



Artificial Intelligence in genetic healthcare: toward adaptive and systems-level genomic medicine

Tupili I^{1*}, Ravipati D^{2*}, Mangu K³, Kour B⁴

Submitted: March 3, 2026, Revised: version 1, May 8, 2026, version 2, June 20, 2026, version 3, July 21, 2026, version 4, September 17, 2026

Accepted: September 25, 2026

Abstract

Artificial intelligence is reshaping genetic healthcare through increasingly sophisticated approaches to variant interpretation, disease prediction, biomarker discovery, gene editing, and precision therapeutics. Yet these applications remain fragmented, often treating genomic and clinical information as static observations rather than components of biological systems that evolve during disease. This review connects current AI capabilities with a next-generation model of adaptive genomic medicine integrating longitudinal multi-omics, clinical and environmental data, explainable AI, federated learning, network-level biomarkers, and continuous outcome feedback. We extend this framework beyond conventional genomic prediction by considering pathway interactions, compensatory biological redundancy, network reorganization, and cross-system changes as potentially informative features of disease progression. We further propose longitudinal approaches for tracking changing biological relationships and resilience before conventional biomarkers become abnormal. Together, these concepts reposition genomic AI from a collection of predictive tools toward a dynamic systems-level framework capable of learning how disease evolves, anticipating state transitions, and supporting increasingly individualized clinical intervention.

Keywords

Artificial Intelligence, Genomic medicine, Precision medicine, Multi-omics, Machine Learning, Clinical genomics, Biomarker discovery, Explainable AI, Systems biology, Redundancy-informed genomic modeling

¹Ishanvi Tupili and ²Deekshitha Ravipati, Indus International School, Kondakal Village, Near Mokila, Shankarpalle, Telangana 501203, India. *Both authors contributed equally

³Co-Corresponding author: Karthik Mangu, National Institute of Pharmaceutical Education and Research NIPER), and Kognovate Education and Research, Bionest, Avishkaran (NIPER), Hyderabad, Telangana 500037, India. karthikmanguias@gmail.com

⁴Corresponding author: Bhumandeep Kour, Kognovate Education and Research, Bionest, Avishkaran (NIPER), Hyderabad, Telangana 500037, India. bhumandeepkour@gmail.com

1. Introduction

The advent of Next-Generation Sequencing (NGS) has substantially increased the volume of genomic data available for genetic research and healthcare, providing new opportunities for personalized medicine. Over the past two decades, significant advancements in genomics and epidemiology, supported by large-scale biobanks such as the UK Biobank, have led to the discovery of numerous genetic variants associated with disease susceptibility and treatment response (1). However, despite these foundational breakthroughs, a persistent challenge remains: the comprehensive identification and interpretation of subtle, polygenic, and context-dependent risk factors underlying complex diseases. This challenge contributes to the phenomenon often termed “missing heritability” (2). Traditional statistical and computational approaches frequently fail to fully capture these non-linear biological interactions and the dynamic interplay between genetic predispositions, environmental factors, and lifestyle choices (3).

In response, AI, encompassing ML, DL, and advanced LLMs, has increasingly been applied to genomic research and healthcare. These approaches can process high-dimensional, multi-modal data, ranging from genomics and transcriptomics to medical imaging and electronic health records, to autonomously uncover hidden patterns and integrate genomic findings with clinical phenotypes (4, 5). AI applications are becoming integral to genetic research and precision medicine; for example, tools for variant calling (e.g., DeepVariant) (6), protein structure prediction (e.g., AlphaFold) (7), and genome editing (e.g., CRISPR-ML) (8). Recent network-based and pathway-centric

studies further indicate that AI can identify hidden regulatory subnetworks, non-differentially expressed but functionally essential genes, and disease-associated redundancy patterns that may be overlooked by conventional analytical frameworks.

Although the use of AI in genetic healthcare has increased, its translation from “bench to bedside” faces critical impediments. Existing literature often dissects AI’s computational principles in genomics but typically lacks a comprehensive, cross-domain synthesis that explicitly connects algorithmic development with real-world clinical applications and patient outcomes. Current genomic interpretation methods predominantly rely on static frameworks for classifying genetic variations, resulting in inconsistent predictions, a high prevalence of “Variants of Uncertain Significance” (VUS) (9), and suboptimal diagnostic yields, particularly for rare diseases where up to 50% of cases remain undiagnosed (10).

A major concern is the pervasive ancestral bias within current genomic datasets, where individuals of European descent are vastly overrepresented. This “diversity gap” leads to AI models with reduced predictive accuracy and poor generalizability across diverse ancestral populations, which consequently exacerbates health disparities (11-14). Furthermore, the fragmented nature of data pipelines, limitations in data diversity, and the absence of standardized evaluation frameworks impede robust clinical deployment. Ethical and regulatory considerations, including model interpretability, algorithmic bias, data privacy, and the need for explainable AI, further

compound these challenges, highlighting the need to evaluate AI systems within patient-centered and outcome-driven models rather than purely computational paradigms (15, 16). The current paradigm largely treats genomic and clinical data as static snapshots, failing to account for the dynamic interplay of genetic predispositions with gene regulation, environmental exposures, cellular states, aging, and disease progression over time. This fragmented understanding may limit the extent to which personalized medicine can be implemented. Moreover, most existing approaches remain insufficiently equipped to model how biological systems dynamically reorganize through compensatory pathways, redundant subnetworks, and adaptive network buffering during disease progression.

The review has three aims. First, it examines recent AI models used for genetic variant detection, interpretation, and validation, including their strengths, limitations, and potential for integration. Second, it elucidates the emerging role of AI in the discovery of novel biomarkers and precision therapeutics, highlighting how these advancements contribute to personalized treatment strategies. Third, it proposes a comprehensive framework that connects AI research with real-world clinical use. The framework incorporates essential ethical, regulatory, and biological validation steps, thereby outlining a roadmap for the safe and effective integration of AI into next-generation precision genetic healthcare. In particular, this review explores the emerging hypothesis that redundancy-informed biological modeling, including overlapping gene clusters, paralog systems, pathway interactions, and differential network

structures, may improve AI-driven interpretation of disease mechanisms and support more context-aware genomic decision-making.

The novelty of this review lies in its systems-level perspective that considers multi-omics integration, clinical translation, and regulatory insight into a single cohesive framework. Unlike existing literature that predominantly focuses on technical model performance, this review emphasizes the translational continuum, from data generation and model training to patient-level outcome assessment. It highlights AI's capacity not only for computational efficiency but also for enabling hyper-personalized medicine, where real-time genomic, behavioral, and environmental data dynamically inform healthcare decisions (17). Furthermore, the review introduces a redundancy-aware and pathway-centric perspective of genomic AI, proposing that adaptive AI frameworks integrating network-level biomarkers, graph-based learning, and multimodal longitudinal data may better capture functional compensation and system-level disease biology than traditional static genomic models. Ultimately, the integration of AI into genetic healthcare may contribute to a transition toward a predictive, preventive, personalized, and participatory (P4) model of medicine (18), ensuring equitable access and improved outcomes across global populations.

2. Review scope and literature search

This article is an analytical review providing an integrated synthesis of AI applications in genetic healthcare. A structured literature search was performed using PubMed and Google Scholar for articles published between

2022 and 2026, supplemented by a few highly cited pre-2020 landmark papers (2016, 2018, and 2019) and recent preprints highly relevant to emerging AI techniques. The review was conducted from January 2025 to January 2026. The keywords and Boolean strings included: “artificial intelligence AND genomics”, “AI in diagnosis”, “Deep Learning genomic analysis”, “polygenic risk score ML”, “AI in biomarker discovery”, “AI in rare diseases”, “AI in personalized medicine”, “personalized treatment”, “explainable AI genetics”, “functional redundancy”, “redundant pathways in disease”, “non-differentially expressed driver genes”, “paralog compensation”, and “network analysis”. The review focused on AI, ML, or DL when applied to genomics, variant calling, risk scoring, pathology, clinical diagnostics, and therapeutic advancement. Studies were included if they were focused on AI or ML applications in genomics, variant detection, biomarker identification, or clinical decision support; were published in English; and presented human genetic or clinical data. The initial search yielded 3900 records. After removing duplicates and screening titles/abstracts for relevance, 364 articles were assessed.

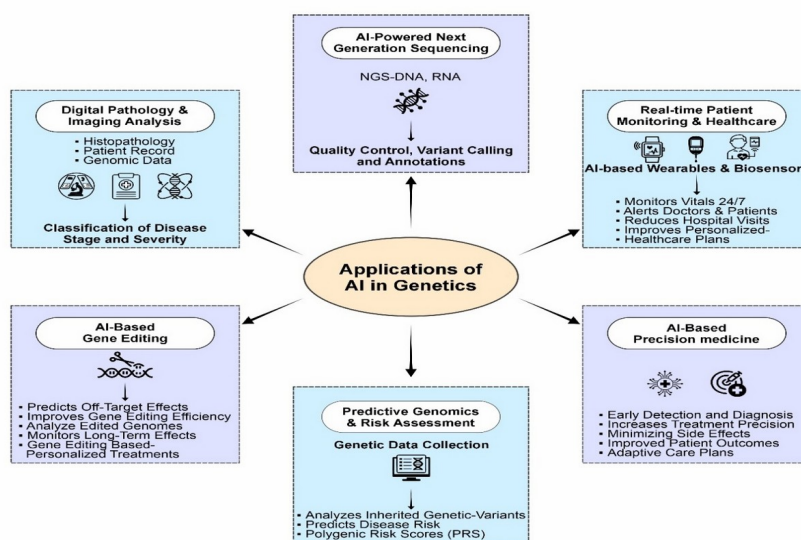


Figure 1. The figure illustrates major applications of AI in different fields of genetics: AI-powered next-generation sequencing (NGS-DNA, RNA) for quality control, variant calling, and annotation; real-time patient monitoring and healthcare using AI-based wearables and biosensors to monitor vitals, alert doctors and patients, reduce hospital visits, and improve personalized healthcare plans; AI-based precision medicine for early detection, increased treatment precision, minimized side effects, improved patient outcomes, and adaptive care plans; predictive genomics and risk assessment for analyzing inherited genetic variants, predicting disease risk, and calculating polygenic risk scores (PRS); AI-based gene editing for predicting off-target effects, improving editing efficiency, analyzing edited genomes, monitoring long-term effects, and enabling personalized treatments; and digital pathology and imaging analysis for integrating histopathology, patient records, and genomic data to classify disease stage and severity.

3. AI in genetic healthcare

AI's potential is harnessed across different

domains of genetic healthcare, including genomic data analysis, diagnosis of complex

and rare diseases, mutation identification, risk assessment, gene editing, targeted drug discovery, novel biomarkers, and precision medicine. These applications can improve clinical insights and healthcare outcomes (Figure 1). AI application in clinical genetics can be conceptualized as a multistep workflow. First, genomic and other relevant clinical or omics data are collected and subjected to quality control, normalization, cleaning, feature extraction, and batch correction. Appropriate AI models (e.g., ML, DL, CNN, RNN, NLP, etc.) are then selected and trained using supervised or unsupervised learning approaches (4, 5). The performance of these

models is subsequently evaluated through cross-validation. The resulting predictions can then be examined using XAI, SHAP, LIME, if-then rules, and transparency frameworks, before being clinically implemented for applications such as drug discovery, rare disease identification, facial pattern recognition, targeted therapies, personalized therapy, treatment optimization, etc. After real-time implementation, the output is analyzed, and clinical decisions are made to formulate preventive strategies and precision medicine. Finally, the feedback is collected from patient outcomes for model refinement to ensure continuous learning (Figure 2).

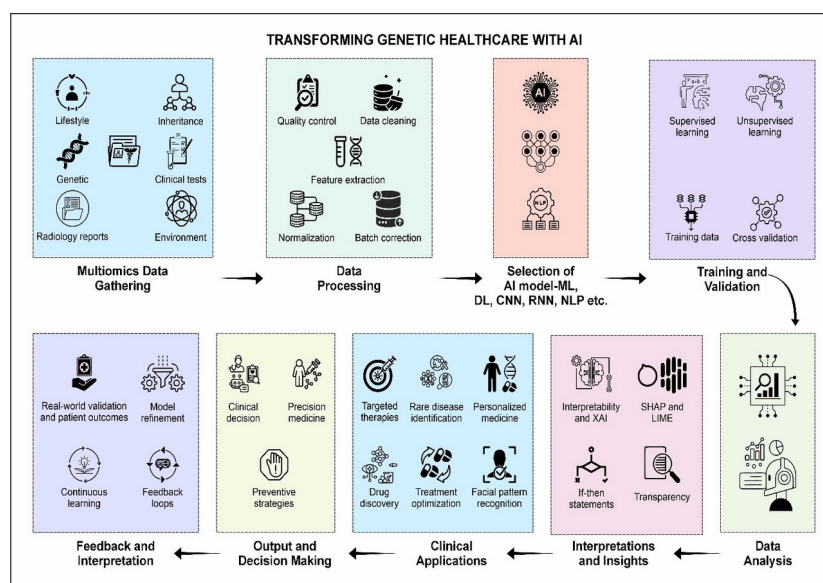


Figure 2. Process of AI-driven genetic healthcare by transforming multi-omics data into actionable clinical insights for precision medicine and developing targeted therapies: data collection (multi-omics data); data processing (quality check, filtering, normalization, and batch correction); model selection (DL, CNN, NLP, LLP, RNN, etc.); training and validation; data analysis and interpretation; clinical application, output, and decision-making; and feedback-driven model interpretation.

3.1 AI in genetic data analysis

AI is employed in the analysis of sequenced data from high-throughput techniques like NGS. The following sections provide a

comprehensive overview of how AI-based tools are being used for genomic data analysis.

3.1.1 Genome sequencing

The increasing volume of data generated by NGS and the demand for more accurate analysis methods have led to a greater reliance on AI and ML. AI can automate genomic analysis pipelines and integrate multiple computational tools. It can also improve quality control and facilitate simultaneous variant calling and identification from raw sequenced data (19). Currently, whole-genome sequencing is not routinely used due to its high cost, time-consuming data interpretation challenges, the limited immediate value of uninterpreted data, the influence of the microenvironment on treatment decisions, and complexities in selecting appropriate target therapies (20). Genomic data analysis pipelines are intricate, and accurately identifying disease-relevant variants from millions of error-prone sequences remains a significant challenge (21). Deep convolutional neural network-based methods, such as DeepVariant, enhance variant identification by analyzing read pileup images, which are graphical representations of sequenced reads aligned to a reference genome. These models estimate statistical relationships for genotype calls, demonstrating the capability of DL to overcome limitations of traditional methods in variant calling (6). DeepVariant is flexible and applicable to various sequencing methods (e.g., Ion Ampliseq exomes, 10x Genomics) and DNA types (human and non-

human) (20). Despite its utility, systemic errors can occur in NGS data analysis, and DeepVariant itself has limitations in annotating variant details.

To address these limitations, tools like Clairvoyante were developed to provide more comprehensive variant details. The tool predicts zygosity, indel length, variant type, and alternate alleles, and processes both long-read sequencing data as well as short-read datasets within two hours. Subsequently, DeepSV, a DL-based tool, was introduced for visualizing sequenced reads to call long deletions and was reported to be more accurate and efficient than DeepVariant and Clairvoyante. This highlights DL's efficacy in complex variant calling beyond single-nucleotide polymorphisms (22). Google also developed DeepTrio, an extension of DeepVariant, to analyze genomic variations in family trios. This tool aids in detecting Mendelian diseases, studying genetic trait transmission, and identifying novel mutations present in a child but absent in parents. However, high computational resource demands, data quality issues, and the need for specialized expertise constrain the application of DeepTrio for interpreting variant calling results (23). Table 1 summarizes key tools used in variant calling.

Table 1. Comparative analysis of standard tools used in variant calling

AI Tools	Application Domain	Input/Output Format	Training Data	Strengths	Limitations	Current use/Development status	Ref.
DeepVariant	Converts aligned reads into images to identify SNPs and indels.	Input: FASTQ/BAM Output: VCF	Genome in a Bottle (GIAB) truth sets; Illumina WGS/WES data	High accuracy, automated variant filtration, Reduces the burden of post-processing, Applicable to both short-read and	Requires high computational processing, Works primarily for human data and very few other genomes, Needs to be	Widely used in genomics pipelines (Illumina, Google Health)	(24)

				long-read sequencing data	trained on other organisms' genomes		
Clairvoyante	An advanced multitask caller used for noisy long-read sequencing data	Input: pileup images (from BAM); Output: VCF	HG001–HG005 GIAB datasets (Illumina, PacBio, ONT)	Specially designed to detect SNPs/indels from data with more errors (e.g., Oxford Nanopore and PacBio long reads) Good performance with F1 score between 87%-99%, Fast, No need to retrain; can work on any genome	Good for SNPs or small indels but struggles to identify multiple indels in the same region, repetitive sequences, and overlapping of multiple mutation types	Research stage; benchmarked in multiple studies	(22)
DeepSV	Variant caller for long deletions greater than 50 base pairs	Input: BAM (short reads); Output: SV VCF	Simulated and GIAB benchmark datasets	More accurate and efficient than traditional SV callers, Combines read depth and split reads	Applicable only for deletions, not for insertions, inversions, and translocations	Research-level; experimental validation ongoing	(25)
DeepTrio	An extension of DeepVariant that calls variants from family trios, both parents, and the child	Input: FASTQ/BAM (trio data) Output: VCF	GIAB trios (Ashkenazim, Chinese families)	By integrating family variant data, ensures more accuracy and identification of novel mutations, and detects Mendelian inherited variants	Requires high computational work and power, limited to humans only, and results require lab validations	Used in Mendelian and clinical variant discovery projects	(26)
GATK	Used for calling haplotypes and as a statistical variant caller for short-read sequenced data	Input: FASTQ/BAM Output: VCF + QC metrics	1000 Genomes & GIAB datasets	Widely used, more advanced, Well-annotated, open source, Fast and offers further filtering and annotations	Low accuracy in regions with multiple mutations and compared to DeepVariant, Requires more manual filtration and threshold selections	Used in validated workflows	(27)
DeepVariant + GLnexus	Creates a unified file of genetic data of a large cohort and calls the variants (Cohort callset)	Input: gVCF (from DeepVariant); Output: combined VCF	GIAB and cohort WGS datasets (NIST, Illumina)	Enhances DeepVariant by combining GLnexus with it, cheaper	Requires large data storage and has scalability issues, Dependent on external tools for post-processing, Chances of false positives	Used in population-scale and research projects (e.g., UK Biobank, All of Us)	(28)

3.1.2 AI-driven SNP and mutation analysis (identification, interpretation, and prediction of genetic variations)

AI has contributed to advances in SNP and mutation analysis, improving the identification, interpretation, and prediction of genetic variations in genomics (19).

3.1.2.1 Identification of genetic variations

The process of SNP and mutation identification begins with variant calling, which identifies differences between the sequenced genome and

a reference genome. AI-based variant calling tools, leveraging ML and DL algorithms, enhance the accuracy, efficiency, and scalability of detecting single nucleotide polymorphisms, insertions/deletions (indels), and structural variants compared to traditional statistical approaches (29). Building on approaches such as DeepVariant and DeepSV described above, recent developments have expanded the application of AI to diverse variant types and sequencing contexts (6, 20). Recent advancements include tools like Fabric

GEM, which accurately identifies structural genetic variants for rare disorders (30), and other DL methods that have improved the detection of both small-scale and large-scale variants (31). DeepVariant-AF was developed to address some of these limitations by incorporating population data. It can improve the detection of low-coverage and rare variants while reducing false-positive and false-negative calls (32). SNooPer, an ML-based tool, identifies somatic variants from low-depth NGS data, especially in cancers. It eliminates user-defined thresholds, which enhances its flexibility and mitigates platform biases (33). This emphasizes the need for AI tools designed to account for genomic diversity and different sequencing platforms.

3.1.2.2 Interpretation and prediction of genetic variations

Once variant calling is completed, variant prioritization is essential for downstream analysis. This process involves filtering variants according to parameters such as read depth, variant allele fraction, read quality, base quality, and mapping scores (34). Heuristic filtering can reduce false positives; however, overly stringent thresholds may exclude true mutations, whereas overly relaxed thresholds may introduce false positives (35). Mutational analysis can then be used to assess the biological relevance or clinical significance of prioritized mutations (36). Predictive models such as Support Vector Machines, decision trees, logistic regression, and Bayesian networks enable accurate forecasting of disease

outcomes and individual risk identification (37).

For interpretation and prediction, AI provides advanced tools for assessing variant pathogenicity, moving beyond linear regressions to sophisticated DL models (38). The American College of Medical Genetics and Genomics and the Association for Molecular Pathology guidelines now integrate computational predictions as crucial evidence for classifying clinical variants (39). For example, Google DeepMind's AlphaMissense predicts the pathogenicity of millions of missense variants, which are critical for understanding disease etiology (40). An AI system like DiagAI also combines ML models with phenotypic information to prioritize causal variants in rare diseases, improving diagnostic yields (10, 41). Despite these breakthroughs, challenges persist regarding the scarcity of high-quality labeled datasets, the omission of genotype-phenotype associations in model training, and the computational resources required for complex analysis (42, 43). Ongoing research continues to refine these AI-driven approaches, focusing on improved interpretability, data sharing, and global collaborations to further enhance precision medicine (44). Tables 2 and 3 provide a comprehensive analysis of different AI-based tools used for variant detection and prioritization. The comparison shows that no single tool performs equally well across all data types, underscoring the need for tailored selection tools.

Table 2. Comparison of variant-detection tools in terms of performance, data requirements, strengths, limitations, and development status

AI Tool	Accuracy	Application Domain	Input/output	Training Data	Strengths	Weakness	Current use/ Development status	Ref.
VarScan 2	More than 89%	Somatic and germline variant calling in tumor–normal samples	Input: SAM/BAM Output: VCF	Benchmark cancer genome datasets (TCGA, GIAB)	Detects high somatic variations and indels, Fewer false positives	Precision in detecting indels varies from moderate to high, with a low F1 score Reliability is poor when applied to population screens or allele frequency studies	Widely used in oncology pipelines.	(45)
VariantTransformer	89.26%	Normalizes and harmonizes variant representations	Input: VCF/JSON Output: standardized variant notation	Ensembl VEP and ClinVar datasets	Flexible, Automated variant prioritization from VCF files with low coverage data in 10x-15x, Trained with large datasets including SNPs and short InDels.	Limited diversity in training data because it was derived primarily from the Genome in a Bottle consortium, Platform dependency (only Illumina sequencing)	Research/ pipeline integration stage	(24)
DeepVariant-AF	-	DL model integrating allele frequency.	Input: FASTQ/BAM; Output: VCF with AF-based scoring	GIAB + population AF data (gnomAD)	Includes allele frequency from the 1000 Genomes Project, Trained for both high and low coverage data of WGS and WES, Can also be implemented for rare variants (homozygous) and clinically verified mutations (disease-associated)	Limited to individuals of European, Ashkenazi, and Asian ancestry Optimized for Illumina sequencing data, and performance might vary for other platforms like PacBio or Oxford Nanopore, Cannot map variants out of 1000Genomes database (allele frequency=0), which might be novel, Lack of diversity in the allele frequency panel	Advanced research tool	(32)
SNooPer		Somatic variant caller using ML ensemble	Input: BAM; Output: VCF	Simulated tumor–normal datasets	Identifies somatic variants from low-depth data Outperformed VarScan2, JointSNVMix and MuTect, Removes platform-specific biases and	Training data bias in prediction (quality and representativeness) The training phase is time-consuming Poorly trained platform bias can persist, Strand bias	Research-level tool	(33)

					systematic errors	Unverified ability to detect the InDels		
Deep SNP	80-92%	Structural variant and CNV detection using deep CNNs	Input: Copy-number or array intensity data Output: variant calls	Public CNV/CMA datasets	Detects breakpoints in cancers	Trained on small training data of SNP array, Doesn't apply to NGS data Requires manual data curation	Experimental validation stage	(46)
NanoSNP	-	Variant calling from Oxford Nanopore sequencing	Input: FAST5 / FASTQ; Output: VCF	Nanopore sequencing (NA12878, GIAB)	Calls low-coverage Oxford Nanopore Technologies (ONT) sequencing data, Predicts zygosity and genotype, and helps to map MHC regions, Haplotype refinement, Includes base count, quality scores, and mapping quality	Not applicable for other sequencing platforms, such as PacBio or Illumina Run time is almost 50% longer than the traditional tools	Research/ clinical trial testing	(47)
DNAscope	On PacBio, >99%, for Illumina, ~57%	ML-enhanced variant caller	Input: BAM/FASTQ Output: VCF	Diverse human WGS/WES datasets	Combines ML-based genotyping and GATK, Rapid runtime, Applicable for long reads	Low efficiency for indels on Illumina F1 score is lower than DeepVariants	Used in commercial genomic pipelines (Sentieon)	(48)
Clair 3	95%	CNN-RNN hybrid variant caller for long reads	Input: BAM (PacBio/ONT) Output: VCF	GIAB + long-read benchmark datasets	Haploid variants Its extension, Clair3-Trio, identifies variants from family trio sequencing (mother, father, and child) Fast pileup and alignments	Clair3 depends on long-read data from Nanopore or PacBio, It lacks optimization for somatic variants	Commercial product for long-read pipelines	(49)
BCFTools	Illumina (95%) and PacBio (81%)	Variant manipulation, filtering, and consensus calling	Input: VCF/BCF Output: VCF	General genomic data	More efficient and faster for Illumina	Slow for a large population study Not applicable for Oxford Nanopore	Research standard	(50)
Platypus	SNVs (98%) and Indels (61%)	Bayesian haplotype-based variant caller	Input: BAM; Output: VCF	Human and cancer genomes (1000G, TCGA)	High efficiency in detecting SNVs Fast	Poor at recalling SNVs and low sensitivity for Indels	Research-level; widely benchmarked	(51)
FreeBayes	99%	Bayesian genetic variant detector	Input: BAM; Output: VCF	1000 Genomes & population datasets	Aligns the reads and identifies haploid variation, SNPs, Indels, and complex variants, Flexible and can be applied to polyploidy	Produces more false positives in non-human Moderate recall for Indels, Too sensitive to depth and quality of data to enhance the performance	Extensively used in research workflows.	(52)

Table 3. Comparative analysis of AI-powered tools used for variant prioritization

AI Tools	Accuracy	Application Domain	Input/output	Training Data	Strengths	Limitations	Current use/ Development status	Ref.
deepCNNvalid	96%	Pathogenicity classification of genetic variants using deep CNNs	Input: Annotated VCF or variant features Output: Pathogenicity score (Benign/Pathogenic)	ClinVar, HGMD, and 1000 Genomes variant databases	Refine somatic variants automatically from tumor samples, Read BAM files	Limited to leukemia samples, not for solid tumors Don't map InDels Accuracy is dependent on manual annotations	Research-level model; benchmarked on public datasets	(53)
EnGenome VarChat	80%	Clinical variant interpretation and report generation	Input: VCF / annotated variants Output: Clinical interpretation report with ACMG classification	Integrated in EnGenome's proprietary knowledge base + ClinVar	Provides literature-mined support, Interactive	Sometimes returns evidence without a clear source and no clinical validation Lacks in-depth querying	Used in diagnostic labs; CE-IVD marked (Europe)	(54)
DeepPVP	92% with ROC-AUC 0.959	Prioritization of pathogenic variants using phenotype-aware DL	Input: Annotated VCF + patient HPO terms Output: Ranked pathogenic variants	ClinVar, OMIM, and HPO phenotype databases	High performance in variant ranking Identifies novel disease associated genes, Relates phenotype data during scoring of pathogenicity	Needs high storage space, Dependent on phenotype data Efficiency gets reduced when bulk analysis is done	Research / translational stage	(55)
Fabric GEM	85%	AI-based clinical decision support for rare disease variant prioritization	Input: VCF + clinical phenotype Output: Gene/disease ranking and diagnostic report	Curated clinical datasets + HPO + OMIM + gnomAD	Helps to prioritize neonatal or pediatric rare diseases, Combines phenotype data (HPO), ClinVar, and OMIM data Identifies SNVs, small Indels	Closed source Dependent on the quality of phenotype data	Commercial decision-support software	(30)
Franklin (Genoox)	64%	Cloud-based AI platform for variant interpretation and data sharing	Input: VCF / panel data Output: Interactive clinical classification (ACMG-guided)	ClinVar, gnomAD, internal consortium data	Can prioritize SNPs, CNVs, and Indels, and highlight their related phenotype Web-based platform that requires no pre-installation	Data updating is manual, which slows down its high-throughput feature, Lacks transparency Has data privacy issues as it stores data in cloud space	Widely adopted in genetic testing labs globally	(56)
SeqOne DiagAI	97%	AI-driven platform for genomic diagnostics and report generation	Input: Raw sequencing data / VCF Output: Automated diagnostic report	Hospital-grade clinical databases and validated pipelines	Helps to prioritize pathogenic variants from WGS or WES Identifies, germline variants Includes XAI Offers cross-validation	Commercial Cloud-based platform that allows license holders to access, Not applicable to somatic mutations, CNVs, structural rearrangements Depends on HPO phenotype quality data	Certified (ISO 13485 compliant; CE-IVD)	(57)

Despite significant improvements over traditional pipelines, current AI-based variant calling and prioritization tools do not fully resolve key challenges in clinical genomics. Most benchmarking studies rely on curated datasets and simulated samples, which fail to capture real-world complexities such as variable sequencing depth, mixed ancestries, mosaicism, and technical artifacts (32, 45–48). As a result, strong performance metrics (e.g., high F1 scores) do not always translate into improved diagnostic yield, particularly in underrepresented populations. Performance also varies across variant types and genomic contexts. While deep learning models perform well for single-nucleotide variants and short indels, they remain less reliable for structural variants, repetitive regions, and complex loci (22, 25, 31, 48). Tools such as DeepSV, DeepSNP, and NanoSNP address some of these gaps but are often platform-specific, limiting generalizability (46, 47). Similarly, allele-frequency-aware models like DeepVariant-AF improve detection of rare variants but depend on reference panels biased toward specific ancestries, reducing accuracy in diverse populations (1, 32).

Variant prioritization introduces additional uncertainty. Many tools rely on phenotype data, curated databases, and proprietary knowledge bases, which may embed historical biases and lack transparency (30, 39, 53–57). Limited high-quality training data and incomplete genotype–phenotype mappings further constrain reliability, especially for novel or population-specific variants (39–43). Overall, AI tools should be viewed as components of integrated diagnostic workflows rather than standalone solutions. Future efforts

should focus on combining complementary methods, evaluating performance across diverse populations, incorporating uncertainty and explainability, and prioritizing clinically meaningful outcomes such as diagnostic yield and reduction in VUS (32, 39, 43).

