Review

Bioinformatics Accelerates the Major Tetrad: A Real Boost for the Pharmaceutical Industry

Tapan Behl [1,*], Ishnoor Kaur [1], Aayush Sehgal [1], Sukhbir Singh [1], Saurabh Bhatia [2,3], Ahmed Al-Harrasi [3], Gokhan Zengin [4], Elena Emilia Babes [5,†], Ciprian Brisc [5,†], Manuela Stoicescu [5,†], Mirela Marioara Toma [6,7,†], Cristian Sava [5] and Simona Gabriela Bungau [6,7,*]

1   Department of Pharmacology, Chitkara College of Pharmacy, Chitkara University, Punjab 140401, India; ishnoorkaur7@gmail.com (I.K.); aayushsehgal00@gmail.com (A.S.); sukhbir.singh@chitkara.edu.in (S.S.)
2   Amity Institute of Pharmacy, Amity University, Gurugram 122413, India; sbsaurabhbhatia@gmail.com
3   Natural & Medical Sciences Research Centre, University of Nizwa, Birkat Al Mauz, Nizwa 616, Oman; aharrasi@unizwa.edu.om
4   Department of Biology, Faculty of Science, Selcuk University Campus, 42130 Konya, Turkey; biyologzengin@gmail.com
5   Department of Medical Disciplines, Faculty of Medicine and Pharmacy, University of Oradea, 410073 Oradea, Romania; babes.emilia@gmail.com (E.E.B.); brisciprian@gmail.com (C.B.); manuela_stoicescu@yahoo.com (M.S.); cristian.sava2004@gmail.com (C.S.)
6   Department of Pharmacy, Faculty of Medicine and Pharmacy, University of Oradea, 410028 Oradea, Romania; mire.toma@yahoo.com
7   Doctoral School of Biomedical Sciences, University of Oradea, 410087 Oradea, Romania
*   Correspondence: tapanbehl31@gmail.com (T.B.); simonabungau@gmail.com (S.G.B.)
†   These authors have equal contribution to the first author.

Abstract: With advanced technology and its development, bioinformatics is one of the avant-garde fields that has managed to make amazing progress in the pharmaceutical–medical field by modeling the infrastructural dimensions of healthcare and integrating computing tools in drug innovation, facilitating prevention, detection/more accurate diagnosis, and treatment of disorders, while saving time and money. By association, bioinformatics and pharmacovigilance promoted both sample analyzes and interpretation of drug side effects, also focusing on drug discovery and development (DDD), in which systems biology, a personalized approach, and drug repositioning were considered together with translational medicine. The role of bioinformatics has been highlighted in DDD, proteomics, genetics, modeling, miRNA discovery and assessment, and clinical genome sequencing. The authors have collated significant data from the most known online databases and publishers, also narrowing the diversified applications, in order to target four major areas (tetrad): DDD, anti-microbial research, genomic sequencing, and miRNA research and its significance in the management of current pandemic context. Our analysis aims to provide optimal data in the field by stratification of the information related to the published data in key sectors and to capture the attention of researchers interested in bioinformatics, a field that has succeeded in advancing the healthcare paradigm by introducing developing techniques and multiple database platforms, addressed in the manuscript.

Keywords: bioinformatics; pharmacovigilance; microbiology; microRNA; COVID-19; public health

1. Introduction

Bioinformatics is a term that refers to the collection and evaluation of scientific data by employing computational techniques, integrating the biological information comprising of proteins, genes, cells, robotics, medical information, and ecosystems with technological mediums such as databases, software, tools, etc. [1–3]. Bioinformatics was completely established as a significant field by the 1990s, with an established role in the scientific paradigm [4]. Numerous economical bioinformatics tools have been established as a result

of accelerating technological advancements [5]. The mRNA profiling and genomic sequencing techniques have transformed the disease detection and management approaches [3]. The conduct of clinical trials requires significant time and massive capital investment in introducing a drug into the market, thereby establishing a dire need for an economical approach to discover and develop drugs, followed by their overall assessment and evaluation [6].

The pharmacological therapies are available for only 30% of the diseases identified, as per the investigations [7], and many biological targets for numerous diseases are yet to be identified [8]. Bioinformatics integrates biostatistical aspects and computational techniques with biomedical sciences, like genetics, proteomics, epidemiology, and genomics [9]. Bioinformatics aims to enable data analysis, management, and interpretation from observational investigations and biological studies, therefore targeting implementation and development of databases, biological inference and interpretation, as well as data evaluation and mining, which is very well managed by the National Center for Biotechnology Information (NBCI), European Bioinformatics Institute (EBI) [9], Riken (Japanese National Research and Development Agencies), etc.

Bioinformatics has been introduced in the pharmaceutical industry on the basis of two significant aspects, computational chemistry and molecular biology, where the latter is a specific and targeted approach, focusing on genetics and sequencing of data, which is a necessary biomedical tool, while the former emphasizes structural analysis more than biological applications, comprising crystallographic techniques, protein-structure determination, etc. [10]. Molecular biologists adopt a sequence-based approach to bioinformatics, whereas computational chemists employ bioinformatics in terms of protein structures [10]. In the pharmaceutical industry, bioinformatics has occupied a fundamental position by facilitating the management and organization of novel developments and sectors, such as pharmacovigilance, which provides a significant basis for drug development and research, and enables the identification, evaluation, prevention, and understanding of the adverse effects or reactions as a result of drug administration [11]. "Biomedical informatics" is defined by the American Medical Informatics as an interdisciplinary approach developed to evaluate and understand the utilization of information and data of biomedical importance, along with knowledge for scientific inquiry, decisive ability, and analytical thinking to solve problems, synergized by steps to enhance human health [12–16]. The techniques associated with bioinformatics have been found to benefit the objectives of the pharmaceutical and medical sector, due to the availability of large-scale datasets, such as methodologies of collecting useful safety information from newer resources, such as electronic health records [12]. Pharmacovigilance has been reported to be influenced by basic principles of bioinformatics, such as storage, decision making, computational skills, data generation, retrieval, use, communication, or sharing, which enable the production of a suitable framework to promote the organization of events associated with pharmacovigilance [13]. The healthcare system has reached the so-called "third wave" of digitalization, according to McKinsey, due to the application of innovative techniques, which are patient-centric, targeting the development of required services, which further propagates more intense innovations in pharmacovigilance and the healthcare sector.

Bioinformatics has been found to play a significant role in case processing in the pharmaceutical industry, using computational collection of data, followed by entry of the complete data, evaluation of expectedness and casualty, quality control (QC), narrative writing, and submission of reports, as well as storage and maintenance of data [14,15]. Additionally, information related to drug safety management, summary of clinical safety information, updated data related to global drug safety regulations, participation in meetings and training sessions of operational staff on issues related to drug safety, adverse drug reaction (ADR) reporting to the regulatory agencies, QC work of other staff, and taking over a task given by the medical director or the manager within the abilities of the drug safety associate are appropriately carried out [14,16].

The present review enlists the tools and databases based upon bioinformatics, in tabular form throughout the body, in order to simplify the applications for the readers, thereby maintaining the interest and relevance of the review throughout. It also throws light on the significance of technical approaches to the management of the current situation of coronavirus disease (COVID-19). Moreover, this review proceeds to attract the focus of global researchers towards the functional aspects of bioinformatic strategies and invokes the need to establish awareness about this field in the healthcare sector in order to accelerate the potential and rate of advancements to enable rapid diagnosis, screening, treatment, and prevention of disorders, along with better and reliable management of problems related to public health.

2. Methodology

The review has incorporated comprehensive data collected from multiple online platforms, publishers, and databases (i.e., Pub Med, MDPI, National Center for Biotechnology Information–NCBI, Google Scholar, Elsevier, Springer, Research Gate, etc.), after thorough research, associated with the role of bioinformatics in the healthcare sector. Competent selection criteria have been considered by selecting articles targeting the role of bioinformatics in the pharmaceutical and medical sectors in particular to provide a localized approach in the present era. The flow chart presented in Figure 1 describes the selection methodology of the articles.
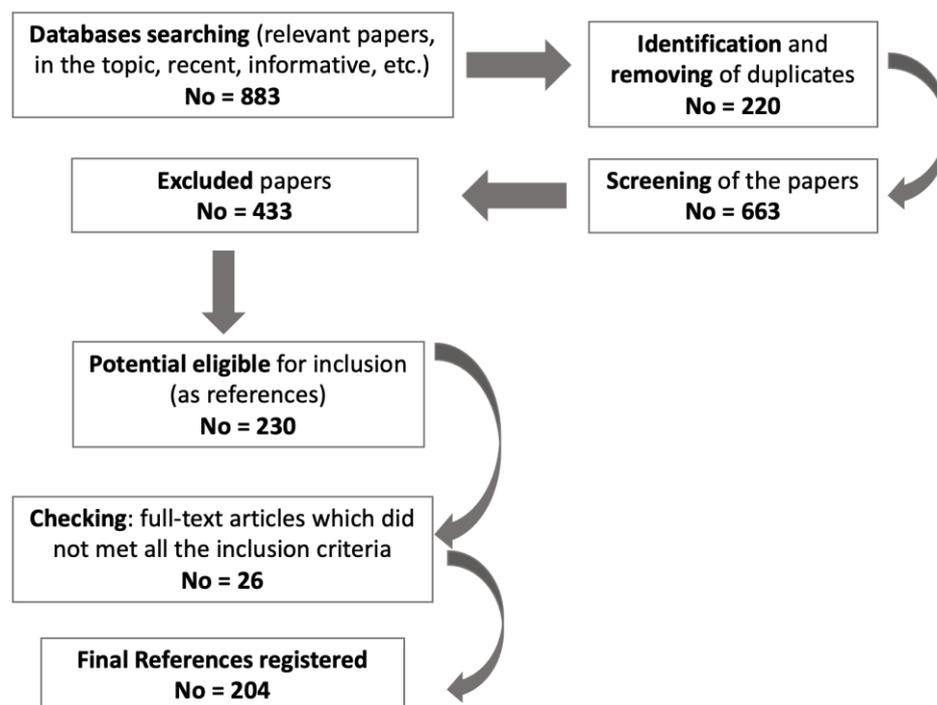

Figure 1. Flow chart presenting the methodology of published data selection.

The authors have referred to multiple articles from some specific journals in bioinformatics (Briefings in Bioinformatics; Network Modeling Analysis in Health Informatics and Bioinformatics; Proteins: Structure, Function, and Bioinformatics; Journal of Biomedical Informatics, Bioinformatics Database, etc.) with year of publication primarily ranging from 2010 to 2021; also, they aimed to design a thorough review, with an objective to expound the existing and future trends of bioinformatics tools and techniques in the pharmaceutical and medicinal sectors, beginning from their descriptive role in the pharmaceutical industry and pharmacovigilance, following which its applications in prime areas has been elaborated. The role of bioinformatics was segregated into four major areas, forming a tetrad in the

healthcare sector as follows: drug discovery and development, anti-microbial research, genomic sequencing, and miRNA research and COVID-19 management.

3. Bioinformatics as an Asset of the Drug Discovery and Development (DDD) Paradigm

The selection of appropriate macromolecular targets is an important facet of the drug discovery process, which is performed after complete knowledge of the disease, which is followed by screen building and selection of lead compounds [10]. Medicinal chemistry approaches are then used for the optimization of leads to develop compounds that can be propagated to clinical trials. The first step of information gathering considerably employs computational methods, which comprise access to online literary databases, in-house collection of chemical structures, and data related to screening and biological assay [17]. In the laboratories, where the access to such tools is limited to just the specialists, bioinformatics approaches can catalyze the flow of these tools towards the bench scientists as well. Prior to the target identification, preliminary investigations are carried out, which comprise experimentation, study of biochemical processes, disease phenotypes, genome information, genetic linkage data, etc., using public domain software and other bioinformatics approaches [18]. A significant advantage is that these processes are dynamic and data collection is rapid, allowing the query to be repeated periodically to obtain complete and necessary information via systems referred to as alerting systems, which are moving forward into literature searching and bioinformatics, contributing to drug discovery [10]. Furthermore, efficient tools should be made available to facilitate the construction of computational models for significant processes by using selected and retrieved data from multiple sources [19]. Bioinformatics shortens the time taken by the preliminary stage of the drug discovery process by enabling browsing of large quantities of systemically evaluated genomic data on the desktop [10].

Study of the mode of action of compounds, resulting from screens, is enabled by the identification of well-defined targets and genes associated with the disease that can be performed by sequence analysis and methods of computational molecular biology, alongside multiple databases of sequences and programs that employed comparison, analysis, and prediction of properties of sequences [20]. Moreover, once the targets and their functions are identified, it is necessary to exhibit their validation as targets. This is enabled by generating computational models associated with the biochemical pathways of the genes/targets in question, which can further make it possible to recognize the most suitable point for therapy [10]. Multiple gene expression techniques can further help to differentiate the pattern of gene expression profiles between healthy and diseased states, which can further be modeled to promote the effect of intervention at multiple points. The following stage promotes the use of a target, which is validated, in a high throughput screen (HTS) to facilitate the recognition of lead compounds by adapting either an existing assay or developing a new one, which requires significant molecular biology and other bioinformatics approaches [10]. Bioinformatics can be significantly employed to guide the chemists involved in the synthesis of combinatorial libraries, which focuses on the appropriate chemical properties of macromolecular targets, to promote the identification of small molecules with relevant activity from the screen. The structural information, available for the target in question, decides the contribution of bioinformatics to the process of rational designing, where the significant properties of the molecules should be evaluated to define their significance and potential in medicine. These properties include toxicity profile, selectivity, potency, and pharmacokinetic parameters [10]. Figure 2 represents the steps in the process of drug discovery and the events that follow.
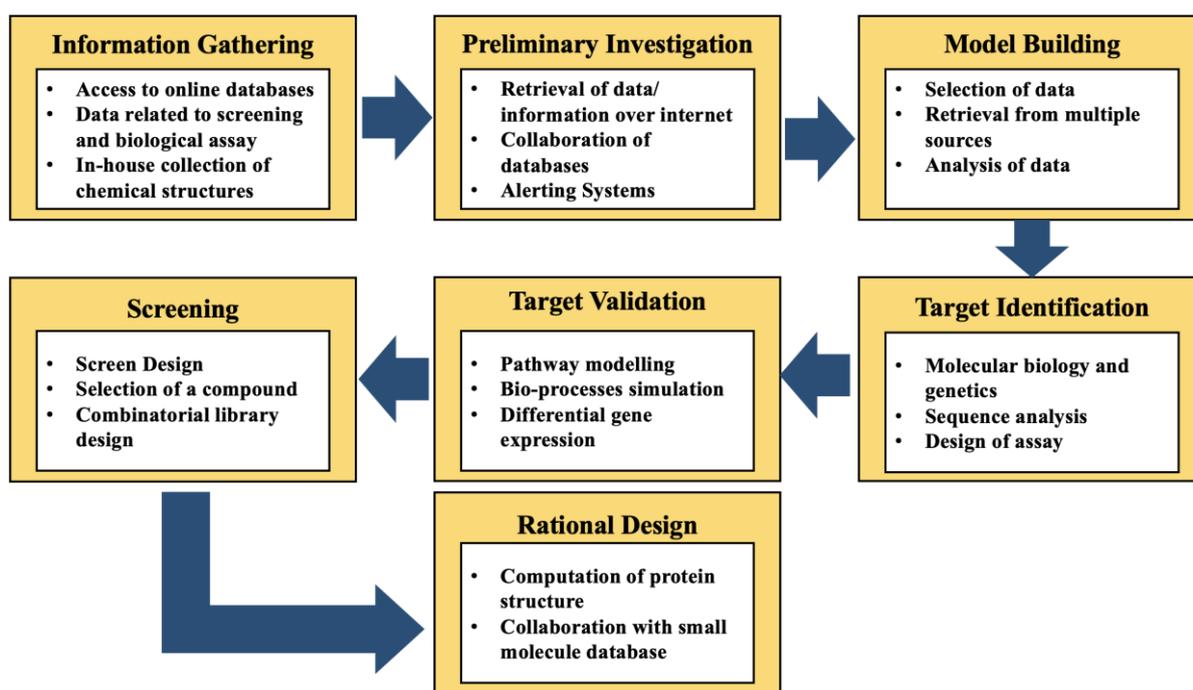
Figure 2. The multiple stages of drug discovery and the events targeted in each of the discussed steps.

