Check for updates

OPEN ACCESS

EDITED BY Lei Wang, Chinese Academy of Sciences (CAS), China

REVIEWED BY N. T. Pramathesh Mishra, Dr. A.P.J. Abdul Kalam Technical University, India Wei Mengmeng, China University of Mining and Technology, China

*CORRESPONDENCE Ruby Srivastava, ⊠ amitruby1@gmail.com

RECEIVED 11 March 2025 ACCEPTED 21 April 2025 PUBLISHED 30 April 2025

CITATION

Srivastava R (2025) Advancing precision oncology with Al-powered genomic analysis. *Front. Pharmacol.* 16:1591696. doi: 10.3389/fphar.2025.1591696

COPYRIGHT

© 2025 Srivastava. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Advancing precision oncology with AI-powered genomic analysis

Ruby Srivastava 🕩 *

Bioinformatics, Centre for Cellular and Molecular Biology-CSIR, Hyderabad, India

Multiomics data integration approaches offer a comprehensive functional understanding of biological systems, with significant applications in disease therapeutics. However, the quantitative integration of multiomics data presents a complex challenge, requiring highly specialized computational methods. By providing deep insights into disease-associated molecular mechanisms, multiomics facilitates precision medicine by accounting for individual omics profiles, enabling early disease detection and prevention, aiding biomarker discovery for diagnosis, prognosis, and treatment monitoring, and identifying molecular targets for innovative drug development or the repurposing of existing therapies. Al-driven bioinformatics plays a crucial role in multiomics by computing scores to prioritize available drugs, assisting clinicians in selecting optimal treatments. This review will explain the potential of Al and multiomics data integration for disease understanding and therapeutics. It highlight the challenges in quantitative integration of diverse omics data and clinical workflows involving AI in cancer genomics, addressing the ethical and privacy concerns related to AI-driven applications in oncology. The scope of this text is broad yet focused, providing readers with a comprehensive overview of how AI-powered bioinformatics and integrative multiomics approaches are transforming precision oncology. Understanding bioinformatics in Genomics, it explore the integrative multiomics strategies for drug selection, genome profiling and tumor clonality analysis with clinical application of drug prioritization tools, addressing the technical, ethical, and practical hurdles in deploying AI-driven genomics tools.

KEYWORDS

genomics, precision oncology, bioinformatics, artificial intelligence, therapeutics and analysis resource

1 Introduction

The emergence of advanced and cost-effective high-throughput technologies (Misra et al., 2018) has generated vast amounts of biological data, ushering in a new era of precision medicine in oncology (Srivastava, 2023a). Precision medicine offers significant potential for cancer treatment and management, enabling oncologists to tailor therapies for individual patients. Precision oncology focuses on treating specific groups of cancer patients by utilizing population-specific diagnostic or prognostic biomarkers. This information is crucial for monitoring disease progression and assessing a patient's response to treatment. Additionally, it helps identify the molecular mechanisms underlying drug resistance, allowing for the targeted inhibition of genes or pathways responsible for resistance. Precision medicine relies on large datasets that must be processed and

analysed to detect molecular patterns and make patient-specific treatment decisions. However, handling these extensive datasets is both costly and time-consuming, a challenge exacerbated by the continuous growth of data due to high-throughput technologies (Misra et al., 2018; Ahmed, 2020; Srivastava, 2024a; Srivastava, 2024b).

Artificial intelligence (AI) and machine learning (ML) provide solutions to these challenges. AI encompasses a range of machinedriven functions, including rule-based logic, machine learning (ML), deep learning (DL), natural language processing (NLP), and computer imaging (Srivastava, 2023a; Ahmed, 2020). The rapid advancement of technologies capable of generating vast amounts of omics data-such as genomic, transcriptomic, proteomic (phenotypic), and epigenomic data-has underscored the necessity of AI in medical data analysis. The surge in genomic and transcriptomic data is primarily attributed to next-generation sequencing (NGS), while the increase in proteomic data results from mass spectrometric analysis (Srivastava, 2023a; Srivastava, 2024a; Srivastava, 2024b). AI can also predict the impact of genetic mutations on protein structure and function. Treatment efficacy and adverse effects can vary based on factors such as age, sex, genetics, and environmental influences, including anthropometric and metabolic status, dietary patterns, and lifestyle choices (Jaccard et al., 2018). Precision medicine aims to design the most effective interventions based on an individual's biological profile (Tebani et al., 2016). Clinical data and omics information can be obtained from databases or collected through screening technologies for various purposes, including disease diagnosis (Menyhárt and Gyorffy, 2021), class prediction (Hasin et al., 2017), biomarker discovery (Sun and Hu, 2016), disease subtyping (Menyhárt and Gyorffy, 2021), enhanced systems biology insights (Dahal et al., 2020), and drug repurposing (Srivastava, 2022; Srivastava, 2023b).

Multiomics refers to the comprehensive analysis of multiple layers of biological data such as genomics (DNA), transcriptomics (RNA), proteomics (proteins), epigenomics (epigenetic modifications), and metabolomics (metabolites) to gain a holistic understanding of biological systems (Chen et al., 2023). Integrating these diverse omics datasets is crucial for precision oncology because cancer is a complex, multi-factorial disease involving alterations at various molecular levels. By combining insights from different omics layers, researchers and clinicians can uncover more accurate biomarkers, better understand tumor heterogeneity, and identify personalized therapeutic targets, ultimately leading to more effective, tailored cancer treatments (Correa-Aguila et al., 2022). The current applications of AI in multiomics data analysis, emphasizes its role in precision oncology and therapeutics. Transcriptomics research has provided insights into the molecular mechanisms underlying both physiological processes (e.g., developmental stages, cell cycle phases) and pathological conditions, leading to clinical applications such as MammaPrint®, a 70-gene panel used to predict the risk of relapse and metastasis in breast cancer (Hamet and Tremblay, 2017). While single-omics analyses have contributed valuable findings, many prevalent diseases with high mortality rates, such as type 2 diabetes and cardiovascular disease, still lack effective therapeutic solutions (Hamet and Tremblay, 2017). This is partly because the functions of genetic variants are not always easily interpretable, often limiting the development of targeted treatments.

The human genome consists of approximately 3 billion base pairs, encompassing both coding and non-coding regions (Makałowski, 2001). A key distinction within the genome is the difference between introns and exons—introns are non-coding segments of genes, whereas exons represent coding regions responsible for protein production. Whole Genome Sequencing (WGS) is a comprehensive technique that sequences an organism's entire DNA, allowing for the identification of genetic variants and providing a complete picture of genomic composition. As mentioned earlier, genomics research focuses on genetic variations such as single nucleotide polymorphisms (SNPs) and larger structural changes that contribute to an organism's genetic makeup. SNPs are the most prevalent form of genetic variation, representing differences in a single nucleotide (Nakagawa et al., 2015; Nakagawa and Fujita, 2018).

Genomic alterations affecting large DNA segments (≥50 bp) are classified as structural variations (SVs). These variations arise from multiple mutational mechanisms, including deletions, insertions, and duplications, which alter the genomic sequence quantity and are collectively referred to as copy-number variations (CNVs) (Escaramıs et al., 2015; Li et al., 2020; Ho et al., 2020). Various techniques have been employed to study CNVs, with Whole Genome Sequencing (WGS) increasingly emerging as the preferred approach due to its declining costs and continuous improvements in variant detection methods (Pos et al., 2021).

CNVs can be analyzed using WGS by identifying genomic regions with an abnormal number of sequencing reads compared to expected levels, a method known as depth of coverage (DOC) analysis. Given its ability to yield reliable results even at shallow sequencing depths (0.19–1.09 coverage of the genome), CNV analysis has become a valuable tool in clinical diagnostics (Dong et al., 2017) Several tools, such as WISECONDORX, have been developed to characterize CNVs and assess their clinical and therapeutic significance (Raman et al., 2019).

While WGS sequences all DNA, including both coding and noncoding regions, Whole Exome Sequencing (WES) specifically targets protein-coding regions of genes (Lelieveld et al., 2015). WES selectively sequences these regions along with approximately 20 nucleotides of adjacent intronic sequences to investigate protein-coding areas in greater detail. The field of genomics was first introduced by American geneticists in 1986 to study the composition, structure, function, localization, and editing of DNA. Today, genomics is used to analyze all genes within an organism, providing insights into their biological significance. Advancements in genomic technologies have made it possible to efficiently analyze whole-genome data, leading to the discovery of genes, proteins, and biological pathways associated with diseases. In drug-target screening, genomic technologies compare DNA sequencing data from tumor and non-malignant tissues to identify key genetic differences. These differential genes can serve as potential drug targets and can be further validated using CRISPR-Cas9 knockout technology, allowing researchers to individually screen and assess their impact (Chan et al., 2022; Yamamoto et al., 2019). Genomic research is divided into three primary areas: structural genomics, functional genomics, and comparative genomics. Structural genomics focuses on analyzing nucleotide sequences through whole-genome sequencing to determine genome composition and gene positioning. Functional genomics

involves modifying gene sequences or their expression within cells to observe resulting phenotypic changes, thereby linking genotype to phenotype and clarifying gene functions. Comparative genomics examines variations in genome structure and function across different species to understand their evolutionary and biological relationships (Haley and Roudnicky, 2020). Functional genomics, which explores gene functions and networks, has become a crucial tool for understanding the complex interactions within human tumors and their microenvironments. Technologies such as RNA interference (Yin and Kassner, 2016; Adams et al., 2016), small interfering RNA (siRNA) (Zhang et al., 2007), short hairpin RNA (shRNA) (Takase et al., 2017), CRISPR interference, and CRISPR inhibition (le Sage et al., 2017; DePristo et al., 2011) are instrumental in drug-target discovery and validation. Bioinformatics plays a critical role in analyzing cancer somatic mutations, as these mutations are key targets for precision therapies that minimize damage to healthy cells. However, germline variants, which influence drug metabolism and substrate interactions, can significantly impact drug efficacy and toxicity. Therefore, considering both somatic and germline variations is essential when developing personalized treatment strategies (Koboldt, 2020).

The performance of different AI algorithms as deep learning (DL) and traditional machine learning (ML) in multi-omics data analysis has been extensively evaluated across various studies, each offering distinct strengths and limitations depending on the nature of the task and dataset (Wei et al., 2023). Traditional machine learning algorithms such as random forests, support vector machines (SVMs), k-nearest neighbors (KNN), and gradient boosting methods (e.g., XGBoost) have been widely used due to their robustness, ease of implementation, and interpretability. These models typically perform well in properly structured, relatively low-dimensional, and preprocessed datasets (Sarker, 2021). They are particularly effective when individual omics layers are analyzed separately or integrated in a relatively simple manner (e.g., feature concatenation). However, their performance may plateau when dealing with complex, high-dimensional multi-omics datasets due to limited capacity to capture nonlinear relationships and cross-omics interactions (Subramanian et al., 2020). In contrast, deep learning algorithms such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), autoencoders, variational autoencoders (VAEs), and graph neural networks (GNNs) excel in modeling complex and high-dimensional multi-omics data (Alzubaidi et al., 2021). DL models are particularly well-suited for capturing intricate, nonlinear interactions between various omics layers (genomics, transcriptomics, proteomics, etc.), making them powerful tools for tasks such as disease classification, biomarker discovery, and survival prediction (Nguyen PHD. et al., 2021). For example, multi-modal deep learning frameworks that use multi-omics integration techniques like late fusion, intermediate fusion, or attention mechanisms often outperform traditional ML approaches in predictive performance (Nakach et al., 2024). However, deep learning requires large amounts of data to avoid overfitting and due to its limited interpretability, it generate significant drawback in clinical contexts. Moreover, ensemble and hybrid approaches which combine both ML and DL techniques are increasingly used to leverage the strengths of each method (Almulihi et al., 2022). For instance, feature representations extracted from deep learning models can be fed into ML classifiers for improved performance and interpretability. Emerging trends also emphasize explainable AI (XAI) to improve the transparency of deep learning models in biomedical applications (Mathew et al., 2025).

The effectiveness of different AI algorithms particularly deep learning (DL) versus traditional machine learning (ML) in analyzing multiomics data across various types of cancer varies depending on cancer type, data complexity, and clinical endpoints. Each algorithm has unique strengths that make it more suitable for certain scenarios in cancer research (Arjmand et al., 2022). In breast cancer, where extensive multiomics datasets (e.g., genomics, transcriptomics, proteomics, and methylation data) are available from sources like TCGA (Tomczak et al, 2015), DL models such as autoencoders and multimodal neural networks have demonstrated superior performance in tasks like subtype classification, survival prediction, and treatment response modeling (Yang et al., 2024). DL excels due to its ability to capture complex nonlinear interactions among heterogeneous data layers. However, ML algorithms like random forests and SVMs still perform competitively when applied to well-engineered features, offering higher interpretability and lower computational costs. In lung cancer, particularly non-small cell lung cancer (NSCLC), ensemble ML methods (e.g., XGBoost and random forests) have shown strong predictive performance in biomarker discovery and prognosis prediction, especially when integrating genomics with transcriptomics or imaging data (Li et al., 2022). DL methods, especially graph neural networks (GNNs), have become effective in modeling protein-protein interaction networks or pathwaylevel features but are limited by smaller cohort sizes in some datasets. In glioblastoma multiforme (GBM), a highly heterogeneous brain tumor, DL particularly variational autoencoders and multi-view DL architectures outperforms traditional ML in integrating multiomics data (e.g., methylation, copy number variation, transcriptomics) for subtype discovery and survival prediction (Poursaeed et al., 2024). DL's capacity to learn latent representations from noisy, high-dimensional data is especially beneficial in such complex cancers. For colorectal and prostate cancers, where the multiomics landscape is less characterized than breast or lung cancer, ML algorithms are often preferred due to smaller sample sizes. Random forests and logistic regression are widely used for classification and feature selection, especially in studies focused on diagnostic and prognostic biomarker discovery (Wei et al., 2022; Hachem et al., 2024; Bao et al., 2024). In pan-cancer studies, where multiomics data across multiple tumor types are analyzed collectively, DL models like multimodal neural networks and transformers have emerged as powerful tools for learning shared and distinct features across cancers, enabling crosscancer subtype clustering and drug response prediction. However, pan-cancer DL models require large, well-curated datasets and advanced techniques to avoid biases introduced by imbalanced data distributions across cancer types (Divate et al., 2022). A hybrid approach, combining the representation power of DL with the interpretability of ML, is increasingly adopted to leverage the strengths of both in cancer multiomics research (Mavaie et al., 2023).

2 Bioinformatics approaches in genomics

Genomic data analysis involves examining single nucleotide polymorphisms (SNPs), copy number variations, gene expression, microRNA expression, protein expression, and other genetic alterations. Precision oncology leverages high-throughput technologies and bioinformatics tools to personalize cancer treatments based on individual genetic profiles. This approach enhances the ability to identify and validate biomarkers crucial for cancer diagnosis, prognosis, and tailored therapeutic strategies. Successfully integrating bioinformatics into precision oncology requires expertise in oncology, bioinformatics, and biostatistics (Li et al., 2018; Szymczak et al., 2009). Whole Genome Sequencing and Whole Exome Sequencing enable rapid, comprehensive analysis of genetic mutations, SNPs, and structural variations within tumors.

Bioinformatics tools play a key role in data integration and interpretation, particularly in variant annotation and functional prediction. Tools like ANNOVAR (Wang et al., 2010) facilitate the identification of actionable mutations by providing scores from multiple predictive models, including SIFT (Ng and SIFT, 2003), PolyPhen-2 (Adzhubei et al., 2013), LRT (Chen et al., 2020), FATHMM (Rogers et al., 2018), MetaSVM and MetaLR (Dong et al. 2015), VEST3 (Carter et al., 2013), and CADD (Kircher et al. 2014). SIFT (Ng and SIFT, 2003) determines whether a variant is deleterious by using PSI-BLAST to assess amino acid conservation across closely related sequences. PolyPhen-2 (Adzhubei et al., 2013) employs a pipeline combining eight sequence-based and three structure-based methods to classify mutations as benign, probably deleterious, or known to be deleterious. The Likelihood Ratio Test (LRT) (Chen et al., 2020) evaluates conservation across closely related species to determine the functional impact of mutations. FATHMM (Rogers et al., 2018) utilizes Hidden Markov Models and sequence conservation to predict the effects of missense mutations on protein function. MetaSVM and MetaLR are ensemble methods that integrate ten predictor scores (SIFT, PolyPhen-2 HDIV, PolyPhen-2 HVAR, GERP++, MutationTaster, Mutation Assessor, FATHMM, LRT, SiPhy, and PhyloP) along with the maximum observed frequency from the 1,000 Genomes Project to predict deleterious variants. MetaSVM is based on Support Vector Machines (SVM), while MetaLR employs Logistic Regression (LR) to generate final variant scores.

