Review

Application of Genomic Data in Translational Medicine During the Big Data Era

Yun Zhang[1], Jian Yu[1], Xiaoyan Xie[1], Feng Jiang[2,]*, Chuyan Wu[1,]*

[1]Department of Rehabilitation Medicine, The First Affiliated Hospital of Nanjing Medical University, 210029 Nanjing, Jiangsu, China
[2]Department of Neonatology, Obstetrics and Gynecology Hospital of Fudan University, 200090 Shanghai, China
*Correspondence: chuyan_w@hotmail.com (Chuyan Wu); dxyjiang@163.com (Feng Jiang)
Academic Editor: Graham Pawelec

Abstract

Advances in gene sequencing technology and decreasing costs have resulted in a proliferation of genomic data as an integral component of big data. The availability of vast amounts of genomic data and more sophisticated genomic analysis techniques has facilitated the transition of genomics from the laboratory to clinical settings. More comprehensive and precise DNA sequencing empowers patients to address health issues at the molecular level, facilitating early diagnosis, timely intervention, and personalized healthcare management strategies. Further exploration of disease mechanisms through identification of associated genes may facilitate the discovery of therapeutic targets. The prediction of an individual's disease risk allows for improved stratification and personalized prevention measures. Given the vast amount of genomic data, artificial intelligence, as a burgeoning technology for data analysis, is poised to make a significant impact in genomics.

Keywords: genomic data; translational medicine; next-generation sequencing; whole-genome association studies; polygenic risk scores; machine learning

1. Introduction

In recent years, genomic data has exploded due to the advancement of high-throughput sequencing technologies, a reduction in sequencing costs, and the proliferation of consumer-oriented sequencing platforms. There are already estimates that genomics will generate more data than social media applications and astronomy within a few years [1]. Genomic data becomes an integral part of Big Data. Gene interactions, gene-environment interactions, and noncoding regions of the genome are studied by genomics. In contrast, genetics focuses on individual traits [2]. A limitless opportunity has been created by genomic big data for the life science field through its efficient use. As big data has taken hold, genomics has witnessed unprecedented progress, including the establishment of large-scale global biobanks that contain genetic and phenotypic information [3–8], as well as a variety of advanced computational and statistical approaches to predict disease risks and map disease genetics [9]. Consequently, genomic data are more accessible and the role of genes in disease and health is better understood. For example, there are more than 7000 single-gene diseases with known molecular etiologies and shapes (https://www.omim.org/statistics/geneMap). Transforming genomics into medicine involves applying a large amount of genomics data to solve healthcare problems.

Translational medicine epitomizes a form of medical inquiry that transmogrifies the products of fundamental research into tangible modalities encompassing ailment abatement, ailment discernment, therapeutics, and prognosis assessment for living patients. Its quintessential trait resides in interdisciplinarity, exhaustive foundational exploration concerning clinical quandaries, and the swift dissemination of research outcomes into practicality. The nucleus of translational medicine resides in the scrutiny of biomarkers. To this end, it encompasses the evolution and utilization of diverse "omics" methodologies and molecular biology repositories to scrutinize a spectrum of biomarkers for estimating ailment susceptibility, ailment diagnosis, and categorization, evaluating responses to therapeutics, prophesying the trajectory of ailment, and the evolution of innovative remedial methodologies and novel pharmacotherapeutics. The translational and clinical sciences, introduced by Zerhouni in 2005, are a significant part of biomedicine and are a promising new interdisciplinary approach for translating superior scientific innovations into health benefits [10]. Within the precincts of translational medicine, the harnessing of voluminous data is deemed a potent cornerstone, bequeathing invaluable insights to fortify the realms of medical inquiry and therapeutic paradigms.

Within the confines of this scrutiny, we have delved into the realm of clinical governance, a sphere profoundly impacted by germane technologies and methodologies within the milieu of genomics and vast datasets. Synthesizing the amalgamation of genomic erudition with pragmatic clinical implementation, alongside the adversities encountered in the analogous realm of genomic application, could serve as a beacon guiding the further transmutation of genomic information. It points out the direction for additional transformation of genomics data (Fig. 1).
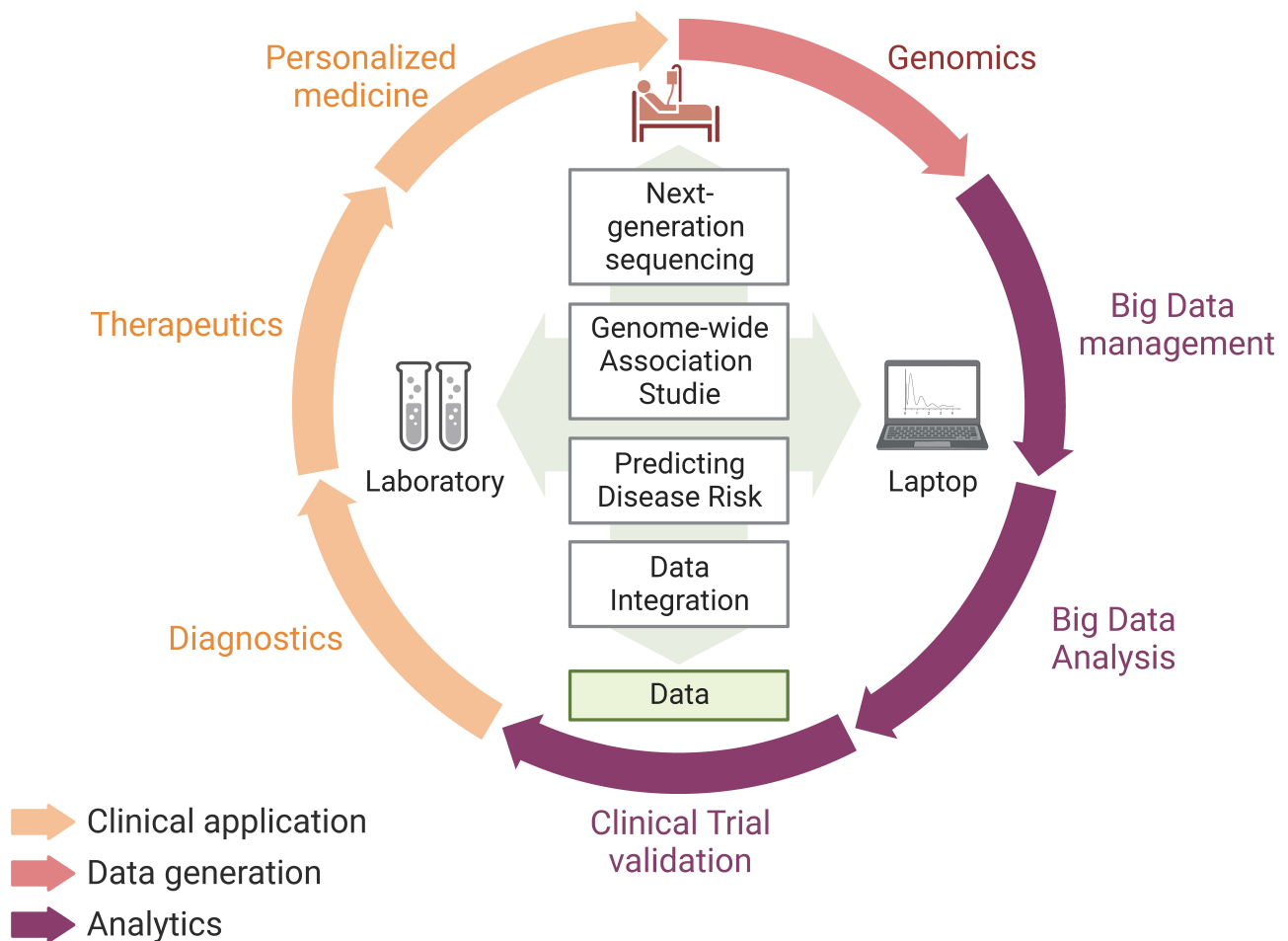
Fig. 1. Application of Genomic Data in Translational Medicine including data generation, analytics and clinical application.

2. The Application of Gene Data in Translational Medicine

2.1 The Next-generation Sequencing Entering the Public Eye and the Clinic

The continuous improvement of genome sequencing technology allows patients to choose different next-generation sequencing (NGS) technologies such as targeted sequencing, whole-exome sequencing, and whole-genome sequencing to obtain their genome information according to their actual needs. This has the potential to change the health outcomes of patients to a great extent for only a few hundred dollars. The major advantage of NGS is its ability to simplify, accelerate, and expand the range and number of sequences that can be assessed in comparison to conventional methods of sequencing. The use of genome sequencing technology is crucial to translational medicine in genomics, as it provides high-resolution genetic information highly relevant to diseases, as well as improving disease screening, molecular diagnosis, treatment, and management [11–14].

2.1.1 NGS in Prevention

Precision prevention is a prevention strategy for primary (disease prevention) and secondary (early detection) prevention of disease by combining non-genetic and genetic characteristics of individuals. Many cancers, including hereditary breast cancer and Lynch syndrome, can be detected and risk assessed using genetic testing. It is recommended that carriers of *BRCA1* or *BRCA2* germline variants undergo prophylactic surgery to mitigate their risk of developing breast cancer, as well as testing relatives at risk and implementing preventive interventions for those who carry variants [15]. An individual who carries cancer susceptibility genes is typically identified based on his family history [16], however, the family-history model has been recognized to have some limitations [17]. Between 50% and 60% of patients miss the opportunity for genetic testing because they do not meet the criteria as assessed by the family history model [18–20]. Hence, in light of the increasing ease of access to NGS technologies, it has been proposed to incorporate them into a population-based screening approach [21]. With population-based genetic testing, it is possible to significantly improve health outcomes for cancer susceptibility carriers while enabling accurate cancer

prevention through risk stratification. Studies have shown that population-based genetic testing reduces cancer incidence without increasing mental stress is cost-effective, and is capable of identifying more mutation carriers [19,21,22]. NGS also plays an irreplaceable role in the screening and diagnosis of additional diseases. For example, its use in sudden cardiac death (SCD), which kills millions of people every year [23]. It is estimated that up to 70 percent of SCD cases in people under 50 years of age are potentially due to genetic causes [24]. As the majority of SCD occurs in the general population rather than in those diagnosed with heart disease [25], a high priority should be given to optimising risk stratification and prevention of SCD in the general population as an essential clinical and public health goal. A study has shown that genetic screening using a targeted NGS panel can identify molecular genetic causes in a significant proportion of patients with suspected inherited heart disease and can be applied to cascade screening of family members of genotype-positive pre-diagnosed individuals [26]. It also guides the choice of preventive measures, such as treatment with antiarrhythmic drugs or implantable cardio verterde fibrillators (ICDs) [26]. With additional genetic and phenotypic correlation studies, NGS is expected to be a rapid and cost-effective molecular diagnostic tool for sudden cardiac death [27] that can reduce the probability of sudden cardiac death in patients through population-based screening.

