological processes.

other hand, consolidates various databases for pathway enrichment and network analysis, while GeneMANIA predicts gene functionality based on coexpression and interaction networks [204, 205]. Collectively, these tools empower researchers to identify biomarkers by systematically analyzing critical network nodes and dysregulated pathways.

3.5 | Validation of Biomarkers

Validating biomarkers is a complex yet crucial process for translating research findings into clinical trials and applications [206]. Biomarker candidates that have successfully completed the initial verification phase and developed precise, robust quantitative assays must then undergo *in silico*, analytical, and clinical validation to ensure reliability and clinical relevance. *In silico* validation, which encompasses computational evaluations, has emerged as an integral part of the validation process for potential biomarkers [207]. This method employs mathematical modeling, simulations, and data analysis from publicly accessible datasets such as TCGA and Gene Expression Omnibus (GEO). These resources offer comprehensive multiomics and gene expression data from thousands of patients across a variety of cancer types. Specifically, TCGA plays a pivotal role in validating genetic and RNA-based biomarkers, while GEO emphasizes RNA-based biomarkers in a range of clinical conditions and cancers [208, 209]. Supplementary databases like ArrayExpress and the International Cancer Genome Consortium (ICGC) provide functional genomics validation through diverse experimental designs, allowing for the assessment of somatic mutations CNVs [210, 211]. Additionally, meta-analytic strategies enhance the validation process by combining data from numerous studies, thereby increasing statistical power and confirming the robustness of potential biomarkers [212, 213].

Bioinformatics tools offer numerous benefits, such as cost efficiency, comprehensive evaluations of potential biomarkers, and improved survival analyses alongside risk assessments in the validation process. Survival analyses, including Kaplan–Meier (KM) curve analyses utilizing TCGA and GEO datasets, allow researchers to link biomarker expression with prognostic results, including overall survival, progression-free survival, and disease-free intervals [214]. SurvExpress, an extensive gene expression database and online biomarker validation resource, enhances risk stratification by combining survival analysis with gene

expression profiles, confirming the predictive value of biomarkers in separate cohorts [215]. For sophisticated statistical modeling, R-based packages like Survminer utilized to generate KM curves and Cox proportional hazards (Coxph) models [216]. The Coxph model is vital for examining the relationships between survival time and their predictors [217]. Web platforms such as GEPIA connect TCGA and GTEx data for gene expression and survival examination, investigating the prognostic significance of particular genes [218]. Additionally, tools like TIMER extend biomarker validation by factoring in immune cell infiltration dynamics with survival information, aiding in the assessment of how immune cells and gene expression affect survival rates [219]. Additionally, specialized bioinformatics platforms like OncoLnc and Sangerbox 3 streamline survival analyses utilizing TCGA-derived molecular profiles (e.g., mRNA, miRNA, lncRNA) to investigate survival outcomes [220]. Prognoscan offers meta-analysis capabilities to assess the connections between survival and gene expression, while the TCGA Biolinks (R package) facilitates large-scale survival research by maintaining reproducible access and analysis of TCGA data, thereby preserving methodological rigor in biomarker validation [221].

One essential aspect of bioinformatics tools involves performing biomarker analyses after an *in vivo* or *in vitro* efficacy study, which assists in validation efforts by correlating targets with the drug's mechanism of action [222]. This process helps predict the response or nonresponse of potential biomarkers. The findings from these analyses can serve as a basis for establishing inclusion and exclusion criteria for future clinical studies and validations. Bioinformatics tools systematically analyze downstream data from extensive *in vitro* assays, incorporating cell lines and genomically annotated tumor organoids, enabling researchers to differentiate between responding and nonresponding cell populations based on unique genomic signatures, thus providing early indications of possible biomarkers. By merging pharmacological response data from these screenings with computational models, researchers can link drug sensitivity or resistance with various genomic features such as gene expression profiles, mutational landscapes, CNVs, and pathway activation states, to formulate robust, multidimensional biomarker hypotheses. Promising candidates identified *in vitro* can progress to *in vivo* validation in preclinical models, where functional and mechanistic studies sharpen their biological significance [222]. Bioinformatics-based approaches play a crucial role in choosing xenograft or *patient-derived xenograft* models that align with a drug's molecular target

The outcome of these combined efforts leads to actionable oncotherapeutic strategies, such as uncovering disease-specific biomarkers, identifying new therapeutic targets, developing personalized medicine approaches, and assessing treatment efficacy and potential toxicity. This comprehensive framework promotes a data-driven and mechanistic understanding of cancer biology, thereby propelling precision oncology through advanced bioinformatics and experimental strategies. The figure is generated using BioRender.com. GEO: Gene Expression Omnibus, ArrayExpress: a public database for functional genomics data, TCGA: The Cancer Genome Atlas, DNAm: DNA methylation, cBioPortal: a tool for bioinformatics and genomics data exploration, DESeq2: differential expression analysis based on *RNA-seq* data, edgeR: a bioconductor package for differential gene expression analysis, limma: linear models for microarray data, Salmon: *RNA-seq* quantification tool, STAR: spliced transcripts alignment to a reference (*RNA-seq* aligner), HTSeq: a Python package for *RNA-seq* data processing, miRBase: a database for microRNA sequences, miRDB: a database for microRNA-target interactions, miRDeep2: a tool for identifying novel microRNAs, MethyC-2.0: a methylation data analysis tool, DNMT3: DNA methyltransferase 3D, PanCan-mQTL: pan-cancer methylation Quantitative Trait Loci, Bismark: a tool for mapping bisulfite-treated sequencing data, MiSeq: a sequencing platform by Illumina, KEGG: Kyoto Encyclopedia of Genes and Genomes, HMDB: Human Metabolome Database, MSEA: metabolic set enrichment analysis, SangerBox3: a software for genomics and bioinformatics analysis, XCMS: a tool for processing liquid chromatography–mass spectrometry (LC–MS) data, LipidBlast: a comprehensive lipid mass spectrometry database, LIP score: a score used for assessing lipidomics-based biomarkers.

or mutation profile, thereby ensuring translational relevance. This iterative process links *in silico* predictions to experimental validation, assisting in the prioritization and confirmation of the most clinically significant biomarkers potential [223].

Advanced analytical validation models, including AI and ML, enhance the validation process through cross-validation, which rigorously assesses the application of biomarkers across various datasets [224]. This enables evidence-based decisions about their clinical usefulness. Such an approach prevents overfitting, where a model too closely aligns with training data, impairing its ability to generalize to new data. Effective analytical validation assures both accuracy and generality of the model. Clinical validation is a crucial aspect of biomarker validation, as it improves the reliability of results and verifies the clinical significance of potential biomarkers by assessing their sensitivity and specificity [225]. Additionally, clinical validation clarifies the relationship between a biomarker and clinical aspects like treatment response, disease stage, and comorbid conditions. This validation relies on both *in silico* and analytical methods, which may involve retrospective reviews of past clinical trial data or new prospective trials. Retrospective reviews serve as a form of external clinical validation, particularly when biomarker assessments were not included in the original study design. In contrast, prospective clinical trials showcase the clinical relevance of a biomarker, acting as a kind of external validation that illustrates how its application can improve health results [226]. Various prospective clinical trial designs aim to confirm the clinical utility of biomarkers. A significant case is the United States Food and Drug Administration (US FDA's) 2017 tissue-agnostic approval of pembrolizumab, the first treatment authorized based on a biomarker rather than tumor location [227]. This decision stemmed from the KEYNOTE-016 study, which indicated higher overall response rates in patients with microsatellite instability-high (MSI-H) tumors treated with pembrolizumab compared with those with microsatellite stable (MSS) tumors, regardless of cancer type. Regulatory approval was based on pooled results from five trials (total $N = 149$), with MSI-H status retrospectively identified in 14 patients from two prospective studies, while 135 patients from three additional trials were prospectively confirmed. The objective response rate for MSI-H patients was 39.6% (including a 7% complete response rate) across 15 tumor types, which is deemed clinically significant. In contrast, patients with MSS tumors in the KEYNOTE-016 trial had a 0% response rate, underscoring the predictive power of the biomarker [228]. Another significant prospective clinical trial is the EORTC trial, which resulted in the US FDA approval of erlotinib as a first-line therapy for metastatic NSCLC with EGFR mutations [229]. Other examples of such trials include the MARVEL trial and SWOG S0819 [230, 231].

4 | AI and ML in Biomarker Discovery

AI and ML have become pivotal technologies in oncology, especially for identifying predictive biomarkers in precision medicine [232–238]. By facilitating the examination of extensive and intricate datasets, AI methods offer robust tools for discovering new biomarkers that aid in diagnosis, prognosis, and treatment choice. This section explores the role of AI and ML in biomarker discovery, emphasizing their methodologies, resources, and potential for the future (Table 2).

4.1 | Overview of AI Techniques in Oncology

AI techniques in biomedical research are categorized into supervised and unsupervised learning methods, both effective for revealing patterns in intricate datasets [256–258]. Supervised learning involves training algorithms using labeled data with known outcomes, which helps models understand the links between inputs (like gene expression profiles) and outputs (such as cancer subtypes). Common approaches include decision trees, random forests, SVMs, and neural networks, which are employed in cancer classification and in forecasting patient outcomes based on biomarker information. In contrast, unsupervised learning tackles unlabeled data to discover hidden patterns or clusters. Techniques like clustering (e.g., k-means and hierarchical clustering) and dimensionality reduction (e.g., PCA and t-SNE) are instrumental in biomarker discovery, identifying patient subgroups, and simplifying data complexity [181].

Deep learning, a branch of ML, plays a crucial role in biomarker discovery by seamlessly extracting hierarchical features from raw data [259]. CNNs and RNNs excel in the analysis of genomic, transcriptomic, and imaging data. These models have proven effective for classifying tumors based on histopathological images and predicting gene mutations from sequencing data, providing powerful tools for biomarker identification and large-scale exploration of biological patterns [260, 261].

4.2 | AI Tools for Predictive Biomarkers

AI-powered tools for biomarker discovery and predictive modeling have expedited oncology research [262]. Predictive modeling utilizes ML techniques to examine the connections between molecular traits and clinical results, including survival rates and therapeutic responses. A prime example is DeepVariant, a deep learning tool created by Google, which identifies genetic variants from NGS data and enhances mutation detection in cancer genomics [263]. PandOmics, a cloud-based software platform that applies AI and bioinformatics approaches to multimodal omics data for biomarker discovery [264]. These tools are capable of predicting prospective biomarkers from a range of omics data (Figure 3).

Numerous specialized platforms have highlighted the critical role of AI in identifying biomarkers. OncoKB, a precision oncology knowledge base, merges genomic sequencing with clinical trial data to offer insights into gene mutations vital for cancer treatment [164]. Open-source tools like DeepLearning4J and Omics Data Science offer deep learning libraries for complex biomarker analysis [265]. Drug Discovery AI predicts effective drug combinations based on tumor biomarkers, thereby promoting progress in personalized oncology treatments.

4.3 | Challenges and Opportunities

AI technologies are significantly impacting biomarker discovery in oncology, yet challenges remain [266]. A key obstacle is the scarcity of high-quality labeled data, particularly for rare cancers or newly identified biomarkers. Training effective AI models with small datasets can lead to overfitting, reducing both

TABLE 2 | AI/ML approaches in cancer biomarker discovery.

Approach	Application	Emerging trends	Practical challenges	Tools	Notable studies	References
DL	Analyzing complex omics and imaging data to identify biomarkers	CNNs for genomic sequence analysis, RNNs for time-series data analysis	Data sparsity, overfitting on small datasets	TensorFlow, Keras, Caffe, PyTorch	Immune gene signatures in ovarian cancer, leveraging DL models	[18]
SL	Predicting cancer subtypes or patient outcomes based on biomarkers	Support vector machines, random forests, gradient boosting	Need for large labeled datasets, class imbalance	Scikit-learn, XGBoost, LightGBM	Seven-gene signature for lung cancer prognosis using supervised ML	[239]
UL	Clustering patients based on omics data, identifying novel cancer subtypes	Dimensionality reduction techniques, clustering algorithms like K-means	Lack of labeled data, interpretation of clusters	PCA, t-SNE, UMAP, k-Means	Clustering BC subtype based on multiomics integration using UL	[240]
FL	Enabling model training across decentralized data sources, preserving privacy	Distributed learning, model aggregation without data transfer	Ensuring privacy and security across datasets	TensorFlow Federated, PySyft	Federated learning-based cancer survival prediction method with privacy protection	[241]
TL	Leveraging pretrained models on smaller datasets to improve biomarker prediction	Fine-tuning models on new datasets for improved accuracy	Requires pretrained models, domain adaptation	Keras, PyTorch	Deep learning for electronic cancer record data with transfer learning	[242]
XAI	Making AI predictions transparent for clinical applications	Feature importance, rule extraction, local explanation models	Lack of trust from clinicians, need for clinical validation	SHAP, LIME	LIME for explaining ML models in healthcare	[243]
MOI	Identifying multiomics signatures for comprehensive biomarker discovery	Integration of genomic, transcriptomic, and proteomic data	Data integration challenges, noise in multiomics data	iClusterPlus, OmicLearn, MOFA, MultiOmics	Integrating multiomics data for precision oncology biomarker discovery	[244]
RL	Optimizing treatment strategies based on patient responses over time	Real-time treatment adaptation, dynamic learning from outcomes	Requires patient-specific long-term data, ethical concerns	OpenAI Gym, TensorFlow Agents	Personalized cancer treatment strategies tailor treatments on the basis of a patient's health status, cancer type, and stage	[245]
EL	Combining multiple models for improved biomarker prediction accuracy	Increased focus on ensemble methods for robust predictions	Complexity in model integration, computational load	RF, Gradient Boosting, Stacking	Ensemble learning for cancer biomarker validation	[246]
NLP	Mining scientific literature to identify potential biomarkers	Text mining for the extraction of novel biomarkers from unstructured data	Ambiguity in language, challenges in processing large volumes of text	SpaCy, BERT, SciSpacy	NLP for cancer biomarker discovery through literature mining	[247]
GA	Feature selection and optimization in biomarker discovery	Evolutionary algorithms to discover optimized biomarker sets	Convergence to local minima, complexity in tuning parameters	DEAP, GALib	Genetic algorithm-based biomarker selection for cancer diagnosis	[248]

(Continues)

TABLE 2 | (Continued)

Approach	Application	Emerging trends	Practical challenges	Tools	Notable studies	References
SVM	Classification of biomarker profiles for disease prediction	Use of kernel methods for nonlinear relationships in biomarker data	Difficulty in scaling to large datasets, computational cost	LIBSVM, Scikit-learn	Biomarker discovery using SVM for cancer prediction	[249]
GNN	Identifying biomarker relationships and gene interactions	Modeling biomarker interactions and pathways using graph structures	Lack of sufficient graph data, scalability	DGL, PyTorch Geometric	Graph neural networks for biomarker discovery in cancer	[250]
RF	Identifying relevant features for biomarker prediction	Leveraging feature importance scores for biomarker prioritization	Difficulty in interpreting large ensemble models	Scikit-learn, R	Random forest for biomarker selection in cancer diagnosis	[251]
BN	Modeling probabilistic relationships between biomarkers	Dynamic modeling of biomarkers with uncertainty quantification	Computational complexity, need for high-quality data	PyMC3, Netica	Bayesian networks for cancer biomarker discovery	[252]
AutoML	Automating the selection of ML models for biomarker prediction	Deployment of automated platforms for personalized biomarker discovery	Limited flexibility, over-reliance on automated processes	Google AutoML, H2O.ai, Auto-sklearn	AutoML for biomarker discovery in cancer	[253]
CA	Unsupervised classification of biomarkers across different conditions	Integration of clustering with single-cell RNA-seq data	Difficulty in interpreting clustering results, computational expense	k-Means, DBSCAN, Hierarchical Clustering	Clustering biomarkers for cancer subtyping	[254]
DR	Reducing high-dimensional omics data for biomarker identification	Multilevel dimensionality reduction methods to handle large omics datasets	Loss of information, difficulty in balancing accuracy and dimensionality	PCA, t-SNE, UMAP	PCA for dimensionality reduction in biomarker discovery	[255]

Artificial intelligence (AI) and machine learning (ML) algorithms, such as deep learning (DL), supervised learning (SL), unsupervised learning (UL), and federated learning (FL), have greatly enhanced cancer research by facilitating the analysis of intricate, high-dimensional omics data. Methods including random forests (RF), support vector machines (SVM), and graph neural networks (GNN) are particularly effective at uncovering hidden biomarkers for cancer diagnosis, prognosis, and treatment responses, outperforming conventional statistical techniques. Multiomics integration (MOI) and explainable AI (XAI) are vital for improving data interpretation and model transparency, effectively addressing black-box decision-making apprehensions. Transfer learning (TL) and reinforcement learning (RL) also contribute to refining predictive models by utilizing prior knowledge and adaptive learning methods. Nonetheless, substantial challenges remain, such as issues with data integration, model interpretability, and the ability to generalize across various cancer subtypes. The intricate nature of multiomics datasets, alongside the need for extensive, well-annotated training data, highlights the necessity for enhanced data quality, algorithm transparency, and computational efficiency. Approaches like AutoML and ensemble learning (EL) are under consideration to simplify model selection and hyperparameter optimization, while genetic algorithms (GA) and Bayesian networks (BN) provide promising solutions for feature selection and probabilistic modeling. Furthermore, dimensionality reduction (DR) and clustering algorithms (CA) are crucial for preprocessing large datasets, ensuring that meaningful biological insights can be drawn. As AI/ML advances, overcoming these challenges will be essential to fully leverage its potential in precision oncology.

generalizability and clinical relevance. Although strategies like data augmentation, transfer learning, and cross-validation can help address these problems, the lack of data continues to restrict AI's potential in biomarker discovery. Another significant challenge involves integrating various data types, such as genomics, transcriptomics, proteomics, and clinical records, which present different formats and noise levels, thereby complicating model training and efficiency [267]. Advancements in data integration techniques are vital to boost AI's capability in identifying clinically meaningful biomarkers.

Despite facing challenges, AI demonstrates significant promise in precision oncology. It facilitates the integration of diverse

datasets, uncovering complex biomarker signatures that consider molecular interactions. This could lead to more accurate biomarker identification, reducing the time and costs associated with discovering predictive markers, and enabling the application of findings in clinical practice. However, the “black box” characteristic of AI models raises interpretability concerns, making clinicians reluctant to adopt these technologies [268]. Exploring explainable AI and visualization techniques is essential for nurturing clinical trust. Additionally, leveraging patient data for AI applications introduces ethical and regulatory dilemmas, necessitating well-defined guidelines to uphold ethical integrity [269]. Collaborative efforts among researchers, clinicians, and data scientists can help tackle existing challenges, while

advancements in AI technologies, like natural language processing and advanced imaging, offer fresh possibilities for biomarker discovery and therapeutic target identification [270].

5 | Public Databases and Resources for Biomarker Discovery

Public databases and online resources play a crucial role in modern oncology research by providing vast datasets that help identify biomarkers for precision oncology [262]. These databases contain high-dimensional data from various omics studies, such as genomics, transcriptomics, proteomics, and metabolomics, enabling researchers to discover potential biomarkers across multiple cancer types. By granting access to diverse datasets, public resources empower scientists to explore genetic, transcriptomic, and proteomic differences, which are essential for identifying biomarkers needed for diagnosis, prognosis, and targeted therapies (Figure 4).

5.1 | Cancer Genomic Databases

Cancer genomic databases such as TCGA, ICGC, COSMIC, Genomic Data Commons (GDC), and cBioPortal are essential for examining genetic mutations (CNVs), and variations in gene expression associated with cancer [271]. TCGA is notable for its vast repository, which includes data from thousands of patients with various tumor types, featuring both genomic sequences and transcriptomic profiles [272]. ICGC focuses on diverse populations, aiding in the identification of genetic alterations relevant to specific demographics [273]. COSMIC specializes in somatic mutations, offering insights into actionable mutations and predictive capabilities [274]. Platforms like GDC and cBioPortal facilitate integrative analyses by combining data from numerous cancer studies and enhancing the connections between genetic alterations and clinical outcomes [274].

