COMPUTATIONAL AND STRUCTURAL BIOTECHNOLOGY JOURNAL

Review

Turning genome-wide association study findings into opportunities for drug repositioning

Alexandria Lau [a], Hon-Cheong So [a,b,c,d,e,f,g,*]

[a] School of Biomedical Sciences, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong SAR, China
[b] KIZ-CUHK Joint Laboratory of Bioresources and Molecular Research of Common Diseases, Kunming Zoology Institute of Zoology and The Chinese University of Hong Kong, Hong Kong SAR, China
[c] Department of Psychiatry, The Chinese University of Hong Kong, Hong Kong SAR, China
[d] Margaret K.L. Cheung Research Centre for Management of Parkinsonism, The Chinese University of Hong Kong, Hong Kong SAR, China
[e] Shenzhen Research Institute, The Chinese University of Hong Kong, Shenzhen, China
[f] Brain and Mind Institute, The Chinese University of Hong Kong, Hong Kong SAR, China
[g] Hong Kong Branch of the Chinese Academy of Sciences Center for Excellence in Animal Evolution and Genetics, The Chinese University of Hong Kong, Hong Kong SAR, China

ARTICLE INFO

ABSTRACT

Drug development is a very costly and lengthy process, while repositioned or repurposed drugs could be brought into clinical practice within a shorter time-frame and at a much reduced cost. Numerous computational approaches to drug repositioning have been developed, but methods utilizing genome-wide association studies (GWASs) data are less explored.

The past decade has observed a massive growth in the amount of data from GWAS; the rich information contained in GWAS has great potential to guide drug repositioning or discovery. While multiple tools are available for finding the most relevant genes from GWAS hits, searching for top susceptibility genes is only one way to guide repositioning, which has its own limitations.

Here we provide a comprehensive review of different computational approaches that employ GWAS data to guide drug repositioning. These methods include selecting top candidate genes from GWAS as drug targets, deducing drug candidates based on drug-drug and disease-disease similarities, searching for reversed expression profiles between drugs and diseases, pathway-based methods as well as approaches based on analysis of biological networks. Each method is illustrated with examples, and their respective strengths and limitations are discussed. We also discussed several areas for future research.

© 2020 The Author(s). Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (http://creative-commons.org/licenses/by-nc-nd/4.0/).

Contents

* Corresponding author at: School of Biomedical Sciences, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong SAR, China.
    E-mail address: hcso@cuhk.edu.hk (H.-C. So).

1. Introduction

Drug development is a very costly and lengthy process, which typically involves multiple processes from drug target validation, clinical trials to final approval by the FDA and other government agencies. The estimated cost of developing a new drug is around USD 2.6 billion [1]. As a result, there has been increased interest in repositioning existing drugs for new usage. Drug repositioning refers to finding new indications of existing or investigational drugs [2]. Repositioned drugs can be brought to clinical practice in a much shorter time-frame and at a much lower cost, as these drugs have gone through pharmacokinetic/pharmacodynamic and safety profiling during development. Besides existing drugs with known indications, drugs that are shelved due to failure in clinical trials may also serve as repositioning candidates. In fact, repurposing these drugs may serve to recover the high cost that went into developing them (see Fig. 1).

In practice, many drugs in wide use today stem from repositioning. Two classic examples are Sildenafil and thalidomide [3], which are now commonly used to treat erectile dysfunction and multiple myeloma, despite not originally designed for these indications. However, these and many other drugs were discovered based on serendipity alone. Computational repositioning approaches offer a more systematic and cost-effective way of discovering such unexpected relationships between drugs and diseases when compared to experimental approaches.

The past few years have witnessed a massive rise in the amount of 'omics' and other forms of biomedical data such as electronic health records, which makes computational approaches an attractive option to prioritize repositioning candidates. One of the fastest growing types of data comes from genome-wide association studies (GWASs), a high-throughput technique that interrogates the whole genome for common genetic variations that contribute to diseases or traits. GWAS has been highly successful in unraveling the genetic basis of many complex traits or diseases, and study scales have increased rapidly [4].

Many statistical/computational methodologies have been developed to improve the power in detecting susceptibility variants in GWAS. These include, for example, entropy-based methods for genetic association [5,6], integrating the trend test and Pearson's test for association [7], and a variety of approaches to select from or integrate different models of inheritance [8–11]. However, from a clinical point of view, one of the most important questions would be: *could GWAS findings be translated into opportunities for drug discoveries or repositioning*? This question calls for more innovative approaches to analyzing GWAS data with a translational focus. In this article, we shall review several categories of methods for prioritizing drug repositioning candidates, highlighting their applications as well as their respective strengths and limitations.

2. Overview of GWAS and its potential in guiding drug discoveries and repositioning

GWAS aims to decipher associations between common genetic variants and disease or disease-related traits. Typically the genetic variants studied are SNPs, and current GWAS arrays allow millions of SNPs across the entire genome to be interrogated at the same time. Variants that are not genotyped can also be imputed with appropriate reference panels [12]. The most common design for GWAS is a population-based case-control study, in which we recruit subjects with and without the disease, and search for SNPs with significant differences in allele frequencies between the two groups. However, GWAS can also be used to study continuous or time-to-event outcomes. A full GWAS workflow tutorial can be found in Ref. [13].

A number of GWAS resources are available online, and one of the largest is the GWAS catalog (https://www.ebi.ac.uk/gwas/). It is a structured repository of summary statistics for a large variety of traits. As of writing, the GWAS Catalog contains 4554 publications and 185,864 associations, showing the popularity of GWAS and the vast amount of data available for mining. Other useful resources include the LD-hub (http://ldsc.broadinstitute.org/ld-hub/), GWAS summary statistics from the UK Biobank (http://www.nealelab.is/uk-biobank) and dbGaP (https://www.ncbi.nlm.nih.gov/gap/) which allows access to individual genomic data to authorized users.

In order to treat a disease, it is critical to understand the causal risk factors, which is one of the main goals of GWAS - to understand the underlying biological mechanisms [14]. Unlike Mendelian diseases where typically only less than a few genes cause the condition, complex diseases arise as combinations of a multitude of genetic and environmental causes [15]. GWAS have a niche for identifying risk loci for a disease without *a priori* hypotheses [16]; hence it is useful in identifying *novel* genetic loci that are beyond our current understanding of the disease. At the same time, the development of drugs with new mechanisms of action has become increasingly difficult; GWAS data therefore holds great potential in guiding drug discovery. Another important advantage is that for many diseases (e.g. psychiatric disorders [17] and cancers [18]), current cell-based or animal models are unable to fully mimic the human condition, limiting the success rate of translating preclinical findings into clinical practice. However, GWASs are based on clinical samples of patients with actual phenotype data, and may more realistically reflect the genetic basis of the condition under study.

In an important study, Nelson et al. [19] showed that the proportion of drugs with direct genetic support increased along the development pipeline, increasing from 2.0% at the preclinical stage
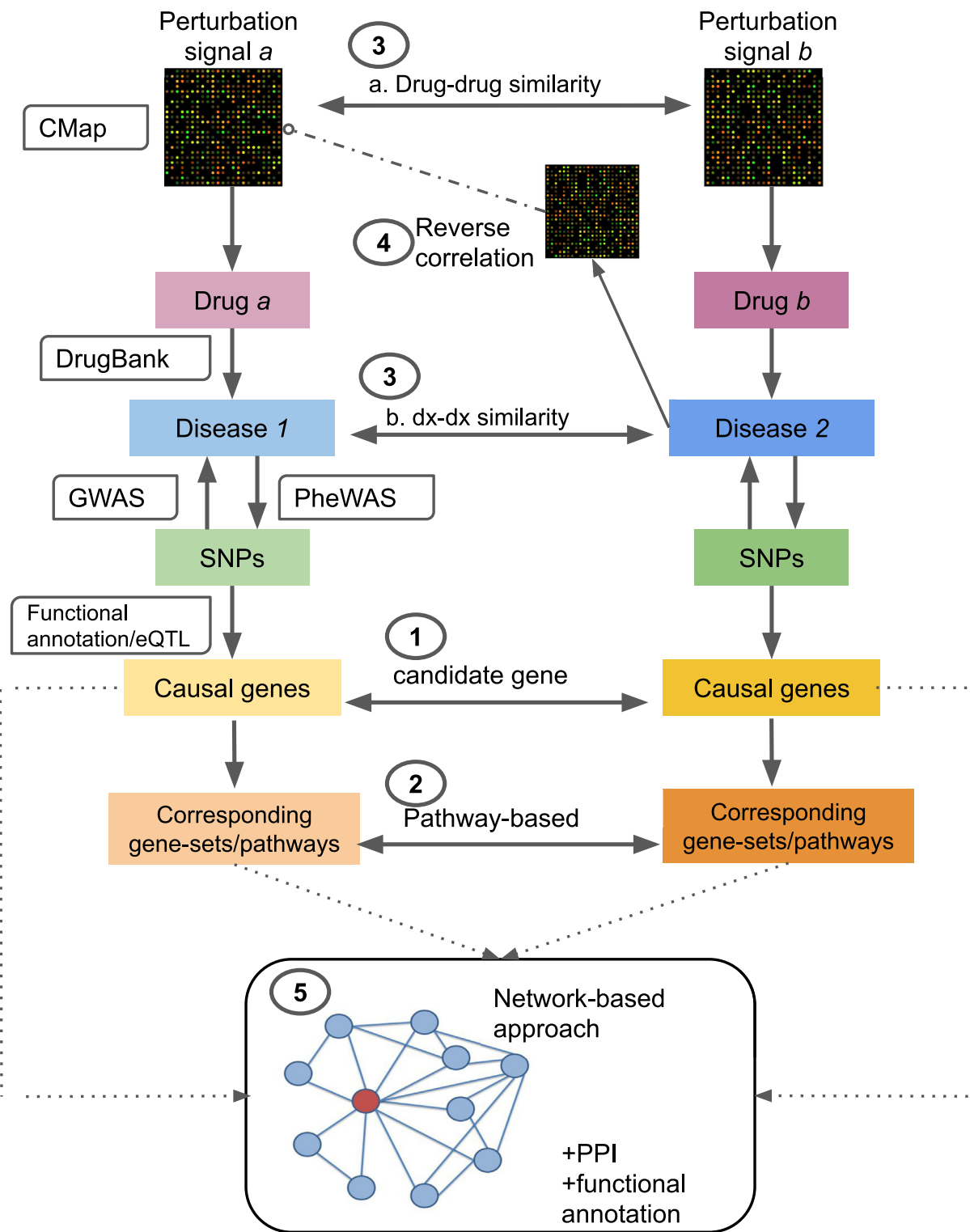
Fig. 1. A schematic diagram showing the five drug repositioning approaches reviewed in this paper. (1) Candidate gene approach: the risk loci from GWAS data can be mapped to the most likely relevant genes with functional annotations and eQTL data. If the identified candidate gene is druggable and the drug is not already indicated for the disease, the drug may serve as a repositioning candidate. (2) Pathway or gene-set analysis approach: the identified candidate gene(s) are placed in the context of its pathways; drugs that target members of the same pathway are potential drug candidates. Alternatively, the entire set of GWAS data may be used to derive gene-based statistics, and enrichment tests performed to look for drugs whose targets/effector genes achieve higher significance (lower p-values) than expected as a whole. (3) Comparing similarities between drugs and diseases (dx): drugs may be repositioned to the indications of another one if the two shared sufficient similarity. For example, one could compare the similarities between the transcriptome of two drugs using cell-line expression data from the Connectivity Map (CMap). In a similar vein, if two diseases are similar, then the drugs used for treating one disease may be repurposed for the other. (4) Looking for reversed expression patterns between drugs and diseases: The core hypothesis is that if a drug produces an expression profile that is opposite to that of a disease, then the drug may be considered a repositioning candidate (due to its potential to 'reverse' disease-related expression profiles). (5) Network-based analysis: Integration of multiple sources of data such as drugs, proteins, genes and diseases relationships to construct biological networks. Further analysis with computational methods such as random walk may reveal novel drug-disease connections. The principle of this method is close to 'similarity-based' methods (see main text) for drug repositioning, but network-based methods usually integrate a greater variety of information. PPI; protein–protein interactions.

