ELSEVIER

Computational drug repositioning for rare diseases in the era of precision medicine

Brian Delavan[1,2], Ruth Roberts[3,4], Ruili Huang[5], Wenjun Bao[6], Weida Tong[1] and Zhichao Liu[1]

[1] National Center for Toxicological Research, US Food and Drug Administration, Jefferson, AR 72079, USA
[2] University of Arkansas at Little Rock, Little Rock, AR 72204, USA
[3] ApconiX, BioHub at Alderley Park, Alderley Edge SK10 4TG, UK
[4] University of Birmingham, Edgbaston, Birmingham B15 2TT, UK
[5] National Center for Advancing Translational Sciences, National Institutes of Health Rockville, MD 20850, USA
[6] SAS Institute Inc., Cary, NC, USA

There are tremendous unmet needs in drug development for rare diseases. Computational drug repositioning is a promising approach and has been successfully applied to the development of treatments for diseases. However, how to utilize this knowledge and effectively conduct and implement computational drug repositioning approaches for rare disease therapies is still an open issue. Here, we focus on the means of utilizing accumulated genomic data for accelerating and facilitating drug repositioning for rare diseases. First, we summarize the current genome landscape of rare diseases. Second, we propose several promising bioinformatics approaches and pipelines for computational drug repositioning for rare diseases. Finally, we discuss recent regulatory incentives and other enablers in rare disease drug development and outline the remaining challenges.

Introduction

Most rare diseases have a genetic etiology, affect a small proportion of the population (usually less than 1/1500 in the USA or 1/2000 in Europe) but are severe and life-threatening [1–3]. Although rare diseases are themselves infrequent by definition, collectively they are a common occurrence. There are more than 7000 rare diseases based on the European Organization for Rare Diseases (EURORDIS) statistics (http://www.eurordis.org/about-rare-diseases). However, there have only been ~600 treatment options available since the Orphan Drug Act of 1983 was passed [4]. The average time to diagnosis of a rare disease is more than 7 years. Over one-third of children with a rare disease will not live more than 5 years, and about 35% of these children will die within the first year of life [5].

The fundamental challenge of orphan drug development is a lack of knowledge about pathophysiology, etiology and the natural history of rare diseases. Few patients are available and, together with their geographical dispersal, clinical trials are often impractical [6]. Also, researchers have great difficulty in gauging the genetic origin of rare diseases [1]. The causative genetic mutations are either hereditary (even when the disease has a late onset in the patient's life) or they are caused by a new mutation (*de novo*) [7]. Like common diseases, heterogeneity also exists in rare diseases, which makes it extremely challenging to distinguish patients with different morphological features or genetic variants and then look for the right treatment options. One example is cystic fibrosis (CF), which is accounted for by the genetic mutation of the transmembrane conductance regulator (CFTR) gene. There are ~2000 identified mutations within the CFTR gene from CF patients. Among the 2000 identified CFTR mutations, F508del and G551D are major mutations that are carried by >90% of CF patients. However, the associated phenotypic outcomes of the two mutations are distinct. The F508del mutation is mainly associated with CFTR folding impairment, stability at the endoplasmic reticulum and plasma

Corresponding authors: Tong, W. (weida.tong@fda.hhs.gov), Liu, Z. (zhichao.liu@fda.hhs.gov)

1359-6446/Published by Elsevier Ltd.
https://doi.org/10.1016/j.drudis.2017.10.009

membrane, and chloride channel gating. The G551D mutation is mainly related to channel gating alternation [8,9]. The only FDA-approved drug, ivacaftor, is only effective in patients with the 33 genetic mutation types such as the G551D mutation, which only covers a total of 6% of CF patients (https://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm559212.htm). Meanwhile, there are still a substantial number of CF patients carrying the F508del mutation without a treatment option.

The advent of next-generation sequencing (NGS) has changed the landscape of rare disease research, presenting the opportunity for the causative genes of rare diseases to be identified at an unprecedented pace and resolution [1]. NGS is also considered as a key technology for advancing precision medicine [10]. Many genetic variants of rare diseases have been detected and the data are publicly accessible. However, there are still many undetected genes associated with rare diseases [7,11,12]. Ongoing efforts are being made and will lead to substantial improvement in our understanding of the genetic origin of rare diseases. For example, the International Rare Diseases Research Consortium (IRDiRC) set a goal of developing the capacity to diagnose all of the rare diseases and to establish 200 new or repurposed therapies for rare diseases by the year 2020 [12].

How to translate the accumulated genetic knowledge to facilitate rare disease treatment development is still an open question [13]. First, to identify and validate therapeutic targets of rare diseases is a great challenge. Even if a causative genetic mutation in a patient with rare diseases is detected, there is no guarantee that a therapeutic option might arise from this knowledge. This is because the mutated protein might be unsuitable as a therapeutic target for a variety of reasons such as inaccessibility or lack of suitability as a small molecule target [14]. In this context, the current drug design paradigm has proved generally successful in inhibiting therapeutic targets in rare diseases with gain-of-function mutations [15]. Rare diseases with gain-of-function mutations, like most common diseases, are defined as the activation of specific pathways or the ectopic activity in relation to the proteins, which aligns well with the current concept of target identification. However, there are many rare diseases that are caused by loss-of-function where the impairment of a particular protein drives the etiology [15]. Therefore, a novel approach for translating knowledge of loss-of-function genetic variants into clinical use is urgently needed.

Drug repositioning that aims to find new uses for existing drugs is considered as an effective and alternative paradigm of drug development [16]. Computational drug repositioning provides a systematic and rational solution for identifying treatment options as compared with conventional drug repositioning approaches arising from serendipity or close clinical observation [17–20]. Linking the genetic findings of rare disease and drug repositioning into the same framework to accelerate drug development for rare diseases is imperative and is also a necessary practice for precision medicine. In this review, we summarize the current progress in research on the genetic origins of rare diseases. Then, we propose several novel strategies to integrate these accumulated genetic findings into computational drug repositioning frameworks for the development of treatments for rare diseases (Fig. 1). Finally, we discuss the remaining challenges and future perspectives in this field.

Genetic landscape of rare diseases

During the past decade, much progress has been made in the detection of the genetic origin of rare diseases even though patient recruitment is a challenge for obtaining samples and for carrying out clinical studies for the development of treatment options. This has resulted from the advancement of new techniques, the assistance of social media and the policy shifts of regulatory agencies [1,21,22]. Particularly, NGS techniques have greatly enabled the detection of the possible genetic basis of rare diseases [23].

To date, the molecular level etiology information of around one-third of rare diseases has been uncovered, although many causative genes of rare diseases remain to be identified [24,25]. Based on the Orphanet data [26], there are a total of 6289 rare diseases with a causative gene relationship, which corresponds to 3343 rare diseases and 3398 genes. Among the 3343 rare diseases, 2442 (2442/3343; 73%) have a single causative gene (Fig. 2a). Among 6289 rare disease and causal gene relationships, 5032 (4171 unclassified + 715 loss-of-function + 146 gain-of-function) belong to germline mutations in the causative genes, which account for >80% of mutation types (Fig. 2b). It can also be seen that 4171 unclassified mutations (4171/5032; 82.9% of total germline mutations) remain to be annotated at the functional level.

Genetic variants have been implicated in mutation functions and phenotypic outcomes. However, genetic variants such as structural variants are still considered as one of most difficult to interpret with regard to their functional consequence [21]. Structural variants comprise different unbalanced forms of variants such as deletion, insertion, reduplication and balanced forms such as translocation and inversion. ClinVar is a database for the clinical significance of mutations [11]. Based on ClinVar, there are a total of 52 944 genetic mutations from 3502 unique rare-disease-associated genes that distribute into different chromosome locations. The types of 52 944 rare-disease-related genetic variants include single nucleotide variant (SNV), deletion, duplication, insertion, undetermined variant, NT expansion, protein only, copy number loss, copy number gain, inversion, short repeat, structural variant. Among 13 mutation types, SNV, deletion and duplication are the three-most-frequent mutation types (Fig. 2c). Fig. 2d shows the SNV distribution across the chromosomes. Chromosomes 17, 13 and X contain more SNVs compared with other chromosomes.