Similarly, NGS has been proposed for newborn screening. Virtual rapid whole-genome sequencing for newborn screening performed in UK Biobank identified 15 disease cases missed by conventional screening. The study concluded that if whole-genome sequencing had been conducted on day 5 of life, symptoms in seven critically ill children could have been completely prevented [28]. Although neonatal genome sequencing is beneficial in detecting many diseases at an early stage, its use as a routine application in NGS requires further research on the long-term impact of preventive genomic screening on families as well as children, the long-term management of genetic sequencing results due to updated knowledge, and the unique ethical challenges that may arise [29]. It is possible to apply genomics to clinical practice as a means of precision prevention through population-based genetic testing, enabling more precise identification of diseased potential groups and earlier clinical intervention, leading to a health strategy that maximizes benefits for more patients.

2.1.2 NGS in Diagnosis

Information from the molecular level of disease can more accurately characterize the disease, improve the efficiency of disease diagnosis, refine disease typing, and provide information for treatment decisions. There has always been a strong focus on genomic research in cancer. Origins of tissues, types of cells, and morphologies are traditionally used to classify tumors. Cancer diagnostics and prognostics

have become increasingly reliant on mutations in the context of cancer's genomic basis. In the molecular classification of endometrial carcinoma by The Cancer Genome Atlas (TCGA), *POLE* gene mutation indicates *POLE* hypermutation, which is associated with a more favorable prognosis. Conversely, the high copy type with *TP53* mutations is linked to the poorest prognosis [30]. As well, breast cancers frequently have somatic mutations in *TP53* and *PIK3C* genes. It should be noted that the frequency of these gene mutations varies depending on the subtype of breast cancer. With NGS, mutational variants can be identified, thus making subtype classification easier [31].

It is challenging to collect samples of tissue from patients suffering from cancer during the early stages of their disease so novel non-invasive biomarkers are being explored. It is possible to use plasma cell-free DNA (cfDNA) genotyping instead of tissue genotyping in circumstances where tissue specimens are insufficient or inaccessible [32,33]. Plasma cfDNA, which is released from cells, contains circulating tumor DNA [34]. NGS technology has been used to detect a variety of mutations associated with cancer in cfDNA from individuals with various types of cancer, including copy number changes, single nucleotide mutations, DNA fragmentation patterns, and methylation changes [35–37]. Leighl *et al.* [38] compared tissue genotyping with cfDNA genotyping by NGS in untreated metastatic non-small-cell lung cancer and found highly consistent results and more efficient cfDNA genotyping. Additionally, cfDNA genotyping can enable targeted treatment matching and monitoring of resistance to therapy to provide personalized treatment [39,40]. In a short time and for a low cost, NGS can quickly identify a wide variety of biomarkers from small biological samples due to its high sensitivity and specificity. As a result, NGS became the technique of choice for analyzing cfDNA, known as liquid biopsy, a new trend in oncology [41].

The implementation of next-generation sequencing in prenatal diagnosis represents a significant achievement in the application of genomics within clinical settings. Conducting noninvasive prenatal screenings (NIPSs) to detect fetal chromosomal defects through the analysis of maternal cfDNA has changed the way chromosomal and other genetic disorders are diagnosed and treated during pregnancy [42,43]. Furthermore, whole-genome sequencing (WGS) has been shown to assist in determining the cause of rare diseases as well as performing accurate molecular diagnoses. In comparison to traditional diagnostic methods, WGS is more efficient in diagnosing, reducing diagnostic odyssey costs, changing medical management, and ultimately benefiting patients and their families [44–46]. There are already established guidelines recommending the utilization of clinical whole genome sequencing as either a primary or secondary diagnostic tool for rare disease sufferers [47,48].

2.1.3 NGS in Personalized Treatment

Personalized therapeutic decisions are increasingly being guided by NGS. Genomic analysis of cancer is crucial to selecting drugs that target somatic gene mutations that drive cancer development or progression. According to research, NGS can help identify at least one actionable target for targeted therapy [49–51]. Further, targeted therapy appears to be associated with higher progression-free survival and response rates [52]. If patients do not benefit from molecularly targeted agents, immunotherapy may also be considered. Patients for whom immunotherapy may be effective can be identified based on genomic information about the tumor obtained through NGS. Recently, a clinical trial demonstrated that patients with *HER2*-negative progressive gastric cancer achieved significantly improved overall survival with tislelizumab, an immune checkpoint inhibitor, compared to the chemotherapy group [53]. The utilization of immunotherapy has expanded the treatment strategies available to cancer patients, and it is increasingly being employed as a first-line therapy for many types of cancer [54].

With the advent of various biotechnologies, such as variant analysis methods, and gene editing, gene therapy has entered a new era. Genetic analysis through NGS facilitates the identification of target mutations that drive disease progression, furnishes highly precise DNA sequence data, and constitutes a crucial tool for enhancing gene therapy. As a result, many novel gene therapy products have been approved for clinical use after successful laboratory testing [55]. Gene therapy is now used not only for single genetic disorders, and cancer [56] but also for common complex diseases such as osteoarthritis [57] and diabetic neuropathy [58]. Gene therapy uses normally functioning genes to repair or replace defective genes to treat genetic diseases that are generally untreatable by drugs and are less likely to face drug resistance problems [59]. However, the field is still in an immature stage, with challenges that include ineffective delivery systems [60], and the inability to ensure continuous and stable gene expression and host immune responses. To facilitate the comprehensive market penetration of gene therapy, it is necessary to deepen basic research, overcome technical difficulties, precise safety supervision to reduce the harm caused by possible side effects [61], and appropriate pricing to increase the accessibility of treatment.

In summary, genome sequencing technology which is the key to facilitating faster translation of genomic knowledge for clinical practice provides genomic information that provides unprecedented insight into the biology and pathogenesis of many diseases [62–64] and has been used in the fields of cancer detection and treatment, assisted reproduction, prenatal and perinatal screening, inpatient management of critically ill infants, management of Mendelian diseases, rare diseases, and other clinical aspects [65–68]. The current NGS still suffers from technical defects such as short long-read and the inability to completely analyze

complex repetitive sequences [69]. In addition to the detection and analysis capabilities of sequencing technology itself, there are still problems other than technology that hinder its application in clinical practice. While NGS has become more affordable, its economic viability isn't always guaranteed in all clinical and research settings [70]. The accumulation of genetic data has produced an increasing proportion of variants of uncertain significance (VUS), which overshadows clinicians' decision-making. There is a need to classify the VUS found and review them regularly to ensure a beneficial impact on patient health guidance.

2.2 Genome-Wide Association Studies (GWAS) Identifying Risk Genes Associated with Disease

The clinical translation of genetic mechanisms for monogenic diseases has been accelerated by the dramatic impact of rare variants identified through linkage analysis. This is exemplified by ongoing clinical trials or FDA approval of gene therapies for rare diseases such as hemophilia, cystic fibrosis, and spinal muscular atrophy sclerosis, which hold significant promise for affected individuals and their families [71]. However, research into common complex diseases remains a formidable challenge. Genetic variation, mutations associated with disease, and genotype-phenotype associations can be studied with a large amount of genomic data [72–75]. Genetic research on complex diseases has ushered in a new era in recent years following the discovery of two single nucleotide polymorphisms (SNPs) by the genome-wide association study (GWAS) associated with age-related macular degeneration [76]. Up to now, GWAS has successfully identified risk genes for a large number of diseases and traits, including Parkinson's disease [77], obesity [78], autoimmune diseases [79], height [80], and others.

2.2.1 From GWAS to Biological Mechanisms of Diseases

Large sample sizes and genomic data from large biobanks have been an important basis for supporting discoveries in GWAS. Among these biobanks, the UK Biobank plays a leading role, containing deep genetic and phenotypic data of 500,000 individuals [74]. Expanding the sample size of GWAS can facilitate the identification of additional risk loci and generate a more comprehensive list for discovering novel therapeutic targets. A genome-wide association analysis with a large sample size of 100,285 subjects revealed novel loci shared by lung function and obesity. However, the complexity of the linkage disequilibrium pattern and deficiencies in the imputation data impeded the identification of true causal variants [81]. Therefore, while we continue to explore novel associations, we must prioritize causal studies of established associations. These studies will undoubtedly provide in-depth insights into the biological mechanisms of disease and enhance clinical translation. In addition, it should be noted that susceptibility loci identified by GWAS do not necessarily correspond to causative genes. Functional genomics studies

are required to accurately map these loci to specific variants and genes. The majority of association signals are located in regions of the genome that are not coding for proteins, which poses a challenge for deciphering the functional role of target genes. Moreover, due to the potent capacity of GWAS to detect subtle effects, the identified genes may not necessarily belong to the core pathways that regulate the phenotype [82]. Therefore, identifying causal relationships between variant genes and phenotypes is extremely complex and challenging. A growing number of functional datasets and genomics resources (e.g., EN-CODE [83] and the GTEx [84], Epigenome RoadMap [85], FANTOM5 [86]) make it possible to combine GWAS findings with functional genomics data to advance variant functional annotations. What's more, improved bioinformatics approaches such as computer annotation of gene regulatory regions [87–89], enrichment of causal variants in epigenome annotation [90,91], colocalization of GWAS and expression quantitative trait locus (eQTL) signals [92,93], gene expression prediction [94,95], and fine mapping of causal variants [96,97] have provided ideas for downstream analysis of GWAS. Several approaches are currently being studied to establish the connection between regulatory elements and their target genes, such as the 3C-based identification of chromatin loops and the CRISPR/Cas9 system [98,99]. These innovative approaches with integration with GWAS will further increase the functional understanding of disease-related genes and facilitate biological discoveries translation. As an illustration, the investigators devised a method to decipher the molecular mechanism of disease-linked variants detected by GWAS. In this study, the Activity-by-Contact model was utilized to generate enhancer-gene profiles in cells and tissues, resulting in the identification of 5026 GWAS signals associated with 2249 genes. For inflammatory bowel disease (IBD), this map has revealed the mechanism of genomic regulation by demonstrating that enhancers harboring IBD risk variants alter PPIF expression, thereby altering immune cell mitochondrial function [100].

