Artificial Intelligence, Healthcare, Clinical Genomics, and Pharmacogenomics Approaches in Precision Medicine

Habiba Abdelhalim[1], Asude Berber[1], Mudassir Lodi[1], Rihi Jain[1], Achuth Nair[1], Anirudh Pappu[1], Kush Patel[1], Vignesh Venkat[1], Cynthia Venkatesan[1], Raghu Wable[1], Matthew Dinatale[1], Allyson Fu[1], Vikram Iyer[1], Ishan Kalove[1], Marc Kleyman[1], Joseph Koutsoutis[1], David Menna[1], Mayank Paliwal[1], Nishi Patel[1], Thirth Patel[1], Zara Rafique[1], Rothela Samadi[1], Roshan Varadhan[1], Shreyas Bolla[1], Sreya Vadapalli[1] and Zeeshan Ahmed[1,2]*

[1]Rutgers Institute for Health, Health Care Policy and Aging Research, Rutgers University, New Brunswick, NJ, United States, [2]Department of Medicine, Rutgers Robert Wood Johnson Medical School, Rutgers Biomedical and Health Sciences, New Brunswick, NJ, United States

Precision medicine has greatly aided in improving health outcomes using earlier diagnosis and better prognosis for chronic diseases. It makes use of clinical data associated with the patient as well as their multi-omics/genomic data to reach a conclusion regarding how a physician should proceed with a specific treatment. Compared to the symptom-driven approach in medicine, precision medicine considers the critical fact that all patients do not react to the same treatment or medication in the same way. When considering the intersection of traditionally distinct arenas of medicine, that is, artificial intelligence, healthcare, clinical genomics, and pharmacogenomics—what ties them together is their impact on the development of precision medicine as a field and how they each contribute to patient-specific, rather than symptom-specific patient outcomes. This study discusses the impact and integration of these different fields in the scope of precision medicine and how they can be used in preventing and predicting acute or chronic diseases. Additionally, this study also discusses the advantages as well as the current challenges associated with artificial intelligence, healthcare, clinical genomics, and pharmacogenomics.

Keywords: artificial intelligence, healthcare, clinical-genomics, pharmacogenomics, precision medicine

INTRODUCTION

Precision medicine is the utilization of healthcare tools to create specialized treatments that consist of optimal actions for the patient, based on the data available (König et al., 2017; Pinho, 2017; Gameiro et al., 2018; Ginsburg and Phillips, 2018; Ahmed et al., 2020a; Ahmed, 2020; Elemento, 2020; Faulkner et al., 2020; Ahmed et al., 2021a). As clinical, genomic, and metabolic data become easier to obtain and interpret in relation to complex and chronic diseases such as cancer, disease treatment will become more effective (McAlister et al., 2017; Pinho, 2017; Ginsburg and Phillips, 2018; Goetz and Schork, 2018; Bilkey et al., 2019; Ahmed et al., 2020a; Ahmed, 2020; Faulkner et al., 2020; Ahmed et al., 2021a). In the current state of healthcare, healthcare professionals tend to divide their attention

and plan treatments based on symptoms (McAlister et al., 2017; Bilkey et al., 2019). However, symptoms like pain, vary from patient-to-patient and may even be absent in life-threatening situations (Lazaridis et al., 2014; McAlister et al., 2017; Pinho, 2017; Goetz and Schork, 2018; Bilkey et al., 2019). Since symptoms can greatly vary between patients, utilizing genomic and metabolic data in conjunction with clinical data from previous patients enables clinicians can prescribe a better, more personalized treatment plan (McAlister et al., 2017; Goetz and Schork, 2018). Thus, the development and implementation of precision medicine should improve the quality of healthcare compared to the conventional system dominated by symptom-driven medicine (Lazaridis et al., 2014; McAlister et al., 2017; Pinho, 2017; Ginsburg and Phillips, 2018; Goetz and Schork, 2018; Bilkey et al., 2019; Ahmed et al., 2020a; Faulkner et al., 2020; Ahmed et al., 2021a).

International interest in precision medicine could be seen as early as 2011, with the American Association for Cancer Research's (AARC) Project GENIE, which utilized several "big data initiatives" such as Genomics Evidence Neoplasia Information Exchange (GENIE) (Sweeney et al., 2017). The aim of this project was to address the challenges that came with sharing large amounts of genomics and clinical data, specifically regarding effective cancer therapies (Micheel et al., 2018). This level of innovation in precision medicine coincided with a decline in costs of DNA-sequencing techniques and a more widespread adoption of electronic medical records, which conveniently allowed for sharing and analysis of data (Bentley, 2006; Shendure et al., 2008; Evans, 2016; Kruse et al., 2016; Garrido-Cardenas et al., 2017; Graber et al., 2017; Howe et al., 2018). One common application of precision medicine in the United States is genetic screening, which is used to predict and diagnose critical conditions, which can reduce rates of morbidity (Smed et al., 2021). Another emerging application is the prescription of drugs based on genetic markers of efficacy. For instance, studies have shown that seizure drug carbamazepine having the HLA-B*1502 gene is highly likely to experience adverse effects. However, the integration of these genetic markers in prescribing drugs requires strong evidence of clinical validity first. However, the integration of these genetic markers in the prescription of drugs first require strong evidence of clinical validity at all stages of the drug product cycle (Sweeney et al., 2017; Ginsburg and Phillips, 2018). Despite pharmacogenomics being in the early stages of development, it shows great promise toward driving patient-specific outcomes.
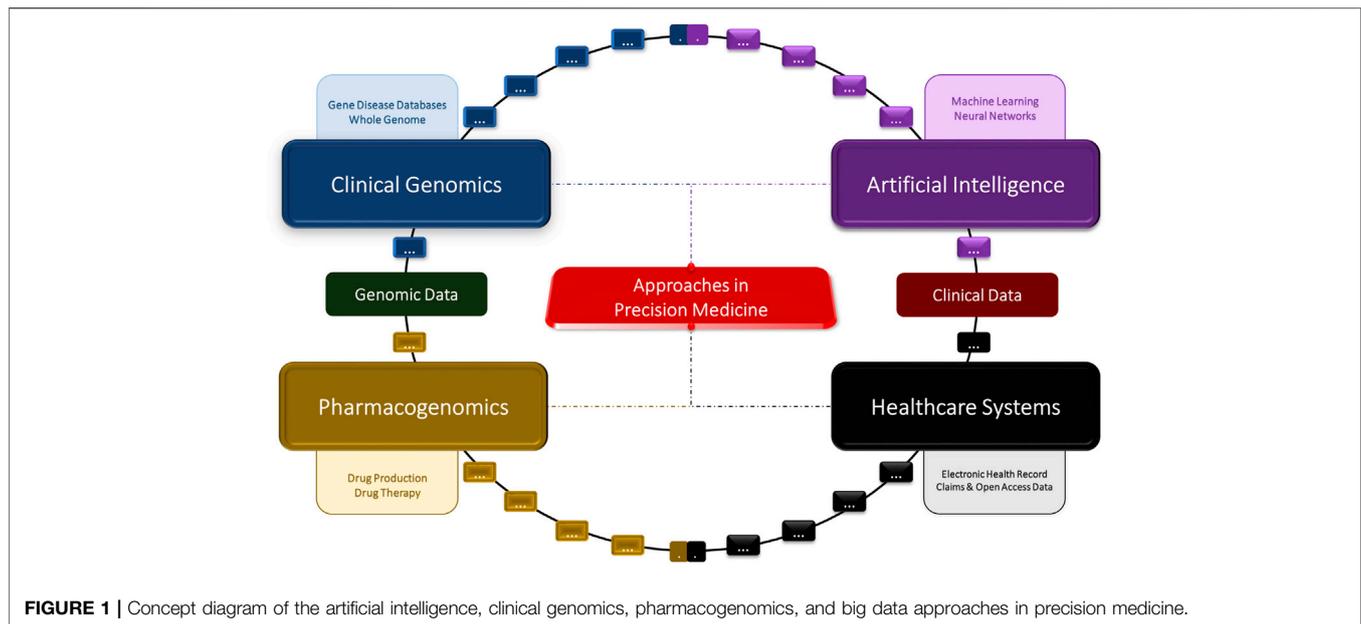
Artificial intelligence (AI) is a general term used to describe the process of using computers and technology to create stimulating software that resembles human-like critical thinking (Ramesh et al., 2004; Amisha Malik et al., 2019). AI is known to utilize many techniques (fuzzy expert system and artificial neural networks, *etc.*) that can be useful when applied to healthcare (Ramesh et al., 2004; Hessler and Baringhaus, 2018; Amisha Malik et al., 2019; Mintz and Brodie, 2019; Hashimoto et al., 2020). AI can be applied to medicine in two different ways: virtually and physically. The uses of the virtual aspect of AI can range from electronic healthcare records (Esteva et al., 2019) to neural networks guiding patient treatments (McDonnell et al.,

2021). The physical subunit of AI involves physical machines like robots assisting in surgeries (Bhandari et al., 2020), or AI-generated prosthetics for the disabled (Bernauer et al., 2021). In addition, ML, or machine learning, is a subunit of artificial intelligence, which utilizes algorithms and code in order to provide personalized experiences, where predictions are backed by mathematical data points (Deo, 2015; Currie et al., 2019). Classical machine learning is heavily dependent on human intervention; however, more unsupervised techniques have been employed in the healthcare field in more recent years (Baştanlar and Özuysal, 2014; Shen et al., 2021). This review focuses on the development of four major fields: AI, healthcare, clinical genomics, and pharmacogenomics, and their overall impacts on precision medicine (**Figure 1**).

HEALTHCARE SYSTEMS

A healthcare system is the combination of institutions, people, and resources that are involved in delivering health services and care to individuals (Kim et al., 2017; Ahmed et al., 2019a; Ho et al., 2020). The two main types of healthcare systems are commercial and academic (Kim et al., 2017; Ahmed et al., 2019a). Commercial healthcare systems (e.g., eClinicalWorks, praxis, and Allscripts) are characterized by continuous data flow and are utilized by clinical staff to implement patient treatment; therefore, no mistakes can be made with these data because this can lead to a detrimental consequence on humans (Kim et al., 2017; Ahmed et al., 2019a). One example of a commercial healthcare system is EPIC, a popular electronic medical record system, which is used by 45 percent of the United States population to store their records (Epic Systems, 2019; Shull, 2019). Their physicians then utilize the software to monitor their patients' healthcare from start to finish (Epic Systems, 2019; Shull, 2019). Academic healthcare systems are research based and can be characterized by limited clinical data management and periodic data flow with the overall goal of improving patient treatment (Rim et al., 2016; Kim et al., 2017; Ahmed et al., 2019a). An electronic healthcare record (EHR) is the collection of patient health information that is digitally stored (Knaup et al., 2007; Wronikowska et al., 2021). The compilation of patient information serves to also create an overall understanding of a given population's health status, for example, the frequency of specific diseases within the population or a specific sub-group (Knaup et al., 2007; Yeh et al., 2020). This information can then be shared across different healthcare settings, and such communication is important toward specific treatment for individuals (Koppel and Lehmann, 2014). Furthermore, an Observational Medical Outcomes Partnership (OMOP) model can be used alongside the electronic healthcare system to efficiently provide data to institutions (Gini et al., 2016; Michael et al., 2020). It was initially launched with the vision of using the system to determine the best practices by using healthcare data and helping patients (Gini et al., 2016; Al-Hanawi et al., 2021).