Table 1
General *in silico* approaches for drug repositioning and relevant reviews.

Category	Examples of approaches/aspects of drug repositioning covered	Relevant reviews
Cheminformatics	Molecular docking; Quantitative structure–activity relationship (QSAR) modeling; repositioning based on shared mechanism of action (MOA) between drugs, target binding pocket similarities and similarities of drug candidates in the chemical space; off-target effects modeling	[31–35]
Network-based	Network analysis of drugs, proteins/genes and diseases	[31,33–36]
Omics data	Modeling drug expression/gene perturbation data based on CMap; assessing links between diseases and drugs based on omics data	[31–35]
Clinical data and text mining	Mining of electronic health records; prediction of drug indications and side effects via text mining and semantic analysis of the literature or other relevant databases	[31,32,35]

Some reviews cover more than one category; the above is just a rough overview of some of the aspects of drug repositioning covered in the reviews. For details please refer to the corresponding references.

to 8.2% among the approved drugs. In a more updated analysis by King et al. [20], they reported similar findings that genetically supported targets were more likely to be successful in Phases II and III clinical trials, especially when the genes implicated are likely causal.

However, GWAS has been criticized for theinterpretability of its findings and small effect sizes of most risk variants. Approximately 90% of the SNPs found by GWAS are located in non-coding regions, suggesting that they may alter transcription of one or more target genes through modification of splice sites, non-codingRNAs, promoter or transcription factor (TF) binding sites [15,21,22]. Also, due to linkage disequilibrium (LD), the association of a disease with a genetic locus does not reveal which variant is the 'causal' variant or which gene is the 'causal' one inside the locus. Another concern is that most susceptibility variants discovered so far only confer small effects on diseases or traits. However, small effect sizes of individual SNPs do not necessarily dictate low drug efficacy when we target the corresponding protein(s). For instance, one of the most successful lipid lowering drugs is statins, which targets HMG Co-A reductase (HMGCR). This target is also suggested by genetic evidence, since GWAS of LDL-cholesterol (LDL-C) also implicated SNPs in this locus. However, the effect size of SNP(s) at *HMGCR* is relatively modest [23,24] (e.g. G allele of rs17238484 was associated with ~0.06 mmol/L reduction in LDL-C), but the effect size of statins is much larger [25]. Therefore, modest effect sizes of genetic variations do not exclude therapeutic potential of the corresponding targets.

3. Approaches to computational drug repositioning using GWAS

Comprehensive reviews of various computational drug repositioning approaches are already available (Table 1), which covers

Table 2
Summary and comparison of drug repositioning approaches using GWAS data.

Approach	Brief description	Pros	Cons	Selected references
1. (i) Using functional annotation for identifying top candidate genes (then link to drugs)	Map GWAS SNPs to corresponding functionally related genes with functional annotation tools	Relatively clear biological interpretation; straightforward computation and low computational cost; multiple databases available for annotation	Usually only single or a few genes are examined; may potentially miss multi-target drugs; directionality of effect may not be clear; functional annotation information for some SNPs may be missing; not all genes are directly druggable	[15,26,29]
1. (ii) Using eQTL data for identifying top candidate genes (then link to drugs)	Map GWAS SNPs to corresponding gene using eQTL information	Directionality indicated; covers SNPs that affects multiple genes at the same time; easy to implement	GTEx data still limited in sample size, and bias towards European population; not all SNPs may affect gene expression	[74–76]
2. Pathway/gene-set analysis	Repositioning drug based on pathway or gene-set analysis of GWAS results	Consideration of drug effect on a genome-wide scale; multi-target drugs included; inclusion of risk loci with small effect size individually but good therapeutic potential when combined as a pathway	Definition of pathways can be complicated; incomplete characterization of pathways for all drugs; directionality of effect may not be clear	[67,77]
3. Similarity-based: Drug-drug or disease-disease similarity	Evaluating similarities between drugs and diseases effect may reveal novel drug-disease relationships	Intuitive in concept; simple computation; less detailed understanding of drug mechanism required	'Similarity' may not be easy to define; difficulty in integrating different sources of similarity measures; relatively hard to uncover drugs with novel mechanisms of actions	[78,79]
4. Reversed expression pattern between drugs and diseases	A drug with expression profile opposite to that of a disease are candidate therapeutic agents	Considers data across many genes instead of the most significant ones; imputed expression readily available for many tissues and from large GWAS samples, and less susceptible to confounding and reverse causality; understanding of drug mechanism not required; relatively better at uncovering drugs of novel mechanism	Expression reversal may not be the only drug mechanism; limitation of cell lines (cannot fully model human conditions); imputation accuracy of some genes may be poor	[80,81]
5. Network-based methods	Integration of multiple sources of data regarding drugs, proteins, genes and diseases relationships to reveal novel drug-disease connections	Flexible; ability to integrate multiple sources of data; well established network analysis methods from other fields	Integrating data with different nature and potentially different kinds of bias is difficult; complicated parameter optimization; difficulty in determining edge strength; relatively less capable of revealing unexpected repositioning candidates	[82–86]

different aspects and approaches of repositioning. However, the current review is distinct in that we particularly focus on the use of human genomic data, especially GWAS, in guiding drug repositioning. Given the wealth of GWAS data available, we believe that many novel drug-disease relationships are yet to be discovered. We note that multiple methods or tools are available for finding relevant genes from GWAS hits [15,26–30]; however searching for top susceptibility genes is only one way to guide repositioning, which has its own limitations (see discussions in later sections).

In this review, we provide (i) a survey of various drug repositioning approaches using GWAS data; (ii) highlights of studies with real disease applications of the above approaches; and (iii) discussions on the limitations of these methods (Table 2). For simplicity, we have categorized GWAS-based repositioning approaches into five main groups, namely (1) candidate gene approaches; (2) pathway-based mapping; (3) investigation of drug-drug and disease-disease similarities; (4) methods based on reversed expression pattern between drugs and diseases; and (5) network-based approaches Fig. 1. However, multiple methods are often used together in studies, therefore the distinction is not always clear.

3.1. Selecting top candidate genes from GWAS as targets for drug repositioning

Perhaps the most intuitive approach to drug repositioning with GWAS is to focus on the top candidate genes identified in the study. We may first map the SNPs onto corresponding genes, preferably with knowledge of the functional roles of the SNP (e.g. whether it affects expression or regulation of a gene). In the next step, we may query these target genes in drug databases where information about drug-gene and drug-disease indications can be retrieved. Finally, drug repositioning opportunities present itself as a 'mismatch' between the drug indication and the disease of interest. For example, we may find a drug that targets the GWAS top gene but it has not been used for treating the disease of interest yet. The target genes retrieved from GWAS serve as a connection between the disease and drugs.

Mapping susceptibility SNPs from GWAS to the corresponding functionally important gene is a fundamental step for this approach of repositioning, and is a topic of active research. However, this can be a challenging process since many SNPs are located in non-coding regions where the functional roles of variants are not fully understood. There are several comprehensive reviews on computational approaches or tools for identifying the most relevant genes from GWAS hits [15,26–30]. We shall highlight a few approaches below in the context of drug repositioning and discuss the limitations of this repositioning methodology.