3.2 AI for functional genomics and multiomic integration

AI has become an important computational approach in functional genomics and multiomic integration, offering unprecedented capabilities for deciphering complex biological systems. The integration of high-throughput sequencing technologies with advanced AI algorithms can provide information across multiple molecular levels and may reveal associations that are not apparent in single-omics analyses. AI methods can process large, complex, and heterogeneous datasets, and facilitate the integration of diverse biological information, from gene expression to metabolic profiles. These integrated datasets can support the construction of predictive models of health and disease.

3.2.1 Transcriptomics and epigenomics analysis

AI and ML can enhance transcriptomic and epigenomic analysis. In transcriptomics, DL is particularly valuable for analyzing the high-dimensional, sparse, and noisy data generated by single-cell and spatial omics. These models can identify meaningful patterns that may be difficult to detect using traditional analytical methods (58). Researchers are developing AI models to analyze large RNA-sequencing datasets and predict RNA structures. These models can also help elucidate gene-expression regulatory mechanisms, supporting advances in functional genomics and personalized medicine

(59). For instance, DL-based feature selection methods improve the analysis of single-cell RNA-sequencing (scRNA-seq) data, which is critical for gene dimensionality reduction and downstream tasks like gene marker identification and cell type classification (60). Explainable AI models are also being developed to identify features and patterns in scRNA-seq data that are associated with disease-related processes (61).

In epigenomics, AI and DL algorithms are used to map epigenetic modifications to their phenotypic manifestations, predicting disease markers, gene expression, enhancer-promoter interactions, and chromatin states (62). This includes rapid, high-resolution analysis of DNA methylation profiles, which is particularly relevant in cancer epigenomics for advancing pan-cancer detection and precision oncology (63). AI also plays a role in analyzing histone modifications and chromatin accessibility, relevant to the characterization of epigenetic regulation of gene expression and identifying therapeutic targets (64). Furthermore, ML and DL methods are employed for predicting 3D genome organization from genomic annotations, DNA sequences, and other genomic properties, offering insights into higher-order chromatin structures (65).

3.2.2 Proteomics and metabolomics integration

The integration of proteomics and metabolomics data through AI and ML provides a comprehensive understanding of biological systems that single-omics approaches cannot achieve (66). In multi-omics studies, AI algorithms have been applied to identify molecular patterns and associations

relevant to disease processes, particularly in cancer research. This integration facilitates the discovery of novel biomarkers and therapeutic targets by effectively managing the complexity and heterogeneity of multi-omics datasets (67).

In metabolomics, the rapid analysis of diverse datasets and identification of disease-related biomarkers have improved personalized medical care (68, 69). DL models further enhance accuracy in biomarker identification and disease detection by integrating supervised and unsupervised learning approaches (69). AI and ML are applied in various stages of metabolomics research, including data pre-processing, feature extraction, pattern recognition, and data integration (68). DL is particularly suitable for multiomic applications because it can analyze large datasets from multiple sources and identify complex patterns. These patterns can be used to develop predictive models for disease diagnosis and prognosis, including COVID-19 and many types of cancers (70–73). Integrating across omics platforms provides a more holistic view of cellular responses, potentially revealing emergent properties not visible in single-omics analysis (74).

3.2.3 Systems biology approaches

AI-based approaches are creating new opportunities to model inter-organ communication and systemic biological processes (75). AI-powered multi-scale modeling frameworks integrate multi-omics data across biological levels and organismal hierarchies. These frameworks can also be applied across species to investigate genotype-environment-phenotype relationships under diverse conditions (76–78). These models can

identify novel molecular targets, biomarkers, pharmaceutical agents, and personalized medicines to address unmet medical needs (78).

AI and ML methods are playing an important role in understanding and predicting complex biological pathways and cellular networks (79, 80). While traditional mechanistic models have been indispensable for testing biological hypotheses, ML is suited to inferring patterns from large, high-dimensional datasets. The integration of ML with mechanistic models, often referred to as Scientific ML, is a promising new field that combines their complementary strengths to model complex biomedical systems (81). This approach is particularly valuable when dealing with sparse or limited datasets and aims to overcome the limitations of purely data-driven models by incorporating physical laws and biological interpretability (78, 82, 83). Therefore, AI-based systems biology approaches may provide additional information for characterizing cellular processes and disease-associated network changes, with ongoing efforts to integrate complex biological data and improve interpretability (84, 85).

3.2.4 *Network-based disease modeling*

Network-based disease modeling leverages AI, especially Graph Neural Networks (GNNs), to represent and analyze the intricate relationships within biological systems. GNNs have emerged as powerful tools for AI-aided drug discovery because they operate directly on molecular graphs. This enables them to learn complex topological and geometric features of drug-like molecules (86–88). This approach supports the characterization of molecular interactions and

the identification of disease modules, which are clusters of genes, proteins, and metabolites that are interconnected and co-regulated (89). AI-driven network pharmacology integrates AI and multi-omics technologies to decipher complex bioactive compound-target-pathway networks. This methodology is well suited to therapeutic approaches involving multiple components, targets, and pathways. It may therefore accelerate drug discovery and reduce experimental costs (90). GNNs are used in drug design processes by accurately modeling molecular structures and interactions with binding targets. They have led to breakthroughs in predicting molecular properties, drug repurposing, toxicity assessment, and novel molecule design (91–93). Network-based approaches are also vital for personalized medicine. By constructing biological networks and identifying disease modules, ML models can predict key genes or regulatory hubs controlling biological pathways, aiding in the discovery of essential genes and driver genes in diseases (94–96). AI has contributed to advances in network biology, which provides a framework for understanding complex human diseases. Current research focuses on multimodal data integration, higher-order network analysis, and network-based approaches to personalized medicine (97). AI in network pharmacology offers insights into the intricate relationships between substances and biological systems, facilitating the customization of treatments based on individual genetic profiles (98).

3.2.5 *Conceptual challenges in multi-omics integration*

Although AI has accelerated multi-omics

integration, most current frameworks remain constrained by three conceptual challenges. First, integration strategies are often chosen for technical convenience rather than clinical interpretability. Early integration approaches concatenate heterogeneous features and can capture complex patterns. However, they may produce latent representations that are difficult to map to specific pathways or cell states. This limitation can reduce their usefulness for mechanistic insight and biomarker development (66–70). Late fusion and hierarchical models improve interpretability by operating on omics-specific summaries, but may miss cross-layer interactions that are critical for understanding disease mechanisms (74, 78).

Second, many models are trained and validated on single-cohort, cross-sectional datasets, which may not generalize across laboratories, platforms, and populations. Batch effects, differences in sample preparation, and variation in clinical phenotyping can alter observed multi-omics relationships. Consequently, models may perform well in their original setting but degrade when transferred to new laboratories, platforms, or populations (66–68, 82–85). This problem is particularly acute when integrating high-dimensional modalities such as single-cell transcriptomic or spatial omics, where technical variation may overshadow subtle biological signals. Systematic external validation and cross-site harmonisation remain the exception rather than the norm (73, 78).

Third, many AI-based multi-omics models are still effectively static. They learn associations from snapshots of molecular and clinical data

but rarely incorporate longitudinal trajectories, treatment exposures, or evolving environmental contexts (17, 75–78). As a result, they are better suited for classification or risk stratification at a single time point than for predicting how molecular networks reorganize during disease progression or under therapy. This is a critical limitation for the adaptive genomic intelligence framework proposed in this review. For future dynamics, context-aware models need to incorporate serial measurements and update risk estimates over time, rather than assuming fixed relationships between omics layers (17, 82, 83).

Addressing these challenges is likely to require standardised benchmarks for cross-cohort multi-omics integration, longitudinal study designs, and closer coupling between mechanistic models and data-driven architectures (75–81). In this context, AI should not only maximise predictive accuracy but also prioritize stability across platforms, robustness to population heterogeneity, and the ability to yield biologically interpretable hypotheses that can be tested experimentally (84, 85).

3.3. AI in rare genetic disorders

AI-based methods have expanded the range of computational approaches from genomic analysis to rare and monogenic diseases. By analyzing large genomic data and genome-phenome interactions, AI can support pattern recognition and diagnostic assessment.

3.3.1 *Role of AI in detecting and diagnosing rare diseases*

An integrated AI and ML approach facilitates not only diagnosis but also the categorization

of patients and the analysis of symptoms. AI leverages DL, such as Convolutional Neural Networks, for analyzing medical images and facial patterns to aid in the diagnosis of genetic diseases (99, 100). For analyzing symptoms and categorizing patients, ML models like Bayesian Networks and Support Vector Machines are employed (101, 102). Studies in precision health have demonstrated how AI can rapidly analyze large datasets to support tailored diagnostics. AI can also facilitate biomarker discovery and the identification of genetic susceptibility for conditions like Amyotrophic Lateral Sclerosis (101, 102). Furthermore, AI utilizes Natural Language Processing for text mining from electronic medical records to extract and analyze patient information, transforming unstructured clinical notes into accessible data for research and patient care (103).

Recent advancements in AI-driven tools further streamline the diagnostic process for rare diseases. For instance, SeqOne DiagAI offers an AI-driven platform for genomic diagnostics and report generation, incorporating Explainable AI to prioritize pathogenic variants from whole-exome or whole-genome sequencing data (41, 57). The CLAIRE tool is an AI-based system designed for rare disease diagnosis and variant prioritization, which can also link genetic variant data with drug responses (104). Additionally, an automated genome interpretation model has been developed for such conditions and is routinely used in clinical genetic laboratories (105).

Second-generation AI systems are distinguished by their patient-centric approach, bridging prognostic, diagnostic, and therapeutic

aspects of care (106, 107). They have been proposed for continuous monitoring of patient progress, as exemplified in Gaucher disease, allowing dynamic adjustments to treatment schedules and providing dosage reminders or recommendations for non-drug therapies based on individual patient response (107, 108). Despite its efficiency, the widespread adoption of AI in rare disease diagnostics faces several challenges, including data privacy, bias, security, and ethical concerns (108). To combat data privacy issues, federated learning systems are being developed. These systems allow institutions to collaboratively analyze electronic health records and genomic data without directly sharing sensitive patient information. This can preserve privacy while increasing the amount of data available for model development (109–112). This approach can support the development of larger datasets necessary for rare disease research, including smaller datasets from diverse ancestries to reduce bias (111). Electronic medical records often contain inconsistent formats and incomplete information. Advanced Natural Language Processing techniques, including hybrid frameworks with LLMs, are being developed to extract relevant information from unstructured clinical notes. These approaches can help harmonize clinical information for rare-disease identification (113–118). Medical imaging data, which typically follows standardized formats, also aids in training AI for rare disease diagnosis (108). Beyond direct diagnosis, AI also supports the rare disease ecosystem through tools like AI-powered chatbots and virtual assistants, which provide preliminary information and answer patient queries, reducing the workload on genetic counselors and improving accessibility to

critical information (119, 120). This allows human counselors to focus on the psychosocial, ethical, and familial complexities inherent in rare disease cases (121).

In the context of rare diseases, where etiologies can be complex and heterogeneous, AI-powered multi-omics integration is well suited to capturing molecular complexity. By combining genomic, transcriptomic, proteomic, and clinical features, these systems can identify previously unrecognized molecular signatures that serve as predictive biomarkers for disease progression or therapy response (102, 122). This holistic approach provides a more comprehensive understanding than single-omics analysis (78).

3.3.2 *AI-based phenotype-genotype correlation analysis*

Establishing a robust correlation between genotype and phenotype is fundamental for advancing precision medicine, disease risk prediction, and the diagnosis of rare diseases. This task, however, is impeded by the difficulty in accessing detailed clinical phenotypic data and the inherent complexity of genotype-phenotype associations (70).

DL models have shown strong potential in SNP-based phenotype predictions for both forensic applications and disease risk assessment (123). Traditional genotype-phenotype studies, such as Polygenic Risk Scores and Genome-Wide Association Studies, primarily focus on linear associations with single SNPs (124). ML methods such as gradient boosting, ensemble learning, and Principal Component Analysis can identify complex relationships among multiple SNPs.

These approaches may also help to uncover novel gene-phenotype associations by analyzing complex non-linear interactions (125). DL models, including Convolutional Neural Networks and Transfer Learning, demonstrate strong potential in predicting phenotypes from genomic data (126). For instance, transfer learning allows models to be trained on large, well-annotated cohorts. These models can then be fine-tuned for smaller or underrepresented populations, potentially improving prediction accuracy across various phenotypes (127, 128).

AI also enables multi-class classification, allowing for the prediction of different subtypes of phenotypes or varying levels of disease risk based on SNP variations, moving beyond simple case/control distinctions (125). This capability is helpful for understanding complex traits like eye color, type 2 diabetes risks, and hair color, where gene-environment interactions are diverse across populations (126, 129). A typical ML pipeline for predicting phenotype-associated genes involves three key steps: classifying individuals based on genotype data, prioritizing genes, and cross-referencing findings (126). Phenotype-genotype correlation studies significantly impact rare diseases, risk assessment, precision medicine, and drug discovery (130).

Various AI-integrated tools are available to enhance the understanding of clinical genotype-phenotype correlations. The GPCards database, which utilizes a transfer learning-based approach, maps genetic variants to phenotypes using patient data, offering subtype classification, variant interpretation, gene prioritization, and disease prediction. Unlike

some proprietary databases that may lack detailed individual-level phenotypic data or have accessibility limitations, GPCards provides free genotype-phenotype analysis. This free genotype-phenotype analysis may contribute to models with improved accuracy, generalizability, and interpretability for precision medicine and cross-species genetic studies (126). Other tools often integrate Human Phenotype Ontology or variant annotation workflows to diagnose rare genetic diseases. Supplementary file Table 1 explains the functionality, strengths, and limitations of some key AI-based tools (131-140).

3.4 Polygenic risk prediction

3.4.1 *AI-powered polygenic risk score (PRS)*

Polygenic Risk Scores are statistical measures used to assess an individual's genetic predisposition to complex diseases such as cardiovascular disease, coronary artery disease, cancers, and type 2 diabetes. They help identify high-risk individuals for preventive strategies (141). Traditionally, PRS calculation relied on Genome-Wide Association Studies (GWAS) to identify individual SNPs associated with disease (142). However, these traditional methods were limited by their focus on linear associations and were predominantly developed using genetic data from European ancestries, leading to accuracy gaps and equity concerns when applied to non-European populations (143).

AI-based PRS approaches have been investigated to improve disease risk prediction by incorporating a broader range of genetic and non-genetic factors (142). These models can use ML and DL methods, such as gradient

boosting machines (e.g., XGBoost) and deep neural networks. These approaches can model complex non-linear interactions among SNPs, including gene-gene interactions like epistasis that may not be captured by conventional linear models (144). They can also help in feature selection, identification and exclusion of irrelevant variants. Such approaches may improve the predictive power of PRS models for diseases such as inflammatory bowel disease, breast cancer, type 2 diabetes, and Alzheimer's disease. Recent advancements in AI-powered PRS further enhance the prediction landscape.

3.4.1.1 *Multi-PRS integration*

ML can combine multiple PRS for different or unrelated phenotypes. This strategy may improve predictions and uncover shared genetic pathways across diseases like breast cancer, coronary artery disease, and type 2 diabetes (142).

3.4.1.2 *Integrated AI frameworks*

Models such as PRSmix and PRSmix+, developed using Elastic Net ML algorithms, combine PRS from correlated genetic diseases to provide a deeper understanding of genetic architecture and improve prediction accuracy for complex traits (144).

3.4.1.3 *Geometric DL*

Novel geometric DL models use GNNs to generate scores from GWAS data. These scores can then be mapped onto protein-protein interaction networks to help interpret the clinical significance of disease-associated genes and pathways (145).

3.4.1.4 *Explainable AI*

Integrating ML models like XGBoost with XAI techniques (e.g., SHAP) offers transparency by identifying which genetic or clinical features contribute most to disease risk (146).

Although AI-powered PRS enables personalized risk assessment and early diagnosis by integrating environmental, clinical, and genetic factors, challenges persist regarding the need for diverse and large datasets, improved interpretability, and stronger safeguards for privacy and ethics. Generative AI holds promise for further enhancing PRS models' accuracy for complex genetic interactions by integrating clinical and phenotypic data (147). Efforts are ongoing to develop robust probabilistic risk prediction models using ML algorithms trained on clinical and genomic data (148–150).

3.4.2 Population-scale genomics

AI is becoming an important component in analyzing and interpreting the vast and complex datasets generated by population-scale genomics, which aims to understand genetic variation and disease susceptibility across large cohorts. AI-driven approaches facilitate the integration and harmonization of genomic data from diverse populations, enabling large-scale analyses of genetic variation and population-specific disease risk. They can also support quality control, advanced variant calling, and the generation of unified callsets from large population datasets (19, 28). The importance of these approaches is particularly evident in variant calling, where tools such as DeepVariant+GLnexus combine genetic data from large cohorts to create unified callsets. However, implementing these approaches requires substantial computational

infrastructure, data storage capacity, and scalable analytical frameworks to manage the increasing volume of genomic data (28). AI-powered statistical genetics is also being applied in large biobanks to enhance the discovery of genetic variants associated with disease susceptibility and treatment response (1). A significant challenge addressed by AI is the historical bias in genomic datasets towards European ancestries, which limits the applicability of genetic risk models to diverse populations. As a mitigation strategy, federated learning systems are being utilized, allowing for privacy-preserving collaborative analysis of multi-institutional genomic data. These systems enable the integration of diverse datasets, including smaller contributions from underrepresented ancestries. This can increase representative sample sizes and improve model accuracy while reducing bias. Importantly, these benefits can be achieved without centralizing sensitive patient information (109–112). Furthermore, AI-powered Polygenic Risk Scores, are being refined through population data to identify low-coverage and rare variants, leading to more accurate and equitable genetic risk assessments and enabling personalized medicine strategies at a population level (32).

4. AI in genetic medicine and therapeutics

4.1 AI in genetic therapies: personalized therapies

4.1.1 Gene editing and CRISPR optimization

The advent of Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR/Cas9) has advanced genome editing by enabling precise genetic modifications. This technology has also facilitated the development of single-base

editing for high-throughput screening (151). AI is transitioning it from a potent molecular tool into a clinically defined and customizable platform for advanced therapies. AI integration with CRISPR has been reported to improve the efficiency, precision, and affordability of gene editing. LLMs, such as CRISPR-GPT, are used for automated design processes, leading to the discovery of novel editors and customized base editing strategies (8, 152, 153). These AI models can be used to find new guide RNAs (gRNAs) and optimize their selection for specified target sequences through tools like DeepCRISPR, DeepHF, and CRISTA (8). ML accelerates variant screening and reduces the time and resources required for validating gene editing results, ensuring both precision and speed in clinical applications (151). A key contribution of AI is its ability to improve prediction accuracy for off-target effects, thereby enabling more precise DNA modifications. This is crucial for developing safe and effective personalized treatments tailored to individual genetic profiles (8).

AI-powered CRISPR is being applied to personalized therapeutic strategies, including immunotherapies like CAR T-cell therapy. AI helps identify targets and ensure the efficacy of knocking out immune checkpoints (154). A notable example of patient-specific gene editing involves the recent use of base editing to treat a rare genetic disorder, carbamoyl-phosphate synthetase 1 deficiency, in newborns. This therapy, delivered *via* nanoparticles, demonstrated significant improvement without severe side effects (155). In cancer research, the combination of AI and CRISPR holds promise for more effective treatments by identifying cancer subtypes

through genomic data analysis and modifying immune cells to target oncogenic genes or pathways. Tools like SPROUT, developed using ML, predict how accurately gene editing repairs cells (8). This integration of AI with CRISPR can assist in optimizing the entire genome editing pipeline. It may provide new approaches for manipulating genes and deepening our understanding of the genetic code, and may contribute to the development of more effective and personalized medical interventions. Supplementary file Table 2 discusses key AI-powered tools assisting in genome editing (153,156-163).

4.1.2 *AI in gene replacement therapy*

AI is enhancing gene replacement therapy by improving the design, precision, and delivery of genetic material. AI algorithms can be used to analyze vast genomic datasets to precisely target and modify genetic material, with the potential to improve the accuracy and safety of treatments (164, 165). Recent advancements highlight AI's role in developing optimal formulations for gene delivery systems, considering factors like particle size, surface charge, and composition through data-driven design (166). Predictive modeling, powered by ML, forecasts the behavior of these delivery systems in biological environments, accelerating development and improving success rates in gene therapy trials (166). Furthermore, AI can be beneficial in personalizing gene therapy by tailoring delivery systems to individual patients based on their genetic backgrounds and disease profiles (166). The integration of AI with CRISPR-based genome editing (including base and prime editing) optimizes these technologies to introduce precise changes without relying on

homology-directed repair, enabling personalized treatments for diseases like cancer, diabetes, and Alzheimer's (8). AI also assists in the discovery and optimization of viral vectors, such as AAV, by analyzing parameters like viral tropism and transduction efficiency, leading to vectors with enhanced targeting capabilities and improved stability (167, 168). This allows for the precise modification of genes, such as in the successful base editing for a rare genetic disorder in a new-born (155).

4.1.3 *AI in CAR T-cell therapy*

AI can support CAR T-cell therapy by optimizing multiple stages, from target identification to enhancing therapeutic efficacy and manufacturing. ML and DL algorithms analyze multi-dimensional clinical and omics data to identify optimal target antigens on cancer cells, predict patient responses, and personalize treatment strategies (169, 170). This includes the ability to discern patterns in data to create CARs with highly efficient single-chain variable fragment (scFv) combinations, improving specificity and affinity (170). AI-driven approaches have been reported to improve the design of new CAR constructs and predict their efficacy, addressing critical aspects such as minimizing severe side effects such as cytokine release syndrome and neurotoxicity (171, 172). Generative AI, for example, is being used to engineer *de novo* binding CAR-T cell therapies (173). Moreover, AI, particularly computer vision, enhances image-based sorting technologies, enabling automated extraction of cellular features for real-time classification, which improves throughput and consistency in CAR T-cell manufacturing (174). AI has also been found

assisting in optimizing production conditions and workflows, critical for ensuring the quality and process control of these complex therapies (175, 176). The integration of AI with CRISPR-Cas9 genome editing offers further advancements by enhancing precision and safety in CAR T-cell engineering (171).

4.1.4 *AI in RNA therapies*

AI-based approaches support the development and optimization of RNA-based therapeutics through data-driven design and predictive modeling. These approaches can also contribute to the development of personalized treatment strategies. AI and ML facilitate the rational design of RNA sequences to maximize target affinity while minimizing immunogenicity and off-target effects (177). DL models, trained on extensive datasets of RNA structure, function, and biological interactions, accelerate candidate optimization and reduce development timelines (177). For instance, DL-based siRNA design platforms have been shown to reduce predicted off-target binding significantly compared to traditional rule-based algorithms (177). Generative AI models are also being applied to optimize mRNA secondary structure, leading to improved protein translation efficiency (177). In mRNA vaccine development, AI assists in sequence optimization, predicting stability, translation efficiency, and formulating delivery systems like lipid nanoparticles, which are critical for effective prophylactic and therapeutic applications (178–182). Generative pre-trained language models like GenerRNA are emerging for *de novo* RNA design, capable of generating RNA with stable secondary structures without prior knowledge (183). The ability of AI to analyze and predict the stability

and function of RNA molecules, including structural interactions and conformational dynamics, can be helpful for designing next-generation mRNA therapeutics (184, 185).

4.1.5 AI in gene silencing

AI contributes to the advancement of gene silencing therapies by improving the design and efficacy prediction of small interfering RNAs (siRNAs) and short hairpin RNAs (shRNAs), and by enhancing our understanding of epigenetic gene regulation. ML strategies are being integrated into siRNA research for design and analysis. These approaches can support the development and optimization of siRNA-mediated post-transcriptional gene silencing (186). AI can determine the on- and off-target efficacy of siRNA-based drugs by considering a multitude of factors, including target and off-target binding sites on mRNA and competition between endogenous molecules (185). Novel approaches, such as AttSiOff, utilize self-attention mechanisms to predict siRNA inhibition and off-target effects with improved accuracy and generalization by analyzing siRNA embeddings and local mRNA sequences (187). DL models also predict the efficacy of shRNAmirs by using sequence and context features, which is critical for designing RNA interference drugs (188). Beyond direct RNA targeting, AI is extensively applied in epigenomics, analyzing DNA methylation patterns, histone modifications, and chromatin accessibility to understand gene expression changes without altering the DNA sequence (8, 62, 189). This allows for the identification of key factors influencing gene activation and silencing, supporting the design of guide RNAs for gene regulation (8) and contributing to precision oncology by identifying biomarkers

and predicting responses to epigenetic therapies (190).

4.1.6 Evidence gaps and translational barriers in AI-enabled genetic therapies

AI has improved design efficiency, off-target prediction, and candidate prioritization across gene editing, gene replacement, CAR T-cell therapy, RNA therapeutics, and gene silencing, particularly in preclinical settings (8, 151–153). However, most evidence derives from *in silico* analysis, high-throughput screens, and limited *in vitro* or animal studies, with relatively few prospective clinical trials validating these approaches (151, 155). While early clinical applications, such as base-editing for carbamoyl-phosphate synthetase 1 deficiency, are promising, they remain isolated cases rather than broadly generalizable standards (155). Safety and manufacturability remain major challenges. AI models may reduce predicted off-target effects, but residual risks persist and require experimental validation, especially in underrepresented genomic or tissue contexts (151, 185–188). Similarly, AI-optimized delivery systems must still address immunogenicity, tissue specificity, and scalability, which are not fully captured computationally (166–168). In CAR T-cell therapy, AI enhances target selection and construct design, yet regulatory approval continues to depend on traditional clinical endpoints such as efficacy and toxicity (169–173). Another limitation is the mismatch between AI optimization targets (e.g., editing efficiency, binding affinity) and clinically meaningful outcomes, which do not always correlate with long-term safety or therapeutic benefit (8, 177–179). Additionally, training datasets are often limited in diversity, raising

concerns about generalizability across populations and disease contexts (111–113, 189). Future priorities include prospective validation studies comparing AI-guided and conventional approaches, expansion of diverse and clinically relevant datasets, and stronger interdisciplinary collaboration to establish regulatory standards and ensure safe clinical translation (111–113, 151, 155).

4.2 AI in clinical implementation of precision/personalized medicine

The clinical implementation of AI is rapidly advancing precision and personalized medicine, fostering "hyper-personalized medicine" by tailoring medical therapies to individual molecular characteristics, lifestyle, and environmental factors (17). AI's capacity to analyze vast, complex health datasets enables dynamic wellness plans and promotes a proactive, preventive healthcare model. This transformation, however, necessitates robust infrastructure and collaborative data-sharing frameworks that prioritize patient privacy and address the unique ethical and regulatory challenges of AI in healthcare (17).

AI enhances genomic interpretation, a foundational aspect of personalized medicine for identifying genetic variations influencing drug responses and disease susceptibility (191). AI, including LLMs, plays an important role in classifying genetic variants. Tools such as AlphaMissense, which leverage data from the AlphaFold Protein Structure Database (7), can score the pathogenicity of missense variants with high accuracy. Combining LLM models with AlphaMissense and similar approaches helps classify VUS using ClinVar/ProteinGym datasets, thus evolving variant interpretation

(192). Further contributing to this area, DL-based tools like REVEL, CADD, and SpliceAI offer improved scalability and accuracy by managing splice variants and scoring pathogenicity in conditions such as cancer predisposition and rare Mendelian disorders (192). AI also accelerates DNA and RNA sequencing analysis and variant calling, which is essential for timely clinical genetic insights. Tools such as DeepVariant and NVIDIA Parabricks significantly accelerate these processes, which supports faster results in clinical genetics (7). Beyond accelerating analysis, AI-powered genomic interpretation improves the diagnosis of rare diseases by processing large datasets to classify variations (193), and automated genome interpretation models are now routinely employed in clinical genetic laboratories (105). ML has gained increasing attention in genomic diagnostics for its high accuracy in pathogenic variant identification and its ability to process large volumes of genomic data. (194). However, despite these advancements, significant challenges persist, including the scarcity of high-quality labeled datasets, the omission of genotype-phenotype associations in model training, and the substantial computational resources required for complex analysis (42, 43). Therefore, critical assessment of variant effect predictors remains a focus to ensure their reliability in clinical settings (39).

AI is greatly impacting pharmacogenomics by enhancing drug response prediction and treatment optimization. ML and DL algorithms navigate the complexity of genomic data to elucidate relationships between genetic factors and drug responses (195). This enables the identification of genetic markers, leading to

comprehensive models that guide treatment decisions, minimize adverse reactions, and optimize drug dosages in clinical settings (195). Tools like PharmCAT, PAnno, and PharmVIP can translate patient genotypes into clinical prescriptions, allowing tailored drug regimens based on patient VCF files (196–198). Furthermore, platforms like Chemistry42 utilize AI to quickly prioritize compounds for personalized drug design (122).

The integration of AI and ML into wearable sensor technologies has improved health data science by enabling continuous monitoring, personalized interventions, and predictive analytics (199, 200). These devices provide individual-level data for disease management, geriatric care, and health optimization (201). However, the rapid advancement also raises critical ethical and regulatory concerns. Data privacy, algorithmic bias, informed consent, and the opacity of automated decision-making are paramount challenges (199, 200, 202). Proactive efforts are needed to minimize disparities and promote inclusive healthcare, ensuring accessibility and addressing barriers like technological literacy and affordability (202). Interdisciplinary collaboration is necessary to ensure wearables are not only functional but also intuitive, adaptive, and trustworthy, guided by principles of privacy, fairness, and transparency (199).

AI is also integrating with advanced clinical technologies to improve patient care and support decision-making. In bioelectronic medicine, AI contributes to precision in neural modulation, which involves using electrical stimulation to regulate nerves and provide external sensory input (203, 204). In healthcare

robotics, wearable robotics and collaborative robots (cobots), enhanced by AI, are transforming assisted healthcare. AI optimizes motion control, decision-making, and real-time adaptability in applications such as rehabilitation, eldercare, and surgical assistance (205–208). This human-centered approach aims to bridge the gap between laboratory models and practical clinical deployment (15, 209). Furthermore, an advanced version of Explainable AI has been proposed, which is expected to enable real-time sequencing with high resolution for complex diseases and cancer, addressing the need for transparency in critical clinical applications (210).

AI's ability to process and learn from vast datasets with accuracy and efficiency beyond traditional methods is contributing to advances in disease diagnosis and risk prediction in precision medicine (211, 212). AI can generate insights from multi-dimensional datasets by integrating various omics data, thereby supporting biomarker discovery and prognostic assessments (213). It can help in predicting drug sensitivity in cancer and informing personalized medicine strategies (214). AI is used in Clinical Decision Support Systems to further improve diagnostic precision, risk stratification, and treatment planning (215, 216). The goal is to build trust and ensure clinical adoption through interpretable models that provide transparent explanations of their predictions (15, 209, 214–217). However, the integration of AI in clinical settings requires careful consideration of data quality, privacy, algorithmic biases, and ethical challenges. Stakeholders express concerns about data security, potential mistakes due to "hallucination", and the threat to clinical jobs

(212). These challenges must be addressed for AI to translate its full potential into tangible clinical impact (213, 218). The underlying algorithms used by these AI tools and the accuracy/metrics are summarized in Supplementary Table 3, with supporting references provided therein.

