life

MDPI

Review

Artificial Intelligence and Cardiovascular Genetics

Chayakrit Krittanawong [1,2,3,4,*], Kipp W. Johnson [3,5], Edward Choi [6], Scott Kaplin [2], Eric Venner [4], Mullai Murugan [7], Zhen Wang [8,9], Benjamin S. Glicksberg [3,5], Christopher I. Amos [10], Michael C. Schatz [11,12] and W. H. Wilson Tang [13,14,15]

1 Section of Cardiology, Baylor College of Medicine, Houston, TX 77030, USA
2 Department of Cardiovascular Medicine, NYU Langone, New York, NY 10016, USA; scott.kaplin@nyulangone.org
3 The Hasso Plattner Institute for Digital Health at the Mount Sinai, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; kipp.johnson@icahn.mssm.edu (K.W.J.); benjamin.glicksberg@mssm.edu (B.S.G.)
4 Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA; venner@bcm.edu
5 Department of Genetics and Genomic Sciences, Institute for Next Generation Healthcare, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
6 Google Health Research, Google, Mountain View, CA 94043, USA; edwardchoi@kaist.ac.kr
7 Human Genome Sequencing Center, Department of Software Development, Baylor College of Medicine, Houston, TX 77030, USA; mullai.murugan@bcm.edu
8 Robert D. and Patricia E. Kern Center for the Science of Health Care Delivery, Mayo Clinic, Rochester, MN 55905, USA; wang.zhen@mayo.edu
9 Division of Health Care Policy and Research, Department of Health Sciences Research, Mayo Clinic, Rochester, MN 55905, USA
10 Dan L Duncan Comprehensive Cancer Center, Baylor College of Medicine, Houston, TX 77030, USA; chris.amos@bcm.edu
11 Department of Computer Science, Johns Hopkins University, Baltimore, MD 21218, USA; mschatz@cs.jhu.edu
12 Department of Biology, Johns Hopkins University, Baltimore, MD 21218, USA
13 Department of Cardiovascular Medicine, Heart and Vascular Institute, Cleveland Clinic, Cleveland, OH 44195, USA; tangw@ccf.org
14 Department of Cellular and Molecular Medicine, Lerner Research Institute, Cleveland, OH 44195, USA
15 Center for Clinical Genomics, Cleveland Clinic, Cleveland, OH 44195, USA
* Correspondence: chayakrit.krittanawong@va.gov; Tel.: +1-713-798-4951

check for updates

Citation: Krittanawong, C.; Johnson, K.W.; Choi, E.; Kaplin, S.; Venner, E.; Murugan, M.; Wang, Z.; Glicksberg, B.S.; Amos, C.I.; Schatz, M.C.; et al. Artificial Intelligence and Cardiovascular Genetics. *Life* **2022**, *12*, 279. https://doi.org/10.3390/life12020279

Academic Editors: Lefteris Koumakis, Pere Puigbò, George Potamias and Alexandros Kanterakis

Received: 30 December 2021
Accepted: 9 February 2022
Published: 14 February 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

Abstract: Polygenic diseases, which are genetic disorders caused by the combined action of multiple genes, pose unique and significant challenges for the diagnosis and management of affected patients. A major goal of cardiovascular medicine has been to understand how genetic variation leads to the clinical heterogeneity seen in polygenic cardiovascular diseases (CVDs). Recent advances and emerging technologies in artificial intelligence (AI), coupled with the ever-increasing availability of next generation sequencing (NGS) technologies, now provide researchers with unprecedented possibilities for dynamic and complex biological genomic analyses. Combining these technologies may lead to a deeper understanding of heterogeneous polygenic CVDs, better prognostic guidance, and, ultimately, greater personalized medicine. Advances will likely be achieved through increasingly frequent and robust genomic characterization of patients, as well the integration of genomic data with other clinical data, such as cardiac imaging, coronary angiography, and clinical biomarkers. This review discusses the current opportunities and limitations of genomics; provides a brief overview of AI; and identifies the current applications, limitations, and future directions of AI in genomics.

Keywords: genomics; AI; genetics; deep learning; cardiovascular disease; cardiology; machine learning; artificial intelligence

1. Introduction

Multiple diseases of the cardiovascular system are associated with genetic polymorphisms including both common conditions, such as hypercholesterolemia [1,2] and less common conditions, such as cardiac channelopathies [3], cardiomyopathies [4], aortopathies [5],

and various structural and congenital diseases of the heart and great vessels [6]. Given that the fields of cardiovascular genetics and precision medicine are rapidly evolving, it is unsurprising that recently published guidelines include an increased focus on genetic testing. The 2020 Scientific Statement From the American Heart Association (AHA) on Genetic Testing for Inherited Cardiovascular Diseases recommended testing specific genes in certain monogenic cardiovascular diseases (CVDs) in appropriate clinical circumstances [7] (e.g., *LDLR*, *APOB*, and *PCSK9* genes for familial hypercholesterolemia, and *TTN*, *LMNA*, *MYH7*, *TNNT2*, *BAG3*, *RBM20*, *TNNC1*, *TNNI3*, *TPM1*, *SCN5A*, and *PLN* genes for dilated cardiomyopathy). The 2021 Scientific Statement from the AHA on Genetic Testing for Heritable Cardiovascular Diseases in Pediatric Patients also recommended cardiovascular genetic testing in children as an important component in determining the risk of developing heritable cardiovascular diseases in adulthood [8]. With advancements in technology, several recent genetic studies have revealed potential targets for CVD screening and therapies. For example, a recent genome-wide association study of 2780 cases and 47,486 controls identified 12 genome-wide susceptibility loci which were significant for hypertrophic cardiomyopathy (HCM), and found that single-nucleotide polymorphism heritability indicated a strong polygenic influence, especially for sarcomere-negative HCM (64% of cases; h2g = 0.34 ± 0.02) [9]. Another recent study of patients with hereditary transthyretin (TTR) cardiac amyloidosis with polyneuropathy showed that administration of NTLA-2001 led to a decrease in serum TTR protein concentrations through targeted knockout of *TTR*. Hence, genetic screening of TTR may, thus, prove to be increasingly useful in the future as it may allow susceptible patients to be identified and treated appropriately at an earlier stage of disease [10]. On the other hand, genetic testing in polygenic CVDs, with their inherently more complicated genetic etiology, remains challenging.

Artificial intelligence (AI) is a discipline of computer science that aims to mimic human thought processes, learning capacity, and knowledge storage [11]. A central tenet of AI is learning the value of potential choices rather than rigidly following predetermined thresholds or procedures, e.g., optimizing the selection of variants to maximize the predictive accuracy for disease risk rather than using a predetermined list. AI involves several components, including machine learning and deep learning, with increasing potential to explore novel CVD genotypes and phenotypes, among many other exciting opportunities. In this review, we summarize several important current limitations of genomics; provide a brief overview of AI; and identify the current applications, limitations, and potential future directions of AI in cardiovascular genetics.

2. Genetic Testing Gap in Cardiovascular Diseases

The majority of CVDs and cardiovascular risk factors have a significant genetic component, which is most commonly polygenic in origin [1,2]. Current clinical practice utilizes a patient's medical history, family history, physical examination, cardiac biomarkers, and various modalities of cardiac imaging to establish diagnoses and to stratify risks. Despite rapid advances and availability of genetic testing panels, clinicians seldom utilize genetic testing as part of their initial patient assessments beyond cases with a known family history of genetic, inherited CVDs (e.g., HCM, arrhythmogenic right ventricular cardiomyopathy (ARVC), long QT syndrome (LQTS), or catecholaminergic polymorphic ventricular tachycardia (CPVT)). This lack of routine testing as part of care pathway creates a "diagnostic gap" (i.e., a delay in time from disease manifestation to establishing a definitive diagnosis) that can lead to inappropriate or ineffective treatment in patients suffering from inherited CVDs. In a recent study from Baylor College of Medicine's Human Genome Sequencing Center, 84% of surveyed physicians reported medical management changes, including specialist referrals, cardiac testing, and medication changes, after receiving the results of a panel of genes associated with CVDs [12].

Despite its demonstrated clinical relevance, current guidelines only recommend genomic testing for a small number of cardiac conditions (e.g., HCM, familial hypercholesterolemia), limited by the relatively few genetic tests that are currently available and the

lack of strong studies in cardiovascular genetics [13,14]. For example, Brugada syndrome has a large number of potentially pathogenic genetic variants (e.g., *CACNA1C*, *GPD1L*, *HEY2*, *PKP2*, *RANGRF*, *SCN10A*, *SCN1B*, *SCN2B*, *SCN3B*, *SLMAP*, and *TRPM4*) but current guidelines continue to recommend a comprehensive genetic analysis for only Brugada syndrome caused by the *SCN5A* genetic variant [15,16]. With advancements in genetic testing technologies, preemptive genetic testing for various cardiomyopathies may be useful in the presence of an asymptomatic type 1 Brugada ECG pattern, family history of dilated cardiomyopathy, or the development of spontaneous coronary artery dissection (SCAD). While a recent study by Murdock and colleagues demonstrated the diagnostic potential of genetics guided coronary artery disease (CAD) risk factor management based on *LPA* polymorphisms and polygenic risk, genetic testing for a selection of well-understood variant–phenotype associations remains very limited (i.e., a "treatment gap") [12]. With further research and development, comprehensive genetic testing could become routinely used in clinical cardiovascular practice and applied to primary disease prevention and the facilitation of precision cardiovascular medicine.

3. Next Generation Sequencing (NGS) in the Modern Clinic

Genomics is becoming nearly ubiquitous in biomedical research [17]. Large-scale sequencing efforts have revolutionized our understanding of the complex genetic interrelationships involved in the pathogenesis of most cardiovascular conditions [18]. The tremendous advancements in genomic research are largely driven by the advent of NGS, which has led to the discovery of novel associations and the ability to more easily assess genetic heterogeneity across patients. Several categories of NGS include: (1) whole genome sequencing (WGS); (2) whole exome sequencing (WES), where the sequencing is concentrated over the protein-coding regions of the genome (~2% of the genome); and (3) gene panels, where very deep coverage (>100× coverage) is generated for a select number of genes. Both WGS and WES allow for the accurate identification of single-nucleotide variants (SNVs), large copy number variations (CNVs), small insertion deletions (InDels), and information on variant frequencies in different populations [19]. Because WGS examines the noncoding regions of the genome, it offers a more comprehensive appraisal of both small and large genomic risk variants for CVDs. However, WGS is more costly and time-consuming than WES, and may be limited by lower depth [20,21]. Conversely, the results of WES, while more limited in scope, are typically viewed as more straightforward to interpret and historically have been a useful method to identify variants causing Mendelian disease. Panel-based NGS relies on high sequencing depth of previously determined important genetic loci, making this kind of testing more resource-efficient. However, the narrow focus of this type of assay results in decreased power to detect novel associations and is often less effective for assessing other types of genetic alterations, such as structural variants. Although NGS is now widely used due to its speed, robustness, and cost-effectiveness, orthogonal confirmation with the traditional Sanger sequencing method is sometimes still required for validation prior to clinical use [22–24].

Nonetheless, the implementation of AI to NGS and genomics has already been shown to accurately predict the consequences of genetic risk factors in CVDs [25,26], show the noncoding-variant effects in CVDs [27,28], find patients with cardiac amyloidosis [29,30], and initiate specific therapies from tumor sequencing [31] by integrating with electronic health records (EHRs) in several academic and medical institutions. Additionally, there are several direct-to-consumer genomics companies that use AI along with WGS and WES; however, to date, these applications have been limited by a lack of transparency in the algorithms they utilize due to their proprietary nature and commercial competition, as well as a lack of a consistent validation cohort, genomic guided clinical trials, and high-quality phenotype data that are consistently encoded and managed (Table 1). Although some direct-to-consumer companies have collaborated with academic institutions and published their methodologies, evidence for their clinical relevance remains scarce.

Table 1. Example of direct-to-consumer genomics companies.

Company	AI Algorithms	Input	Database	Limitations	More Information	Example Diseases
23andMe	ML models	Genetic variants	In-house 23andMe database and public databases (e.g., UK Biobank)	Heterogeneity of data (phenotypes, QC control for genetics) between UK Biobank and 23andMe	Map the impact of individuals' genetic material on phenotypes https://research.23andme.com/publications/ (accessed on 8 February 2022)	Weight pharmacogenetic testing
AncestryDNA	Not specified	Genotype samples on the Illumina OmniExpress platforms	AncestryDNA database	Serious privacy concerns	https://support.ancestry.com/s/article/AncestryDNA-White-Papers (accessed on 8 February 2022)	
Atomwise	ANN model	Gene targets and drug discovery	Public databases and proprietary sources	NA	Predict novel binding compounds; drug discovery ANN model runs an SBVS, which works well with convolution's ability of extracting local feature clusters from multidimensional input.	Prevent drug related cardiac toxicity
ATUM	ML to develop its Leap-In transposase technology	DNA synthesis Protein Antibody	Protein engineering (ProteinGPS) platform, public domain genetic databases, and proprietary platforms	NA	Enables any recombinant DNA sequence to behave as a transposon (a DNA sequence that can change its position within a genome altering the cell's genetic identity and genomic size) https://www.atum.bio/resources/archive/presentation-publications (accessed on 8 February 2022)	NA
BenevolentAI	Several models: BioNLP, BERT, deep learning, GuacaMol, Monte Carlo tree search, and symbolic AI		The Reaxys The Chemistry database The ChEMBL database The ZINC database	NA	Understanding the disease mechanisms at the earliest stage of our programs; identify the patients who are likely to respond to a treatment; identify drug targets that control these mechanism(s); and make drugs to correct them https://benevolent.ai/publications (accessed on 8 February 2022)	NA
Calico (Calico Life Sciences LLC)		Proteome Analysis GWAS	AncestryDNA database UK Biobank	NA	www.calicolabs.com/publications/ (accessed on 8 February 2022)	NA