2.2.2 Drug Repurposing and Adverse Reaction Prediction

GWAS is gaining attention in the pharmaceutical industry. Researchers have discovered that there is a likelihood of at least twice as much in receiving approval for drugs supported by GWAS, especially in cases where causal genes have been identified [101,102]. The prolonged duration and exorbitant cost of novel drug development and clinical trials have prompted researchers to shift their attention towards exploring alternative indications for existing drugs. GWAS has been used successfully to discover drug repurposing opportunities [103,104]. Although GWAS do not typically provide direct information regarding causative genes or disease mechanisms, the identification of causal variants and biological pathways associated with diseases can be achieved through a combination of GWAS findings,

subsequent bioinformatics approaches, and functional experiments. This strategy enables the matching of new drug-disease relationships using a catalog of known drug targets and the disease gene associations identified by GWAS. The high efficiency of drug repurposing satisfies the urgent demand for Corona Virus Disease 2019 (COVID-19) treatment. The *TYK2* gene, which has been identified by GWAS as associated with host-driven inflammatory lung injury, is causally linked to critical illness COVID-19 and represents a promising therapeutic target for this disease [105]. *TYK2* belongs to the Janus kinase (JAK) family, and baricitinib is a JAK inhibitor that specifically inhibits *TYK2* expression [106]. Baricitinib was initially approved for the treatment of rheumatoid arthritis in 2018 [107]. After two phase III clinical trials that were randomized, double-blind, and placebo-controlled confirmed the significant reduction in mortality of patients hospitalized with COVID-19 by baricitinib, it has now been approved by the Food and Drug Administration (FDA) for treating severe cases of COVID-19 [108–110]. Furthermore, GWAS aids in the prediction of adverse drug reactions and enables the screening of safer drug targets, thereby mitigating foreseeable safety events and enhancing the efficiency of drug development [111].

2.2.3 Enabling Precision Medicine Programs

GWAS is anticipated to offer valuable guidance for precise therapy by identifying genetic variants that are associated with drug response. The observation that patients with the same cancer type and receiving the same chemotherapeutic agent often exhibit varying degrees of response suggests that drug regimens based solely on disease phenotype are not only inefficient in terms of drug utilization but may also exacerbate patient outcomes due to delayed treatment. Accurately identifying patients who are sensitive to a specific chemotherapeutic agent is crucial for improving clinical outcomes. A genome-wide association study (GWAS) has identified two variants in *ADCY1* that may affect the responsiveness of non-small cell lung cancer patients to platinum-based chemotherapy. *In vitro* cellular experiments have confirmed that high expression of *ADCY1* is associated with increased sensitivity to cisplatin [112]. More evidence is required to elucidate the mechanisms underlying the modulation of chemotherapeutic drug sensitivity by the two SNPs and to validate the clinical utility of genotype-guided chemotherapeutic drug selection. Additionally, a GWAS study identified a *CYP2C19* gene variant associated with poor clopidogrel efficacy [113]. Moreover, clinical trials designed based on this research finding have demonstrated the benefits of antiplatelet therapy regimens formulated according to the *CYP2C19* genotype for patients undergoing percutaneous coronary intervention [114]. The Clinical Pharmacogenetics Implementation Consortium (CPIC) has developed a series of dosing guidelines that incorporate genetic variants associated with drug response in an evidence-based and rigorous manner, aimed

at assisting clinicians in comprehending the significance of genetic test results and optimizing drug therapy [115]. These guidelines include the clopidogrel regimen based on the *CYP2C19* genotype, as previously mentioned. Therefore, GWAS provides a novel approach to achieve precise drug selection in an agnostic manner and allows the integration of research findings into clinical practice.

2.2.4 Disease Risk Prediction

The genetic variants identified through GWAS may be valuable in discerning individuals who have an elevated risk of specific diseases. According to a study, the *LOXL1* gene contains two nonsynonymous SNPs that contribute 99% of the population-attributable risk for exfoliative glaucoma [116]. It is generally the case, however, that individual common variants exhibit low effect sizes and collectively they account for a moderate portion of heritability [117]. The persistent issue of "missing heritability" hinders the clinical applicability of GWAS in predicting diseases. Understanding the origins of deletion heritability is crucial for investigating genetic architecture and phenotypes of complex diseases. This can be attributed to several factors: (1) rare variants serve as the primary source of deletion heritability [118]; (2) neglecting the contribution of small genetic loci that are not captured by GWAS to the phenotype, and (3) intricate gene-environment interactions [119]. When considering associations beyond those of genome-wide significance, polygenic risk scores (PRS), derived from GWAS are anticipated to address the limitations of GWAS in predicting phenotypes through genetic effects.

2.3 Predicting Disease Risk Based on Genetic Variation Effects

PRS can improve disease risk prediction, guide treatment decisions, and even refine prognosis when we are not limited to variants of genome-wide significance. PRS quantifies the contribution of the genome to complex disease risk assessment by combining the cumulative effects of genome variants [120]. For coronary artery disease (CAD), PRS has the potential to identify a significantly larger number of individuals at equivalent or higher risk than rare single mutation carriers [121], suggesting that it may offer substantial benefits for high-risk individuals through the promotion of healthy lifestyles and treatment with statins [122,123]. Damask *et al*. [124] found that high PRS in CAD was correlated to an elevated risk of recurrent major adverse cardiovascular events (MACE) following acute coronary syndrome. The reduction of MACE risk by alirocumab treatment was more pronounced in patients with higher PRS than those with lower PRS. However, there are challenges to incorporating PRS into medical decision-making in the real world. Complex diseases arise from the interplay of genetic and environmental factors. Therefore, to maximize predictive value, risk models should incorporate not only genetic but also environmental and other relevant factors that contribute to the phenotype. Modifiable environmental factors are precisely one breakthrough point in reducing disease susceptibility. For instance, environmental factors like activity levels, economic status, and dietary composition can modify the epistatic effects of pathogenic mutations on obesity as well as the methylation patterns of obesity genes [125]. Obesity, being a contributor to cardiovascular disease and diabetes, can be prevented or treated through lifestyle management interventions. Several scholars have suggested an analytical pipeline incorporating evaluations of SNP post-translational effects based on tissues and cell types, along with clinical phenotypes, into PRS calculations. This offers a practical method for enhancing the predictive accuracy of PRS [126]. Considering the current limited predictive diagnostic capacity of the PRS, it is not recommended to use it as a standalone diagnostic tool [127]. However, incorporating the PRS into an overall individual assessment strategy can provide a more comprehensive evaluation of disease risk. A successful example of considering multiple factors is the Canrisk tool for risk prediction of breast and ovarian cancer. This tool incorporates PRS, rare pathogenic variants in susceptibility genes, lifestyle factors, and clinical phenotypes to enhance the accuracy of disease prediction models [128]. In addition, whole genome sequencing (WGS) captures a greater amount of total genetic variation as well as rare variants compared to SNP genotyping [129]. Therefore, future larger scale GWAS based on the genomic data generated by WGS will significantly enhance the predictive power of PRS. Moreover, it is imperative to acknowledge that PRS tends to lose predictive accuracy when used across populations as a result of multiple factors, such as allele frequency differences and changes in polymorphism effect sizes [130]. Most PRS are currently calculated using information on individuals of European ancestry, with limited transferability across populations, which is likely to exacerbate health inequities in the genomic era [131,132]. In a study, researchers integrated genomic data from multiple biobanks to investigate the PRS of lifespan biomarkers across diverse ethnic populations. The impact of body mass index (BMI) on lifespan reduction in the Japanese population differs from that observed in the European population, indicating a need for further exploration into this racial disparity regarding obesity-related health burdens. Therefore, enhancing the diversity of study populations is crucial for advancing our comprehension of human genetic variation [133,134] and promoting healthcare equity.

2.4 A More Efficient Approach to Data Integration—Artificial Intelligence and Machine Learning

Identifying the right therapeutic targets requires an understanding of multilevel pathogenic mechanisms, including gene regulation in noncoding regions [135], DNA methylation status [136], and RNA splicing [137]. Therefore, to achieve better clinical transformation of genomics, it is not enough to be limited to genomics itself, but also to

combine other omics knowledge, such as transcriptomics, epigenomics, etc., and accurate disease subtypes are also individualized medically necessary. The development of electronic medical records, the popularization of wearable devices, and the explosive growth of high-throughput sequencing have generated an increasing amount of medical big data. Instead of being overwhelmed by piles of data, we should explore the mysteries behind a large amount of heterogeneous data to serve the cause of human health. Artificial intelligence (AI) and machine learning (ML) present the possibility to achieve this goal.