Pathway analysis tools, such as Ingenuity Pathway Analysis (IPA) (Krämer et al., 2014) and Gene Set Enrichment Analysis (GSEA) (Mootha et al., 2003), play a crucial role in identifying disrupted biological pathways and networks, offering valuable insights into tumorigenesis. Bioinformatics has been instrumental in pinpointing immunotherapy targets like Programmed Death Ligand 1 (PD-L1) (Han et al., 2020) and in identifying biomarkers for epidermal growth factor receptor (EGFR) inhibitors in non-small cell lung cancer (Prabhakar, 2015), as well as poly (ADP-ribose) polymerase (PARP) inhibitors for cancers with Breast Cancer Gene 1/2 (BRCA1/2) mutations (Faraoni and Graziani, 2018). These biomarkers undergo rigorous validation to ensure their accuracy and clinical relevance.

As biomarker identification techniques improve and genomic data repositories continue to expand, the sheer complexity and

volume of data necessitate more sophisticated analytical tools. This growing demand has fueled the increasing reliance on machine learning (ML) and predictive algorithms. ML techniques excel at handling large, high-dimensional datasets, uncovering patterns, and establishing relationships within the data. By employing dimensionality reduction and feature selection, ML algorithms enhance the efficiency of biological data analysis, allowing for the evaluation of disease mechanisms and the identification of potential biomarkers. To enhance cancer patient care, precision treatment should include monitoring and managing Quality of Life (QoL) data collected in the patient's home environment, along with its integration and analysis. Recent advanced technologies has facilitated the development of smartphone devices that support both patients and clinicians by consolidating all relevant patient data and assisting with patientreported outcomes (Srivastava, 2023b; Srivastava, 2023c). Genomewide association studies (GWAS) have also generated vast amounts of genomic data for cancer research. The successful application of patient-specific data in precision medicine hinges on the accurate integration, analysis, and interpretation of these datasets to provide a comprehensive overview of gene expression changes in individual cancer patients (Li et al., 2018; Szymczak et al., 2009; Telenti et al., 2018). Such analyses can reveal alterations in metabolic and signaling pathways specific to a patient, paving the way for highly personalized treatment plans. This multidimensional approach offers significant advantages over traditional singlelayer analyses, which focus on isolated features (Chari et al., 2010; Wang et al., 2014). However, for AI to effectively interpret such data, it must first be trained to recognize key features and patterns.

Over the past decade, large-scale cancer research initiatives have emerged to streamline the analysis of omics data. Projects such as The Cancer Genome Atlas (TCGA) [27], the International Cancer Genome Consortium (ICGC) (ICGC, 2022 Hudson et al., 2010), COSMIC (Tate et al., 2019), TARGET, and the German Cancer Consortium (DKTK) (Joos et al., 2019), along with platforms like the Genomic Data Commons (GDC) [32], cBioPortal (Cerami et al., 2012), UCSC Genome Browser (Rosenbloom et al., 2013), Array Express (Parkinson et al., 2007), and Gene Expression Omnibus (GEO), have significantly contributed to this effort. The list of various databases, their links and types of data analyses of human tumors and tumor cell lines are given in Table 1.

For instance, a study on lung adenocarcinoma by Gillette et al. (2020) [37] refined tumor classification by dividing the proximalproliferative cluster using transcriptomic data, deep-scale proteomic profiling, and post-translational modifications. Despite these advancements, several challenges persist in integrative analysis, including (i) the high dimensionality of data, which complicates inference; (ii) inherent heterogeneity across different technical platforms, reducing biological signal clarity; (iii) the diversity of data types, making it unlikely that a single analytical method will be applicable across all omics layers; and (iv) the difficulty of interpretation due to the sheer volume of information, which can obscure meaningful conclusions. To address these challenges, various integrative approaches have been developed, focusing on (i) patient stratification, (ii) clinical outcome prediction, and (iii) identifying molecular mechanisms that operate across different biological layers (Srivastava, 2024a; Srivastava, 2024b). Recent

SN	Database	Link	Types of data/analyses
1.	TCGA	https://www.cancer.gov/ccg/research/genome-sequencing/ tcga	20,000 primary cancer and matched normal samples spanning 33 cancer types
2.	ICGC-AGRO	https://www.icgc-argo.org/	ICGC ARGO will uniformly analyze specimens from 100,000 cancer patients
3.	GDC	https://portal.gdc.cancer.gov/	Explore and analyse clinical and genomic data from cancer genomics studies
4.	GEO	https://www.ncbi.nlm.nih.gov/geo/	GEO is a public functional genomics data repository supporting MIAME- compliant data submissions.
5.	Array Express	https://www.ebi.ac.uk/biostudies/arrayexpress	Stores high-throughput genomics data.
6.	ENA	https://www.ebi.ac.uk/ena/browser/home	Stores raw data files in Fastq format.
7.	Sequence Read Archive	https://www.ncbi.nlm.nih.gov/sra	Stores raw data files in SRA format.
8.	Dryad Digital Repository	https://datadryad.org/stash	Open access repository of medical research data
9.	Figshare	https://figshare.com/	Cross-disciplinary open-access repository for academic research
10.	Harvard Dataverse Network	https://dataverse.harvard.edu/	Multi-disciplinary data storage center
11.	Kaggle	https://www.kaggle.com/	Platform for data science training, competitions, and datasets
12.	Network Data Exchange	https://home.ndexbio.org/about-ndex/	Repository for network biology data
13.	Open Science Framework	https://osf.io/	Platform for collaborating on research projects
14.	GenoVault	https://github.com/bioinformaticscdac/GenoVault	Cloud-based repository for NGS data
15.	UK Biobank	https://www.ukbiobank.ac.uk/	Large-scale biomedical research database
16.	cBioPortal	https://www.cBioPortal.org/	Visualizations, analysis, cancer genomics projects
17.	COSMIC	https://cancer.sanger.ac.uk/cosmic	Database of somatic mutations in cancer.
18.	IGV	https://software.broadinstitute.org/software/igv/	High-performance genome browser for visualizing and analyzing large- scale genomic data.
19.	Regulome Explorer	https://explorer-cancerregulome.systemsbiology.net/	Exploring and analyzing regulatory elements in the genome.
20.	UCSC Genome Browser	https://genome.ucsc.edu/	Provides access to a vast collection of genomic data and annotations
21.	Bioconductor	https://www.bioconductor.org/	Open-source software project for the analysis and comprehension of high- throughput genomics data.
22.	Cytoscape	https://cytoscape.org/	Network analysis and visualization tool
23.	Gene Ontology	http://geneontology.org/	Standardized system for annotating genes and their functions in different organisms.
24.	UALCAN	https://ualcan.path.uab.edu/	Web portal for in-depth analysis of cancer transcriptome data.
25.	DAVID	https://david.ncifcrf.gov/	Functional annotation and enrichment analysis of gene lists
26.	HumanBase (GIANT)	https://hb.flatironinstitute.org/	Exploring human genomic data and conducting large-scale integrative analysis.
27.	CEDER	https://pmc.ncbi.nlm.nih.gov/articles/PMC3488134/	Detection of differentially expressed genes
28.	CPTRA	https://gdc.cancer.gov/about-gdc/contributed-genomic- data-cancer-research/clinical-proteomic-tumor-analysis- consortium-cptac	Package for analyzing transcriptome sequencing data
29.	Bioconductor	https://www.bioconductor.org/	Open-source software for genomic data analysis
30.	LIMMA	https://bioconductor.org/packages/release/bioc/html/limma. html	Statistical package for the analysis of microarray and RNA-seq data.
31.	CARET	https://cran.r-project.org/web/packages/caret/index.html	R package for training and evaluating ML models.

TABLE 1 The list of various databases, their links and types of data analyses of human tumors and tumor cell lines.

(Continued on following page)

SN	Database	Link	Types of data/analyses
32.	netClass	https://doi.org/10.1093/bioinformatics/btu025	A tool for classifying biological samples using network-based features.
33.	WGCNA	https://horvath.genetics.ucla.edu/html/ CoexpressionNetwork/Rpackages/WGCNA/	Identifying gene modules and their relationships in high-throughput data.
34.	MyCancerGenome	https://www.mycancergenome.org/	Understanding cancer genomics and personalized cancer treatment options.
35.	CIViC	https://civicdb.org/welcome	Treatment options for cancer patients based on their unique tumor DNA Molecular characterization
36.	TARGET	https://www.cancer.gov/ccg/research/genome-sequencing/ target	Molecular characterization
37.	CGI	https://www.genomicinterpretation.org/	Genomic alterations in cancer and their potential clinical relevance
38.	ClinicalTrials.gov	https://www.clinicaltrials.gov/	An online database that provides information on clinical trials
39.	EUCTR	https://www.clinicaltrialsregister.eu/	Database containing information on clinical trials conducted in the European Union
40.	OncoKB	https://www.oncokb.org/	Mutations, CNVs, fusions
41.	DepMap	https://depmap.org/portal/	Genetic loss-of-function screening, pharmacologic dependencies, CCLE omics characterizations
42.	canSAR.ai	https://cansar.ai/	Integrates biology, chemistry, pharmacology, structural biology, cellular networks and clinical annotations, and applies machine learning approaches to develop predictions useful in drug discovery
43.	ClinGen	https://clinicalgenome.org/	Clinical relevance of genes and variants
44.	ClinVar	www.ncbi.nlm.nih.gov/clinvar/intro/	Database of genomic variants with public submissions of variant interpretations and disease relations.
45.	Cosmic	https://cancer.sanger.ac.uk/cosmic	Catalogue Of Somatic Mutations In Cancer
46.	dbSNP	https://www.ncbi.nlm.nih.gov/snp/	Contains human single nucleotide variations, microsatellites, and small- scale insertions and deletions
47.	Ensembl	https://www.ensembl.org/index.html	Genome browser for vertebrate genomes that supports research in comparative genomics, evolution, sequence variation and transcriptional regulation
48.	Find Zebra	https://www.findzebra.com/	Tool for helping diagnosis of rare diseases. It uses freely available high quality curated information on rare diseases
49.	Genomics England	https://www.genomicsengland.co.uk/	Comprehensive site describing the progress of the UK sequencing initiative. Site contains usefull overviews over gene panels and diseases.
50.	gnomAD	https://gnomad.broadinstitute.org	Exome and genome sequencing data with allele frequencies from a wide variety of large-scale sequencing projects
51.	GTEX	https://gtexportal.org/home/	Comprehensive public resource to study tissue-specific gene expression and regulation. Samples were collected from 54 non-diseased tissue sites across nearly 1000 individuals, primarily for molecular assays including WGS, WES, and RNA-Seq.
52.	HGMD	https://www.hgmd.cf.ac.uk/ac/index.php	Collate all known (published) gene lesions responsible for human inherited disease
53.	Human Phenotype Ontology (HPO)	https://hpo.jax.org/app/	Provides a standardized vocabulary of phenotypic abnormalities encountered in human disease
54.	Matchmaker Exchange	https://www.matchmakerexchange.org	Genomic discovery through the exchange of phenotypic and genotypic profiles
55.	MaveDB	https://www.mavedb.org/	Collection, distribution, and analysis of variant effect maps
56.	MedGen	https://www.ncbi.nlm.nih.gov/medgen/	Organizes information related to human medical genetics, such as attributes of conditions with a genetic contribution
57.	NCBI	https://www.ncbi.nlm.nih.gov/	The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information

TABLE 1 (Continued) The list of various databases, their links and types of data analyses of human tumors and tumor cell lines.

(Continued on following page)

SN	Database	Link	Types of data/analyses
58.	OMIM	https://www.omim.org/	Compendium of human genes and genetic phenotypes
59.	RefSeq	https://www.ncbi.nlm.nih.gov/refseq/	A comprehensive, integrated, non-redundant, well-annotated set of reference sequences including genomic, transcript, and protein.
60.	Uniprot	https://www.uniprot.org/	Comprehensive and freely accessible resource of protein sequence and functional information.
61.	RefSeq	https://www.ncbi.nlm.nih.gov/refseq/	A comprehensive, integrated, non-redundant, well-annotated set of reference sequences including genomic, transcript, and protein.

TABLE 1 (Continued) The list of various databases, their links and types of data analyses of human tumors and tumor cell lines.

Abbreviations: TCGA, The Cancer Genome Atlas Program; ICGCAGRO, The International Cancer Genome Consortium Accelerating Research in Genomic Oncology; GDC, Genomic Data Commons; GEO, Gene Expression Omnibus; ENA, The European Nucleotide Archive; COSMIC, Catalogue Of Somatic Mutations In Cancer; IGV, Integrative Genome Viewer; UALCAN, University of Alabama at Birmingham Cancer Data Analysis Portal; DAVID, Database for Annotation, Visualization, and Integrated Discovery; CEDER, Cancer Epitope Database and Analysis Resource; CPTRA, cross-platform transcriptome analysis; Limma, Linear Models for Microarray and Omics Data; CARET, Carotene and Retinol Efficacy Trial; CIViC, Clinical Interpretation of Variants in Cancer; TARGET, *Therapeutically Applicable Research to Generate Effective Treatments*; WGCNA, Weighted correlation network analysis; CGI, Cancer Genome Interpreter; EUCTR, EU Clinical Trials Register; GTEx, Genotype-Tissue Expression; HGMD, Human Gene Mutation Database; NCBI, The National Center for Biotechnology Information, OMIM, *Online Mendelian Inheritance in Man*; CCLE, Cancer Cell Line Encyclopedia; CNV, copy-number variation; DepMap, Dependency Map.



studies have applied causal inference techniques to either validate (Ber existing biological relationships (Dugourd et al., 2021) or infer stable connections across multiple experimental conditions without prior knowledge (Meinshausen et al., 2016). Different classification frameworks have been proposed based on application type (unsupervised vs. supervised, with the latter further divided into predictive and explanatory models), strategy (early, intermediate, or late integration), and methodology. The literature commonly categorizes six major families of integrative methods: matrix factorization, Bayesian approaches, multiple kernel learning, ensemble learning, deep learning, and network-based methods

(Bersanelli et al., 2016; Huang et al., 2017; Krassowski et al., 2020; Vahabi and Michailidis, 2022; Picard et al., 2021; Subramanian et al., 2020; Broad Institute, 2018).

The bioinformatics pipeline developed and implemented at the Utah Public Health Laboratory (UPHL) consists of eight key steps: (1) read quality control, (2) reference strain identification, (3) read mapping to the reference strain, (4) detection of single nucleotide polymorphisms and small insertions or deletions (INDELs), (5) *de novo* genome assembly, (6) genome annotation, (7) phylogenetic tree construction, and (8) phylogenetic analysis. While these processes are standard, multiple software tools are available to

perform each step (Srivastava, 2023b). The first step in genomicsbased drug selection involves identifying clinically relevant alterations in cancer patients through variant calling analysis. According to the Genome Analysis Toolkit (GATK) (DePristo et al., 2011), the general workflow of variant calling includes nine steps: quality control (QC) and trimming, alignment, marking duplicates, local realignment of INDELs, base quality score recalibration (BQSR), variant calling, filtering, and annotation of variants (Koboldt, 2020). The basic steps for application of bioinformatics in whole genome sequencing are given in Figure 1.

Following sample QC and trimming, raw sequencing reads are aligned to the reference genome using tools like BWA-MEM (Li, 2013). Duplicate reads are then removed with PICARD. GATK tools are employed to minimize alignment artifacts and improve sequencing quality estimates. Variant calling is performed using tools such as MUTECT2, HAPLOTYPECALLER (McKenna et al., 2010), VARSCAN 2 (Koboldt et al., 2012), VARDICT (Lai et al., 2016), or SOMATICSNIPER (Larson et al., 2012), which identify short variants, including single nucleotide variants (SNVs) and insertions or deletions (INDELs) of less than 50 base pairs (bp). The identified variants undergo filtering to eliminate low-quality calls, followed by annotation to determine their biological impact, population frequency, and clinical relevance. This analysis primarily focuses on somatic variants in coding regions. Nonsynonymous SNVs are considered more detrimental, as they alter the final protein sequence, potentially affecting its folding and function (Sun et at., 2019).

Somatic genomic alterations are classified based on their population frequency as either rare variants or polymorphisms. Variants with a high frequency (>1%) are generally deemed clinically benign. In most patients, at least one detected somatic alteration holds clinical significance (Bieg-Bourne et al., 2017; Sanchez-Vega et al., 2018), as it may influence gene function, suggest preventive surveillance, aid in diagnosis, impact prognosis, or guide treatment selection. Several automated variant annotation tools exist to streamline this process. SNPEFF (Cingolani et al., 2012) assesses the biological impact of candidate variants, while ANNOVAR (Wang et al., 2021) and the VARIANT EFFECT PREDICTOR (VEP) (McLaren et al., 2016) provide additional information on variant population frequency. The Variant Caller with Multinomial Probabilistic Model (VCMM) detects SNVs and INDELs from whole exome sequencing (WES) and whole genome sequencing (WGS) studies by using a multinomial probabilistic model with quality score and strand bias filters. VCMM reduces false-positive and false-negative variant calls compared to GATK and SAMtools, improving the accuracy of variant detection (Shigemizu et al., 2013).