EHRs can assist healthcare providers in efficiently diagnosing specific rare medical cases that may not be encountered often

FIGURE 1 | Concept diagram of the artificial intelligence, clinical genomics, pharmacogenomics, and big data approaches in precision medicine.

(Knaup et al., 2007; Yeh et al., 2020; Wronikowska et al., 2021). In recent years, it has become more difficult to improve the health outcomes for a patient by a drastic margin without raising their out-of-pocket costs (Al-Hanawi et al., 2021), thus making it difficult for timely progress toward better patient care (Pastorino et al., 2019; Al-Hanawi et al., 2021). However, these obstacles can be overcome by using Big Data, which relies on its ability to recognize patterns and convert extremely high volumes of data into usable knowledge in the field of precision medicine (Ahmed et al., 2019a; Cammarota et al., 2020; Seyed Tabib et al., 2020). This data can be analyzed in a high-performance computing (HPC) environment where highly intensive computing experiments can be performed (Castrignanò et al., 2020). EHR plays a significant role, as it allows for the communication of patient data across different platforms (Knaup et al., 2007; Yeh et al., 2020), thus maximizing the chance of an effective treatment being developed (Pastorino et al., 2019; Seyed Tabib et al., 2020; Al-Hanawi et al., 2021).

Claims data refers to health insurance claims, which can contain a large variety of information about a specific patient, including details about patient diagnoses, treatments required, and finances (Stein et al., 2014; Moscatelli et al., 2018; Prosperi et al., 2018). In addition, claims data reduce selection bias by giving physicians an overall view of the patient's continued use of the healthcare system (Stein et al., 2014). These features make claims data desirable in the field of precision medicine (Prosperi et al., 2018). This is due to the availability of a larger subset of data, and the greater the number of individuals the data is pulled from, the more likely it is for a claim made based on that data to be accurate. Complex, multivariable modeling is also important because much of precision medicine is based on the fact that there are many different factors/determinants of an individual's health (Moscatelli et al., 2018). Clinical data are the compilation of patient data, scattered in numerous databases throughout the healthcare system (McGinnis et al., 2011). Such data often

address and assess genetic and metabolomic considerations, individual health, health behaviors, such as lifestyle, environmental factors, and healthcare financing, defines as the patient's management of funds toward the medical area. Clinical data also address the responses and medical procedures taken toward the health concern (McGinnis et al., 2011; Bram et al., 2015).

GENOMICS

Genomics is a subfield of molecular biology concerned with mapping structure and function genomes. The differences between individual genomes are due to the unique biological DNA they are composed of (Woese, 1969; International Human Genome Sequencing Consortium, 2001; Venter et al., 2001; Schneider et al., 2017; Roth, 2019; Zeeshan et al., 2020). Humans have between 20,000 and 25,000 genes, with each gene consisting of between a few hundred to 2 million DNA bases (Woese, 1969; International Human Genome Sequencing Consortium, 2001; Schneider et al., 2017). With the completion of the mapping of the human genome by the Human Genome Project in 2003 (Collins et al., 1998), many new and exciting applications of genomics in the medical field have been made possible. One of these applications can be seen in pharmacogenomics, which allows specialists to assign medication and corresponding dosage based on the patient's genetic markers (Wake et al., 2019; Cecchin and Stocco, 2020). Another application can be seen in clustered regularly interspaced short palindromic repeats (CRISPR), which allows for efficient gene modification in a variety of organisms. CRISPR-Cas9 has aided in understanding of the disease process through establishing genetically variable disease models (Ma et al., 2014; Cho et al., 2018; Manghwar et al., 2019). Advancements in using CRIPSR-Cas9 has made it possible to potentially treat

chronic diseases such as cancers (Chen et al., 2019), leukemia (Tzelepis et al., 2016), HIV (Xiao et al., 2019), β-thalassemia (Frangoul et al., 2021), and sickle cell anemia (Demirci et al., 2019; Frangoul et al., 2021).

Constituents of the Genome

The phenotypic display of humans is dependent upon the expression of genes, which is affected by numerous factors (Woese, 1969; International Human Genome Sequencing Consortium, 2001; Venter et al., 2001; Schneider et al., 2017; Roth, 2019; Zeeshan et al., 2020). The bridge between genotype and phenotype is proteins, and gene expressions are broken down into 2 stages: transcription (De Klerk and 't Hoen, 2015; Polychronopoulos et al., 2017; Dieci et al., 2013) and translation (Simpson et al., 2020; Lyu et al., 2021). RNA splicing, which takes place between the transcription and translation processes, results in large portions of the RNA molecule being removed, with the remaining strands being reconnected (Montes et al., 2019; Zhao, 2019; Wang and Aifantis, 2020; Xu et al., 2021). The sequences that are cut out are known as introns, non-coding intervening sequences within the primary transcript, while the other strands are known as exons, sequences within a primary transcript that remain after RNA processing (Montes et al., 2019; Xu et al., 2021). Due to the presence of introns, a single gene can encode for more than one kind of polypeptide depending on which sequences are treated as exons (Montes et al., 2019; Ule and Blencowe, 2019; Zhao, 2019; Wang and Aifantis, 2020; Xu et al., 2021; Xu et al., 2021). This process is commonly known as alternative RNA splicing and leads to increased diversity in protein coding (Ule and Blencowe, 2019; Zhao, 2019; Xu et al., 2021). However, it is not extensively regulated, and mis-splicing or mutations can lead to diseases such as muscular dystrophy (Takeshima et al., 2010; Fletcher et al., 2013; Scotti and Swanson, 2016) and premature aging dystrophy (Niblock and Gallo, 2012; Scotti and Swanson, 2016). Protein-coding DNA accounts for only 1.5% of the human genome. However, 75% of the genome is transcribed at some point, indicating that a significant amount of the genome is transcribed into non-protein-coding RNAs (ncRNAs) (Mattick and Makunin, 2006; Mishra et al., 2016; Slack and Chinnaiyan, 2019). Non-coding DNA including introns can play a role in the regulation of gene expression such as transcription initiation and termination (Hombach and Kretz, 2016; Montes et al., 2019; Wang and Aifantis, 2020; Xu et al., 2021; Lyu et al., 2021). In addition, some non-coding DNA is transcribed into function non-coding RNA molecules like tRNA (Phizicky and Hopper, 2010), rRNA (Yan et al., 2019), and regulatory RNAs which help with the processes of transcription and translation (Panni et al., 2020). According to RefSeq (O'Leary et al., 2016), a database run by the NCBI, there are currently 20,203 protein-coding genes and 17,871 non-coding genes (Polychronopoulos et al., 2017).

Next-Generation Sequencing

Next-generation sequencing (NGS) refers to the genome sequencing technologies that began to rapidly develop in the early 2000s (Sanger et al., 1977). Prior to this, Sanger sequencing, a type of sequencing created in 1975, was the primary technique used to sequence DNA (Sanger et al., 1977). With the later development of polymerase chain reaction (Green and Sambrook, 2018) in addition to automated DNA sequencing using fluorescent tags (Metzker et al., 1994; Welch and Burgess, 1999), DNA sequencing became powerful enough to create the first draft of the human genome in 2001 (International Human Genome Sequencing Consortium, 2001; Venter et al., 2001). Currently, Illumina sequencing is the most popular sequencing technology due to its accuracy, cost, and speed (Liu et al., 2012). Illumina sequencing belongs to a family of NGS technology that produces short reads (50–300 base pairs), with the most notable other technology in this category being Ion Torrent sequencing (Rhoads and Au, 2015; De Coster and Van Broeckhoven, 2019). Long-read sequencing technologies created by Oxford Nanopore (Green and Sambrook, 2018) and Pacific Biosciences (Rhoads and Au, 2015) generate reads that are thousands of base pairs long, but their reads are generally lower quality than those created by short-read sequencing (Rhoads and Au, 2015; Jain et al., 2016; Lahens et al., 2017). However, due to great interest in structural variants (genomic alterations that can be thousands of base pairs long) and their effects on diseases, improvements need to be made to long-read sequencing to avoid the costs that short-read sequencing reads suffer from such as low sensitivity (30–70%) when detecting structural variants (Liu et al., 2012; De Coster and Van Broeckhoven, 2019; Levy and Boone, 2019). After sequencing data is collected, the FASTQ file format is the traditional format used to share this information. FASTQ is a type of plain text files formatted such that each sequence has four corresponding lines of text. These lines contain information such as the sequence identifier, nucleotide sequence, a "+" sign to indicate the end of the sequence, and a line of quality values corresponding to the sequence of bases recorded in the second line (Cock et al., 2010). For storing the reference genome, FASTA, another text-based file format, is commonly used. Using the reads from the sequencing and the reference genome, algorithms map the reads to the reference genome, and these results are stored in either a Sequence Alignment Map (SAM) or its binary equivalent (BAM) file (Li et al., 2009; Hoogstrate et al., 2021). A SAM file is readable by humans, but since file sizes are so large, BAM files are used to compress the data (Yohe and Thyagarajan, 2017). Finally, since genetic variation is of interest in research, variant call format (VCF) files are commonly generated based on this data, which are files that describe the sequence variations, insertions, and deletions found in the samples along with rich annotations (Zhang, 2016; Morash et al., 2018).

NGS data can supplement other genomic sequencing methods and further develop the capabilities of DNA sequencing, which can significantly improve the effectiveness of precision medicine (Yohe and Thyagarajan, 2017; Morash et al., 2018). In conjunction with whole-genome sequencing (WGS), which has become more affordable, NGS can improve disease risk detection with improved methods for analyzing sequenced data. This assists in the development of precision medicine (Yohe and Thyagarajan, 2017; Zeeshan et al., 2020; Caspar et al., 2021). Reporting on NGS utilization, several studies have shown its efficiency in identifying actionable genetic mutations in cancer

patients, developing drugs to target tumors, and utilizing sequencing results to match patients to therapy methods (Yohe and Thyagarajan, 2017; Morash et al., 2018; Nakagawa and Fujita, 2018; Morganti et al., 2019; Zhao et al., 2019; Caspar et al., 2021). While the current effectiveness of NGS data in precision medicine remains controversial, primarily due to the experimental design of studies determining the efficacy of NGS data, and there is still immense potential for development (Yohe and Thyagarajan, 2017; Nakagawa and Fujita, 2018; Morganti et al., 2019).