Table 1. *Cont.*

Company	AI Algorithms	Input	Database	Limitations	More Information	Example Diseases
Color Genomics	ML models		Inhouse and industry (e.g., Agilent, Illumina and Hamilton)	No detail of ML model provided	https://www.color.com/wp-content/uploads/2019/12/Color-Hereditary-Heart-Health_WP_v3A.pdf (accessed on 8 February 2022)	Long QT syndrome (LQTS):Left ventricular noncompaction cardiomyopathy Fabry disease
CZ Biohub	ML models	Biochips embedded with human cells	Transcriptome data from animal model	NA	https://www.czbiohub.org/projects/ (accessed on 8 February 2022)	NA
Deep Genomics	Deep Learning	Several types of genetic data	European Genome-Phenome Archive	No detail of DL model provided	Identifying one or more genes responsible for a disease, potential drug therapies for an individual based on genome https://www.deepgenomics.com/platform/ (accessed on 8 February 2022)	Spinal muscular atrophy, nonpolyposis colorectal cancer, and autism
DNAnexus	DeepVariant	NGS data	Public database such as UK Biobank	NA	https://www.dnanexus.com/resources/case-studies (accessed on 8 February 2022)	NA
Fabric Genomics	Proprietary algorithms	NGS	Public database such as gnomAD (gnomad.broadinstitute.org/)	Proprietary model	A proprietary set of algorithms; The Variant Annotation, Analysis and Search Tool (AAST) and Phevor (Phenotype Driven Variant Ontological Re-ranking tool) https://fabricgenomics.com/resources/ (accessed on 8 February 2022)	NA
Freenome	Standard ML models such as logistic regression, principal component analysis (PCA) and support vector machine (SVM)	Whole-genome sequencing, cfDNA, cfRNA, and protein data	Proprietary sources and public database (e.g., NIH Roadmap Epigenome Mapping Consortium)	Proprietary sources	AI-EMERGE (NCT03688906)	NA
Futura Genetics		DNA from saliva		NA	APEX (arrayed primer extension) technology for detecting SNPs	NA

Table 1. *Cont.*

Company	AI Algorithms	Input	Database	Limitations	More Information	Example Diseases
Genoox	AI-based variant classification (aiVCE)	NGS	In-house exome database; public and in-house variant databases	NA	Diagnosis and treatment of genetic disorders and cancer, as well as new drug discovery and family planning; automated classification engine based on ACMG guidelines https://www.genoox.com/publications/ (accessed on 8 February 2022)	NA
Grail				NA	The Circulating Cell-free Genome Atlas (CCGA) Study The STRIVE Study SUMMIT Study https://grail.com/science/publications/ (accessed on 8 February 2022)	NA
IBM Watson for Genomics	NLP for several different predictive models	VCFs, CNV, and gene expression data abstracts and full-text articles	In-house hospital, PubMed and ClinicalTrials.gov	NA	Driver alterations, actionable variants, VUS, relevant therapies, and potential clinical trials https://www.ibm.com/us-en/marketplace/watson-for-genomics (accessed on 8 February 2022)	glioblastoma
Illumina	SpliceAI PrimateAI: deep residual neural network	NGS	Public databases (e.g., the ExAC/gnomAD database; the Single-Nucleotide Polymorphism Database (dbSNP); and ClinVar database	NA	Distinguish a handful of disease-causing mutations in patients with rare genetic diseases from a large number of benign variants present in healthy people https://www.illumina.com/science/publication-reviews.html (accessed on 8 February 2022)	NA
Karius	Proprietary Karius AI technology	blood test based on next-generation sequencing	NA	Proprietary model	https://www.kariusdx.com/clinical-data#publications (accessed on 8 February 2022)	endocarditis

Table 1. *Cont.*

Company	AI Algorithms	Input	Database	Limitations	More Information	Example Diseases
Nvidia and Scripps Research Translational Institute	Deep Learning	Development phase	NA	Still in development phase and not many details disclosed	Blood pressure monitoring; blood glucose genomics; digital wearable data	NA
Quest Diagnostics	Watson's cognitive computing and hc1's machine learning technology	Genome sequencing	In-house	No detail of ML model provided	https://www.hc1.com/blog/tag/quest-diagnostics/ (accessed on 8 February 2022)	NA
SOPHiA Genetics	Proprietary and standard algorithms (e.g., hidden Markov model algorithm)	NGS data	In-house and public databases (e.g., ClinVar, ExAC, and dbSNP)	NA	SNVs, Indels and CNVs detection, ALU insertions, Pseudogene variants differentiation and variant annotation https://www.sophiagenetics.com/en_US/hospitals/solutions/solutions/CAS.html (accessed on 8 February 2022)	arrhythmias (e.g., Long/Short QT syndrome or Brugada syndrome) and cardiomyopathies
Synpromics	ML models	Gene promoter design, a novel genomics-based platform	BIOBASE Biological Databases, UCSC GoldenPath, European Bioinformatics Institute	No detail of ML model provided	Predict the genomic sequences that are involved in cell type-specific regulation of gene expression	Design of Synthetic Mammalian Promoters
Verge Genomics	AI in pharmacogenomics	microRNA (miRNA)	Academic databases, research centers, and public databases (e.g., the NCBI database and the Molecular Signatures Database (MSigDB))	Proprietary AI model	AI-generated therapies for ALS and Parkinson by screening thousands genes https://www.vergegenomics.com/publications (accessed on 8 February 2022)	NA
Verily	DeepMass Project Baseline Health Study Status	Protein signals, genomics, and transcriptomics	Identify and quantify proteins	No validation	Integrate protein signals with other biomolecular data, such as genomics and transcriptomics, as well as with device measurements and disease status, to find out how genetics and behavior affect protein profiles https://blog.verily.com/2019/05/deepmass-new-machine-learning-method.html (accessed on 8 February 2022)	NA

Table 1. *Cont.*

Company	AI Algorithms	Input	Database	Limitations	More Information	Example Diseases
Veritas Genetics	ML models and AI Arvados Data Platform	Whole Genome Sequencing and Whole Exome Sequencing	Internal databases of two clinical testing laboratories (Laboratory for Molecular Medicine and Veritas Genetics) and public databases (e.g., ClinVar)	NA	https://www.veritasgenetics.com/in-the-news (accessed on 8 February 2022)	NA
Viome	Watson machine-learning	Gut microbiome	NA	No publications seen in Pubmed	https://www.viome.com/our-science (accessed on 8 February 2022)	NA

4. Introduction of AI to Clinical Cardiovascular Genetics

AI encompasses a broad range of applications for automated reasoning and inference, and is starting to have a major impact on clinical assessment and diagnosis. For example, in both United States of America (US) and United Kingdom (UK) datasets, AI outperformed human radiologists in screening mammography (greater than the AUC-ROC for the average radiologist by an absolute margin of 11.5%) and significantly reduced false positives and false negatives [32]. The most widely used groups of methods for pattern recognition in genomics include machine learning (ML) and deep learning (DL). Other AI approaches, for example natural language processing (NLP) and cognitive computing, are also starting to play a role in cardiovascular clinical care to enable more natural interactions between clinicians and computational systems [33–35]. Notably, the Food and Drug Administration (FDA) has been rapidly approving AI/ML-based medical devices and algorithms. Therefore, it is crucial for medical professionals to understand how best to utilize them. In a recent study using a web-based search for announcements of FDA approvals of AI/ML-based medical devices and algorithms, of the 64 found, 30 (46.9%), 16 (25.0%), and 10 (15.6%) were developed for the fields of radiology, cardiology, and internal medicine/general practice, respectively [36]. These AI approaches fundamentally work to train programs to recognize relationships within data. Table 2 provides examples of variant calling, reporting, and interpretation AI. Figure 1 demonstrates the potential of AI in cardiovascular genetics.

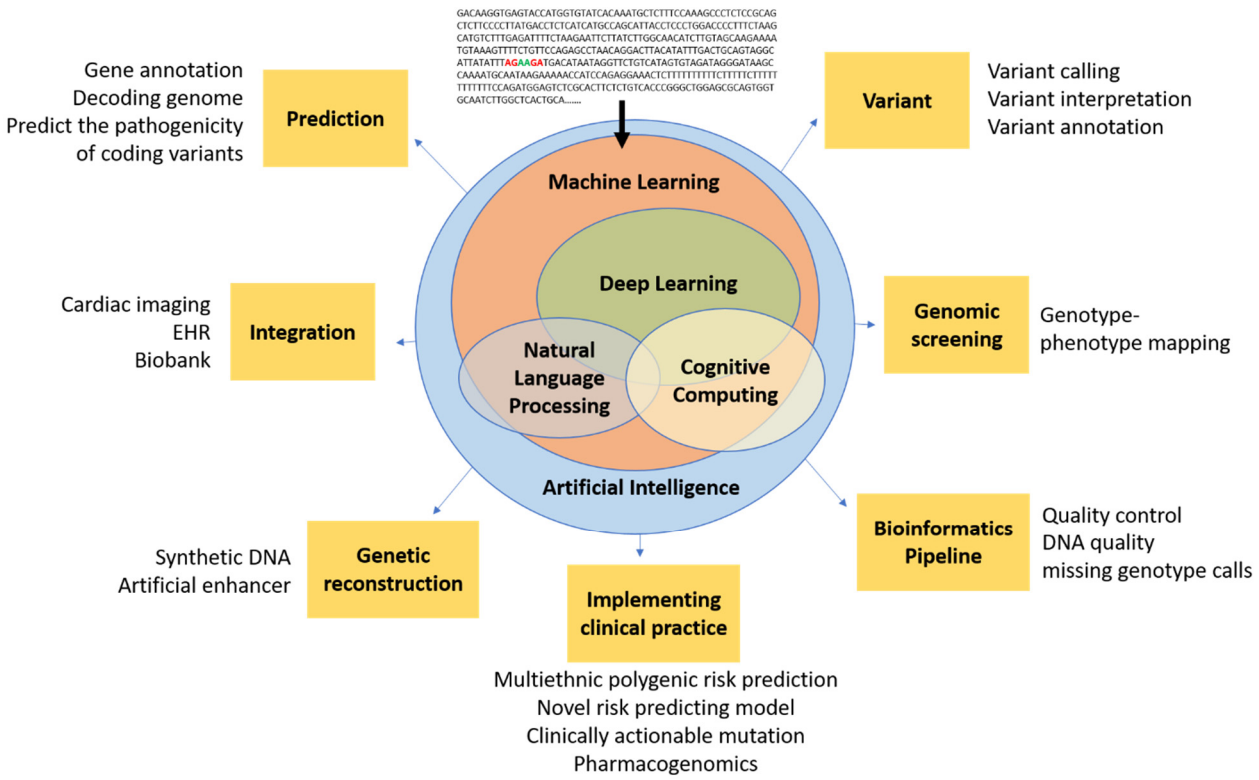

Figure 1. Conceptual schematic for artificial intelligence in cardiovascular genetics. Artificial intelligence encompasses a spectrum of concepts, including machine learning, NLP, and cognitive computing, which are generally enabled by deep learning and could ultimately be used in cardiovascular genomics for prediction, integration, reconstruction, bioinformatic techniques (e.g., pipeline, screening, variant analysis), and clinical practice. Artificial intelligence has the potential to filter raw genetic data into novel insights that could inform future clinical trials and, ultimately, clinical practice.

Table 2. Examples of variant calling, reporting, and interpretation AI.

Name	Algorithms	Example Function
DeepVariant [37]	Deep convolutional neural network (CNN)	Variant calling from short-read sequencing by reconstructing DNA alignments as an image
Clairvoyante [38]	A multi-task convolutional deep neural network	(1) Variant calling in single molecule sequencing (2) Predicts variant types (SNP or indel), zygosity, and alleles at the same time
Skyhawk [39]	Neural network	Mimics the process of expert review for clinically significant genomics variants identification
DeepBind [40]	Deep CNN	Predicts the binding sites of DNA-binding proteins and RBPs
iDeep [41]	Deep belief networks (DBN) and CNN	Cross-domain features and sequence information
DeepSEA [42]	Deep CNN	Predicts functional consequences of noncoding variants
DeepNano [43]	Recurrent neural networks (RNN)	Base calling in MinION nanopore reads
SpliceAI [44]	Deep neural network (DNN)	(1) Predicts splice junctions from an arbitrary pre-mRNA transcript sequence (2) Predicts noncoding genetic variants that cause cryptic splicing
DeepGestalt [45]	DNN	Distinguishes more than 200 rare diseases based on patient face images, which could also separate different genetic subtypes (e.g., Noonan syndrome)
DeepPVP [46]	DNN	Variant prioritization by integrating patients' phenotype information
DeepSVR [47]	Deep learning and random forest models	Predicts somatic variants confirmed by orthogonal validation sequencing data
DeepGene [48]	DNN	Extracts the high-level features between combinatorial somatic point mutations and cancer types. Classify cancer type
Deep AE [49]	Autoencoder	gene expression data
DeepMethyl [50]		Predicts methylation states of DNA CpG dinucleotides
BioVec [51]		Feature representation
DeepMotif [52]	Deep convolutional/highway MLP framework	Sequential data about gene regulation
DeepChrome [53]	Deep CNN	Sequential data about gene regulation Classifies gene expression using histone modification data as input.
Chiron [54]	Deep learning model	Translates the raw signal to DNA sequence
Variational Autoencoders [55]	Autoencoder	Predicts drug response
GARFIELD-NGS [56]	Deep CNN	Dissects false and true variants in exome sequencing
DeepGS [57]	Deep CNN	Predicts phenotypes from genotypes
DANN [58]	DNN	Predicts deleterious annotation or pathogenicity of genetic variants
DanQ [59]	Hybrid model Deep RNN and CNN	Quantifies the function of non-coding DNA
ProLanGO [60]	RNN	Protein function prediction
BCC-NER [61]	NLP	Bidirectional and contextual clues named entity tagger for gene/protein mention recognition
BioNLP [62]	NLP	Gene regulation network
SpaCy [63]	NLP	Tagging, parsing, and entity recognition

4.1. Machine Learning and Deep Learning

Since it is origins in the 1940s, ML has used algorithmic and statistical techniques to process data for a variety of purposes and applications [64,65]. ML concepts, such as supervised machine learning (e.g., support vector machines to distinguish between cases and controls) and unsupervised machine learning (e.g., a variety of models to reduce highly dimensional data into lower dimensional space), are common tools in genome-wide association studies (GWAS). In contrast to these types of ML, DL is a time- and resource-intensive subtype of ML that can achieve higher performance via its ability to learn complex representations from the data, depending on the task. Recently, advancements in computational power have enabled the application of DL onto large data sets (i.e., "big data") to build extremely expressive and complex multi-layer artificial neural networks (ANN) [66]. The initial success of DL began in image processing and recognition, where it can be used to recognize objects without explicitly defining the relevant features. For example, instead of trying to identify the specific contours of the nose, eyes, or mouth, the DL algorithm categorizes an object as a "face", which is recognized through a more abstract representation automatically learned from prior training on a dataset. In CVD, DL has been applied to non-imaging data, improving the accuracy of patient risk stratification and relationship prediction in comparison to traditional models, such as the Framingham Risk Score; although, typically, DL outperforms other models only on non-tabular data where there are complex nonlinear features that can be learned in a highly connected model [67,68].