4.3 AI for targeted therapies based on genetic profiles

The traditional process of targeted drug discovery is typically time-consuming, expensive, and high-risk, often taking over a decade and billions of dollars to bring a novel drug to market, with high failure rates in clinical trials due to lack of efficacy (228). AI offers more efficient alternatives to these labor-intensive experimental approaches by providing computational approaches for the identification of drug targets and novel therapeutic options. Computational methods, including pharmacophore screening, reverse docking, and AI-driven ML, are now central to this process (228). In precision oncology, the integration of genomic profiling with AI is becoming more central to developing tailored therapies for patients. AI tools are capable of analyzing genetic alterations, clinical histories, and even digital imaging data to identify optimal targets for treatment design (229).

AI has demonstrated potential in predicting patient responses to treatments, such as immunotherapies or targeted drugs, and can even detect early signs of drug resistance. For instance, AI platforms like Deep 6, Triomics, and Massive Bio can swiftly analyze both genetic profiles and clinical histories, and then accurately match patients to appropriate clinical trials within minutes (230–232). This

automation addresses a significant bottleneck in traditional approaches, where only a small percentage of patients typically participate in clinical trials due to time-consuming manual screening and limited awareness among clinicians. Automated matching may reduce the time required for trial screening and has been associated with improved matching performance in selected studies, particularly in rare disease trials, potentially increasing the likelihood of successful drug development and approval (230–232). Beyond clinical trial matching, AI-based approaches have been proposed for constructing “digital twin” models by combining genomic, expression, proteomic, and immune response data from mini-tumors or organoids. This would help to predict relapse risk and optimal combination therapies for cancers (229). While the clinical integration of such platforms is still evolving, they require high-quality data and robust validation.

AI is also driving drug discovery for genetic disorders, a pressing need given the lack of effective treatments for many rare diseases due to limited genetic and molecular data. It reduces timelines and costs by rapidly screening large chemical libraries and designing novel therapeutic compounds using models such as Support Vector Machines and Random Forests (23). Quantitative Structure-Activity Relationship (QSAR) modeling can identify potential drug candidates, while generative AI explores new chemical spaces to design novel therapeutic compounds, streamlining data analysis and accelerating treatment discovery (23). For example, the Harvard Medical School in 2024 developed TxGNN, an AI model that identified potential therapeutic candidates for 17,000 genetic and

rare diseases, including those without existing treatments, by studying drug-gene interactions and molecular features (233, 234). Another repurposing tool, DeepDrug, forms effective combinations of existing approved drugs for complex conditions like Alzheimer's disease by integrating curated domain-specific knowledge and generating biomedical graphs using GNN (235, 236). The discovery of new molecules is further aided by AlphaFold-based pipelines such as Chemistry42 and PandaOmics, which rely on protein structure prediction (237). Hybrid AI models, which combined genetic algorithms (SELFIES) and DL, identified novel

tyrosine kinase inhibitors by optimizing drug-like properties and predicting their bioactivity with CNN, achieving an AUC of 0.93. This demonstrates the potential to accelerate targeted therapies and reduce the need for extensive wet lab screening (238). These AI-driven approaches offer a more efficient and generalized path for drug discovery across various disease conditions, reducing development timelines from years to months, though challenges related to data integrity and the lack of functional or kinetic annotations remain. An overview of tools assisting in drug identification is presented in Table 4.

Table 4. Comparative analysis of AI-powered tools that assist in the identification of novel drugs

AI Tools	Application Domain	Input/Output	Training Data	Strengths	Limitations	Current use/ Development status	Ref.
Deep 6	AI-driven patient–trial matching for clinical research	Input: EHRs, structured/unstructured clinical notes Output: Matched patient–trial lists	ClinicalTrials.gov, EMR data, investigator feedback	Identifies appropriate trials by matching with patients' both structured and unstructured medical records	Lacks Validation	Deployed in >200 U.S. hospitals; validated in multicenter studies	(228)
Triomics	Oncology-focused clinical and molecular trial matching	Input: Genomic reports (VCF), EMRs, trial databases Output: Ranked trial eligibility results	Public oncology trial datasets + real-world hospital EMR data	Identify appropriate trials by matching with patients' real-world studies	Under validation, and its performance depends on the quality of data	Commercially deployed; used by oncology networks in India, U.S.	(230)
Massive Bio	Precision oncology decision support and trial enrollment	Input: EHR + pathology + genomic profiles; Output: Personalized treatment or trial recommendation	Curated clinical trial registries, NGS test data	Identify appropriate trials by matching with the patient's medical record and genomic data.	Lacks transparency in decision-making	Commercial pilot; operational in >20 countries	(232)
DarwinHealth	Systems-biology–based drug discovery and target identification	Input: Tumor transcriptomics and proteomics; Output: Oncogenic master regulator network models	TCGA, GTEx, proprietary drug–target networks	Evaluates protein activity by analyzing expression data to identify key regulators/biomarkers, treatment response, and helps with gene therapy	Accuracy depends on high-quality transcriptomic data	Translational stage; multiple pharma partnerships	(239)

Insilico Medicine	End-to-end AI platform for drug target discovery and de novo molecule design	Input: Multi-omics + chemical structures; Output: Novel drug candidates	PubChem, ChEMBL, clinical omics datasets	Identifies novel drug targets and designs molecules by integrating multi-omics data	Focuses more on drug discovery and less on clinical implementations	Phase I trial (INS018_055); peer-reviewed preclinical validation	(240)
TrialMatchAI	Large language model-based clinical trial eligibility and recruitment tool	Input: Trial protocols (text) + patient health records; Output: Trial-patient match ranking	ClinicalTrials.gov, EHRs, PubMed corpora	Identify appropriate trials by matching with the patient's medical record.	It specifically focuses on matching trials, but it can't help in predicting treatment.	Prototype / early clinical pilot phase	(241)
CLAIRE	AI-based rare disease diagnosis and variant prioritization	Input: VCF files + clinical phenotype (HPO terms); Output: Prioritized disease and gene candidates with explanation	ClinVar, OMIM, HGMD, and curated rare disease datasets	It helps to predict beneficial cancer therapy by linking genetic variant data and drug responses	Under Clinical validation	Validated in research and pilot studies (BMC Med Genomics, 2023)	(104)
XRare	Automated clinical interpretation and report generation for rare disease WES/WGS data	Input: VCF or FASTQ + HPO phenotype data; Output: ACMG-classified variants and diagnostic report	ClinVar, gnomAD, OMIM, HPO databases	Helps in the diagnosis of rare diseases and decision-making for appropriate precision therapy selection	Lacks transparency in decision-making, is not applicable for other diseases, has less generalizability, and is under validation	Deployed in diagnostic labs; CE-IVD certification (Europe, 2023)	(240, 140)

4.4 Adaptive genomic intelligence: proposed design principles

AI is increasingly integrated across genomic medicine, including variant interpretation, pharmacogenomics, clinical trial matching, and drug discovery (30, 191–198). However, most current systems remain fragmented, optimized for isolated tasks, trained on static datasets, and evaluated mainly using surrogate metrics such as AUC or enrichment scores. To address these limitations, we propose an adaptive genomic intelligence framework based on four complementary design principles. First, the genomic AI should adopt longitudinal and multimodal learning rather than relying on single time-point data. Integrating serial omics measurements, treatment history, environmental exposures, imaging, and EHR time-series data would enable models to capture dynamic biological changes and

disease progression over time (17). Second, ancestry-aware fairness and population diversity should be core design objectives. Existing genomic datasets remain heavily biased toward European ancestry, limiting model generalizability across populations (109–113). Federated learning offers a practical strategy for integrating geographically and ancestrally diverse datasets while preserving privacy, though model performance should still be routinely evaluated across ancestry-specific cohorts and clinical contexts (109–112).

Third, pathway- and network-level interpretation should complement single-gene analysis. Many biologically relevant signals emerge from pathway interactions, regulatory modules, and redundant subnetworks rather than isolated variants (66, 73). Accordingly, adaptive systems should incorporate pathway

activity scores, network centrality measures, and redundancy-aware biomarkers to better capture system-level disease mechanisms and support experimental validation. Finally, evaluation should prioritize clinically meaningful outcomes in addition to conventional machine learning metrics. Relevant endpoints include diagnostic yield, reduction in VUS, time-to-diagnosis, recruitment diversity in clinical trials, and prospective success of AI-prioritized therapeutic candidates (39, 105, 191–194). Collectively, these principles provide a framework for developing adaptive, interpretable, and clinically relevant genomic AI systems.

4.5 AI-driven drug discovery for genetic disorders

AI has enhanced disease analysis and understanding of disease mechanisms by uncovering hidden patterns in vast data, which might not be identifiable by human researchers (232). Clinical trials of AI-aided discovered drugs are growing in popularity; examples include INS018_055 (for idiopathic pulmonary fibrosis) and EXS-21546 (for solid tumors) (242). DL models like generative adversarial networks (GANs), transfer learning, and neural networks are being used in aging research, drug prediction based on transcriptional data, and small molecule design. For instance, in the case of amyotrophic lateral sclerosis (ALS), a combination of different bioinformatic tools and DL models was trained with multiomic data, which identifies potential genetic targets by prioritizing druggable genes (228). To identify genes linked to multiple age-related diseases, the models learned and extracted patterns from gene and protein ontology,

protein-protein interactions, and biological pathways. A DL ensemble trained on transcriptomic data from 12,000 embryonic and adult cells was used to identify key targets for controlling embryonic-fetal transitions, tissue regeneration, and cancer. Also, the LLMs in biomedical text mining models such as BioGPT and ChatPandaGPT can analyze a vast amount of literature to identify associations among diseases, genes, and biological processes. Such models can help to create hypotheses for complex diseases like cancer, autoimmune disorders, and neurodegenerative disorders. However, they may perpetuate biases and lack novelty. Combining LLM-based models with experimental validation can assess these hypotheses and support the development of reliable solutions (228, 243–245). AI can also analyze genomic data of virus outbreaks to identify important viral proteins and host factors that can be targeted for inhibiting viral replication and developing effective antiviral drugs (246). Collectively, these applications demonstrated the potential in developing synergistic therapeutic targets for complex diseases involving multiple pathways (228).

Beyond target selection, AI-based approaches have also been applied to identify therapeutic candidates for genetic and rare diseases. For example, QSAR-based models have been applied to identify potential therapeutic candidates such as riluzole for ALS, while AI models can also predict physicochemical and ADMET properties of candidate compounds. These approaches also streamline data analysis, reduce trial-and-error, and speed up treatment discovery, such as for COVID-19 (23). To date, tofacitinib, Niraparib, Empagliflozin, Doxercalciferol, and Baricitinib were five

identified compounds that can target three key mechanisms of AD: mitochondrial dysfunction, glucose metabolism, and neuroinflammation. Such models offer a new direction towards a more practical approach through designing multi-targeting treatments in complex diseases (235, 236).

5. AI in clinical genetic practice

5.1 AI in diagnostics and risk prediction

AI has expanded the capabilities of medical diagnostics and risk prediction by facilitating precise and efficient analyses and identifying subtle indicators of disease across various domains. AI can process and learn from vast datasets with accuracy far exceeding that of traditional methods. This is revolutionizing how diseases are diagnosed and how individual risks are assessed (211, 212).

AI-based image-analysis methods can extract quantitative and morphological features from medical images. DL, particularly Convolutional Neural Networks, is employed to predict genetic alterations, tumor mutation burden scores, and survival status directly from hematoxylin and eosin-stained histopathology slides. This capability has enabled the detection of morphological patterns indicative of genetic mutations in cancers, such as lung and colorectal cancers, and facilitates molecular subtyping, including predictions for homologous recombination deficiency and alterations in mesothelioma (22, 99). Beyond pathology, AI's diagnostic utility is evident in radiology, where ML algorithms provide statistical predictions from image classifications, such as tools that can interpret chest radiographs to detect multiple diseases

(247). The first FDA-approved AI-based tool in ophthalmology, designed for the diagnosis of diabetic retinopathy, further highlights AI's impact in this area (247).

For genetic and rare disease diagnosis, AI addresses the challenges posed by limited data and complex clinical presentations. Tools like GEM assist clinicians by analyzing whole-exome sequencing data, correlating it with patient medical records, and using reliable genetic disease databases (e.g., OMIM, HPO, ClinVar) to accurately predict disease-associated genes. This system can filter errors and unrelated mutations, is flexible enough to handle large structural genetic changes to small SNVs, and can even operate when family data is unavailable, making rare genetic disease prediction faster, cheaper, and more reliable (30). LLMs are also being used in analyzing rare disease text data, uncovering insights and patterns critical for diagnosis (116). Furthermore, AI-powered genomic interpretation is improving the diagnosis of rare diseases by processing large datasets to classify variations, with automated genome interpretation models routinely employed in clinical genetic laboratories (105, 193).

In biomarker discovery and risk prediction, AI integrates multi-omic datasets (genomic, proteomic, metabolomic) to identify previously unrecognized molecular signatures that serve as predictive biomarkers for disease progression or therapy response (77, 78, 122, 213). For instance, DeepDeeProM, a DL model applied to pan-cancer proteomic datasets, examines protein-level data to uncover markers associated with drug response and key genes for cancer cell survival, offering deeper

insights into disease activity than traditional expression-based models (248). In metabolomics, AI and ML applications, such as the ML-based XGBoost classifier used to identify biomarkers (e.g., LYVE1, REG1A, REG1B, TFF1, plasma CA19-9) for pancreatic cancer, demonstrate high accuracy (89%) and sensitivity (91%) compared to conventional statistical techniques in differentiating between healthy controls, benign cases, and ductal adenocarcinomas (69, 249, 250). For immediate risk stratification, such as during pandemics like COVID-19, Random Forest models efficiently and accurately identify biomarkers to prioritize treatment and improve outcomes in emergency settings (251). AI also aids in the classification and diagnosis of neurological disorders like Amyotrophic Lateral Sclerosis using CNN and SVM tools, leveraging complex pipelines involving GANs for data preprocessing, AlphaFold for protein structure prediction, and IBM Watson for omics and gene-level analysis (101). For early cancer detection, AI-based screening can improve accuracy in breast cancer biopsy screening, DL models process MRI data to differentiate between benign and malignant brain tumors, SVM models predict ovarian cancer risk using clinical and genetic data, and Bayesian models assess lung cancer risk based on factors like smoking, age, and genetics (191). AI further facilitates biomarker identification for various diagnostic purposes, including predicting a patient's ancestry and even diagnosing certain genetic diseases through facial recognition (100, 123).

Although AI-based approaches show potential for improving diagnostic and risk-prediction workflows, their reliability for widespread

clinical use requires continuous optimization, addressing core aspects like data quality, interpretability, and ethical considerations (100). Nevertheless, the integration of these advanced technologies with healthcare may make diagnoses faster, more efficient, and ultimately improve overall health outcomes.

5.2 Role of AI in identifying disease associated biomarkers

Traditional biomarker identification methods encounter limitations in pinpointing genetic susceptibility loci for complex diseases, leaving some genetic risk factors unexplained (252). Although GWAS have been instrumental in identifying genetic variants, they can overlook variants that do not meet statistical thresholds and may fail to capture complex interactions among single nucleotide polymorphisms that influence complex traits. The SNP-set analysis method was developed to enhance GWAS results and address this "missing heritability". SNP-set analysis successfully identified significant and biologically relevant SNPs that GWAS had previously overlooked due to stringent statistical cutoffs. This approach, by grouping similar variants rather than analyzing SNPs individually, effectively reduced the risk of false negatives. This approach can also be applied to other complex conditions, including cancer, diabetes, or autoimmune disorders, as well as combinations of these diseases. However, it also highlighted the challenges regarding SNP selection, refining clustering techniques, and mitigating population structure biases, often requiring principal component analysis to minimize errors (253).

Metabolomics provides a comprehensive

understanding of an organism's metabolite profile, revealing variations influenced by environment, lifestyle, and genetics (254, 255). Metabolomics involves measuring metabolite levels in blood, urine, or tissues to identify alterations as biomarkers (256). However, traditional methods such as ELISA, Western blotting, immunohistochemistry, and microarrays are often time-consuming and labor-intensive (257). AI has shown potential in metabolomics by rapidly analyzing large datasets to identify disease-related biomarkers, thereby leading to more precise and personalized medical care. DL models further enhance accuracy in biomarker identification and disease detection through the integration of supervised and unsupervised learning approaches. Metabolomics plays an important role in the study of various diseases, including cancer, diabetes, cardiovascular diseases, neurodegenerative diseases, and infectious diseases such as tuberculosis and COVID-19 (69).

5.3 AI in pedigree analysis and genetic counseling

The integration of AI into pedigree analysis and genetic counseling addresses several challenges arising from intricate family structures, varying data availability, and the complexity of genetic information. Pedigree analysis is a longstanding approach fundamental to understanding inheritance patterns, identifying genetic disorders, and mapping disease loci (258). However, the accurate construction of these charts, particularly in large cohorts or with complex family structures, is challenging. Manually constructing family trees can be tedious and prone to errors, often resulting in valuable data

remaining confined to paper records (259–261). Traditional methods frequently prove insufficient for these complex relational structures, especially when dealing with uncertain relationships, limited familial data, and the exponential complexity involved in exploring numerous possible ancestries (234). Furthermore, low-coverage sequencing data, data sparsity, and various biases can hamper the accuracy of computational methods for pedigree inference, impacting their ability to leverage probabilistic genetic information effectively (262–264). The "curse of dimensionality," stemming from the vast number of features in genomic data, also poses a significant challenge for predictive modeling in this domain (265).

ML and graph-based DL algorithms have been adopted to address these complexities, enabling the modeling of intricate relationships, inference of genetic probabilities, and accurate reconstruction of pedigree structures, even in the presence of genetic recombination, mutations, and missing data (261, 266–268). These AI-driven approaches are being refined to improve parentage analysis and reconstruct pedigrees by evaluating genomic relationship coefficients and identifying inconsistencies, allowing for a more precise understanding of inheritance dynamics (269–271).

The integration of AI into genetic counseling is improving efficiency, scalability, and accessibility of services (272–274). Digital platforms, such as "Genome Access," are being developed to streamline various stages of the genetic counseling process, offering features like informed consent, educational modules, and explanations of DNA test results (274,

275). AI-powered chatbots and virtual assistants, including Clear Genetics' GIA, provide preliminary information, answer common patient queries, and offer initial education and risk assessments (119, 120, 276–278). These tools can also assist in scheduling appointments and collecting family history, which are often routine but time-consuming aspects of the pre-test counseling workflow (273). By automating these tasks, AI reduces the burden on human genetic counselors, allowing them to allocate more time to the psychosocial, ethical, and familial complexities inherent in genetic conditions, and to effectively triage high-risk cases (119–121, 276). The integration of these digital tools with electronic health records further enables automatic risk stratification and personalized genomic assessments. Additionally, telemedicine (often incorporating video-based consultations) has improved access to genetic counseling for patients in remote areas and offers greater scheduling flexibility (121).

A notable example of a digital tool is MyKinMatters, a web application designed to facilitate the communication of genetic test results to at-risk relatives (279, 280). This platform allows individuals who have undergone genetic testing (proband) to create a family tree. Clinicians can then utilize this family tree to identify and indicate which relatives should be contacted. The proband can upload the proband's test results to the platform, enabling the electronic sharing of these results with the identified relatives. MyKinMatters provides clear guidance on "which relatives to inform and what information to share," thereby ensuring effective communication and reducing the

potential for psychological burden on the proband (279). The software can present family relationships in both genealogical and pedigree formats, with clinicians typically using the pedigree format for risk assessment, aiming to empower relatives to make informed decisions about their own genetic testing, surveillance, or preventive options (280, 281).

AI also helps in family risk modeling and the clinical implementation of Polygenic Risk Scores in genetic counseling. ML models are being developed for the risk assessment and prediction of Mendelian genetic disorders in children, leveraging clinical history and pedigree analysis (266, 268). For instance, ML approaches can improve variant assessment and guide clinical decisions by providing refined, individualized disease risk estimates through penetrance modeling (282). PRSs are increasingly recognized for personalized risk assessments for common, complex conditions by combining genetic risk factors with environmental exposures and behaviors (267, 283–288). AI and ML algorithms can capture complex causal relationships between heterogeneous risk factors, move beyond simple additive models, and identify highly predictive features within high-dimensional genetic data (265, 287, 289–291). Such models are important for integrating genomic and non-genomic information, optimizing predictions for diverse genetic ancestries, and ensuring clear communication of complex risk information to both clinicians and patients (283, 284, 292–294). However, the clinical implementation of PRSs requires careful consideration and educational resources for clinicians, particularly regarding the interpretation and effective communication of

results and the establishment of robust regulatory frameworks for disease prevention strategies (149, 267, 283, 285).

Despite the potential benefits of AI in enhancing genetic services, its widespread integration in genetic counseling and pedigree analysis raises important ethical and practical considerations. These include concerns regarding data access, privacy and protection, algorithmic bias, responsibility and accountability, the impact on the patient-professional relationship, justice and equity, patient well-being, and transparency and explainability (295–297). Ensuring proper data governance, avoiding health inequities, and securing patient rights are paramount, especially given that generative AI models might operate by sharing data in non-restricted environments (272, 298, 299). Therefore, a human-centered approach remains indispensable, ensuring that AI functions as a decision-support tool that complements human expertise in navigating the sensitive and complex landscape of genetic health rather than replacing it (277, 297).

6. Challenges in AI use for genetic healthcare and proposed solutions

AI holds significant promise for advancing genetics, but its integration into healthcare faces substantial challenges in maintaining data quality, ensuring interpretability, safeguarding privacy, navigating regulatory landscapes, and streamlining clinical workflows. The current inflexibility and lack of standardized formats for integrating multi-omic data, imaging data, and electronic health records pose significant issues, potentially leading to false diagnoses and inaccurate risk predictions (98). While DL-

based genetic mutation detection may provide faster, less invasive, and more cost-effective alternatives to traditional NGS, further validation is required for broader accessibility and integration into clinical workflows.

DL has advanced many genomic tasks, but recent literature highlights two fundamental, recurring limitations: prediction errors (including poor out-of-sample generalization) and a disproportionate dependence on large, well-annotated datasets (128). AI models frequently report high accuracy on internal benchmarks but often show reduced performance when applied to external cohorts or different sequencing/assay conditions (24, 48). For example, polygenic and variant-effect predictors trained on one ancestry or biobank often show systematic declines in accuracy across diverse genetic ancestries, suggesting that model errors often stem from training-data mismatch rather than model expressivity (292, 294). Data heterogeneity (due to varying platforms, ancestries, and phenotype definitions) and limited sample sizes for many populations are primary causes of unreliable predictions in genomics (58). Methods aiming to reduce these errors, such as evolutionary-inspired augmentation to increase sequence diversity or federated learning to pool multi-center signals without centralizing raw genomes, show promise but are still in early validation stages (111). Moreover, parallels from medical imaging indicate that DL pipelines often lack rigorous uncertainty quantification and standardized external validation, increasing the risk of harmful mispredictions in clinical settings. In summary, despite the power of DL, high prediction error rates and substantial data-scale requirements

remain major obstacles to safe and equitable clinical deployment in genetic healthcare. To address the primary challenges that limit every AI model from being adopted in healthcare, valuable solutions have been proposed (Figure 3).

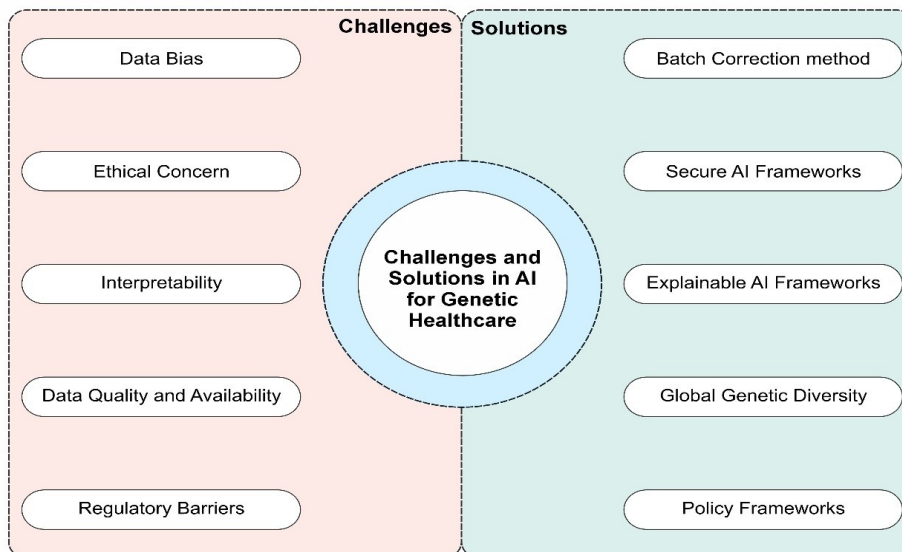


Figure 3. This figure highlights core challenges, including data heterogeneity and bias, limited model interpretability, ethical and regulatory constraints, and clinical integration barriers. Corresponding solutions illustrated include federated learning and standardized pipelines for data quality, XAI for transparency, privacy-preserving computation for governance, and clinician-in-the-loop validation for workflow adoption. Together, these elements highlight the multidimensional approach required for safe and effective clinical translation of AI systems.

6.1 Improving AI models with high-quality data

The effectiveness of AI models in genetic healthcare is largely dependent on the quality and representativeness of the training data. Addressing dataset limitations and biases is essential for developing robust and generalizable AI applications.

6.1.1 Strategies to overcome dataset limitations and biases

6.1.1.1 Data heterogeneity

Genetic datasets often suffer from variations due to different sequencing platforms, experimental protocols, and population

ancestries, leading to models that perform poorly when applied to data from different sources than those used for training, undermining generalizability and fairness (179, 292, 294). Standardizing data collection methods and formats is essential. Techniques like batch correction and transfer learning are employed to mitigate differences between datasets (127). Transfer learning allows knowledge from larger, well-studied populations to be applied to smaller, underrepresented groups, with the potential to improve prediction accuracy (128). While these strategies help, complete elimination of heterogeneity is difficult, and the performance of transfer learning can still be limited by the

degree of dissimilarity between source and target domains.

6.1.1.2 Interdependent data structure

Genomic data often exhibit complex interdependencies, such as linkage disequilibrium and gene networks, which traditional AI models might struggle to capture, leading to overlooked key biological relationships or incorrect assumptions about feature independence (145). GNNs and other network-based approaches are used to model these complex relationships, integrating genomic and proteomic interaction networks (95, 97, 145). Multi-omics integration also captures a more holistic view of biological systems by considering interactions across different molecular levels (66, 73). GNNs require specialized expertise and computational resources. Constructing accurate biological networks from noisy data is itself a challenge.

6.1.1.3 Confounding variables

Unaccounted confounding factors (e.g., environmental influences, lifestyle choices, unmeasured genetic variants) can lead to spurious associations and biased model predictions, obscuring true genetic signals (17). Advanced statistical methods, careful study design, and the integration of diverse data types (genomic, environmental, clinical, lifestyle) can help control for confounders. AI models designed to learn causal relationships rather than just correlations are an active area of research (78). Identifying and accurately measuring all potential confounders is often impossible. Causal inference with AI in complex biological systems remains a significant research challenge.

6.1.1.4 Data leakage

Data leakage occurs when information from the test set inadvertently leaks into the training set, leading to overly optimistic performance estimates and models that fail to generalize to new, unseen data, compromising clinical reliability (115). Data leakage can be difficult to detect in complex biological datasets due to their intricate dependencies (115). Strict adherence to data partitioning protocols (e.g., proper cross-validation, independent validation sets) and careful feature engineering are essential to prevent leakage. Best practice guidelines promote the avoidance and recognition of data leakage to prevent biased results (115). Addressing leaky pre-processing, where common transformations unintentionally introduce information from the test set into the training process, is also critical (115). Subtle forms of data leakage can be difficult to detect, especially in complex genomic datasets with shared biological contexts, and the exponentially large space of potential user prompts can make it nearly impossible to anticipate all forms of data leakage in generative AI models (115).

6.1.1.5 Class imbalance in disease genomics

Rare diseases or specific genetic variants are often underrepresented in datasets, leading to AI models that are biased towards common conditions or variants and perform poorly on minority classes (109, 115). This bias impacts the diagnosis and treatment of rare genetic disorders (115). Techniques such as oversampling (e.g., SMOTE), undersampling, and cost-sensitive learning are used to overcome class imbalance (115). Specialized algorithms for rare variant analysis and methods that sample positive and negative

examples to ensure the model is not biased toward any class are also being developed (115). Federated learning can also help by aggregating data from multiple institutions without centralizing rare disease data, increasing the effective sample size (110, 111). Oversampling techniques can lead to over fitting, and under sampling can discard potentially valuable information (115). Specialized algorithms still require some representation of the rare class to learn effectively.

6.1.1.6 DNA-based computing

While DNA computing offers potential for massive parallelism and solving complex optimization problems, current implementations face challenges with scalability, error rates, and the vast number of DNA molecules required for complex computations (115). Research continues on evolutionary DNA algorithms to overcome parallelism limitations and implement molecular model training in ML (115). DNA computing is being explored for logical and neural network computation, with complex designs demonstrating its potential for self-organization and self-learning capabilities (115). Practical applications in genetic healthcare are still in their early stages, with significant hurdles in feasible circuit size, function diversity, and computing efficiency (115).

6.1.2 Importance of global genetic diversity in AI training

The lack of global genetic diversity in AI training datasets is a critical challenge, as most genomic data originate from individuals of European descent, leading to models that

perform poorly when applied to diverse populations, exacerbating health disparities (111, 294). Initiatives to increase the representation of diverse ancestries in genomic databases are paramount. Federated learning offers a privacy-preserving approach to integrate smaller datasets from different ancestries, thereby reducing bias and improving model accuracy without centralizing sensitive patient data (111, 112). Building globally diverse datasets is a massive undertaking requiring international collaboration, significant funding, and careful consideration of ethical and social implications. Even with federated learning, biases can persist if the underlying data from contributing institutions are themselves biased.

6.2. Enhancing explainability and transparency

The "black-box" nature of many advanced AI models, particularly DL, limits their clinical adoption due to a lack of trust and understanding among healthcare professionals and regulators.