Variant Discovery

WGS and WES are two different types of NGS, which are more efficient and accurate methods of DNA sequencing than traditional Sanger sequencing (Ahmed et al., 2019b). These two techniques can be used to find different variants in an individual's DNA that may be of clinical significance (may relate to the appearance of a specific disease). WGS involves the sequencing of an individual's entire genome, including both exons (protein-coding regions) and introns (non-protein coding regions) (Meienberg et al., 2016; Petersen et al., 2017). WES, however, only involves the sequencing of exons, or protein-coding regions (Petersen et al., 2017; Ahmed et al., 2019b). Variants in the human genome are changes in the DNA/nucleotide sequence that may or may not result in changes to the protein-encoding transcript and protein-building process as a whole (Petersen et al., 2017). There are several different types of variants that can be detected by the WGS and WES processes, such as single-nucleotide polymorphisms (SNP) and short or long insertions or deletions polymorphism (indels) (Petersen et al., 2017; Ahmed et al., 2021b). Variant calling or the identification of different types of variants is important because it can help researchers develop a greater understanding of the genetic components of various diseases, which can aid in future diagnoses and disease prevention (Ahmed et al., 2021b). There are three main types of WGS and WES pipelines: cloud-computing, centralized, and standalone (Ahmed et al., 2021b). The main differences between these three pipelines are the environments in which they are applied. Cloud-computing pipelines are used in environments with, "on-demand compute resources managed and provided by external vendors" (Petersen et al., 2017; Ahmed et al., 2021b). Centralized pipelines are used in local computers (Petersen et al., 2017). Standalone pipelines are mainly used in high-performance computing environments (Petersen et al., 2017). These pipelines are designed to effectively collect and process the data from WGS or WES in a way so that it can be used by researchers or medical professionals to best recognize the links between genetic variants and diseases (Petersen et al., 2017; Ahmed et al., 2021b; Ahmed et al., 2021c).

CLINICAL GENOMICS

The gene–disease relationship involves the detection of diseases associated with numerous genes with the help of sequencing techniques (Ahmed et al., 2020b). Applications can be used to help organize and assimilate information from the genomic data with the phenotypic data. By doing so, the gene–disease relationship can improve precision while detecting abnormalities in patients. They can also predict patient susceptibility to a particular disease and open the possibility of treatment options of rare diseases (Strande et al., 2017; Hu et al., 2021; Megías-Vericat et al., 2021). The study of this association can also help elucidate gene function (Van Dam et al., 2018), estimate the prevalence of genes in populations (Zhou and Skolnick, 2016), differentiate among subtypes of diseases (Nakatsuka et al., 2017) and trace how genes may predispose to (Sørlie et al., 2003) or protect against illnesses (Pirmohamed, 2006), and improve medical intervention (Ahmed et al., 2020b; Wickenhagen et al., 2021).

Gene–disease databases are essential due to their unique display of information regarding the "exchange and reporting of actionable genetic variants and associated phenotype" (Ahmed et al., 2020b). This information is vital because it can be used to aid in treatment for complex diseases like cancer, where detection of just a few vital variants can help warn about the development of a tumor (Orkin and Bauer, 2019). An example of gene–disease databases is a database of disease–gene associations with annotated relationships among genes (eDGAR), which collects and arranges data related to gene/disease associations along with interactions in heterogenous and polygenic diseases (Huang et al., 2018). With a focus on "the structural and functional annotations of the genes" (Huang et al., 2018), the database provides the cytogenetic location of a gene. When multiple sets of genes are prevalent in the same disease, the data are organized in such a way that it allows for customized data search (Huang et al., 2018). This is important because it allows for multiplatform usage and comparability so users can get the information that they need as accurately as possible in order to tailor procedures specific to their patients or research (Huang et al., 2018).

As the technology for genomic analysis advances, more reports of gene–disease relationships are beginning to expand, facilitating the need for accessibility to information storage (Babbi et al., 2017; Ahmed et al., 2020b). The ability to analyze a database containing information about genetic diseases and cross referencing that data with patient records has the potential to treat a genetic disease before it proliferates (Babbi et al., 2017; Huang et al., 2018). There are approximately 18,000 gene–disease databases that collect data (Huang et al., 2018). Out of these 18,000, there are approximately 50 authentic databases. However, there are no existing databases that cover the entire human genome. Examples of these gene–disease databases are Ensembl (Landrum et al., 2016), GenCode (Aken et al., 2016), ClinVar (Frankish et al., 2019), GeneCards (Landrum et al., 2016), HGMD (Stelzer et al., 2016), OMIM (Stenson et al., 2012), Orphanet (Amberger et al., 2015), SwissProt (INSERM, 1997), and LncRNADisease (Consortium, 2015). **Table 1** further details the extensive availability of various databases that contain raw and multi-omics data. Very few of these gene–disease databases are approved by the American College of Medical Genetic and Genomics (ACMG), a medical organization that focuses on improving health through medical genetics and genomics.

TABLE 1 | Multi-omics/genomics databases: Ensembl, GENCODE, ClinVar, GeneCards, HGMD, OMIM, Orphanet, SwissProt, TCGA, GenBank, EMBL, InterPro, Reactome, 1000 Genomes, European Nucleotide Archive, Sequence Read Archive, United Kingdom Biobank, and TOPMed.

Database	Description
Ensembl	Ensembl genome browser, provides genomics sequence variation, annotates genes, and computes multiple alignments
GENCODE	Generalized Coding, identification of protein-coding genes
ClinVar	NCBI ClinVar, integration of genomic variation and human health
GeneCards	Human genes integration database
Human Gene Mutation Database (HGMD)	Human Gene Mutation Database, aligning gene lesions for human-inherited diseases
Online Mendelian Inheritance in Man (OMIM)	Online Mendelian Inheritance in man, a database on human genes and genetic phenotypes and identification of gene–disease association
Orphanet	Orphanet Rare Disease Ontology (ORDO), analyzing rare diseases and gene–disease relationship
SwissProt	Protein sequence database, information about high level of annotations
The Cancer Genome Atlas (TCGA)	The Cancer Genome Atlas, analyzation of primary cancer samples that genomically sequenced and molecularly characterized
GenBank	National Center for Biotechnology Information for the primary database
The EMBL Nucleotide Sequence Database	Molecular Biology Primary Database, in-depth data collection of primary nucleotide sequences
InterPro	Protein-based secondary database, provides protein functionality analysis and information about classification of them into families, domains, and sites
Reactome	A specialized database for metabolic and biological pathways
1000 Genomes	A comprehensive catalog of human genetic variation
European Nucleotide Archive	European Nucleotide Archive, provides nucleotide sequence information and raw sequencing data
Sequence Read Archive	Sequence Read Archive, stores raw sequencing data and alignment information
United Kingdom Biobank	Biomedical database, provides profoundly genetic and health information
Trans-Omics for Precision Medicine (TOPMed)	NHLBI Trans-Omics for Precision Medicine, provides whole-genome sequencing (WGS) and other omics data

Despite the plethora of gene–disease databases available to researchers and the healthcare industry, there are quite a few shortcomings within the databases (Huang et al., 2018). The primary issue stems from the fact that there is no standardization within these databases, as no singular database contains all the data on the currently available genome, diseases, and drugs in the market (INSERM, 1997; Stenson et al., 2012; Amberger et al., 2015; Consortium, 2015; Aken et al., 2016; Cardon and Harris, 2016; Landrum et al., 2016; Stelzer et al., 2016; Babbi et al., 2017; Huang et al., 2018; Frankish et al., 2019). In addition, these databases are often out-of-date or do not provide relevant essential information regarding diseases, reducing the practicality and usability of these databases (Chen et al., 2012; Ahmed et al., 2020b). To resolve these issues, the IOS application PAS-Gen or PROMIS-APP-SUITE has been developed to provide an accessible central database for genomic and disease information, potentially accelerating medical discoveries in the future (Stenson et al., 2017). The purpose of the app is to provide researchers and employees in the healthcare industry with information about genes that could result in the development of certain diseases for educational and non-commercial uses. In addition, the user-friendly interface of the application enables accessibility to a wide variety of users and for the app to continuously receive updates if needed (Stenson et al., 2017). This application includes a total of 59,293 genes (19,989 protein-coding and 39,304 non-protein-coding), and "is composed of 98,064 gene–disease combinations reported from 809 distinct sources" (Stenson et al., 2017). These features make the PAS App comprehensive, easily accessible, and suited for the future of personalized medicine. This centralized database and application could prove to be tremendously useful in the future, boosting the abilities of healthcare researchers to understand the human genome and its implications in the development of diseases (Stenson et al., 2017).

PHARMACOGENOMICS

Pharmacogenomics is the analysis of how an individual's genome (their unique genetic makeup) influences their reaction to certain medications prescribed to them (Wake et al., 2019; Karol and Yang, 2020). Protein-coding genes could influence the treatment of the drug either by breaking it down, absorbing it, or transporting the drug to desired (or undesired) locations (Hoehe and Morris-Rosendahl, 2018; Venkatachalapathy et al., 2021). Grouping people that have similar genetic variations will allow for better observation of how these groups may have a common treatment response to a given treatment (Hoehe and Morris-Rosendahl, 2018; Wake et al., 2019b; Karol and Yang, 2020; Venkatachalapathy et al., 2021). One example of pharmacogenomics is thiopurine methyltransferase testing to determine candidates for thiopurine drug therapy, which are used to treat autoimmune disorders like Crohn's disease or rheumatoid arthritis (Wake et al., 2019).

Pharmacogenomics has a multitude of applications for both precision medicine and related fields, allowing for the improvement of drugs in all stages of the process, from production to consumption (Wake et al., 2019; Kim et al., 2021). One such application comes in the field of drug research and production, where information from pharmacogenomics practices can help assist in how best to research potential medicine for patients for certain diseases (Penny and McHale, 2005; Payami and Factor, 2014; Relling and Evans, 2015). There have been genome-wide association

studies conducted to measure responses to certain drugs, which have allowed for the identification of groups of people in which a drug may be more effective than it would be in other groups (Kim et al., 2021). Pharmacogenomics also has applications in a later step of the drug manufacturing and distribution process, primarily in the production and labeling of drugs. Pharmacogenomics information (PGx) has been increasingly included in the labels of new drug approvals, with most of these being clinically actionable, allowing for better testing and utilization in clinical practices (Cheng et al., 2021; Kim et al., 2021). A third application comes in the prescription of these drugs. For instance, the dosage and selection of drugs can be optimized with data regarding the effects of nucleotide polymorphisms on drug efficacy (Chung et al., 2020; Megías-Vericat et al., 2021).