The research process transits to the development phase when it is no longer possible to further improve or modify the molecule, where the evaluation of safety and efficacy parameters become a significant consideration, prior to becoming a medicine. Bioinformatics is also a pivotal tool in the development phase, in addition to the research, where it accelerates the understanding of the genetic profile of the individuals receiving the drug before its administration, which would make it necessary to genotype the subjects to be sure that they will benefit from the medicine [10].

3.1. Translational Bioinformatics and Systems Biology in DDD

The discovery and development of newer drugs with greater efficacy in complex disorders is a significant application of translational bioinformatics and systems biology [21]. Certain drawbacks are associated with conventional methods of DDD (post-market surveillance and clinical trials), such as the small size of samples, limited clinical data, and biased analysis [22]. Such challenges and issues can be resolved by progressive advancements in high-throughput analysis (HTP) and functional genomics by introducing better and more convenient techniques in the processes of drug designing [23]. A prospective and retrospective evaluation occurs as a result of applying computational and experimental methods in systems biology, facilitating the identification of more efficacious and personalized options in medications [23].

The complex network of protein–protein interactions is called the "interactome" [24], which can be elucidated by systems biology approaches, such as translational bioinformatics, and HTP technology of multiparametric data sets, in order to aid the development and targeting of more reliable drug candidates. Such techniques have exhibited useful applications in portraying the interactive motifs in disorders such as neuronal degeneration diseases [25]. Furthermore, the cellular alterations can be examined and described in detail at the cellular level using high content screen-based systems biology strategies [26], where the information obtained can be collated in proteomic and transcriptomic profiles, along with the relationship to phenotypic disorders and drug reactions.

Certain obstacles need to be overcome in translational bioinformatics in order to facilitate effective drug discovery, such as execution and compilation of multiple types of data from studies associated with systems biology, the interrelationship among complex

attributes and entities, and the need for data mining with greater efficiency and more reliable decision support tools, as well as more comprehensive data designs [21]. From the "translational" perspective, the strategies of systems biology can be employed in the process of drug discovery by wrapping the network-depending designs, targeting molecular and cellular interactions, which may facilitate more rapid target validation and ameliorated attrition rates [27]. This system-based strategy may be useful in the discovery of multidrug therapies and combination of drugs [21].

Drug repositioning is a convenient tool for drug discovery that, with the identification of patient subgroups as well as novel combinations of conventional drug candidates, may enhance the effectiveness of therapy and abbreviate the zooming expenses in the healthcare system [21]. The drug repositioning approach can be supported by the translational bioinformatics approach to facilitate analysis of transcriptomic data and identification of the biological relationship between drug candidates and health disorders, which is a pivotal step towards personalized medicine [21]. Such systems-based designs may also be used in the treatment of cancer and cardiovascular disorders by targeting the network interactions in migration and proliferation of cells, as well as drug resistance [21].

Further, from a "bioinformatics" perspective, effective management of a massive number of drug interactions and combinations is a significant task. More reliable screening and decision-making for combination of drugs are enabled by integrated applications of HTP technologies, systems biology, and translational bioinformatics approaches [21]. Therefore, systems biology and bioinformatics techniques account for multilevel designing of functional genomics and signaling pathways [28]. Strategies, such as models based upon statistical association, would be applicable for recognizing biomarkers, depending upon the network and "interactome" signatures for optimum targets, further enabling computer-aided drug design (CADD) and poly-pharmacological approaches [29].

3.2. Discussion of Methods, Tools, and Databases Based on Translational Bioinformatics

Certain bioinformatics approaches and databases can be employed in translational studies associated with drug discovery and development, such as absorption of the drug, distribution profile, metabolic processes, excretion, and toxicity profile (ADMET) along with an integrated database for ADMET, as well as adverse effect predictive modeling (IDAAPM), which provides predictive designing, alongside evaluation and analysis of drug information provided by the Food and Drug Administration (FDA) [29]. The analysis of drug targets and discovery is found to be promoted by the drug-minded protein interaction database (DrumPID) [30]. Significant information related to structural inter-atomics is collected by CREDO to aid in the development of drugs [31]. Moreover, a protein–drug interaction database (PDID) is a database associated with an interactive relationship between proteins and drugs in the human proteome [32]. A manually and carefully selected database, the Orphan Nuclear Receptor Ligand Binding Database (ONRLDB), employs ligand moieties for orphan nuclear receptors for drug modeling [33]. A platform integrates genomic information with drug response data, known as the Mutations and Drug Portal (MDP) [34]. A Virtually Aligned Matched Molecular Pairs Including Receptor Environment (VAMMPIRE) database is associated with similar molecular pairs to aid in the development of drug models as well as optimization based upon structural elucidations [35].

Other databases also exist such as the cancer drug resistance (CancerDR) database (associated with cancer drug resistance), ChEMBL (drug model promoting large-scale bioactivity database), the metabolism and drug interaction database (associated with the evaluation of drug interactions), and PROMISCUOUS (concerned with drug repositioning based upon network) [36]. The treatment portfolio primarily focuses on the later stages of pathology with low predictive values and potential rates [37], but investigations associated with environmental and genomic interactions facilitate the development of more reliable preventive approaches. This process is aided by translational bioinformatics, which is collated with the clinical drug discovery and research data towards the development of a personalized approach to medicine [21]. The datasets for disease classification, discovery

of biomarkers, drug targeting, network modeling, and drug repositioning can be effectively organized by translational bioinformatics. The disease models, based upon multiscale networks (possessing great predictive values), can be collated with genomic data related to clinical characteristics and other aspects [21]. The prime driving agents and biomarkers of the disease can be recognized by causal network interference models [37].

It is noteworthy that the information from proteomics and transcriptomics data, as well as HTP technologies, is effectively interpreted by data integration, which forms a potential aspect of computational drug designing and translational bioinformatics [38]. Furthermore, data warehouses, federated databases, as well as semantic technology would promote retrieval of data, discovery of novel drugs, and clinical screening of disease. Not only the biomedical aspects but also the translational bioinformatics approaches aid in overpowering the challenges associated with the pharmaceutical industry [39]. The methylated genes related to drug resistance associated with ovarian cancer were subjected to integrative analysis [40]. Bioinformatics studies focused on protein interactions, annotations, and enrichment of biological processes. The investigation depicted a blunt correlation between the phosphatase and tensin (PTEN) homolog gene, along with other genes, targeting the primary regulatory responsibilities of PTEN [21].

Furthermore, the investigation portrayed the importance of genes with the methyl group in the management of resistant ovarian cancer. Such an outcome might further aid in the prognosis of ovarian cancer [21]. Moreover, in the case of off-label selection of drug amongst patients suffering from triple-negative breast cancer (TNBC), a personalized medication approach was developed, along with collation of cancer drugs, knowledge resources and drug target databases, supporting selection of targets, followed by analysis of patient information in the Cancer Genome Altar [21]. A bioinformatics-based approach was established for the selection of tumor drugs, via involving personal molecular profile data, such as mutations, copy number alterations as well as genomic expressions. The investigation recognized certain extra targets, such as protein tyrosine kinase 6 and gamma-glutamyl hydrolase, which had not been fully studied in TNBC patients [21].

The genetic expression profiles in osteoporosis were developed from information in Gene Expression Omnibus, in an investigation associated with osteoporosis-related drug targets [41]. A classical t-test method has been employed for the analysis of differentially expressed genes (DEGs). The categories of dysregulated gene ontology and dysfunctional pathways were identified by the functional pathway enrichment analysis [21]. The compounds that promoted inverse gene alterations were identified using a connectivity map. The investigation revealed enrichment of DEGs in nine pathways, including the mitogen-activated protein kinase (MAPK) signaling pathway. Additionally, sanguinarine was reported to be a potent therapeutic drug [21].

The translational bioinformatics-based strategies have been reported to be applicable for repositioning of drugs associated with organ transplantation, approved by the FDA [42]. The molecular processes were identified by applying meta-analysis of genomic information and drug databases, along with bioinformatics approaches. For instance, crucial enhancement was reported for the interleukin-17 (IL-17) process [21]. The techniques constituted analysis as well as microarray data set profiling from allograft biopsies in human kidneys, which permitted a drug repositioning technique by employing the drugs available, which would abbreviate the costs [21].

Translational bioinformatics approaches may also aid in the discovery of drugs and repositioning of existing drug candidates, in the case of inflammatory bowel disease (IBD) and other autoimmune disorders [43]. Identification of gene and microRNA biomarkers is enabled by employing bioinformatics techniques and HTP computations [21]. The drugs for repositioning can be discovered with the help of appropriate gene-level profiling of IBD subtypes, on a clinical level, and their relationship with autoimmune disorders. The IBD genes (highly expressed) might suitably target drugs for viral infections, gastrointestinal tract (GIT) cancers, and autoimmune disorders [21].

BioHCVKD is a bioinformatics tool and knowledge discovery system that was developed for the annotation and mining of suitable information, in a study, related to hepatitis C virus (HCV) [44], that collaborated with the conditional random field-based gene mention tagger and dictionary-based filtering and may aid in the identification of ligands, active residues, and proteins to potentiate the drug discovery process. A bioinformatics approach, based upon signal processing, was employed in the examination of protein residues, in the evaluation of therapies for human immune-deficiency virus/acquired immune-deficiency syndrome (HIV/AIDS) and drug resistance, by applying digital signal processing methods, such as the informational spectrum method (ISM) [45], which collated ISM, information associated with protein sequences, as well as other suitable data.

The digital technique to evaluate resistance towards a drug can be employed in other drug resistance studies to develop a computer-aided drug resistance calculator. Databases can be developed to interpret molecular associations in order to evaluate drug abuse and Neuro-AIDS [46]. Effective evaluation of gene expression interactions can aid in better understanding the depth and intensity of the problems. The potent database systems may constitute large data sets, as well as function as a database for a public domain, to facilitate queries, deposition, and review of data [21].

Numerous public databases, such as GLYCAN, Kyoto Encyclopedia of Genes and Genomes (KEGG), Consortium for Functional Glycomics, and glycoSCIENCES.de, can be employed for glycome informatics, which has been reported to aid in the analysis of data associated with the structure of the glycan [21]. Table 1 enlists the methods of translational bioinformatics for drug discovery and development.

Table 1. Techniques of translational bioinformatics for discovery and development of drugs.

Diseased States	Translational Bioinformatics Techniques	Refs.
Osteoporosis drug targets	Functional pathway enrichment, genetic expression profiles from GEO, dysfunctional pathways	[41]
Neuro AIDS and drug abuse	A public domain database, molecular relationship-evaluating database	[46]
Repositioning of drugs	Evaluation of transcriptomic information for relationship between drugs and disease	[47]
Drug resistance in ovarian cancer therapy	Evaluation of protein interactions, evaluation of methylated genes, related to drug resistance, enrichment of biological process	[40]
Drugs for AIDS and drug resistance	Calculator of resistance to drugs, evaluation of residues of protein with digital signaling processing	[45]
Repositioning of drugs in transplantation of organs	Microarray dataset profiling, meta-analysis of genomic drugs as well as information, recognizing redundant molecular processes	[42]
HCV drug discovery	Collation of filtering, based upon dictionary, and gene mention tagger, knowledge discovery process for literature mining	[44]
Off-label selection of drugs for TNBC	Evaluation of TNBC patient information, collaboration of databases of cancer drugs and respective targets, evaluation of personal molecular profiles	[48]
Glycomics and drug targets	Tree-based algorithmic models for glycan structure data, collation of data, public databases for glycome informatics, such as KEGG	[49]

Legend: AIDS—acquired immune-deficiency syndrome; GEO—gene expression omnibus; HCV—Hepatitis C virus; KEGG—Kyoto Encyclopedia of Genes and Genomes; TNBC—triple-negative breast cancer.

4. Optimization of Anti-Microbial Research by Bioinformatics Approaches

The research in the antimicrobial paradigm has resulted in the discovery of potential anti-microbial drug candidates; however, the elevating number of antimicrobial resistance bacteria has triggered the need to develop more efficient and novel antimicrobial drug candidates. The novel genetic data can cause changes in the protein structure, which impacts the ability to carry antibiotics, enzyme-mediated inactivation of drugs, and structural alterations during interactions between bacteria and drugs [50]. Furthermore, numerous natural compounds can be used to fight against such infections, on account of their antimicrobial properties, which are referred to as antimicrobial peptides (AMPs).

Bioinformatics-associated advancements in bacterial transcriptome provide a greater understanding of varying microbial adaptations in conditions of environmental stress, which will aid in the development of novel AMPs [50]. The tools and techniques associated with bioinformatics ameliorate and short-list the total number of lead candidates to be employed as drugs and recognizes the efficient therapeutic agents. Additionally, there are numerous forms of advantages of bioinformatics in the microbiology field [50] (Figure 3).
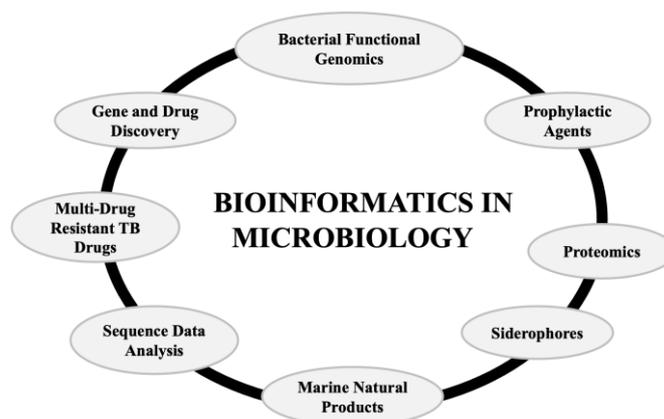

Figure 3. Bioinformatics in multiple assets of microbiology—proteomics, bacterial functional genomics, gene and drug discovery, siderophores, marine natural products, sequence data analysis, multi-drug-resistant tuberculosis (TB) drugs, and prophylactic agents.

Amongst the prime areas of bioinformatics, metagenomic shotgun sequencing (MSS) is rapidly growing, in association with mathematics, biology, and computational techniques [50]. Shotgun sequencing can be employed to attain high-resolution taxonomic composition as well as genetic profiles of metagenome samples [50]. MSS has reported limited gene richness and retarded butyrate-producing bacteria in the gut microbiome of individuals with obesity [51]. Rapid taxonomic profiling of metagenomic data is done by a bioinformatics tool, MetaPhlAn [52], by employing a database of particular marker genes of taxonomical importance, shortlisted from 3000 microbial reference genomes. The functional profiling of metagenomics information is in high demand to overcome the reduced specificity and sensitivity due to long runs in the mapping of deoxyribonucleic acid (DNA) reads and pseudohits to non-associated proteins [50].

An online service was provided by Argonne National Laboratory, namely, metagenomics rapid annotation using subsystem technology (MG-RAST), which facilitates automatic metagenomic functional profiles [53]. DNA read alignment to protein database is enabled by DIAMOND, which exhibits 20,000 times greater speed compared to the basic local alignment search tool (BLAST) × on short reads, where the sensitivity of the two is the same [54]. The protein–protein alignments are enabled by MMseq2, which is 400 times more rapid as compared to PSI-BLAST [55].

Novel approaches by employing rapid DNA–DNA alignment with Bowtie 2 [55–57] to the metal gear solid (MGS) database, resulting in more specific alignments. Alignment of metagenomic reads is enabled by HUMAnN2, to NCBI UniRef microbial genomes [50]. Another metagenomic tool, MGS-Fast, aids in DNA alignment, and its gene annotation is related to KEGG and IGC [50]. Multiple operational taxonomic units (OTU), which depict bacteria, which are uncultured, are developed by direct polymerase chain reaction (PCR)-amplified 16S gene sequencing [50]. About 1,719,541 16S rRNA sequences of bacteria are contained in rRNA database, SILVA, with 99% level of identity into 645,151 representative sequences [56].