Both ML and DL have their advantages for clinical genetics and carry the potential to improve the capabilities of cardiovascular genetics. As mentioned above, ML and DL can be further classified into supervised [69] and unsupervised [70,71] approaches. In a supervised approach, a classifier learns to predict known outcomes (e.g., predict the effect of a *LAMP2* mutation and understand its relationship to the phenotype of Danon Disease), while an unsupervised approach learns to infer relationships within the dataset (e.g., to identify subsets of patients who may carry similar genetic features or disease risk factors). ML has also been applied for several different tasks in NGS [72]. Support vector machine (SVM) models (learning methods used for classification, regression, and outlier detection) are used in high dimension datasets, similar to those used for predicting polygenic risk factors for hypertension [73] or inherited arrhythmias [74]. More complex ANN models have been used to predict advanced coronary artery calcium through a large-scale GWAS [75] and inheritable dilated cardiomyopathy through SNVs [76]. These ML models can also be used to cluster low-expression genes in pulmonary arterial hypertension [77].

The complexity of DL architecture creates challenges when analyzing large genomic data. There are several steps to analyzing genomic data using DL. First, before performing DL analysis, genomic data must be transformed into an appropriate data set for analysis and the network architecture should be designed to solve the specific cardiovascular task. "One-hot encoding" is a vector-based approach that has emerged as the most common method to represent genomic sequences for DL analysis, although other numerical representations (e.g., vectors, matrices, or tensors) and image-based approaches (e.g., DeepVariant transforming BAM files to images) have been proposed [78]. The second step is to design the network architecture. The major components of network architecture design include the type and resolution of the input filters and layers, the depth and density of the network, and a decision on the loss function regularization strategy. Once the genomic data and network architecture parameters are set, training the network with back propagation can begin [79].

The next step is to train the network. During training, the model parameters are learned by the network from the training data provided relative to the labeled examples using backpropagation and other related gradient descent learning techniques. The major challenge of this task is collecting enough training data and optimizing the hyperparameters (e.g., initialization strategy, learning rate, regularization techniques) so that the network can learn a robust set of parameters for the given prediction. It may also be necessary to

reconsider the overall network architecture if the performance remains low. Importantly, given enough training data, sufficient computational resources, and an appropriate network architecture, nearly any mathematical function can be learned, including highly abstract functions from genomics data or image data to a disease state.

Once training is complete, the main task of prediction can begin (e.g., predicting gene function [80], pharmacogenomics outcome [19], or variant detection [81]) using supervised learning for genotype–phenotype mapping (e.g., SNV variations with phenotypes) or to apply the learned models (if the data are labeled) to novel datasets. This task is particularly challenging in cardiology because many cardiovascular conditions are heterogeneous and not well-defined. For instance, heart failure classification is largely based on ejection fraction (HFrEF, HFpEF, and HFmrEF) but ejection fraction assessment can be affected by angle-dependent and interoperability issues. Furthermore, current cardiovascular genetic datasets restrict access and contain a homogeneous population. The Million Veteran Program, one of the largest genetic and CVD datasets assembled, limits access to its data, and most other major public CVD genetic data sets are largely based on UK Biobank samples, which are from a largely Caucasian British population (94% Caucasian).

Once training is complete, the creation and analysis of artificial nucleotide sequences, such as the creation of artificial human genomes [82] or artificial enhancers ("synthetic DNA"), using approaches such as generative adversarial networks (GANs), can be considered [83]. GANs are DL models that include two primary components: a generator and a discriminator. Generated DNA sequences are used as inputs for the discriminator to analyze if the model has generated a convincingly real biological sequence. This feedback is used to iteratively train the generator model to produce artificial sequences with increasingly realistic properties. For example, a recent study used a type of GAN (an auxiliary classifier generative adversarial network) to generate synthetic participants that closely resembled real participants from the SPRINT trial (Systolic Blood Pressure Trial) to facilitate exploratory analyses [84].

Using these techniques, DL has been successfully applied within genomics in several major projects, including DeepSEA (a DL-based sequence analyzer that can predict the epigenetic state in multiple cell types), and a subsequent DragoNN primer online training in academic institutions globally [27,85]. To date, convolutional neural networks (CNNs), recurrent neural networks (RNNs), autoencoders, and GANs have been the primary DL techniques used in genomics (Table 2). These approaches have been implemented for several tasks, including functional assessments of variants [28], AI-guided multiethnic polygenic risk score (PRS) generation [86]. and variant calling optimization [87]. Interestingly, the number of layers within DL architectures used in genomics has generally been far less than those used for image recognition, and, thus far, typically consist of only a few layers [27,79,88] with many hundreds to thousands of parameters [89].

Given the broad variety of potential genomic data types (e.g., genetic variants, DNA methylation, gene expression, miRNA expression data, transcription factor binding, chromatin state, etc.), there is a growing trend to use DL to perform multi-faceted biological data integration. This strategy could be used to classify new CVD genotype–phenotype relationships, which could then result in the identification of novel therapeutic targets (e.g., new therapies based on genetic loci and left ventricular mass to volume ratio from cardiovascular magnetic resonance imaging, left ventricular end-diastolic pressure from echocardiography, or novel strain patterns from strain imaging) [25,90]. Using DL-guided WES in clinical practice to bridge the phenotype–genotype gap also shows promising utility [91]. DL could be used to reduce sequencing biases known to affect WES data analysis (e.g., coverage biases [92] or GC content bias [93,94]). Figure 2 demonstrates a typical DL model used in genomics. We have previously described several major DL libraries [65,66] and DL guidelines in cardiovascular medicine [66]. In addition, new open-source genomics libraries, such as Nucleus, which builds on top of TensorFlow, may be used for future DL in genomic research. At least one clinical trial (NCT03877614) is underway using DL in genetics and CVDs, including CAD, HFrEF, HCM, atrial fibrillation, pulmonary hyperten-

sion, and Fabry's disease, compared to a healthy to low risk control group (atherosclerotic cardiovascular disease score <10%). In the future, DL could potentially be used to predict the future development of many CVDs using genomic findings as inputs.

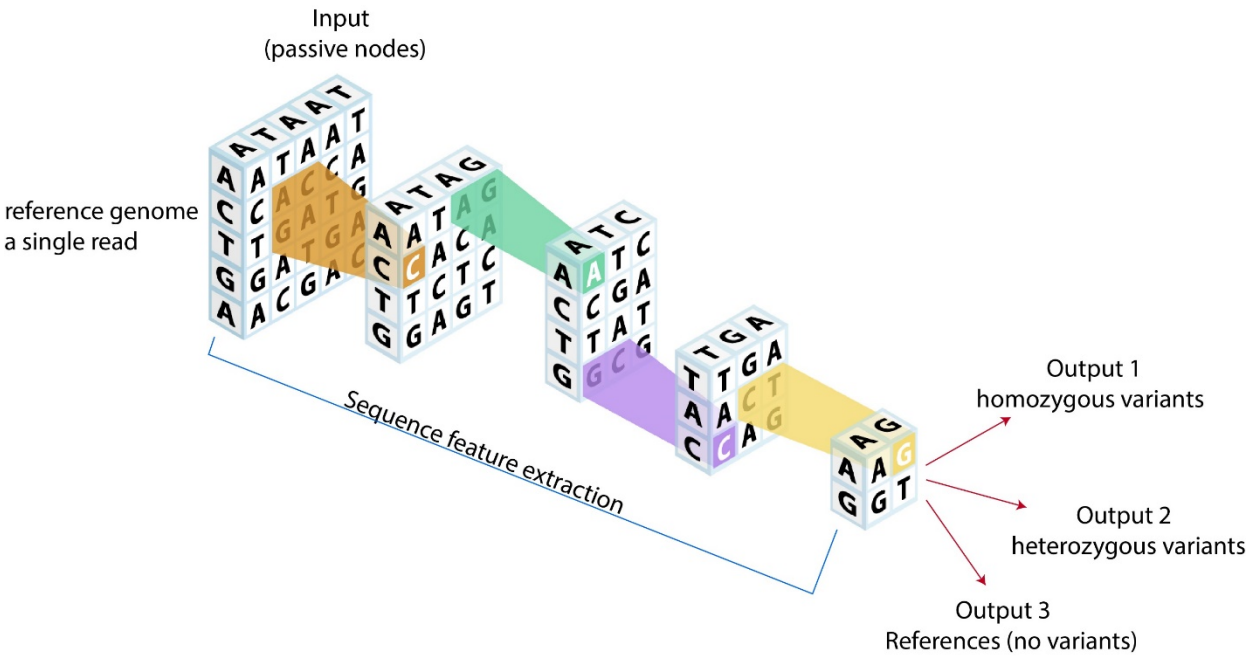

Figure 2. Potential analytic models for cardiovascular genomics. Reference genome or a single read could be fed into neural network models using convolutional genetic coding based on genetic structures. After neural network processing, outputs can be categorized into homozygous variants, heterozygous variants, and references (no variants), which could ultimately provide novel clinical genetic insights.

4.2. Natural Language Processing

NLP is a set of computational methods that are able to understand language by analyzing its syntax and semantics. Major applications of NLP within medicine include analyzing progress notes [95], identifying critical illness [96], de-identifying patient records [97], reducing human workload of literature reviews [98], and predicting readmission from discharge summaries [99]. Within genomics, NLP has been used for gene recognition or normalization [100] and identifying gene–disease associations in heart failure [101]. Interestingly, NLP has also been used to predict genes for CAD [102,103], while other techniques rely on a combination of ML, DL, and NLP to predict gene alterations [63,64].

Advancements in NLP may incorporate clinical guidelines to automatically generate appropriate recommendations for CVD prevention in a patient's discharge summary. For example, based on the current literature and the level I evidence available, NLP could recommend the most appropriate anticoagulation treatment for patients with a left ventricular thrombus. Another example would be NLP of admission notes to determine possible necessity for genetic screening. However, NLP must first understand the relevant clinical semantics (e.g., analyzing all literature in PubMed and clinical notes in EHRs) in order to provide appropriate clinical recommendations. Although ML algorithms are more often used for predictive analyses, ML algorithms are also able to perform NLP tasks using ML-based NLP models [104]. For example, the implementation of NLP-DL to review genes related to clinically actionable mutations is feasible [105]. Advanced AI techniques, such as deep reinforcement techniques, can be a powerful approach for NLP tasks for heterogeneous CVDs and genomics [106]. Deep reinforcement-based NLP models could, for example, potentially enhance traditional algorithms to identify mutations by working to rule out read errors.

5. Current Limitations in Genomics and Potential Solutions with AI

Below we describe the limitations in current genomic research and discuss how AI implementation can address these limitations and advance the field (Figure 3).

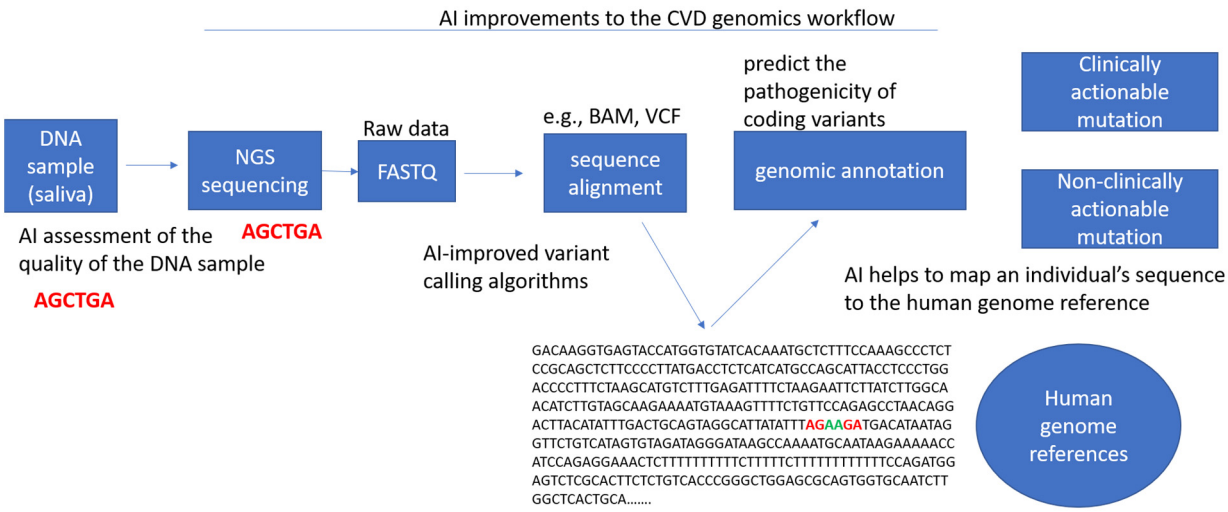

Figure 3. Potential artificial intelligence improvements to the workflow in cardiovascular genomics. This includes the assessment of the quality of genetic samples obtained (e.g., DNA, RNA, exome), the improvement of informatics pipelines for variant calling, the translation of clinical guidelines for variant interpretation, the transformation of genetic files (e.g., VCF to BAM, VCF to PED), the prediction of variant pathogenicity, the mapping of an individual's sequence to genome references, and the identification of any clinically actionable mutations.

5.1. Lack of Clinical and Technical Guidelines for Cardiovascular Genetics

Currently in clinical cardiovascular genetics, the guidelines do not specify which genes should be tested or how to validate the results. For example, the 2019 HRS Expert Consensus Statement on Evaluation, Risk Stratification, and Management of Arrhythmogenic Cardiomyopathy did not define how genetic testing should be validated or carried out in ARVC and other arrhythmogenic cardiomyopathies [107]. Similarly, the 2020 and 2021 scientific statements from the AHA on Genetic Testing for Heritable Cardiovascular Diseases in adult and pediatric patients did not specify how genetic testing should be validated or carried out in heritable cardiovascular diseases [7,8].

At a more rudimentary level, the Clinical Laboratory Improvement Amendment (CLIA) and the College of American Pathologists (CAP) have left many inconsistencies and regulatory gaps in their guidance for wet and dry labs [108], resulting in heterogeneous variant reporting. Moreover, CAP/CLIA regulations only require that validation is performed in the production environment, which may lead to unexpected errors in the production phase. Bioinformatics pipelines should be validated and tested for how precisely and sensitively variants are called in wet labs. Technical variability in the QC process, such as consistency of sequencing [109], QC standardization [110], and DNA quality [111,112], has been highly problematic; however, with current technologies, the accuracy of SNV is generally very robust (particularly if 30x or greater sequencing coverage is available). However, despite the advances in SNV analysis, structural variation calling continues to be highly variable and problematic. Automated QC systems using AI may decrease these issues by recognizing outliers and inconsistent data, identifying structural variations or small mutations from random errors and complex variants from long-read sequencing [113], and improving missing genotypes imputation [114]. While few studies have developed NLP-guided bioinformatics pipelines [115–117], ML-based pipelines have been more widely reported [118,119]. Unfortunately, most of these ML-based pipelines are not well validated across different databases, which may introduce population-specific biases. Given the variety of DL architectures (e.g., convolutional networks or encoders), DL

models may be able to target and improve existing bioinformatics pipelines and variant classifications [120,121].

