Functional & Integrative Genomics (2024) 24:138
https://doi.org/10.1007/s10142-024-01417-9

REVIEW

The impact and future of artificial intelligence in medical genetics and molecular medicine: an ongoing revolution

Firat Ozcelik[1] · Mehmet Sait Dundar[2] · A. Baki Yildirim[1] · Gary Henehan[3] · Oscar Vicente[4] · José A. Sánchez-Alcázar[5] · Nuriye Gokce[1] · Duygu T. Yildirim[1] · Nurdeniz Nalbant Bingol[6] · Dijana Plaseska Karanfilska[7] · Matteo Bertelli[8] · Lejla Pojskic[9] · Mehmet Ercan[1] · Miklos Kellermayer[10] · Izem Olcay Sahin[1] · Ole K. Greiner-Tollersrud[11] · Busra Tan[1] · Donald Martin[12] · Robert Marks[13] · Satya Prakash[14] · Mustafa Yakubi[1] · Tommaso Beccari[15] · Ratnesh Lal[16] · Sehime G. Temel[6,17,18] · Isabelle Fournier[19] · M. Cerkez Ergoren[20] · Adam Mechler[21] · Michel Salzet[19] · Michele Maffia[22] · Dancho Danalev[23] · Qun Sun[24] · Lembit Nei[25] · Daumantas Matulis[26] · Dana Tapaloaga[27] · Andres Janecke[28,29] · James Bown[30] · Karla Santa Cruz[31] · Iza Radecka[32] · Celal Ozturk[33] · Ozkan Ufuk Nalbantoglu[34] · Sebnem Ozemri Sag[17] · Kisung Ko[35] · Reynir Arngrimsson[36] · Isabel Belo[37] · Hilal Akalin[1] · Munis Dundar[1]

Received: 2 July 2024 / Revised: 1 August 2024 / Accepted: 5 August 2024 / Published online: 16 August 2024

Abstract

Artificial intelligence (AI) platforms have emerged as pivotal tools in genetics and molecular medicine, as in many other fields. The growth in patient data, identification of new diseases and phenotypes, discovery of new intracellular pathways, availability of greater sets of omics data, and the need to continuously analyse them have led to the development of new AI platforms. AI continues to weave its way into the fabric of genetics with the potential to unlock new discoveries and enhance patient care. This technology is setting the stage for breakthroughs across various domains, including dysmorphology, rare hereditary diseases, cancers, clinical microbiomics, the investigation of zoonotic diseases, omics studies in all medical disciplines. AI's role in facilitating a deeper understanding of these areas heralds a new era of personalised medicine, where treatments and diagnoses are tailored to the individual's molecular features, offering a more precise approach to combating genetic or acquired disorders. The significance of these AI platforms is growing as they assist healthcare professionals in the diagnostic and treatment processes, marking a pivotal shift towards more informed, efficient, and effective medical practice. In this review, we will explore the range of AI tools available and show how they have become vital in various sectors of genomic research supporting clinical decisions.

Keywords Artificial intelligence · Molecular genetics · Diagnostic · Therapy · Omics

Introduction

The human genome still holds many mysteries more than 20 years after the initial results of the Human Genome Project were obtained (Marian 2014). More than 7,000 rare diseases have been defined in the literature, and it has been estimated that between 40 and 80% of these disorders have genetic aetiology (Ferreira 2019). Moreover, many genetic diseases have overlapping features regarding their clinical, radiological and laboratory findings. Dysmorphological findings may accompany some phenotypes and their value in diagnosing rare syndromes is undeniable. This complex architecture of genetic diseases poses significant challenges to medical geneticists and other clinicians in diagnosing and treating these conditions. Next-generation sequencing (NGS) methods have made an impactful contribution to diagnosing complex genetic disorders. The massive genomic data they generated helped define genotype-phenotype relations and diagnose many more patients. These advancements enabled defining phenotypes from genotype data, a process known

Firat Ozcelik and Mehmet Sait Dundar contributed equally to this work.

Extended author information available on the last page of the article

as "reverse phenotyping". Furthermore, long-read sequencing methods such as Oxford Nanopore Sequencing and PacBio Sequencing have been developed. These methods offer real-time sequencing of single molecules, are faster and more cost-effective, and overcome certain shortcomings of generating shorter reads (Athanasopoulou et al. 2022). Nonetheless, the diagnostic workup may remain inconclusive even after all available tests, and the contribution from other disciplines, such as Artificial Intelligence (AI), could be valuable (Davenport and Kalakota 2019).

Other than rare hereditary diseases, genomic changes also occur in cancers, as well as other complex and multifactorial diseases. Furthermore, changes in the RNA, protein, and metabolite levels and epigenetic modifications may provide valuable insights into the etiopathogenesis of hereditary and acquired disorders (Sun and Hu 2016). The massive progress in omics studies made it possible to unravel molecular mechanisms of diseases and cellular processes and offered revolutionary approaches in personalized medicine. This leap in omics studies generated large amounts of data, which required bioinformatical computerized approaches to analyse (Sun and Hu 2016).

In its simplest definition, AI is the capability of a computer to perform tasks that normally require human intelligence, including speech recognition, natural language comprehension, and decision-making processes. Deep learning (DL) and Machine Learning (ML) are subcategories of AI. ML is inspired by the functioning of the human brain. Neurons interact in networks to form the basis of all decision-making based on gathered information. Similarly, ML techniques systematically process available data using algorithms to learn, analyse, and make decisions. A subset of ML, Deep Learning, utilises neural networks to process significantly larger datasets to derive insights (Soori et al. 2023). In areas with large amounts of complex data, AI platforms are crucial for quickly analysing and evaluating these data.

The intersection of AI and genetics presents an opportunity to achieve profound insights and enhanced treatments for a wide array of inherited disorders, cancers, and also multifactorial diseases (Modi et al. 2023). The growing accessibility of omics data thanks to the large databases and biobanks, alongside advancements in AI, has facilitated this convergence, offering novel methodologies for data analysis, interpretation, and utilisation (Fig. 1). The importance of integrating AI into genetics extends beyond merely augmenting the accuracy or efficiency of genetic analyses. It also includes the identification of novel gene associations, therapeutic targets, and personalised medicine concepts (Abdallah et al. 2023).

Amalgamating diverse data sources marks a pivotal shift towards data-driven discovery. Furthermore, obtaining big data could be used to generate new hypotheses, which could then be tested experimentally (Hulsen et al. 2019). AI algorithms, powered by machine learning models and deep learning networks, transcend pattern recognition to analyse genetic data with a capability far surpassing human analytical potential (Table 1). Such capabilities are indispensable for deciphering the complexities of the human genome, uncovering the genetic underpinnings of diseases, developing targeted therapies, studying drug resistance in infectious diseases, and predicting drug response using pharmacogenetics. AI applications in genetics also significantly enhance the precision of genetic testing, disease prediction, and the determination of optimal treatments. Gene editing and, synthetic biology applications are other research fields that could greatly benefit from AI-driven tools (Gartland et al. 2017; García Martín et al. 2024).

The benefits of AI in molecular research and diagnostics are not limited to human genetics. AI provides valuable contributions to omics studies across all organisms. Genomic, transcriptomic, proteomic, metabolomic, epigenomic, and multi-omic studies utilizing AI in other animals, plants, and microorganisms are becoming more and more common (Cembrowska-Lech et al. 2023). Therefore, AI is being incorporated into the disciplines of infectious diseases and microbiomics, too (Smith et al. 2020).

In this article, our primary objectives are to explore the utilization of AI platforms across various sectors of molecular research and diagnostics and to spotlight some of the most prominent AI platforms employed in genetics and molecular medicine. Even though our review mostly covers topics related to medical genetics, we also discuss AI's implications related to molecular research in infectious diseases, microbiomics, and synthetic biology The topics discussed below are based on a consensus among the authors. Our article provides a perspective on the current state of AI in molecular medicine and medical genetics, its limitations and its potential implications for the future.

Dysmorphology

Many genetic syndromes, especially neurodevelopmental disorders, present with characteristic craniofacial features (Carrer et al. 2023). For instance, 30–40% of known inherited diseases have identifiable craniofacial, oral, or dental features (Hart and Hart 2009). Additionally, various teratogens may also give rise to congenital malformations and craniofacial findings (Gilbert-Barness 2010; Saal 2016). Dysmorphology explores patterns of human morphologic development, structural defects, and their association with syndromes (Hunter 2002; Basel 2020). In over 1500 different genetic syndromes, dysmorphological facial features

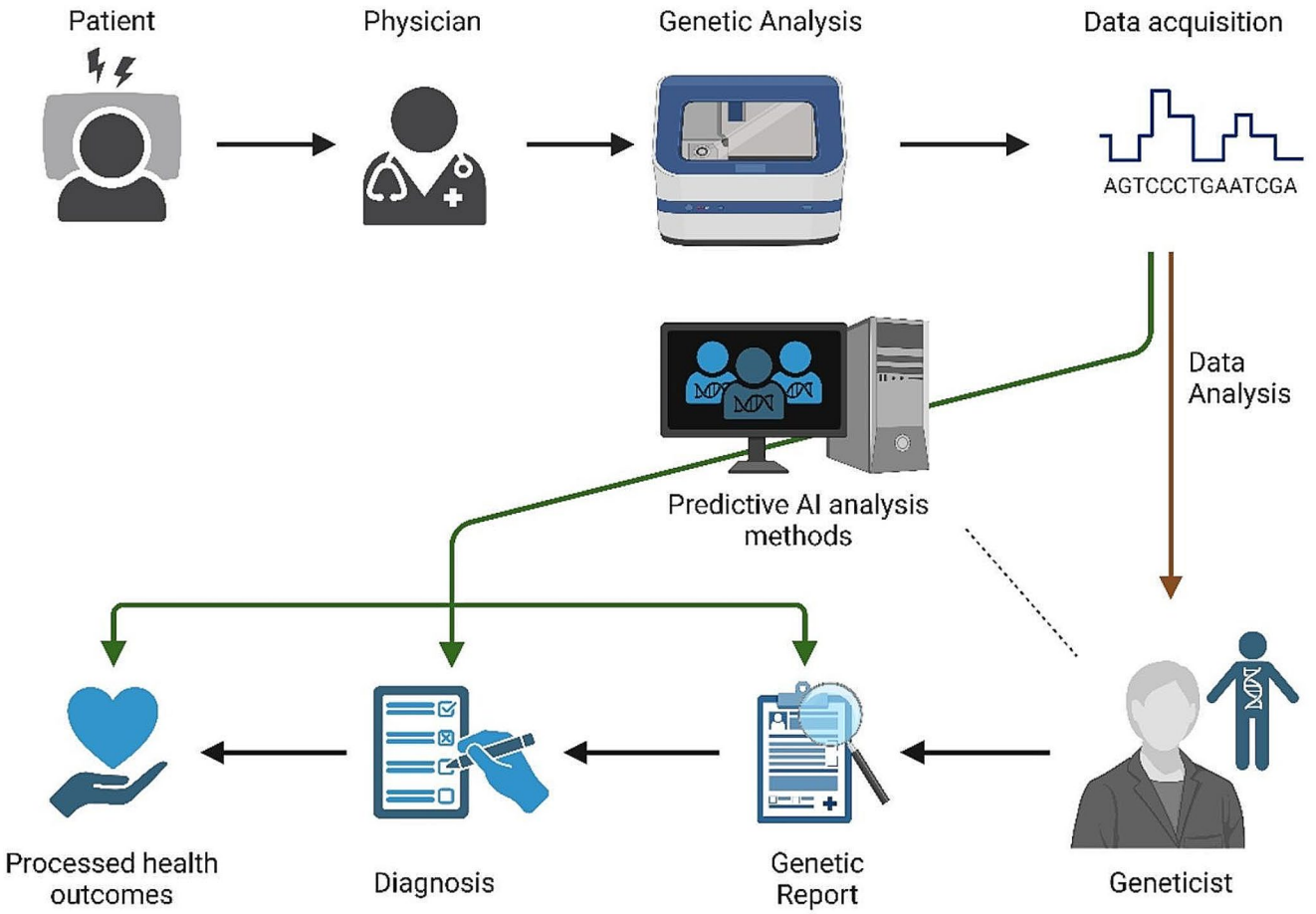

Fig. 1 An illustration of a diagnostic genetic testing workflow. Using AI algorithms to compare and analyse patients' genetic data accelerates the development of diagnosis and treatment plans

specific to the syndrome are observed (Bannister et al. 2020). Such a large number of dysmorphological facial findings has led to the development of phenotype databases and automated assessment tools (Gurovich et al. 2019a). Human phenotype ontology (HPO) is a comprehensive source of the phenotypes associated with human disease, which is also useful for electronic health record analyses and machine learning approaches (Gargano et al. 2024). Phenomizer is a tool that uses HPO data and semantic similarity measures to generate p values for relevant diseases when the presenting clinical findings and dysmorphic features are given. Therefore, the most likely preliminary diagnoses may be obtained (Posey et al. 2017). Furthermore, visual data are also incorporated into the diagnostic process of dysmorphic patients. The most commonly used face recognition application is Face2Gene (FDNA Inc, USA). Face2Gene uses a deep convolutional neural network called DeepGestalt to analyse patients' images and identify the characteristic patterns of the syndrome (Carrer et al. 2023). The DeepGestalt algorithm was initially "trained" with more than 17,000 images from people with 200 different syndromes (Gurovich et al. 2019b). Face2Gene contains a databank of images with dysmorphic facial features characteristic of different genetic syndromes (Mishima et al. 2019). The facial image of the person under examination is uploaded to the algorithm and evaluated by comparing it with the images in the database. In a study using facial images of individuals with Silver-Russell (SRS) and Prader-Willi (PWS) syndromes, the first syndrome suggested in 41% of patients with SRS was SRS, whereas in PWS, the algorithm suggested the correct diagnosis as the most likely syndrome in 76% of the cases (Ciancia et al. 2023). Another AI-driven tool used in dysmorphology based on DeepGestalt framework is GestaltMatcher. GestaltMatcher finds phenotypic resemblences between the uploaded cases and matches them. Therefore, it can contribute to the diagnoses of ultra-rare diseases even if the algorithm was not trained for that particular diagnosis (Hsieh et al. 2021). An important issue concerning both deep learning algorithms and clinical judgement in dysmorphology is that ethnicity and the age of the patient may alter the diagnostic yield. Furthermore, the experience of the clinician and the extent of the available dataset and training of

Table 1 Leading artificial intelligence algorithms used in various areas of omics research