Antimicrobial resistance in pathogens has triggered the development of novel antimicrobial candidates. Siderophores provided newer approaches for the establishment of suitable targets for antibiotic discovery [57]. The siderophore receptors are located on the cell membrane of the pathogen and facilitate entry of the antibiotic and produce a black

hole due to deficiency of iron. Detailed study of biosynthetic pathways of siderophores permits the development of significant targets to facilitate hindering of these siderophores in pathogenic agents, which regulate pathogenic virulence, referred to as the Trojan Horse Strategy, which prevents the pathogen from becoming resistant to the drug candidate [50].

Antimicrobial resistance is accelerating from proto resistance to uncurable clinical pathogens. The effective therapy of resistant infections and discovery of novel drugs is facilitated by genotype data [50]. For instance, the estimation of resistance phenotype from genotype is enabled by the Comprehensive Antibiotic Resistance Database (CARD), which acquires curated mechanisms for resistance to data, genes as well as their targets, for resource establishment for the generation of an algorithm for the estimation of resistance to antibiotics [50]. The Resistance Gene Identifier (RGI) in the CARD presently gives an estimation of resistant genes, evaluates genome assemblies, and gives a comprehensive account of estimated genes resistant to antibiotics as well as targeted groups of drugs [50].

Furthermore, bioinformatics approaches are also employed in the development of multidrug-resistant tuberculosis (TB) drugs. The management of TB can be accelerated by using -omics technologies, where some drugs, approved by the FDA, are being validated for repurposing, followed by the development of more effective drugs with the ability to reduce tolerance towards a drug or altering the immune system response of the host [58]. The bioinformatics tools for genotyping and drug-resistant TB are listed in Table 2.

Table 2. Bioinformatics tools for genotyping and drug-resistant TB.

Bioinformatics Tools	Type	Genotyping
PhyTB	Online	SNP
CASTB	Online	4a
TGS-TB	Online	4b
KvarQ	Stand-alone	SNP/Spol
TB-Profiler	Online	SNP
PhyResSE	Online	SNP

Legend: PhyTB—phylogenetic tree visualization and sample positioning for *Mycobacterium tuberculosis*; CASTB—comprehensive analysis server for the complex *M. tuberculosis*; TGS-TB—Total Genotyping Solution for *M. tuberculosis*; KvarQ—tool that directly scans fastq files of bacterial genome sequences for known variants; TB-Profiler—Profiling tool for *M. tuberculosis* to detect drug resistance and lineage from Whole-Genome Sequencing Data; PhyResSE—a Web Tool Delineating *M. tuberculosis* Antibiotic Resistance and Lineage from Whole-Genome Sequencing Data; SNP—single nucleotide polymorphisms; Spol—database.

Advancements in bioinformatics and next-generation sequencing (NGS) techniques facilitate the evaluation and characterization of novel chemicals in marine natural products [59]. Bioinformatics approaches also aid in the development of prophylactic agents [50].

Bioinformatics tools are employed in the study of bacterial functional genomics. A reliable overview of the metabolic processes and phylogenetic diversity in microbial organisms is provided by integrated genome comparison systems, which comprise the prediction tools and protein functional classification systems [50]. A detailed account of the biochemical functions, metabolic processes, knockout phenotypes, inhibitors, and substrates is collated into a prioritization tool [60].

Furthermore, bioinformatics has brought about a revolution in the field of pharmacy by aiding in advancing the areas of drug discovery and target validation. This approach involves the employment of chemo-informatics in antimicrobial study, programs, machine learning approaches, fuzzy logic modeling, artificial neural networks (ANNs), genomics, and target discovery as well as molecular dynamics and simulations [50]. Numerous bioinformatics-based tools have been developed in microbiology to facilitate delivery and genes and drugs in organisms, as enlisted in Table 3.

Table 3. Bioinformatics tools in microbiological paradigm of gene and drug delivery.

Bioinformatics Tools	Role	Refs.
AntiBP server	Anti-bacterial peptides prediction in a protein sequence	[61]
BACTIBASE incorporated with MODELER	Prediction of 3D structure of user peptide by homology to known bacteriocins; such relational database is applicable for in silico designing of newer AMPs.	[50]
Titanium	Capable of making 400–600 million bases/run with 400 base pairs' read lengths	[50]
454 sequencing technology	Employed in studies related to 16S profile of microbiomes	[62]
Illumina sequencing platform	Rapid outputs due to more accurate outcomes and read lengths	[50]
MiSeq and HiSeq	Production of sequencing data of 100 GB in just 6 days	[63]
ANNs	Mathematical modeling algorithms, which provide an effective and reliable method for in silico detection of newer AMPs	[64]
Fuzzy logic modeling	Accurate evaluation of AMPs, enables antimicrobial research and development in quantitative structure–activity relationship (QSAR)	[65]
Machine learning	Evaluation of high-dimensional gene expression datasets for selection of gene	[66]
CHARMM	Program employed to facilitate activation of AMP interactions with lipid bilayers	[67]
GROMACS	Freeware tool employed in MD studies to develop trajectories	[68]

Legend: AMPs—antimicrobial peptides; ANNs—artificial neural networks; AntiBP—anti-bacterial peptides; BACTIBASE—database dedicated to bacteriocins; CHARMM—biomolecular simulation program; GB—gigabyte; GROMACS—Groningen machine for chemical simulations; MD—medical doctor; MiSeq and HiSeq—sequencing technologies developed by Illumina; QSAR—quantitative structure-activity relationship.

Furthermore, bioinformatics databases and tools have been found to aid in big data analytics, comprising the generation of a 3D homology model (Figure 3). Targeting of the NGS platform has been modified to attain diagnostic techniques at a molecular level, as a result of a proper understanding of molecular epidemiology and resistance genotype [69]. Novel data sharing prototypes have to be developed by CARD that aid in clinical research and diagnostics, using AMR computational techniques [69]. Moreover, proteomics is an accelerating field that is expected to significantly contribute to bioinformatics in the future. The computational techniques have already replaced methods such as protein microarrays, 2D gel electrophoresis, and mass spectrometry in protein structure interpretation [50]. The bioinformatics tools aid in phylogenetic profiling, metabolic pathway mapping, and expression profiling [50,70–72] (Figure 4).
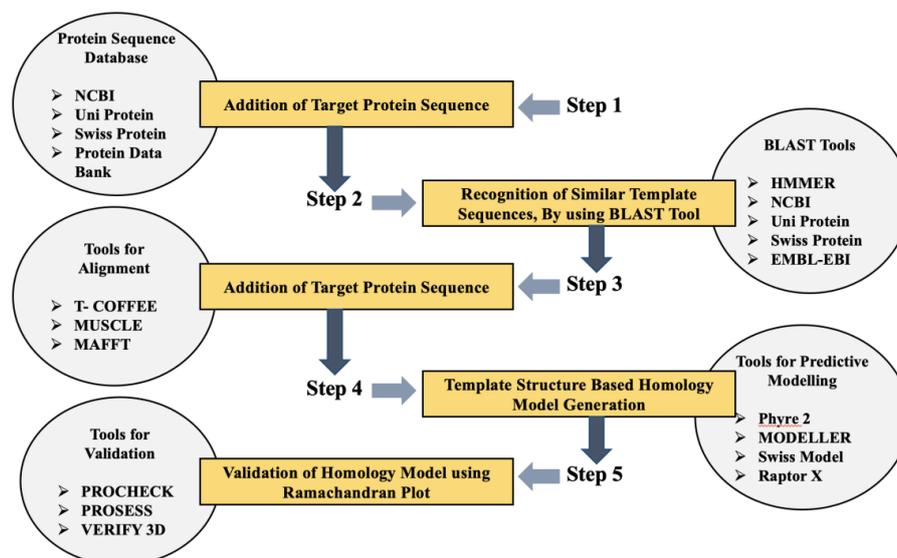

Figure 4. Use of bioinformatics-based tools, for alignment, validation, prediction modeling, BLAST tools, and protein sequence databases, in multiple steps of 3D homology model generation. Legend: BLAST—basic local alignment search tool; EBI—European Bioinformatics Institute; EMBL—European Molecular Biology laboratory; HMMER—biosequence analysis using profile hidden Markov models; MAFFT—Multiple Alignment using Fast Fourier Transform; MUSCLE—multiple sequence comparison by long expectation; NCBI—National Center for Biotechnology Information; Phyre 2—Protein Homology/AnalogY Recognition Engine; T-COFFEE—tree-based consistency objective function for alignment evaluation.

5. An Overview of Bioinformatics in Clinical Genomic Sequencing and MicroRNA Research

The bioinformatics approaches have interlinked the biological science with computational approaches, which has proven to be revolutionary in the field of medical sciences [71]. The bioinformatics approaches have been developed for mining, regulating, and processing as well as evaluating raw information, which has been accumulated from modernized approaches, such as the next-generation sequencing technique, as well as published studies [73–75]. Moreover, the bioinformatics tools based upon miRNAs are further divided into primary groups, namely, miRNA and target gene, metabolic pathway of miRNA, miRNAs identification, miRNA and transcription factors, miRNA–miRNA functional network, miRNA and mutations, miRNA and diseases, etc. [74,75]. The ease of accessibility influences the significance of the database amongst the customers, whereas superior quality evaluation of data creates more favorable tools of the software [73–75].

5.1. Bioinformatics Tools in MicroRNA Research

MicroRNAs (miRNAs) are RNA entities that are non-coding, single-stranded molecules that significantly contribute to the systems regulating complex genetic expression [70]. They aid in numerous metabolic and physiological processes, namely, proliferation, evolution, apoptotic pathway, differentiation, aging, and pathological processes [71]. Multiple human genomic areas, other than chromosome Y, comprise miRNA encoding genes, either individually or as clusters [72]. Effective profiling of miRNA expression is necessary in clinical healthcare and research, as altered miRNAs contribute to multiple body disorders [73,74]. Their significance as effective biomarkers for clinical use depends upon their expression profile in body fluids. Furthermore, the techniques associated with the quantification of miRNA and collection of standardized samples are simple [75]. The development of highly effective novel techniques, such as next-generation sequencing, results in the elevation of a number of raw biodatas. Moreover, advanced integrated computational techniques, such as artificial intelligence (AI), data processing, computer science, etc., are employed in the management and evaluation of large-scale biodata [71]. The field of bioinformatics research constitutes the -omics approaches, such as transcriptomics, genomics, metabolomics, and proteomics [76]. Furthermore, numerous bioinformatics techniques are presently available for investigations related to miRNAs, as well as in silico detection of miRNA biomarkers. Different bioinformatics tools have been investigated in the microRNA field, targeting the associated pathways, mutations, targets, biomarker discovery, and diseases [71].

MicroRNAs have also been identified to be used as biomarkers in the detection and treatment of diseases, which has been investigated over the previous years [71]. The miRNAs have been found to play a significant role in disorders such as cardiovascular complications, neuronal disorders, autoimmune diseases, cancer, and viral diseases [77]. The unreliable treatment approaches and problems faced in accomplishing desired expression profiles in diabetes, hepatic disorders, and cancers have diverted medical attention to the significance of miRNAs as fundamental elements that are required to be studied further [71]. The proprietary bioinformatics approaches are required to facilitate the recognition of miRNAs as biomarkers [78,79]. Firstly, detailed information about a specific miRNA is provided by microRNA and intragenic data (MIRIAD) [80].

The tools based upon bioinformatics facilitate the evaluation of the expression pathway of biomarkers, analysis of target genes, and sequence analysis. The second step in biomarker selection constitutes recognition of cell-line and tissue-specific microRNAs, where the tissue-specific miRNA expression profile depicts bioinformatics-based tools, namely the microRNA body map and miRmine, depending upon the transcriptomic results obtained from micro-arrays pr RT-PCR techniques [71]. The tools aim to introduce the functional miRNAs that are revealed in a specific tissue [81]. With the growing significance of circulating biomarkers, blood miRNAs are considered to be effective tools in interpreting the status of biomarker agents, depending upon their cellular origin in the blood stream [82]. The ability of miRNAs as biomarkers can be evaluated by MIRUMIR, which is a tool, based

upon bioinformatics, and is employed in numerous types of cancers [83]. Similarly, such tools are available for other disorders, such as cardiovascular complications, as is evident from Cardio_ncRNA database [84]. Furthermore, the triplexRNA tool facilitates monitoring of 2D interaction between target genes and biomarkers [85]. Similarly, miRNA-drug and miRNA-virus interactions are determined by bioinformatics tools such as the VIRBase and mammalian transcriptomic database (mTD), which facilitates the selection of biomarkers of therapeutic importance [86,87].

Different approaches based upon bioinformatics offer multiple facilities, which have been further elaborated. The manuscript targets the text mining and meta-databases, based upon bioinformatics, such as OMIC tools [88] and Tools4miRs [89], which were employed to attain numerous bioinformatics tools associated with microRNAs. Different bioinformatics tools belonging to multiple categories have been evaluated, as listed in Table 4 below.

Table 4. Bioinformatics-based databases and software for discovery and evaluation of miRNAs.

Tool	Category	Significance	Refs.
DIANAmicroT-CDS	Predicted miRNA target evaluation	A web-based application that facilitates data interpretation, evaluates functions exhibited by ncRNAs in body processes and diseases, and scrutinizes expression regulation datasets and miRNA regulatory elements.	[90]
MiRBase	Search for miRNAs	Introduces miRNA-based novel genes, provides comprehensive data on immature and mature miRNAs, provides immediate access to all the miRNA-related published data and resources	[91]
MiRDB	Predicted miRNA target evaluation	Predicts and evaluates the role of target genes, offers comprehensive data, provides screening options to facilitate role-based anticipation of different miRNAs, facilitates alignment of sequences	[92]
MiRscan	Search for miRNAs	A web-based application that recognizes and contrasts miRNA genes in the genetic sequence of greater than one organism	[93]
microTar	Predicted miRNA target evaluation	A windows application that evaluates the effect of miRNA binding on the whole mRNA molecule	[94]
MiReader	Search for miRNAs	A Linux and Windows application that recognizes the sequences of mature miRNAs without the requirement for reference genome sequences	[95]
miRmap	Predicted miRNA target evaluation	A windows/web-based application that is a mixture of characteristics from PITA, TargetScan, PACMIT, and miRanda, offers user-friendly approaches for operating precomputed predictions and modeling of miRNA targets	[96]
MiRanalyzer	Search for miRNAs	A windows/web-based application that is an sRNA toolbox, which recognizes or evaluates miRNA features based upon the outcome of the next-generation sequencing approaches	[97]
MiRNAPath	miRNAs and metabolic pathway	Depicts the inter-association between gene, miRNA, and metabolic pathway inputs, thus used to study the miRNA-based metabolic pathways	[98]
MiRmaid	Search for miRNAs	A web-based application that comprises all the characteristics of the miRNA database	[99]
MiRecords	Assessment of confirmed and estimated targets	Search and review of targets for miRNAs	[100]
Pharmaco-miR	MiRNAs, genes, and drugs	Significant in pharmacogenomics; integrates function status to the expression profile of miRNA via validated experimental proofs and computational techniques	[101]
MiRwalk	Assessment of confirmed and estimated targets	Validation of newer MiRNA targets, comprehensive database	[102]
ViTa	Viruses and miRNAs	Curates the target sites of miRNAs in chicken, mice, humans, or rats, with miRBase-derived known viral miRNA genes	[103]

Table 4. *Cont.*

Tool	Category	Significance	Refs.
TMREC	MiRNA regulatory network	Evaluation of role of regulatory processes, controlled by interactions between transcription factors and miRNAs in diseased states	[104]
MiR2GO	Mutations and miRNAs	Evaluation of impact of mutations and alterations in single nucleotide in central core miRNA sequence, coupling with target mRNA on their function; comparison and assessment of similarity percentage of miRNA pair functions at a cellular and molecular level	[105]
DAVID	MiRNA regulatory network	Effective interpretation of changes in a large number of genes; evaluation of procedure of functional product generation and genetic expression	[106]
PolymiRTS	Mutations and miRNAs	Evaluation of genetic polymorphisms in central core binding or target pairing site	[107]
PhenomiR	Diseases and miRNAs	Relationship between miRNA and disease	[108]
MiREnvironment	Environment and miRNAs	Interaction between environmental factors and miRNAs	[71]
MiRcancer	Diseases and miRNAs	Relationship between miRNA and cancer	[109]
CircuitsDB	Transcription factors and miRNAs	Interaction between miRNA and transcription factors to facilitate regulation of joint target gene	[110]
MiR2disease	Diseases and miRNAs	Portrays validated and estimated gene targets upon miRNA changes in diseases	[111]
PutmiR	Transcription factors and miRNAs	Interaction between miRNA and transcription factors to facilitate regulation of gene expression	[112]
MiRgator	MiRNA-miRNA interaction	Deep sequencing miRNA database that facilitates massive data analysis, comprehensive evaluation of target genes, expression profiles of miRNA–miRNA interactions, proper representation of miRNAs genes chromosomal region	[113]
DIANA-mirExTra	MiRNA-miRNA interaction	Functional evaluation of expression profiles and targets of miRNAs	[114]

Legend: Bioinformatics-based databases and software for discovery and evaluation of miRNAs. Significance of tools targeting the following categories: 1. Predicted miRNA target evaluation, 2. miRNA search, 3. miRNA metabolic pathway, 4. Confirmed and estimated targets assessment, 5. miRNAs, genes, and drugs, 6. Viruses and miRNAs, 7. miRNA regulatory network, 8. Mutations and miRNAs, 9. Disease and miRNAs, 10. Environment and miRNAs, 11. Transcription factors and miRNAs, 12. miRNA-miRNA interaction. DIANA—displacement analyzer; MiR—micro ribo-nucleic acid; MiRDB—miRNA database; MicroTAR—miRNA target prediction program; miRNAPath—miRNA and metabolic pathway; Pharmaco-miR—Pharmacogenomics and miRNA; ViTa—visual interpretations with three-dimensional annotations; TMREC—database for transcription factor and miRNA regulatory cascades in human diseases; miR2GO—comparative functional analysis for microRNAs; DAVID—the database for annotation, visualization, and integrated discovery; PolymiRTs—polymorphisms in microRNA target site; Phenotypic analysis of miRNA; PutmiR—putative transcription factor and micro RNA.