Public data repositories provide valuable resources for annotating candidate somatic variants by linking them to drugs and their interconnections. Some key databases include ClinVar (Landrum et al., 2018), which catalogs genetic variants and their clinical significance; the Catalogue of Somatic Mutations in Cancer (COSMIC) (Tate et al., 2019), which compiles information on the impact of somatic mutations in cancer; OncoKB (Chakravarty et al., 2017) and CIViC (Griffith et al., 2017), which associate somatic cancer variants with clinical and therapeutic implications; and DGIdb (Cotto et al., 2018), a database of gene–drug interactions. See Table 1. These patient-centered tools analyze somatic variants in tumors and can be categorized based on the type of input data required. For instance, with a list of available variants, resources like MTB-REPORT (Perera-Bel et al., 2018), the Cancer Genome Interpreter (CGI) (Tamborero et al., 2018), the Variant Interpretation for Cancer Consortium Meta-Knowledgebase (VICC METAKB) (Wagner et al., 2020), PREMEDKB (Yu et al., 2019), and the SMART Cancer Navigator (Warner et al., 2018) can be useful. Some tools also accept disease or drug-related queries. If a variant calling file (VCF) is available, platforms such as MTBP (Tamborero et al., 2022) and PANDRUGS offer additional support, with PANDRUGS accommodating both gene and drug queries. In addition to guiding therapy selection at the individual level, broader approaches have been developed to analyze treatment trends across different tumor types at a larger scale (Rubio-Perez et al., 2015). While most of these methods prioritize drug selection based solely on somatic variants, germline variants also play a crucial role in drug metabolism, influencing treatment effectiveness and potential toxicity (Menden et al., 2018). As a result, patients may exhibit varied responses to the same therapy, ranging from high efficacy to ineffectiveness or even adverse drug reactions (ADRs). ADRs are significant contributors to morbidity and mortality and pose a financial burden on healthcare systems (Khalil and Huang, 2020). Variability in drug response primarily stems from genetic differences in genes encoding drug substrates or those involved in xenobiotic metabolism and transport (Roden et al., 2019).

To optimize drug selection based on germline variants, pharmacogenomic databases such as DrugBank (Wishart et al., 2018), PharmGKB (Whirl-Carrillo et al., 2021), and the Table of Pharmacogenomic Biomarkers in Drug Labeling (https://www.fda.gov/media/124784/) can be leveraged to prioritize effective drugs while avoiding those that may be ineffective or cause ADRs. Tools like PHARMCAT (Sangkuhl et al., 2020) enable the development of personalized treatments based on germline variants found in VCF files. Additionally, some platforms, such as MTBP, integrate both germline and somatic variant data for a more comprehensive approach to treatment selection.

Numerous knowledge bases and bioinformatics tools are available for variant annotation, biomarker identification, drug prioritization, and response prediction, serving as essential resources to help clinicians determine the best treatment options for their patients. One emerging biomarker in this field is tumor mutational burden (TMB), which has shown promise in identifying patients most likely to benefit from immunotherapy across various cancer types (Hellmann and Paz-Ares, 2018). TMB is determined by calculating the total number of somatic mutations per megabase (Mbp) of sequenced DNA. However, the lack of standardization in TMB assessment remains a challenge, limiting its universal applicability as a biomarker. While high TMB is generally linked to better immunotherapy responses, its predictive value is not consistent across all cancer types (McGrail et al., 2021). Tumor Mutational Burden (TMB) has emerged as a promising genomic biomarker for identifying patients likely to benefit from immunotherapy across various cancer types (Hellmann and Paz-Ares, 2018). However, a lack of standardization in TMB assessment complicates its use as a universally reliable biomarker. While high TMB is generally linked to better immunotherapy responses, its predictive value varies across different cancer types (McGrail et al., 2021). Advances in bioinformatics now enable in-depth TMB

analysis, allowing for *in silico* hypothesis generation that goes beyond simple TMB-based patient stratification. By leveraging these insights, targeted therapies can be prioritized based on mutations with established treatment options. Bioinformatics advancements now allow for an in-depth analysis of TMB, generating *in silico* hypotheses that extend beyond simple TMBbased patient stratification. These insights can be used to prioritize targeted therapies based on mutations with known treatment options. For instance, PANDRUGS [24] is a platform that ranks drug treatments based on actionable mutations found in TMB, enabling a more precise approach to therapy selection.

Additionally, numerous bioinformatics tools and AI-based methodologies have been designed to facilitate the interpretation of cancer-related variants and suggest potential treatment options based on prior evidence (Cotto et al., 2018). These patient-centered resources rely on tumor-specific somatic variants and can be categorized based on the type of input data they require. For instance, if a list of variants is available, tools such as MTB-REPORT (Perera-Bel et al., 2018), the Cancer Genome Interpreter (CGI) (Tamborero et al., 2018), the Variant Interpretation for Cancer Consortium Meta-Knowledgebase (VICC METAKB) (Wagner et al., 2020), PREMEDKB (Yu et al., 2019), and the SMART Cancer Navigator (Warner et al., 2018) can provide useful insights. Some tools also accept disease- or drugrelated queries, while those with access to a variant calling file (VCF) may utilize platforms such as MTBP (Tamborero et al., 2022) or PANDRUGS (Pineiro-Yanez et al., 2018), with the latter supporting both gene and drug queries.

Beyond individual patient analysis, large-scale approaches have been developed to guide treatment selection across various tumor types and identify broader trends in therapy response (Rubio-Perez et al., 2015). While many of these methods prioritize drug selection based on somatic variants alone, germline variants are equally important, as they play a crucial role in drug metabolism, effectiveness, and potential toxicity (Menden et al., 2018). This genetic variability leads to differing patient responses, ranging from positive therapeutic outcomes to ineffectiveness or even adverse drug reactions (ADRs). ADRs significantly contribute to morbidity, mortality, and increased healthcare costs (Khalil and Huang, 2020).

Differences in drug response are largely attributed to genetic variations in genes encoding drug substrates or those involved in xenobiotic metabolism and transport (Roden et al., 2019). By leveraging pharmacogenomic databases such as DrugBank (Wishart et al., 2018), PharmGKB (Whirl-Carrillo et al., 2021), and the Table of Pharmacogenomic Biomarkers in Drug Labeling (https://www.fda.gov/media/124784/), effective drugs can be prioritized over those that are ineffective or may cause ADRs. Tools like PHARMCAT (Sangkuhl et al., 2020) provide tailored treatment recommendations based on germline variants found in VCF files. Furthermore, platforms like MTBP integrate both germline and somatic variant data for a more comprehensive approach to personalized medicine (Borchert et al., 2021; Yao et al., 2020).

A crucial aspect of studying mutational events is the ability to differentiate significant mutations from those commonly found in the healthy population (Zarrei et al., 2015). Mutation detection approaches can be categorized into two main types: reference-free and reference-based methods (Raman et al., 2019). Reference-free methods normalize samples using inherent genomic features such as GC content and mappability, while reference-based tools rely on either a single normal sample matched to the sample of interest or a Panel of Normals (PON) (PON, 2021). The inclusion of normal samples helps eliminate variations introduced by experimental factors such as sample handling, preparation, and sequencing technology.

structural variants, including those that cause copy-number variations (CNVs), can be highly complex in their impact on the genome (Schutte et al, 2019; Baca et al., 2013). Their characterization depends on various techniques, including paired-read and split-read analysis, as well as *de novo* genome assembly of the sample. However, the short read length of next-generation sequencing (NGS) imposes limitations on these analyses. The advent of advanced bioinformatics tools and long-read sequencing technologies has addressed these challenges, providing deeper insights into SVs (Cameron et al., 2021). Nanopore-based sequencers, for example, offer advantages such as portability and real-time data analysis. Additionally, bioinformatics tools facilitate the clinical characterization of SVs for diagnostic applications (Valle-Inclan et al., 2021).

SVs influence both germline and somatic genomic instability, contributing to disease development and potentially guiding therapy selection and drug response prediction. Some bioinformatics platforms designed for drug prioritization based on small variants also accept CNVs (Perera-Bel et al., 2018) and gene fusions (Wagner et al., 2020) as inputs. More sophisticated diagnostic approaches leverage shallow whole-genome sequencing (sWGS) for CNV analysis, aiming to establish CNV-based signatures that enable more precise diagnostics and treatment selection (Macintyre et al., 2018; van Belzen et al., 2021).

Mutational signatures in genomic DNA provide insights into the mutational processes driving cancer progression (Greenman et al., 2007; Degasperi et al., 2022). These signatures can be characterized by different mutation types, including single base substitutions (SBS), doublet base substitutions (DBS), insertions and deletions (indels), CNVs, and genomic rearrangements (Alexandrov et al., 2020. The identification of mutational signatures may aid in detecting therapeutically actionable biomarkers, supporting their use in personalized medicine. While over 30 mutational signatures have been identified, many remain of unknown origin. Some, however, have clear clinical relevance, such as those linked to tobacco exposure, ultraviolet (UV) radiation, and defects in DNA repair mechanisms, including mismatch repair and double-strand break repair. Studies have shown that tumors with DNA damage repair deficiencies exhibit therapeutic sensitivity to DNA-damaging agents and immunotherapy (Waddell et al., 2015; Ma J. et al., 2018; Ma X. et al., 2018; Connor et al., 2017). For example, a mutational signature associated with pathogenic BRCA1 and BRCA2 mutations in breast and ovarian cancers suggests homologous recombination (HR) deficiency, indicating sensitivity to PARP inhibitors (Lord and Ashworth, 2016) Conversely, prior exposure to DNA-damaging chemotherapy agents has been linked to drug resistance (Levatic et al., 2022). Mutational signatures also serve as molecular footprints of cancer therapies, helping estimate their contribution to tumor mutational burden (TMB) and revealing their long-term genomic effects (Pich et al., 2019). Tumor Mutational Burden (TMB) has

emerged as a significant biomarker for predicting responses to immunotherapy across various cancer types. However, the assessment of TMB has faced challenges due to variability in measurement techniques, leading to efforts aimed at standardizing its evaluation. TMB measurement can vary significantly across different cancer types and sequencing platforms (Jardim et al., 2021). Various factors including differences in panel size, gene content, and bioinformatics pipelines contribute to this variability. For example, certain cancers like uterine, bladder, and colon cancers exhibit greater variability in panel TMB values compared to lung and head and neck cancers. This variability underscores the necessity for standardized methodologies to ensure consistent and reliable TMB assessment. To address these challenges, initiatives such as Friends of Cancer Research (Friends) TMB Harmonization Project is undertaken in which significant strides are made in unifying TMB measurement across various laboratories. By identifying approaches to enhance consistency in evaluating the genetic mutations of tumors, the initiative aims to improve the reliability of TMB as a biomarker (Merino et al., 2020) Various guidelines are proposed to harmonize TMB quantification across different diagnostic platforms. These recommendations focus on standardizing TMB reporting, aligning analytical validation studies, and ensuring consistent methodologies in clinical samples. Consistent measurement methodologies enable more accurate predictions of patient responses to immune checkpoint inhibitors, thereby informing treatment decisions across various cancer types (Huang et al., 2021). The standardized TMB evaluation facilitates the comparison of clinical trial results and supports the broader application of TMB as a predictive biomarker in oncology. (Sha et al., 2020) Computational methods for mutational signature analysis vary in their mathematical frameworks and fall into two main categories: de novo discovery of novel signatures and refitting methods for detecting known signatures (Baez-Ortega and Gori, 2019; Omichessan et al., 2019). Tools such as SIGPROFILER (Bergstrom et al., 2019; Islam et al., 2020; Bergstrom et al., 2020; Kim et al., 2016), previously used in the COSMIC database, and SIGNATUREANALYZER (Kim et al., 2016; Kasar et al., 2015; Haradhvala et al., 2018; Degasperi et al., 2020) were instrumental in analyzing large cancer genome datasets from PCAWG, TCGA, and ICGC projects. SIGNAL, a web-based tool, not only identifies mutational signatures but also links them to gene drivers, potentially revealing novel therapeutic dependencies (Degasperi et al., 2020). Additionally, HRDETECT predicts HR deficiency, helping to stratify patients based on their likely response to PARP inhibitors (Davies et al., 2017). In large-scale genomic studies such as the 100,000 Genomes Project, somatic variant data-including sequencing coverage, small variants, and structural variations-is visually represented using a Circos plot, offering an intuitive overview of genomic alterations (Srivastava, 2024a). Advancements in high-throughput technologies with ML based approaches have enabled the generation of large-scale human gut microbiota profiles, driving growing interest in uncovering the links between the gut microbiome and complex human diseases. Results indicated accuracy in identifying individuals at high risk by extracting and integrating insights from complex microbiome datasets with challenges in managing the heterogeneity and sparsity of microbial features and in capturing the underlying

relationships among various human diseases (Huang et al., 2024). Data-tool such as scPriorGraph is used to construct biosemantic cell-cell graphs with prior gene set selection for cell type identification from scRNA-seq data (Cao et al., 2024).

3 Integrative multiomics strategies for drug selection

Advancements in high-throughput technologies have enabled the integration of multiple omics layers, providing a deeper understanding of biological systems (Hasin et al., 2017; do Valle et al., 2018). Tools such as PANOPLY (Kalari et al., 2018) and MOalmanac (Reardon et al., 2021) combine genomic and transcriptomic data to identify and prioritize potential drug targets. The Cancer Druggable Gene Atlas (TCDA) (Jiang et al., 2022) database compiles information on genomic alterations, including short variants, copy-number variations (CNVs), and gene fusions, along with gene expression, dependencies, and druggability.

DRUGCOMBOEXPLORER (Huang et al., 2019) incorporates DNA sequencing, gene copy number, methylation, and gene expression data from cancer patients to (a) identify key driver signaling pathways and (b) suggest effective anticancer drug combinations. Additionally, transcriptomic networks can be further enhanced with other omics layers, providing broader functional insights. For example, COSMOS (Dugourd et al., 2021) integrates phosphoproteomics, transcriptomics, and metabolomics to infer kinase and transcription factor activity. Deep learning algorithms are gaining popularity for multi-omics integration due to their ability to capture complex nonlinear and hierarchical relationships (Kang et al., 2022). One such tool, DEEPDRK (Wang et al., 2021), utilizes genomics, transcriptomics, epigenomics, and chemical compound properties to predict drug susceptibility in cancer cell lines and patients (Keskin et al., 2019).

Neoantigen prediction pipelines, such as PVACTOOLS (Hundal et al., 2020) incorporate computational tools to detect neoantigens from tumor DNA-seq and RNA-seq data. These tools also estimate an individual's HLA class and rank neoantigens based on their molecular compatibility with the patient's major histocompatibility complex (MHC) and other relevant parameters (Hackl et al., 2016). Furthermore, tools like CIBERSORTX (Newman et al., 2019) and MCP-COUNTER (Becht et al., 2016) analyze expression data to infer the presence of immune infiltrates in tumor tissue. Understanding the immune composition of a tumor, alongside tumor mutational burden (TMB) values, can aid in treatment selection. However, only a limited number of these tools currently prioritize drug treatments or neoantigen selection based on TMB content in clinical trials (Keskin et al., 2019). Intratumoural heterogeneity (ITH) within individual tumors is driven by a combination of somatic single nucleotide variants (SNVs), structural variations (SVs), transcriptomic and epigenetic modifications affecting gene expression, the tumor microenvironment (TME), and the antitumor immune response (Black and McGranahan, 2021; Nguyen TM. et al., 2021). ITH can be spatial, occurring in distinct tumor regions, or temporal, evolving over time through clonal progression. Understanding the extent of ITH and characterizing clonal subpopulations based on their unique mutational or transcriptomic profiles can be valuable for prioritizing drug treatments and predicting tumor response to therapy. This section provides an overview of key methodologies for dissecting ITH to guide drug selection.