Another major barrier to current cardiovascular genetic research is the lack of professional recommendations for the clinical integration of genomics. Several clinical research projects using different genomics databases (e.g., UK Biobank [67], MESA [122], and ARIC [123]) have demonstrated accurate ML model discrimination and calibration (e.g., Brier score) for CVD risk prediction using genetics, but there are as yet no specific guidelines for genetic testing in clinical practice or regulatory guidance for direct-to-consumer products. This has also led to a lack of reimbursements for testing and a lack of incentives for routine testing. While most direct-to-consumer genetic testing companies are CAP/CLIA-certified, the lack of transparency and validation of these company's tests and results poses a challenge for effective integration into clinical practice. Although the 2019 AHA Scientific Statement initiated the AHA Cardiovascular Genome–Phenome initiative, the guidelines for genomic processing or genetic testing in clinical practice remain poorly defined [124]. Through analyzing genes related to particular heritable conditions and improving prediction models, AI has the potential to facilitate efficient testing of family members and implement precise medicine-based care rather than the current standard practice of diagnosis and treatment based on broad population guidelines.

5.2. Variant Calling, Reporting, and Interpretation

Variant calling is used to identify the differences between an individual genome and a reference genome. Despite CLIA approval, there are no guidelines for approval of informatics pipelines for variant calling. There are several variant-related tasks (e.g., read alignment, variant calling, reporting, and interpretation) currently used in genomics screening, the identification of probands, and cascade testing in CVD where AI could be applied. The discrepancies in variant calling between labs, largely because of the lack of clear guidelines, are magnified when undertaking the task of distinguishing true genetic variants from spurious differences introduced by sequencing errors, alignments errors, and other technical artifacts. Other limitations of variant calling include a lack of consensus between variant calling pipelines when analyzing the same data [125], variable accuracies of variant calling algorithms when using different AI technologies, and comparison sequencing of only a limited gene panel. Importantly, AI-driven software, such as DeepVariant, Clairvoyante [38], and Skyhawk [39], have already been used to automatically recognize and prioritize variants with substantially improved accuracy when compared to more traditional statistical models. For example, Google's DeepVariant uses image recognition techniques and pre-trained models (e.g., inception-v3, variants of CNN model [87]) to pre-process inputs, make inferences, call variants, and then output variant calling format (VCF) files with the variant information. This represents a potential AI solution to the current inconsistencies in variant calling.

Once variants are identified, AI can also help with the interpretation and impact of these variants in clinical practice [126]. For example, SpliceAI [44], DeepBind [127], and DeepSEA [27] can predict the outcomes from different variants with respect to alternative splicing, transcription factor binding, or epigenetic changes, respectively. Additionally, NLP tools have been used in both direct and indirect genetics extraction. For example, BCC-NER [128], and BioNLP [129] have been used for automated extraction of gene and genetic variants or the identification of targeted genes from published literature (Table 2). In CVD specifically, indirect extraction using a family history of sudden cardiac death or HCM using NLP holds promise for better and more efficient management of HCM patients [130]. Most importantly, emerging hybrid models, such as a combination of DL-NLP and deep reinforcement learning, capsule learning, or meta-learning, may overcome the limited knowledge that is currently available to support genomic research. However, a validation of those algorithms is needed first. AI can also be used to collect all clinically relevant information from Medline, the AHA precision medicine platform, or genomic datasets using pre-trained models. However, before that can become reality, a trial of

different pre-trained architectures for improved accuracy in variant calling within noisy and imbalanced sequencing data will be needed.

Variant reporting and interpretation are challenging tasks in clinical cardiovascular practice because, like for variant calling, there are currently few published guidelines in cardiovascular genetics [131]. There are some specific guidelines available, but they only apply to specific genes (e.g., myh7) [132] and are, therefore, not useful in the majority of situations. It is not unreasonable to expect greater guidance in variant interpretation for cardiovascular clinical practice, as other organizations have already released guidelines. For example, the 2015 updated standards and guidelines from the American College of Medical Genetics and Genomics (ACMG) and the Association for Molecular Pathology (AMP) recommended 28 criteria for the clinical interpretation of sequence variants with respect to human diseases. The AHA and ACC should follow this example and develop a statement for genetic testing and a variant interpretation strategy in cardiovascular genetics.

5.3. Combining Genomics with Other Clinical Data Types

Cardiovascular genetics is challenging because both the clinical variables associated with CVDs and the genomics data are heterogeneous and often involve complex interactions between a patient's genetics and environmental factors. This challenge is largely why applying AI to these multiple types of data is a very promising research direction, and may be especially useful in classifying genome-phenome relationships in CVD using EHRs [133]. For example, combining genomic data describing different septal morphologies of HCM [134,135] with clinical information from echocardiography and angiography could help personalize therapy for individual patients (e.g., deciding if a particular HCM patient needs an ICD). Echo-guided genetic testing or genetic-guided PCI [136] and DAPT duration (e.g., high- vs. low-risk bleeding loci) would also be useful applications of this technology. Another potential application worth researching is the diagnosis of diastolic dysfunction using a combination of echo parameters (e.g., LAVI, E/A ratio, annular e' velocity, and peak TR velocity) and genetic predispositions since normal diastolic function changes with age [70,71,137]. Precision statin therapy is another potential application for the integration of multiple data types by AI. For instance, in a young female without traditional atherosclerotic risk factors, a combination of genetic testing (e.g., Lp (a), apo C genes) and cardiac imaging (e.g., coronary CT) may reveal a clinical need for preventative statin therapy, which would otherwise never be considered.

The technical aspects of integrating clinical and genomic data rely on data transformations [138] which convert data into a common vector-matrix format prior to processing using a kernel function. However, this is not the only way to harmonize different data types and modalities. In cardiac amyloidosis, for example, data transformation can be used on echocardiography parameters, immunofixation electrophoresis, and *MAGE CT* genes, and then an ANN can identify the suitability of gene-targeted therapy for patients with equivocal biopsy results. Future research in gene editing therapies for cardiac amyloidosis could be heavily aided and accelerated by AI. In another example, Ross et al. used data transformations to combine 10 SNVs, clinical variables, and laboratory imaging data to predict mortality in peripheral artery diseases using elastic net regression and random forest models [139].

AI models can also combine genomic data with data drawn from the EHR and combine them into a unified matrix for clinical analysis. While this strategy is not yet routinely performed, several studies have shown its power and promise. EHR-based phenotyping algorithms have been able to identify familial hypercholesterolemia [140], significant carotid stenosis [141], and the relative prevalence of CAD among different cohorts [142]. Recently, IBM Watson (an automated NLP based algorithm), the Broad Institute of MIT, and Harvard have partnered with the aim of developing AI-based PRS models using population- and hospital-based biobank data, genomic information, and EHRs to identify patients at serious risk for CVD. In addition, ML models have been applied to integrate genetics, cardiac imaging [143], biobank data, and clinical information from EHRs [144]

for high-throughput mapping of genotype–phenotype associations to predict diabetes, titin-truncating variants related to DCM [145], and CAD [146]. Another ML study using the Framingham Heart Study cohort used a combination of clinical and genotype data (56 SNPs) for predictive modeling of advanced coronary calcium [75]. By using these examples as a foundation, more advanced studies can be performed with even greater amounts of multidimensional data.

Ultimately, the pipelines of clinical data convergence lie in the ability of AI to unlock multidimensional complex interactions (e.g., gene–environment or gene–behavioral interactions) beyond simply studying gene–gene interactions or host–gut microbiome interactions [147]. For example, air pollutant exposure could lead to changes in DNA methylation and gene silencing without altering the actual DNA sequence [148]. AI could potentially identify relationships between air pollution or zip codes and genes related to detoxification (GSTM1 and GSTT1) or iron processing (HFE), and then generate individualized healthcare recommendations [149]. The combined analysis of these multi-omics data using AI has the potential to provide an improved overall picture of the characteristics of heterogeneous CVDs and, therefore, aid our understanding of their molecular underpinnings.

5.4. Lack of Population Specific Analysis Tools

Across all fields of medicine and research, population-specific analysis tools and databases that can detect population-specific risk factors are urgently needed. Unfortunately, in most cases, including in CV research, significant disparities in research for different ethnicities remain. The pooled cohort equations (PCE) is the cornerstone for atherosclerotic cardiovascular disease (ASCVD) risk stratification and statin treatment decisions [14]. However, the PCE computation mainly focuses on the Caucasian population and overestimates ASCVD risk in Asian and Hispanic populations. Although PCE computations exclude genetic components, the ethnicity disparity is not limited to cardiovascular genetic research [150]. While genomic research in Asian ancestry and African ancestry has increased in recent times [151,152], more than 90% of genomic research has been conducted in patients of mainly European ancestry [153,154]. Furthermore, while most GWAS attempts can control bias of population stratification, fully correcting for population stratification can be challenging and the lack of ethnic diversity included can affect the analysis of gene–environment interactions [155]. Therefore, a major challenge for applying AI more widely is the lack of publicly available non-European genetic databases. In addition, PRS is an emerging technique for assigning genetic risk to individual outcomes that outperforms traditional risk scores [156], but the performance of translating PRS from European ancestry to different ethnicities is largely unknown and not validated [157]. The AI technique of transfer learning could potentially be used to bridge this gap.

A recent study showed that polygenic risk powerfully modifies the risk conferred by monogenic risk variants [158]. However, incorporating these loci into clinical practice is not well established and PRS has limitations in complex disease predictions because of its dependency on linear regression, a lack of phenotype differentiation [159], and a variation in the numbers of SNVs in PRS [160]. A recent quantitative experiment demonstrated some improvement in prediction accuracy using multi-ethnic PRS (mixing training data from Europeans, South Asians, and Africans) [86]. Zhao and Zou investigated PRS, both empirically and theoretically, and found that accuracy can vary dramatically depending on how sparse true genetic signals are [161]. Therefore, an important future research direction is to use AI to explore non-linear PRS relationships, handle interactive high-dimensional data, and randomize selection of SNVs and genetic signals. AI could also be applied to multi-ethnic cohorts to elucidate the role that PRS and ML models, such as GANs, and could potentially play a role in creating a multi-ethnicity PRS. Despite the challenges, some steps have been taken to increase the diversity of WES and WGS samples with efforts such as the Trans-Omics for Precision Medicine (TOPMed) Program, the Million Veteran Program, the Atherosclerosis Risk in Communities (ARIC) Study, the MultiEthnic Study of

Atherosclerosis (MESA) [162], and the Multiethnic Variation in Recovery: Role of Gender on Outcomes of Young AMI Patients (VIRGO) Study [163]. Nevertheless, the "unknown unknowns" of modifier genes or polygenic influences in CVDs remains to be explored. Although ML models in the PRS field remain in a developmental phase and have not yet been clinically tested in cardiovascular genetics, AI is poised to overcome current challenges by integrating ethnicity into genomic research.

6. Current Limitations in AI Cardiovascular Genetics

Despite steadfast advances, implementing AI in cardiovascular genomics still faces several challenges, including generalizability of results, the required construction of large genomic datasets, and limited computing power. Ultimately, the largest barrier remains the ability of clinicians to implement findings from AI studies.

The first challenge that plagues AI is overfitting an algorithm to a dataset that may adversely affect the generalizability of the results. Generalizability can be partially assessed by evaluating the overfitting of a new dataset. For instance, the results of applying DL models to diabetic retinopathy could not be replicated in different datasets [164,165], and AI methods lack validation data when applied to disease-associated non-coding variants [166,167]. Moreover, many of these mutations have a very small effect on disease risk, even when their combined effects can be clinically relevant. In addition to the ethnic bias discussed previously, AI methods, such as DL, can inadvertently integrate other forms of bias attributed to the training dataset (e.g., bias in word embeddings or variability in extraction algorithms) [168], which represents another challenge in implementing and generalizing results from DL [169,170].

Despite the promise of various AI methods, genomic datasets themselves have built-in limitations: the costs incurred remains a large barrier to performing thorough studies; heterogeneous genetic conditions, such as dilated cardiomyopathy, lack known outputs; and the rarity of specific conditions results in unbalanced case-control studies. These are important limitations when considering the construction of a genomics dataset. Currently, there is not a consensus or indication for genetic testing across several entities within CVD. For patients who undergo genetic testing, the sample can undergo a variety of sequencing techniques that differ between vendors, affecting the quality of the resulting data and confounding interpretation. Moreover, strong evidence of treatment data in cardiovascular genetics is lacking. Premature CAD, for example, with a known or novel actionable mutation may still be treated the same as CAD in older adults by using a high intensity statin, ezetimibe, and/or PCSK9 inhibitor. Identifying a confounder from the CVD-causing relevant environmental factors themselves in genomic data is also challenging, and current DL algorithms have difficulty identifying them as well. Although some DL algorithms can be used in confounder filtering, they cannot be used effectively to control population stratification in GWAS [171,172]. Relatedly, evaluating simulated data or partitioning existing datasets into smaller groups to try and limit confounders may not capture the complexity of genetic data sets and may generate substantially different results.

An equally important barrier to integrating AI study results into clinical practice is the fact that physicians currently lack the necessary access as well as education and training to interpret results from AI studies on genomic data [173,174]. To facilitate clinical adoption, AI can fill the gap in knowledge in clinical practice with automated analysis to detect clinically actionable mutations. However, there is a figurative territorial embargo which limits medical genetics to trained specialists because of the complexity of handling genomic data, rather than a democratization and availability of this technology to all clinicians and patients. Emerging technology, such as homomorphic encryption or blockchains, which can provide an immediate and transparent exchange of encrypted data simultaneously to multiple parties, may be able to fill this gap by at least ensuring data security in handling genomic data. However, there is no process for lifelong interrogation of such data, nor is there specialty infrastructure or funding processes capable of handling that. Most importantly, the main challenge is "trust" in data stewardship. AI has the promise to do

automated analyses, but there is no agreement over the format, interpretation, reliability, or reproducibility of the results.

Despite tremendous recent advances, current quantum or cognitive computing application is still in its infancy. For example, the IBM Watson system has been tasked with identifying and interpreting clinically actionable mutations [175], but still heavily relies on human supervision. Watson's limitations are likely due to difficulty in integrating with EHRs, too many reported options, and a lack of clinical trials [104,176]. Most importantly, reports from Watson for genomics are based on single-centered studies with weak evidence; they are not based on guidelines and may or may not be beneficial in certain populations or conditions [177]. Notably, the general lack of software infrastructure for genomics-oriented research (e.g., quantum computing, cloud services, supercomputers, or cognitive computing workstations) in cardiology and genetics departments limits the power of AI, even among experts with current access to the data.