DISCUSSION

Clinical genomics, AI, big data, and pharmacogenomics are essential in the future development of precision medicine (**Figure 1**). Precision medicine has many advantages, such as providing the means for physicians and other healthcare professionals to make more accurate diagnoses, giving healthcare officials and researchers easier access to larger amounts of medical data, and leading to a greater understanding of diseases and their underlying causes (McAlister et al., 2017; Pinho, 2017; Ginsburg and Phillips, 2018; Goetz and Schork, 2018; Bilkey et al., 2019; Ahmed et al., 2020a; Ahmed, 2020; Faulkner et al., 2020; Ahmed et al., 2021a). At present, the most prominent challenge in precision medicine is identifying which approaches to implement when working with different types of medical data (Ahmed, 2020). In addition, there is no current system in place that allows for the comparison of multi-omics patient data to provide accurate predictive and personalized results. A more user-friendly interface would be required for effective implementation of precision medicine in the healthcare field. While there have been many developments in this field in recent years, there are also many ethical and logistical challenges that are need to be addressed in clinical genomics, big data, and pharmacogenomics implementation. Furthermore, the prematurity of some these approaches, such as the pharmacogenomics field, in the scope of precision medicine present a challenge in providing reliable and accurate results.

With the rapid development of many new approaches in precision medicine, the integration of two or more fields is currently being explored. Each type of approach offers a unique understanding and perspective of the data. The integration of multiple approaches allows for a more holistic and comprehensive approach to precision medicine, incorporating information from several different fields. In precision medicine, integration of these approaches provides researchers a comprehensive understanding of a given medical case, from the cause of the disease to the relevant mechanisms and interactions associated with the disease (Hasin et al., 2017). This, in turn, allows for a more accurate selection of treatment method

for a given disease. One example of the integration of these approaches in precision medicine is the application of AI to genomics. Extensive developments have been made in the intersection of the field of these two fields. AI has aided in the process of combining different algorithms to make disease analysis and predictions more efficient (Xu et al., 2019). Genomic analysis using AI is mainly focused on gene expression, DNA methylation, somatic point mutation, and copy number alteration (Xu et al., 2019). By employing machine and deep learning approaches among many others, predictive models are built that can utilize genomic and other multi-omics data to speedup the process of data analysis thus, resulting in faster decision-making (Xu et al., 2019). Different integration methods of multi-omics datasets are essential when it comes to capturing the complexity of each omics approach (Picard et al., 2021). While recent advancements have been observed in the genomics and multi-omics field, more benchmark studies are needed to choose the ideal machine-learning strategy to be implemented (Picard et al., 2021; Reel et al., 2021). Multi-omics integrative models can aid in understanding the intricacies of disease abnormalities, which is not always possible with only genomic or other single-omics analysis (Reel et al., 2021).

Although big data have some advantages in medical research, such as reduction of medical error and enablement of the correct treatment approach to a disease, the need to standardize the data content and clinical definition presents a significant disadvantage (Xu et al., 2019). Despite their own advantages in precision medicine, there are also some limitations in open access clinical data and claims data in precision medicine. One significant limitation is that the health care-specific information included in the claims can be incomplete, inaccurate, or altogether missing. This originates difficulties in determining how patient-specific treatment is appropriate or effective (Stein et al., 2014). Clinical genomics is also an emerging approach in precision medicine and involves using clinical patient data to derive relationships between genes and diseases. These gene–disease relationships are typically stored in genomic databases. These databases enable the collection and analysis of genomic data, which allows for personalized treatment of the disease before significant proliferation. Similar to open access clinical data, one major limitation of such databases is that they are not standardized, meaning data cannot be easily transferred from database-to-database. This complicates the process of cross-referencing data between different databases and can prove to be an obstacle to efficient patient treatment. Pharmacogenomics is similar to clinical genomics in that it is a branch of genomics (analysis of an individual genome). However, it focuses more on the individual's reaction to specific treatments and medications rather than the disease itself. This is essential in precision medicine, as such study provides an insight into how patients respond to specific treatments. Such correlations can be derived between a patient's genomic makeup and their reaction to treatments. This then allows for the grouping of patients based on similar genetic variations and allow for more precise and personalized prescription of treatment. However, pharmacogenomics is still a new field of study and has not been developed enough to be reliably utilized.

The advancements in AI, healthcare, clinical genomics, and pharmacogenomics have resulted in a substantial volume of data generated. However, this large influx of data presents an issue, as no reliable or standardized means of analysis has been developed. Such data are too large to be analyzed through common visual analysis or statistical correlation methods (Álvarez-Machancoses et al., 2020; Caudai et al., 2021). The use of AI and ML techniques alleviates this issue by allowing for the efficient management of data and providing the ability to recognize patterns in complex datasets (Caudai et al., 2021). In addition, the AI and ML techniques do not require explicit programming to complete specific tasks, as they are able to independently detect and analyze patterns within the data (Caudai et al., 2021). The implementation of such techniques and methods also provides the ability to predict an attribute based on correlated data. For example, this is especially beneficial in clinical settings, in which such techniques can be implemented to predict the pharmaceutical properties of drug targets and drug candidates (König et al., 2021). Overall, the AI and ML methods are one promising solution to the rapidly increasing volume of high-throughput data (Vadapalli et al., 2022).

In this study, we hypothesize that the continued integration of pharmacogenomics, clinical genomics, AI, and healthcare will allow for increased potential in patient diagnosis and treatment. A standardized model linking and translating all these different variables needs to be implemented at the clinical level in order to further progress in the precision medicine field. The application of clinical genomics, pharmacogenomics, artificial intelligence algorithms, and big data analytics in precision medicine across different types of patient data (multi-omics) will allow for better prognosis and prediction of disease states and their mechanisms.

REFERENCES

Ahmed, Z., Zeeshan, S., Xiong, R., and Liang, B. T. (2019). Debutant iOS App and Gene-Disease Complexities in Clinical Genomics and Precision Medicine. *Clin. Transl. Med.* 8 (1), 26–11. doi:10.1186/s40169-019-0243-8

Ahmed, Z., Kim, M., and Liang, B. T. (2019). MAV-clic: Management, Analysis, and Visualization of Clinical Data. *JAMIA open* 2 (1), 23–28. doi:10.1093/jamiaopen/ooy052

Ahmed, Z., Mohamed, K., Zeeshan, S., and Dong, X. (2020). Artificial Intelligence with Multi-Functional Machine Learning Platform Development for Better Healthcare and Precision Medicine. *Database J. Biol. databases curation* 2020, baaa010. doi:10.1093/database/baaa010

Ahmed, Z. (2020). Practicing Precision Medicine with Intelligently Integrative Clinical and Multi-Omics Data Analysis. *Hum. Genomics* 14 (1), 35. doi:10.1186/s40246-020-00287-z

Ahmed, Z., Renart, E. G., Mishra, D., and Zeeshan, S. (2021). JWES: a New Pipeline for Whole Genome/exome Sequence Data Processing, Management, and Gene-variant Discovery, Annotation, Prediction, and Genotyping. *FEBS Open bio* 11 (9), 2441–2452. doi:10.1002/2211-5463.13261

Ahmed, Z., Renart, E. G., and Zeeshan, S. (2021). Genomics Pipelines to Investigate Susceptibility in Whole Genome and Exome Sequenced Data for Variant Discovery, Annotation, Prediction and Genotyping. *PeerJ* 9, e11724. doi:10.7717/peerj.11724

Ahmed, Z., Zeeshan, S., Foran, D. J., Kleinman, L. C., Wondisford, F. E., and Dong, X. (2021). Integrative Clinical, Genomics and Metabolomics Data Analysis for Mainstream Precision Medicine to Investigate COVID-19. *BMJ Innov.* 7 (1), 6–10. doi:10.1136/bmjinnov-2020-000444

Ahmed, Z., Zeeshan, S., Mendhe, D., and Dong, X. (2020). Human Gene and Disease Associations for Clinical-genomics and Precision Medicine Research. *Clin. Transl. Med.* 10 (1), 297–318. doi:10.1002/ctm2.28

Aken, B. L., Ayling, S., Barrell, D., Clarke, L., Curwen, V., Fairley, S., Fernandez Banet, J., Billis, K., García Girón, C., Hourlier, T., Howe, K., Kähäri, A., Kokocinski, F., Martin, F. J., Murphy, D. N., Nag, R., Ruffier, M., Schuster, M., Tang, Y. A., Vogel, J.-H., White, S., Zadissa, A., Flicek, P., and Searle, S. M. J. (2016). The Ensembl Gene Annotation System. *Database* 2016, baw093. doi:10.1093/database/baw093

Al-Hanawi, M. K., Mwale, M. L., and Qattan, A. M. N. (2021). Health Insurance and Out-Of-Pocket Expenditure on Health and Medicine: Heterogeneities along Income. *Front. Pharmacol.* 12. doi:10.3389/fphar.2021.638035

Álvarez-Machancoses, Ó., DeAndrés Galiana, E. J., Cernea, A., Fernández Sánchez de la Viña, J., and Fernández-Martínez, J. L. (2020). On the Role of Artificial Intelligence in Genomics to Enhance Precision Medicine. *Pgpm* Vol. 13, 105–119. doi:10.2147/PGPM.S205082

Amberger, J. S., Bocchini, C. A., Schiettecatte, F., Scott, A. F., and Hamosh, A. (2015). OMIM.org: Online Mendelian Inheritance in Man (OMIM), an Online Catalog of Human Genes and Genetic Disorders. *Nucleic acids Res.* 43 (D1), D789–D798. doi:10.1093/nar/gku1205

Amisha Malik, P., Pathania, M., and Rathaur, V. (2019). Overview of Artificial Intelligence in Medicine. *J. Fam. Med. Prim. Care* 8 (7), 2328–2331. doi:10.4103/jfmpc.jfmpc_440_19

Babbi, G., Martelli, P. L., Profiti, G., Bovo, S., Savojardo, C., and Casadio, R. (2017). Edgar: A Database of Disease-Gene Associations with Annotated Relationships Among Genes. *BMC Genomics* 18 (S5), 554. doi:10.1186/s12864-017-3911-3

Baştanlar, Y., and Özuysal, M. (2014). Introduction to Machine Learning. *Methods Mol. Biol. Clift. N.J.)* 1107, 105–128. doi:10.1007/978-1-62703-748-8_7

Bentley, D. R. (2006). Whole-genome Re-sequencing. *Curr. Opin. Genet. Dev.* 16 (6), 545–552. doi:10.1016/j.gde.2006.10.009

Bernauer, S. A., Zitzmann, N. U., and Joda, T. (2021). The Use and Performance of Artificial Intelligence in Prosthodontics: A Systematic Review. *Sensors* 21 (19), 6628. doi:10.3390/s21196628

Bhandari, M., Zeffiro, T., and Reddiboina, M. (2020). Artificial Intelligence and Robotic Surgery. *Curr. Opin. urology* 30 (1), 48–54. doi:10.1097/MOU.0000000000000692

Bilkey, G. A., Burns, B. L., Coles, E. P., Mahede, T., Baynam, G., and Nowak, K. J. (2019). Optimizing Precision Medicine for Public Health. *Front. Public Health* 7, 42. doi:10.3389/fpubh.2019.00042