Artificial intelligence tool	Main purpose	References
Dysmorphology and Clinical Genetics Practice		
Phenomizer	Clinical dysmorphology, uses HPO phenotypes	(Posey et al. 2017)
DeepGestalt	Clinical dysmorphology, uses images	(Carrer et al. 2023)
Face2Gene	Clinical dysmorphology, uses images, based on DeepGestalt	(Mishima et al. 2019)
GestaltMatcher	Clinical dysmorphology, uses images, based on DeepGestalt	(Hsieh et al. 2021)
Eye2Gene	Inherited retinal disorders diagnosis, uses retinal images	(PONTIKOS et al. 2022)
Molecular Cancer Research		
Lung-CLiP	ctDNA detection	(Chabon et al. 2020)
CancerSEEK	early cancer detection	(Cohen et al. 2018)
CancerVar	variant interpretation for cancer	(Li et al. 2022a)
DeepPATH	analysing Whole Slide Images	(Coudray et al. 2018)
PREDICT	drug repurposing	(Bhinder et al. 2021)
Rare Hereditary Diseases		
CADD	prediction of variant pathogenicity	(Schubach et al. 2024)
REVEL	prediction of variant pathogenicity	(Ioannidis et al. 2016)
MutationTaster	prediction of variant pathogenicity	(Schwarz et al. 2014)
Exomiser	exome or genome sequencing analysis using HPO phenotypes	(Smedley et al. 2015)
Fabric GEM	exome or genome sequencing analysis using HPO phenotypes	(De La Vega et al. 2021)
PhenIX	exome or genome sequencing analysis using HPO phenotypes	(Zemojtel et al. 2014)
Three-dimensional genomic research		
DeepC	predicts DNA folding in three dimensions	(Schwessinger et al. 2020)
ChINN	predicts chromatin interactions	(Cao et al. 2021)
Akita	predicts DNA folding in three dimensions	(Fudenberg et al. 2020)
Orca	predicts DNA folding in three dimensions	(Zhou 2022)
DeepCTCFLoop	predicts chromatin interactions	(Kuang and Wang 2021)
Epigenetics		
CHICAGO	investigates chromatin interactions	(McMaster et al. 2018)
Transcriptomics		
Cibersort /Cibersortx	determines cell type abundance and expression from bulk tissue analyses	(Newman et al. 2015, 2019)
SEDR	spatial transcriptomics	(Li et al. 2022c; Erfanian et al. 2023)
SpaCell	spatial transcriptomics	(Li et al. 2022c; Erfanian et al. 2023)
Tangram	spatial transcriptomics	(Li et al. 2022c; Erfanian et al. 2023)
DEEPsc	spatial transcriptomics	(Li et al. 2022c; Erfanian et al. 2023)
GCNG	spatial transcriptomics	(Li et al. 2022c; Erfanian et al. 2023)
DeepST	spatial transcriptomics	(Li et al. 2022c; Erfanian et al. 2023)
SPANR	predicts genetic variant effects on splicing	(Wang et al. 2023)
MMSplice	predicts genetic variant effects on splicing	(Wang et al. 2023)
MTSplice	predicts genetic variant effects on splicing	(Wang et al. 2023)
SpliceAI	predicts genetic variant effects on splicing	(Wang et al. 2023)
Pangolin	predicts genetic variant effects on splicing	(Wang et al. 2023)
Proteomics		
AlphaFold	predicts three dimensional structures of proteins	(Jumper et al. 2021)
BIAS	proteomics and spatial information	(Mund et al. 2022)
DeepNovo	de novo peptide sequencing	(Li et al. 2022b)
pDeep	spectrum prediction of peptides	(Li et al. 2022b)
DeepRT	peptide retention time prediction	(Li et al. 2022b)
DeepIso	peptide feature detection	(Li et al. 2022b)
DeepBind	nucleic acid binding protein identification	(Li et al. 2022b)
PoxiPred	design multi-epitope vaccines	(Martinez et al. 2024)
DISCO-MS	spatial proteomics	(Bhatia et al. 2022)
MHCpLogics	immunopeptidome analysis	(Shahbazy et al. 2024)
Drug design and discovery		

Table 1 (continued)

Artificial intelligence tool	Main purpose	References
BenevolentAI	drug discovery	(Richardson et al. 2023)
URSAHD	disease classification, drug repurposing	(Brasil et al. 2019; Lee et al. 2019)
mediKanren	rare disease drug repurposing	(Foksinska et al. 2022)
Metabolomics		
MS2Mol	Metabolomics	(Aksenov et al. 2021)
Gene editing		
DeepCRISPR	Gene Editing	(Chuai et al. 2018)
CRISPRLearner	Gene Editing	(Dimauro et al. 2019)
CRISPR-GA	Gene Editing	(Güell et al. 2014)

the DL solutions may modify the success rate (Lumaka et al. 2017; Mishima et al. 2019). These issues could be handled with broadening the datasets with more patients from different populations and age groups.

Another widely used application that uses visual input and helps making differential diagnoses in genetic diseases is Eye2Gene, an AI tool used in retinal disease diagnosis. Inherited Retinal Disorders (IRD) may be influenced by more than 260 genes. The gene defects lead to functional disruptions in the structure of proteins that play a role in the retina (Duncan et al. 2018). This AI tool is particularly useful due to the often limited clinical expertise in the diagnosis of IRD (Poplin et al. 2018). In a study of the 36 most common causes of Retinal Disorders that evaluated the effectiveness of Eye2Gene, the top five diagnosis accuracy was 85.6% (PONTIKOS et al. 2022).

Genetic counselling – chatbots

The field of genetic counselling has experienced significant growth in recent years, with an increasing demand for genetic counselling services across various healthcare settings. As with other medical disciplines, AI and other technologies are increasingly shaping the practice of genetic counselling (Gordon et al. 2018).

One application of AI in the genetic counseling field is a chatbot, a virtual assistant designed to simulate conversation with human users, typically through text or voice interfaces (Kearney et al. 2020). Gia (Genetic Information Assistant) is a chatbot designed to assist genetic counsellors and patients in navigating genetic information and counselling processes. Gia assists in obtaining patient consent, offers post-genetic test result follow-up, and includes a sharing feature to aid discussions with potentially affected family members. Feedback from patient focus groups highlighted their satisfaction by Gia's convenience and pace (Nazareth et al. 2021). Another chatbot, Rosa, was developed in Norway to provide easily accesible and reliable genetic counselling to the patients regarding the hereditary breast and ovarian cancer risk (Siglen et al. 2023). To improve the communication about additional findings in genetic tests, a chatbot called Edna was developed (Ireland et al. 2021).

Undoubtedly, AI-driven chatbots hold great promise for advancing genetic counselling services by improving their accessibility, efficiency, accuracy, and personalisation (Fig. 2). However, the responsible and equitable integration of AI into genetic counselling practice requires addressing the potential risks and challenges related to privacy, bias, communication, regulation and ethical considerations (Dundar and Emirogullari 2012; Gordon et al. 2018). The recent developments in the generative AI applications may facilitate answering the patients' questions more reliably and aid the physicians or genetic counsellors provide optimal care.

Molecular cancer research

AI applications in cancer research have led to major breakthroughs in identifying gene alterations and classifying various types of cancers (Fig. 3). Molecular alterations in cancer-related genes may occur as germline or somatic variations. Germline variants may confer susceptibility to developing certain types of cancers and influence the interactions with the anticancer drugs. On the other hand, somatic variants may develop in the cancer genome and accumulate during carcinogenesis. While the development of driver mutations result in an increase in the replication rate of cancer cells and provide a "fitness" advantage, passenger mutations occur in cells with driver mutations without such an effect (Ostroverkhova et al. 2023). Both germline and somatic variations may influence the treatment strategies amd provide valuable information in the management of cancer patients. Data derived from next-generation sequencing (NGS) technologies, including whole exome sequencing, whole genome sequencing, targeted panels, as well as transcription profiles from microarray, RNA-seq, and microRNAs, along with methylation profiles, are invaluable resources for the diagnosis of cancer, classification of tumours into distinct subtypes, and identification of actionable targets For this purpose, blood samples, tumour samples and formalin-fixed paraffin-embedded (FFPE)

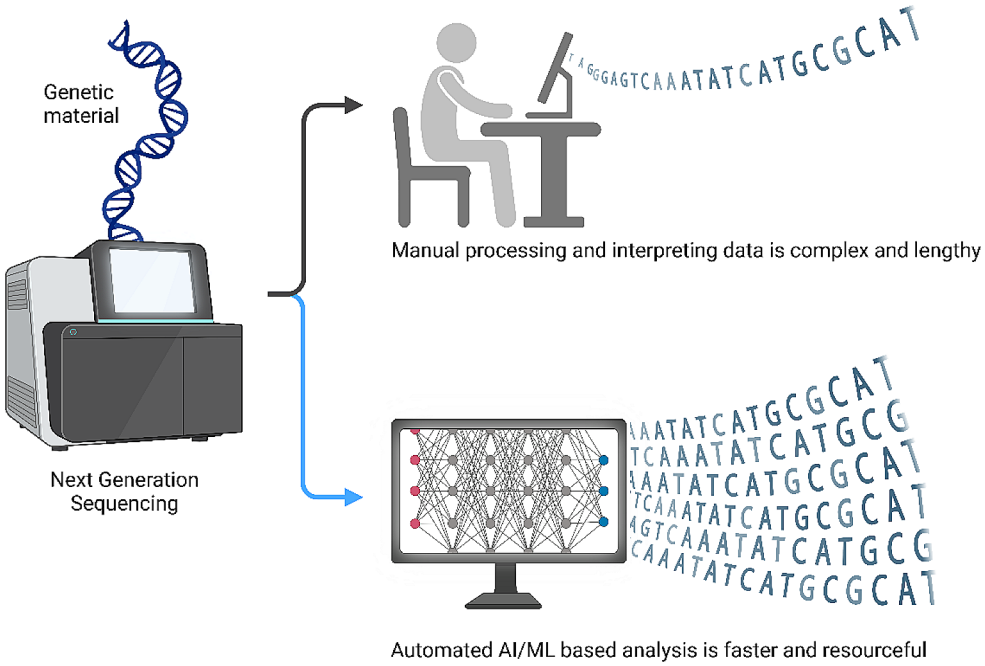

Fig. 2 AI offers significant advantages over traditional methods in its ability to analyse high volumes of genetic data quickly and efficiently. While traditional methods take time for manual processing and data interpretation, AI-based analysis is faster and resource efficient

blocks, liquid biopsies and cell lines can be used (Dlamini et al. 2020). The vast amount of generated data may be incorporated into large databases like The Cancer Genome Atlas (TCGA). TCGA is a landmark project initiated in 2006 by the National Cancer Institute and the National Human Genome Research Institute. It creates a comprehensive atlas of cancer genomic profiles by molecularly characterizing over 20,000 primary cancer and matched normal samples across 33 cancer types (The Cancer Genome Atlas Program (TCGA). Machine learning (ML) techniques have been instrumental in analyzing TCGA's comprehensive dataset, encompassing omics information across various cancer types, employing both supervised and unsupervised ML approaches to decipher complex patterns, aiding in the identification of novel biomarkers, understanding tumor heterogeneity, and predicting patient outcomes (Liñares-Blanco et al. 2021).

Circulating tumour DNA (ctDNA) analyses detect somatic variations in cancer and offer possibilities for early diagnosis, relapse monitoring, and treatment guidance (Bhinder et al. 2021). Various machine learning models have been utilized to analyse the biologically complex composition of cell free DNA (cfDNA) samples. These tools provided valuable information regarding disease subtyping, predicting chemotherapy response, early cancer detection, and detection of new biomarkers in different studies (Moser et al. 2023). A notable development in this area is the Lung-CLiP method, an ML-based approach that assesses the likelihood of ctDNA's presence in lung cancer patients' blood samples (Chabon et al. 2020). The method uses an elastic net model and an ensemble classifier with five algorithms, combining cfDNA mutation data and copy number scores to predict the presence of ctDNA. The model's predictive accuracy, with area under the curve (AUC) scores ranging from 0.69 to 0.98, differed across various cancer stages (Chabon et al. 2020). CancerSEEK, a blood test developed for early cancer detection, can identify eight common cancer types, including ovary, liver, stomach, pancreas, oesophagus, colorectum, lung, or breast (Cohen et al. 2018). In a study, a random forest-based classifier qualified with cfDNA fragment size features was shown to be able to predict the presence of ctDNA in blood in different cancer types with high accuracy (Bhinder et al. 2021). In another research involving 1005 patients, CancerSEEK demonstrated a median detection success of 70% for various cancer types, achieving sensitivity levels between 69 and 98% specifically for cancers (ovarian, liver, stomach, pancreatic, and oesophageal).

AI-driven platforms have changed how oncogenomics works by analysing complex datasets for gene mutations and identifying abnormal protein interactions that guide timely diagnosis, treatment, and control of cancer. For example, the CancerVar algorithm offers multiple query options at the variant, gene, and copy number change levels across 30 cancer types and two reference genome versions (Li et al. 2022a).

Interestingly, AI has even been used to accurately predict some key mutations from histopathological images (Bhinder et al. 2021). DeepPATH algorithm was used to classify lung cancer based on the TCGA criteria. Six key mutations (*TP53*, *KRAS*, *SETBPI*, *FAT1*, *EGFR and STK11*) associated with lung cancer (as detected by whole exomes)