AI refers to computer output generated by imitating human behavior. ML is a subset of artificial intelligence. The essence of it lies in utilizing algorithms to learn from vast amounts of data, from which a model is constructed, and subsequently verifying and refining the model. A probability distribution can be used to estimate the most probable successful decision, therefore facilitating the identification and prediction of new data [138]. The strengths of machine learning lie in its versatility, extensibility, automatability, and capability to deal with complex dimensional datasets. ML models are primarily categorized into three types: supervised learning, unsupervised learning, and reinforcement learning. The process of supervised learning involves utilizing labeled data as the training target to facilitate algorithmic training and model fitting, ultimately enabling the model to accurately predict outcomes for new datasets [139]. For example, ML may predict the postoperative survival of cancer patients based on the patient's genetic and phenotypic characteristics. On the contrary, unsupervised learning is used to process unlabeled data, by identifying patterns and regularities in data, clustering unlabeled data, discovering new associations, and data dimensionality reduction [140]. Reinforcement learning learns from trial and error and aims to reduce prediction errors, similar to conditioning mechanisms in psychology [141]. Notably, deep learning is gaining increasing popularity, using artificial neural networks to automatically extract data features, surpassing traditional machine learning applied to natural language processing [142] and computer vision [143].

2.4.1 Exploration of Disease Mechanisms with ML

Variants associated with diseases are typically located in non-coding genes, and it has been challenging to develop methods that can account for the impact of mutations in these genes. Machine learning provides novel insights for predicting the impact of noncoding mutations on disease. In a study, researchers utilized a deep learning model to analyze genome-wide data from 1790 families with autism spectrum disorder (ASD) and identified the contribution of novel noncoding variants on ASD pathogenesis by comparing anticipated transcriptional and post-transcriptional regulatory consequences of these variants in probands versus unaffected matched siblings [144]. In addition, Jaganathan *et al.* [145] successfully trained a deep learning model to identify non-coding mutations in rare genetic diseases by predicting the splicing of pre-mRNAs from genomic sequences. This shows that machine learning may help increase our understanding of the mechanisms of disease occurrence.

It has been shown that one-third of the genes in the human genome are co-regulated by microRNAs (miRNAs) [146]. MiRNAs are loaded into the Argonaute (AGO) family of proteins to generate miRNA-induced silencing complexes (miRISCs), which are base-paired with target mRNAs and regulate genes through mRNA cleavage or translational repression expression [146]. MiRNAs contribute to the pathogenesis of complex human diseases through post-transcriptional regulation of gene expression, thus serving as potential biomarkers for disease treatment [147]. The first step in clarifying miRNA pathogenesis is to accurately identify miRNA associations with specific diseases through experimental methods; however, this can be time-consuming and expensive given the numerous miRNA disease combinations. Computational modeling addresses this problem by identifying the most likely relevant miRNA-disease associations (MDA) for validation in biological experiments, thereby accelerating new MDA discovery. To fully utilize large-scale multi-source heterogeneous datasets and improve the accuracy of MDA prediction, researchers have started to focus on the potential of machine learning-based methods [148], including various classifiers (e.g., decision trees, support vector machines, plain Bayes, and neural networks) and matrix decomposition techniques, which simplify high-dimensional matrices into a few low-dimensional matrices. Computer-based miRNA analysis tools can predict the association of miRNAs with cellular functions and diseases and predict miRNA target genes and binding sites at the molecular level. For example, experimentally validated reports of miRNA-target interactions can be obtained using MiRTarBase [149]; miRDB, constructed by machine learning algorithms based on support vector machines, is an online resource for miRNA target prediction and functional annotation [150]. Numerous studies have previously described the available tools [151,152]. In recent years, deep learning methods have emerged as promising new approaches to improve miRNA target group prediction. One study reported deep learning-based miRAW, which predicts miRNA targets by analyzing the entire miRNA transcripts and outperforms existing prediction methods in comparisons using independent datasets [153]. Machine learning-based research on the pathogenic mechanism of miRNA has been applied not only to cancer but also shown good prediction performance in other common diseases, such as coronary heart disease and dementia [154]. The investigators analyzed plasma extracellular vesicles (EVs) miRNA sequencing data from SCD patients and normal patients using bioinformatics online tools and statistical methods. The investigators used TargetScan and miRanda to identify EVs-miRNA targets, and functional enrichment analysis using

Metascape to screen plasma EVsmiR-208b-3p and miR-143-3p as promising biomarkers for predicting SCD in patients with acute coronary syndrome (ACS) [155]. In a recently published article, the concept of theranomiRNAs, miRNAs for diagnostic and therapeutic purposes, was mentioned for the first time [156], which reaffirms the role of miRNAs in the diagnosis and treatment of diseases, and is a new direction in the translation of basic research into clinical practice, which is expected to play an influential role in precision medicine in the future.

2.4.2 Diagnosing Diseases with ML

63.4% of the untranslated region (UTR) variants in the ClinVar database are classified as "variants of uncertain significance" (VUS) [157]. In this context, in addition to functional annotation via databases, it is essential to assess the pathogenicity of genetic variants in non-coding regions. Constructing a pathogenicity prediction framework using non-coding variants in the human genome could provide a more comprehensive understanding of disease biology and reveal opportunities to develop new therapeutic targets. Deep learning has become a powerful tool for the functional study of non-coding mutations because genomics research is inherently characterized by features such as sequence local dependence and long-range correlation, and its large-scale and deep data characteristics fit well with the logic of the work of convolutional neural networks algorithms (CNNs). Currently, tools based on the CNN framework for prioritizing mutations in non-coding regions include DeepBind, DeepSEA, Basset, DanQ, Basenji, etc [158]. In 2020, the DeepFun model integrates data from ENCODE and Roadmap on top of existing CNN models to present a dense human-to-human epigenome mapping [159]. Successive upgrades of the model have contributed to increased accuracy in predicting the pathogenicity of non-coding region variants. The recently proposed "Junk" Annotation genome-wide Residual Variation Intolerance Score (JARVIS), which captures previously unavailable human pedigree constraint information, outperforms other human pedigree-specific scores [160]. JARVIS introduces a genome-wide residual variant intolerance score (gwRVIS) and incorporates primary genome sequence information and additional functional genome annotations to prioritize regions in the non-coding genome that may be more likely to be associated with clinically relevant effects when mutated.

Through the integration of genetic data with medical imaging, which is the largest source of data in the healthcare system, we can gain a more profound understanding of how genes influence organ morphology or function. Hallgrímsson et al. [161] used machine learning models to identify 3D facial images and thereby automatically diagnose the syndrome. The approach yielded a balanced accuracy rate of 73% when applied to a sample size of 7057 subjects. ML is capable of assisting in diagnosing diseases by utilizing multi-omics data. Khanna et al. [162] have developed a multivariate multimodal model that utilizes genomic data, imaging data, and biomarkers to predict the time of Alzheimer's disease diagnosis. In this model, a Bayesian network model (an algorithm of machine learning) aims to reveal the interactions between genetic variants, biological pathways, and imaging-related features across biological scales.

2.5 How Can Genomic Data Translate into Clinical Practice?

Clinical transformation of genomics translates discoveries of chromosome/chromatin, DNA and RNA sequences, structure, and function into clinical applications for predicting, diagnosing, monitoring human disease-specific phenomena, while characterizing severity, duration, stage, and response to therapy. Progress in basic research discovery and translation has remained closely linked to clinical observations for most of its history, but there has been a serious mismatch between the phenomenal advances in basic biomedical sciences and the slow onset of translational medicine in the past 40 years. As a result, biomedical research in the third decade of the 21st century has been plagued by a disconnect between basic biomedical science and clinical practice. Although encouraging advances have been made, for example, in 2004, studies demonstrated the efficacy of gefitinib, an EGFR kinase inhibitor, in non-small cell lung cancer with specific gene mutations [163], and in 2011, ivacaftor, a cystic fibrosis transmembrane conductance regulator enhancer, was shown to be useful in the treatment of cystic fibrosis [164]. In complex polygenic disease disorders, the association of low density lipoprotein (LDL) cholesterol and triglyceride levels with CAD has been found to reflect a causal relationship through genomic data analysis methods, becoming a clinical marker and preventive therapeutic target for CAD [165]. Notably, the ability to effectively treat diseases as we now understand them is still limited: Of the approximately 8000 diseases that affect humans, fewer than 600 have any regulatory-approved treatments, most of which merely relieve symptoms. Effectively translating ideas from the laboratory into interventions for clinical practice currently takes more than 20 years and has a success rate of less than 1 percent [166]. Therefore, the use of data analysis methods in the big data era and the vast amount of genomics data from the level of the molecular mechanisms of disease will hopefully facilitate the efficient translation of genomics data.

As a method of medical practice that is data-driven, precision medicine is a field that takes into account relevant medical, genetic, behavioral, and environmental information about individuals to accurately predict disease risk in healthy populations while providing targeted treatment options for patients. Genomics big data is an important tool to achieve precision medicine. Variants in the human genome such as SNPs, insertions and deletions, structural variants, and copy number variants play an impor-

tant role in disease onset, progression, and performance status. Therefore, the clinical translation of genomics big data reflects the goal of precision medicine, which usually relies on genomics analysis tools to mine the information behind the data that is closely related to disease diagnosis and treatment. Genomics technology and science are closely linked to the extent to which genomic information is used in medicine. Advances in high-throughput sequencing have generated vast amounts of genomic data for scientists to study gene structure and function variations. It can also serve as a medium for genomic insights to penetrate the clinic and become an essential tool for personalized therapy. A genome-wide association study uses extensive genomic data to investigate genotype-phenotype associations, and the findings may provide powerful support for drug discovery. Polygenic risk scores, developed based on GWAS, offer new ideas for disease diagnosis and personalized therapy. In the face of increasing big data, there is a need for high-performance data analysis methods to integrate a great deal of multidimensional data, and to discover patterns across different types of data to guide the improvement of high medical standards. Artificial intelligence and machine learning methods are helping data scientists to overcome this challenge.