Additional resources will significantly enhance biomarker research beyond just genome-level changes. The Cancer Mutation Census gathers information on mutations in genes related to cancer, facilitating the identification of common mutations [275]. Although not cancer-specific, ExAC offers data on rare variants, contributing to a better understanding of cancer susceptibility [276]. The GEO contains extensive high-throughput gene expression data, including essential cancer studies that are vital for comparing expression patterns in cancerous and normal tissues [277]. Together, these resources create a strong foundation for biomarker discovery, encouraging further investigation into molecular variations in cancer.

5.2 | Transcriptomic Databases

Transcriptomic databases are essential for providing insights into gene expression, allowing researchers to probe gene regulation and identify DEGs that serve as biomarkers. GEO is a key public repository that offers high-throughput gene expression data from various genomic techniques, such as microarray and RNA-seq. It supports investigations into differential expression related to different cancer types, treatment responses, and dis-

ease stages. Similarly, ArrayExpress provides curated datasets from microarray and sequencing studies, enabling researchers to explore transcriptional changes linked to specific cancer phenotypes [278]. These transcriptomic resources are vital for precision oncology, assisting identification of critical biomarkers for diagnosis, prognosis, and treatment responses [278].

5.3 | Proteomics and Metabolomics Databases

Proteomics and metabolomics databases provide useful insights into protein expression, PTMs, and metabolite profiles, facilitating the identification of biomarkers at both protein and metabolite levels. The PRoteomics IDentifications database serves as a vital resource for MS-based proteomics, offering essential data on protein expression, identification, and modifications that are crucial for discovering oncology biomarkers [279]. Furthermore, PeptideAtlas compiles high-confidence protein and peptide identifications, enhancing biomarker discovery by providing information on protein abundance, modifications, and cancer-related interactions [280]. Other essential proteomic databases and repositories include GPMDB, MassIVE, PASSEL, SRMatlas, and Panorama utilized for several studies including biomarker discovery [281]. MetaboLights offers small-molecule profiles from various biological samples, enabling researchers to explore metabolic changes in cancer and identify metabolite biomarkers reflective of tumor metabolism and its influence on cancer progression [282].

5.4 | Integrated Multiomics Databases

Integrated multiomics databases improve the simultaneous analysis of genomic, transcriptomic, proteomic, and metabolomic data, providing a thorough understanding of cancer biology. These platforms facilitate the integration of various data types, helping researchers identify biomarkers across different molecular layers. cBioPortal, a key open-access platform, merges genomic, transcriptomic, and clinical data from large cancer studies like TCGA and ICGC, allowing for the investigation of mutations, expression patterns, and patient outcomes to uncover biomarkers. Another essential tool, Xenabrowser, offers interactive access to multiomics data from TCGA and ICGC, combining genomic, epigenomic, and transcriptomic data to identify clinically relevant biomarkers [23, 283].

6 | Case Studies: Bioinformatics-Driven Biomarker Discovery

Bioinformatics plays a vital role in identifying and validating key biomarkers in oncology, significantly impacting clinical decisions and personalized treatment strategies. This section features case studies that highlight successful biomarker discoveries enabled by bioinformatics workflows, emphasizing their therapeutic effects and contributions to advancements in targeted therapies within clinical oncology (Table 3).

Integration of AI in precision oncology

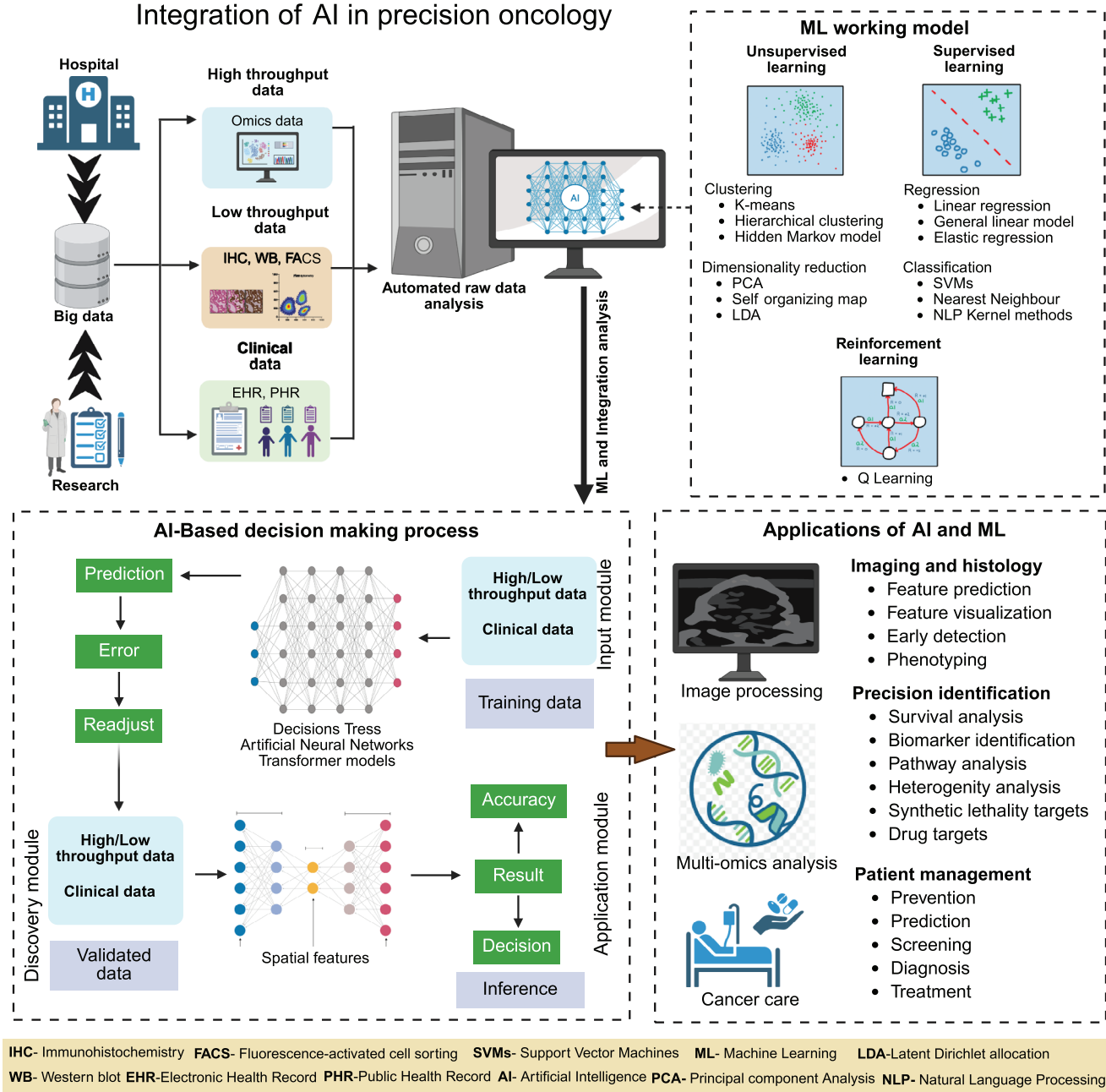


FIGURE 3 | An overview of integrating artificial intelligence (AI) and machine learning (ML) into precision oncology, highlighting AI's pivotal role in cancer research and patient care. This figure depicts a comprehensive approach to precision medicine that integrates various data sources for deep phenotyping. Data collection includes public annotation databases alongside patient-specific information, which encompasses genomic, proteomic, microbiome, clinical, epigenomic, and metabolomic datasets. Additionally, high-throughput omics data are complemented by low-throughput data from methods like IHC and FACS, along with clinical data sourced from EHR and PHR. The analysis of this data is powered by various machine learning algorithms, which fall into three main learning categories: unsupervised learning (e.g., clustering methods such as K-Means and hierarchical clustering, along with dimensionality reduction techniques like PCA and LDA), supervised learning (e.g., regression models and classification algorithms, including SVMs and nearest neighbor techniques), and reinforcement learning (e.g., Q-learning). The AI-driven decision-making process is illustrated as an ongoing cycle involving prediction, error analysis, and adjustments, thus improving model accuracy through the incorporation of high-throughput and validated clinical data. As highlighted in the figure, the key applications of AI and ML in oncology include image processing, multimomics analysis, early cancer detection, phenotype characterization, biomarker discovery, pathway analysis, and holistic patient management encompassing prevention, prediction, screening, diagnosis, and treatment. The figure highlights the transformative potential of AI-powered tools, such as decision trees, artificial neural networks, and transformer models, in enhancing precision oncology through personalized treatment strategies and improving cancer care outcomes. The figure is generated using BioRender.com. IHC: Immunohistochemistry, FACS: fluorescence-activated cell sorting, SVMs: support vector machines, ML: machine learning, PCA: principal component analysis, NLP: natural language processing, EHR: electronic health record, PHR: public health record, AI: artificial intelligence, WB: Western blot. LDA: linear discriminant analysis, KMeans: K-means clustering, ANNs: artificial neural networks, KNN: K-nearest neighbors.

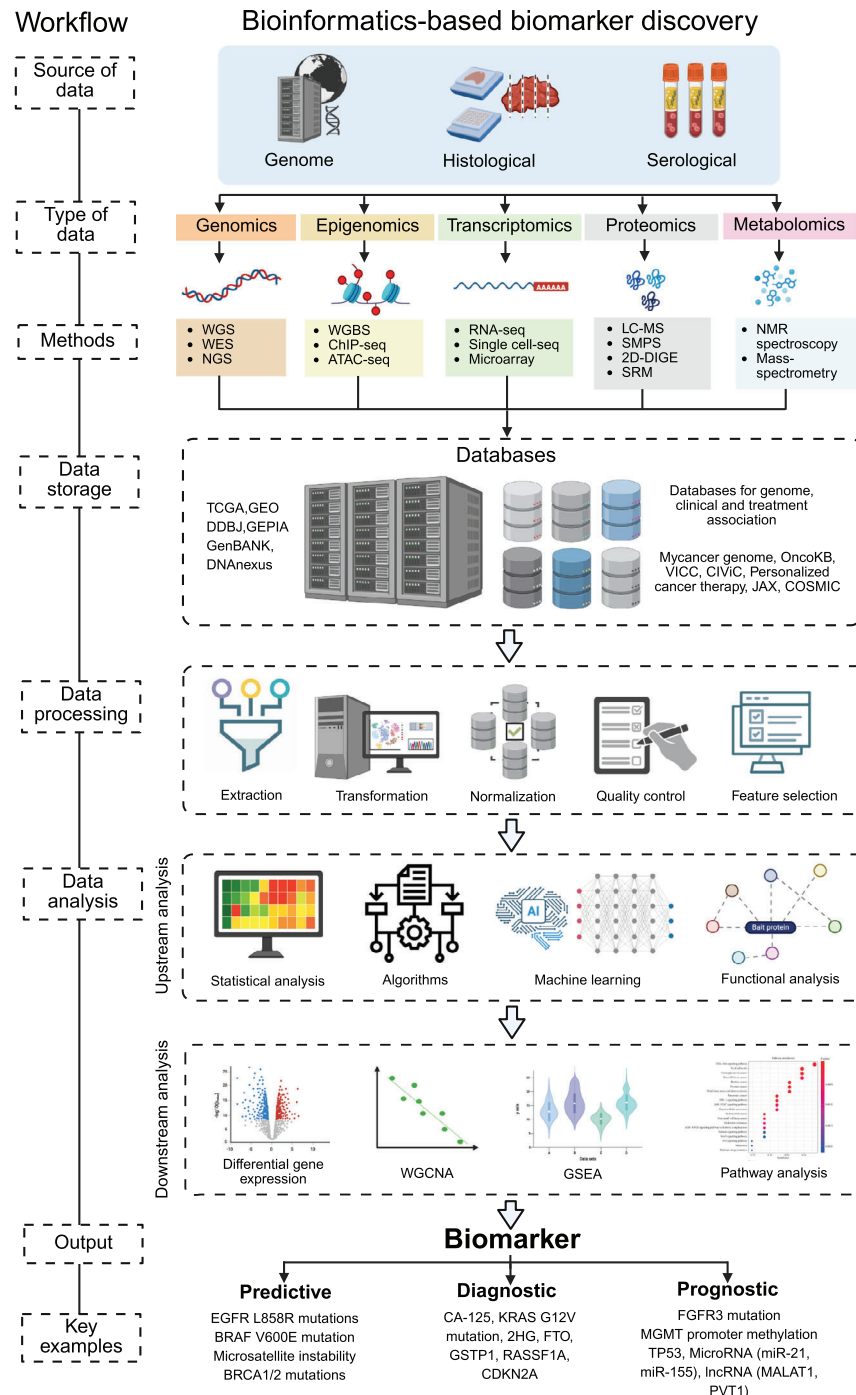


FIGURE 4 | Bioinformatics-driven workflow for biomarker discovery and validation in precision oncology. This schematic representation outlines a comprehensive bioinformatics-driven workflow for biomarker discovery and validation in precision oncology, emphasizing the integration of diverse data sources and multiomics platforms. The workflow begins with the collection of various data types, including genomic, histopathological, and serological information derived from high-throughput datasets and patient-derived tissue and blood samples. These datasets incorporate multiomics approaches—genomics, transcriptomics, proteomics, and metabolomics—to identify key biomolecules such as DNA, RNA, proteins, and metabolites associated with cancer progression. Advanced technologies, including whole-genome sequencing (WGS), RNA sequencing (RNA-seq), whole-exome sequencing (WES), whole-genome bisulfite sequencing, chromatin immunoprecipitation sequencing (ChIP-seq), assay for transposase-accessible chromatin sequencing (ATAC-seq), liquid chromatography–tandem mass spectrometry (LC–MS), single-molecule protein sequencing (SMPS), two-dimensional difference gel electrophoresis (2D-DIGE), selected reaction monitoring (SRM), and nuclear magnetic resonance (NMR), facilitate deep phenotyping and high-resolution molecular characterization. The data generated are systematically stored and retrieved from key repositories such as The Cancer Genome Atlas (TCGA), Gene Expression Omnibus (GEO), GenBank, DNA Data Bank of Japan (DDBJ), Gene Expression Profiling Interactive Analysis (GEPIA), Variant Interpretation for Cancer Consortium (VICC), Clinical Interpretation of Variants in Cancer (CIViC), and the Catalogue of Somatic Mutations in Cancer (COSMIC). Following data acquisition, rigorous preprocessing steps—including data extraction, transformation, normalization, and quality control—ensure reliability and reproducibility. Subsequent computational analyses employ sophisticated statistical methods,

TABLE 3 | Bioinformatics-driven cancer biomarker discovery: Identification, methodology, clinical significance, and applications.

Cancer	Biomarker	Discovery method	Clinical significance	Potential applications	References
BC	HER2 (ERBB2)	Gene amplification studies, FISH, IHC	Overexpression in 20–30% of breast cancers; associated with aggressive disease and poor prognosis	Targeted therapy with trastuzumab, pertuzumab; prognostic biomarker for treatment stratification	[284]
PC	PSA	Blood testing, ELISA	Elevated serum levels indicate prostate cancer presence; widely used for screening and prognosis assessment	Early diagnosis, monitoring therapy response, detection of recurrence	[285]
OC	CA-125	Enzyme-linked immunosorbent assay	Elevated levels associated with ovarian cancer; used to monitor disease progression and recurrence	Screening in high-risk populations, monitoring therapeutic response, recurrence detection	[286]
LC	EGFR L858R mutations	PCR-based sequencing, NGS	Activating mutations in EGFR are present in 10–15% of NSCLC, predictive of response to EGFR inhibitors like gefitinib	Predictive biomarker for targeted therapy with EGFR inhibitors (e.g., gefitinib, erlotinib)	[287]
CRC	KRAS G12V mutation	PCR-based sequencing, NGS	Mutations in KRAS (in 40–50% of cases) predict lack of response to anti-EGFR monoclonal antibodies	Identifying candidates for anti-EGFR therapies (cetuximab, panitumumab)	[288]
Mel	BRAF V600E mutation	PCR-based sequencing, NGS	Found in ~50% of melanomas; associated with worse prognosis and response to targeted therapies like BRAF inhibitors	Predictive biomarker for BRAF inhibitors (e.g., vemurafenib, dabrafenib)	[289]
HCC	AFP	Blood test, immunoassays	Elevated AFP levels indicate HCC; also used to monitor treatment response and recurrence	Screening for HCC, monitoring therapy effectiveness, prognosis	[290]
PC	CA 19-9	Enzyme-linked immunosorbent assay	Elevated in pancreatic cancer, especially in advanced stages; linked to poor prognosis	Early diagnosis, monitoring disease progression, therapeutic response assessment	[291]
GC	MSI	PCR, NGS	MSI status indicates defective mismatch repair system; predictive of better response to immune checkpoint inhibitors	Predictive biomarker for immune therapy (pembrolizumab, nivolumab)	[292]
BLC	FGFR3 S249C mutation	PCR-based sequencing, NGS	Found in 60–70% of nonmuscle-invasive bladder cancers; associated with less aggressive tumors	Prognostic marker, target for FGFR inhibitors (e.g., erdafitinib)	[293]
Leuk	PML-RARA fusion gene	Fluorescence in situ hybridization	Specific to acute promyelocytic leukemia; associated with abnormal proliferation and differentiation	Target for all-trans retinoic acid (ATRA) and arsenic trioxide therapy	[294]
GBM	MGMT promoter methylation	Methylation-specific PCR, pyrosequencing	Predictive of response to temozolomide chemotherapy; methylation inactivates MGMT gene and sensitizes tumors to chemotherapy	Prognostic marker, therapeutic guidance for chemotherapy decision-making (temozolomide)	[295]

(Continues)

TABLE 3 | (Continued)

Cancer	Biomarker	Discovery method	Clinical significance	Potential applications	References
CC	HPV	PCR, hybrid capture, sequencing	Detection of high-risk HPV genotypes (e.g., HPV16, HPV18) strongly associated with cervical carcinoma	Screening, vaccination, prevention, and monitoring recurrence	[296]
NSCLC	ALK rearrangements	FISH, RT-PCR, NGS	ALK rearrangements (e.g., EML4–ALK) found in a subset of NSCLC; associated with sensitivity to ALK inhibitors	Targeted therapy with ALK inhibitors (e.g., crizotinib, alectinib)	[297]
MM	Beta-2 Microglobulin	Blood test, immunoassays	Elevated levels associated with poor prognosis, correlates with tumor burden and kidney function	Prognosis assessment, monitoring disease progression, therapy evaluation	[298]
BC	BRCA1 185delAG, BRCA2 6174delT mutations	Genetic testing, NGS	Inherited mutations increase the risk of breast and ovarian cancers; predictive of response to PARP inhibitors	Risk assessment, predictive biomarker for PARP inhibitors (e.g., olaparib, talazoparib)	[299]

Abbreviations: FISH, fluorescence in situ hybridization; IHC, immunohistochemistry; HPV, human papillomavirus DNA; NGS, next-generation sequencing; PCR, polymerase chain reaction; MSI, microsatellite instability; AFP, alpha-fetoprotein; PSA, prostate-specific antigen; NSCLC, non-small cell lung cancer; BC, breast cancer; PC, prostate cancer; OC, ovarian cancer; LC, lung cancer; CRC, colorectal cancer; Mel, melanoma; HCC, hepatocellular carcinoma; PC, pancreatic cancer; GC, gastric cancer; BLC, bladder cancer; Leuk, leukemia; GBM, glioblastoma; CC, cervical cancer; MM, multiple myeloma; ELISA, enzyme-linked immunosorbent assay; REF, references [333–348].