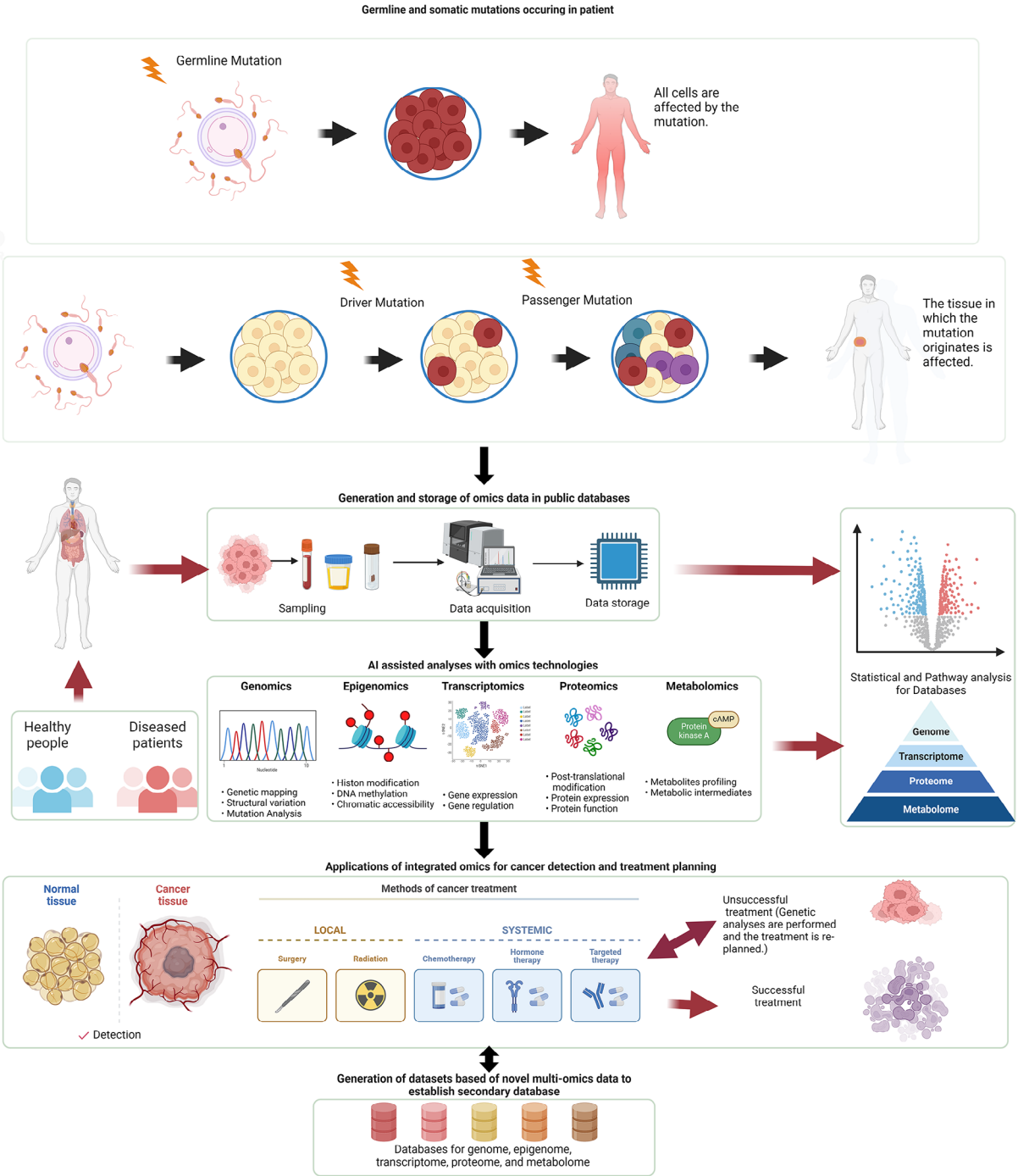

Fig. 3 A general flowchart of the integration of AI and multi-omics technologies in cancer research. The upper part shows the process of sampling healthy individuals and diseased patients, collecting data, and storing these data in public databases. Germline mutations are genetic changes found in the reproductive cells of an individual that can be passed on from generation to generation. Driver mutations are genetic changes that play a direct role in cancer development and cause uncontrolled growth of cells. Passenger mutations are genetic changes that are frequently found in cancer cells but do not directly contribute to the progression of cancer. AI-enabled analyses cover genomic, epigenomic, transcriptomic, proteomic and metabolomic data, allowing a better understanding of diseases. These analyses enable the data to be used for statistical and pathway analysis, thereby enabling the development of strategies for cancer detection and treatment. **In the middle section**, treatments related to cancer progression are included. Here, various treatment modalities such as local treatment (e.g. radiation), advanced treatment (e.g. Chimeric Antigen Receptor-T (CAR-T) cell therapy, chemotherapy (e.g. Doxotaxel) and metastasis as well as resistance to treatment and apoptosis (programmed cell death) of cancer cells are described. **In the lower part**, the generation of datasets based on new multi-omics data and to create secondary databases for the genome, epigenome, transcriptome, proteome and metabolome is described

were predicted directly from the whole slide imagess of 59 patients, with AUCs ranging from 0.73 to 0.85 (Coudray et al. 2018).

In the future, more advanced technologies, which shed light on the effects of mutations, could be used extensively in research and clinical settings (Dundar et al. 2019, 2024). For example, the effects of non-coding mutations on gene expression, epigenetic mechanisms and disease risk can be predicted by AI technology (Deeni et al. 2014; Cohen et al. 2018; Aguet et al. 2019).

Besides aiding mutation detection and diagnosis in cancer genomics, AI is also helpful making clinical decisions, determining prognosis and treatment strategies (Dlamini et al. 2020). Currently, DepMap data, which contains both cell line-specific and gene-specific data, is used by the ECLIPSE machine learning tool to predict cancer-specific drug targets (Bhinder et al. 2021). In a different study, DepMap data was analyzed and the proteomics data showed high predictive power in cancer cell line dependencies (Chabon et al. 2020). In another study, drug targets for liver cancer were predicted using a one-class support vector machine that utilised protein interaction networks and clinical data with gene expression profiles (AUC, 0.88) (Tong et al. 2019). In addition, the study by Lopez-Cortes et al., documented many different candidates that have the potential to become drug targets or biomarkers for breast cancer by using a breast cancer-specific deep learning-based classification approach to predict proteins associated with breast cancer by incorporating information from various databases such as TCGA, Cancer Genome Interpreter, PharmGKB and other databases (Klein et al. 2001; Ding et al. 2018; López-Cortés et al. 2018, 2020; Tamborero et al. 2018). Another active area of AI research is the search for repurposing candidates. In one study, similarities between diseases and drugs were integrated to predict different indications for drugs using the PREDICT computational pipeline (Gottlieb et al. 2011). The literature shows that PREDICT has successfully identified numerous new indications for common therapies. For example, progesterone can be used in the treatment of renal cell carcinoma (Bhinder et al. 2021).

Predicting the treatment responses in cancer patients has been another focus of AI research.

In a study conducted by Sun et al., DNN was trained with features obtained from copy-number alterations, gene expressions and clinical profiles of patients diagnosed with breast cancer to predict the prognosis of patients after treatment with different indications (Sun et al. 2019). Similarly, survival of patients with brain tumors was predicted based on pathway and gene expression profiles and for liver cancer based on methylation, miRNA and gene expression profiles using omics-based approaches with DNNs (Chaudhary et al. 2018; Hao et al. 2018). In a study by Liu et al., a logistic regression-based classifier trained on clinical features, treatment-free transcriptomes and genomic profiles were used to predict resistance to PD-1 inhibitors in patients with progressive melanoma (AUC, 0.73–0.83) (Liu et al. 2019).

Genome analysis and diagnosis of Rare Hereditary diseases

With the help of Next-generation sequencing methods lowering the cost and increasing the availability of exome and genome sequencing, massive amounts of patient data are continuously being generated. Clinical information, dysmorphic findings, laboratory and imaging results, and family history are evaluated together with each variant's specific features in a rigorous effort (Solomon et al. 2023). AI may play a crucial role in the diagnosis and management of genetic diseases as it can quickly detect the disease-causing genetic changes and match these with the presenting phenotypes.

There are important AI algorithms that evaluate the impact of variants preexisting in the human genome. The importance of these algorithms are reflected on the guidelines for the interpretation of sequence variants by American College of Medical Genetics and Genomics (ACMG) and the Association for Molecular Pathology (AMP). PP3 and BP4 criteria in these guidelines mention the evidence provided by computational tools for the interpretation of variants (Richards et al. 2015). Computational tools like Combined Annotation Dependent Depletion (CADD), REVEL, MutationTaster, PolyPhen-2, SpliceAI, and many others make it easier to predict the deleteriousness or functional impact of genetic variants (Adzhubei et al. 2010; Schwarz et al. 2014; Ioannidis et al. 2016; Brasil et al. 2019; Jaganathan et al. 2019; Schubach et al. 2024). Furthermore, AI-based applications may assist or even perform high-throughput sequencing data analysis as they may evaluate the clinical data of the patients the potential effect of each variant and the available literature. Some of these tools include Exomiser, Fabric GEM, PhenIX (Zemojtel et al. 2014; Smedley et al. 2015; De La Vega et al. 2021). A noteworthy limitation for the genomic variant analysis is that certain populations are poorly represented in population databases and the reference genome. A pangenomic approach could reflect the genomic diversity better. These developments could facilitate classifying new mutations in some populations and more accurate automated analyses (Wang et al. 2022; Liao et al. 2023).Unraveling the non-coding regions of the DNA is another significant challenge in genomics. DNABERT is a deep learning method excelling in the prediction of promoter regions, transcription factor binding sites, and splice sites, as well as identification of functional variants (Ji et al. 2021).

Three-dimensional genomic research

Comprehending the three-dimensional (3D) configuration of the genome and its impact on gene expression and cellular function became a highly exciting field (Mohanta et al. 2021; Chen et al. 2023). This area of research has the capacity to clarify the fundamental mechanisms of genetic illnesses and create novel treatments. Nevertheless, the intricacy of 3D genome mapping necessitates using sophisticated analysis methods to extract this information. AI and ML algorithms are utilised in this context [3]. These algorithms are specifically developed to address the difficulties encountered in the processing and interpreting genetic information. Advanced AI algorithms like DeepC, ChINN, Akita, Orca and DeepCTCFLoop have demonstrated remarkable performance when applied to 3D genome data (Zhang 2017; Yang and Ma 2022; Hovenga et al. 2023).

DeepC is a deep neural network that uses transfer learning to predict the folding of 3D genomes, from DNA sequences on a megabase scale, with high accuracy (Schwessinger et al. 2020). The objective is to analyse non-coding genetic differences within the framework of three-dimensional (3D) genome architecture and accurately forecast the impact of these changes on genome folding.

Conventional approaches to estimating the structure of 3D genomes either concentrate on making predictions at a detailed, base-pair level or deal with broader genomic settings without adequately incorporating these smaller scales (Whalen et al. 2016; Buckle et al. 2018). DeepC addresses this deficiency by integrating information across extensive genomic regions using high-resolution data, resulting in predictions of topological association domains (TAD) and their bounds.

The predictions made by DeepC have been verified and demonstrate strong concurrence with actual biophysical limits (Schwessinger et al. 2020).

Chromatin Interaction Neural Network (ChINN) is a machine learning methodology that uses DNA sequences to predict chromatin interactions (Cao et al. 2021). ChINN offers a significant advantage in situations where the production of comprehensive chromatin interaction data is limited by high costs and technological difficulties. ChINN has the ability to predict chromatin interactions associated with CCCTC-binding factor (CTCF) and RNA polymerase II, as well as interactions detected by Hi-C methods, demonstrating its flexibility to handle various types of chromatin interactions (Cao et al. 2021). This method is advantageous for investigating chromatin interactions in many cell types or situations without the need for distinct interaction datasets for each one.

The technique has been utilised on samples from patients with chronic lymphocytic leukemia (CLL) and has showcased its promise in biomedical research by uncovering extensive variability in chromatin connections among CLL samples (Rendeiro et al. 2016; Cao et al. 2021).

Akita tool uses a specialised type of AI called a convolutional neural network (CNN) specifically created to make accurate predictions about the DNA folding in three dimensions (3D) (Fudenberg et al. 2020). This technology is crucial for comprehending the three-dimensional structure of the genome and enables us to ascertain how the DNA sequence encodes a specific pattern of folding at a particular place. Akita has diverse uses, including in silico saturation mutagenesis, eQTL interpretation, structural variant prediction, and cross-species genome folding analysis (Aguet et al. 2019; Fudenberg et al. 2020).

Orca is a framework that utilises a sophisticated deep learning model to accurately forecast the complete three-dimensional (3D) structure of the genome, ranging from the individual DNA sequence to the kilobase to chromosomal size (Zhou 2022). Orca can detect and analyse different structures within cells, including chromatin compartments and TADs (Zhou 2022). It can also identify various interactions, such as those mediated by CTCF, enhancer-promoter contacts, and Polycomb-mediated interactions. At the level of submegabases, it is anticipated that distinct transcription factor motifs are responsible for cell type-specific genome connections. At the level of individual compartments, the virtual scan proposes the modelling of sequencing actions on chromatin compartments, with a particular emphasis on the significant function of transcription start sites (TSSs) (Zhou 2022).

DeepCTCFLoop is a sophisticated neural network model designed to identify the specific sequence patterns that drive CTCF-mediated chromatin interactions, which play a crucial role in the three-dimensional (3D) organisation of the human genome (Kuang and Wang 2021). While the significance of CTCF in chromatin interactions is well-established, the specific sequence patterns that govern the looping of CTCF motif pairs in chromatin remain uncertain. DeepCTCFLoop utilises the DNA sequences of near or tandem CTCF motifs and their surrounding regions to predict the formation of a chromatin loop. The findings demonstrate that DeepCTCFLoop effectively differentiates between CTCF motif pairs that create chromatin loops and those that do not. In addition, DeepCTCFLoop also uncovers intricate sequence patterns involved in the creation of CTCF-mediated chromatin loops. These findings indicate that specific patterns in some transcription factors, including ZNF384, ZNF263, ASCL1, SP1, and ZEB1, bind to DNA regions and may play a crucial role in the creation of chromatin loops (Kuang and Wang 2021).

Epigenetics

Epigenetics is defined as the study of hereditary changes in gene expression without a change in the DNA sequence (Wolffe and Guschin 2000). Complex interactions between an individual's genotype and lifestyle factors such as age, diet, alcohol consumption, and smoking influence epigenetic changes (Feil and Fraga 2012). Epigenetics involves three major inherited regulatory systems that determine chromatin remodelling and gene transcription regulation. These are DNA methylation, non-coding RNAs and chromatin remodelling (histone modifications and histone variants) (Al Aboud and Jialal 2018).

AI and machine learning (ML) are finding applications in epigenetics in a variety of diseases (Hull et al. 2023; Vinciguerra 2023). Deep neural network-based prediction models have been used to improve performance in diagnosing Alzheimer's disease (AD) using multi-omic datasets (Chen et al. 2022). Also, a machine learning model to improve Epigenome wide association studies (EWAS) was developed. This approach, named EWASplus, could discover new candidate genes and loci regarding AD (Huang et al. 2021).

Epigenetic complexity plays a major role in the aetiology of schizophrenia (SZ) (Vinciguerra 2023). One study demonstrated the potential of blood DNA methylation data as a marker for SZ risk prediction (Chen et al. 2020). AI/ML-based approaches have the potential to accurately classify SZ patients based on systemic epigenetic variants, surpassing traditional models based on polygenic risk scores (Gunasekara et al. 2021).

The challenges of cell type specificity of epigenetic data can be overcome by developing models that integrate gene expression and DNA methylation data to increase predictive accuracy for these diseases (Vinciguerra 2023). Using AI algorithms such as deep learning to identify epigenetic patterns in cancer cells has yielded successful results (Hull et al. 2023). These patterns can be used in critical decisions such as early diagnosis, disease prognosis, survival prediction and determination of treatment options (Hull et al. 2023).

For example, in a study by McMaster et al., CHICAGO, a convolution background model, was used to investigate key chromatin interactions in patients with Waldenström's macroglobulinemia and healthy controls in a GWAS consisting of two stages (McMaster et al. 2018).

Transcriptomics

RNA-based studies have been a vital intrument investigating human diseases at the molecular level. Transcriptomics is the study of all the RNA species produced by the organism. Microarrays and NGS methods for bulk RNA sequencing (RNA-seq) and more recently for single cell RNA-seq have been used to generate transcriptome data. In the last decades, these data have made significant impacts and contributions to our understanding of disease mechanisms and altered cellular pathways, as well as the discovery of new biomarkers and therapeutic targets (Byron et al. 2016; Giudice et al. 2021). The massive amount of data generated from RNA-seq led to the emergence of many AI tools in transcriptome analysis (Giudice et al. 2021; Li et al. 2022c). For RNA-seq analysis, numerous computational tools are available for data preprocessing steps like normalisation, data correction, clustering and cell annotations as well as downstream analyses (Erfanian et al. 2023).

Obtaining gene expression information at single cell level provides insights on heterogeneity between cells in a tissue or sample, however, single-cell RNA-seq techniques are expensive and not easily applicable to many routinely collected specimens. Cibersort and Cibersortx are machine learning based tools enabling cell type abundance and cell-type-specific gene expression estimations from bulk tissue transcriptomes (Newman et al. 2015, 2019).

