



Review

Integrating Artificial Intelligence in Next-Generation Sequencing: Advances, Challenges, and Future Directions

Konstantina Athanasopoulou , Vasiliki-Ioanna Michalopoulou, Andreas Scorilas and
Panagiotis G. Adamopoulos *

Department of Biochemistry and Molecular Biology, Faculty of Biology, National and Kapodistrian University of Athens, 15701 Athens, Greece; konnath@biol.uoa.gr (K.A.); vasilinamichalopoulou@gmail.com (V.-I.M.); ascorilas@biol.uoa.gr (A.S.)

* Correspondence: padamopoulos@biol.uoa.gr; Tel.: +30-210-72-74-686

Abstract: The integration of artificial intelligence (AI) into next-generation sequencing (NGS) has revolutionized genomics, offering unprecedented advancements in data analysis, accuracy, and scalability. This review explores the synergistic relationship between AI and NGS, highlighting its transformative impact across genomic research and clinical applications. AI-driven tools, including machine learning and deep learning, enhance every aspect of NGS workflows—from experimental design and wet-lab automation to bioinformatics analysis of the generated raw data. Key applications of AI integration in NGS include variant calling, epigenomic profiling, transcriptomics, and single-cell sequencing, where AI models such as CNNs, RNNs, and hybrid architectures outperform traditional methods. In cancer research, AI enables precise tumor subtyping, biomarker discovery, and personalized therapy prediction, while in drug discovery, it accelerates target identification and repurposing. Despite these advancements, challenges persist, including data heterogeneity, model interpretability, and ethical concerns. This review also discusses the emerging role of AI in third-generation sequencing (TGS), addressing long-read-specific challenges, like fast and accurate basecalling, as well as epigenetic modification detection. Future directions should focus on implementing federated learning to address data privacy, advancing interpretable AI to improve clinical trust and developing unified frameworks for seamless integration of multi-modal omics data. By fostering interdisciplinary collaboration, AI promises to unlock new frontiers in precision medicine, making genomic insights more actionable and scalable.

Keywords: NGS; AI; machine learning; deep learning; genomics; transcriptomics; precision medicine; cancer research; drug discovery; data integration



Academic Editor: Gaurav Tripathi

Received: 20 May 2025

Revised: 12 June 2025

Accepted: 17 June 2025

Published: 19 June 2025

Citation: Athanasopoulou, K.; Michalopoulou, V.-I.; Scorilas, A.; Adamopoulos, P.G. Integrating Artificial Intelligence in Next-Generation Sequencing: Advances, Challenges, and Future Directions. *Curr. Issues Mol. Biol.* **2025**, *47*, 470. <https://doi.org/10.3390/cimb47060470>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The 20th century marked a pivotal revolution in our understanding of the genetic foundations of life. The discovery of the double-helix structure of DNA and the elucidation of the central dogma of molecular biology laid the groundwork for a new era in human biology and genomics [1,2]. These discoveries redefined biology as a science fundamentally driven by the genetic alphabet of adenine, thymine, cytosine, and guanine. However, deciphering this “book of life” required not only biochemical ingenuity but also significant technological innovation. This transition was accelerated by the completion of the Human Genome Project (HGP), which catalyzed rapid advances in sequencing technology and computational methods, ultimately reshaping the landscape of genomic research [3,4]. Despite the early contributions of Sanger sequencing, the HGP completion did not mark an

endpoint but, rather, the beginning of a new era, unveiling the complexity of the human genome and highlighting the limitations of manual data analysis [5].

Following the completion of the HGP, sequencing technologies demonstrated significant improvements at an exponential rate. Next-generation sequencing (NGS) and third-generation sequencing (TGS) platforms, such as those developed by Illumina® and Oxford Nanopore Technologies (ONT), respectively, revolutionized genomic data acquisition by offering high-throughput and cost-effective sequencing solutions. These innovations enabled the sequencing of entire genomes within hours, significantly reducing costs and accelerating both research and clinical applications [6]. Concurrently, genome editing tools evolved remarkably, with the CRISPR/Cas9 system transforming genome editing from a theoretical concept into a precise and programmable tool. This breakthrough enabled targeted manipulations of DNA sequences with unprecedented versatility and accuracy, revolutionizing genetic research and therapeutic development [7].

Furthermore, the genomic data volume, and thus the amount of information per experiment, has increased exponentially, surpassing in many cases the capabilities of traditional computational approaches. Despite these advances, NGS data analysis continues to face major challenges. These include the sheer volume of sequencing data, the complexity and variability of biological signals, and the prevalence of technical artifacts such as amplification bias, batch effects, and sequencing errors. Traditional computational tools often struggle with these issues, which motivates the integration of artificial intelligence (AI) approaches that can model nonlinear patterns, automate feature extraction, and improve interpretability across large-scale datasets. This bottleneck catalyzed the adoption and integration of AI into genomics. Machine learning (ML) and deep learning (DL) algorithms, powered by advancements in neural networks and cloud computing, emerged as indispensable tools. Advanced DL models are now being employed to predict the structural and functional characteristics of genes from DNA or protein sequences, revealing intricate patterns beyond human discernment and effectively bridging the gap between data generation and biological interpretation [8].

Today, AI and genomics operate within a dynamic feedback loop. AI enhances genomic research by streamlining experimental design, simulating outcomes through predictive models, automating laboratory procedures to reduce manual work, and facilitating complex data analysis. Conversely, the vast repositories of genomic data enhance AI systems, refining their capacity to emulate complex biological reasoning. However, this synergy also raises significant ethical concerns, including cognitive offloading, algorithmic biases, privacy issues, as well as the profound moral implications of editing the fundamental code of life. These challenges necessitate urgent and ongoing discourse [9,10].

In this review, we explore the synergistic interplay between AI and genomics, with a focused exploration on AI's contributions to NGS applications and data analysis. We highlight AI's transformative impact across multiple NGS workflows—from experimental design to automated library preparation and implementation of AI-driven pipelines. Special attention is given to how ML enhances the accuracy of NGS data interpretation, enables predictive modeling of editing outcomes, and accelerates the discovery of novel findings. The discussion extends to current challenges in AI-assisted genomic analysis, including scalability, bias mitigation, and ethical considerations in precision genome editing. Through this discussion, we aim to highlight the capabilities of AI-based tools that can be utilized to overcome impediments in NGS-based research and clinical applications.

2. AI in the Laboratory: Enhancing Research from Plan to Data Analysis

In recent years, the convergence of AI and genomics has garnered significant academic and research attention, as well as substantial investments from universities, research insti-

tutions, hospitals, and pharmaceutical companies. This integration has fully transformed the landscape of genomic research, incorporating AI-driven methodologies—including ML and DL—that enhance our capacity to analyze complex biological data, yielding more accurate results [11,12]. This evolution represents a paradigm shift in how genomic information is interpreted and applied, reflecting AI's transformative impact across the field of sequencing.

2.1. The Pre-Wet-Lab Phase

The pre-wet-lab phase in genomics has traditionally been characterized by manual experimental design, which was heavily reliant on prior knowledge, empirical guidelines, and trial-and-error methods. With the advent of AI, pre-wet-lab procedures have undergone a radical transformation [13]. Today, AI-driven computational tools play a pivotal role in the strategic planning of experiments, assisting researchers in predicting outcomes, optimizing protocols, and anticipating potential challenges prior to initiating wet-lab work (Figure 1).

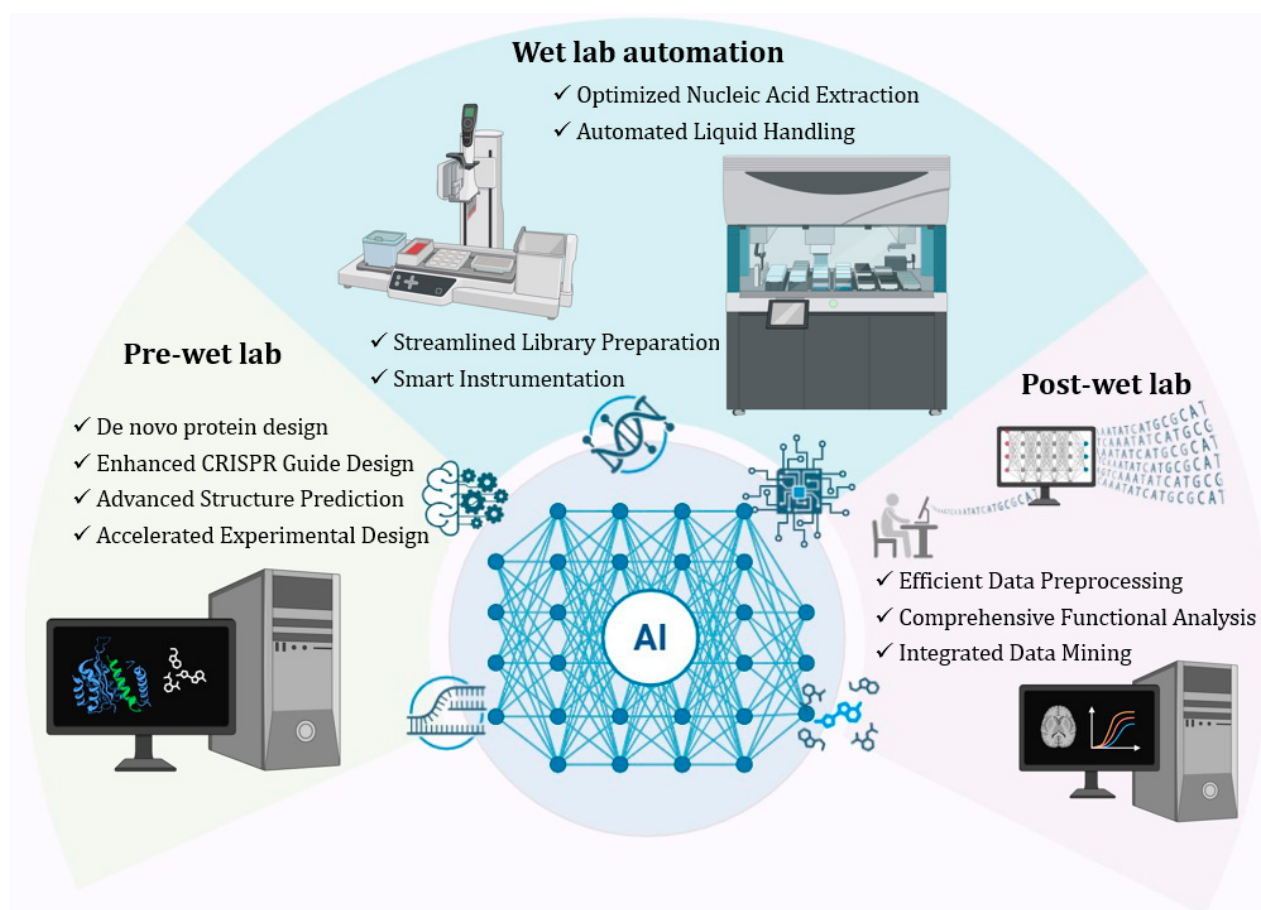


Figure 1. Artificial intelligence enhances NGS-based workflows across pre-wet lab, wet-lab automation, and post-wet-lab phases.

Briefly, several AI tools can be used to enhance the pre-wet-lab phase by improving experimental design and simulating outcomes. Benchling is a cloud-based platform that integrates AI to help researchers efficiently design experiments, optimize protocols, and manage lab data [14]. Similarly, DeepGene, an AI-powered tool, uses advanced deep neural networks to predict gene expression and assess experimental conditions, offering insight into the expected outcomes before laboratory work begins [15]. Of note, platforms such as Labster provide interactive virtual labs that simulate experimental setups, enabling researchers

to visualize outcomes and troubleshoot potential failures in a risk-free environment. In addition, tools, including Indigo AI (<https://www.fintechscotland.com/fintech/indigo-ai/>, accessed on 16 June 2025) and LabGPT (<https://chatgpt.com/g/g-3eIYfoFVJ-labgpt>, accessed on 16 June 2025), offer generative AI capabilities for automated protocol generation and experimental planning, while Synthace enables AI-powered protocol optimization and lab automation. Colabra supports collaborative planning and experiment tracking using AI-enhanced project workflows, and DeepChem offers ML frameworks for molecular property prediction, aiding compound screening and hypothesis formulation. These AI-driven innovations not only improve the efficiency of the experimental design process but also help to identify and mitigate potential issues before they arise in the wet lab, thus enhancing the overall success of genomic research.

2.2. The Wet-Lab Phase

AI's impact extends into the wet-lab phase, transforming laboratory workflows through automation, optimization, and real-time analysis. AI-driven automation technologies have streamlined traditional labor-intensive procedures, significantly improving reproducibility, scalability, and data quality (Figure 1). For instance, the Tecan Fluent systems are highly modular, deck-based liquid handling workstations that can be tailored to automate plate- or tube-based assays [16]. These platforms are particularly effective in automating tasks such as PCR and qPCR setup, NGS library preparation, nucleic acid extractions, and CRISPR workflows, utilizing AI algorithms to detect worktable and pipetting errors [17]. More precisely, in CRISPR workflows, AI-powered platforms have emerged to streamline both experimental design and validation. For instance, Synthego's CRISPR Design Studio offers automated gRNA design, editing outcome prediction, and end-to-end workflow planning, while tools like DeepCRISPR use DL to maximize editing efficiency and minimize off-target effects [18]. Additionally, R-CRISPR is used for gRNA design and combines convolutional neural networks (CNNs) and recurrent neural networks (RNNs) to predict off-target effects with mismatch or indels [19]. In proteomics and immunoassay workflows, these platforms facilitate the automation of ELISA, bead-based immunoassays, as well as protein digestion for mass spectrometry. Additionally, in cell-based assays, the systems can handle automated cell seeding, cell culture media change, staining and fixing steps, and cytotoxicity assays [20].

Beyond automation, AI enhances laboratory workflows through real-time monitoring and feedback control. For instance, a recent study integrated the AI-powered YOLOv8 model with the Opentrons OT-2 liquid handling robot to provide real-time quality control. This system enables precise detection of pipette tips and liquid volumes, offering immediate feedback to correct errors, including missing tips or incorrect liquid volumes, thereby ensuring experimental accuracy [21]. Such AI-driven solutions can significantly facilitate laboratory automation, making advanced capabilities accessible even in resource-limited settings.