How to interpret and annotate the genotype information in the context of rare disease phenotypic outcome is key in applying the genomic findings to the clinical practice [27]. Much effort has been made to standardize, integrate and associate the rare-disease-related genotypic and phenotypic relationship [28–32]. Table 1 summarizes the publicly available resources and efforts of genotype and phenotype information about rare diseases. For example, Human Phenotype Ontology (HPO, http://human-phenotype-ontology.github.io/about.html) [28,33–35] consists of >110 000 phenotypic annotations for >10 000 common and rare diseases. HPO is widely used to compare the phenotypic overlap between rare and common diseases based on their common genetic variants [36–39]. Another example is the Monarch Initiative (https://monarchinitiative.org/) [29], which was developed by a consortium effort and aims to integrate diverse rare-disease-related genotypic and phenotypic information, including OMIM, Orphanet, HPO, among others, to accelerate un-

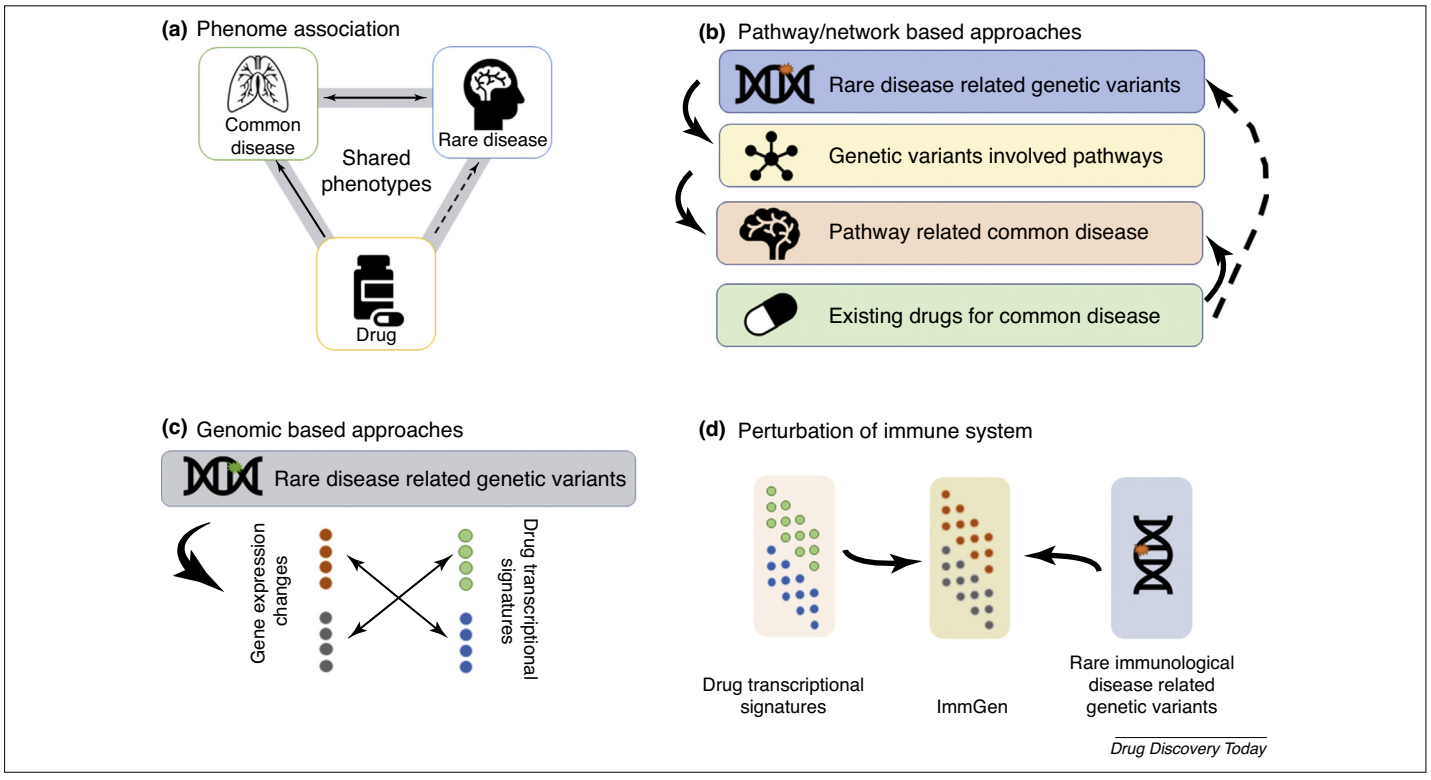

FIGURE 1

The proposed computational drug repositioning approaches for rare disease therapy. **(a)** Phenotype association. The shared phenotype information between rare and common diseases implies the repositioning opportunities for rare disease by using the drugs originally developed for its associated common disease. **(b)** The rationale behind pathway/network approaches is the shared pathways and biological process between rare and common diseases create the opportunity for rare disease therapy by using drugs perturbing the shared pathways or involving the shared biological processes. **(c)** Genomic-based drug repositioning looks for the repositioning candidate with transcriptional signature that is reversely correlated with rare-disease-related gene expression. **(d)** Immune-related drug repositioning aims to seek the rare-immunological-disease-related gene expression profiles caused by genetic variants. Then, the drug transcriptional signature reversely correlates with the immune-related gene expression and could be potentially used for therapy development.

derstanding disease mechanisms and to promote diagnosis and therapy development. Furthermore, Mammalian Phenotype Ontology (MPO, http://www.informatics.jax.org/vocab/mp_ontology/) [31] is a well-established ontology, centralizing the phenotypic information on rare and common diseases across different species. MPO provides a cross-species phenotypic mapping strategy to translate the phenotypic findings from animal models to clinical use in humans.

Application of NGS for rare diseases

NGS techniques have enabled the comprehensive sequencing of DNA that is relevant to rare diseases at a much higher throughput and much lower costs than previously possible. Compared with the conventional genetic mutation detection methodologies [40–44], NGS techniques provide more-detailed and high-resolution information of genetic variants (see Fig. S1 in Supplementary material online). There are three major NGS techniques: whole genome sequencing (WGS), whole exome sequencing (WES) and targeted sequencing (TS) for rare disease diagnosis. WGS and WES mainly focus on exploring the genetic basis without any prior knowledge and hypothesis. In WES the ~1% protein-coding region of the human genome as a target is enriched by different capture strategies before being sequenced by NGS. Because most genetic mutations of rare diseases happen in the protein-coding region, WES has rapidly become one of the main tools for studying the genetic causes of Mendelian disease. The initial successful cases for identification of genes responsible for several rare diseases by using WES include recessive Miller syndrome [45], the dominant Freeman–Sheldon syndrome [46] and dominant Schinzel–Giedion syndrome [47]. Unlike exome sequencing, WGS provides a much broader coverage of the genome with decreasing cost. Belkadi *et al.* [48] performed a comprehensive comparison for WES and WGS in six unrelated Caucasian patients with isolated congenital asplenia. WGS could generate more and much higher quality SNVs and indels in the exome regions regarding coverage depth, genotype quality and minor read ratio. Furthermore, WES is not capable of effectively identifying copy number variants (CNVs) because the CNV location is beyond the targeted region. Targeted sequencing was a hypothesis-driven approach and aims to provide much deeper detection power of genetic variants within predefined gene panels [49,50]. Stessman *et al.* [50] applied targeted sequencing to detect the gene-disruptive mutations for neurodevelopmental disorders (NDDs). Within a panel of a 208 candidate genes for targeted sequencing, a list of 91 genes (38 new NDD genes) and their mutations were identified based on a larger cohort of NDD patients (>11 730 cases vs >2867 controls).

Although NGS techniques are becoming routine genetic detection tools for rare disease diagnosis, data analysis still poses great challenges and bioinformatics plays an important part in harmonizing this process. There are five major steps for NGS data analy-