3.1.1. An overview of SNP-to-gene mapping

Following Edwards et al. [26], mapping GWAS variants to the 'target' (i.e. functionally relevant) genes can be broken down into several steps. The first step involves fine mapping of the SNPs associated with traits of interest. Briefly, fine mapping is a process of identifying the most probable causal candidate SNPs within the identified genetic loci [15,37,38]. Note that GWAS chips are not designed to necessarily sequence the functional SNPs; rather SNPs that are the most representative (i.e. 'tag' SNPs) are usually chosen [26]. Tools for fine-mapping include, for example, FINEMAP [39], PAINTOR [40], CAVIAR [41], CAVIAR Bayes factor [42] and more recently CAUSALdb [43]. The second step involves *in silico* methods for annotation and characterization of the functional impact of identified SNPs. Functional significance of the fine-mapped loci can be investigated with information such as chromatin accessibility, TF binding, DNA protein interactions, histone modification and DNA methylation and chromatin interactions. Another common approach is to look for overlap of the identified variants with

expression- or other types of quantitative trait loci (QTL). Finally, if resources allow, one may perform further experiments in cell lines or animal models to ascertain the roles of the target variants or genes.

3.1.2. Using functional annotations to map associated SNPs to genes

Given that most associated SNPs are located in non-coding regions, it is logical to hypothesize that these SNPs may regulate gene expression in certain ways to affect disease risks. The Encyclopedia of DNA elements (ENCODE) is one of the earliest initiatives to systematically characterize functional elements in the human genome. ENCODE aims to extensively characterize multiple genetic elements, examples of which include TF binding regions, chromatin and DNA accessibility, histone modification, epigenetics and 3D chromatin interactions. The project employs a variety of techniques including RNA-seq, DNase-seq, FAIRE-seq and ChIP-seq etc., which provide very rich data for functional annotation. Other useful resources or tools for functional annotation include modENCODE (modencode.org), the NIH Roadmap Epigenomics Project, GWAS3D [44] and its successor GWAS4D [45], HaploReg [46], Mutation Enrichment Gene set Analysis of Variants (MEGA-V) [47], SNPinfo [48], FUMA [49] and CAUSALdb [43]. One point to note is that the genes closest to the associated SNP may not necessarily be the most functionally relevant gene; a recent study suggested that the likely causative genes are often >2Mbp from the index SNP [50]. To improve the reliability of gene mapping, it is advisable to employ a variety of annotation methods to prioritize the best genes as drug target candidates.

3.1.3. Using expression-QTL (eQTL) to map associated SNPs to genes

Besides prioritizing the corresponding 'target' gene(s) for the associated SNPs, it is preferable to also determine the directionality of such relationships to facilitate drug repositioning. One important question is whether the SNPs cause changes in gene expressions, and if so, what is the direction of change and which tissues are involved. For example, if the identified SNP causes an upregulation of gene *X* leading to increased risk of a disease, then an inhibitor of its protein product may be considered a repositioning candidate. One of the largest eQTL resources is the Genotype-Tissue Expression (GTEx) project (gtexportal.org), which includes eQTL data from 49 tissues of over 800 subjects. However, most of the subjects are Europeans with male predominance (~67%), and the sample size may still be insufficient to detect eQTL with modest effects. Another related approach is to use tools such as PrediXcan [51] to *impute* the expression changes based on raw genotype or GWAS summary data.

3.1.4. Querying drug databases for repositioning candidates

After identifying the most relevant genes from associated SNPs, finding candidate drugs that target the selected genes is relatively straightforward. For instance, in a recent attempt to repurpose drugs for inflammatory bowel disease (IBD), Grenier et al. [52] first selected the most likely causal SNPs by Bayes factor within each locus, then these SNPs were mapped to relevant genes by functional annotations. Finally, drugs that may be repositioned for IBD were derived from these genes using a web-based tool Gene2-Drug [53] (see descriptions below). In another study, Okada et al. [40] performed a GWAS for rheumatoid arthritis (RA) and discovered 42 additional significant loci. They found that drug targets from approved RA drugs showed significant overlap with RA risk genes or their interacting partners. Tragante et al. [54] made use of data from 49 coronary artery disease (CAD) and myocardial infarction (MI) GWAS to prioritize 153 risk loci. Leveraging drug-gene information from DGIdb, they identified three repositioning candidates and re-discovered several existing treatments [55]. Grover et al. [56] suggested 981 novel therapeutic agents for CAD

based on 647 candidate genes they identified from the Wellcome Trust Case Control Consortium (WTCCC) GWAS. They also identified additional candidate genes with Gentrepid [57], which expanded the list of genes using pathway, protein–protein interaction (PPI) and protein domain homology information. Other reports pulled drug data from multiple databases including TTD, Pandrugs, PharmGKB and DrugBank to derive drug-target relationships [56,58–63].

3.1.5. PheWAS

Phenome-wide association studies (PheWAS) is another way to reveal drug repositioning opportunities. Unlike GWAS, PheWAS evaluates the 'opposite' direction of association, examining what diseases or traits are associated with specific genetic variants. For example, variants in drug target genes may be subject to PheWAS, hence enabling the discovery of additional drug-disease links.

In one of the earliest large-scale PheWAS, Denny et al. made use of electronic medical records (EMRs) to examine phenotype associations of ~3000 SNPs previously implicated in GWAS. They found that PheWAS successfully replicated 66% of previous GWAS associations and revealed new pleiotropic associations that were subsequently replicated. These findings provided support to the validity of the PheWAS approach. In a subsequent study, Rastegar-Mojarad et al. [58] proposed that PheWAS may be further utilized to reveal drug repositioning candidates. The authors discovered 52,966 drug-disease pairs via PheWAS, among which ~28% were supported by the literature or tested in clinical trials. It was proposed that the rest of the drug-disease pairs may provide new repositioning opportunities. For a further discussion of the complementary nature between GWAS and PheWAS, please also refer to Robinson et al. [64].

3.1.6. Limitations

There are several limitations of using the top candidate genes for direct drug repositioning. Firstly, the top genes identified from GWAS may not be easily druggable. Finan et al. performed a comprehensive analysis on the druggability of genes [65]. They estimated that only 4479 (22%) of the 20,300 protein coding genes are druggable (or already targeted by a drug). Secondly, focusing on the effect of the top SNPs may miss biologically meaningful target genes with small effect sizes [66–68]. Third, focusing on a single candidate gene may miss *multi*-target drugs, which could be more effective than single-target ones for some conditions [69]. Recently increasing attention has been placed on development of multi-target drugs [70,71]. Fourth, as discussed earlier, due to the complexity of the human genome, there is no perfect way to proper annotation. As a result, different studies may have employed different (sometimes incomplete) annotation procedures, and integrating various annotation approaches is not straightforward. As for the limitations of PheWAS, it is still not as well-developed as GWAS, given that systematically curating EMR and establishing standardized nomenclature are not straightforward and require a huge concerted effort. Also, it remains an open question on how to model high dependence between phenotypes derived from EMR. A more detailed discussion on the strengths and limitations of PheWAS can be found in [72,73].

3.2. Drug repositioning based on pathway or gene-set analysis

Pathway or gene-set analysis (as opposed to single gene or SNP-based studies) offers a more macroscopic view of the biological processes underlying diseases and drug effects. The key idea behind pathway analysis is to organize various functionally or biologically relevant genes together and consider their overall effect. As mentioned earlier, approaches that consider only individual SNPs or genes may miss biologically meaningful associations of modest effect sizes [87]. On the other hand, 'gene-sets' can be

any set of functionally related genes or a set based on arbitrary criteria set by the researcher (in this context, drug targets or 'effector genes'). For the purpose of this review, we refer to both of these grouping methods as 'pathways'. Several examples of applications are discussed below.

De Jong et al. [79] studied repositioning candidates for schizophrenia by a pathway analysis based on GWAS summary statistics, where each drug pathway is defined by pharmacological profiles and chemical binding affinities. The analysis highlighted several candidates reaching a suggestive level of significance, including two dopamine receptor antagonists and a tyrosine kinase inhibitor. In another recent work, So et al. [88] investigated whether findings from GWAS may be used to guide drug repositioning for depression and anxiety disorders. Drug-effector gene-sets were extracted with DSigDB and gene-based significance from GWAS was computed by FASTBAT [89]. Then pathway analyses (following the principle of MAGMA [90]) were conducted to look for enrichment of GWAS results for specific drugs. Interestingly, the repositioning hits were largely enriched for known psychiatric drugs or those included in clinical trials. Enrichment was seen for antidepressants and anxiolytics but also for antipsychotics. The study also revealed other repositioning candidates with literature support. Note that the above two studies did not only focus on the top genes but also considered the actual significance level of each gene.

Another approach was presented by Jhamb et al. [91]. The authors first obtained GWAS data from STOPGAP [92]. After SNPs with $p < 5e{-}8$ are mapped to genes, these genes were mapped to their respective pathways with MetaBase, a manually curated software suite that contains interaction data. The main novelty is that the authors considered an expanded set of GWAS 'hits' by including genes in related pathways as well. Finally a 2 by 2 table is constructed for each disease to look for over-representation of the (expanded) GWAS hits among drug targets indicated for each disease.

Yet another work by Gaspar et al. investigated drug repositioning opportunities for schizophrenia using GWAS data [93]. They also proposed a new visualization approach and studied how increased sample size of the original GWAS may improve the yield of drug repositioning. Pathway analysis can also be combined with other methods. Mäkinen et al. [66] first mapped cardiovascular disease (CVD) associated-SNPs to genes with eQTL, then investigated the enrichment of these 'e-SNPs' among known biological pathways. After obtaining the CVD pathway gene-set, the pathways were augmented with co-expression analysis and subsequently placed in a Bayesian network model of gene-gene interactions. By integrating a larger variety of sources of omics data, the identified genes may serve as better targets for repositioning.