6.1 | Examples of Successful Biomarker Discovery

Numerous key biomarkers discovered through bioinformatics have transformed precision oncology. These biomarkers enhance cancer diagnosis and prognosis and guide therapies by identifying patients likely to respond to specific treatments. One significant example is PD-L1 (programmed death-ligand 1), which allows tumor cells to evade immune responses by attaching to the PD-1 receptor [300]. Bioinformatics methods, including differential gene expression analysis and pathway enrichment, have played a vital role in identifying PD-L1 as an immune checkpoint biomarker [301]. Examining high-throughput RNA-seq data with tools like DESeq2 and EdgeR have shown PD-L1 upregulation in various cancers [21]. By merging transcriptomic data with clinical outcomes and immune-related pathways through tools such as GSEA and IPA, researchers have gained a deeper understanding of PD-L1’s role in immune evasion [302]. PD-L1 testing is instrumental in selecting patients who could benefit from immune checkpoint inhibitors like pembrolizumab and nivolumab, with bioinformat-

ics crucial in associating PD-L1 expression with responses to immunotherapy [300].

A significant breakthrough in biomarker research is the discovery of *BRCA1* and *BRCA2* germline mutations, which strongly indicate hereditary breast and ovarian cancers [303]. Breast and ovarian cancers are responsible for the majority of the cancer-related deaths among women around the globe [304–306]. Bioinformatics analyses of extensive sequencing data have validated these mutations as highly penetrant biomarkers. Mutation detection tools, such as Mutect and GATK, were utilized to analyze whole-exome sequencing (WES) and NGS data to identify *BRCA mutations* [307]. Additionally, variant annotation tools like ANNOVAR and SnpEff, combined with computational predictions, have assessed the functional implications of these mutations [308]. The clinical impact is significant as BRCA testing now serves a vital role in evaluating cancer risk and shaping treatment decisions, especially with PARP inhibitors like olaparib, which target *BRCA-mutated* cancers [309]. Genetic testing for BRCA mutations has become the standard of care

machine learning algorithms, and functional annotation techniques to derive biologically meaningful insights. Key downstream analyses, such as differential gene expression analysis, weighted gene coexpression network analysis (WGCNA), and gene set enrichment analysis (GSEA), enable the identification of putative biomarkers with potential clinical utility. The outputs of this workflow encompass predictive, diagnostic, and prognostic biomarkers, reinforcing the crucial role of bioinformatics in advancing personalized medicine and precision oncology. Identified biomarkers, including EGFR (epidermal growth factor receptor), BRCA (breast cancer gene), CA-125 (cancer antigen 125), 2HG (2-hydroxyglutarate), FTO (fat mass and obesity associated gene), GSTP1 (glutathione S-transferase P), RASSF1A (Ras association domain family member 1), CDKN2A (cyclin-dependent kinase inhibitor 2A), and FGFR3 (fibroblast growth factor receptor 3), undergo rigorous validation through multiomics integration and bioinformatics pipelines. This process ensures the identification of high-quality biomarkers for improved cancer diagnostics, prognostic assessments, and the development of targeted treatment strategies, ultimately driving advancements in precision medicine. The figure is generated using BioRender.com.

in oncology, particularly for patients diagnosed with breast and ovarian cancer [310].

6.2 | Therapeutic Implications

The identification of predictive biomarkers significantly enhances our understanding of cancer biology and guides treatment decisions. These biomarkers enable the selection of targeted therapies, thereby facilitating customized treatments personalized to patients' molecular profiles to attain optimal effectiveness.

EGFR mutations are associated with increased sensitivity to EGFR inhibitors such as erlotinib and gefitinib, especially in NSCLC [311]. Leveraging bioinformatic tools like VarScan and OncoKB is essential for identifying these mutations as actionable biomarkers [312]. By analyzing *EGFR mutations* via sequencing data, links to downstream pathways have been revealed, which assist in selecting targeted therapies and results in better clinical outcomes [313].

Rearrangements of anaplastic lymphoma kinase (ALK) in lung cancer have spurred the development of ALK inhibitors, such as crizotinib and alectinib [314, 315]. Bioinformatics tools, including FusionCatcher and STAR-Fusion, evaluate RNA-seq data for ALK fusions and network-based tools like Cytoscape assist in validating ALK partners [316]. Testing for ALK rearrangements now determines eligibility for ALK inhibitors, thereby enhancing progression-free survival rates in lung cancer [317].

TFF3 and pBAD are emerging as important targets in several malignancies including mammary, lung, liver, pancreatic, colorectal, ovarian, and endometrial due to their roles in tumor growth and treatment responses [318–337]. They are vital for epithelial cell regeneration and are linked to poor patient outcomes. By targeting TFF3, it may be possible to hinder tumor development by influencing the tumor microenvironment [338]. Tools such as GEPIA and DAVID are useful for assessing TFF3 expression and its related pathways. Moreover, pBAD, recognized for its ability to promote apoptosis, could be harnessed to increase cell death in tumors, making them more susceptible to chemotherapy [327]. Analyzing TFF3 and pBAD expression through bioinformatics reveals their contributions to cancer progression and supports innovative treatment strategies.

7 | Challenges in Bioinformatics-Driven Biomarker Discovery

While bioinformatics has significantly transformed the process of biomarker discovery, challenges persist in translating computational results into clinically applicable biomarkers [339]. Principal issues include data integration, computational constraints, and the transition from discovery to clinical application. Addressing these challenges is crucial for improving bioinformatics methodologies and their impact on precision oncology.

7.1 | Data Integration and Heterogeneity

One of the significant challenges in biomarker discovery driven by bioinformatics is integrating various omics data types—such as genomics, transcriptomics, proteomics, epigenomics, and metabolomics—into a unified framework [340]. Each omics layer offers unique insights while differing in structure, scale, and complexity; for example, genomic data are typically static, whereas transcriptomic and proteomic data are more dynamic. Advanced computational tools are required to tackle issues such as discrepancies in data types, noise, and missing values to effectively integrate these datasets. Despite continuous efforts using tools like iCluster and SNF, integrating multiomics data remain a significant challenge due to a lack of standardization, variations in sample collection methods, and disparate data processing pipelines [192, 340]. Additionally, the heterogeneity of cancer complicates biomarker discovery even further, as differences among patients and tumor subtypes imply that signatures may not be applicable across the board [262].

7.2 | Interpreting Big Data

The massive amount of data produced by high-throughput technologies poses a considerable challenge [341]. Current sequencing technologies generate terabytes of data, requiring advanced computational infrastructure for processing, storage, and analysis. Many bioinformatics tools demand significant computational power, which can be a barrier for researchers without access to high-performance systems [342]. The necessity for real-time analysis in clinical settings further increases the complexity. Proper long-term storage of omics data, particularly in large cohort studies, requires robust data management systems, raising issues around data retrieval, sharing, and reproducibility. Additionally, the increase in data volume raises the likelihood of false positives and misleading correlations, especially in omics studies with multiple variables are simultaneously tested [343]. Addressing these challenges requires enhanced interdisciplinary collaboration and more effective validation and regulatory processes to facilitate the integration of bioinformatics-driven biomarkers into clinical practice.

7.3 | Translating Findings to Clinical Practice

Bioinformatics has significant potential for the discovery of biomarkers; however, applying these findings in clinical settings remains challenging [339]. Many biomarkers identified through bioinformatics do not transition to clinical practice due to issues with validation, regulatory hurdles, and the necessity to align computational findings with patient care. Comprehensive experimental validation, typically requiring large and well-documented patient cohorts, is crucial for bioinformatics insights, yet such groups are frequently scarce [344]. Additionally, obtaining US FDA regulatory approval is a lengthy and expensive endeavor that involves multiple clinical trials. The gap between bioinformatics tools used for discovery and those available in clinical laboratories adds further challenges, as computational methods need to be reproducible, transparent, and pertinent for clinical application. This situation is exacerbated by insufficient collaboration between bioinformaticians and clinicians, hindering the effective

prioritization of clinically significant discoveries. To overcome these obstacles, enhanced interdisciplinary collaboration and more efficient validation and regulatory processes are essential for successfully integrating bioinformatics-derived biomarkers into clinical practice.

7.4 | Clinical Integration of Omics Data in Precision Oncology

The integration of single-cell and spatial omics data alongside clinical information represents a significant advancement in personalized cancer therapy [33]. This approach correlates molecular profiles with clinical outcomes, facilitating the discovery of new biomarkers and significantly enhancing the prediction of treatment responses. By utilizing various data sources, the researchers can generate actionable insights that directly inform targeted treatment strategies, ultimately resulting to more effective and customized patient interventions.

To fully utilize the potential of emerging technologies in precision oncology, it is crucial to pursue ongoing advancements in bioinformatics tools and analytical frameworks. Future research should focus on key areas such as integrating multiomics data, improving visualization techniques, and applying ML and AI [262]. Specifically, integrating multiomics data require creating advanced computational methods that combine information from genomics, transcriptomics, proteomics, and metabolomics, thus building a comprehensive model of cellular function and disease progression. This integrated approach deepens our understanding of cancer biology and helps identify clinically relevant biomarkers. Moreover, enhancing visualization capabilities is vital for interpreting high-dimensional and complex datasets, making them more understandable and accessible for researchers and clinicians. By advancing visualization and computational methods, bioinformatics tools can yield deeper biological insights, improving patient stratification, diagnostic precision, and therapeutic decision-making. Tackling these bioinformatics challenges will enhance our grasp of cancer mechanisms and accelerate the translation of multiomics findings into practical clinical applications, thereby driving the progress of personalized healthcare [23].

Researchers in precision medicine emphasize the necessity of standardized frameworks and forward-looking clinical studies for ensuring reproducibility, quality assurance, and clinical relevance in bioinformatics methods [345]. By combining various data types, such as genomics, imaging, proteomics, and electronic health records (EHRs), there is a considerable opportunity to create and validate AI-based medical models that improve predictive diagnostics, disease categorization, and treatment strategies [346]. These standardized frameworks enhance reproducibility by setting validated protocols for bioinformatics processes and AI models, while ensuring adherence to regulations and reliability in evaluations. Furthermore, incorporating diverse data strengthens the generalizability of AI, mitigates bias, and facilitates the transfer of computational insights into practical clinical applications. However, challenges remain, including the intricacy and heterogeneity of multimodal data, the requirement for comprehensive, well-annotated clinical datasets, and the significant computational power needed for integration and

model training [347]. Ethical and privacy issues regarding patient data also demand secure methods of sharing, like federated learning, to safeguard sensitive information while promoting collaborative research [269]. Additionally, regulatory challenges pose significant barriers to the validation and application of AI-powered clinical decision-support tools, making it necessary for collaborative efforts to standardize approaches across institutions and regulatory agencies [348].

Despite the existence of these challenges, addressing them through interdisciplinary collaboration among clinicians, bioinformaticians, data scientists, and regulatory experts is necessary to develop robust, interpretable, and clinically useful AI models [349]. While hurdles continue to persist, the advantages of precision medicine grounded in bioinformatics, such as enhanced patient stratification, biomarker identification, and personalized treatment options, are substantial. The ongoing innovation and advancement of these methodologies will continue to drive progress in cancer diagnostics and treatment strategies, ultimately transforming the future of personalized oncology care.

7.5 | Ethical and Privacy Considerations in the Practical Application of Bioinformatics Tools in Precision Oncology

The integration of bioinformatics with AI and ML algorithms in clinical practice raises several vital ethical and privacy concerns [350]. Addressing these issues is crucial for protecting patient welfare, ensuring equitable access to care, and promoting the responsible use of advanced technologies. Ethical challenges in bioinformatics include data privacy, the importance of informed consent, data sharing, potential misuse of genetic data, the necessity for transparency in result interpretation, and algorithmic bias in AI and ML systems [351]. Establishing these ethical frameworks is critical for the effective use of biomarkers and designing personalized medicine while maintaining ethical standards.

Advanced sequencing and bioinformatics technologies produce massive amounts of genomic and clinical information, which risks confidentiality breaches if not properly safeguarded [352]. To preserve patient privacy and secure sensitive information, it is essential to anonymize or de-identify data before sharing or publication [353]. However, the unique nature of genomic data raises concerns about achieving complete anonymization. This challenge requires the implementation of robust protocols and advanced computational tools to thwart reidentification attempts. Data encryption is another vital measure, ensuring safe transmission and storage of patient information [354]. Additionally, encryption tools help guard against unauthorized access by staff or external parties. To enhance these efforts, enforcing strict access controls, using secure servers, following regulatory frameworks like GDPR (General Data Protection Regulation) and HIPAA (Health Insurance Portability and Accountability Act), and allowing only authorized personnel access to sensitive information can significantly mitigate the risk of privacy violations [355].

Informed consent is a key ethical issue in biomedical research [352]. Patients whose data are used in such studies must receive

detailed information regarding the study's purpose, the use of their biological samples and personal data, and any associated risks and benefits [356]. In bioinformatics, obtaining meaningful consent requires addressing the complexities of information sharing, broad consent, and the need for re-consent [357]. Effectively communicating complex genomic data and analysis to participants can be challenging. Thus, researchers should create explicit, accessible consent materials free of jargon, ensuring that participants fully understand the study's objectives and methodologies. Some researchers may opt for broad consent or move to dynamic consent, which supports long-term scientific development and enables the future use of participants' data in new research [358]. Furthermore, this approach can enhance scientific exploration, while necessitating a strong ethical commitment to respect participants' autonomy, including their right to withdraw consent at any moment during the study.

Alongside privacy concerns, sharing data and ensuring equitable access to precision oncology studies and analyses are crucial [359]. Data sharing is vital in bioinformatics but requires solid agreements to maintain ethical standards and protect participants' rights [360]. Repositories such as dbGap (database of genotypes and phenotypes) offer dual access: open-access data for transparency and a controlled-access system for participant privacy protection [361]. Controlled access allows researchers to responsibly use sensitive data while ensuring participant confidentiality, thus aligning the benefits of shared data with ethical responsibilities. Precision oncology must ensure that innovations based on biomarkers are available to all patient demographics, including those from low socioeconomic or underrepresented backgrounds [362]. Social justice should guide funding and resource allocation decisions to prevent healthcare disparities. Additionally, predictive models, including AI and ML algorithms, play a vital role in bioinformatics, yet they also raise significant ethical issues that require careful consideration [363]. A key problem with AI systems is the biases inherent in training data, which can involve missing information, unrecognized patients (algorithmic bias), and sample size challenges and misclassification. For example, vulnerable populations, such as those with low socioeconomic status, psychosocial obstacles, and immigrants, often experience nonrandom data omissions in healthcare systems, leading to gaps in EHRs (e.g., absence of diagnostic tests, chronic illness medications, or social factors like housing instability) [351, 363]. Such deficiencies can lead ML algorithms to misinterpret existing data or exclude at-risk individuals from clinical tools designed for early intervention. Furthermore, EHRs might miss data on specific elements critical for enhancing health outcomes in these populations. Tackling these biases necessitates employing fairness-aware algorithms, utilizing diverse training datasets, and intentionally integrating metadata to prevent discriminatory treatment outcomes that could worsen socioeconomic healthcare disparities or disproportionately affect certain groups [363, 364]. Key steps include identifying or defining the target population, choosing training/testing datasets that embody this diversity, developing and validating algorithms across varied healthcare contexts, rigorously examining potential discriminatory patterns during data processing, establishing feedback mechanisms for validating output, and prioritizing clinically meaningful results instead of just performance benchmarks. These strategies are intended to promote equitable healthcare delivery and ensure that ML-based

algorithms and tools effectively serve all populations, especially those typically marginalized in data representation [360].

The ethical aspects of bioinformatics are complex, influenced by technical details and societal duties. A key concern is the lack of clarity in AI models, which can hide decision-making processes. It is crucial to ensure transparency and accountability in AI-driven decisions to uphold trust, particularly given bioinformatics' impact on research outcomes. Additionally, ethical considerations require careful attention to essential elements of ethical bioinformatics practices, including data privacy, informed consent, and responsible data sharing, alongside measures to prevent the misuse of participants' genetic information [352, 360].

8 | Future Directions

The prospective trajectory of bioinformatics research in precision oncology holds transformative possibilities, fueled by progress in computational power, AI, ML, and collective data-sharing platforms. Significant advancements are driven by refined computational methods, the integration of AI and ML, and strengthened collaboration to establish a cohesive framework for discovering biomarkers and strategies in precision medicine [365].

With the increasing reliance on data in cancer research, the future of biomarker discovery and therapeutic advancements depends on utilizing advanced computational methods and state-of-the-art technologies. Scalable cloud computing solutions like AWS, Google Cloud, and Microsoft Azure, coupled with distributed systems for parallel processing, will significantly enhance the efficiency of large-scale omics data analysis. User-friendly tools such as Galaxy, Seven Bridges, and DNAnexus are poised to optimize complex data workflows, while innovative single-cell technologies and spatial omics approaches offer exceptional insights into cellular dynamics and the tumor microenvironment, opening new avenues for discovering biomarkers and drug targets [366].

Rapid advancements in ML and AI are transforming the future of precision oncology [367]. Advanced deep learning models, utilizing multiomics datasets, are anticipated to uncover new biomarkers, predict responses to treatment, and identify actionable therapeutic targets with greater precision. By integrating AI into clinical workflows, oncologists will be equipped to make personalized, data-driven treatment choices that reflect each patient's molecular profile [368]. Additionally, the role in enhancing prevention, screening, and prognostic predictions for cancers holds significant potential for early intervention and better patient outcomes. Collaborative efforts and open data-sharing approaches are critical for progressing cancer research. Public databases such as TCGA, ICGC, and GEO will continue to enable cross-cohort validation and reproducibility, while initiatives like the Global Alliance for Genomics and Health and ELIXIR will develop strong data-sharing standards [369, 370]. Platforms like cBioPortal and XenaBrowser are poised to improve data visualization and collaborative research, encouraging cooperation among academic scientists, industry innovators, and healthcare professionals [283, 371].

In the future, the integration of advanced computational technologies, AI/ML advancements, and collaborative environments will transform biomarker discovery and treatment strategies in precision oncology. These innovations will significantly enhance our understanding of cancer biology, increase the reliability of biomarkers, and enhance patient outcomes with more effective, personalized therapies, leading to a transformative phase in cancer care.

9 | Limitations of Bioinformatics Tools in Precision Oncology

Bioinformatics tools utilized across diverse omics levels in precision oncology face many limitations that negatively impact their effectiveness in biomarker discovery [16]. Genomic instruments like DESeq2 for gene expression analysis and GATK for variant calling require considerable computational resources and show performance variability, especially at extreme expression levels. Similarly, functional genomics tools such as GSEA and KEGG often struggle with context specificity due to differing pathway relevance, which is affected by gene annotations and the limitations of related databases.

Epigenomic tools, such as Bismark for DNA methylation and MACS for chromatin accessibility are sensitive to sequencing errors. They face challenges in repetitive regions and situations with low-signal data. Similarly, proteomics and metabolomics tools, such as MaxQuant and XCMS, struggle to detect low-abundance analytes and rely on high-quality reference libraries. On the other hand, lipidomics tools like LipidSearch require specialized standards to manage the complexity of lipid diversity effectively [31, 372].

Transcriptomic tools for both bulk and single-cell RNA-seq, including Salmon and Seurat, encounter difficulties due to noise and dropout rates, which hinder accurate quantification and annotation [373]. Integrative multiomics tools are crucial for combining diverse data but demand significant computational resources and sophisticated interpretation for effective integration. Furthermore, cancer-specific tools such as cBioPortal and TIMER, which play a crucial role in biomarker identification and immune infiltration analysis, often struggle with limited dataset availability. They may also miss rare mutations and tumor heterogeneity, key factors essential for personalized medicine oncology [23].