5.2. Bioinformatics Process in Clinical Genomic Sequencing

The molecular genetic pathological paradigm has been transformed from a visual approach to an informatics-dependent strategy with the advent of high-throughput sequencing technologies. The advantages, such as enhanced efficiency, ameliorated costs of sequencers, data storage, and data computation, have permitted access to large-scale sequencing (genomes and exomes) to a broad range of patients [115]. These consist of deep evaluation of tumor-normal pairs, detection and screening of rare genetic maladies, and diagnosis of healthy subjects. The elevating amount of information has enhanced the significance of the field of clinical bioinformatics [115]. Therefore, it is necessary to understand the importance of clinical bioinformatics in the present molecular genetic workflow pattern, where this section aims to provide a general outlook process, required to put clinical bioinformatics into action in the field of genomic sequencing. Informatics is fundamental in numerous spheres of laboratory testing, but this section accounts for computational approaches, parsing, and interpretation from sequencing tools, via a set of variants [115].

Figure 4 portrays a typical bioinformatics process in clinical genomic sequencing where the whole pathway is segregated into the primary, secondary, and tertiary interpretation. The algorithms related to the sequencing tools, which deal with transforming raw sequence reads into a string of arsenic (As) and cesium (Cs), constitute the primary analysis [115]. The alignment or mapping of the sequence reads on the reference genomic

sequence is detailed in the secondary analysis, followed by variant calling or recognition of variations between the reference genome and individual's sequence [115]. All the steps that are required to analyze the recognized sequence variants are included in tertiary analysis, which comprises filtering and annotating the recognized variants in order to find variations that are clinically relevant. The steps or phases of the process of clinical bioinformatics are followed by the quality control process, which portrays the robustness, completeness, and reliability of the generated data [115] (Figure 5).
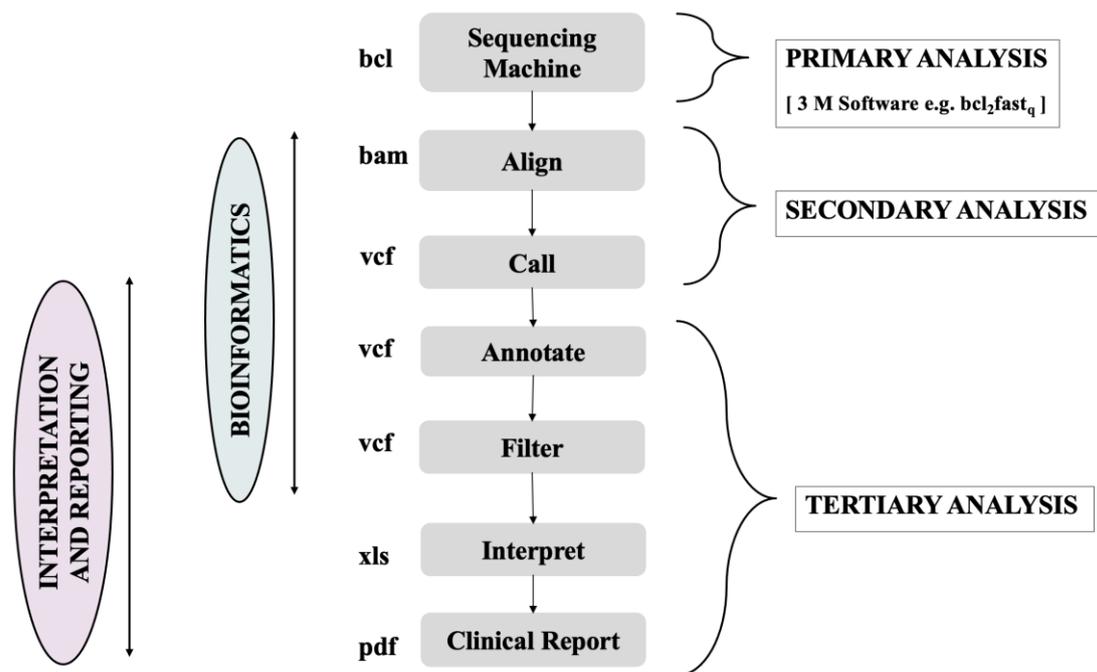

Figure 5. Role of bioinformatics in primary, secondary, and tertiary analysis in clinical genome sequencing, employing tools in sequence machine, alignment, calling, annotation, filtering, interpretation, and preparation of clinical reports.

Genome-scale sequencing is consistently advancing in both detection and screening of healthy individuals as well as newborn babies [115]. More rapid genome sequencing will be enabled, with greater accuracy, on account of multiple advancements in technological parameters, algorithms, and computation, which include [115]:

- A reference genome based upon graphical elucidation, which will promote a complete genome pathway with greater accuracy in alignment, especially for ethnicities that are under-represented.
- A longer read sequencing, which will promote greater genome resolution with greater accuracy in alignment.
- A software based upon a graphics processing unit (GPU), which will promote elevated parallelization and rapid computation.

Genomic sequences have begun to utilize basic computational advances, especially around big data. The accuracy of variant interpretations has been reported to be elevated by using natural language processing and machine learning, which is also employed in variant calling [116]. Full genome interpretation will be enabled by data integration, which will not be limited to the coding areas that are the prime target of the present clinical interpretation workflows. In the combined form, these technological tools will help to expose the potential of the genetic sequence of an individual within the area of clinical practice [115] (Table 5).

Table 5. Softwares used for secondary analysis and variant annotation in clinical genomic sequencing.

Name	Application	Refs.
1.	**For secondary analysis**	
GATK-UnifiedGenotyper	Consistently proceeds towards HalotypeCaller	[117]
BWA-SW	Shorter reads alignment	[118]
Freebayes	Based upon Bayesian haplotype, efficient in evaluating genomes with specific properties	[119]
Novoalign	Alignment software for commercial use	[120]
BCF tools and VCF tools	Distinct analytical characteristics, yet some common functions	[119,121]
BWA-MEM	Longer reads alignment, with bp > 100	[118,122]
FastQC	Evaluation of sequencing quality	[123]
Picard Tools	Provides QC evaluation for multiple secondary analysis stages	[124]
GenomeStrip	Evaluation of read length, read depth, and read mate pairing	[125]
verifybamID	Detection of sample contamination	[126]
BreakDancer	SV detection	[127]
Vcfeval	Comparison of two distinct variant call files, significant for validation	[128]
Pindel	Detection of large deletions and limited insertions	[121]
Bedtools	Manipulation of multiple bed files	[129]
Manta	Germline and somatic evaluation	[130]
VisCap	CNV calling for panel information	[131]
XHMM	CNV calling for exome information	[132]
2.	**For variant annotation**	
VEP	Annotation and prediction of impact of variant on genes	[133]
WGSA	Integration of results from SnpEff, VEP, and ANNOVAR, for annotation based upon gene modeling, integration of numerous epigenomics projects, integration of conservation scores, database associated with the disease, multiple prediction scores, and allele frequencies for SNV-centric resources	[134]
DANN	Integration of multiple annotations into one metric, annotation of non-coding and coding variants	[135]
CADD	Integration of multiple annotations into one metric, annotation of non-coding and coding variants	[136]
ANNOVAR	Annotation based upon gene, region, and filter	[137]
Oncotator	Aggregation of annotations from genomic, cancer variants, protein, and non-cancer variant annotations	[138]
SnpEff	Annotation and prediction of the impact of variants on genes	[139]

Legend: Software used in the process of clinical genomic sequencing, targeting secondary analysis (detailing the alignment or mapping of the sequence reads on the reference genomic sequence) and variant annotation (deals with assigning data to DNA variants). GATK—genome analysis tool kit; BWA-SW—Burrows–Wheeler aligner software; BCF—binary variant call format; VCF—variant call format; FastQC—fast quality control software; XHMM—eXome-Hidden Markov Model; VEP—variant effect predictor; WGSA—whole-genome sequencing annotator; DANN—deleterious annotation of genetic variants using neural networks; CADD—computer-aided drug design; ANNOVAR—annotate variation; SnpEff—single nucleotide polymorphism annotator.

6. The COVID-19 Pandemic Giving an Impulse to the Bioinformatics Approaches

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is responsible for the coronavirus disease 2019 (COVID-19), which has put a tremendous amount of pressure on the global healthcare system and has taken millions of lives across the globe [140]. Multiple novel vaccines have been developed by different areas of the healthcare sector, constituting medicine, public health, biology, computer science researchers, and bioinformatics [141]. Multiple themes are associated with COVID-19 and SARS-CoV-2 research, from a biological perspective, comprising HTP technologies such as Next-generation sequencing for detection of SARS-CoV-2 genome, databases for storage of SARS-CoV-2 variants and genomes, and databases and software tools based upon bioinformatics approaches for the evaluation and investigation of interactions between host and virus [142].

In particular with the therapeutic interventions, the major research themes comprise the identification of biomarkers associated with COVID-19, discovery of drug targets, and bioinformatics methods for drug repurposing, which is employing already-developed and available drugs for COVID-19 management [143]. The primary research themes at the public health and epidemiological level comprising ordered assemblage and release of information about the infection spread, such as regular reports of number of cases, number of deaths, number of patients hospitalized and in ICU, etc., may aid the general population to realize the severity of the infection [144], followed by biological tests for testing and evaluation and computational tools and techniques to track the infected individuals, storage of large amount of clinical data in electronic health records of infected patients [145], evaluation and assessment of the impact and effects of complete lockdown

at a socioeconomic level, and techniques to aid quarantined people by using local services, virtual assistants, and advanced technologies such as robots [143].

Complex techniques are used in the case of viruses to potentiate the coding ability of genomes due to the small size of the genome [146]. Bioinformatics, in collaboration with genomics, has provided a significant amount of information to understand the pathogenetic mechanisms as well as the spread of anti-microbial resistance to the immune responses of host cells in infectious diseases [147]. The size of the genetic sequence of novel severe acute respiratory syndrome 2 (SARS-CoV-2) has been reported to vary from 29.8 to 29.9 kb, with sequence variation as compared to the human coronaviruses, such as SARS and Middle East Respiratory Syndrome (MERS), recognized earlier [148]. However, it is essential to investigate the virological, epidemiological, and pathogenic data of SARS-CoV-2 in order to evaluate newer therapies and to promote the development of efficacious strategies to prevent the spread of the disease [149–151]. Bioinformatics strategies and tools have been utilized in the investigation of SARS-CoV-2 in order to minimize and control the negative impact of this pandemic, which has been observed to deteriorate the health of the population as well as the global socio-economic status [142].

6.1. Bioinformatics Tools and Techniques for COVID-19 Research

As a result of significant research in this regard, many scientific publications have been made available, along with novel tools and techniques to study the COVID-19 literature, such as data mining as well as natural language processing, to extract appropriate information [6]. The genomic data of SARS-CoV-2 can be assessed and detected by using next-generation sequencing that gives access to general information about the virus, followed by computation of the data and extending the knowledge by using bioinformatics pipelines as well as biological databases related to the interactions between host cell and virus.

Bioinformatics tools have also been developed for the detection, evaluation, and treatment of the disease exhibition fundamental actions such as SARS-CoV-2 detection, sequencing data analysis, tracing and containing the spread of the pandemic, the study of the evolution of the coronavirus, effective drug target discovery, and therapeutic approaches [143]. Data integration has been reported to aid in the assessment and identification of SARS-CoV-2 genome sequences and metadata, along with host–pathogen integrated datasets and integrative surveillance mechanisms [143]. Furthermore, the role of pathway enrichment analysis (PEA) in identifying suitable viral targets in biological processes in the host has also been reported. The online resources and Web tools focus on significant parameters of the virus, comprising genomics, interactomics, epidemiology, and pharmacology [143].

The COVID-19 pandemic has also provided an impulse for drug repurposing and 3D modeling, where the former deals with discovering the applications and roles of already-available drugs, resulting in amelioration of cost and time consumption, and the latter comprises oligomer and protomer models, SARS-CoV-2 human protein interactions, protein-ligand docking, and the effect of mutations [143]. This section primarily focuses on three bioinformatics approaches, next-generation sequencing (NGS), genome-wide association study (GWAS), and computer-aided drug design (CADD), as discussed further. Table 6 highlights multiple tools and databases, based upon bioinformatics, that can be employed in the management of the current pandemic situation.

Table 6. Bioinformatics-based tools and databases with the potential to combat the COVID-19 pandemic.