2.3. The Post-Wet-Lab Phase

Following experimental implementation, the post-wet-lab phase has traditionally involved intensive and complex data analysis, a process frequently hindered by the complexity of genomic datasets (Figure 1). AI has dramatically accelerated this phase by providing tools that streamline bioinformatics and enhance data interpretation. Platforms such as Illumina BaseSpace Sequence Hub and DNAnexus enable bioinformatics analyses without requiring advanced programming skills or command-line tools. These cloud-based environments have recently incorporated bioinformatics tools that leverage AI/ML to perform analysis of complex genomic and biomedical data. Their user-friendly graphical interfaces often support custom pipeline construction through intuitive drag-and-drop

features. DL models trained on large-scale biological datasets can identify subtle patterns, predict biological functions, and suggest mechanistic hypotheses that might elude traditional statistical approaches. Tools like DeepVariant apply deep neural networks to improve the accuracy of variant calling from sequencing data, surpassing traditional heuristic-based approaches [22]. As for CRISPR post-experiment analysis, tools such as CRISPResso2, GUIDE-seq, and Cas-OFFinder are widely used to detect and quantify genome-wide off-target edits [23–25]. More advanced methods, including CRISPRitz and AI-enhanced scoring models in DeepCRISPR, provide flexible and rapid off-target prediction, improving both accuracy and safety in gene-editing applications [26,27].

The availability of extensive online biological databases has become essential to post-wet-lab research. Public repositories, such as NCBI's Gene Expression Omnibus (GEO), ENCODE, TCGA, and EMBL-EBI, provide access to large-scale, curated datasets across genomics, transcriptomics, and epigenomics [28–31]. AI and computational platforms now assist in dataset selection and functional comparison. OMICtools provides a searchable catalog of bioinformatics tools and knockout-specific datasets [32]. MAGECK enables normalization, ranking, and comparison of gene dependencies from pooled CRISPR screens across multiple cell lines [33]. Graph-based systems like KnetMiner support gene-centric queries, retrieving knockout experiments, and comparing affected pathways and phenotypes across species [34]. After dataset retrieval, analysis systems such as ShinyGO, ReactomeGSA, and Metascape facilitate cross-study enrichment and network comparison, revealing shared and divergent biological impacts of gene knockouts in various models [35–37].

Moreover, novel AI-driven discovery frameworks are now applied to gene regulatory networks (GRNs). These tools use curiosity-driven exploration algorithms to probe the range of stable “goal states” that GRNs can robustly reach—effectively mapping a “behavioral catalog” for each network. For instance, in one study spanning 432 GRNs from plants, bacteria, rodents, and humans, natural networks exhibited orders of magnitude higher behavioral versatility than matched random networks [38]. Additionally, MICRAT is a specialized algorithm for inferring GRNs from time-series gene expression data, enhancing directionality and causal insight by combining MIC with conditional entropy and temporal information [39]. These tools enable quantitative comparisons of network robustness and versatility across organisms and can identify clinically relevant states and facilitate therapeutic strategy design.

AI-driven search and meta-analysis tools enhance the utility of these databases by enabling cross-study comparisons, predicting functional associations, and detecting subtle patterns that may not be apparent through conventional methods [40]. Recently, the generative AI models of ChatGPT (GPT-4o), DeepSeek (version V3), and BioGPT (<https://github.com/microsoft/>) have emerged as powerful tools for scientific interpretation and experimental design refinement [41–43]. These language models assist researchers in interpreting complex datasets, summarizing extensive experimental results, and proposing biological insights based on multi-omics data. However, while AI performs highly sophisticated analyses, it is crucial to validate findings through orthogonal methods and maintain a critical perspective. Automated interpretation, although powerful, is not immune to biases or data artifacts, emphasizing the need for human oversight.

3. AI in NGS Data Analysis

The analysis of NGS data has been radically transformed by the application of AI models, which provide powerful tools for pattern recognition, classification, and exploratory analysis. These approaches are particularly valuable for their interpretability and computational efficiency, especially in cases of genomic and transcriptomic datasets. To efficiently

manage and retrieve specific output files from large-scale NGS datasets, several workflow management systems and automation tools have been developed (Table 1). Nextflow is one such widely adopted workflow management system that enables scalable, reproducible data analysis across diverse computing environments. It streamlines complex pipelines by automating file handling, software dependencies, and output organization using features like the publishDir directive [44].

Table 1. Overview of Nextflow-based pipelines facilitating scalable and reproducible NGS workflows.

Pipeline	Application	Features	Reference
nf-core/circrna	circRNA and miRNA analysis	circRNA quantification, miRNA target prediction, differential analysis	[45]
nf-rnaSeqCount	RNA-seq quantification	QC, alignment, count quantification, MultiQC reporting	[46]
GeneTEFlow	Gene + TE expression	STAR/RSEM + SQuIRE quantification, DE analysis, Docker containerization	[47]
scATACpipe	scATAC-seq processing	Preprocessing, fragment/BED/BAM outputs, ArchR clustering, interactive HTML report	[48]
polishCLR	Genome assembly polishing	PacBio CLR polishing with Illumina data, haplotig purging, outputs quality-checked assemblies and logs	[49]
nf-gwas-pipeline	GWAS analysis	Automates genotype QC, population structure correction, association, and visualization	[50]
StellarPGx	Pharmacogenomics	CYP star allele calling, variant annotation, phasing and clinical interpretation	[51]
NFTest	Pipeline testing	Automated functional testing of Nextflow workflows using synthetic test cases	[52]
PoSeiDon	Positive selection and recombination analysis	Runs PAML CODEML for detecting positive selection and recombination in nucleotide alignments; configurable for HPC/Singularity deployments	[53]
Pipeliner	Sequencing data preprocessing	Modular processing for bulk and single-cell RNA-Seq; leverages Nextflow + Conda for reproducible workflows	[54]
FA-nf	Functional annotation	Nextflow-based functional annotation of novel genomes using Pfam, GO terms, database scaffolding	[55]
Geniac	Nextflow add-on	Auto-generates config files + containers; linter enforces standardized outputs	[56]
DolphinNext	Workflow management	GUI platform built on Nextflow; drag-and-drop pipeline design, monitoring, containerized reproducibility	[57]

3.1. Machine Learning Approaches

Supervised ML algorithms have become essential for NGS applications requiring precise classification or regression. Random forests (RFs) and support vector machines (SVMs) represent the most widely adopted methods due to their robustness against overfitting and their ability to handle high-throughput sequencing data. In variant calling, RF-based pipelines, like GATK (© Broad Institute, USA) (Genome Analysis Toolkit), demonstrate superior performance in distinguishing true genetic variants from sequencing artifacts by integrating multiple quality metrics [58,59]. Gradient boosted machines (GBMs) have also proven effective for predicting variant pathogenicity when trained on curated clinical datasets [60]. For transcriptomic data, supervised learning enables disease subtyping and biomarker discovery [61]. The PAM50 classifier, which uses an SVM-based approach to identify breast cancer subtypes from RNA-Seq data, has been widely known as a clinical standard [62]. Similarly, ML models have been successfully applied to develop prognostic signatures for various cancers by selecting target genes from expression profiles [63].

Unsupervised techniques are indispensable for extracting patterns from unlabeled NGS data. Principal component analysis (PCA), while a statistical rather than a classical ML technique, is widely used for dimensionality reduction in genomic datasets to implement clustering and visualization. PCA-based tools enable fast and memory-efficient analysis of single-nucleotide polymorphism (SNP) variations in population-scale studies [64]. Notably,

more advanced methods, such as t-distributed stochastic neighbor embedding (t-SNE) [65] and uniform manifold approximation and projection (UMAP) [66], facilitate the visualization of high-dimensional datasets. These techniques have proven transformative in single-cell RNA sequencing (scRNA-seq) studies, enabling the discovery of novel cell states and populations.

While ML methods offer advantages in interpretability and present relatively low computational requirements, they can be limited in handling high-dimensional NGS data. Feature selection becomes critical to mitigate the “curse of dimensionality,” and for more complex applications, integration with DL techniques may offer improved performance. Finally, careful attention must be paid to batch effects and technical artifacts, which can confound ML-driven analyses if not properly examined.

3.2. Deep Learning Approaches

DL architectures such as CNNs, RNNs, and transformers are particularly well-suited for NGS data due to their ability to model distinct data characteristics. CNNs excel at identifying localized patterns indicative of mutations or sequence motifs, making them ideal for tasks such as variant calling and transcription factor binding site prediction [67]. RNNs, including long short-term memory networks (LSTMs), capture sequential dependencies and are powerful for modeling long genomic sequences. Transformer-based models leverage self-attention mechanisms to capture both local and long-range dependencies efficiently, which is critical for interpreting regulatory elements and 3D genomic interactions.

DL has emerged as a powerful approach in NGS data analysis, addressing key limitations of traditional machine learning by automatically extracting hierarchical representations from raw, high-dimensional data [68]. Neural network architectures eliminate the need for manual feature engineering and exhibit exceptional capabilities in modeling complex, nonlinear biological interactions. CNNs have become fundamental in sequence-based analyses due to their ability to capture local sequence motifs [69,70]. Early applications demonstrated CNNs’ effectiveness in predicting transcription factor binding sites directly from DNA sequences [71,72]. In clinical genomics, CNN-based variant calling models have substantially improved performance by interpreting sequencing reads as image-like data, exhibiting notably higher accuracy than traditional heuristic methods [73]. Furthermore, CNN architectures have been utilized to identify the functional effects of non-coding variants by predicting chromatin accessibility across hundreds of cell types and annotating disease-associated SNPs [74].

RNNs, particularly LSTMs, have been employed to genomic sequence data analysis due to their ability to capture sequential dependencies and long-range interactions [75]. RNNs are designed to capture time-series data or sequence information, making them powerful tools for multiple genomic and transcriptomic tasks that include chromatin accessibility analysis [76], identification of RNA splicing events [75], variant interpretation [77], and metagenomic analysis [78]. LSTMs, a type of RNN, have been specifically designed to address the issue of vanishing gradients by maintaining long-term dependencies, which is critical when analyzing long genomic sequences. LSTM models have been successfully employed to predict splicing alterations in RNA sequences, enabling the identification of pathogenic variants in rare diseases with remarkable accuracy [79]. Bidirectional LSTMs (BLSTMs) model long-range dependencies in nucleotide sequences by processing them in both forward and reverse directions using gating mechanisms that retain contextual memory [77]. They are widely used in tasks such as transcription factor binding prediction and variant calling, often in combination with CNNs, as exemplified by the DAVI model [70,80].

Beyond basic sequence modeling, RNNs with attention mechanisms have further improved the interpretability and performance of multiple models. Attention mechanisms allow the model to focus on the most relevant parts of the sequence, making it more

efficient at predicting complex biological phenomena. One noteworthy application is the prediction of gene expression based on histone modification patterns, where RNNs with attention mechanisms outperformed traditional classifiers in terms of AUC (area under the curve) [81]. The attention-based approach helps identify key regulatory regions that are crucial for understanding gene regulation, providing a more refined model for genomic analysis.

3.3. Hybrid and Ensemble Approaches

Hybrid and ensemble methods address NGS challenges by strategically combining algorithms or data modalities [82]. In the context of genomics, hybrid models often integrate ML and DL methods to manage the complexity of high-dimensional data and overcome drawbacks inherent in single-algorithm models. Hybrid CNN-RNN approaches are employed for sequence-based tasks where both local motifs and long-range dependencies are important, such as variant calling and splicing prediction. Indicatively, Clair3 [83] and PEPPER-Margin-DeepVariant [84] represent hybrid architectures by combining CNN-based feature extraction with sequence modeling for long-read variant detection. Ensemble methods, on the other hand, aggregate predictions from multiple base learners—such as decision trees, support vector machines, or neural networks—using techniques like bagging, boosting, or stacking. Ensemble techniques like XGBoost-RF hybrids (e.g., in ClinPred [85]) integrate tree-based models to improve pathogenic variant classification, outperforming single-model approaches in clinical datasets. These ensemble models significantly outperform individual algorithms in tasks like missense variant interpretation and cancer gene prioritization. Although these approaches increase computational complexity, they provide a pathway towards more accurate and generalizable AI models, which are essential for the efficient clinical translation of genomic analyses.

3.4. Deploying AI Within Clinical NGS Workflows

The integration of AI into clinical NGS workflows presents several challenges. Data heterogeneity, including batch effects and class imbalance in rare disease datasets, significantly reduces model accuracy and reliability [86]. Moreover, the substantial computational requirements—such as high GPU memory and large-scale storage for whole-genome data—pose logistical constraints. Reproducibility remains a persistent limitation, with many models exhibiting reduced performance upon external validation. Variability across sequencing platforms further compromises generalizability. Emerging solutions like federated learning offer improved cross-institutional performance, while cloud-based APIs help reduce local computational demands. Continuous learning systems that incorporate clinician feedback enhance adaptability, though they address only part of the deployment gap. Ultimately, aligning research with clinical application will require advances in model transparency, standardization, and infrastructure.

4. AI in NGS Applications

The integration of AI into NGS applications is revolutionizing the landscape of genomics, epigenomics, transcriptomics, and clinical diagnostics (Figure 2). By leveraging advanced techniques such as NNs and language models, AI enables the classification of raw sequencing data with unprecedented precision. This allows for more accurate variant detection, comprehensive functional annotation of complex genomic regions, as well as the discovery of robust biomarkers. In oncology, AI-driven NGS analysis can efficiently contribute to tumor subtyping and therapy selection, while in the context of rare diseases, ML enhances the prioritization of pathogenic variants from whole-genome sequencing data. Furthermore, the fusion of multi-omics data strengthens predictive models, ad-

vancing personalized medicine by offering deeper insights into disease mechanisms and treatment responses.

