



Review

# WGS Data Collections: How Do Genomic Databases Transform Medicine?

Zbigniew J. Król <sup>1,\*</sup> , Paula Dobosz <sup>1</sup>, Antonina Ślubowska <sup>2</sup> and Magdalena Mroczek <sup>3</sup>

<sup>1</sup> Central Clinical Hospital of Ministry of the Interior and Administration in Warsaw, 02-507 Warsaw, Poland

<sup>2</sup> Department of Biostatistics and Research Methodology, Faculty of Medicine, Collegium Medicum, Cardinal Stefan Wyszyński University in Warsaw, Wóycickiego 1/3, 01-938 Warsaw, Poland

<sup>3</sup> Center for Cardiovascular Genetics and Gene Diagnostics, Foundation for People with Rare Diseases, Wagistrasse 25, 8952 CH-Schlieren-Zurich, Switzerland

\* Correspondence: zbigniew.j.krol@cskmswia.gov.pl

**Abstract:** As a scientific community we assumed that exome sequencing will elucidate the basis of most heritable diseases. However, it turned out it was not the case; therefore, attention has been increasingly focused on the non-coding sequences that encompass 98% of the genome and may play an important regulatory function. The first WGS-based datasets have already been released including underrepresented populations. Although many databases contain pooled data from several cohorts, recently the importance of local databases has been highlighted. Genomic databases are not only collecting data but may also contribute to better diagnostics and therapies. They may find applications in population studies, rare diseases, oncology, pharmacogenetics, and infectious and inflammatory diseases. Further data may be analysed with AI technologies and in the context of other omics data. To exemplify their utility, we put a highlight on the Polish genome database and its practical application.

**Keywords:** COVID-19; database; genomic databases; whole-genome sequencing; WGS



**Citation:** Król, Z.J.; Dobosz, P.; Ślubowska, A.; Mroczek, M. WGS Data Collections: How Do Genomic Databases Transform Medicine? *Int. J. Mol. Sci.* **2023**, *24*, 3031. <https://doi.org/10.3390/ijms24033031>

Academic Editor: Ancha Baranova

Received: 30 December 2022

Revised: 23 January 2023

Accepted: 26 January 2023

Published: 3 February 2023



**Copyright:** © 2023 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/>).

## 1. Introduction

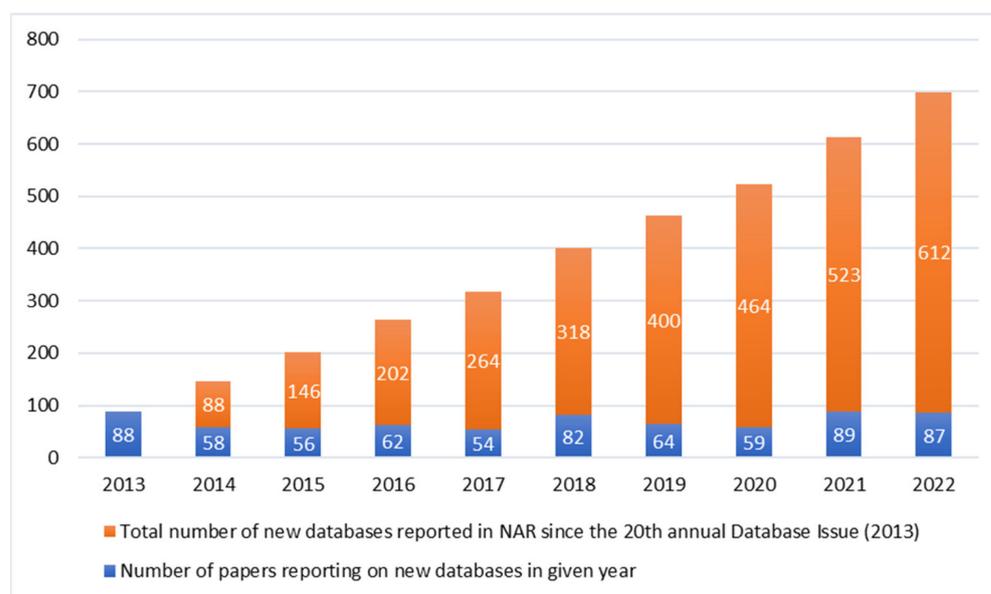
It has been over 20 years since the very first version of the entire human genome was released [1]. Today, although advanced sequencing methods are available at a reasonable price and the role of significant genetic variants localised along the whole genome is quite well defined, the clinical implementation of whole-genome sequencing (WGS) in diagnosis and treatment remains in its infancy [2–4]. Whole exome sequencing was thought to elucidate the genetic background of most of the inherited diseases. However, it was not the case, which is why other omics technologies, such as WGS, RNA-Seq, Epigenomics and Metabolomics, gained importance [5–7]. Here, we would like to highlight the importance of creating WGS databases. With an increasing number of individuals included in the databases, it became apparent that genetic variation differs significantly across ethnic groups [8]. Therefore, it is necessary to create local genomic databases that mirror the smaller and sometimes even endemic population structure.

Oncology remains the major field that can benefit the most from whole genome sequencing since cancer develops because of the changes in the DNA inside cells [9] and also given the number of individuals affected. Thus, we should not hesitate to say that cancer actually is a disease of our genome. Traditionally, most of the studies were focused on the identification of cancer mutations solely in protein-coding genes, ignoring the remaining 99% of the genome dubbed as “the junk DNA”, partially because very few tools for big data analysis were available, if any at all [10]. As a result, the first collection of more than 350 cancer-related genes, protein-coding genes to be more specific, has been created with new genes being added over time [11]. Currently used cancer diagnostic panels span more than 500 genes forming the modern foundation of cancer diagnosis [12]. However,

gene panels identify mutations only in 0.01–0.10% of the genome, at best. The identification of variants scattered in the whole sequence of the genome may provide significantly more precision in cancer diagnosis and better treatment options.

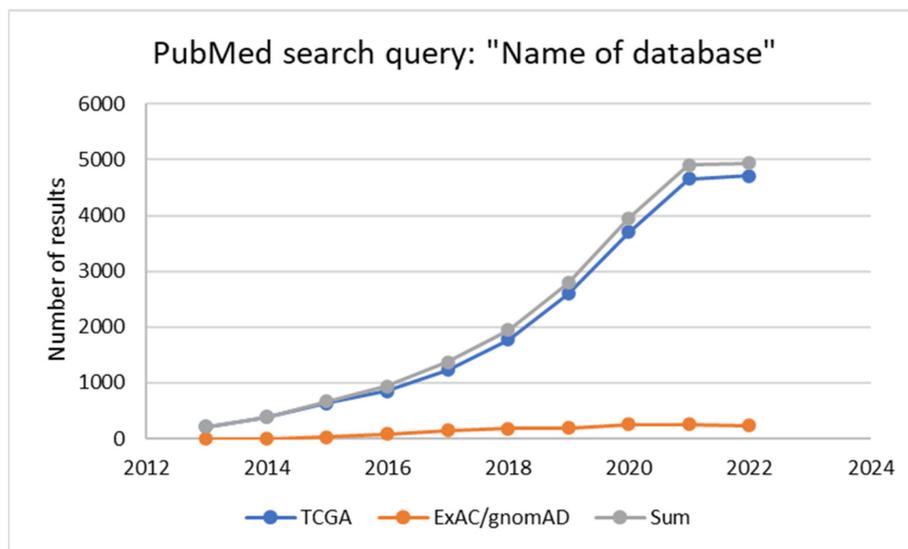
The remaining 99% of the total DNA sequence, colloquially called “dark matter of the DNA” is much less characterised, but also holds disease-relevant changes [10]. Multiple types of RNA produced from those regions regulate gene expression at many levels and this knowledge dramatically changed our understanding of how disease arises and progresses [12,13]. Whole Genome Sequence (WGS)-based analyses of thousands of genomes representing various cancer tissues revealed multiple cancer-driver events localised in non-coding regions of DNA such as promoters, enhancers, or miRNA coding genes [14], to name just a few. Such events include not only single nucleotide variations (SNV) but also small indels and larger structural changes [14]. Although the driver mutations identified in non-coding regions are less frequent (13%) than in protein-coding genes (87%), these variant numbers will grow with more cancer genomes sequenced [14]. Moreover, in terms of exome region sequencing, WGS is more powerful than WES [15], also in terms of detecting structural variants and exome coverage [16]. Fine-tuning cell functioning using this whole new category of personalised therapies, or targeting specific targets in non-coding regions, might have tremendous results, arming us with new powerful tools in intervening and treating human diseases, cancer in particular. Big data analysis is a game-changer, and we can be certain that the remaining unknown chunk of the DNA is important, even if we are at the very beginning of the road to fully understand it.

The number of already existing genomic databases is increasing. The journal *Nucleic Acids Research* (NAR) publishes annually a special issue on Molecular Biology Databases, a considerable proportion of which is related to genomics. The number of NAR’s papers reporting new databases in the last ten years has reached nearly 700 (Figure 1) [17–26]. Most of the databases, except from a few exceptions, contain pooled data not only from genomes, but also from exomes, RNA-Seq or epigenomic data. Additionally, they differ in terms of informed consent given by the participant [27].