3. Conclusions

Advances in genomic science technology, and research have enabled the translation of genomics from basic to clinical settings. The interplay of sequencing technologies, functional genomics, and other emerging technologies (e.g., artificial intelligence) has provided new perspectives and new understandings for the exploration of disease pathogenesis. Currently, some genomics data have been successfully applied in various aspects of clinical screening, diagnosis, and treatment to provide decision support for personalized healthcare issues. Therefore, there is a growing expectation to integrate genomics into mainstream medicine. Nevertheless, we need to clearly understand the challenges facing broader clinical integration of genomics. Genomic data of individuals hide private information related to life and health, which becomes a limiting factor for data sharing. A global medical data-sharing mechanism should be established under the premise of fully safeguarding data privacy and security, developing data science technology with data protection capability, and promoting the maximum utilization of medical data value. Genomics research results should be validated by clinical trials to confirm their true clinical utility. For example, attention should be paid to whether technologies with disease-predictive capabilities actually promote healthier changes in patient behavior or just increase excessive anxiety. Finally, improving the genomics literacy of health industry practitioners and promoting the mass dissemination and popularization of genomics knowledge is crucial to the industrial implementation of genomics data in the medical field. Over-

all, medical practice should always be patient-centred, and genomic-level data information can help researchers to be more focused, rigorous, and scientific in their experimental design. In the process of clinical implementation, researchers should pay attention to the feasibility (especially the cost and invasive operation that patients are concerned about), effectiveness, and universality (to minimise the inequality of medical treatment among races) of the research results. The collaboration of experts across diverse industries is imperative in fostering the harmonious advancement of fundamental biological research and translational medicine practice, thereby making significant contributions to the pursuit of human well-being.

Author Contributions

YZ and JY contributed to the formal analysis, methodology, investigation, and writing the original draft. XX participated in conceptualization, and writing review & editing. FJ and CW designed the work, reviewed it critically, contributed to supervision and funding acquisition. All authors read and approved the final manuscript. All authors have participated sufficiently in the work and agreed to be accountable for all aspects of the work.

Ethics Approval and Consent to Participate

Not applicable.

Acknowledgment

Not applicable.

Funding

Conflict of Interest

The authors declare no conflict of interest. Given his/her role as Guest Editor, Feng Jiang had no involvement in the peer-review of this article and has no access to information regarding its peer review. Full responsibility for the editorial process for this article was delegated to Graham Pawelec.

References

[1] Stephens ZD, Lee SY, Faghri F, Campbell RH, Zhai C, Efron MJ, et al. Big Data: Astronomical or Genomical? PLoS Biology. 2015; 13: e1002195.

[2] Khodadadian A, Darzi S, Haghi-Daredeh S, Sadat Eshaghi F, Babakhanzadeh E, Mirabutalebi SH, et al. Genomics and Transcriptomics: The Powerful Technologies in Precision Medicine. International Journal of General Medicine. 2020; 13: 627–640.

[3] Agerbo E, Sullivan PF, Vilhjálmsson BJ, Pedersen CB, Mors O, Børglum AD, et al. Polygenic Risk Score, Parental Socioeconomic Status, Family History of Psychiatric Disorders, and the

Risk for Schizophrenia: A Danish Population-Based Study and Meta-analysis. JAMA Psychiatry. 2015; 72: 635–641.

[4] Chen Z, Chen J, Collins R, Guo Y, Peto R, Wu F, *et al*. China Kadoorie Biobank of 0.5 million people: survey methods, baseline characteristics and long-term follow-up. International Journal of Epidemiology. 2011; 40: 1652–1666.

[5] Collins R. What makes UK Biobank special? Lancet (London, England). 2012; 379: 1173–1174.

[6] Metspalu A, Köhler F, Laschinski G, Ganten D, Roots I. The Estonian Genome Project in the context of European genome research. Deutsche Medizinische Wochenschrift (1946). 2004; 129: S25–S28.

[7] Nagai A, Hirata M, Kamatani Y, Muto K, Matsuda K, Kiyohara Y, *et al*. Overview of the BioBank Japan Project: Study design and profile. Journal of Epidemiology. 2017; 27: S2–S8.

[8] Stark Z, Dolman L, Manolio TA, Ozenberger B, Hill SL, Caulfied MJ, *et al*. Integrating Genomics into Healthcare: A Global Responsibility. American Journal of Human Genetics. 2019; 104: 13–20.

[9] Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, *et al*. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature. 2020; 581: 434–443.

[10] Zerhouni EA. Translational and clinical science–time for a new vision. The New England Journal of Medicine. 2005; 353: 1621–1623.

[11] Besser J, Carleton HA, Gerner-Smidt P, Lindsey RL, Trees E. Next-generation sequencing technologies and their application to the study and control of bacterial infections. Clinical Microbiology and Infection: the Official Publication of the European Society of Clinical Microbiology and Infectious Diseases. 2018; 24: 335–341.

[12] Mellis R, Chandler N, Chitty LS. Next-generation sequencing and the impact on prenatal diagnosis. Expert Review of Molecular Diagnostics. 2018; 18: 689–699.

[13] Pipis M, Rossor AM, Laura M, Reilly MM. Next-generation sequencing in Charcot-Marie-Tooth disease: opportunities and challenges. Nature Reviews. Neurology. 2019; 15: 644–656.

[14] Spaulding TP, Stockton SS, Savona MR. The evolving role of next generation sequencing in myelodysplastic syndromes. British Journal of Haematology. 2020; 188: 224–239.

[15] Menko FH, Monkhorst K, Hogervorst FBL, Rosenberg EH, Adank MA, Ruijs MWG, *et al*. Challenges in breast cancer genetic testing. A call for novel forms of multidisciplinary care and long-term evaluation. Critical Reviews in Oncology/hematology. 2022; 176: 103642.

[16] National Collaborating Centre for Cancer (UK). Familial Breast Cancer: Classification and Care of People at Risk of Familial Breast Cancer and Management of Breast Cancer and Related Risks in People with a Family History of Breast Cancer. 2013.

[17] Griffith GL, Edwards RT, Gray J. Cancer genetics services: a systematic review of the economic evidence and issues. British Journal of Cancer. 2004; 90: 1697–1703.

[18] Gabai-Kapara E, Lahad A, Kaufman B, Friedman E, Segev S, Renbaum P, *et al*. Population-based screening for breast and ovarian cancer risk due to BRCA1 and BRCA2. Proceedings of the National Academy of Sciences of the United States of America. 2014; 111: 14205–14210.

[19] Manchanda R, Loggenberg K, Sanderson S, Burnell M, Wardle J, Gessler S, *et al*. Population testing for cancer predisposing BRCA1/BRCA2 mutations in the Ashkenazi-Jewish community: a randomized controlled trial. Journal of the National Cancer Institute. 2014; 107: 379.

[20] Møller P, Hagen AI, Apold J, Maehle L, Clark N, Fiane B, *et al*. Genetic epidemiology of BRCA mutations–family history detects less than 50% of the mutation carriers. European Journal of Cancer (Oxford, England: 1990). 2007; 43: 1713–1717.

[21] Petelin L, Hossack L, Shanahan M, Mitchell G, Liew D, James PA, *et al*. Cost-effectiveness of long-term clinical management of BRCA pathogenic variant carriers. Genetics in Medicine: Official Journal of the American College of Medical Genetics. 2020; 22: 831–839.

[22] Manchanda R, Burnell M, Gaba F, Desai R, Wardle J, Gessler S, *et al*. Randomised trial of population-based BRCA testing in Ashkenazi Jews: long-term outcomes. BJOG: an International Journal of Obstetrics and Gynaecology. 2020; 127: 364–375.

[23] Empana JP, Lerner I, Valentin E, Folke F, Böttiger B, Gislason G, *et al*. Incidence of Sudden Cardiac Death in the European Union. Journal of the American College of Cardiology. 2022; 79: 1818–1827.

[24] Lynge TH, Nielsen JL, Risgaard B, van der Werf C, Winkel BG, Tfelt-Hansen J. Causes of sudden cardiac death according to age and sex in persons aged 1-49 years. Heart Rhythm. 2023; 20: 61–68.

[25] Lynge TH, Risgaard B, Banner J, Nielsen JL, Jespersen T, Stampe NK, *et al*. Nationwide burden of sudden cardiac death: A study of 54,028 deaths in Denmark. Heart Rhythm. 2021; 18: 1657–1665.

[26] Broendberg AK, Christiansen MK, Nielsen JC, Pedersen LN, Jensen HK. Targeted next generation sequencing in a young population with suspected inherited malignant cardiac arrhythmias. European Journal of Human Genetics: EJHG. 2018; 26: 303–313.

[27] Chanavat V, Janin A, Millat G. A fast and cost-effective molecular diagnostic tool for genetic diseases involved in sudden cardiac death. Clinica Chimica Acta; International Journal of Clinical Chemistry. 2016; 453: 80–85.

[28] Kingsmore SF, Smith LD, Kunard CM, Bainbridge M, Batalov S, Benson W, *et al*. A genome sequencing system for universal newborn screening, diagnosis, and precision medicine for severe genetic diseases. American Journal of Human Genetics. 2022; 109: 1605–1619.

[29] Newson AJ. The promise of public health ethics for precision medicine: the case of newborn preventive genomic sequencing. Human Genetics. 2022; 141: 1035–1043.

[30] Cancer Genome Atlas Research Network, Kandoth C, Schultz N, Cherniack AD, Akbani R, Liu Y, *et al*. Integrated genomic characterization of endometrial carcinoma. Nature. 2013; 497: 67–73.

[31] Yang L, Ye F, Bao L, Zhou X, Wang Z, Hu P, *et al*. Somatic alterations of TP53, ERBB2, PIK3CA and CCND1 are associated with chemosensitivity for breast cancers. Cancer Science. 2019; 110: 1389–1400.

[32] Cheema PK, Gomes M, Banerji S, Joubert P, Leighl NB, Melosky B, *et al*. Consensus recommendations for optimizing biomarker testing to identify and treat advanced *EGFR*-mutated non-small-cell lung cancer. Current Oncology (Toronto, Ont.). 2020; 27: 321–329.

[33] Mosele F, Remon J, Mateo J, Westphalen CB, Barlesi F, Lolkema MP, *et al*. Recommendations for the use of next-generation sequencing (NGS) for patients with metastatic cancers: a report from the ESMO Precision Medicine Working Group. Annals of Oncology: Official Journal of the European Society for Medical Oncology. 2020; 31: 1491–1505.