A variety of user-friendly, web-based tools, such as Paintomics 4 (Liu et al. 2022), 3Omics (Kuo et al., 2013), and Galaxy (Galaxy, 2024), enable easy analysis with only a basic understanding of the underlying methodologies. More advanced tools, including integrOmics (Cao et al., 2009), SteinerNet (Tuncbag et al., 2012), Omics Integrator (Tuncbag et al., 2016), and MixOmics (Rohart et al. 2017), require programming expertise and offer customizable parameters for greater control over data analysis. Metabolomics datasets can be analyzed using the XCMSOnline (Tautenhahn et al., 2012) web tool, which integrates metabolomics data with genomic and proteomic information. A novel equivariant 3D-conditional diffusion model, called DiffFBDD, has been developed to generate new pharmaceutical compounds based on the 3D geometric structure of specific target protein pockets. DiffFBDD addresses the common underutilization of geometric information by leveraging an equivariant graph neural network to integrate detailed atomic-level data from protein pockets down to their backbone atoms (Zheng et al., 2025). AI-driven drug prioritization relies on a synergy between predictive modeling, network analysis, and knowledgebase integration, enabling personalized and data-driven therapeutic decision-making in oncology. Supervised learning models are widely used to correlate genomic alterations with drug response data. These models are trained on large pharmacogenomics datasets like GDSC (Genomics of Drug Sensitivity in Cancer) (Cokelaer et al., 2018) and CCLE (Cancer Cell Line Encyclopedia) to learn patterns between molecular features (e.g., gene expression, mutations) and drug sensitivity (Barretina et al., 2012). Deep learning based tools are particularly effective in integrating multi-layered omics data and capturing nonlinear interactions between genes, pathways, and drugs. Such as DeepDR (Jiang and Li, 2024), DeepSynergy (Preuer et al., 2018), and GraphDRP (Nguyen PHD. et al., 2021) utilize these architectures to predict drug response or synergistic drug combinations with higher accuracy. The drug prioritization scores can be computed using network-based methods, where biological networks (e.g., protein-protein interaction networks) are analyzed to identify key driver genes or pathways affected in a patient. These are then matched with known drug-target relationships using databases like DrugBank (https://go.drugbank.com/), DGIdb (Drug-Gene Interaction database) (https://dgidb.org/), and LINCS (Library of Integrated Network-Based Cellular Signatures) (Koleti et al., 2018). Other frameworks, such as OncoKB (Chakravarty et al., 2017), iCAGES (Dong et al., 2016), and PANOPLY (Mani et al., 2021), combine multiomics data with curated clinical and molecular knowledgebases to rank drugs based on patient-specific molecular alterations, mutation impact, and druggability. These tools not only improve treatment efficacy but also assist clinicians in identifying repurposable drugs and novel therapeutic strategies tailored to each patient's molecular landscape. AI systems typically generate a drug prioritization score based on predicted sensitivity (e.g., IC50 or AUC values), Drug-target interactions and pathway relevance, Molecular similarity between tumor and drug response signatures and Integration of clinical trial or approved drug data (Paul et al., 2021). These scores are then ranked to help clinicians identify the most promising therapies tailored to an individual's molecular cancer profile.

4 Genome profiling for tumor clonality

Tumors contain both clonal mutations, which are present in all cells, and subclonal mutations, which are restricted to specific subpopulations. The prevalence of subclonal mutations provides insight into tumor phylogeny, allowing researchers to identify active subclones and their evolutionary relationships. Cancer subclones undergo Darwinian evolution, where each subclone exhibits a distinct fitness level that can be inherited by daughter cells. Studies have shown that increased levels of CNVs may confer a selective advantage to certain subclones, enabling them to outcompete neighboring populations (Salehi et al., 2021).

Administration of anticancer drugs creates selective pressure that impacts subclonal fitness. Drug-sensitive cells are eliminated, but some subclones-often a minority-may acquire resistance through pre-existing mutations or de novo drug-induced mutations in drug-tolerant cells. These resistant subclones can subsequently expand, leading to tumor relapse. For example, research by Xie et al. identified a subgroup of quiescent cells (CSCs) glioblastoma cancer stem that survived antiproliferative chemotherapy, later re-entered the cell cycle, and contributed to tumor regrowth, ultimately causing treatment failure and relapse (Xie et al., 2022). Other studies have suggested combining multiregion sampling to analyze spatial ITH with the monitoring of circulating tumor DNA (ctDNA) through liquid biopsies to track clonal evolution in real time and adjust therapies accordingly (Amirouchene-Angelozzi et al., 2017; Siravegna et al., 2017). A Bayesian evolutionary framework has also been applied to investigate the spatiotemporal dynamics of cancer subclones within individual patients (Alves et al., 2019).

Subclone identification can be performed using various approaches, including genome profiling and single-cell sequencing. Genome profiling remains the primary strategy for studying clonal evolution. Several bioinformatics tools have been developed to infer cancer subclones based on SNV allele frequencies, CNV profiles, and tumor purity measures, including PYCLONE-VI (Gillis and Roth, 2020), PHYLOWGS (Deshwar et al., 2015), FASTCLONE (Xiao et al., 2020), SCICLONE (Miller et al., 2014), and MOBSTER (Caravagna et al., 2020). However, this approach has limitations. It primarily detects mutations present in most or all tumor cells, while stromal contamination can influence mutation frequency estimates. Additionally, many prior inference steps in these tools may introduce errors, which can propagate through subsequent analyses (Turajlic et al., 2019).

The concept of clonetherapy has emerged, aiming to optimize treatment regimens that account for ITH by targeting all subclones, including minor populations with relapse potential (Jiménez-Santos et al., 2022). Several computational tools support this approach, such as OmicsTIDE (Harbig, 2023), which enables interactive exploration of multi-omics data trends; FORALL, (Aswad and Jafari, 2023), an interactive Shiny/R web portal for navigating high-throughput multi-omics data in pediatric acute lymphoblastic leukemia;



the patient, which includes the implications of the findings, their impact on the patient's condition, and recommended next steps. If the initial analysis does not identify a disease-causing variant, the stored WGS data is periodically re-analyzed (inner grey arrow). The continuous process allows for the incorporation of new scientific discoveries, potentially leading to a diagnosis without requiring further hospitalization or additional sampling. Additionally, other clinically relevant insights, such as pharmacogenetic data can be extracted from the WGS data to enhance patient care. This figure is adapted from Bagger, FO. et al. BMC Med Genomics 17, 39 (2024).

MMDRP (Taj and Stein, 2024) which applies multi-modal deep learning for drug response prediction and biomarker discovery; and iCluF (Shakyawar et al., 2024), an unsupervised iterative cluster-fusion method for patient stratification using multi-omics data.

5 Incorporating drug prioritisation tools into the clinical practice

Bioinformatics-driven therapy selection remains in its early stages, with drug prioritization methods still facing significant technical and biological challenges that hinder their routine clinical application. However, considerable progress has been made to integrate these methodologies into medical practice for patient benefit. Cancer care spans multiple stages, from disease prevention and early detection to diagnosis, treatment, and followup. To determine the most effective treatment options, physicians require integrated patient information presented in a clear and interpretable format through clinical decision support systems. These systems must efficiently access electronic medical records containing diverse data types, including genomic information collected at different stages of a patient's journey. See Figure 2.

Next-generation sequencing (NGS) data analysis, including drug prioritization algorithms, will be incorporated into clinical decision support systems, necessitating broad interoperability across data, metadata, research software, and computational infrastructure. This requires standardized nomenclature, well-annotated genomic datasets linked to clinicopathological information, and efficient data-sharing mechanisms. To achieve this, multimodal cancer data must be meaningfully integrated, highlighting the importance of data harmonization and standardization. Several initiatives are actively addressing this challenge. The Findable, Accessible, Interoperable, and Reusable (FAIR) principles facilitate efficient clinical data exchange (Kush et al., 2020). Data harmonization efforts include the NIH Data Commons (https://commonfund.nih.gov/commons) and the Cancer Research Data Commons (CRDC) (https://datacommons.cancer.gov/). Additionally, ICGC-ARGO (https://platform.icgc-argo.org/) is working to collect comprehensive cancer genomic datasets enriched with clinical information, health records, and treatment response data (ICGC, 2022). The Beyond 1 Million Genomes (B1MG) initiative is also advancing efforts in this direction. The integration of multiomics approaches and in silico drug prioritization tools into routine clinical practice requires continued development within healthcare systems. These tools would significantly benefit from extensive, standardized clinical, pathological, and genomic annotations within a federated datasharing model that preserves patient privacy while storing retrospective treatment response data. Such a framework would facilitate benchmarking, training, and validation of novel drug response prediction models and support the identification of new predictive biomarkers based on historical data (Rajpurkar et al., 2022). The Global Alliance for Genomics and Health (GA4GH) provides international policies and standards to ensure responsible access to genomic and health-related data (Rehm et al., 2021). Projects such as the GA4GH Genome Beacons have pioneered bioinformatics frameworks that allow hospitals to query clinicogenomic datasets while maintaining data privacy and ownership (Fiume et al., 2019).

Artificial Intelligence (AI)-driven Clinical Decision Support Systems (CDSS) have been implemented in various real-world medical settings, demonstrating their potential to enhance diagnostic accuracy, streamline workflows, and improve patient outcomes. The notable case studies include Instant Skin Cancer Diagnosis in NHS Hospitals, at Chelsea and Westminster Hospital in London, an AI tool named 'Derm' is being utilized for rapid, autonomous skin cancer assessments. Healthcare professionals use an iPhone equipped with a magnifying lens to capture images of suspicious moles, which the AI app analyzes within seconds. The 'Derm' system boasts a 99.9% accuracy rate in ruling out melanoma and has significantly reduced waiting lists by enabling doctors to focus on more severe cases. Currently adopted by 20 NHS hospitals, this technology has detected approximately 13,000 cancer cases to date. (https://www.chelwest. nhs.uk/about-us/news/chelsea-and-westminster-hospital-leads-theway-with-autonomous-ai-technology-to-speed-up-life-saving-skincancer-checks, 2025) In another study, AI Assisted in Radiology Diagnoses in South Australia, South Australian Medical Imaging (SAMI) has integrated AI to assist in interpreting chest X-rays across multiple hospitals. This AI functions as a "spell checker" for radiologists, highlighting areas of interest and suggesting potential SAMI approximately diagnoses. performs 700,000 radiological examinations annually (https://www.sahealth. sa.gov.au/). Enhanced Cancer Detection Rates in General Practices has been carried out in England, in which the 'C the Signs' AI tool has been deployed in around 1,400 general practices to analyze patient medical records for hidden patterns indicative of cancer risk. Its implementation led to an increase in cancer detection rates from 58.7% to 66.0% with identification of over 50 different types of cancer, ensuring faster and earlier diagnoses (Bakshi et al., 2024). A regional hospital implemented an AI-based CDSS aimed at reducing 30-day all-cause hospital readmission rates. The AI tool combined clinical and non-clinical data to predict patients' risk of readmission and provided recommendations to mitigate this risk (Romero-Brufau et al., 2020). Another study involving the Watson for Oncology (WfO) AI-based CDSS assessed its influence on treatment decisions for complex breast cancer cases (Xu et al., 2020). These case studies illustrate the tangible benefits and effectiveness of AI-driven CDSS in diverse medical environments, highlighting their role in improving diagnostic accuracy, patient management, and overall healthcare delivery.

6 Challenges

The process of drug prioritization in cancer research is hindered by several biological and technical challenges. A major obstacle is the shortage of experts with specialized knowledge in multiomics analysis, bioinformatics, and clinical interpretation. Furthermore, the accessibility and availability of clinical samples remain problematic, exacerbated by the lack of standardized protocols for sample processing, which can result in inconsistencies in data quality and reliability. Scalability is another pressing issue, as translating multiomics findings into clinical applications requires a robust infrastructure capable of managing large-scale data generation and analysis. The absence of standardized, highquality reference datasets for training and validating genomic analysis methods further complicates efforts to ensure accuracy and reproducibility. Additionally, many healthcare institutions face computational limitations, making it difficult to efficiently process and integrate large-scale omics data. Strict data privacy regulations add another layer of complexity, as maintaining patient data security and confidentiality is crucial. AI-driven genomics offers powerful tools for precision oncology, but it also faces several technical and computational challenges. A key concern is model interpretability. Various deep learning models function as "black boxes," making it difficult to understand how specific features influence predictions, which is critical in clinical settings. Additionally, data heterogeneity across omics platforms, patient populations, and sequencing technologies complicates data integration and can lead to inconsistent results. Biases in AI models, often stemming from imbalanced training data or underrepresented subpopulations, can result in skewed predictions that may not generalize well across diverse patient groups. Addressing these challenges is essential for building reliable, equitable, and clinically useful AI tools in genomics (Dias and Torkamani, 2019).

Moreover, the ethical and legal implications of using omics data in clinical settings must be carefully addressed to establish a comprehensive regulatory framework. While implementing AIdriven genomics tools in precision oncology, several ethical, legal, and social issues must be addressed to ensure safe, fair, and responsible use (Farasati, 2023). These issues are the Data Ownership, Informed Consent, Algorithmic Bias, Transparency and Accountability, and Data Privacy and Security. The Patients often do not have direct control over their genomic data once it is collected, especially when stored in centralized or commercial databases, due to Institutional vs. personal ownership. Further lack of standardization around who can access, use, or profit from genomic data can hinder trust and data-sharing (Gerke et al., 2020). Patients may not fully understand how their data will be analyzed, integrated, or reused over time. The one-time consent forms are not sufficient for evolving AI applications. The

patients may not have consented to secondary use for unrelated research or algorithm training. AI models trained on predominantly western, caucasian, or male genomic datasets which may yield inaccurate predictions for diverse populations. Sometimes there are biased outcome predictions (Cross et al., 2024). Algorithms may perform well in controlled settings but fail in real-world, and heterogeneous clinical populations. Many models, particularly deep learning networks lack explainability, making it difficult for clinicians and patients to trust decisions. The guidelines for validation, certification, or clinical approval of AI-driven genomics tools are still emerging. The genomic databases are high-value targets for breaches, with implications for patient confidentiality and potential misuse (Bonomi et al., 2020). Even de-identified genomic data can be re-identified due to its unique nature. And most importantly, the accountability for errors remains unclear (Martinez-Martin and Magnus, 2019).

Biologically, a key challenge lies in the incomplete understanding of inter and intratumor heterogeneity and the somatic evolutionary processes driving cancer progression. The relationship between clonal expansion and cancer initiation remains unclear, as do the intricate topological interactions between tumors and the tumor microenvironment (TME), including cell-cell communication. Another critical issue is the progressive exhaustion of antitumor immunity, which limits therapeutic efficacy. Furthermore, the mechanisms underlying the emergence and expansion of drug-resistant subclones have not been fully elucidated. There is also a lack of comprehensive characterization of genetic and epigenetic alterations-such as structural variations, and transcriptional driver mutations-and their impact on drug response. The interplay between aging, cellular senescence, and drug efficacy remains poorly understood, further complicating treatment approaches. Additionally, information on the influence of germline variants on adverse drug reactions (ADRs) in many anticancer therapies is still insufficient.

On the technical side, the use of formalin-fixed, paraffinembedded (FFPE) sample preparation can cause DNA fragmentation and degradation, making it difficult to distinguish true variants from artifacts in genomic analyses. Another challenge is the trade-off between sequencing scope and read depth-whole genome sequencing (WGS) provides a broad view but at lower coverage, whereas targeted sequencing offers deeper reads but a narrower scope. Similarly, in single-cell sequencing, increasing the number of cells analyzed reduces the read depth per cell. Multialignment reads pose difficulties due to repetitive genomic regions, complicating accurate variant calling. While short-read sequencing is widely used, it struggles to detect large structural variations, whereas long-read sequencing, despite its ability to identify these variations, has a higher error rate. Additionally, the lack of standardized guidelines for analyzing spatial data in single-cell technologies presents a significant challenge. Finally, predicting toxic interactions and synergistic effects in combination therapies remains a major hurdle in optimizing cancer treatment strategies. These computational, biological and technical limitations collectively hinder the accurate prioritization of drugs for cancer treatment, highlighting the need for continued advancements in genomics, immunology, and computational biology. Addressing these challenges is essential for the successful integration of multiomics approaches into personalized medicine and routine patient care. Numerous studies have shown that AI can surpass human capabilities in interpreting the vast amounts of data associated with complex diseases like cancer. However, AI should be seen as a tool to enhance human intelligence rather than replace it. Any analysis conducted by AI must be reviewed and validated by domain experts. Additionally, machine learning (ML) and deep learning (DL) models require oversight from specialists in bioinformatics and programming to ensure their reliability and accuracy.