Finally, the quality of genomic data between direct-to-consumer companies and clinical or academic institutions may affect the availability and accuracy of "raw data" for AI to analyze. Genotyping data from direct-to-consumer companies, even those that are CLIA certified, contain errors and potentially high false-positive rates (up to 40%) [178]. For example, there is inconsistent labelling of *COL3A1* and *COL5A1* mutations (known to be associated with Ehlers–Danlos syndrome and SCAD) between laboratories [178]. Therefore, standard measures for correlating and combining data from direct-to-consumer and data from clinical or academic institutions are urgently needed. Beyond the technical issues of how variants are reported, there are also substantial privacy concerns involved when sharing genetic data with a direct-to-consumer company. As a minimum, advanced encryption is certainly required to maintain patient privacy.

7. Conclusions

The major barriers to AI-aided genomics reaching widespread clinical practice are fundamentally related to the relative newness of the field itself. Namely, a lack of deep understanding of AI by clinicians, a lack of standardized bioinformatics pipelines, a lack of transparency in AI models, difficulties interpreting the limitations of DL (compared to traditional statistical inferences), problems in structural variations and other complex variant types, unsatisfactory predictive performances in real world genomic problems, a lack of good phenotype data, and poor genomic data quality. In addition, the use of AI-aided genomics research in CVD is also challenged by the heterogeneity of genetic and environmental risk factors. However, with time and further research, these barriers will be overcome, and combinations of AI models will lead to increasingly sophisticated interpretations that may eventually enhance clinical decision making in cardiovascular clinical genetics. Lifestyle data from wearable technology combined with clinical data from EHR and genetic data could tailor treatment towards personalized medicine, ideally identifying CVD at an early stage when it can be more efficiently treated and create a larger improvement in quality of life. In the era of big data, AI-guided studies will translate into increasingly complex genomic datasets, resulting in more sophisticated clinical treatments and improvements in precision medicine.

Author Contributions: Conceptualization, C.K.; writing—original draft preparation, C.K.; writing—review and editing, K.W.J., E.C., S.K., E.V., M.M., Z.W., B.S.G., C.I.A., M.C.S. and W.H.W.T.; supervision, C.I.A., M.C.S. and W.H.W.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Acknowledgments: The authors would like to thank Ishan Kamat and Jennifer Wilcox for their suggestions and comments on this article.

Conflicts of Interest: Krittanawong discloses the following relationships–Member of the American College of Cardiology Solution Set Oversight Committee, the American Heart Association Committee of the Council on Genomic and Precision Medicine, and the American College of Cardiology/American Heart Association (ACC/AHA) Task Force on Performance Measures, The Lancet Digital Health (Advisory Board), European Heart Journal Digital Health (Editorial board), Journal of the American Heart Association (Editorial board), JACC: Asia (Section Editor), The Journal of Scientific Innovation in Medicine (Associate Editor), and Frontiers in Cardiovascular Medicine (Associate Editor). Other authors have no disclosure.

Abbreviations

ARVC	arrhythmogenic right ventricular cardiomyopathy
AI	artificial intelligence
ANN	artificial neural network
CVD	cardiovascular disease
CLIA	Clinical Laboratory Improvement Amendment
CAP	College of American Pathologists
CNN	convolutional neural network
CNV	copy number variation
CAD	coronary artery disease
DL	deep learning
DCM	dilated cardiomyopathy
DAPT	dual antiplatelet therapy
DNN	deep neural network
EHR	electronic health record
FDA	Food and Drug Administration
GAN	generative adversarial network
GWAS	genome-wide association studies
HFmrEF	heart failure with midrange ejection fraction
HFpEF	heart failure with preserved ejection fraction
HFrEF	heart failure with reduced ejection fraction
HCM	hypertrophic cardiomyopathy
ICD	implantable cardioverter defibrillator
LAVI	left atrial volume index
ML	machine learning
NLP	natural language processing
NGS	next generation sequencing
PCI	percutaneous coronary intervention
PRS	polygenic risk score
QC	quality control
RNN	recurrent neural network
SNP	single nucleotide polymorphisms
SNV	single nucleotide variant
SCAD	spontaneous coronary artery dissection
TR	tricuspid regurgitation
UK	United Kingdom
USA	United States of America
VCF	Variant Calling Format
WES	whole exome sequencing
WGS	whole genome sequencing

References

1. Bertolini, S.; Pisciotta, L.; Di Scala, L.; Langheim, S.; Bellocchio, A.; Masturzo, P.; Cantafora, A.; Martini, S.; Averna, M.; Pes, G.M.; et al. Genetic polymorphisms affecting the phenotypic expression of familial hypercholesterolemia. *Atherosclerosis* **2004**, *174*, 57–65. [CrossRef] [PubMed]
2. Krittanawong, C.; Khawaja, M.; Rosenson, R.S.; Amos, C.I.; Nambi, V.; Lavie, C.J.; Virani, S.S. Association of PCSK9 Variants with the Risk of Atherosclerotic Cardiovascular Disease and Variable Responses to PCSK9 Inhibitor Therapy. *Curr. Probl. Cardiol.* **2021**, 101043. [CrossRef] [PubMed]

3. Campuzano, O.; Beltrán-Álvarez, P.; Iglesias, A.; Scornik, F.; Pérez, G.; Brugada, R. Genetics and cardiac channelopathies. *Genet. Med.* **2010**, *12*, 260–267. [CrossRef] [PubMed]
4. Bleumink, G.S.; Schut, A.F.; Sturkenboom, M.C.; Deckers, J.W.; van Duijn, C.M.; Stricker, B.H. Genetic polymorphisms and heart failure. *Genet. Med.* **2004**, *6*, 465–474. [CrossRef] [PubMed]
5. Vecoli, C.; Borghini, A.; Turchi, S.; Mercuri, A.; Andreassi, M.G. Genetic polymorphisms of miRNA machinery genes in bicuspid aortic valve and associated aortopathy. *Pers. Med.* **2021**, *18*, 21–29. [CrossRef]
6. Girdauskas, E.; Geist, L.; Disha, K.; Kazakbaev, I.; Groß, T.; Schulz, S.; Ungelenk, M.; Kuntze, T.; Reichenspurner, H.; Kurth, I. Genetic abnormalities in bicuspid aortic valve root phenotype: Preliminary results†. *Eur. J. Cardio-Thorac. Surg.* **2017**, *52*, 156–162. [CrossRef] [PubMed]
7. Musunuru, K.; Hershberger, R.E.; Day, S.M.; Klinedinst, N.J.; Landstrom, A.P.; Parikh, V.N.; Prakash, S.; Semsarian, C.; Sturm, A.C.; American Heart Association Council on Genomic and Precision Medicine; et al. Genetic Testing for Inherited Cardiovascular Diseases: A Scientific Statement From the American Heart Association. *Circ. Genom. Precis. Med.* **2020**, *13*, e000067. [CrossRef] [PubMed]
8. Landstrom, A.P.; Kim, J.J.; Gelb, B.D.; Helm, B.M.; Kannankeril, P.J.; Semsarian, C.; Sturm, A.C.; Tristani-Firouzi, M.; Ware, S.M.; on behalf of the American Heart Association Council on Genomic and Precision Medicine; et al. Genetic Testing for Heritable Cardiovascular Diseases in Pediatric Patients: A Scientific Statement From the American Heart Association. *Circ. Genom. Precis. Med.* **2021**, *14*, e000086. [CrossRef]
9. Harper, A.R.; Goel, A.; Grace, C.; Thomson, K.L.; Petersen, S.E.; Xu, X.; Waring, A.; Ormondroyd, E.; Kramer, C.M.; Ho, C.Y.; et al. Common genetic variants and modifiable risk factors underpin hypertrophic cardiomyopathy susceptibility and expressivity. *Nat. Genet.* **2021**, *53*, 135–142. [CrossRef] [PubMed]
10. Gillmore, J.D.; Gane, E.; Taubel, J.; Kao, J.; Fontana, M.; Maitland, M.L.; Seitzer, J.; O'Connell, D.; Walsh, K.R.; Wood, K.; et al. CRISPR-Cas9 In Vivo Gene Editing for Transthyretin Amyloidosis. *N. Engl. J. Med.* **2021**, *385*, 493–502. [CrossRef]
11. Krittanawong, C.; Zhang, H.; Wang, Z.; Aydar, M.; Kitai, T. Artificial Intelligence in Precision Cardiovascular Medicine. *J. Am. Coll. Cardiol.* **2017**, *69*, 2657–2664. [CrossRef] [PubMed]
12. Murdock, D.R.; Venner, E.; Muzny, D.M.; Metcalf, G.A.; Murugan, M.; Hadley, T.D.; Chander, V.; de Vries, P.S.; Jia, X.; Hussain, A.; et al. Genetic testing in ambulatory cardiology clinics reveals high rate of findings with clinical management implications. *Genet. Med.* **2021**, *23*, 2404–2414. [CrossRef]
13. Ommen, S.R.; Mital, S.; Burke, M.A.; Day, S.M.; Deswal, A.; Elliott, P.; Evanovich, L.L.; Hung, J.; Joglar, J.A.; Kantor, P.; et al. 2020 AHA/ACC Guideline for the Diagnosis and Treatment of Patients With Hypertrophic Cardiomyopathy. *Circulation* **2020**, *142*, e558–e631. [PubMed]
14. Grundy, S.M.; Stone, N.; Bailey, A.L.; Beam, C.; Birtcher, K.K.; Blumenthal, R.S.; Braun, L.T.; De Ferranti, S.; Faiella-Tommasino, J.; Forman, D.E.; et al. 2018 AHA/ACC/AACVPR/AAPA/ABC/ACPM/ADA/AGS/APhA/ASPC/NLA/PCNA Guideline on the Management of Blood Cholesterol: Executive Summary: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *J. Am. Coll. Cardiol.* **2019**, *73*, 3168–3209. [CrossRef] [PubMed]
15. Brugada, J.; Campuzano, O.; Arbelo, E.; Sarquella-Brugada, G.; Brugada, R. Present Status of Brugada Syndrome. *J. Am. Coll. Cardiol.* **2018**, *72*, 1046–1059. [CrossRef] [PubMed]
16. Al-Khatib, S.M.; Stevenson, W.G.; Ackerman, M.J.; Bryant, W.J.; Callans, D.J.; Curtis, A.B.; Deal, B.J.; Dickfeld, T.; Field, M.E.; Fonarow, G.C.; et al. 2017 AHA/ACC/HRS Guideline for Management of Patients With Ventricular Arrhythmias and the Prevention of Sudden Cardiac Death. *Circulation* **2018**, *138*, e272–e391. [PubMed]
17. McKusick, V.A.; Ruddle, F.H. Toward a complete map of the human genome. *Genomics* **1987**, *1*, 103–106. [CrossRef]
18. Novelli, G.; Predazzi, I.M.; Mango, R.; Romeo, F.; Mehta, J.L. Role of genomics in cardiovascular medicine. *World J. Cardiol.* **2010**, *2*, 428–436. [CrossRef] [PubMed]
19. Tang, J.; Liu, R.; Zhang, Y.-L.; Liu, M.-Z.; Hu, Y.-F.; Shao, M.-J.; Zhu, L.-J.; Xin, H.-W.; Feng, G.-W.; Shang, W.-J.; et al. Application of Machine-Learning Models to Predict Tacrolimus Stable Dose in Renal Transplant Recipients. *Sci. Rep.* **2017**, *7*, 42192. [CrossRef] [PubMed]
20. Belkadi, A.; Bolze, A.; Itan, Y.; Cobat, A.; Vincent, Q.B.; Antipenko, A.; Shang, L.; Boisson, B.; Casanova, J.-L.; Abel, L. Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 5473–5478. [CrossRef]
21. Boyle, E.A.; Li, Y.I.; Pritchard, J.K. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* **2017**, *169*, 1177–1186. [CrossRef] [PubMed]
22. Sanger, F.; Nicklen, S.; Coulson, A.R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **1977**, *74*, 5463–5467. [CrossRef]
23. Lincoln, S.E.; Truty, R.; Lin, C.-F.; Zook, J.M.; Paul, J.; Ramey, V.H.; Salit, M.; Rehm, H.L.; Nussbaum, R.L.; Lebo, M.S. A Rigorous Interlaboratory Examination of the Need to Confirm Next-Generation Sequencing–Detected Variants with an Orthogonal Method in Clinical Genetic Testing. *J. Mol. Diagn.* **2019**, *21*, 318–329. [CrossRef] [PubMed]
24. Hume, S.; Nelson, T.N.; Speevak, M.; McCready, E.; Agatep, R.; Feilotter, H.; Parboosingh, J.; Stavropoulos, D.J.; Taylor, S.; Stockley, T.L. CCMG practice guideline: Laboratory guidelines for next-generation sequencing. *J. Med. Genet.* **2019**, *56*, 792–800. [CrossRef] [PubMed]