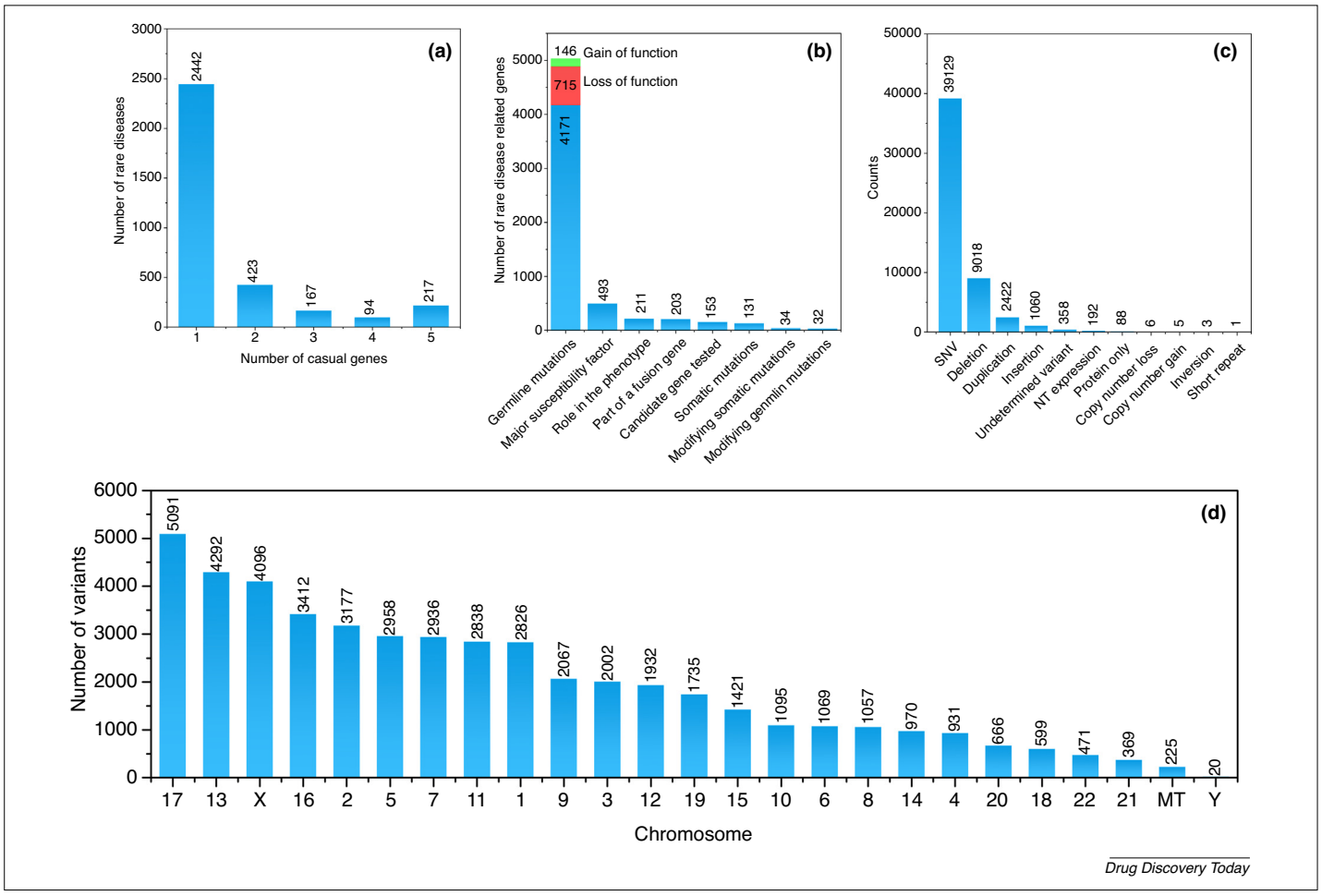

FIGURE 2

The statistics of rare disease genetic information. **(a)** The relationship between rare disease and its causal genes based on Orphadata. The number of rare diseases was counted by their causal genes and the number of rare diseases was plotted versus the number of causal genes. **(b)** The known mutation origin and functions of rare diseases based on Orphadata (http://www.orphadata.org/cgi-bin/index.php/). The number of rare disease causal genes was plotted versus different genetic mutation types and functions. **(c)** The structure variant distribution of rare diseases based on ClinVar. The bar chart was drawn based on the number of genetic variants types. **(d)** The genetic variant distribution across different chromosomes based on ClinVar database. The genetic variants were counted by each chromosome and plotted. The Orphadata data are downloaded from http://www.orphadata.org/cgi-bin/index.php/and the ClinVar data were retrieved and extracted from //ftp.ncbi.nlm.nih.gov/pub/clinvar/tab_delimited/.

sis: quality control of raw readers, alignment, variant calling, variant annotation and variant evaluation [51]. A plethora of tools were developed for different analysis steps that provide a wide range of selection based on different purpose and rare disease categories. How to select the right tool to construct the NGS data analysis pipeline is still elusive. Unlike the array-based techniques with predefined probes such as microarrays and genotyping arrays, NGS techniques result in an uncontrolled number of variants. Consequently, when the different bioinformatics pipelines are employed, a different number of variants is generated. Therefore, the confidence level of detected variants should be defined based on statistical and biological relevance. For example, Genome in a Bottle (GIAB) Consortium developed a WGS reference sample NA12878 for generating a benchmark genotype data including SNVs and small indels by integrating the variant-calling results from different algorithms. Furthermore, the high-confidence regions for variant calling were also defined, which provides gold standard genotype datasets for further evaluating the newly developed calling algorithm and calibrating the novel sequencing techniques [52]. GIAB datasets have been applied widely for germline mutation and somatic mutation calling methodology comparison [53–55].

Paths toward rare disease therapy

The emerging techniques have accelerated the pace of the identification of rare disease genetic variants [1]. However, the majority of detected variants remains to be translated into treatment options. Here, we summarize and propose several computational drug repositioning approaches for facilitating this process (Fig. 1).

Phenotype associations between rare and common diseases

The candidate gene studies and genome-wide association studies (GWAS) have identified a large number of single nucleotide polymorphism (SNP) trait and common and complex disease relationships [56], which can be used to prioritize genetic findings and further identify therapeutic targets [57–59]. Sanseau *et al.* [60] assessed the utility of GWAS for identifying alternative uses of existing drugs. It was found that a list of 155 genes identified from

TABLE 1

Publicly available resources and efforts in genotype and phenotype information of rare disease

Databases/consortiums	Web link	Remarks
Public databases		
Orphanet	http://www.orphadata.org/cgi-bin/index.php	Orphanet is a comprehensive resource on rare diseases that provides rare disease information including rare-disease-associated genes, clinical signs, epidemiological data and rare disease classification
Online Mendelian Inheritance in Man (OMIM)	http://www.omim.org/	OMIM is a comprehensive, authoritative compendium of human genes and genetic phenotypes that contains the information about all the Mendelian diseases and over 15 000 genes and their variants
ClinVar	http://www.ncbi.nlm.nih.gov/clinvar/intro/	ClinVar provides freely accessible, reported relationships among human variations and phenotypes, with supporting evidence. The human variants information is also linked to Orphanet and OMIM databases
COSMIC	http://cancer.sanger.ac.uk/cosmic	COSMIC is designed to store and display somatic mutation information and related details and contains information relating to human cancers. The somatic mutation on rare cancers could be retrieved from COSMIC
Database of genomic variation and phenotype in humans using ensemble resources (DECIPHER)	https://decipher.sanger.ac.uk/	DECIPHER is an interactive web-based database that incorporates a suite of tools designed to aid the interpretation of genomic variants of rare disease
The NHGRI GWAS Catalog	http://www.genome.gov/gwastudies	A catalog of published genome-wide association studies that provides SNP–trait associations
DisGeNET	http://www.disgenet.org/web/DisGeNET/menu	DisGeNET is a curated effort and aims to integrate the disease and gene relationship from public databases and literature mining
Consortium efforts		
Care for Rare	http://care4rare.ca/	CARE for RARE is a Canadian nationwide research program focusing on the improvement of the diagnosis and treatment of rare diseases. Currently, their research embraces 637 different rare disease studies with >1000 rare disease patients with 81 novel rare diseases causing genes identified
Finding of Rare Disease Genes in Canada (FORGE Canada)	http://www.genomebc.ca	FORGE Canada is a national consortium of clinicians and scientists using next-generation sequencing technology to identify genes responsible for a wide spectrum of rare pediatric-onset disorders present in the Canadian population
The Centers for Mendelian Genomics (CMG)	http://www.mendelian.org/	The CMG aims to discover genetic basis of Mendelian disorders in two main ways: applying novel sequencing techniques for rare disease research and collaboration with other rare disease research consortiums
The Global Alliance for Genomics and Health	https://genomicsandhealth.org/	The Global Alliance for Genomics and Health (Global Alliance) was formed to help accelerate the potential of genomic medicine to advance human health by using emerging sequencing techniques
The UK 100 000 Genomes Project	http://www.genomicsengland.co.uk/	The project will sequence 100 000 genomes from around 70 000 people. Participants are NHS patients with a rare disease, plus their families and patients with cancer
Deciphering Developmental Disorders (DDD)	http://www.ddduk.org/	The DDD study aims to advance clinical genetic practice for children with developmental disorders by the systematic application of the latest microarray and sequencing methods while addressing the new ethical challenges raised
The International Rare Diseases Research Consortium (IRDiRC)	http://www.irdirc.org/	IRDiRC teams up researchers and organizations investing in rare disease research to achieve two main objectives by the year 2020, namely to deliver 200 new therapies for rare diseases and the means to diagnose the most rare diseases
The Genetic Disorders of Mucociliary Clearance Consortium (GDMCC)	http://www.rarediseasesnetwork.org/cms/GDMCC	The Genetic Disorders Of Mucociliary Clearance Consortium is a clinical research network created to improve the diagnostic testing and treatment of rare airway diseases, including primary ciliary dyskinesia (PCD), variant forms of cystic fibrosis (CF), pseudohypoaldosteronism (PHA) and now idiopathic bronchiectasis and NTM pulmonary disease
Phenotype-related resources		
Human Phenotype Ontology (HPO)	http://human-phenotype-ontology.github.io/about.html	HPO aims to provide a standardized phenotypic abnormalities vocabulary in human diseases, which includes ~11 000 terms and >115 000 annotations to rare diseases
Monarch	https://monarchinitiative.org/page/about	Monarch aims to integrate the phenotypic information by using semantics, which enables phenotypic-based comparison between the diseases within and across species
CentoMD®	https://www.centogene.com/mutation-database-centomd.html	CentoMD® is a commercial database that offers the genotype and phenotype relationships based on the >4.5 million variants from individuals by using exome sequencing techniques
Encyclopedia of Rare Disease Annotation for Precision Medicine (eRAM)	http://www.pediascape.org/eram/	eRAM offers *in silico* annotations for ~16 000 rare diseases, yielding 6147 human-disease-related phenotype terms, >30 000 mammalian phenotype terms, 10 202 phenotypes from UMLS, 18 815 genes and 92 580 genotypes, which could not only provide researchers new information about the mechanism of a rare disease but also facilitate clinical diagnosis and therapy development of rare diseases

TABLE 1 (*Continued*)