6.2.1 Development of interpretable AI models in genetics

Complex AI models often make decisions through intricate, non-linear computations that are difficult for humans to understand, making it hard to verify their reasoning or identify potential errors (81, 209, 300). Explainable AI techniques are being developed to provide insights into how models arrive at their predictions. This includes methods like SHAP and LIME, which help attribute model outputs to specific input features (61, 146). Rule-based models and attention mechanisms in neural networks also enhance interpretability (179, 187). In genetics, XAI can help identify key

genetic or clinical features driving disease risk predictions (146). XAI methods can sometimes offer simplified or *post-hoc* explanations that may not fully capture the model's internal workings. The trade-off between model performance and interpretability often remains a design challenge.

6.2.2 Strategies for better clinical validation and trust building

Without clear explanations, clinicians may be hesitant to trust AI-driven insights, and regulatory bodies struggle to approve models whose decision-making processes are opaque. Rigorous external validation of AI models on independent clinical cohorts is essential to demonstrate real-world performance (300). Collaborative validation frameworks involving clinicians, geneticists, and AI developers can build trust. Implementing AI systems that offer clear, human-understandable rationales for their recommendations, such as by highlighting relevant genetic variants or clinical data, can increase adoption in clinical practice (15, 209). Integrating AI into clinical decision support systems also allows for human oversight and intervention. Clinical validation is time-consuming and expensive. Achieving consensus on what constitutes "sufficient" explainability and trust for clinical deployment is ongoing.

6.3 Addressing ethical and privacy concerns

Genetic data are among the most sensitive types of personal information, raising significant ethical and privacy concerns when processed by AI.

6.3.1 Secure AI frameworks for handling genetic data

The risk of data breaches, re-identification, and misuse of genetic information is high, especially with cloud-based AI platforms (298, 299). Traditional data centralization for AI training can expose sensitive patient data (115). Secure AI frameworks employ privacy-enhancing technologies such as federated learning, homomorphic encryption, and differential privacy (109, 110). Federated learning allows AI models to be trained on distributed datasets without the raw data ever leaving the local institution, protecting patient privacy (111). Blockchain technology is also being explored for secure sharing and management of health data. Implementing these advanced security measures can add computational overhead and complexity. Homomorphic encryption is computationally intensive, and differential privacy can sometimes reduce model accuracy by adding noise.

6.3.2 Ethical guidelines for AI-driven genetic diagnostics

The rapid development of AI in genetics outpaces the establishment of comprehensive ethical guidelines, leading to potential challenges with informed consent, algorithmic bias, equity of access, and the impact on the patient-professional relationship (85, 277, 295–297). Developing clear, comprehensive ethical guidelines and regulatory frameworks will be important for responsible implementation. These guidelines should address data ownership, consent for data use in AI, responsible disclosure of AI-generated insights, and mechanisms for accountability when errors occur (200, 296). Collaboration between ethicists, clinicians, patients, and AI developers is vital to create human-centered AI that

prioritizes patient well-being and autonomy (15). Ethical considerations are complex and context-dependent. Translating broad ethical principles into specific, enforceable guidelines for diverse AI applications in genetics is challenging.

6.4 Regulatory improvements and clinical adoption

For AI tools to have a tangible impact in clinical genetics, robust regulatory frameworks and clear pathways for adoption are necessary.

6.4.1 Standardizing AI-based genetic tools for healthcare

A lack of standardized evaluation metrics, data formats, and reporting requirements for AI tools hinders their comparability, validation, and integration into clinical practice (115). The diverse range of AI algorithms and their specific requirements make standardization difficult. Regulatory agencies are expanding their mandates to include AI oversight and certification (115). International organizations like ISO/TC 215/SC 1 Genomics Informatics are actively developing standards for genomics informatics, covering data elements, metadata, quality control metrics for DNA sequencing, and structured clinical reports (115). These efforts aim to harmonize data formats and access protocols, ensuring consistency and reliability of clinical and omics data (115). The fast pace of AI innovation often outstrips the ability of regulatory bodies to create and update standards. Achieving international harmonization of standards is also a significant hurdle.

6.4.2 Policy frameworks for safe AI integration in genetics

Existing policy frameworks may not adequately address the unique challenges posed by AI in genetics, such as the potential for algorithmic bias, the need for continuous model monitoring, and the implications of AI-driven decisions for medical liability (115). New policy frameworks are being developed to ensure the safe, ethical, and effective integration of AI in genetic healthcare. These frameworks typically focus on areas such as data governance, transparency, accountability, and equity of access (98). Policies that promote public trust, ensure patient autonomy, and support ongoing research into AI safety are critical. A consortium model, similar to GA4GH, could be adapted for Genome Language Models to address the provision of AI-informed outputs for patient care (115). Policy development can be slow and may struggle to keep pace with technological advancements. Balancing innovation with safety and ethical concerns is an ongoing challenge, requiring adaptable and forward-looking regulatory approaches.

7. Adaptive and context-aware genomic intelligence in clinical medicine

The advent of NGS has substantially increased the scale of genomic data available for genetic healthcare and may contribute to more individualized approaches to genomic medicine. AI, particularly ML and DL, is changing how this data is analyzed and interpreted, influencing variant analysis, diagnostic interpretation, and predictive and prognostic medicine (301, 302). AI applications in genomics span across various domains, including genomic variant interpretation (9, 303), biomarker discovery, risk scoring (194), and multi-omics data

integration (5, 304). Advanced AI models are increasingly utilized for tasks such as pathogenic variant identification, offering faster and more cost-effective solutions compared to traditional methods (194).

Despite these advancements, significant challenges persist. Current genomic medicine largely relies on static frameworks for interpreting genetic variation. Variant classification systems and predictive models are often built from limited datasets and updated periodically, leading to inconsistencies and a high prevalence of VUS (9). This issue is particularly acute in rare diseases, where diagnostic yields, despite technological advances, remain between 35% and 55% (10). The complexity of genomic data, including the need to evaluate variants without fully incorporating biological contexts such as gene regulation, environmental exposure, cellular state, aging, or disease progression, further complicates interpretation (5).

A critical limitation of current AI models in genomics is bias arising from limited diversity in training datasets, which can reduce the generalizability of models across ancestral populations (11–14). Such limitations can contribute to disparities in genomic prediction and interpretation (305). While DL models offer strong predictive capabilities, their performance may also be affected by overfitting to characteristics of the training data, limiting generalization when applied to datasets that differ from those used for model development (11–13).

To address these limitations, emerging AI

approaches are gaining traction. Multimodal learning, which integrates diverse data types such as genomic information, electronic health records, medical imaging, and laboratory results, offers a promising solution to capture the complexity and high dimensionality of biomedical data (5, 304, 306). Such models have the potential to revolutionize personalized medicine by understanding interactions within biological systems and predicting complex diseases (306). Furthermore, federated learning (FL) is emerging as a critical privacy-preserving approach. FL allows AI models to be collaboratively trained on distributed datasets across multiple institutions without centralizing sensitive raw data, thereby mitigating privacy concerns, addressing data silos, and enabling larger, more diverse training cohorts (111, 112, 307–309). This decentralized training paradigm can achieve predictive accuracy comparable to models trained on centrally aggregated data (112, 309), making it invaluable for integrating multi-institutional genetic datasets (306, 307).

However, an important gap remains in how genomic and clinical information is integrated across time and changing biological states. While AI has improved individual-variant analysis, it remains challenging to integrate heterogeneous, longitudinal data sources to create continuously evolving, patient-specific genomic models. This limitation makes it difficult to account for the dynamic interplay of genetic predispositions with gene regulation, environmental exposures, cellular states, aging processes, and disease progression over time. Consequently, genomic interpretation may remain associated with persistent VUS, inconsistent predictive performance across

diverse populations due to training biases, and a fragmented understanding of genomic influences on health and disease progression, limiting the potential for truly personalized medicine. The field lacks a robust, adaptive AI framework capable of learning from a constantly expanding and diversifying data landscape to refine genomic interpretations in real-time. This highlights the need for context-aware genomic intelligence systems that can integrate evolving biological, clinical, and environmental factors to support continuously updated and personalized genomic insights (304).

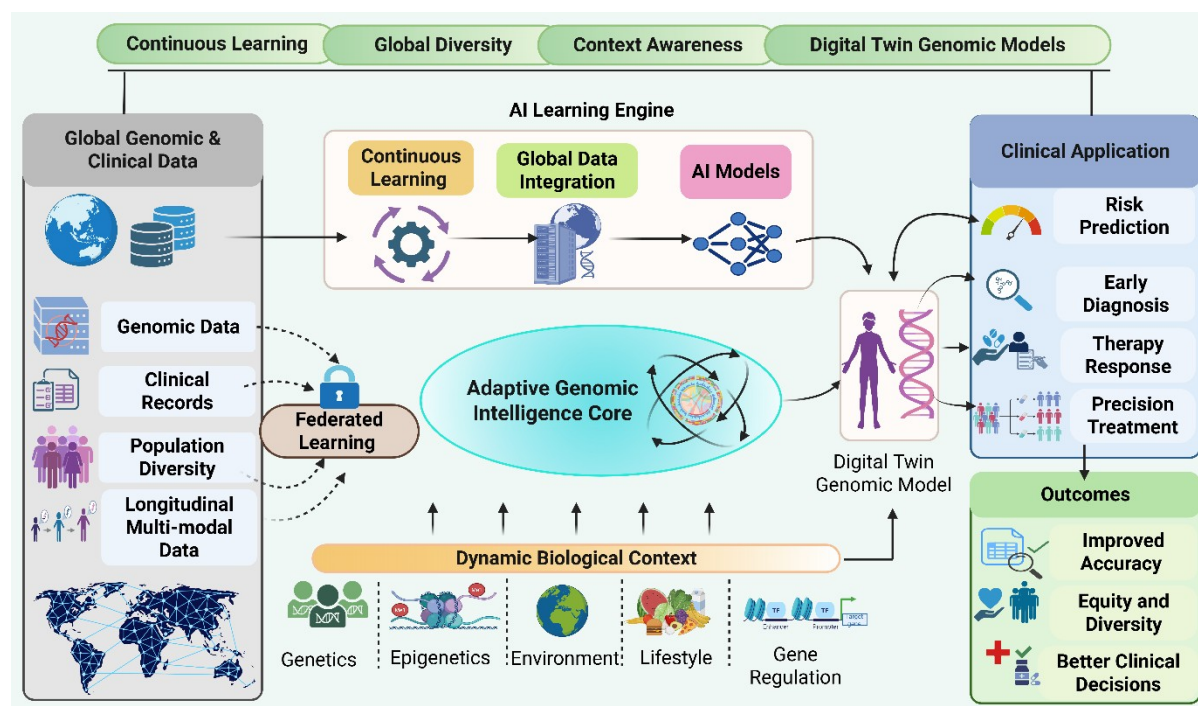


Figure 4. Integrated Adaptive Genomic Intelligence Framework. This illustration presents a framework describing how globally diverse genomic and clinical datasets are integrated with continuously learning AI (adaptive AI) to improve genomic medicine. Multimodal inputs, including genomic data, clinical records, population diversity, and longitudinal information, could be integrated through federated learning and AI-driven data integration. The adaptive genomic intelligence core would incorporate dynamic biological context, such as genetics, epigenetics, environment, lifestyle, and gene regulation, to generate patient-specific digital twin models. These models could support clinical applications including risk prediction, early diagnosis, therapy response assessment, and precision treatment. Clinical outcomes could provide a feedback loop that continuously refines the system, potentially improving diagnostic accuracy, equity across populations, and clinical decision-making.

We hypothesize that an AI framework designed to continuously integrate globally diverse genomic datasets with longitudinal, multi-modal clinical and environmental data within a federated learning framework could generate adaptive, patient-specific genomic models. We also hypothesize that such a model may evolve, potentially improving diagnostic accuracy and therapeutic guidance in clinical medicine. The conceptual model of the proposed adaptive genomic intelligence system is presented in Figure 4.

Based on this hypothesis, several specific and

testable predictions can be formulated:

[1] *Improved diagnostic yield*: The adaptive AI system is expected to improve diagnostic yield for rare genetic diseases and reduce the proportion of cases that remain undiagnosed compared with current static interpretation methods.

[2] *Reduced variants of uncertain significance*: The continuous learning and contextual integration capabilities of the AI system may substantially decrease the number of VUS across diverse patient populations by providing dynamic evidence for reclassification.

[3] *Improved cross-population prediction and fairness*: Models within the adaptive AI system, trained on globally diverse and continuously updated datasets, are expected to demonstrate more consistent predictive performance across major ancestral populations and may help reduce existing ancestry-related biases.

[4] *Enhanced clinical decision support*: The adaptive AI system may provide more precise, context-aware, and actionable clinical decision support for treatment selection and disease management, reflecting the patient's evolving genomic and phenotypic state over time.

7.1 Potential experimental approaches

Evaluating the proposed hypothesis would require a multi-phase, longitudinal study comparing an adaptive, context-aware AI system against current standard-of-care genomic interpretation methods.

7.1.1 System development and baseline establishment

Phase A: Develop an AI platform incorporating multimodal data integration capabilities (genomics, EHR, wearables, environmental

data), a Federated Learning architecture for multi-institutional data sharing, and continuous learning algorithms. This platform would prioritize interpretability through integrated Explainable AI techniques (310).

Phase B: Establish a large, prospectively collected cohort of patients undergoing genomic sequencing for diagnostic or risk prediction purposes across multiple, geographically diverse clinical sites. Train a "static" AI model (representing current best practice) and conduct Standard of Care (SOC) interpretation on baseline data.

7.1.2 Longitudinal comparative study

This comparative study would directly assess the efficacy and superiority of the adaptive, context-aware AI system against current SOC genomic interpretation methods over a defined period.

7.1.2.1 Study design and patient randomization

Newly recruited patients undergoing genomic sequencing for diagnostic clarification or disease risk assessment across multiple, geographically diverse clinical sites would be prospectively enrolled. Eligibility criteria would be stringent, focusing on conditions where genomic insights are clinically actionable or diagnostic challenges are prevalent (e.g., rare diseases, complex hereditary conditions). Following informed consent, eligible patients would be randomized (e.g., 1:1 ratio) into two distinct study arms.

7.1.2.2 Control Arm

Patients in this arm would receive genomic interpretation and clinical decision-making

based entirely on the current SOC methods. This typically involves:

[1] *Manual Interpretation*: Detailed analysis of genetic variants by board-certified geneticists and genetic counselors, utilizing established guidelines (e.g., ACMG/AMP criteria for variant classification) (36, 39), existing variant databases (e.g., ClinVar, HGMD) (56, 126), and literature review.

[2] *Static AI Tools*: Use of any currently approved or widely adopted static AI-based tools, e.g., specific variant callers like DeepVariant for initial variant detection (28, 48), or diagnostic prioritization tools like Fabric GEM for rare diseases (30) that are not dynamically updated with continuous, multi-modal data.

[3] *Data Silo Approach*: Clinical management and genetic counseling would not integrate continuous, longitudinal multi-modal data in an automated, adaptive manner.

7.1.2.3 Intervention Arm

Patients in this arm would have their genomic interpretation and clinical decision support augmented by the adaptive, context-aware AI system developed in Phase A. The AI system would include:

[1] *Augmented Interpretation*: Provide geneticists and clinicians with AI-generated insights, dynamic variant reclassifications (e.g., VUS re-evaluations (9)), and personalized therapeutic recommendations, directly informing clinical decision-making. These insights would be delivered *via* a dedicated clinical decision support interface or platform, integrated into the existing electronic health record system, designed for intuitive interpretation and interaction. The human clinician retains ultimate decision authority,

integrating AI outputs as advanced decision support (209).

[2] *Continuous Data Integration and Learning*: The AI system would continuously and securely integrate new patient data generated during the study period. This includes:

[2a] *Longitudinal Clinical Data*: Updates from electronic health records, including new diagnoses, treatment responses, laboratory results, and imaging studies.

[2b] *Updated Multi-omics Data*: Any subsequent omics data (e.g., repeat transcriptomics, metabolomics profiles) (73, 79).

[2c] *Environmental Exposures/Lifestyle*: Data from patient questionnaires, wearables, or other sources capturing environmental factors and lifestyle changes (17, 199).

[2d] *Federated Learning*: The AI model's underlying algorithms would be continuously refined across the federated network of participating institutions, drawing insights from the aggregated (but never centrally stored raw) data of all intervention arm patients, enhancing model generalizability and predictive power without compromising individual patient privacy (111, 112, 307).

7.1.2.4 Study duration

The longitudinal nature of the study would require a defined observation period (e.g., 3-5 years) to adequately capture the evolving dynamics of patient health, the impact of continuous learning, and the long-term clinical utility of adaptive genomic models.

7.1.3 Outcome measurement

[1] *Diagnostic Yield*: Compare the rate of definitive genetic diagnoses in the intervention arm versus the control arm.

[2] *VUS Reclassification*: Track the number and impact of VUS reclassifications over time in both arms.

[3] *Predictive Accuracy*: For relevant conditions, compare the accuracy of risk prediction and treatment response prediction across diverse populations.

[4] *Clinical Decision Impact*: Assess the influence of AI-generated insights on treatment modifications, patient outcomes, and physician satisfaction through surveys and clinical record review.

7.1.4 *Ethical and privacy considerations*

Implement robust ethical oversight, informed consent protocols, and privacy-preserving technologies (e.g., homomorphic encryption alongside federated learning) to ensure responsible data handling and patient protection throughout the study (110).

7.2 Implications for genetic healthcare

If validated, the proposed framework could represent an important advance in genetic healthcare. Adaptive and context-aware genomic intelligence would move beyond static interpretations to provide dynamic, evolving insights into an individual's health trajectory.

7.2.1 *Precision diagnostics and prophylaxis*

Earlier and more accurate diagnosis of genetic conditions, particularly rare diseases, could support proactive management of health risks based on evolving genomic and environmental factors.

7.2.2 *Equitable healthcare delivery*

By mitigating ancestral biases through diverse, continuously updated datasets, this framework could contribute to more equitable access to the

benefits of genomic medicine across global populations and potentially help reduce existing health disparities.

7.2.3 *Enhanced clinical workflow*

Geneticists and clinicians could be provided with intelligent tools that reduce the burden of VUS, streamline variant interpretation, and provide continuously refined decision support, potentially contributing to more precise and personalized patient management.

7.2.4 *Accelerated research and discovery*

The continuous aggregation and analysis of diverse, longitudinal data could generate new insights into gene-environment interactions and disease mechanisms, fostering a learning healthcare system.

7.2.5 *Ethical and trustworthy AI*

By emphasizing transparency, explainability, and privacy-preserving federated approaches, this paradigm could promote greater trust in AI systems within the medical community and among patients, paving the way for broader clinical adoption. Overall, this adaptive framework could represent a critical step towards realizing the full promise of personalized genomic medicine, creating a dynamic ecosystem where AI models can be updated as new data and evidence become available, supporting patient-specific care.

8. AI-enabled discovery of system-level biomarkers: cross-pathway redundancy

Conventional Differential Expression (DE) analysis effectively identifies genes with large mean expression differences between conditions but fails to capture key biological phenomena, including (i) functional

redundancy among pathways, (ii) rewiring of interaction networks without significant expression changes, and (iii) the role of non-differentially expressed (non-DE) genes that remain essential for maintaining information flow and phenotype (311–315). Increasing evidence suggests that disease progression may involve a loss of functional specialization across biological modules, whereby pathways become more correlated or exhibit altered dependency structures (316). Biological redundancy, in which multiple genes or pathways perform overlapping functions, is a fundamental property of living systems that promotes robustness through duplicated genes, parallel pathways, and overlapping regulatory mechanisms (317–319). In pathway enrichment analysis, redundancy manifests as overlapping gene sets, often producing correlated statistical signals and inflated lists of apparently significant pathways that do not represent independent biological processes (312, 313, 320, 321). Standard DE methods, such as DESeq2 and edgeR, evaluate genes independently using generalized linear models, rendering them insensitive to (a) regulatory genes with stable expression but altered connectivity, (b) compensatory effects among partially redundant pathways, and (c) nonlinear gene interactions that drive phenotypes, particularly in machine learning-defined subgroups. Consequently, pathway-level redundancy and non-DE functional drivers are frequently overlooked in conventional DE-based pipelines (314, 315). Evidence from human genetic disorders highlights the biological and clinical relevance of redundancy. In Spinal Muscular Atrophy (SMA), caused by loss of SMN1, the paralog SMN2 partially compensates by producing functional protein, with SMN2 copy number strongly correlating with disease severity and serving as a clinically established biomarker (322). Similarly, in Darier's disease, caused by ATP2A2 (SERCA2) deficiency, compensatory upregulation of the homologous Ca^{2+} pump ATP2C1 (SPCA1) helps buffer calcium homeostasis, demonstrating functional redundancy, although its diagnostic utility remains unexplored (323). Redundant gene clusters can also influence disease risk. For example, increased copy number of the trypsinogen genes PRSS1 and PRSS2 is associated with higher pancreatitis risk, supported by both human genetic studies and transgenic murine models, linking gene dosage within redundant systems to disease phenotype (324). In cancer, redundancy has been extensively investigated through CRISPR-based screens. Tumor cells often tolerate loss of a single paralog due to compensatory mechanisms, while combined loss can result in synthetic lethality (325, 326). Large-scale studies have revealed that such interactions are relatively rare and highly context-dependent. Proteomic analyses further show that loss of one gene can induce compensatory upregulation of its paralog, which may stabilize cellular networks while also creating new dependencies that could be therapeutically exploited (325). Collectively, these findings demonstrate that redundant gene systems play a critical role in disease modulation, although their application as clinical biomarkers remains underexplored.

Conventional methodologies can detect redundancy in gene systems using several approaches. Single and double knockout experiments reveal backup mechanisms when

only combined gene loss causes a phenotype, though they are time-consuming and limited in humans. CRISPR combinatorial screens enable high-throughput identification of redundant or synthetic lethal gene pairs but are mostly restricted to *in vitro* contexts (326). Transcriptomic and proteomic analyses identify compensatory changes (such as upregulation of related genes), though these findings are often correlative rather than definitive proof of functional redundancy (327). Some investigators have proposed that system-level redundancy could serve as an indicator of disease. In practice, this means determining how different gene modules co-vary and suggesting that redundancy shapes disease, but the approach is not yet standardized. Careful controls are needed to distinguish true redundancy from mere correlation, and to assess predictive value in cohorts. Despite conceptual interest, the gaps are clear: no standard metric of redundancy has been validated as a biomarker in patient studies. Most evidence is limited to model systems or retrospective genetic analysis. It remains unknown whether redundancy increases uniformly with disease or is specific to certain conditions.

Recent primary research has addressed these gaps using network-based, pathway-centric, and redundancy-aware methods, some explicitly quantifying redundancy across pathways, others inferring hidden regulators and rewired subnetworks that conventional DE overlooks. Collectively, these studies show that much of the functionally relevant variation occurs at the level of interaction structure and pathway usage rather than per-gene expression levels alone, but they also highlight limitations

such as dependence on incomplete interaction databases, sensitivity to parameter choices, and relatively sparse experimental validation of predicted non-DE drivers. The evidence supports using redundancy-aware and differential-network approaches as complements, not replacements, for standard DE, especially in complex traits such as cancer and host–pathogen interactions (314, 315, 328). This section establishes the biological basis for redundancy. Section 10 extends this conceptualization toward a broader mathematical formulation.

9. Redundancy control within pathway databases

9.1 ReCiPa: overlap-controlled pathway sets

ReCiPa was introduced as an application for quantifying and controlling pathway overlap in databases such as KEGG and Reactome prior to enrichment analysis. It demonstrated that naively using raw pathway collections yields large clusters of highly overlapping pathways whose significance is driven by shared genes rather than distinct biology, statistically manifesting as correlated p-values and overdispersion. ReCiPa removes overlapping genes to create cleaner, non-redundant pathway sets, improving interpretability and highlighting disease-relevant pathways (312). However, it only addresses database-level redundancy and relies on standard DE inputs without capturing condition-specific redundancy patterns (312).

Another approach to control was Set-theoretic minimization of pathway redundancy, where Stoney et al. (2018) used set theory to derive minimal, non-redundant pathway subsets that

preserve gene coverage while reducing overlap, improving interpretability of enrichment analysis (313). Applied across major databases, this approach minimizes redundant pathway clusters without losing biological information. However, it addresses structural redundancy in pathway databases rather than condition-specific functional redundancy or network dynamics (313). Applied to KEGG, Reactome, and Consensus PathDB, this approach achieved large reductions in pathway redundancy without discarding genes, and enabled more interpretable downstream visualization and enrichment results (313).

9.2 DINA and related differential network analysis

Gambardella et al. (2013) developed DINA (Differential Network Analysis) to identify condition-specific gene co-regulation by comparing gene–gene correlations rather than marginal expression, enabling detection of pathways with altered connectivity and their regulatory transcription factors (315). Applied to tissue-specific networks, DINA revealed metabolic pathway rewiring and identified regulators such as nuclear receptors and YEATS2, supported by functional validation. However, it remains sensitive to correlation estimation and data quality (315). Grimes et al. further showed that comparing regulatory networks between conditions uncovers group-specific interactions and pathway activity missed by DE, highlighting shifts in functional connectivity and compensation across pathways (329).

Similarly, Guzzi et al. (2024) introduced a non-parametric framework capturing non-Gaussian correlation changes and sex-specific

subnetworks in diseases like diabetes and atherosclerosis, demonstrating that differential network approaches detect subtle, context-dependent molecular signatures beyond traditional DE analysis (330).

9.3 Network-based identification of condition-specific pathway activity

Differential network analysis can identify disease-associated pathway changes that are not apparent from marginal gene-expression differences alone. Gambardella et al. developed DINA to identify condition-specific pathway activity and regulation by examining changes in gene–gene relationships across conditions (315). Such approaches complement conventional DE by shifting attention from individual expression changes toward alterations in network organization and pathway connectivity.

9.4 Explicit identification of non-DE but functionally essential genes

The DENetwork identifies functionally important non-DE genes using a network-based information-flow framework that integrates DE data with receptor–ligand interactions, signaling pathways, TF–target links, and PPIs (314). It scores signaling paths and applies *in silico* gene knockouts to quantify each gene's contribution to maintaining network flow, revealing key non-DE regulators such as receptors and TFs. Across disease datasets, these genes showed strong enrichment for known disease mechanisms and improved pathway coherence over standard methods. While robust to noise, the approach depends on prior interaction data and limited experimental validation, leaving many predictions as hypotheses (314).

9.5 Network-based inclusion of non-differentially active nodes

KeyPathwayMiner identifies disease- or condition-associated connected subnetworks enriched for differentially active genes while permitting a user-defined number of exception nodes that need not themselves be differentially active (331, 332). Such nodes can preserve network connectivity and may therefore identify potentially important components that would be excluded by strict DE filtering.

9.6 Evidence that DE overlooks nonlinear and redundant patterns

Studies have shown that conventional DE analysis may fail to identify features that drive nonlinear machine-learning predictions, whereas explainable-AI approaches such as LIME and SHAP can recover model-relevant features missed by DE analysis (333). Such discrepancies could affect subsequent pathway analysis and suggest that conventional DE may overlook combinatorial biological signals; however, pathway-level redundancy was not directly evaluated in this study.

9.7 Biological case studies of functional redundancy invisible to bulk DE

Developmental genetics studies show that “differential redundancy” can arise from subtle regulatory changes not detectable by bulk DE analysis. For example, paralog-specific enhancer rewiring in BOP1/BOP2 genes demonstrates spatially controlled redundancy through altered cis-regulation rather than expression magnitude changes (321). Similarly, overlapping enhancers at the *En1* locus exhibit lineage-specific redundancy, buffering phenotypic effects and highlighting that compensatory regulatory architectures are often

invisible to conventional DE approaches (334). Major gaps include the lack of standard measures for “system redundancy” in patient data, and the need to test whether redundancy metrics predict disease state or outcome. Future experiments could use targeted double-knockout or knockdown in relevant human cell models (with patient mutations) and measure phenotypes, or apply multi-omics time-series to see if coherence among pathways increases before clinical decline. CRISPR-based screens in patient-derived cells could map paralog dependencies. In sum, redundancy is an intriguing idea but demands careful validation.

Given these considerations, redundancy-based biomarkers can be defined at the cluster level. Measurable biomarkers could include: (a) a composite genetic risk score for a gene cluster (e.g., total deleterious mutation count in any gene of the cluster); (b) a pathway activity score reflecting co-expression or combined dysregulation of the cluster; or (c) a network-based metric (e.g., graph centrality or “crosstalk” score across redundant pathways). For example, one might compute the Pearson correlation of expression among genes in two normally independent pathways; a rising correlation in disease could signal pathological coupling. Importantly, this is not mere co-expression analysis, but specifically focuses on functionally linked but distinct modules becoming simultaneously perturbed. To avoid conflating simple co-expression with functional redundancy, cluster definitions could be anchored using curated databases (KEGG, Reactome, GO) and empirical paralog/gene family relationships (335, 336). By focusing on clusters, the biomarker is the aggregate state of the system, not any single gene. This may

increase robustness; for example, if one gene harbors a VUS, its clinical impact may be interpreted by observing the rest of the genes within the cluster. In practice, clusters could be designed *via* a combination of: [1] biological databases pathway and network modules and [2] data-driven modules (co-function networks, graph clustering). Each cluster could yield a numerical feature (e.g., sum of variant impact scores, cluster-level expression change) that feeds into AI models.

Collectively, these studies demonstrate that biologically meaningful signals often emerge from network structure, pathway interactions, and redundancy patterns rather than individual gene expression changes alone. However, current approaches remain limited by static analysis frameworks, lack of longitudinal integration, and inability to model redundancy dynamically at the patient level. Building on these observations, the proposed adaptive and context-aware genomic AI framework could be extended by incorporating redundancy-informed biological modeling, where gene clusters and pathways with overlapping functions are treated as dynamic biomarkers. While the proposed hypothesis describes a framework that enables continuous learning from multimodal and longitudinal data, incorporating redundancy would allow the system to capture functional compensation and systems-level robustness that are not evident in single-gene analysis. Thus, AI models may not merely predict outcomes, but also identify patterns associated with biological reorganization and loss of buffering capacity during disease progression. This could improve diagnostic interpretation, reduce VUS, and

enable more mechanistically informed clinical decision-making.

Machine learning models, including pathway-guided deep neural networks, could be used to capture nonlinear dependencies and latent cluster-level features across genomic, transcriptomic, proteomic, and clinical datasets, improving over traditional linear approaches (336, 337). GNNs could incorporate biological interaction networks (e.g., STRING, BioGRID) to detect functionally redundant subnetworks, supported by explainability tools such as GNNExplainer (336). To ensure scalability and privacy, federated learning could enable decentralized model training across institutions, maintaining performance comparable to centralized models while preserving data security (112). Multimodal learning architectures would integrate diverse data types, including longitudinal EHRs, imaging, and environmental data, to capture context-specific gene–environment interactions (336). Explainable AI methods (e.g., SHAP, LRP) would provide interpretability by linking predictions to specific genes and pathways, enhancing clinical trust (336). Additionally, bias-mitigation strategies could be incorporated to address ancestral imbalance and ensure equitable model performance across populations (338). Collectively, Sections 8 and 9 establish the biological foundations of redundancy through pathway overlap, compensatory mechanisms, and network-level interactions. Building on these observations, the following section explores how redundancy may be conceptualized as a dynamic and systems-level property of disease biology, extending beyond individual pathways to

interactions among tissues, organs, and physiological systems.