**Figure 1.** New databases reported in *Nucleic Acids Research* database issue in the last 10 years.

They find their application in various fields of medicine and their use in research generally increases over time. This phenomenon can be illustrated with a comparison of the number of results of the PubMed search query: “Name of database” in the last ten years for two databases described in the next part of the review dedicated to oncology (Figure 2). The first one is The Cancer Genome Atlas (TCGA). The second one, the Exome Aggregation Consortium (ExAC), transformed into the Genome Aggregation Database (gnomAD), is widely used in the population studies of all types, cancer studies included.



**Figure 2.** Number of results of the PubMed search query “Name of database” for two databases described in the review.

In this review, we present the examples of the existing and currently being developed genomic databases and their possible use. We also discuss challenges and possible limitations to these global efforts and future means of improvement, to be implemented not only in oncology or infectious diseases, but also other areas of medicine. We strongly believe that modern medicine cannot exist without genomics.

## 2. Genomic Databases: Global vs. Local Initiatives

Several efforts focused on creating large aggregating genome databases gathering data across different population of diverse ancestries. Some projects gathered a considerable number of samples, such as the TopMed Program with 53831 genome-sequenced samples [28] or gnomAD with 15,708 genomes in addition to exomes [29] (Figure 3). One of the biggest databases that is publicly accessible gathering both phenotypical and genetic databases is UK biobank, which released WGS data from nearly 200.000 individuals, mainly British, but also from Asian and African origin [30]. On the pan-European level a well-established initiative is 1000 Genomes Project (1kGP), which collected 3202 genomes from diverse ancestries, among them 602 trios [31]. A bigger project, the 1+ Million Genomes Initiative, is coordinating the efforts to provide a proper infrastructure and framework according to local regulations to work on the genomic data [32]. On the global level The National Center for Biotechnology Information (NCBI) curates a large genomic database, Human Genome Resources [33], encompassing small variations, structural genomic changes and information on the relation of genomic variation to human health. Additionally, investigating infectious diseases such as COVID required pooling of the large genetic data from many populations and across diverse ancestries. One such initiative is the COVID Human Genetic Effort (HGE), collecting the genomic data across the populations for the scope of identifying variants influencing the disease course [34–36].

The non-European genomic databases are still underrepresented in the global databases with pooled data and a study from 2009 showed that 96% of genetic data are from studies performed solely on Europeans [37]. However, several of the gaps have been covered by local projects (Figure 3). Recently, very ambitious efforts in terms of WGS databases have been undertaken in Africa and Asia. In Africa a large consortium plans to WGS sequence 3 million genomes [38]. The initiative will be a response to public health challenges, such as parasitic and infectious diseases, especially HIV. The Genome Asia 100K project plans to sequence 100,000 individuals from 12 South Asian, and at least 7 North and East Asian countries [39]. A genome sequencing initiative for Japan has also already been implemented [40]. In the Middle East, more than 7000 individuals have been genome-sequenced in recent years [41]. Importantly, it revealed that polygenic scores (PGS) have reduced predictive performance in the Qatari population, postulating the necessity of including population-specific data in the PGS studies.

Although some genomic data are aggregated as part of the biggest projects, still there is a need to create smaller, population-specific databases that can reveal a different genetic landscape on the local level than a pan-European project. Given the ancestry component in medical genetics, it is important that huge datasets reflect the populations of interest. One such genomic database considering the structure and background of the investigated subjects is PGG.Population [42]. The database gathered >7000 genomes from 356 global populations and is designed for population genetic studies. On the European level, several European countries started their own biobanks, including genomic data. The leaders in this field are Iceland [43], Finland [44] and Estonia [45]; however, the majority of European countries is curating its own genomic database, in addition to their contribution to global efforts [46]. Slavic genomes are often underrepresented [47,48]; therefore, a recent Polish initiative to create a database of 1076 unrelated Poles is an important step forward into including less represented European populations [48]. Several other initiatives for Slavic genomes have been undertaken, including in the Ukraine [47]. We have a personal experience on this project created initially to investigate the course of COVID disease. As the participants consented to the scientific use of their anonymised data, the database can be used as a reference database for the Polish population. The data of the participants that consented might be deposited as part of the Genome of Europe project.

Furthermore, some of the analyses, especially evolutionary events, justify the creation of smaller local databases. A Japanese genomic database revealed evolution traits related to alcohol or nutrition [49]. In Asia local projects sequencing local populations have arisen, such as KOVA 2, collecting almost 2000 genome-sequenced Korean individuals, surpassing, together with WES data collected in KOVA 2, the data on Korean individuals collected in gnomAD [50]. The whole genome sequencing of Ethiopian highlanders delineated genomic regions that may have an impact on hypoxia tolerance [51]. In Europe there is a tendency to create local genomic databases that are especially important in interpreting VUS and in the application of PGS on the local level [46].

	Oncology	COVID-19	Infectious diseases	Other
Local	National biobanks, case-patients cohorts	Thousand Polish Genomes	Whole-genome sequences of pathogens/case-patients	Smaller (specific) populations databases
Global	TCGA, ICGC, ExAC/gnomAD, COSMIC, CGC	HGI, HGE	NCBI database, CARD, VFDB	UK biobank, 1+ Million Genomes, Human Genome Resources (NCBI), GenomeAsia 100K
Possible use	Prevention, detection and individualised treatment	Identification of variants influencing the disease course	Identification of resistance and virulence biomarkers, treatment targets and genetic factors of severity	Finding local genetic adaptations, comparing data in large-scale studies

**Figure 3.** Sources of genomic data (existing databases) described in the review. Note that not all of the databases mentioned above contain WGS-derived data; some of them were created on the basis of other techniques.

### 3. Genomic Databases in Oncology

As previously mentioned, thanks to genomics, our knowledge of cancer biology has expanded considerably in the past few years. The rapid development of genomic research in this field would be impossible without the joint efforts of the scientific community to generate public databases, which have been extensively used as a tool for further studies on various aspects of oncology.

By the end of 2005, the US National Institutes of Health launched The Cancer Genome Atlas (TCGA) project. In 12 years, TCGA characterised more than 20,000 samples from 33 different cancer types, generating over 2.5 petabytes of genomic, epigenomic, transcriptomic and proteomic data [52]. As it soon turned out that characterising a higher number of tumour samples from different cancer types would require international cooperation, the International Cancer Genome Consortium (ICGC) was initiated in 2008 to coordinate large-scale cancer genome sequencing studies in 50 different tumour types “that are of clinical and societal importance across the globe” [53]. It is the most significant cancer genome sequencing project to date. Over 80 million somatic mutations have been identified in this dataset. Both TCGA and ICGC were mainly focused on the exome. However, several studies have shown the important role of non-coding and regulatory regions in carcinogenesis. That is why the Pan-Cancer Analysis of Whole Genomes (PCAWG) initiative within the ICGC was established to identify common patterns of mutation in more than 2600 cancer whole genomes. According to the flagship paper of the TCGA/ICGC PCAWG consortium published in 2020, the majority of cancer genomes contain a few driver mutations in both coding and non-coding regions but in about 5% of them, no known mutation was identified, which leaves room for speculation. Is the catalogue of cancerogenic mutations incomplete or do other processes have more impact in these cases? The new phase of the project Accelerating Research in Genomic Oncology (ARGO) started in 2019. Its main goal is to improve the outcome of cancer patients. It will analyse 100,000 samples in comparison with clinical data to find out how to best use genomic knowledge in the prevention, detection and treatment of cancers [52,54].

Projects, as described above, have enabled the creation of such data collections as the Catalogue of Somatic Mutations in Cancer (COSMIC), the world’s largest and most detailed resource for exploring the effect of somatic mutations in human cancer, and the Cancer Gene Census (CGC) [55,56]. COSMIC covers all the known genetic mechanisms by which somatic mutations promote cancer such as coding and non-coding mutations, gene fusions, copy-number variants, and drug-resistance mutations, whereas CGC is an expert-curated catalogue of the genes driving human cancer that is used as a standard in cancer genetics across basic research, medical reporting and pharmaceutical development. It also includes functional descriptions of how each gene contributes to disease generation [55,56].

Another large database widely used in oncological research, as well as in other domains, is the Genome Aggregation Database (gnomAD), originally launched in 2014 as the Exome Aggregation Consortium (ExAC) [57]. It contains over 125,000 exome and 15,000 whole genome sequences from European, Latino African and African American, South Asian, East Asian, Ashkenazi Jewish and other populations (<https://gnomad.broadinstitute.org> (accessed on 10 December 2022)). All the data were contributed to the project from independent large-scale human sequencing studies led by more than 100 investigators, then processed into summary high-quality variant data and made available for the wider scientific community. The gnomAD papers report 241 million small genetic variants and over 335 thousand structural variants [57]. Even though this database is widely used in oncology, it remains a valuable and broad population database with many significant applications outside medicine too.

In addition to the already listed, numerous smaller, more specific databases have been created. Some of the interesting examples include a database of extrachromosomal circular DNA (eccDNAdb), which seems to play a crucial role in oncogene amplification and tumour progression [58]; single nucleotide polymorphisms (SNPs) databases, (SNPs can influence methylation and participate in signalling pathway degeneracy in cancer) [59]

and upstream open reading frames (uORFs) databases. Genetic defects in the last ones have been linked to the development of various diseases, including cancer [60].