25. Aung, N.; Vargas, J.D.; Yang, C.; Cabrera, C.P.; Warren, H.R.; Fung, K.; Tzanis, E.; Barnes, M.R.; Rotter, J.I.; Taylor, K.D.; et al. Genome-Wide Analysis of Left Ventricular Image-Derived Phenotypes Identifies Fourteen Loci Associated With Cardiac Morphogenesis and Heart Failure Development. *Circulation* **2019**, *140*, 1318–1330. [CrossRef] [PubMed]
26. Amarbayasgalan, T.; Park, K.H.; Lee, J.Y.; Ryu, K.H. Reconstruction error based deep neural networks for coronary heart disease risk prediction. *PLoS ONE* **2019**, *14*, e0225991. [CrossRef] [PubMed]
27. Zhou, J.; Troyanskaya, O.G. Predicting effects of noncoding variants with deep learning–based sequence model. *Nat. Methods* **2015**, *12*, 931–934. [CrossRef]
28. Jaganathan, K.; Panagiotopoulou, S.K.; McRae, J.F.; Darbandi, S.F.; Knowles, D.; Li, Y.I.; Kosmicki, J.A.; Arbelaez, J.; Cui, W.; Schwartz, G.B.; et al. Predicting Splicing from Primary Sequence with Deep Learning. *Cell* **2019**, *176*, 535–548.e24. [CrossRef] [PubMed]
29. Rossi, A.; Voigtlaender, M.; Janjetovic, S.; Thiele, B.; Alawi, M.; März, M.; Brandt, A.; Hansen, T.; Radloff, J.; Schön, G.; et al. Mutational landscape reflects the biological continuum of plasma cell dyscrasias. *Blood Cancer J.* **2017**, *7*, e537. [CrossRef] [PubMed]
30. Kufova, Z.C.; Sevcikova, T.; Januska, J.; Vojta, P.; Boday, A.; Vanickova, P.; Filipova, J.; Growkova, K.; Jelinek, T.; Hajduch, M.; et al. Newly designed 11-gene panel reveals first case of hereditary amyloidosis captured by massive parallel sequencing. *J. Clin. Pathol.* **2018**, *71*, 687–694. [CrossRef]
31. Caravagna, G.; Giarratano, Y.; Ramazzotti, D.; Tomlinson, I.; Graham, T.A.; Sanguinetti, G.; Sottoriva, A. Detecting repeated cancer evolution from multi-region tumor sequencing data. *Nat. Methods* **2018**, *15*, 707–714. [CrossRef] [PubMed]
32. McKinney, S.M.; Sieniek, M.; Godbole, V.; Godwin, J.; Antropova, N.; Ashrafian, H.; Back, T.; Chesus, M.; Corrado, G.S.; Darzi, A.; et al. International evaluation of an AI system for breast cancer screening. *Nature* **2020**, *577*, 89–94. [CrossRef] [PubMed]
33. Krittanawong, C.; Johnson, K.W.; Hershman, S.G.; Tang, W. Big data, artificial intelligence, and cardiovascular precision medicine. *Expert Rev. Precis. Med. Drug Dev.* **2018**, *3*, 305–317. [CrossRef]
34. Johnson, K.; Shameer, K.; Glicksberg, B.; Readhead, B.; Sengupta, P.P.; Björkegren, J.L.; Kovacic, J.C.; Dudley, J.T. Enabling Precision Cardiology Through Multiscale Biology and Systems Medicine. *JACC Basic Transl. Sci.* **2017**, *2*, 311–327. [CrossRef] [PubMed]
35. Johnson, K.; Soto, J.T.; Glicksberg, B.; Shameer, K.; Miotto, R.; Ali, M.; Ashley, E.; Dudley, J.T. Artificial Intelligence in Cardiology. *J. Am. Coll. Cardiol.* **2018**, *71*, 2668–2679. [CrossRef] [PubMed]
36. Benjamens, S.; Dhunnoo, P.; Meskó, B. The state of artificial intelligence-based FDA-approved medical devices and algorithms: An online database. *NPJ Digit. Med.* **2020**, *3*, 118. [CrossRef] [PubMed]
37. Poplin, R.; Chang, P.-C.; Alexander, D.; Schwartz, S.; Colthurst, T.; Ku, A.; Newburger, D.; Dijamco, J.; Nguyen, N.; Afshar, P.T.; et al. A universal SNP and small-indel variant caller using deep neural networks. *Nat. Biotechnol.* **2018**, *36*, 983–987. [CrossRef] [PubMed]
38. Luo, R.; Sedlazeck, F.J.; Lam, T.-W.; Schatz, M.C. A multi-task convolutional deep neural network for variant calling in single molecule sequencing. *Nat. Commun.* **2019**, *10*, 998. [CrossRef]
39. Luo, R.; Lam, T.-W.; Schatz, M.C. Skyhawk: An Artificial Neural Network-based discriminator for reviewing clinically significant genomic variants. *bioRxiv* **2019**, *13*, 311985. [CrossRef]
40. Hassanzadeh, H.R.; Wang, M.D. DeeperBind: Enhancing prediction of sequence specificities of DNA binding proteins. In Proceedings of the 2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Shen Zhen, China, 15–18 December 2016; pp. 178–183.
41. Pan, X.; Shen, H.-B. RNA-protein binding motifs mining with a new hybrid deep learning based cross-domain knowledge integration approach. *BMC Bioinform.* **2017**, *18*, 136. [CrossRef] [PubMed]
42. DeepSea. Available online: https://hb.flatironinstitute.org/deepsea/ (accessed on 8 February 2022).
43. Boža, V.; Brejova, B.; Vinař, T. DeepNano: Deep recurrent neural networks for base calling in MinION nanopore reads. *PLoS ONE* **2017**, *12*, e0178751. [CrossRef]
44. SpliceAI: Predicting Splicing from Primary Sequence with Deep Learning. Available online: https://hpc.nih.gov/apps/SpliceAI.html (accessed on 8 February 2022).
45. Gurovich, Y.; Hanani, Y.; Bar, O.; Nadav, G.; Fleischer, N.; Gelbman, D.; Basel-Salmon, L.; Krawitz, P.M.; Kamphausen, S.B.; Zenker, M.; et al. Identifying facial phenotypes of genetic disorders using deep learning. *Nat. Med.* **2019**, *25*, 60–64. [CrossRef] [PubMed]
46. PhenomeNet Variant Predictor (PVP). Available online: https://github.com/bio-ontology-research-group/phenomenet-vp (accessed on 8 February 2022).
47. Ainscough, B.J.; Barnell, E.K.; Ronning, P.; Campbell, K.M.; Wagner, A.H.; Fehniger, T.A.; Dunn, G.P.; Uppaluri, R.; Govindan, R.; Rohan, T.E.; et al. A deep learning approach to automate refinement of somatic variant calling from cancer sequencing data. *Nat. Genet.* **2018**, *50*, 1735–1743. [CrossRef] [PubMed]
48. Yuan, Y.; Shi, Y.; Li, C.; Kim, J.; Cai, W.; Han, Z.; Feng, D.D. DeepGene: An advanced cancer type classifier based on deep learning and somatic point mutations. *BMC Bioinform.* **2016**, *17*, 243–256. [CrossRef] [PubMed]
49. Xie, R.; Wen, J.; Quitadamo, A.; Cheng, J.; Shi, X. A deep auto-encoder model for gene expression prediction. *BMC Genom.* **2017**, *18*, 39–49. [CrossRef] [PubMed]

50. Wang, Y.; Liu, T.; Xu, D.; Shi, H.; Zhang, C.; Mo, Y.-Y.; Wang, Z. Predicting DNA Methylation State of CpG Dinucleotide Using Genome Topological Features and Deep Networks. *Sci. Rep.* **2016**, *6*, 19598. [CrossRef] [PubMed]
51. Abrahamsson, E.; Plotkin, S.S. BioVEC: A program for Biomolecule Visualization with Ellipsoidal Coarse-graining. *J. Mol. Graph. Model.* **2009**, *28*, 140–145. [CrossRef]
52. Lanchantin, J.; Singh, R.; Wang, B.; Qi, Y. Deep Motif Dashboard: Visualizing and Understanding Genomic Sequences Using Deep Neural Net-Works. *Pac. Symp. Biocomput.* **2017**, *22*, 254–265. [PubMed]
53. Singh, R.; Lanchantin, J.; Robins, G.; Qi, Y. DeepChrome: Deep-learning for predicting gene expression from histone modifications. *Bioinformatics* **2016**, *32*, i639–i648. [CrossRef] [PubMed]
54. Teng, H.; Cao, M.D.; Hall, M.B.; Duarte, T.; Wang, S.; Coin, L.J.M. Chiron: Translating nanopore raw signal directly into nucleotide sequence using deep learning. *GigaScience* **2018**, *7*, giy037. [CrossRef] [PubMed]
55. Way, G.P.; Greene, C.S. Extracting a biologically relevant latent space from cancer transcriptomes with variational autoencoders. *Pac. Symp. Biocomput.* **2018**, *23*, 80–91.
56. Ravasio, V.; Ritelli, M.; Legati, A.; Giacopuzzi, E. GARFIELD-NGS: Genomic vARiants FIltering by dEep Learning moDels in NGS. *Bioinformatics* **2018**, *34*, 3038–3040. [CrossRef] [PubMed]
57. Lin, X.; Zhao, K.; Xiao, T.; Quan, Z.; Wang, Z.-J.; Yu, P.S. DeepGS: Deep Representation Learning of Graphs and Sequences for Drug-Target Binding Affinity Prediction. *arXiv* **2020**, arXiv:2003.13902.
58. Quang, D.; Chen, Y.; Xie, X. DANN: A deep learning approach for annotating the pathogenicity of genetic variants. *Bioinformatics* **2015**, *31*, 761–763. [CrossRef] [PubMed]
59. Quang, D.; Xie, X. DanQ: A hybrid convolutional and recurrent deep neural network for quantifying the function of DNA sequences. *Nucleic Acids Res.* **2016**, *44*, e107. [CrossRef]
60. Cao, R.; Freitas, C.; Chan, L.; Sun, M.; Jiang, H.; Chen, Z. ProLanGO: Protein Function Prediction Using Neural Machine Translation Based on a Recurrent Neural Network. *Molecules* **2017**, *22*, 1732. [CrossRef]
61. BCC-NER Gene/Protein Mention Tagger. Available online: http://www.biominingbu.org:8080/BCC-NER/ (accessed on 8 February 2022).
62. Provoost, T.; Moens, M.-F. Semi-supervised Learning for the BioNLP Gene Regulation Network. *BMC Bioinform.* **2015**, *16*, S4. [CrossRef] [PubMed]
63. Ramachandran, R.; Arutchelvan, K. Named entity recognition on bio-medical literature documents using hybrid based approach. *J. Ambient Intell. Humaniz. Comput.* **2021**, *10*, 1–10. [CrossRef]
64. Topol, E.J. High-performance medicine: The convergence of human and artificial intelligence. *Nat. Med.* **2019**, *25*, 44–56. [CrossRef]
65. Krittanawong, C.; Bomback, A.S.; Baber, U.; Bangalore, S.; Messerli, F.H.; Tang, W.H.W. Future Direction for Using Artificial Intelligence to Predict and Manage Hypertension. *Curr. Hypertens. Rep.* **2018**, *20*, 75. [CrossRef]
66. Krittanawong, C.; Johnson, K.; Rosenson, R.S.; Wang, Z.; Aydar, M.; Baber, U.; Min, J.K.; Tang, W.H.W.; Halperin, J.L.; Narayan, S.M. Deep learning for cardiovascular medicine: A practical primer. *Eur. Hear. J.* **2019**, *40*, 2058–2073. [CrossRef] [PubMed]
67. Alaa, A.M.; Bolton, T.; Angelantonio, E.D.; Rudd, J.H.F.; van der Schaar, M. Cardiovascular disease risk prediction using automated machine learning: A prospective study of 423,604 UK Biobank participants. *PLoS ONE* **2019**, *14*, e0213653. [CrossRef] [PubMed]
68. Amarbayasgalan, T.; Van Huy, P.; Ryu, K.H. Comparison of the Framingham Risk Score and Deep Neural Network-Based Coronary Heart Disease Risk Prediction. In *Advances in Intelligent Information Hiding and Multimedia Signal Processing*; Springer Science and Business Media LLC: Berlin/Heidelberg, Germany, 2020; pp. 273–280.
69. Carter, R.J.; Dubchak, I.; Holbrook, S.R. A computational approach to identify genes for functional RNAs in genomic sequences. *Nucleic Acids Res.* **2001**, *29*, 3928–3938. [CrossRef] [PubMed]
70. Bockhorst, J.; Craven, M.; Page, D.; Shavlik, J.; Glasner, J. A Bayesian network approach to operon prediction. *Bioinformatics* **2003**, *19*, 1227–1235. [CrossRef] [PubMed]
71. Cawley, S.L.; Pachter, L. HMM sampling and applications to gene finding and alternative splicing. *Bioinformatics* **2003**, *19*, ii36–ii41. [CrossRef]
72. Pounraja, V.K.; Jayakar, G.; Jensen, M.; Kelkar, N.; Girirajan, S. A machine-learning approach for accurate detection of copy-number variants from exome sequencing. *Genome Res.* **2018**, *29*, 460931. [CrossRef] [PubMed]
73. Li, C.; Sun, D.; Liu, J.; Li, M.; Zhang, B.; Liu, Y.; Wang, Z.; Wen, S.; Zhou, J. A Prediction Model of Essential Hypertension Based on Genetic and Environmental Risk Factors in Northern Han Chinese. *Int. J. Med Sci.* **2019**, *16*, 793–799. [CrossRef] [PubMed]
74. Juhola, M.; Joutsijoki, H.; Penttinen, K.; Aalto-Setälä, K. Detection of genetic cardiac diseases by Ca^{2+} transient profiles using machine learning methods. *Sci. Rep.* **2018**, *8*, 9355. [CrossRef] [PubMed]
75. Oguz, C.; Sen, S.K.; Davis, A.R.; Fu, Y.-P.; O'Donnell, C.J.; Gibbons, G.H. Genotype-driven identification of a molecular network predictive of advanced coronary calcium in ClinSeq(R) and Fram-ingham Heart Study cohorts. *BMC Syst. Biol.* **2017**, *11*, 99. [CrossRef]
76. Burghardt, T.P.; Ajtai, K. Neural/Bayes network predictor for inheritable cardiac disease pathogenicity and phenotype. *J. Mol. Cell. Cardiol.* **2018**, *119*, 19–27. [CrossRef] [PubMed]
77. Cui, S.; Wu, Q.; West, J.; Bai, J. Machine learning-based microarray analyses indicate low-expression genes might collectively influence PAH disease. *PLoS Comput. Biol.* **2019**, *15*, e1007264. [CrossRef] [PubMed]