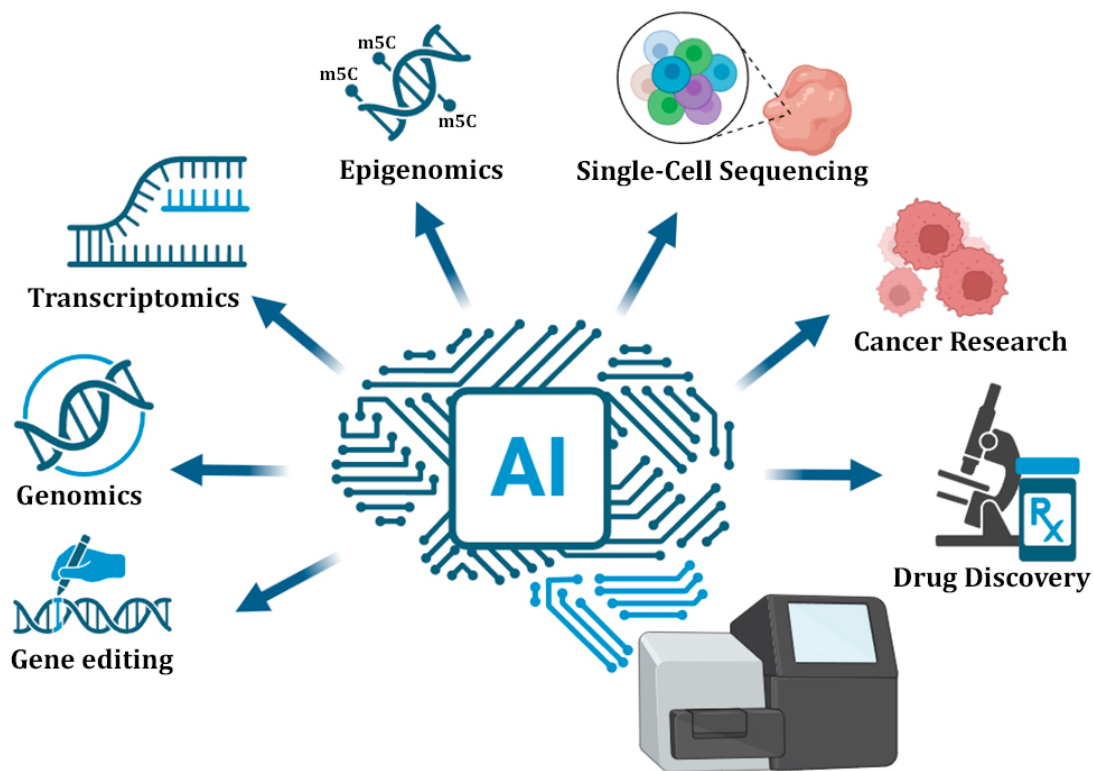


Figure 2. The integration of AI into NGS applications for both research and clinical purposes. AI has revolutionized the landscape of gene editing approaches, genomics, epigenomics, transcriptomics, as well as clinical applications, including cancer research and drug discoveries.

4.1. Genomics and Epigenomics

In the last few years, AI-powered genomic analysis has become indispensable for interpreting the functional impact of genetic variants and enabling precision medicine. Firstly, the process of identifying genetic variants from NGS datasets has been significantly enhanced by AI-based tools, since they offer improved accuracy, efficiency, and scalability as compared to traditional statistical methods. Two noteworthy DL models, DeepVariant [22] and DeepFilter [87], apply CNNs to distinguish mutations from sequencing artifacts in challenging genomic regions, including homopolymers and segmental duplications, significantly improving accuracy. DeepVariant reduces false positives by analyzing sequencing reads as pileup images, while DeepFilter enhances the precision of VarDict calls by postprocessing labeled variant data. For pathogenicity prediction, ensemble methods integrate evolutionary conservations, protein structures, and clinical annotations. Another alternative and very recent approach for variant calling is DNAscope, which combines GATK's HaplotypeCaller with an AI-based genotyping model for efficient processing of large datasets [88]. Additionally, frameworks such as REVEL aggregate outputs from multiple individual predictors using RFs, achieving high accuracy for deleterious missense variants [89]. Recent transformer-based models like DNABERT capture long-range dependencies to interpret non-coding variants, identifying pathogenic regulatory mutations in previously undiagnosed rare disease cases [90].

In addition to genomics, AI has radically transformed epigenomic analysis by enabling accurate and scalable prediction of DNA and/or RNA methylation patterns, chromatin accessibility, and histone modification dynamics from sequencing data. In the context of chromatin state analysis, tools like ChromHMM use histone mark profiles and hidden

Markov models to segment the genome into active (euchromatin) and repressive (heterochromatin) regions across various cell types [91,92]. More recently, DL frameworks such as EPInformer have leveraged DL to predict chromatin accessibility from DNA sequences, enabling insights into chromatin state dynamics [93]. Additionally, DeepChrome and DeepDiff use CNNs and LSTMs, accordingly, to integrate histone modification data and predict gene expression, offering indirect inference of chromatin openness [94,95].

DL models leverage CNNs and RNNs to capture sequence context and long-range dependencies, outperforming traditional statistical methods and thus enhancing the identification accuracy of CpG methylation at single-base resolution [96,97]. Of note, the AI-driven DNA methylation tool DeepMethyl [98] analyzes whole-genome bisulfite sequencing (WGBS) data to identify differentially methylated regions (DMRs) with increased sensitivity in cancer epigenomes. In addition, MethylNet and MethylSPWNet further extend this capability by incorporating multi-layer perceptrons and Bayesian priors for improved generalization across sample types [99,100]. It should be noted that among the most recent advancements are MethylGPT, a foundation model for the DNA methylome, and DiffuCpG, a cutting-edge generative AI model designed to address missing data in high-throughput methylation studies. MethylGPT stands out for its ability to integrate complex DNA methylation data, while DiffuCpG is particularly effective in leveraging both short- and long-range interactions, including three-dimensional genome architecture, to improve dataset accuracy, scalability, and versatility [96,97]. Both models represent novel strides in epigenetic research, with DiffuCpG demonstrating substantial improvements across multiple tissue types, cancers, and sequencing technologies (Table 2).

Table 2. List of AI-based tools designed for NGS and TGS applications.

NGS/TGS Application	Tool/Platform	Functionality	Algorithm Type
Variant Calling	DeepVariant	SNP and indel calling	CNN
	DNAScope	High-accuracy genotyping	ML
	Clair3, Clairvoyante	Long-read variant calling	CNN
	Clair3-RNA	Small variant caller for long-read RNA sequencing data	DL
	PEPPER-Margin-DeepVariant	Alignment and consensus-based variant calling	DL
	ClinPred	Pathogenicity prediction for missense variants	Ensemble (XGBoost + RF)
	REVEL	Aggregated predictions for the pathogenicity of variant effects	Ensemble
Splicing Prediction	DeepFilter	Variant call filtering	DL
	SpliceAI	Splice-disruptive variants prediction	CNN
	Splam	Splice junctions in DNA prediction	DL/CNN
Transcriptomics	Bambu	Transcript discovery and quantification from long-read RNA-Seq data	ML
	Deep Count Autoencoder (DCA)	Gene expression denoising	DL
	scVI/scVAE	Batch correction, embedding, differential expression	DL
	MultiPLIER	Interpretable feature extraction across large transcriptomic cohorts	ML
	Cox-nnet	Patient prognosis prediction from high-throughput RNA-Seq data	ANN
	BABEL	Cross-modality translation between multi-omic profiles	DL
Genomics	DNABERT	Genome-wide prediction of promoters, splice sites, and transcription factor binding sites	Bidirectional encoder representation from transformers

Table 2. Cont.

NGS/TGS Application	Tool/Platform	Functionality	Algorithm Type
Methylation Analysis	MethylNet	DNA methylation prediction	DL
	MethylSPWNet	Classification of CpGs into biologically relevant capsules	DL
	DeepMethyl	CpG methylation prediction	DL
	DiffuCpG	Methylation imputation	DL
	MethylGPT	Methylation value prediction	DL
Chromatin Accessibility	Basset	Prediction of accessible chromatin regions	CNN
	Enformer	Prediction of variant effects on gene expression	DL
	scATAC-pro	Quality assessment, analysis, and visualization of single-cell chromatin accessibility sequencing data	VAE
Histone Modifications	DeepHistone	Histone modification patterns prediction	NN/DL
	dHICA	Histone mark imputation and prediction of their modifications	DL
Single-Cell Analysis	CopyKAT	CNV inference from scRNA-seq	Integrative Bayesian segmentation approach
	scGraphformer	Unveiling cellular heterogeneity and interactions in scRNA-seq data	Transformer-based GNN
	TotalVI	Multi-modal data analysis/joint analysis of CITE-seq data	VAE
	scANVI	Cell state/transcriptomics data annotation	DL
	scShaper	Accurate linear trajectory inference	Ensemble
3D Genome Structure	ChromoGen	Single-cell chromatin conformation modeling	Generative model
Cancer Research	PyClone-VI	Inference of clonal population structures using whole genome data	Bayesian statistical method
	HoneyBADGER	Identification of CNVs and heterozygosity loss in individual cells from single-cell RNA-Seq	HMM-integrated Bayesian hierarchical model
	CUPLR	Tissue of origin classification for cancer of unknown primary diagnostics	ML
	CUP-AI-DX	Inference of cancer tissue of origin and molecular subtyping using gene expression data	CNN
Drug Discovery	DeepDR	Drug repurposing	DL
	DeepPurpose	Drug–target interaction prediction	DL
	MolTrans	Drug–target interaction prediction	DL
	MatchMaker	Drug synergy prediction	DL
Clinical Diagnostics	AlphaMissense	Variant pathogenicity prediction	ML
Multi-Omics Integration	OmiEmbed	Multi-omics data analysis	DL
	MOGONET	Patient classification and biomarker identification	GNN

Chromatin accessibility prediction has progressed significantly with DL models such as Basset [101], which applies CNNs to associate DNA sequence motifs with open chromatin regions from ATAC-seq and DNase-seq data. More recently, transformer-based models, such as Enformer, have outperformed earlier architectures by capturing long-range genomic interactions, enhancing the prediction of distal enhancers and regulatory elements [102]. In single-cell epigenomics, tools like scATAC-pro employ variational autoencoders and probabilistic graphical models to resolve cell-type-specific accessibility in complex tissues [103] (Table 2). Moreover, scGraphformer represents a graph-transformer-based model that captures both local chromatin accessibility and global regulatory interac-

tions in single-cell ATAC-seq data, enabling accurate cell-type annotation and regulatory network reconstruction [104].

Regarding histone modification analysis, a recent tool called DeepHistone integrates ChIP-seq profiles and DNA sequence features to predict the genome-wide distribution of canonical histone marks [105]. Recently, the transformer-based model dHICA introduced another robust framework for accurate histone mark imputation from chromatin accessibility data, demonstrating how AI can infer missing epigenomic layers and uncover gene regulatory relationships [106].

At the level of 3D genome organization, generative models are emerging as powerful tools. ChromoGen, a diffusion-based model, predicts single-cell chromatin conformations by learning from paired chromatin accessibility and genomic contact data, addressing variability and sparsity in single-cell 3D genome measurements [107] (Table 2). A broader review of DL approaches also highlights AI's growing role in decoding chromatin interactions and integrating diverse epigenomic profiles for comprehensive regulatory inference [108]. Collectively, these advancements demonstrate how AI-driven tools are reshaping our understanding of the chromatin landscape at both bulk and single-cell resolution [109].

4.2. Transcriptomics

AI has already been widely integrated in transcriptomic analysis by enabling precise identification of gene expression patterns and alternative splicing events from RNA-Seq data, offering high-resolution snapshots of mRNA exon–intron boundaries and expression. In addition to the fact that up until today, noise, batch effects, and complex experimental designs often hinder accurate interpretation, DL models have provided innovative solutions across the transcriptomics workflow—from normalization and feature selection to differential expression analysis (DEA) and splicing prediction [110]. CNNs outperform traditional statistical methods in detecting subtle expression changes, particularly in low-abundance transcripts. A newly introduced tool, deep count autoencoder (DCA) applies deep autoencoder networks to denoise and compress gene expression data, improving downstream analyses and visualization [111]. AI models, particularly DL architectures, have shown remarkable success in cell type and subtype identification based on scRNA-seq data. Both scVI and scVAE represent two recent variational autoencoders capable of correcting batch effects, while preserving biological variation, enabling large-scale atlas integration, unsupervised clustering, trajectory inference, and DEA with minimal manual intervention [112–114] (Table 2). Additionally, numerous CNN-based tools and/or frameworks have been developed to predict the functional impact of non-coding variants on gene expression and chromatin features, including DeepSEA [72], Basset [101], and DeepFIGV [115].

The limited accuracy of short-read alignment is overcome by graph neural networks (GNNs), which are effective in modeling transcriptome complexity, enabling improved reconstruction and quantification. Tools such as Bambu utilize splice graph representations to reconstruct full-length transcripts with high precision, significantly outperforming traditional linear models in capturing isoform diversity [116]. For alternative splicing analysis, SpliceAI [117–119] and Splam employ DL models trained directly on raw genomic sequences to predict branchpoints and splice variants directly from sequence context with exceptional performance [120].

4.3. Single-Cell Sequencing

AI plays a pivotal role in advancing the field of single-cell sequencing analysis, addressing the unique challenges associated with this approach. ML techniques are extensively used for various aspects of single-cell sequencing data, including imputing missing gene

expression values often caused by technical noise, removing batch effects that arise from processing samples at different times or in different laboratories, identifying distinct cell types and their states, estimating copy number variations within individual cells, inferring cellular differentiation trajectories, analyzing cell–cell interactions, and reconstructing regulatory networks [121]. These advances are driven by three key applications: cell type identification, trajectory inference, and heterogeneity analysis.

Cell type identification benefits from NNs that integrate gene expression with epigenetic data. scANVI, a semi-supervised extension of the variational autoencoder framework, enables precise annotation of rare and ambiguous cell types with noteworthy precision, even in minimally labeled datasets. Multi-modal models like TotalVI, which utilize CITE-seq data, can further improve classification accuracy, particularly in immune profiling tasks [122] (Table 2). Trajectory inference has advanced through neural ordinary differential equations (ODEs) and generative models [123]. Deep generative models incorporating neural ODEs reconstruct branching and cyclic trajectories, capturing progenitor states previously undetected in a subset of datasets [124,125]. Among the analysis solutions AI can offer today, scShaper offers benchmarking for many trajectory inference algorithms [126], while CopyKAT infers large-scale copy number alterations and malignant clones from scRNA-seq with high concordance to DNA sequencing [127] (Table 2).

4.4. Cancer Research

AI-based NGS analysis is transforming cancer research by enabling precise molecular subtyping, biomarker discovery, and personalized treatment strategies [128]. DL models now classify tumor subtypes with high accuracy using whole exome sequencing data, outperforming traditional pathology in multicenter validations [129]. Tumor heterogeneity analysis particularly benefits from AI tools that resolve clonal architecture from bulk and single-cell sequencing [130]. PyClone-VI, which employs variational Bayesian inference, detects resistant clones present at extremely low frequencies [131] (Table 2). Graph neural networks have been applied to reconstruct phylogenetic trees of tumor progression, showing high concordance with longitudinal clinical data and revealing metastatic drivers in a substantial subset of analyzed cases [132]. In the single-cell context, HoneyBADGER allows simultaneous analysis of copy number variation and gene expression to identify genetic subclones with distinct transcriptional programs, as demonstrated in progressive multiple myeloma [133]. While HoneyBADGER does not utilize DL or neural networks, it integrates probabilistic modeling and statistical inference—core components of AI methodologies—to interpret complex biological data.