All these resources are used in cancer-related analyses. They allow detection of viral sequences in cancer tissues, e.g., herpesvirus family or HPV in bladder cancers [61]; finding new genetic markers to diagnose and treat diseases with relatively poor prognosis such as liver and oesophageal cancer [62,63]; examining the tumour microenvironment, which is thought to be essential, e.g., for breast cancer progression and metastasis [64].

Beyond questions, the role of non-coding variants in cancer genome is significant and should be incorporated into diagnostic and treatment procedures, which in fact is being preceded by several guidelines-producing bodies including ACMG (for example [65–67]). WGS of cancer genome allows to characterise the whole profile of genetic variants and assign them to a proper cancer signature or specific feature. Each of the more than a hundred signatures identified up to date across human cancers indicates a specific mechanism of cancer development [68]. Most of the signatures can be associated with a defective DNA maintenance process and a precisely pinpointed disrupted pathway, which brings us to the point where specific treatment may be administered on the basis of this information, such as PARP inhibitors.

PARP inhibitors (poly-(ADP-ribose)-polymerase inhibitors) are ground-breaking agents, effective in treating several cancer types including breast, ovarian, prostate and even pancreatic cancer [69–74]. Multiple randomised clinical trials have demonstrated their efficacy and the PARPi drug family constantly expands, comprising such agents as Olaparib, Niraparib, Rucaparib and Talazoparib, with many more under clinical trials around the world [75–77]. However, it remains worrisome that only a subset of cancer patients treated with PARPi actually benefit from the therapy [78,79].

The biomarker currently used for PARPi administration is far from being perfect: the *BRCA1* and *BRCA2* gene mutations [80,81]. Even though they are an excellent indicator of Homologous Recombination Deficiency (HRD), they are not the only hallmarks of HRD disruption [82]. However, clinical trials have clearly demonstrated that patients without *BRCA1* nor *BRCA2* mutations can also benefit from PARPi therapy [83]. For example, the PRIMA trial (PRIMA/ENGOT-OV26/GOG-3012 trial results presented at the European Society for Medical Oncology (ESMO) Congress in 2019) showed that assessing HRD status with the aid of computer algorithms may allow more cancer patients with no *BRCA1* and *BRCA2* mutations to undergo a successful PARPi treatment [83]. Thus, many more patients without *BRCA1* and *BRCA2* do respond to PARP inhibitors and therefore may benefit from the treatment.

In fact, the most advanced clinical application originating from cancer signatures, not only mutated genes, relates to Homologous Recombination Deficiency and PARPi [84,85]. WGS is being used in a couple of commercially available cancer diagnostics; for example, Illumina Comprehensive Genomic Profiling considers Tumour Mutation Burden (TMB) or Microsatellite Instability (MSI) or MyChoiceCDx created by Myriad Genetics Inc. The diagnosis and treatment based on advanced machine learning algorithms, such as HRDetect or myChoice already show promising results: several clinical trials of the drugs based on PARPi (such as Niraparib, most recently) were effective in reducing the risk of ovarian cancer progression by 38% [85–87]. AI-based computer algorithms are created to screen WGS data for rare and common variants potentially significant in pharmacogenomics, leading to new applications of the drugs already existing in the market, but also identification of novel regulatory variants located in non-coding parts of the genome and their function, patient stratification and, in some cases, even the mechanistic prediction of drug targets, response and their interactions [88,89]. Some cancer databases are designed to find patient target genes and potential treating molecules [90]. Although datasets contain various omics datasets, such as mRNA and epigenomics, WGS data are still the core of such databases. As a result, a hit containing a list of potential drugs targeting a particular genetic sequence is returned.

Regional databases also play an important role in cancer research. Numerous studies are focused on specific populations, such as 237 patients from a reported population-based south Swedish triple-negative breast cancer cohort profiled by RNA sequencing and whole-genome sequencing included in “Molecular analyses of triple-negative breast cancer in the young and elderly” or a population-based Estonian biobank (over 150,000) and breast cancer-affected cases from Latvia chosen to assess the spectrum and frequency of CHEK2 variants in the breast cancer-affected and general population in the Baltic states region [91,92].

#### 4. Genomic Databases in Infectious Diseases

The same is true for many other human threats including infectious diseases. It has been long known that not only can we track pathogens’ routes of transmission or evolutionary development, as it has been done for MRSA strains [93,94] or cholera outbreaks in Haiti [95,96], but also genomic regions in human DNA connected with susceptibility or resistance to a certain pathogen, such as norovirus infections [97–99]. More recently, this phenomenon was beautifully depicted by the global cooperation established at the very early days of the COVID-19 pandemic, namely the COVID-19 Host Genetics Initiative (HGI) and the COVID Human Genetic Effort (HGE). These global initiatives aimed at understanding the disease enabled worldwide genomic sample collection, used further by us and others, and resulted in enormous datasets suitable for AI- and ML-based algorithms (exemplified by the HGI and HGE consortia findings described in [100,101]). Such great databases provide evidence that, as a scientific community worldwide, we are already very good at collecting data, but the time has come to share these datasets more eagerly. Especially in case of the genomic datasets, it may not be feasible nor technically doable for a single team to analyse and interpret properly whole genome sequences of such a huge and expanding collection.

It is worth emphasizing that all the genomic data collected during the COVID-19 pandemic can be used not only in the infectious context. Our project “Search for Genomic Markers Predicting the Severity of the Response to COVID-19” may be taken as an example. Between April 2020 and April 2021, we collected samples from 1222 Poles to study their genetic susceptibility to COVID-19 infections. We analysed the whole genomes to identify and genotype a wide spectrum of genomic variation, such as small and structural variants, runs of homozygosity, mitochondrial haplogroups and de novo variants. This study is the biggest whole-genome screening of the Slavic and Central Europe populations done to date. The allele frequencies, calculated for 1076 unrelated individuals, were released as a publicly available resource, the Thousand Polish Genomes database. The Polish population, highly homogenous and sedentary by its nature, is unique and can serve as a genetic reference for the Slavic nations that account for over 4.5% of world inhabitants. The Thousand Polish Genomes database contributes to the worldwide genomic resources accessible to researchers and clinicians. It lays the foundation for further studies in the population history and epidemiology of diseases caused by mutations in the autosomal-recessive genes, as well as creates opportunities for tailoring NGS-based genetic screening tests and guidelines for clinical geneticists in Poland [48].

Genomic databases in infectious diseases can play multiple roles not only in relation to COVID [102]. They may help in identification of resistance biomarkers and treatment targets. This seems to be particularly crucial in Africa, especially for the detection and surveillance of malaria, HIV and drug-resistant tuberculosis [103].

As communicable diseases quite often have a localised character, creating small, local databases might be particularly useful in their case. During the 2019–2020 dengue fever epidemic in the Dominican Republic, a study on 488 children with a confirmed disease was conducted to find the genetic factors of its severity in this group [104]. On the African continent, there is a need to investigate tropical arboviruses with described zoonotic potential. The whole-genome sequencing using novel technological approaches allows a better understanding of their genetic diversity and distribution that may help to reduce the

threat they pose to human and animal health [105]. The large international databases are also frequently used in this domain. In Asia, 10 *Pasteurella canis* and 16 *Pasteurella multocida* whole-genome sequences from National Center of Biotechnology database were selected to perform a comparative analysis of virulence factors (VFs) between two species that both cause zoonotic infections [106]. The collections such as the Comprehensive Antibiotic Resistance Database (CARD) or the Virulence Factor Database (VFDB) are used to identify the genes responsible for drug-resistance or virulence and characterise local pathogens. It was done recently in the case of multidrug-resistant *Staphylococcus hominis* isolated in Malaysia [107].

### 5. Genomic Databases in Rare Diseases

Rare diseases were one of the areas that profited from the WGS technology at first. Moreover, in terms of standardisation and guidelines, WGS in rare diseases is well established [108]. One of the most known and pioneering initiative is 100,000 Human Genomes, a project targeted at sequencing NHS patients affected with rare diseases [109]. The preliminary results gave a diagnostic yield of 35% for likely monogenic disorders and 11% for likely complex disorders [109]. In the US, Centers for Mendelian Genomics are pioneering institutions that use WGS in rare disease cohorts [110]. In Canada a centralized WGS database for rare diseases has been introduced to facilitate cooperation and new gene discovery [111]. On the European level several EU-founded projects, such as Solve-RD and ERN, implemented WGS as part of their workups [112]. In the recent years also regional initiatives, such as the Brazilian Rare Genomes Programme [113] and the Initiative on Rare and Undiagnosed Diseases in Japan [114], have successfully been implemented.

### 6. Genomic Databases as a Fuel for AI-Driven Algorithms

Furthermore, it is worth noting that AI-based methods already are and will remain an integral component of every modern WGS-based procedure. We can develop AI-driven algorithms to extract crucial information from the patient's genome and include them into prediction or prognostic tests even without full understanding of the region itself. This is a major breakthrough seriously challenging our perception of the scientific method. So far, the sequence of diagnostic and therapeutic actions was preceded by deep understanding of the target itself, its structure and function, such as the gene. Using various techniques, we have been studying genes and their role in model organisms for years before transferring this knowledge into human beings. However, now AI-driven algorithms may pinpoint genomic regions or clusters of unknown function, yet crucial for the early prediction, advanced diagnosis, or effective treatment. Although AI technologies have raised an initial enthusiasm, there are also critical voices. For example, in a review article on AI methods in diagnosing COVID 19, the authors found methodological flaws and biases leading to an optimistic performance. The authors advise standardisation of methodology on several levels [115].