With the huge leaps in single cell RNA-seq analyses, a new concept called spatial transcriptomics (ST) has emerged. Contrary to single cell RNA-seq, ST analysis methods provide spatial information. Adding the spatial context of the tissue to the transcriptomic data produces significant advancements in interpretation of cell type distribution, cell function and cell-cell interaction (Li et al. 2022c; Du et al. 2023). AI tools in single cell RNA-seq and/or ST include SEDR, SpaCell, Tangram, DEEPsc, GCNG, DeepST, and many others (Li et al. 2022c; Erfanian et al. 2023).

Besides their use in cancer and other common acquired disorders, RNA-seq analyses may also solve rare disease cases when routine testing approaches such as exome sequencing yield no diagnosis. Pre-mRNA splicing defects are significant causes of rare genetic diseases. Machine learning algorithms like SPANR, MMSplice, MTSplice, SpliceAI, and Pangolin have been designed to accurately predict the splice-altering variants (Wang et al. 2023).

Metabolomics

Metabolomics enables the study of metabolites at the body scale, investigating the changes in small molecules under the influences of diseases and various internal and external factors. AI technologies play an important role in analysing data in this field (Galal et al. 2022). As metabolomics data is complex, AI programs are increasingly being used to analyse, interpret and make predictions.

Coronary artery disease, diabetes mellitus type 2, infectious diseases like influenza and COVID-19, and various cancer types have been investigated at the metabolomic level using machine learning algorithms (Galal et al. 2022). AI contributes to the metabolome studies in various steps such as experiment design, data conversion, data pre-processing, data correction, and metabolite annotation (Coler et al. 2024).

For the deconvolution step of gas chromatography-mass spectrometry, a machine learning based tool called MSHub was developed (Aksenov et al. 2021).

The network inference technique approach extracts structural similarities between molecules based on spectral similarities and thus provides ease of analysis by enabling visualisation of inter-metabolite relationships and interactions (Aksenov et al. 2021). In particular, the MS2Mol algorithm, which increases the ability to accurately predict molecular structures from mass spectrometry data, provides great convenience to researchers in molecular structure predictions (Aksenov et al. 2021).

Liang and colleagues explored the variations in metabolite levels throughout pregnancy. The research group successfully developed a linear regression model that correlated certain plasma metabolites with gestational age, showing strong concordance with ultrasound findings (Liang et al. 2020).

Proteomics

Proteomics is the study of the structure, quantity, activity, modifications, localisation, and interactions of all proteins in the genome. Data obtained from proteomic research are complex and multidimensional. Taking advantage of rapid developments in bioinformatics, AI-supported data analysis programmes are being developed to analyse these complex datasets (Fig. 4). One of the most prominent developments of proteomic research using AI occurred with AlphaFold, a highly accurate neural network-based model predicting the structures of proteins (Jumper et al. 2021).

Mund and colleagues have integrated sub-micron resolution imaging, AI-based single-cell phenotyping for image analysis, and ultra-sensitive proteomic workflow to analyse proteomes in their native cellular context. To accurately identify single-cell boundaries and cell classes and convert them into proteomic samples, they developed an AI-based software called "BIAS" that coordinates scanning and laser microdissection microscopes. This software seamlessly combines processes such as image analysis, cell segmentation, and cell type identification. This approach allows for obtaining molecular insights at the phenotypic level in proteomic analysis and preserving complete spatial information, thereby enabling a deeper understanding of cellular diversity and organisation (Mund et al. 2022).

In mass spectrometry-based proteomics, the DeepNovo algorithm has been developed to de novo sequence peptide segments, while the pDeep algorithm predicts the secondary spectra of peptide segments. Additionally, in liquid

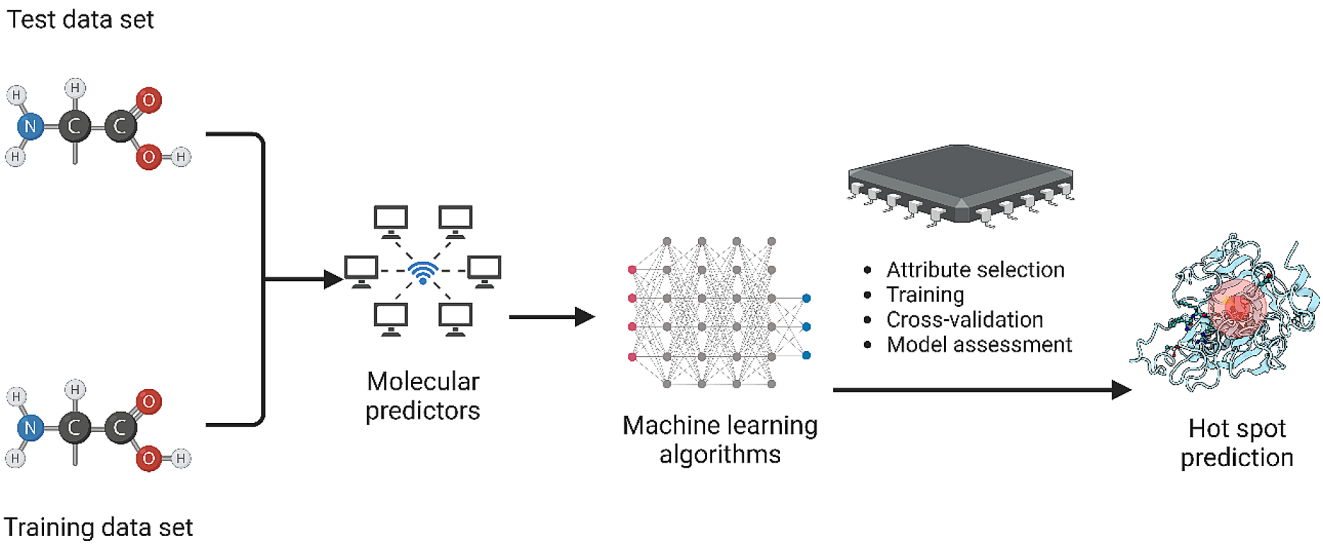

Fig. 4 A process in which artificial intelligence algorithms are used to predict hot spots within protein-protein interactions. The flowchart in the image shows that molecular predictors are obtained from both test and training datasets. These predictors are fed into machine learning algorithms and then processed through a process that involves steps such as feature selection, training, cross-validation and model evaluation of the algorithm. In the final stage, the learnt model is used to predict hotspots in the interaction domains between proteins. These predictions are important for identifying biologically important areas in the functional regions of protein structures. Hotspots often play key roles in protein interactions and are of strategic importance in drug design, disease understanding and biomolecular research

chromatography-mass spectrometry tandem analysis, algorithms based on deep learning methods, such as DeepRT, have been developed for predicting the retention time of peptides. Similarly, algorithms like DeepIso, based on convolutional neural networks (CNN), have been used for extracting peptide chromatographic and mass spectrometric features. They have also developed deep learning-based approaches, such as DeepBind, for identifying nucleic acid-binding proteins. The combination of proteomic techniques like mass spectrometry (MS) and machine learning is widely used for biomarker screening (Li et al. 2022b).

Martinez and colleagues have developed immunoinformatics and AI-based tools called PoxiPred to design multi-epitope vaccines against newly emerging infectious diseases, and they have reported that this tool successfully identified 3191 antigen proteins from 25 different poxviruses (Martinez et al. 2024).

AI-assisted algorithms have been found to be effective in deriving the biomedical significance and applications of human pluripotent stem cell (hPSC)-derived organoids, facilitating the development of new antibody therapeutics for adaptive immune response applications (Maramraju et al. 2024; Townsend et al. 2024).

Bhatia and colleagues have developed a method to characterise the proteome of samples isolated from transparent tissues by combining ultra-high sensitivity mass spectrometry-based proteomic analysis with whole-organ and tissue clearing and imaging procedures. Named DISCO-MS, this method enabled the exploration of spatial-molecular profiles of initiating pathological events in various disease models by successfully analysing small target regions isolated from mouse or human organs (Bhatia et al. 2022).

Shahbazy and colleagues introduced an interactive and fully unsupervised data visualisation and cluster analysis tool called MHCpLogics. They used this tool to analyse large and complex datasets of HLA peptides derived from cells expressing various alleles. MHCpLogics is used for parsing, querying, and analysing immunopeptidomes. It is compatible with different MS data processing and computational software tools and can handle large datasets without the need for preprocessing steps, filtering them using different approaches and performing various analyses. Users can manage and utilise various scoring functions depending on research purposes and data types (Shahbazy et al. 2024).

Monsalve and colleagues highlight the presence of antibiotic-resistance genes in ancient soil samples alongside the antibiotics developed based on compounds produced by environmental microbes. By employing in silico strategies to discover new antimicrobial peptides in proteomes, they evaluated the proteomes of eight organisms and obtained 513,909 peptides. Nine peptides with the highest antimicrobial scores were selected and validated using machine learning algorithms and laboratory analyses. These peptides were effective against bacteria, fungi, and cancer cell lines (Monsalve et al. 2024).

Pharmacogenomics

Pharmacogenomics aims to determine inter-individual DNA variations that are associated with drug response. Genetic variations can determine whether a drug is safe for one person but can be harmful to another (Fig. 5). The overall frequency of hospitalizations due to adverse drug reactions is around 6.5% in adults (Pirmohamed et al. 2004) and exceeds 15% in patients with multiple comorbidities (Osanlou et al. 2022; Pirmohamed 2023). It is important that adverse drug reactions can be eliminated through pharmacogenetics by stratifying patients who are at risk in order to apply personalized medicine. For example, a drug can be tailored based on a person's unique genetic makeup.

In the field of pharmacogenetics, supervised machine learning (SML) techniques are more frequenty uesd than unsupervised machine learning (UML) methods, as they perform better in predicting drug response (Cilluffo et al. 2021a). SML methods have been used to analyse the response to anti-cancer drugs. In this context, random forest (RF), SVM and elastic-net (EN) methods showed high accuracy, transferability and generalizability (Ding et al. 2018; Kim et al. 2019; Su et al. 2019).

Machine learning mehods have also been used to study treatments not related with cancer. In a study conducted to evaluate the pharmacogenetics of antidepressant response, different supervised techniques such as random forests, gradient boosted machine, neural network (NN), learning vector quantization and recursive partitioning were compared. The analyzed data included 671 adult patients from three European studies on major depression. When comparing all models, the NN showed the highest accuracy (Fabbri et al. 2018). In another study involving 186 patients with major depression, the performance of support vector machine (SVM) and regression tree (RT) was compared and the response to antidepressants was predicted. The results of the study showed that SVM had the highest accuracy in predicting response to antidepressants (Maciukiewicz et al. 2018).

Since the main goal of pharmacogenetics is to predict drug response, few studies have reported the results of UML techniques. For instance, a study that aimed to analyze the association between clinical parameters and genetic polymorphisms on treatment outcome in treatment-resistant depression using a two-stage ML approach. In this study, first, the RF algorithm was used to analyze the patients. Second, a cluster analysis was performed to group the data. Five clusters of patients who responded significantly to treatment

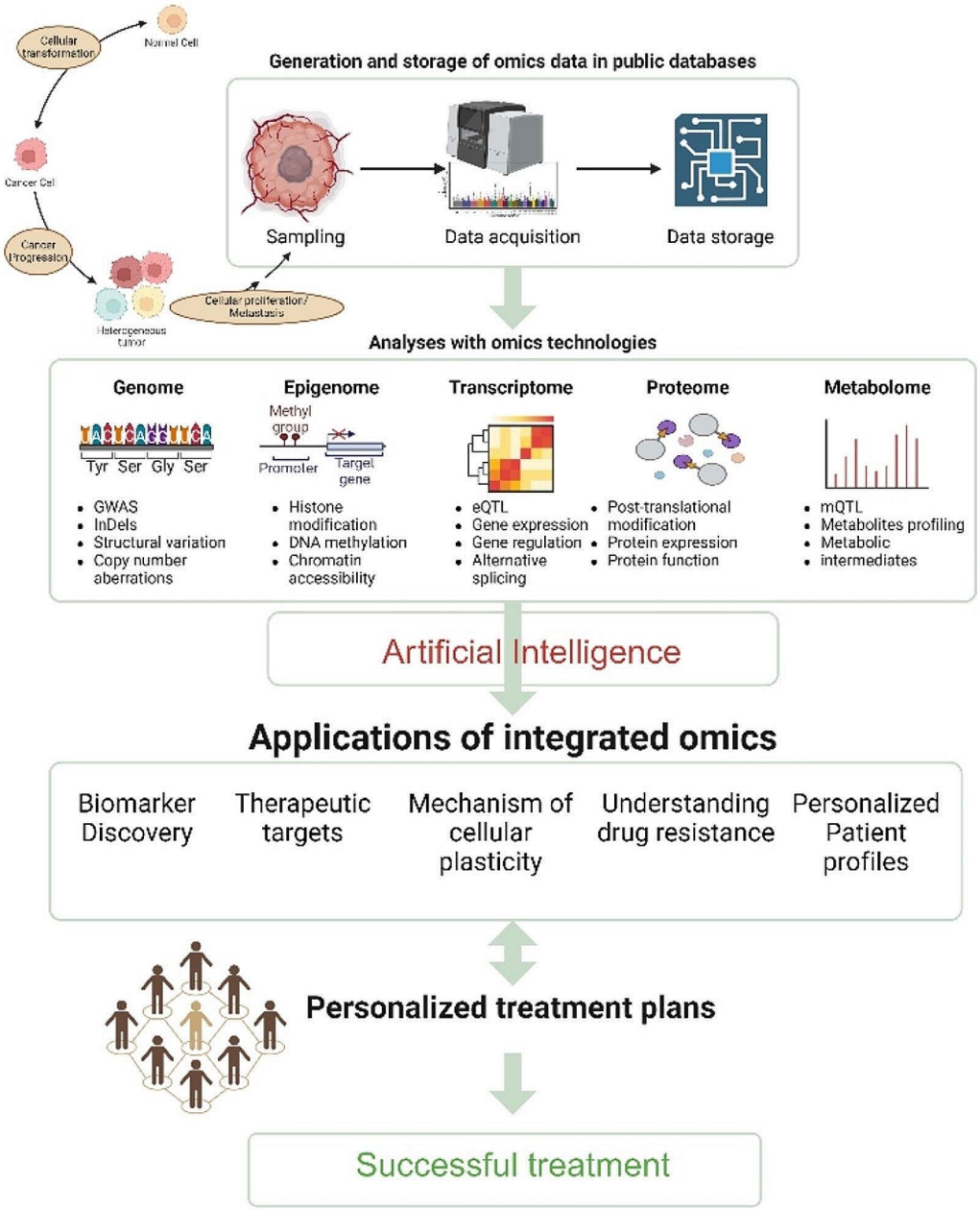

Fig. 5 Use of integrated omics technologies and artificial intelligence (AI) for personalised cancer therapy. The process of transformation of normal cells into cancer cells and the heterogeneous character of cancer progression in this process are emphasised. Following sampling, the obtained data are stored in public databases. Under the analysis of omics technologies, different layers of the genome, epigenome, transcriptome, proteome and metabolome are described. These analyses include components such as large-scale combined genomic studies (GWAS), structural variations, histone modifications, DNA methylation, gene expression, protein function and metabolite profiling. Artificial intelligence is a fundamental tool for the integration and analysis of this multi-omics data and applications include biomarker discovery, identification of therapeutic targets, mechanisms of cellular plasticity, understanding mechanisms of drug resistance and personalised patient profiles. At the bottom, a model of how personalised treatment plans lead to successful treatment as a result of this integrated approach is shown

were identified by the cluster analysis (Kautzky et al. 2015; Cilluffo et al. 2021b).