Biomarker discovery leverages multi-modal integration of genomic, transcriptomic, and proteomic data. In immuno-oncology, the TIDE framework analyzes T-cell dysfunction and exclusion signatures to stratify immune checkpoint inhibitor responders [134]. CUPLR [135] and CUP-AI-Dx [136] further improve tissue-of-origin classification and microenvironment profiling, particularly in rare cancers. Multimodal AI models leverage variational autoencoders to jointly model gene expression, protein abundance, and methylation status, extracting latent representations predictive of clinical outcomes [137]. Similarly, MultiPLIER expands the PLIER framework to enable interpretable feature extraction across large transcriptomic cohorts like The Cancer Genome Atlas (TCGA) [138]. Another tool that should be mentioned is BABEL, which enables cross-modality translation by predicting gene expression from chromatin accessibility or protein abundance, allowing integrated analysis across scRNA-seq, ATAC-seq, and CITE-seq modalities [139].

Pharmacogenomic models that match tumor-specific mutations with drug sensitivity profiles have improved response rates [140]. Predictive platforms like Cox-nnet integrate neural networks with Cox regression to prioritize prognostic markers, improving five-year

survival prediction [141]. SNP-informed pharmacogenomic tools are increasingly used in clinical oncology to tailor drug dosing and minimize toxicity risk [142]. Furthermore, liquid biopsy analysis using DL techniques now enables the detection of circulating tumor DNA (ctDNA) at extremely low allele frequencies, enabling early diagnosis and real-time monitoring of disease progression [143]. AI-enhanced tumor mutational burden (TMB) calculators have also demonstrated improved stratification of patients likely to benefit from immune checkpoint blockade [144,145].

4.5. Drug Discovery

AI has accelerated the process of identifying drug targets, repurposing existing drugs, and designing personalized therapies by leveraging high-throughput sequencing data. DL models now predict drug–target interactions with unprecedented accuracy [146], significantly accelerating preclinical development. Interestingly, transformer-based frameworks like MolTrans predict binding affinities directly from sequence and structural representations [147] (Table 2). These approaches have successfully identified previously overlooked targets when applied to cancer genomics datasets.

Drug repurposing has particularly benefited from AI integration in NGS data analysis. For instance, the DeepDR platform combines multiple drug-related networks—including transcriptomics, chemical structure, and disease association data—to identify candidates for existing medications, outperforming traditional repurposing strategies [148]. Another recent tool, MatchMaker, incorporates gene expression and chemical fingerprint data to identify cross-indication therapeutic candidates [149]. In precision oncology and rare disease contexts, pharmacogenomic models built from whole-exome sequencing can now predict optimal drug combinations with high accuracy. Platforms like DeepPurpose employ DL models to match genetic variants to compound libraries, expanding personalized treatment options [150] (Table 2). Furthermore, real-time NGS integration in adaptive clinical trials is now guiding therapy selection dynamically based on tumor evolution, improving patient stratification and trial efficiency [151].

5. AI-Driven Multi-Omics Integration and Clinical Translation

AI is becoming essential for leveraging multi-omics data and translating insights into clinical applications. By integrating genomics, transcriptomics, epigenomics, proteomics, and metabolomics, AI models uncover biological insights that transcend the limitations of single-omics approaches. Tools like OmiEmbed apply DL frameworks to capture shared biological variation across multiple omics layers, enabling a more comprehensive understanding of disease mechanisms [152].

Multi-omics integration strategies typically fall into three categories: early integration, where features from different omics are concatenated into a single input matrix; intermediate integration, which uses modality-specific encoders and combines latent representations; and late integration, where predictions from individual omics models are fused via ensemble methods. Among these, intermediate integration has become the most widely adopted in DL-based frameworks due to its flexibility and improved performance in heterogeneous datasets. Graph-based approaches have demonstrated strong performance in this space [153,154]. For example, MOGONET employs graph neural networks (GNNs) to learn patient similarity networks within each omics modality [155]. These graphs are then integrated through an attention-based mechanism that enhances classification and biomarker identification, particularly in cancer subtyping. Similarly, the integrative graph convolutional network (IGCN) builds sample-level graphs from high-dimensional input and applies convolutional operations to extract cross-omic interactions and shared regulatory patterns, facilitating accurate disease subtype discovery and survival prediction [156,157].

Despite these promising advances, several limitations remain. Multi-omics datasets often suffer from missing data, modality imbalance, and batch effects, which can distort model training. Additionally, the computational demands of training deep integration frameworks on high-dimensional inputs are substantial. Interpretability also poses a challenge, as black-box architectures may obscure biological mechanisms underlying predictions. Furthermore, the lack of standardized benchmarking datasets hinders rigorous comparisons between integration models, particularly in clinical contexts.

In clinical contexts, AI models are now central to personalized pipelines medicine. Platforms such as PathAI combine genomic, transcriptomic, histopathological, and clinical data to assist in disease diagnosis, predict patient outcomes, and suggest tailored therapeutic strategies, thereby accelerating data-to-decision workflows in precision oncology [158]. A major leap in variant interpretation is represented by DeepMind's AlphaMissense, a ML model trained to predict the pathogenicity of missense mutations in human genomes [159] (Table 2). By analyzing protein structure and evolutionary data, the model assesses whether a given amino acid substitution is likely to cause disease, thereby providing functional annotations for variants of uncertain significance [160]. Despite these advances, translating AI tools into clinical practice presents significant challenges. Clinical validation requires extensive wet-lab confirmation, external dataset benchmarking, and diverse cohort representation to ensure model robustness. Regulatory approval remains complex, especially for "black-box" models where explainability and reproducibility are essential for trust and compliance with standards like FDA or CE marking. Furthermore, institutional inertia, integration with existing electronic health records (EHRs), and computational infrastructure gaps can delay real-world deployment. Nonetheless, tools like DeepVariant are being adopted in clinical diagnostic labs, and federated learning models are currently undergoing pilot testing in multi-center hospital networks, suggesting that regulatory and infrastructural paths are slowly being paved for AI-driven personalized medicine.

6. Challenges and Limitations

The integration of AI, ML, and DL into NGS analysis holds great promise, yet several challenges must be addressed to fully realize its transformative potential. First, the quality and quantity of sequencing data remain critical concerns. Technical artifacts, such as noise, amplification bias, and platform variability, can significantly affect model performance and reproducibility [121,161]. Moreover, limited sample sizes and class imbalance, particularly common in rare diseases and specific clinical cohorts, reduce model robustness and generalizability [157].

A major challenge lies in the interpretability of complex "black-box" models. Despite high predictive accuracy, the biological meaning of many AI outputs remains unclear. This lack of transparency hinders clinical adoption [162]. Explainable AI methods—such as SHAP analysis—are increasingly necessary to trace predictions back to interpretable molecular features [163]. Another significant issue is cross-platform variability. AI models trained on one sequencing technology often underperform when employed to datasets generated from other platforms [164]. While sequencing platform variability challenges model generalizability, the Protein Data Bank (PDB) provides a valuable example of successful cross-platform integration, standardizing protein structures resolved by X-ray crystallography, NMR, and cryo-EM through universal coordinate systems and file formats [165,166]. In contrast, a unified repository for DNA 3D structure remains lacking: although projects like the 4DN Data Portal compile uniformly processed datasets from Hi-C and related assays [167], and resources such as GSDB and 3DGD provide reconstructed 3D models from Hi-C, there is still no widely adopted standard file format or coordinate system for DNA architectures [168,169]. This absence of universal representation hinders AI model

interoperability and cross-study validation and highlights the need for community-driven frameworks to support DNA structural genomics.

Data privacy and ethical concerns are of utmost importance in clinical and cross-institutional settings. Federated learning frameworks allow collaborative AI model development without centralized data sharing, thus preserving patient privacy while enhancing training data diversity [170]. These models are increasingly used in diagnostics and drug discovery to overcome institutional barriers to data access. Clinical translation of AI models remains an ongoing challenge. Even highly accurate predictions often require extensive wet-lab validation. Interpretability is crucial for gaining approval from regulatory authorities, especially in diagnostics [171]. Furthermore, population biases in training datasets can distort outcomes and limit applicability. Multimodal integration frameworks—combining methylation, transcriptomics, and chromatin accessibility—are helping address cell-type heterogeneity and improve signal resolution. Meanwhile, transformer-based models enable unified analysis across multiple omic layers, though clinical readiness requires more rigorous benchmarking [172,173]. In drug discovery, limited training data for novel or rare targets continues to constrain performance. Generative AI models show promise in designing new compounds from multi-omic profiles, and some AI-guided therapies are now entering clinical trials [174].

In summary, while AI has made remarkable progress critical challenges remain in data quality, interpretability, clinical validation, and privacy [175]. Continued collaboration between computational scientists, biologists, and clinicians will be key to navigating this evolving landscape.

7. Future Perspectives—AI Integration into Third-Generation Sequencing

Although NGS altered genomic research, it still presents numerous limitations due to the generation of short reads, which are surmounted with the advent of TGS. The revolutionary long-read technologies introduced by Pacific Biosciences® (PacBio) and Oxford Nanopore Technologies® (ONT) have transformed genome and transcriptome sequencing. TGS enables single-molecule, real-time sequencing and has introduced novel chemistries that produce long reads averaging 10 kb, compared to NGS's 600 nt, thus allowing better analysis of complex genomic regions and improving the quality of the sequencing results [176]. By omitting the need for PCR amplification, TGS reduces biases that complicate computational analysis, allowing for faster, simpler, and more accurate decoding of nucleic acids [177]. Nevertheless, persistent challenges—such as high error rates and complex data interpretation—remain for both TGS and NGS. In this context, AI has become essential in addressing these limitations and enhancing data analysis.

AI was initially applied to the challenging yet crucial process of basecalling, enhancing the accuracy of converting raw signal data into nucleotide sequences. Notably, while early ONT basecallers (i.e., Metrichor) relied on hidden Markov models, later tools like Nanonet and DeepNano adopted recurrent neural networks [178,179]. The research basecaller Bonito combines CNNs with connectionist temporal classification (CTC) decoders [180], while the main ONT basecaller, Guppy, integrates CNNs with conditional random field (CRF) decoders [181]. In addition to ONT, AI has been also introduced to PacBio technology as errors in HiFi reads are reduced using DeepConsensus—a gap-aware sequence transformer encoder that produces high-accuracy consensus sequences [182].

While alignment of TGS data typically does not use AI approaches, variant calling increasingly relies on them. DL-based tools are gradually replacing traditional statistical methods to enhance the accuracy of long-read variant identification. Models like MAMnet and BreakNet, both CNN-based, predict indels from long-read data, while SVision is preferred for studying structural variants [183]. Other DL-based variant callers include

Clairvoyante, Clair, and Nanocaller (pileup-based); PEPPER-Margin-DeepVariant (full alignment-based); Medaka (consensus-based); Clair3, which combines pileup and full-alignment algorithms [83]; and Clair3-RNA, a small variant caller for long-read RNA sequencing data for both ONT and PacBio platforms [184].

Of note, TGS is increasingly used to detect epigenetic modifications in DNA and RNA, offering many advantages over NGS. Recently, AI approaches have been developed to interpret the generated raw signal data that are crucial for this type of analysis. PacBio® uses interpulse durations (IPDs) and pulse widths to infer DNA base modifications, while ONT detects modified bases through shifts in electrical current as DNA or RNA translocates through a nanopore. AI methods for identifying DNA modifications include Nanopolish, which uses HMMs to detect m5C; DeepSignal, which applies CNNs to detect m5C and m6A; and mCaller, a neural network classifier for identifying 6mA [180,185–187]. Additionally, SignalAlign combines HMMs with hierarchical Dirichlet process (HDP) models to correlate ionic current shifts with DNA modifications (m5C, hm5C, m6A) [188]. ONT's direct RNA sequencing bypasses reverse transcription, enabling real-time detection of RNA modifications—an approach not possible with NGS. To analyze RNA methylation, MINES, a random forest classifier, assigns m6A to DRACH motifs [189], while EpiNano and Nanom6A focus on RRACH motifs [190,191]. Other tools, such as Nanocompare, use statistical models and ML for modified RNA base detection [192]; nanoDoc combines unsupervised DL with CNNs for RNA post-transcriptional modification analysis [193], and xPore, based on Bayesian deep learning, identifies differential RNA modification sites [194].

8. Conclusions

The integration of AI into NGS workflows has revolutionized genomics by enabling faster, more accurate, and scalable analysis of complex biological data. Through both ML and DL methods, AI facilitates tasks ranging from basecalling and variant calling to multi-omics integration, cancer research, and drug discovery. Notably, AI-based tools have demonstrated superior performance in detecting rare variants, reconstructing full-length transcripts, interpreting non-coding regions, and predicting treatment outcomes, especially in cancer genomics and single-cell studies. Ensemble and hybrid models further enhance prediction accuracy and generalizability by leveraging complementary algorithmic strengths. However, significant challenges remain in model interpretability, data heterogeneity, clinical validation, and privacy compliance. The success of AI applications in genomics depends not only on technical advancements but also on fostering interdisciplinary collaborations between data scientists, biologists, and clinicians. As long-read sequencing technologies and multi-modal datasets continue to evolve, AI will be pivotal in translating genomic insights into clinical practice. Future directions should prioritize the implementation of federated learning frameworks to preserve data privacy in cross-institutional AI training, the development of explainable AI models to enhance clinical trust and regulatory approval, and the seamless integration of TGS data to improve variant detection and epigenetic profiling. Robust benchmarking, regulatory validation pipelines, and interdisciplinary collaboration will be key to ensuring trustworthy AI deployment in diagnostic and therapeutic settings. Ultimately, AI promises to usher in a new era of precision medicine that is data-driven, individualized, and scalable.