Although several multiple large-scale whole genome sequencing projects have been launched globally, and the results obtained so far are important both scientifically and clinically, the clinical implementation of these data is for the most part lagging behind. Most of the projects are focused on rare diseases and clinical WGS was primarily used as a rare disease's diagnostics. Together with whole-exome sequencing (WES), WGS has been introduced into diagnostic procedures of many clinical centres, such as Genomic Medicine Centre Karolinska-Rare Diseases in Stockholm or Genomics England centres across the UK. There are many more similar institutions and programs worldwide focusing on clinical WGS in rare disease diagnostics, such as several National Institutes of Health grant programs in the US, the Clinical Sequencing Exploratory Research Consortium, the Centres for Mendelian Genomics, and the Undiagnosed Diseases Program and Network [110,116]. The advantage of WGS is estimated as the 7.5–30% increase in diagnostic yield. However, WGS seems to be promising because of the diversity of variants detected, difficult to find

using other available methods, including CNVs, balanced structural variants, short tandem repeats and runs of homozygosity [117].

Another challenge is the identification of common and rare disease genetic variants in genome-wide association studies. WGS together with dedicated AI-driven algorithms was shown to increase the mapping precision for rare and low frequency variants. More and more, WGS is being performed in a variety of different populations, supporting the notion that WGS of related cohorts improves the power to identify genetic associations [43,118–121].

The most significant drawback when using AI-harnessing algorithms, apart from the costly IT infrastructure, remains the huge dataset necessary for the proper learning process. Although several mathematical models can overcome this problem, at least to some extent, the reliability of the results and their clinical implementation should be strongly considered and properly validated. Perhaps the next generation of AI-based genomic-analysing tools are required and thus, should emerge from interdisciplinary close cooperation.

Finally, another objection—the time required to complete the process—is no longer valid. Although our project was performed using a “traditional” short-reads approach and Illumina pipeline for the WGS data, there are already other methods which might be better, especially given the narrow time constraints in the case of some rapidly progressing diseases. One of the most amazing examples of interdisciplinary cooperation on the ground of AI-driven tools implementation in clinical genomics was a recent world record in whole-genome sequencing speed, counted from the moment of sample arrival till results delivery. A Stanford University research team led by Dr Euan Ashley, in collaboration with such technological giants as Nvidia, Oxford Nanopore Technologies, Google, as well as the medical world-famous Baylor College of Medicine and the University of California, managed to complete the process in just five hours and two minutes [122].

Although there is a lot of hope in the AI-based methods, too early translation to the clinics may lead to wrong conclusions and failures in treatment. All algorithms are often trained on a single centre’s data and may be biased. For example, surgical skin markings confused a deep learning algorithm for melanoma detection in which it classified benign nevi as malignant [123]. As another example, an AI system recommended “unsafe and incorrect” cancer treatments [124]. Similarly, a sepsis prediction algorithm implemented in a widely used EHR system performed poorly in practice. [125] A special care should be given as most of the projects are targeting clinical research [126].

## 7. Conclusions

Genomic databases are not only collecting data but may also contribute to better diagnostics and therapies. Genomic databases play a special role in infectious diseases, as well as rare or heritable diseases. As a medical community, we should make the most of what we have already achieved in genomics to effectively treat cancer patients. Offering the most advanced diagnostic methods or early detection tests today, we should continuously participate in building more accurate prediction models especially for early detection or targeted therapies.

**Author Contributions:** Conceptualization, P.D., M.M., methodology P.D., M.M., A.Ś., formal analysis: Z.J.K., P.D., data curation: Z.J.K., P.D., writing—original draft preparation: Z.J.K., M.M., A.Ś., P.D., writing—review and editing: A.Ś., M.M., visualization: A.Ś., funding acquisition, Z.J.K. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was partially funded by the The National Centre for Research and Development project Szpitala Jednoimienne/02/2020 “Development of an innovative diagnostic test to assess the course of COVID-19 and post-death complications with the aid of whole-genome analysis”, as well as The Medical Research Agency project 2020/ABM/COVID19/0022 “A clinical trial in the search for genetic markers responsible for the intensity of the course of the COVID-19 disease, with particular emphasis on patients with accompanying cardiopulmonary diseases”.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare that they have no competing interest.

## References

1. Lander, E.S.; Linton, L.M.; Birren, B.; Nusbaum, C. Initial sequencing and analysis of the human genome. *Nature* **2001**, *409*, 860–921. [CrossRef]
2. Lerner-Ellis, J.P. The clinical implementation of whole genome sequencing: A conversation with seven scientific experts. *J. Inherit. Metab. Dis.* **2012**, *35*, 689–693. [CrossRef] [PubMed]
3. Rossen, J.; Friedrich, A.; Moran-Gilad, J. Practical issues in implementing whole-genome-sequencing in routine diagnostic microbiology. *Clin. Microbiol. Infect.* **2018**, *24*, 355–360. [CrossRef] [PubMed]
4. Battke, F.; Schulte, B.; Schulze, M.; Biskup, S. The question of WGS's clinical utility remains unanswered. *Eur. J. Hum. Genet.* **2021**, *29*, 722–723. [CrossRef]
5. Dillon, O.J.; Melbourne Genomics Health Alliance; Lunke, S.; Stark, Z.; Yeung, A.; Thorne, N.; Gaff, C.; White, S.M.; Tan, T.Y. Exome sequencing has higher diagnostic yield compared to simulated disease-specific panels in children with suspected monogenic disorders. *Eur. J. Hum. Genet.* **2018**, *26*, 644–651. [CrossRef]
6. Kerr, K.; McAnaney, H.; Smyth, L.; Bailie, C.; McKee, S.; McKnight, A.J. A scoping review and proposed workflow for multi-omic rare disease research. *Orphanet J. Rare Dis.* **2020**, *15*, 1–18. [CrossRef] [PubMed]
7. Colin, E.; Duffourd, Y.; Tisserant, E.; Relator, R.; Bruel, A.-L.; Mau-Them, F.T.; Denomme-Pichon, A.-S.; Safraou, H.; Delanne, J.; Jean-Marçais, N.; et al. OMIXCARE: OMICS technologies solved about 33% of the patients with heterogeneous rare neuro-developmental disorders and negative exome sequencing results and identified 13% additional candidate variants. *Front. Cell Dev. Biol.* **2022**, *10*, 1021785. [CrossRef]
8. Zhang, C.; Hansen, M.E.; Tishkoff, S.A. Advances in integrative African genomics. *Trends Genet.* **2021**, *38*, 152–168. [CrossRef]
9. Stratton, M.R.; Campbell, P.J.; Futreal, P.A. The cancer genome. *Nature* **2009**, *458*, 719–724. [CrossRef]
10. Carey, N. *Junk DNA: A Journey through the Dark Matter of the Genome*; Columbia University Press: New York, NY, USA, 2015.
11. Futreal, P.A.; Coin, L.; Marshall, M.; Down, T.; Hubbard, T.; Wooster, R.; Rahman, N.; Stratton, M.R. A census of human cancer genes. *Nat. Rev. Cancer* **2004**, *4*, 177–183. [CrossRef]
12. UCSF Health Center for Clinical Genetics and Genomics. UCSF 500 Cancer Gene Panel Test (UCSF500/UC500). Available online: <https://genomics.ucsf.edu/content/ucsf-500-cancer-gene-panel-test-ucsf500-uc500> (accessed on 10 December 2022).
13. Peng, Y.; Croce, C.M. The role of MicroRNAs in human cancer. *Signal Transduct. Target. Ther.* **2016**, *1*, 15004. [CrossRef]
14. Anastasiadou, E.; Jacob, L.S.; Slack, F.J. Non-coding RNA networks in cancer. *Nat. Rev. Cancer* **2018**, *18*, 5–18. [CrossRef]
15. 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]
16. Meienberg, J.; Bruggmann, R.; Oexle, K.; Matyas, G. Clinical sequencing: Is WGS the better WES? *Hum. Genet.* **2016**, *135*, 359–362. [CrossRef]
17. Rigden, D.J.; Fernández, X.M. The 2022 *Nucleic Acids Research* database issue and the online molecular biology database collection. *Nucleic Acids Res.* **2021**, *50*, D1–D10. [CrossRef] [PubMed]
18. Rigden, D.J.; Fernández, X.M. The 2021 *Nucleic Acids Research* database issue and the online molecular biology database collection. *Nucleic Acids Res.* **2020**, *49*, D1–D9. [CrossRef] [PubMed]
19. Rigden, D.J.; Fernández, X.M. The 27th annual *Nucleic Acids Research* database issue and molecular biology database collection. *Nucleic Acids Res.* **2019**, *48*, D1–D8. [CrossRef]
20. Rigden, D.J.; Fernández, X.M. The 26th annual *Nucleic Acids Research* database issue and Molecular Biology Database Collection. *Nucleic Acids Res.* **2018**, *47*, D1–D7. [CrossRef] [PubMed]
21. Rigden, D.J.; Fernandez, X. The 2018 *Nucleic Acids Research* database issue and the online molecular biology database collection. *Nucleic Acids Res.* **2017**, *46*, D1–D7. [CrossRef]
22. Galperin, M.Y.; Fernández-Suárez, X.M.; Rigden, D.J. The 24th annual *Nucleic Acids Research* database issue: A look back and upcoming changes. *Nucleic Acids Res.* **2016**, *45*, D1–D11. [CrossRef] [PubMed]
23. Rigden, D.J.; Fernández-Suárez, X.M.; Galperin, M.Y. The 2016 database issue of *Nucleic Acids Research* and an updated molecular biology database collection. *Nucleic Acids Res.* **2015**, *44*, D1–D6. [CrossRef] [PubMed]
24. Galperin, M.Y.; Rigden, D.J.; Fernández-Suárez, X.M. The 2015 *Nucleic Acids Research* Database Issue and Molecular Biology Database Collection. *Nucleic Acids Res.* **2015**, *43*, D1–D5. [CrossRef] [PubMed]
25. Fernández-Suárez, X.M.; Rigden, D.J.; Galperin, M.Y. The 2014 *Nucleic Acids Research* Database Issue and an updated NAR online Molecular Biology Database Collection. *Nucleic Acids Res.* **2013**, *42*, D1–D6. [CrossRef] [PubMed]
26. Fernandez, X.; Galperin, M.Y. The 2013 *Nucleic Acids Research* Database Issue and the online Molecular Biology Database Collection. *Nucleic Acids Res.* **2012**, *41*, D1–D7. [CrossRef]