Another tool combining pathway analysis and perturbation library is Gene2Drug [53], which is used by Grenier et al. [52] (see discussion above). Gene2Drug takes a single gene as input, and aims to discover drugs linked to the input gene. The program extracts drug-related pathways based on perturbation data from CMap (see section below), then generates the subset of pathways that include the input gene.

Pathway-based analysis is a flexible approach and several extensions are possible. For example, in clinical practice, comorbid diseases are very common, and it will be preferable to find drugs that target both diseases at the same time. A recent study proposed a pathway-based approach to address this problem. Wong et al. [94] first employed a false discovery rate (FDR)-based approach to uncover genetic loci *shared* between depression/anxiety and cardiometabolic diseases from GWAS summary data. The shared loci were then subject to pathway analysis to uncover repositioning candidates, many of which are supported by the literature. Also, while the focus of this review is on repositioning using GWAS, the pathway-based approach can be readily extended to handle other types of human genomic data. For example, a recent study

showed by pathway analysis that *de novo* mutations may be used to guide drug discoveries for neuro-psychiatric disorders [95]. Readers may also refer to Ref. [96] for further discussions on the potential of GWAS data in drug discovery for neuro-psychiatric disorders.

3.2.1. Limitations

First, defining a pathway is complicated because feedback inhibition and compensatory mechanisms are almost ubiquitous in biology [67]. Also, there is an emphasis on the known pathways in pathway-based approaches [91], therefore the applicability of this approach may be limited by our current knowledge of biology. The mechanisms of many drugs are not entirely clear so their 'pathways' or gene-sets may be incompletely characterized. Finally, the directionality of effect may not be clear. However, there are tools that aid the estimating of directionalities in pathway analyses [97–99], and commercially available tools such as QIAGEN's Ingenuity Pathway Analysis (IPA) (digitalinsights.qiagen.com) and MetaCore (portal.genego.com/).

3.3. Similarity-based drug repositioning

Similarity between drug-drug and disease-disease pairs may be leveraged for drug repositioning. One of the characteristics of such similarity-based approaches is that they generally do not require detailed understanding of drug or disease mechanisms, and are able to consider high-throughput omics data across many genes without restricting to the most significant ones. Note that similarity-based approaches are closely related to network-based approaches, however some algorithms have been specifically designed for the former and network-based methods usually involve modeling a larger variety of information (apart from similarity between drugs and diseases).

3.3.1. Drug-drug similarity match

Drug-drug similarity analysis begins with the comparison between the chemical and/or biological profiles of different drugs, after which drugs can be repositioned to the indications of another one if the two shared sufficient similarity. An example to illustrate this idea is the work by Napolitano et al. [100]. The authors first generated a drug similarity matrix by integrating multiple sources of information, such as similarity in transcriptomic changes after drug administration (from CMap), as well as similarity in drug target proteins and chemical structures. The authors then integrated this information via a machine learning (ML) framework and proposed candidates for repositioning based on similarity to the known drugs. Ferrero and Agarwal [80] proposed another approach to repositioning based on the overlap between the mechanisms of actions (MOA) of different drugs. The main idea is to compare the downstream genes that are perturbed by the drugs, which may reflect the underlying MOA.

3.3.2. Disease-disease similarity match

It is reasonable to hypothesize that if two diseases are similar enough, then the drugs that are indicated for one disease may be repositioned to treat the other. This concept has been used in a number of studies in drug repositioning. For instance, Gottlieb et al. [101] proposed an algorithm (PREDICT) to combine multiple measures of drug-drug similarity and disease-disease similarity for repositioning. Later, Wang et al. [102] combined the approach in PREDICT with drug and disease data from DrugBank, CMap, OFF-SIDE (a side effect database from PharmGKB) and literature text mining to create a recommendation system for novel drug candidates. While the aforementioned studies did not directly utilize GWAS data, GWAS data may be used to measure disease-disease

similarity as well. For example, as a proof of concept, Li et al. [78] examined pairs of genetic variants shared between diseases while considering eQTL information at the same time. They found that pairs of diseases that were genetically similar were also more likely to be comorbid in clinical samples. The study suggests that genetic data from GWAS may also be used to assess 'similarity', which might have implications in drug repositioning.

3.3.3. Limitations

Similarities between drugs and diseases may not be easy to define. For example, some studies used CMap to ascertain drug-drug similarity, but there are limitations of the CMap database (please also refer to discussions in the next section), such as lack of appropriate tissue-specific cell lines in some occasions. How to integrate different sources of data to define similarity remains an open question. By its nature, this approach may tend to uncover candidates that are similar to the known drugs, and is less capable of revealing drugs with novel mechanisms of actions. It is possible that the drug candidates are already suggested by clinical experience (or prescribed off-label) due to their similarity with the known ones. Network-based approaches may also share similar shortcomings, due to similarity in their principles.

3.4. Searching for reversed expression profiles between drugs and diseases

A more distinct approach to repositioning is to compare the expression profiles of drugs against those of specific diseases. The core hypothesis is that if a drug produces an expression profile that is opposite to that of a disease, then the drug may be considered a repositioning candidate (due to its potential to 'reverse' disease-related expression profiles). This approach does not require knowledge of the drug mechanisms or even drug targets; it can be applied as long as drug-induced expression profiles are available. In addition, it utilizes GWAS data across multiple genes instead of focusing on only the top significant ones. As described later, the method is also readily available to any complex diseases or traits with GWAS summary statistics available. Compared to similarity or network-based methods, this approach may have a greater chance of uncovering drugs with new MOA.

To perform this kind of analysis, one would need the following components: (1) Drug-induced (differential) expression profile; (2) Disease-induced (differential) expression profile; and (3) algorithms to determine the correlations between (1) and (2).

Publicly available data are available for drug-induced expression profiles. Popular choices include the Connectivity Map (CMap) [103] and its successor Library of Integrated Network-based Cellular Signatures (LINCS) L1000 data [104]. CMap is a collection of expression profiles by applying ~1300 compounds on 5 human cancer cell lines, with expression profiles for ~7000 genes as of Build 2. LINCS (L1000) is a database of a much larger scale that contains about 1,328,098 gene expression profiles as a result from the applications of ~42,553 perturbagens. A set of 1000 'landmark' transcripts was directly measured while the expression levels of other genes were mainly computed by imputation. A comprehensive review about these resources is given by Musa et al. [105].

For the purpose of this review, we shall mainly discuss GWAS-derived data for component (2). While one may use RNA-seq or microarray data in patients to directly estimate component (2), as argued in Ref. [81], there are several advantages of using imputed expression profiles from GWAS. For example, patients are often medicated before their samples are collected, which may affect their expression profiles. GWAS-imputed transcriptome is much less susceptible to confounding by medication or other environmental factors. In addition, expression can be imputed for a large variety of tissues, even for those (e.g. brain) which are dif-

ficult to access. The sample size of GWAS is also often much larger than standard RNA-seq or microarray studies on clinical samples. This approach of comparing GWAS-imputed expression against drug transcriptomes was proposed and adopted by So et al. [81] to prioritize repositioning candidates for psychiatric disorders. The method revealed numerous candidates supported by previous preclinical or clinical studies, and the approach was able to 'rediscover' known psychiatric medications for the respective disorders, despite having no prior knowledge of the drug indications.

Imputation of gene expression profiles can be performed by various tools, the most popular being PrediXcan [51] or its successor S-PrediXcan [106]. The latter is able to impute expression changes by GWAS summary statistics alone, which enhances the applicability of the above repositioning methodology as summary statistics are now widely available. Briefly, these tools first learn how SNPs may dictate gene expression levels from training data extracted from datasets such as GTEx, GEUVADIS and DGN. Based on the prediction model built, the tools can then predict or impute expression levels when presented with genotypes from new samples.

Finally, for component (3) of the analysis, we need to assess the (anti-)correlation between the expression profiles from drugs and diseases. For example, So et al. [81] adopted Spearman and Pearson correlation as well as Kolmogorov–Smirnov (KS) test to evaluate patterns of reversed expression.

3.4.1. Limitations

There are several limitations of this approach, many of which are also discussed in Ref. [81]. First, perturbation libraries such as CMap and L1000 are usually based on cell line data, meaning that all the limitations of cell line-based experiments may also apply. For instance, cell lines cannot model complex cell-cell or cell-ECM (extracellular matrix) interactions, and pharmacokinetic properties and dosage control are often dependent on experiment protocols [107]. Second, many of the cell lines are cancer cell lines; using them for expression profiling may not be optimal when the aim is to prioritize drugs for other diseases, such as psychiatric disorders or CVD. Third, not all therapeutically important drugs may work by reversing gene expression profiles [108], although the approach appears to work reasonably well at least in psychiatric disorders. Finally, imputation of expression profiles from GWAS data may not be as accurate as directly measuring the actual tissue-specific expression profiles from patient samples, although the latter is often difficult to collect. Some genes may be poorly imputed due to large environmental influence on gene expression [109], or that the original sample is not powerful enough to detect the eQTLs. Limitations of the GTEx sample discussed earlier also apply here.

3.5. Network-based approaches

3.5.1. Overview

Network-based drug repositioning approaches aim to uncover novel drug-disease relationships by integrating a wide variety of biological information. A typical network consists of nodes (genes, proteins, diseases/traits, compounds) and edges (often weighted in biological settings) connecting the nodes. This approach is a popular and well-established drug repositioning technique, which offers high flexibility, as it allows for consideration of multiple dimensions of data sources. The types of biological networks that are useful for drug repositioning include, for example, gene regulatory, gene-gene interaction, metabolic, drug-target interaction (DTI) and protein–protein interaction (PPI) [84] networks. It is preferable to integrate multiple biological networks in order to reduce noise and improve biological relevance [83,85]. After preparing a curated network of information, analyzing the graphs usually involves tools developed from graph theory. For drug repositioning, two types of network analysis approaches might be useful: clustering

and propagation [36]. Briefly, clustering is a way of discovering subnetworks by the similarity of its elements, since biologically related entities should intuitively share a handful of underlying connections [110,111]. Clustering may reveal subnetworks and new relationships between drugs and diseases, fostering the discovery of drug candidates. Propagation approaches, as the name suggests, models the propagation of information from a source node to its surroundings [112]. Approaches such as random walk [113] may be used to assess the distance between a drug and a disease, thereby prioritizing drug candidates. For the purpose of this reivew, we shall not focus on the general approaches of network-based drug repositioning which has already been extensively reviewed [82–86,114–117]. Instead, we shall focus on the role and contribution of GWAS data towards this type of repositioning.