10. Future perspectives: dynamic and systems-level redundancy in disease biology

The preceding sections reviewed current applications of AI in genetic healthcare, including variant interpretation, biomarker discovery, precision medicine, explainable AI, federated learning, and clinical translation. The following section presents a forward-looking conceptual perspective on how future AI systems may leverage dynamic redundancy, longitudinal network organization, and systems-level biological interactions to improve disease prediction and monitoring. Although these concepts are informed by emerging evidence from systems biology and network science, they remain largely hypothetical and require further empirical validation.

10.1 Dynamic redundancy as a systems-level biomarker of disease progression

Recent studies suggest that many diseases arise not from isolated gene-level alterations, but from the gradual reorganization of biological networks over time. Rather than viewing diseases as static events, current research recognizes them as dynamic processes in which interactions among genes, pathways, and cellular systems continuously evolve. For example, the Dynamical Network Biomarker (DNB) framework identifies early warning signs of disease by detecting groups of genes that begin to fluctuate more and become more coordinated just before a critical transition occurs (339-341). Similarly, network medicine proposes that disease progression involves the gradual rewiring of molecular interaction

networks as biological systems move from healthy to diseased states (342). Marku et al. (2026) empirically observed module shifts (cytokine signaling, metabolic reprogramming) in leukemia cells under immune stimulation, illustrating system reconfiguration (341). In both acute lung injury and chronic liver disease studies, researchers observed that groups of genes started behaving differently before any obvious signs of disease appeared. The genes became more tightly coordinated with each other, suggesting that the biological system was losing its normal stability. These changes were associated with early warning signals of an impending transition to disease. Once the disease was fully established, the signal diminished, indicating that network instability is a temporary but potentially informative marker of early disease development (339). Furthermore, Keikha et al. (2025) developed multilayer network models from single-cell data to identify genes whose regulatory roles change across different stages of disease. Other studies have used temporal gene-expression networks to track shifts in pathway activity and regulatory programs as disease develops (343-345). Importantly, time-resolved network analysis can identify temporal changes in network organization that may not be apparent in static analyses. Interactions between genes may appear or disappear, and groups of functionally related genes may merge, separate, or reorganize as disease progresses (346, 347). Accordingly, approaches such as sliding-window correlation, change-point detection, multilayer modeling, GNNs, and dynamical systems-based methods are used to characterize disease trajectories (343). However, most current approaches focus on local network perturbations and rely on limited longitudinal

multi-omics data. Consequently, no unified framework exists to track how redundancy evolves across biological systems over time. Longitudinal redundancy metrics could therefore provide early indicators of declining resilience and impending disease transitions.

10.2 Cross-system redundancy and organism-level disease dynamics

Traditional disease models have focused on abnormalities within individual genes, pathways, tissues, or organs. However, increasing evidence suggests that many diseases arise from disrupted coordination across interconnected biological systems. Genes, cells, tissues, organs, and physiological networks operate as hierarchical systems that exchange information across multiple spatial and temporal scales, making disease progression a consequence of altered relationships among biological components rather than isolated defects. This systems-level view is supported by network biology, complex systems theory, and network physiology. Montgomery (2025) describes biological organization as hierarchical and modular, with bidirectional information flow between molecular, cellular, tissue, and organismal levels. Because functional units remain interconnected, perturbations can propagate across tissues and organs, generating coordinated system-wide responses (348). Such cross-scale coupling provides a framework for understanding disease as a progressive loss of modular independence and increasing cross-system coordination (348). Network physiology further demonstrates that physiological states such as sleep, stress, and recovery emerge from interactions among multiple organs rather than single systems

(349, 350). Similarly, diseases including diabetes, cardiovascular disorders, and neurodegeneration involve extensive crosstalk among metabolic, neural, immune, and vascular systems (342). Communication through neural, endocrine, immune, and circulatory pathways links organs such as the heart, brain, and gut into integrated physiological networks (351, 352). At the molecular level, disease progression may also involve increasing similarity between previously distinct biological systems. Izgi et al. (2022) identified a “Divergence-then-Convergence” (DiCo) pattern in aging, where tissue-specific gene-expression programs progressively converge, reflecting loss of tissue identity and specialization (353). Likewise, Deng et al. (2025) developed the LEGEND framework, integrating single-cell and spatial transcriptomics to identify gene modules shared across tissues and cell types while revealing condition-specific rewiring of gene–gene interactions. Taken together, these findings suggest that disease progression may be driven by increasing coordination and redundancy across biological systems in addition to abnormalities confined to individual pathways (354). A related perspective is provided by Thomas et al. (2026), who challenge the traditional view that pathology is anchored to a single lesion or organ. Instead, they propose that many diseases persist as distributed pathological systems maintained through interactions among multiple tissues, pathological sites, and feedback loops. Conditions such as metastatic cancer, amyloidosis, fibrosis, autoimmune disorders, and endometriosis are described as networks of semi-autonomous pathological units that collectively generate organism-level effects. In

this framework, local pathology can alter systemic physiology, which in turn promotes the emergence and persistence of additional pathological sites, creating self-reinforcing disease networks (355). Consequently, disease progression becomes a property of the entire system rather than any single pathological focus. This distributed view of pathology suggests that disease persistence may rely on redundant interactions among tissues, pathways, and pathological sites. Therefore, targeting a single lesion may be insufficient, as compensatory mechanisms maintain network function. In this framework, redundancy is proposed as a property of functional coupling across biological systems rather than simple pathway overlap.

Evidence for such cross-system dependence is also observed in mitochondrial biology. Studies of mitochondrial membrane organization demonstrate that mitochondrial function depends on coordinated interactions among lipid composition, membrane architecture, oxidative phosphorylation, energy metabolism, and organelle dynamics (356). Alterations in any one component can propagate throughout the broader cellular network, contributing to disorders ranging from neurodegeneration and diabetes to cardiovascular disease and cancer. Similarly, systems immunology and metabolomics reveal extensive integration between immune and metabolic processes through interconnected immunometabolic networks. However, most existing studies remain conceptual or rely on static measurements, and few explicitly model how relationships between biological systems evolve. Collectively, these findings suggest that organism-level disease progression may be

characterized by increasing coupling among metabolic, immune, neural, vascular, and structural systems. Such convergence may manifest as shared transcriptional programs, synchronized physiological responses, coordinated pathological feedback loops, or increasing correlation between molecular networks across tissues.

10.3 Longitudinal redundancy matrices for early disease detection

Longitudinal data analysis seeks to identify biological signals that change before clinically detectable symptoms emerge. One common approach is to build network matrices at different time points and track how they evolve over time. For example, in neuroscience, researchers use dynamic functional connectivity, which represents time-varying correlation networks between brain regions calculated over sliding time windows to study cognitive decline in diseases like Alzheimer's and cognitive impairment (357-359). In genomics, similar ideas have been used to build time-based gene co-expression networks during development or in response to treatment, although they are less frequently applied to disease onset. A related idea is the redundancy matrix, in which each entry represents the amount of information shared between two biological pathways or modules. Similar concepts already exist in mutual-information methods (354) and Gaussian graphical models (360, 361). The proposed extension is to calculate these redundancy matrices repeatedly over time (or across clinical visits) and then study how they change, for example, whether off-diagonal values gradually increase, suggesting growing system-wide coupling or loss of independence between pathways.

Related approaches, like Dynamical Network Biomarker methods, already track simpler summaries such as matrix norms or the largest eigenvalues of correlation networks as early warning signs of disease (339, 344). Therefore, extending these approaches to full redundancy matrices over time is a logical next step in the same direction. Several important challenges remain. First, there is no clear standard for defining or reliably estimating redundancy matrices from noisy, high-dimensional longitudinal multi-omics data. Second, we still lack robust statistical tools to detect meaningful temporal changes in these matrices, such as shifts in eigenvalues, increasing network density, or loss of modular structure. Third, it is unclear how many time points are needed for reliable inference, especially when clinical sampling is irregular. Another limitation is that most current longitudinal biomarkers (like proteins or imaging features) are analyzed individually, while redundancy matrices are inherently high-dimensional. A key challenge is linking network-level redundancy changes to biological and clinical outcomes and distinguishing true loss of resilience from normal variability. Progress would require synchronized longitudinal multi-omics and health-monitoring data, combined with systems biology, statistics, and dynamical systems approaches.

10.4 Beyond functional redundancy: emergent mathematical redundancy across biological systems

An important consideration is that redundancy does not always arise from known biological pathways but can also emerge statistically. Two biologically unrelated pathways may become statistically or informationally redundant if

their activity patterns increasingly co-vary over time, even without a known direct mechanistic connection. Information theory supports this concept, showing that mutual information between otherwise unrelated groups can increase through network-wide effects, generating emergent redundancy. Dynamic network studies further demonstrate that large-scale biological reorganization can create new dependencies among system components. For example, Deng et al. (2025) showed that integrating multiple data modalities can uncover hidden co-expression clusters in healthy tissues that are not explained by known biological functions, suggesting that coordinated behavior can emerge beyond established pathway maps (354). However, most biological studies still view redundancy as a structural feature, such as gene similarity families or overlapping pathways that provide functional backup. Although Tononi et al. (1999) distinguished redundancy from degeneracy conceptually; practical applications in molecular biology largely focus on established biological relationships (362). Consequently, the broader concept of statistical or emergent redundancy remains underexplored despite its potential to capture system-level behavior. Existing approaches measure redundancy between gene sets using similarity metrics such as Jaccard and Ochiai indices (363, 364). Deng et al. (2025) further introduced a gene–gene redundancy matrix within the LEGEND framework, using shared expression patterns as a covariance surrogate for network modeling (354). Similarly, neuroscience studies of redundancy–synergy gradients and network-science measures of edge covariance and network similarity capture related manifestations of emergent functional

redundancy. A major challenge is the lack of standardized methods to detect and interpret emergent redundancy. It remains unclear how to define statistical thresholds for redundancy or distinguish meaningful biological convergence from random correlations. Current predictive models rarely incorporate redundancy directly and typically rely on predefined network edges rather than shared information patterns. Multi-scale redundancy, involving groups of pathways rather than pairwise relationships, is also largely unexplored. Furthermore, system-level metrics linking redundancy to disease states are absent, and computational methods for quantifying complex redundancy–synergy relationships remain underdeveloped in biology.

10.5 AI-Based quantification of emergent mathematical redundancy (ERI/DRI)

Building on the concept of emergent mathematical redundancy across biological systems, a promising future direction is the incorporation of an Emergent Redundancy Intelligence (ERI) module within the adaptive and context-aware genomic AI framework proposed in Section 7. Rather than identifying redundancy from predefined pathway annotations or overlapping biological functions, the proposed ERI module would estimate redundancy as an emergent, dynamic property of biological organization by learning evolving statistical dependencies among genes, pathways, tissues, organs, and physiological systems. Using longitudinal multi-omics, physiological, imaging, wearable, microbiome, lifestyle, and clinical data, the framework would construct temporal multilayer biological networks at successive clinical visits, with nodes representing biological entities and edges

representing learned information-sharing relationships rather than predefined biological interactions. During an initial healthy period, the ERI module would estimate an individual's baseline redundancy landscape, representing normal coordination among biological systems. At subsequent time points, temporal GNN, transformer-based models, self-supervised learning, information-theoretic measures, and variational autoencoders would continuously reconstruct updated redundancy networks and compare them with this personalized baseline. Instead of detecting disease through abnormal expression of individual genes or pathways, the framework would identify progressive increases/decreases in cross-system coordination, loss of modular independence, declining buffering capacity, and network convergence, even when individual molecular biomarkers remain within normal ranges. These deviations would be summarized as a Dynamic Redundancy Index (DRI), defined as the distance between an individual's current redundancy landscape and their healthy baseline. If changes in redundancy relationships precede overt molecular dysregulation and clinical abnormalities, changing DRI values could potentially serve as early indicators of declining biological resilience and impending disease-state transitions. By transforming emergent mathematical redundancy into a measurable longitudinal biomarker, this framework would provide an actionable strategy for preclinical disease prediction, continuous health monitoring, and next-generation predictive, preventive, personalized, and participatory (P4) medicine.

11. Conclusion

Artificial intelligence in genetic healthcare is moving beyond isolated tasks such as variant calling, risk prediction, and biomarker discovery toward a more integrated model of precision medicine. However, greater predictive accuracy alone will not deliver this transition. Genomic information is embedded within biological systems that change over time, compensate for disruption, reorganize during disease, and interact with clinical and environmental influences. AI models that treat these processes as static or independent risk missing precisely the changes that precede disease progression.

The next advance should therefore be from static genomic prediction to adaptive systems-level intelligence. Longitudinal multi-omics, explainable and federated AI, network-level biomarkers, and continuous clinical feedback provide the foundations for models that can follow disease rather than merely classify it. Incorporating pathway interactions, compensatory redundancy, network rewiring, and cross-system coupling may further reveal loss of biological resilience and emerging disease-state transitions before conventional biomarkers become abnormal. These concepts remain to be rigorously validated, but they

define testable directions for the next generation of genomic AI.

The ultimate goal is not simply an AI system that predicts or classifies disease more accurately. It is one that can determine how an individual's biological state is changing, why it is changing, and when that trajectory becomes clinically consequential. Achieving this would transform genomic AI from a collection of powerful analytical tools into a dynamic framework for earlier detection, mechanistically informed risk assessment, and increasingly individualized intervention.

Acknowledgements

We gratefully acknowledge Avishkaran (NIPER), Hyderabad, Telangana, India, for providing resources. Figures 1, 2 & 3 were created with CorelDRAW.com (<https://www.coreldraw.com/>). The intellectual content and conceptual interpretations were entirely developed by the authors. Figure 4 was created with BioRender (<https://www.biorender.com/>). The authors acknowledge the assistance of the AI tool ChatGPT for language refinement and structural editing.

Author contributions

Ishanvi Tupili: Writing Original Draft, Visualization. Deekshitha Ravipati: Writing Original Draft. Karthik Mangu: Review, Editing, and Investigation. Bhumandeep Kour: Conceptualization, Review, Editing, Supervision, Formal analysis.

References

1. Narita, A., Ueki, M., & Tamiya, G. (2020). Artificial intelligence powered statistical genetics in biobanks. *Journal of Human Genetics*, 66, 61–65. <https://doi.org/10.1038/s10038-020-0822-y>

2. Singh, R. S. (2021). Decoding ‘Unnecessary Complexity’: A Law of Complexity and a Concept of Hidden Variation Behind “Missing Heritability” in Precision Medicine. *Journal of Molecular Evolution*, 89(8), 513. <https://doi.org/10.1007/s00239-021-10023-3>
3. Herrera-Luis, E., Benke, K. S., Volk, H. E., Ladd-Acosta, C., & Wojcik, G. L. (2024). Gene–environment interactions in human health. *Nature Reviews Genetics*, 25(11), 768. <https://doi.org/10.1038/s41576-024-00731-z>
4. Sarumi, O. A., & Heider, D. (2024). Large language models and their applications in bioinformatics. *Computational and structural biotechnology journal*, 23, 3498-3505. <https://doi.org/10.1016/j.csbj.2024.09.031>
5. Li, T., Shi, W., Isgut, M., Zhong, Y., Lais, P., Gloster, L., et al. (2023). Integrating Multi-Omics Data With EHR for Precision Medicine Using Advanced Artificial Intelligence. *IEEE Reviews in Biomedical Engineering*, 17, 80. <https://doi.org/10.1109/rbme.2023.3324264>
6. Junjun, R., Zhengqian, Z., Ying, W., Wang, J., & Liu, Y. (2024). A comprehensive review of deep learning-based variant calling methods. *Briefings in Functional Genomics*, 23(4), 303-313. <https://doi.org/10.1093/bfpg/elae003>
7. Váradi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., Yuan, D., Stroe, O., Wood, G., Laydon, A., Židek, A., Green, T., Tunyasuvunakool, K., Petersen, S., Jumper, J., Clancy, E., Green, R., Vora, A., Lutfi, M.,... Velankar, S. (2021). AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic acids research*, 50(D1), D439-D444. <https://doi.org/10.1093/nar/gkab1061>
8. Dixit, S., Kumar, A., Srinivasan, K., Vincent, P. M. D. R., & Krishnan, N. R. (2024). Advancing genome editing with artificial intelligence: opportunities, challenges, and future directions. *Frontiers in bioengineering and biotechnology*, 11, 1335901. <https://doi.org/10.3389/fbioe.2023.1335901>
9. Pakpahan, I., Sihombing, M., Liu, H., Wang, M., Su, Z., & Fang, M. (2026). Harnessing artificial intelligence for genomic variant prediction: advances, challenges, and future directions. *GigaScience*, 15. <https://doi.org/10.1093/gigascience/giag004>
10. Zucca, S., Nicora, G., Paoli, F. D., Carta, M. G., Bellazzi, R., Magni, P., et al. (2024). An AI-based approach driven by genotypes and phenotypes to uplift the diagnostic yield of genetic diseases. *Human Genetics*, 144(2), 159-171 <https://doi.org/10.1007/s00439-023-02638-x>

11. Martin, A. R., Kanai, M., Kamatani, Y., Okada, Y., Neale, B. M., & Daly, M. J. (2019). Clinical use of current polygenic risk scores may exacerbate health disparities. *Nature genetics*, 51(4), 584–591. <https://doi.org/10.1038/s41588-019-0379-x>
12. Gyawali, P. K., Le Guen, Y., Liu, X., Belloy, M. E., Tang, H., Zou, J., & He, Z. (2023). Improving genetic risk prediction across diverse population by disentangling ancestry representations. *Communications Biology*, 6(1), 964. <https://doi.org/10.1038/s42003-023-05352-6>
13. Popejoy, A. B., & Fullerton, S. M. (2016). Genomics is failing on diversity. *Nature*, 538(7624), 161–164. <https://doi.org/10.1038/538161a>
14. Corpas, M., Pius, M., Poburennaya, M., Guio, H., Dwek, M., Nagaraj, S. H., López-Correa, C., et al. (2024). Bridging genomics' greatest challenge: The diversity gap. *Cell Genomics*, 5(1), 100724. <https://doi.org/10.1016/j.xgen.2024.100724>
15. Sassi, Z., Eickmann, S., Roller, R., Osmanodja, B., Burchardt, A., Tretter, M., Samhammer, D., Dabrock, P., Möller, S., Budde, K., & Herrmann, A. (2025). Human-centered AI in healthcare: empowering patients and support persons in clinical decision-making. *BMC Medical Informatics and Decision Making*, 25(1), 431. <https://doi.org/10.1186/s12911-025-03298-9>
16. Johnson, K. B., Wei, W.-Q., Weeraratne, D., Frisse, M. E., Misulis, K., Rhee, K., Zhao, J., & Snowdon, J. L. (2020). Precision Medicine, AI, and the Future of Personalized Health Care. *Clinical and translational science*, 14(1), 86-93. <https://doi.org/10.1111/cts.12884>
17. Tan, M. J. T., Kasireddy, H. R., Satriya, A. B., Karim, H. A., & AlDahoul, N. (2025). Health is beyond genetics: on the integration of lifestyle and environment in real-time for hyper-personalized medicine. *Frontiers in Public Health*, 12, 1522673. <https://doi.org/10.3389/fpubh.2024.1522673>
18. Tian, Q., Price, N. D., & Hood, L. (2012). Systems cancer medicine: towards realization of predictive, preventive, personalized and participatory (P4) medicine. *Journal of internal medicine*, 271(2), 111-121. <https://doi.org/10.1111/j.1365-2796.2011.02498.x>
19. Chandrashekar, K., Niranjana, V., Vishal, A., & Setlur, A. S. (2024). Integration of Artificial Intelligence, Machine Learning and Deep Learning Techniques in Genomics: Review on Computational Perspectives for NGS Analysis of DNA and RNA Seq Data. *Current Bioinformatics*, 19(9), 825-844. <https://doi.org/10.2174/0115748936284044240108074937>

20. Alagarwamy, K., Shi, W., Boini, A., Messaoudi, N., Grasso, V., Cattabiani, T., et al. (2024). Should AI-Powered Whole-Genome Sequencing Be Used Routinely for Personalized Decision Support in Surgical Oncology—A Scoping Review. *BioMedInformatics*, 4(3), 1757. <https://doi.org/10.3390/biomedinformatics4030096>
21. Berger, B., & Yu, Y. W. (2022). Navigating bottlenecks and trade-offs in genomic data analysis. *Nature Reviews Genetics*, 24(4), 235-250. <https://doi.org/10.1038/s41576-022-00551-z>
22. Luo, R., Sedlazeck, F. J., Lam, T., & Schatz, M. C. (2019). A multi-task convolutional deep neural network for variant calling in single molecule sequencing. *Nature communications*, 10(1), 998. <https://doi.org/10.1038/s41467-019-09025-z>
23. Gangwal, A., & Lavecchia, A. (2024). AI-Driven Drug Discovery for Rare Diseases. *Journal of Chemical Information and Modeling*, 65(5), 2214-2231. <https://doi.org/10.1021/acs.jcim.4c01966>
24. Abdelwahab, O., & Torkamaneh, D. (2024). Refinement of genetic variants needs attention. arXiv preprint arXiv:2408.00659. <https://doi.org/10.48550/arxiv.2408.00659>
25. Cai, L., Wu, Y., & Gao, J. (2019). DeepSV: accurate calling of genomic deletions from high-throughput sequencing data using deep convolutional neural network. *BMC Bioinformatics*, 20(1), 665. <https://doi.org/10.1186/s12859-019-3299-y>
26. Ip, E., Hadinata, C., Ho, J. W. K., & Giannoulatou, E. (2020). dv-trio: a family-based variant calling pipeline using DeepVariant. *Bioinformatics*, 36(11), 3549. <https://doi.org/10.1093/bioinformatics/btaa116>
27. Brouard, J., Schenkel, F. S., Marete, A., & Bissonnette, N. (2019). The GATK joint genotyping workflow is appropriate for calling variants in RNA-seq experiments. *Journal of Animal Science and Biotechnology/Journal of Animal Science and Biotechnology*, 10(1), 44. <https://doi.org/10.1186/s40104-019-0359-0>
28. Yun, T., Li, H., Chang, P.-C., Lin, M., Carroll, A., & McLean, C. Y. (2020). Accurate, scalable cohort variant calls using DeepVariant and GLnexus. *Bioinformatics*, 36(24), 5582. <https://doi.org/10.1093/bioinformatics/btaa1081>
29. Abdelwahab, O., & Torkamaneh, D. (2025). Artificial intelligence in variant calling: a review. *Frontiers in Bioinformatics*, 5. <https://doi.org/10.3389/fbinf.2025.1574359>

30. Vega, F. M. D. L., Chowdhury, S., Moore, B., Frise, E., McCarthy, J., Hernández, E. J., et al. (2021). Artificial intelligence enables comprehensive genome interpretation and nomination of candidate diagnoses for rare genetic diseases. *Genome Medicine*, 13(1), 153. <https://doi.org/10.1186/s13073-021-00965-0>
31. Wang, S., & Ye, K. (2024). Deep-learning based representation and recognition for genome variants—from SNVs to structural variants. *National Science Review*, 11(11). <https://doi.org/10.1093/nsr/nwae335>
32. Chen, N.-C., Kolesnikov, A., Goel, S., Yun, T., Chang, P.-C., & Carroll, A. (2023). Improving variant calling using population data and deep learning. *BMC bioinformatics*, 24(1), 197. <https://doi.org/10.1186/s12859-023-05294-0>
33. Spinella, J.-F., Mehanna, P., Vidal, R., Saillour, V., Cassart, P., Richer, C., Ouimet, M., Healy, J., & Sinnett, D. (2016). SNooPer: a machine learning-based method for somatic variant identification from low-pass next-generation sequencing. *BMC genomics*, 17(1), 912. <https://doi.org/10.1186/s12864-016-3281-2>
34. Dashti, M. J. S., & Gamiieldien, J. (2017). A Practical Guide to Filtering and Prioritizing Genetic Variants. *Biotechniques*, 62(1), 18-30. <https://doi.org/10.2144/000114492>
35. Koboldt, D. C. (2020). Best practices for variant calling in clinical sequencing. *Genome medicine*, 12(1), 91. *Genome medicine*, 12(1), 91. <https://doi.org/10.1186/s13073-020-00791-w>
36. Duzkale, H., Shen, J., McLaughlin, H. M., Alfares, A., Kelly, M., Pugh, T. J., et al. (2013). A systematic approach to assessing the clinical significance of genetic variants. *Clinical genetics*, 84(5), 453-463. <https://doi.org/10.1111/cge.12257>
37. Çifçi, D., Foersch, S., & Kather, J. N. (2022). Artificial intelligence to identify genetic alterations in conventional histopathology. *The Journal of pathology*, 257(4), 430-444. <https://doi.org/10.1002/path.5898>
38. Larrea-Sebal, A., Benslaiman, S. J., García, U. G., Jose-Urteaga, A. S., Uribe, K. B., Benito-Vicente, A., & Martín, C. (2023). Predictive Modeling and Structure Analysis of Genetic Variants in Familial Hypercholesterolemia: Implications for Diagnosis and Protein Interaction Studies. *Current Atherosclerosis Reports*, 25(11), 839. <https://doi.org/10.1007/s11883-023-01154-7>

39. Rastogi, R., Chung, R., Li, S., Li, C., Lee, K., Woo, J., et al. (2025). Critical assessment of missense variant effect predictors on disease-relevant variant data. *Human Genetics*, 144(2), 281-93. <https://doi.org/10.1007/s00439-025-02732-2>
40. Maslej, N., Fattorini, L., Perrault, R., Parli, V., Reuel, A., Brynjolfsson, E., et al. (2024). Artificial Intelligence Index Report 2024. arXiv preprint arXiv:2405.19522. <https://doi.org/10.48550/arxiv.2405.19522>
41. Růžicka, J., Ravel, J., Audoux, J., Boulat, A., Thévenon, J., Yaury, K., et al. (2025). Clinical Evaluation of an AI System for Streamlined Variant Interpretation in Genetic Testing. *bioRxiv*. <https://doi.org/10.1101/2025.02.04.25321641>
42. Pushkaran, A. C., & Arabi, A. A. (2024). From understanding diseases to drug design: can artificial intelligence bridge the gap? *Artificial Intelligence Review*, 57(4). <https://doi.org/10.1007/s10462-024-10714-5>
43. Kong, S. W., Lee, I., Collen, L. V., Manrai, A. K., Snapper, S. B., & Mandl, K. D. (2024). Discordance between a deep learning model and clinical-grade variant pathogenicity classification in a rare disease cohort. *medRxiv*. <https://doi.org/10.1101/2024.05.22.24307756>
44. Abbas, S., Abbas, Z., Zahir, A., & Lee, S. W. (2025). Advancing genome-based precision medicine: a review on machine learning applications for rare genetic disorders. *Briefings in Bioinformatics*, 26(4). <https://doi.org/10.1093/bib/bbaf329>
45. Koboldt, D. C., Zhang, Q., Larson, D. E., Shen, D., McLellan, M. D., Lin, L., Miller, C. A., Mardis, E. R., Ding, L., & Wilson, R. K. (2012). VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Research*, 22(3), 568. <https://doi.org/10.1101/gr.129684.111>
46. Eghbal-zadeh, H., Fischer, L., Popitsch, N., Kromp, F., Taschner-Mandl, S., Gerber, T., et al. (2018). DeepSNP: An End-to-End Deep Neural Network with Attention-Based Localization for Breakpoint Detection in Single-Nucleotide Polymorphism Array Genomic Data. *Journal of Computational Biology*, 26(6), 572. <https://doi.org/10.1089/cmb.2018.0172>
47. Huang, N., Xu, M., Nie, F., Ni, P., Xiao, C., Luo, F., & Wang, J. (2022). NanoSNP: a progressive and haplotype-aware SNP caller on low-coverage nanopore sequencing data. *Bioinformatics*, 39(1). <https://doi.org/10.1093/bioinformatics/btac824>

48. Abdelwahab, O., Belzile, F., & Torkamaneh, D. (2023). Performance analysis of conventional and AI-based variant callers using short and long reads. *BMC Bioinformatics*, 24(1), 472. <https://doi.org/10.1186/s12859-023-05596-3>
49. Su, J., Zheng, Z., Ahmed, S. S., Lam, T., & Luo, R. (2022). Clair3-trio: high-performance Nanopore long-read variant calling in family trios with trio-to-trio deep neural networks. *Briefings in Bioinformatics*, 23(5). <https://doi.org/10.1093/bib/bbac301>
50. Genovese, G., Rockweiler, N. B., Gorman, B. R., Bigdeli, T. B., Pato, M. T., Pato, C. N., et al. (2024). BCFtools/liftover: an accurate and comprehensive tool to convert genetic variants across genome assemblies. *Bioinformatics*, 40(2). <https://doi.org/10.1093/bioinformatics/btae038>
51. Rimmer, A., Phan, H., Mathieson, I., Iqbal, Z., Twigg, S. R. F., Wilkie, A. O. M., et al. (2014). Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. *Nature Genetics*, 46(8), 912. <https://doi.org/10.1038/ng.3036>
52. Richter, F., Morton, S. U., Qi, H., Kitaygorodsky, A., Wang, J., Homsy, J., et al. (2020). Whole Genome De Novo Variant Identification with FreeBayes and Neural Network Approaches. *BioRxiv*, 2020-03. <https://doi.org/10.1101/2020.03.24.994160>
53. Vaisband, M., Schubert, M., Gassner, F. J., Geisberger, R., Greil, R., Zaborsky, N., et al. (2023). Validation of genetic variants from NGS data using deep convolutional neural networks. *BMC Bioinformatics*, 24(1), 158. <https://doi.org/10.1186/s12859-023-05255-7>
54. Paoli, F. D., Berardelli, S., Limongelli, I., Rizzo, E., & Zucca, S. (2024). VarChat: the generative AI assistant for the interpretation of human genomic variations. *Bioinformatics*, 40(4). <https://doi.org/10.1093/bioinformatics/btae183>
55. Boudelloua, I., Kulmanov, M., Schofield, P. N., Gkoutos, G. V., & Hoehndorf, R. (2019). DeepPVP: phenotype-based prioritization of causative variants using deep learning. *BMC Bioinformatics*, 20(1), 65. <https://doi.org/10.1186/s12859-019-2633-8>
56. Mahajan, R., Frangione, E., Aujla, N., Casalino, S., Jayachandran, L., Morgan, G., et al. (2024). P733: Comparative analysis of DNA variant classifications between the GENCOV COVID-19 genome study and the ClinVar database. *Genetics in Medicine Open*, 2, 101637. <https://doi.org/10.1016/j.gimo.2024.101637>
57. Bastida-Lertxundi, N., Martí, I., Laña-Ruiz, B., Barbosa, O. M.-M., Muguerza-Iraola, R., Sáez-Villaverde, R., et al. (2025). Evaluating seven bioinformatics platforms for tertiary analysis of genomic data from whole exome sequencing in a pilot group of patients. *Advances in*

Laboratory Medicine / Avances En Medicina de Laboratorio, 6(1), 28.