Author Contributions: K.A. and V.-I.M. drafted the text of the present review article; P.G.A. provided additional text and important discussions for the enhancement of the text. A.S. and P.G.A. performed a critical review of the manuscript, providing significant corrections and additional text. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial intelligence
ANN	Artificial neural network
ATAC-seq	Assay for transposase-accessible chromatin using sequencing
AUC	Area under the curve
CITE-seq	Cellular indexing of transcriptomes and epitopes by sequencing
CNN	Convolutional neural network
CNV	Copy number variation
CRF	Conditional random field
CTC	Connectionist temporal classification
ctDNA	Circulating tumor DNA
DL	Deep learning
DEA	Differential expression analysis
DMR	Differentially methylated region
EMBL-EBI	European Bioinformatics Institute
GATK	Genome Analysis Toolkit
GBM	Gradient boosting machine
GEO	Gene Expression Omnibus
GNN	Graph neural network
GRN	Gene regulatory network
HDP	Hierarchical Dirichlet process
HGP	Human Genome Project
HMM	Hidden Markov model
IGCN	Integrative graph convolutional network
IPD	Interpulse duration
LSTM	Long short-term memory
ML	Machine learning
m5C	5-methylcytosine
m6A	N6-methyladenosine
hm5C	5-hydroxymethylcytosine
ML	Machine learning
mRNA	Messenger RNA
NGS	Next-generation sequencing
ONT	Oxford Nanopore Technologies
PCA	Principal component analysis
PCR	Polymerase chain reaction
RF	Random forest
RNN	Recurrent neural network
scRNA-seq	Single-cell RNA sequencing
SdAs	Stacked denoising autoencoders
SNP	Single nucleotide polymorphism

SVM	Support vector machine
TCGA	The Cancer Genome Atlas
TGS	Third-generation sequencing
t-SNE	t-distributed stochastic neighbor embedding
UMAP	Uniform manifold approximation and projection
VAE	Variational autoencoder
WGBS	Whole-genome bisulfite sequencing

References

1. Watson, J.D.; Crick, F.H. The structure of DNA. *Cold Spring Harb. Symp. Quant. Biol.* **1953**, *18*, 123–131. [[CrossRef](#)] [[PubMed](#)]
2. Crick, F. Central dogma of molecular biology. *Nature* **1970**, *227*, 561–563. [[CrossRef](#)] [[PubMed](#)]
3. Collins, F.S.; Morgan, M.; Patrinos, A. The human genome project: Lessons from large-scale biology. *Science* **2003**, *300*, 286–290. [[CrossRef](#)] [[PubMed](#)]
4. Green, E.D.; Watson, J.D.; Collins, F.S. Human genome project: Twenty-five years of big biology. *Nature* **2015**, *526*, 29–31. [[CrossRef](#)]
5. Giani, A.M.; Gallo, G.R.; Gianfranceschi, L.; Formenti, G. Long walk to genomics: History and current approaches to genome sequencing and assembly. *Comput. Struct. Biotechnol. J.* **2020**, *18*, 9–19. [[CrossRef](#)]
6. Caudai, C.; Galizia, A.; Geraci, F.; Le Pera, L.; Morea, V.; Salerno, E.; Via, A.; Colombo, T. Ai applications in functional genomics. *Comput. Struct. Biotechnol. J.* **2021**, *19*, 5762–5790. [[CrossRef](#)]
7. Dixit, S.; Kumar, A.; Srinivasan, K.; Vincent, P.; Ramu Krishnan, N. Advancing genome editing with artificial intelligence: Opportunities, challenges, and future directions. *Front. Bioeng. Biotechnol.* **2023**, *11*, 1335901. [[CrossRef](#)]
8. Koh, E.; Sunil, R.S.; Lam, H.Y.I.; Mutwil, M. Confronting the data deluge: How artificial intelligence can be used in the study of plant stress. *Comput. Struct. Biotechnol. J.* **2024**, *23*, 3454–3466. [[CrossRef](#)]
9. Farhud, D.D.; Zokaei, S. Ethical issues of artificial intelligence in medicine and healthcare. *Iran. J. Public Health* **2021**, *50*, i–v. [[CrossRef](#)]
10. Elendu, C.; Amaechi, D.C.; Elendu, T.C.; Jingwa, K.A.; Okoye, O.K.; John Okah, M.; Ladele, J.A.; Farah, A.H.; Alimi, H.A. Ethical implications of ai and robotics in healthcare: A review. *Medicine* **2023**, *102*, e36671. [[CrossRef](#)]
11. Serrano, D.R.; Luciano, F.C.; Anaya, B.J.; Ongoren, B.; Kara, A.; Molina, G.; Ramirez, B.I.; Sanchez-Guirales, S.A.; Simon, J.A.; Tomietto, G.; et al. Artificial intelligence (ai) applications in drug discovery and drug delivery: Revolutionizing personalized medicine. *Pharmaceutics* **2024**, *16*, 1328. [[CrossRef](#)] [[PubMed](#)]
12. Jamialahmadi, H.; Khalili-Tanha, G.; Nazari, E.; Rezaei-Tavirani, M. Artificial intelligence and bioinformatics: A journey from traditional techniques to smart approaches. *Gastroenterol. Hepatol. Bed Bench* **2024**, *17*, 241–252.
13. Harrer, S.; Rane, R.V.; Speight, R.E. Generative ai agents are transforming biology research: High resolution functional genome annotation for multiscale understanding of life. *EBioMedicine* **2024**, *109*, 105446. [[CrossRef](#)]
14. Ghose, A.K.; Abdullah, S.N.A.; Md Hatta, M.A.; Megat Wahab, P.E. DNA free crispr/dcas9 based transcriptional activation system for ugt76g1 gene in stevia rebaudiana bertonii protoplasts. *Plants* **2022**, *11*, 2393. [[CrossRef](#)]
15. Yuan, Y.; Shi, Y.; Li, C.; Kim, J.; Cai, W.; Han, Z.; Feng, D.D. Deepgene: An advanced cancer type classifier based on deep learning and somatic point mutations. *BMC Bioinform.* **2016**, *17*, 476. [[CrossRef](#)]
16. Tegally, H.; San, J.E.; Giandhari, J.; de Oliveira, T. Unlocking the efficiency of genomics laboratories with robotic liquid-handling. *BMC Genom.* **2020**, *21*, 729. [[CrossRef](#)]
17. Ng, N.; Gately, R.; Ooi, L. Automated liquid handling for microplate assays: A simplified user interface for the hamilton microlab star. *J. Appl. Bioanal.* **2020**, *7*, 11–18. [[CrossRef](#)]
18. Chuai, G.; Ma, H.; Yan, J.; Chen, M.; Hong, N.; Xue, D.; Zhou, C.; Zhu, C.; Chen, K.; Duan, B.; et al. Deepcrispr: Optimized crispr guide rna design by deep learning. *Genome Biol.* **2018**, *19*, 80. [[CrossRef](#)]
19. Niu, R.; Peng, J.; Zhang, Z.; Shang, X. R-crispr: A deep learning network to predict off-target activities with mismatch, insertion and deletion in crispr-cas9 system. *Genes* **2021**, *12*, 1878. [[CrossRef](#)]
20. Truong, V.; Viken, K.; Geng, Z.; Barkan, S.; Johnson, B.; Ebeling, M.C.; Montezuma, S.R.; Ferrington, D.A.; Dutton, J.R. Automating human induced pluripotent stem cell culture and differentiation of ipsc-derived retinal pigment epithelium for personalized drug testing. *SLAS Technol.* **2021**, *26*, 287–299. [[CrossRef](#)]
21. Khan, S.; Møller, V.; Frandsen, R.; Mansourvar, M. Real-time ai-driven quality control for laboratory automation: A novel computer vision solution for the opentrons ot-2 liquid handling robot. *Appl. Intell.* **2025**, *55*, 524. [[CrossRef](#)]
22. Poplin, R.; Chang, P.C.; Alexander, D.; Schwartz, S.; Colthurst, T.; Ku, A.; Newburger, D.; Dijamco, J.; Nguyen, N.; Afshar, P.T.; et al. A universal snp and small-indel variant caller using deep neural networks. *Nat. Biotechnol.* **2018**, *36*, 983–987. [[CrossRef](#)]

23. Clement, K.; Rees, H.; Canver, M.C.; Gehrke, J.M.; Farouni, R.; Hsu, J.Y.; Cole, M.A.; Liu, D.R.; Joung, J.K.; Bauer, D.E.; et al. Crispresso2 provides accurate and rapid genome editing sequence analysis. *Nat. Biotechnol.* **2019**, *37*, 224–226. [\[CrossRef\]](#)
24. Tsai, S.Q.; Zheng, Z.; Nguyen, N.T.; Liebers, M.; Topkar, V.V.; Thapar, V.; Wyvekens, N.; Khayter, C.; Iafrate, A.J.; Le, L.P.; et al. Guide-seq enables genome-wide profiling of off-target cleavage by crispr-cas nucleases. *Nat. Biotechnol.* **2015**, *33*, 187–197. [\[CrossRef\]](#)
25. Bae, S.; Park, J.; Kim, J.S. Cas-offinder: A fast and versatile algorithm that searches for potential off-target sites of cas9 rna-guided endonucleases. *Bioinformatics* **2014**, *30*, 1473–1475. [\[CrossRef\]](#)
26. Cancellieri, S.; Canver, M.C.; Bombieri, N.; Giugno, R.; Pinello, L. Crispritz: Rapid, high-throughput and variant-aware in silico off-target site identification for crispr genome editing. *Bioinformatics* **2020**, *36*, 2001–2008. [\[CrossRef\]](#)
27. Gao, L.; Yuan, J.; Hong, K.; Ma, N.L.; Liu, S.; Wu, X. Technological advancement spurs komagataella phaffii as a next-generation platform for sustainable biomanufacturing. *Biotechnol. Adv.* **2025**, *82*, 108593. [\[CrossRef\]](#)
28. Clough, E.; Barrett, T.; Wilhite, S.E.; Ledoux, P.; Evangelista, C.; Kim, I.F.; Tomashevsky, M.; Marshall, K.A.; Phillippy, K.H.; Sherman, P.M.; et al. Ncbi geo: Archive for gene expression and epigenomics data sets: 23-year update. *Nucleic Acids Res.* **2024**, *52*, D138–D144. [\[CrossRef\]](#)
29. Thakur, M.; Brooksbank, C.; Finn, R.D.; Firth, H.V.; Foreman, J.; Freeberg, M.; Gurwitz, K.T.; Harrison, M.; Hulcoop, D.; Hunt, S.E.; et al. EMBL's European Bioinformatics Institute (EMBL-EBI) in 2024. *Nucleic Acids Res.* **2025**, *53*, D10–D19. [\[CrossRef\]](#)
30. Sloan, C.A.; Chan, E.T.; Davidson, J.M.; Malladi, V.S.; Strattan, J.S.; Hitz, B.C.; Gabdank, I.; Narayanan, A.K.; Ho, M.; Lee, B.T.; et al. Encode data at the encode portal. *Nucleic Acids Res.* **2016**, *44*, D726–D732. [\[CrossRef\]](#)
31. Tomczak, K.; Czerwinski, P.; Wiznerowicz, M. The cancer genome atlas (TCGA): An immeasurable source of knowledge. *Contemp. Oncol.* **2015**, *19*, A68–A77. [\[CrossRef\]](#)
32. Perrin, H.; Denorme, M.; Grosjean, J.; OMICtools Community; Dynomant, E.; Henry, V.; Pichon, F.; Darmoni, S.; Desfeux, A.; Gonzalez, B. Omictools: A community-driven search engine for biological data analysis. *arXiv* **2017**, arXiv:1707.03659.
33. Li, W.; Xu, H.; Xiao, T.; Cong, L.; Love, M.I.; Zhang, F.; Irizarry, R.A.; Liu, J.S.; Brown, M.; Liu, X.S. Mageck enables robust identification of essential genes from genome-scale crispr/cas9 knockout screens. *Genome Biol.* **2014**, *15*, 554. [\[CrossRef\]](#)
34. Hassani-Pak, K.; Singh, A.; Brandizi, M.; Hearnshaw, J.; Parsons, J.D.; Amberkar, S.; Phillips, A.L.; Doonan, J.H.; Rawlings, C. Knetminer: A comprehensive approach for supporting evidence-based gene discovery and complex trait analysis across species. *Plant Biotechnol. J.* **2021**, *19*, 1670–1678. [\[CrossRef\]](#)
35. Ge, S.X.; Jung, D.; Yao, R. Shinygo: A graphical gene-set enrichment tool for animals and plants. *Bioinformatics* **2020**, *36*, 2628–2629. [\[CrossRef\]](#)
36. Gretnier, A.; Ragueneau, E.; Gong, C.; Prinz, A.; Gansberger, S.; Oyarzun, I.; Hermjakob, H.; Griss, J. ReactomeGSA: New features to simplify public data reuse. *Bioinformatics* **2024**, *40*, btac338. [\[CrossRef\]](#)
37. Zhou, Y.; Zhou, B.; Pache, L.; Chang, M.; Khodabakhshi, A.H.; Tanaseichuk, O.; Benner, C.; Chanda, S.K. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* **2019**, *10*, 1523. [\[CrossRef\]](#)
38. Etcheverry, M.; Moulin-Frier, C.; Oudeyer, P.Y.; Levin, M. AI-driven automated discovery tools reveal diverse behavioral competencies of biological networks. *eLife* **2025**, *13*, RP92683. [\[CrossRef\]](#)
39. Yang, B.; Xu, Y.; Maxwell, A.; Koh, W.; Gong, P.; Zhang, C. Micrat: A novel algorithm for inferring gene regulatory networks using time series gene expression data. *BMC Syst. Biol.* **2018**, *12*, 115. [\[CrossRef\]](#)
40. Pop, M.; Attwood, T.K.; Blake, J.A.; Bourne, P.E.; Conesa, A.; Gaasterland, T.; Hunter, L.; Kingsford, C.; Kohlbacher, O.; Lengauer, T.; et al. Biological databases in the age of generative artificial intelligence. *Bioinform. Adv.* **2025**, *5*, vbaf044. [\[CrossRef\]](#)
41. Peng, Y.; Malin, B.A.; Rousseau, J.F.; Wang, Y.; Xu, Z.; Xu, X.; Weng, C.; Bian, J. From GPT to DeepSeek: Significant gaps remain in realizing AI in healthcare. *J. Biomed. Inform.* **2025**, *163*, 104791. [\[CrossRef\]](#)
42. Luo, R.; Sun, L.; Xia, Y.; Qin, T.; Zhang, S.; Poon, H.; Liu, T.Y. Biogpt: Generative pre-trained transformer for biomedical text generation and mining. *Brief. Bioinform.* **2022**, *23*, bbac409. [\[CrossRef\]](#)
43. Peng, Y.; Chen, Q.; Shih, G. Deepseek is open-access and the next AI disrupter for radiology. *Radiol. Adv.* **2025**, *2*, umaf009. [\[CrossRef\]](#)
44. Di Tommaso, P.; Chatzou, M.; Floden, E.W.; Barja, P.P.; Palumbo, E.; Notredame, C. Nextflow enables reproducible computational workflows. *Nat. Biotechnol.* **2017**, *35*, 316–319. [\[CrossRef\]](#)
45. Digby, B.; Finn, S.P.; Broin, P.O. Nf-core/circrna: A portable workflow for the quantification, miRNA target prediction and differential expression analysis of circular RNAs. *BMC Bioinform.* **2023**, *24*, 27. [\[CrossRef\]](#)
46. Mpangase, P.T.; Frost, J.; Tikly, M.; Ramsay, M.; Hazelhurst, S. Nf-rnaseqcount: A nextflow pipeline for obtaining raw read counts from RNA-seq data. *S. Afr. Comput. J.* **2021**, *33*, 2. [\[CrossRef\]](#)
47. Liu, X.; Bienkowska, J.R.; Zhong, W. Geneteflow: A nextflow-based pipeline for analysing gene and transposable elements expression from RNA-seq data. *PLoS ONE* **2020**, *15*, e0232994. [\[CrossRef\]](#)
48. Hu, K.; Liu, H.; Lawson, N.D.; Zhu, L.J. Scatacpipeline: A nextflow pipeline for comprehensive and reproducible analyses of single cell ATAC-seq data. *Front. Cell Dev. Biol.* **2022**, *10*, 981859. [\[CrossRef\]](#)