27. Ayuso, C.; Millán, J.M.; Mancheño, M.; Dal-Ré, R. Informed consent for whole-genome sequencing studies in the clinical setting. Proposed recommendations on essential content and process. *Eur. J. Hum. Genet.* **2013**, *21*, 1054–1059. [[CrossRef](#)] [[PubMed](#)]
28. Taliun, D.; NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium; Harris, D.N.; Kessler, M.D.; Carlson, J.; Szpiech, Z.A.; Torres, R.; Taliun, S.A.G.; Corvelo, A.; Gogarten, S.M.; et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* **2021**, *590*, 290–299. [[CrossRef](#)]
29. Karczewski, K.J.; Francioli, L.C.; Tiao, G.; Cummings, B.B.; Alfoldi, J.; Wang, Q.; Collins, R.L.; Laricchia, K.M.; Ganna, A.; Birnbaum, D.P.; et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **2020**, *581*, 434–443. [[CrossRef](#)]
30. Halldorsson, B.V.; Eggertsson, H.P.; Moore, K.H.S.; Hauswedell, H.; Eiriksson, O.; Ulfarsson, M.O.; Palsson, G.; Hardarson, M.T.; Oddsson, A.; Jensson, B.O.; et al. The sequences of 150,119 genomes in the UK Biobank. *Nature* **2022**, *607*, 732–740. [[CrossRef](#)]
31. Byrská-Bishop, M.; Evani, U.S.; Zhao, X.; Basile, A.O.; Abel, H.J.; Regier, A.A.; Corvelo, A.; Clarke, W.E.; Musunuri, R.; Nagulapalli, K.; et al. High-coverage whole-genome sequencing of the expanded 1000 Genomes Project cohort including 602 trios. *Cell* **2022**, *185*, 3426–3440.e19. [[CrossRef](#)]
32. European “1+ Million Genomes” Initiative | Shaping Europe’s Digital Future. Available online: <https://digital-strategy.ec.europa.eu/en/policies/1-million-genomes> (accessed on 28 December 2022).
33. Human Genome Resources at NCBI. Available online: <https://www.ncbi.nlm.nih.gov/genome/guide/human/> (accessed on 28 December 2022).
34. Zhang, Q.; Bastard, P.; Liu, Z.; Le Pen, J.; Moncada-Velez, M.; Chen, J.; Ogishi, M.; Sabli, I.K.D.; Hodeib, S.; Korol, C.; et al. Inborn errors of type I IFN immunity in patients with life-threatening COVID-19. *Science* **2020**, *370*, eabd4570. [[CrossRef](#)]
35. Bastard, P.; Rosen, L.B.; Zhang, Q.; Michailidis, E.; Hoffmann, H.-H.; Zhang, Y.; Dorgham, K.; Philippot, Q.; Rosain, J.; Béziat, V.; et al. Auto-antibodies against type I IFNs in patients with life-threatening COVID-19. *Science* **2020**, *370*, eabd4585. [[CrossRef](#)] [[PubMed](#)]
36. Du Bois, H.; Heim, T.A.; Rahman, S.A.; Yagnik, B.; Bastard, P.; Gervais, A.; Le Voyer, T.; Rosain, J.; Philippot, Q.; Manry, J.; et al. Autoantibodies neutralizing type I IFNs are present in ~4% of uninfected individuals over 70 years old and account for ~20% of COVID-19 deaths. *Sci. Immunol.* **2021**, *6*, eabl4340. [[CrossRef](#)]
37. Need, A.C.; Goldstein, D.B. Next generation disparities in human genomics: Concerns and remedies. *Trends Genet.* **2009**, *25*, 489–494. [[CrossRef](#)] [[PubMed](#)]
38. Wonkam, A. Sequence three million genomes across Africa. *Nature* **2021**, *590*, 209–211. [[CrossRef](#)]
39. GenomeAsia100K Consortium; Wall, J.D.; Stawiski, E.W.; Ratan, A.; Kim, H.L.; Kim, C.; Gupta, R.; Suryamohan, K.; Gusareva, E.S.; Purbojati, R.W.; et al. The GenomeAsia 100K Project enables genetic discoveries across Asia. *Nature* **2019**, *576*, 106–111. [[CrossRef](#)] [[PubMed](#)]
40. Yamaguchi-Kabata, Y.; Nariyai, N.; Kawai, Y.; Sato, Y.; Kojima, K.; Tateno, M.; Katsuoka, F.; Yasuda, J.; Yamamoto, M.; Nagasaki, M. iJGVD: An integrative Japanese genome variation database based on whole-genome sequencing. *Hum. Genome Var.* **2015**, *2*, 15050. [[CrossRef](#)]
41. Thareja, G.; Al-Sarraj, Y.; Belkadi, A.; Almotawa, M.; Ismail, S.; Al-Muftah, W.; Badji, R.; Mbarek, H.; Darwish, D.; Fadl, T.; et al. Whole genome sequencing in the Middle Eastern Qatari population identifies genetic associations with 45 clinically relevant traits. *Nat. Commun.* **2021**, *12*, 1–10. [[CrossRef](#)]
42. Zhang, C.; Gao, Y.; Liu, J.; Xue, Z.; Lu, Y.; Deng, L.; Tian, L.; Feng, Q.; Xu, S. PGG.Population: A database for understanding the genomic diversity and genetic ancestry of human populations. *Nucleic Acids Res.* **2017**, *46*, D984–D993. [[CrossRef](#)]
43. Gudbjartsson, D.F.; Helgason, H.; Gudjonsson, S.A.; Zink, F.; Oddsson, A.; Gylfason, A.; Besenbacher, S.; Magnusson, G.; Halldórsson, B.; Hjartarson, E.; et al. Large-scale whole-genome sequencing of the Icelandic population. *Nat. Genet.* **2015**, *47*, 435–444. [[CrossRef](#)]
44. Chheda, H.; Palta, P.; Pirinen, M.; McCarthy, S.; Walter, K.; Koskinen, S.; Salomaa, V.; Daly, M.; Durbin, R.; Palotie, A.; et al. Whole-genome view of the consequences of a population bottleneck using 2926 genome sequences from Finland and United Kingdom. *Eur. J. Hum. Genet.* **2017**, *25*, 477–484. [[CrossRef](#)]
45. Leitsalu, L.; Haller, T.; Esko, T.; Tammesoo, M.-L.; Alavere, H.; Snieder, H.; Perola, M.; Ng, P.C.; Mägi, R.; Milani, L.; et al. Cohort Profile: Estonian Biobank of the Estonian Genome Center, University of Tartu. *Leuk. Res.* **2014**, *44*, 1137–1147. [[CrossRef](#)] [[PubMed](#)]
46. Smetana, J.; Brož, P. National Genome Initiatives in Europe and the United Kingdom in the Era of Whole-Genome Sequencing: A Comprehensive Review. *Genes* **2022**, *13*, 556. [[CrossRef](#)] [[PubMed](#)]
47. Oleksyk, T.K.; Wolfsberger, W.W.; Schubelka, K.; Mangul, S.; O’Brien, S.J. The Pioneer Advantage: Filling the blank spots on the map of genome diversity in Europe. *Gigascience* **2022**, *11*, giac081. [[CrossRef](#)]
48. Kaja, E.; Lejman, A.; Sielski, D.; Sypniewski, M.; Gambin, T.; Dawidziuk, M.; Suchocki, T.; Golik, P.; Wojtaszewska, M.; Mroczek, M.; et al. The Thousand Polish Genomes—A Database of Polish Variant Allele Frequencies. *Int. J. Mol. Sci.* **2022**, *23*, 4532. [[CrossRef](#)]
49. Okada, Y.; Momozawa, Y.; Sakaue, S.; Kanai, M.; Ishigaki, K.; Akiyama, M.; Kishikawa, T.; Arai, Y.; Sasaki, T.; Kosaki, K.; et al. Deep whole-genome sequencing reveals recent selection signatures linked to evolution and disease risk of Japanese. *Nat. Commun.* **2018**, *9*, 1–10. [[CrossRef](#)]