Databases/consortiums	Web link	Remarks
Mammalian Phenotype Ontology (MP)	http://www.informatics.jax.org/vocab/mp_ontology/	MP aims to classify and organize phenotypic information related to the mouse and other mammalian species. MP ontology allows comparing of data from diverse sources, across mammalian species, facilitating identifying appropriate experimental disease models, candidate disease genes and molecular signaling pathways discovery
Disease Ontology (DO)	http://www.disease-ontology.org	DO offers a standardized ontology for human rare and common disease and aims to provide the biomedical community with consistent, reusable and sustainable descriptions of human disease terms, phenotype characteristics and related medical vocabulary disease concepts

GWAS had been targeted by at least one existing drug or candidate in clinical trials. For 92 of 155 genes, the suggested drug indication was different from the original disease trait identified by GWAS, which implies that these new drug-indication pairs should be further verified for identifying new disease treatment options [59,61]. However, results from GWAS contain a high false-positive rate owing to the limitations posed by either technique or sample size [62,63]. Integration of electronic health records (EHRs) of various disease types from different ethnic groups to the dense genomic information presents a new vision of precision medicine [64]. Denny *et al.* [65] reported a novel paradigm named phenome-wide association study (PheWAS), which incorporated SNP–trait relationship identified from GWAS with the electronic medical records of genetic scanning from a large cohort of people with European ancestry. The PheWAS not only provided an extra verification of the results from GWAS but also revealed some potentially interesting associations. The PheWAS tremendously expanded the scale of SNP–trait relationship and provided more opportunities for looking for new uses of existing drugs. Rastegar–Mojarad *et al.* [66] combined PheWAS and DrugBank [67] to identify repositioning candidates for rare and common diseases. A total of 52 966 drug–disease pairs were enriched by the approach. Approximately 30% of 52 966 drug pairs were verified for known drug–disease relationships, ongoing clinical trials or literature reports. About 70% of drug pairs could be candidates for drug repositioning.

PheWAS-based drug repositioning techniques are mainly used for common and complex diseases. How to extrapolate the PheWAS approach for rare disease therapy is an illusion. Considering the limited sample size and heterogeneity nature of rare disease patients, the genetic information from genome-sequencing studies needs to be prioritized. However, the EHR data for rare diseases is not abundant enough to generate the statistical power needed to implement PheWAS approaches. Thus, the standardized rare disease phenotype ontologies are urgently needed to cluster rare-disease-related phenotype information together based on shared genotype information, which could precisely and accurately prioritize the rare disease phenotype and genotype associations. Efforts have been made to develop rare-disease-related phenotypic ontologies including HPO [28,34], the Monarch Initiative [29] and MPO [31]. These phenotype ontologies are widely used to accommodate coding rare-disease-related phenotypes derived from different sources, including EHRs, patient narratives and clinical interpretations. Furthermore, like the EHR data in the PheWAS approaches, phenotype ontologies could enhance rare disease gene identification [27]. Akawi *et al.* [36] developed a novel statistical approach to integrate phenotypic information from HPO and genotypic information from genome-sequencing 4125 family trios to improve the discovery rate of new causal genes for rare recessive disease. Furthermore, the rare and common phenotype ontologies could be used to explore potential repositioning opportunities. The rationale behind phenotype association is to link rare and common diseases based on their shared phenotypic information. The common phenotypic information between rare and common diseases indicates the genetic similarity between these diseases and further suggests plausible interventions for rare diseases by using a drug that was originally designed for a common disease to treat the rare disease (Fig. 1). Hoehndorf *et al.* [39] applied text-mining strategies to mine the phenotypic information, including signs and symptoms of >6000 rare and common diseases, by integrating the phenotypic-related ontologies such as disease ontology (DO), HPO and MPO. It was found that rare and common diseases with the same signs and symptoms were clustered together with many shared genes. This sharing of genes provides the potential opportunity for repurposing drugs designed for common diseases to rare disease therapy. The Centre for Therapeutic Target Validation (CTTV), initialized between the European Bioinformatics Institute (EMBL-EBI), GlaxoSmithKline (GSK) and the Wellcome Trust Sanger Institute (WTSI), aims to better understand the relationship between rare and common diseases. CTTV developed experimental factor ontology (EFO) by integrating rare and common disease-related phenotype and genotype ontologies and identified 20 common diseases and 85 rare diseases that share similar phenotypes [68].

Pathway- or network-based approaches

Genes with genetic variants might not be suitable 'druggable' targets. However, pathway or network approaches can be helpful in finding genes involved in general signaling networks or biological pathways, and could provide a list of proteins for therapeutic target identification [69]. For example, the Ras/MAPK syndromes (Noonan, LEOPARD, Costello and cardio-facio-cutaneous) are a class of rare developmental disorders caused by germline mutations of genes including PTPN11, SOS1, RASA1, NF1, KRAS, HRAS, NRAS, BRAF, RAF1, MAP2K1, MAP2K2, SPRED1, RIT1, SHOC2 and CBL. Ras/MAPK signaling pathways deregulated by cancerous somatic mutations exist in approximately one-third of all cancer types [70,71]. Naturally, it is assumed oncology drugs that could inhibit the Ras/MAPK signaling pathway components could be used to treat RASopathy-related rare development disorders. A mouse model was developed for verification of the oncology drug rapamycin for treating LEOPARD syndrome (LS) [72]. Specifically,

mice carrying the *ptpn11* mutation developed LS symptoms, and experiments verified that the mammalian target of rapamycin (mTOR) inhibitor: rapamycin, could reverse some of these, such as hypertrophic cardiomyopathy (HCM).

Linking the common disease with the rare disease based on a shared gene is an idea originally proposed by Goh *et al*. [73] who developed as a concept to identify the disease–disease relationship based on their shared pathways [74]. However, there is little knowledge about the underlying molecular mechanism of the influence of genetic variants on the pathways. This knowledge is crucial to understanding the pathogenesis of diseases. Kiel *et al*. [75] developed a structure–energy-based prediction and network modeling framework to uncover the different degrees of perturbation of the Ras/MAPK pathway by germline mutations and somatic mutations. By measuring quantitative activity changes in the pathway based on mutated 3D protein structure, the difference between germline RASopathy mutation and cancer mutations could be explained by switching the genes on and off and assessing the degree of protein–protein interactions. Furthermore, the binding constants and affinities could be different for the same protein with different disease-related mutations. In addition, the energy change noted in a pathway was higher with a somatic mutation compared with a germline mutation. Overall, these pathway- or network-based methodologies and conclusions are of great value in uncovering the impact of genetic variants on pathways, further facilitating target identification and subsequent treatment development for rare diseases.

Genomic data integration

Deciphering the effect of genetic variants on cellular processes such as gene expression at the cellular or organism level is crucial in dissecting genetic contributions to phenotypic endpoints [76,77]. This also paves the way for linking genetic variants to treatment development because vast amounts of drug transcriptome data in different cell types and organisms are publicly available [78–80]. The correlation between genetic variants and gene expression has been discussed and applied in the cancer genome field (Table 2). Although the proposed approaches are tailored to driver gene enrichment and patient survival, it also could be applied for treatment development. For example, Masica and Karchin [81] proposed a statistical strategy with network analysis for correlating somatic mutation and gene expression and applied it to 149 human glioblastoma (GBM) samples. They found that somatic mutations of 41 genes were highly related to GBM progression and patient survival. Bertrand *et al*. [82] developed a network approach by integrating SNP, CNV and gene expression for driver-gene enrichment. The proposed methodology was also applied to GBM and a novel driver gene TRIM24 was found and experimentally verified. In addition, the methodology was used for >1000 tumor samples from five different cancer types for identifying modes of synergistic action, which could be potentially used for combination drug design for cancer treatment. Ping *et al*. [83] developed a hybrid integrative approach named CMDD by combining partial least squares regression and network methods covering multiple omics profiles such as CNV, DNA methylation, miRNA and gene expression. CMDD was also applied to GBM, and six other cancer types and the genes involved in the enriched modules were correlated with overall patient survival. Ding *et al*. [84] presented a novel hierarchical Bayes graphic modeling approach for symmetrically qualifying the effect of somatic mutation on gene expression across 12 cancers. Some very interesting conclusions were drawn: (i) the patients carried the same somatic mutations, which influenced different downstream gene expression; (ii) some somatic mutations are conserved across cancer types. Gerstung *et al*. [85] developed a computational approach for detecting the phenotypic heterogeneity caused by distinct genotype and applied it to 124 patients with myelodysplastic syndrome (a rare cancer) and with TCGA acute myeloid leukemia (AML). It was found that one or more genetic variants were correlated with ~20% of all genes, which dictated 20–65% of

TABLE 2

Data integration strategy for correlating genetic mutation and gene expression

Data profiles	Diseases	Methodology	Notes	PubMed ID
Somatic mutation and gene expression	Glioblastoma (GBM)	Fisher's exact test with network analysis	The somatic mutation and gene expression are needed for each patient	21555372
SNP, CNV and gene expression	GBM and five other cancer types	Network analysis for integrative data including SNP, CNV and gene expression for driver gene enrichment	OncoIMPACT is developed and source code is available from http://sourceforge.net/projects/oncoimpact	25572314
CNV, methylation, miRNA and gene expression	GBM and six other cancer types	Partial least squares regression and network analysis	The results were verified by survival analysis and a core gene module of 17 genes was enriched for candidate GBM driver genes	25653168
Somatic mutation and gene expression	12 pan cancers	Hierarchical Bayes statistical model http://compbio.bccrc.ca/software/xseq/	Patient genetic heterogeneity was observed and some mutation types were conserved across cancer types	26436532
Germline mutation, miRNA, transcription factor, gene expression	Cystic fibrosis	miRNA transcription factor feed-forward loop construction by cumulative hypergeometric test	The 48 repurposing candidates were enriched for cystic fibrosis treatment, 26 of 48 candidates were verified by literature survey or existing clinical trials	25484921
Somatic mutation and gene expression	Myelodysplastic syndromes and acute myeloid leukemia (AML)	Principal component analysis (PCA) with schematic linear decomposition	One or more genetic lesions correlates with expression levels of ~20% of all genes, explaining 20–65% of observed expression variability. Differential expression patterns vary between mutations and reflect the underlying biology	25574665

gene expression variability. These proposed methodologies have been successfully used to uncover genetic mutation and gene expression relationships for common or rare cancers. It is worth investigating the utility of these approaches in the rare diseases field to decipher germline mutations and their influence on gene expression profiles.

