We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

154
Countries delivered to

Our authors are among the

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS

BOOK CITATION INDEX

INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Chapter

Gene Signature-Based Drug Repositioning

*Zhilong Jia, Xinyu Song, Jinlong Shi, Weidong Wang
and Kunlun He*

Abstract

With the advent of dynamical omics technology, especially the transcriptome and proteome, a huge amount of data related to various diseases and approved drugs are available under multi global projects or researches with their interests. These omics data and new machine learning technology largely promote the translation of drug research into clinical trials. We will cover the following topics in this chapter. 1) An introduction to the basic discipline of gene signature-based drug repurposing; 2) databases of genes, drugs and diseases; 3) gene signature databases of the approved drugs; 4) gene signature databases of various diseases; 5) gene signature-based methods and tools for drug repositioning; 6) new omics technology for drug repositioning; 7) drug repositioning examples with reproducible code. And finally, discuss the future trends and conclude.

Keywords: transcriptome, databases, drug repurposing, mode of action, reproducible study

1. Introduction

Drug repositioning is to identify new indications of the approved drugs. It has lower risk, less human resources, lower cost, and shorter developmental period, compared with traditional drug development. Sir James Black, a Nobel Prize laureate, originally stated that "The most fruitful basis for the discovery of a new drug is to start with an old drug", largely promoting the concept of drug repositioning [1]. There are huge examples of drug repositioning as described in the book. Multinational pharmaceutical companies, such as AstraZeneca and GSK, also showed their great interest in drug repurposing approaches [2, 3].

In this chapter, we focus on gene signature-based drug repositioning. The idea could date from 2000 year. Hughes et al. built a prototypical library of the microarray-based gene expression signatures of Yeast with about 300 diverse gene mutations and the treatment of 13 drugs with known molecular targets by keeping other experimental conditions consistent [4]. They identified a new target of the drug dyclonine by comparing the signatures of genes and drugs via pattern matching [4]. This article opened a door for gene signature-based drug repositioning [5].

A comprehensive gene signature library of genes, diseases and perturbations plays a fundamental role in gene-signature-based drug repositioning. From the genes' view, the knocking down, knocking out, knocking in genes could be achieved to represent the expression signatures of genes with the advances of molecular biology, especially the emergence of the RNAi and CRISPR/Cas9 technology [6].

From the diseases' view, modeling disease in a cell or animal experimental assay would make it possible to produce the gene signatures of various diseases via the quantification of molecular phenotypes. It should be noted that modeling various diseases in parallel and high throughput ways are relatively difficult so far as the condition of modeling various diseases is disease-specific or unclear due to the complexity and our little understanding of some diseases. However, with the development of the pathogenesis of various diseases, it will be efficient to model cellular and animal models of various diseases by magic genome editing using CRISPR/Cas9 technology [7].

Finally, from a drugs' view, there are thousands of approved drugs available so far. Lots of the bioactive compounds, besides the approved drugs, were also tested to obtain their gene signatures. Particularly, the connectivity map (CMap) [8] and Library of Integrated Network-based Cellular Signatures (LINCS) program [9, 10] largely promoted the rapid development of drug repositioning as they provided a huge of gene signatures of drugs and compounds freely available to the scientific community.