Treatment of genetic and/or Rare diseases

Drug design and discovery

Treatment of rare diseases is a major challenge due to their mostly genetic or degenerative etiology, a smaller number of patients needing treatment, and unsatisfactory responses (Visibelli et al. 2023). These challenges led treatments for these diseases to be named "orphan drugs". So far, approximately 5% of rare diseases have a treatment. AI has great potential for the treatment of rare diseases, as encouraging results have been achieved in the treatment of common diseases (Brasil et al. 2019; Lee et al. 2019; Visibelli et al. 2023). AI may facilitate different steps in treatment optimization and drug discovery, such as biomarker discovery, drug design and high-throughput drug screening (Boniolo et al. 2021). Clinical trials (CTs) play a crucial role in drug development. Nevertheless, clinical trials only exist for a small number of rare diseases due to difficulties in patient recruitment, data collection and analysis in small populations, and the number of reliable biomarkers. Importantly, AI can provide solutions to the problems associated with CTs in RDs (Crow et al. 2018; Day et al. 2018; Logviss et al. 2018; Wojtara et al. 2023). One of the AI algorithms enabling drug design and optimisation is CADD, as mentioned earlier. In the case of neurological diseases such as Spinal Muscular Atrophy, where a single drug is not sufficient for treatment, CADD reduces the workload by shortening the drug testing process by reducing the number of molecules tested (Chong et al. 2021).

Another example of AI-guided drug discovery process is related with splicing defects, a pathophysiologic mechanism caused by pathogenic variations in genes leading to altered pre-mRNA splicing, disrupted or dysregulated protein expression and disease phenotypes. An ML approach identified that BPN-15,477, a splicing modulator compound that can correct the splicing of *ELP1* exon 20, could also correct certain splicing defects in *CFTR*, *LIPA*, *MLH1*, and *MAPT* (Gao et al. 2021).

Repurposing of drugs

Pharmaceutical companies struggle to recoup investments in rare disease drugs due to their limited patient population. This raises ethical concerns, as the society may need to approve public reimbursement for high-cost drugs to provide affordable treatment for a small number of patients. Repurposing drugs can reduce costs for rare disease development by identifying new uses for existing ones. Exploring previously approved drugs can save time and money. AI can analyse drug activity data to identify new applications in treating rare diseases. Repurposing approaches often benefit from data enriched with omics information to gain insights into target mechanisms and ensure candidate safety and efficiency (Alves et al. 2022).

An excellent example of AI being used in drug repurposing occurred during the COVID-19 pandemic. In early 2020, BenevolentAI suggested Janus kinase (JAK) inhibitor baricitinib as a combination anti-inflammatory and antiviral for treating hospitalised COVID-19 patients, which later proved to be effective in clinical studies (Kalil et al. 2021; Richardson et al. 2023).

Considering the mechanistic similarities and similar processes between rare and common diseases, URSAHD (unveiling RNA sample annotation for human diseases) has predicted that cisplatin could be used to treat refractory anemia characterized by excess blasts. In addition, resveratrol has been shown to be a candidate for the treatment of sideroblastic anemia. It was noted that the algorithms used for prediction in both studies need to be experimentally tested and validated (Brasil et al. 2019; Lee et al. 2019). Another tool developed for drug repurposing for genetic diseases is mediKanren, which was able to recommend treatments for *TMLHE* and *RHOBTB2* associated diseases with anecdotal improvement, necessitating further studies and clinical trials (Foksinska et al. 2022).

Population genetics, evolution and archeogenetics

Studying human (sub)-population genetic structure within spatial and temporal dimensions has always been a valuable source of knowledge on human microevolutionary changes and the distribution of human variation. Large genomic datasets available from the human populations of various historical periods allow studies on human migration patterns, but also the origins of mutations and genetic disease progeny and distribution. A vast amount of human genome data ensures a transition from probabilistic to data-driven theories of evolutionary processes. Furthermore, exact data enables the development of more suitable statistical approaches and computational algorithms (Korfmann et al. 2023).

The most immanent characteristic of humankind is its huge genetic heterogeneity coming from the admixture of a certain number of local ancestries. This fact affects one's ability to easily explore causative factors or identify, i.e. disease-causing mechanisms (mutations) using direct genetic association studies. Deep learning algorithms and neural network analysis approaches can help utilise greater statistical power for discovery in representative multicentric

datasets from populations of various ancestries in a quest for disease-associated variants (Mester et al. 2023).

Medical microbiology, zoonotic disease research and its genetic implications

AI has become a pivotal tool in zoonotic disease research, enabling a nuanced analysis of pathogens capable of crossing species barriers and their transmission dynamics. Recent studies underscore the utility of AI in elucidating the complex interactions between environmental changes, such as agricultural intensification and biodiversity loss, and the emergence of zoonotic diseases, which are significantly influenced by factors like host genetic diversity and human population densities (Jones et al. 2013; Keesing and Ostfeld 2021; Pillai et al. 2022).

The genomic studies of zoonotic pathogens like Giardiasis and avian influenza have been instrumental in revealing genetic adaptations that may either facilitate or inhibit cross-species transmission, highlighting the potential for AI to contribute to the development of targeted interventions and vaccines (Yaoyu and Xiao 2011; Horman et al. 2018; Pillai et al. 2022).

Moreover, AI's capacity to analyse genetic diversity within pathogens can provide insights into the mechanisms underlying disease virulence and spread, thereby informing public health strategies aimed at mitigating zoonotic risks (Yaoyu and Xiao 2011; Chu et al. 2018; Horman et al. 2018; Pillai et al. 2022).

Clinical microbiomics

AI is significantly enhancing the field of clinical microbiomics, which focuses on understanding microbial communities in the human body and their impact on health and disease. Disturbances in the delicate balance of the microbiome are known as dysbiosis and contribute to a variety of diseases, including metabolic disorders, autoimmune conditions, neurological diseases, and even some cancers. The complexity of microbial data, including the diversity of species, their interactions, and their influence on the host's genetic makeup, poses substantial challenges for researchers. AI and machine learning technologies are instrumental in navigating these challenges by efficiently analysing vast microbiomic datasets. These technologies help identify patterns, predict disease associations, and provide new insights into the microbiome's role in human health (Rhoads 2020).

Specific applications of AI in clinical microbiomics include predictive modelling for disease risk based on microbiome composition, personalised medicine approaches tailored to an individual's microbiome, and AI-driven drug discovery targeting microbiome-related pathways (Smith et al. 2020). The integration of AI in this field has the potential to transform healthcare by improving diagnostic accuracy, enabling personalised treatment plans, and facilitating the development of novel therapeutics targeting the microbiome. Some of the popular disorders and topics in this field are Irritable Bowel Syndrome, obesity, cancer, personalised nutrition, senescence, and probiotic production.

The gut-brain axis, a bi-directional communication pathway between the gut and the brain, is an exciting frontier of AI-driven microbiome research. Studies suggest that AI-based analysis of microbiome-related BBB dysfunction may enable early detection of neurological diseases such as Alzheimer's or Parkinson's and reveal potential therapeutic targets (Zhu et al. 2020); Kenanoglu et al. 2022).

Future directions for AI in clinical microbiomics include ongoing research, potential breakthroughs, and the importance of integrating AI with microbiomic research to advance personalised medicine and public health, echoing its transformative impact on genetic research and disease management (Zeng et al. 2021).

In this evolving landscape, the development of online tools such as proteome2pathogen.com and its extension, proteome2virus, represent a significant leap forward (Balvers et al. 2023). These platforms leverage AI to refine the process of pathogen identification from proteomic data, directly impacting the study of microbial communities' role in health and disease. By enabling the precise identification of bacteria and viruses down to the species level, these tools facilitate a deeper understanding of the microbial contributions to infectious diseases (Balvers et al. 2023).

Synthetic biology and, gene editing

Synthetic biology (Synbio) aims to develop biological systems that perform a specific task (e.g. cells that produce a certain amount of biofuel or respond specifically to external stimuli) (Cameron et al. 2014). To achieve these goals, engineering design principles are required to control complicated biological systems. In particular, these engineering principles consist of the Design-Build-Test-Learn (DBTL) cycle and the standardized genetic parts used to achieve a desired outcome. Synbio has had a revolutionary impact on materials, medicine, food, energy, and climate (Bilitchenko et al. 2011). Anticancer and antimalaria drugs, the production of insulin, and renewable biofuels are a few examples of Synbio's impact (Eslami et al. 2022).

An essential example of artificial intelligence use in synthetic biology is mRNA therapeutics (Castillo-Hair and Seelig 2022; Hınçer et al. 2023). mRNA therapeutics have a wide range of applications, such as vaccines for infectious diseases, cancer immunotherapy, and regenerative medicine. Machine learning algorithms are used to develep predictive

models and sequence design algorithms (Castillo-Hair and Seelig 2022).

Gene editing technologies are methods developed to make specific changes to DNA sequences, and AI is used in various ways in these processes (Fig. 6). The most frequently used genome editing technology is Clustered regularly interspaced short palindromic repeat (CRISPR)-CRISPR-associated proteins (Cas) system (149). The predictions of artificial intelligence algorithms, which are often used to predict the activity of guide RNA (gRNA) in CRISPR-Cas systems, are key factors in determining the efficiency and specificity of gene-editing processes (Lee 2023). Deep learning models can predict both on-target and off-target effects by modelling the interactions of the gRNA with the target DNA sequence. These models have become indispensable tools for accelerating the gRNA design and optimisation process. The principal artificial intelligence programs employed in this domain include DeepCRISPR (Chuai et al. 2018), CRISPRLearner (Dimauro et al. 2019), and CRISPR-GA (Güell et al. 2014).

Limitations

At the current stage, one of the leading limitations and concerns is trust. Many AI models lack transparency and the reasoning behind their decisions are not always known. This issue is referred as the "black box" of AI. In clinically relevant conditions, wrong decisions caused by this opaqueness becomes a major threat. Due to this concern, interpretable and explainable AI models have been developed (Karim et al. 2023). These methods are already being used in genomics and personalized medicine, however, there is still room for improvement (He et al. 2024). Also, constructing better models and increasing repeatability is an unmet need (He et

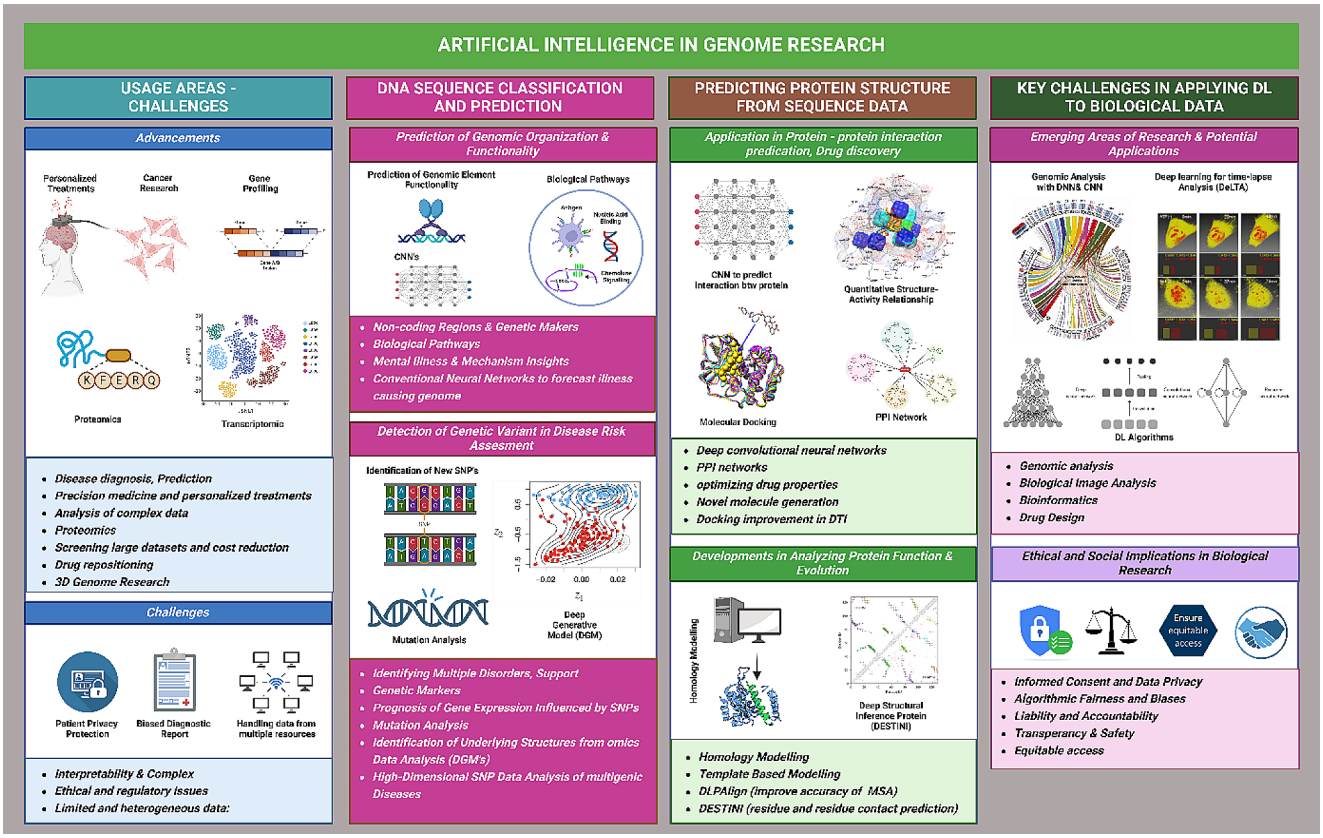

Fig. 6 The use of artificial intelligence (AI) in genome research. Under the title "Usage Areas - Challenges", the advantages offered by AI in areas such as personalised treatments, cancer research, gene profiling, proteomics and transcriptomics are shown. At the same time, issues such as disease diagnosis, precision medicine and personalised treatments, complex data analysis and cost reduction were particularly emphasised. In the "DNA Sequence Classification and Prediction" section, topics such as genomic organisation and functionality predictions, biological pathways, determination of the role of genetic variants in disease risk, and techniques such as artificial neural networks and convolutional neural networks (CNNs) used in this process are at the forefront. In the "Predicting Protein Structure from Sequence Data" section, AI techniques used in the prediction of protein structures and protein-protein interactions, molecular binding, and developments for drug discovery are shown. In particular, deep learning-based artificial neural networks, protein-protein interaction networks (PPI), new molecule generation, and improvements in drug design are focused. In the "Key Challenges in Applying DL to Biological Data" section, the ethical and social implications of applying DL to fields such as genomic analysis, biological image analysis, bioinformatics and drug design are presented

al. 2023). Even when these issues are correctly addressed, the quality and consistency of the obtained data and the researchers' knowledge of the clinical characteristics of the study population may influence the applicability of AI models in molecular studies (Gomes and Ashley 2023). To improve the data quality and increase the representation of different populations, different researchers and institutions could combine their data in rare diseases and molecular medicine. Another issue regarding the interpretation of AI-based tools' results is the need for the experimental validation before reaching definitive conclusions (Brasil et al. 2019). Furthermore, data regulation, possible ethical challenges, and data and machine bias are other significant concerns (Dias and Torkamani 2019).