Furthermore, dysfunctional noncoding RNA such as miRNA and lncRNA in different biological processes often leads to disease [86]. The genetic mutation can change the binding affinity to miRNA, impairing gene expression and contributing to the phenotypic expression of the diseases [87]. Liu *et al.* [88] introduced a feed-forward loop concept into the drug repositioning field and applied it for the development of treatment for CF by integrating information including germline mutation, miRNA, transcription factors (TFs) and gene expression. Then, 15 CF-specific miRNA-TF feed-forward loops were enriched by using a cumulative hypergeometric test. Finally, by investigating the perturbation of obtained CF-specific feed-forward loops with small molecules, a list of 48 CF-repurposed candidates was proposed. Among the 48 repurposed candidates for CF, 26 candidates were verified by literature survey and existing clinical trials.

Once the correlation between genetic variants and gene expression is established, drug transcriptome data could be applied to look for repositioning opportunities (Table 3). The Connectivity Map (CMap) [80] as the key source has been widely applied to drug repositioning fields [89,90]. For example, Dudley *et al.* [91] proposed a novel approach that aims to look for inverse drug–disease relationships by comparing the disease signature generated from the Gene Expression Omnibus (GEO) databases [92] and drug signatures obtained from CMap. They found several repurposing candidates for treating inflammatory bowel disease (IBD) and these were verified by in *vitro* assays.

Besides CMap, several large toxicogenomics efforts such as TG-GATEs [79] and DrugMatrix [78] have accumulated hundreds of drug transcriptome data profiles at multiple time, dose, assay type points. Iskar *et al.* [93] identified a large set of drug-induced transcriptional modules with CMap and DrugMatrix data that are from human cancer cell lines and from rat liver *in vivo*. They found that 70% of drug-induced transcriptional modules were conserved in both assay types, which suggests that toxicogenomics data could also be used for drug repositioning, although further comprehensive assessment is needed. Furthermore, miRNAs have been considered as novel, and promising therapeutic targets against various diseases [94,95]; several miRNA and small-molecule relationship databases such as SM2miR [96] and Pharmaco-miR [97] were constructed by curation from the literature or *in silico* prediction.

Perturbation of the immune system

There are ~300 immunological rare diseases based on Orphanet [98]. One example is Kostmann syndrome, which affects myelopoiesis and causes severe congenital neutropenia. Kostmann syndrome is usually with life-threatening bacterial infections in infancy [99]. There is currently no effective treatment option for Kostmann syndrome. The regular option including filgrastim aims to improve neutrophil counts and immune function, which increases risk of AML for patients [100]. A better understanding of how the existing drugs affect or trigger the immune system could pave a way to develop the treatment of immunological rare diseases.

During the past decade, a lot of large data compendia on the immune system have been generated and publicly available [101]. The National Institute for Allergy and Infectious Disease (NIAID) in the NIH implemented an ImmPort (https://immport.niaid.nih.gov) to share the molecular and clinical data of immune-related studies including population genetics analysis about immune systems, HLA region genomics in immune-related diseases [102]. The Immunological Genome Project Consortium developed the pilot studies named ImmGen (http://www.immgen.org), which aims to provide a microarray gene expression data atlas of mouse immune-related cell lines [103]. Those datasets could be applied to verify different hypotheses of repurposing existing drugs for rare immunological disease treatment development. Kidd *et al.* [104] mapped transcriptional signatures from CMap to 304 immune-cell state-change signatures in the ImmGen compendium, which generates 69 995 interaction pairs. The interaction could be further linked to different kinds of rare immunological diseases by comparing their shared genomic similarity, which could generate a prioritized repurposing drug list for further wet lab evaluation.

TABLE 3

Drug–transcriptome and drug–miRNA relationship resources

Databases	Web link	Remarks
The Connectivity Map (CMap)	https://www.broadinstitute.org/cmap/	Provides a comprehensive drug transcriptional responses of 1309 drugs or lead compounds in the clinical trials to six or seven different cancer cell lines
Open TG-GATEs	http://www.toxico.nibiohn.go.jp/english/	TG-GATEs consists of the comprehensive toxicogenomic profiles of 170 compounds with four different assay types (human/rat *in vitro*/*in vivo*) and multiple time and dose points in rat liver and kidney. The histopathological profiles for compounds are also available
DrugMatrix	https://ntp.niehs.nih.gov/drugmatrix/index.html	DrugMatrix contains toxicogenomic profiles for 638 different compounds from Codelink and Affymetrix platforms, covering multiple organisms including liver, kidney, heart, bone-marrow, spleen and skeletal muscle
SM2miR	http://www.bioinfo.hrbmu.edu.cn/SM2miR	SM2miR is a manual curated database that collects and incorporates the experimentally validated small molecules and miRNA relationship from around 20 species by literature survey
Pharmaco-miR	http://www.pharmaco-mir.org/	Pharmaco-miR identifies associations of miRNAs, genes and drugs by integrating PharmaGKB database and *in silico* prediction

Other approaches

Although the advances in genomics technology have innovated the way and tremendously accelerated the pace to unravel etiology and pathogenesis of rare diseases, many rare diseases still have limited genetic and genomics information available. Therefore, other *in silico* approaches, including chemical-based approaches and text-mining-based approaches, provide alternative means to enhance drug repositioning candidates for rare disease therapy [105].

Knowledge-based drug repositioning aims to uncover the 'hidden' knowledge on disease mechanism and utilize off-target effects of drugs by using available information on drugs, diseases and targets by employing bioinformatics or chemoinformatics approaches [20,106]. For example, the Harriet Lane Handbook (HLH), edited by Johns Hopkins School of Medicine, is widely used as a manual for pediatric information. Blatt and Corey [107] combined HLH and PubMed literature to explore potential repositioning candidates for rare diseases. In addition, Liu *et al*. [108] proposed a concept to reuse oncology drugs for rare disease treatment development. The rationale behind the study is three layers of similarity between rare diseases and cancers, including rare disease cancer predisposition, shared genes and pathways, germline mutation and somatic mutation continuum.

Structure-based drug repositioning, such as molecular docking, is widely applied in the drug discovery process [109–111]. There are two kinds of docking strategies. For known rare-disease-related targets, the small molecules of existing drugs could be docked to a crystallized structure or a structure from a homolog model. Alternatively, approved drugs and lead compounds could be inversely virtually screened against all the rare-disease-related targets. Many rare-disease-related protein structures are crystallized and stored in the Protein Data Bank (PDB, https://www.rcsb.org/pdb) [112]. These crystallized proteins could be used to implement docking approaches for rare disease repositioning candidate discovery.

Text-mining-based drug repositioning takes full advantage of growing public biomedical resources, EHR data and patent and healthcare-related social media to mine hidden relationships between diseases and existing drugs [113–116]. For example, Elsevier® developed a sophisticated platform to seamlessly link the PubMed literature, semantic ontologies and advanced visualized tools to research treatment options for rare diseases. For example, oncology drugs such as rapamycin were identified for potential use to treat congenital hyperinsulinism (CHI) (https://www.elsevier.com/connect/repurposing-drugs-helps-patients-with-rare-diseases). Another example is Linguamatics, which employs the I2E techniques to search domain knowledge bases by scanning literature and other text-based biomedical documents to explore novel drug and disease relationships (https://www.linguamatics.com/life-sciences-applications/drug-repurposing).

Right tools for the right purpose

Different drug repositioning approaches have been developed based on different kinds of data profiles. Considering the limited information for rare diseases, it is necessary to wisely choose the computational drug repositioning approach. Here, we compare the *in silico* drug repositioning approaches in the context of rare diseases, aiming to facilitate method selection (Table 4). For example, the genetic information from Orphanet [24] or OMIM [7] can be easily mapped to the PDB to see which crystalized protein structures are available, setting the foundation for molecular docking approaches. Meanwhile, the functions of genetic variants of rare diseases have been accumulated in databases such as ClinVar [11], which divides the function of genetic variants into categories such as gain-of-function or loss-of-function. This division could be used to further filter for molecular docking approaches because we know the loss-of-function is not suitable for docking approaches. In addition, some rare-disease-related proteins function in the cell rather than on the protein surface, which also creates hurdles for applying docking approaches. Another example is CMap-based approaches, which requires comparing the rare disease transcriptomic signatures with the drug transcriptomic signatures. Therefore, we can first check whether the genomics data for rare diseases are available in public genomics repositories such as Gene Expression Omnibus (GEO) [117] or ArrayExpress [118]. By contrast, there is a large amount of text-based information disseminated in the public BioMed resources, such as PubMed and across social media data healthcare forums. The text-mining strategy could be employed to mine the rare disease and drug association, such as off-label uses, which could generate the new hypothesis for rare disease therapy development. It is important to note that no single computational drug repositioning could be 'magic' all the time. In this review, we suggest a consensus strategy by integrating different approaches under the same framework to provide more-confident repositioning candidates for rare disease therapy.