49. Engelberg, A.B.; Avorn, J.; Kesselheim, A.S. A new way to contain unaffordable medication costs—Exercising the government’s existing rights. *N. Engl. J. Med.* **2022**, *386*, 1104–1106. [\[CrossRef\]](#)
50. Song, Z.; Gurinovich, A.; Federico, A.; Monti, S.; Sebastiani, P. Nf-gwas-pipeline: A nextflow genome-wide association study pipeline. *J. Open Source Softw.* **2021**, *6*, 2957. [\[CrossRef\]](#)
51. Twesigomwe, D.; Drogemoller, B.I.; Wright, G.E.B.; Siddiqui, A.; da Rocha, J.; Lombard, Z.; Hazelhurst, S. Stellarpgx: A nextflow pipeline for calling star alleles in cytochrome p450 genes. *Clin. Pharmacol. Ther.* **2021**, *110*, 741–749. [\[CrossRef\]](#)
52. Patel, Y.; Zhu, C.; Yamaguchi, T.N.; Bugh, Y.Z.; Tian, M.; Holmes, A.; Fitz-Gibbon, S.T.; Boutros, P.C. Nftest: Automated testing of nextflow pipelines. *Bioinformatics* **2024**, *40*, btae081. [\[CrossRef\]](#)
53. Holzer, M.; Marz, M. Poseidon: A nextflow pipeline for the detection of evolutionary recombination events and positive selection. *Bioinformatics* **2021**, *37*, 1018–1020. [\[CrossRef\]](#)
54. Federico, A.; Karagiannis, T.; Karri, K.; Kishore, D.; Koga, Y.; Campbell, J.D.; Monti, S. Pipeliner: A nextflow-based framework for the definition of sequencing data processing pipelines. *Front. Genet.* **2019**, *10*, 614. [\[CrossRef\]](#)
55. Vlasova, A.; Hermoso Pulido, T.; Camara, F.; Ponomarenko, J.; Guigo, R. Fa-nf: A functional annotation pipeline for proteins from non-model organisms implemented in nextflow. *Genes* **2021**, *12*, 1645. [\[CrossRef\]](#)
56. Allain, F.; Romejon, J.; La Rosa, P.; Jarlier, F.; Servant, N.; Hupe, P. Geniac: Automatic configuration generator and installer for nextflow pipelines. *Open Res. Eur.* **2021**, *1*, 76. [\[CrossRef\]](#)
57. Yukselen, O.; Turkyilmaz, O.; Ozturk, A.R.; Garber, M.; Kucukural, A. Dolphinnext: A distributed data processing platform for high throughput genomics. *BMC Genom.* **2020**, *21*, 310. [\[CrossRef\]](#)
58. Babadi, M.; Fu, J.M.; Lee, S.K.; Smirnov, A.N.; Gauthier, L.D.; Walker, M.; Benjamin, D.I.; Zhao, X.; Karczewski, K.J.; Wong, I.; et al. Gatk-gcnv enables the discovery of rare copy number variants from exome sequencing data. *Nat. Genet.* **2023**, *55*, 1589–1597. [\[CrossRef\]](#)
59. Brouard, J.S.; Bissonnette, N. Variant calling from rna-seq data using the gatk joint genotyping workflow. *Methods Mol. Biol.* **2022**, *2493*, 205–233.
60. Tong, S.Y.; Fan, K.; Zhou, Z.W.; Liu, L.Y.; Zhang, S.Q.; Fu, Y.; Wang, G.Z.; Zhu, Y.; Yu, Y.C. Mvppt: A highly efficient and sensitive pathogenicity prediction tool for missense variants. *Genom. Proteom. Bioinform.* **2023**, *21*, 414–426. [\[CrossRef\]](#)
61. Ng, S.; Masarone, S.; Watson, D.; Barnes, M.R. The benefits and pitfalls of machine learning for biomarker discovery. *Cell Tissue Res.* **2023**, *394*, 17–31. [\[CrossRef\]](#) [\[PubMed\]](#)
62. Jaber, M.I.; Song, B.; Taylor, C.; Vaske, C.J.; Benz, S.C.; Rabizadeh, S.; Soon-Shiong, P.; Szeto, C.W. A deep learning image-based intrinsic molecular subtype classifier of breast tumors reveals tumor heterogeneity that may affect survival. *Breast Cancer Res.* **2020**, *22*, 12. [\[CrossRef\]](#)
63. Tang, Y.; Li, S.; Zhu, L.; Yao, L.; Li, J.; Sun, X.; Liu, Y.; Zhang, Y.; Fu, X. Improve clinical feature-based bladder cancer survival prediction models through integration with gene expression profiles and machine learning techniques. *Heliyon* **2024**, *10*, e38242. [\[CrossRef\]](#) [\[PubMed\]](#)
64. He, W.; Xu, L.; Wang, J.; Yue, Z.; Jing, Y.; Tai, S.; Yang, J.; Fang, X. Vcf2pccluster: A simple, fast and memory-efficient tool for principal component analysis of tens of millions of snps. *BMC Bioinform.* **2024**, *25*, 173. [\[CrossRef\]](#) [\[PubMed\]](#)
65. Cieslak, M.C.; Castelfranco, A.M.; Roncalli, V.; Lenz, P.H.; Hartline, D.K. T-distributed stochastic neighbor embedding (t-sne): A tool for eco-physiological transcriptomic analysis. *Mar. Genom.* **2020**, *51*, 100723. [\[CrossRef\]](#)
66. Li, T.; Zou, Y.; Li, X.; Wong, T.K.F.; Rodrigo, A.G. Mugen-umap: Umap visualization and clustering of mutated genes in single-cell DNA sequencing data. *BMC Bioinform.* **2024**, *25*, 308. [\[CrossRef\]](#)
67. Schmidt, B.; Hildebrandt, A. Deep learning in next-generation sequencing. *Drug Discov. Today* **2021**, *26*, 173–180. [\[CrossRef\]](#)
68. Hwang, H.; Jeon, H.; Yeo, N.; Baek, D. Big data and deep learning for rna biology. *Exp. Mol. Med.* **2024**, *56*, 1293–1321. [\[CrossRef\]](#)
69. Zhao, X.; Wang, L.; Zhang, Y.; Han, X.; Deveci, M.; Parmar, M. A review of convolutional neural networks in computer vision. *Artif. Intell. Rev.* **2024**, *57*, 99. [\[CrossRef\]](#)
70. Alharbi, W.S.; Rashid, M. A review of deep learning applications in human genomics using next-generation sequencing data. *Hum. Genom.* **2022**, *16*, 26. [\[CrossRef\]](#)
71. Vaz, J.M.; Balaji, S. Convolutional neural networks (cnns): Concepts and applications in pharmacogenomics. *Mol. Divers.* **2021**, *25*, 1569–1584. [\[CrossRef\]](#) [\[PubMed\]](#)
72. Zhou, J.; Troyanskaya, O.G. Predicting effects of noncoding variants with deep learning-based sequence model. *Nat. Methods* **2015**, *12*, 931–934. [\[CrossRef\]](#) [\[PubMed\]](#)
73. Krishnamachari, K.; Lu, D.; Swift-Scott, A.; Yeraliyev, A.; Lee, K.; Huang, W.; Leng, S.; Jacobsen Skanderup, A. Accurate somatic variant detection using weakly supervised deep learning. *Nat. Commun.* **2022**, *13*, 4248. [\[CrossRef\]](#) [\[PubMed\]](#)
74. Lai, B.; Qian, S.; Zhang, H.; Zhang, S.; Kozlova, A.; Duan, J.; Xu, J.; He, X. Annotating functional effects of non-coding variants in neuropsychiatric cell types by deep transfer learning. *PLoS Comput. Biol.* **2022**, *18*, e1010011. [\[CrossRef\]](#)

75. Jaganathan, K.; Kyriazopoulou Panagiotopoulou, S.; McRae, J.F.; Darbandi, S.F.; Knowles, D.; Li, Y.I.; Kosmicki, J.A.; Arbelaez, J.; Cui, W.; Schwartz, G.B.; et al. Predicting splicing from primary sequence with deep learning. *Cell* **2019**, *176*, 535–548.e24. [\[CrossRef\]](#)
76. Beknazarov, N.; Poptsova, M. Deepz: A deep learning approach for z-DNA prediction. *Methods Mol. Biol.* **2023**, *2651*, 217–226.
77. Mienye, D.; Swart, T.; Obaido, G. Recurrent neural networks: A comprehensive review of architectures, variants, and applications. *Information* **2024**, *15*, 517. [\[CrossRef\]](#)
78. Liu, F.; Miao, Y.; Liu, Y.; Hou, T. Rnn-virseeker: A deep learning method for identification of short viral sequences from metagenomes. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2022**, *19*, 1840–1849. [\[CrossRef\]](#)
79. Paggi, J.M.; Bejerano, G. A sequence-based, deep learning model accurately predicts rna splicing branchpoints. *RNA* **2018**, *24*, 1647–1658. [\[CrossRef\]](#)
80. Gupta, G.; Saini, S. Davi: Deep learning based tool for alignment and single nucleotide variant identification. *bioRxiv* **2019**, 778647.
81. Chen, Y.; Xie, M.; Wen, J. Predicting gene expression from histone modifications with self-attention based neural networks and transfer learning. *Front. Genet.* **2022**, *13*, 1081842. [\[CrossRef\]](#) [\[PubMed\]](#)
82. Aburass, S.; Dorgham, O.; Al Shaqsi, J. A hybrid machine learning model for classifying gene mutations in cancer using lstm, bilstm, cnn, gru, and glove. *Syst. Soft Comput.* **2024**, *6*, 200110. [\[CrossRef\]](#)
83. Zheng, Z.; Li, S.; Su, J.; Leung, A.W.; Lam, T.W.; Luo, R. Symphonizing pileup and full-alignment for deep learning-based long-read variant calling. *Nat. Comput. Sci.* **2022**, *2*, 797–803. [\[CrossRef\]](#) [\[PubMed\]](#)
84. Shafin, K.; Pesout, T.; Chang, P.-C.; Nattestad, M.; Kolesnikov, A.; Goel, S.; Baid, G.; Kolmogorov, M.; Eizenga, J.M.; Miga, K.H.; et al. Haplotype-aware variant calling with pepper-margin-deepvariant enables high accuracy in nanopore long-reads. *Nat. Methods* **2021**, *18*, 1322–1332. [\[CrossRef\]](#)
85. Alirezaie, N.; Kernohan, K.D.; Hartley, T.; Majewski, J.; Hocking, T.D. Clinpred: Prediction tool to identify disease-relevant nonsynonymous single-nucleotide variants. *Am. J. Hum. Genet.* **2018**, *103*, 474–483. [\[CrossRef\]](#)
86. Choon, Y.W.; Choon, Y.F.; Nasarudin, N.A.; Al Jasmi, F.; Remli, M.A.; Alkayali, M.H.; Mohamad, M.S. Artificial intelligence and database for ngs-based diagnosis in rare disease. *Front. Genet.* **2023**, *14*, 1258083. [\[CrossRef\]](#)
87. Zhang, H.; Yin, Z.; Wei, Y.; Schmidt, B.; Liu, W. Deepfilter: A deep learning based variant filter for vardict. *Tsinghua Sci. Technol.* **2023**, *28*, 665–672. [\[CrossRef\]](#)
88. Freed, D.; Pan, R.; Chen, H.; Li, Z.; Hu, J.; Aldana, R. Dnascope: High accuracy small variant calling using machine learning. *bioRxiv* **2022**.
89. Ioannidis, N.M.; Rothstein, J.H.; Pejaver, V.; Middha, S.; McDonnell, S.K.; Baheti, S.; Musolf, A.; Li, Q.; Holzinger, E.; Karyadi, D.; et al. Revel: An ensemble method for predicting the pathogenicity of rare missense variants. *Am. J. Hum. Genet.* **2016**, *99*, 877–885. [\[CrossRef\]](#)
90. Ji, Y.; Zhou, Z.; Liu, H.; Davuluri, R.V. Dnabert: Pre-trained bidirectional encoder representations from transformers model for DNA-language in genome. *Bioinformatics* **2021**, *37*, 2112–2120. [\[CrossRef\]](#)
91. Thurman, R.E.; Rynes, E.; Humbert, R.; Vierstra, J.; Maurano, M.T.; Haugen, E.; Sheffield, N.C.; Stergachis, A.B.; Wang, H.; Vernot, B.; et al. The accessible chromatin landscape of the human genome. *Nature* **2012**, *489*, 75–82. [\[CrossRef\]](#) [\[PubMed\]](#)
92. Ernst, J.; Kellis, M. Chromatin-state discovery and genome annotation with chromhmm. *Nat. Protoc.* **2017**, *12*, 2478–2492. [\[CrossRef\]](#) [\[PubMed\]](#)
93. Lin, J.; Luo, R.; Pinello, L. Epinformer: A scalable deep learning framework for gene expression prediction by integrating promoter-enhancer sequences with multimodal epigenomic data. *bioRxiv* **2024**.
94. Singh, R.; Lanchantin, J.; Robins, G.; Qi, Y. Deepchrome: Deep-learning for predicting gene expression from histone modifications. *Bioinformatics* **2016**, *32*, i639–i648. [\[CrossRef\]](#) [\[PubMed\]](#)
95. Sekhon, A.; Singh, R.; Qi, Y. Deepdiff: Deep-learning for predicting differential gene expression from histone modifications. *Bioinformatics* **2018**, *34*, i891–i900. [\[CrossRef\]](#)
96. Yan, F.; Telonis, A.G.; Yang, Q.; Jiang, L.; Garrett-Bakelman, F.E.; Sekeres, M.A.; Santini, V.; Ceccarelli, M.; Goel, N.; Garcia-Martinez, L.; et al. Genome-wide methylome modeling via generative ai incorporating long- and short-range interactions. *Sci. Adv.* **2025**, *11*, eadt4152. [\[CrossRef\]](#)
97. Ying, K.; Song, J.; Cui, H.; Zhang, Y.; Li, S.; Chen, X.; Liu, H.; Eames, A.; McCartney, D.L.; Marioni, R.E.; et al. Methylgpt: A foundation model for the DNA methylome. *bioRxiv* **2024**.
98. Wang, Y.; Liu, T.; Xu, D.; Shi, H.; Zhang, C.; Mo, Y.Y.; Wang, Z. Predicting DNA methylation state of cpg dinucleotide using genome topological features and deep networks. *Sci. Rep.* **2016**, *6*, 19598. [\[CrossRef\]](#)
99. Levy, J.J.; Titus, A.J.; Petersen, C.L.; Chen, Y.; Salas, L.A.; Christensen, B.C. Methylnet: An automated and modular deep learning approach for DNA methylation analysis. *BMC Bioinform.* **2020**, *21*, 108. [\[CrossRef\]](#)
100. Levy, J.J.; Chen, Y.; Azizgolshani, N.; Petersen, C.L.; Titus, A.J.; Moen, E.L.; Vaickus, L.J.; Salas, L.A.; Christensen, B.C. Methylsp-net and methylcapsnet: Biologically motivated organization of dnam neural networks, inspired by capsule networks. *NPJ Syst. Biol. Appl.* **2021**, *7*, 33. [\[CrossRef\]](#)