One of the primary challenges in applying AI and DL to cancer diagnosis, prognosis, and treatment is the "black box" problem. This refers to the lack of transparency regarding how AI systems process information and arrives at conclusions. When AI operates autonomously with minimal human oversight, it may become unclear how it selects features or makes decisions, potentially leading to skepticism about its predictions. This uncertainty could force clinicians and researchers to accept AI-generated results on "blind faith" (Sorell et al., 2021). In response, researchers have been working to develop AI systems that provide explainable insights for physicians and clinicians. To ensure data availability and sharing while protecting patient privacy in multiomics cancer data, several technical and strategic approaches are employed such as Federated Learning (FL) (Saha et al., 2024), Differential Privacy (DP), Secure Multi-Party Computation (SMPC) (Zhou et al., 2024), Homomorphic Encryption (Ogburn et al., 2013), Trusted Research Environments (TREs) (Kavianpour et al., 2022), Data Deidentification and Anonymization (Chevrier et al., 2019), Synthetic Data Generation, Standardized Data Use Agreements (DUAs), Dynamic and Informed Consent Models (Wendland et al., 2022) and Adherence to FAIR and CARE Principles (Carroll et al., 2021). These approaches collectively support secure, ethical, and effective sharing of multiomics cancer data, facilitating advances in research and personalized medicine. For example, Kwong et al. (2022) designed an AI model using ML to predict whether prostate cancer patients would benefit from nerve-sparing radical prostatectomy by assessing the likelihood of tumor extension beyond the prostate. The AI's decision-making process was made interpretable using a publicly available web application, Shapley Additive exPlanations (SHAP) (Kwong et al., 2022). Deep learning also requires vast amounts of data to develop robust algorithms applicable to new datasets. Consequently, cancer research studies must collect multiple samples to serve as training data (Hussain et al., 2017). Furthermore, the use of AI and big data raises ethical concerns, particularly regarding patient data privacy. In some cases, patient data is used for purposes beyond direct medical care, and this may occur without the patient's explicit consent (Rigby, 2019). A significant challenge in multiomics integration is missing data, as not all biomolecules are measured across all samples. This can be due to financial constraints, instrument sensitivity, or other experimental limitations, leading to incomplete datasets for certain omics technologies. While recent advancements in AI and statistical learning have greatly improved multiomics data analysis, many techniques still assume the presence of fully observed data. However, new approaches are being developed

to address the issue of missing data, allowing for more effective utilization of incomplete datasets and paving the way for improved precision oncology in the future. Handling missing values in multi-omics data is a critical step for improving the accuracy and robustness of downstream analysis, as such datasets often suffer from incomplete measurements across different omics layers due to technical limitations or sample variability. Several specific methods and tools have been developed to address this challenge effectively such as Statistical Imputation Techniques, Matrix Factorization Methods, Bayesian and Probabilistic Models, Machine Learning-Based Imputation and Deep Learning Methods (Huang et al., 2023). There are various Multi-Omics-Specific Tools such as MOFA (Multi-Omics Factor Analysis) (Hama et al., 2023) and MAGIC (Markov Affinity-based Graph Imputation of Cells (van Dijk et al., 2018) and impute omics designed for imputing missing values in multiomics datasets using joint matrix completion and feature correlations across omics types. Methods like multi-view learning and tensor factorization integrate data from multiple omics layers simultaneously, enabling imputation that leverages crossomics relationships. Thus selecting an appropriate imputation method depends on the extent and pattern of missingness, the data type (e.g., continuous vs. categorical), and the structure of the dataset. More advanced, multiomics-aware approaches especially those based on probabilistic modeling and deep learning are increasingly favored for their ability to preserve biological signals and improve the accuracy of downstream analyses such as clustering, classification, and biomarker discovery (Jadhav et al., 2019).

7 Future prospects

The advancement of bioinformatics tools and platforms will be essential for the future of multiomics research. These tools must enable seamless integration and analysis of diverse omics datasets, including genomics, transcriptomics, proteomics, and metabolomics. Enhancements in computational power, data storage, and cloud computing will further support large-scale multi-omics data processing. Additionally, blockchain technology presents a promising solution for data management, ensuring integrity, security, and patient privacy. By providing a transparent and tamper-proof system for storing and sharing multi-omics data, blockchain can help build trust and encourage collaboration within the healthcare community. The future of AIdriven bioinformatics in cancer treatment will likely involve increased interdisciplinary collaboration, drawing from bioinformatics, systems biology, computational biology, and clinical research. These collaborative efforts will be instrumental in addressing complex biological questions and developing comprehensive disease models. Strong partnerships between data scientists and clinicians will also help bridge the gap between intricate multi-omics data analysis and its practical applications in healthcare, ensuring that insights are both clinically relevant and actionable.

Despite its potential, multi-omics research still faces several challenges that must be overcome to fully realize its benefits. One

key issue is the standardization of data collection and analysis across different omics layers and research institutions. Establishing standardized protocols and quality control measures will be critical to ensuring the reliability and reproducibility of multiomics studies. Additionally, the complexity and high dimensionality of multi-omics data necessitate the development of advanced statistical methods and sophisticated algorithms for accurate interpretation and meaningful conclusions.

Multiomics approaches are expected to play a pivotal role in the advancement of personalized medicine. By integrating genetic, transcriptomic, proteomic, and metabolomic data, healthcare providers can tailor treatments to each patient's unique characteristics, improving outcomes and minimizing adverse effects. Whole genome sequencing (WGS) and whole exome sequencing (WES) also hold great promise for enhancing our understanding of complex diseases such as cancer, cardiovascular conditions, and neurodegenerative disorders. By uncovering the molecular mechanisms driving these diseases, multiomics studies can help identify novel biomarkers for early diagnosis and potential drug targets. Over the next decade, significant advancements in multiomics integration with other emerging technologies could lead to the development of personalized virtual models of patients. These models would allow for in silico testing of treatments and interventions before their application in real-life clinical settings. Additionally, improvements in multiomics data visualization tools will enhance researchers' and clinicians' ability to interpret complex datasets, facilitating the translation of AI-driven genomic insights into clinical practice.

The future of AI in genomics is promising, with numerous emerging trends, technological breakthroughs, and interdisciplinary approaches driving innovation in precision medicine. Addressing current challenges and exploring new applications will be essential for unlocking the full potential of this research. Continued investment, along with collaborative efforts across various fields, will ensure that AI-powered genomics remains at the forefront of scientific and medical progress.

8 Conclusion

The integration of intelligent computing in genomics for cancer research represents a crucial step toward unlocking the full potential of precision medicine. With approximately 14 billion laboratory tests conducted annually, clinical laboratories contribute to nearly 70% of medical decisions, highlighting the necessity for accurate and comprehensive data. The incorporation of AI in developing health indices, predicting health trajectories, and combining advanced statistical modeling with digital twins (DTs) showcases the transformative potential of these technologies in revolutionizing healthcare delivery. However, overcoming key challenges-such as generating actionable and concise metrics from omics data and establishing meaningful intra-level comparators-is essential for progress. As these advancements reshape the medical landscape, ethical considerations remain paramount to ensuring that technology complements, rather than replaces, the human touch in healthcare. The future of medicine depends on a deep understanding of patient journeys and care pathways, ensuring that AI-driven innovations align seamlessly with the complexities of individual wellbeing while enhancing patient-centered care.

Author contributions

RS: Conceptualization, Formal Analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Visualization, Writing – original draft, Writing – review and editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. DST/WIDUSHI-B/CS/2024/35.

Acknowledgments

RS is thankful to her CoInvestigator Prof. Gopalan Rajaraman, Indian Institute of Technology, Mumbai, India for the support.

References

Adams, R., Steckel, M., Nicke, B., and Pohlenz, H.-D. (2016). RNAi as a tool for target discovery in early pharmaceutical research. *Pharm.-Int. J. Pharm. Sci.* 71, 35–42.

Adzhubei, I., Jordan, D. M., and Sunyaev, S. R. (2013). Predicting functional effect of human missense mutations using PolyPhen-2. *Curr. Protoc. Hum. Genet.* 7. doi:10. 1002/0471142905

Ahmed, Z. (2020). Practicing precision medicine with intelligently integrative clinical and multi-omics data analysis. *Hum. Genomics* 14, 35. doi:10.1186/s40246-020-00287-z

Alexandrov, L. B., Kim, J., Haradhvala, N. J., Huang, M. N., Tian Ng, A. W., Wu, Y., et al. (2020). The repertoire of mutational signatures in human cancer. *Nature* 578, 94–101. doi:10.1038/s41586-020-1943-3

Almulihi, A., Saleh, H., Hussien, A. M., Mostafa, S., El-Sappagh, S., Alnowaiser, K., et al. (2022). Ensemble learning based on hybrid deep learning model for heart disease early prediction. *Diagn. (Basel)* 12 (12), 3215. doi:10.3390/diagnostics12123215

Alves, J. M., Prado-Lopez, S., Cameselle-Teijeiro, J. M., and Posada, D. (2019). Rapid evolution and biogeographic spread in a colorectal cancer. *Nat. Commun.* 10, 5139. doi:10.1038/s41467-019-12926-8

Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., et al. (2021). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *J. Big Data* 8, 53. doi:10.1186/s40537-021-00444-8

Amirouchene-Angelozzi, N., Swanton, C., and Bardelli, A. (2017). Tumor evolution as a therapeutic target. *Cancer Discov.* 7, 805–817. doi:10.1158/2159-8290.CD-17-0343

Arjmand, B., Hamidpour, S. K., Tayanloo-Beik, A., Goodarzi, P., Aghayan, H. R., Adibi, H., et al. (2022). Machine learning: a new prospect in multi-omics data analysis of cancer. *Front. Genet.* 13, 824451. doi:10.3389/fgene.2022.824451

Aswad, L., and Jafari, R. (2023). FORALL: an interactive shiny/R web portal to navigate multi-omics high-throughput data of pediatric acute lymphoblastic leukemia. *Bioinform Adv.* 3 (1), vbad143. doi:10.1093/bioadv/vbad143

Baca, S. C., Prandi, D., Lawrence, M. S., Mosquera, J. M., Romanel, A., Drier, Y., et al. (2013). Punctuated evolution of prostate cancer genomes. *Cell* 153, 666–677. doi:10. 1016/j.cell.2013.03.021

Baez-Ortega, A., and Gori, K. (2019). Computational approaches for discovery of mutational signatures in cancer. *Brief. Bioinform* 20, 77–88. doi:10.1093/bib/bbx082

Bakshi, B., Dadhania, S., Holloway, P., Corbett, C., and Payling, M. (2024). Using an artificial intelligence platform to enhance cancer detection rates in primary care. *J. Clin. Oncol.*, 42, (16), 1560. doi:10.1200/JCO.2024.42.16_suppl.1560

Bao, X., Li, Q., Chen, D., Dai, X., Liu, C., Tian, W., et al. (2024). A multiomics analysis-assisted deep learning model identifies a macrophage-oriented module as a potential therapeutic target in colorectal cancer. *Cell Rep. Med.* 5 (2), 101399. doi:10.1016/j.xcrm.2024.101399

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author declares that Generative AI was used in the creation of this manuscript. Generative AI was used for the creation of this manuscript. Author used ChatGPT 4.0 to grammatically edit this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A. A., Kim, S., et al. (2012). The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 483 (7391), 603–607. doi:10.1038/nature11003

Becht, E., Giraldo, N. A., Lacroix, L., Buttard, B., Elarouci, N., Petitprez, F., et al. (2016). Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol.* 17, 218. doi:10.1186/s13059-016-1070-5

Bergstrom, E. N., Barnes, M., Martincorena, I., and Alexandrov, L. B. (2020). Generating realistic null hypothesis of cancer mutational landscapes using SigProfilerSimulator. *BMC Bioinforma*. 21, 438. doi:10.1186/s12859-020-03772-3

Bergstrom, E. N., Huang, M. N., Mahto, U., Barnes, M., Stratton, M. R., Rozen, S. G., et al. (2019). SigProfilerMatrixGenerator: a tool for visualizing and exploring patterns of small mutational events. *BMC Genomics* 20, 685. doi:10.1186/s12864-019-6041-2

Bersanelli, M., Mosca, E., Remondini, D., Giampieri, E., Sala, C., Castellani, G., et al. (2016). Methods for the integration of multi-omics data: mathematical aspects. *BMC Bioinforma*. 17, S15. doi:10.1186/s12859-015-0857-9

Bieg-Bourne, C. C., Millis, S. Z., Piccioni, D. E., Fanta, P. T., Goldberg, M. E., Chmielecki, J., et al. (2017). Next-generation sequencing in the clinical setting clarifies patient characteristics and potential actionability. *Cancer Res.* 77, 6313–6320. doi:10. 1158/0008-5472.CAN-17-1569

Black, J. R. M., and McGranahan, N. (2021). Genetic and non-genetic clonal diversity in cancer evolution. *Nat. Rev. Cancer* 21, 379–392. doi:10.1038/s41568-021-00336-2

Bonomi, L., Huang, Y., and Ohno-Machado, L. (2020). Privacy challenges and research opportunities for genomic data sharing. *Nat. Genet.* 52 (7), 646–654. doi:10.1038/s41588-020-0651-0

Borchert, F., Mock, A., Tomczak, A., Hugel, J., \in Alkarkoukly, S., Knurr, A., et al. (2021). Knowledge bases and software support for variant interpretation in precision oncology. *Brief. Bioinform* 22, bbab134. doi:10.1093/bib/bbab134

Broad Institute (2018). Picard toolkit. Available online at: https://github.com/ broadinstitute/picard.

Cameron, D. L., Baber, J., Shale, C., Valle-Inclan, J. E., Besselink, N., van Hoeck, A., et al. (2021). GRIDSS2: comprehensive characterisation of somatic structural variation using single breakend variants and structural variant phasing. *Genome Biol.* 22, 202. doi:10.1186/s13059-021-02423-x

Cao, K.-A.Lê, González, I., and Déjean, S. (2009). integrOmics: an R package to unravel relationships between two omics datasets. *Bioinformatics* 25 (Issue 21), 2855–2856. doi:10.1093/bioinformatics/btp515

Cao, X., Huang, Y. A., You, Z. H., Shang, X., Hu, L., Hu, P. W., et al. (2024). scPriorGraph: constructing biosemantic cell-cell graphs with prior gene set selection for cell type identification from scRNA-seq data. *Genome Biol.* 25 (1), 207. doi:10.1186/s13059-024-03357-w

Caravagna, G., Sanguinetti, G., Graham, T. A., and Sottoriva, A. (2020). The MOBSTER R package for tumour subclonal deconvolution from bulk DNA wholegenome sequencing data. *BMC Bioinforma*. 21, 531. doi:10.1186/s12859-020-03863-1

Carroll, S. R., Herczog, E., Hudson, M., Russell, K., and Stall, S. (2021). Operationalizing the CARE and FAIR principles for indigenous data futures. *Sci. Data* 8, 108. doi:10.1038/s41597-021-00892-0

Carter, H., Douville, C., Stenson, P. D., Cooper, D. N., and Karchin, R. (2013). Identifying Mendelian disease genes with the variant effect scoring tool. *BMC Genomics* 14 (Suppl. 3), S3. doi:10.1186/1471-2164-14-S3-S3

Cerami, E., Gao, J., Dogrusoz, U., Gross, B. E., Sumer, S. O., Aksoy, B. A., et al. (2012). The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* 2 (5), 401–404. doi:10.1158/2159-8290.CD-12-0095

Chakravarty, D., Gao, J., Phillips, S. M., Kundra, R., Zhang, H., Wang, J., et al. (2017). OncoKB: a precision oncology knowledge base. *JCO Precis. Oncol.* 2017, 1–16. doi:10. 1200/PO.17.00011

Chan, Y.-T., Lu, Y., Wu, J., Zhang, C., Tan, H.-Y., Bian, Z.-X., et al. (2022). CRISPR-Cas9 library screening approach for anti-cancer drug discovery: overview and perspectives. *Theranostics* 12, 3329–3344. doi:10.7150/thno.71144

Chari, R., Coe, B. P., Vucic, E. A., Lockwood, W. W., and Lam, W. L. (2010). An integrative multi-dimensional genetic and epigenetic strategy to identify aberrant genes and pathways in cancer. *BMC Syst. Biol.* 4, 67. doi:10.1186/1752-0509-4-67

Chen, C., Wang, J., Pan, D., Wang, X., Xu, Y., Yan, J., et al. (2023). Applications of multi-omics analysis in human diseases. *MedComm* 4 (4), e315. doi:10.1002/mco2.315

Chen, Y., Moustaki, I., and Zhang, H. (2020). A note on likelihood Ratio tests for models with latent variables. *Psychometrika* 85 (4), 996–1012. doi:10.1007/s11336-020-09735-0

Chevrier, R., Foufi, V., Gaudet-Blavignac, C., Robert, A., and Lovis, C. (2019). Use and understanding of anonymization and de-identification in the biomedical literature: scoping review. J. Med. Internet Res. 21 (5), e13484. doi:10.2196/13484

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., et al. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* 6, 80–92. doi:10.4161/fly.19695

Cokelaer, T., Chen, E., Iorio, F., Menden, M. P., Lightfoot, H., Saez-Rodriguez, J., et al. (2018). GDSCTools for mining pharmacogenomic interactions in cancer. *Bioinformatics* 34 (7), 1226–1228. doi:10.1093/bioinformatics/btx744