Verification of repurposing candidates

Computational drug repositioning provides a rapid turnaround list of repositioning candidate drugs. The challenge is to experimentally verify the efficacy and safety of these and to move the drugs forward into clinical trials. Currently, most *in silico* drug repositioning approaches are verified by animal-based *in vitro* or *in vivo* models [90,91,119–121]. Moving these *in silico* findings toward clinical application is challenging owing to difficulties in patient recruitment, which are especially hard with patients with rare diseases. About 30% of clinical Phase III studies fail as a result of patient enrollment problems [122]. Therefore, a lot of proposed repositioning candidates remain at the report or literature level. Patient registries that have been created by patient advocacy groups, non-profit organizations, government agencies and companies facilitate progress in the enrollment and retention of patients with rare diseases. For example, the NIH established the Rare Diseases Clinical Research Network I (RDCRN I, http://www.rarediseasesnetwork.org/) to address the unique challenges of research on rare diseases. RDCRN studies >90 rare diseases at ~100 academic institutions. Patient advocacy groups actively participate in the research.

Concluding remarks

The NGS technologies have driven a dramatic shift in our understanding of rare diseases at a genome-wide scale [123]. Bioinformatics plays a central part and has become an important component in NGS data analysis, generating many algorithms and workflows. However, building a standard bioinformatics solution for NGS analysis and application to clinical practice remains to be carried out. Accurate and reliable NGS analysis ensures

TABLE 4

Comparison of computational drug repositioning approaches for rare disease therapies

Approaches	Selected resources	Web link	Advantages and disadvantages	
Molecular docking	AutoDock & AutoDock Vina	http://autodock.scripps.edu/	Molecular docking is an economic approach to carry out virtual screening for a large number of small molecules	(i) Limited number of rare-disease-related protein structures available in the PDB database (ii) High false-positive rate (iii) Not suitable of rare diseases owing to loss-of-function (iv) Not suitable rare disease causal protein functioning inside the cell membrane
	Schrödinger	https://www.schrodinger.com/		
	DOCK 3.7	http://dock.compbio.ucsf.edu/		
Pathway- or network-based approaches	KEGG pathways	http://www.genome.jp/kegg/pathway.html	Pathway- or network-based drug repositioning is to look for the rare and common disease shared pathways or involving the same biological process. Then, the drug for common disease could be repositioned to its associated rare diseases	(i) Only suitable for rare disease with known causal genes or proteins and their involved pathways (ii) Therapeutic target identification is still needed
	DAVID	https://david.ncifcrf.gov/		
	STRING 10	https://string-db.org/		
CMap-based approaches	The Connectivity Map (CMAP)	CMAP 02: https://portals.broadinstitute.org/cmap/	CMap-based approaches fully take advantage of massive amount of public transcriptional data to explore new drug indication relationship	(i) Availability of rare disease transcriptional data is a prerequisite (ii) The drug and disease transcriptional datasets are from different experimental settings including cell types, platforms, sites, which can cause 'batch effect' (iii) The drug transcriptional signatures are generated based on collapsed and ranked list of other real cell-specific 'treat vs control' signature, which can cause misleading
		L1000 assay: https://clue.io/cmap		
	Gene Expression Omnibus (GEO)	https://www.ncbi.nlm.nih.gov/geo/		
Text-mining-based approaches	Elsevier®	https://www.elsevier.com/connect/repurposing-drugs-helps-patients-with-rare-diseases	Vast amount of biomedical and pharmaceutical knowledge available in the literature, patents and databases contain a lot of hidden drug and diseases relationships	
	Linguamatics	https://www.linguamatics.com/life-sciences-applications/drug-repurposing	Sophisticated semantic and rare-disease-related ontologies are available to carry out text-mining-based repositioning	
	PubMed	https://www.ncbi.nlm.nih.gov/pubmed/	(i) Hypothesis-driven approach and more downstream verification process is needed (ii) How to establish the statistical confidence based on limited rare-disease-related case reports is an open question	
Knowledge-based approaches	Therapeutic Target Database (TTD)	http://bidd.nus.edu.sg/group/cjttd	Data integration of known drug–target disease to deduce or predict the drug repositioning opportunity The approaches combining different machine learning methodologies and taking advantage of different data profiles develop systems-biology-based platform for drug repositioning	(i) How to link the sparse rare-disease-related information and build a fit-for-purpose prediction model is a great challenge (ii) Domain experts and close clinical observation are needed to interpret and verify the findings
	DrugBank	https://www.drugbank.ca/		
	Clinical Outcomes Search Space (COSS™) (commercial license)	https://www.biovista.com/technology/		

patients with rare diseases receive the correct diagnosis and further facilitates the practice of precision medicine. However, inaccurate NGS testing can lead to poor or misleading results. Therefore, the drug makers, scientific researchers and reviewers need collaboratively to standardize NGS techniques application and performance evaluation strategies. The Precision FDA (https://precision.fda.gov/) program and the NIH Precision Medicine Initiative Cohort Program (https://www.nih.gov/precision-medicine-initiative-cohort-program) have been created to provide insightful vision on precision medicine taking advantage of emerging techniques.

Government-sponsored initiatives and accompanying policy shifts have also had a great impact on the development of treatments for rare diseases. For example, the FDA awarded 18 new research grants for the development of rare disease products or biomarkers or to defray the cost of clinical trials (http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm463539.htm). So far, 40 orphan disease products have been partially funded by grants from the 'orphan products grant program'. Furthermore, the FDA has developed four distinct and promising routes, enacting faster drug review and approval, shortening rare disease therapy development. In addition, there are other government-sponsored initiatives such as the Medical Research Council in the UK and the NIH National Center for Advancing Translational Sciences (NCATS). The NIH has established partnerships among public funders, the pharmaceutical industry and academic investigators, which will also be beneficial for the development of therapies for rare diseases [124]. In summary, under the precision medicine umbrella, the landscape of rare diseases has been redrawn by applying NGS techniques. The accumulated genomic data provide great opportunities for the development of treatments for rare diseases by providing insight into the possibility of drug repositioning. Enabling the translation of these novel findings to clinical practice of rare disease treatment development is the real practice of precision medicine. Several promising bioinformatics approaches, as summarized, have shown great potential in tailoring genomic findings to developing therapies for rare diseases. Combined with other established drug repositioning approaches and efforts from scientific communities, government agencies and pharmaceutical companies, the timing is excellent for furthering the development of innovative approaches and clinical practice toward precision medicine for rare diseases.

Conflicts of interest

Dr Ruth Roberts is co-founder and co-director of Apconix, an integrated toxicology and ion channel company that provides expert advice on nonclinical aspects of drug discovery and drug development to academia, industry and not-for-profit organizations.

The views presented in this article do not necessarily reflect current or future opinion or policy of the FDA or NIH. Any mention of commercial products is for clarification and not intended as endorsement.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.drudis.2017.10.009.