101. Kelley, D.R.; Snoek, J.; Rinn, J.L. Basset: Learning the regulatory code of the accessible genome with deep convolutional neural networks. *Genome Res.* **2016**, *26*, 990–999. [[CrossRef](#)] [[PubMed](#)]
102. Avsec, Z.; Agarwal, V.; Visentin, D.; Ledsam, J.R.; Grabska-Barwinska, A.; Taylor, K.R.; Assael, Y.; Jumper, J.; Kohli, P.; Kelley, D.R. Effective gene expression prediction from sequence by integrating long-range interactions. *Nat. Methods* **2021**, *18*, 1196–1203. [[CrossRef](#)] [[PubMed](#)]
103. Yu, W.; Uzun, Y.; Zhu, Q.; Chen, C.; Tan, K. Scatac-pro: A comprehensive workbench for single-cell chromatin accessibility sequencing data. *Genome Biol.* **2020**, *21*, 94. [[CrossRef](#)] [[PubMed](#)]
104. Fan, X.; Liu, J.; Yang, Y.; Gu, C.; Han, Y.; Wu, B.; Jiang, Y.; Chen, G.; Heng, P.A. Scgraphformer: Unveiling cellular heterogeneity and interactions in scRNA-seq data using a scalable graph transformer network. *Commun. Biol.* **2024**, *7*, 1463. [[CrossRef](#)]
105. Yin, Q.; Wu, M.; Liu, Q.; Lv, H.; Jiang, R. DeepHistone: A deep learning approach to predicting histone modifications. *BMC Genom.* **2019**, *20*, 193. [[CrossRef](#)]
106. Wen, W.; Zhong, J.; Zhang, Z.; Jia, L.; Chu, T.; Wang, N.; Danko, C.G.; Wang, Z. Dhica: A deep transformer-based model enables accurate histone imputation from chromatin accessibility. *Brief. Bioinform.* **2024**, *25*, bbae459. [[CrossRef](#)]
107. Schuette, G.; Lao, Z.; Zhang, B. Chromogen: Diffusion model predicts single-cell chromatin conformations. *Sci. Adv.* **2025**, *11*, eadr8265. [[CrossRef](#)]
108. Wang, Y.; Kong, S.; Zhou, C.; Wang, Y.; Zhang, Y.; Fang, Y.; Li, G. A review of deep learning models for the prediction of chromatin interactions with DNA and epigenomic profiles. *Brief. Bioinform.* **2024**, *26*, bbae651. [[CrossRef](#)]
109. Brozek, A.; Theodoris, C.V. Ai learns from chromatin data to uncover gene interactions. *Nature* **2025**, *637*, 799–800. [[CrossRef](#)]
110. Saadh, M.J.; Ahmed, H.H.; Kareem, R.A.; Yadav, A.; Ganesan, S.; Shankhyan, A.; Sharma, G.C.; Naidu, K.S.; Rakhmatullaev, A.; Sameer, H.N.; et al. Advanced machine learning framework for enhancing breast cancer diagnostics through transcriptomic profiling. *Discov. Oncol.* **2025**, *16*, 334. [[CrossRef](#)]
111. Eraslan, G.; Simon, L.M.; Mircea, M.; Mueller, N.S.; Theis, F.J. Single-cell RNA-seq denoising using a deep count autoencoder. *Nat. Commun.* **2019**, *10*, 390. [[CrossRef](#)] [[PubMed](#)]
112. Svensson, V.; Gayoso, A.; Yosef, N.; Pachter, L. Interpretable factor models of single-cell RNA-seq via variational autoencoders. *Bioinformatics* **2020**, *36*, 3418–3421. [[CrossRef](#)] [[PubMed](#)]
113. Gronbech, C.H.; Vording, M.F.; Timshel, P.N.; Sonderby, C.K.; Pers, T.H.; Winther, O. Scvae: Variational auto-encoders for single-cell gene expression data. *Bioinformatics* **2020**, *36*, 4415–4422. [[CrossRef](#)] [[PubMed](#)]
114. Lopez, R.; Regier, J.; Cole, M.B.; Jordan, M.I.; Yosef, N. Deep generative modeling for single-cell transcriptomics. *Nat. Methods* **2018**, *15*, 1053–1058. [[CrossRef](#)]
115. Hoffman, G.E.; Bendl, J.; Girdhar, K.; Schadt, E.E.; Roussos, P. Functional interpretation of genetic variants using deep learning predicts impact on chromatin accessibility and histone modification. *Nucleic Acids Res.* **2019**, *47*, 10597–10611. [[CrossRef](#)]
116. Chen, Y.; Sim, A.; Wan, Y.K.; Yeo, K.; Lee, J.J.X.; Ling, M.H.; Love, M.I.; Goke, J. Context-aware transcript quantification from long-read RNA-seq data with bambu. *Nat. Methods* **2023**, *20*, 1187–1195. [[CrossRef](#)]
117. de Sainte Agathe, J.M.; Filser, M.; Isidor, B.; Besnard, T.; Gueguen, P.; Perrin, A.; Van Goethem, C.; Verebi, C.; Masingue, M.; Rendu, J.; et al. Spliceai-visual: A free online tool to improve spliceai splicing variant interpretation. *Hum. Genom.* **2023**, *17*, 7. [[CrossRef](#)]
118. Barbosa, P.; Savisaar, R.; Carmo-Fonseca, M.; Fonseca, A. Computational prediction of human deep intronic variation. *Gigascience* **2022**, *12*, giad085. [[CrossRef](#)]
119. Strauch, Y.; Lord, J.; Niranjana, M.; Baralle, D. Ci-spliceai-improving machine learning predictions of disease causing splicing variants using curated alternative splice sites. *PLoS ONE* **2022**, *17*, e0269159. [[CrossRef](#)]
120. Chao, K.H.; Mao, A.; Salzberg, S.L.; Pertea, M. Splam: A deep-learning-based splice site predictor that improves spliced alignments. *Genome Biol.* **2024**, *25*, 243. [[CrossRef](#)]
121. Erfanian, N.; Heydari, A.A.; Feriz, A.M.; Ianez, P.; Derakhshani, A.; Ghasemigol, M.; Farahpour, M.; Razavi, S.M.; Nasser, S.; Safarpour, H.; et al. Deep learning applications in single-cell genomics and transcriptomics data analysis. *Biomed. Pharmacother.* **2023**, *165*, 115077. [[CrossRef](#)] [[PubMed](#)]
122. Gayoso, A.; Steier, Z.; Lopez, R.; Regier, J.; Nator, K.L.; Streets, A.; Yosef, N. Joint probabilistic modeling of single-cell multi-omic data with totalvi. *Nat. Methods* **2021**, *18*, 272–282. [[CrossRef](#)] [[PubMed](#)]
123. Fu, Y.; Qu, H.; Qu, D.; Zhao, M. Trajectory inference with cell-cell interactions (ticci): Intercellular communication improves the accuracy of trajectory inference methods. *Bioinformatics* **2025**, *41*, btad027. [[CrossRef](#)] [[PubMed](#)]
124. Fang, M.; Gorin, G.; Pachter, L. Trajectory inference from single-cell genomics data with a process time model. *PLoS Comput. Biol.* **2025**, *21*, e1012752. [[CrossRef](#)]
125. Huguet, G.; Magruder, D.S.; Tong, A.; Fasina, O.; Kuchroo, M.; Wolf, G.; Krishnaswamy, S. Manifold interpolating optimal-transport flows for trajectory inference. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 29705–29718.
126. Smolander, J.; Junttila, S.; Venalainen, M.S.; Elo, L.L. Scshaper: An ensemble method for fast and accurate linear trajectory inference from single-cell RNA-seq data. *Bioinformatics* **2022**, *38*, 1328–1335. [[CrossRef](#)]

127. Gao, R.; Bai, S.; Henderson, Y.C.; Lin, Y.; Schalck, A.; Yan, Y.; Kumar, T.; Hu, M.; Sei, E.; Davis, A.; et al. Delineating copy number and clonal substructure in human tumors from single-cell transcriptomes. *Nat. Biotechnol.* **2021**, *39*, 599–608. [[CrossRef](#)]
128. Alum, E.U. Ai-driven biomarker discovery: Enhancing precision in cancer diagnosis and prognosis. *Discov. Oncol.* **2025**, *16*, 313. [[CrossRef](#)]
129. Unger, M.; Kather, J.N. Deep learning in cancer genomics and histopathology. *Genome Med.* **2024**, *16*, 44. [[CrossRef](#)]
130. Tellez-Gabriel, M.; Ory, B.; Lamoureux, F.; Heymann, M.F.; Heymann, D. Tumour heterogeneity: The key advantages of single-cell analysis. *Int. J. Mol. Sci.* **2016**, *17*, 2142. [[CrossRef](#)]
131. Gillis, S.; Roth, A. Pyclone-vi: Scalable inference of clonal population structures using whole genome data. *BMC Bioinform.* **2020**, *21*, 571. [[CrossRef](#)] [[PubMed](#)]
132. Lu, B. Cancer phylogenetic inference using copy number alterations detected from DNA sequencing data. *Cancer Pathog. Ther.* **2025**, *3*, 16–29. [[CrossRef](#)] [[PubMed](#)]
133. Fan, J.; Lee, H.O.; Lee, S.; Ryu, D.E.; Lee, S.; Xue, C.; Kim, S.J.; Kim, K.; Barkas, N.; Park, P.J.; et al. Linking transcriptional and genetic tumor heterogeneity through allele analysis of single-cell rna-seq data. *Genome Res.* **2018**, *28*, 1217–1227. [[CrossRef](#)] [[PubMed](#)]
134. Jiang, P.; Gu, S.; Pan, D.; Fu, J.; Sahu, A.; Hu, X.; Li, Z.; Traugh, N.; Bu, X.; Li, B.; et al. Signatures of t cell dysfunction and exclusion predict cancer immunotherapy response. *Nat. Med.* **2018**, *24*, 1550–1558. [[CrossRef](#)]
135. Nguyen, L.; Van Hoeck, A.; Cuppen, E. Machine learning-based tissue of origin classification for cancer of unknown primary diagnostics using genome-wide mutation features. *Nat. Commun.* **2022**, *13*, 4013. [[CrossRef](#)]
136. Zhao, Y.; Pan, Z.; Namburi, S.; Pattison, A.; Posner, A.; Balachander, S.; Paisie, C.A.; Reddi, H.V.; Rueter, J.; Gill, A.J.; et al. Cup-ai-dx: A tool for inferring cancer tissue of origin and molecular subtype using rna gene-expression data and artificial intelligence. *EBioMedicine* **2020**, *61*, 103030. [[CrossRef](#)]
137. Wu, Y.; Xie, L. Ai-driven multi-omics integration for multi-scale predictive modeling of genotype-environment-phenotype relationships. *Comput. Struct. Biotechnol. J.* **2025**, *27*, 265–277. [[CrossRef](#)]
138. Taroni, J.N.; Grayson, P.C.; Hu, Q.; Eddy, S.; Kretzler, M.; Merkel, P.A.; Greene, C.S. Multiplier: A transfer learning framework for transcriptomics reveals systemic features of rare disease. *Cell Syst.* **2019**, *8*, 380–394.e4. [[CrossRef](#)]
139. Wu, K.E.; Yost, K.E.; Chang, H.Y.; Zou, J. Babel enables cross-modality translation between multiomic profiles at single-cell resolution. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2023070118. [[CrossRef](#)]
140. Truesdell, P.; Chang, J.; Coto Villa, D.; Dai, M.; Zhao, Y.; McIlwain, R.; Young, S.; Hiley, S.; Craig, A.W.; Babak, T. Pharmacogenomic discovery of genetically targeted cancer therapies optimized against clinical outcomes. *NPJ Precis. Oncol.* **2024**, *8*, 186. [[CrossRef](#)]
141. Ching, T.; Zhu, X.; Garmire, L.X. Cox-nnet: An artificial neural network method for prognosis prediction of high-throughput omics data. *PLoS Comput. Biol.* **2018**, *14*, e1006076. [[CrossRef](#)] [[PubMed](#)]
142. Mhandire, D.Z.; Goey, A.K.L. The value of pharmacogenetics to reduce drug-related toxicity in cancer patients. *Mol. Diagn. Ther.* **2022**, *26*, 137–151. [[CrossRef](#)] [[PubMed](#)]
143. Ma, L.; Guo, H.; Zhao, Y.; Liu, Z.; Wang, C.; Bu, J.; Sun, T.; Wei, J. Liquid biopsy in cancer current: Status, challenges and future prospects. *Signal Transduct. Target. Ther.* **2024**, *9*, 336. [[CrossRef](#)] [[PubMed](#)]
144. Norouzkhani, N.; Mobaraki, H.; Varmazyar, S.; Zaboli, H.; Mohamadi, Z.; Nikeghbali, G.; Bagheri, K.; Marivany, N.; Najafi, M.; Nozad Varjovi, M.; et al. Artificial intelligence networks for assessing the prognosis of gastrointestinal cancer to immunotherapy based on genetic mutation features: A systematic review and meta-analysis. *BMC Gastroenterol.* **2025**, *25*, 310. [[CrossRef](#)]
145. Kong, J.; Zhao, X.; Singhal, A.; Park, S.; Bachelder, R.; Shen, J.; Zhang, H.; Moon, J.; Ahn, C.; Ock, C.Y.; et al. Prediction of immunotherapy response using mutations to cancer protein assemblies. *Sci. Adv.* **2024**, *10*, eado9746. [[CrossRef](#)]
146. You, Y.; Lai, X.; Pan, Y.; Zheng, H.; Vera, J.; Liu, S.; Deng, S.; Zhang, L. Artificial intelligence in cancer target identification and drug discovery. *Signal Transduct. Target. Ther.* **2022**, *7*, 156. [[CrossRef](#)]
147. Huang, K.; Xiao, C.; Glass, L.M.; Sun, J. Moltrans: Molecular interaction transformer for drug-target interaction prediction. *Bioinformatics* **2021**, *37*, 830–836. [[CrossRef](#)]
148. Zeng, X.; Zhu, S.; Liu, X.; Zhou, Y.; Nussinov, R.; Cheng, F. Deepdr: A network-based deep learning approach to in silico drug repositioning. *Bioinformatics* **2019**, *35*, 5191–5198. [[CrossRef](#)]
149. Kuru, H.I.; Tastan, O.; Cicek, A.E. Matchmaker: A deep learning framework for drug synergy prediction. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2022**, *19*, 2334–2344. [[CrossRef](#)]
150. Huang, K.; Fu, T.; Glass, L.M.; Zitnik, M.; Xiao, C.; Sun, J. Deeppurpose: A deep learning library for drug-target interaction prediction. *Bioinformatics* **2021**, *36*, 5545–5547. [[CrossRef](#)]
151. Duan, X.P.; Qin, B.D.; Jiao, X.D.; Liu, K.; Wang, Z.; Zang, Y.S. New clinical trial design in precision medicine: Discovery, development and direction. *Signal Transduct. Target. Ther.* **2024**, *9*, 57. [[CrossRef](#)] [[PubMed](#)]
152. Zhang, X.; Xing, Y.; Sun, K.; Guo, Y. Omiembed: A unified multi-task deep learning framework for multi-omics data. *Cancers* **2021**, *13*, 3047. [[CrossRef](#)] [[PubMed](#)]