<https://doi.org/10.1515/almed-2025-0101>

58. Ge, S., Sun, S., Xu, H., Cheng, Q., & Ren, Z. (2025). Deep learning in single-cell and spatial transcriptomics data analysis: advances and challenges from a data science perspective. *Briefings in Bioinformatics*, 26(2), bbaf136. <https://doi.org/10.1093/bib/bbaf136>

59. Ali, M. I., Shabbir, K., Ali, S., Mohsin, M., Kumar, A., Aziz, M., et al. (2024). A New Era of Discovery: How Artificial Intelligence has Revolutionized the Biotechnology. *Nepal Journal of Biotechnology*, 12(1). <https://doi.org/10.54796/njb.v12i1.312>

60. Huang, H., Liu, C., Wagle, M. M., & Yang, P. (2023). Evaluation of deep learning-based feature selection for single-cell RNA sequencing data analysis. *Genome Biology*, 24(1). <https://doi.org/10.1186/s13059-023-03100-x>

61. Usman, M., Varea, O., Radeva, P., Canals, J. M., Abante, J., & Ortiz, D. G. (2025). Explainable AI model reveals disease-related mechanisms in single-cell RNA-seq data. *arXiv preprint arXiv:2501.03923*. <https://doi.org/10.48550/arxiv.2501.03923>

62. Tahir, M., Norouzi, M., Khan, S. S., Davie, J., Yamanaka, S., & Ashraf, A. (2024). Artificial intelligence and deep learning algorithms for epigenetic sequence analysis: A review for epigeneticists and AI experts. *Computers in Biology and Medicine*, 183, 109302. <https://doi.org/10.1016/j.combiomed.2024.109302>

63. Sahoo, K., Lingasamy, P., Khatun, M., Lulu, S., Salumets, A., Sundararajan, V., & Modhukur, V. (2025). Artificial Intelligence in cancer epigenomics: a review on advances in pan-cancer detection and precision medicine. *Epigenetics & Chromatin*, 18(1), 35. <https://doi.org/10.1186/s13072-025-00595-5>

64. Saraswat, A., & Roopesh, S. (2024). Machine Learning in Genomic Data Analysis for Personalized Medicine. *International Journal for Research in Applied Science and Engineering Technology*, 12(8), 614. <https://doi.org/10.22214/ijraset.2024.63975>

65. Wall, B. P. G., Nguyen, M., Harrell, J. C., & Dozmorov, M. G. (2024). Machine and Deep Learning Methods for Predicting 3D Genome Organization. *Methods in Molecular Biology*, 2856, 357. https://doi.org/10.1007/978-1-0716-4136-1_22

66. Sanches, P. H. G., Melo, N., Porcari, A. M., & Carvalho, L. M. de. (2024). Integrating Molecular Perspectives: Strategies for Comprehensive Multi-Omics Integrative Data Analysis

and Machine Learning Applications in Transcriptomics, Proteomics, and Metabolomics. *Biology*, 13(11), 848. <https://doi.org/10.3390/biology13110848>

67. Shoaib, A. S. M., Nishat, N., Raasetti, M., & Arif, I. (2024). Integrative machine learning approaches for multi-omics data analysis in cancer research. *Deleted Journal*, 1(2), 26. <https://doi.org/10.62304/ijhm.v1i2.149>

68. Hai, L., Jiang, Z., Zhang, H., & Sun, Y. (2024). From multi-omics to predictive biomarker: AI in tumor microenvironment. *Frontiers in Immunology*, 15. <https://doi.org/10.3389/fimmu.2024.1514977>

69. Qureshi, M. A., Ramzan, M., Amjad, F., & Haider, N. (2024). Artificial Intelligence in Metabolomics for Disease Profiling: A Machine Learning Approach to Biomarker Discovery. *Indus Journal of Bioscience Research*, 2(2), 87. <https://doi.org/10.70749/ijbr.v2i02.146>

70. Licciardi, A., Fiannaca, A., Rosa, M. L., Urso, A., & Paglia, L. L. (2024). A Deep Learning Multi-omics Framework to Combine Microbiome and Metabolome Profiles for Disease Classification. In: *Artificial Neural Networks and Machine Learning – ICANN 2024*. Cham: Springer Nature, 3-14. https://doi.org/10.1007/978-3-031-72353-7_1

71. Lázari, L. C., Oliveira, G. S. de, Macedo-da-Silva, J., Rosa-Fernandes, L., & Palmisano, G. (2023). Mass spectrometry and machine learning in the identification of COVID-19 biomarkers. *Frontiers in Analytical Science*, 3, 1119438. <https://doi.org/10.3389/frans.2023.1119438>

72. Xu, Y., Cao, L., Chen, Y., Zhang, Z., Liu, W., Li, H., et al. (2024). Integrating Machine Learning in Metabolomics: A Path to Enhanced Diagnostics and Data Interpretation. *Small Methods*, 8(12). <https://doi.org/10.1002/smt.202400305>

73. Ballard, J. L., Wang, Z., Li, W., Shen, L., & Long, Q. (2024). Deep learning-based approaches for multi-omics data integration and analysis. *BioData Mining*, 17(1), 38. <https://doi.org/10.1186/s13040-024-00391-z>

74. Sun, Y., Ai, J., Liu, Z., Sun, R., Qian, L., Payne, S., et al. (2025). Strategic priorities for transformative progress in advancing biology with proteomics and artificial intelligence. *arXiv*. <https://doi.org/10.48550/arxiv.2502.15867>

75. Sun, N. (2025). Challenges and potential applications of AI in systems biology. *Nature Reviews Molecular Cell Biology*, 27(2), 89-90. <https://doi.org/10.1038/s41580-025-00934-0>

76. Wu, Y., & Xie, L. (2025). AI-driven multi-omics integration for multi-scale predictive modeling of genotype-environment-phenotype relationships. *Computational and Structural Biotechnology Journal*, 27, 265-277. <https://doi.org/10.1016/j.csbj.2024.12.030>
77. Baião, A. R., Cai, Z., Poulos, R. C., Robinson, P. J., Reddel, R. R., et al. (2025). A technical review of multi-omics data integration methods: from classical statistical to deep generative approaches. *Briefings in bioinformatics*, 26(4), bbaf355. <https://doi.org/10.1093/bib/bbaf355>
78. Benkirane, H., Pradat, Y., Michiels, S., & Cournède, P. H. (2023). CustOmics: A versatile deep-learning based strategy for multi-omics integration. *PLOS Computational Biology*, 19(3), e1010921. <https://doi.org/10.1371/journal.pcbi.1010921>
79. Ali, A. M., & Mohammed, M. A. (2023). A Comprehensive Review of Artificial Intelligence Approaches in Omics Data Processing: Evaluating Progress and Challenges. *International Journal of Mathematics Statistics and Computer Science*, 2, 114. <https://doi.org/10.59543/ijmscs.v2i.8703>
80. Yetgin, A. (2025). Revolutionizing multi-omics analysis with artificial intelligence and data processing. *Quantitative Biology*, 13(3). <https://doi.org/10.1002/qub2.70002>
81. Noordijk, B., Garcia Gomez, M. L., ten Tusscher, K. H. W. J., de Ridder, D., van Dijk, A. D. J., & Smith, R. W. (2024). The rise of scientific machine learning: a perspective on combining mechanistic modelling with machine learning for systems biology. *Frontiers in Systems Biology*, 4, 1407994. <https://doi.org/10.3389/fsysb.2024.1407994>
82. Ahmadi, N., Cao, Q., Humphrey, J. D., & Karniadakis, G. E. (2025). Physics-Informed Machine Learning in Biomedical Science and Engineering. *arXiv*. <https://doi.org/10.48550/arxiv.2510.05433>
83. Voit, E. O., Shah, A. M., Olivença, D. V., & Vodovotz, Y. (2023). What's next for computational systems biology?. *Frontiers in Systems Biology*, 3, 1250228. <https://doi.org/10.3389/fsysb.2023.1250228>
84. Silva-Sousa, T., Usuda, J. N., Al-Arawe, N., Hinterseher, I., Catar, R., Luecht, et al. (2025). Artificial intelligence and systems biology analysis in stem cell research and therapeutics development. *Stem Cells Translational Medicine*, 14(10). <https://doi.org/10.1093/stcltm/szaf037>
85. Hein, Z. M., Guruparan, D., Okunsai, B., Nassir, C. M. N. C. M., Ramli, M. D. C., & Kumar, S. (2025). AI and Machine Learning in Biology: From Genes to Proteins. *Biology*, 14(10), 1453. <https://doi.org/10.3390/biology14101453>

86. Fang, Z., Zhang, X., Zhao, A., Li, X., Chen, H., & Li, J. (2025). Recent Developments in GNNs for Drug Discovery. arXiv. <https://doi.org/10.48550/arxiv.2506.01302>
87. Choubisa, D. (2024). An overview of graph neural networks for molecular biology: Challenges and solutions. *International Journal of Science and Research Archive*, 13(1), 2670. <https://doi.org/10.30574/ijrsra.2024.13.1.1985>
88. Zhang, O., Lin, H., Zhang, X., Wang, X., Wu, Z., Ye, Q., Zhao, W., Wang, J., Ying, K., Kang, Y., Hsieh, C.-Y., & Hou, T. (2025). Graph Neural Networks in Modern AI-Aided Drug Discovery. *Chemical Reviews*, 125(20), 10001. <https://doi.org/10.1021/acs.chemrev.5c00461>
89. de Weerd, H. (2023). Novel methods and software for disease module inference. [dissertation]. Linköping: Linköping University Electronic Press, <https://doi.org/10.3384/9789180750097>
90. Yang, L., Wang, H., Zhu, Z., Yang, Y., Xiong, Y., Cui, X., & Liu, Y. (2025). Network Pharmacology-Driven Sustainability: AI and Multi-Omics Synergy for Drug Discovery in Traditional Chinese Medicine. *Pharmaceuticals*, 18(7), 1074. <https://doi.org/10.3390/ph18071074>
91. Wang, R., & Zhuang, C. (2025). Graph neural networks driven acceleration in drug discovery. *Acta Pharmaceutica Sinica B*, 15(12), 6163. <https://doi.org/10.1016/j.apsb.2025.10.011>
92. Huang, K., Chandak, P., Wang, Q., Havaladar, S., Vaid, A., Leskovec, J., Nadkarni, G. N., Glicksberg, B. S., Gehlenborg, N., & Žitnik, M. (2024). A foundation model for clinician-centered drug repurposing. *Nature Medicine*, 30(12), 3601. <https://doi.org/10.1038/s41591-024-03233-x>
93. Ferreira, F. J. N., & Carneiro, A. S. (2025). AI-Driven Drug Discovery: A Comprehensive Review. *ACS Omega*, 10(23), 23889. <https://doi.org/10.1021/acsomega.5c00549>
94. Alberuni, S., & Ray, S. (2024). Integration of biological data via NMF for identification of human disease-associated gene modules through multi-label classification. *PLoS ONE*, 19(12). <https://doi.org/10.1371/journal.pone.0305503>
95. Charles, S., & Natarajan, J. (2023). Identification of Key Gene Modules and Novel Transcription Factors in Tetralogy of Fallot Using Machine Learning and Network Topological Features. *Deleted Journal*, 1(1), 27. <https://doi.org/10.47852/bonviewmedin32021554>

96. Subirana-Granés, M., Hoffman, J., Zhang, H., Akirtava, C., Nandi, S., Fotso, K., & Pividori, M. (2025). Genetic studies through the lens of gene networks. *Annual review of biomedical data science*, 8(1), 125-47. <https://doi.org/10.1146/annurev-biodatasci-103123-095355>
97. Žitnik, M., Li, M. M., Wells, A. V., Glass, K., Gysi, D. M., Krishnan, A., et al. (2024). Current and future directions in network biology. *Bioinformatics Advances*, 4(1). <https://doi.org/10.1093/bioadv/vbae099>
98. Babu, B. K. (2024). Revolutionizing Herbal Medicine: The Role of E-health Informatics and Network Pharmacology in Personalized Herbal Therapies. *International Journal of Pharmaceutical Investigation*, 15(1), 219. <https://doi.org/10.5530/ijpi.20251754>
99. Coudray, N., Ocampo, P., Sakellaropoulos, T., Narula, N., Snuderl, M., Fenyő, D., et al. (2018). Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nature medicine*, 24(10), 1559-1567. <https://doi.org/10.1038/s41591-018-0177-5>
100. Alkuraya, I. F. (2023). Is artificial intelligence getting too much credit in medical genetics? *In American Journal of Medical Genetics*, 19(3), e32062. <https://doi.org/10.1002/ajmg.c.32062>
101. Kitaoka, Y., Uchihashi, T., Kawata, S., Nishiura, A., Yamamoto, T., Hiraoka, S., et al. (2025). Role and Potential of Artificial Intelligence in Biomarker Discovery and Development of Treatment Strategies for Amyotrophic Lateral Sclerosis. *International Journal of Molecular Sciences*, 26(9), 4346. <https://doi.org/10.3390/ijms26094346>
102. Reyes, S. O. L. (2020). Artificial intelligence in precision health: Systems in practice. In *Artificial intelligence in precision health* (pp. 499-519). Academic Press. <https://doi.org/10.1016/b978-0-12-817133-2.00021-5>
103. Pereira, L. A. M., Rijo, R., Silva, C., & Martinho, R. (2015). Text Mining Applied to Electronic Medical Records. *International Journal of E-Health and Medical Communications (IJEHMC)*, 6(3), 1-18. <https://doi.org/10.4018/ijehmc.2015070101>
104. Hruska, M., Džubák, P., & Hajdúch, M. (2024). Claire: a search engine leveraging deep databases for reliable detection of peptides, with a special focus on proteogenomic applications. *Cancer Research*, 84(6), 7414-7414. <https://doi.org/10.1158/1538-7445.am2024-7414>
105. Meng, L., Attali, R., Talmy, T., Regev, Y., Mizrahi, N., Smirin-Yosef, P., et al. (2023). Evaluation of an automated genome interpretation model for rare disease routinely used in a

clinical genetic laboratory. *Genetics in Medicine*, 25(6), 100830.

<https://doi.org/10.1016/j.gim.2023.100830>

106. Ilan, Y. (2020). Second-Generation Digital Health Platforms: Placing the Patient at the Center and Focusing on Clinical Outcomes. *Frontiers in Digital Health*, 2, 569178.

<https://doi.org/10.3389/fdgth.2020.569178>

107. Hurvitz, N., Azmanov, H., Kesler, A., & Ilan, Y. (2021). Establishing a second-generation artificial intelligence-based system for improving diagnosis, treatment, and monitoring of patients with rare diseases. *European Journal of Human Genetics*, 29(10), 1485-1490.

<https://doi.org/10.1038/s41431-021-00928-4>

108. Abdallah, S., Sharifa, M., Almadhoun, M. K. I. K., Khawar, M. M., Shaikh, U., Balabel, K. M., Saleh, I., Manzoor, A., Mandal, A. K., Ekomwereren, O., Khine, W. M., & Oyelaja, O. T. (2023). The Impact of Artificial Intelligence on Optimizing Diagnosis and Treatment Plans for Rare Genetic Disorders. *Cureus*, 15(10). <https://doi.org/10.7759/cureus.46860>

109. Montalvo, N., Requena, F., Capriotti, E., & Rausell, A. (2025). Federated learning for the pathogenicity annotation of genetic variants in multi-site clinical settings. *Bioinformatics*, 41(10), btaf523. <https://doi.org/10.1093/bioinformatics/btaf523>

110. Meduri, K., Nadella, G. S., Yadulla, A. R., Kasula, V. K., Maturi, M. H., Brown, S., et al. (2025). Leveraging Federated Learning for Privacy-Preserving Analysis of Multi-Institutional Electronic Health Records in Rare Disease Research. *Journal of Economy and Technology*, 3, 177-189. <https://doi.org/10.1016/j.ject.2024.11.001>

111. Kolobkov, D., Sharma, S. M., Medvedev, A., Lebedev, M., Kosaretskiy, E., & Vakhitov, R. (2024). Efficacy of federated learning on genomic data: a study on the UK Biobank and the 1000 Genomes Project. *Frontiers in Big Data*, 7. <https://doi.org/10.3389/fdata.2024.1266031>

112. Raimondi, D., Chizari, H., Verplaetse, N., Löscher, B.-S., Franke, A., & Moreau, Y. (2023). Genome interpretation in a federated learning context allows the multi-center exome-based risk prediction of Crohn's disease patients. *Scientific Reports*, 13(1). <https://doi.org/10.1038/s41598-023-46887-2>

113. Shyr, C., Hu, Y., Tinker, R. J., Cassini, T. A., Byram, K. W., Hamid, R., et al. (2026). An artificial intelligence framework for end-to-end rare disease phenotyping from clinical notes using large language models. *arXiv preprint arXiv:2602.20324*.

<https://doi.org/10.48550/arXiv.2602.20324>

114. Wu, J., Dong, H., Li, Z., Wang, H., Li, R., Patra, A., Dai, C., Ali, W., Scordis, P., & Wu, H. (2024). A hybrid framework with large language models for rare disease phenotyping. *BMC Medical Informatics and Decision Making*, 24(1), 289. <https://doi.org/10.1186/s12911-024-02698-7>
115. Rolando, M., Raggio, V., Naya, H., Spangenberg, L., & Cagnina, L. (2024). RDCorpus: labeled medical records for the timely detection of rare diseases. *Research Square*. <https://doi.org/10.21203/rs.3.rs-4795232/v1>
116. Kim, S., Zhou, Y., Guo, Y., Xiao, C., & Zheng, K. (2026). Applying natural language processing and large language models to clinical notes for phenotyping and diagnosing rare diseases: a systematic review. *Journal of the American Medical Informatics Association*, 33(6), 1225-1235. <https://doi.org/10.1093/jamia/ocag045>
117. Shyr, C., Hu, Y., Bastarache, L., Cheng, A., Hamid, R., Harris, P., & Xu, H. (2024). Identifying and extracting rare diseases and their phenotypes with large language models. *Journal of Healthcare Informatics Research*, 8(2), 438-461. <https://doi.org/10.1007/s41666-023-00155-0>
118. Xi, N. M., Deng, Y., Wang, L. (2026). Leveraging large language models for rare disease named entity recognition. *PLOS Digital Health*, 5(2), e0001242. <https://doi.org/10.1371/journal.pdig.0001242>
119. Gordon, E. J., Gacki-Smith, J., Gooden, M. J., Waite, P., Yacat, R., Abubakari, Z. R., et al. (2024). Development of a culturally targeted chatbot to inform living kidney donor candidates of African ancestry about APOL1 genetic testing: a mixed methods study. *Journal of Community Genetics*, 15(2), 205-216. <https://doi.org/10.1007/s12687-024-00698-8>
120. Sato, A., Haneda, E., Hiroshima, Y., & Narimatsu, H. (2024). Preliminary Screening for Hereditary Breast and Ovarian Cancer Using an AI Chatbot as a Genetic Counselor: Clinical Study. *Journal of Medical Internet Research*, 26, e48914. <https://doi.org/10.2196/48914>
121. Kohut, K., Limb, S., & Crawford, G. (2019). The Changing Role of the Genetic Counsellor in the Genomics Era. *Current Genetic Medicine Reports*, 7(2), 75-84. <https://doi.org/10.1007/s40142-019-00163-w>
122. Ahmed, Z., Mohamed, K. G., Zeeshan, S., & Dong, X. (2020). Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine. *Database*, 2020, baaa010. <https://doi.org/10.1093/database/baaa010>

123. Kayser, M., Branicki, W., Parson, W., & Phillips, C. (2023). Recent advances in Forensic DNA Phenotyping of appearance, ancestry and age. *Forensic Science International: Genetics*, 65, 102870. <https://doi.org/10.1016/j.fsigen.2023.102870>
124. Crouch, D. J. M., & Bodmer, W. F. (2020). Polygenic inheritance, GWAS, polygenic risk scores, and the search for functional variants. *Proceedings of the National Academy of Sciences*, 117(32), 18924-18933. <https://doi.org/10.1073/pnas.2005634117>
125. Manthena, V., Jarquín, D., & Howard, R. (2023). Integrating and optimizing genomic, weather, and secondary trait data for multiclass classification. *Frontiers in Genetics*, 13, 1032691. <https://doi.org/10.3389/fgene.2022.1032691>
126. Li, B., Wang, Z., Chen, Q., Li, K., Wang, X., Wang, Y., Zeng, Q., Han, Y., Lu, B., Zhao, Y., Zhang, R., Jiang, L., Pan, H., Luo, T., Zhang, Y., Fang, Z., Xiao, X., Zhou, X., Wang, R.,... Li, J. (2021). GPCards: An integrated database of genotype–phenotype correlations in human genetic diseases. *Computational and Structural Biotechnology Journal*, 19, 1603-1611. <https://doi.org/10.1016/j.csbj.2021.03.011>
127. Muneeb, M., Feng, S. F., & Henschel, A. (2022). Transfer learning for genotype–phenotype prediction using deep learning models. *BMC bioinformatics*, 23(1), 511. <https://doi.org/10.1186/s12859-022-05036-8>
128. Sharma, A., Lysenko, A., Jia, S., Boroevich, K. A., & Tsunoda, T. (2024). Advances in AI and machine learning for predictive medicine. *Journal of Human Genetics*, 69(10), 487-497. <https://doi.org/10.1038/s10038-024-01231-y>
129. Kukla-Bartoszek, M., Teisseyre, P., Pośpiech, E., Karłowska-Pik, J., Zieliński, P., Woźniak, A., Boroń, M., Dąbrowski, M., Zubańska, M., Jarosz, A., Płoski, R., Grzybowski, T., Spólnicka, M., Mielniczuk, J., & Branicki, W. (2021). Searching for improvements in predicting human eye colour from DNA. *International Journal of Legal Medicine*, 135(6), 2175-2187. <https://doi.org/10.1007/s00414-021-02645-5>
130. Díaz-Santiago, E., Jabato, F. M., Rojano, E., Seoane, P., Pazos, F., Perkins, J. R., et al. (2020). Phenotype-genotype comorbidity analysis of patients with rare disorders provides insight into their pathological and molecular bases. *PLoS genetics*, 16(10), e1009054. <https://doi.org/10.1371/journal.pgen.1009054>
131. Smedley, D., Jacobsen, J. O. B., Jäger, M., Köhler, S., Holtgrewe, M., Schubach, M., Siragusa, E., Zemojtel, T., Buske, O. J., Washington, N. L., Bone, W. P., Haendel, M. A., &

- Robinson, P. N. (2015). Next-generation diagnostics and disease-gene discovery with the Exomiser. *Nature Protocols*, 10(12), 2004. <https://doi.org/10.1038/nprot.2015.124>
132. Robinson, P. N., Ravanmehr, V., Jacobsen, J. O. B., Daniš, D., Zhang, X., Carmody, L., et al. (2020). Interpretable Clinical Genomics with a Likelihood Ratio Paradigm. *The American Journal of Human Genetics*, 107(3), 403. <https://doi.org/10.1016/j.ajhg.2020.06.021>
133. Zhai, W., Huang, X., Shen, N., & Zhu, S. (2023). Phen2Disease: a phenotype-driven model for disease and gene prioritization by bidirectional maximum matching semantic similarities. *Briefings in Bioinformatics*, 24(4). <https://doi.org/10.1093/bib/bbad172>
134. Zhao, M., Havrilla, J. M., Li, F., Chen, Y., Peng, J., Liu, C., et al. (2020). Phen2Gene: rapid phenotype-driven gene prioritization for rare diseases. *NAR Genomics and Bioinformatics*, 2(2). <https://doi.org/10.1093/nargab/lqaa032>
135. Singleton, M., Guthery, S. L., Voelkerding, K. V., Chen, K., Kennedy, B., Margraf, R. L., et al. (2014). Phevor Combines Multiple Biomedical Ontologies for Accurate Identification of Disease-Causing Alleles in Single Individuals and Small Nuclear Families. *The American Journal of Human Genetics*, 94(4), 599. <https://doi.org/10.1016/j.ajhg.2014.03.010>
136. Pearson, N. M., Stolte, C., Shi, K., Beren, F., Abul-Husn, N. S., Bertier, G., Brown, K., Diaz, G. A., Odgis, J. A., Suckiel, S. A., Horowitz, C. R., Wasserstein, M., Gelb, B. D., & Kenny, E. E. (2021). GenomeDiver: a platform for phenotype-guided medical genomic diagnosis. *Genetics in Medicine*, 23(10), 1998. <https://doi.org/10.1038/s41436-021-01219-5>
137. Birgmeier, J., Haeussler, M., Deisseroth, C. A., Steinberg, E. H., Jagadeesh, K. A., Ratner, A. J., Guturu, H., Wenger, A. M., Diekhans, M. E., Stenson, P. D., Cooper, D. N., Ré, C., Beggs, A. H., Bernstein, J. A., & Bejerano, G. (2020). AMELIE speeds Mendelian diagnosis by matching patient phenotype and genotype to primary literature. *Science Translational Medicine*, 12(544). <https://doi.org/10.1126/scitranslmed.aau9113>
138. Sobreira, N., Schiettecatte, F., Valle, D., & Hamosh, A. (2015). GeneMatcher: A Matching Tool for Connecting Investigators with an Interest in the Same Gene. *Human Mutation*, 36(10), 928. <https://doi.org/10.1002/humu.22844>
139. Pantel, J. T., Hajjir, N., Danyel, M., Elsner, J., Abad-Perez, A. T., Hansen, P., Mundlos, S., Spielmann, M., Horn, D., Ott, C. E., & Mensah, M. A. (2020). Efficiency of Computer-Aided Facial Phenotyping (DeepGestalt) in Individuals With and Without a Genetic Syndrome: Diagnostic Accuracy Study. *Journal of Medical Internet Research*, 22(10), 0. <https://doi.org/10.2196/19263>

140. Li, Q.-G., Zhao, K., Bustamante, C. D., Ma, X., & Wong, W. H. (2019). Xrare: a machine learning method jointly modeling phenotypes and genetic evidence for rare disease diagnosis. *Genetics in Medicine*, 21(9), 2126. <https://doi.org/10.1038/s41436-019-0439-8>
141. Khanna, N. N., Singh, M., Maindarkar, M., Kumar, A., Johri, A. M., Mentella, L., et al. (2023). Polygenic Risk Score for Cardiovascular Diseases in Artificial Intelligence Paradigm: A Review. *Journal of Korean medical science*, 38(46). <https://doi.org/10.3346/jkms.2023.38.e395>
142. Klau, J. H., Maj, C., Klinkhammer, H., Krawitz, P., Mayr, A., Hillmer, A. M., Schumacher, J., & Heider, D. (2023). AI-based multi-PRS models outperform classical single-PRS models. *Frontiers in genetics*, 14, 1217860. <https://doi.org/10.3389/fgene.2023.1217860>
143. Huang, Y., Chen, C., & Yang, H. (2024). AI-enhanced integration of genetic and medical imaging data for risk assessment of Type 2 diabetes. *Nature communications*, 15(1), 4230. <https://doi.org/10.1038/s41467-024-48618-1>
144. Truong, B., Hull, L. E., Ruan, Y., Huang, Q., Hornsby, W., Martin, H. C., et al. (2024). Integrative polygenic risk score improves the prediction accuracy of complex traits and diseases. *Cell Genomics*, 4(4), 100523 <https://doi.org/10.1016/j.xgen.2024.100523>
145. Li, H., Zeng, J., Snyder, M., & Zhang, S. (2024). Modeling gene interactions in polygenic prediction via geometric deep learning. *Genome Research*, 35(1), 178-187. <https://doi.org/10.1101/gr.279694.124>
146. Gao, X., Chiariglione, M., Qin, K., Nuytemans, K., Scharre, D. W., Li, Y., & Martin, E. R. (2023). Explainable machine learning aggregates polygenic risk scores and electronic health records for Alzheimer's disease prediction. *Scientific reports*, 13(1), 450. <https://doi.org/10.1038/s41598-023-27551-1>
147. Gunter, N. B., Gebre, R. K., Graff-Radford, J., Heckman, M. G., Jack, C. R., Jr., Lowe, V. J., Knopman, D. S., Petersen, R. C., Ross, O. A., Vemuri, P., Ramanan, V. K., & Alzheimer's Disease Neuroimaging Initiative. (2024). Machine Learning Models of Polygenic Risk for Enhanced Prediction of Alzheimer Disease Endophenotypes. *Neurology: Genetics*, 10(1), e200120. <https://doi.org/10.1212/nxg.0000000000200120>
148. Jalalinajafabadi, F., Stadler, M. A., Dand, N., Jadon, D. R., Soomro, M., Ho, P., et al. (2021). Application of information theoretic feature selection and machine learning methods for the development of genetic risk prediction models. *Scientific reports*, 11(1), 23335. <https://doi.org/10.1038/s41598-021-00854-x>

149. Kour, B., Shukla, N., Bhargava, H., Sharma, D. K., Sharma, A., Singh, A., Valadi, J., Sadasukhi, T. C., Vuree, S., & Suravajhala, P. (2023). Identification of Plausible Candidates in Prostate Cancer Using Integrated Machine Learning Approaches. *Current Genomics*, 24(5), 287-306. <https://doi.org/10.2174/0113892029240239231109082805>
150. Goldstein, B. A., Návar, A. M., Pencina, M., & Ioannidis, J. P. A. (2016). Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review. *Journal of the American Medical Informatics Association: JAMIA*, 24(1), 198. <https://doi.org/10.1093/jamia/ocw042>
151. Dai, M., Li, X., Zhāng, Q., Liang, T., Huang, X., & Fu, Q. (2024). Health research in the era of artificial intelligence: Advances in gene-editing study. *Medicine plus*, 1(2), 100027. <https://doi.org/10.1016/j.medp.2024.100027>
152. Callaway, E. (2024). ‘ChatGPT for CRISPR’ creates new gene-editing tools. *Nature*, 629(8011), 272-272. <https://doi.org/10.1038/d41586-024-01243-w>
153. Qu, Y., Huang, K., Yin, M., Zhan, K., Liu, D., Yin, D., Cousins, H. C., Johnson, W. A., Wang, X., Shah, M., Altman, R. B., Zhou, D., Wang, M., & Cong, L. (2026). CRISPR-GPT for agentic automation of gene-editing experiments. *Nature Biomedical Engineering*, 10(2), 245-258. <https://doi.org/10.1038/s41551-025-01463-z>
154. Lei, T., Wang, Y., Zhang, Y., Yang, Y., Cao, J., Huang, J., Chen, J., Chen, H., Zhang, J., Wang, L., Xu, X., Gale, R. P., & Wang, L. (2024). Leveraging CRISPR gene editing technology to optimize the efficacy, safety and accessibility of CAR T-cell therapy. *Leukemia*, 38(12), 2517-2543. <https://doi.org/10.1038/s41375-024-02444-y>
155. Musunuru, K., Grandinette, S., Wang, X., Hudson, T. R., Briseno, K., Berry, A. M., et al. (2025). Patient-Specific In Vivo Gene Editing to Treat a Rare Genetic Disease. *New England Journal of Medicine*, 392(22), 2235-2243. <https://doi.org/10.1056/NEJMoa2504747>
156. Chuai, G., Ma, H., Yan, J., Chen, M., Hong, N., Xue, D., et al. (2018). DeepCRISPR: optimized CRISPR guide RNA design by deep learning. *Genome Biology*, 19(1), 80. <https://doi.org/10.1186/s13059-018-1459-4>
157. Jasieniecka, A., & Domingues, I. (2025). CRISPR-Cas9 and Its Bioinformatics Tools: A Systematic Review. *Current Issues in Molecular Biology*, 47(5), 307. <https://doi.org/10.3390/cimb47050307>