10 | Conclusion

Bioinformatics is crucial for discovering biomarkers and developing therapeutic strategies in precision oncology by allowing the analysis of intricate, high-throughput datasets [16]. By integrating multiomics data—including genomics, transcriptomics, proteomics, epigenomics, and metabolomics—researchers can decipher molecular mechanisms underlying cancer and identify clinically significant biomarkers for diagnosis and prognosis. For example, genomic biomarkers such as *EGFR* and *BRCA1/2* mutations guide targeted therapies like *EGFR* inhibitors (erlotinib, gefitinib) for NSCLC and PARP inhibitors (olaparib, niraparib) for *BRCA1/2*-mutated breast and ovarian cancers [14, 374].

Additionally, epigenetic biomarkers like *MLH1* methylation assist in optimizing treatment selection [12]. Advances in these areas have also enhanced immunotherapy, with immune checkpoint inhibitors (pembrolizumab, nivolumab) targeting the *PD-1/PD-L1* axis to enhance antitumor responses [375]. HER2-targeted therapies, such as trastuzumab and pertuzumab, have been instrumental in treating HER2-amplified breast and gastric cancers by blocking HER2 signaling and inhibiting tumor growth [376].

In addition to single-agent therapies, bioinformatics insights have facilitated the development of combination strategies that enhance efficacy by simultaneously targeting multiple pathways [377]. For example, pembrolizumab is administered with chemotherapy for NSCLC, trastuzumab is combined with pertuzumab for HER2-positive breast cancer, bevacizumab is combined with chemotherapy for metastatic colorectal cancer, and bortezomib is used with dexamethasone for multiple myeloma [378–381]. These cotargeting methods utilize network-based analyses to enhance treatment regimens and improve clinical outcomes. The transformative role of bioinformatics is further strengthened by ML and AI, which automate complex data analyses, uncover hidden patterns, and enhance predictive biomarker discovery [382]. For instance, the overexpression of secreted oncoproteins like human growth hormone (*hGH*) and trefoil factor family (*TFF*) proteins has been recognized as a significant factor in cancer progression through ML-based pattern recognition [328]. Integrating multiomics platforms and employing network-based modeling has facilitated the discovery of robust biomarkers across various molecular layers and informed therapeutic strategies by connecting these biomarkers to actionable pathways [383]. An illustrative example is the use of synthetic lethality-based approaches to exploit weaknesses in DNA repair pathways with PARP inhibitors (olaparib, rucaparib, niraparib, and talazoparib) in cancers harboring *BRCA* or *ATM* mutations [384]. Additionally, emerging technologies such as cloud computing, distributed systems, single-cell analysis, and spatial omics continue to enhance tumor profiling and heterogeneity evaluation [385, 386]. Platforms like TCGA and cBioPortal harness cloud-based bioinformatics, while FireCloud supports distributed computing for extensive data analysis [283]. Techniques like single-cell RNA-seq (*scRNA-seq*) and spatial transcriptomics (e.g., CyTOF) enable detailed tumor mapping, allowing for the observation of cellular interactions within the tumor microenvironment [387].

Looking ahead, ML and AI are expected to transform biomarker discovery and therapeutic design by improving our understanding of tumor evolution, drug response, and resistance mechanisms. With the rise of collaborative efforts and open data-sharing initiatives, the clinical application of bioinformatics-driven discoveries is expected to accelerate, ultimately advancing precision oncology through the development of more effective, personalized therapies for patient care.

Author Contributions

T.W., V.B., and V.P. contributed to the conceptualization, data curation, and formal analysis of the review. T.W. and V.B. were responsible for writing the original draft preparation, visualization, and methodology. V.P. reviewed and edited the manuscript. V.P. provided supervision, project

administration, and resources and secured funding for the study. All the authors have read and approved the final version of the manuscript.

Acknowledgments

We acknowledge the Institute of Biopharmaceutical and Health Engineering, Tsinghua Shenzhen International Graduate School, China for administrative support. This work utilized Grammarly (www.grammarly.com) for language editing to ensure grammatical precision and improve overall readability. BioRender.com has been utilized for figure preparation.

Conflicts of Interest

The authors declare that this research was conducted in the absence of any commercial or financial relationships that could be construed as potential conflicts of interest.

Ethics Statement

The authors have nothing to report.

Data Availability Statement

The data that support the findings of this study are openly available in toothgennet-anterior at <https://github.com/superbijk/toothgennet-anterior/>.

References

1. G. Molla, M. Bitew, “Revolutionizing Personalized Medicine: Synergy With Multi-Omics Data Generation, Main Hurdles, and Future Perspectives,” *Biomedicines* 12, no. 12 (2024): 2750.
2. R. C. Wang, Z. Wang, “Precision Medicine: Disease Subtyping and Tailored Treatment,” *Cancers* 15, no. 15 (2023): 3837.
3. R. Vitorino, “Transforming Clinical Research: The Power of High-Throughput Omics Integration,” *Proteomes* 12, no. 3 (2024): 25.
4. P. Krzyszczyk, A. Acevedo, E. J. Davidoff, et al., “The Growing Role of Precision and Personalized Medicine for Cancer Treatment,” *Technology (Singap World Science)* 6, no. 3-4 (2018): 79-100.
5. A. Passaro, M. Al Bakir, E. G. Hamilton, “Cancer Biomarkers—Emerging Trends and Clinical Implications for Personalized Treatment,” *Cell* 187, no. 7 (2024): 1617-1635.
6. H. Xu, Z. Jia, F. Liu, et al., “Biomarkers and Experimental Models for Cancer Immunology Investigation,” *MedComm* 4, no. 6 (2023): e437.
7. Y. Zhou, L. Tao, J. Qiu, et al., “Tumor Biomarkers for Diagnosis, Prognosis and Targeted Therapy,” *Signal Transduction and Targeted Therapy* 9, no. 1 (2024): 132.
8. Unravelling the Role of Biomarker in Cancer Detection: An In-Depth Review | Current Pharmacology Reports. Accessed April 29, 2025. <https://link.springer.com/article/10.1007/s40495-025-00413-2>
9. R. U. Quraish, T. Hirahata, A. U. Quraish, “Ul Quraish S. An Overview: Genetic Tumor Markers for Early Detection and Current Gene Therapy Strategies,” *Cancer Inform* 22 (2023): 11769351221150772.
10. C. Sha, P. C. Lee, “EGFR-Targeted Therapies: A Literature Review,” *Journal of Clinical Medicine* 13, no. 21 (2024): 6391.
11. K. Kamińska, E. Nalejska, M. Kubiak, et al., “Prognostic and Predictive Epigenetic Biomarkers in Oncology,” *Molecular Diagnosis & Therapy* 23, no. 1 (2019): 83-95.
12. T. Wolde, J. Huang, P. Huang, V. Pandey, P. Qin, “Depleted-MLH1 Expression Predicts Prognosis and Immunotherapeutic Efficacy in Uterine Corpus Endometrial Cancer: An in Silico Approach,” *BioMedInformatics* 4, no. 1 (2024): 326-346.
13. Y. Zhang, J. Wu, C. Zhao, S. Zhang, J. Zhu, “Recent Advancement of PD-L1 Detection Technologies and Clinical Applications in the Era of Precision Cancer Therapy,” *Journal of Cancer* 14, no. 5 (2023): 850-873.
14. A. Ragupathi, M. Singh, A. M. Perez, D. Zhang, “Targeting the BRCA1/2 Deficient Cancer With PARP Inhibitors: Clinical Outcomes and Mechanistic Insights,” *Frontiers in Cell and Developmental Biology* 11 (2023): 1133472.
15. A. M. Bode, Z. Dong, “Recent Advances in Precision Oncology Research,” *Npj Precision Oncology* 2, no. 1 (2018): 1-6.
16. A. J. Clark, J. W. Lillard, “A Comprehensive Review of Bioinformatics Tools for Genomic Biomarker Discovery Driving Precision Oncology,” *Genes (Basel)* 15, no. 8 (2024): 1036.
17. S. Dotolo, R. Esposito Abate, C. Roma, “Bioinformatics: From NGS Data to Biological Complexity in Variant Detection and Oncological Clinical Practice,” *Biomedicines* 10, no. 9 (2022): 2074.
18. T. Wolde, V. Bhardwaj, M. Reyad-UI-Ferdous, P. Qin, V. Pandey, “The Integrated Bioinformatic Approach Reveals the Prognostic Significance of LRP1 Expression in Ovarian Cancer,” *International Journal of Molecular Sciences* 25, no. 14 (2024): 7996.
19. H. Zhao, X. Zhang, L. Guo, S. Shi, C. Lu, “A Robust Seven-Gene Signature Associated with Tumor Microenvironment to Predict Survival Outcomes of Patients with Stage III–IV Lung Adenocarcinoma,” *Frontiers in Genetics* 12 (2021): 684281.
20. V. P. Snijesh, S. Krishnamurthy, V. Bhardwaj, et al., “SHH Signaling as a Key Player in Endometrial Cancer: Unveiling the Correlation With Good Prognosis, Low Proliferation, and Anti-Tumor Immune Milieu,” *International Journal of Molecular Sciences* 25, no. 19 (2024): 10443.
21. D. Rosati, M. Palmieri, G. Brunelli, et al., “Differential Gene Expression Analysis Pipelines and Bioinformatic Tools for the Identification of Specific Biomarkers: A Review,” *Computational and Structural Biotechnology Journal* 23 (2024): 1154-1168.
22. M. Adnane, A. M. de Almeida, A. Chapwanya, “Unveiling the Power of Proteomics in Advancing Tropical Animal Health and Production,” *Tropical Animal Health and Production* 56, no. 5 (2024): 182.
23. C. Chen, J. Wang, D. Pan, et al., “Applications of Multi-omics Analysis in human Diseases,” *MedComm* 4, no. 4 (2023): e315.
24. I. Subramanian, S. Verma, S. Kumar, A. Jere, K. Anamika, “Multi-omics Data Integration, Interpretation, and Its Application,” *Bioinform Biol Insights* 14 (2020): 1177932219899051.
25. C. J. Cremin, S. Dash, X. Huang, “Big Data: Historic Advances and Emerging Trends in Biomedical Research,” *Current Research in Biotechnology* 4 (2022): 138-151.
26. R. J. Woodman, A. A. Mangoni, “A Comprehensive Review of Machine Learning Algorithms and Their Application in Geriatric Medicine: Present and Future,” *Aging Clinical and Experimental Research* 35, no. 11 (2023): 2363-2397.
27. H. Shimizu, K. I. Nakayama, “Artificial Intelligence in Oncology,” *Cancer Science* 111, no. 5 (2020): 1452-1460.
28. M. V. C. Santos, A. S. Feltrin, I. C. Costa-Amaral, et al., “Network Analysis of Biomarkers Associated With Occupational Exposure to Benzene and Malathion,” *International Journal of Molecular Sciences* 24, no. 11 (2023): 9415.
29. R. Asimwe, S. Lam, Leung, et al., “From Biobank and Data Silos Into a Data Commons: Convergence to Support Translational Medicine,” *Journal of Translational Medicine* 19 (2021): 493.
30. V. Brancato, G. Esposito, L. Coppola, et al., “Standardizing Digital Biobanks: Integrating Imaging, Genomic, and Clinical Data for Precision Medicine,” *Journal of Translational Medicine* 22, no. 1 (2024): 136.
31. H. Satam, K. Joshi, U. Mangrolia, et al., “Next-Generation Sequencing Technology: Current Trends and Advancements,” *Biology (Basel)* 12, no. 7 (2023): 997.
32. E. Papadaki, I. Kakkos, P. Vlamos, et al., “Recent Web Platforms for Multi-Omics Integration Unlocking Biological Complexity,” *Applied Sciences* 15, no. 1 (2025): 329.

33. A. Jose, P. Kulkarni, J. Thilakan, et al., "Integration of Pan-omics Technologies and Three-dimensional in Vitro Tumor Models: An Approach Toward Drug Discovery and Precision Medicine," *Molecular Cancer* 23, no. 1 (2024): 50.
34. Y. Qin, Y. Liu, X. Xiang, et al., "Cuproptosis Correlates With Immunosuppressive Tumor Microenvironment Based on Pan-cancer Multiomics and Single-cell Sequencing Analysis," *Molecular Cancer* 22, no. 1 (2023): 59.
35. F. Menghi, E. T. Liu, "Functional Genomics of Complex Cancer Genomes," *Nature Communications* 13, no. 1 (2022): 5908.
36. S. E. Levy, B. E. Boone, "Next-Generation Sequencing Strategies," *Cold Spring Harbor Perspectives in Medicine* 9, no. 7 (2019): a025791.
37. Y. Tang, Y. Fan, "Combined KRAS and TP53 Mutation in Patients With Colorectal Cancer Enhance Chemoresistance to Promote Postoperative Recurrence and Metastasis," *BMC Cancer* 24 (2024): 1155.
38. C. Peretto, M. Aprile, S. Cataldi, E. Giovannetti, V. Costa, "Unraveling BRAF Alterations: Molecular Insights to Circumvent Therapeutic Resistance Across Cancer Types," *Cancer Drug Resistance* 8 (2025): 14.
39. A. McKenna, M. Hanna, E. Banks, et al., "The Genome Analysis Toolkit: A MapReduce Framework for Analyzing next-generation DNA Sequencing Data," *Genome Research* 20, no. 9 (2010): 1297-1303.
40. K. Cibulskis, M. S. Lawrence, S. L. Carter, et al., "Sensitive Detection of Somatic Point Mutations in Impure and Heterogeneous Cancer Samples," *Nature Biotechnology* 31, no. 3 (2013): 213-219.
41. K. Wang, M. Li, H. Hakonarson, "ANNOVAR: Functional Annotation of Genetic Variants From High-throughput Sequencing Data," *Nucleic Acids Research* 38, no. 16 (2010): e164.
42. M. Arigoni, M. L. Ratto, F. Riccardo, et al., "A Single Cell RNAseq Benchmark Experiment Embedding "Controlled" Cancer Heterogeneity," *Scientific Data* 11, no. 1 (2024): 159.
43. D. Sun, M. Liu, F. Huang, F. Huang, "[Bioinformatics analysis of expression and function of EXD3 gene in gastric cancer]," *Nan Fang Yi Ke Da Xue Xue Bao* 39, no. 2 (2019): 215-221.
44. K. Yan, Q. Z. Liu, R. R. Huang, et al., "Spatial Transcriptomics Reveals Prognosis-associated Cellular Heterogeneity in the Papillary Thyroid Carcinoma Microenvironment," *Clinical and Translational Medicine* 14, no. 3 (2024): e1594.
45. Y. Ma, X. Zhou, "Spatially Informed Cell Type Deconvolution for Spatial Transcriptomics," *Nature Biotechnology* 40, no. 9 (2022): 1349-1359. Published online May 2, 2022.
46. S. Zhao, W. P. Fung-Leung, A. Bittner, K. Ngo, X. Liu, "Comparison of RNA-Seq and Microarray in Transcriptome Profiling of Activated T Cells," *PLoS ONE* 9, no. 1 (2014): e78644.
47. A. Mortazavi, B. A. Williams, K. McCue, L. Schaeffer, B. Wold, "Mapping and Quantifying Mammalian Transcriptomes by RNA-Seq," *Nature Methods* 5, no. 7 (2008): 621-628.
48. B. Zhao, B. Li, Y. Chen, et al., "CDKN2A is a Promising Diagnostic and Prognostic Biomarker and Associations With Immune Infiltrates in Colorectal Cancer," *Heliyon* Published online March 6, 2025:e43049, <https://doi.org/10.1016/j.heliyon.2025.e43049>.
49. H. El Ahanidi, M. El Azzouzi, B. Addoum, "STAT1 and STAT4 Expression as Prognostic Biomarkers in Patients With Bladder Cancer," *Molecular and Clinical Oncology* 22, no. 4 (2025): 1-7.
50. S. Zhang, K. Liu, Y. Liu, X. Hu, X. Gu, "The Role and Application of Bioinformatics Techniques and Tools in Drug Discovery," *Frontiers in Pharmacology* 16 (2025): 1547131.
51. A. Siavoshi, M. Taghizadeh, E. Dookhe, M. Piran, "Gene Expression Profiles and Pathway Enrichment Analysis to Identification of Differentially Expressed Gene and Signaling Pathways in Epithelial Ovarian Cancer Based on High-throughput RNA-seq Data," *Genomics* 114, no. 1 (2022): 161-170.
52. F. Finotello, E. Lavezzo, L. Bianco, et al., "Reducing Bias in RNA Sequencing Data: A Novel Approach to Compute Counts," *BMC Bioinformatics [Electronic Resource]* 15, no. Suppl 1 (2014): S7.
53. D. Chen, L. Xu, H. Xing, et al., "Sangerbox 2: Enhanced Functionalities and Update for a Comprehensive Clinical Bioinformatics Data Analysis Platform," *Imeta* 3, no. 5 (2024): e238.
54. F. Rohart, B. Gautier, A. Singh, K. A. Lê Cao, "mixOmics: An R Package for 'Omics Feature Selection and Multiple Data Integration," *Plos Computational Biology* 13, no. 11 (2017): e1005752.
55. R. Argelaguet, B. Velten, D. Arnol, et al., "Multi-Omics Factor Analysis-a Framework for Unsupervised Integration of Multi-omics Data Sets," *Molecular Systems Biology* 14, no. 6 (2018): e8124.
56. T. Stuart, A. Butler, P. Hoffman, et al., "Comprehensive Integration of Single-Cell Data," *Cell* 177, no. 7 (2019): 1888-1902. e21.
57. F. A. Wolf, P. Angerer, F. J. Theis, "SCANPY: Large-scale Single-cell Gene Expression Data Analysis," *Genome Biology* 19 (2018): 15.
58. Y. Cui, S. Zhang, Y. Liang, X. Wang, T. N. Ferraro, Y. Chen, "Consensus Clustering of Single-cell RNA-seq Data by Enhancing Network Affinity," *Briefings in Bioinformatics* 22, no. 6 (2021): bbab236.
59. M. Valerio, A. Inno, S. Gori, "pyBioPortal: A Python Package for Simplifying cBioPortal Data Access in Cancer Research," *JAMIA Open* 8, no. 1 (2025): ooa4146.
60. P. Brlek, A. Kafka, A. Bukovac, "Pecina-Šlaus N. Integrative cBioPortal Analysis Revealed Molecular Mechanisms That Regulate EGFR-PI3K-AKT-mTOR Pathway in Diffuse Gliomas of the Brain," *Cancers* 13, no. 13 (2021): 3247.
61. Patro R., G. Duggal, M. I. Love, R. A. Irizarry, C. Kingsford, "Salmon Provides Fast and Bias-aware Quantification of Transcript Expression," *Nature Methods* 14, no. 4 (2017): 417-419.
62. C. Robert, M. Watson, "Errors in RNA-Seq Quantification Affect Genes of Relevance to human Disease," *Genome Biology* 16, no. 1 (2015): 177.
63. S. Anders, P. T. Pyl, W. Huber, "HTSeq—a Python Framework to Work With High-throughput Sequencing Data," *Bioinformatics* 31, no. 2 (2015): 166-169.
64. L. A. Corchete, E. A. Rojas, D. Alonso-López, et al., "Systematic Comparison and Assessment of RNA-seq Procedures for Gene Expression Quantitative Analysis," *Scientific Reports* 10, no. 1 (2020): 19737.
65. S. J. Fleming, M. D. Chaffin, A. Arduini, et al., "Unsupervised Removal of Systematic Background Noise From Droplet-based Single-cell Experiments Using CellBender," *Nature Methods* 20, no. 9 (2023): 1323-1335.
66. J. Ding, A. Regev, "Deep Generative Model Embedding of Single-cell RNA-Seq Profiles on Hyperspheres and Hyperbolic Spaces," *Nature Communications* 12, no. 1 (2021): 2554.
67. R. Cannoodt, W. Saelens, L. Deconinck, Y. Saeys, "Spearheading Future Omics Analyses Using dyngen, a Multi-modal Simulator of Single Cells," *Nature Communications* 12, no. 1 (2021): 3942.
68. R. Cannoodt, W. Saelens, L. Deconinck, Y. Saeys. dyngen: a multi-modal simulator for spearheading new single-cell omics analyses. Published online September 14, 2020:2020.02.06.936971. <https://doi.org/10.1101/2020.02.06.936971>
69. Y. Cao, J. Zhu, P. Jia, Z. Zhao, "scRNASeqDB: A Database for RNA-Seq Based Gene Expression Profiles in Human Single Cells," *Genes (Basel)* 8, no. 12 (2017): 368.
70. X. Cui, F. Qin, X. Yu, F. Xiao, G. Cai, "SCISSORTM: A Single-cell Inferred Site-specific Omics Resource for Tumor Microenvironment Association Study," *NAR Cancer* 3, no. 3 (2021): zcab037.
71. M. Su, Z. Zhang, L. Zhou, C. Han, C. Huang, E. C. Nice, "Proteomics, Personalized Medicine and Cancer," *Cancers* 13, no. 11 (2021): 2512.