50. Lee, J.; Lee, J.; Jeon, S.; Lee, J.; Jang, I.; Yang, J.O.; Park, S.; Lee, B.; Choi, J.; Choi, B.-O.; et al. A database of 5305 healthy Korean individuals reveals genetic and clinical implications for an East Asian population. *Exp. Mol. Med.* **2022**, *54*, 1862–1871. [[CrossRef](#)]
51. Udpa, N.; Ronen, R.; Zhou, D.; Liang, J.; Stobdan, T.; Appenzeller, O.; Yin, Y.; Du, Y.; Guo, L.; Cao, R.; et al. Whole genome sequencing of Ethiopian highlanders reveals conserved hypoxia tolerance genes. *Genome Biol.* **2014**, *15*, R36. [[CrossRef](#)] [[PubMed](#)]
52. Fonseca-Montaña, M.A.; Blancas, S.; Herrera-Montalvo, L.A.; Hidalgo-Miranda, A. Cancer Genomics. *Arch. Med Res.* **2022**, *53*, 723–731. [[CrossRef](#)]
53. Hudson, T.J.; Anderson, W.; Artez, A.; Barker, A.D.; Bell, C.; Bernabé, R.R.; Bhan, M.K.; Calvo, F.; Eerola, I.; Gerhard, D.S.; et al. International Network of Cancer Genome Projects. *Nature* **2010**, *464*, 993–998. [[CrossRef](#)]
54. Sondka, Z.; Bamford, S.; Cole, C.G.; Ward, S.A.; Dunham, I.; Forbes, S.A. The COSMIC Cancer Gene Census: Describing genetic dysfunction across all human cancers. *Nat. Rev. Cancer* **2018**, *18*, 696–705. [[CrossRef](#)]
55. Tate, J.G.; Bamford, S.; Jubb, H.C.; Sondka, Z.; Beare, D.M.; Bindal, N.; Boutselakis, H.; Cole, C.G.; Creatore, C.; Dawson, E.; et al. COSMIC: The Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res.* **2019**, *47*, D941–D947. [[CrossRef](#)] [[PubMed](#)]
56. About gnomAD. Available online: <https://gnomad.broadinstitute.org/about> (accessed on 10 December 2022).
57. The Genome Aggregation Database (gnomAD). Available online: <https://www.nature.com/immersive/d42859-020-00002-x/index.html> (accessed on 12 December 2022).
58. Peng, L.; Zhou, N.; Zhang, C.-Y.; Li, G.-C.; Yuan, X.-Q. eccDNAdb: A database of extrachromosomal circular DNA profiles in human cancers. *Oncogene* **2022**, *41*, 2696–2705. [[CrossRef](#)] [[PubMed](#)]
59. Samy, M.D.; Yavorski, J.M.; Mauro, J.A.; Blanck, G. Impact of SNPs on CpG Islands in the MYC and HRAS oncogenes and in a wide variety of tumor suppressor genes: A multi-cancer approach. *Cell Cycle* **2016**, *15*, 1572–1578. [[CrossRef](#)] [[PubMed](#)]
60. Manske, F.; Ogoniak, L.; Jürgens, L.; Grundmann, N.; Makalowski, W.; Wethmar, K. The new uORFdb: Integrating literature, sequence, and variation data in a central hub for uORF research. *Nucleic Acids Res.* **2022**, *51*, D328–D336. [[CrossRef](#)]
61. Cantalupo, P.G.; Katz, J.P.; Pipas, J.M. Viral sequences in human cancer. *Virology* **2017**, *513*, 208–216. [[CrossRef](#)] [[PubMed](#)]
62. Liu, W.; Zheng, L.; Zhang, R.; Hou, P.; Wang, J.; Wu, L.; Li, J. Circ-ZEB1 promotes PIK3CA expression by silencing miR-199a-3p and affects the proliferation and apoptosis of hepatocellular carcinoma. *Mol. Cancer* **2022**, *21*, 1–15. [[CrossRef](#)]
63. Kang, N.; Ou, Y.; Wang, G.; Chen, J.; Li, D.; Zhan, Q. miR-875-5p exerts tumor-promoting function via down-regulation of CAPZA1 in esophageal squamous cell carcinoma. *PeerJ* **2021**, *9*, e10020. [[CrossRef](#)]
64. Xu, M.; Li, Y.; Li, W.; Zhao, Q.; Zhang, Q.; Le, K.; Huang, Z.; Yi, P. Immune and Stroma Related Genes in Breast Cancer: A Comprehensive Analysis of Tumor Microenvironment Based on the Cancer Genome Atlas (TCGA) Database. *Front. Med.* **2020**, *7*, 64. [[CrossRef](#)]
65. Miller, D.T.; Lee, K.; Gordon, A.S.; Amendola, L.M.; Adelman, K.; Bale, S.J.; Chung, W.K.; Gollob, M.H.; Harrison, S.M.; Herman, G.E.; et al. Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2021 update: A policy statement of the American College of Medical Genetics and Genomics (ACMG). *Anesthesia Analg.* **2021**, *23*, 1391–1398. [[CrossRef](#)]
66. 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 American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Anesthesia Analg.* **2015**, *17*, 405–424. [[CrossRef](#)]
67. Cristofoli, F.; Sorrentino, E.; Guerri, G.; Miotto, R.; Romanelli, R.; Zulian, A.; Cecchin, S.; Paolacci, S.; Miertus, J.; Bertelli, M.; et al. Variant Selection and Interpretation: An Example of Modified VarSome Classifier of ACMG Guidelines in the Diagnostic Setting. *Genes* **2021**, *12*, 1885. [[CrossRef](#)]
68. Rheinbay, E.; Nielsen, M.M.; Abascal, F.; Wala, J.A.; Shapira, O.; Tiao, G.; Hornshøj, H.; Hess, J.M.; Juul, R.I.; Lin, Z.; et al. Analyses of non-coding somatic drivers in 2,658 cancer whole genomes. *Nature* **2020**, *578*, 102–111. [[CrossRef](#)] [[PubMed](#)]
69. Curtin, N.J.; Szabo, C. Poly(ADP-ribose) polymerase inhibition: Past, present and future. *Nat. Rev. Drug Discov.* **2020**, *19*, 711–736. [[CrossRef](#)]
70. Konstantinopoulos, P.A.; Lheureux, S.; Moore, K.N. PARP Inhibitors for Ovarian Cancer: Current Indications, Future Combinations, and Novel Assets in Development to Target DNA Damage Repair. *Am. Soc. Clin. Oncol. Educ. Book* **2020**, *40*, e116–e131. [[CrossRef](#)] [[PubMed](#)]
71. Nizialek, E.; Antonarakis, E.S. PARP Inhibitors in Metastatic Prostate Cancer: Evidence to Date. *Cancer Manag. Res.* **2020**, *12*, 8105–8114. [[CrossRef](#)] [[PubMed](#)]
72. Teyssonneau, D.; Margot, H.; Cabart, M.; Anonnay, M.; Sargos, P.; Vuong, N.-S.; Soubeyran, I.; Sevenet, N.; Roubaud, G. Prostate cancer and PARP inhibitors: Progress and challenges. *J. Hematol. Oncol.* **2021**, *14*, 1–19. [[CrossRef](#)] [[PubMed](#)]
73. Zhu, H.; Wei, M.; Xu, J.; Hua, J.; Liang, C.; Meng, Q.; Zhang, Y.; Liu, J.; Zhang, B.; Yu, X.; et al. PARP inhibitors in pancreatic cancer: Molecular mechanisms and clinical applications. *Mol. Cancer* **2020**, *19*, 1–15. [[CrossRef](#)] [[PubMed](#)]
74. Chi, J.; Chung, S.Y.; Parakrama, R.; Fayyaz, F.; Jose, J.; Saif, M.W. The role of PARP inhibitors in BRCA mutated pancreatic cancer. *Ther. Adv. Gastroenterol.* **2021**, *14*. [[CrossRef](#)] [[PubMed](#)]
75. Arora, S.; Balasubramaniam, S.; Zhang, H.; Berman, T.; Narayan, P.; Suzman, D.; Bloomquist, E.; Tang, S.; Gong, Y.; Sridhara, R.; et al. FDA Approval Summary: Olaparib Monotherapy or in Combination with Bevacizumab for the Maintenance Treatment of Patients with Advanced Ovarian Cancer. *Oncol.* **2020**, *26*, e164–e172. [[CrossRef](#)]