Because AI tools are being incorporated in real life decisions for human subjects, the aforementioned limitations are major issues. Some of the AI tools discussed in the manuscript are yet to be supported with extensive clinical studies proving their reliability. This concern must be kept in mind while using these tools to make clinically impactful decisions and giving genetic counselling.

The future of AI in genomics

Artificial intelligence is and will be one of the leading concepts that transform medicine and healthcare to a more personalised and precise approach (Xu et al. 2019). The massive and difficult-to-analyse data generated by single-cell sequencing, multi-omics and spatial transcriptomics approaches will keep AI an undispensable tool in genomic diagnostics and research (Brlek et al. 2024). The increase in the generated data will lead to more biobanks and databases with more comprehensive and detailed information. The power of machine learning could be immense integrating different datasets and also multi-omics data. Integrating and translating multi-omics data into actionable insights could be a giant leap forward for preventive and predictive strategies, disease monitoring and subtyping, and developing new treatments (Kang et al. 2022). We can predict that it may be necessary for genetic professionals to adapt to a more automated genomic diagnostic process. AI will play an expanding role in drug discovery and in drug repurposing providing novel treatment strategies for rare disorders by analysing multi-omic databases (Boniolo et al. 2021).

The potential of combining AI with genetic engineering and gene editing is vast and has the capacity to revolutionise the treatment of hereditary and acquired disorders. Furthermore, AI can effectively combine genetic data and patient health records to forecast hereditary risks and pharmacological reactions for individuals with greater accuracy. All the aforementioned developments could lead to improved personalised treatment strategies. AI excels in tasks that are repetitive or require some sort of classification. AI tools were thought to lack the skills to adapt to different situations that are encountered in clinical settings, as well as empathy (Greatbatch et al. 2019; Kearney et al. 2020). However, more recent studies have shown promising results. For example, a recent study reported that chatbots performed better than physicians in terms of both quality and empathy (Ayers et al. 2023). Therefore, we may expect AI tools to be more incorporated into genomic field, helping geneticists, other clinicians, and researchers in their routine, without replacing them.

Conclusion

To conclude, artificial intelligence and its applications are rapidly being incorporated in molecular diagnostics and research across many clinical disciplines. Even though patient care and physical examination are indispensable features of clinical practice, all clinicians, geneticists and other health professionals must be aware of and ready to benefit from the current and potential impacts of AI technologies in this highly dynamic field. Our review humbly portraits the advancements, potentials and limitations of the AI platforms and tools in a wide variety of molecular research fields without solely focusing on a specific discipline. We believe this rather unique approach contributes to the existing literature in this exciting and popular subject that is revolutionizing medicine.

Abbreviations

ACMG	American College of Medical Genetics and Genomics
AD	Alzheimer's Disease
AI	Artificial Intelligence
AMP	Association for Molecular Pathology
AUC	Area Under the Curve
CADD	Combined Annotation Dependent Depletion
CAR-T	Chimeric Antigen Receptor-T cells
cfDNA	Cell-Free DNA
ChINN	Chromatin Interaction Neural Network
CLL	Chronic Lymphocytic Leukemia
CNNs	Convolutional Neural NetworksX
CNVs	Copy Number Variations
CRISPR	Clustered Regularly İnterspaced Short Palindromic Repeat
CTCF	CCCTC-binding factor
ctDNA	Circulating Tumour DNA
DBTL	Design-Build-Test-Learn
DL	Deep Learning
eQTL	Expression Quantitative Trait Loci
EWAS	Epigenome Wide Association Study

FFPE	Formalin-Fixed Paraffin-Embedded
GWAS	Large-Scale Combined Genomic Studies
HLA	Human Leukocyte Antigens
HMES	Human Embryonic Stem Cells
HPO	Human Phenotype Ontology
IRD	Inherited Retinal Disorders
JAK	Janus Kinase
ML	Machine Learning
MS	Mass Spectrometry
NGS	Next-Generation Sequencing
PWS	Prader-Willi Syndrome
REVEL	Rare Exome Variant Ensemble Learner
RNA-seq	RNA Sequencing
SNVs	Single Nucleotide Variations
SRS	Silver-Russell Syndrome
SVM	Support Vector Machine
SZ	Schizophrenia
TAD	Topologically Associating Domain
TCGA	The Cancer Genome Atlas
TSSs	Transcription Start Sites
UML	Unsupervised Machine Learning
WES	Whole Exome Sequencing
WGS	Whole-Genome Sequencing

Author contributions MSD, FO, ABY wrote the article. GH, OV, JASA, NG, DTY, NNB, DPK, MB, LP, ME, MK, IOS, OKGT, BT, DM, RM, SP, MY, TB, RL, SGT, IF, MCE, AM, MS, MM, DD, QS, LN, DM, DT, AJ, JB, KSC, IR, CO and OUN did the necessary research for the article. SOS and KK created figures and tables. IB and RA edited the article. HA and MD controlled and coordinated the whole process. All authors reviewed the manuscript.

Funding The authors received no financial support for the research of this article.

Data availability No datasets were generated or analysed during the current study.

Declarations

Competing interests The authors declare no competing interests.

References

Abdallah S, Sharifa M, ALMADHOUN IKH, MK et al (2023) The impact of Artificial Intelligence on optimizing diagnosis and treatment plans for Rare Genetic disorders. Cureus 15. https://doi.org/10.7759/cureus.46860

Adzhubei IA, Schmidt S, Peshkin L et al (2010) A method and server for predicting damaging missense mutations. Nat Methods 7

Aguet F, Barbeira AN, Bonazzola R et al (2019) The GTEx Consortium atlas of genetic regulatory effects across human tissues. bioRxiv 787903. https://doi.org/10.1101/787903

Aksenov AA, Laponogov I, Zhang Z et al (2021) Auto-deconvolution and molecular networking of gas chromatography–mass spectrometry data. Nat Biotechnol 39:169–173. https://doi.org/10.1038/s41587-020-0700-3

Al Aboud NM, Jialal I (2018) Genetics, Epigenetic Mechanism

Alves VM, Korn D, Pervitsky V et al (2022) Knowledge-based approaches to drug discovery for rare diseases. Drug Discov Today 27:490–502

Athanasopoulou K, Boti MA, Adamopoulos PG et al (2022) Third-generation sequencing: the spearhead towards the radical transformation of modern genomics. Life 12. https://doi.org/10.3390/life12010030

Ayers JW, Poliak A, Dredze M et al (2023) Comparing physician and Artificial Intelligence Chatbot responses to patient questions posted to a Public Social Media Forum. JAMA Intern Med 183:589–596. https://doi.org/10.1001/jamainternmed.2023.1838

Balvers M, Gordijn IF, Voskamp-Visser IAI et al (2023) Proteome2virus: shotgun mass spectrometry data analysis pipeline for virus identification. J Clin Virol Plus 3:100147. https://doi.org/10.1016/j.jcvp.2023.100147

Bannister JJ, Crites SR, Aponte JD et al (2020) Fully automatic landmarking of syndromic 3D facial surface scans using 2D images. Sens (Switzerland) 20:1–14. https://doi.org/10.3390/s20113171

Basel D (2020) Dysmorphology in a genomic era. Clin Perinatol 47:15–23

Bhatia HS, Brunner AD, Öztürk F et al (2022) Spatial proteomics in three-dimensional intact specimens. Cell 185:5040–5058. .e19

Bhinder B, Gilvary C, Madhukar NS, Elemento O (2021) Artifi Cial intelligence in cancer research and precision medicine. Cancer Discov 11:900–915. https://doi.org/10.1158/2159-8290.CD-21-0090

Bilitchenko L, Liu A, Densmore D (2011) The Eugene language for synthetic biology. In: Methods in Enzymology. pp 153–172

Boniolo F, Dorigatti E, Ohnmacht AJ et al (2021) Artificial intelligence in early drug discovery enabling precision medicine. Expert Opin Drug Discov 16:991–1007

Brasil S, Pascoal C, Francisco R et al (2019) Artificial intelligence (AI) in rare diseases: Is the future brighter? Genes (Basel). 10

Brlek P, Bulić L, Bračić M et al (2024) Implementing whole genome sequencing (WGS) in clinical practice: advantages, challenges, and future perspectives. Cells 13. https://doi.org/10.3390/cells13060504

Buckle A, Brackley CA, Boyle S et al (2018) Polymer simulations of heteromorphic chromatin predict the 3D folding of Complex genomic loci. Mol Cell 72:786–797e11. https://doi.org/10.1016/j.molcel.2018.09.016

Byron SA, Van Keuren-Jensen KR, Engelthaler DM et al (2016) Translating RNA sequencing into clinical diagnostics: opportunities and challenges. Nat Rev Genet 17

Cameron DE, Bashor CJ, Collins JJ (2014) A brief history of synthetic biology. Nat Rev Microbiol 12:381–390

Cao F, Zhang Y, Cai Y et al (2021) Chromatin interaction neural network (ChINN): a machine learning-based method for predicting chromatin interactions from DNA sequences. Genome Biol 22:1–25. https://doi.org/10.1186/s13059-021-02453-5

Carrer A, Romaniello MG, Calderara ML et al (2023) Application of the Face2Gene tool in an Italian dysmorphological pediatric clinic: retrospective validation and future perspectives. Am J Med Genet Part A 194. https://doi.org/10.1002/ajmg.a.63459

Castillo-Hair SM, Seelig G (2022) Machine learning for Designing Next-Generation mRNA therapeutics. Acc Chem Res 55:24–34. https://doi.org/10.1021/acs.accounts.1c00621

Cembrowska-Lech D, Krzemińska A, Miller T et al (2023) An Integrated Multi-omics and Artificial Intelligence Framework for Advance Plant phenotyping in horticulture. Biology (Basel) 12. https://doi.org/10.3390/biology12101298

Chabon JJ, Hamilton EG, Kurtz DM et al (2020) Integrating genomic features for non-invasive early lung cancer detection. Nature 580. https://doi.org/10.1038/s41586-020-2140-0

Chaudhary K, Poirion OB, Lu L, Garmire LX (2018) Deep learning–based multi-omics integration robustly predicts survival in liver cancer. Clin Cancer Res 24. https://doi.org/10.1158/1078-0432.CCR-17-0853

Chen J, Zang Z, Braun U et al (2020) Association of a reproducible Epigenetic Risk Profile for Schizophrenia with Brain methylation and function. JAMA Psychiatry 77. https://doi.org/10.1001/jamapsychiatry.2019.4792

Chen L, Saykin AJ, Yao B, Zhao F (2022) Multi-task deep autoencoder to predict Alzheimer's disease progression using temporal DNA methylation data in peripheral blood. Comput Struct Biotechnol J 20. https://doi.org/10.1016/j.csbj.2022.10.016

Chen M, Liu X, Liu Q et al (2023) 3D genomics and its applications in precision medicine. Cell Mol Biol Lett 28:1–26

Chong LC, Gandhi G, Lee JM et al (2021) Drug discovery of spinal muscular atrophy (sma) in computational perspective: a comprehensive review. Int J Mol Sci 22

Chu DKW, Hui KPY, Perera RAPM et al (2018) MERS coronaviruses from camels in Africa exhibit region-dependent genetic diversity. Proc Natl Acad Sci U S A 115:3144–3149. https://doi.org/10.1073/pnas.1718769115

Chuai G, Ma H, Yan J et al (2018) DeepCRISPR: optimized CRISPR guide RNA design by deep learning. Genome Biol. https://doi.org/10.1186/s13059-018-1459-4. 19:

Ciancia S, Goedegebuure WJ, Grootjen LN et al (2023) Computer-aided facial analysis as a tool to identify patients with silver–Russell syndrome and prader–Willi syndrome. Eur J Pediatr 182:2607–2614. https://doi.org/10.1007/s00431-023-04937-x

Cilluffo G, Fasola S, Ferrante G et al (2021a) Machine learning: an overview and applications in pharmacogenetics. Genes (Basel) 12:1511

Cilluffo G, Fasola S, Ferrante G et al (2021b) Machine learning: an overview and applications in pharmacogenetics. Genes (Basel) 12:1–12. https://doi.org/10.3390/genes12101511

Cohen JD, Li L, Wang Y et al (2018) Detection and localization of surgically resectable cancers with a multi-analyte blood test. Science (80-) 359:. https://doi.org/10.1126/science.aar3247

Coler EA, Chen W, Melnik AV et al (2024) Metabolomics in the era of artificial intelligence. https://doi.org/10.1530/mah-23-0017. Microbiota Host 2:

Coudray N, Ocampo PS, Sakellaropoulos T et al (2018) Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning. Nat Med 24. https://doi.org/10.1038/s41591-018-0177-5

Crow RA, Hart KA, McDermott MP et al (2018) A checklist for clinical trials in rare disease: obstacles and anticipatory actions-lessons learned from the FOR-DMD trial. Trials 19

Davenport T, Kalakota R (2019) The potential for artificial intelligence in healthcare. Futur Healthc J 6:94–98. https://doi.org/10.7861/futurehosp.6-2-94

Day S, Jonker AH, Lau LPL et al (2018) Recommendations for the design of small population clinical trials. Orphanet J Rare Dis 13. https://doi.org/10.1186/s13023-018-0931-2

De La Vega FM, Chowdhury S, Moore B et al (2021) Artificial intelligence enables comprehensive genome interpretation and nomination of candidate diagnoses for rare genetic diseases. Genome Med 13:1–19. https://doi.org/10.1186/s13073-021-00965-0

Deeni Y, Beccari T, Dundar M et al (2014) Novel technologies and their applications in biotechnology and the life sciences. J Biotechnol 185:S12. https://doi.org/10.1016/j.jbiotec.2014.07.043

Dias R, Torkamani A (2019) Artificial intelligence in clinical and genomic diagnostics. Genome Med. 11

Dimauro G, Colagrande P, Carlucci R et al (2019) Crisprlearner: a deep learning-based system to predict crispr/cas9 sgrna on-target cleavage efficiency. Electron 8. https://doi.org/10.3390/electronics8121478

Ding L, Bailey MH, Porta-Pardo E et al (2018) Perspective on oncogenic processes at the end of the beginning of Cancer Genomics. https://doi.org/10.1016/j.cell.2018.03.033. Cell 173:

Dlamini Z, Francies FZ, Hull R, Marima R (2020) Artificial intelligence (AI) and big data in cancer and precision oncology. Comput Struct Biotechnol J 18