Connor, A. A., Denroche, R. E., Jang, G. H., Timms, L., Kalimuthu, S. N., Selander, I., et al. (2017). Association of distinct mutational signatures with correlates of increased immune activity in pancreatic ductal adenocarcinoma. *JAMA Oncol.* 3, 774–783. doi:10. 1001/jamaoncol.2016.3916

Correa-Aguila, R., Alonso-Pupo, N., and Hernández-Rodríguez, E. W. (2022). Multiomics data integration approaches for precision oncology. *Mol. Omics* 18 (6), 469–479. doi:10.1039/d1mo00411e

Cotto, K. C., Wagner, A. H., Feng, Y.-Y., Kiwala, S., Coffman, A. C., Spies, G., et al. (2018). DGIdb 3.0: a redesign and expansion of the drug-gene interaction database. *Nucleic Acids Res.* 46, D1068–D1073. doi:10.1093/nar/gkx1143

Cross, J. L., Choma, M. A., and Onofrey, J. A. (2024). Bias in medical AI: implications for clinical decision-making. *PLOS Digit. Health* 3 (11), e0000651. doi:10.1371/journal. pdig.0000651

Dahal, S., Yurkovich, J. T., Xu, H., Palsson, B. O., and Yang, L. (2020). Synthesizing systems biology knowledge from omics using genome-scale models. *Proteomics* 20, e1900282. doi:10.1002/pmic.201900282

Davies, H., Glodzik, D., Morganella, S., Yates, L. R., Staaf, J., Zou, X., et al. (2017). HRDetect is a predictor of BRCA1 and BRCA2 deficiency based on mutational signatures. *Nat. Med.* 23, 517-525. doi:10.1038/nm.4292

Degasperi, A., Amarante, T. D., Czarnecki, J., Shooter, S., Zou, X., Glodzik, D., et al. (2020). A practical framework and online tool for mutational signature analyses show intertissue variation and driver dependencies. *Nat. Cancer* 1, 249–263. doi:10.1038/ s43018-020-0027-5

Degasperi, A., Zou, X., Dias Amarante, T., MartinezMartinez, A., Koh, G. C. C., Dias, J. M. L., et al. (2022). Substitution mutational signatures in whole-genome-sequenced cancers in the UK population. *Science* 376. doi:10.1126/science.abl9283

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* 43, 491–498. doi:10.1038/ng.806

Deshwar, A. G., Vembu, S., Yung, C. K., Jang, G. H., Stein, L., and Morris, Q. (2015). PhyloWGS: reconstructing subclonal composition and evolution from whole-genome sequencing of tumors. *Genome Biol.* 16, 35. doi:10.1186/s13059-015-0602-8 Dias, R., and Torkamani, A. (2019). Artificial intelligence in clinical and genomic diagnostics. *Genome Med.* 11, 70. doi:10.1186/s13073-019-0689-8

Divate, M., Tyagi, A., Richard, D. J., Prasad, P. A., Gowda, H., and Nagaraj, S. H. (2022). Deep learning-based pan-cancer classification model reveals tissue-of-origin specific gene expression signatures. *Cancers (Basel)* 14 (5), 1185. doi:10.3390/ cancers14051185

Dong, C., Guo, Y., Yang, H., He, Z., Liu, X., and Wang, K. (2016). iCAGES: integrated CAncer GEnome Score for comprehensively prioritizing driver genes in personal cancer genomes. *Genome Med.* 8 (1), 135. doi:10.1186/s13073-016-0390-0

Dong, C., Wei, P., Jian, X., Gibbs, R., Boerwinkle, E., Wang, K., et al. (2015). Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Hum. Mol. Genet.* 24, 2125–2137. doi:10.1093/hmg/ddu733

Dong, Z., Xie, W., Chen, H., Xu, J., Wang, H., Li, Y., et al. (2017). Copy-number variants detection by low-pass wholegenome sequencing. *Curr. Protoc. Hum. Genet.* 94 (8.17), 1–16. doi:10.1002/cphg.43

do Valle, I. F., Menichetti, G., Simonetti, G., Bruno, S., Zironi, I., Durso, D. F., et al. (2018). Network integration of multi-tumour omics data suggests novel targeting strategies. *Nat. Commun.* 9, 4514. doi:10.1038/s41467-018-06992-7

Dugourd, A., Kuppe, C., Sciacovelli, M., Gjerga, E., Gabor, A., Emdal, K. B., et al. (2021). Causal integration of multi-omics data with prior knowledge to generate mechanistic hypotheses. *Mol. Syst. Biol.* 17, e9730. doi:10.15252/msb. 20209730

Escaramis, G., Docampo, E., and Rabionet, R. (2015). A decade of structural variants: description, history and methods to detect structural variation. *Brief. Funct. Genomics* 14, 305–314. doi:10.1093/bfgp/elv014

Faraoni, I., and Graziani, G. (2018). Role of BRCA mutations in cancer treatment with poly(ADP-ribose) polymerase (PARP) inhibitors. *Cancers (Basel)* 10 (12), 487. doi:10. 3390/cancers10120487

Farasati, F. B. (2023). Artificial intelligence ethics in precision oncology: balancing advancements in technology with patient privacy and autonomy. *Explor Target Antitumor Ther.* 4 (4), 685–689. doi:10.37349/etat.2023.00160

Fiume, M., Cupak, M., Keenan, S., Rambla, J., de la Torre, S., Dyke, S. O. M., et al. (2019). Federated discovery and sharing of genomic data using Beacons. *Nat. Biotechnol.* 37, 220–224. doi:10.1038/s41587-019-0046-x

Galaxy (2024). The Galaxy platform for accessible, reproducible, and collaborative data analyses: 2024 update. *Nucleic Acids Res.* 52 (W1), W83–W94. doi:10.1093/nar/gkae410

Gerke, S., Minssen, T., and Cohen, G. (2020). Ethical and legal challenges of artificial intelligence-driven healthcare. *Artif. Intell. Healthc.*, 295–336. doi:10.1016/B978-0-12-818438-7.00012-5

Gillette, M. A., Satpathy, S., Cao, S., Dhanasekaran, S. M., Carr, S. A., Krug, K., et al. (2020). Proteogenomic characterization reveals therapeutic vulnerabilities in lung adenocarcinoma. *Cell* 182 (1), 200–225.e35. doi:10.1016/j.cell.2020.06.013

Gillis, S., and Roth, A. (2020). PyClone-VI: scalable inference of clonal population structures using whole genome data. *BMC Bioinforma*. 21, 571. doi:10.1186/s12859-020-03919-2

Greenman, C., Stephens, P., Smith, R., Dalgliesh, G. L., Hunter, C., Bignell, G., et al. (2007). Patterns of somatic mutation in human cancer genomes. *Nature* 446, 153–158. doi:10.1038/nature05610

Griffith, M., Spies, N. C., Krysiak, K., McMichael, J. F., Coffman, A. C., Danos, A. M., et al. (2017). CIViC is a community knowledgebase for expert crowdsourcing the clinical interpretation of variants in cancer. *Nat. Genet.* 49, 170–174. doi:10.1038/ng. 3774

Hachem, S., Yehya, A., El Masri, J., Mavingire, N., Johnson, J. R., Dwead, A. M., et al. (2024). Contemporary update on clinical and experimental prostate cancer biomarkers: a multi-omics-focused approach to detection and risk stratification. *Biol. (Basel).* 13 (10), 762. doi:10.3390/biology13100762

Hackl, H., Charoentong, P., Finotello, F., and Trajanoski, Z. (2016). Computational genomics tools for dissecting tumourimmune cell interactions. *Nat. Rev. Genet.* 17, 441–458. doi:10.1038/nrg.2016.67

Haley, B., and Roudnicky, F. (2020). Functional genomics for cancer drug target discovery. *Cancer Cell* 38, 31-43. doi:10.1016/j.ccell.2020.04.006

Hama, H., Seymen, N., Gerlevik, S., Kaya, D. E., Napolitani, G., Ogawa, S., et al. (2023). Multi-omics factor analysis (MOFA) identifies transposable element expression as a risk factor and inflammaging as a protective factor in myelodysplastic syndromes. *Blood* 142 (Suppl. 1), 6450. doi:10.1182/blood-2023-177683

Hamet, P., and Tremblay, J. (2017). Artificial intelligence in medicine. *Metabolism* 69s, S36-s40. doi:10.1016/j.metabol.2017.01.011

Han, Y., Liu, D., and Li, L. (2020). PD-1/PD-L1 pathway: current researches in cancer. *Am. J. Cancer Res.* 10 (3), 727–742. PMID: 32266087.

Haradhvala, N. J., Kim, J., Maruvka, Y. E., Polak, P., Rosebrock, D., Livitz, D., et al. (2018). Distinct mutational signatures characterize concurrent loss of polymerase proofreading and mismatch repair. *Nat. Commun.* 9, 1746. doi:10.1038/s41467-018-04002-4

Harbig, T. A., Fratte, J., Krone, M., and Nieselt, K. (2023). OmicsTIDE: interactive exploration of trends in multi-omics data. *Bioinforma. Adv.* 3 (1), vbac093. doi:10.1093/bioadv/vbac093

Hasin, Y., Seldin, M., and Lusis, A. (2017). Multi-omics approaches to disease. Genome Biol. 18, 83. doi:10.1186/s13059-017-1215-1

Hellmann, M. D., Paz-Ares, L., and Costa, D. B. (2018). Lung cancer with a high tumor mutational burden. N. Engl. J. Med. 379, 1093–1094. doi:10.1056/NEJMc1808566

Ho, S. S., Urban, A. E., and Mills, R. E. (2020). Structural variation in the sequencing era. *Nat. Rev. Genet.* 21, 171–189. doi:10.1038/s41576-019-0180-9

Huang, L., Brunell, D., Stephan, C., Mancuso, J., Yu, X., He, B., et al. (2019). Driver network as a biomarker: systematic integration and network modeling of multi-omics data to derive driver signaling pathways for drug combination prediction. *Bioinformatics* 35, 3709–3717. doi:10.1093/bioinformatics/btz109

Huang, L., Song, M., Shen, H., Hong, H., Gong, P., Deng, H. W., et al. (2023). Deep learning methods for omics data imputation. *Biol. (Basel).* 12 (10), 1313. doi:10.3390/biology12101313

Huang, S., Chaudhary, K., and Garmire, L. X. (2017). More is better: recent progress in multi-omics data integration methods. *Front. Genet.* 8, 84. doi:10.3389/fgene.2017. 00084

Huang, T., Chen, X., Zhang, H., Liang, Y., Li, L., Wei, H., et al. (2021). Prognostic role of tumor mutational burden in cancer patients treated with immune checkpoint inhibitors: a systematic review and meta-analysis. *Front. Oncol.* 11, 706652. doi:10. 3389/fonc.2021.706652

Huang, Z. A., Hu, P., Hu, L., You, Z. H., Tan, K. C., and Huang, Y. A. (2024). Toward multilabel classification for multiple disease prediction using gut microbiota profiles. *IEEE Trans. Neural Netw. Learn Syst.* 12, 1–14. doi:10.1109/TNNLS.2024. 3453967

Hudson, C. T. J., Anderson, W., Aretz, A., Artez, A., Barker, A. D., Bell, C., et al. (2010). International network of cancer genome projects. *Nature* 464, 993–998. doi:10. 1038/nature08987

Hundal, J., Kiwala, S., McMichael, J., Miller, C. A., Xia, H., Wollam, A. T., et al. (2020). pVACtools: a computational toolkit to identify and visualize cancer neoantigens. *Cancer Immunol. Res.* 8, 409–420. doi:10.1158/2326-6066.CIR-19-0401

Hussain, Z., Gimenez, F., Yi, D., and Rubin, D. (2017). Differential data Augmentation techniques for medical imaging classification tasks. *AMIA Annu. Symp. Proc.* 2017, 979–984.

ICGC (2022). ARGO data platform. Available online at: https://platform.icgc-argo. org/215ELIXIREurope.Beyond1milliongenomes.

Islam, S. M. A., Diaz-Gay, M., Wu, Y., Barnes, M., Vangara, R., Bergstrom, E. N., et al. (2020). Uncovering novel mutational signatures by *de novo* extraction with SigProfilerExtractor. *bioRxiv*. doi:10.1101/2020.12.13.422570

Jaccard, E., Cornuz, J., Waeber, G., and Guessous, I. (2018). Evidence-based precision medicine is needed to move toward general internal precision medicine. *J. Gen. Intern Med.* 33, 11–12. doi:10.1007/s11606-017-4149-0

Jadhav, A., Pramod, D., and Ramanathan, K. (2019). Comparison of performance of data imputation methods for numeric dataset. *Appl. Artif. Intell.* 33 (10), 913–933. doi:10.1080/08839514.2019.1637138

Jardim, D. L., Goodman, A., de Melo Gagliato, D., and Kurzrock, R. (2021). The challenges of tumor mutational burden as an immunotherapy biomarker. *Cancer Cell* 39 (2), 154–173. doi:10.1016/j.ccell.2020.10.001

Jiang, J., Yuan, J., Hu, Z., Zhang, Y., Zhang, T., Xu, M., et al. (2022). Systematic illumination of druggable genes in cancer genomes. *Cell Rep.* 38, 110400. doi:10.1016/j. celrep.2022.110400

Jiang, Z., and Li, P. (2024). DeepDR: a deep learning library for drug response prediction. *Bioinformatics* 40 (12), btae688. doi:10.1093/bioinformatics/btae688

Jiménez-Santos, M. J., García-Martín, S., Fustero-Torre, C., Di Domenico, T., Gómez-López, G., and Al-Shahrour, F. (2022). Bioinformatics roadmap for therapy selection in cancer genomics. *Mol. Oncol.* 16 (21), 3881–3908. doi:10.1002/1878-0261.13286

Joos, S., Nettelbeck, D. M., Reil-Held, A., Engelmann, K., Moosmann, A., Eggert, A., et al. (2019). German Cancer Consortium (DKTK) - a national consortium for translational cancer research. *Mol. Oncol.* 13 (3), 535–542. doi:10.1002/1878-0261. 12430

Kalari, K. R., Sinnwell, J. P., Thompson, K. J., Tang, X., Carlson, E. E., Yu, J., et al. (2018). PANOPLY: omics-guided drug prioritization method tailored to an individual patient. *JCO Clin. Cancer Inf.* 2, 1–11. doi:10.1200/CCI.18.00012

Kang, M., Ko, E., and Mersha, T. B. (2022). A roadmap for multiomics data integration using deep learning. *Brief. Bioinform* 23, bbab454. doi:10.1093/bib/bbab454

Kasar, S., Kim, J., Improgo, R., Tiao, G., Polak, P., Haradhvala, N., et al. (2015). Whole-genome sequencing reveals activation-induced cytidine deaminase signatures during indolent chronic lymphocytic leukaemia evolution. *Nat. Commun.* 6, 8866. doi:10.1038/ncomms9866

Kavianpour, S., Sutherland, J., Mansouri-Benssassi, E., Coull, N., and Jefferson, E. (2022). Next-generation capabilities in trusted research environments: interview study. *J. Med. Internet Res.* 24 (9), e33720. doi:10.2196/33720 Keskin, D. B., Anandappa, A. J., Sun, J., Tirosh, I., Mathewson, N. D., Li, S., et al. (2019). Neoantigen vaccine generates intratumoral T cell responses in phase Ib glioblastoma trial. *Nature* 565, 234–239. doi:10.1038/s41586-018-0792-9

Khalil, H., and Huang, C. (2020). Adverse drug reactions in primary care: a scoping review. *BMC Health Serv. Res.* 20, 5. doi:10.1186/s12913-019-4651-7

Kim, J., Mouw, K. W., Polak, P., Braunstein, L. Z., Kamburov, A., Kwiatkowski, D. J., et al. (2016). Somatic ERCC2 mutations are associated with a distinct genomic signature in urothelial tumors. *Nat. Genet.* 48, 600–606. doi:10.1038/ng.3557

Kircher, M., Witten, D. M., Jain, P., O'Roak, B. J., Cooper, G. M., and Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* 46, 310–315. doi:10.1038/ng.2892

Koboldt, D. C. (2020). Best practices for variant calling in clinical sequencing. *Genome Med.* 12, 91. doi:10.1186/s13073-020-00791-w

Koboldt, D. C., Zhang, Q., Larson, D. E., Shen, D., McLellan, M. D., Lin, L., et al. (2012). VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* 22, 568–576. doi:10.1101/gr.129684.111

Koleti, A., Terryn, R., Stathias, V., Chung, C., Cooper, D. J., Turner, J. P., et al. (2018). Data Portal for the Library of Integrated Network-based Cellular Signatures (LINCS) program: integrated access to diverse large-scale cellular perturbation response data. *Nucleic Acids Res.* 46 (D1), D558–D566. doi:10.1093/nar/gkx1063