Bioinformatics-Based Tools/Databases	Role	Refs.
SRA database	High-throughput sequencing data repository	[152]
Fast QC	Quality control check on raw sequences in WGS	[153]
AUGUSTUS	Gene prediction in eukaryotic genome sequencing	[154]
MaSuRCA	Assembly of genome	[155]
Prokka	Prokaryotic genome annotation in WGS	[156]
Cutadapt	Recognizing and eliminating adaptor sequences, poly A tails, primer, and other unrequired sequences in WGS and metagenomics	[157]
Ragout	Reference-assisted assembly tool in WGS	[158]
Gene expression omnibus (GEO) database	Repository of data related to functional genomics	[159]
dbSNP	Repository for single-base nucleotide substitutions in SNP discovery	[160]
UCSC genome browser	Collection and analysis of model organism annotations in genomics	[161]
PROVEAN	Estimations of effect of substitution of amino acid on biological role of protein in SNP discovery	[162]
Kyoto encyclopaedia of genes and genome (KEGG)	Analysis of metabolic pathway	[163]
SIFT	Estimation of amino acid substitution on functional role of proteins	[164]
Conserved domain (CD) Search	Sequence alignment	[165]
NCBI gene database	Genetic data repository	[166]
PAUP	Evaluation of phylogenetic relationship between molecular sequences by using parsimony method	[167]
UniProt	Stores functional data on proteins	[168]
PopArt	Phylogenetic evaluation with visualization of haplotype diversity network	[169]
Molecular evolutionary genetics analysis (MEGA)	Alignment of multiple sequences, generation and statistical evaluation of phylogenetic relationships	[170]
Primer 3	Primer design in high-throughput genomics	[171]
PubChem	Chemical structure database for drug designing	[172]
Basic local alignment search tool (BLAST)	Finding similarity between sequences	[173]
AutoDock, Patch dock, Swiss dock, Zdock	Molecular docking tools	[142]
Protein databank (PDB)	3D-protein structure database	[174]
Drug bank	Comprises of data on FDA approved drugs	[142]
Modeler	Homology of 3D protein structures	[175]
PyMol	Editing and visualization of molecular structure	[176]
GROMACS	Tool for simulation of molecular dynamics	[142]
NAMD	Parallel molecular dynamics code	[142]
Open Babel	Chemical toolbox aiding in drug designing	[177]
VMD	Built-in scripting and 3D graphics-based visualization program	[178]

Legend: Bioinformatics-based tools and databases with the potential to combat the COVID-19 pandemic. Techniques and databases associated with high-throughput sequencing, whole-genome sequencing, data repositories, drug designing, molecular docking, phylogenetic evaluation, metabolic pathway analysis, sequence alignment, 3D homology, simulation of molecular dynamics, etc. SRA—sequence read archive; Fast QC—fast quality control; Ragout—Reference-assisted genome ordering utility; Prokka—prokaryotic genome annotation; Cutadapt—cutting adaptor sequences; GEO—gene expression omnibus; dpSNP—single nucleotide polymorphism database; UCSC Genome Browser—The University of California Santa Cruz Genome Browser; PROVEAN—protein variant effect analyzer; KEGG—Kyoto encyclopedia of genes and genome; SIFT—sorting intolerant from tolerant; CD—conserved domain; NCBI—National Centre for Biotechnology Information; PAUP—phylogenetic analysis using parsimony; UniProt—Universal Protein resource; PopArt—population analysis with reticulate tress; MEGA—Molecular evolutionary genetics analysis; BLAST—basic local alignment search tool; PDB—protein databank; PyMol—Python using molecular graphics tool; GROMACS—Groningen machine for chemical simulations; NAMD—nanoscale molecular dynamics; VMD—visual molecular dynamics.

6.1.1. Next-Generation Sequencing (NGS)

The advancement NGS has optimized the intensity and scale of biomedical sciences. In the healthcare paradigm, during the condition of an outbreak or a pandemic, it is essential to effectively and rapidly identify the causative agent/pathogen responsible along with epidemiological surveys and analysis, which are essential to promote disease control reaction [142]. Furthermore, whole-genome sequencing (WGS), metagenomics strategy, and other high-throughput sequencing techniques promote the opportunity to obtain the full sequence of the disease-causing genome [142]. Metagenomics is a simple and economical technique that is independent of the reference sequence for analysis, unlike the in silico virus sequencing [142]. It has been identified as a powerful tool for the identification of pathogens from environmental samples, followed by direct genomic evaluation of the organism during the conditions of pandemics or outbreaks [179].

Similarly, in the current COVID-19 pandemic, metagenomics has been applied to provide significant novel data related to SARS-CoV-2, along with rapid recognition and

characterization of the initial few COVID-19 cases [180,181], to promote the examination of SARS-CoV-2 parallel to the other co-infections in nazo-pharyngeal throat swab samples of patients [182], the intermediate recognition of the host responsible for infection transference to the human body [183], SARS-CoV-2 homologous sequence screening in other organisms [184], the impact of SARS-CoV-2 on changes in fecal microbiota in humans [185], the clinical infection of SARS-CoV-2 along with co-infections of bacteria [179], and so on. Such observations aid the clinicians and researchers to understand and isolate the COVID-19 patients, depicting different symptoms in a better way.

The applications of metagenomic strategies can be interpreted by numerous databases and software [142]. The limited load of the viral genome, unlike the host DNA, and problems associated with accurate genome assembly pose a significant challenge to directly obtaining the genome sequence of the virus from clinical samples [142]. The study of evolution and the genetic relationship of the virus to diseases as well as tracking the outbreaks is enabled by the WGS technique, which is considered an efficient strategy on account of the intensity of the sequencing data and quality of the sequences obtained [186].

Samples were collected from various countries globally by employing NGS techniques such as Roche, Illumina MiSeq, etc., and WGS of SARS-CoV-2 was carried out in order to facilitate early diagnosis and interpretation of COVID-19 disease [187–189]. Nanopore sequencing was employed for genetic material sequencing of SARS-CoV-2 [189]. The whole-genome sequence, available on different online databases as well as data analysis software, has been reported to optimize the genomic data analysis and offers the administration of better medications to the patients [142].

6.1.2. Genome-Wide Association Study (GWAS)

The GWAS has established a significant relationship between complex characteristics of humans and disorders. Translational genomic research requires effective and comprehensive identification of variants from WGS [190]. GWAS is associated with the detection of variants across the genomic sequences of various individuals to detect the relationship between genotype and phenotype [142]. These genetic variants identified are employed in the recognition of individuals susceptible to deadly disorders, which affects early screening and prevention of illness [191]. GWAS is a potential genetic analytical technique of observable alleles related to the disorder in the host cell, as single nucleotide polymorphisms (SNPs) [192]. GWAS is associated with certain applications, such as genetic or nucleotide alterations in the form of SNPs, sequence analysis, alignment, structural changes in the genomic sequence, primer design, and so on, which have developed newer approaches in SARS-CoV-2 investigations by effectively identifying as well as quantifying rare viral variants in the species [148,193]. The haplotype diversity analysis has been incorporated along with phylogenetic analysis in research analysis, associated with SARS-CoV-2, in order to investigate the population demography and evolution of SARS-CoV-2 across the globe [194,195]. The phylogenetic study has been reported to be applicable to the investigation of evolutionary and molecular association of SARS-CoV-2 with other coronavirus species, which further provided necessary information for efficient evaluation of genomic sequence of SARS-CoV-2 [193,196–198].

Certain primers are developed via in silico algorithms by focusing on conserved segments in the viral genetic material in order to ameliorate false-positive results during COVID-19 testing via a real-time polymerase chain reaction (rtPCR) as well as retard the requirement for standardization across varying PCR protocols [199,200]. These data related to infectious genes of SARS-CoV-2 are aiding the researchers across the globe in the generation of a vaccine to combat the virus, as per the detected virus genes coding regions, molecular alterations, and genetic sequence variations between the isolated species globally [142]. The genomic analysis and experimental investigations comprising phylogenetic analysis, single nucleotide polymorphism (SNP) study, primer designing, and so on have been conducted via high-throughput bioinformatics strategies and technologies, which are reported to be effective in data annotation and analysis [142].

6.1.3. Computer-Aided Drug Design (CADD)

The drug design process is considered to be a challenging approach that is cost-ineffective as well as time-consuming [201]. The drug design is incorporated with bioinformatics, which has become a significant part of this process and plays an essential role in drug target validation. It aids in understanding the complexity of biological pathways in order to enhance the drug discovery process [202]. CADD has been reported as a dominant process on account of its appropriate algorithms, comprising digital repositories' establishment, for the investigation of associations of chemical interactions [142].

Computational techniques for designing compounds with unusual physico-chemical properties along with assessment tools for the evaluation of potential lead candidates and so on aid in the discovery and development of drugs [203]. This method is associated with further advantages, such as cost-effectiveness, insight knowledge of interaction between drug and receptor, time to market, and an accelerated drug discovery and development process, which elevates its popularity in the research field associated with science [194].

7. Future Prospects and Conclusions

The bioinformatics assets can be modified to further improve the diagnostic and detection criteria and procedures in the healthcare sector. The employment of bioinformatics tools in the growing fields of pharmacovigilance and genomic sequencing holds great future benefits. Future transformations in the techniques and tools of bioinformatics can aid in better understanding drug resistance and microbial virulence, which can facilitate effective management of viral infections. There are numerous health disorders for which proper and reliable treatments are not yet available, such as cancer, HIV-AIDS, neurodegenerative diseases, etc. There is an absence of a significant research model in pharmacovigilance that would aid in providing a focused direction and scope of this field in the future, resulting in potential benefits [204]. The bioinformatics-based computational techniques can facilitate the acceleration of the drug development criteria, which can further result in the development of more active therapeutic candidates with limited toxicity profiles. Furthermore, the decision-making tools and comprehensive models can aid in transforming the conventional processes of drug delivery from single target to 'function first' as well as phenotypic selection methods, targeting systematic networks [21]. Taking into account the pandemic situation, the researchers can investigate and evaluate the SNPs associated with the affected host body, and computational primer design algorithms can be used to design modified forms of newer primers of genes or nucleotides [142]. Therefore, the bioinformatics approaches would facilitate simulation, identification, and prediction of the progression of the disease and responses of the drug candidates, for elevating the uses, safety profiles, and impact of newer and existing drugs, thereby strengthening the healthcare system.

Bioinformatics has accelerated the field of biomedical sciences and has potentiated the clinical and general aspects of the healthcare system. The review focuses on giving a detailed account of the significance of bioinformatics in the pharmaceutical industry and pharmacovigilance, followed by the fundamental assets of bioinformatics-based tools and databases in drug discovery and development. The value of translational bioinformatics approaches intensifies the development and discovery of suitable drug candidates, where the manuscript enlists numerous databases in tabular forms. Moreover, bioinformatics has been reported to be employed to accelerate microRNA research and clinical genomic sequencing, where it is assessed for its potential in -omic technologies and studies. The role of bioinformatics approaches in microbiology is explained owing to its significance in gene and drug discovery, proteomics, sequence data analysis, bacterial functional genomics, and the development of multi-drug resistant TB drugs, prophylactic agents, marine natural products, siderophores, and so on. In addition, the authors reveal the positive outcomes of this tech-driven strategy in the management of the current COVID-19 pandemic, where its involvement in next-generation sequencing, genome-wide association study, and computer-aided drug design tends to strengthen the fight against the pandemic. In addition, the

review aims to provide a clarified image of bioinformatics and invokes the necessity of technological tools, databases, and software, thereby attracting the readers and researchers to exhibit future assessment of this field to facilitate acceleration and elevated efficiency of the healthcare system.

Author Contributions: Conceptualization, T.B. and I.K.; methodology, T.B. and A.S.; software, S.S., S.B. and A.A.-H.; validation, T.B. and S.G.B.; formal analysis, T.B.; investigation, T.B., I.K., G.Z., E.E.B., C.B., M.S., M.M.T., C.S. and S.B.; resources, T.B. and. S.G.B.; data curation T.B., E.E.B., C.B. ans M.S.; writing—original draft preparation, T.B. and S.G.B.; writing—review and editing, T.B. and S.G.B.; visualization, S.G.B.; supervision, T.B. and S.G.B.; project administration, T.B.; funding acquisition, T.B. and M.M.T. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Gilbert, D. Bioinformatics software resources. *Brief. Bioinform.* **2004**, *5*, 300–304. [CrossRef] [PubMed]
2. Huang, D.W.; Sherman, B.T.; Lempicki, R.A. Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **2009**, *37*, 1–13. [CrossRef] [PubMed]
3. Teufel, A.; Krupp, M.; Weinmann, A.; Galle, P.R. Current bioinformatics tools in genomic biomedical research. *Int. J. Mol. Med.* **2006**, *17*, 967–973. [CrossRef] [PubMed]
4. Luscombe, N.; Greenbaum, D.; Gerstein, M. Review: What is bioinformatics. *An introduction and overview. Yearb. Med. Inform.* **2001**, *1*, 2.
5. Mandal, S.; Mandal, S.K. Rational drug design. *Eur. J. Pharmacol.* **2009**, *625*, 90–100. [CrossRef] [PubMed]
6. Katara, P. Role of bioinformatics and pharmacogenomics in drug discovery and development process. *Netw. Model. Anal. Health Inform. Bioinform.* **2013**, *2*, 225–230. [CrossRef]
7. Thibaut, U. Bioinformatics and rational drug design: Tools for discovery and better understanding of biological targets and mode of action of drugs. *Scand. J. Gastroenterol.* **2002**, *37*, 95–99. [CrossRef] [PubMed]
8. Ferrer-Costa, C.; Orozco, M.; de la Cruz, X. Use of bioinformatics tools for the annotation of disease-associated mutations in animal models. *Proteins Struct. Funct. Bioinform.* **2005**, *61*, 878–887. [CrossRef]
9. Moore, J.H.; Rhodes, C.H. Integration of molecular and cellular pathogenesis: A bioinformatics approach. In *Molecular Pathology*; Elsevier: Amsterdam, The Netherlands, 2009; pp. 219–224.
10. Lyall, A. Bioinformatics in the pharmaceutical industry. *Trends Biotechnol.* **1996**, *14*, 308–312. [CrossRef]
11. WHO (World Health Organization). The Importance of Pharmacovigilance. Available online: http://apps.who.int/medicinedocs/en/d/Js4893e/ (accessed on 20 April 2021).
12. Dal Pan, G.J. Ongoing challenges in pharmacovigilance. *Drug Saf.* **2014**, *37*, 1–8. [CrossRef]
13. Definition of Biomedical Informatics. American Medical Informatics Association. Available online: https://www.amia.org/biomedical-informatics-core-competencies (accessed on 15 March 2021).
14. Bhangale, R.; Vaity, S.; Kulkarni, N. A day in the life of a pharmacovigilance case processor. *Perspect. Clin. Res.* **2017**, *8*, 192. [CrossRef] [PubMed]
15. Bungau, C.; Blaga, F.; Gherghea, C. *Kaizen Implementation for Cost Reduction in Manufacturing Process Product "Driver Control Board"*; University of Oradea: Oradea, Romania, 2014; pp. 55–58.
16. Daina, L.G.; Sabau, M.; Daina, C.M.; Neamțu, C.; Tit, D.M.; Buhaș, C.L.; Bungau, C.; Aleya, L.; Bungau, S. Improving performance of a pharmacy in a Romanian hospital through implementation of an internal management control system. *Sci. Total Environ.* **2019**, *675*, 51–61. [CrossRef]
17. Grewal, A.; Kataria, H.; Drawan, I. Literature search for research planning and identification of research problem. *Indian J. Anaesth.* **2016**, *60*, 635–639. [CrossRef] [PubMed]
18. Murray-Rust, P. Bioinformatics and drug discovery. *Curr. Opin. Biotechnol.* **1994**, *5*, 648–653. [CrossRef]
19. Gaasterland, T.; Selkov, E. Reconstruction of metabolic networks using incomplete information. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **1995**, *3*, 127–135. [PubMed]

20. Bishop, M.J.; Plank, G.; Burton, R.A.B.; Schneider, J.E.; Gavaghan, D.J.; Grau, V.; Kohl, P. Development of an anatomically detailed MRI-derived rabbit ventricular model and assessment of its impact on simulations of electrophysiological function. *Am. J. Physiol. Heart Circ. Physiol.* **2010**, *298*, H699–H718. [CrossRef] [PubMed]

21. Yan, Q. *Translational Bioinformatics and Systems Biology Methods for Personalized Medicine*; Academic Press: Cambridge, MA, USA, 2017.