Bram, J. T., Warwick-Clark, B., Obeysekare, E., and Mehta, K. (2015). Utilization and Monetization of Healthcare Data in Developing Countries. *Big Data* 3 (2), 59–66. doi:10.1089/big.2014.0053

Cammarota, G., Ianiro, G., Ahern, A., Carbone, C., Temko, A., Claesson, M. J., et al. (2020). Gut Microbiome, Big Data and Machine Learning to Promote Precision Medicine for Cancer. *Nat. Rev. Gastroenterol. Hepatol.* 17 (10), 635–648. doi:10.1038/s41575-020-0327-3

Cardon, L. R., and Harris, T. (2016). Precision Medicine, Genomics and Drug Discovery: Table 1. *Hum. Mol. Genet.* 25 (R2), R166–R172. doi:10.1093/hmg/ddw246

Caspar, S. M., Schneider, T., Stoll, P., Meienberg, J., and Matyas, G. (2021). Potential of Whole-Genome Sequencing-Based Pharmacogenetic Profiling. *Pharmacogenomics* 22 (03), 177–190. doi:10.2217/pgs-2020-0155

Castrignanò, T., Gioiosa, S., Flati, T., Cestari, M., Picardi, E., Chiara, M., et al. (2020). ELIXIR-IT HPC@CINECA: High Performance Computing Resources for the Bioinformatics Community. *BMC Bioinforma.* 21 (Suppl. 10), 352. doi:10.1186/s12859-020-03565-8

Caudai, C., Galizia, A., Geraci, F., Le Pera, L., Morea, V., Salerno, E., et al. (2021). AI Applications in Functional Genomics. *Comput. Struct. Biotechnol. J.* 19, 5762–5790. doi:10.1016/j.csbj.2021.10.009

Cecchin, E., and Stocco, G. (2020). Pharmacogenomics and Personalized Medicine. *Genes.* 11 (6), 679. doi:10.3390/genes11060679

Chen, G., Wang, Z., Wang, D., Qiu, C., Liu, M., Chen, X., et al. (2012). LncRNADisease: a Database for Long-Non-Coding RNA-Associated Diseases. *Nucleic acids Res.* 41 (D1), D983–D986. doi:10.1093/nar/gks1099

Chen, M., Mao, A., Xu, M., Weng, Q., Mao, J., and Ji, J. (2019). CRISPR-Cas9 for Cancer Therapy: Opportunities and Challenges. *Cancer Lett.* 447, 48–55. doi:10.1016/j.canlet.2019.01.017

Cheng, C. M., So, T. W., and Bubp, J. L. (2021). Characterization of Pharmacogenetic Information in Food and Drug Administration Drug Labeling and the Table of Pharmacogenetic Associations. *Ann. Pharmacother.* 55 (10), 1185–1194. doi:10.1177/1060028020983049

Cho, S., Shin, J., and Cho, B.-K. (2018). Applications of CRISPR/Cas System to Bacterial Metabolic Engineering. *Ijms* 19 (4), 1089. doi:10.3390/ijms19041089

Chung, J. E., Yee, J., Hwang, H. S., Park, J. Y., Lee, K. E., Kim, Y. J., et al. (2020). Influence of GRK5 Gene Polymorphisms on Ritodrine Efficacy and Adverse Drug Events in Preterm Labor Treatment. *Sci. Rep.* 10 (1), 1351. doi:10.1038/s41598-020-58348-1

Cock, P. J. A., Fields, C. J., Goto, N., Heuer, M. L., and Rice, P. M. (2010). The Sanger FASTQ File Format for Sequences with Quality Scores, and the Solexa/Illumina FASTQ Variants. *Nucleic acids Res.* 38 (6), 1767–1771. doi:10.1093/nar/gkp1137

Collins, F. S., Patrinos, A., Jordan, E., Chakravarti, A., Gesteland, R., and Walters, L. (1998). New Goals for the U.S. Human Genome Project: 1998-2003. *Science* 282 (5389), 682–689. doi:10.1126/science.282.5389.682

Consortium, U. (2015). UniProt: a Hub for Protein Information. *Nucleic Acids Res.* 43 (D1), D204–D212. doi:10.1093/nar/gku989

Currie, G., Hawk, K. E., Rohren, E., Vial, A., and Klein, R. (2019). Machine Learning and Deep Learning in Medical Imaging: Intelligent Imaging. *J. Med. imaging Radiat. Sci.* 50 (4), 477–487. doi:10.1016/j.jmir.2019.09.005

De Coster, W., and Van Broeckhoven, C. (2019). Newest Methods for Detecting Structural Variations. *Trends Biotechnol.* 37 (9), 973–982. doi:10.1016/j.tibtech.2019.02.003

De Klerk, E., and 't Hoen, P. A. C. (2015). Alternative mRNA Transcription, Processing, and Translation: Insights from RNA Sequencing. *Trends Genet.* 31 (3), 128–139. doi:10.1016/j.tig.2015.01.001

Demirci, S., Leonard, A., Haro-Mora, J. J., Uchida, N., and Tisdale, J. F. (2019). CRISPR/Cas9 for Sickle Cell Disease: Applications, Future Possibilities, and Challenges. *Cell. Biol. Transl. Med.* 5, 37–52. doi:10.1007/5584_2018_331

Deo, R. C. (2015). Machine Learning in Medicine. *Circulation* 132 (20), 1920–1930. doi:10.1161/CIRCULATIONAHA.115.001593

Dieci, G., Bosio, M. C., Fermi, B., and Ferrari, R. (2013). Transcription Reinitiation by RNA Polymerase III. *Biochimica Biophysica Acta (BBA) - Gene Regul. Mech.* 1829 (3-4), 331–341. doi:10.1016/j.bbagrm.2012.10.009

Elemento, O. (2020). The Future of Precision Medicine: towards a More Predictive Personalized Medicine. *Emerg. Top. life Sci.* 4 (2), 175–177. doi:10.1042/ETLS20190197

Epic Systems (2019). Epic. Available at https://www.epic.com/about.

Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., et al. (2019). A Guide to Deep Learning in Healthcare. *Nat. Med.* 25 (1), 24–29. doi:10.1038/s41591-018-0316-z

Evans, R. S. (2016). Electronic Health Records: Then, Now, and in the Future. *Yearb. Med. Inf.* 25 (S 01), S48–S61. doi:10.15265/iys-2016-s006

Faulkner, E., Holtorf, A.-P., Walton, S., Liu, C. Y., Lin, H., Biltaj, E., et al. (2020). Being Precise about Precision Medicine: What Should Value Frameworks Incorporate to Address Precision Medicine? A Report of the Personalized Precision Medicine Special Interest Group. *Value Health* 23 (5), 529–539. doi:10.1016/j.jval.2019.11.010

Fletcher, S., Meloni, P. L., Johnsen, R. D., Wong, B. L., Muntoni, F., and Wilton, S. D. (2013). Antisense Suppression of Donor Splice Site Mutations in the Dystrophin Gene Transcript. *Mol. Genet. Genomic Med.* 1 (3), 162–173. doi:10.1002/mgg3.19

Frangoul, H., Altshuler, D., Cappellini, M. D., Chen, Y.-S., Domm, J., Eustace, B. K., Foell, J., de la Fuente, J., Grupp, S., Handgretinger, R., Ho, T. W., Kattamis, A., Kernytsky, A., Lekstrom-Himes, J., Li, A. M., Locatelli, F., Mapara, M. Y., de Montalembert, M., Rondelli, D., Sharma, A., Sheth, S., Soni, S., Steinberg, M. H., Wall, D., Yen, A., and Corbacioglu, S. (2021). CRISPR-Cas9 Gene Editing for Sickle Cell Disease and β-Thalassemia. *N. Engl. J. Med.* 384 (3), 252–260. doi:10.1097/01.ogx.0000754392.61396.7910.1056/nejmoa2031054

Frankish, A., Diekhans, M., Ferreira, A.-M., Johnson, R., Jungreis, I., Loveland, J., Mudge, J. M., Sisu, C., Wright, J., Armstrong, J., Barnes, I., Berry, A., Bignell, A., Carbonell Sala, S., Chrast, J., Cunningham, F., Di Domenico, T., Donaldson, S., Fiddes, I. T., García Girón, C., Gonzalez, J. M., Grego, T., Hardy, M., Hourlier, T., Hunt, T., Izuogu, O. G., Lagarde, J., Martin, F. J., Martínez, L., Mohanan, S., Muir, P., Navarro, F. C. P., Parker, A., Pei, B., Pozo, F., Ruffier, M., Schmitt, B. M., Stapleton, E., Suner, M.-M., Sycheva, I., Uszczynska-Ratajczak, B., Xu, J., Yates, A., Zerbino, D., Zhang, Y., Aken, B., Choudhary, J. S., Gerstein, M., Guigó, R., Hubbard, T. J. P., Kellis, M., Paten, B., Reymond, A., Tress, M. L., and Flicek, P. (2019). GENCODE Reference Annotation for the Human and Mouse Genomes. *Nucleic acids Res.* 47 (D1), D766–D773. doi:10.1007/978-981-16-5812-9_110.1093/nar/gky955

Gameiro, G. R., Sinkunas, V., Liguori, G. R., and Auler-Júnior, J. O. C. (2018). Precision Medicine: Changing the Way We Think about Healthcare. *Clinics* 73, e723. doi:10.6061/clinics/2017/e723

Garrido-Cardenas, J., Garcia-Maroto, F., Alvarez-Bermejo, J., and Manzano-Agugliaro, F. (2017). DNA Sequencing Sensors: an Overview. *Sensors* 17 (3), 588. doi:10.3390/s17030588

Gini, R., Schuemie, M., Brown, J., Ryan, P., Vacchi, E., Coppola, M., et al. (2016). Data Extraction and Management in Networks of Observational Health Care Databases for Scientific Research: A Comparison Among EU-ADR, OMOP, Mini-Sentinel and MATRICE Strategies. *eGEMs* 4 (1), 2. doi:10.13063/2327-9214.1189

Ginsburg, G. S., and Phillips, K. A. (2018). Precision Medicine: From Science to Value. *Health Aff.* 37 (5), 694–701. doi:10.1377/hlthaff.2017.1624

Goetz, L. H., and Schork, N. J. (2018). Personalized Medicine: Motivation, Challenges, and Progress. *Fertil. Steril.* 109 (6), 952–963. doi:10.1016/j.fertnstert.2018.05.006

Graber, M. L., Byrne, C., and Johnston, D. (2017). The Impact of Electronic Health Records on Diagnosis. *Diagnosis* 4 (4), 211–223. doi:10.1515/dx-2017-0012

Green, M. R., and Sambrook, J. (2018). The Basic Polymerase Chain Reaction (PCR). *Cold Spring Harb. Protoc.* 2018 (5), pdb.prot095117. doi:10.1101/pdb.prot095117