Du J, Yang YC, An ZJ et al (2023) Advances in spatial transcriptomics and related data analysis strategies. J Transl Med 21

Duncan JL, Pierce EA, Laster AM et al (2018) Inherited retinal degenerations: current landscape and knowledge gaps. Transl Vis Sci Technol 7:6–6. https://doi.org/10.1167/tvst.7.4.6

Dundar M, Emirogullari EF (2012) Biotechnology, cloning and ethics. Hum Evol 27:179–182

Dundar M, Prakash S, Lal R, Martin DK (2019) Future biotechnology. Eurobiotech J 3:53–56

Dundar MS, Yildirim AB, Yildirim DT et al (2024) Artificial cells: a potentially groundbreaking field of research and therapy. Eurobiotech J 8:55–64. https://doi.org/10.2478/ebtj-2024-0006

Erfanian N, Heydari AA, Feriz AM et al (2023) Deep learning applications in single-cell genomics and transcriptomics data analysis. Biomed Pharmacother 165

Eslami M, Adler A, Caceres RS et al (2022) Artificial intelligence for synthetic biology. Commun ACM 65:88–97. https://doi.org/10.1145/3500922

Fabbri C, Corponi F, Albani D et al (2018) Pleiotropic genes in psychiatry: calcium channels and the stress-related FKBP5 gene in antidepressant resistance. Prog Neuro-Psychopharmacology Biol Psychiatry 81:203–210. https://doi.org/10.1016/j.pnpbp.2017.10.005

Feil R, Fraga MF (2012) Epigenetics and the environment: emerging patterns and implications. Nat Rev Genet 13

Ferreira CR (2019) The burden of rare diseases. Am J Med Genet Part A 179:885–892

Foksinska A, Crowder CM, Crouse AB et al (2022) The precision medicine process for treating rare disease using the artificial intelligence tool mediKanren. Front Artif Intell 5. https://doi.org/10.3389/frai.2022.910216

Fudenberg G, Kelley DR, Pollard KS (2020) Predicting 3D genome folding from DNA sequence with Akita. Nat Methods 17:1111–1117. https://doi.org/10.1038/s41592-020-0958-x

Galal A, Talal M, Moustafa A (2022) Applications of machine learning in metabolomics: Disease modeling and classification. Front Genet 13

Gao D, Morini E, Salani M et al (2021) A deep learning approach to identify gene targets of a therapeutic for human splicing disorders. Nat Commun 12. https://doi.org/10.1038/s41467-021-23663-2

García Martín H, Mazurenko S, Zhao H (2024) Special issue on Artificial Intelligence for Synthetic Biology. ACS Synth Biol 13

Gargano MA, Matentzoglu N, Coleman B et al (2024) The human phenotype ontology in 2024: phenotypes around the world. Nucleic Acids Res 52. https://doi.org/10.1093/nar/gkad1005

Gartland KMA, Dundar M, Beccari T et al (2017) Advances in biotechnology: Genomics and genome editing. EuroBiotech J 1:2–9. https://doi.org/10.24190/issn2564-615x/2017/01.02

Gilbert-Barness E (2010) Teratogenic causes of malformations. Ann Clin Lab Sci 40:99–114

Giudice M, Del, Peirone S, Perrone S et al (2021) Artificial intelligence in bulk and single-cell rna-sequencing data to foster precision oncology. Int J Mol Sci 22

Gomes B, Ashley EA (2023) Artificial Intelligence in Molecular Medicine. N Engl J Med 388:2456–2465. https://doi.org/10.1056/nejmra2204787

Gordon ES, Babu D, Laney DA (2018) The future is now: Technology's impact on the practice of genetic counseling. Am. J. Med. Genet. Part C Semin. Med. Genet. 178

Gottlieb A, Stein GY, Ruppin E, Sharan R (2011) PREDICT: a method for inferring novel drug indications with application to personalized medicine. Mol Syst Biol 7. https://doi.org/10.1038/msb.2011.26

Greatbatch O, Garrett A, Snape K (2019) The impact of artificial intelligence on the current and future practice of clinical cancer genomics. Genet Res (Camb) 101:e9

Güell M, Yang L, Church GM (2014) Genome editing assessment using CRISPR Genome Analyzer (CRISPR-GA). Bioinformatics 30:2968–2970. https://doi.org/10.1093/bioinformatics/btu427

Gunasekara CJ, Hannon E, MacKay H et al (2021) A machine learning case–control classifier for schizophrenia based on DNA methylation in blood. Transl Psychiatry 11. https://doi.org/10.1038/s41398-021-01496-3

Gurovich Y, Hanani Y, Bar O et al (2019a) Identifying facial phenotypes of genetic disorders using deep learning. Nat Med 25:60–64. https://doi.org/10.1038/s41591-018-0279-0

Gurovich Y, Hanani Y, Bar O et al (2019b) Identifying facial phenotypes of genetic disorders using deep learning. Nat Med 25:60–64. https://doi.org/10.1038/s41591-018-0279-0

Hao J, Kim Y, Kim TK, Kang M (2018) PASNet: pathway-associated sparse deep neural network for prognosis prediction from high-throughput data. BMC Bioinformatics 19. https://doi.org/10.1186/s12859-018-2500-z

Hart TC, Hart PS (2009) Genetic studies of craniofacial anomalies: clinical implications and applications. Orthod Craniofac Res 12:212–220

He X, Liu X, Zuo F et al (2023) Artificial intelligence-based multi-omics analysis fuels cancer precision medicine. Semin Cancer Biol 88:187–200

He D, Wang R, Xu Z et al (2024) The use of artificial intelligence in the treatment of rare diseases: a scoping review. Intractable Rare Dis Res 13:12–22

Hınçer A, Ahan RE, Aras E, Şeker UÖŞ (2023) Making the next generation of therapeutics: mRNA meets Synthetic Biology. ACS Synth Biol 12:2505–2515

Horman WSJ, Nguyen THO, Kedzierska K et al (2018) The drivers of pathology in zoonotic avian influenza: the interplay between host and pathogen. Front Immunol 9

Hovenga V, Kalita J, Oluwadare O (2023) HiC-GNN: a generalizable model for 3D chromosome reconstruction using graph convolutional neural networks. Comput Struct Biotechnol J 21:812–836. https://doi.org/10.1016/j.csbj.2022.12.051

Hsieh T-C, Bar-Haim A, Moosa S et al (2021) GestaltMatcher: Overcoming the limits of rare disease matching using facial phenotypic descriptors. medRxiv

Huang Y, Sun X, Jiang H et al (2021) A machine learning approach to brain epigenetic analysis reveals kinases associated with Alzheimer's disease. Nat Commun 12. https://doi.org/10.1038/s41467-021-24710-8

Hull R, Ramagaga S, Nkosi N et al (2023) Epigenetics Analysis Using Artificial Intelligence in the era of Precision Oncology. Artificial Intelligence and Precision Oncology. Bridging Cancer Research and Clinical Decision Support

Hulsen T, Jamuar SS, Moody AR et al (2019) From big data to precision medicine. Front Med 6

Hunter AGW (2002) Medical genetics: 2. The diagnostic approach to the child with dysmorphic signs. C Can Med Assoc J 167:367–372

Ioannidis NM, Rothstein JH, Pejaver V et al (2016) REVEL: an Ensemble Method for Predicting the pathogenicity of rare missense variants. Am J Hum Genet 99. https://doi.org/10.1016/j.ajhg.2016.08.016

Ireland D, Bradford DK, Szepe E et al (2021) Introducing Edna: a trainee chatbot designed to support communication about additional (secondary) genomic findings. Patient Educ Couns 104:739–749. https://doi.org/10.1016/j.pec.2020.11.007

Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF et al (2019) Predicting Splicing from primary sequence with deep learning. https://doi.org/10.1016/j.cell.2018.12.015. Cell 176:

Ji Y, Zhou Z, Liu H, Davuluri RV (2021) DNABERT: pre-trained bidirectional encoder representations from transformers model for DNA-language in genome. Bioinformatics 37:2112–2120. https://doi.org/10.1093/bioinformatics/btab083

Jones BA, Grace D, Kock R et al (2013) Zoonosis emergence linked to agricultural intensification and environmental change. Proc Natl Acad Sci U S A 110:8399–8404. https://doi.org/10.1073/pnas.1208059110

Jumper J, Evans R, Pritzel A et al (2021) Highly accurate protein structure prediction with AlphaFold. Nature 596:583–589. https://doi.org/10.1038/s41586-021-03819-2

Kalil AC, Patterson TF, Mehta AK et al (2021) Baricitinib plus Remdesivir for hospitalized adults with Covid-19. N Engl J Med 384:795–807. https://doi.org/10.1056/nejmoa2031994

Kang M, Ko E, Mersha TB (2022) A roadmap for multi-omics data integration using deep learning. Brief Bioinform 23

Karim MR, Islam T, Shajalal M et al (2023) Explainable AI for Bioinformatics: methods, tools and applications. Brief Bioinform 24. https://doi.org/10.1093/bib/bbad236

Kautzky A, Baldinger P, Souery D et al (2015) The combined effect of genetic polymorphisms and clinical parameters on treatment outcome in treatment-resistant depression. Eur Neuropsychopharmacol 25:441–453. https://doi.org/10.1016/j.euroneuro.2015.01.001

Kearney E, Wojcik A, Babu D (2020) Artificial intelligence in genetic services delivery: Utopia or apocalypse? J Genet Couns 29. https://doi.org/10.1002/jgc4.1192

Keesing F, Ostfeld RS (2021) Impacts of biodiversity and biodiversity loss on zoonotic diseases. Proc. Natl. Acad. Sci. U. S. A. 118

Kenanoglu S, Gokce N, Akalin H et al (2022) Implication of the Mediterranean diet on the human epigenome. J Prev Med Hyg 63:E44–E55

Kim YR, Kim D, Kim SY (2019) Prediction of acquired taxane resistance using a personalized pathway-based machine learning method. Cancer Res Treat 51:672–684. https://doi.org/10.4143/crt.2018.137

Klein TE, Chang JT, Cho MK et al (2001) Integrating genotype and phenotype information: an overview of the PharmGKB project. Pharmacogenomics J 1:167–170. https://doi.org/10.1038/sj.tpj.6500035

Korfmann K, Gaggiotti OE, Fumagalli M (2023) Deep learning in Population Genetics. Genome Biol Evol 15

Kuang S, Wang L (2021) Deep learning of sequence patterns for CCCTC-Binding factor-mediated chromatin Loop formation. J Comput Biol 28:133–145. https://doi.org/10.1089/cmb.2020.0225

Lee M (2023) Deep learning in CRISPR-Cas systems: a review of recent studies. Front Bioeng Biotechnol 11

Lee Ysuk, Krishnan A, Oughtred R et al (2019) A computational Framework for Genome-wide characterization of the Human Disease Landscape. Cell Syst 8:152–162e6. https://doi.org/10.1016/j.cels.2018.12.010

Li Q, Ren Z, Cao K et al (2022a) CancerVar: an artificial intelligence–empowered platform for clinical interpretation of somatic mutations in cancer. Sci Adv 8. https://doi.org/10.1126/sciadv.abj1624

Li R, Li L, Xu Y, Yang J (2022b) Machine learning meets omics: applications and perspectives. Brief Bioinform 23

Li Y, Stanojevic S, Garmire LX (2022c) Emerging artificial intelligence applications in spatial transcriptomics analysis. Comput Struct Biotechnol J 20:2895–2908

Liang L, Rasmussen MLH, Piening B et al (2020) Metabolic Dynamics and Prediction of gestational age and Time to delivery in pregnant women. Cell 181:1680–1692e15. https://doi.org/10.1016/j.cell.2020.05.002

Liao WW, Asri M, Ebler J et al (2023) A draft human pangenome reference. Nature 617:312–324. https://doi.org/10.1038/s41586-023-05896-x

Liñares-Blanco J, Pazos A, Fernandez-Lozano C (2021) Machine learning analysis of TCGA cancer data. PeerJ Comput Sci 7. https://doi.org/10.7717/PEERJ-CS.584

Liu D, Schilling B, Liu D et al (2019) Integrative molecular and clinical modeling of clinical outcomes to PD1 blockade in patients with metastatic melanoma. Nat Med 25:1916–1927. https://doi.org/10.1038/s41591-019-0654-5

Logviss K, Krievins D, Purvina S (2018) Characteristics of clinical trials in rare vs. common diseases: a register-based Latvian study. PLoS ONE 13. https://doi.org/10.1371/journal.pone.0194494

López-Cortés A, Paz-y-Miño C, Cabrera-Andrade A et al (2018) Gene prioritization, communality analysis, networking and metabolic integrated pathway to better understand breast cancer pathogenesis. Sci Rep 8. https://doi.org/10.1038/s41598-018-35149-1

López-Cortés A, Cabrera-Andrade A, Vázquez-Naya JM et al (2020) Prediction of breast cancer proteins involved in immunotherapy, metastasis, and RNA-binding using molecular descriptors and artificial neural networks. Sci Rep 10. https://doi.org/10.1038/s41598-020-65584-y

Lumaka A, Cosemans N, Lulebo Mampasi A et al (2017) Facial dysmorphism is influenced by ethnic background of the patient and of the evaluator. Clin Genet 92:166–171. https://doi.org/10.1111/cge.12948

Maciukiewicz M, Marshe VS, Hauschild AC et al (2018) GWAS-based machine learning approach to predict duloxetine response in major depressive disorder. J Psychiatr Res 99:62–68. https://doi.org/10.1016/j.jpsychires.2017.12.009

Maramraju S, Kowalczewski A, Kaza A et al (2024) AI-organoid integrated systems for biomedical studies and applications. Bioeng Transl Med 9

Marian AJ (2014) Sequencing your genome: what does it mean? Methodist Debakey Cardiovasc. J 10:3–6

Martinez GS, Dutt M, Kelvin DJ, Kumar A (2024) PoxiPred: an Artificial-Intelligence-based method for the prediction of potential antigens and epitopes to accelerate Vaccine Development efforts against Poxviruses. Biology (Basel) 13:125. https://doi.org/10.3390/biology13020125

McMaster ML, Berndt SI, Zhang J et al (2018) Two high-risk susceptibility loci at 6p25.3 and 14q32.13 for Waldenström macroglobulinemia. Nat Commun 9. https://doi.org/10.1038/s41467-018-06541-2

Mester R, Hou K, Ding Y et al (2023) Impact of cross-ancestry genetic architecture on GWASs in admixed populations. Am J Hum Genet 110:927–939. https://doi.org/10.1016/j.ajhg.2023.05.001

Mishima H, Suzuki H, Doi M et al (2019) Evaluation of Face2Gene using facial images of patients with congenital dysmorphic syndromes recruited in Japan. J Hum Genet 64:789–794. https://doi.org/10.1038/s10038-019-0619-z

Modi K, Singh I, Kumar Y (2023) A Comprehensive Analysis of Artificial Intelligence Techniques for the prediction and prognosis of Lifestyle diseases. Arch Comput Methods Eng 30:4733–4756

Mohanta TK, Mishra AK, Al-Harrasi A (2021) The 3d genome: from structure to function. Int J Mol Sci 22