76. Rose, M.; Burgess, J.T.; O'Byrne, K.; Richard, D.J.; Bolderson, E. PARP Inhibitors: Clinical Relevance, Mechanisms of Action and Tumor Resistance. *Front. Cell Dev. Biol.* **2020**, *8*, 564601. [CrossRef]
77. Murthy, P.; Muggia, F. PARP inhibitors: Clinical development, emerging differences, and the current therapeutic issues. *Cancer Drug Resist* **2019**, *2*, 665–679. [CrossRef]
78. Yi, M.; Dong, B.; Qin, S.; Chu, Q.; Wu, K.; Luo, S. Advances and perspectives of PARP inhibitors. *Exp. Hematol. Oncol.* **2019**, *8*, 1–12. [CrossRef]
79. Kim, D.-S.; Camacho, C.V.; Kraus, W.L. Alternate therapeutic pathways for PARP inhibitors and potential mechanisms of resistance. *Exp. Mol. Med.* **2021**, *53*, 42–51. [CrossRef]
80. Dziadkowiec, K.N.; Gąsiorowska, E.; Nowak-Markwitz, E.; Jankowska, A. PARP inhibitors: Review of mechanisms of action and BRCA1/2 mutation targeting. *Menopausal Rev.* **2016**, *15*, 215–219. [CrossRef]
81. Lee, J.-M.; Ledermann, J.A.; Kohn, E.C. PARP Inhibitors for BRCA1/2 mutation-associated and BRCA-like malignancies. *Ann. Oncol.* **2013**, *25*, 32–40. [CrossRef] [PubMed]
82. Toh, M.; Ngeow, J. Homologous Recombination Deficiency: Cancer Predispositions and Treatment Implications. *Oncol.* **2021**, *26*, e1526–e1537. [CrossRef] [PubMed]
83. González-Martín, A.; Pothuri, B.; Vergote, I.; DePont Christensen, R.; Graybill, W.; Mirza, M.R.; McCormick, C.; Lorusso, D.; Hoskins, P.; Freyer, G.; et al. Niraparib in Patients with Newly Diagnosed Advanced Ovarian Cancer. *N. Engl. J. Med.* **2019**, *381*, 2391–2402. [CrossRef] [PubMed]
84. Alexandrov, L.B.; Kim, J.; Haradhvala, N.J.; Huang, M.N.; Ng, A.W.T.; Wu, Y.; Boot, A.; Covington, K.R.; Gordenin, D.A.; Bergstrom, E.N.; et al. The repertoire of mutational signatures in human cancer. *Nature* **2020**, *578*, 94–101. [CrossRef]
85. Davies, H.; Glodzik, D.; Morganella, S.; Yates, L.R.; Staaf, J.; Zou, X.; Ramakrishna, M.; Martin, S.; Boyault, S.; Sieuwerts, A.M.; et al. HRDetect is a predictor of BRCA1 and BRCA2 deficiency based on mutational signatures. *Nat. Med.* **2017**, *23*, 517–525. [CrossRef] [PubMed]
86. PRIMA Trial Reports Benefit with Niraparib Across Ovarian Cancer Subsets—The ASCO Post. Available online: <https://ascopost.com/issues/september-10-2020-supplement-gynecologic-cancers-almanac/prima-trial-reports-benefit-with-niraparib-across-ovarian-cancer-subsets/> (accessed on 10 December 2022).
87. MyChoice CDx | Myriad Genetics. Myriad Oncology. Available online: <https://myriad.com/oncology/mychoice-cdx/> (accessed on 10 December 2022).
88. Schrijver, I.; Wiita, A.P. Clinical application of high throughput molecular screening techniques for pharmacogenomics. *Pharmacogenomics Pers. Med.* **2011**, *4*, 109–121. [CrossRef]
89. Rezayi, S.; Kalhori, S.R.N.; Saeedi, S. Effectiveness of Artificial Intelligence for Personalized Medicine in Neoplasms: A Systematic Review. *BioMed Res. Int.* **2022**, *2022*, 1–34. [CrossRef]
90. Belizário, J.E.; Sangiuliano, B.A.; Perez-Sosa, M.; Neyra, J.M.; Moreira, D.F. Using Pharmacogenomic Databases for Discovering Patient-Target Genes and Small Molecule Candidates to Cancer Therapy. *Front. Pharmacol.* **2016**, *7*, 312. [CrossRef]
91. Aine, M.; Boyaci, C.; Hartman, J.; Häkkinen, J.; Mitra, S.; Campos, A.B.; Nimeus, E.; Ehinger, A.; Vallon-Christersson, J.; Borg, Å.; et al. Molecular analyses of triple-negative breast cancer in the young and elderly. *Breast Cancer Res.* **2021**, *23*, 1–19. [CrossRef]
92. Pavlovica, K.; Irmejs, A.; Noukas, M.; Palover, M.; Kals, M.; Tonisson, N.; Metspalu, A.; Gronwald, J.; Lubinski, J.; Murmane, D.; et al. Spectrum and frequency of CHEK2 variants in breast cancer affected and general population in the Baltic states region, initial results and literature review. *Eur. J. Med Genet.* **2022**, *65*, 104477. [CrossRef] [PubMed]
93. Larsen, J.; Raisen, C.L.; Ba, X.; Sadgrove, N.J.; Padilla-González, G.F.; Simmonds, M.S.J.; Loncaric, I.; Kerschner, H.; Apfalter, P.; Hartl, R.; et al. Emergence of methicillin resistance predates the clinical use of antibiotics. *Nature* **2022**, *602*, 135–141. [CrossRef] [PubMed]
94. Coll, F.; Harrison, E.M.; Toleman, M.S.; Reuter, S.; Raven, K.E.; Blane, B.; Palmer, B.; Kappeler, A.R.M.; Brown, N.M.; Török, M.E.; et al. Longitudinal genomic surveillance of MRSA in the UK reveals transmission patterns in hospitals and the community. *Sci. Transl. Med.* **2017**, *9*, eaak9745. [CrossRef] [PubMed]
95. Blane, B.; E Raven, K.; Leek, D.; Brown, N.; Parkhill, J.; Peacock, S.J. Rapid sequencing of MRSA direct from clinical plates in a routine microbiology laboratory. *J. Antimicrob. Chemother.* **2019**, *74*, 2153–2156. [CrossRef]
96. Eppinger, M.; Pearson, T.; Koenig, S.S.K.; Pearson, O.; Hicks, N.; Agrawal, S.; Sanjar, F.; Galens, K.; Daugherty, S.; Crabtree, J.; et al. Genomic Epidemiology of the Haitian Cholera Outbreak: A Single Introduction Followed by Rapid, Extensive, and Continued Spread Characterized the Onset of the Epidemic. *Mbio* **2014**, *5*, e01721-14. [CrossRef]
97. Orata, F.; Keim, P.S.; Boucher, Y. The 2010 Cholera Outbreak in Haiti: How Science Solved a Controversy. *PLoS Pathog.* **2014**, *10*, e1003967. [CrossRef]
98. Nordgren, J.; Sharma, S.; Kambhampati, A.; Lopman, B.; Svensson, L. Innate Resistance and Susceptibility to Norovirus Infection. *PLoS Pathog.* **2016**, *12*, e1005385. [CrossRef]
99. Nordgren, J.; Svensson, L. Genetic Susceptibility to Human Norovirus Infection: An Update. *Viruses* **2019**, *11*, 226. [CrossRef] [PubMed]