Hashimoto, D. A., Witkowski, E., Gao, L., Meireles, O., and Rosman, G. (2020). Artificial Intelligence in Anesthesiology. *Anesthesiology* 132 (2), 379–394. doi:10.1097/ALN.0000000000002960

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

Hessler, G., and Baringhaus, K.-H. (2018). Artificial Intelligence in Drug Design. *Molecules* 23 (10), 2520. doi:10.3390/molecules23102520

Ho, D., Quake, S. R., McCabe, E. R. B., Chng, W. J., Chow, E. K., Ding, X., et al. (2020). Enabling Technologies for Personalized and Precision Medicine. *Trends Biotechnol.* 38 (5), 497–518. doi:10.1016/j.tibtech.2019.12.021

Hoehe, M. R., and Morris-Rosendahl, D. J. (2018). The Role of Genetics and Genomics in Clinical Psychiatry. *Dialogues Clin. Neurosci.* 20 (3), 169–177. doi:10.31887/DCNS.2018.20.3/mhoehe

Hombach, S., and Kretz, M. (2016). Non-coding RNAs: Classification, Biology and Functioning. *Non-coding RNAs colorectal cancer* 937, 3–17. doi:10.1007/978-3-319-42059-2_1

Hoogstrate, Y., Jenster, G., and van de Werken, H. J. G. (2021). FASTAFS: File System Virtualisation of Random Access Compressed FASTA Files. *BMC Bioinforma.* 22 (1), 1–12. doi:10.1101/2020.11.11.377689

Howe, J. L., Adams, K. T., Hettinger, A. Z., and Ratwani, R. M. (2018). Electronic Health Record Usability Issues and Potential Contribution to Patient Harm. *Jama* 319 (12), 1276–1278. doi:10.1001/jama.2018.1171

Hu, J., Lepore, R., Dobson, R. J. B., Al-Chalabi, A., Bean, D. M., and Iacoangeli, A. (2021). DGLinker: Flexible Knowledge-Graph Prediction of Disease-Gene Associations. *Nucleic acids Res.* 49 (W1), W153–W161. doi:10.1093/nar/gkab449

Huang, D., Sun, W., Zhou, Y., Li, P., Chen, F., Chen, H., et al. (2018). Mutations of Key Driver Genes in Colorectal Cancer Progression and Metastasis. *Cancer Metastasis Rev.* 37 (1), 173–187. doi:10.1007/s10555-017-9726-5

INSERM (1997). Orphanet: an Online Database of Rare Diseases and Orphan Drugs. Available at http://www.orpha.net.

International Human Genome Sequencing Consortium (2001). Initial Sequencing and Analysis of the Human Genome. *Nature* 412 (6846), 565. doi:10.1038/35057062

Jain, M., Olsen, H. E., Paten, B., and Akeson, M. (2016). The Oxford Nanopore MinION: Delivery of Nanopore Sequencing to the Genomics Community. *Genome Biol.* 17 (1), 1–11. doi:10.1186/s13059-016-1103-0

Karol, S. E., and Yang, J. J. (2020). Pharmacogenomics and ALL Treatment: How to Optimize Therapy. *Seminars Hematol.* 57 (3), 130–136. doi:10.1053/j.seminhematol.2020.10.001

Kim, J. A., Ceccarelli, R., and Lu, C. Y. (2021). Pharmacogenomic Biomarkers in US FDA-Approved Drug Labels (2000-2020). *Jpm* 11 (3), 179. doi:10.3390/jpm11030179

Kim, M. O., Coiera, E., and Magrabi, F. (2017). Problems with Health Information Technology and Their Effects on Care Delivery and Patient Outcomes: a Systematic Review. *J. Am. Med. Inf. Assoc.* 24 (2), 246–250. doi:10.1093/jamia/ocw154

Knaup, P., Leiner, F., and Haux, R. (2007). On Educating about Medical Data Management. *Methods Inf. Med.* 46 (01), 74–79. doi:10.1055/s-0038-1628137

König, H., Frank, D., Baumann, M., and Heil, R. (2021). AI Models and the Future of Genomic Research and Medicine: True Sons of Knowledge?: Artificial Intelligence Needs to Be Integrated with Causal Conceptions in Biomedicine to Harness its Societal Benefits for the Field. *BioEssays* 43, e2100025. doi:10.1002/bies.202100025

König, I. R., Fuchs, O., Hansen, G., von Mutius, E., and Kopp, M. V. (2017). What Is Precision Medicine? *Eur. Respir. J.* 50 (4), 1700391. doi:10.1183/13993003.00391-2017

Koppel, R., and Lehmann, C. U. (2014). Implications of an Emerging EHR Monoculture for Hospitals and Healthcare Systems. *J. Am. Med. Inf. Assoc.* 22 (2), 465–471. doi:10.1136/amiajnl-2014-003023

Kruse, C. S., Kristof, C., Jones, B., Mitchell, E., and Martinez, A. (2016). Barriers to Electronic Health Record Adoption: a Systematic Literature Review. *J. Med. Syst.* 40 (12), 1–7. doi:10.1007/s10916-016-0628-9

Lahens, N. F., Ricciotti, E., Smirnova, O., Toorens, E., Kim, E. J., Baruzzo, G., Hayer, K. E., Ganguly, T., Schug, J., and Grant, G. R. (2017). A Comparison of Illumina and Ion Torrent Sequencing Platforms in the Context of Differential Gene Expression. *BMC genomics* 18 (1), 1–13. doi:10.1186/s12864-017-4011-0

Landrum, M. J., Lee, J. M., Benson, M., Brown, G., Chao, C., Chitipiralla, S., Gu, B., Hart, J., Hoffman, D., Hoover, J., Jang, W., Katz, K., Ovetsky, M., Riley, G., Sethi, A., Tully, R., Villamarin-Salomon, R., Rubinstein, W., and Maglott, D. R. (2016). ClinVar: Public Archive of Interpretations of Clinically Relevant Variants. *Nucleic Acids Res.* 44 (D1), D862–D868. doi:10.3410/f.725948565.79353107610.1093/nar/gkv1222

Lazaridis, K. N., McAllister, T. M., Babovic-Vuksanovic, D., Beck, S. A., Borad, M. J., Bryce, A. H., et al. (2014). Implementing Individualized Medicine into the Medical Practice. *Am. J. Med. Genet.* 166 (1), 15–23. doi:10.1002/ajmg.c.31387

Levy, S. E., and Boone, B. E. (2019). Next-Generation Sequencing Strategies. *Cold Spring Harb. Perspect. Med.* 9 (7), a025791. doi:10.1101/cshperspect.a025791

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/map Format and SAMtools. *Bioinformatics* 25 (16), 2078–2079. doi:10.1093/bioinformatics/btp352

Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., et al. (2012). Comparison of Next-Generation Sequencing Systems. *J. Biomed. Biotechnol.* 2012, 1–11. doi:10.1155/2012/251364

Lyu, X., Yang, Q., Zhao, F., and Liu, Y. (2021). Codon Usage and Protein Length-dependent Feedback from Translation Elongation Regulates Translation Initiation and Elongation Speed. *Nucleic acids Res.* 49 (16), 9404–9423. doi:10.1093/nar/gkab729

Ma, Y., Zhang, L., and Huang, X. (2014). Genome Modification by CRISPR/Cas9. *Febs J.* 281 (23), 5186–5193. doi:10.1111/febs.13110

Manghwar, H., Lindsey, K., Zhang, X., and Jin, S. (2019). CRISPR/Cas System: Recent Advances and Future Prospects for Genome Editing. *Trends plant Sci.* 24 (12), 1102–1125. doi:10.1016/j.tplants.2019.09.006

Mattick, J. S., and Makunin, I. V. (2006). Non-coding RNA. *Hum. Mol. Genet.* 15 (Suppl. l_1), R17–R29. doi:10.1093/hmg/ddl046

McAlister, F. A., Laupacis, A., and Armstrong, P. W. (2017). Finding the Right Balance between Precision Medicine and Personalized Care. *Cmaj* 189 (33), E1065–E1068. doi:10.1503/cmaj.170107

McDonnell, J. M., Evans, S. R., McCarthy, L., Temperley, H., Waters, C., Ahern, D., et al. (2021). The Diagnostic and Prognostic Value of Artificial Intelligence and Artificial Neural Networks in Spinal Surgery. *bone & Jt. J.* 103-B (9), 1442–1448. doi:10.1302/0301-620X.103B9.BJJ-2021-0192.R1

J. M. McGinnis, L. Olsen, W. A. Goolsby, and C. Grossmann (Editors) (2011). *Clinical Data as the Basic Staple of Health Learning: Creating and Protecting a Public Good: Workshop Summary* (Washington, DC: National Academies Press).

Megías-Vericat, J. E., Martínez-Cuadrón, D., Herrero, M. J., Rodríguez-Veiga, R., Solana-Altabella, A., Boluda, B., et al. (2021). Impact of Combinations of Single-Nucleotide Polymorphisms of Anthracycline Transporter Genes upon the Efficacy and Toxicity of Induction Chemotherapy in Acute Myeloid Leukemia. *Leukemia lymphoma* 62 (3), 659–668. doi:10.1080/10428194.2020.1839650

Meienberg, J., Bruggmann, R., Oexle, K., and Matyas, G. (2016). Clinical Sequencing: Is WGS the Better WES? *Hum. Genet.* 135 (3), 359–362. doi:10.1007/s00439-015-1631-9

Metzker, M. L., Raghavachari, R., Richards, S., Jacutin, S. E., Civitello, A., Burgess, K., et al. (1994). Termination of DNA Synthesis by Novel 3'-modifieddeoxyribonucleoside 5'-triphosphates. *Nucl. Acids Res.* 22, 4259–4267. doi:10.1093/nar/22.20.4259

Michael, C. L., Sholle, E. T., Wulff, R. T., Roboz, G. J., and Campion, T. R., Jr (2020). Mapping Local Biospecimen Records to the OMOP Common Data Model. *AMIA Jt. Summits Transl. Sci. Proc.* 2020, 422–429.