3.5.2. Using GWAS data in network-based drug repositioning

GWAS can reveal associations between a genetic variant and multiple diseases, hence providing new repositioning candidates or drug targets. In a recent work, Gaspar et al. built bipartite drug-target networks by leveraging gene-based statistics from MAGMA and S-PrediXcan, accounting for both SNP-level associations and imputed transcriptomic changes. They also built an online tool Drug Targetor (drugtargetor.com), which visualizes the resulting drug-target network. The authors built a network from GWAS of major depressive disorder, and suggested potential new drug candidates and their modes of actions [118].

In another application, a network-based approach was used to analyze GWAS results of CVD [119]. The study began with pathway analysis of the significant SNPs from 16 GWAS datasets. The identified pathways were then mapped to the PPI network InWeb and analyzed using random walk. Next, the authors examined the topological properties of the identified clusters and prioritized CVD-associated genes that displayed high centrality and betweenness, which may be prioritized as potential drug targets. In another report, Shu et al. [120] studied shared genetic networks and key 'driver genes' for both CVD and T2D using a systematic approach. They first considered co-expression modules for CVD and type 2 diabetes (T2D), and incorporated gene regulatory networks from GIANT [121] and Bayesian networks constructed from CVD and T2D related tissues to prioritize key drivers. The potential driver genes or key regulators were then further validated by cell culture and animal models. These genes might serve as useful targets for drug discovery or repositioning. While the above two examples were not solely aimed at drug repositioning, they demonstrated how network-based methods may shed light on biological processes and help prioritize drug target genes. The application of GWAS data in network-based repositioning has also been reviewed by Nabirotchkin et al. [122] recently.

3.5.3. Limitations

First, integrating different sources of networks can be tedious and challenging in terms of data cleaning, and this might require a certain degree of understanding of the data and their biological meanings. Second, network analysis heavily relies on the similarity or closeness between different entities [86], therefore the repositioning results may be concentrated around the 'nearby pharmacological space'; it may be relatively difficult to uncover drugs with novel mechanisms. Also, the strengths of similarity between nodes (i.e. edge strengths) are often hard to define. In addition, limitations in data sources could affect the performance of network-based analysis methods, for example the reliability and completeness of interactome data [123] and the difficulty in constructing accurate PPI networks [124]. Finally, the complicated nature of network structures makes parameter optimization a key issue when analyzing the data with network-based algorithms [125].

4. Conclusion and future directions

Despite the fact that more than 4000 GWAS have been conducted thus far, comparatively few studies have systematically analyzed the potential of GWAS in drug discovery or repositioning. With growing resources from biobanks (e.g. the UK Biobank) and increasing availability of GWAS data, such data should provide a very rich resource for guiding drug discovery/repositioning. Indeed, as discussed above, a number of studies have provided early evidence that human genomics data from GWAS might improve the success of drug development, or repositioning strategies based on GWAS are able to 're-discover' known drugs for diseases and/or suggest reasonable new candidates.

We focused on computational approaches to *prioritize* drug candidates in this review, but we should emphasize that further experimental and clinical studies are necessary to confirm the findings. Computational or bioinformatics approaches help to narrow down the search space and improve the success rate of drug development by prioritizing the best candidates, but they are not designed to provide confirmatory evidence. Nevertheless, given the huge cost and long time involved in developing a new drug, even a tiny improvement in success rate would translate to very substantial savings in absolute terms.

We hereby make a few comments to highlight several general limitations and directions for further research. First, to improve the accuracy of GWAS-based drug-repositioning studies, there is a need for more high-quality data, including GWAS with large cohort sizes and richer phenotypes, GTEx or other eQTL projects on more tissues and larger sample sizes, perturbation libraries that involve more comprehensive drug testing models etc. Second, new statistical methods have emerged to ascertain not only associations but also *causal* relationships between exposure (or risk factors) and outcomes; one of the most prominent methods is Mendelian Randomization (MR) [126]. A few studies have suggested that MR may be used to model the effects or side-effects of drugs with known targets. This may have implications for drug discovery or repositioning, although further development in methodology may be required [127–129]. Third, machine learning (ML) and artificial intelligence are among the fastest growing areas in recent years. ML approaches such as deep learning and other methods hold great promise for accelerating drug discovery/repositioning, as they may be able to discover and predict with higher accuracy the complex patterns and relationships between genes, drugs and diseases [130–133]. For example, ML methods may be used to capture complex relationships between drug transcriptome and the drug's treatment potential for specific diseases [134]. For diseases with high heterogeneity, a drug may only be useful for a subgroup of patients [135]. Unsupervised learning methods may help to subtype diseases more accurately, enhancing the success rate of drug development [136]. Finally, GWAS is a very rich source of data, but integration with other forms of human omics data, such as exome sequencing, transcriptomics and epigenomics studies will further improve the reliability of repositioning. In the same vein, another important direction is the integration of large-scale EMR with multi-omics data in drug development [58,64]. As emphasized in this review, each repositioning method has its own strengths and limitations; an important future direction is to integrate different methodologies for optimal prediction of drug candidates.

CRediT authorship contribution statement

Alexandria Li-Ching Lau: Conceptualization, Writing - original draft. **Hon-Cheong So:** Supervision, Conceptualization, Writing - original draft.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

References

[1] DiMasi JA, Grabowski HG, Hansen RW. Innovation in the pharmaceutical industry: new estimates of R&D costs. J Health Econ 2016;47:20–33. https://doi.org/10.1016/j.jhealeco.2016.01.012.

[2] Ashburn TT, Thor KB. Drug repositioning: identifying and developing new uses for existing drugs. Nat Rev Drug Discov 2004;3:673–83. https://doi.org/10.1038/nrd1468.

[3] Shim JS, Liu JO. Recent advances in drug repositioning for the discovery of new anticancer drugs. Int J Biol Sci 2014;10:654–63. https://doi.org/10.7150/ijbs.9224.

[4] Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, et al. 10 Years of GWAS discovery: biology, function, and translation. Am J Hum Genet 2017;101:5–22. https://doi.org/10.1016/j.ajhg.2017.06.005.

[5] Cui Y, Kang G, Sun K, Qian M, Romero R, Fu W. Gene-centric genomewide association study via entropy. Genetics 2008;179:637–50. https://doi.org/10.1534/genetics.107.082370.

[6] Ruiz-Marín M, Matilla-García M, Cordoba JAG, Susillo-González JL, Romo-Astorga A, González-Pérez A, et al. An entropy test for single-locus genetic association analysis. BMC Genet 2010;11:19. https://doi.org/10.1186/1471-2156-11-19.

[7] Joo J, Kwak M, Ahn K, Zheng G. A robust genome-wide scan statistic of the wellcome trust case-control consortium. Biometrics 2009;65:1115–22. https://doi.org/10.1111/j.1541-0420.2009.01185.x.

[8] Joo J, Kwak M, Zheng G. Improving power for testing genetic association in case-control studies by reducing the alternative space. Biometrics 2010;66:266–76. https://doi.org/10.1111/j.1541-0420.2009.01241.x.

[9] Bi W, Kang G, Pounds SB. Statistical selection of biological models for genome-wide association analyses. Methods 2018;145:67–75. https://doi.org/10.1016/j.ymeth.2018.05.019.

[10] Freidlin B, Zheng G, Li Z, Gastwirth JL. Trend tests for case-control studies of genetic markers: power, sample size and robustness. Hum Hered 2002;53:146–52. https://doi.org/10.1159/000064976.

[11] So H-C, Sham PC. Robust association tests under different genetic models, allowing for binary or quantitative traits and covariates. Behav Genet 2011;41:768–75. https://doi.org/10.1007/s10519-011-9450-9.

[12] Das S, Abecasis GR, Browning BL. Genotype imputation from large reference panels. Annu Rev Genom Hum Genet 2018;19:73–96. https://doi.org/10.1146/annurev-genom-083117-021602.

[13] Marees AT, de Kluiver H, Stringer S, Vorspan F, Curis E, Marie-Claire C, et al. A tutorial on conducting genome-wide association studies: quality control and statistical analysis. Int J Methods Psychiatr Res 2018;27. https://doi.org/10.1002/mpr.1608.

[14] Bush WS, Moore JH. Chapter 11: Genome-wide association studies. PLoS Comput Biol 2012;8. https://doi.org/10.1371/journal.pcbi.1002822.

[15] Gallagher MD, Chen-Plotkin AS. The post-GWAS era: from association to function. Am J Hum Genet 2018;102:717–30. https://doi.org/10.1016/j.ajhg.2018.04.002.

[16] Need AC, Goldstein DB. Whole genome association studies in complex diseases: where do we stand?. Dialog Clin Neurosci 2010;12:37–46.

[17] Nestler EJ, Hyman SE. Animal models of neuropsychiatric disorders. Nat Neurosci 2010;13:1161–9. https://doi.org/10.1038/nn.2647.

[18] Mak IW, Evaniew N, Ghert M. Lost in translation: animal models and clinical trials in cancer treatment. Am J Transl Res 2014;6:114–8.