100. Pairo-Castineira, E.; Clohisey, S.; Klaric, L.; Bretherick, A.D.; Rawlik, K.; Pasko, D.; Walker, S.; Parkinson, N.; Fourman, M.H.; Russell, C.D.; et al. Genetic mechanisms of critical illness in COVID. *Nature* **2021**, *591*, 92–98. [[CrossRef](#)]
101. Andreacos, E.; Abel, L.; Vinh, D.C.; Kaja, E.; Drolet, B.A.; Zhang, Q.; O'Farrelly, C.; Novelli, G.; Rodríguez-Gallego, C.; Haerynck, F.; et al. A global effort to dissect the human genetic basis of resistance to SARS-CoV-2 infection. *Nat. Immunol.* **2021**, *23*, 159–164. [[CrossRef](#)]
102. Liu, Y.-T. Infectious Disease Genomics. In *Genetics and Evolution of Infectious Diseases*; Elsevier: Amsterdam, The Netherlands, 2017; pp. 211–225. [[CrossRef](#)]
103. Inzaule, S.C.; Tessema, S.K.; Kebede, Y.; Ouma, A.E.O.; Nkengasong, J.N. Genomic-informed pathogen surveillance in Africa: Opportunities and challenges. *Lancet Infect. Dis.* **2021**, *21*, e281–e289. [[CrossRef](#)]
104. Simpson, B.N.; Sang, M.E.M.; Puello, Y.C.; Brockmans, E.J.D.; Soto, M.F.D.; Defilló, S.M.R.; Cruz, K.M.T.; Pérez, J.O.S.; Husami, A.; E Day, M.; et al. The 2019–2020 Dengue Fever Epidemic: Genomic Markers Indicating Severity in Dominican Republic Children. *J. Pediatr. Infect. Dis. Soc.* **2022**, piac136. [[CrossRef](#)]
105. Schulz, A.; Sadeghi, B.; Stoek, F.; King, J.; Fischer, K.; Pohlmann, A.; Eiden, M.; Groschup, M.H. Whole-Genome Sequencing of Six Neglected Arboviruses Circulating in Africa Using Sequence-Independent Single Primer Amplification (SISPA) and MinION Nanopore Technologies. *Pathogens* **2022**, *11*, 1502. [[CrossRef](#)] [[PubMed](#)]
106. Yoshida, H.; Kim, J.-M.; Maeda, T.; Goto, M.; Tsuyuki, Y.; Shibata, S.; Shizuno, K.; Okuzumi, K.; Kim, J.-S.; Takahashi, T. Virulence-associated Genome Sequences of *Pasteurella canis* and Unique Toxin Gene Prevalence of *P. canis* and *Pasteurella multocida* Isolated from Humans and Companion Animals. *Ann. Lab. Med.* **2022**, *43*, 263–272. [[CrossRef](#)] [[PubMed](#)]
107. Al-Trad, E.I.; Hamzah, A.M.C.; Puah, S.M.; Chua, K.H.; Kwong, S.M.; Yeo, C.C.; Chew, C.H. Comparative Genomic Analysis of a Multidrug-Resistant *Staphylococcus hominis* ShoR14 Clinical Isolate from Terengganu, Malaysia, Led to the Discovery of Novel Mobile Genetic Elements. *Pathogens* **2022**, *11*, 1406. [[CrossRef](#)]
108. Souche, E.; Beltran, S.; Brosens, E.; Belmont, J.W.; Fossum, M.; Riess, O.; Gilissen, C.; Ardeshirdavani, A.; Houge, G.; van Gijn, M.; et al. Recommendations for whole genome sequencing in diagnostics for rare diseases. *Eur. J. Hum. Genet.* **2022**, *30*, 1017–1021. [[CrossRef](#)]
109. Smedley, D.; Smith, K.R.; Martin, A.; Thomas, E.A.; McDonagh, E.M.; Cipriani, V.; Ellingford, J.M.; Arno, G.; Tucci, A.; Vandrovicova, J.; et al. 100,000 Genomes Pilot on Rare-Disease Diagnosis in Health Care — Preliminary Report. *N. Engl. J. Med.* **2021**, *385*, 1868–1880. [[CrossRef](#)]
110. Bamshad, M.J.; Shendure, J.A.; Valle, D.; Hamosh, A.; Lupski, J.R.; Gibbs, R.A.; Boerwinkle, E.; Lifton, R.P.; Gerstein, M.; Gunel, M.; et al. The Centers for Mendelian Genomics: A new large-scale initiative to identify the genes underlying rare Mendelian conditions. *Am. J. Med Genet. Part A* **2012**, *158A*, 1523–1525. [[CrossRef](#)]
111. Faraji, S.; Patrinos, D.; Hagan, J.; Knoppers, B.M. A centralized rare disease database and whole-genome sequencing as a standard of care: Two essential implementations for the future of health. *Facets* **2021**, *6*, 1831–1834. [[CrossRef](#)]
112. Zurek, B.; Ellwanger, K.; Vissers, L.E.L.M.; Schüle, R.; Synofzik, M.; Töpf, A.; de Voer, R.M.; Laurie, S.; Matalonga, L.; Gilissen, C.; et al. Solve-RD: Systematic pan-European data sharing and collaborative analysis to solve rare diseases. *Eur. J. Hum. Genet.* **2021**, *29*, 1325–1331. [[CrossRef](#)]
113. Coelho, A.V.C.; Mascaro-Cordeiro, B.; Lucon, D.R.; Nóbrega, M.S.; Reis, R.D.S.; de Alexandre, R.B.; Moura, L.M.S.; de Oliveira, G.S.; Guedes, R.L.M.; Caraciolo, M.P.; et al. The Brazilian Rare Genomes Project: Validation of Whole Genome Sequencing for Rare Diseases Diagnosis. *Front. Mol. Biosci.* **2022**, *9*, 821582. [[CrossRef](#)]
114. Takahashi, Y.; Date, H.; Oi, H.; Adachi, T.; Imanishi, N.; Kimura, E.; Takizawa, H.; Kosugi, S.; Matsumoto, N.; Kosaki, K.; et al. Six years' accomplishment of the Initiative on Rare and Undiagnosed Diseases: Nationwide project in Japan to discover causes, mechanisms, and cures. *J. Hum. Genet.* **2022**, *67*, 505–513. [[CrossRef](#)]
115. Roberts, M.; Driggs, D.; Thorpe, M.; Gilbey, J.; Yeung, M.; Ursprung, S.; Aviles-Rivero, A.I.; Etmann, C.; McCague, C.; Beer, L.; et al. Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans. *Nat. Mach. Intell.* **2021**, *3*, 199–217. [[CrossRef](#)]
116. Green, R.C.; Goddard, K.A.; Jarvik, G.P.; Amendola, L.M.; Appelbaum, P.S.; Berg, J.S.; Bernhardt, B.A.; Biesecker, L.G.; Biswas, S.; Blout, C.L.; et al. Clinical Sequencing Exploratory Research Consortium: Accelerating Evidence-Based Practice of Genomic Medicine. *Am. J. Hum. Genet.* **2016**, *98*, 1051–1066. [[CrossRef](#)]
117. Stranneheim, H.; Lagerstedt-Robinson, K.; Magnusson, M.; Kvarnung, M.; Nilsson, D.; Lesko, N.; Engvall, M.; Anderlid, B.-M.; Arnell, H.; Johansson, C.B.; et al. Integration of whole genome sequencing into a healthcare setting: High diagnostic rates across multiple clinical entities in 3219 rare disease patients. *Genome Med.* **2021**, *13*, 1–15. [[CrossRef](#)]
118. Gilly, A.; Suveges, D.; Kuchenbaecker, K.; Pollard, M.; Southam, L.; Hatzikotoulas, K.; Farmaki, A.-E.; Bjornland, T.; Waples, R.; Appel, E.V.R.; et al. Cohort-wide deep whole genome sequencing and the allelic architecture of complex traits. *Nat. Commun.* **2018**, *9*, 1–9. [[CrossRef](#)]
119. Boomsma, D.I.; Wijmenga, C.; Slagboom, E.P.; Swertz, M.A.; Karssen, L.C.; Abdellaoui, A.; Ye, K.; Guryev, V.; Vermaat, M.; van Dijk, F.; et al. The Genome of the Netherlands: Design, and project goals. *Eur. J. Hum. Genet.* **2013**, *22*, 221–227. [[CrossRef](#)]
120. Telenti, A.; Pierce, L.C.T.; Biggs, W.H.; di Iulio, J.; Wong, E.H.M.; Fabani, M.M.; Kirkness, E.F.; Moustafa, A.; Shah, N.; Xie, C.; et al. Deep sequencing of 10,000 human genomes. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 11901–11906. [[CrossRef](#)]

121. Besenbacher, S.; Liu, S.; Izarzugaza, J.M.G.; Grove, J.; Belling, K.; Bork-Jensen, J.; Huang, S.; Als, T.D.; Li, S.; Yadav, R.; et al. Novel variation and de novo mutation rates in population-wide de novo assembled Danish trios. *Nat. Commun.* **2015**, *6*, 5969. [[CrossRef](#)]
122. Gorzynski, J.E.; Goenka, S.D.; Shafin, K.; Jensen, T.D.; Fisk, D.G.; Grove, M.E.; Spiteri, E.; Pesout, T.; Monlong, J.; Baid, G.; et al. Ultrarapid Nanopore Genome Sequencing in a Critical Care Setting. *N. Engl. J. Med.* **2022**, *386*, 700–702. [[CrossRef](#)]
123. Winkler, J.K.; Fink, C.; Toberer, F.; Enk, A.; Deinlein, T.; Hofmann-Wellenhof, R.; Thomas, L.; Lallas, A.; Blum, A.; Stolz, W.; et al. Association Between Surgical Skin Markings in Dermoscopic Images and Diagnostic Performance of a Deep Learning Convolutional Neural Network for Melanoma Recognition. *JAMA Dermatol.* **2019**, *155*, 1135–1141. [[CrossRef](#)]
124. Ross CIBM's Watson Supercomputer Recommended "Unsafe Incorrect" Cancer Treatments Internal Documents Show, S.T.A.T. Available online: <https://www.statnews.com/2018/07/25/ibm-watson-recommended-unsafe-incorrect-treatments/> (accessed on 12 December 2022).
125. Wong, A.; Otles, E.; Donnelly, J.P.; Krumm, A.; McCullough, J.; DeTroyer-Cooley, O.; Pestrue, J.; Phillips, M.; Konye, J.; Penozza, C.; et al. External Validation of a Widely Implemented Proprietary Sepsis Prediction Model in Hospitalized Patients. *JAMA Intern. Med.* **2021**, *181*, 1065–1070. [[CrossRef](#)]
126. Bernstam, E.V.; Shireman, P.K.; Meric-Bernstam, F.; Zozus, M.N.; Jiang, X.; Brimhall, B.B.; Windham, A.K.; Schmidt, S.; Visweswaran, S.; Ye, Y.; et al. Artificial intelligence in clinical and translational science: Successes, challenges and opportunities. *Clin. Transl. Sci.* **2021**, *15*, 309–321. [[CrossRef](#)]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.