22. Boland, M.R.; Jacunski, A.; Lorberbaum, T.; Romano, J.D.; Moskovitch, R.; Tatonetti, N.P. Systems biology approaches for identifying adverse drug reactions and elucidating their underlying biological mechanisms. *Wiley Interdiscip. Rev.* **2016**, *8*, 104–122. [CrossRef] [PubMed]

23. Prathipati, P.; Mizuguchi, K. Systems biology approaches to a rational drug discovery paradigm. *Curr. Top. Med. Chem.* **2016**, *16*, 1009–1025. [CrossRef]

24. Chautard, E.; Thierry-Mieg, N.; Ricard-Blum, S. Interaction networks: From protein functions to drug discovery. A review. *Pathol. Biol.* **2009**, *57*, 324–333. [CrossRef] [PubMed]

25. Vlasblom, J.; Jin, K.; Kassir, S.; Babu, M. Exploring mitochondrial system properties of neurodegenerative diseases through interactome mapping. *J. Proteom.* **2014**, *100*, 8–24. [CrossRef]

26. A Dunn, D.; Apanovitch, D.; Follettie, M.; He, T.; Ryan, T. Taking a systems approach to the identification of novel therapeutic targets and biomarkers. *Curr. Pharm. Biotechnol.* **2010**, *11*, 721–734. [CrossRef]

27. Vandamme, D.; Minke, B.A.; Fitzmaurice, W.; Kholodenko, B.N.; Kolch, W. Systems biology-embedded target validation: Improving efficacy in drug discovery. *Wiley Interdiscip. Rev.* **2014**, *6*, 1–11. [CrossRef] [PubMed]

28. Ryall, K.A.; Tan, A.C. Systems biology approaches for advancing the discovery of effective drug combinations. *J. ChemInform.* **2015**, *7*, 7. [CrossRef]

29. Wathieu, H.; T Issa, N.; W Byers, S.; Dakshanamurthy, S. Harnessing polypharmacology with computer-aided drug design and systems biology. *Curr. Pharm. Des.* **2016**, *22*, 3097–3108. [CrossRef] [PubMed]

30. Kunz, M.; Liang, C.; Nilla, S.; Cecil, A.; Dandekar, T. The drug-minded protein interaction database (DrumPID) for efficient target analysis and drug development. *Database* **2016**, *2016*. [CrossRef] [PubMed]

31. Schreyer, A.M.; Blundell, T.L. CREDO: A structural interactomics database for drug discovery. *Database* **2013**, *2013*. [CrossRef] [PubMed]

32. Wang, C.; Hu, G.; Wang, K.; Brylinski, M.; Xie, L.; Kurgan, L. PDID: Database of molecular-level putative protein–drug interactions in the structural human proteome. *Bioinformatics* **2016**, *32*, 579–586. [CrossRef] [PubMed]

33. Nanduri, R.; Bhutani, I.; Somavarapu, A.K.; Mahajan, S.; Parkesh, R.; Gupta, P. ONRLDB—Manually curated database of experimentally validated ligands for orphan nuclear receptors: Insights into new drug discovery. *Database* **2015**, *2015*. [CrossRef]

34. Taccioli, C.; Sorrentino, G.; Zannini, A.; Caroli, J.; Beneventano, D.; Anderlucci, L.; Lolli, M.; Bicciato, S.; Del Sal, G. MDP, a database linking drug response data to genomic information, identifies dasatinib and statins as a combinatorial strategy to inhibit YAP/TAZ in cancer cells. *Oncotarget* **2015**, *6*, 38854. [CrossRef]

35. Weber, J.; Achenbach, J.; Moser, D.; Proschak, E. VAMMPIRE: A matched molecular pairs database for structure-based drug design and optimization. *J. Med. Chem.* **2013**, *56*, 5203–5207. [CrossRef] [PubMed]

36. Von Eichborn, J.; Murgueitio, M.S.; Dunkel, M.; Koerner, S.; Bourne, P.E.; Preissner, R. PROMISCUOUS: A database for network-based drug-repositioning. *Nucleic Acids Res.* **2010**, *39*, D1060–D1066. [CrossRef] [PubMed]

37. Readhead, B.; Dudley, J. Translational bioinformatics approaches to drug development. *Adv. Wound Care* **2013**, *2*, 470–489. [CrossRef] [PubMed]

38. A Seoane, J.; Aguiar-Pulido, V.; R Munteanu, C.; Rivero, D.; R Rabunal, J.; Dorado, J.; Pazos, A. Biomedical data integration in computational drug design and bioinformatics. *Curr. Comput. Aided Drug Des.* **2013**, *9*, 108–117. [CrossRef] [PubMed]

39. Buchan, N.S.; Rajpal, D.K.; Webster, Y.; Alatorre, C.; Gudivada, R.C.; Zheng, C.; Sanseau, P.; Koehler, J. The role of translational bioinformatics in drug discovery. *Drug Discov. Today* **2011**, *16*, 426–434. [CrossRef]

40. Yan, B.; Yin, F.; Wang, Q.; Zhang, W.; Li, L. Integration and bioinformatics analysis of DNA-methylated genes associated with drug resistance in ovarian cancer. *Oncol. Lett.* **2016**, *12*, 157–166. [CrossRef]

41. Yu, G.; Wang, L.; Li, Y.; Ma, Z.; Li, Y. Identification of drug candidate for osteoporosis by computational bioinformatics analysis of gene expression profile. *Eur. J. Med. Res.* **2013**, *18*, 5. [CrossRef]

42. Roedder, S.; Kimura, N.; Okamura, H.; Hsieh, S.-C.; Gong, Y.; Sarwal, M.M. Significance and suppression of redundant IL17 responses in acute allograft rejection by bioinformatics based drug repositioning of fenofibrate. *PLoS ONE* **2013**, *8*, e56657. [CrossRef] [PubMed]

43. Clark, P.M.; Dawany, N.; Dampier, W.; Byers, S.W.; Pestell, R.G.; Tozeren, A. Bioinformatics analysis reveals transcriptome and microRNA signatures and drug repositioning targets for IBD and other autoimmune diseases. *Inflamm. Bowel Dis.* **2012**, *18*, 2315–2333. [CrossRef] [PubMed]

44. Abul Seoud, R.A. BioHCVKD: A bioinformatics knowledge discovery system for HCV drug discovery–identifying proteins, ligands and active residues, in biological literature. *Int. J. Bioinform. Res. Appl.* **2011**, *7*, 317–333.

45. Nwankwo, N.; Seker, H. A signal processing-based Bioinformatics approach to assessing drug resistance: Human Immunodeficiency Virus as a case study. In Proceedings of the 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology, Buenos Aires, Argentina, 31 August–4 September 2010; pp. 1836–1839.

46. Shapshak, P.; Duncan, R.; Turchan, J.; Nath, A.; Minagar, A.; Kangueane, P.; Davis, W.; Chiappelli, F.; Elkomy, F.; Seth, R. Bioinformatics models in drug abuse and Neuro-AIDS: Using and developing databases. *Bioinformation* **2006**, *1*, 86. [CrossRef] [PubMed]

47. Setoain, J.; Franch, M.; Martínez, M.; Tabas-Madrid, D.; Sorzano, C.O.; Bakker, A.; Gonzalez-Couto, E.; Elvira, J.; Pascual-Montano, A. NFFinder: An online bioinformatics tool for searching similar transcriptomics experiments in the context of drug repositioning. *Nucleic Acids Res.* **2015**, *43*, W193–W199. [CrossRef] [PubMed]

48. Cheng, L.; Schneider, B.P.; Li, L. A bioinformatics approach for precision medicine off-label drug drug selection among triple negative breast cancer patients. *J. Am. Med Inform. Assoc.* **2016**, *23*, 741–749. [CrossRef] [PubMed]

49. Aoki-Kinoshita, K.F.; Kanehisa, M. Bioinformatics approaches in glycomics and drug discovery. *Curr. Opin. Mol. Ther.* **2006**, *8*, 514–520. [PubMed]

50. Koduru, S.K. The Impact of Bioinformatics Tools in the Development of Antimicrobial Drugs and Other Agents. In *Recent Developments in Applied Microbiology and Biochemistry*; Elsevier: Amsterdam, The Netherlands, 2019; pp. 335–347.

51. Le Chatelier, E.; Nielsen, T.; Qin, J.; Prifti, E.; Hildebrand, F.; Falony, G.; Almeida, M.; Arumugam, M.; Batto, J.-M.; Kennedy, S. Richness of human gut microbiome correlates with metabolic markers. *Nature* **2013**, *500*, 541–546. [CrossRef]

52. Segata, N.; Waldron, L.; Ballarini, A.; Narasimhan, V.; Jousson, O.; Huttenhower, C. Metagenomic microbial community profiling using unique clade-specific marker genes. *Nat. Methods* **2012**, *9*, 811–814. [CrossRef] [PubMed]

53. Glass, E.M.; Wilkening, J.; Wilke, A.; Antonopoulos, D.; Meyer, F. Using the metagenomics RAST server (MG-RAST) for analyzing shotgun metagenomes. *Cold Spring Harb. Protoc.* **2010**, *2010*, pdb-prot5368. [CrossRef]

54. Buchfink, B.; Xie, C.; Huson, D.H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **2015**, *12*, 59–60. [CrossRef] [PubMed]

55. Steinegger, M.; Söding, J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.* **2017**, *35*, 1026–1028. [CrossRef]

56. Pruesse, E.; Peplies, J.; Glöckner, F.O. SINA: Accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* **2012**, *28*, 1823–1829. [CrossRef]

57. Dhusia, K.; Bajpai, A.; Ramteke, P. Overcoming antibiotic resistance: Is siderophore Trojan horse conjugation an answer to evolving resistance in microbial pathogens? *J. Control. Release Off. J. Control. Release Soc.* **2017**, *269*, 63–87. [CrossRef]

58. Swaminathan, S.; Sundaramurthi, J.C.; Palaniappan, A.N.; Narayanan, S. Recent developments in genomics, bioinformatics and drug discovery to combat emerging drug-resistant tuberculosis. *Tuberculosis* **2016**, *101*, 31–40. [CrossRef]

59. Zhang, G.; Li, J.; Zhu, T.; Gu, Q.; Li, D. Advanced tools in marine natural drug discovery. *Curr. Opin. Biotechnol.* **2016**, *42*, 13–23. [CrossRef]

60. Freiberg, C. Novel computational methods in anti-microbial target identification. *Drug Discov. Today* **2001**, *6*, 72–80. [CrossRef]

61. Lata, S.; Mishra, N.K.; Raghava, G.P. AntiBP2: Improved version of antibacterial peptide prediction. *BMC Bioinform.* **2010**, *11*, 1–7. [CrossRef] [PubMed]

62. Bowman, J.S.; Rasmussen, S.; Blom, N.; Deming, J.W.; Rysgaard, S.; Sicheritz-Ponten, T. Microbial community structure of Arctic multiyear sea ice and surface seawater by 454 sequencing of the 16S RNA gene. *ISME J.* **2012**, *6*, 11–20. [CrossRef]

63. Caporaso, J.G.; Lauber, C.L.; Walters, W.A.; Berg-Lyons, D.; Huntley, J.; Fierer, N.; Owens, S.M.; Betley, J.; Fraser, L.; Bauer, M. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* **2012**, *6*, 1621–1624. [CrossRef] [PubMed]

64. Fjell, C.D.; Jenssen, H.; Hilpert, K.; Cheung, W.A.; Pante, N.; Hancock, R.E.; Cherkasov, A. Identification of novel antibacterial peptides by chemoinformatics and machine learning. *J. Med. Chem.* **2009**, *52*, 2006–2015. [CrossRef]

65. Mikut, R.; Hilpert, K. Interpretable features for the activity prediction of short antimicrobial peptides using fuzzy logic. *Int. J. Pept. Res. Ther.* **2009**, *15*, 129–137. [CrossRef]

66. Hao, Y.; Weiss, G.M. Gene Selection from Microarray Data for Age-related Macular Degeneration by Data Mining. 2016. Available online: https://storm.cis.fordham.edu/~{}gweiss/papers/DMIN-2016.pdf (accessed on 20 April 2021).

67. Brooks, B.R.; Brooks, C.L., III; Mackerell, A.D., Jr.; Nilsson, L.; Petrella, R.J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S. CHARMM: The biomolecular simulation program. *J. Comput. Chem.* **2009**, *30*, 1545–1614. [CrossRef]

68. Hess, B.; Kutzner, C.; Van Der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435–447. [CrossRef] [PubMed]

69. Alam, M.T.; Petit, R.A., III; Crispell, E.K.; Thornton, T.A.; Conneely, K.N.; Jiang, Y.; Satola, S.W.; Read, T.D. Dissecting vancomycin-intermediate resistance in Staphylococcus aureus using genome-wide association. *Genome Biol. Evol.* **2014**, *6*, 1174–1185. [CrossRef]

70. Ma, R.; Jiang, T.; Kang, X. Circulating microRNAs in cancer: Origin, function and application. *J. Exp. Clin. Cancer Res.* **2012**, *31*, 38. [CrossRef]

71. Pourteymourfard-Tabrizi, Z.; Jami, M.-S. Selection of suitable bioinformatic tools in micro-RNA research. *Gene Rep.* **2020**, *21*, 100893. [CrossRef]

72. Ghorai, A.; Ghosh, U. miRNA gene counts in chromosomes vary widely in a species and biogenesis of miRNA largely depends on transcription or post-transcriptional processing of coding genes. *Front. Genet.* **2014**, *5*, 100. [CrossRef]

73. Mahmoudian-Sani, M.-R.; Jami, M.-S.; Mahdavinezhad, A.; Amini, R.; Farnoosh, G.; Saidijam, M. The effect of the microrna-183 family on hair cell-specific markers of human bone marrow-derived mesenchymal stem cells. *Audiol. Neurotol.* **2018**, *23*, 208–215. [CrossRef] [PubMed]

74. Mehri-Ghahfarrokhi, A.; Pourteymourfard-Tabrizi, Z.; Farrokhi, E.; Chaleshtori, M.H.; Jami, M.-S. Increased levels of miR-124 in human dental pulp stem cells alter the expression of neural markers. *J. Otol.* **2019**, *14*, 121–127. [CrossRef] [PubMed]

75. Witwer, K.W. Circulating microRNA biomarker studies: Pitfalls and potential solutions. *Clin. Chem.* **2015**, *61*, 56–63. [CrossRef]

76. Mahmoudian Sani, M.R.; Mehri-Ghahfarrokhi, A.; Hashemzadeh-Chaleshtori, M.; Saidijam, M.; Jami, M.-S. Comparison of three types of mesenchymal stem cells (bone marrow, adipose tissue, and umbilical cord-derived) as potential sources for inner ear regeneration. *Int. Tinnitus J.* **2017**, *21*, 122–127. [CrossRef] [PubMed]

77. Larrea, E.; Sole, C.; Manterola, L.; Goicoechea, I.; Armesto, M.; Arestin, M.; Caffarel, M.M.; Araujo, A.M.; Araiz, M.; Fernandez-Mercado, M. New concepts in cancer biomarkers: Circulating miRNAs in liquid biopsies. *Int. J. Mol. Sci.* **2016**, *17*, 627. [CrossRef] [PubMed]

78. Bhomia, M.; Balakathiresan, N.S.; Wang, K.K.; Papa, L.; Maheshwari, R.K. A panel of serum MiRNA biomarkers for the diagnosis of severe to mild traumatic brain injury in humans. *Sci. Rep.* **2016**, *6*, 28148. [CrossRef] [PubMed]

79. Zhang, W.; Zang, J.; Jing, X.; Sun, Z.; Yan, W.; Yang, D.; Guo, F.; Shen, B. Identification of candidate miRNA biomarkers from miRNA regulatory network with application to prostate cancer. *J. Transl. Med.* **2014**, *12*, 66. [CrossRef] [PubMed]

80. Panwar, B.; Omenn, G.S.; Guan, Y. miRmine: A database of human miRNA expression profiles. *Bioinformatics* **2017**, *33*, 1554–1560. [CrossRef] [PubMed]

81. Mestdagh, P.; Lefever, S.; Pattyn, F.; Ridzon, D.; Fredlund, E.; Fieuw, A.; Ongenaert, M.; Vermeulen, J.; De Paepe, A.; Wong, L. The microRNA body map: Dissecting microRNA function through integrative genomics. *Nucleic Acids Res.* **2011**, *39*, e136. [CrossRef] [PubMed]

82. Juzenas, S.; Venkatesh, G.; Hübenthal, M.; Hoeppner, M.P.; Du, Z.G.; Paulsen, M.; Rosenstiel, P.; Senger, P.; Hofmann-Apitius, M.; Keller, A. A comprehensive, cell specific microRNA catalogue of human peripheral blood. *Nucleic Acids Res.* **2017**, *45*, 9290–9301. [CrossRef]