Micheel, C. M., Sweeney, S. M., LeNoue-Newton, M. L., André, F., Bedard, P. L., Guinney, J., et al.AACR Project GENIE Consortium (2018). American Association for Cancer Research Project Genomics Evidence Neoplasia Information Exchange: From Inception to First Data Release and Beyond-Lessons Learned and Member Institutions' Perspectives. *JCO Clin. cancer Inf.* 2, 1–14. doi:10.1200/CCI.17.00083

Mintz, Y., and Brodie, R. (2019). Introduction to Artificial Intelligence in Medicine. *Minim. Invasive Ther. Allied Technol.* 28 (2), 73–81. doi:10.1080/13645706.2019.1575882

Mishra, S., Yadav, T., and Rani, V. (2016). Exploring miRNA Based Approaches in Cancer Diagnostics and Therapeutics. *Crit. Rev. oncology/hematology* 98, 12–23. doi:10.1016/j.critrevonc.2015.10.003

Montes, M., Sanford, B. L., Comiskey, D. F., and Chandler, D. S. (2019). RNA Splicing and Disease: Animal Models to Therapies. *Trends Genet.* 35 (1), 68–87. doi:10.1016/j.tig.2018.10.002

Morash, M., Mitchell, H., Beltran, H., Elemento, O., and Pathak, J. (2018). The Role of Next-Generation Sequencing in Precision Medicine: A Review of Outcomes in Oncology. *Jpm* 8 (3), 30. doi:10.3390/jpm8030030

Morganti, S., Tarantino, P., Ferraro, E., D'Amico, P., Duso, B. A., and Curigliano, G. (2019). Next Generation Sequencing (NGS): a Revolutionary Technology in Pharmacogenomics and Personalized Medicine in Cancer, *Adv. Exp. Med. Biol.* 1168, 9–30. doi:10.1007/978-81-322-1184-6_310.1007/978-3-030-24100-1_2

Moscatelli, M., Manconi, A., Pessina, M., Fellegara, G., Rampoldi, S., Milanesi, L., et al. (2018). An Infrastructure for Precision Medicine through Analysis of Big Data. *BMC Bioinforma.* 19 (Suppl. 10), 351. doi:10.1186/s12859-018-2300-5

Nakagawa, H., and Fujita, M. (2018). Whole Genome Sequencing Analysis for Cancer Genomics and Precision Medicine. *Cancer Sci.* 109 (3), 513–522. doi:10.1111/cas.13505

Nakatsuka, N., Moorjani, P., Rai, N., Sarkar, B., Tandon, A., Patterson, N., Bhavani, G. S., Girisha, K. M., Mustak, M. S., Srinivasan, S., Kaushik, A., Vahab, S. A., Jagadeesh, S. M., Satyamoorthy, K., Singh, L., Reich, D., and Thangaraj, K. (2017). The Promise of Discovering Population-specific Disease-Associated Genes in South Asia. *Nat. Genet.* 49 (9), 1403–1407. doi:10.1038/ng.3917

Niblock, M., and Gallo, J.-M. (2012). Tau Alternative Splicing in Familial and Sporadic Tauopathies. *Biochem. Soc. Trans.* 40 (4), 677–680. doi:10.1042/bst20120091

O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., Astashyn, A., Badretdin, A., Bao, Y., Blinkova, O., Brover, V., Chetvernin, V., Choi, J., Cox, E., Ermolaeva, O., Farrell, C. M., Goldfarb, T., Gupta, T., Haft, D., Hatcher, E., Hlavina, W., Joardar, V. S., Kodali, V. K., Li, W., Maglott, D., Masterson, P., McGarvey, K. M., Murphy, M. R., O'Neill, K., Pujar, S., Rangwala, S. H., Rausch, D., Riddick, L. D., Schoch, C., Shkeda, A., Storz, S. S., Sun, H., Thibaud-Nissen, F., Tolstoy, I., Tully, R. E., Vatsan, A. R., Wallin, C., Webb, D., Wu, W., Landrum, M. J., Kimchi, A., Tatusova, T., DiCuccio, M., Kitts, P., Murphy, T. D., and Pruitt, K. D. (2016). Reference Sequence (RefSeq) Database at NCBI: Current Status, Taxonomic Expansion, and Functional Annotation. *Nucleic Acids Res.* 44 (D1), D733–D745. doi:10.1093/nar/gkv1189

Orkin, S. H., and Bauer, D. E. (2019). Emerging Genetic Therapy for Sickle Cell Disease. *Annu. Rev. Med.* 70, 257–271. doi:10.1146/annurev-med-041817-125507

Panni, S., Lovering, R. C., Porras, P., and Orchard, S. (2020). Non-coding RNA Regulatory Networks. *Biochimica Biophysica Acta (BBA) - Gene Regul. Mech.* 1863 (6), 194417. doi:10.1016/j.bbagrm.2019.194417

Pastorino, R., De Vito, C., Migliara, G., Glocker, K., Binenbaum, I., Ricciardi, W., et al. (2019). Benefits and Challenges of Big Data in Healthcare: an Overview of the European Initiatives. *Eur. J. public health* 29 (Suppl. ment_3), 23–27. doi:10.1093/eurpub/ckz168

Payami, H., and Factor, S. A. (2014). Promise of Pharmacogenomics for Drug Discovery, Treatment and Prevention of Parkinson's Disease. A Perspective. *Neurotherapeutics* 11 (1), 111–116. doi:10.1007/s13311-013-0237-y

Penny, M. A., and McHale, D. (2005). Pharmacogenomics and the Drug Discovery Pipeline. *Am. J. PharmacoGenomics* 5 (1), 53–62. doi:10.2165/00129785-200505010-00005

Petersen, B.-S., Fredrich, B., Hoeppner, M. P., Ellinghaus, D., and Franke, A. (2017). Opportunities and Challenges of Whole-Genome and -exome Sequencing. *BMC Genet.* 18 (1), 1–13. doi:10.1186/s12863-017-0479-5

Phizicky, E. M., and Hopper, A. K. (2010). tRNA Biology Charges to the Front. *Genes. Dev.* 24 (17), 1832–1860. doi:10.1101/gad.1956510

Picard, M., Scott-Boyer, M.-P., Bodein, A., Périn, O., and Droit, A. (2021). Integration Strategies of Multi-Omics Data for Machine Learning Analysis. *Comput. Struct. Biotechnol. J.* 19, 3735–3746. doi:10.1016/j.csbj.2021.06.030

Pinho, J. R. R. (2017). Precision Medicine. *Einstein (São Paulo)* 15 (1), VII–X. doi:10.1590/S1679-45082017ED4016

Pirmohamed, M. (2006). Genetic Factors in the Predisposition to Drug-Induced Hypersensitivity Reactions. *Aaps J.* 8 (1), E20–E26. doi:10.1208/aapsj080103

Polychronopoulos, D., King, J. W. D., Nash, A. J., Tan, G., and Lenhard, B. (2017). Conserved Non-coding Elements: Developmental Gene Regulation Meets Genome Organization. *Nucleic acids Res.* 45 (22), 12611–12624. doi:10.1093/nar/gkx1074

Prosperi, M., Min, J. S., Bian, J., and Modave, F. (2018). Big Data Hurdles in Precision Medicine and Precision Public Health. *BMC Med. Inf. Decis. Mak.* 18 (1), 1–15. doi:10.1186/s12911-018-0719-2

Ramesh, A., Kambhampati, C., Monson, J., and Drew, P. (2004). Artificial Intelligence in Medicine. *Ann. R. Coll. Surg. Engl.* 86 (5), 334–338. doi:10.1308/147870804290

Reel, P. S., Reel, S., Pearson, E., Trucco, E., and Jefferson, E. (2021). Using Machine Learning Approaches for Multi-Omics Data Analysis: A Review. *Biotechnol. Adv.* 49, 107739. doi:10.1016/j.biotechadv.2021.107739

Relling, M. V., and Evans, W. E. (2015). Pharmacogenomics in the Clinic. *Nature* 526 (7573), 343–350. doi:10.1038/nature15817

Rhoads, A., and Au, K. F. (2015). PacBio Sequencing and its Applications. *Genomics, proteomics Bioinforma.* 13 (5), 278–289. doi:10.1016/j.gpb.2015.08.002

Rim, M. H., Smith, L., and Kelly, M. (2016). Implementation of a Patient-Focused Specialty Pharmacy Program in an Academic Healthcare System. *Am. J. Health-System Pharm.* 73 (11), 831–838. doi:10.2146/ajhp150947

Roth, S. C. (2019). What Is Genomic Medicine? *jmla* 107 (3), 442 448. doi:10.5195/jmla.2019.604

Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA Sequencing with Chain-Terminating Inhibitors. *Proc. Natl. Acad. Sci. U.S.A.* 74 (12), 5463–5467. doi:10.1073/pnas.74.12.5463

Schneider, V. A., Graves-Lindsay, T., Howe, K., Bouk, N., Chen, H.-C., Kitts, P. A., Murphy, T. D., Pruitt, K. D., Thibaud-Nissen, F., Albracht, D., Fulton, R. S., Kremitzki, M., Magrini, V., Markovic, C., McGrath, S., Steinberg, K. M., Auger, K., Chow, W., Collins, J., Harden, G., Hubbard, T., Pelan, S., Simpson, J. T., Threadgold, G., Torrance, J., Wood, J. M., Clarke, L., Koren, S., Boitano, M., Peluso, P., Li, H., Chin, C.-S., Phillippy, A. M., Durbin, R., Wilson, R. K., Flicek, P., Eichler, E. E., and Church, D. M. (2017). Evaluation of GRCh38 and De Novo Haploid Genome Assemblies Demonstrates the Enduring Quality of the Reference Assembly. *Genome Res.* 27 (5), 849–864. doi:10.1101/gr.213611.116

Scotti, M. M., and Swanson, M. S. (2016). RNA Mis-Splicing in Disease. *Nat. Rev. Genet.* 17 (1), 19–32. doi:10.1038/nrg.2015.3

Seyed Tabib, N. S., Madgwick, M., Sudhakar, P., Verstockt, B., Korcsmaros, T., and Vermeire, S. (2020). Big Data in IBD: Big Progress for Clinical Practice. *Gut* 69 (8), 1520–1532. doi:10.1136/gutjnl-2019-320065

Shen, J., Li, W., Deng, S., and Zhang, T. (2021). Supervised and Unsupervised Learning of Directed Percolation. *Phys. Rev. E* 103 (5-1), 052140. doi:10.1103/PhysRevE.103.052140

Shendure, J. A., Porreca, G. J., and Church, G. M. (2008). Overview of DNA Sequencing Strategies. *Curr. Protoc. Mol. Biol.* 81 (1), 7–1. doi:10.1002/0471142727.mb0701s81

Shull, J. G. (2019). Digital Health and the State of Interoperable Electronic Health Records. *JMIR Med. Inf.* 7 (4), e12712. doi:10.2196/12712

Simpson, L. J., Tzima, E., and Reader, J. S. (2020). Mechanical Forces and Their Effect on the Ribosome and Protein Translation Machinery. *Cells* 9 (3), 650. doi:10.3390/cells9030650

Slack, F. J., and Chinnaiyan, A. M. (2019). The Role of Non-coding RNAs in Oncology. *Cell.* 179 (5), 1033–1055. doi:10.1016/j.cell.2019.10.017

Smed, V. M., Petersen, O. B. B., Gerdes, A. A., Diness, B. R., and Roos, L. S. (2021). Genetic Screening of Prospective Parents. *Ugeskr. Laeger* 183 (13). doi:10.1016/s0300-7073(11)70116-5