[19] Nelson MR, Tipney H, Painter JL, Shen J, Nicoletti P, Shen Y, et al. The support of human genetic evidence for approved drug indications. Nat Genet 2015;47:856–60. https://doi.org/10.1038/ng.3314.

[20] King EA, Davis JW, Degner JF. Are drug targets with genetic support twice as likely to be approved? Revised estimates of the impact of genetic support for drug mechanisms on the probability of drug approval: supplementary methods and results. Genetics 2019. https://doi.org/10.1101/513945.

[21] Schaub MA, Boyle AP, Kundaje A, Batzoglou S, Snyder M. Linking disease associations with regulatory information in the human genome. Genome Res 2012;22:1748–59. https://doi.org/10.1101/gr.136127.111.

[22] Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, et al. Systematic localization of common disease-associated variation in regulatory DNA. Science 2012;337:1190–5. https://doi.org/10.1126/science.1222794.

[23] Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M, et al. Biological, clinical and population relevance of 95 loci for blood lipids. Nature 2010;466:707–13. https://doi.org/10.1038/nature09270.

[24] Swerdlow DI, Preiss D, Kuchenbaecker KB, Holmes MV, Engmann JEL, Shah T, et al. HMG-coenzyme A reductase inhibition, type 2 diabetes, and bodyweight: evidence from genetic analysis and randomised trials. The Lancet 2015;385:351–61. https://doi.org/10.1016/S0140-6736(14)61183-1.

[25] Law MR. Quantifying effect of statins on low density lipoprotein cholesterol, ischaemic heart disease, and stroke: systematic review and meta-analysis. BMJ 2003;326:1423. https://doi.org/10.1136/bmj.326.7404.1423.

[26] Edwards SL, Beesley J, French JD, Dunning AM. Beyond GWASs: illuminating the dark road from association to function. Am J Hum Genet 2013;93:779–97. https://doi.org/10.1016/j.ajhg.2013.10.012.

[27] Pritchard JE, O'Mara TA, Glubb DM. Enhancing the promise of drug repositioning through genetics. Front Pharmacol 2017;8:896. https://doi.org/10.3389/fphar.2017.00896.

[28] Nishizaki SS, Boyle AP. Mining the unknown: assigning function to noncoding single nucleotide polymorphisms. Trends Genet 2017;33:34–45. https://doi.org/10.1016/j.tig.2016.10.008.

[29] Miller JE, Veturi Y, Ritchie MD. Innovative strategies for annotating the "relationSNP" between variants and molecular phenotypes. BioData Min 2019;12:10. https://doi.org/10.1186/s13040-019-0197-9.

[30] Spain SL, Barrett JC. Strategies for fine-mapping complex traits. Hum Mol Genet 2015;24:R111–9. https://doi.org/10.1093/hmg/ddv260.

[31] Pushpakom S, Iorio F, Eyers PA, Escott KJ, Hopper S, Wells A, et al. Drug repurposing: progress, challenges and recommendations. Nat Rev Drug Discov 2019;18:41–58. https://doi.org/10.1038/nrd.2018.168.

[32] Li J, Zheng S, Chen B, Butte AJ, Swamidass SJ, Lu Z. A survey of current trends in computational drug repositioning. Brief Bioinform 2016;17:2–12. https://doi.org/10.1093/bib/bbv020.

[33] Dudley JT, Deshpande T, Butte AJ. Exploiting drug–disease relationships for computational drug repositioning. Brief Bioinform 2011;12:303–11. https://doi.org/10.1093/bib/bbr013.

[34] Liu Z, Fang H, Reagan K, Xu X, Mendrick DL, Slikker W, et al. In silico drug repositioning – what we need to know. Drug Discov Today 2013;18:110–5. https://doi.org/10.1016/j.drudis.2012.08.005.

[35] Hurle MR, Yang L, Xie Q, Rajpal DK, Sanseau P, Agarwal P. Computational drug repositioning: from data to therapeutics. Clin Pharmacol Ther 2013;93:335–41. https://doi.org/10.1038/clpt.2013.1.

[36] Xue H, Li J, Xie H, Wang Y. Review of drug repositioning approaches and resources. Int J Biol Sci 2018;14:1232–44. https://doi.org/10.7150/ijbs.24612.

[37] Schaid DJ, Chen W, Larson NB. From genome-wide associations to candidate causal variants by statistical fine-mapping. Nat Rev Genet 2018;19:491–504. https://doi.org/10.1038/s41576-018-0016-z.

[38] Benner C, Havulinna AS, Järvelin M-R, Salomaa V, Ripatti S, Pirinen M. Prospects of fine-mapping trait-associated genomic regions by using summary statistics from genome-wide association studies. Am J Hum Genet 2017;101:539–51. https://doi.org/10.1016/j.ajhg.2017.08.012.

[39] Benner C, Spencer CCA, Havulinna AS, Salomaa V, Ripatti S, Pirinen M. FINEMAP: efficient variable selection using summary data from genome-wide association studies. Bioinformatics 2016;32:1493–501. https://doi.org/10.1093/bioinformatics/btw018.

[40] Kichaev G, Yang W-Y, Lindstrom S, Hormozdiari F, Eskin E, Price AL, et al. Integrating functional data to prioritize causal variants in statistical fine-mapping studies. PLoS Genet 2014;10. https://doi.org/10.1371/journal.pgen.1004722.

[41] Hormozdiari F, Kostem E, Kang EY, Pasaniuc B, Eskin E. Identifying causal variants at loci with multiple signals of association. Genetics 2014;198:497–508. https://doi.org/10.1534/genetics.114.167908.

[42] Chen W, Larrabee BR, Ovsyannikova IG, Kennedy RB, Haralambieva IH, Poland GA, et al. Fine mapping causal variants with an approximate bayesian method using marginal test statistics. Genetics 2015;200:719–36. https://doi.org/10.1534/genetics.115.176107.

[43] Wang J, Huang D, Zhou Y, Yao H, Liu H, Zhai S, et al. CAUSALdb: a database for disease/trait causal variants identified using summary statistics of genome-wide association studies. Nucleic Acids Res 2020;48:D807–16. https://doi.org/10.1093/nar/gkz1026.

[44] Li MJ, Wang LY, Xia Z, Sham PC, Wang J. GWAS3D: detecting human regulatory variants by integrative analysis of genome-wide associations, chromosome interactions and histone modifications. Nucleic Acids Res 2013;41:W150–8. https://doi.org/10.1093/nar/gkt456.

[45] Huang D, Yi X, Zhang S, Zheng Z, Wang P, Xuan C, et al. GWAS4D: multidimensional analysis of context-specific regulatory variant for human complex diseases and traits. Nucleic Acids Res 2018;46:W114–20. https://doi.org/10.1093/nar/gky407.

[46] Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res 2012;40:D930–4. https://doi.org/10.1093/nar/gkr917.

[47] Gambardella G, Cereda M, Benedetti L, Ciccarelli FD. MEGA-V: detection of variant gene sets in patient cohorts. Bioinformatics 2017;33:1248–9. https://doi.org/10.1093/bioinformatics/btw809.

[48] Xu Z, Taylor JA. SNPinfo: integrating GWAS and candidate gene information into functional SNP selection for genetic association studies. Nucleic Acids Res 2009;37:W600–5. https://doi.org/10.1093/nar/gkp290.

[49] Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. Nat Commun 2017;8:1–11. https://doi.org/10.1038/s41467-017-01261-5.

[50] Brodie A, Azaria JR, Ofran Y. How far from the SNP may the causative genes be?. Nucleic Acids Res 2016;44:6046–54. https://doi.org/10.1093/nar/gkw500.

[51] Gamazon ER, Wheeler HE, Shah KP, Mozaffari SV, Aquino-Michaels K, Carroll RJ, et al. A gene-based association method for mapping traits using reference transcriptome data. Nat Genet 2015;47:1091–8. https://doi.org/10.1038/ng.3367.

[52] Grenier L, Hu P. Computational drug repurposing for inflammatory bowel disease using genetic information. Comput Struct Biotechnol J 2019;17:127–35. https://doi.org/10.1016/j.csbj.2019.01.001.

[53] Napolitano F, Carrella D, Mandriani B, Pisonero-Vaquero S, Sirci F, Medina DL, et al. Gene2drug: a computational tool for pathway-based rational drug repositioning. Bioinformatics 2018;34:1498–505. https://doi.org/10.1093/bioinformatics/btx800.

[54] Tragante V, Hemerich D, Alshabeeb M, Brænne I, Lempiainen H, Patel R, et al. Druggability of coronary artery disease risk loci. Circ Genom Precis Med 2018;11:. https://doi.org/10.1161/CIRCGEN.117.001977e001977.

[55] Cotto KC, Wagner AH, Feng Y-Y, Kiwala S, Coffman AC, Spies G, et al. DGIdb 3.0: a redesign and expansion of the drug–gene interaction database. Nucleic Acids Res 2018;46:D1068–73. https://doi.org/10.1093/nar/gkx1143.

[56] Grover MP, Ballouz S, Mohanasundaram KA, George RA, Goscinski A, Crowley TM, et al. Novel therapeutics for coronary artery disease from genome-wide association study data. BMC Med Genomics 2015;8:1–11. https://doi.org/10.1186/1755-8794-8-S2-S1.

[57] Ballouz S, Liu JY, Oti M, Gaeta B, Fatkin D, Bahlo M, et al. Candidate disease gene prediction using Gentrepid: application to a genome-wide association study on coronary artery disease. Mol Genet Genomic Med 2014;2:44–57. https://doi.org/10.1002/mgg3.40.

[58] Rastegar-Mojarad M, Ye Z, Kolesar JM, Hebbring SJ, Lin SM. Opportunities for drug repositioning from phenome-wide association studies. Nat Biotechnol 2015;33:342–5. https://doi.org/10.1038/nbt.3183.