72. Z. Zhou, R. Zhang, A. Zhou, et al., "Proteomics Appending a Complementary Dimension to Precision Oncotherapy," *Computational and Structural Biotechnology Journal* 23 (2024): 1725-1739.
73. D. M. Berney, "Biomarkers for Prostate Cancer Detection and Progression: Beyond Prostate-specific Antigen," *Drug News & Perspectives* 23, no. 3 (2010): 185-194.
74. G. J. S. Rustin, M. E. L. van der Burg, C. L. Griffin, et al., "Early versus Delayed Treatment of Relapsed Ovarian Cancer (MRC OV05/EORTC 55955): A Randomised Trial," *Lancet* 376, no. 9747 (2010): 1155-1163.
75. S. Wu, S. Zhang, C. M. Liu, A. R. Fernie, S. Yan, "Recent Advances in Mass Spectrometry-Based Protein Interactome Studies," *Molecular & Cellular Proteomics* 24, no. 1 (2025): 100887.
76. J. Li, H. J. Zhu, "Liquid Chromatography-Tandem Mass Spectrometry (LC-MS/MS)-Based Proteomics of Drug-Metabolizing Enzymes and Transporters," *Molecules (Basel, Switzerland)* 25, no. 11 (2020): 2718.
77. O. Stoevesandt, M. J. Taussig, M. He, "Protein Microarrays: High-throughput Tools for Proteomics," *Expert Review of Proteomics* 6, no. 2 (2014): 145-157.
78. A. Palomba, M. Abbondio, G. Fiorito, S. Uzzau, D. Pagnozzi, A. Tanca, "Comparative Evaluation of MaxQuant and Proteome Discoverer MS1-Based Protein Quantification Tools," *Journal of Proteome Research* 20, no. 7 (2021): 3497-3507.
79. B. MacLean, D. M. Tomazela, N. Shulman, et al., "Skyline: An Open Source Document Editor for Creating and Analyzing Targeted Proteomics Experiments," *Bioinformatics* 26, no. 7 (2010): 966-968.
80. B. C. Orsburn, "Proteome Discoverer-A Community Enhanced Data Processing Suite for Protein Informatics," *Proteomes* 9, no. 1 (2021): 15.
81. Y. M. Farag, C. Horro, M. Vaudel, H. Barsnes, "PeptideShaker Online: A User-Friendly Web-Based Framework for the Identification of Mass Spectrometry-Based Proteomics Data," *Journal of Proteome Research* 20, no. 12 (2021): 5419-5423.
82. T. D. Müller, A. Siraj, A. Walter, et al., "OpenMS WebApps: Building User-Friendly Solutions for MS Analysis," *Journal of Proteome Research* 24, no. 2 (2025): 940-948. Published online January 30, 2025.
83. OpenMS. Accessed April 30, 2025. <https://openms.de/>
84. D. Szklarczyk, A. L. Gable, D. Lyon, et al., "STRING v11: Protein-protein Association Networks With Increased Coverage, Supporting Functional Discovery in Genome-wide Experimental Datasets," *Nucleic Acids Research* 47, no. Database issue (2019): D607-D613.
85. N. T. Doncheva, J. H. Morris, J. Gorodkin, L. J. Jensen, "Cytoscape StringApp: Network Analysis and Visualization of Proteomics Data," *Journal of Proteome Research* 18, no. 2 (2019): 623-632.
86. R. Oughtred, C. Stark, B. J. Breitkreutz, et al., "The BioGRID Interaction Database: 2019 Update," *Nucleic Acids Research* 47, no. D1 (2019): D529-D541.
87. P. V. Hornbeck, J. M. Kornhauser, S. Tkachev, et al., "PhosphoSitePlus: A Comprehensive Resource for Investigating the Structure and Function of Experimentally Determined Post-translational Modifications in man and Mouse," *Nucleic Acids Research* 40, no. Database issue (2012): D261-270.
88. S. Rosignoli, M. Pacelli, F. Manganiello, A. Paiardini, "An Outlook on Structural Biology After AlphaFold: Tools, Limits and Perspectives," *FEBS Open Bio* 15, no. 2 (2025): 202-222.
89. J. Yang, Y. Zhang, "Protein Structure and Function Prediction Using I-TASSER," *CP in Bioinformatics* 52 (2015): 5.8.1-5.8.15.
90. Y. Chen, H. Zhang, W. Wang, Y. Shen, Z. Ping, "Rapid Generation of High-quality Structure Figures for Publication With PyMOL-PUB," *Bioinformatics* 40, no. 3 (2024): btae139.
91. H. Takeda, Y. Matsuzawa, M. Takeuchi, et al., "MS-DIAL 5 Multimodal Mass Spectrometry Data Mining Unveils Lipidome Complexities," *Nature Communications* 15, no. 1 (2024): 9903.
92. V. Demichev, L. Szyrwił, F. Yu, et al., "dia-PASEF Data Analysis Using FragPipe and DIA-NN for Deep Proteomics of Low Sample Amounts," *Nature Communications* 13, no. 1 (2022): 3944.
93. N. H. Tran, X. Zhang, L. Xin, B. Shan, M. Li, "De Novo Peptide Sequencing by Deep Learning," *Proceedings of the National Academy of Sciences* 114, no. 31 (2017): 8247-8252.
94. S. H. Yu, D. Ferretti, J. P. Schessner, J. D. Rudolph, G. H. H. Börner, J. Cox, "Expanding the Perseus Software for Omics Data Analysis with Custom Plugins," *Current Protocols in Bioinformatics* 71, no. 1 (2020): e105.
95. Y. Chen, L. Chen, A. T. L. Lun, P. L. Baldoni, G. K. Smyth, "edgeR v4: Powerful Differential Analysis of Sequencing Data With Expanded Functionality and Improved Support for Small Counts and Larger Datasets," *Nucleic Acids Research* 53, no. 2 (2025): gkaf018.
96. M. Gu, B. Ren, Y. Fang, et al., "Epigenetic Regulation in Cancer," *MedComm* 5, no. 2 (2024): e495.
97. K. Struhl, "The Distinction Between Epigenetics and Epigenomics," *Trends in Genetics* 40, no. 12 (2024): 995-997.
98. A. P. Feinberg, M. A. Koldobskiy, A. Göndör, "Epigenetic Modulators, Modifiers and Mediators in Cancer Aetiology and Progression," *Nature Reviews Genetics* 17, no. 5 (2016): 284-299.
99. W. Zhang, L. Qi, Z. Liu, et al., "Integrated Multiomic Analysis and High-throughput Screening Reveal Potential Gene Targets and Synergetic Drug Combinations for Osteosarcoma Therapy," *MedComm* 4, no. 4 (2023): e317.
100. X. Chen, H. Xu, X. Shu, C. X. Song, "Mapping Epigenetic Modifications by Sequencing Technologies," *Cell Death and Differentiation* 32, no. 1 (2025): 56-65.
101. M. Kernaléguen, C. Daviaud, Y. Shen, et al., "Whole-Genome Bisulfite Sequencing for the Analysis of Genome-Wide DNA Methylation and Hydroxymethylation Patterns at Single-Nucleotide Resolution," *Methods in Molecular Biology* 1767 (2018): 311-349.
102. C. C. Pritchard, H. H. Cheng, M. Tewari, "MicroRNA Profiling: Approaches and Considerations," *Nature Reviews Genetics* 13, no. 5 (2012): 358-369.
103. T. S. Furey, "ChIP-seq and Beyond: New and Improved Methodologies to Detect and Characterize Protein-DNA Interactions," *Nature Reviews Genetics* 13, no. 12 (2012): 840-852.
104. R. Prabhakaran, R. Thamarai, S. Sivasamy, et al., "Epigenetic Frontiers: MiRNAs, Long Non-coding RNAs and Nanomaterials Are Pioneering to Cancer Therapy," *Epigenetics & Chromatin* 17, no. 1 (2024): 31.
105. V. Bhardwaj, Y. Q. Tan, M. M. Wu, et al., "Long Non-coding RNAs in Recurrent Ovarian Cancer: Theranostic Perspectives," *Cancer Letters* 502 (2021): 97-107.
106. M. M. Kadhim, A. A. Ramírez-Coronel, A. T. Jalil, et al., "Autophagy as a Self-digestion Signal in human Cancers: Regulation by microRNAs in Affecting Carcinogenesis and Therapy Response," *Pharmacological Research* 189 (2023): 106695.
107. S. Kansara, V. Pandey, P. E. Lobie, G. Sethi, M. Garg, A. K. Pandey, "Mechanistic Involvement of Long Non-Coding RNAs in Oncotherapeutics Resistance in Triple-Negative Breast Cancer," *Cells* 9, no. 6 (2020): 1511.
108. M. A. Zandieh, M. H. Farahani, R. Rajabi, "Epigenetic Regulation of Autophagy by Non-coding RNAs in Gastrointestinal Tumors: Biological Functions and Therapeutic Perspectives," *Pharmacological Research* 187 (2023): 106582.
109. M. Garg, G. Sethi, "Emerging Role of Long Non-coding RNA (lncRNA) in human Malignancies: A Unique Opportunity for Precision Medicine," *Cancer Letters* 519 (2021): 1.
110. R. Elango, V. Radhakrishnan, S. Rashid, et al., "Long Noncoding RNA Profiling Unveils LINC00960 as Unfavorable Prognostic Biomarker Pro-

- moting Triple Negative Breast Cancer Progression,” *Cell Death Discovery* 10, no. 1 (2024): 333.
111. F. Krueger, S. R. Andrews, “Bismark: A Flexible Aligner and Methylation Caller for Bisulfite-Seq Applications,” *Bioinformatics* 27, no. 11 (2011): 1571-1572.
112. A. Akalin, M. Kormaksson, S. Li, et al., “methylKit: A Comprehensive R Package for the Analysis of Genome-wide DNA Methylation Profiles,” *Genome Biology* 13, no. 10 (2012): R87.
113. H. Feng, K. N. Conneely, H. Wu, “A Bayesian Hierarchical Model to Detect Differentially Methylated Loci From Single Nucleotide Resolution Sequencing Data,” *Nucleic Acids Research* 42, no. 8 (2014): e69.
114. Y. Zhang, T. Liu, C. A. Meyer, et al., “Model-based Analysis of ChIP-Seq (MACS),” *Genome Biology* 9, no. 9 (2008): R137.
115. C. Zang, D. E. Schones, C. Zeng, K. Cui, K. Zhao, W. Peng, “A Clustering Approach for Identification of Enriched Domains From Histone Modification ChIP-Seq Data,” *Bioinformatics* 25, no. 15 (2009): 1952-1958.
116. J. Ernst, M. Kellis, “ChromHMM: Automating Chromatin-state Discovery and Characterization,” *Nature Methods* 9, no. 3 (2012): 215-216.
117. L. Shen, N. Y. Shao, X. Liu, I. Maze, J. Feng, E. J. Nestler, “diffReps: Detecting Differential Chromatin Modification Sites From ChIP-seq Data With Biological Replicates,” *PLoS ONE* 8, no. 6 (2013): e65598.
118. M. E. Ritchie, B. Phipson, D. Wu, et al., “limma Powers Differential Expression Analyses for RNA-sequencing and Microarray Studies,” *Nucleic Acids Research* 43, no. 7 (2015): e47.
119. M. D. Robinson, D. J. McCarthy, G. K. Smyth, “edgeR: A Bioconductor Package for Differential Expression Analysis of Digital Gene Expression Data,” *Bioinformatics* 26, no. 1 (2010): 139-140.
120. F. Sarno, G. Benincasa, M. List, et al., “Clinical Epigenetics Settings for Cancer and Cardiovascular Diseases: Real-life Applications of Network Medicine at the Bedside,” *Clinical Epigenetics* 13, no. 1 (2021): 66.
121. M. Friedemann, F. Horn, K. Gutewort, et al., “Increased Sensitivity of Detection of RASSF1A and GSTP1 DNA Fragments in Serum of Prostate Cancer Patients: Optimisation of Diagnostics Using OBBPA-ddPCR,” *Cancers (Basel)* 13, no. 17 (2021): 4459.
122. M. K. Skinner, “Epigenetic Biomarkers for Disease Susceptibility and Preventative Medicine,” *Cell Metabolism* 36, no. 2 (2024): 263-277.
123. X. Li, Y. Yang, B. Zhang, et al., “Lactate Metabolism in human Health and Disease,” *Signal Transduction and Targeted Therapy* 7, no. 1 (2022): 305.
124. S. Qiu, Y. Cai, H. Yao, et al., “Small Molecule Metabolites: Discovery of Biomarkers and Therapeutic Targets,” *Signal Transduction and Targeted Therapy* 8, no. 1 (2023): 132.
125. D. R. Schmidt, R. Patel, D. G. Kirsch, C. A. Lewis, M. G. Vander Heiden, J. W. Locasale, “Metabolomics in Cancer Research and Emerging Applications in Clinical Oncology,” *CA: A Cancer Journal for Clinicians* 71, no. 4 (2021): 333-358.
126. L. Dang, D. W. White, S. Gross, et al., “Cancer-associated IDH1 Mutations Produce 2-hydroxyglutarate,” *Nature* 462, no. 7274 (2009): 739-744.
127. J. H. Wang, W. L. Chen, J. M. Li, et al., “Prognostic Significance of 2-hydroxyglutarate Levels in Acute Myeloid Leukemia in China,” *Proceedings of the National Academy of Sciences* 110, no. 42 (2013): 17017-17022.
128. S. Pan, L. Yin, J. Liu, et al., “Metabolomics-driven Approaches for Identifying Therapeutic Targets in Drug Discovery,” *MedComm* 5, no. 11 (2024): e792.
129. W. Fu, A. Sun, H. Dai, “Lipid Metabolism Involved in Progression and Drug Resistance of Breast Cancer,” *Genes & Diseases* 12, no. 4 (2024): 101376.
130. W. Wang, Z. Rong, G. Wang, Y. Hou, F. Yang, M. Qiu, “Cancer Metabolites: Promising Biomarkers for Cancer Liquid Biopsy,” *Biomarker Research* 11, no. 1 (2023): 66.
131. F. Bhinderwala, N. Wase, C. DiRusso, R. Powers, “Combining Mass Spectrometry and NMR Improves Metabolite Detection and Annotation,” *Journal of Proteome Research* 17, no. 11 (2018): 4017-4022.
132. A. H. Emwas, R. Roy, R. T. McKay, et al., “NMR Spectroscopy for Metabolomics Research,” *Metabolites* 9, no. 7 (2019): 123.
133. A. H. M. Emwas, Z. A. Al-Talla, Y. Yang, N. M. Kharbatia, “Gas Chromatography-mass Spectrometry of Biofluids and Extracts,” *Methods in Molecular Biology* 1277 (2015): 91-112.
134. M. Ciborowski, A. Lipska, J. Godzien, et al., “Combination of LC-MS- and GC-MS-based Metabolomics to Study the Effect of Ozonated Autohemotherapy on human Blood,” *Journal of Proteome Research* 11, no. 12 (2012): 6231-6241.
135. H. P. Benton, D. M. Wong, S. A. Trauger, G. Siuzdak, “XCMS2: Processing Tandem Mass Spectrometry Data for Metabolite Identification and Structural Characterization,” *Analytical Chemistry* 80, no. 16 (2008): 6382.
136. W. Niu, E. Knight, Q. Xia, B. D. McGarvey, “Comparative Evaluation of Eight Software Programs for Alignment of Gas Chromatography-mass Spectrometry Chromatograms in Metabolomics Experiments,” *Journal of Chromatography A* 1374 (2014): 199-206.
137. Z. Pang, J. Chong, G. Zhou, et al., “MetaboAnalyst 5.0: Narrowing the Gap Between Raw Spectra and Functional Insights,” *Nucleic Acids Research* 49, no. W1 (2021): W388-W396.
138. Z. Pang, Y. Lu, G. Zhou, et al., “MetaboAnalyst 6.0: Towards a Unified Platform for Metabolomics Data Processing, Analysis and Interpretation,” *Nucleic Acids Research* 52, no. W1 (2024): W398-W406.
139. C. Ruttkies, E. L. Schymanski, S. Wolf, J. Hollender, S. Neumann, “MetFrag Relaunched: Incorporating Strategies Beyond in Silico Fragmentation,” *Journal of Cheminformatics* 8 (2016): 3.
140. L. F. Nothias, D. Petras, R. Schmid, et al., “Feature-Based Molecular Networking in the GNPS Analysis Environment,” *Nature Methods* 17, no. 9 (2020): 905-908.
141. B. Zhang, S. Hu, E. Baskin, A. Patt, J. K. Siddiqui, E. A. Mathé, “RaMP: A Comprehensive Relational Database of Metabolomics Pathways for Pathway Enrichment Analysis of Genes and Metabolites,” *Metabolites* 8, no. 1 (2018): 16.
142. D. S. Wishart, A. Guo, E. Oler, et al., “HMDB 5.0: The Human Metabolome Database for 2022,” *Nucleic Acids Research* 50, no. D1 (2022): D622-D631.
143. M. R. Belhaj, N. G. Lawler, N. J. Hoffman, “Metabolomics and Lipidomics: Expanding the Molecular Landscape of Exercise Biology,” *Metabolites* 11, no. 3 (2021): 151.
144. B. Xi, H. Gu, H. Baniyadi, D. Raftery, “Statistical Analysis and Modeling of Mass Spectrometry-Based Metabolomics Data,” *Methods in Molecular Biology* 1198 (2014): 333-353.
145. P. Castellano-Escuder, R. González-Domínguez, F. Carmona-Pontaque, C. Andrés-Lacueva, A. Sánchez-Pla, “POMAShiny: A User-friendly Web-based Workflow for Metabolomics and Proteomics Data Analysis,” *Plos Computational Biology* 17, no. 7 (2021): e1009148.
146. M. Lefouili, K. Nam, “The Evaluation of Bcftools Mpileup and GATK HaplotypeCaller for Variant Calling in Non-human Species,” *Scientific Reports* 12, no. 1 (2022): 11331.
147. C. Xu, “A Review of Somatic Single Nucleotide Variant Calling Algorithms for next-generation Sequencing Data,” *Computational and Structural Biotechnology Journal* 16 (2018): 15-24.
148. A. L. Gable, D. Szklarczyk, D. Lyon, J. F. Matias Rodrigues, C. von Mering, “Systematic Assessment of Pathway Databases, Based on a Diverse Collection of User-submitted Experiments,” *Brief Bioinform* 23, no. 5 (2022): bbac355.
149. D. Masood, L. Ren, C. Nguyen, et al., “Evaluation of Somatic Copy Number Variation Detection by NGS Technologies and Bioinformatics Tools on a Hyper-diploid Cancer Genome,” *Genome Biology* 25, no. 1 (2024): 163.