References

1 Boycott, K.M. *et al.* (2013) Rare-disease genetics in the era of next-generation sequencing: discovery to translation. *Nat. Rev. Genet.* 14, 681–691
2 Basch, E. (2010) The missing voice of patients in drug-safety reporting. *N. Engl. J. Med.* 362, 865–869
3 Lesko, L.J. and Atkinson, A.J. (2001) Use of biomarkers and surrogate endpoints in drug development and regulatory decision making: criteria, validation, strategies 1. *Ann. Rev. Pharmacol. Toxicol.* 41, 347–366
4 Mullard, A. (2013) 2012 FDA drug approvals. *Nat. Rev. Drug Discov.* 12, 87–90
5 Schadow, G. (2007) Assessing the impact of HL7/FDA structured product label (SPL) content for medication knowledge management. *AMIA Annual Symposium Proceedings*, American Medical Informatics Association pp. 646
6 Haffner, M.E. (2006) Adopting orphan drugs — two dozen years of treating rare diseases. *N. Engl. J. Med.* 354, 445–447
7 Hamosh, A. *et al.* (2005) Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res.* 33 (Suppl. 1), D514–517
8 Rowe, S.M. and Verkman, A.S. (2013) Cystic fibrosis transmembrane regulator correctors and potentiators. *Cold Spring Harb. Perspect. Med.* 3, a009761
9 Rowe, S.M. *et al.* (2005) Cystic fibrosis. *N. Engl. J. Med.* 352, 1992–2001
10 Sboner, A. and Elemento, O. (2016) A primer on precision medicine informatics. *Brief. Bioinform.* 17, 145–153
11 Landrum, M.J. *et al.* (2014) ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res.* 42, D980–985
12 Potter, B.K. *et al.* (2016) Translating rare-disease therapies into improved care for patients and families: what are the right outcomes, designs, and engagement approaches in health-systems research? *Genet. Med.* 18, 117–123
13 Jamuar, S.S. and Tan, E.-C. (2015) Clinical application of next-generation sequencing for Mendelian diseases. *Human Genom.* 9, 10
14 Briggs, M.D. *et al.* (2015) New therapeutic targets in rare genetic skeletal diseases. *Expert Opin. Orphan Drugs* 3, 1137–1154
15 Segalat, L. (2007) Loss-of-function genetic diseases and the concept of pharmaceutical targets. *Nat. Rev. Genet.* http://dx.doi.org/10.1186/1750-1172-2-30
16 Ashburn, T.T. and Thor, K.B. (2004) Drug repositioning: identifying and developing new uses for existing drugs. *Nat. Rev. Drug Discov.* 3, 673–683
17 Liu, Z. *et al.* (2013) *In silico* drug repositioning — what we need to know. *Drug Discov. Today* 18, 110–115
18 Li, Y. and Jones, S. (2012) Drug repositioning for personalized medicine. *Genome Med.* 4, 27
19 Ekins, S. *et al.* (2011) *In silico* repositioning of approved drugs for rare and neglected diseases. *Drug Discov. Today* 16, 298–310
20 Sardana, D. *et al.* (2011) Drug repositioning for orphan diseases. *Brief. Bioinform.* 12, 346–356
21 Weischenfeldt, J. *et al.* (2013) Phenotypic impact of genomic structural variation: insights from and for human disease. *Nat. Rev. Genet.* 14, 125–138
22 Schumacher, K.R. *et al.* (2014) Social media methods for studying rare diseases. *Pediatrics* 133, e1345–1353
23 van Dijk, E.L. *et al.* (2014) Ten years of next-generation sequencing technology. *Trends Genet.* 30, 418–426
24 Aymé, S. (2003) Orphanet, an information site on rare diseases. *Soins* 672, 46
25 Wei, C.-Y. *et al.* (2012) Pharmacogenomics of adverse drug reactions: implementing personalized medicine. *Hum. Mol. Genet.* 21, R58–65
26 Frueh, F.W. *et al.* (2008) Pharmacogenomic biomarkerinformation in drug labels approved by the United States Food and Drug Administration: prevalence of related drug use. *Pharmacotherapy* 28, 992–998
27 Vissers, L.E.L.M. and Veltman, J.A. (2015) Standardized phenotyping enhances Mendelian disease gene identification. *Nat. Genet.* 47, 1222–1224
28 Köhler, S. *et al.* (2014) The Human Phenotype Ontology project: linking molecular biology and disease through phenotype data. *Nucleic Acids Res.* 42, D966–974
29 Mungall, C.J. *et al.* (2017) The Monarch Initiative: an integrative data and analytic platform connecting phenotypes to genotypes across species. *Nucleic Acids Res.* 45, D712–722
30 Trujillano, D. *et al.* (2017) A comprehensive global genotype–phenotype database for rare diseases. *Mol. Genet. Genom. Med.* 5, 66–75
31 Smith, C.L. and Eppig, J.T. (2009) The Mammalian Phenotype Ontology: enabling robust annotation and comparative analysis. *Wiley Interdiscip. Rev. Syst. Biol. Med.* 1, 390–399
32 Kibbe, W.A. *et al.* (2015) Disease Ontology 2015 update: an expanded and updated database of human diseases for linking biomedical knowledge through disease data. *Nucleic Acids Res.* 43, D1071–1078
33 Brookes, A.J. and Robinson, P.N. (2015) Human genotype-phenotype databases: aims, challenges and opportunities. *Nat. Rev. Genet.* 16, 702
34 Köhler, S. *et al.* (2017) The human phenotype ontology in 2017. *Nucleic Acids Res.* 45, D865–876