83. Antonov, A.; Knight, R.; Melino, G.; Barlev, N.; Tsvetkov, P. MIRUMIR: An online tool to test microRNAs as biomarkers to predict survival in cancer using multiple clinical data sets. *Cell Death Differ.* **2013**, *20*, 367. [CrossRef]

84. Li, Y.; Huo, C.; Pan, T.; Li, L.; Jin, X.; Lin, X.; Chen, J.; Zhang, J.; Guo, Z.; Xu, J. Systematic review regulatory principles of non-coding RNAs in cardiovascular diseases. *Brief. Bioinform.* **2019**, *20*, 66–76. [CrossRef]

85. Schmitz, U.; Lai, X.; Winter, F.; Wolkenhauer, O.; Vera, J.; Gupta, S.K. Cooperative gene regulation by microRNA pairs and their identification using a computational workflow. *Nucleic Acids Res.* **2014**, *42*, 7539–7552. [CrossRef]

86. Chen, X.; Xie, W.-B.; Xiao, P.-P.; Zhao, X.-M.; Yan, H. mTD: A database of microRNAs affecting therapeutic effects of drugs. *J. Genet. Genom.* **2017**, *44*, 269–271. [CrossRef] [PubMed]

87. Li, Y.; Wang, C.; Miao, Z.; Bi, X.; Wu, D.; Jin, N.; Wang, L.; Wu, H.; Qian, K.; Li, C. ViRBase: A resource for virus–host ncRNA-associated interactions. *Nucleic Acids Res.* **2015**, *43*, D578–D582. [CrossRef] [PubMed]

88. Henry, V.J.; Bandrowski, A.E.; Pepin, A.-S.; Gonzalez, B.J.; Desfeux, A. OMICtools: An informative directory for multi-omic data analysis. *Database* **2014**, *2014*. [CrossRef]

89. Lukasik, A.; Wójcikowski, M.; Zielenkiewicz, P. Tools4miRs–one place to gather all the tools for miRNA analysis. *Bioinformatics* **2016**, *32*, 2722–2724. [CrossRef]

90. Reczko, M.; Maragkakis, M.; Alexiou, P.; Grosse, I.; Hatzigeorgiou, A.G. Functional microRNA targets in protein coding sequences. *Bioinformatics* **2012**, *28*, 771–776. [CrossRef]

91. Kozomara, A.; Griffiths-Jones, S. miRBase: Annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.* **2014**, *42*, D68–D73. [CrossRef]

92. Wong, N.; Wang, X. miRDB: An online resource for microRNA target prediction and functional annotations. *Nucleic Acids Res.* **2015**, *43*, D146–D152. [CrossRef] [PubMed]

93. Lim, L.P.; Lau, N.C.; Weinstein, E.G.; Abdelhakim, A.; Yekta, S.; Rhoades, M.W.; Burge, C.B.; Bartel, D.P. The microRNAs of Caenorhabditis elegans. *Genes Dev.* **2003**, *17*, 991–1008. [CrossRef]

94. Thadani, R.; Tammi, M.T. MicroTar: Predicting microRNA targets from RNA duplexes. In *BMC Bioinformatics*; BioMed Centra: London, UK, 2006; pp. 1–9.

95. Jha, A.; Shankar, R. miReader: Discovering novel miRNAs in species without sequenced genome. *PLoS ONE* **2013**, *8*, e66857.

96. Vejnar, C.E.; Blum, M.; Zdobnov, E.M. miRmap web: Comprehensive microRNA target prediction online. *Nucleic Acids Res.* **2013**, *41*, W165–W168. [CrossRef] [PubMed]

97. Rueda, A.; Barturen, G.; Lebrón, R.; Gómez-Martín, C.; Alganza, Á.; Oliver, J.L.; Hackenberg, M. sRNAtoolbox: An integrated collection of small RNA research tools. *Nucleic Acids Res.* **2015**, *43*, W467–W473. [CrossRef]

98. Chiromatzo, A.; Oliveira, T.; Pereira, G.; Costa, A.; Montesco, C.; Gras, D.; Yosetake, F.; Vilar, J.; Cervato, M.; Prado, P. miRNApath: A database of miRNAs, target genes and metabolic pathways. *Genet. Mol. Res.* **2007**, *6*, 859–865.

99. Jacobsen, A.; Krogh, A.; Kauppinen, S.; Lindow, M. miRMaid: A unified programming interface for microRNA data resources. *BMC Bioinform.* **2010**, *11*, 1–8. [CrossRef] [PubMed]

100. Xiao, F.; Zuo, Z.; Cai, G.; Kang, S.; Gao, X.; Li, T. miRecords: An integrated resource for microRNA–target interactions. *Nucleic Acids Res.* **2009**, *37*, D105–D110. [CrossRef] [PubMed]

101. Rukov, J.L.; Wilentzik, R.; Jaffe, I.; Vinther, J.; Shomron, N. Pharmaco-miR: Linking microRNAs and drug effects. *Brief. Bioinform.* **2014**, *15*, 648–659. [CrossRef] [PubMed]

102. Dweep, H.; Gretz, N. miRWalk2. 0: A comprehensive atlas of microRNA-target interactions. *Nat. Methods* **2015**, *12*, 697. [CrossRef] [PubMed]

103. Hsu, P.W.-C.; Lin, L.-Z.; Hsu, S.-D.; Hsu, J.B.-K.; Huang, H.-D. ViTa: Prediction of host microRNAs targets on viruses. *Nucleic Acids Res.* **2007**, *35*, D381–D385. [CrossRef]

104. Wang, S.; Li, W.; Lian, B.; Liu, X.; Zhang, Y.; Dai, E.; Yu, X.; Meng, F.; Jiang, W.; Li, X. TM REC: A Database of Transcription Factor and MiRNA Regulatory Cascades in Human Diseases. *PLoS ONE* **2015**, *10*, e0125222.

105. Bhattacharya, A.; Cui, Y. miR2GO: Comparative functional analysis for microRNAs. *Bioinformatics* **2015**, *31*, 2403–2405. [CrossRef] [PubMed]

106. Jiao, X.; Sherman, B.T.; Huang, D.W.; Stephens, R.; Baseler, M.W.; Lane, H.C.; Lempicki, R.A. DAVID-WS: A stateful web service to facilitate gene/protein list analysis. *Bioinformatics* **2012**, *28*, 1805–1806. [CrossRef]

107. Bhattacharya, A.; Ziebarth, J.D.; Cui, Y. PolymiRTS Database 3.0: Linking polymorphisms in microRNAs and their target sites with human diseases and biological pathways. *Nucleic Acids Res.* **2014**, *42*, D86–D91. [CrossRef] [PubMed]

108. Ruepp, A.; Kowarsch, A.; Schmidl, D.; Buggenthin, F.; Brauner, B.; Dunger, I.; Fobo, G.; Frishman, G.; Montrone, C.; Theis, F.J. PhenomiR: A knowledgebase for microRNA expression in diseases and biological processes. *Genome Biol.* **2010**, *11*, R6. [CrossRef]

109. Xie, B.; Ding, Q.; Han, H.; Wu, D. miRCancer: A microRNA–cancer association database constructed by text mining on literature. *Bioinformatics* **2013**, *29*, 638–644. [CrossRef] [PubMed]

110. Friard, O.; Re, A.; Taverna, D.; De Bortoli, M.; Corá, D. CircuitsDB: A database of mixed microRNA/transcription factor feed-forward regulatory circuits in human and mouse. *BMC Bioinform.* **2010**, *11*, 435. [CrossRef]

111. Jiang, Q.; Wang, Y.; Hao, Y.; Juan, L.; Teng, M.; Zhang, X.; Li, M.; Wang, G.; Liu, Y. miR2Disease: A manually curated database for microRNA deregulation in human disease. *Nucleic Acids Res.* **2009**, *37*, D98–D104. [CrossRef]

112. Bandyopadhyay, S.; Bhattacharyya, M. PuTmiR: A database for extracting neighboring transcription factors of human microRNAs. *BMC Bioinform.* **2010**, *11*, 190. [CrossRef]

113. Cho, S.; Jang, I.; Jun, Y.; Yoon, S.; Ko, M.; Kwon, Y.; Choi, I.; Chang, H.; Ryu, D.; Lee, B. MiRGator v3. 0: A microRNA portal for deep sequencing, expression profiling and mRNA targeting. *Nucleic Acids Res.* **2012**, *41*, D252–D257. [CrossRef]

114. Vlachos, I.S.; Vergoulis, T.; Paraskevopoulou, M.D.; Lykokanellos, F.; Georgakilas, G.; Georgiou, P.; Chatzopoulos, S.; Karagkouni, D.; Christodoulou, F.; Dalamagas, T. DIANA-mirExTra v2.0: Uncovering microRNAs and transcription factors with crucial roles in NGS expression data. *Nucleic Acids Res.* **2016**, *44*, W128–W134. [CrossRef] [PubMed]

115. Lebo, M.S.; Hao, L.; Lin, C.-F.; Singh, A. Bioinformatics in clinical genomic sequencing. *Adv. Mol. Pathol.* **2018**, *1*, 9–26. [CrossRef]

116. Poplin, R.; Newburger, D.; Dijamco, J.; Nguyen, N.; Loy, D.; Gross, S.S.; McLean, C.Y.; DePristo, M.A. Creating a universal SNP and small indel variant caller with deep neural networks. *bioRxiv* **2016**. [CrossRef]

117. McKenna, A.; Hanna, M.; Banks, E.; Sivachenko, A.; Cibulskis, K.; Kernytsky, A.; Garimella, K.; Altshuler, D.; Gabriel, S.; Daly, M. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **2010**, *20*, 1297–1303. [CrossRef] [PubMed]

118. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **2009**, *25*, 1754–1760. [CrossRef] [PubMed]

119. Garrison, E.; Marth, G. Haplotype-based variant detection from short-read sequencing. *arXiv* **2012**, arXiv:1207.3907.

120. NovoAlign | Novocraft. Available online: http://www.novocraft.com/products/novoalign/ (accessed on 15 March 2021).

121. Ye, K.; Schulz, M.H.; Long, Q.; Apweiler, R.; Ning, Z. Pindel: A pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics* **2009**, *25*, 2865–2871. [CrossRef] [PubMed]

122. Zhang, L.; Liu, C.; Dong, S. Pipe MEM: A Framework to Speed up BWA-MEM in Spark with Low Overhead. *Genes* **2019**, *10*, 886. [CrossRef] [PubMed]

123. Babraham Bioinformatics—FastQC a Quality Control Tool for High Throughput Sequence Data. Available online: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (accessed on 16 April 2021).

124. Picard Tools—By Broad Institute. Available online: https://broadinstitute.github.io/picard/ (accessed on 16 April 2021).

125. Handsaker, R.E.; Van Doren, V.; Berman, J.R.; Genovese, G.; Kashin, S.; Boettger, L.M.; McCarroll, S.A. Large multiallelic copy number variations in humans. *Nat. Genet.* **2015**, *47*, 296–303. [CrossRef]

126. Jun, G.; Flickinger, M.; Hetrick, K.N.; Romm, J.M.; Doheny, K.F.; Abecasis, G.R.; Boehnke, M.; Kang, H.M. Detecting and estimating contamination of human DNA samples in sequencing and array-based genotype data. *Am. J. Hum. Genet.* **2012**, *91*, 839–848. [CrossRef] [PubMed]

127. Fan, X.; Abbott, T.E.; Larson, D.; Chen, K. BreakDancer: Identification of genomic structural variation from paired-end read mapping. *Curr. Protoc. Bioinform.* **2014**, *45*, 15–16. [CrossRef]

128. Cleary, J.G.; Braithwaite, R.; Gaastra, K.; Hilbush, B.S.; Inglis, S.; Irvine, S.A.; Jackson, A.; Littin, R.; Rathod, M.; Ware, D. Comparing variant call files for performance benchmarking of next-generation sequencing variant calling pipelines. *bioRxiv* **2015**, 023754. [CrossRef]

129. Quinlan, A.R. BEDTools: The Swiss-army tool for genome feature analysis. *Curr. Protoc. Bioinform.* **2014**, *47*, 11–12. [CrossRef] [PubMed]

130. Chen, X.; Schulz-Trieglaff, O.; Shaw, R.; Barnes, B.; Schlesinger, F.; Källberg, M.; Cox, A.J.; Kruglyak, S.; Saunders, C.T. Manta: Rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics* **2016**, *32*, 1220–1222. [CrossRef] [PubMed]

131. Pugh, T.J.; Amr, S.S.; Bowser, M.J.; Gowrisankar, S.; Hynes, E.; Mahanta, L.M.; Rehm, H.L.; Funke, B.; Lebo, M.S. VisCap: Inference and visualization of germ-line copy-number variants from targeted clinical sequencing data. *Genet. Med.* **2016**, *18*, 712–719. [CrossRef] [PubMed]

132. Fromer, M.; Purcell, S.M. Using XHMM software to detect copy number variation in whole-exome sequencing data. *Curr. Protoc. Hum. Genet.* **2014**, *81*, 7–23. [CrossRef] [PubMed]

133. McLaren, W.; Gil, L.; Hunt, S.E.; Riat, H.S.; Ritchie, G.R.; Thormann, A.; Flicek, P.; Cunningham, F. The ensembl variant effect predictor. *Genome Biol.* **2016**, *17*, 122. [CrossRef] [PubMed]

134. Liu, X.; White, S.; Peng, B.; Johnson, A.D.; Brody, J.A.; Li, A.H.; Huang, Z.; Carroll, A.; Wei, P.; Gibbs, R. WGSA: An annotation pipeline for human genome sequencing studies. *J. Med. Genet.* **2016**, *53*, 111–112. [CrossRef] [PubMed]

135. 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]

136. Kircher, M.; Witten, D.M.; Jain, P.; O'Roak, B.J.; Cooper, G.M.; Shendure, J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* **2014**, *46*, 310–315. [CrossRef]

137. Wang, K.; Li, M.; Hakonarson, H. ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **2010**, *38*, e164. [CrossRef]

138. Ramos, A.H.; Lichtenstein, L.; Gupta, M.; Lawrence, M.S.; Pugh, T.J.; Saksena, G.; Meyerson, M.; Getz, G. Oncotator: Cancer variant annotation tool. *Hum. Mutat.* **2015**, *36*, E2423–E2429. [CrossRef]

139. Cingolani, P.; Platts, A.; Wang, L.L.; Coon, M.; Nguyen, T.; Wang, L.; Land, S.J.; Lu, X.; Ruden, D.M. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly* **2012**, *6*, 80–92. [CrossRef]

140. Hu, B.; Guo, H.; Zhou, P.; Shi, Z.L. Characteristics of SARS-CoV-2 and COVID-19. *Nat. Rev. Microbiol.* **2021**, *19*, 141–154. [CrossRef]

141. Forni, G.; Mantovani, A.; Moretta, L.; Rappuoli, R.; Rezza, G.; Bagnasco, A.; Barsacchi, G.; Bussolati, G.; Cacciari, M.; Cappuccinelli, P.; et al. COVID-19 vaccines: Where we stand and challenges ahead. *Cell Death Differ.* **2021**, *28*, 626–639. [CrossRef]

142. Ray, M.; Sable, M.N.; Sarkar, S.; Hallur, V.K. Essential interpretations of bioinformatics in COVID-19 pandemic. *Meta Gene* **2020**, 100844.

143. Cannataro, M.; Harrison, A. Bioinformatics helping to mitigate the impact of COVID-19–Editorial. *Brief. Bioinform.* **2021**, *22*, 613–615. [CrossRef]

144. Dong, E.; Du, H.; Gardner, L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect. Dis.* **2020**, *20*, 533–534. [CrossRef]

145. Weber, G.M.; Hong, C.; Palmer, N.P.; Avillach, P.; Murphy, S.N.; Gutiérrez-Sacristán, A.; Xia, Z.; Serret-Larmande, A.; Neuraz, A.; Omenn, G.S.; et al. International comparisons of harmonized laboratory value trajectories to predict severe covid-19: Leveraging the 4ce collaborative across 342 hospitals and 6 countries: A retrospective cohort study. *medRxiv* **2021**. [CrossRef]

146. Gautam, A.; Tiwari, A.; Malik, Y.S. Bioinformatics applications in advancing animal virus research. In *Recent Advances in Animal Virology*; Springer: Berlin, Germany, 2019; pp. 447–471.