[34] Jahr S, Hentze H, Englisch S, Hardt D, Fackelmayer FO, Hesch RD, *et al*. DNA fragments in the blood plasma of cancer patients: quantitations and evidence for their origin from apoptotic and necrotic cells. Cancer Research. 2001; 61: 1659–1665.

[35] Bak SE, Kim H, Ho JY, Cho EH, Lee J, Youn SM, *et al*. Shallow Whole-Genome Sequencing of Cell-Free DNA (cfDNA) Detects Epithelial Ovarian Cancer and Predicts Patient Prognosis. Cancers. 2023; 15: 530.

[36] Sun X, Feng W, Cui P, Ruan R, Ma W, Han Z, *et al*. Detection and monitoring of HBV-related hepatocellular carcinoma from plasma cfDNA fragmentation profiles. Genomics. 2022;

114: 110502.

[37] Wu TH, Hsiue EHC, Yang JCH. Opportunities of circulating tumor DNA in lung cancer. Cancer Treatment Reviews. 2019; 78: 31–41.

[38] Leighl NB, Page RD, Raymond VM, Daniel DB, Divers SG, Reckamp KL, et al. Clinical Utility of Comprehensive Cell-free DNA Analysis to Identify Genomic Biomarkers in Patients with Newly Diagnosed Metastatic Non-small Cell Lung Cancer. Clinical Cancer Research: an Official Journal of the American Association for Cancer Research. 2019; 25: 4691–4700.

[39] Rolfo C, Mack PC, Scagliotti GV, Baas P, Barlesi F, Bivona TG, et al. Liquid Biopsy for Advanced Non-Small Cell Lung Cancer (NSCLC): A Statement Paper from the IASLC. Journal of Thoracic Oncology: Official Publication of the International Association for the Study of Lung Cancer. 2018; 13: 1248–1268.

[40] Nabet BY, Esfahani MS, Moding EJ, Hamilton EG, Chabon JJ, Rizvi H, et al. Noninvasive Early Identification of Therapeutic Benefit from Immune Checkpoint Inhibition. Cell. 2020; 183: 363–376.e13.

[41] Yamamoto G, Kikuchi M, Kobayashi S, Arai Y, Fujiyoshi K, Wakatsuki T, et al. Routine genetic testing of lung cancer specimens derived from surgery, bronchoscopy and fluid aspiration by next generation sequencing. International Journal of Oncology. 2017; 50: 1579–1589.

[42] Dar P, Shani H, Evans MI. Cell-free DNA: Comparison of Technologies. Clinics in Laboratory Medicine. 2016; 36: 199–211.

[43] Wapner RJ, Babiarz JE, Levy B, Stosic M, Zimmermann B, Sigurjonsson S, et al. Expanding the scope of noninvasive prenatal testing: detection of fetal microdeletion syndromes. American Journal of Obstetrics and Gynecology. 2015; 212: 332.e1–9.

[44] 100,000 Genomes Project Pilot Investigators, Smedley D, Smith KR, Martin A, Thomas EA, McDonagh EM, et al. 100,000 Genomes Pilot on Rare-Disease Diagnosis in Health Care - Preliminary Report. The New England Journal of Medicine. 2021; 385: 1868–1880.

[45] Splinter K, Adams DR, Bacino CA, Bellen HJ, Bernstein JA, Cheatle-Jarvela AM, et al. Effect of Genetic Diagnosis on Patients with Previously Undiagnosed Disease. The New England Journal of Medicine. 2018; 379: 2131–2139.

[46] Taylor JC, Martin HC, Lise S, Broxholme J, Cazier JB, Rimmer A, et al. Factors influencing success of clinical genome sequencing across a broad spectrum of disorders. Nature Genetics. 2015; 47: 717–726.

[47] Boycott K, Hartley T, Adam S, Bernier F, Chong K, Fernandez BA, et al. The clinical application of genome-wide sequencing for monogenic diseases in Canada: Position Statement of the Canadian College of Medical Geneticists. Journal of Medical Genetics. 2015; 52: 431–437.

[48] Manickam K, McClain MR, Demmer LA, Biswas S, Kearney HM, Malinowski J, et al. Exome and genome sequencing for pediatric patients with congenital anomalies or intellectual disability: an evidence-based clinical guideline of the American College of Medical Genetics and Genomics (ACMG). Genetics in Medicine: Official Journal of the American College of Medical Genetics. 2021; 23: 2029–2037.

[49] Johnson DB, Dahlman KH, Knol J, Gilbert J, Puzanov I, Means-Powell J, et al. Enabling a genetically informed approach to cancer medicine: a retrospective evaluation of the impact of comprehensive tumor profiling using a targeted next-generation sequencing panel. The Oncologist. 2014; 19: 616–622.

[50] Meric-Bernstam F, Brusco L, Shaw K, Horombe C, Kopetz S, Davies MA, et al. Feasibility of Large-Scale Genomic Testing to Facilitate Enrollment Onto Genomically Matched Clinical Trials. Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology. 2015; 33: 2753–2762.

[51] Tan O, Shrestha R, Cunich M, Schofield DJ. Application of next-generation sequencing to improve cancer management: A re-view of the clinical effectiveness and cost-effectiveness. Clinical Genetics. 2018; 93: 533–544.

[52] Schwaederle M, Zhao M, Lee JJ, Lazar V, Leyland-Jones B, Schilsky RL, et al. Association of Biomarker-Based Treatment Strategies With Response Rates and Progression-Free Survival in Refractory Malignant Neoplasms: A Meta-analysis. JAMA Oncology. 2016; 2: 1452–1459.

[53] Yoshinami Y, Shoji H. Recent advances in immunotherapy and molecular targeted therapy for gastric cancer. Future Science OA. 2023; 9: FSO842.

[54] Tewari KS, Monk BJ, Vergote I, Miller A, de Melo AC, Kim HS, et al. Survival with Cemiplimab in Recurrent Cervical Cancer. The New England Journal of Medicine. 2022; 386: 544–555.

[55] Arabi F, Mansouri V, Ahmadbeigi N. Gene therapy clinical trials, where do we go? An overview. Biomedicine & Pharmacotherapy. 2022; 153: 113324.

[56] Junghans RP. The challenges of solid tumor for designer CAR-T therapies: a 25-year perspective. Cancer Gene Therapy. 2017; 24: 89–99.

[57] Evans CH, Ghivizzani SC, Robbins PD. Osteoarthritis gene therapy in 2022. Current Opinion in Rheumatology. 2023; 35: 37–43.

[58] Kessler JA, Smith AG, Cha BS, Choi SH, Wymer J, Shaibani A, et al. Double-blind, placebo-controlled study of HGF gene therapy in diabetic neuropathy. Annals of Clinical and Translational Neurology. 2015; 2: 465–478.

[59] Sayed N, Allawadhi P, Khurana A, Singh V, Navik U, Pasumarthi SK, et al. Gene therapy: Comprehensive overview and therapeutic applications. Life Sciences. 2022; 294: 120375.

[60] Shim G, Kim D, Le QV, Park GT, Kwon T, Oh YK. Nonviral Delivery Systems for Cancer Gene Therapy: Strategies and Challenges. Current Gene Therapy. 2018; 18: 3–20.

[61] Jafarlou M, Baradaran B, Saedi TA, Jafarlou V, Shanehbandi D, Maralani M, et al. An overview of the history, applications, advantages, disadvantages and prospects of gene therapy. Journal of Biological Regulators and Homeostatic Agents. 2016; 30: 315–321.

[62] Onoyama T, Ishikawa S, Isomoto H. Gastric cancer and genomics: review of literature. Journal of Gastroenterology. 2022; 57: 505–516.

[63] Rojas EA, Gutiérrez NC. Genomics of Plasma Cell Leukemia. Cancers. 2022; 14: 1594.

[64] Ruan H, Wang PC, Han L. Characterization of circular RNAs with advanced sequencing technologies in human complex diseases. Wiley Interdisciplinary Reviews. RNA. 2023; 14: e1759.

[65] Caceres V, Murray T, Myers C, Parbhoo K. Prenatal Genetic Testing and Screening: A Focused Review. Seminars in Pediatric Neurology. 2022; 42: 100976.

[66] Hou W, Shi G, Ma Y, Liu Y, Lu M, Fan X, et al. Impact of preimplantation genetic testing on obstetric and neonatal outcomes: a systematic review and meta-analysis. Fertility and Sterility. 2021; 116: 990–1000.

[67] Sullivan JA, Schoch K, Spillmann RC, Shashi V. Exome/Genome Sequencing in Undiagnosed Syndromes. Annual Review of Medicine. 2023; 74: 489–502.

[68] Xiao F, Yan K, Tang M, Ji X, Hu L, Yang L, et al. Diagnostic utility of rapid sequencing in critically ill infants: a systematic review and meta-analysis. Expert Review of Molecular Diagnostics. 2022; 22: 833–840.

[69] Tan D, Ou T. Research progress and clinical application of the third- generation sequencing techniques. Sheng Wu Gong Cheng Xue Bao. 2022; 38: 3121–3130. (In Chinese)

[70] Pereira R, Oliveira J, Sousa M. Bioinformatics and Computational Tools for Next-Generation Sequencing Analysis in Clinical Genetics. Journal of Clinical Medicine. 2020; 9: 132.

[71] Wijekoon, N., Gonawala, L., Ratnayake, P., Sirisena, D., Gunasekara, H., Dissanayake, A., et al. Gene therapy for selected

neuromuscular and trinucleotide repeat disorders - An insight to subsume South Asia for multicenter clinical trials. IBRO Neuroscience Reports. 2023; 14: 146–153.

[72] ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium. Pan-cancer analysis of whole genomes. Nature. 2020; 578: 82–93.

[73] 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, *et al*. A global reference for human genetic variation. Nature. 2015; 526: 68–74.

[74] Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, *et al*. The UK Biobank resource with deep phenotyping and genomic data. Nature. 2018; 562: 203–209.

[75] Kaplanis J, Samocha KE, Wiel L, Zhang Z, Arvai KJ, Eberhardt RY, *et al*. Evidence for 28 genetic disorders discovered by combining healthcare and research data. Nature. 2020; 586: 757–762.