Krämer, A., Green, J., Pollard, J., Jr., and Tugendreich, S. (2014). Causal analysis approaches in ingenuity pathway analysis. *Bioinformatics* 30 (4), 523–530. doi:10.1093/ bioinformatics/btt703

Krassowski, M., Das, V., Sahu, S. K., and Misra, B. B. (2020). State of the field in multiomics research: from computational needs to data mining and sharing. *Front. Genet.* 11, 610798. doi:10.3389/fgene.2020.610798

Kuo, T. C., Tian, T. F., and Tseng, Y. J. (2013). 3Omics: a web-based systems biology tool for analysis, integration and visualization of human transcriptomic, proteomic and metabolomic data. *BMC Syst. Biol.* 7, 64. doi:10.1186/1752-0509-7-64

Kush, R. D., Warzel, D., Kush, M. A., Sherman, A., Navarro, E. A., Fitzmartin, R., et al. (2020). FAIR data sharing: the roles of common data elements and harmonization. *J. Biomed. Inf.* 107, 103421. doi:10.1016/j.jbi.2020.103421

Kwong, J. C. C., Khondker, A., Tran, C., Evans, E., Cozma, A. I., Javidan, A., et al. (2022). Explainable artificial intelligence to predict the risk of sidespecific extraprostatic extension in pre-prostatectomy patients. *Can. Urol. Assoc. J.* 16, 213–221. doi:10.5489/ cuaj.7473

Lai, Z., Markovets, A., Ahdesmaki, M., Chapman, B., Hofmann, O., McEwen, R., et al. (2016). VarDict: a novel and versatile variant caller for next-generation sequencing in cancer research. *Nucleic Acids Res.* 44, e108. doi:10.1093/nar/gkw227

Landrum, M. J., Lee, J. M., Benson, M., Brown, G. R., Chao, C., Chitipiralla, S., et al. (2018). ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res.* 46, D1062–D1067. doi:10.1093/nar/gkx1153

Larson, D. E., Harris, C. C., Chen, K., Koboldt, D. C., Abbott, T. E., Dooling, D. J., et al. (2012). SomaticSniper: identification of somatic point mutations in whole genome sequencing data. *Bioinformatics* 28, 311–317. doi:10.1093/bioinformatics/btr665

Lelieveld, S. H., Spielmann, M., Mundlos, S., Veltman, J. A., and Gilissen, C. (2015). Comparison of exome and genome sequencing technologies for the complete capture of protein-coding regions. *Hum. Mutat.* 36 (8), 815–822. doi:10.1002/humu.22813

le Sage, C., Lawo, S., Panicker, P., Scales, T. M., Rahman, S. A., Little, A. S., et al. (2017). Dual direction CRISPR transcriptional regulation screening uncovers gene networks driving drug resistance. *Sci. Rep.* 7, 17693. doi:10.1038/s41598-017-18172-6

Levatic, J., Salvadores, M., Fuster-Tormo, F., and Supek, F. (2022). Mutational signatures are markers of drug sensitivity of cancer cells. *Nat. Commun.* 13, 2926. doi:10.1038/s41467-022-30582-3

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:13033997v2 [q-bioGN]. doi:10.48550/arXiv.1303.3997

Li, Y., Roberts, N. D., Wala, J. A., Shapira, O., Schumacher, S. E., Kumar, K., et al. (2020). Patterns of somatic structural variation in human cancer genomes. *Nature* 578, 112–121. doi:10.1038/s41586-019-1913-9

Li, Y., Shi, W., and Wasserman, W. W. (2018). Genome-wide prediction of cisregulatory regions using supervised deep learning methods. *BMC Bioinf* 19 (1), 202. doi:10.1186/s12859-018-2187-1

Li, Y., Wu, X., Yang, P., Jiang, G., and Luo, Y. (2022). Machine learning for lung cancer diagnosis, treatment, and prognosis. *Genomics Proteomics Bioinforma*. 20 (5), 850–866. doi:10.1016/j.gpb.2022.11.003

Liu, T., Salguero, P., Petek, M., Martinez-Mira, C., Balzano-Nogueira, L., Ramšak, Ž., et al. (2022). PaintOmics 4: new tools for the integrative analysis of multi-omics datasets supported by multiple pathway databases. *Nucleic Acids Res.* 50 (W1), W551–W559. doi:10.1093/nar/gkac352

Lord, C. J., and Ashworth, A. (2016). BRCAness revisited. Nat. Rev. Cancer 16, 110-120. doi:10.1038/nrc.2015.21

Ma, J., Setton, J., Lee, N. Y., Riaz, N., and Powell, S. N. (2018a). The therapeutic significance of mutational signatures from DNA repair deficiency in cancer. *Nat. Commun.* 9, 3292. doi:10.1038/s41467-018-05228-y

Ma, X., Yu, L., Liu, Y., Alexandrov, L. B., Edmonson, M. N., Gawad, C., et al. (2018b). Pan-cancer genome and transcriptome analyses of 1,699 paediatric leukaemias and solid tumours. *Nature* 555:371–376. doi:10.1038/nature25795

Macintyre, G., Goranova, T. E., De Silva, D., Ennis, D., Piskorz, A. M., Eldridge, M., et al. (2018). Copy number signatures and mutational processes in ovarian carcinoma. *Nat. Genet.* 50, 1262–1270. doi:10.1038/s41588-018-0179-8

Makałowski, W. (2001). The human genome structure and organization. Acta Biochim. Pol. 48, 587–598. doi:10.18388/abp.2001_3893

Mani, D. R., Maynard, M., Kothadia, R., Krug, K., Christianson, K. E., Heiman, D., et al. (2021). PANOPLY: a cloud-based platform for automated and reproducible proteogenomic data analysis. *Nat. Methods* 18 (6), 580–582. doi:10.1038/s41592-021-01176-6

Martinez-Martin, N., and Magnus, D. (2019). Privacy and ethical challenges in nextgeneration sequencing. *Expert Rev. Precis. Med. Drug Dev.* 4 (2), 95–104. doi:10.1080/ 23808993.2019.1599685

Mathew, D. E., Ebem, D. U., Ikegwu, A. C., Ukeoma, P. E., and Dibiaezue, N. F. (2025). Recent emerging techniques in explainable artificial intelligence to enhance the interpretable and understanding of AI models for human. *Neural Process Lett.* 57, 16. doi:10.1007/s11063-025-11732-2

Mavaie, P., Holder, L., and Skinner, M. K. (2023). Hybrid deep learning approach to improve classification of low-volume high-dimensional data. *BMC Bioinforma*. 24, 419. doi:10.1186/s12859-023-05557-w

McGrail, D. J., Pilie, P. G., Rashid, N. U., Voorwerk, L., Slagter, M., Kok, M., et al. (2021). High tumor mutation burden fails to predict immune checkpoint blockade response across all cancer types. *Ann. Oncol.* 32, 661–672. doi:10.1016/j.annonc.2021. 02.006

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing nextgeneration DNA sequencing data. *Genome Res.* 20, 1297–1303. doi:10.1101/gr. 107524.110

McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R. S., Thormann, A., et al. (2016). The Ensembl variant effect predictor. *Genome Biol.* 17, 122. doi:10.1186/s13059-016-0974-4

Meinshausen, N., Hauser, A., Mooij, J. M., Peters, J., Versteeg, P., and Bühlmann, P. (2016). Methods for causal inference from gene perturbation experiments and validation. *Proc. Natl. Acad. Sci.* 113, 7361–7368. doi:10. 1073/pnas.1510493113

Menden, M. P., Casale, F. P., Stephan, J., Bignell, G. R., Iorio, F., McDermott, U., et al. (2018). The germline genetic component of drug sensitivity in cancer cell lines. *Nat. Commun.* 9, 3385. doi:10.1038/s41467-018-05811-3

Menyhárt, O., and Gyorffy, B. (2021). Multi-omics approaches in cancer research with applications in tumor subtyping, prognosis, and diagnosis. *Comput. Struct. Biotechnol. J.* 19, 949–960. doi:10.1016/j.csbj.2021.01.009

Merino, D. M., McShane, L. M., Fabrizio, D., Funari, V., Chen, S. J., White, J. R., et al. (2020). Establishing guidelines to harmonize tumor mutational burden (TMB): *in silico* assessment of variation in TMB quantification across diagnostic platforms: phase I of the Friends of cancer research TMB harmonization project. *J. Immunother. Cancer* 8 (1), e000147. doi:10.1136/jitc-2019-000147

Miller, C. A., White, B. S., Dees, N. D., Griffith, M., Welch, J. S., Griffith, O. L., et al. (2014). SciClone: inferring clonal architecture and tracking the spatial and temporal patterns of tumor evolution. *PLoS Comput. Biol.* 10, e1003665. doi:10.1371/journal.pcbi. 1003665

Misra, B. B., Langefeld, C. D., Olivier, M., and Cox, L. A. (2018). Integrated omics: tools, advances, and future approaches. *J. Mol. Endocrinol.* 62, R21–R45. doi:10.1530/JME18-0055

Mootha, V. K., Lindgren, C. M., Eriksson, K. F., Subramanian, A., Sihag, S., Lehar, J., et al. (2003). PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat. Genet.* 34, 267–273. doi:10.1038/ng1180

Nakach, F. Z., Idri, A., and Goceri, E. (2024). A comprehensive investigation of multimodal deep learning fusion strategies for breast cancer classification. *Artif. Intell. Rev.* 57, 327. doi:10.1007/s10462-024-10984-z

Nakagawa, H., and Fujita, M. (2018). Whole genome sequencing analysis for cancer genomics and precision medicine. *Cancer Sci.* 109, 513–522. doi:10.1111/cas. 13505

Nakagawa, H., Wardell, C. P., Furuta, M., Taniguchi, H., and Fujimoto, A. (2015). Cancer whole-genome sequencing: present and future. *Oncogene* 34, 5943–5950. doi:10. 1038/onc.2015.90

Newman, A. M., Steen, C. B., Liu, C. L., Gentles, A. J., Chaudhuri, A. A., Scherer, F., et al. (2019). Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat. Biotechnol.* 37, 773–782. doi:10.1038/s41587-019-0114-2

Ng, P. C., and Sift, H. S. (2003). Predicting amino acid changes that affect protein function. *Nucleic Acids Res.* 31 (13), 3812-3814. doi:10.1093/nar/gkg509

Nguyen, P. H. D., Ma, S., Phua, C. Z. J., Kaya, N. A., Lai, H. L. H., Lim, C. J., et al. (2021a). Intratumoural immune heterogeneity as a hallmark of tumour evolution and progression in hepatocellular carcinoma. Nat. Commun. 12, 227. doi:10.1038/s41467-020-20171-7

Nguyen, T., Nguyen, G. T. T., Nguyen, T., and Le, D. H. (2022). Graph convolutional networks for drug response prediction. *IEEE/ACM Trans. Comput. Biol. Bioinform* 19 (1), 146–154. doi:10.1109/TCBB.2021.3060430

Nguyen, T. M., Kim, N., Kim, D. H., Le, H. L., Piran, M. J., Um, S. J., et al. (2021b). Deep learning for human disease detection, subtype classification, and treatment response prediction using epigenomic data. *Biomedicines* 9 (11), 1733. doi:10.3390/biomedicines9111733

Ogburn, M., Turner, C., Dahal, P., and Encryption, H. (2013). Homomorphic encryption. *Procedia Comput. Sci.* 20, 502–509. doi:10.1016/j.procs.2013.09.310

Omichessan, H., Severi, G., and Perduca, V. (2019). Computational tools to detect signatures of mutational processes in DNA from tumours: a review and empirical comparison of performance. *PLoS One* 14, e0221235. doi:10.1371/journal.pone.0221235

Parkinson, H., Kapushesky, M., Shojatalab, M., Abeygunawardena, N., Coulson, R., Farne, A., et al. (2007). ArrayExpress-a public database of microarray experiments and gene expression profiles. *Nucleic Acids Res.* 35 (Database issue), D747–D750. doi:10. 1093/nar/gkl995

Paul, D., Sanap, G., Shenoy, S., Kalyane, D., Kalia, K., and Tekade, R. K. (2021). Artificial intelligence in drug discovery and development. *Drug Discov. Today* 26 (1), 80–93. doi:10.1016/j.drudis.2020.10.010

Perera-Bel, J., Hutter, B., Heining, C., Bleckmann, A., Frohlich, M., Fr \in ohling, S., et al. (2018). From somatic variants \in towards precision oncology: evidence-driven reporting of treatment options in molecular tumor boards. *Genome Med.* 10, 18. doi:10. 1186/s13073-018-0529-2

Picard, M., Scott-Boyer, M.-P., Bodein, A., Périn, O., and Droit, A. (2021). Integration strategies of multi-omics data for machine learning analysis. *Comput. Struct. Biotechnol. J.* 19, 3735–3746. doi:10.1016/j.csbj.2021.06.030

Pich, O., Muinos, F., Lolkema, M. P., Steeghs, N., ~ Gonzalez-Perez, A., and Lopez-Bigas, N. (2019). The mutational footprints of cancer therapies. *Nat. Genet.* 51, 1732–1740. doi:10.1038/s41588-019-0525-5

Pineiro-Yanez, E., Reboiro-Jato, M., G ~ omez-L opez, G., Perales-Paton, J., Troul, e K., Rodrguez, J. M., et al. (2018). PanDrugs: a novel method to prioritize anticancer drug treatments according to individual genomic data. *Genome Med.* 10, 41. doi:10. 1186/s13073-018-0546-1

PON (2021). Panel of normals (PON). Available online at: https://gatk.broadinstitute. org/hc/enus/articles/360035890631-Panel-of-Normals-PON.

Pos, O., Radvanszky, J., Styk, J., P \in os, Z., Bugly, \in o G., Kajsik, M., et al. (2021). Copy number variation: methods and clinical applications. *NATO Adv. Sci. Inst. Ser. E Appl. Sci.* 11, 819. doi:10.3390/app11020819

Poursaeed, R., Mohammadzadeh, M., and Safaei, A. A. (2024). Survival prediction of glioblastoma patients using machine learning and deep learning: a systematic review. *BMC Cancer* 24 (1), 1581. doi:10.1186/s12885-024-13320-4

Prabhakar, C. N. (2015). Epidermal growth factor receptor in non-small cell lung cancer. *Transl. Lung Cancer Res.* 4 (2), 110–118. doi:10.3978/j.issn.2218-6751.2015. 01.01

Preuer, K., Lewis, R. P. I., Hochreiter, S., Bender, A., Bulusu, K. C., and Klambauer, G. (2018). DeepSynergy: predicting anti-cancer drug synergy with Deep Learning. *Bioinformatics* 34 (9), 1538–1546. doi:10.1093/bioinformatics/btx806

Rajpurkar, P., Chen, E., Banerjee, O., and Topol, E. J. (2022). AI in health and medicine. *Nat. Med.* 28, 31-38. doi:10.1038/s41591-021-01614-0

Raman, L., Dheedene, A., De Smet, M., Van Dorpe, J., and Menten, B. (2019). WisecondorX: improved copy number detection for routine shallow whole-genome sequencing, *Nucleic Acids Res.*, 47, 4, 1605–1614. doi:10.1093/nar/gky1263

Reardon, B., Moore, N. D., Moore, N. S., Kofman, E., AlDubayan, S. H., Cheung, A. T. M., et al. (2021). Integrating molecular profiles into clinical frameworks through the Molecular Oncology Almanac to prospectively guide precision oncology. *Nat. Cancer* 2 (10), 1102–1112. doi:10.1038/s43018-021-00243-3

Rehm, H. L., Page, A. J. H., Smith, L., Adams, J. B., Alterovitz, G., Babb, L. J., et al. (2021). GA4GH: international policies and standards for data sharing across genomic research and healthcare. *Cell Genom* 1, 100029. doi:10.1016/j.xgen.2021.100029

Rigby, M. J. J. (2019). Ethical dimensions of using artificial intelligence in health care. A.J.o. E. 21 (2), 121–124.