[59] Sanseau P, Agarwal P, Barnes MR, Pastinen T, Richards JB, Cardon LR, et al. Use of genome-wide association studies for drug repositioning. Nat Biotechnol 2012;30:317–20. https://doi.org/10.1038/nbt.2151.

[60] Wang Z-Y, Zhang H-Y. Rational drug repositioning by medical genetics. Nat Biotechnol 2013;31:1080–2. https://doi.org/10.1038/nbt.2758.

[61] Grover MP, Ballouz S, Mohanasundaram KA, George RA, Sherman CDH, Crowley TM, et al. Identification of novel therapeutics for complex diseases from genome-wide association data. BMC Med Genomics 2014;7:S8. https://doi.org/10.1186/1755-8794-7-S1-S8.

[62] Quan Y, Liu M-Y, Liu Y-M, Zhu L-D, Wu Y-S, Luo Z-H, et al. Facilitating anti-cancer combinatorial drug discovery by targeting epistatic disease genes. Molecules 2018;23:736. https://doi.org/10.3390/molecules23040736.

[63] Piñeiro-Yáñez E, Reboiro-Jato M, Gómez-López G, Perales-Patón J, Troulé K, Rodríguez JM, et al. PanDrugs: a novel method to prioritize anticancer drug treatments according to individual genomic data. Genome Med 2018;10:41. https://doi.org/10.1186/s13073-018-0546-1.

[64] Robinson JR, Denny JC, Roden DM, Driest SLV. Genome-wide and phenome-wide approaches to understand variable drug actions in electronic health records. Clin Transl Sci 2018;11:112–22. https://doi.org/10.1111/cts.12522.

[65] Finan C, Gaulton A, Kruger FA, Lumbers RT, Shah T, Engmann J, et al. The druggable genome and support for target identification and validation in drug development. Sci Transl Med 2017;9:eaag1166. https://doi.org/10.1126/scitranslmed.aag1166.

[66] Mäkinen V-P, Civelek M, Meng Q, Zhang B, Zhu J, Levian C, et al. Integrative genomics reveals novel molecular pathways and gene networks for coronary artery disease. PLOS Genet 2014;10:. https://doi.org/10.1371/journal.pgen.1004502e1004502.

[67] White MJ, Yaspan BL, Veatch OJ, Goddard P, Risse-Adams OS, Contreras MG. Strategies for pathway analysis using GWAS and WGS data. Curr Protoc Hum Genet 2019;100:. https://doi.org/10.1002/cphg.79e79.

[68] Sakagami M. Systemic delivery of biotherapeutics through the lung: opportunities and challenges for improved lung absorption. Ther Deliv 2013;4:1511–25. https://doi.org/10.4155/tde.13.119.

[69] Talevi A. Multi-target pharmacology: possibilities and limitations of the "skeleton key approach" from a medicinal chemist perspective. Front Pharmacol 2015. https://doi.org/10.3389/fphar.2015.00205.

[70] Bang S, Son S, Kim S, Shin H. Disease pathway cut for multi-target drugs. BMC Bioinf 2019;20:74. https://doi.org/10.1186/s12859-019-2638-3.

[71] Lu J-J, Pan W, Hu Y-J, Wang Y-T. Multi-target drugs: the trend of drug research and development. PLoS One 2012;7:. https://doi.org/10.1371/journal.pone.0040262e40262.

[72] Hebbring SJ. The challenges, advantages and future of phenome-wide association studies. Immunology 2014;141:157–65. https://doi.org/10.1111/imm.12195.

[73] Verma A, Ritchie MD. Current scope and challenges in phenome-wide association studies. Curr Epidemiol Rep 2017;4:321–9. https://doi.org/10.1007/s40471-017-0127-7.

[74] Huang QQ, Ritchie SC, Brozynska M, Inouye M. Power, false discovery rate and Winner's Curse in eQTL studies. Nucleic Acids Res 2018;46:e133. https://doi.org/10.1093/nar/gkv780.

[75] Heinig M. Using gene expression to annotate cardiovascular GWAS loci. Front Cardiovasc Med 2018;5. https://doi.org/10.3389/fcvm.2018.00059.

[76] Zeng B, Lloyd-Jones LR, Holloway A, Marigorta UM, Metspalu A, Montgomery GW, et al. Constraints on eQTL fine mapping in the presence of multisite local regulation of gene expression. G3 Genes Genomes Genet 2017;7:2533–44. https://doi.org/10.1534/g3.117.043752.

[77] Pan Y, Cheng T, Wang Y, Bryant SH. Pathway analysis for drug repositioning based on public database mining. J Chem Inf Model 2014;54:407–18. https://doi.org/10.1021/ci4005354.

[78] Li H, Fan J, Vitali F, Berghout J, Aberasturi D, Li J, et al. Novel disease syndromes unveiled by integrative multiscale network analysis of diseases sharing molecular effectors and comorbidities. BMC Med Genomics 2018;11:112. https://doi.org/10.1186/s12920-018-0428-9.

[79] de Jong S, Vidler LR, Mokrab Y, Collier DA, Breen G. Gene-set analysis based on the pharmacological profiles of drugs to identify repurposing opportunities in schizophrenia. J Psychopharmacol (Oxf) 2016;30:826–30. https://doi.org/10.1177/0269881116653109.

[80] Ferrero E, Agarwal P. Connecting genetics and gene expression data for target prioritisation and drug repositioning. BioData Min 2018;11:7. https://doi.org/10.1186/s13040-018-0171-y.

[81] So H-C, Chau CK-L, Chiu W-T, Ho K-S, Lo C-P, Yim SH-Y, et al. Analysis of genome-wide association data highlights candidates for drug repositioning in psychiatry. Nat Neurosci 2017;20:1342–9. https://doi.org/10.1038/nn.4618.

[82] Luo Y, Zhao X, Zhou J, Yang J, Zhang Y, Kuang W, et al. A network integration approach for drug-target interaction prediction and computational drug repositioning from heterogeneous information. Nat Commun 2017;8:1–13. https://doi.org/10.1038/s41467-017-00680-8.

[83] Lotfi Shahreza M, Ghadiri N, Mousavi SR, Varshosaz J, Green JR. A review of network-based approaches to drug repositioning. Brief Bioinform 2018;19:878–92. https://doi.org/10.1093/bib/bbx017.

[84] Alaimo S, Pulvirenti A. Network-based drug repositioning: approaches, resources, and research directions. In: Vanhaelen Q, editor. Comput. Methods Drug Repurposing. New York, NY: Springer; 2019. p. 97–113. https://doi.org/10.1007/978-1-4939-8955-3_6.

[85] Wu Z, Wang Y, Chen L. Network-based drug repositioning. Mol Biosyst 2013;9:1268–81. https://doi.org/10.1039/C3MB25382A.

[86] Yu D, Kim M, Xiao G, Hwang TH. Review of biological network data and its applications. Genomics Inform 2013;11:200–10. https://doi.org/10.5808/GI.2013.11.4.200.

[87] Stringer S, Wray NR, Kahn RS, Derks EM. Underestimated effect sizes in GWAS: fundamental limitations of single SNP analysis for dichotomous phenotypes. PLoS One 2011;6. https://doi.org/10.1371/journal.pone.0027964.

[88] So H-C, Chau CK-L, Lau A, Wong S-Y, Zhao K. Translating GWAS findings into therapies for depression and anxiety disorders: gene-set analyses reveal enrichment of psychiatric drug classes and implications for drug repositioning. Psychol Med 2018:1–17. https://doi.org/10.1017/S0033291718003641.

[89] Bakshi A, Zhu Z, Vinkhuyzen AAE, Hill WD, McRae AF, Visscher PM, et al. Fast set-based association analysis using summary data from GWAS identifies novel gene loci for human complex traits. Sci Rep 2016;6:32894. https://doi.org/10.1038/srep32894.

[90] de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. PLOS Comput Biol 2015;11:. https://doi.org/10.1371/journal.pcbi.1004219e1004219.

[91] Jhamb D, Magid-Slav M, Hurle MR, Agarwal P. Pathway analysis of GWAS loci identifies novel drug targets and repurposing opportunities. Drug Discov Today 2019;24:1232–6. https://doi.org/10.1016/j.drudis.2019.03.024.

[92] Shen J, Song K, Slater AJ, Ferrero E, Nelson MR. STOPGAP: a database for systematic target opportunity assessment by genetic association predictions. Bioinformatics 2017;33:2784–6. https://doi.org/10.1093/bioinformatics/btx274.

[93] Gaspar HA, Breen G. Drug enrichment and discovery from schizophrenia genome-wide association results: an analysis and visualisation approach. Sci Rep 2017;7:1–9. https://doi.org/10.1038/s41598-017-12325-3.

[94] Wong BC, Chau CK, Ao F, Mo C, Wong S, Wong Y, et al. Differential associations of depression-related phenotypes with cardiometabolic risks: polygenic analyses and exploring shared genetic variants and pathways. Depress Anxiety 2019;36:330–44. https://doi.org/10.1002/da.22861.

[95] So H-C, Wong Y-H. Implications of de novo mutations in guiding drug discovery: a study of four neuropsychiatric disorders. J Psychiatr Res 2019;110:83–92. https://doi.org/10.1016/j.jpsychires.2018.12.015.

[96] Breen G, Li Q, Roth BL, O'Donnell P, Didriksen M, Dolmetsch R, et al. Translating genome-wide association findings into new therapeutics for psychiatry. Nat Neurosci 2016;19:1392–6. https://doi.org/10.1038/nn.4411.

[97] Garcia-Albornoz M, Nielsen J. Finding directionality and gene-disease predictions in disease associations. BMC Syst Biol 2015;9:35. https://doi.org/10.1186/s12918-015-0184-9.