Sørlie, T., Tibshirani, R., Parker, J., Hastie, T., Marron, J. S., Nobel, A., Deng, S., Johnsen, H., Pesich, R., Geisler, S., Demeter, J., Perou, C. M., Lønning, P. E., Brown, P. O., Børresen-Dale, A.-L., and Botstein, D. (2003). Repeated Observation of Breast Tumor Subtypes in Independent Gene Expression Data Sets. *Proc. Natl. Acad. Sci. U.S.A.* 100 (14), 8418–8423. doi:10.1073/pnas.0932692100

Stein, J. D., Lum, F., Lee, P. P., Rich, W. L., III, and Coleman, A. L. (2014). Use of Health Care Claims Data to Study Patients with Ophthalmologic Conditions. *Ophthalmology* 121 (5), 1134–1141. doi:10.1016/j.ophtha.2013.11.038

Stelzer, G., Rosen, N., Plaschkes, I., Zimmerman, S., Twik, M., Fishilevich, S., Stein, T. I., Nudel, R., Lieder, I., Mazor, Y., Kaplan, S., Dahary, D., Warshawsky, D., Guan-Golan, Y., Kohn, A., Rappaport, N., Safran, M., and Lancet, D. (2016). The GeneCards Suite: from Gene Data Mining to Disease Genome Sequence Analyses. *Curr. Protoc. Bioinforma.* 54 (1), 1–30. doi:10.1002/cpbi.5

Stenson, P. D., Ball, E. V., Mort, M., Phillips, A. D., Shaw, K., and Cooper, D. N. (2012). The Human Gene Mutation Database (HGMD) and its Exploitation in

the Fields of Personalized Genomics and Molecular Evolution. *Curr. Protoc. Bioinforma.* 39 (1), 1–13. doi:10.1002/0471250953.bi0113s39

Stenson, P. D., Mort, M., Ball, E. V., Evans, K., Hayden, M., Heywood, S., et al. (2017). The Human Gene Mutation Database: towards a Comprehensive Repository of Inherited Mutation Data for Medical Research, Genetic Diagnosis and Next-Generation Sequencing Studies. *Hum. Genet.* 136 (6), 665–677. doi:10.1136/jmg.2007.05521010.1007/s00439-017-1779-6

Strande, N. T., Riggs, E. R., Buchanan, A. H., Ceyhan-Birsoy, O., DiStefano, M., Dwight, S. S., Goldstein, J., Ghosh, R., Seifert, B. A., Sneddon, T. P., Wright, M. W., Milko, L. V., Cherry, J. M., Giovanni, M. A., Murray, M. F., O'Daniel, J. M., Ramos, E. M., Santani, A. B., Scott, A. F., Plon, S. E., Rehm, H. L., Martin, C. L., and Berg, J. S. (2017). Evaluating the Clinical Validity of Gene-Disease Associations: an Evidence-Based Framework Developed by the Clinical Genome Resource. *Am. J. Hum. Genet.* 100 (6), 895–906. doi:10.1016/j.ajhg.2017.04.015

Sweeney, C., Baras, A., Pugh, T., Schultz, N., Stricker, T., Lindsay, J., et al. (2017). AACR Project Genie: Powering Precision Medicine through an International Consortium. *Cancer Discov.* 7 (8), 818–831. doi:10.1158/2159-8290.CD-17-0151

Takeshima, Y., Yagi, M., Okizuka, Y., Awano, H., Zhang, Z., Yamauchi, Y., Nishio, H., and Matsuo, M. (2010). Mutation Spectrum of the Dystrophin Gene in 442 Duchenne/Becker Muscular Dystrophy Cases from One Japanese Referral Center. *J. Hum. Genet.* 55 (6), 379–388. doi:10.1038/jhg.2010.49

Tzelepis, K., Koike-Yusa, H., De Braekeleer, E., Li, Y., Metzakopian, E., Dovey, O. M., Mupo, A., Grinkevich, V., Li, M., Mazan, M., Gozdecka, M., Ohnishi, S., Cooper, J., Patel, M., McKerrell, T., Chen, B., Domingues, A. F., Gallipoli, P., Teichmann, S., Ponstingl, H., McDermott, U., Saez-Rodriguez, J., Huntly, B. J. P., Iorio, F., Pina, C., Vassiliou, G. S., and Yusa, K. (2016). A CRISPR Dropout Screen Identifies Genetic Vulnerabilities and Therapeutic Targets in Acute Myeloid Leukemia. *Cell. Rep.* 17 (4), 1193–1205. doi:10.1016/j.celrep.2016.09.079

Ule, J., and Blencowe, B. J. (2019). Alternative Splicing Regulatory Networks: Functions, Mechanisms, and Evolution. *Mol. Cell.* 76 (2), 329–345. doi:10.1016/j.molcel.2019.09.017

Vadapalli, S., Abdelhalim, H., Zeeshan, S., and Ahmed, Z. (2022). Artificial Intelligence and Machine Learning Approaches Using Gene Expression and Variant Data for Personalized Medicine. *Briefings Bioinforma.*, bbac191. doi:10.1093/bib/bbac191

Van Dam, S., Võsa, U., van der Graaf, A., Franke, L., and de Magalhães, J. P. (2018). Gene Co-expression Analysis for Functional Classification and Gene-Disease Predictions. *Brief. Bioinform* 19 (4), bbw139–592. doi:10.1093/bib/bbw139

Venkatachalapathy, P., Padhilahouse, S., Sellappan, M., Subramanian, T., Kurian, S. J., Miraj, S. S., et al. (2021). Pharmacogenomics and Personalized Medicine in Type 2 Diabetes Mellitus: Potential Implications for Clinical Practice. *Pgpm* Vol. 14, 1441–1455. doi:10.2147/PGPM.S329787

Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., et al. (2001). The Sequence of the Human Genome. *science* 291 (5507), 1304–1351. doi:10.1016/s0002-9394(01)01077-710.1126/science.1058040

Wake, D. T., Ilbawi, N., Dunnenberger, H. M., and Hulick, P. J. (2019). Pharmacogenomics. *Med. Clin. N. Am.* 103 (6), 977–990. doi:10.1016/j.mcna.2019.07.002

Wang, E., and Aifantis, I. (2020). RNA Splicing and Cancer. *Trends Cancer* 6 (8), 631–644. doi:10.1016/j.trecan.2020.04.011

Welch, M. B., and Burgess, K. (1999). Synthesis of Fluorescent, Photolabile 3′-O-Protected Nucleoside Triphosphates for the Base Addition Sequencing Scheme. *Nucleosides Nucleotides* 18 (2), 197–201. doi:10.1080/15257779908043067

Wickenhagen, A., Sugrue, E., Lytras, S., Kuchi, S., Noerenberg, M., Turnbull, M. L., Loney, C., Herder, V., Allan, J., Jarmson, I., Cameron-Ruiz, N., Varjak, M., Pinto, R. M., Lee, J. Y., Iselin, L., Palmalux, N., Stewart, D. G., Swingler, S., Greenwood, E. J. D., Crozier, T. W. M., Gu, Q., Davies, E. L., Clohisey, S., Wang, B., Trindade Maranhão Costa, F., Freire Santana, M., de Lima Ferreira, L. C., Murphy, L., Fawkes, A., Meynert, A., Grimes, G., au, fnm., Da Silva Filho, J. L., Marti, M., Hughes, J., Stanton, R. J., Wang, E. C. Y., Ho, A., Davis, I., Jarrett, R.

F., Castello, A., Robertson, D. L., Semple, M. G., Openshaw, P. J. M., Palmarini, M., Lehner, P. J., Baillie, J. K., Rihn, S. J., and Wilson, S. J. (2021). A Prenylated dsRNA Sensor Protects against Severe COVID-19. *Science* 374 (6567), eabj3624. doi:10.1126/science.abj3624

Woese, C. R. (1969). "The Biological Significance of the Genetic Code," in *Progress in Molecular and Subcellular Biology* (Berlin, Heidelberg: Springer), 5–46. doi:10.1007/978-3-642-46200-9_2

Wronikowska, M. W., Malycha, J., Morgan, L. J., Westgate, V., Petrinic, T., Young, J. D., et al. (2021). Systematic Review of Applied Usability Metrics within Usability Evaluation Methods for Hospital Electronic Healthcare Record Systems. *J. Eval. Clin. Pract.* 27 (6), 1403–1416. doi:10.1111/jep.13582

Xiao, Q., Guo, D., and Chen, S. (2019). Application of CRISPR/Cas9-based Gene Editing in HIV-1/AIDS Therapy. *Front. Cell. Infect. Microbiol.* 9, 69. doi:10.3389/fcimb.2019.00069

Xu, B., Meng, Y., and Jin, Y. (2021). RNA Structures in Alternative Splicing and Back-splicing. *WIREs RNA* 12 (1), e1626. doi:10.1002/wrna.1626

Xu, J., Yang, P., Xue, S., Sharma, B., Sanchez-Martin, M., Wang, F., et al. (2019). Translating Cancer Genomics into Precision Medicine with Artificial Intelligence: Applications, Challenges and Future Perspectives. *Hum. Genet.* 138 (2), 109–124. doi:10.1007/s00439-019-01970-5

Yan, Q., Zhu, C., Guang, S., and Feng, X. (2019). The Functions of Non-coding RNAs in rRNA Regulation. *Front. Genet.* 10, 290. doi:10.3389/fgene.2019.00290

Yeh, V. M., Bergner, E. M., Bruce, M. A., Kripalani, S., Mitrani, V. B., Ogunsola, T. A., et al. (2020). Can Precision Medicine Actually Help People like Me? African American and Hispanic Perspectives on the Benefits and Barriers of Precision Medicine. *Ethn. Dis.* 30 (Suppl. 1), 149–158. doi:10.18865/ed.30.S1.149

Yohe, S., and Thyagarajan, B. (2017). Review of Clinical Next-Generation Sequencing. *Archives pathology laboratory Med.* 141 (11), 1544–1557. doi:10.5858/arpa.2016-0501-ra

Zeeshan, S., Xiong, R., Liang, B. T., and Ahmed, Z. (2020). 100 Years of Evolving Gene-Disease Complexities and Scientific Debutants. *Briefings Bioinforma.* 21 (3), 885–905. doi:10.1093/bib/bbz038

Zhang, H. (2016). "Overview of Sequence Data Formats," in *Statistical Genomics* (New York, NY: Humana Press), 3–17. doi:10.1007/978-1-4939-3578-9_1

Zhao, E. Y., Jones, M., and Jones, S. J. M. (2019). Whole-genome Sequencing in Cancer. *Cold Spring Harb. Perspect. Med.* 9 (3), a034579. doi:10.1101/cshperspect.a034579

Zhao, S. (2019). Alternative Splicing, RNA-Seq and Drug Discovery. *Drug Discov. Today* 24 (6), 1258–1267. doi:10.1016/j.drudis.2019.03.030

Zhou, H., and Skolnick, J. (2016). A Knowledge-Based Approach for Predicting Gene-Disease Associations. *Bioinformatics* 32 (18), 2831–2838. doi:10.1093/bioinformatics/btw358

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

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