Roden, D. M., McLeod, H. L., Relling, M. V., Williams, M. S., Mensah, G. A., Peterson, J. F., et al. (2019). Pharmacogenomics. *Lancet.* 394, 521–532. doi:10.1016/S0140-6736(19)31276-0

Rogers, M. F., Shihab, H. A., Mort, M., Cooper, D. N., Gaunt, T. R., and Campbell, C. (2018). FATHMM-XF: accurate prediction of pathogenic point mutations via extended features. *Bioinformatics* 34 (3), 511–513. doi:10.1093/bioinformatics/btx536

Rohart, F., Gautier, B., Singh, A., and Lê Cao, K. A. (2017). mixOmics: an R package for 'omics feature selection and multiple data integration. *PLoS Comput. Biol.* 13 (11), e1005752. doi:10.1371/journal.pcbi.1005752

Romero-Brufau, S., Wyatt, K. D., Boyum, P., Mickelson, M., Moore, M., and Cognetta-Rieke, C. (2020). Implementation of artificial intelligence-based clinical

decision support to reduce hospital readmissions at a regional hospital. Appl. Clin. Inf. 11 (4), 570–577. doi:10.1055/s-0040-1715827

Rosenbloom, K. R., Sloan, C. A., Malladi, V. S., Dreszer, T. R., Learned, K., Kirkup, V. M., et al. (2013). ENCODE data in the UCSC Genome Browser: year 5 update. *Nucleic Acids Res.* 41 (Database issue), D56–D63. doi:10.1093/nar/gks1172

Rubio-Perez, C., Tamborero, D., Schroeder, M. P., Antolin, A. A., Deu-Pons, J., Perez-Llamas, C., et al. (2015). *In silico* prescription of anticancer drugs to cohorts of 28 tumor types reveals targeting opportunities. *Cancer Cell* 27, 382–396. doi:10.1016/j.ccell.2015. 02.007

Saha, S., Hota, A., Chattopadhyay, A. K., Nag, A., and Nandi, S. (2024). A multifaceted survey on privacy preservation of federated learning: progress, challenges, and opportunities. *Artif. Intell. Rev.* 57, 184. doi:10.1007/s10462-024-10766-7

Salehi, S., Kabeer, F., Ceglia, N., Andronescu, M., Williams, M. J., Campbell, K. R., et al. (2021). Clonal fitness inferred from time-series modelling of single-cell cancer genomes. *Nature* 595, 585–590. doi:10.1038/s41586-021-03648-3

Sanchez-Vega, F., Mina, M., Armenia, J., Chatila, W. K., Luna, A., La, K. C., et al. (2018). Oncogenic signaling pathways in the cancer genome atlas. *Cell* 173, 321–337.e10. doi:10.1016/j.cell.2018.03.035

Sangkuhl, K., Whirl-Carrillo, M., Whaley, R. M., Woon, M., Lavertu, A., Altman, R. B., et al. (2020). Pharmacogenomics clinical annotation tool (PharmCAT). *Clin. Pharmacol. Ther.* 107, 203–210. doi:10.1002/cpt.1568

Sarker, I. H. (2021). Machine learning: algorithms, real-world applications and research directions. SN Comput. Sci. 2, 160. doi:10.1007/s42979-021-00592-x

Schutte, J., Reusch, J., Khandanpour, C., and Eisfeld, C. (2019). Structural variants as a basis for targeted therapies in hematological malignancies. *Front. Oncol.* 9, 839. doi:10. 3389/fonc.2019.00839

Sha, D., Jin, Z., Budczies, J., Kluck, K., Stenzinger, A., and Sinicrope, F. A. (2020). Tumor mutational burden as a predictive biomarker in solid tumors. *Cancer Discov.* 10 (12), 1808–1825. doi:10.1158/2159-8290.CD-20-0522

Shakyawar, S. K., Sajja, B. R., Patel, J. C., and Guda, C. (2024). *iCluF*: an unsupervised iterative cluster-fusion method for patient stratification using multiomics data. *Bioinform Adv.* 4 (1), vbae015. doi:10.1093/bioadv/vbae015

Shigemizu, D., Fujimoto, A., Akiyama, S., Abe, T., Nakano, K., Boroevich, K. A., et al. (2013). A practical method to detect SNVs and indels from whole genome and exome sequencing data. *Sci. Rep.* 3, 2161. doi:10.1038/srep02161

Siravegna, G., Marsoni, S., Siena, S., and Bardelli, A. (2017). Integrating liquid biopsies into the management of cancer. *Nat. Rev. Clin. Oncol.* 14, 531–548. doi:10.1038/nrclinonc.2017.14

Sorell, T., Rajpoot, N., and Verrill, C. (2021). Ethical issues in computational pathology. J. Med. Ethics 48, 278-284. doi:10.1136/medethics-2020-107024

Srivastava, R. (2022). Role of transcriptomics in precision oncology. *Rep. Radiother.* Oncol. 9 (1), e142195. doi:10.5812/rro-142195

Srivastava, R. (2023a). Applications of artificial intelligence multiomics in precision oncology. J. Cancer Res. Clin. Oncol. 149, 503–510. doi:10.1007/s00432-022-04161-4

Srivastava, R. (2023b). Role of smartphone devices in precision oncology. J. Cancer Res. Clin. Oncol. 149 (1), 393-400. doi:10.1007/s00432-022-04413-3

Srivastava, R. (2023c). Multiomics analysis on the clinical treatment for multiple myeloma (MM). J. Hematol. Mult. Myeloma 6 (1), 1028.

Srivastava, R. (2024a). Applications of artificial intelligence in medicine. *Explor Res. Hypothesis Med.* 9 (2), 138–146. doi:10.14218/ERHM.2023.00048

Srivastava, R. (2024b). Artificial intelligence multiomics in precision oncology. Cambridge Scholars Publishing Newcastle upon Tyne, United Kingdom.

Subramanian, I., Verma, S., Kumar, S., Jere, A., and Anamika, K. (2020). Multi-omics data integration, interpretation, and its application. *Bioinf Biol. Insights* 14, 1177932219899051. doi:10.1177/1177932219899051

Sun, H., and Yu, G. (2019). New insights into the pathogenicity of non-synonymous variants through multi-level analysis. *Sci. Rep.* 9, 1667. doi:10.1038/s41598-018-38189-9

Sun, Y. V., and Hu, Y.-J. (2016). Integrative analysis of multi-omics data for discovery and functional studies of complex human diseases. *Adv. Genet.* 93, 147–190. doi:10. 1016/bs.adgen.2015.11.004

Szymczak, S., Biernacka, J. M., Cordell, H. J., González-Recio, O., König, I. R., Zhang, H., et al. (2009). Machine learning in genome-wide association studies. *Genet. Epidemiol.* 33 (Suppl. 1), S51–S57. doi:10.1002/gepi.20473

Taj, F., and Stein, L. D. (2024). MMDRP: drug response prediction and biomarker discovery using multi-modal deep learning. *Bioinform Adv.* 4 (1), vbae010. doi:10.1093/bioadv/vbae010

Takase, S., Kurokawa, R., Arai, D., Kanemoto Kanto, K., Okino, T., Nakao, Y., et al. (2017). A quantitative shRNA screen identifies ATP1A1 as a gene that regulates cytotoxicity by aurilide B. *Sci. Rep.* 7, 2002. doi:10.1038/s41598-017-02016-4

Tamborero, D., Dienstmann, R., Rachid, M. H., Boekel, J., Lopez-Fernandez, A., Jonsson, M., et al. (2022). The Molecular Tumor Board Portal supports clinical decisions and automated reporting for precision oncology. *Nat. Cancer* 3, 251–261. doi:10.1038/s43018-022-00332-x

Tamborero, D., Rubio-Perez, C., Deu-Pons, J., Schroeder, M. P., Vivancos, A., Rovira, A., et al. (2018). Cancer Genome Interpreter annotates the biological and clinical relevance of tumor alterations. *Genome Med.* 10, 25. doi:10.1186/s13073-018-0531-8

Tate, J. G., Bamford, S., Jubb, H. C., Sondka, Z., Beare, D. M., Bindal, N., et al. (2019). COSMIC: the Catalogue of somatic mutations in cancer. *Nucleic Acids Res.* 47 (D1), D941–D947. doi:10.1093/nar/gky1015

Tautenhahn, R., Patti, G. J., Rinehart, D., and Siuzdak, G. (2012). XCMS Online: a web-based platform to process untargeted metabolomic data. *Anal. Chem.* 84 (11), 5035–5039. doi:10.1021/ac300698c

Tebani, A., Afonso, C., Marret, S., and Bekri, S. (2016). Omics-based strategies in precision medicine: toward a paradigm shift in inborn errors of metabolism investigations. *Int. J. Mol. Sci.* 17, 1555. doi:10.3390/ijms17091555

Telenti, A., Lippert, C., Chang, P. C., and DePristo, M. (2018). Deep learning of genomic variation and regulatory network data. *Hum. Mol. Genet.* 27 (R1), R63-r71. doi:10.1093/hmg/ddy115

Tomczak, K., Czerwinska, P., and Wiznerowicz, M. (2015). Review the cancer genome atlas (TCGA): an immeasurable source of knowledge. *Contemp. oncology/Współczesna Onkol.* 2015: 68–77. doi:10.5114/wo.2014.47136

Tuncbag, N., Gosline, S. J., Kedaigle, A., Soltis, A. R., Gitter, A., and Fraenkel, E. (2016). Network-based interpretation of diverse high-throughput datasets through the Omics Integrator software package. *PLoS Comput. Biol.* 12, e1004879. doi:10.1371/journal.pcbi.1004879

Tuncbag, N., McCallum, S., Huang, S. S., and Fraenkel, E. (2012). SteinerNet: a web server for integrating 'omic' data to discover hidden components of response pathways. *Nucleic Acids Res.* 40 (Web Server issue), W505–W509. doi:10.1093/nar/gks445

Turajlic, S., Sottoriva, A., Graham, T., and Swanton, C. (2019). Resolving genetic heterogeneity in cancer. *Nat. Rev. Genet.* 20, 404–416. doi:10.1038/s41576-019-0114-6

Vahabi, N., and Michailidis, G. (2022). Unsupervised multi-omics data integration methods: a comprehensive review. *Front. Genet.* 13, 854752. doi:10.3389/fgene.2022. 854752

Valle-Inclan, J. E., Stangl, C., de Jong, A. C., van Dessel, L. F., van Roosmalen, M. J., Helmijr, J. C. A., et al. (2021). Optimizing nanopore sequencing-based detection of structural variants enables individualized circulating tumor DNA-based disease monitoring in cancer patients. *Genome Med.* 13, 86. doi:10.1186/s13073-021-00899-7

van Belzen, IAEM, Schonhuth, A., Kemmeren, P., and Hehir- € Kwa, J. Y. (2021). Structural variant detection in cancer genomes: computational challenges and perspectives for precision oncology. *NPJ Precis. Oncol.* 5, 15. doi:10.1038/s41698-021-00155-6

van Dijk, D., Sharma, R., Nainys, J., Yim, K., Kathail, P., Carr, A. J., et al. (2018). Recovering gene interactions from single-cell data using data diffusion. *Cell* 174, 716–729.e27. doi:10.1016/j.cell.2018.05.061

Waddell, N., Pajic, M., Patch, A.-M., Chang, D. K., Kassahn, K. S., Bailey, P., et al. (2015). Whole genomes redefine the mutational landscape of pancreatic cancer. *Nature* 518, 495–501. doi:10.1038/nature14169

Wagner, A. H., Walsh, B., Mayfield, G., Tamborero, D., Sonkin, D., Krysiak, K., et al. (2020). A harmonized metaknowledgebase of clinical interpretations of somatic genomic variants in cancer. *Nat. Genet.* 52, 448–457. doi:10.1038/s41588-020-0603-8

Wang, B., Mezlini, A. M., Demir, F., Fiume, M., Tu, Z., Brudno, M., et al. (2014). Similarity network fusion for aggregating data types on a genomic scale. *Nat. Methods* 11 (3), 333–337. doi:10.1038/nmeth.2810

Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from highthroughput sequencing data. *Nucleic Acids Res.* 38, e164. doi:10.1093/nar/gkq603

Wang, Y., Yang, Y., Chen, S., and Wang, J. (2021). DeepDRK: a deep learning framework for drug repurposing through kernel-based multi-omics integration. *Brief. Bioinform* 22, bbab048. doi:10.1093/bib/bbab048

Warner, J. L., Prasad, I., Bennett, M., Arniella, M., BeeghlyFadiel, A., Mandl, K. D., et al. (2018). SMART Cancer Navigator: a framework for implementing ASCO workshop recommendations to enable precision cancer medicine. *JCO Precis. Oncol.* 2018, 1–14. doi:10.1200/PO.17.00292

Wei, L., Niraula, D., Gates, E. D. H., Fu, J., Luo, Y., Nyflot, M. J., et al. (2023). Artificial intelligence (AI) and machine learning (ML) in precision oncology: a review on enhancing discoverability through multiomics integration. *Br. J. Radiol.* 96 (1150), 20230211. doi:10.1259/bjr.20230211

Wei, Z., Han, D., Zhang, C., Wang, S., Liu, J., Chao, F., et al. (2022). Deep learningbased multi-omics integration robustly predicts relapse in prostate cancer. *Front. Oncol.* 12, 893424. doi:10.3389/fonc.2022.893424

Wendland, P., Birkenbihl, C., Gomez-Freixa, M., Sood, M., Kschischo, M., and Fröhlich, H. (2022). Generation of realistic synthetic data using multimodal neural ordinary differential equations. *NPJ Digit. Med.* 5 (1), 122. doi:10.1038/s41746-022-00666-x

Whirl-Carrillo, M., Huddart, R., Gong, L., Sangkuhl, K., Thorn, C. F., Whaley, R., et al. (2021). An evidence-based framework for evaluating pharmacogenomics knowledge for personalized medicine. *Clin. Pharmacol. Ther.* 110, 563–572. doi:10.1002/cpt.2350

Wishart, D. S., Feunang, Y. D., Guo, A. C., Lo, E. J., Marcu, A., Grant, J. R., et al. (2018). DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* 46, D1074–D1082. doi:10.1093/nar/gkx1037

Xiao, Y., Wang, X., Zhang, H., Ulintz, P. J., Li, H., and Guan, Y. (2020). FastClone is a probabilistic tool for deconvoluting tumor heterogeneity in bulk-sequencing samples. *Nat. Commun.* 11, 4469. doi:10.1038/s41467-020-18169-2

Xie, X. P., Laks, D. R., Sun, D., Ganbold, M., Wang, Z., Pedraza, A. M., et al. (2022). Quiescent human glioblastoma cancer stem cells drive tumor initiation, expansion, and recurrence following chemotherapy. *Dev. Cell* 57, 32–46.e8. doi:10.1016/j.devcel.2021. 12.007

Xu, F., Sepúlveda, M. J., Jiang, Z., Wang, H., Li, J., Liu, Z., et al. (2020). Effect of an artificial intelligence clinical decision support system on treatment decisions for complex breast cancer. *JCO Clin. Cancer Inf.* 4, 824–838. doi:10.1200/CCI.20. 00018

Yamamoto, T. N., Kishton, R. J., and Restifo, N. P. (2019). Developing neoantigentargeted T cell-based treatments for solid tumors. *Nat. Med.* 25, 1488–1499. doi:10. 1038/s41591-019-0596-y

Yang, S., Wang, Z., Wang, C., Li, C., and Wang, B. (2024). Comparative evaluation of machine learning models for subtyping triple-negative breast cancer: a deep learningbased multi-omics data integration approach. *J. Cancer* 15 (12), 3943–3957. doi:10. 7150/jca.93215 Yao, H., Liang, Q., Qian, X., Wang, J., Sham, P. C., and Li, M. J. (2020). Methods and resources to access mutation-dependent effects on cancer drug treatment. *Brief. Bioinform* 21, 1886–1903. doi:10.1093/bib/bbz109

Yin, H., and Kassner, M. (2016). "In vitro high-throughput RNAi screening to accelerate the process of target identification and drug development," in *High-throughput RNAi screening: methods and protocols* (Berlin/Heidelberg, Germany: Springer), 137–149.

Yu, Y., Wang, Y., Xia, Z., Zhang, X., Jin, K., Yang, J., et al. (2019). PreMedKB: an integrated precision medicine knowledgebase for interpreting relationships between diseases, genes, variants and drugs. *Nucleic Acids Res.* 47, D1090–D1101. doi:10.1093/nar/gky1042

Zarrei, M., MacDonald, J. R., Merico, D., and Scherer, S. W. (2015). A copy number variation map of the human genome. *Nat. Rev. Genet.* 16, 172–183. doi:10.1038/nrg3871

Zhang, Q., Major, M. B., Takanashi, S., Camp, N. D., Nishiya, N., Peters, E. C., et al. (2007). Small-molecule synergist of the Wnt/beta-catenin signaling pathway. *Proc. Natl. Acad. Sci. U. S. A.* 104, 7444–7448. doi:10.1073/pnas.0702136104

Zheng, J., Yi, H. C., and You, Z. H. (2025). Equivariant 3D-conditional diffusion model for *de novo* drug design. *IEEE J. Biomed. Health Inf.* 29 (3), 1805–1816. doi:10. 1109/JBHI.2024.3491318

Zhou, F. T., Piccardi, M., Abolhasan, M., Franklin, D., and Lipman, J. (2024). Secure multi-party computation for machine learning: a survey. *IEEE Access* 12, 53881–53899. doi:10.1109/ACCESS.2024.3388992