158. Wang, D., Zhang, C., Wang, B., Li, B., Wang, Q., Liu, D., et al. (2019). Optimized CRISPR guide RNA design for two high-fidelity Cas9 variants by deep learning. *Nature Communications*, 10(1), 4284. <https://doi.org/10.1038/s41467-019-12281-8>
159. Bao, C., & Liu, F. (2024, November). Prediction of CRISPR on-target effects via deep learning. In *Proceedings of the 1st workshop on NLP for science*, 9-15. <https://doi.org/10.18653/v1/2024.nlp4science-1.2>
160. Graham, J. P., Zhang, Y., He, L., & Gonzalez-Fernandez, T. (2024). CRISPR-GEM: A Novel Machine Learning Model for CRISPR Genetic Target Discovery and Evaluation. *ACS Synthetic Biology*, 13(10), 3413. <https://doi.org/10.1021/acssynbio.4c00473>
161. Lee, J. S., Abdin, O., & Kim, P. M. (2025). Language models for protein design. *Current Opinion in Structural Biology*, 92, 103027. <https://doi.org/10.1016/j.sbi.2025.103027>
162. Du, W., Zhao, L., Diao, K., Zheng, Y., Yang, Q., Zhu, Z., Zhu, X., & Tang, D. (2025). A versatile CRISPR/Cas9 system off-target prediction tool using language model. *Communications Biology*, 8(1), 882. <https://doi.org/10.1038/s42003-025-08275-6>
163. Cui, Z., Lin, L., Zong, Y., Chen, Y., & Wang, S. (2024). Precision gene editing using deep learning: A case study of the CRISPR-Cas9 editor. *Applied and Computational Engineering*, 64(1), 133. <https://doi.org/10.54254/2755-2721/64/20241357>
164. Wu, Y.-C., Tuo, N., Shi, G., Li, K. L., Mu, S., & Li, Y. (2025). The Evolving Role of Artificial Intelligence in Medical Genetics: Advancing Healthcare, Research, and Biosafety Management. *Genes*, 17(1), 6. <https://doi.org/10.3390/genes17010006>
165. Sarabi, P. A., Shabanpouremam, M., Eghtedari, A. R., Barat, M., Moshiri, B., et al. (2024). AI-Based solutions for current challenges in regenerative medicine. *European Journal of Pharmacology*, 984, 177067. <https://doi.org/10.1016/j.ejphar.2024.177067>
166. Gong, D., Ben-Akiva, E., Singh, A., Yamagata, H., Est-Witte, S., Shade, J. K., et al. (2022). Machine learning guided structure function predictions enable in silico nanoparticle screening for polymeric gene delivery. *Acta biomaterialia*, 154, 349-358. <https://doi.org/10.1016/j.actbio.2022.09.072>
167. Singh, K., Jain, D., Sethi, P., Gupta, J. K., Dubey, A., Noman, A. A., et al. (2025). Advances in viral vector-based delivery systems for gene therapy: a comprehensive review. *3 Biotech*, 15(7), 196. <https://doi.org/10.1007/s13205-025-04366-7>

168. Gao, Z. (2024). Strategies for enhanced gene delivery to the central nervous system. *Nanoscale Advances*, 6(12), 3009. <https://doi.org/10.1039/d3na01125a>
169. Dagar, G., Gupta, A., Masoodi, T., Nisar, S., Merhi, M., Hashem, S., et al. (2023). Harnessing the potential of CAR-T cell therapy: progress, challenges, and future directions in hematological and solid tumor treatments. *Journal of Translational Medicine*, 21(1). <https://doi.org/10.1186/s12967-023-04292-3>
170. Anurogo, D., Luthfiana, D., Anripa, N., Fauziah, A. I., Soleha, M., Rahmah, L., et al. (2024). The Art of Bioimmunogenomics (BIGs) 5.0 in CAR-T Cell Therapy for Lymphoma Management. *Advanced Pharmaceutical Bulletin*, 14(2), 314. <https://doi.org/10.34172/apb.2024.034>
171. Boretti, A. (2024). The transformative potential of AI-driven CRISPR-Cas9 genome editing to enhance CAR T-cell therapy. *Computers in Biology and Medicine*, 182, 109137. <https://doi.org/10.1016/j.combiomed.2024.109137>
172. Cui, J. J., Zhang, D., & Wang, G. (2025). Combinatorial application of artificial intelligence and CRISPR/Cas9 on the next-generation CAR-T immunotherapy. *AIMS Molecular Science*, 12(3), 292. <https://doi.org/10.3934/molsci.2025018>
173. Mergen, M., Abele, D., Koleci, N., Fernandez, A. S., Sugden, M., Holzleitner, N., et al. (2024). Engineering de novo binder CAR-T cell therapies with generative AI. *bioRxiv*, 2024-11. <https://doi.org/10.1101/2024.11.25.625151>
174. Baena, J. C., Victoria, J., Toro-Pedroza, A., Aragón, C. C., Ortiz-Guzman, J., García-Robledo, J. E., et al. (2025). Smart CAR-T Nanosymbionts: archetypes and proto-models. *Frontiers in Immunology*, 16. <https://doi.org/10.3389/fimmu.2025.1635159>
175. Karsten, H., Matrisch, L., Cichutek, S., Fiedler, W., Alsdorf, W., & Block, A. (2023). Broadening the horizon: potential applications of CAR-T cells beyond current indications. *Frontiers in Immunology*, 14. <https://doi.org/10.3389/fimmu.2023.1285406>
176. Bäckel, N., Hort, S., Kis, T., Nettleton, D., Egan, J. R., Jacobs, J. J. L., et al. (2023). Elaborating the potential of Artificial Intelligence in automated CAR-T cell manufacturing. *Frontiers in Molecular Medicine*, 3. <https://doi.org/10.3389/fmmed.2023.1250508>
177. Makkar, S. K. (2025). Advances in RNA-based therapeutics: current breakthroughs, clinical translation, and future perspectives. *Frontiers in Genetics*, 16. <https://doi.org/10.3389/fgene.2025.1675209>

178. Gu, R., Si, D., Lei, D., Xie, X., Li, Y., & Wen, H. (2026). Accelerate the Highly Efficient Development of mRNA Vaccines Through Advanced Computational Methods. *MedComm*, 7(2). <https://doi.org/10.1002/mco2.70612>
179. Zhang, Y., Li, J., Wang, Z., Kuang, Y., Li, S., & Wang, X. (2025). Precision-engineered mRNA vaccines: antigen design, structural optimization, and programmable delivery for emerging pathogens. *Animal Diseases*, 5(1). <https://doi.org/10.1186/s44149-025-00186-7>
180. Zhang, H., Liu, H., Xu, Y., Huang, H., Liu, Y., Wang, J., et al. (2025). Deep generative models design mRNA sequences with enhanced translational capacity and stability. *Science*, 390(6773). <https://doi.org/10.1126/science.adr8470>
181. Soon, H. I., Abdullah, A. A., Nishizaki, H., Mashor, M. Y., Kamarudin, L. M., Mohamed-Hussein, Z., et al. (2024). Optimizing hybrid neural networks for precise COVID-19 mRNA vaccine degradation prediction. *International journal of advanced and applied sciences*, 11(7), 87-100. <https://doi.org/10.21833/ijaas.2024.07.011>
182. Maharjan, R., Lee, J., Lee, K., Han, H., Kim, K. H., & Jeong, S. H. (2023). Recent trends and perspectives of artificial intelligence-based machine learning from discovery to manufacturing in biopharmaceutical industry. *Journal of Pharmaceutical Investigation*, 53(6), 803. <https://doi.org/10.1007/s40005-023-00637-8>
183. Zhao, Y., Oono, K., Takizawa, H., & Kotera, M. (2024). GenerRNA: A generative pre-trained language model for de novo RNA design. *PLoS ONE*, 19(10). <https://doi.org/10.1371/journal.pone.0310814>
184. Anand, R., Joshi, C. K., Morehead, A., Jamasb, A. R., Harris, C. B., Mathis, S. V., et al. (2024). RNA-FrameFlow: Flow matching for de novo 3D RNA Backbone Design. *arXiv*. <https://doi.org/10.48550/arXiv.2406.13839>
185. Kozlov, D., Rodimova, S., Филатов, П., Mozherov, A., Timashev, P., Zyuzin, M. V., & Kuznetsova, D. (2025). Genomic medicine in hepatology: mechanisms and liver treatment strategies. *Molecular Medicine*, 31(1). <https://doi.org/10.1186/s10020-025-01358-4>
186. Lee, M. (2023). Machine learning for small interfering RNAs: a concise review of recent developments. *Frontiers in Genetics*, 14. <https://doi.org/10.3389/fgene.2023.1226336>

187. Liu, B., Yuan, Y., Pan, X., Shen, H., & Jin, C. (2024). AttSiOff: a self-attention-based approach on siRNA design with inhibition and off-target effect prediction. *Med-X*, 2(1). <https://doi.org/10.1007/s44258-024-00019-1>
188. Hwang, H., Jeon, H., Yeo, N., & Baek, D. (2024). Big data and deep learning for RNA biology. *Experimental & Molecular Medicine*, 56(6), 1293. <https://doi.org/10.1038/s12276-024-01243-w>
189. Hawkins-Hooker, A., Visonà, G., Narendra, T., Rojas-Carulla, M., Schölkopf, B., & Schweikert, G. (2023). Getting personal with epigenetics: towards individual-specific epigenomic imputation with machine learning. *Nature Communications*, 14(1). <https://doi.org/10.1038/s41467-023-40211-2>
190. Srivastava, R. (2026). Artificial intelligence meets epigenetics: a new frontier in precision oncology. *Epigenetics Reports*, 4(1). <https://doi.org/10.1080/28361512.2026.2614541>
191. Maqsood, K., Hagrass, H., & Zabet, N. R. (2024). An overview of artificial intelligence in the field of genomics. *Discover Artificial Intelligence*, 4(1), 9. <https://doi.org/10.1007/s44163-024-00103-w>
192. Gillies, C. E., Mbatchou, J., Habegger, L., Kessler, M. D., Bao, S., Balasubramanian, S., et al. (2025). Variant Classification Using Proteomics-Informed Large Language Models Increases Power of Rare Variant Association Studies and Enhances Target Discovery. *Genetic Epidemiology*, 49(8), e70023. <https://doi.org/10.1002/gepi.70023>
193. Choon, Y. W., Choon, Y. F., Nasarudin, N. A., Jasmi, F. A., Remli, M. A., Alkayali, M. H., et al. (2024). Artificial intelligence and database for NGS-based diagnosis in rare disease. *Frontiers in Genetics*, 14. <https://doi.org/10.3389/fgene.2023.1258083>
194. Prodduturi, V. R. (2025). Machine learning in genomic diagnostics for precision medicine. *International Journal of Science and Research Archive*, 14(1), 1758. <https://doi.org/10.30574/ijrsra.2025.14.1.0282>
195. Taherdoost, H., & Ghofrani, A. (2024). AI's role in revolutionizing personalized medicine by reshaping pharmacogenomics and drug therapy. *Intelligent Pharmacy*, 2(5), 643. <https://doi.org/10.1016/j.ipha.2024.08.005>
196. Piriyapongsa, J., Sukritha, C., Kaewprommal, P., Intarat, C., Triparn, K., Phornsiricharoenphant, K., et al. (2021). PharmVIP: A Web-Based Tool for Pharmacogenomic

Variant Analysis and Interpretation. *Journal of Personalized Medicine*, 11(11), 1230.

<https://doi.org/10.3390/jpm11111230>

197. Tippenhauer, K., Philips, M., Largiadèr, C. R., & Sariyar, M. (2023). Using the PharmCAT tool for Pharmacogenetic clinical decision support. *Briefings in bioinformatics*, 25(1), bbad452.

<https://doi.org/10.1093/bib/bbad452>

198. Li, B., Sangkuhl, K., Keat, K., Whaley, R., Woon, M., Verma, S. S., et al. (2022). How to Run the Pharmacogenomics Clinical Annotation Tool (PharmCAT). *Clinical Pharmacology & Therapeutics*, 11(5), 1036. <https://doi.org/10.1002/cpt.2790>

199. Pearl, D., Intriligator, J., & Liu, W. J. (2025). Seamless Integration: The Evolution, Design and Future Impact of Wearable Technology. *arXiv*. <https://doi.org/10.48550/arxiv.2502.05797>

200. Radanliev, P. (2025). Privacy, ethics, transparency, and accountability in AI systems for wearable devices. *Frontiers in Digital Health*, 7, 1431246.

<https://doi.org/10.3389/fdgth.2025.1431246>

201. Padhan, S., Mohapatra, A., Kumar, S., & Agrawal, S. (2023). Artificial Intelligence (AI) and Robotics in Elderly Healthcare: Enabling Independence and Quality of Life. *Cureus*, 15(8), e42905. <https://doi.org/10.7759/cureus.42905>

202. Adeghe, E. P., Okolo, C. A., & Ojeyinka, O. T. (2024). A review of wearable technology in healthcare: Monitoring patient health and enhancing outcomes. *Open Access Research Journal of Multidisciplinary Studies*, 7(1), 142. <https://doi.org/10.53022/oarjms.2024.7.1.0019>

203. Serrano, R. R., Troughton, J. G., Mirkhani, N., Martínez, N., Mariello, M., Tsigarides, J., et al. (2025). From Neural Sensing to Stimulation: An Interdisciplinary Roadmap for Neurotechnology. *arXiv*, <https://doi.org/10.48550/arxiv.2510.07116>

204. González-González, M. A., Conde, S. V., Latorre, R., Thébault, S., Pratelli, M., Spitzer, N. C., et al. (2024). Bioelectronic Medicine: a multidisciplinary roadmap from biophysics to precision therapies. *Frontiers in Integrative Neuroscience*, 18, 1321872.

<https://doi.org/10.3389/fnint.2024.1321872>

205. Trigili, E., & Hirche, S. (2023). Wearable robotics in the rehabilitation continuum of care: assessment, treatment and home assistance. *Frontiers in Neurorobotics*, 17, 1305786.

<https://doi.org/10.3389/fnbot.2023.1305786>

206. Mobedi, E., Solak, G., & Ajoudani, A. (2025). A Framework for Adaptive Load Redistribution in Human-Exoskeleton-Cobot Systems. *IEEE Robotics and Automation Letters*, 10(6), 5927-5934. <https://doi.org/10.48550/arxiv.2504.10066>
207. Saeidi, H., Opfermann, J. D., Kam, M., Wei, S., Leonard, S., Hsieh, M. H., Kang, J. U., & Krieger, A. (2022). Autonomous robotic laparoscopic surgery for intestinal anastomosis. *Science robotics*, 7(62), eabj2908. <https://doi.org/10.1126/scirobotics.abj2908>
208. Cao, Y., Xu, B., Li, B., & Fu, H. (2024). Advanced Design of Soft Robots with Artificial Intelligence. *Nano-micro letters*, 16(1), 214. <https://doi.org/10.1007/s40820-024-01423-3>
209. Sokol, K., Fackler, J. C., & Vogt, J. E. (2025). Artificial intelligence should genuinely support clinical reasoning and decision making to bridge the translational gap. *npj Digital Medicine*, 8(1), 345. <https://doi.org/10.1038/s41746-025-01725-9>
210. Angelov, P., Soares, E., Jiang, R., Arnold, N. I., & Atkinson, P. M. (2021). Explainable artificial intelligence: an analytical review. *Wiley Interdisciplinary Reviews Data Mining and Knowledge Discovery*, 11(5). <https://doi.org/10.1002/widm.1424>
211. Mumtaz, H., Saqib, M., Jabeen, S., Muneeb, M., Mughal, W., Sohail, H., et al. (2023). Exploring alternative approaches to precision medicine through genomics and artificial intelligence – a systematic review. *Frontiers in Medicine*, 10. <https://doi.org/10.3389/fmed.2023.1227168>
212. O'Connor, O., & McVeigh, T. (2025). Increasing use of artificial intelligence in genomic medicine for cancer care- the promise and potential pitfalls. *BJC Reports*, 3(1), 20. <https://doi.org/10.1038/s44276-025-00135-4>
213. Carini, C., & Seyhan, A. A. (2024). Tribulations and future opportunities for artificial intelligence in precision medicine. *Journal of Translational Medicine*, 22(1). <https://doi.org/10.1186/s12967-024-05067-0>
214. Pant, L. P., Mukaddim, A. A., Rahman, M., Sayeed, A. A., Hossain, Md. S., Khan, M. T., et al. (2024). Genomic predictors of drug sensitivity in cancer: Integrating genomic data for personalized medicine in the USA. *Computer Science & IT Research Journal*, 5(12), 2682. <https://doi.org/10.51594/csitrj.v5i12.1760>
215. Abbas, Q., Jeong, W.-M., & Lee, S. W. (2025). Explainable AI in Clinical Decision Support Systems: A Meta-Analysis of Methods, Applications, and Usability Challenges. *Healthcare*, 13(17), 2154. <https://doi.org/10.3390/healthcare13172154>

216. Noor, A. A., Manzoor, A., Mazhar Qureshi, M. D., Qureshi, M. A., & Rashwan, W. (2025). Unveiling explainable AI in healthcare: Current trends, challenges, and future directions. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 15(2), e70018. <https://doi.org/10.1002/widm.70018>
217. Allen, B. (2024). The Promise of Explainable AI in Digital Health for Precision Medicine: A Systematic Review. *Journal of Personalized Medicine*, 14(3), 277. <https://doi.org/10.3390/jpm14030277>
218. Dai, J., Xu, H., Chen, T., Huang, T., Liang, W., Zhang, R., et al. (2025). Artificial intelligence for medicine 2025: Navigating the endless frontier. *The Innovation Medicine*, 3(1), 100120. <https://doi.org/10.59717/j.xinn-med.2025.100120>
219. Poplin, R., Chang, P.-C., Alexander, D. H., Schwartz, S., Colthurst, T., Ku, A., et al. (2018). A universal SNP and small-indel variant caller using deep neural networks. *Nature Biotechnology*, 36(10), 983. <https://doi.org/10.1038/nbt.4235>
220. Brambrink, L., Kolesnikov, A., Goel, S., Nattestad, M., Yun, T., Baid, G., et al. (2021). DeepTrio: variant calling in families using deep learning. *BioRxiv*, 2021-04. <https://doi.org/10.1101/2021.04.05.438434>
221. Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., Angel, G. del, Levy-Moonshine, A., et al. (2013). From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. *Current Protocols in Bioinformatics*, 43(1). <https://doi.org/10.1002/0471250953.bi1110s43>
222. Krøigård AB, Thomassen M, Lænkholm AV, Kruse TA, Larsen MJ. Evaluation of nine somatic variant callers for detection of somatic mutations in exome and targeted deep sequencing data. *PLoS One*. 2016;11(3):e0151664. <https://doi.org/10.1371/journal.pone.0151664>
223. Abadi, S., Yan, W. X., Amar, D., & Mayrose, I. (2017). A machine learning approach for predicting CRISPR-Cas9 cleavage efficiencies and patterns underlying its mechanism of action. *PLoS Computational Biology*, 13(10). <https://doi.org/10.1371/journal.pcbi.1005807>
224. Qu, Y., Huang, K., Yin, M., Zhan, K., Liu, D., Yin, D., et al. (2025). CRISPR-GPT for agentic automation of gene-editing experiments. *Nature Biomedical Engineering*, 10(2), 245. <https://doi.org/10.1038/s41551-025-01463-z>

225. Ruffolo, J. A., Nayfach, S., Gallagher, J. P., Bhatnagar, A., Beazer, J., Hussain, R., et al. (2025). Design of highly functional genome editors by modelling CRISPR–Cas sequences. *Nature*, 645(8080), 518. <https://doi.org/10.1038/s41586-025-09298-z>
226. Jacobsen, J. O. B., Kelly, C., Cipriani, V., Consortium, G. E. R., Mungall, C., Reese, J., Daniš, D., & Robinson, P. N. (2022). Phenotype-driven approaches to enhance variant prioritization and diagnosis of rare disease. *Human Mutation*, 43(8), 1071. <https://doi.org/10.1002/humu.24380>
227. Sherif, F. F., Tawfik, N., Mousa, D., Abdallah, M. S., & Cho, Y. I. (2024). Automated Multi-Class Facial Syndrome Classification Using Transfer Learning Techniques. *Bioengineering*, 11(8), 827. <https://doi.org/10.3390/bioengineering11080827>
228. Pun, F. W., Ozerov, I. V., & Zhavoronkov, A. (2023). AI-powered therapeutic target discovery. *Trends in pharmacological sciences*, 44(9), 561-572 <https://doi.org/10.1016/j.tips.2023.06.010>
229. Calvino, G., Farro, J., Zampatti, S., Peconi, C., Megalizzi, D., Trastulli, G., et al. (2025). From Genomics to AI: Revolutionizing Precision Medicine in Oncology. *Applied Sciences*, 15(12), 6578. <https://doi.org/10.3390/app15126578>
230. Nievas, M., Basu, A., Wang, Y., & Singh, H. (2024). Distilling large language models for matching patients to clinical trials. *Journal of the American Medical Informatics Association*, 31(9), 1953-1963. <https://doi.org/10.1093/jamia/ocae073>
231. Kaskovich, S., Wyatt, K. D., Oliwa, T., Graglia, L., Furner, B., Lee, J., et al. (2023). Automated Matching of Patients to Clinical Trials: A Patient-Centric Natural Language Processing Approach for Pediatric Leukemia. *JCO Clinical Cancer Informatics*, 7, e2300009. <https://doi.org/10.1200/cci.23.00009>
232. Flora, D. B. (2025). The Top 10 AI in Precision Oncology Stories of 2024: A Year of Breakthrough Moments. *AI in Precision Oncology*, 2(1), 1-5. <https://doi.org/10.1089/aipo.2025.ed>
233. Nishan, M. N. H. (2025). AI-powered drug discovery for neglected diseases: accelerating public health solutions in the developing world. *Journal of Global Health*, 15, 03002. <https://doi.org/10.7189/jogh.15.03002>
234. Bessadok, A., & Grisoni, F. (2024). An explainable foundation model for drug repurposing. *Nature medicine*, 30(12), 3422-3423. <https://doi.org/10.1038/s41591-024-03333-8>

235. Yin, Q., Fan, R., Cao, X., Liu, Q., Jiang, R., & Zeng, W. (2023). DeepDrug: A general graph-based deep learning framework for drug-drug interactions and drug-target interactions prediction. *Quantitative Biology*, 11(3), 260-274. <https://doi.org/10.15302/j-qb-022-0320>
236. Li, V. O. K., Yang, H., Kaistha, T., Zhang, Q., Downey, J., Gozes, I., et al. (2025). DeepDrug as an expert guided and AI driven drug repurposing methodology for selecting the lead combination of drugs for Alzheimer's disease. *Scientific Reports*, 15(1), 2093. <https://doi.org/10.1038/s41598-025-85947-7>
237. Ren, F., Ding, X., Zheng, M., Korzinkin, M., Cai, X., Zhu, W., Mantsyzov, A. B., et al. (2023). AlphaFold accelerates artificial intelligence powered drug discovery: efficient discovery of a novel CDK20 small molecule inhibitor. *Chemical science*, 14(6), 1443-1452. <https://doi.org/10.1039/d2sc05709c>
238. Romero, R. (2024). Integration of Genetic Algorithms and Deep Learning for the Generation and Bioactivity Prediction of Novel Tyrosine Kinase Inhibitors. *arXiv*, <https://doi.org/10.48550/arxiv.2408.07155>
239. Laise, P., Bosker, G., Califano, A., & Alvarez, M. J. (2022). A Patient-to-Model-to-Patient (PMP) Cancer Drug Discovery Protocol for Identifying and Validating Therapeutic Agents Targeting Tumor Regulatory Architecture. *Current Protocols*, 2(9). <https://doi.org/10.1002/cpz1.544>
240. Ren, F., Aliper, A., Chen, J. et al. (2025). A small-molecule TNIK inhibitor targets fibrosis in preclinical and clinical models. *NatBiotechnol*, 43, 63–75. <https://doi.org/10.1038/s41587-024-02143-0>
241. Nagaiah, B. (2025). Use of Smart contracts and AI agents in clinical trials. *International Journal on Science and Technology*, 16(4). <https://doi.org/10.71097/ijst.v16.i4.8751>
242. Ren, F., Aliper, A., Chen, J., Zhao, H., Rao, S., Kuppe, C., et al. (2024). A small-molecule TNIK inhibitor targets fibrosis in preclinical and clinical models. *Nature Biotechnology*, 43(1), 63. <https://doi.org/10.1038/s41587-024-02143-0>
243. Azam, M. S., Chen, Y., Arowolo, M. O., Liu, H., Popescu, M., & Xu, D. (2024). A comprehensive evaluation of large language models in mining gene relations and pathway knowledge. *Quantitative Biology*, 12(4), 360. <https://doi.org/10.1002/qub2.57>

244. Luo, R., Sun, L., Xia, Y., Qin, T., Zhang, S., Poon, H., & Liu, T. (2022). BioGPT: generative pre-trained transformer for biomedical text generation and mining. *Briefings in Bioinformatics*, 23(6). <https://doi.org/10.1093/bib/bbac409>
245. Sherkatghanad, Z., Abdar, M., Charlier, J., & Makarenkov, V. (2023). Using traditional machine learning and deep learning methods for on- and off-target prediction in CRISPR/Cas9: A review. *Briefings in Bioinformatics*, 24(3), bbad131. <https://doi.org/10.1093/bib/bbad131>
246. Xu, T., Xu, M., Zhang, Q. et al. Leveraging viral genome sequences and machine learning models for identification of potentially selective antiviral agents. *Commun Chem*, 8, 189 (2025). <https://doi.org/10.1038/s42004-025-01583-2>
247. Kearney, E., Wojcik, A., & Babu, D. (2019). Artificial intelligence in genetic services delivery: Utopia or apocalypse? *Journal of genetic counseling*, 29(1), 8-17. <https://doi.org/10.1002/jgc4.1192>
248. Ocaña, A., Pandiella, A., Privat, C., Bravo, I., Luengo-Oroz, M., Amir, E., et al. (2025). Integrating artificial intelligence in drug discovery and early drug development: a transformative approach. *Biomarker research*, 13(1), 45. <https://doi.org/10.1186/s40364-025-00758-2>
249. Shin, D. W., Cho, J., Cho, S., Youn, Y., & Hwang, J.-H. (2025). Development of a serum protein biomarker panel for the diagnosis of pancreatic ductal adenocarcinoma using a machine learning approach. *Scientific Reports*, 15(1), 35659. <https://doi.org/10.1038/s41598-025-19631-1>
250. Zardab, M., Balarajah, V., Banerjee, A., Stasinos, K., Sâad, A., Imrali, A., et al. (2022). Differentiating Ductal Adenocarcinoma of the Pancreas from Benign Conditions Using Routine Health Records: A Prospective Case-Control Study. *Cancers*, 15(1), 280. <https://doi.org/10.3390/cancers15010280>
251. Garrido, N. J., González-Martínez, F., Torres, A. M., Blasco-Segura, P., Losada, S., Plaza, A., et al. (2025). Role of Artificial Intelligence in Identifying Vital Biomarkers with Greater Precision in Emergency Departments During Emerging Pandemics. *International Journal of Molecular Sciences*, 26(2), 722. <https://doi.org/10.3390/ijms26020722>
252. Alemu, R., Sharew, N. T., Arsano, Y. Y., Ahmed, M., Tekola-Ayele, F., Mersha, T. B., & Amare, A. T. (2025). Multi-omics approaches for understanding gene-environment interactions in noncommunicable diseases: techniques, translation, and equity issues. *Human Genomics*, 19(1), 8. <https://doi.org/10.1186/s40246-025-00718-9>

253. Silva, P. P., Gaudillo, J., Vilela, J., Roxas-Villanueva, R. M. L., Tiangco, B., Domingo, M., et al. (2022). A machine learning-based SNP-set analysis approach for identifying disease-associated susceptibility loci. *Scientific Reports*, 12(1), 15817. <https://doi.org/10.1038/s41598-022-19708-1>
254. Cox, D. G., Oh, J., Keasling, A. W., Colson, K., & Hamann, M. T. (2014). The utility of metabolomics in natural product and biomarker characterization. *Biochimica et Biophysica Acta (BBA)-General Subjects*, 1840(12), 3460-3474. <https://doi.org/10.1016/j.bbagen.2014.08.007>
255. Mukherjee, A., Abraham, S., Singh, A., Balaji, S., & Mukunthan, K. S. (2024). From Data to Cure: A Comprehensive Exploration of Multi-omics Data Analysis for Targeted Therapies. *Molecular biotechnology*, 67(4), 1269-1289. <https://doi.org/10.1007/s12033-024-01133-6>
256. Wanders, R. J. A., & Wevers, R. A. (2024). Strategies in Biochemical Genetics. In *Laboratory Guide to the Methods in Biochemical Genetics*, 1-14. https://doi.org/10.1007/978-3-031-58819-8_1
257. Lawrence, S. R., & Shah, K. M. (2024). Prospects and Current Challenges of Extracellular Vesicle-Based Biomarkers in Cancer. *Biology*, 13(9), 694. <https://doi.org/10.3390/biology13090694>
258. Філіпцова, О. В., Набока, О. И., Kryskiv, O. S., Zakhariev, A., Мартынюк, Т. В., Хохленкова, Н. В., et al. (2023). Problems of collecting genetic history, interpretation of pedigrees with common and rare dermatological, skeletal, dental and other peculiarities that affect the appearance and health of a human. *ScienceRise Biological Science*, 4. <https://doi.org/10.15587/2519-8025.2023.298576>
259. Li, W., Murray, M. F., & Giovanni, M. A. (2019). Obtaining a Genetic Family History Using Computer-Based Tools. *Current protocols in human genetics*, 100(1), e72. <https://doi.org/10.1002/cphg.72>
260. Pedersen, E. M., Steinbach, J., Pedersen, C. B., Schork, A. J., Krebs, M. D., Vilhjálmsson, B. J., & Privé, F. (2025). Automatic detection of n-degree family members. *Frontiers in Genetics*, 16, 1708315. <https://doi.org/10.3389/fgene.2025.1708315>
261. Conte, L., Rizzo, E., Grassi, T., Bagordo, F., Matteis, E. D., & Nunzio, G. D. (2024). Artificial Intelligence Techniques and Pedigree Charts in Oncogenetics: Towards an Experimental Multioutput Software System for Digitization and Risk Prediction. *Computation*, 12(3), 47. <https://doi.org/10.3390/computation12030047>