150. A. Merkel, M. Fernández-Callejo, E. Casals, et al., “gemBS: High Throughput Processing for DNA Methylation Data From Bisulfite Sequencing,” *Bioinformatics* 35, no. 5 (2019): 737-742.
151. S. Ma, B. Zhang, L. M. LaFave, et al., “Chromatin Potential Identified by Shared Single-Cell Profiling of RNA and Chromatin,” *Cell* 183, no. 4 (2020): 1103-1116. e20.
152. G. M. Richter, J. Kruppa, H. G. Keceli, et al., “Epigenetic Adaptations of the Masticatory Mucosa to Periodontal Inflammation,” *Clinical Epigenetics* 13, no. 1 (2021): 203.
153. P. L. Baldoni, N. U. Rashid, J. G. Ibrahim, “Efficient Detection and Classification of Epigenomic Changes Under Multiple Conditions,” *Biometrics* 78, no. 3 (2022): 1141-1154.
154. M. Vaudel, J. M. Burkhart, R. P. Zahedi, et al., “PeptideShaker Enables Reanalysis of MS-derived Proteomics Data Sets,” *Nature Biotechnology* 33, no. 1 (2015): 22-24.
155. D. Szklarczyk, A. L. Gable, K. C. Nastou, et al., “The STRING Database in 2021: Customizable Protein-protein Networks, and Functional Characterization of User-uploaded Gene/Measurement Sets,” *Nucleic Acids Research* 49, no. D1 (2021): D605-D612.
156. L. Chen, Q. Li, K. F. A. Nasif, et al., “AI-Driven Deep Learning Techniques in Protein Structure Prediction,” *International Journal of Molecular Sciences* 25, no. 15 (2024): 8426.
157. S. Hemmer, S. K. Manier, S. Fischmann, F. Westphal, L. Wagmann, M. R. Meyer, “Comparison of Three Untargeted Data Processing Workflows for Evaluating LC-HRMS Metabolomics Data,” *Metabolites* 10, no. 9 (2020): 378.
158. C. Ruttkies, S. Neumann, S. Posch, “Improving MetFrag With Statistical Learning of Fragment Annotations,” *BMC Bioinformatics [Electronic Resource]* 20, no. 1 (2019): 376.
159. A. Marco-Ramell, M. Palau-Rodríguez, A. Alay, et al., “Evaluation and Comparison of Bioinformatic Tools for the Enrichment Analysis of Metabolomics Data,” *BMC Bioinformatics [Electronic Resource]* 19, no. 1 (2018), <https://doi.org/10.1186/s12859-017-2006-0>.
160. C. Calderón, L. Rubarth, M. Cebo, I. Merfort, M. Lämmerhofer, “Lipid Atlas of Keratinocytes and Betulin Effects on Its Lipidome Profiled by Comprehensive UHPLC-MS/MS With Data Independent Acquisition Using Targeted Data Processing,” *Proteomics* 20, no. 11 (2020): e1900113.
161. D. Wang, Y. Liu, Y. Zhang, et al., “A Real-world Multi-center RNA-seq Benchmarking Study Using the Quartet and MAQC Reference Materials,” *Nature Communications* 15, no. 1 (2024): 6167.
162. G. X. Y. Zheng, J. M. Terry, P. Belgrader, et al., “Massively Parallel Digital Transcriptional Profiling of Single Cells,” *Nature Communications* 8 (2017): 14049.
163. I. Kalvari, E. P. Nawrocki, J. Argasinska, et al., “Non-Coding RNA Analysis Using the Rfam Database,” *CP in Bioinformatics* 62, no. 1 (2018): e51.
164. D. Chakravarty, J. Gao, S. M. Phillips, et al., “OncoKB: A Precision Oncology Knowledge Base,” *JCO Precision Oncology* 2017 (2017): PO.17.00011.
165. N. Feizi, S. K. Nair, P. Smirnov, et al., “PharmacoDB 2.0: Improving Scalability and Transparency of in Vitro Pharmacogenomics Analysis,” *Nucleic Acids Research* 50, no. D1 (2022): D1348-D1357.
166. X. Wang, L. Chen, W. Liu, et al., “TIMEDB: Tumor Immune Micro-environment Cell Composition Database With Automatic Analysis and Interactive Visualization,” *Nucleic Acids Research* 51, no. D1 (2023): D1417-D1424.
167. J. Singer, A. Irmisch, H. J. Ruscheweyh, et al., “Bioinformatics for Precision Oncology,” *Briefings in Bioinformatics* 20, no. 3 (2019): 778-788.
168. D. Fernández-Orth, M. Rueda, B. Singh, et al., “A Quality Control Portal for Sequencing Data Deposited at the European Genome-phenome Archive,” *Brief Bioinform* 23, no. 3 (2022): bbac136.
169. G. de Sena Brandine, A. D. Smith, “Falco: High-speed FastQC Emulation for Quality Control of Sequencing Data,” *F1000Res* 8 (2021): 1874.
170. A. M. Bolger, M. Lohse, B. Usadel, “Trimmomatic: A Flexible Trimmer for Illumina Sequence Data,” *Bioinformatics* 30, no. 15 (2014): 2114-2120.
171. X. Zhang, “Highly Effective Batch Effect Correction Method for RNA-seq Count Data,” *Computational and Structural Biotechnology Journal* 27 (2024): 58-64.
172. J. T. Leek, W. E. Johnson, H. S. Parker, A. E. Jaffe, J. D. Storey, “The sva Package for Removing Batch Effects and Other Unwanted Variation in High-throughput Experiments,” *Bioinformatics* 28, no. 6 (2012): 882-883.
173. J. T. Leek, “svaseq: Removing Batch Effects and Other Unwanted Noise From Sequencing Data,” *Nucleic Acids Research* 42, no. 21 (2014): e161.
174. L. Liu, J. Gao, G. Beasley, S. H. Jung, “LASSO and Elastic Net Tend to Over-Select Features,” *Mathematics* 11, no. 17 (2023): 3738.
175. R. Tibshirani, “Regression Shrinkage and Selection via the Lasso,” *Journal of the Royal Statistical Society: Series B (Methodological)* 58, no. 1 (1996): 267-288.
176. H. Zou, T. Hastie, “Regularization and Variable Selection via the Elastic Net,” *Journal of the Royal Statistical Society Series B: Statistical Methodology* 67, no. 2 (2005): 301-320.
177. S. Ng, S. Masarone, D. Watson, M. R. Barnes, “The Benefits and Pitfalls of Machine Learning for Biomarker Discovery,” *Cell and Tissue Research* 394, no. 1 (2023): 17-31.
178. S. Uddin, A. Khan, M. E. Hossain, M. A. Moni, “Comparing Different Supervised Machine Learning Algorithms for Disease Prediction,” *BMC Medical Informatics and Decision Making* 19, no. 1 (2019): 281.
179. S. Huang, N. Cai, P. P. Pacheco, S. Narrandes, Y. Wang, W. Xu, “Applications of Support Vector Machine (SVM) Learning in Cancer Genomics,” *Cancer Genomics & Proteomics* 15, no. 1 (2018): 41-51.
180. J. Martorell-Marugán, S. Tabik, Y. Benhammou, et al. Deep Learning in Omics Data Analysis and Precision Medicine. In: Husi H, ed. *Computational* (Codon Publications, 2019), Accessed May 1, 2025. <http://www.ncbi.nlm.nih.gov/books/NBK550335/>.
181. J. Jovel, R. Greiner, “An Introduction to Machine Learning Approaches for Biomedical Research,” *Frontiers in Medicine* 8 (2021), <https://doi.org/10.3389/fmed.2021.771607>.
182. M. H. Ferrato, A. G. Marsh, K. R. Franke, et al., “Machine Learning Classifier Approaches for Predicting Response to RTK-type-III Inhibitors Demonstrate High Accuracy Using Transcriptomic Signatures and Ex Vivo Data,” *Bioinformatics Advances* 3, no. 1 (2023): vbad034.
183. F. Pedregosa, G. Varoquaux, A. Gramfort, et al. Scikit-learn: Machine Learning in Python. Published online June 5, 2018. doi:10.48550/arXiv.1201.0490
184. O. C. Novac, M. C. Chirodea, C. M. Novac, et al., “Analysis of the Application Efficiency of TensorFlow and PyTorch in Convolutional Neural Network,” *Sensors (Basel)* 22, no. 22 (2022): 8872.
185. L. Rampasek, A. Goldenberg, “TensorFlow: Biology’s Gateway to Deep Learning?,” *Cell Systems* 2, no. 1 (2016): 12-14.
186. L. J. Marcos-Zambrano, V. M. López-Molina, B. Bakir-Gungor, “A Toolbox of Machine Learning Software to Support Microbiome Analysis,” *Frontiers in Microbiology* 14 (2023): 1250806.
187. L. Chen, C. Wang, H. Sun, et al., “The Bioinformatics Toolbox for circRNA Discovery and Analysis,” *Brief Bioinform* 22, no. 2 (2020): 1706-1728.
188. T. D. Tran, M. T. Nguyen, “C-Biomarker.Net: A Cytoscape App for the Identification of Cancer Biomarker Genes From Cores of Large Biomolecular Networks,” *Bio Systems* 226 (2023): 104887.

189. G. Su, J. H. Morris, B. Demchak, G. D. Bader, "Biological Network Exploration With Cytoscape 3," *CP in Bioinformatics* 47 (2014): 8.13.1-24.
190. C. Jia, T. Wang, D. Cui, et al., "A Metagene Based Similarity Network Fusion Approach for Multi-omics Data Integration Identified Novel Subtypes in Renal Cell Carcinoma," *Brief Bioinform* 25, no. 6 (2024): bbae606.
191. M. Chierici, N. Bussola, A. Marcolini, et al., "Integrative Network Fusion: A Multi-Omics Approach in Molecular Profiling," *Frontiers in Oncology* 10 (2020): 1065.
192. Q. Mo, R. Shen, C. Guo, M. Vannucci, K. S. Chan, S. G. Hilsenbeck, "A Fully Bayesian Latent Variable Model for Integrative Clustering Analysis of Multi-type Omics Data," *Biostatistics (Oxford, England)* 19, no. 1 (2018): 71-86.
193. G. Zhang, Z. Peng, C. Yan, J. Wang, J. Luo, H. Luo, ". MultiGATAE: A Novel Cancer Subtype Identification Method Based on Multi-Omics and Attention Mechanism," *Frontiers in Genetics* 13 (2022): 855629.
194. E. M. Liu, A. Luna, G. Dong, C. Sander, "netboxr: Automated Discovery of Biological Process Modules by Network Analysis in R," *PLoS ONE* 15, no. 11 (2020): e0234669.
195. J. Reimand, R. Isser, V. Voisin, et al., "Pathway Enrichment Analysis and Visualization of Omics Data Using G:Profiler, GSEA, Cytoscape and EnrichmentMap," *Nature Protocols* 14, no. 2 (2019): 482-517.
196. G. Dennis, B. T. Sherman, D. A. Hosack, et al., "DAVID: Database for Annotation, Visualization, and Integrated Discovery," *Genome Biology* 4, no. 5 (2003): P3.
197. DAVID Functional Annotation Bioinformatics Microarray Analysis. Accessed May 1, 2025. <https://davidbioinformatics.nih.gov/>
198. A. Fabregat, K. Sidiropoulos, G. Viteri, et al., "Reactome Pathway Analysis: A High-performance in-memory Approach," *BMC Bioinformatics [Electronic Resource]* 18, no. 1 (2017): 142.
199. Home—Reactome Pathway Database. Accessed May 1, 2025. <https://reactome.org/>
200. A. Krämer, J. Green, J. Pollard, S. Tugendreich, "Causal Analysis Approaches in Ingenuity Pathway Analysis," *Bioinformatics* 30, no. 4 (2014): 523-530.
201. R. Oughtred, J. Rust, C. Chang, et al., "The BioGRID Database: A Comprehensive Biomedical Resource of Curated Protein, Genetic, and Chemical Interactions," *Protein Science* 30, no. 1 (2021): 187-200.
202. T. Wu, E. Hu, S. Xu, et al., "clusterProfiler 4.0: A Universal Enrichment Tool for Interpreting Omics Data," *Innovation (Camb)* 2, no. 3 (2021): 100141.
203. S. Patkar, A. Magen, R. Sharan, S. Hannenhalli, "A Network Diffusion Approach to Inferring Sample-specific Function Reveals Functional Changes Associated With Breast Cancer," *PLOS Computational Biology* 13, no. 11 (2017): e1005793.
204. J. Montojo, K. Zuberi, H. Rodriguez, et al., "GeneMANIA Cytoscape Plugin: Fast Gene Function Predictions on the Desktop," *Bioinformatics* 26, no. 22 (2010): 2927-2928.
205. Y. Zhou, B. Zhou, L. Pache, et al., "Metascape Provides a Biologist-oriented Resource for the Analysis of Systems-level Datasets," *Nature Communications* 10, no. 1 (2019): 1523.
206. K. D. Davis, N. Aghaeepour, A. H. Ahn, et al., "Discovery and Validation of Biomarkers to Aid the Development of Safe and Effective Pain Therapeutics: Challenges and Opportunities," *Nature Reviews Neurology* 16, no. 7 (2020): 381-400.
207. A. Sadlon, "In Silico Models to Validate Novel Blood-Based Biomarkers," *Methods in Molecular Biology* 2785 (2024): 321-344.
208. L. Venkataramana, S. G. Jacob, S. Saraswathi, D. Venkata Vara Prasad, "Identification of Common and Dissimilar Biomarkers for Different Cancer Types From Gene Expressions of RNA-sequencing Data," *Gene Reports* 19 (2020): 100654.
209. Z. Huang, T. Lan, J. Wang, Z. Chen, X. Zhang, "Identification and Validation of Seven RNA Binding Protein Genes as a Prognostic Signature in Oral Cavity Squamous Cell Carcinoma," *Bioengineered* 12, no. 1 (2021): 7248-7262.
210. R. Edgar, M. Domrachev, A. E. Lash, "Gene Expression Omnibus: NCBI Gene Expression and Hybridization Array Data Repository," *Nucleic Acids Research* 30, no. 1 (2002): 207-210.
211. J. Zhang, J. Baran, A. Cros, et al., "International Cancer Genome Consortium Data Portal—a One-stop Shop for Cancer Genomics Data," *Database (Oxford)* 2011 (2011): bar026.
212. S. Kim, C. W. Lin, Tseng GeorgeC, "MetaKTSP: A Meta-analytic Top Scoring Pair Method for Robust Cross-study Validation of Omics Prediction Analysis," *Bioinformatics* 32, no. 13 (2016): 1966-1973.
213. S. U. Khan, "Composite to Clarity: Shifting from Combined to Individual Endpoints in Meta-Analyses of Cardiovascular Outcome Trials," *JACC: Advances* 2, no. 7 (2023): 100548.
214. X. Cheng, Y. Liu, J. Wang, et al., "cSurvival: A Web Resource for Biomarker Interactions in Cancer Outcomes and in Cell Lines," *Brief Bioinform* 23, no. 3 (2022): bbac090.
215. R. Aguirre-Gamboa, H. Gomez-Rueda, E. Martínez-Ledesma, "SurvExpress: An Online Biomarker Validation Tool and Database for Cancer Gene Expression Data Using Survival Analysis," *PLoS ONE* 8, no. 9 (2013): e74250.
216. CRAN: Package survminer. Accessed May 1, 2025. <https://cran.r-project.org/web/packages/survminer/index.html>
217. K. Jóźwiak, V. H. Nguyen, L. Sollfrank, S. C. Linn, M. Hauptmann, "Cox Proportional Hazards Regression in Small Studies of Predictive Biomarkers," *Scientific Reports* 14, no. 1 (2024): 14232.
218. Z. Tang, B. Kang, C. Li, T. Chen, Z. Zhang, "GEPIA2: An Enhanced Web Server for Large-scale Expression Profiling and Interactive Analysis," *Nucleic Acids Research* 47, no. W1 (2019): W556-W560.
219. T. Li, J. Fan, B. Wang, et al., "TIMER: A Web Server for Comprehensive Analysis of Tumor-infiltrating Immune Cells," *Cancer Research* 77, no. 21 (2017): e108-e110.
220. J. Anaya, "OncoLnc: Linking TCGA Survival Data to mRNAs, miRNAs, and lncRNAs," *PeerJ Computer Science* 2 (2016): e67.
221. H. Mizuno, K. Kitada, K. Nakai, A. Sarai, "PrognScan: A New Database for Meta-analysis of the Prognostic Value of Genes," *BMC Medical Genomics* 2 (2009): 18.
222. Biomarker Discovery and Validation Using a Combination of In Vitro and In Vivo Studies. Accessed May 1, 2025. <https://blog.crownbio.com/biomarker-discovery-and-validation-combining-in-vitro-and-in-vivo-studies>
223. M. A. Hossain, M. Sohel, M. H. Rahman, et al., "Bioinformatics and in Silico Approaches to Identify Novel Biomarkers and Key Pathways for Cancers That Are Linked to the Progression of Female Infertility: A Comprehensive Approach for Drug Discovery," *PLoS ONE* 18, no. 1 (2023): e0265746.
224. M. Mann, C. Kumar, W. F. Zeng, M. T. Strauss, "Artificial Intelligence for Proteomics and Biomarker Discovery," *Cell Systems* 12, no. 8 (2021): 759-770.
225. F. S. Ou, S. Michiels, Y. Shyr, A. A. Adjei, A. L. Oberg, "Biomarker Discovery and Validation: Statistical Considerations," *Journal of Thoracic Oncology* 16, no. 4 (2021): 537-545.
226. O. Alter, S. P. Ponnappalli, J. W. Tsai, et al., "Prospective Validation From a Retrospective Trial That Validated an AI/ML-derived Whole-genome Biomarker as the Most Accurate and Precise Predictor of Survival and Response to Treatment in Glioblastoma," *JCO* 42, no. 16_suppl (2024): e14028-e14028.
227. M. M. Boyiadzis, J. M. Kirkwood, J. L. Marshall, C. C. Pritchard, N. S. Azad, J. L. Gulley, "Significance and Implications of FDA