78. Carroll, A.; Chang, P. Improving the Accuracy of Genomic Analysis with DeepVariant 1.0. Available online: https://ai.googleblog.com/2020/09/improving-accuracy-of-genomic-analysis.html (accessed on 8 February 2022).
79. Zou, J.; Huss, M.; Abid, A.; Mohammadi, P.; Torkamani, A.; Telenti, A. A primer on deep learning in genomics. *Nat. Genet.* **2019**, *51*, 12–18. [CrossRef] [PubMed]
80. Park, C.; Kim, J.; Kim, J.; Park, S. Machine learning-based identification of genetic interactions from heterogeneous gene expression profiles. *PLoS ONE* **2018**, *13*, e0201056. [CrossRef]
81. Li, J.; Jew, B.; Zhan, L.; Hwang, S.; Coppola, G.; Freimer, N.B.; Sul, J.H. ForestQC: Quality control on genetic variants from next-generation sequencing data using random forest. *PLoS Comput. Biol.* **2019**, *15*, e1007556. [CrossRef] [PubMed]
82. Yelmen, B.; Decelle, A.; Ongaro, L.; Marnetto, D. Creating Artificial Human Genomes Using Generative Models. *bioRxiv* **2019**, 769091. [CrossRef]
83. Gupta, A.; Zou, J. Feedback GAN (FBGAN) for DNA: A Novel Feedback-Loop Architecture for Optimizing Protein Functions. *arXiv* **2018**, arXiv:1804.01694.
84. Beaulieu-Jones, B.K.; Wu, Z.S.; Williams, C.; Lee, R.; Bhavnani, S.P.; Byrd, J.B.; Greene, C.S. Privacy-preserving generative deep neural networks support clinical data sharing. *Circ. Cardiovasc. Qual. Outcomes* **2019**, *12*, e005122. [CrossRef] [PubMed]
85. Available online: http://kundajelab.github.io/dragonn/ (accessed on 8 February 2022).
86. Márquez-Luna, C.; Loh, P.; South Asian Type 2 Diabetes (SAT2D) Consortium; The SIGMA Type 2 Diabetes Consortium; Price, A.L. Multiethnic polygenic risk scores improve risk prediction in diverse populations. *Genet. Epidemiol.* **2017**, *41*, 811–823. [CrossRef] [PubMed]
87. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the Inception Architecture for Computer Vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016.
88. Kelley, D.R.; Snoek, J.; Rinn, J.L. Basset: Learning the regulatory code of the accessible genome with deep convolutional neural net-works. *Genome Res.* **2016**, *26*, 990–999. [CrossRef] [PubMed]
89. Angermueller, C.; Pärnamaa, T.; Parts, L.; Stegle, O. Deep learning for computational biology. *Mol. Syst. Biol.* **2016**, *12*, 878. [CrossRef] [PubMed]
90. Rappoport, N.; Shamir, R. Multi-omic and multi-view clustering algorithms: Review and cancer benchmark. *Nucleic Acids Res.* **2018**, *46*, 10546–10562. [CrossRef]
91. Ferreira, C.R.; Altassan, R.; Marques-Da-Silva, D.; Francisco, R.; Jaeken, J.; Morava, E. Recognizable phenotypes in CDG. *J. Inherit. Metab. Dis.* **2018**, *41*, 541–553. [CrossRef] [PubMed]
92. Clark, M.J.; Chen, R.; Lam, H.Y.K.; Karczewski, J.; Chen, R.; Euskirchen, G.; Butte, A.J.; Snyder, M. Performance comparison of exome DNA sequencing technologies. *Nat. Biotechnol.* **2011**, *29*, 908–914. [CrossRef]
93. Chilamakuri, C.S.R.; Lorenz, S.; Madoui, M.-A.; Vodák, D.; Sun, J.; Hovig, E.; Myklebost, O.; A Meza-Zepeda, L. Performance comparison of four exome capture systems for deep sequencing. *BMC Genom.* **2014**, *15*, 449. [CrossRef] [PubMed]
94. Zhang, Y.-Z.; Yamaguchi, R.; Imoto, S.; Miyano, S. Sequence-specific bias correction for RNA-seq data using recurrent neural networks. *BMC Genom.* **2017**, *18*, 1044. [CrossRef] [PubMed]
95. Sheikhalishahi, S.; Miotto, R.; Dudley, J.T.; Lavelli, A.; Rinaldi, F.; Osmani, V. Natural Language Processing of Clinical Notes on Chronic Diseases: Systematic Review. *JMIR Med. Inform.* **2019**, *7*, e12239. [CrossRef] [PubMed]
96. Weissman, G.E.; Harhay, M.O.; Lugo, R.M.; Fuchs, B.D.; Halpern, S.D.; Mikkelsen, M.E. Natural Language Processing to Assess Documentation of Features of Critical Illness in Discharge Documents of Acute Respiratory Distress Syndrome Survivors. *Ann. Am. Thorac. Soc.* **2016**, *13*, 1538–1545. [CrossRef] [PubMed]
97. Calapodescu, I.; Rozier, D.; Artemova, S.; Bosson, J.-L. Semi-Automatic De-identification of Hospital Discharge Summaries with Natural Language Processing: A Case-Study of Performance and Real-World Usability. In Proceedings of the 2017 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), Institute of Electrical and Electronics Engineers (IEEE), Exeter, UK, 21–23 June 2017; pp. 1106–1111.
98. Deng, Z.; Yin, K.; Bao, Y.; Armengol, V.D.; Wang, C.; Tiwari, A.; Barzilay, R.; Parmigiani, G.; Braun, D.; Hughes, K.S. Validation of a Semiautomated Natural Language Processing–Based Procedure for Meta-Analysis of Cancer Susceptibility Gene Penetrance. *JCO Clin. Cancer Inform.* **2019**, *3*, 1–9. [CrossRef] [PubMed]
99. Rumshisky, A.A.; Ghassemi, M.; Naumann, T.; Szolovits, P.; Castro, V.M.; McCoy, T.; Perlis, R.H. Predicting early psychiatric readmission with natural language processing of narrative discharge summaries. *Transl. Psychiatry* **2016**, *6*, e921. [CrossRef] [PubMed]
100. Morgan, A.A.; Hirschman, L.; Colosimo, M.; Yeh, A.S.; Colombe, J.B. Gene name identification and normalization using a model organism database. *J. Biomed. Inform.* **2004**, *37*, 396–410. [CrossRef] [PubMed]
101. Gligorijevic, D.; Stojanovic, J.; Djuric, N.; Radosavljevic, V.; Grbovic, M.; Kulathinal, R.J.; Obradovic, Z. Large-Scale Discovery of Disease-Disease and Disease-Gene Associations. *Sci. Rep.* **2016**, *6*, 32404. [CrossRef] [PubMed]
102. Buchan, K.; Filannino, M.; Uzuner, Ö. Automatic prediction of coronary artery disease from clinical narratives. *J. Biomed. Inform.* **2017**, *72*, 23–32. [CrossRef]
103. Arvind, P.; Jayashree, S.; Jambunathan, S.; Nair, J.; Kakkar, V.V. Understanding gene expression in coronary artery disease through global profiling, network analysis and independent validation of key candidate genes. *J. Genet.* **2015**, *94*, 601–610. [CrossRef] [PubMed]

104. Young, T.; Hazarika, D.; Poria, S.; Cambria, E. Recent trends in deep learning based natural language processing. *IEEE Comput. Intell. Mag.* **2018**, *13*, 55–75. [CrossRef]
105. Zomnir, M.G.; Lipkin, L.; Pacula, M.; Meneses, E.D.; MacLeay, A.; Duraisamy, S.; Nadhamuni, N.; Al Turki, S.H.; Zheng, Z.; Rivera, M.; et al. Artificial Intelligence Approach for Variant Reporting. *JCO Clin. Cancer Inform.* **2018**, *2*, CCI.16.00079. [CrossRef]
106. Luketina, J.; Nardelli, N.; Farquhar, G.; Foerster, J.; Andreas, J.; Grefenstette, E.; Whiteson, S.; Rocktäschel, T. A Survey of Reinforcement Learning Informed by Natural Language. *arXiv* **2019**, arXiv:1906.03926.
107. Towbin, J.A.; McKenna, W.J.; Abrams, D.J.; Ackerman, M.J.; Calkins, H.; Darrieux, F.C.C.; Daubert, J.P.; de Chillou, C.; DePasquale, E.C.; Desai, M.Y.; et al. 2019 HRS expert consensus statement on evaluation, risk stratification, and management of arrhythmogenic cardio-myopathy. *Heart Rhythm* **2019**, *16*, e301–e372. [CrossRef] [PubMed]
108. Tavtigian, S.V.; Greenblatt, M.S.; Harrison, S.M.; Nussbaum, R.L.; Prabhu, S.A.; Boucher, K.M.; Biesecker, L.G. Modeling the ACMG/AMP variant classification guidelines as a Bayesian classification framework. *Genet. Med.* **2018**, *20*, 1054–1060. [CrossRef] [PubMed]
109. Phelan, J.; O'Sullivan, D.M.; Machado, D.; Ramos, J.; Whale, A.S.; O'Grady, J.; Dheda, K.; Campino, S.; McNerney, R.; Viveiros, M.; et al. The variability and reproducibility of whole genome sequencing technology for detecting resistance to anti-tuberculous drugs. *Genome Med.* **2016**, *8*, 132. [CrossRef] [PubMed]
110. Traore, K.; Bull, S.; Niare, A.; Konate, S.; Thera, M.A.; Kwiatkowski, D.; Parker, M.; Doumbo, O.K. Understandings of genomic research in developing countries: A qualitative study of the views of MalariaGEN participants in Mali. *BMC Med. Ethic.* **2015**, *16*, 42. [CrossRef]
111. The Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3000 shared controls. *Nature* **2007**, *447*, 661. [CrossRef] [PubMed]
112. Clayton, D.G.; Walker, N.M.; Smyth, D.J.; Pask, R.; Cooper, J.D.; Maier, L.M.; Smink, L.J.; Lam, A.C.; Ovington, N.R.; Stevens, H.E.; et al. Population structure, differential bias and genomic control in a large-scale, case-control association study. *Nat. Genet.* **2005**, *37*, 1243–1246. [CrossRef] [PubMed]
113. Sedlazeck, F.J.; Rescheneder, P.; Smolka, M.; Fang, H.; Nattestad, M.; von Haeseler, A.; Schatz, M.C. Accurate detection of complex structural variations using single-molecule sequencing. *Nat. Methods* **2018**, *15*, 461–468. [CrossRef] [PubMed]
114. Romagnoni, A.; Jégou, S.; Van Steen, K.; Wainrib, G.; Hugot, J.-P.; International Inflammatory Bowel Disease Genetics Consortium (IIBDGC). Comparative performances of machine learning methods for classifying Crohn Disease patients using genome-wide genotyping data. *Sci. Rep.* **2019**, *9*, 10351. [CrossRef] [PubMed]
115. Ptaszynski, M.; Rzepka, R.; Araki, K.; Momouchi, Y. Language Combinatorics: A Sentence Pattern Extraction Architecture Based on Combinatorial Explosion. *Int. J. Comput. Linguist. Res.* **2011**, *2*, 24–36.
116. Xing, W.; Qi, J.; Yuan, X.; Li, L.; Zhang, X.; Fu, Y.; Xiong, S.; Hu, L.; Peng, J. A gene-phenotype relationship extraction pipeline from the biomedical literature using a representation learning approach. *Bioinformatics* **2018**, *34*, i386–i394. [CrossRef] [PubMed]
117. Tseytlin, E.; Mitchell, K.J.; Legowski, E.; Corrigan, J.; Chavan, G.; Jacobson, R.S. NOBLE–Flexible concept recognition for large-scale biomedical natural language processing. *BMC Bioinform.* **2016**, *17*, 32. [CrossRef] [PubMed]
118. Ceballos, D.; López-Alvarez, D.C.; Isaza, G.; Tabares-Soto, R.; Orozco-Arias, S.; Ferrin, C.D. A Machine Learning-based Pipeline for the Classification of CTX-M in Metagenomics Samples. *Processes* **2019**, *7*, 235. [CrossRef]
119. Guzzetta, G.; Jurman, G.; Furlanello, C. A machine learning pipeline for quantitative phenotype prediction from genotype data. *BMC Bioinform.* **2010**, *11*, S3. [CrossRef] [PubMed]
120. Kalkatawi, M.; Magana-Mora, A.; Jankovic, B.; Bajic, V.B. DeepGSR: An optimized deep-learning structure for the recognition of genomic signals and regions. *Bioinformatics* **2019**, *35*, 1125–1132. [CrossRef] [PubMed]
121. Boudellioua, I.; Kulmanov, M.; Schofield, P.N.; Gkoutos, G.V.; Hoehndorf, R. DeepPVP: Phenotype-based prioritization of causative variants using deep learning. *BMC Bioinform.* **2019**, *20*, 65. [CrossRef]
122. Ambale-Venkatesh, B.; Yang, X.; Wu, C.O.; Liu, K.; Hundley, W.G.; McClelland, R.; Gomes, A.S.; Folsom, A.R.; Shea, S.; Guallar, E.; et al. Cardiovascular Event Prediction by Machine Learning: The Multi-Ethnic Study of Atherosclerosis. *Circ. Res.* **2017**, *121*, 1092–1101. [CrossRef] [PubMed]
123. Zhuang, X.; Sun, X.; Zhong, X.; Zhou, H.; Zhang, S.; Liao, X. Deep phenotyping and prediction of long-term heart failure by machine learning. *J. Am. Coll. Cardiol.* **2019**, *73*, 690. [CrossRef]
124. Ahmad, F.; McNally, E.M.; Ackerman, M.J.; Baty, L.C.; Day, S.M.; Kullo, I.J.; Madueme, P.C.; Maron, M.S.; Martinez, M.W.; Salberg, L.; et al. Establishment of specialized clinical cardiovascular genetics programs: Recognizing the need and meeting standards: A scientific statement from the American Heart Association. *Circ. Genom. Precis. Med.* **2019**, *12*, e000054. [CrossRef]
125. O'Rawe, J.; Jiang, T.; Sun, G.; Wu, Y.; Wang, W.; Hu, J.; Bodily, P.; Tian, L.; Hakonarson, H.; Johnson, W.E.; et al. Low concordance of multiple variant-calling pipelines: Practical implications for exome and genome sequencing. *Genome Med.* **2013**, *5*, 28. [CrossRef] [PubMed]
126. Dias, R.; Torkamani, A. Artificial intelligence in clinical and genomic diagnostics. *Genome Med.* **2019**, *11*, 70. [CrossRef] [PubMed]
127. Alipanahi, B.; Delong, A.; Weirauch, M.T.; Frey, B.J. Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. *Nat. Biotechnol.* **2015**, *33*, 831–838. [CrossRef] [PubMed]
128. Murugesan, G.; Abdulkadhar, S.; Bhasuran, B.; Natarajan, J. BCC-NER: Bidirectional, contextual clues named entity tagger for gene/protein mention recognition. *EURASIP J. Bioinform. Syst. Biol.* **2017**, *2017*, 7. [CrossRef]