262. Popli, D., Peyrégne, S., & Peter, B. M. (2023). KIN: a method to infer relatedness from low-coverage ancient DNA. *Genome Biology*, 24(1), 10. <https://doi.org/10.1186/s13059-023-02847-7>
263. Mostad, P., Tillmar, A., & Kling, D. (2023). Improved computations for relationship inference using low-coverage sequencing data. *BMC Bioinformatics*, 24(1), 90. <https://doi.org/10.1186/s12859-023-05217-z>
264. Huang, M., Liu, M., Li, H., King, J. L., Smuts, A., Budowle, B., & Ge, J. (2022). A machine learning approach for missing persons cases with high genotyping errors. *Frontiers in Genetics*, 13, 971242. <https://doi.org/10.3389/fgene.2022.971242>
265. Pudjihartono, N., Fadason, T., Kempa-Liehr, A. W., & O'Sullivan, J. M. (2022). A Review of Feature Selection Methods for Machine Learning-Based Disease Risk Prediction. *Frontiers in Bioinformatics*, 2, 927312. <https://doi.org/10.3389/fbinf.2022.927312>
266. Ramaswamy, M., Senthilnathan, S., Saravanan, A. R. I., Snehal, M., & Siva, K. (2025). Leveraging machine learning for risk assessment and prediction of inheriting genetic disorders. *Computational and Structural Biotechnology Reports*, 2, 100038. <https://doi.org/10.1016/j.csbr.2025.100038>
267. Wand, H., Kalia, S. S., Helm, B. M., Suckiel, S. A., Brockman, D., Vriesen, N., et al. (2023). Clinical genetic counseling and translation considerations for polygenic scores in personalized risk assessments: A Practice Resource from the National Society of Genetic Counselors. *Journal of Genetic Counseling*, 32(3), 558. <https://doi.org/10.1002/jgc4.1668>
268. Rahman, A. U., Nasir, M. U., Gollapalli, M., Alsaif, S. A., Almadhor, A. S., Mehmood, S., et al. (2022). IoMT-Based Mitochondrial and Multifactorial Genetic Inheritance Disorder Prediction Using Machine Learning. *Computational intelligence and neuroscience*, 2022(1), 2650742. <https://doi.org/10.1155/2022/2650742>
269. Ziadi, C., Demyda-Peyrás, S., Valera, M., Perdomo-González, D. I., Laseca, N., Terreros, A. R.-S. de los, Encina, A., et al. (2025). Comparative Analysis of Genomic and Pedigree-Based Approaches for Genetic Evaluation of Morphological Traits in Pura Raza Española Horses. *Genes*, 16(2), 131. <https://doi.org/10.3390/genes16020131>
270. Nani, J. P., Jansen, G. B., Zaabza, H. B., & VanRaden, P. M. (2025). Improved computation of genomic and pedigree inbreeding and relationships accounting for the X chromosome. *Journal of Dairy Science*, 108(6), 6190. <https://doi.org/10.3168/jds.2024-26056>

271. Guardado, M., Pérez, C., Campana, S., Rojas, B. C., Magaña, J., Jackson, S., et al. (2025). py_ped_sim: a flexible forward pedigree and genetic simulator for complex family pedigree analysis. BMC Bioinformatics, 26(1), 122. <https://doi.org/10.1186/s12859-025-06142-z>
272. Basharat, S., Smith, A., Darvesh, N., & Rader, T. (2023). 2023 Watch List: Top 10 Precision Medicine Technologies and Issues. Canadian Journal of Health Technologies, 3(3). <https://doi.org/10.51731/cjht.2022.590>
273. Suckiel, S. A., Kelly, N., Odgis, J. A., Gallagher, K. M., Sebastin, M., Bonini, K. E., et al. (2023). The NYCKidSeq randomized controlled trial: Impact of GUÍA digitally enhanced genetic counseling in racially and ethnically diverse families. medRxiv. <https://doi.org/10.1101/2023.07.05.23292193>
274. Mighton, C., Jan, A., Lee, L. et al. (2026). Impact of a digital platform on genetic counselling encounters in the screening context. European Journal of Human Genetics, 34(5), 715-722. <https://doi.org/10.1038/s41431-026-02029-6>
275. Cazzaniga, A., Plebani, M., & Crimi, M. (2022). Genome access and other web-based IT solutions: Genetic counseling in the digital era. Frontiers in Public Health, 10, 1035316. <https://doi.org/10.3389/fpubh.2022.1035316>
276. Biesecker, B. B. (2019). Genetic Counseling and the Central Tenets of Practice. Cold Spring Harbor Perspectives in Medicine, 10(3), a038968. <https://doi.org/10.1101/cshperspect.a038968>
277. Wong, K. K., Pan, V., & Caleshu, C. (2025). Generative AI: a partner for genetic counseling? Trends in Genetics, 41(9), 725. <https://doi.org/10.1016/j.tig.2025.07.002>
278. Webster, E., Ahsan, M. D., Perez, L., Levi, S. R., Thomas, C., Christos, P. J., et al. (2023). Chatbot Artificial Intelligence for Genetic Cancer Risk Assessment and Counseling: A Systematic Review and Meta-Analysis. JCO Clinical Cancer Informatics, 7(7). <https://doi.org/10.1200/cci.23.00123>
279. Ballard, L., Band, R., & Lucassen, A. (2023). Interventions to support patients with sharing genetic test results with at-risk relatives: a synthesis without meta-analysis (SWiM). European Journal of Human Genetics, 31(9), 988. <https://doi.org/10.1038/s41431-023-01400-1>
280. Ballard, L. M., Horton, R. H., Fenwick, A., & Lucassen, A. M. (2020). Genome sequencing in healthcare: understanding the UK general public's views and implications for clinical practice. European journal of human genetics, 28(2), 155-164. <https://doi.org/10.1038/s41431-019-0504-4>

281. Konstantinidis, G., Chapman, A., Weal, M., Alzubaidi, A., Ballard, L., & Lucassen, A. (2020). The Need for Machine-Processable Agreements in Health Data Management. *Algorithms*, 13(4), 87. <https://doi.org/10.3390/a13040087>
282. Forrest, I. S., Vy, H. M. T., Rocheleau, G., Jordan, D. M., Petrazzini, B. O., Nadkarni, G. N., et al. (2025). Machine learning–based penetrance of genetic variants. *Science*, 389(6763). <https://doi.org/10.1126/science.adm7066>
283. Lennon, N. J., Kottyan, L. C., Kachulis, C., Abul-Husn, N. S., Arias, J., Belbin, G. M., et al. (2024). Selection, optimization and validation of ten chronic disease polygenic risk scores for clinical implementation in diverse US populations. *Nature Medicine*, 30(2), 480. <https://doi.org/10.1038/s41591-024-02796-z>
284. Young, M., Yanes, T., Cust, A. E., Dunlop, K., Limb, S., Newson, A. J., et al. (2023). Human Genetics Society of Australasia Position Statement: Use of Polygenic Scores in Clinical Practice and Population Health. *Twin Research and Human Genetics*, 26(1), 40. <https://doi.org/10.1017/thg.2023.10>
285. Roberts, E., Flaum, N., & Evans, D. G. (2025). Clinical implementation of polygenic risk scores. *European Journal of Human Genetics*, 1-3. <https://doi.org/10.1038/s41431-025-01931-9>
286. Shaheen, S. O., & Ghaly, M. (2025). Stakeholder experiences and perspectives of genetic and genomic screening procedures in the Gulf Cooperation Council (GCC) region: a qualitative exploration. *Journal of Community Genetics*, 16(6), 813-825 <https://doi.org/10.1007/s12687-025-00819-x>
287. Dias, R., & Torkamani, A. (2019). Artificial intelligence in clinical and genomic diagnostics. *Genome Medicine*, 11(1). <https://doi.org/10.1186/s13073-019-0689-8>
288. Alsaedi, S. B., Ogasawara, M., Alarawi, M., Gao, X., & Gojobori, T. (2025). AI-powered precision medicine: utilizing genetic risk factor optimization to revolutionize healthcare. *NAR Genomics and Bioinformatics*, 7(2). <https://doi.org/10.1093/nargab/lqaf038>
289. Busby, G. B., Kulm, S., Bolli, A., Kintzle, J., Domenico, P. D., & Bottà, G. (2023). Ancestry-specific polygenic risk scores are risk enhancers for clinical cardiovascular disease assessments. *Nature communications*, 14(1), 7105. <https://doi.org/10.1038/s41467-023-42897-w>

290. Astrologo, N. C., Gaudillo, J., Albia, J. R., & Roxas-Villanueva, R. M. (2023). Genetic risk assessment based on association and prediction studies. *Scientific Reports*, 13(1).
<https://doi.org/10.1038/s41598-023-41862-3>
291. Ho, D., Schierding, W., Wake, M., Saffery, R., & O'Sullivan, J. M. (2019). Machine Learning SNP-Based Prediction for Precision Medicine. *Frontiers in Genetics*, 10, 267.
<https://doi.org/10.3389/fgene.2019.00267>
292. Chen, X., Wang, F., Zhao, H., Hao, J., Yunga, A., Yang, X., et al. (2026). Comprehensive benchmarking single and multiancestry polygenic score methods with the PGS-hub platform. *Nature Communications*, 17(1), 2014. <https://doi.org/10.1038/s41467-026-68599-7>
293. Bhak, Y., Guthrie, B., & Albert, T. (2026). Genetic insights into number of long-term conditions and their relationship with lifespan. *PLoS ONE*, 21(1).
<https://doi.org/10.1371/journal.pone.0340181>
294. Cole, J. M., Rybacki, S., Smith, S. P., Smith, O. S., & Harpak, A. (2026). Representation in genetic studies affects inference about genetic architecture. *bioRxiv*, 2026-01.
<https://doi.org/10.64898/2026.01.12.699135>
295. Coghlan, S., Gyngell, C., & Vears, D. F. (2023). Ethics of artificial intelligence in prenatal and pediatric genomic medicine. *Journal of Community Genetics*, 15(1), 13.
<https://doi.org/10.1007/s12687-023-00678-4>
296. Lingen, M. N. van, Giesbertz, N. A. A., & Jongsma, K. (2025). Digital technologies in genetic counseling: Recommendations for a morally sound integration. *Genetics in Medicine*, 27(4), 101370. <https://doi.org/10.1016/j.gim.2025.101370>
297. Meekins-Doherty, L., Dive, L., McEwen, A., & Sexton, A. (2025). Generative AI and the profession of genetic counseling. *Journal of Genetic Counseling*, 34(2).
<https://doi.org/10.1002/jgc4.2009>
298. Murugan, M., Yuan, B., Venner, E., Ballantyne, C. M., Robinson, K. M., Coons, J. C., et al. (2024). Empowering personalized pharmacogenomics with generative AI solutions. *Journal of the American Medical Informatics Association*, 31(6), 1356.
<https://doi.org/10.1093/jamia/ocae039>
299. Nemecek, A., Li, W., Jiang, X., Vaidya, J., & Ayday, E. (2026). Quantifying Memorization and Privacy Risks in Genomic Language Models. *arXiv preprint arXiv:2603.08913*.
<https://doi.org/10.48550/arXiv.2603.08913>

300. Albahri, A. S., Duham, A. M., Fadhel, M. A., Alnoor, A., Baqer, N. S., Alzubaidi, L., et al. (2023). A systematic review of trustworthy and explainable artificial intelligence in healthcare: Assessment of quality, bias risk, and data fusion. *Information fusion*, 96, 156-191. <https://doi.org/10.1016/j.inffus.2023.03.008>
301. Duong, D., & Solomon, B. D. (2025). Artificial intelligence in clinical genetics. *European Journal of Human Genetics*, 33(3), 281. <https://doi.org/10.1038/s41431-024-01782-w>
302. Lawrence, J. (2026). Current and Future Applications of Artificial Intelligence in Clinical Genetics. *Cureus*, 18(1). <https://doi.org/10.7759/cureus.100581>
303. Jain, S., Bakolitsa, C., Brenner, S. E., Radivojac, P., Moul, J., Repo, S., et al. (2024). CAGI, the Critical Assessment of Genome Interpretation, establishes progress and prospects for computational genetic variant interpretation methods. *Genome biology*, 25(1). <https://doi.org/10.1186/s13059-023-03113-6>
304. Zhuang, H. (2025). How genomics and multi-modal AI are reshaping precision medicine. *Frontiers in Medicine*, 12, 1660889. <https://doi.org/10.3389/fmed.2025.1660889>
305. Chen, J., Huang, Y., Fan, H., Wang, M., He, G., & Yan, J. (2025). Integrated genetic and geographic ancestry prediction via large-scale genomic data and machine learning. *Human Genomics*, 19(1). <https://doi.org/10.1186/s40246-025-00837-3>
306. Madduri, R., Li, Z., Nandi, T., Kim, K., Ryu, M., & Rodriguez, A. (2024). Advances in privacy preserving federated learning to realize a truly learning healthcare system. In 2024 IEEE 6th International Conference on Trust, Privacy and Security in Intelligent Systems, and Applications (TPS-ISA) (pp. 273-279). IEEE. <https://doi.org/10.1109/TPS-ISA62245.2024.00039>
307. Calvino, G., Peconi, C., Strafella, C., Trastulli, G., Megalizzi, D., Andreucci, S., et al. (2024). Federated Learning: Breaking Down Barriers in Global Genomic Research. *Genes*, 15(12), 1650. <https://doi.org/10.3390/genes15121650>
308. Danek, B., Makarious, M. B., Dadu, A., Vitale, D., Lee, P. S., Singleton, A., et al. (2024). Federated learning for multi-omics: A performance evaluation in Parkinson's disease. *Patterns*, 5(3), 100945. <https://doi.org/10.1016/j.patter.2024.100945>

309. Li, X., Gu, Y., Dvornek, N., Staib, L. H., Ventola, P., & Duncan, J. S. (2020). Multi-site fMRI analysis using privacy-preserving federated learning and domain adaptation: ABIDE results. *Medical Image Analysis*, 65, 101765. <https://doi.org/10.1016/j.media.2020.101765>
310. Doig, K., Perera, R., Kankanige, Y., Fellowes, A., Li, J., Lupat, R., et al. (2025). Using artificial intelligence (AI) to model clinical variant reporting for next generation sequencing (NGS) oncology assays. *BioData Mining*, 18(1), 74. <https://doi.org/10.1186/s13040-025-00489-y>
311. Kaushik, A., Ali, S., & Gupta, D. (2017). Altered pathway analyzer: A gene expression dataset analysis tool for identification and prioritization of differentially regulated and network rewired pathways. *Scientific reports*, 7(1), 40450. <https://doi.org/10.1038/srep40450>
312. Vivar, J. C., Pemu, P., McPherson, R., & Ghosh, S. (2013). Redundancy control in pathway databases (ReCiPa): an application for improving gene-set enrichment analysis in Omics studies and “Big data” biology. *Omics: a journal of integrative biology*, 17(8), 414-422. <https://doi.org/10.1089/omi.2012.0083>
313. Stoney, R. A., Schwartz, J. M., Robertson, D. L., & Nenadic, G. (2018). Using set theory to reduce redundancy in pathway sets. *BMC bioinformatics*, 19(1), 386. <https://doi.org/10.1186/s12859-018-2355-3>
314. Zhao, B., Su, T. Y., Wang, J., Islam, Q. S., Song, K., Huang, S. K., et al. (2025). DENetwork unveils non-differentially expressed genes with functional relevance across conditions through information flow perturbation. *Nucleic acids research*, 53(22), gkaf1350. <https://doi.org/10.1093/nar/gkaf1350>
315. Gambardella, G., Moretti, M. N., De Cegli, R., Cardone, L., Peron, A., & Di Bernardo, D. (2013). Differential network analysis for the identification of condition-specific pathway activity and regulation. *Bioinformatics*, 29(14), 1776-1785. <https://doi.org/10.1093/bioinformatics/btt290>
316. Ma, X., Gao, L., & Tan, K. (2014). Modeling disease progression using dynamics of pathway connectivity. *Bioinformatics*, 30(16), 2343-2350. <https://doi.org/10.1093/bioinformatics/btu298>
317. Liu, X., Maiorino, E., Halu, A., Glass, K., Prasad, R. B., Loscalzo, J., et al. (2020). Robustness and lethality in multilayer biological molecular networks. *Nature communications*, 11(1), 6043. <https://doi.org/10.1038/s41467-020-19841-3>

318. Jiang, Y., Che, L., & Li, S. C. (2026). Deciphering the personalized functional redundancy hierarchy in the gut microbiome. *Microbiome*, 14(1), 17. <https://doi.org/10.1186/s40168-025-02273-w>
319. Lachowiec, J., Mason, G. A., Schultz, K., & Queitsch, C. (2018). Redundancy, feedback, and robustness in the *Arabidopsis thaliana* BZR/BEH gene family. *Frontiers in genetics*, 9, 523. <https://doi.org/10.3389/fgene.2018.00523>
320. Macneil, L. T., & Walhout, A. J. (2011). Gene regulatory networks and the role of robustness and stochasticity in the control of gene expression. *Genome research*, 21(5), 645–657. <https://doi.org/10.1101/gr.097378.109>
321. Tran, T. C., Mähl, K., Kappel, C., Dakhiya, Y., Sampathkumar, A., Sicard, A., et al. (2024). Altered interactions between cis-regulatory elements partially resolve BLADE-ON-PETIOLE genetic redundancy in *Capsella rubella*. *The Plant cell*, 36(10), 4637–4657. <https://doi.org/10.1093/plcell/koae232>
322. Cuscó, I., Barceló, M. J., Rojas-García, R., Illa, I., Gámez, J., Cervera, C., et al. (2006). SMN2 copy number predicts acute or chronic spinal muscular atrophy but does not account for intrafamilial variability in siblings. *Journal of neurology*, 253(1), 21–25. <https://doi.org/10.1007/s00415-005-0912-y>
323. Zhang, D., Li, X., Wang, Z., Zhang, Y., Guo, K., Wang, S., Tu, C., et al. (2015). Hailey-Hailey disease: investigation of a possible compensatory SERCA2 up-regulation and analysis of SPCA1, p63, and IRF6 expression. *Archives of dermatological research*, 307(2), 143–149. <https://doi.org/10.1007/s00403-014-1506-2>
324. Zou, W. B., Cooper, D. N., Masson, E., Pu, N., Liao, Z., Férec, C., et al. (2022). Trypsinogen (PRSS1 and PRSS2) gene dosage correlates with pancreatitis risk across genetic and transgenic studies: a systematic review and re-analysis. *Human genetics*, 141(8), 1327–1338. <https://doi.org/10.1007/s00439-022-02436-x>
325. Venkatesh, A., Quinn, N., Upadhy, S. R. et al. (2025). Proteomic compensation by paralogs preserves protein interaction networks after gene loss in cancer. *Molecular Systems Biology*, 21, 1090–1118. <https://doi.org/10.1038/s44320-025-00122-4>
326. Flister, M. J., Verduzco, D., Petiwala, S., Carneiro-Lobo, T. C., Gekas, C., Shi, X., et al. (2025). Genetic mapping and predictive modeling of paralog synthetic lethality. *Cell reports*, 44(11), 116512. <https://doi.org/10.1016/j.celrep.2025.116512>

327. Minei, R., Aoki, H., Ogura, A. et al. (2023). Compensatory gene expression potentially rescues impaired brain development in Kit mutant mice. *Scientific reports*, 13, 4166.
<https://doi.org/10.1038/s41598-023-30032-0>
328. Shojaie, A. (2021). Differential network analysis: A statistical perspective. *Wiley Interdisciplinary Reviews: Computational Statistics*, 13(2), e1508.
<https://doi.org/10.1002/wics.1508>
329. Grimes, T., Potter, S. S. & Datta, S. (2019). Integrating gene regulatory pathways into differential network analysis of gene expression data. *Scientific reports*, 9, 5479.
<https://doi.org/10.1038/s41598-019-41918-3>
330. Guzzi, P. H., Roy, A., Milano, M., & Veltri, P. (2024). Non parametric differential network analysis: a tool for unveiling specific molecular signatures. *BMC bioinformatics*, 25(1), 359.
<https://doi.org/10.1186/s12859-024-05969-2>
331. Alcaraz, N., Pauling, J., Batra, R., Barbosa, E., Junge, A., Christensen, A. G. L., Azevedo, V., Ditzel, H. J., & Baumbach, J. (2014). KeyPathwayMiner 4.0: Condition-specific pathway analysis by combining multiple omics studies and networks with Cytoscape. *BMC Systems Biology*, 8, 99. <https://doi.org/10.1186/s12918-014-0099-x>
332. Alcaraz, N., List, M., Dissing-Hansen, M., Rehmsmeier, M., Tan, Q., Mollenhauer, J., Ditzel, H. J., & Baumbach, J. (2016). Robust de novo pathway enrichment with KeyPathwayMiner 5. *F1000Research*, 5, 1531. <https://doi.org/10.12688/f1000research.9054.1>
333. Sabbatini, G., & Manganaro, L. (2022). On potential limitations of differential expression analysis with non-linear machine learning models. *EMBnet. journal*, 28, 1035.
<https://doi.org/10.14806/ej.28.0.1035>
334. Aldea, D., Kokalari, B., Atsuta, Y., Dingwall, H. L., Zheng, Y., Nace, A., Cotsarelis, G., & Kamberov, Y. G. (2023). Differential modularity of the mammalian Engrailed 1 enhancer network directs sweat gland development. *PLoS genetics*, 19(2), e1010614.
<https://doi.org/10.1371/journal.pgen.1010614>
335. Cantini, L., Calzone, L., Martignetti, L., Rydenfelt, M., Blüthgen, N., Barillot, E., & Zinovyev, A. (2018). Classification of gene signatures for their information value and functional redundancy. *NPJ systems biology and applications*, 4(1), 2. <https://doi.org/10.1038/s41540-017-0038-8>

336. Zhou, Q., Madala, N. S., & Huang, C. (2025). Pathway-guided architectures for interpretable AI in biological research. *Computational and structural biotechnology journal*, 27, 4779–4791. <https://doi.org/10.1016/j.csbj.2025.10.063>
337. Zhou, Q., Ye, W., Yu, X., & Bao, Y. J. (2024). A pathway-based computational framework for identification of a new modal of multi-omics biomarkers and its application in esophageal cancer. *Computer Methods and Programs in Biomedicine*, 247, 108077. <https://doi.org/10.1016/j.cmpb.2024.108077>
338. Smith, L. A., Cahill, J. A., Lee, JH. et al. (2025). Equitable machine learning counteracts ancestral bias in precision medicine. *Nature Communications*, 16, 2144. <https://doi.org/10.1038/s41467-025-57216-8>
339. Chen, L., Liu, R., Liu, ZP. et al. (2012). Detecting early-warning signals for sudden deterioration of complex diseases by dynamical network biomarkers. *Scientific reports*, 2(1), 342. <https://doi.org/10.1038/srep00342>
340. Han, C., Zhong, J., Zhang, Q., Hu, J., Liu, R., Liu, H., et al. (2022). Development of a dynamic network biomarkers method and its application for detecting the tipping point of prior disease development. *Computational and structural biotechnology journal*, 20, 1189-1197. <https://doi.org/10.1016/j.csbj.2022.02.019>
341. Marku, M., Chenel, H., Bordenave, J. et al. (2026). Data driven network inference and longitudinal transcriptomics unveil dynamic regulation in Chronic Lymphocytic Leukaemia models. *NPJ systems biology and applications*, 12(1), 24. <https://doi.org/10.1038/s41540-025-00645-4>
342. Barabási, AL., Gulbahce, N. & Loscalzo, J. (2011). Network medicine: a network-based approach to human disease. *Nature Reviews Genetics*, 12(1), 56–68. <https://doi.org/10.1038/nrg2918>
343. Keikha, F., Wang, C., Yang, Z., & Liu, Z. P. (2025). TransMarker: Unveiling dynamic network biomarkers in cancer progression through cross-state graph alignment and optimal transport. *PLOS Computational Biology*, 21(11), e1013743. <https://doi.org/10.1371/journal.pcbi.1013743>
344. Lyu, C., Chen, L., & Liu, X. (2024). Detecting tipping points of complex diseases by network information entropy. *Briefings in Bioinformatics*, 25(4), bbae311. <https://doi.org/10.1093/bib/bbae311>

345. Rimban, E. L. (2025). Dynamic network biology: moving beyond static representations to model temporal complexity. *International Journal of Molecular Biology: Open Access*, 8(1), 71–77. [10.15406/ijmboa.2025.08.00200](https://doi.org/10.15406/ijmboa.2025.08.00200)
346. Khazaal, A., & Vafaee, F. (2026). Temporal network analysis in systems biology: concepts, inference, and validation. *Frontiers in Bioinformatics*, 6, 1822526. <https://doi.org/10.3389/fbinf.2026.1822526>
347. Holme, P., & Saramäki, J. (2012). Temporal networks. *Physics reports*, 519(3), 97-125. <https://doi.org/10.1016/j.physrep.2012.03.001>
348. Montgomery, R. M. (2025). Topological Dynamics in Complex Biological Systems: A Unified Mathematical Framework Integrating Topology, Statistical Mechanics, and Neural Networks. *Scottish Science Society*, 1(4), 1-49. <https://doi.org/10.62162/SSS1061014>
349. Bartsch, R. P., Liu, K. K., Bashan, A., & Ivanov, P. C. (2015). Network physiology: how organ systems dynamically interact. *PloS one*, 10(11), e0142143. <https://doi.org/10.1371/journal.pone.0142143>
350. Bashan, A., Bartsch, R., Kantelhardt, J. et al. (2012). Network physiology reveals relations between network topology and physiological function. *Nature Communications*, 3, 702. <https://doi.org/10.1038/ncomms1705>
351. Valenza, G., Matic, Z. & Catrambone, V. (2025). The brain–heart axis: integrative cooperation of neural, mechanical and biochemical pathways. *Nature Reviews Cardiology*, 22(8), 537–550. <https://doi.org/10.1038/s41569-025-01140-3>
352. Che, H., Gao, Y., Xu, Y. et al. (2026). Organ cross-talk: molecular mechanisms, biological functions, and therapeutic interventions for diseases. *Signal Transduction Targeted Therapy*, 11(1), 8. <https://doi.org/10.1038/s41392-025-02329-1>
353. Izgi, H., Han, D., Isildak, U., Huang, S., Kocabiyik, E., Khaitovich, P., Somel, M., & Dönertaş, H. M. (2022). Inter-tissue convergence of gene expression during ageing suggests age-related loss of tissue and cellular identity. *Elife*, 11, e68048. <https://doi.org/10.7554/eLife.68048>
354. Deng, T., Huang, M., Xu, K., Lu, Y., Xu, Y., Chen, S., et al. (2025). LEGEND: Identifying co-expressed genes in multimodal transcriptomic sequencing data. *Genomics, Proteomics & Bioinformatics*, 23(4), qzaf056. <https://doi.org/10.1093/gpbjnl/qzaf056>

355. Thomas, F., Dujon, A. M., Vaiman, D., Eberl, G., Alix-Panabières, C., Pujol, P., Ujvari, B., Meliani, J., Nedelcu, A. M., & Capp, J.-P. (2026). Selection for Function in Complex Distributed Pathological Systems. *Evolutionary Applications*, 19(2), e70202.

<https://doi.org/10.1111/eva.70202>

356. Plouzenec, S., de la Barca, J. M. C., & Chevrollier, A. (2025). The role of phospholipids in mitochondrial dynamics and associated diseases. *Frontiers in Bioscience-Landmark*, 30(8), 27634. <https://doi.org/10.31083/FBL27634>

357. Mitra, S., BP, K., C R, S., Saikumar, N. V., Philip, P., & Narayanan, M. (2024). Alzheimer's disease rewires gene coexpression networks coupling different brain regions. *NPJ systems biology and applications*, 10(1), 50. <https://doi.org/10.1038/s41540-024-00376-y>

358. Paranjpe, M. D., Belonwu, S., Wang, J. K., Oskotsky, T., Gupta, A., Taubes, A., et al. (2021). Sex-specific cross tissue meta-analysis identifies immune dysregulation in women with Alzheimer's disease. *Frontiers in Aging Neuroscience*, 13, 735611. <https://doi.org/10.3389/fnagi.2021.735611>

359. Langella, S., Sadiq, M. U., Mucha, P. J. et al. (2021). Lower functional hippocampal redundancy in mild cognitive impairment. *Translational psychiatry*, 11(1), 61. <https://doi.org/10.1038/s41398-020-01166-w>

360. Erman, B. (2026). The Gaussian network model as a framework for allosteric analysis: dynamic distance, edge centrality, and entropy sensitivity in KRAS. *Physical Biology*, 23(1), 016005. <https://doi.org/10.1088/1478-3975/ae3e49>

361. Taheriyoun, A. R., Ross, A., Safikhani, A., Soudbakhsh, D., & Rahnavard, A. (2026). Longitudinal omics data analysis: approaches and applications. *Computational and Structural Biotechnology Journal*, 201-315. <https://doi.org/10.1016/j.csbj.2026.01.001>

362. Tononi, G., Sporns, O., & Edelman, G. M. (1999). Measures of degeneracy and redundancy in biological networks. *Proceedings of the National Academy of Sciences*, 96(6), 3257-3262. <https://doi.org/10.1073/pnas.96.6.32>

363. Zucknick, M., Richardson, S., & Stronach, E. A. (2008). Comparing the characteristics of gene expression profiles derived by univariate and multivariate classification methods. *Statistical applications in genetics and molecular biology*, 7(1), 7. <https://doi.org/10.2202/1544-6115.1307>

364. Barido-Sottani, J., Chapman, S. D., Kosman, E. et al. (2019). Measuring similarity between gene interaction profiles. *BMC Bioinformatics*, 20(1), 435. <https://doi.org/10.1186/s12859-019-3024-x>