Monsalve D, Mesa A, Mira LM et al (2024) Antimicrobial peptides designed by computational analysis of proteomes. Antonie Van Leeuwenhoek. Int J Gen Mol Microbiol 117. https://doi.org/10.1007/s10482-024-01946-0

Moser T, Kühberger S, Lazzeri I et al (2023) Bridging biological cfDNA features and machine learning approaches. Trends Genet 39

Mund A, Coscia F, Kriston A et al (2022) Deep visual proteomics defines single-cell identity and heterogeneity. Nat Biotechnol 40:1231–1240. https://doi.org/10.1038/s41587-022-01302-5

Nazareth S, Nussbaum RL, Siglen E, Wicklund CA (2021) Chatbots & artificial intelligence to scale genetic information delivery. In: Journal of Genetic Counseling

Newman AM, Liu CL, Green MR et al (2015) Robust enumeration of cell subsets from tissue expression profiles. Nat Methods 12:453–457. https://doi.org/10.1038/nmeth.3337

Newman AM, Steen CB, Liu CL et al (2019) Determining cell type abundance and expression from bulk tissues with digital cytometry. Nat Biotechnol 37. https://doi.org/10.1038/s41587-019-0114-2

Osanlou R, Walker L, Hughes DA et al (2022) Adverse drug reactions, multimorbidity and polypharmacy: a prospective analysis of 1 month of medical admissions. BMJ Open 12:e055551. https://doi.org/10.1136/BMJOPEN-2021-055551

Ostroverkhova D, Przytycka TM, Panchenko AR (2023) Cancer driver mutations: predictions and reality. Trends Mol Med 29:554–566

Pillai N, Ramkumar M, Nanduri B (2022) Artificial Intelligence Models for Zoonotic Pathogens: A Survey. Microorganisms 10

Pirmohamed M (2023) Pharmacogenomics: current status and future perspectives. Nat Rev Genet 24:350–362

Pirmohamed M, James S, Meakin S et al (2004) Adverse drug reactions as cause of admission to hospital: prospective analysis of 18 820 patients. Br Med J 329:15–19. https://doi.org/10.1136/bmj.329.7456.15

PONTIKOS N, Woof W, Krawitz P et al (2022) Eye2Gene: prediction of causal inherited retinal disease gene from multimodal imaging using AI. Invest Ophthalmol Vis Sci 63:1161. https://doi.org/10.21203/RS.3.RS-2110140/V1

Poplin R, Chang PC, Alexander D et al (2018) A universal snp and small-indel variant caller using deep neural networks. Nat Biotechnol 36:983. https://doi.org/10.1038/nbt.4235

Posey JE, Harel T, Liu P et al (2017) Resolution of Disease Phenotypes resulting from Multilocus genomic variation. N Engl J Med 376. https://doi.org/10.1056/nejmoa1516767

Rendeiro AF, Schmidl C, Strefford JC et al (2016) Chromatin accessibility maps of chronic lymphocytic leukaemia identify subtype-specific epigenome signatures and transcription regulatory networks. Nat Commun 7. https://doi.org/10.1038/NCOMMS11938

Rhoads DD (2020) Computer vision and artificial intelligence are emerging diagnostic tools for the clinical microbiologist. J Clin Microbiol 58

Richards S, Aziz N, Bale S et al (2015) Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med 17:405–424. https://doi.org/10.1038/gim.2015.30

Richardson PJ, Smith DP, de Giorgio A et al (2023) Janus kinase inhibitors are potential therapeutics for amyotrophic lateral sclerosis. Transl Neurodegener 12

Saal HM (2016) Genetic evaluation for Craniofacial conditions. Facial plast. Surg Clin North Am 24:405–425

Schubach M, Maass T, Nazaretyan L et al (2024) CADD v1.7: using protein language models, regulatory CNNs and other nucleotide-level scores to improve genome-wide variant predictions. Nucleic Acids Res 52:D1143–D1154. https://doi.org/10.1093/nar/gkad989

Schwarz JM, Cooper DN, Schuelke M, Seelow D (2014) Mutationtaster2: mutation prediction for the deep-sequencing age. Nat Methods 11

Schwessinger R, Gosden M, Downes D et al (2020) DeepC: predicting 3D genome folding using megabase-scale transfer learning. Nat Methods 17:1118–1124. https://doi.org/10.1038/s41592-020-0960-3

Shahbazy M, Ramarathinam SH, Li C et al (2024) MHCpLogics: an interactive machine learning-based tool for unsupervised data visualization and cluster analysis of immunopeptidomes. Brief Bioinform 25. https://doi.org/10.1093/bib/bbae087

Siglen E, Vetti HH, Augestad M et al (2023) Evaluation of the Rosa Chatbot Providing Genetic Information to patients at risk of Hereditary breast and ovarian Cancer: qualitative interview study. J Med Internet Res 25. https://doi.org/10.2196/46571

Smedley D, Jacobsen JOB, Jäger M et al (2015) Next-generation diagnostics and disease-gene discovery with the Exomiser. Nat Protoc 10. https://doi.org/10.1038/nprot.2015.124

Smith KP, Wang H, Durant TJS et al (2020) Applications of Artificial Intelligence in Clinical Microbiology Diagnostic Testing. Clin Microbiol Newsl 42. https://doi.org/10.1016/j.clinmicnews.2020.03.006

Solomon BD, Adam MP, Fong CT et al (2023) Perspectives on the future of dysmorphology. Am J Med Genet Part A 191. https://doi.org/10.1002/ajmg.a.63060

Soori M, Arezoo B, Dastres R (2023) Artificial intelligence, machine learning and deep learning in advanced robotics, a review. Cogn Robot 3:54–70

Su R, Liu X, Wei L, Zou Q (2019) Deep-Resp-Forest: a deep forest model to predict anti-cancer drug response. Methods 166:91–102. https://doi.org/10.1016/j.ymeth.2019.02.009

Sun YV, Hu YJ (2016) Adv Genet 93:147–190. https://doi.org/10.1016/bs.adgen.2015.11.004. Integrative Analysis of Multi-omics Data for Discovery and Functional Studies of Complex Human Diseases

Sun D, Wang M, Li A (2019) A multimodal deep neural network for human breast Cancer Prognosis Prediction by integrating Multi-dimensional Data. IEEE/ACM Trans Comput Biol Bioinforma 16:841–850. https://doi.org/10.1109/TCBB.2018.2806438

Tamborero D, Rubio-Perez C, Deu-Pons J et al (2018) Cancer Genome Interpreter annotates the biological and clinical relevance of tumor alterations. Genome Med 10. https://doi.org/10.1186/s13073-018-0531-8

Tong Z, Zhou Y, Wang J (2019) Identifying potential drug targets in hepatocellular carcinoma based on network analysis and one-class support vector machine. Sci Rep 9. https://doi.org/10.1038/s41598-019-46540-x

Townsend DR, Towers DM, Lavinder JJ, Ippolito GC (2024) Innovations and trends in antibody repertoire analysis. Curr Opin Biotechnol 86

Vinciguerra M (2023) The potential for Artificial Intelligence Applied to Epigenetics. Mayo Clin Proc Digit Heal. https://doi.org/10.1016/j.mcpdig.2023.07.005. 1:

Visibelli A, Roncaglia B, Spiga O, Santucci A (2023) The Impact of Artificial Intelligence in the Odyssey of Rare Diseases. Biomedicines 11

Wang T, Antonacci-Fulton L, Howe K et al (2022) The human pangenome project: a global resource to map genomic diversity. Nature 604:437–446

Wang R, Helbig I, Edmondson AC et al (2023) Splicing defects in rare diseases: transcriptomics and machine learning strategies towards genetic diagnosis. Brief Bioinform 24

Whalen S, Truty RM, Pollard KS (2016) Enhancer-promoter interactions are encoded by complex genomic signatures on looping chromatin. Nat Genet 48:488–496. https://doi.org/10.1038/ng.3539

Wojtara M, Rana E, Rahman T et al (2023) Artificial intelligence in rare disease diagnosis and treatment. Clin Transl Sci 16:2106–2111

Wolffe AP, Guschin D (2000) Review: chromatin structural features and targets that regulate transcription. J Struct Biol 129. https://doi.org/10.1006/jsbi.2000.4217

Xu J, Yang P, Xue S et al (2019) Translating cancer genomics into precision medicine with artificial intelligence: applications, challenges and future perspectives. Hum Genet 138:109–124

Yang M, Ma J (2022) Machine learning methods for exploring sequence determinants of 3D Genome Organization. J Mol Biol 434:167666

Yaoyu F, Xiao L (2011) Zoonotic potential and molecular epidemiology of Giardia species and giardiasis. Clin Microbiol Rev 24:110–140. https://doi.org/10.1128/CMR.00033-10

Zemojtel T, Köhler S, Mackenroth L et al (2014) Effective diagnosis of genetic disease by computational phenotype analysis of the disease-associated genome. Sci Transl Med 6. https://doi.org/10.1126/scitranslmed.3009262

Zeng J, Xue A, Jiang L et al (2021) Widespread signatures of natural selection across human complex traits and functional genomic categories. Nat Commun 12. https://doi.org/10.1038/s41467-021-21446-3

Zhang Y (2017) Investigate genomic 3D structure using deep neural network. PhD Thesis

Zhou J (2022) Sequence-based modeling of three-dimensional genome architecture from kilobase to chromosome scale. Nat Genet 54:725–734. https://doi.org/10.1038/s41588-022-01065-4

Zhu S, Jiang Y, Xu K et al (2020) The progress of gut microbiome research related to brain disorders. J Neuroinflammation 17

Authors and Affiliations

Firat Ozcelik[1] · Mehmet Sait Dundar[2] · A. Baki Yildirim[1] · Gary Henehan[3] · Oscar Vicente[4] · José A. Sánchez-Alcázar[5] · Nuriye Gokce[1] · Duygu T. Yildirim[1] · Nurdeniz Nalbant Bingol[6] · Dijana Plaseska Karanfilska[7] · Matteo Bertelli[8] · Lejla Pojskic[9] · Mehmet Ercan[1] · Miklos Kellermayer[10] · Izem Olcay Sahin[1] · Ole K. Greiner-Tollersrud[11] · Busra Tan[1] · Donald Martin[12] · Robert Marks[13] · Satya Prakash[14] · Mustafa Yakubi[1] · Tommaso Beccari[15] · Ratnesh Lal[16] · Sehime G. Temel[6,17,18] · Isabelle Fournier[19] · M. Cerkez Ergoren[20] · Adam Mechler[21] · Michel Salzet[19] · Michele Maffia[22] · Dancho Danalev[23] · Qun Sun[24] · Lembit Nei[25] · Daumantas Matulis[26] · Dana Tapaloaga[27] · Andres Janecke[28,29] · James Bown[30] · Karla Santa Cruz[31] · Iza Radecka[32] · Celal Ozturk[33] · Ozkan Ufuk Nalbantoglu[34] · Sebnem Ozemri Sag[17] · Kisung Ko[35] · Reynir Arngrimsson[36] · Isabel Belo[37] · Hilal Akalin[1] · Munis Dundar[1]

✉ Hilal Akalin
h_akaln@yahoo.com

✉ Munis Dundar
dundar@erciyes.edu.tr

[1] Department of Medical Genetics, Faculty of Medicine, Erciyes University, Kayseri, Turkey

[2] Department of Electrical and Computer Engineering, Graduate School of Engineering and Sciences, Abdullah Gul University, Kayseri, Turkey

[3] School of Food Science and Environmental Health, Technological University of Dublin, Dublin, Ireland

[4] Institute for the Conservation and Improvement of Valencian Agrodiversity (COMAV), Universitat Politècnica de València, Valencia, Spain

[5] Centro de Investigación Biomédica en Red: Enfermedades Raras, Centro Andaluz de Biología del Desarrollo (CABD-CSIC-Universidad Pablo de Olavide), Instituto de Salud Carlos III, Sevilla, Spain

[6] Department of Translational Medicine, Institute of Health Sciences, Bursa Uludag University, Bursa, Turkey

[7] Research Centre for Genetic Engineering and Biotechnology, Macedonian Academy of Sciences and Arts, Skopje, Macedonia

[8] MAGI'S LAB, Overeto, Italy

[9] Institute for Genetic Engineering and Biotechnology, University of Sarajevo, Sarajevo, Bosnia and Herzegovina

[10] Department of Biophysics and Radiation Biology, Faculty of Medicine, Semmelweis University, Budapest, Hungary

[11] Institute of Medical Biology, UiT The Arctic University of Norway, Tromsø, Norway

[12] University Grenoble Alpes, CNRS, TIMC-IMAG/SyNaBi (UMR 5525), Grenoble, France

[13] Avram and Stella Goldstein-Goren Department of Biotechnology Engineering, Ben-Gurion University of the Negev, Be'er Sheva, Israel

[14] Department of Biomedical Engineering, University of McGill, Montreal, QC, Canada

[15] Department of Pharmeceutical Sciences, University of Perugia, Perugia, Italy

[16] Neuroscience Research Institute, University of California, Santa Barbara, USA

[17] Department of Medical Genetics, Bursa Uludag University Faculty of Medicine, Bursa, Turkey

[18] Department of Histology and Embryology, Faculty of Medicine, Bursa Uludag University, Bursa, Turkey

[19] Réponse Inflammatoire et Spectrométrie de Masse-PRISM, University of Lille, Lille, France

[20] Department of Medical Genetics, Near East University Faculty of Medicine, Nicosia, Cyprus

[21] Department of Chemistry, La Trobe Institute for Molecular Science, La Trobe University, Melbourne, VIC, Australia

[22] Department of Experimental Medicine, University of Salento, Via Lecce-Monteroni, Lecce 73100, Italy

[23] University of Chemical Technology and Metallurgy, Sofia, Bulgaria

[24] Department of Food Science and Technology, Sichuan University, Chengdu, China

[25] School of Engineering Tallinn University of Technology, Tartu College, Tartu, Estonia

[26] Department of Biothermodynamics and Drug Design, Institute of Biotechnology, Life Sciences Center, Vilnius University, Vilnius, Lithuania

[27] Faculty of Veterinary Medicine, University of Agronomic Sciences and Veterinary Medicine of Bucharest, Bucharest, Romania

[28] Department of Paediatrics I, Medical University of Innsbruck, Innsbruck, Austria

[29] Division of Human Genetics, Medical University of Innsbruck, Innsbruck, Austria

[30] School of Science, Engineering and Technology, Abertay University, Dundee, UK

[31] Department of Physics, University of Sonora, Hermosillo, Mexico

[32] School of Science, Faculty of Science and Engineering, University of Wolverhampton, Wolverhampton, UK

[33] Department of Software Engineering, Erciyes University, Kayseri, Turkey

[34] Department of Computer Engineering, Engineering Faculty, Erciyes University, Kayseri, Turkey

[35] Department of Medicine, College of Medicine, Chung-Ang University, Seoul, Korea

[36] Iceland Landspitali University Hospital, University of Iceland, Reykjavik, Iceland

[37] Centre of Biological Engineering, University of Minho, Braga, Portugal