[98] Melott JM, Weinstein JN, Broom BM. PathwaysWeb: a gene pathways API with directional interactions, expanded gene ontology, and versioning. Bioinformatics 2016;32:312–4. https://doi.org/10.1093/bioinformatics/btv554.

[99] Luo W, Pant G, Bhavnasi YK, Blanchard SG, Brouwer C. Pathview Web: user friendly pathway visualization and data integration. Nucleic Acids Res 2017;45:W501–8. https://doi.org/10.1093/nar/gkx372.

[100] Napolitano F, Zhao Y, Moreira VM, Tagliaferri R, Kere J, D'Amato M, et al. Drug repositioning: a machine-learning approach through data integration. J Cheminformat 2013;5:30. https://doi.org/10.1186/1758-2946-5-30.

[101] Gottlieb A, Stein GY, Ruppin E, Sharan R. PREDICT: a method for inferring novel drug indications with application to personalized medicine. Mol Syst Biol 2011;7:496. https://doi.org/10.1038/msb.2011.26.

[102] Wang H, Gu Q, Wei J, Cao Z, Liu Q. Mining drug–disease relationships as a complement to medical genetics-based drug repositioning: where a recommendation system meets genome-wide association studies. Clin Pharmacol Ther 2015;97:451–4. https://doi.org/10.1002/cpt.82.

[103] Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ, et al. The connectivity map: using gene-expression signatures to connect small molecules, genes, and disease. Science 2006;313:1929–35. https://doi.org/10.1126/science.1132939.

[104] Keenan AB, Jenkins SL, Jagodnik KM, Koplev S, He E, Torre D, et al. The library of integrated network-based cellular signatures NIH program: system-level cataloging of human cells response to perturbations. Cell Syst 2018;6:13–24. https://doi.org/10.1016/j.cels.2017.11.001.

[105] Musa A, Ghoraie LS, Zhang S-D, Glazko G, Yli-Harja O, Dehmer M, et al. A review of connectivity map and computational approaches in pharmacogenomics. Brief Bioinform 2017;19:506–23. https://doi.org/10.1093/bib/bbw112.

[106] Barbeira AN, Dickinson SP, Bonazzola R, Zheng J, Wheeler HE, Torres JM, et al. Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. Nat Commun 2018;9:1–20. https://doi.org/10.1038/s41467-018-03621-1.

[107] Langhans SA. Three-dimensional in vitro cell culture models in drug discovery and drug repositioning. Front Pharmacol 2018.. https://doi.org/10.3389/fphar.2018.00006.

[108] Chen B, Ma L, Paik H, Sirota M, Wei W, Chua M-S, et al. Reversal of cancer gene expression correlates with drug efficacy and reveals therapeutic targets. Nat Commun 2017;8:1–12. https://doi.org/10.1038/ncomms16022.

[109] Ioannidis NM, Wang W, Furlotte NA, Hinds DA, Bustamante CD, Jorgenson E, et al. Gene expression imputation identifies candidate genes and susceptibility loci associated with cutaneous squamous cell carcinoma. Nat Commun 2018;9:1–9. https://doi.org/10.1038/s41467-018-06149-6.

[110] Brohée S, Faust K, Lima-Mendez G, Vanderstocken G, van Helden J. Network analysis tools: from biological networks to clusters and pathways. Nat Protoc 2008;3:1616–29. https://doi.org/10.1038/nprot.2008.100.

[111] Sharma A, Ali HH. Analysis of clustering algorithms in biological networks. In: 2017 IEEE Int. Conf. Bioinforma. Biomed. BIBM. p. 2303–5. https://doi.org/10.1109/BIBM.2017.8218036.

[112] Picart-Armada S, Barrett SJ, Willé DR, Perera-Lluna A, Gutteridge A, Dessailly BH. Benchmarking network propagation methods for disease gene identification. PLOS Comput Biol 2019;15:. https://doi.org/10.1371/journal.pcbi.1007276e1007276.

[113] Luo H, Wang J, Li M, Luo J, Peng X, Wu F-X, et al. Drug repositioning based on comprehensive similarity measures and bi-random walk algorithm. Bioinformatics 2016;32:2664–71. https://doi.org/10.1093/bioinformatics/btw228.

[114] Ma X, Gao L. Biological network analysis: insights into structure and functions. Brief Funct Genomics 2012;11:434–42. https://doi.org/10.1093/bfgp/els045.

[115] Ideker T, Nussinov R. Network approaches and applications in biology. PLOS Comput Biol 2017;13:. https://doi.org/10.1371/journal.pcbi.1005771e1005771.

[116] Charitou T, Bryan K, Lynn DJ. Using biological networks to integrate, visualize and analyze genomics data. Genet Sel Evol 2016;48:27. https://doi.org/10.1186/s12711-016-0205-1.

[117] Zhu X, Gerstein M, Snyder M. Getting connected: analysis and principles of biological networks. Genes Dev 2007;21:1010–24. https://doi.org/10.1101/gad.1528707.

[118] Gaspar HA, Gerring Z, Hübel C. Major depressive disorder working group of the psychiatric genomics consortium, Middeldorp CM, Derks EM, et al. Using genetic drug-target networks to develop new drug hypotheses for major depressive disorder. Transl Psychiatry 2019;9:117. https://doi.org/10.1038/s41398-019-0451-4.

[119] Sujoy Ghosh, Juan Vivar, Nelson Christopher P, Christina Willenborg, Segrè Ayellet V, Ville-Petteri Mäkinen, et al. Systems genetics analysis of genome-wide association study reveals novel associations between key biological processes and coronary artery disease. Arterioscler Thromb Vasc Biol 2015;35:1712–22. https://doi.org/10.1161/ATVBAHA.115.305513.

[120] Shu L, Chan KHK, Zhang G, Huan T, Kurt Z, Zhao Y, et al. Shared genetic regulatory networks for cardiovascular disease and type 2 diabetes in multiple populations of diverse ethnicities in the United States. PLOS Genet 2017;13:. https://doi.org/10.1371/journal.pgen.1007040e1007040.

[121] Greene CS, Krishnan A, Wong AK, Ricciotti E, Zelaya RA, Himmelstein DS, et al. Understanding multicellular function and disease with human tissue-specific networks. Nat Genet 2015;47:569–76. https://doi.org/10.1038/ng.3259.

[122] Nabirotchkin S, Peluffo AE, Rinaudo P, Yu J, Hajj R, Cohen D. Next-generation drug repurposing using human genetics and network biology. Curr Opin Pharmacol 2020. https://doi.org/10.1016/j.coph.2019.12.004.

[123] Dai Y-F, Zhao X-M. A Survey on the Computational Approaches to Identify Drug Targets in the Postgenomic Era. BioMed Res Int 2015. https://doi.org/10.1155/2015/239654.

[124] Radivojac P, Clark WT, Oron TR, Schnoes AM, Wittkop T, Sokolov A, et al. A large-scale evaluation of computational protein function prediction. Nat Methods 2013;10:221–7. https://doi.org/10.1038/nmeth.2340.

[125] Wu Z, Wang Y, Chen L. Drug repositioning framework by incorporating functional information. IET Syst Biol Stevenage 2013;7:188–94.

[126] Burgess S, Small DS, Thompson SG. A review of instrumental variable estimators for Mendelian randomization. Stat Methods Med Res 2017;26:2333–55. https://doi.org/10.1177/0962280215597579.

[127] Mokry LE, Ahmad O, Forgetta V, Thanassoulis G, Richards JB. Mendelian randomisation applied to drug development in cardiovascular disease: a review. J Med Genet 2015;52:71–9. https://doi.org/10.1136/jmedgenet-2014-102438.

[128] Walker VM, Davey Smith G, Davies NM, Martin RM. Mendelian randomization: a novel approach for the prediction of adverse drug events and drug repurposing opportunities. Int J Epidemiol 2017;46:2078–89. https://doi.org/10.1093/ije/dyx207.

[129] Hon-Cheong S, Chau CK, Zhao K. Exploring repositioning opportunities and side-effects of statins: a Mendelian randomization study of HMG-CoA reductase inhibition with 55 complex traits. Genetics 2017. https://doi.org/10.1101/170241.

[130] Zhao K, So H-C. Using drug expression profiles and machine learning approach for drug repurposing. In: Vanhaelen Q, editor. Comput. Methods Drug Repurposing. New York, NY: Springer; 2019. p. 219–37. https://doi.org/10.1007/978-1-4939-8955-3_13.

[131] Vamathevan J, Clark D, Czodrowski P, Dunham I, Ferran E, Lee G, et al. Applications of machine learning in drug discovery and development. Nat Rev Drug Discov 2019;18:463–77. https://doi.org/10.1038/s41573-019-0024-5.

[132] Chen H, Engkvist O, Wang Y, Olivecrona M, Blaschke T. The rise of deep learning in drug discovery. Drug Discov Today 2018;23:1241–50. https://doi.org/10.1016/j.drudis.2018.01.039.

[133] Yella J, Yaddanapudi S, Wang Y, Jegga A. Changing trends in computational drug repositioning. Pharmaceuticals 2018;11:57. https://doi.org/10.3390/ph11020057.

[134] Zhao K, So H-C. Drug repositioning for schizophrenia and depression/anxiety disorders: a machine learning approach leveraging expression data. IEEE J Biomed Health Inform 2019;23:1304–15. https://doi.org/10.1109/JBHI.2018.2856535.

[135] Dugger SA, Platt A, Goldstein DB. Drug development in the era of precision medicine. Nat Rev Drug Discov 2018;17:183–96. https://doi.org/10.1038/nrd.2017.226.

[136] Yin L, Chau CKL, Sham P-C, So H-C. Uncovering complex disease subtypes by integrating clinical data and imputed transcriptome from genome-wide association studies: applications in psychiatry and cardiology. Am J Hum Genet, 105(6);1193–212. doi: 10.1016/j.ajhg.2019.10.012.