35 Groza, T. *et al.* (2015) The Human Phenotype Ontology: semantic unification of common and rare disease. *Am. J. Hum. Genet.* 97, 111–124
36 Akawi, N. *et al.* (2015) Discovery of four recessive developmental disorders using probabilistic genotype and phenotype matching among 4,125 families. *Nat. Genet.* 47, 1363–1369
37 Zhou, X.Z. *et al.* (2014) Human symptoms-disease network. *Nat. Commun.* 5, 10
38 Melamed, R.D. *et al.* (2015) Genetic similarity between cancers and comorbid Mendelian diseases identifies candidate driver genes. *Nat. Commun.* 6, 10
39 Hoehndorf, R. *et al.* (2015) Analysis of the human diseasome using phenotype similarity between common, genetic, and infectious diseases. *Sci. Rep.* 5, 10888
40 Jegga, A.G. (2014) Candidate gene discovery and prioritization in rare diseases. In *Clinical Bioinformatics* (Trent, R., ed.), pp. 295–312, Springer, New York
41 Amyere, M. *et al.* (2014) Common somatic alterations identified in Maffucci syndrome by molecular karyotyping. *Mol. Syndromol.* 5, 259–267
42 Kerem, B. *et al.* (1989) Identification of the cystic fibrosis gene: genetic analysis. *Science* 245, 1073–1080
43 Lander, E. and Botstein, D. (1987) Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. *Science* 236, 1567–1570
44 Blauw, H.M. *et al.* (2008) Copy-number variation in sporadic amyotrophic lateral sclerosis: a genome-wide screen. *Lancet Neurol.* 7, 319–326
45 Ng, S.B. *et al.* (2010) Exome sequencing identifies the cause of a mendelian disorder. *Nat. Genet.* 42, 30–35
46 Ng, S.B. *et al.* (2009) Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* 461, 272–276
47 Hoischen, A. *et al.* (2010) *De novo* mutations of SETBP1 cause Schinzel–Giedion syndrome. *Nat. Genet.* 42, 483–485
48 Belkadi, A. *et al.* (2015) Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. *Proc. Natl. Acad. Sci. U. S. A.* 112, 5473–5478
49 Mercer, T.R. *et al.* (2012) Targeted RNA sequencing reveals the deep complexity of the human transcriptome. *Nat. Biotechnol.* 30, 99–104
50 Stessman, H.A.F. *et al.* (2017) Targeted sequencing identifies 91 neurodevelopmental-disorder risk genes with autism and developmental-disability biases. *Nat. Genet.* 49, 515–526
51 Pabinger, S. *et al.* (2014) A survey of tools for variant analysis of next-generation genome sequencing data. *Brief. Bioinform.* 15, 256–278
52 Zook, J.M. *et al.* (2014) Integrating human sequence data sets provides a resource of benchmark SNP and indel genotype calls. *Nat. Biotechnol.* 32, 246–251
53 Xu, H. *et al.* (2014) Comparison of somatic mutation calling methods in amplicon and whole exome sequence data. *BMC Genom.* 15, 244
54 Cornish, A. and Guda, C. (2015) A comparison of variant calling pipelines using Genome in a Bottle as a reference. *Biomed. Res. Int.* 2015, 456479
55 Hwang, S. *et al.* (2015) Systematic comparison of variant calling pipelines using gold standard personal exome variants. *Sci. Rep.* 5, 17875
56 Hindorff, L.A. *et al.* (2009) Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl. Acad. Sci. U. S. A.* 106, 9362–9367
57 Plenge, R.M. *et al.* (2013) Validating therapeutic targets through human genetics. *Nat. Rev. Drug Discov.* 12, 581–594
58 Wang, Z.-Y. and Zhang, H.-Y. (2013) Rational drug repositioning by medical genetics. *Nat. Biotechnol.* 31, 1080–1082
59 Nelson, M.R. *et al.* (2015) The support of human genetic evidence for approved drug indications. *Nat. Genet.* 47, 856–860
60 Sanseau, P. *et al.* (2012) Use of genome-wide association studies for drug repositioning. *Nat. Biotechnol.* 30, 317–320
61 Hurle, M.R. *et al.* (2013) Computational drug repositioning: from data to therapeutics. *Clin. Pharmacol. Ther.* 93, 335–341
62 Korte, A. and Farlow, A. (2013) The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods* 9, 1–9
63 Bush, W.S. *et al.* (2016) Unravelling the human genome-phenome relationship using phenome-wide association studies. *Nat. Rev. Genet.* 17, 129–145
64 Roden, D.M. and Denny, J.C. (2016) Integrating electronic health record genotype and phenotype datasets to transform patient care. *Clin. Pharmacol. Ther.* 99, 298–305
65 Denny, J.C. *et al.* (2013) Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. *Nat. Biotechnol.* 31, 1102–1111
66 Rastegar-Mojarad, M. *et al.* (2015) Opportunities for drug repositioning from phenome-wide association studies. *Nat. Biotechnol.* 33, 342–345
67 Law, V. *et al.* (2014) DrugBank 4.0: shedding new light on drug metabolism. *Nucleic Acids Res* 42, D1091–1097
68 Sarntivijai, S. *et al.* (2016) Linking rare and common disease: mapping clinical disease-phenotypes to ontologies in therapeutic target validation. *J. Biomed. Seman.* 7, 8
69 Jin, G. and Wong, S.T. (2014) Toward better drug repositioning: prioritizing and integrating existing methods into efficient pipelines. *Drug Discov. Today* 19, 637–644
70 Dhillon, A.S. *et al.* (2007) MAP kinase signalling pathways in cancer. *Oncogene* 26, 3279–3290
71 Kandoth, C. *et al.* (2013) Mutational landscape and significance across 12 major cancer types. *Nature* 502, 333–339
72 Marin, T.M. *et al.* (2011) Rapamycin reverses hypertrophic cardiomyopathy in a mouse model of LEOPARD syndrome–associated PTPN11 mutation. *J. Clin. Investig.* 121, 1026–1043
73 Goh, K.-I. *et al.* (2007) The human disease network. *Proc. Natl. Acad. Sci. U. S. A.* 104, 8685–8690
74 Li, Y. and Agarwal, P. (2009) A pathway-based view of human diseases and disease relationships. *PLoS One* 4, e0004346
75 Kiel, C. and Serrano, L. (2014) Structure-energy-based predictions and network modelling of RASopathy and cancer missense mutations. *Mol. Syst. Biol.* 10, 727
76 Rockman, M.V. and Kruglyak, L. (2006) Genetics of global gene expression. *Nat. Rev. Genet.* 7, 862–872
77 Montgomery, S.B. and Dermitzakis, E.T. (2009) The resolution of the genetics of gene expression. *Hum. Mol. Genet.* 18, R211–215
78 Natsoulis, G. *et al.* (2008) The liver pharmacological and xenobiotic gene response repertoire. *Mol. Syst. Biol.* 4, 12
79 Igarashi, Y. *et al.* (2015) Open TG-GATEs: a large-scale toxicogenomics database. *Nucleic Acids Res.* 43, D921–927
80 Lamb, J. *et al.* (2006) The connectivity map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* 313, 1929–1935
81 Masica, D.L. and Karchin, R. (2011) Correlation of somatic mutation and expression identifies genes important in human glioblastoma progression and survival. *Cancer Res.* 71, 4550–4561
82 Bertrand, D. *et al.* (2015) Patient-specific driver gene prediction and risk assessment through integrated network analysis of cancer omics profiles. *Nucleic Acids Res.* 43, e44
83 Ping, Y. *et al.* (2015) Identifying core gene modules in glioblastoma based on multilayer factor-mediated dysfunctional regulatory networks through integrating multi-dimensional genomic data. *Nucleic Acids Res.* 43, 1997–2007
84 Ding, J. *et al.* (2015) Systematic analysis of somatic mutations impacting gene expression in 12 tumour types. *Nat. Commun.* 6, 8554
85 Gerstung, M. *et al.* (2015) Combining gene mutation with gene expression data improves outcome prediction in myelodysplastic syndromes. *Nat. Commun.* 6, 5901
86 Soifer, H.S. *et al.* (2007) MicroRNAs in disease and potential therapeutic applications. *Mol. Ther.* 15, 2070–2079
87 Amato, F. *et al.* (2013) Gene mutation in microRNA target sites of CFTR gene: a novel pathogenetic mechanism in cystic fibrosis? *PLoS One* 8, e60448
88 Liu, Z. *et al.* (2014) Deciphering miRNA transcription factor feed-forward loops to identify drug repurposing candidates for cystic fibrosis. *Genome Med.* 6, 94
89 Qu, X.A. and Rajpal, D.K. (2012) Applications of Connectivity Map in drug discovery and development. *Drug Discov. Today* 17, 1289–1298
90 Iorio, F. *et al.* (2010) Discovery of drug mode of action and drug repositioning from transcriptional responses. *Proc. Natl. Acad. Sci. U. S. A.* 107, 14621–14626
91 Dudley, J.T. *et al.* (2011) Computational repositioning of the anticonvulsant topiramate for inflammatory bowel disease. *Sci. Transl. Med.* 3, 6
92 Barrett, T. *et al.* (2007) NCBI GEO: mining tens of millions of expression profiles — database and tools update. *Nucleic Acids Res.* 35, D760–765
93 Iskar, M. *et al.* (2013) Characterization of drug-induced transcriptional modules: towards drug repositioning and functional understanding. *Mol. Syst. Biol.* 9, 13
94 Ling, H. *et al.* (2013) MicroRNAs and other non-coding RNAs as targets for anticancer drug development. *Nat. Rev. Drug Discov.* 12, 847–865
95 Li, Z. and Rana, T.M. (2014) Therapeutic targeting of microRNAs: current status and future challenges. *Nat. Rev. Drug Discov.* 13, 622–638
96 Liu, X.Y. *et al.* (2013) SM2miR: a database of the experimentally validated small molecules' effects on microRNA expression. *Bioinformatics* 29, 409–411
97 Rukov, J.L. *et al.* (2014) Pharmaco-miR: linking microRNAs and drug effects. *Brief. Bioinform.* 15, 648–659
98 Orphadata: Free access data from Orphanet. INSERM (1997) Available at: http://www.orphadata.org
99 Boztug, K. *et al.* (2009) A novel syndrome with congenital neutropenia caused by mutations in G6PC3. *N. Engl. J. Med.* 360, 32–43
100 Zeidler, C. *et al.* (2000) Management of kostmann syndrome in the g-csf era. *Br. J. Haematol.* 109, 490–495

101 Sparks, R. *et al.* (2016) Expanding the immunology toolbox: embracing public-data reuse and crowdsourcing. *Immunity* 45, 1191–1204
102 Bhattacharya, S. *et al.* (2014) ImmPort: disseminating data to the public for the future of immunology. *Immunol. Res.* 58, 234–239
103 Heng, T.S.P. *et al.* (2008) The Immunological Genome Project: networks of gene expression in immune cells. *Nat. Immunol.* 9, 1091–1094
104 Kidd, B.A. *et al.* (2016) Mapping the effects of drugs on the immune system. *Nat. Biotechnol.* 34, 47–54
105 Li, J. *et al.* (2016) A survey of current trends in computational drug repositioning. *Brief. Bioinform.* 17, 2–12
106 Jin, G. and Wong, S.T.C. (2014) Toward better drug repositioning: prioritizing and integrating existing methods into efficient pipelines. *Drug Discov. Today* 19, 637–644
107 Blatt, J. and Corey, S.J. (2013) Drug repurposing in pediatrics and pediatric hematology oncology. *Drug Discov. Today* 18, 4–10
108 Liu, Z. *et al.* (2016) Potential reuse of oncology drugs in the treatment of rare diseases. *Trends Pharmacol. Sci.* 37, 843–857
109 Ekins, S. *et al.* (2007) In silico pharmacology for drug discovery: methods for virtual ligand screening and profiling. *Br. J. Pharmacol.* 152, 9–20
110 Chen, Y.Z. and Zhi, D.G. (2001) Ligand–protein inverse docking and its potential use in the computer search of protein targets of a small molecule. *Proteins Struct. Funct. Bioinform.* 43, 217–226
111 Kolb, P. *et al.* (2009) Docking and chemoinformatic screens for new ligands and targets. *Curr. Opin. Biotechnol.* 20, 429–436
112 Berman, H.M. *et al.* (2000) The Protein Data Bank. *Nucleic Acids Res.* 28, 235–242
113 Andronis, C. *et al.* (2011) Literature mining, ontologies and information visualization for drug repurposing. *Brief. Bioinform.* 12, 357–368
114 Agarwal, P. and Searls, D.B. (2008) Literature mining in support of drug discovery. *Brief. Bioinform.* 9, 479–492
115 Su, E.W. and Sanger, T.M. (2017) Systematic drug repositioning through mining adverse event data in ClinicalTrials.gov. *PeerJ* 5, e3154
116 Yang, H.-T. *et al.* (2017) Literature-based discovery of new candidates for drug repurposing. *Brief. Bioinform.* 18, 488–497
117 Barrett, T. *et al.* (2013) NCBI GEO: archive for functional genomics data sets–update. *Nucleic Acids Res.* 41, D991–995
118 Parkinson, H. *et al.* (2005) ArrayExpress–a public repository for microarray gene expression data at the EBI. *Nucleic Acids Res.* 33 (1), D553–555
119 Panic, G. *et al.* (2015) Activity profile of an FDA-approved compound library against *Schistosoma mansoni*. *PLoS Negl. Trop. Dis.* 9, 15
120 Campillos, M. *et al.* (2008) Drug target identification using side-effect similarity. *Science* 321, 263–266
121 Clohessy, J.G. and Pandolfi, P.P. (2015) Mouse hospital and co-clinical trial project–from bench to bedside. *Nat. Rev. Clin. Oncol* 12, 491–498
122 Lynam, E.B. *et al.* (2012) A patient focused solution for enrolling clinical trials in rare and selective cancer indications: a landscape of haystacks and needles. *Drug Inf. J.* 46, 472–478
123 Koboldt, D.C. *et al.* (2013) The next-generation sequencing revolution and its impact on genomics. *Cell* 155, 27–38
124 Frail, D.E. *et al.* (2015) Pioneering government-sponsored drug repositioning collaborations: progress and learning. *Nat. Rev. Drug Discov.* 14, 833–841