- Approval of pembrolizumab for Biomarker-defined Disease,” *Journal for ImmunoTherapy of Cancer* 6, no. 1 (2018): 35.
228. D. T. Le, J. N. Uram, H. Wang, et al., “Programmed Death-1 Blockade in Mismatch Repair Deficient Colorectal Cancer,” *Journal of Clinical Oncology* 34, no. 15 (2016), https://doi.org/10.1200/JCO.2016.34.15_suppl.103. Published online May 20, 2016.
229. R. Rosell, E. Carcereny, R. Gervais, et al., “Erlotinib versus Standard Chemotherapy as First-line Treatment for European Patients With Advanced EGFR Mutation-positive Non-small-cell Lung Cancer (EUR-TAC): A Multicentre, Open-label, Randomised Phase 3 Trial,” *The Lancet Oncology* 13, no. 3 (2012): 239-246.
230. National Cancer Institute (NCI). *MARVEL: Marker Validation of Erlotinib in Lung Cancer- A Phase III Biomarker Validation Study of Second-Line Therapy in Patients with Advanced Non-Small Cell Lung Cancer (NSCLC) Randomized to Pemetrexed versus Erlotinib*. *Clinicaltrials.gov*; 2015. Accessed May 1, 2025. <https://clinicaltrials.gov/study/NCT00738881>
231. R. S. Herbst, M. W. Redman, E. S. Kim, et al., “Cetuximab plus Carboplatin and Paclitaxel With or Without Bevacizumab versus Carboplatin and Paclitaxel With or Without Bevacizumab in Advanced NSCLC (SWOG S0819): A Randomised, Phase 3 Study,” *The Lancet Oncology* 19, no. 1 (2018): 101-114.
232. J. Li, S. Long, Z. Yang, et al., “Single-cell Transcriptomics Reveals IRF7 Regulation of the Tumor Microenvironment in Isocitrate Dehydrogenase Wild-type Glioma,” *MedComm* 5, no. 11 (2024): e754.
233. V. Janitri, K. N. ArulJothi, V. M. Ravi Mythili, et al., “The Roles of Patient-derived Xenograft Models and Artificial Intelligence Toward Precision Medicine,” *MedComm* 5, no. 10 (2024): e745.
234. H. Xu, Z. Fan, S. Jiang, et al., “Integrating Multiplex Immunohistochemistry and Machine Learning for Glioma Subtyping and Prognosis Prediction,” *MedComm* 6, no. 5 (2025): e70138.
235. A. J. Bagchee-Clark, E. J. Mucaki, T. Whitehead, P. K. Rogan, “Pathway-extended Gene Expression Signatures Integrate Novel Biomarkers That Improve Predictions of Patient Responses to Kinase Inhibitors,” *MedComm* 1, no. 3 (2020): 311-327.
236. V. Bhardwaj, A. Sharma, S. V. Parambath, et al., “Machine Learning for Endometrial Cancer Prediction and Prognostication,” *Frontiers in Oncology* 12 (2022): 852746.
237. G. Tanzhu, L. Chen, J. Ning, et al., “Metastatic Brain Tumors: From Development to Cutting-edge Treatment,” *MedComm* 6, no. 1 (2024): e70020.
238. Y. Yu, G. Cai, R. Lin, et al., “Multimodal Data Fusion AI Model Uncovers Tumor Microenvironment Immunotyping Heterogeneity and Enhanced Risk Stratification of Breast Cancer,” *MedComm* 5, no. 12 (2024): e70023.
239. A. Al-Dherasi, Q. T. Huang, Y. Liao, et al., “A Seven-gene Prognostic Signature Predicts Overall Survival of Patients With Lung Adenocarcinoma (LUAD),” *Cancer Cell International* 21, no. 1 (2021): 294.
240. Y. Huang, P. Zeng, C. Zhong, “Classifying Breast Cancer Subtypes on Multi-omics Data via Sparse Canonical Correlation Analysis and Deep Learning,” *BMC Bioinformatics [Electronic Resource]* 25, no. 1 (2024): 132.
241. H. Chai, Y. Huang, L. Xu, X. Song, M. He, Q. Wang, “A Decentralized Federated Learning-based Cancer Survival Prediction Method With Privacy Protection,” *Heliyon* 10, no. 11 (2024): e31873.
242. F. Zhu, R. Zhong, F. Li, et al., “Development and Validation of a Deep Transfer Learning-based Multivariable Survival Model to Predict Overall Survival in Lung Cancer,” *Translational Lung Cancer Research* 12, no. 3 (2023): 471-482.
243. T. Khater, A. Hussain, S. Mahmoud, S. Yassen. Explainable AI for Breast Cancer Detection: A LIME-Driven Approach. *2023 16th International Conference on Developments in eSystems Engineering (DeSE)*. Published online December 18, 2023:540-545. <https://doi.org/10.1109/DeSE60595.2023.10469341>
244. M. Arnold, “Integrating Multi-omics Data for Target and Biomarker Discovery,” *Alzheimers Dement* 20, no. Suppl 1 (2025): e086331.
245. Z. Wang, X. Sui, W. Song, et al., “Reinforcement Learning for Individualized Lung Cancer Screening Schedules: A Nested Case-control Study,” *Cancer Medicine* 13, no. 13 (2024): e7436.
246. A. Mahmoud, E. Takaoka, “An Enhanced Machine Learning Approach With Stacking Ensemble Learner for Accurate Liver Cancer Diagnosis Using Feature Selection and Gene Expression Data,” *Healthcare Analytics* 7 (2025): 100373.
247. W. Gu, X. Yang, M. Yang, K. Han, W. Pan, Z. Zhu, “MarkerGenie: An NLP-enabled Text-mining System for Biomedical Entity Relation Extraction,” *Bioinformatics Advances* 2, no. 1 (2022): vbac035.
248. Z. Wang, Y. Zhou, T. Takagi, J. Song, Y. S. Tian, T. Shibuya, “Genetic Algorithm-based Feature Selection With Manifold Learning for Cancer Classification Using Microarray Data,” *BMC Bioinformatics [Electronic Resource]* 24, no. 1 (2023): 139.
249. Z. Momeni, E. Hassanzadeh, M. Saniee Abadeh, R. Bellazzi, “A Survey on Single and Multi Omics Data Mining Methods in Cancer Data Classification,” *Journal of Biomedical Informatics* 107 (2020): 103466.
250. Y. Wang, W. Hou, N. Sheng, et al., “Graph Pooling in Graph Neural Networks: Methods and Their Applications in Omics Studies,” *Artificial Intelligence Review* 57, no. 11 (2024): 294.
251. T. Kolisnik, F. Keshavarz-Rahaghi, R. V. Purcell, A. N. H. Smith, O. K. Silander, “pyRforest: A Comprehensive R Package for Genomic Data Analysis Featuring scikit-learn Random Forests in R,” *Briefings in Functional Genomics* 24 (2025): elae038.
252. D. Ruiz-Perez, J. Lugo-Martinez, N. Bourguignon, et al., “Dynamic Bayesian Networks for Integrating Multi-omics Time Series Microbiome Data,” *Msystems* 6, no. 2 (2021), <https://doi.org/10.1128/mSystems.01105-20>. e01105-20.
253. F. H. Yagin, R. El Shawi, A. Algarni, C. Colak, F. Al-Hashem, L. P. Ardigo, “Metabolomics Biomarker Discovery to Optimize Hepatocellular Carcinoma Diagnosis: Methodology Integrating AutoML and Explainable Artificial Intelligence,” *Diagnostics (Basel)* 14, no. 18 (2024): 2049.
254. R. Cuevas-Diaz-Duran, J. C. González-Orozco, I. Velasco, J. Q. Wu, “Single-cell and Single-nuclei RNA Sequencing as Powerful Tools to Decipher Cellular Heterogeneity and Dysregulation in Neurodegenerative Diseases,” *Frontiers in Cell and Developmental Biology* 10 (2022): 884748.
255. N. Erfanian, A. A. Heydari, A. M. Feriz, et al., “Deep Learning Applications in Single-cell Genomics and Transcriptomics Data Analysis,” *Biomedicine & Pharmacotherapy* 165 (2023): 115077.
256. D. M. Koh, N. Papanikolaou, U. Bick, et al., “Artificial Intelligence and Machine Learning in Cancer Imaging,” *Communications Medicine (London)* 2 (2022): 133.
257. B. Yuan, D. Yang, B. E. G. Rothberg, H. Chang, T. Xu, “Unsupervised and Supervised Learning With Neural Network for human Transcriptome Analysis and Cancer Diagnosis,” *Scientific Reports* 10, no. 1 (2020): 19106.
258. D. Bertsimas, H. Wiberg, “Machine Learning in Oncology: Methods, Applications, and Challenges,” *JCO Clinical Cancer Informatics* 4 (2020), <https://doi.org/10.1200/CCI.20.00072>. CCI.20.00072.
259. S. Steyaert, M. Pizurica, D. Nagaraj, et al., “Multimodal Data Fusion for Cancer Biomarker Discovery With Deep Learning,” *Nature Machine Intelligence* 5, no. 4 (2023): 351-362.
260. C. Greeley, L. Holder, E. E. Nilsson, M. K. Skinner, “Scalable Deep Learning Artificial Intelligence Histopathology Slide Analysis and Validation,” *Scientific Reports* 14, no. 1 (2024): 26748.
261. C. Wemmert, J. Weber, F. Feuerhake, G. Forestier. Deep Learning for Histopathological Image Analysis. In: Elloumi M, ed. “*Deep Learning for Biomedical Data Analysis: Techniques, Approaches, and Applications*” (Springer International Publishing, 2021): 153-169, https://doi.org/10.1007/978-3-030-71676-9_7.

262. E. U. Alum, "AI-driven Biomarker Discovery: Enhancing Precision in Cancer Diagnosis and Prognosis," *Discover Oncology* 16, no. 1 (2025): 313.
263. M. Kumaran, U. Subramanian, B. Devarajan, "Performance Assessment of Variant Calling Pipelines Using human Whole Exome Sequencing and Simulated Data," *BMC Bioinformatics [Electronic Resource]* 20, no. 1 (2019): 342.
264. P. Kamya, I. V. Ozerov, F. W. Pun, et al., "PandaOmics: An AI-Driven Platform for Therapeutic Target and Biomarker Discovery," *Journal of Chemical Information and Modeling* 64, no. 10 (2024): 3961-3969. Published online February 26, 2024.
265. T. M. Nguyen, N. Kim, D. H. Kim, et al., "Deep Learning for Human Disease Detection, Subtype Classification, and Treatment Response Prediction Using Epigenomic Data," *Biomedicines* 9, no. 11 (2021): 1733.
266. W. Lotter, M. J. Hassett, N. Schultz, K. L. Kehl, E. M. Van Allen, E. Cerami, "Artificial Intelligence in Oncology: Current Landscape, Challenges, and Future Directions," *Cancer Discovery* 14, no. 5 (2024): 711-726.
267. W. L. Bi, A. Hosny, M. B. Schabath, et al., "Artificial Intelligence in Cancer Imaging: Clinical Challenges and Applications," *CA: A Cancer Journal for Clinicians* 69, no. 2 (2019): 127-157.
268. M. Ennab, H. Mcheick, "Enhancing Interpretability and Accuracy of AI Models in Healthcare: A Comprehensive Review on Challenges and Future Directions," *Front Robot AI* 11 (2024): 1444763.
269. C. Mennella, U. Maniscalco, G. De Pietro, M. Esposito, "Ethical and Regulatory Challenges of AI Technologies in Healthcare: A Narrative Review," *Heliyon* 10, no. 4 (2024): e26297.
270. S. Yu, S. S. Lee, H. Hwang, "The Ethics of Using Artificial Intelligence in Medical Research," *KMJ* 39, no. 4 (2024): 229-237.
271. C. J. Creighton, "Making Use of Cancer Genomic Databases," *Current Protocols in Molecular Biology* 121 (2018): 19.14.1-19.14.13.
272. K. Tomczak, P. Czerwińska, M. Wiznerowicz, "The Cancer Genome Atlas (TCGA): An Immeasurable Source of Knowledge," *Contemporary Oncology (Pozn)* 19, no. 1A (2015): A68-A77.
273. International Network of Cancer Genome Projects. *Nature* 2010;464(7291):993-998.
274. A. A. Gazola, W. Lautert-Dutra, L. F. Archangelo, R. B. dos Reis, J. A. Squire, "Precision Oncology Platforms: Practical Strategies for Genomic Database Utilization in Cancer Treatment," *Molecular Cytogenetics* 17, no. 1 (2024): 28.
275. J. G. Tate, S. Bamford, H. C. Jubb, et al., "COSMIC: The Catalogue of Somatic Mutations in Cancer," *Nucleic Acids Research* 47, no. Database issue (2019): D941-D947.
276. M. Tarailo-Graovac, J. Y. A. Zhu, A. Matthews, C. D. M. van Karnebeek, W. W. Wasserman, "Assessment of the ExAC Data Set for the Presence of Individuals With Pathogenic Genotypes Implicated in Severe Mendelian Pediatric Disorders," *Genetics in Medicine* 19, no. 12 (2017): 1300-1308.
277. W. Zhang, X. Xie, Z. Huang, et al., "The Integration of Single-cell Sequencing, TCGA, and GEO Data Analysis Revealed That PRRT3-AS1 Is a Biomarker and Therapeutic Target of SKCM," *Frontiers in immunology* 13 (2022): 919145.
278. R. Lowe, N. Shirley, M. Bleackley, S. Dolan, T. Shafee, "Transcriptomics Technologies," *Plos Computational Biology* 13, no. 5 (2017): e1005457.
279. Y. Perez-Riverol, J. Bai, C. Bandla, et al., "The PRIDE Database Resources in 2022: A Hub for Mass Spectrometry-based Proteomics Evidences," *Nucleic Acids Research* 50, no. D1 (2022): D543-D552.
280. S. Killcoyne, E. W. Deutsch, J. Boyle, "Mining PeptideAtlas for Biomarkers and Therapeutics in human Disease," *Current Pharmaceutical Design* 18, no. 6 (2012): 748-754. Accessed May 2, 2025. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4361040/>.
281. M. Vaudel, K. Verheggen, A. Csordas, et al., "Exploring the Potential of Public Proteomics Data," *Proteomics* 16, no. 2 (2016): 214-225.
282. O. Yurekten, T. Payne, N. Tejera, et al., "MetaboLights: Open Data Repository for Metabolomics," *Nucleic Acids Research* 52, no. D1 (2024): D640-D646.
283. M. J. Goldman, B. Craft, M. Hastie, et al., "Visualizing and Interpreting Cancer Genomics Data via the Xena Platform," *Nature Biotechnology* 38, no. 6 (2020): 675-678.
284. M. V. Dieci, F. Miglietta, G. Griguolo, V. Guarneri, "Biomarkers for HER2-positive Metastatic Breast Cancer: Beyond Hormone Receptors," *Cancer Treatment Reviews* 88 (2020): 102064.
285. S. Garg, A. Sachdeva, M. Peeters, J. McClements, "Point-of-Care Prostate Specific Antigen Testing: Examining Translational Progress Toward Clinical Implementation," *ACS Sens* 8, no. 10 (2023): 3643-3658.
286. P. Charkhchi, C. Cybulski, J. Gronwald, F. O. Wong, S. A. Narod, M. R. Akbari, "CA125 and Ovarian Cancer: A Comprehensive Review," *Cancers (Basel)* 12, no. 12 (2020): 3730.
287. P. T. Harrison, S. Vyse, P. H. Huang, "Rare Epidermal Growth Factor Receptor (EGFR) Mutations in Non-small Cell Lung Cancer," *Seminars in Cancer Biology* 61 (2020): 167-179.
288. L. Tembuysen, M. J. L. Ligtenberg, N. Normanno, S. Delen, J. H. van Krieken, E. M. C. Dequeker, "Higher Quality of Molecular Testing, an Unfulfilled Priority: Results From External Quality Assessment for KRAS Mutation Testing in Colorectal Cancer," *The Journal of Molecular Diagnostics* 16, no. 3 (2014): 371-377.
289. G. Castellani, M. Buccarelli, M. B. Arasi, et al., "BRAF Mutations in Melanoma: Biological Aspects, Therapeutic Implications, and Circulating Biomarkers," *Cancers (Basel)* 15, no. 16 (2023): 4026.
290. H. Hanif, M. J. Ali, A. T. Susheela, et al., "Update on the Applications and Limitations of Alpha-fetoprotein for Hepatocellular Carcinoma," *World Journal of Gastroenterology* 28, no. 2 (2022): 216-229.
291. R. Tenchov, A. K. Sapra, J. Sasso, et al., "Biomarkers for Early Cancer Detection: A Landscape View of Recent Advancements, Spotlighting Pancreatic and Liver Cancers," *ACS Pharmacology & Translational Science* 7, no. 3 (2024): 586-613.
292. A. Ooki, H. Osumi, K. Yoshino, K. Yamaguchi, "Potent Therapeutic Strategy in Gastric Cancer With Microsatellite Instability-high and/or Deficient Mismatch Repair," *Gastric Cancer* 27, no. 5 (2024): 907-931.
293. S. H. Yu, S. S. Kim, S. Kim, H. Lee, T. W. Kang, "FGFR3 Mutations in Urothelial Carcinoma: A Single-Center Study Using Next-Generation Sequencing," *Journal of Clinical Medicine* 13, no. 5 (2024): 1305.
294. E. De Braekeleer, N. Douet-Guilbert, M. De Braekeleer, "RARA Fusion Genes in Acute Promyelocytic Leukemia: A Review," *Expert Review of Hematology* 7, no. 3 (2014): 347-357.
295. S. Brandner, A. McAleenan, C. Kelly, et al., "MGMT Promoter Methylation Testing to Predict Overall Survival in People With Glioblastoma Treated With Temozolomide: A Comprehensive Meta-analysis Based on a Cochrane Systematic Review," *Neuro-Oncology* 23, no. 9 (2021): 1457-1469.
296. M. Bartosik, L. Moranova, N. Izadi, et al., "Advanced Technologies towards Improved HPV Diagnostics," *Journal of Medical Virology* 96, no. 2 (2024): e29409.
297. G. Rosas, R. Ruiz, J. M. Araujo, J. A. Pinto, L. Mas, "ALK Rearrangements: Biology, Detection and Opportunities of Therapy in Non-small Cell Lung Cancer," *Critical Reviews in Oncology/Hematology* 136 (2019): 48-55.
298. D. De Novellis, R. Fontana, A. Carobene, et al., "Serum Free Light-Chain Ratio at Diagnosis Is Associated With Early Renal Damage in Multiple Myeloma: A Case Series Real-World Study," *Biomedicines* 10, no. 7 (2022): 1657.
299. B. Arun, F. J. Couch, J. Abraham, N. Tung, P. A. Fasching, "BRCA-mutated Breast Cancer: The Unmet Need, Challenges and Therapeutic Benefits of Genetic Testing," *British Journal of Cancer* 131, no. 9 (2024): 1400-1414.