153. Hu, J.; Li, X.; Coleman, K.; Schroeder, A.; Ma, N.; Irwin, D.J.; Lee, E.B.; Shinohara, R.T.; Li, M. Spagcn: Integrating gene expression, spatial location and histology to identify spatial domains and spatially variable genes by graph convolutional network. *Nat. Methods* **2021**, *18*, 1342–1351. [[CrossRef](#)] [[PubMed](#)]
154. Jia, S.; Jiang, S.; Zhang, S.; Xu, M.; Jia, X. Graph-in-graph convolutional network for hyperspectral image classification. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**, *35*, 1157–1171. [[CrossRef](#)]
155. Wang, T.; Shao, W.; Huang, Z.; Tang, H.; Zhang, J.; Ding, Z.; Huang, K. Mogonet integrates multi-omics data using graph convolutional networks allowing patient classification and biomarker identification. *Nat. Commun.* **2021**, *12*, 3445. [[CrossRef](#)]
156. Zhang, L.; Song, R.; Tan, W.; Ma, L.; Zhang, W. Igcn: A provably informative gcen embedding for semi-supervised learning with extremely limited labels. *IEEE Trans. Pattern Anal. Mach. Intell.* **2024**, *46*, 8396–8409. [[CrossRef](#)]
157. Tran, A.T.; Zeevi, T.; Payabvash, S. Strategies to improve the robustness and generalizability of deep learning segmentation and classification in neuroimaging. *BioMedInformatics* **2025**, *5*, 20. [[CrossRef](#)]
158. Soliman, A.; Li, Z.; Parwani, A.V. Artificial intelligence's impact on breast cancer pathology: A literature review. *Diagn. Pathol.* **2024**, *19*, 38. [[CrossRef](#)]
159. Minton, K. Predicting variant pathogenicity with alphamissense. *Nat. Rev. Genet.* **2023**, *24*, 804. [[CrossRef](#)]
160. Caswell, R.C.; Gunning, A.C.; Owens, M.M.; Ellard, S.; Wright, C.F. Assessing the clinical utility of protein structural analysis in genomic variant classification: Experiences from a diagnostic laboratory. *Genome Med.* **2022**, *14*, 77. [[CrossRef](#)]
161. Quazi, S. Artificial intelligence and machine learning in precision and genomic medicine. *Med. Oncol.* **2022**, *39*, 120. [[CrossRef](#)] [[PubMed](#)]
162. Wadden, J.J. Defining the undefinable: The black box problem in healthcare artificial intelligence. *J. Med. Ethics* **2021**, *48*, 107529. [[CrossRef](#)] [[PubMed](#)]
163. Linardatos, P.; Papastefanopoulos, V.; Kotsiantis, S. Explainable ai: A review of machine learning interpretability methods. *Entropy* **2020**, *23*, 18. [[CrossRef](#)] [[PubMed](#)]
164. Vadapalli, S.; Abdelhalim, H.; Zeeshan, S.; Ahmed, Z. Artificial intelligence and machine learning approaches using gene expression and variant data for personalized medicine. *Brief. Bioinform.* **2022**, *23*, bbac191. [[CrossRef](#)]
165. Burley, S.K.; Berman, H.M.; Kleywegt, G.J.; Markley, J.L.; Nakamura, H.; Velankar, S. Protein data bank (pdb): The single global macromolecular structure archive. *Methods Mol. Biol.* **2017**, *1607*, 627–641.
166. Karuppasamy, M.P.; Venkateswaran, S.; Subbiah, P. Pdb-2-pbv3.0: An updated protein block database. *J. Bioinform. Comput. Biol.* **2020**, *18*, 2050009. [[CrossRef](#)]
167. Reiff, S.B.; Schroeder, A.J.; Kirli, K.; Cosolo, A.; Bakker, C.; Mercado, L.; Lee, S.; Veit, A.D.; Balashov, A.K.; Vitzthum, C.; et al. The 4d nucleome data portal as a resource for searching and visualizing curated nucleomics data. *Nat. Commun.* **2022**, *13*, 2365. [[CrossRef](#)]
168. Oluwadare, O.; Highsmith, M.; Turner, D.; Lieberman Aiden, E.; Cheng, J. Gsdb: A database of 3D chromosome and genome structures reconstructed from hi-c data. *BMC Mol. Cell Biol.* **2020**, *21*, 60. [[CrossRef](#)]
169. Li, C.; Dong, X.; Fan, H.; Wang, C.; Ding, G.; Li, Y. The 3DGD: A database of genome 3D structure. *Bioinformatics* **2014**, *30*, 1640–1642. [[CrossRef](#)]
170. HariPriya, R.; Khare, N.; Pandey, M. Privacy-preserving federated learning for collaborative medical data mining in multi-institutional settings. *Sci. Rep.* **2025**, *15*, 12482. [[CrossRef](#)]
171. Cesaro, A.; Hoffman, S.C.; Das, P.; de la Fuente-Nunez, C. Challenges and applications of artificial intelligence in infectious diseases and antimicrobial resistance. *NPJ Antimicrob. Resist.* **2025**, *3*, 2. [[CrossRef](#)]
172. Acosta, J.N.; Falcone, G.J.; Rajpurkar, P.; Topol, E.J. Multimodal biomedical ai. *Nat. Med.* **2022**, *28*, 1773–1784. [[CrossRef](#)] [[PubMed](#)]
173. Tarazona, S.; Arzalluz-Luque, A.; Conesa, A. Undisclosed, unmet and neglected challenges in multi-omics studies. *Nat. Comput. Sci.* **2021**, *1*, 395–402. [[CrossRef](#)] [[PubMed](#)]
174. Gangwal, A.; Ansari, A.; Ahmad, I.; Azad, A.K.; Kumarasamy, V.; Subramaniyan, V.; Wong, L.S. Generative artificial intelligence in drug discovery: Basic framework, recent advances, challenges, and opportunities. *Front. Pharmacol.* **2024**, *15*, 1331062. [[CrossRef](#)] [[PubMed](#)]
175. Goktas, P.; Grzybowski, A. Shaping the future of healthcare: Ethical clinical challenges and pathways to trustworthy ai. *J. Clin. Med.* **2025**, *14*, 1605. [[CrossRef](#)]
176. Scarano, C.; Veneruso, I.; De Simone, R.R.; Di Bonito, G.; Secondino, A.; D'Argenio, V. The third-generation sequencing challenge: Novel insights for the omic sciences. *Biomolecules* **2024**, *14*, 568. [[CrossRef](#)]
177. Goodwin, S.; McPherson, J.D.; McCombie, W.R. Coming of age: Ten years of next-generation sequencing technologies. *Nat. Rev. Genet.* **2016**, *17*, 333–351. [[CrossRef](#)]
178. Wang, Y.; Zhao, Y.; Bollas, A.; Wang, Y.; Au, K.F. Nanopore sequencing technology, bioinformatics and applications. *Nat. Biotechnol.* **2021**, *39*, 1348–1365. [[CrossRef](#)]
179. Boza, V.; Brejova, B.; Vinar, T. Deepnano: Deep recurrent neural networks for base calling in minion nanopore reads. *PLoS ONE* **2017**, *12*, e0178751. [[CrossRef](#)]

180. Wan, Y.K.; Hendra, C.; Pratanwanich, P.N.; Goke, J. Beyond sequencing: Machine learning algorithms extract biology hidden in nanopore signal data. *Trends Genet.* **2022**, *38*, 246–257. [[CrossRef](#)]
181. Pages-Gallego, M.; de Ridder, J. Comprehensive benchmark and architectural analysis of deep learning models for nanopore sequencing basecalling. *Genome Biol.* **2023**, *24*, 71. [[CrossRef](#)] [[PubMed](#)]
182. Baid, G.; Cook, D.E.; Shafin, K.; Yun, T.; Llinares-López, F.; Berthet, Q.; Belyaeva, A.; Töpfer, A.; Wenger, A.M.; Rowell, W.J.; et al. Deepconsensus improves the accuracy of sequences with a gap-aware sequence transformer. *Nat. Biotechnol.* **2023**, *41*, 232–238. [[CrossRef](#)] [[PubMed](#)]
183. Ahsan, M.U.; Liu, Q.; Perdomo, J.E.; Fang, L.; Wang, K. A survey of algorithms for the detection of genomic structural variants from long-read sequencing data. *Nat. Methods* **2023**, *20*, 1143–1158. [[CrossRef](#)] [[PubMed](#)]
184. Zheng, Z.; Yu, X.; Chen, L.; Lee, Y.L.; Xin, C.; Wong, A.O.K.; Jain, M.; Kesharwani, R.K.; Sedlazeck, F.J.; Luo, R. Clair3-rna: A deep learning-based small variant caller for long-read rna sequencing data. *bioRxiv* **2025**.
185. Senol Cali, D.; Kim, J.S.; Ghose, S.; Alkan, C.; Mutlu, O. Nanopore sequencing technology and tools for genome assembly: Computational analysis of the current state, bottlenecks and future directions. *Brief. Bioinform.* **2019**, *20*, 1542–1559. [[CrossRef](#)]
186. Ni, P.; Huang, N.; Zhang, Z.; Wang, D.P.; Liang, F.; Miao, Y.; Xiao, C.L.; Luo, F.; Wang, J. DeepSignal: Detecting DNA methylation state from nanopore sequencing reads using deep-learning. *Bioinformatics* **2019**, *35*, 4586–4595. [[CrossRef](#)]
187. McIntyre, A.B.R.; Alexander, N.; Grigorev, K.; Bezdan, D.; Sichtig, H.; Chiu, C.Y.; Mason, C.E. Single-molecule sequencing detection of n6-methyladenine in microbial reference materials. *Nat. Commun.* **2019**, *10*, 579. [[CrossRef](#)]
188. Rand, A.C.; Jain, M.; Eizenga, J.M.; Musselman-Brown, A.; Olsen, H.E.; Akeson, M.; Paten, B. Mapping DNA methylation with high-throughput nanopore sequencing. *Nat. Methods* **2017**, *14*, 411–413. [[CrossRef](#)]
189. Lorenz, D.A.; Sathe, S.; Einstein, J.M.; Yeo, G.W. Direct rna sequencing enables m(6)a detection in endogenous transcript isoforms at base-specific resolution. *RNA* **2020**, *26*, 19–28. [[CrossRef](#)]
190. Liu, H.; Begik, O.; Lucas, M.C.; Ramirez, J.M.; Mason, C.E.; Wiener, D.; Schwartz, S.; Mattick, J.S.; Smith, M.A.; Novoa, E.M. Accurate detection of m(6)a rna modifications in native rna sequences. *Nat. Commun.* **2019**, *10*, 4079. [[CrossRef](#)]
191. Gao, Y.; Liu, X.; Wu, B.; Wang, H.; Xi, F.; Kohnen, M.V.; Reddy, A.S.N.; Gu, L. Quantitative profiling of n(6)-methyladenosine at single-base resolution in stem-differentiating xylem of populus trichocarpa using nanopore direct rna sequencing. *Genome Biol.* **2021**, *22*, 22. [[CrossRef](#)] [[PubMed](#)]
192. Leger, A.; Amaral, P.P.; Pandolfini, L.; Capitanchik, C.; Capraro, F.; Miano, V.; Migliori, V.; Toolan-Kerr, P.; Sideri, T.; Enright, A.J.; et al. Rna modifications detection by comparative nanopore direct rna sequencing. *Nat. Commun.* **2021**, *12*, 7198. [[CrossRef](#)] [[PubMed](#)]
193. Ueda, H. Nanodoc: Rna modification detection using nanopore raw reads with deep one-class classification. *bioRxiv* **2021**.
194. Pratanwanich, P.N.; Yao, F.; Chen, Y.; Koh, C.W.Q.; Wan, Y.K.; Hendra, C.; Poon, P.; Goh, Y.T.; Yap, P.M.L.; Chooi, J.Y.; et al. Identification of differential rna modifications from nanopore direct rna sequencing with xpore. *Nat. Biotechnol.* **2021**, *39*, 1394–1402. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.