[76] Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, Haynes C, *et al*. Complement factor H polymorphism in age-related macular degeneration. Science (New York, N.Y.). 2005; 308: 385–389.

[77] Fernández-Santiago R, Sharma M. What have we learned from genome-wide association studies (GWAS) in Parkinson's disease? Ageing Research Reviews. 2022; 79: 101648.

[78] Mahmoud R, Kimonis V, Butler MG. Genetics of Obesity in Humans: A Clinical Review. International Journal of Molecular Sciences. 2022; 23: 11005.

[79] Gerussi A, Soskic B, Asselta R, Invernizzi P, Gershwin ME. GWAS and autoimmunity: What have we learned and what next. Journal of Autoimmunity. 2022; 133: 102922.

[80] Galván-Femenía I, Obón-Santacana M, Piñeyro D, Guindo-Martinez M, Duran X, Carreras A, *et al*. Multitrait genome association analysis identifies new susceptibility genes for human anthropometric variation in the GCAT cohort. Journal of Medical Genetics. 2018; 55: 765–778.

[81] Zhu Z, Li J, Si J, Ma B, Shi H, Lv J, *et al*. A large-scale genome-wide association analysis of lung function in the Chinese population identifies novel loci and highlights shared genetic aetiology with obesity. The European Respiratory Journal. 2021; 58: 2100199.

[82] Boyle EA, Li YI, Pritchard JK. An Expanded View of Complex Traits: From Polygenic to Omnigenic. Cell. 2017; 169: 1177–1186.

[83] ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012; 489: 57–74.

[84] GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. Nature Genetics. 2013; 45: 580–585.

[85] Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, *et al*. Integrative analysis of 111 reference human epigenomes. Nature. 2015; 518: 317–330.

[86] Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, *et al*. An atlas of active enhancers across human cell types and tissues. Nature. 2014; 507: 455–461.

[87] Amariuta T, Ishigaki K, Sugishita H, Ohta T, Koido M, Dey KK, *et al*. Improving the trans-ancestry portability of polygenic risk scores by prioritizing variants in predicted cell-type-specific regulatory elements. Nature Genetics. 2020; 52: 1346–1354.

[88] Amariuta T, Luo Y, Gazal S, Davenport EE, van de Geijn B, Ishigaki K, *et al*. IMPACT: Genomic Annotation of Cell-State-Specific Regulatory Elements Inferred from the Epigenome of Bound Transcription Factors. American Journal of Human Genetics. 2019; 104: 879–895.

[89] Zhou J, Troyanskaya OG. Predicting effects of noncoding variants with deep learning-based sequence model. Nature Methods. 2015; 12: 931–934.

[90] Trynka G, Sandor C, Han B, Xu H, Stranger BE, Liu XS, *et al*. Chromatin marks identify critical cell types for fine mapping complex trait variants. Nature Genetics. 2013; 45: 124–130.

[91] Ulirsch JC, Lareau CA, Bao EL, Ludwig LS, Guo MH, Benner C, *et al*. Interrogation of human hematopoiesis at single-cell and single-variant resolution. Nature Genetics. 2019; 51: 683–693.

[92] Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, *et al*. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genetics. 2014; 10: e1004383.

[93] Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE, *et al*. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. Nature Genetics. 2016; 48: 481–487.

[94] Gamazon ER, Wheeler HE, Shah KP, Mozaffari SV, Aquino-Michaels K, Carroll RJ, *et al*. A gene-based association method for mapping traits using reference transcriptome data. Nature Genetics. 2015; 47: 1091–1098.

[95] Gusev A, Ko A, Shi H, Bhatia G, Chung W, Penninx BWJH, *et al*. Integrative approaches for large-scale transcriptome-wide association studies. Nature Genetics. 2016; 48: 245–252.

[96] Huang H, Fang M, Jostins L, Umićević Mirkov M, Boucher G, Anderson CA, *et al*. Fine-mapping inflammatory bowel disease loci to single-variant resolution. Nature. 2017; 547: 173–178.

[97] Weissbrod O, Hormozdiari F, Benner C, Cui R, Ulirsch J, Gazal S, *et al*. Functionally informed fine-mapping and polygenic localization of complex trait heritability. Nature Genetics. 2020; 52: 1355–1363.

[98] Gasperini M, Hill AJ, McFaline-Figueroa JL, Martin B, Kim S, Zhang MD, *et al*. A Genome-wide Framework for Mapping Gene Regulation via Cellular Genetic Screens. Cell. 2019; 176: 377–390.e19.

[99] Mumbach MR, Rubin AJ, Flynn RA, Dai C, Khavari PA, Greenleaf WJ, *et al*. HiChIP: efficient and sensitive analysis of protein-directed genome architecture. Nature Methods. 2016; 13: 919–922.

[100] Nasser J, Bergman DT, Fulco CP, Guckelberger P, Doughty BR, Patwardhan TA, *et al*. Genome-wide enhancer maps link risk variants to disease genes. Nature. 2021; 593: 238–243.

[101] King EA, Davis JW, Degner JF. Are drug targets with genetic support twice as likely to be approved? Revised estimates of the impact of genetic support for drug mechanisms on the probability of drug approval. PLoS Genetics. 2019; 15: e1008489.

[102] Nelson MR, Tipney H, Painter JL, Shen J, Nicoletti P, Shen Y, *et al*. The support of human genetic evidence for approved drug indications. Nature Genetics. 2015; 47: 856–860.

[103] Pushpakom S, Iorio F, Eyers PA, Escott KJ, Hopper S, Wells A, *et al*. Drug repurposing: progress, challenges and recommendations. Nature Reviews. Drug Discovery. 2019; 18: 41–58.

[104] Tachmazidou I, Hatzikotoulas K, Southam L, Esparza-Gordillo J, Haberland V, Zheng J, *et al*. Identification of new therapeutic targets for osteoarthritis through genome-wide analyses of UK Biobank data. Nature Genetics. 2019; 51: 230–236.

[105] Pairo-Castineira E, Clohisey S, Klaric L, Bretherick AD, Rawlik K, Pasko D, *et al*. Genetic mechanisms of critical illness in COVID-19. Nature. 2021; 591: 92–98.

[106] Zhang K, Ye K, Tang H, Qi Z, Wang T, Mao J, *et al*. Development and Therapeutic Implications of Tyrosine Kinase 2 Inhibitors. Journal of Medicinal Chemistry. 2023; 66: 4378–4416.

[107] Roskoski R, Jr. Properties of FDA-approved small molecule protein kinase inhibitors: A 2023 update. Pharmacological Research. 2023; 187: 106552.

[108] Marconi VC, Ramanan AV, de Bono S, Kartman CE, Krishnan V, Liao R, *et al*. Efficacy and safety of baricitinib for the treatment of hospitalised adults with COVID-19 (COV-BARRIER): a randomised, double-blind, parallel-group, placebo-controlled phase 3 trial. The Lancet. Respiratory Medicine. 2021; 9: 1407–1418.

[109] nullCOVID-19 update: Baricitinib (Olumiant) FDA-approved for treatment of COVID-19. The Medical Letter on Drugs and

Therapeutics. 2022; 64: e2–e3.

[110] Kalil AC, Patterson TF, Mehta AK, Tomashek KM, Wolfe CR, Ghazaryan V, *et al*. Baricitinib plus Remdesivir for Hospitalized Adults with Covid-19. The New England Journal of Medicine. 2021; 384: 795–807.

[111] Nguyen PA, Born DA, Deaton AM, Nioi P, Ward LD. Phenotypes associated with genes encoding drug targets are predictive of clinical trial side effects. Nature Communications. 2019; 10: 1579.

[112] Mao C, Chen J, Zou T, Zhou Y, Liu J, Li X, *et al*. Genome-wide analysis identify novel germline genetic variations in *ADCY1* influencing platinum-based chemotherapy response in non-small cell lung cancer. Acta Pharmaceutica Sinica. B. 2022; 12: 1514–1522.

[113] Shuldiner AR, O'Connell JR, Bliden KP, Gandhi A, Ryan K, Horenstein RB, *et al*. Association of cytochrome P450 2C19 genotype with the antiplatelet effect and clinical efficacy of clopidogrel therapy. JAMA. 2009; 302: 849–857.

[114] Claassens DMF, Vos GJA, Bergmeijer TO, Hermanides RS, van 't Hof AWJ, van der Harst P, *et al*. A Genotype-Guided Strategy for Oral P2Y_12 Inhibitors in Primary PCI. The New England Journal of Medicine. 2019; 381: 1621–1631.

[115] Caudle KE, Klein TE, Hoffman JM, Muller DJ, Whirl-Carrillo M, Gong L, *et al*. Incorporation of pharmacogenomics into routine clinical practice: the Clinical Pharmacogenetics Implementation Consortium (CPIC) guideline development process. Current Drug Metabolism. 2014; 15: 209–217.

[116] Thorleifsson G, Magnusson KP, Sulem P, Walters GB, Gudbjartsson DF, Stefansson H, *et al*. Common sequence variants in the LOXL1 gene confer susceptibility to exfoliation glaucoma. Science (New York, N.Y.). 2007; 317: 1397–1400.

[117] Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, *et al*. Finding the missing heritability of complex diseases. Nature. 2009; 461: 747–753.

[118] Wainschtein P, Jain D, Zheng Z, TOPMed Anthropometry Working Group, NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium, Cupples LA, *et al*. Assessing the contribution of rare variants to complex trait heritability from whole-genome sequence data. Nature Genetics. 2022; 54: 263–273.

[119] Génin E. Missing heritability of complex diseases: case solved? Human Genetics. 2020; 139: 103–113.

[120] Chatterjee N, Shi J, García-Closas M. Developing and evaluating polygenic risk prediction models for stratified disease prevention. Nature Reviews. Genetics. 2016; 17: 392–406.

[121] Khera AV, Chaffin M, Aragam KG, Haas ME, Roselli C, Choi SH, *et al*. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. Nature Genetics. 2018; 50: 1219–1224.