300. D. B. Doroshow, S. Bhalla, M. B. Beasley, et al., "PD-L1 as a Biomarker of Response to Immune-checkpoint Inhibitors," *Nature reviews Clinical oncology* 18, no. 6 (2021): 345-362.
301. M. Huang, S. Yang, W. C. S. Tai, et al., "Bioinformatics Identification of Regulatory Genes and Mechanism Related to Hypoxia-Induced PD-L1 Inhibitor Resistance in Hepatocellular Carcinoma," *International Journal of Molecular Sciences* 24, no. 10 (2023): 8720.
302. D. Hammerl, J. W. M. Martens, M. Timmermans, et al., "Spatial Immunophenotypes Predict Response to Anti-PD1 Treatment and Capture Distinct Paths of T Cell Evasion in Triple Negative Breast Cancer," *Nature Communications* 12, no. 1 (2021): 5668.
303. N. M. Tung, J. E. Garber, "BRCA1/2 testing: Therapeutic Implications for Breast Cancer Management," *British Journal of Cancer* 119, no. 2 (2018): 141-152.
304. R. L. Siegel, A. N. Giaquinto, A. Jemal, "Cancer Statistics, 2024," *CA: A Cancer Journal for Clinicians* 74, no. 1 (2024): 12-49.
305. A. N. Giaquinto, H. Sung, L. A. Newman, et al., "Breast Cancer Statistics 2024," *CA: A Cancer Journal for Clinicians* 74, no. 6 (2024): 477-495.
306. V. Bhardwaj, X. Zhang, V. Pandey, M. Garg, "Neo-vascularization-based Therapeutic Perspectives in Advanced Ovarian Cancer," *Biochimica et Biophysica Acta: Reviews on Cancer* 1878, no. 3 (2023): 188888.
307. Á. Bartha, B. Györfy, "Comprehensive Outline of Whole Exome Sequencing Data Analysis Tools Available in Clinical Oncology," *Cancers (Basel)* 11, no. 11 (2019): 1725.
308. W. McLaren, L. Gil, S. E. Hunt, et al., "The Ensembl Variant Effect Predictor," *Genome Biology* 17, no. 1 (2016): 122.
309. L. Incorvaia, C. Marchetti, C. Brando, et al., "BRCA Functional Domains Associated With High Risk of Multiple Primary Tumors and Domain-related Sensitivity to olaparib: The Prometheus Study," *ESMO Open* 10, no. 2 (2025): 104076.
310. S. Obayashi, M. Aoki, K. Tanabe, et al., "The Role of BRCA1/2 Genetic Testing in Perioperative Breast Cancer Management: Advancing Shared Decision-making and Personalized Care," *International Journal of Clinical Oncology* (2025). Published online April 29, <https://doi.org/10.1007/s10147-025-02773-7>.
311. B. Zou, V. H. F. Lee, H. Yan, "Prediction of Sensitivity to Gefitinib/Erlotinib for EGFR Mutations in NSCLC Based on Structural Interaction Fingerprints and Multilinear Principal Component Analysis," *BMC Bioinformatics [Electronic Resource]* 19, no. 1 (2018): 88.
312. J. Martín-Arana, F. Gimeno-Valiente, T. V. Henriksen, et al., "Whole-exome Tumor-agnostic ctDNA Analysis Enhances Minimal Residual Disease Detection and Reveals Relapse Mechanisms in Localized Colon Cancer," *Nat Cancer* (2025): 1-17, Published online April 29, <https://doi.org/10.1038/s43018-025-00960-z>.
313. S. Halder, S. Basu, S. Lal, A. K. Ganti, S. K. Batra, P. Seshacharyulu, "Targeting the EGFR Signaling Pathway in Cancer Therapy: What's New in 2023?," *Expert Opinion on Therapeutic Targets* 27, no. 4-5 (2023): 305-324.
314. G. Hamilton, B. Rath, A. Plangger, M. Hochmair, "Implementation of Functional Precision Medicine for Anaplastic Lymphoma Kinase-rearranged Non-small Lung Cancer," *Precision Cancer Medicine* 2, no. 0 (2019), doi:10.21037/pcm.2019.05.03.
315. I. B. Muller, Langen AJ de, E. Giovannetti, G. J. Peters, "Anaplastic Lymphoma Kinase Inhibition in Metastatic Non-small Cell Lung Cancer: Clinical Impact of alectinib," *Ottawa* 10 (2017): 4535-4541.
316. G. Zakharova, M. Suntsova, E. Rabushko, et al., "A New Approach of Detecting ALK Fusion Oncogenes by RNA Sequencing Exon Coverage Analysis," *Cancers* 16, no. 22 (2024): 3851.
317. G. Chazan, F. Franchini, R. Shah, et al., "Real-World Treatment and Outcomes in ALK-Rearranged NSCLC: Results from a Large U.S.-Based Database," *JTO Clinical and Research Reports* 5, no. 8 (2024): 100662.
318. X. N. Wang, S. J. Wang, V. Pandey, et al., "Trefoil Factor 3 as a Novel Biomarker to Distinguish between Adenocarcinoma and Squamous Cell Carcinoma," *Medicine* 94, no. 20 (2015): e860.
319. V. Pandey, Z. S. Wu, M. Zhang, et al., "Trefoil Factor 3 Promotes Metastatic Seeding and Predicts Poor Survival Outcome of Patients With Mammary Carcinoma," *Breast Cancer Research* 16 (2014): 429.
320. F. Cheng, X. Wang, Y. S. Chiou, et al., "Trefoil Factor 3 Promotes Pancreatic Carcinoma Progression via WNT Pathway Activation Mediated by Enhanced WNT Ligand Expression," *Cell Death & Disease* 13, no. 3 (2022): 265.
321. V. Pandey, M. Zhang, Q. Y. Chong, et al., "Hypomethylation Associated Enhanced Transcription of Trefoil Factor-3 Mediates Tamoxifen-stimulated Oncogenicity of ER+ Endometrial Carcinoma Cells," *Oncotarget* 8, no. 44 (2017): 77268-77291.
322. M. L. You, Y. J. Chen, Q. Y. Chong, et al., "Trefoil Factor 3 Mediation of Oncogenicity and Chemoresistance in Hepatocellular Carcinoma Is AKT-BCL-2 Dependent," *Oncotarget* 8, no. 24 (2017): 39323-39344.
323. N. L. C. Bui, V. Pandey, T. Zhu, L. Ma, P. E. Basappa Lobie, "Bad Phosphorylation as a Target of Inhibition in Oncology," *Cancer Letters* 415 (2018): 177-186.
324. V. Pandey, B. Wang, C. D. Mohan, et al., "Discovery of a Small-molecule Inhibitor of Specific Serine Residue BAD Phosphorylation," *PNAS* 115, no. 44 (2018): E10505-E10514.
325. R. M. Chen, Y. S. Chiou, Q. Y. Chong, et al., "Pharmacological Inhibition of TFF3 Enhances Sensitivity of CMS4 Colorectal Carcinoma to 5-Fluorouracil Through Inhibition of p44/42 MAPK," *International Journal of Molecular Sciences* 20, no. 24 (2019): 6215.
326. M. Zhang, B. Wang, Q. Y. Chong, et al., "A Novel Small-molecule Inhibitor of Trefoil Factor 3 (TFF3) Potentiates MEK1/2 Inhibition in Lung Adenocarcinoma," *Oncogenesis* 8, no. 11 (2019): 65.
327. Y. Wang, Y. S. Chiou, Q. Y. Chong, et al., "Pharmacological Inhibition of BAD Ser99 Phosphorylation Enhances the Efficacy of Cisplatin in Ovarian Cancer by Inhibition of Cancer Stem Cell-Like Behavior," *ACS Pharmacology & Translational Science* 3, no. 6 (2020): 1083-1099.
328. V. Pandey, X. Zhang, H. M. Poh, et al., "Monomerization of Homodimeric Trefoil Factor 3 (TFF3) by an Aminonitrile Compound Inhibits TFF3-Dependent Cancer Cell Survival," *ACS Pharmacology & Translational Science* 5, no. 9 (2022): 761-773.
329. X. Zhang, L. Wang, S. Chen, et al., "Combined Inhibition of BADSer99 Phosphorylation and PARP Ablates Models of Recurrent Ovarian Carcinoma," *Communications Medicine (London)* 2 (2022): 82.
330. H. Guo, Y. Q. Tan, X. Huang, et al., "Small Molecule Inhibition of TFF3 Overcomes Tamoxifen Resistance and Enhances Taxane Efficacy in ER+ Mammary Carcinoma," *Cancer Letters* 579 (2023): 216443.
331. X. Zhang, P. Huang, L. Wang, et al., "Inhibition of BAD-Ser99 Phosphorylation Synergizes With PARP Inhibition to Ablate PTEN-deficient Endometrial Carcinoma," *Cell Death & Disease* 13, no. 6 (2022): 558.
332. Y. Q. Tan, Y. S. Chiou, H. Guo, et al., "Vertical Pathway Inhibition of Receptor Tyrosine Kinases and BAD With Synergistic Efficacy in Triple Negative Breast Cancer," *Npj Precision Oncology* 8 (2024): 8.
333. L. Wang, X. Zhang, S. Chen, et al., "Combining Mitomycin C With Inhibition of BAD Phosphorylation Enhances Apoptotic Cell Death in Advanced Cervical Cancer," *Translational Oncology* 49 (2024): 102103.
334. Y. Q. Tan, B. Sun, X. Zhang, et al., "Concurrent Inhibition of pBADs99 Synergistically Improves MEK Inhibitor Efficacy in KRASG12D-mutant Pancreatic Ductal Adenocarcinoma," *Cell Death & Disease* 15, no. 2 (2024): 173.
335. C. He, X. Wang, Y. S. Chiou, et al., "Inhibition of TFF3 Synergizes With c-MET Inhibitors to Decrease the CSC-Like Phenotype and Metastatic Burden in ER+HER2+ Mammary Carcinoma," *Cell Death & Disease* 16, no. 1 (2025): 76.

336. S. Chen, X. Zhang, B. Basappa, T. Zhu, V. Pandey, P. E. Lobie, "TFF3 facilitates Dormancy of Anti-estrogen Treated ER+ Mammary Carcinoma," *Communications Medicine (London)* 5 (2025): 45.
337. S. Zhang, Y. Q. Tan, X. Zhang, et al., "TFF3 drives Hippo Dependent EGFR-TKI Resistance in Lung Adenocarcinoma," *Oncogene* 44, no. 11 (2025): 753-768.
338. P. Huang, T. Wolde, V. Bhardwaj, X. Zhang, V. Pandey, "TFF3 and PVRL2 co-targeting Identified by multi-omics Approach as an Effective Cancer Immunosuppression Strategy," *Life Sciences* 357 (2024): 123113.
339. M. Arbitrio, M. Milano, M. Lucibello, et al., "Bioinformatic Challenges for Pharmacogenomic Study: Tools for Genomic Data Analysis," *Frontiers in Pharmacology* 16 (2025): 1548991.
340. Y. Luo, C. Zhao, F. Chen, "Multiomics Research: Principles and Challenges in Integrated Analysis," *BioDesign Research* 6 (2024): 0059.
341. X. Yang, K. Huang, D. Yang, W. Zhao, X. Zhou, "Biomedical Big Data Technologies, Applications, and Challenges for Precision Medicine: A Review," *Global Challenges* 8, no. 1 (2024): 2300163.
342. E. Pérez-Wohlfeil, O. Torreno, L. J. Bellis, P. L. Fernandes, B. Leskosek, O. Trelles, "Training Bioinformaticians in High Performance Computing," *Heliyon* 4, no. 12 (2018): e01057.
343. J. Rahnenführer, R. De Bin, A. Benner, et al., "Statistical Analysis of High-dimensional Biomedical Data: A Gentle Introduction to Analytical Goals, Common Approaches and Challenges," *BMC Medicine* 21, no. 1 (2023): 182.
344. T. C. Dakal, R. Dhakar, A. Beura, et al., "Emerging Methods and Techniques for Cancer Biomarker Discovery," *Pathology—Research and Practice* 262 (2024): 155567.
345. G. Gómez-López, J. Dopazo, J. C. Cigudosa, A. Valencia, F. Al-Shahrour, "Precision Medicine Needs Pioneering Clinical Bioinformaticians," *Brief Bioinform* 20, no. 3 (2019): 752-766.
346. M. Khalifa, M. Albadawy, "Artificial Intelligence for Clinical Prediction: Exploring Key Domains and Essential Functions," *Computer Methods and Programs in Biomedicine Update* 5 (2024): 100148.
347. M. Martínez-García, E. Hernández-Lemus, "Data Integration Challenges for Machine Learning in Precision Medicine," *Frontiers in Medicine (Lausanne)* 8 (2022): 784455.
348. G. D. Giebel, P. Raszke, H. Nowak, et al., "Problems and Barriers Related to the Use of AI-Based Clinical Decision Support Systems: Interview Study," *Journal of Medical Internet Research* 27, no. 1 (2025): e63377.
349. F. Mirakhori, S. K. Niazi, "Harnessing the AI/ML in Drug and Biological Products Discovery and Development: The Regulatory Perspective," *Pharmaceuticals (Basel)* 18, no. 1 (2025): 47.
350. D. D. Farhud, S. Zokaei, "Ethical Issues of Artificial Intelligence in Medicine and Healthcare," *Iranian Journal of Public Health* 50, no. 11 (2021): i-v.
351. M. Dara, N. Azarpira, "Ethical Considerations Emerge From Artificial Intelligence (AI) in Biotechnology," *Iranian Journal of Public Health* 17, no. 1 (2025): 80-81.
352. S. Wang, X. Jiang, S. Singh, et al., "Genome Privacy: Challenges, Technical Approaches to Mitigate Risk, and Ethical Considerations in the United States," *Annals of the New York Academy of Sciences* 1387, no. 1 (2017): 73-83.
353. R. Chevrier, V. Foufi, C. Gaudet-Blavignac, A. Robert, C. Lovis, "Use and Understanding of Anonymization and De-Identification in the Biomedical Literature: Scoping Review," *Journal of Medical Internet Research [Electronic Resource]* 21, no. 5 (2019): e13484.
354. F. J. Jaime, A. Muñoz, F. Rodríguez-Gómez, A. Jerez-Calero, "Strengthening Privacy and Data Security in Biomedical Microelectromechanical Systems by IoT Communication Security and Protection in Smart Healthcare," *Sensors (Basel)* 23, no. 21 (2023): 8944.
355. P. F. Edemekong, P. Annamaraju, M. Afzal, M. J. Haydel, "Health Insurance Portability and Accountability Act (HIPAA) Compliance," *StatPearls* (StatPearls Publishing, 2025). Accessed May 3, 2025. <http://www.ncbi.nlm.nih.gov/books/NBK500019/>
356. L. M. Beskow, E. Dean, "Informed Consent for Biorepositories: Assessing Prospective Participants' understanding and Opinions," *Cancer Epidemiology and Prevention Biomarkers* 17, no. 6 (2008): 1440-1451.
357. C. Grady, L. Eckstein, B. Berkman, et al., "Broad Consent for Research with Biological Samples: Workshop Conclusions," *The American Journal of Bioethics* 15, no. 9 (2015): 34-42.
358. D. T. Stein, S. F. Terry, "Reforming Biobank Consent Policy: A Necessary Move away From Broad Consent Toward Dynamic Consent," *Genetic Testing and Molecular Biomarkers* 17, no. 12 (2013): 855-856.
359. D. Horgan, M. Tanner, C. Aggarwal, et al., "Precision Oncology: A Global Perspective on Implementation and Policy Development," *JCO Global Oncology* 11 (2025): e2400416. Published online January 2025.
360. N. Khalid, A. Qayyum, M. Bilal, A. Al-Fuqaha, J. Qadir, "Privacy-preserving Artificial Intelligence in Healthcare: Techniques and Applications," *Computers in Biology and Medicine* 158 (2023): 106848.
361. K. M. Wong, K. Langlais, G. S. Tobias, et al., "The dbGaP Data Browser: A New Tool for Browsing dbGaP Controlled-access Genomic Data," *Nucleic Acids Research* 45, no. D1 (2017): D819-D826.
362. M. B. Davis, R. Martini, "Precision Oncology and Genetic Ancestry: The Science Behind Population-based Cancer Disparities," *Cancer Cell* 43, no. 4 (2025): 619-622.
363. M. Harishbhai Tilala, P. Kumar Chenchala, A. Choppadandi, et al., "Ethical Considerations in the Use of Artificial Intelligence and Machine Learning in Health Care: A Comprehensive Review," *Cureus* 16, no. 6: e62443.
364. J. W. Anderson, S. Visweswaran, "Algorithmic Individual Fairness and Healthcare: A Scoping Review," *JAMIA Open* 8, no. 1 (2025): ooae149.
365. J. Yates, E. M. Van Allen, "New Horizons at the Interface of Artificial Intelligence and Translational Cancer Research," *Cancer Cell* 43, no. 4 (2025): 708-727.
366. P. Moreno, N. Huang, J. R. Manning, et al., "User-friendly, Scalable Tools and Workflows for Single-cell RNAseq Analysis," *Nature Methods* 18, no. 4 (2021): 327-328.
367. K. B. Johnson, W. Q. Wei, D. Weeraratne, et al., "Precision Medicine, AI, and the Future of Personalized Health Care," *Clinical and Translational Science* 14, no. 1 (2021): 86-93.
368. K. S. Kumar, V. Miskovic, A. Blasiak, et al., "Artificial Intelligence in Clinical Oncology: From Data to Digital Pathology and Treatment," *American Society of Clinical Oncology Educational Book* 43 (2023): e390084. Published online May 2023.
369. A. Thorogood, H. L. Rehm, P. Goodhand, et al., "International federation of Genomic Medicine Databases Using GA4GH Standards," *Cell Genomics* 1, no. 2 (2021): 100032.
370. H. L. Rehm, A. J. H. Page, L. Smith, et al., "GA4GH: International Policies and Standards for Data Sharing Across Genomic Research and Healthcare," *Cell Genomics* 1, no. 2 (2021): 100029.
371. Z. Zhang, H. Li, S. Jiang, et al., "A Survey and Evaluation of Web-based Tools/Databases for Variant Analysis of TCGA Data," *Brief Bioinform* 20, no. 4 (2019): 1524-1541.
372. T. Züllig, M. Trötzmüller, H. C. Köfeler, "Lipidomics From Sample Preparation to Data Analysis: A Primer," *Analytical and Bioanalytical Chemistry* 412, no. 10 (2020): 2191-2209.
373. J. A. Tzec-Interián, D. González-Padilla, E. B. Góngora-Castillo, "Bioinformatics Perspectives on Transcriptomics: A Comprehensive Review of Bulk and Single-cell RNA Sequencing Analyses," *Quantitative Biology* 13 (2025): e78.

374. N. Shaban, D. Kamashev, A. Emelianova, A. Buzdin, "Targeted Inhibitors of EGFR: Structure, Biology, Biomarkers, and Clinical Applications," *Cells* 13, no. 1 (2023): 47.
375. S. A. Javed, A. Najmi, W. Ahsan, K. Zoghebi, "Targeting PD-1/PD-L1 Immune Checkpoint Inhibition for Cancer Immunotherapy: Success and Challenges," *Frontiers in Immunology* 15 (2024): 1383456.
376. J. Zhao, Z. Zhou, P. E. Saw, E. Song, "Silver Jubilee of HER2 Targeting: A Clinical Success in Breast Cancer," *Journal of the National Cancer Center* (2025). Published online February 12, <https://doi.org/10.1016/j.jncc.2024.12.008>.
377. J. Chen, A. Lin, A. Jiang, et al., "Computational Frameworks Transform Antagonism to Synergy in Optimizing Combination Therapies," *npj Digital Medicine* 8 (2025): 44.
378. S. M. Gadgeel, D. Rodríguez-Abreu, B. Halmos, et al., "Pembrolizumab plus Chemotherapy for Metastatic NSCLC with Programmed Cell Death Ligand 1 Tumor Proportion Score Less than 1%: Pooled Analysis of Outcomes after Five Years of Follow-Up," *Journal of Thoracic Oncology* 19, no. 8 (2024): 1228-1241.
379. J. J. Maly, E. R. Macrae, "Pertuzumab in Combination With Trastuzumab and Chemotherapy in the Treatment of HER2-Positive Metastatic Breast Cancer: Safety, Efficacy, and Progression Free Survival," *Breast Cancer (Auckl)* 8 (2014): 81-88.
380. L. Xiong, Y. Lou, L. Wang, "Effect of bevacizumab Combined With First-line Chemotherapy on Metastatic Colorectal Cancer," *American Journal of Translational Research* 13, no. 4 (2021): 3609-3617. Accessed May 3, 2025. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8129318/>
381. M. S. Hassan Zafar, A. A. Khan, S. Aggarwal, M. Bhargava, "Efficacy and Tolerability of Bortezomib and Dexamethasone in Newly Diagnosed Multiple Myeloma," *South Asian Journal of Cancer* 7, no. 1 (2018): 58-60.
382. H. Jamialahmadi, G. Khalili-Tanha, E. Nazari, M. Rezaei-Tavirani, "Artificial Intelligence and Bioinformatics: A Journey From Traditional Techniques to Smart Approaches," *Gastroenterol Hepatol Bed Bench* 17, no. 3 (2024): 241-252.
383. W. Jiang, W. Ye, X. Tan, Y. J. Bao, "Network-based Multi-omics Integrative Analysis Methods in Drug Discovery: A Systematic Review," *BioData Mining* 18 (2025): 27.
384. S. Kulkarni, J. Brownlie, J. N. Jeyapalan, N. P. Mongan, E. A. Rakha, S. Madhusudan, "Evolving DNA Repair Synthetic Lethality Targets in Cancer," *Bioscience Reports* 42, no. 12 (2022): BSR20221713.
385. A. Guan, C. Quek, "Single-Cell Multi-Omics: Insights Into Therapeutic Innovations to Advance Treatment in Cancer," *International Journal of Molecular Sciences* 26, no. 6 (2025): 2447.
386. S. Fang, B. Chen, Y. Zhang, et al., "Computational Approaches and Challenges in Spatial Transcriptomics," *Genomics, Proteomics & Bioinformatics* 21, no. 1 (2023): 24-47.
387. G. Molla Desta, A. G. Birhanu, "Advancements in Single-cell RNA Sequencing and Spatial Transcriptomics: Transforming Biomedical Research," *Acta Biochimica Polonica* 72 (2025): 13922.