129. Bossy, R.; Golik, W.; Ratkovic, Z.; Valsamou, D.; Bessières, P.; Nédellec, C. Overview of the gene regulation network and the bacteria biotope tasks in BioNLP'13 shared task. *BMC Bioinform.* **2015**, *16* (Suppl. 10), S1. [CrossRef]
130. Moon, S.; Liu, S.; Scott, C.G.; Samudrala, S.; Abidian, M.M.; Geske, J.B.; Noseworthy, P.A.; Shellum, J.L.; Chaudhry, R.; Ommen, S.R.; et al. Automated extraction of sudden cardiac death risk factors in hypertrophic cardiomyopathy patients by natural language processing. *Int. J. Med. Inform.* **2019**, *128*, 32–38. [CrossRef] [PubMed]
131. Richards, S.; Aziz, N.; Bale, S.; Bick, D.; Das, S.; Gastier-Foster, J.; Grody, W.W.; Hegde, M.; Lyon, E.; Spector, E.; et al. Standards and guidelines for the interpretation of sequence variants: A joint consensus recommendation of the Ameri-can College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Circ. Res.* **2015**, *17*, 405–424. [CrossRef] [PubMed]
132. Kelly, M.A.; Caleshu, C.; Morales, A.; Buchan, J.; Wolf, Z.; Harrison, S.M.; Cook, S.; Dillon, M.W.; Garcia, J.; Haverfield, E.; et al. Adaptation and validation of the ACMG/AMP variant classification framework for MYH7-associated inherited cardiomyopathies: Recommendations by ClinGen's Inherited Cardiomyopathy Expert Panel. *Genet. Med.* **2018**, *20*, 351–359. [CrossRef] [PubMed]
133. Glicksberg, B.; Johnson, K.; Dudley, J.T. The next generation of precision medicine: Observational studies, electronic health records, biobanks and continuous monitoring. *Hum. Mol. Genet.* **2018**, *27*, R56–R62. [CrossRef] [PubMed]
134. Solomon, S.D.; Wolff, S.; Watkins, H.; Ridker, P.M.; Come, P.; McKenna, W.J.; Seidman, C.E.; Lee, R.T. Left ventricular hypertrophy and morphology in familial hypertrophic cardiomyopathy associated with mutations of the beta-myosin heavy chain gene. *J. Am. Coll. Cardiol.* **1993**, *22*, 498–505. [CrossRef]
135. Binder, J.; Ommen, S.R.; Gersh, B.J.; Van Driest, S.L.; Tajik, A.J.; Nishimura, R.A.; Ackerman, M.J. Echocardiography-Guided Genetic Testing in Hypertrophic Cardiomyopathy: Septal Morphological Features Predict the Presence of Myofilament Mutations. *Mayo Clin. Proc.* **2006**, *81*, 459–467. [CrossRef]
136. Claassens, D.M.F.; Vos, G.J.; Bergmeijer, T.O.; Hermanides, R.S.; Hof, A.W.V.T.; Van Der Harst, P.; Barbato, E.; Morisco, C.; Gin, R.M.T.J.; Asselbergs, F.W.; et al. A Genotype-Guided Strategy for Oral P2Y12 Inhibitors in Primary PCI. *N. Engl. J. Med.* **2019**, *381*, 1621–1631. [CrossRef] [PubMed]
137. Nagueh, S.F.; Smiseth, O.A.; Appleton, C.P.; Byrd, B.F.; Dokainish, H.; Edvardsen, T.; Flachskampf, F.A.; Gillebert, T.C.; Klein, A.L.; Lancellotti, P.; et al. Recommendations for the Evaluation of Left Ventricular Diastolic Function by Echocardiography: An Update from the American Society of Echocardiography and the European Association of Cardiovascular Imaging. *Eur. Hear. J. Cardiovasc. Imaging* **2016**, *17*, 1321–1360. [CrossRef] [PubMed]
138. Castillo, L.P.; Tasan, M.; Myers, C.L.; Lee, H.; Joshi, T.; Zhang, C.; Guan, Y.; Leone, M.; Pagnani, A.; Kim, W.; et al. A critical assessment of Mus musculus gene function prediction using integrated genomic evidence. *Genome Biol.* **2008**, *9*, S2. [CrossRef]
139. Ross, E.G.; Shah, N.H.; Dalman, R.L.; Nead, K.T.; Cooke, J.; Leeper, N.J. The use of machine learning for the identification of peripheral artery disease and future mortality risk. *J. Vasc. Surg.* **2016**, *64*, 1515–1522.e3. [CrossRef]
140. Safarova, M.; Liu, H.; Kullo, I.J. Rapid identification of familial hypercholesterolemia from electronic health records: The SEARCH study. *J. Clin. Lipidol.* **2016**, *10*, 1230–1239. [CrossRef]
141. Mowery, D.L.; Chapman, B.E.; Conway, M.; South, B.R.; Madden, E.; Keyhani, S.; Chapman, W.W. Extracting a stroke phenotype risk factor from Veteran Health Administration clinical reports: An information content analysis. *J. Biomed. Semant.* **2016**, *7*, 1–12. [CrossRef] [PubMed]
142. Liao, K.P.; Ananthakrishnan, A.N.; Kumar, V.; Xia, Z.; Cagan, A.; Gainer, V.S.; Goryachev, S.; Chen, P.; Savova, G.K.; Agniel, D.; et al. Methods to Develop an Electronic Medical Record Phenotype Algorithm to Compare the Risk of Coronary Artery Disease across 3 Chronic Disease Cohorts. *PLoS ONE* **2015**, *10*, e0136651. [CrossRef] [PubMed]
143. Biffi, C.; De Marvao, A.; Attard, M.I.; Dawes, T.J.W.; Whiffin, N.; Bai, W.; Shi, W.; Francis, C.; Meyer, H.; Buchan, R.; et al. Three-dimensional cardiovascular imaging-genetics: A mass univariate framework. *Bioinformatics* **2017**, *34*, 97–103. [CrossRef]
144. Zhao, J.; Feng, Q.; Wu, P.; Lupu, R.A.; Wilke, R.A.; Wells, Q.S.; Denny, J.C.; Wei, W.-Q. Learning from Longitudinal Data in Electronic Health Record and Genetic Data to Improve Cardiovascular Event Prediction. *Sci. Rep.* **2019**, *9*, 717. [CrossRef] [PubMed]
145. Schafer, S.; de Marvao, A.; Adami, E.; Fiedler, L.R.; Ng, B.; Khin, E.; Rackham, O.J.L.; van Heesch, S.; Pua, C.J.; Kui, M.; et al. Titin-truncating variants affect heart function in disease cohorts and the general population. *Genet. Med.* **2017**, *49*, 46–53. [CrossRef]
146. Dogan, M.V.; Grumbach, I.M.; Michaelson, J.J.; Philibert, R.A. Integrated genetic and epigenetic prediction of coronary heart disease in the Framingham Heart Study. *PLoS ONE* **2018**, *13*, e0190549. [CrossRef] [PubMed]
147. Tang, W.W.; Kitai, T.; Hazen, S.L. Gut Microbiota in Cardiovascular Health and Disease. *Circ. Res.* **2017**, *120*, 1183–1196. [CrossRef] [PubMed]
148. Holloway, J.W.; Francis, S.S.; Fong, K.; Yang, I. Genomics and the respiratory effects of air pollution exposure. *Respirology* **2012**, *17*, 590–600. [CrossRef] [PubMed]
149. Ward-Caviness, C.K. A review of gene-by-air pollution interactions for cardiovascular disease, risk factors, and biomarkers. *Qual. Life Res.* **2019**, *138*, 547–561. [CrossRef]
150. Rodriguez, F.; Chung, S.; Blum, M.R.; Coulet, A.; Basu, S.; Palaniappan, L.P. Atherosclerotic Cardiovascular Disease Risk Prediction in Disaggregated Asian and Hispanic Subgroups Using Elec-tronic Health Records. *J. Am. Heart Assoc.* **2019**, *8*, e011874. [CrossRef] [PubMed]
151. Popejoy, A.B.; Fullerton, S.M. Genomics is failing on diversity. *Nature* **2016**, *538*, 161–164. [CrossRef]

152. Ng, M.C.Y.; Shriner, D.; Chen, B.H.; Li, J.; Chen, W.-M.; Guo, X.; Liu, J.; Bielinski, S.J.; Yanek, L.R.; Nalls, M.A.; et al. Meta-analysis of genome-wide association studies in African Americans provides insights into the genetic architecture of type 2 diabetes. *PLoS Genet.* **2014**, *10*, e1004517. [CrossRef] [PubMed]
153. Bustamante, C.D.; Burchard, E.G.; De la Vega, F.M. Genomics for the world. *Nature* **2011**, *475*, 163–165. [CrossRef] [PubMed]
154. Need, A.C.; Goldstein, D.B. Next generation disparities in human genomics: Concerns and remedies. *Trends Genet.* **2009**, *25*, 489–494. [CrossRef] [PubMed]
155. Shi, M.; Umbach, D.M.; Weinberg, C.R. Family-based gene-by-environment interaction studies: Revelations and remedies. *Epidemiology* **2011**, *22*, 400–407. [CrossRef] [PubMed]
156. Inouye, M.; Abraham, G.; Nelson, C.P.; Wood, A.M.; Sweeting, M.J.; Dudbridge, F.; Lai, F.Y.; Kaptoge, S.; Brozynska, M.; Wang, T.; et al. Genomic Risk Prediction of Coronary Artery Disease in 480,000 Adults. *J. Am. Coll. Cardiol.* **2018**, *72*, 1883. [CrossRef] [PubMed]
157. Martin, A.R.; Kanai, M.; Kamatani, Y.; Okada, Y.; Neale, B.M.; Daly, M.J. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat. Genet.* **2019**, *51*, 584–591. [CrossRef] [PubMed]
158. Fahed, A.C.; Wang, M.; Homburger, J.R.; Patel, A.P.; Bick, A.G.; Neben, C.L.; Lai, C.; Brockman, D.; Philippakis, A.; Ellinor, P.T.; et al. Polygenic background modifies penetrance of monogenic variants conferring risk for coronary artery disease, breast cancer, or colorectal cancer. *Nat. Commun.* **2019**, *11*, 3635. [CrossRef] [PubMed]
159. Ghaleb, Y.; Elbitar, S.; El Khoury, P.; Bruckert, E.; Carreau, V.; Carrié, A.; Moulin, P.; Di Filippo, M.; Charriere, S.; Iliozer, H.; et al. Usefulness of the genetic risk score to identify phenocopies in families with familial hypercholesterolemia? *Eur. J. Hum. Genet.* **2018**, *26*, 570–578. [CrossRef]
160. Dudbridge, F. Power and Predictive Accuracy of Polygenic Risk Scores. *PLoS Gene.* **2013**, *9*, e1003348.
161. Zhao, B.; Zou, F. Is Polygenic Risk Scores Prediction Good? *bioRxiv* **2018**, 447797. [CrossRef]
162. Natarajan, P.; NHLBI TOPMed Lipids Working Group; Peloso, G.M.; Zekavat, S.M.; Montasser, M.; Ganna, A.; Chaffin, M.; Khera, A.V.; Zhou, W.; Bloom, J.M.; et al. Deep-coverage whole genome sequences and blood lipids among 16,324 individuals. *Nat. Commun.* **2018**, *9*, 3391. [CrossRef] [PubMed]
163. Khera, A.V.; Chaffin, M.; Zekavat, S.M.; Collins, R.L.; Roselli, C.; Natarajan, P.; Lichtman, J.H.; D'Onofrio, G.; Mattera, J.; Dreyer, R.; et al. Whole Genome Sequencing to Characterize Monogenic and Polygenic Contributions in Patients Hospitalized with Early-Onset Myocardial Infarction. *Circulation* **2019**, *139*, 1593–1602. [CrossRef] [PubMed]
164. Gulshan, V.; Peng, L.; Coram, M.; Stumpe, M.C.; Wu, D.; Narayanaswamy, A.; Venugopalan, S.; Widner, K.; Madams, T.; Cuadros, J.; et al. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *JAMA* **2016**, *316*, 2402–2410. [CrossRef] [PubMed]
165. Voets, M. *Deep Learning: From Data Extraction to Large-Scale Analysis*; UiT Norges Arktiske Universitet: Alta, Norway, 2018.
166. Schubach, M.; Re, M.; Robinson, P.N.; Valentini, G. Imbalance-Aware Machine Learning for Predicting Rare and Common Disease-Associated Non-Coding Variants. *Sci. Rep.* **2017**, *7*, 2959. [CrossRef] [PubMed]
167. Giral, H.; Landmesser, U.; Kratzer, A. Into the Wild: GWAS Exploration of Non-coding RNAs. *Front. Cardiovasc. Med.* **2018**, *5*, 181. [CrossRef] [PubMed]
168. Ghorbani, A.; Abid, A.; Zou, J. Interpretation of Neural Networks Is Fragile. *Proc. Conf. AAAI Artif. Intell.* **2019**, *33*, 3681–3688. [CrossRef]
169. De Fauw, J.; Ledsam, J.R.; Romera-Paredes, B.; Nikolov, S.; Tomasev, N.; Blackwell, S.; Askham, H.; Glorot, X.; O'Donoghue, B.; Visentin, D.; et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nat. Med.* **2018**, *24*, 1342–1350. [CrossRef]
170. Tan, S.; Caruana, R.; Hooker, G.; Lou, Y. Distill-and-Compare: Auditing Black-Box Models Using Transparent Model Distillation. *arXiv* **2017**, arXiv:1710.06169.
171. Wang, H.; Wu, Z.; Xing, E.P. Fair Deep Learning Prediction for Healthcare Applications with Confounder Filtering. *arXiv* **2018**, arXiv:1803.07276.
172. Wang, H.; Meghawat, A.; Morency, L.-P.; Xing, E.P. Select-additive learning: Improving generalization in multimodal sentiment analysis. In Proceedings of the 2017 IEEE International Conference on Multimedia and Expo (ICME), Institute of Electrical and Electronics Engineers (IEEE), Hong Kong, China, 10–14 July 2017; pp. 949–954.
173. Klitzman, R.; Chung, W.; Marder, K.; Shanmugham, A.; Chin, L.J.; Stark, M.; Leu, C.-S.; Appelbaum, P.S. Attitudes and Practices Among Internists Concerning Genetic Testing. *J. Genet. Couns.* **2013**, *22*, 90–100. [CrossRef] [PubMed]
174. Giardiello, F.M.; Brensinger, J.D.; Petersen, G.M.; Luce, M.C.; Hylind, L.M.; Bacon, J.A.; Booker, S.V.; Parker, R.D.; Hamilton, S.R. The use and interpretation of commercial APC gene testing for familial adenomatous polyposis. *N. Engl. J. Med.* **1997**, *336*, 823–827. [CrossRef] [PubMed]
175. Wrzeszczynski, K.O.; Frank, M.O.; Koyama, T.; Rhrissorrakrai, K.; Robine, N.; Utro, F.; Emde, A.-K.; Chen, B.-J.; Arora, K.; Shah, M.; et al. Comparing sequencing assays and human-machine analyses in actionable genomics for glioblastoma. *Neurol. Genet.* **2017**, *3*, e164. [CrossRef]
176. Lohr, S. What Ever Happened to IBM's Watson? Available online: https://www.nytimes.com/2021/07/16/technology/what-happened-ibm-watson.html (accessed on 8 February 2022).

177. West, H.J. No Solid Evidence, Only Hollow Argument for Universal Tumor Sequencing: Show Me the Data. *JAMA Oncol.* **2016**, *2*, 717–718. [CrossRef] [PubMed]
178. Tandy-Connor, S.; Guiltinan, J.; Krempely, K.; LaDuca, H.; Reineke, P.; Gutierrez, S.; Gray, P.; Davis, B.T. False-positive results released by direct-to-consumer genetic tests highlight the importance of clinical confirma-tion testing for appropriate patient care. *Gene. Med.* **2018**, *20*, 1515–1521. [CrossRef] [PubMed]