147. Bah, S.Y.; Morang'a, C.M.; Kengne-Ouafo, J.A.; Amenga-Etego, L.; Awandare, G.A. Highlights on the application of genomics and bioinformatics in the fight against infectious diseases: Challenges and opportunities in Africa. *Front. Genet.* **2018**, *9*, 575. [CrossRef]

148. Khailany, R.A.; Safdar, M.; Ozaslan, M. Genomic characterization of a novel SARS-CoV-2. *Gene Rep.* **2020**, *19*, 100682. [CrossRef] [PubMed]

149. Messina, F.; Giombini, E.; Agrati, C.; Vairo, F.; Bartoli, T.A.; Al Moghazi, S.; Piacentini, M.; Locatelli, F.; Kobinger, G.; Maeurer, M. COVID-19: Viral–host interactome analyzed by network based-approach model to study pathogenesis of SARS-CoV-2 infection. *J. Transl. Med.* **2020**, *18*, 233. [CrossRef] [PubMed]

150. Behl, T.; Kaur, I.; Bungau, S.; Kumar, A.; Uddin, M.S.; Kumar, C.; Pal, G.; Shrivastava, K.; Zengin, G.; Arora, S. The dual impact of ACE2 in COVID-19 and ironical actions in geriatrics and pediatrics with possible therapeutic solutions. *Life Sci.* **2020**, *257*, 118075. [CrossRef] [PubMed]

151. Kabir, M.T.; Uddin, M.S.; Hossain, M.F.; Abdulhakim, J.A.; Alam, M.A.; Ashraf, G.M.; Bungau, S.G.; Bin-Jumah, M.N.; Abdel-Daim, M.M.; Aleya, L. nCOVID-19 Pandemic: From Molecular Pathogenesis to Potential Investigational Therapeutics. *Front. Cell Dev. Biol.* **2020**, *8*, 616. [CrossRef] [PubMed]

152. Leinonen, R.; Sugawara, H.; Shumway, M.; Collaboration, I.N.S.D. The sequence read archive. *Nucleic Acids Res.* **2010**, *39*, D19–D21. [CrossRef] [PubMed]

153. Brown, J.; Pirrung, M.; McCue, L.A. FQC Dashboard: Integrates FastQC results into a web-based, interactive, and extensible FASTQ quality control tool. *Bioinformatics* **2017**, *33*, 3137–3139. [CrossRef] [PubMed]

154. Stanke, M.; Morgenstern, B. AUGUSTUS: A web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res.* **2005**, *33*, W465–W467. [CrossRef] [PubMed]

155. Zimin, A.V.; Marçais, G.; Puiu, D.; Roberts, M.; Salzberg, S.L.; Yorke, J.A. The MaSuRCA genome assembler. *Bioinformatics* **2013**, *29*, 2669–2677. [CrossRef] [PubMed]

156. Seemann, T. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* **2014**, *30*, 2068–2069. [CrossRef]

157. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* **2011**, *17*, 10–12. [CrossRef]

158. Kolmogorov, M.; Raney, B.; Paten, B.; Pham, S. Ragout—a reference-assisted assembly tool for bacterial genomes. *Bioinformatics* **2014**, *30*, i302–i309. [CrossRef] [PubMed]

159. Clough, E.; Barrett, T. The gene expression omnibus database. In *Statistical Genomics*; Springer: Berlin, Germany, 2016; pp. 93–110.

160. Sherry, S.T.; Ward, M.-H.; Kholodov, M.; Baker, J.; Phan, L.; Smigielski, E.M.; Sirotkin, K. dbSNP: The NCBI database of genetic variation. *Nucleic Acids Res.* **2001**, *29*, 308–311. [CrossRef]

161. Karolchik, D.; Hinrichs, A.S.; Kent, W.J. The UCSC genome browser. *Curr. Protoc. Bioinform.* **2012**, *40*, 1–4. [CrossRef]

162. Ray, M.; Mishra, J.; Priyadarshini, A.; Sahoo, S. In silico identification of potential drug target and analysis of effective single nucleotide polymorphisms for autism spectrum disorder. *Gene Rep.* **2019**, *16*, 100420. [CrossRef]

163. Kanehisa, M.; Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **2000**, *28*, 27–30. [CrossRef]

164. Sim, N.-L.; Kumar, P.; Hu, J.; Henikoff, S.; Schneider, G.; Ng, P.C. SIFT web server: Predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res.* **2012**, *40*, W452–W457. [CrossRef]

165. Ray, M.; Sarkar, S.; Rath, S.N. Druggability for COVID-19: In silico discovery of potential drug compounds against nucleocapsid (N) protein of SARS-CoV-2. *Genom. Inform.* **2020**, *18*, e43. [CrossRef]

166. Brown, G.R.; Hem, V.; Katz, K.S.; Ovetsky, M.; Wallin, C.; Ermolaeva, O.; Tolstoy, I.; Tatusova, T.; Pruitt, K.D.; Maglott, D.R. Gene: A gene-centered information resource at NCBI. *Nucleic Acids Res.* **2015**, *43*, D36–D42. [CrossRef] [PubMed]

167. Wilgenbusch, J.; Swofford, D. Inferring evolutionary trees with PAUP*. *Curr. Protoc. Bioinform.* **2003**. [CrossRef]

168. Consortium, U. The universal protein resource (UniProt). *Nucleic Acids Res.* **2007**, *36*, D190–D195. [CrossRef]

169. Leigh, J.W.; Bryant, D. popart: Full-feature software for haplotype network construction. *Methods Ecol. Evol.* **2015**, *6*, 1110–1116. [CrossRef]

170. Kumar, S.; Nei, M.; Dudley, J.; Tamura, K. MEGA: A biologist-centric software for evolutionary analysis of DNA and protein sequences. *Brief. Bioinform.* **2008**, *9*, 299–306. [CrossRef] [PubMed]

171. Untergasser, A.; Cutcutache, I.; Koressaar, T.; Ye, J.; Faircloth, B.C.; Remm, M.; Rozen, S.G. Primer3—New capabilities and interfaces. *Nucleic Acids Res.* **2012**, *40*, e115. [CrossRef]

172. Kim, S.; Thiessen, P.A.; Cheng, T.; Yu, B.; Shoemaker, B.A.; Wang, J.; Bolton, E.E.; Wang, Y.; Bryant, S.H. Literature information in PubChem: Associations between PubChem records and scientific articles. *J. Cheminform.* **2016**, *8*, 32. [CrossRef]

173. Boratyn, G.M.; Camacho, C.; Cooper, P.S.; Coulouris, G.; Fong, A.; Ma, N.; Madden, T.L.; Matten, W.T.; McGinnis, S.D.; Merezhuk, Y. BLAST: A more efficient report with usability improvements. *Nucleic Acids Res.* **2013**, *41*, W29–W33. [CrossRef] [PubMed]

174. Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. The protein data bank. *Nucleic Acids Res.* **2000**, *28*, 235–242. [CrossRef]

175. Eswar, N.; Webb, B.; Marti-Renom, M.A.; Madhusudhan, M.; Eramian, D.; Shen, M.y.; Pieper, U.; Sali, A. Comparative protein structure modeling using Modeller. *Curr. Protoc. Bioinform.* **2006**, *15*, 5–6. [CrossRef]

176. Seeliger, D.; de Groot, B.L. Ligand docking and binding site analysis with PyMOL and Autodock/Vina. *J. Comput. Aided Mol. Des.* **2010**, *24*, 417–422. [CrossRef] [PubMed]

177. O'Boyle, N.M.; Banck, M.; James, C.A.; Morley, C.; Vandermeersch, T.; Hutchison, G.R. Open Babel: An open chemical toolbox. *J. Cheminform.* **2011**, *3*, 33. [CrossRef]

178. Hsin, J.; Arkhipov, A.; Yin, Y.; Stone, J.; Schulten, K. Using VMD: An Introductory Tutorial. *Curr. Protoc. Bioinform.* **2008**, *24*, 5–7. [CrossRef] [PubMed]

179. Peddu, V.; Shean, R.C.; Xie, H.; Shrestha, L.; Perchetti, G.A.; Minot, S.S.; Roychoudhury, P.; Huang, M.-L.; Nalla, A.; Reddy, S.B. Metagenomic analysis reveals clinical SARS-CoV-2 infection and bacterial or viral superinfection and colonization. *Clin. Chem.* **2020**, *66*, 966–972. [CrossRef] [PubMed]

180. Chen, L.; Liu, W.; Zhang, Q.; Xu, K.; Ye, G.; Wu, W.; Sun, Z.; Liu, F.; Wu, K.; Zhong, B. RNA based mNGS approach identifies a novel human coronavirus from two individual pneumonia cases in 2019 Wuhan outbreak. *Emerg. Microbes Infect.* **2020**, *9*, 313–319. [CrossRef]

181. Manning, J.E.; Bohl, J.A.; Lay, S.; Chea, S.; Sovann, L.; Sengdoeurn, Y.; Heng, S.; Vuthy, C.; Kalantar, K.; Ahyong, V. Rapid metagenomic characterization of a case of imported COVID-19 in Cambodia. *bioRxiv* **2020**. [CrossRef]

182. Van Tan, L.; Man, D.; Hang, V.; Khanh, P.; Xuan, T.; Phong, N.; Tu, T.; Hien, T.; Hung, L.; Truong, N. SARS-CoV-2 and co-infections detection in nasopharyngeal throat swabs of COVID-19 patients by metagenomics. *J. Infect.* **2020**, *81*, e175–e177. [CrossRef] [PubMed]

183. Lam, T.T.-Y.; Jia, N.; Zhang, Y.-W.; Shum, M.H.-H.; Jiang, J.-F.; Zhu, H.-C.; Tong, Y.-G.; Shi, Y.-X.; Ni, X.-B.; Liao, Y.-S. Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins. *Nature* **2020**, *583*, 282–285. [CrossRef]

184. Wahba, L.; Jain, N.; Fire, A.Z.; Shoura, M.J.; Artiles, K.L.; McCoy, M.J.; Jeong, D.-E. An extensive meta-metagenomic search identifies SARS-CoV-2-homologous sequences in pangolin lung viromes. *MSphere* **2020**, *5*, e00160-20. [CrossRef]

185. Zuo, T.; Liu, Q.; Zhang, F.; Lui, G.C.-Y.; Tso, E.Y.; Yeoh, Y.K.; Chen, Z.; Boon, S.S.; Chan, F.K.; Chan, P.K. Depicting SARS-CoV-2 faecal viral activity in association with gut microbiota composition in patients with COVID-19. *Gut* **2021**, *70*, 276–284.

186. Kremer, F.S.; McBride, A.J.A.; Pinto, L.D.S. Approaches for in silico finishing of microbial genome sequences. *Genet. Mol. Biol.* **2017**, *40*, 553–576. [CrossRef]

187. Sah, R.; Rodriguez-Morales, A.J.; Jha, R.; Chu, D.K.; Gu, H.; Peiris, M.; Bastola, A.; Lal, B.K.; Ojha, H.C.; Rabaan, A.A. Complete genome sequence of a 2019 novel coronavirus (SARS-CoV-2) strain isolated in Nepal. *Microbiol. Resour. Announc.* **2020**, *9*, e00169-20. [CrossRef] [PubMed]

188. Yadav, P.D.; Potdar, V.A.; Choudhary, M.L.; Nyayanit, D.A.; Agrawal, M.; Jadhav, S.M.; Majumdar, T.D.; Shete-Aich, A.; Basu, A.; Abraham, P.; et al. Full-genome sequences of the first two SARS-CoV-2 viruses from India. *Indian J. Med. Res.* **2020**, *151*, 200–209.

189. Caly, L.; Druce, J.; Roberts, J.; Bond, K.; Tran, T.; Kostecki, R.; Yoga, Y.; Naughton, W.; Taiaroa, G.; Seemann, T. Isolation and rapid sharing of the 2019 novel coronavirus (SARS-CoV-2) from the first patient diagnosed with COVID-19 in Australia. *Med. J. Aust.* **2020**, *212*, 459–462. [CrossRef] [PubMed]

190. Hwang, K.-B.; Lee, I.-H.; Li, H.; Won, D.-G.; Hernandez-Ferrer, C.; Negron, J.A.; Kong, S.W. Comparative analysis of whole-genome sequencing pipelines to minimize false negative findings. *Sci. Rep.* **2019**, *9*, 3219. [CrossRef] [PubMed]

191. Tam, V.; Patel, N.; Turcotte, M.; Bossé, Y.; Paré, G.; Meyre, D. Benefits and limitations of genome-wide association studies. *Nat. Rev. Genet.* **2019**, *20*, 467–484. [PubMed]

192. Patron, J.; Serra-Cayuela, A.; Han, B.; Li, C.; Wishart, D.S. Assessing the performance of genome-wide association studies for predicting disease risk. *PLoS ONE* **2019**, *14*, e0220215.

193. Ray, M.; Sarkar, S.; Rath, S.N.; Sable, M.N. Elucidation of Genome Polymorphisms in Emerging SARS-CoV-2. *bioRxiv* **2020**. [CrossRef]

194. Ramírez, J.D.; Muñoz, M.; Hernández, C.; Flórez, C.; Gomez, S.; Rico, A.; Pardo, L.; Barros, E.C.; Paniz-Mondolfi, A.E. Genetic diversity among SARS-CoV2 strains in South America may impact performance of molecular detection. *Pathogens* **2020**, *9*, 580. [CrossRef]

195. Fang, B.; Liu, L.; Yu, X.; Li, X.; Ye, G.; Xu, J.; Zhang, L.; Zhan, F.; Liu, G.; Pan, T. Genome-wide data inferring the evolution and population demography of the novel pneumonia coronavirus (SARS-CoV-2). *bioRxiv* **2020**. [CrossRef]

196. Joshi, A.; Paul, S. Phylogenetic analysis of the novel coronavirus reveals important variants in Indian strains. *bioRxiv* **2020**. [CrossRef]

197. Zhou, Y.; Zhang, S.; Chen, J.; Wan, C.; Zhao, W.; Zhang, B. Analysis of variation and evolution of SARS-CoV-2 genome. *Nan Fang Yi Ke Da Xue Xue Bao J. South. Med. Univ.* **2020**, *40*. [CrossRef]

198. Lopes, L.R.; de Mattos Cardillo, G.; Paiva, P.B. Molecular evolution and phylogenetic analysis of SARS-CoV-2 and hosts ACE2 protein suggest Malayan pangolin as intermediary host. *Braz. J. Microbiol.* **2020**, *51*, 1593–1599. [CrossRef] [PubMed]

199. Lanza, D.C.; Lima, J.P.; Jerônimo, S.M. Design and in silico validation of polymerase chain reaction primers to detect severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). *Bioinformatics* **2020**. [CrossRef]

200. Toms, D.; Li, J.; Cai, H.Y. Evaluation of WHO listed COVID-19 qPCR primers and probe in silico with 375 SERS-CoV-2 full genome sequences. *medRxiv* **2020**. [CrossRef]

201. Bisht, N.; Singh, B. Role of computer aided drug design in drug development and drug discovery. *Int. J. Pharm. Sci. Res.* **2018**, *9*, 1405–1415.

202. Choudhury, M.; Saikia, R. Essential basic protocol in computer aided drug designing: Efficiency and challenges. *Int. J. Biotechnol. Bioeng.* **2018**, *4*, 77–80.

203. Song, C.M.; Lim, S.J.; Tong, J.C. Recent advances in computer-aided drug design. *Brief. Bioinform.* **2009**, *10*, 579–591.

204. Beninger, P.; Ibara, M.A. Pharmacovigilance and biomedical informatics: A model for future development. *Clin. Ther.* **2016**, *38*, 2514–2525. [CrossRef]