[122] Mega JL, Stitziel NO, Smith JG, Chasman DI, Caulfield M, Devlin JJ, *et al*. Genetic risk, coronary heart disease events, and the clinical benefit of statin therapy: an analysis of primary and secondary prevention trials. Lancet (London, England). 2015; 385: 2264–2271.

[123] Natarajan P, Young R, Stitziel NO, Padmanabhan S, Baber U, Mehran R, *et al*. Polygenic Risk Score Identifies Subgroup With Higher Burden of Atherosclerosis and Greater Relative Benefit From Statin Therapy in the Primary Prevention Setting. Circulation. 2017; 135: 2091–2101.

[124] Damask A, Steg PG, Schwartz GG, Szarek M, Hagström E, Badimon L, *et al*. Patients With High Genome-Wide Polygenic Risk Scores for Coronary Artery Disease May Receive Greater Clinical Benefit From Alirocumab Treatment in the ODYSSEY OUTCOMES Trial. Circulation. 2020; 141: 624–636.

[125] Reddon H, Guéant JL, Meyre D. The importance of gene-environment interactions in human obesity. Clinical Science (London, England: 1979). 2016; 130: 1571–1597.

[126] Vlachakis D, Papakonstantinou E, Sagar R, Bacopoulou F,

Exarchos T, Kourouthanassis P, *et al*. Improving the Utility of Polygenic Risk Scores as a Biomarker for Alzheimer's Disease. Cells. 2021; 10: 1627.

[127] Koch S, Schmidtke J, Krawczak M, Caliebe A. Clinical utility of polygenic risk scores: a critical 2023 appraisal. Journal of Community Genetics. 2023. (online ahead of print)

[128] Carver T, Hartley S, Lee A, Cunningham AP, Archer S, Babb de Villiers C, *et al*. CanRisk Tool-A Web Interface for the Prediction of Breast and Ovarian Cancer Risk and the Likelihood of Carrying Genetic Pathogenic Variants. Cancer Epidemiology, Biomarkers & Prevention: a Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology. 2021; 30: 469–473.

[129] Kierczak M, Rafati N, Höglund J, Gourlé H, Lo Faro V, Schmitz D, *et al*. Contribution of rare whole-genome sequencing variants to plasma protein levels and the missing heritability. Nature Communications. 2022; 13: 2532.

[130] Martin AR, Gignoux CR, Walters RK, Wojcik GL, Neale BM, Gravel S, *et al*. Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations. American Journal of Human Genetics. 2017; 100: 635–649.

[131] Bustamante CD, Burchard EG, De la Vega FM. Genomics for the world. Nature. 2011; 475: 163–165.

[132] Petrovski S, Goldstein DB. Unequal representation of genetic variation across ancestry groups creates healthcare inequality in the application of precision medicine. Genome Biology. 2016; 17: 157.

[133] Bigdeli TB, Genovese G, Georgakopoulos P, Meyers JL, Peterson RE, Iyegbe CO, *et al*. Contributions of common genetic variants to risk of schizophrenia among individuals of African and Latino ancestry. Molecular Psychiatry. 2020; 25: 2455–2467.

[134] Qi Q, Stilp AM, Sofer T, Moon JY, Hidalgo B, Szpiro AA, *et al*. Genetics of Type 2 Diabetes in U.S. Hispanic/Latino Individuals: Results From the Hispanic Community Health Study/Study of Latinos (HCHS/SOL). Diabetes. 2017; 66: 1419–1425.

[135] Rinn JL, Chang HY. Genome regulation by long noncoding RNAs. Annual Review of Biochemistry. 2012; 81: 145–166.

[136] Singal R, Ginder GD. DNA methylation. Blood. 1999; 93: 4059–4070.

[137] Diner I, Hales CM, Bishof I, Rabenold L, Duong DM, Yi H, *et al*. Aggregation properties of the small nuclear ribonucleoprotein U1-70K in Alzheimer disease. The Journal of Biological Chemistry. 2014; 289: 35296–35313.

[138] Ngiam KY, Khor IW. Big data and machine learning algorithms for health-care delivery. The Lancet. Oncology. 2019; 20: e262–e273.

[139] Alloghani M, Al-Jumeily D, Mustafina J, Hussain A, Aljaaf AJ. A systematic review on supervised and unsupervised machine learning algorithms for data science. Supervised and Unsupervised Learning for Data Science. 2020; 3–21.

[140] Krishnan R, Rajpurkar P, Topol EJ. Self-supervised learning in medicine and healthcare. Nature Biomedical Engineering. 2022; 6: 1346–1352.

[141] Laurent PA. The emergence of saliency and novelty responses from Reinforcement Learning principles. Neural Networks: the Official Journal of the International Neural Network Society. 2008; 21: 1493–1499.

[142] Lan Z, Chen M, Goodman S, Gimpel K, Sharma P, Soricut R. Albert: A lite bert for self-supervised learning of language representations. arXiv. 2019. (Preprint)

[143] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition. 2016; 770–778.

[144] Zhou J, Park CY, Theesfeld CL, Wong AK, Yuan Y, Scheckel C, *et al*. Whole-genome deep-learning analysis identifies contribution of noncoding mutations to autism risk. Nature Genetics. 2019; 51: 973–980.

IMR Press

[145] Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF, Darbandi SF, Knowles D, Li YI, *et al*. Predicting Splicing from Primary Sequence with Deep Learning. Cell. 2019; 176: 535–548.e24.

[146] Hammond SM. An overview of microRNAs. Advanced Drug Delivery Reviews. 2015; 87: 3–14.

[147] Hosseinahli N, Aghapour M, Duijf PHG, Baradaran B. Treating cancer with microRNA replacement therapy: A literature review. Journal of Cellular Physiology. 2018; 233: 5574–5588.

[148] Kumar A, Singh SS, Singh K, Biswas B. Link prediction techniques, applications, and performance: A survey. Physica A: Statistical Mechanics and its Applications. 2020; 553: 124289.

[149] Huang HY, Lin YCD, Li J, Huang KY, Shrestha S, Hong HC, *et al*. miRTarBase 2020: updates to the experimentally validated microRNA-target interaction database. Nucleic Acids Research. 2020; 48: D148–D154.

[150] Chen Y, Wang X. miRDB: an online database for prediction of functional microRNA targets. Nucleic Acids Research. 2020; 48: D127–D131.

[151] Nazarov PV, Kreis S. Integrative approaches for analysis of mRNA and microRNA high-throughput data. Computational and Structural Biotechnology Journal. 2021; 19: 1154–1162.

[152] Khatun MS, Alam MA, Shoombuatong W, Mollah MNH, Kurata H, Hasan MM. Recent Development of Bioinformatics Tools for microRNA Target Prediction. Current Medicinal Chemistry. 2022; 29: 865–880.

[153] Pla A, Zhong X, Rayner S. miRAW: A deep learning-based approach to predict microRNA targets by analyzing whole microRNA transcripts. PLoS Computational Biology. 2018; 14: e1006185.

[154] Christou CD, Mitsas AC, Vlachavas I, Tsoulfas G. The Use of Machine Learning in MicroRNA Diagnostics: Current Perspectives. MicroRNA (Shariqah, United Arab Emirates). 2022; 11: 175–184.

[155] Huang S, Zhang J, Wan H, Wang K, Wu J, Cao Y, *et al*. Plasma extracellular vesicles microRNA-208b-3p and microRNA-143-3p as novel biomarkers for sudden cardiac death prediction in acute coronary syndrome. Molecular Omics. 2023; 19: 262–273.

[156] Sessa F, Salerno M, Esposito M, Cocimano G, Pomara C. miRNA Dysregulation in Cardiovascular Diseases: Current Opinion and Future Perspectives. International Journal of Molecular Sciences. 2023; 24: 5192.

[157] Ellingford JM, Ahn JW, Bagnall RD, Baralle D, Barton S, Campbell C, *et al*. Recommendations for clinical interpretation of variants found in non-coding regions of the genome. Genome Medicine. 2022; 14: 73.

[158] Tabarini N, Biagi E, Uva P, Iovino E, Pippucci T, Seri M, *et al*. Exploration of Tools for the Interpretation of Human Non-Coding Variants. International Journal of Molecular Sciences. 2022; 23: 12977.

[159] Pei G, Hu R, Jia P, Zhao Z. DeepFun: a deep learning sequence-based model to decipher non-coding variant effect in a tissue- and cell type-specific manner. Nucleic Acids Research. 2021; 49: W131–W139.

[160] Vitsios D, Dhindsa RS, Middleton L, Gussow AB, Petrovski S. Prioritizing non-coding regions based on human genomic constraint and sequence context with deep learning. Nature Communications. 2021; 12: 1504.

[161] Hallgrímsson B, Aponte JD, Katz DC, Bannister JJ, Riccardi SL, Mahasuwan N, *et al*. Automated syndrome diagnosis by three-dimensional facial imaging. Genetics in Medicine: Official Journal of the American College of Medical Genetics. 2020; 22: 1682–1693.

[162] Khanna S, Domingo-Fernández D, Iyappan A, Emon MA, Hofmann-Apitius M, Fröhlich H. Using Multi-Scale Genetic, Neuroimaging and Clinical Data for Predicting Alzheimer's Disease and Reconstruction of Relevant Biological Mechanisms. Scientific Reports. 2018; 8: 11173.

[163] Paez JG, Jänne PA, Lee JC, Tracy S, Greulich H, Gabriel S, *et al*. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. Science (New York, N.Y.). 2004; 304: 1497–1500.

[164] Ramsey BW, Davies J, McElvaney NG, Tullis E, Bell SC, Dřevínek P, *et al*. A CFTR potentiator in patients with cystic fibrosis and the G551D mutation. The New England Journal of Medicine. 2011; 365: 1663–1672.

[165] Do R, Willer CJ, Schmidt EM, Sengupta S, Gao C, Peloso GM, *et al*. Common variants associated with plasma triglycerides and risk for coronary artery disease. Nature Genetics. 2013; 45: 1345–1352.

[166] Austin CP. Opportunities and challenges in translational science. Clinical and Translational Science. 2021; 14: 1629–1647.