The core principle of gene signature-based drug repositioning is that the candidate drugs should revert the gene signature of the disease of interest, which is changed by the disease, compared with the controls (**Figure 1**). The reversion could be characterized by anti-correlation, distance, similarity and metrics produced by machine learning models. A derivative principle is that the similarity of two drugs could reveal similar indications of the two drugs. In detail, if drug A could be used to treat disease C, and the other drug B is similar to drug A based on their gene signatures, then drug B could also be used to treat disease C. This idea should come from chemoinformatics as the principle that similar drugs based on chemical structures should have similar functions is widely used in the field of drug research and development, especially the development of me-too drugs [11]. Importantly,
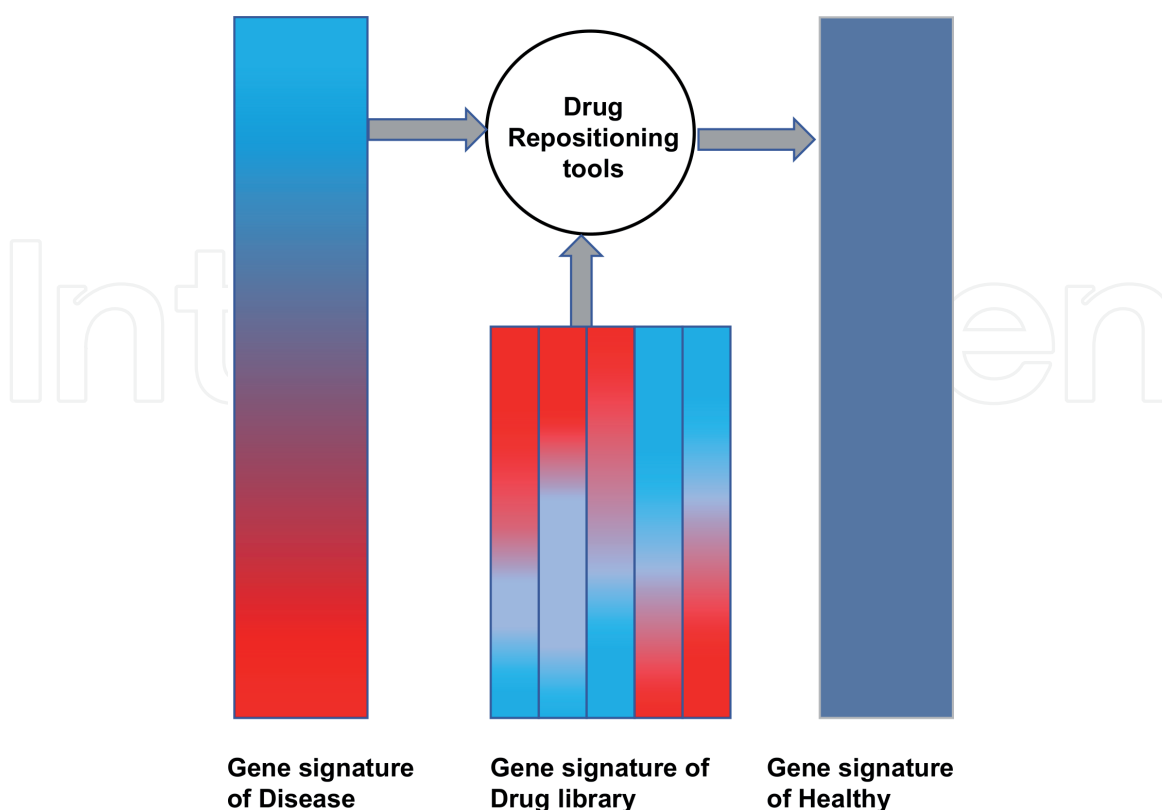

Figure 1.
The core idea of gene signature-based drug repositioning. Drug repositioning tools search the gene signatures of a drug library to identify which signature is "opposite" to the gene signature of disease, reverting the state of disease to the healthy state.

several researchers have developed or detailed this principle from different perspectives, making this idea efficient to implement and use.

The gene signatures are the molecular phenotype, revealing the molecular landscape of genes, diseases or drugs. In general, the gene signatures are the expression profiles or changes of RNA measured by RNASeq-based transcriptome via microarray, Next-Generation Sequencing or Third-Generation Sequencing [5, 8]. More broadly, the gene signature could be the abundance profiles or changes of proteins qualified by the antibody-based or tandem mass spectrometry (MS/MS)-based proteome. The reason why is that the principle of gene signature-based drug repositioning is suitable to any molecular phenotype, such as the transcriptome and proteome. Moreover, in machine learning models, the tabular data of transcriptome and proteome is similar to a great extent as they are features of samples in a high-level and united view.

In summary, with the rapid advance of various omic technology, a huge amount of public available omic data related to molecules, drugs, diseases and genes, computational resources and efficient deep learning algorithms make the field of drug repositioning vigorous. There will be increasing therapeutic applications of drug repositioning. In the following sections, we will introduce the databases related to genes, pathways, drugs and diseases, providing the resources for gene signature-based drug repositioning, then describe key tools for web servers for drug repositioning with a highlight on the new powerful and easy-to-use methods, show examples for drug repositioning for several diseases with reproducible code, convenient to the readers to follow. Finally, we will summarize the ongoing challenges, unmet needs, future trends and conclude.

2. Databases of genes, pathways and drugs for drug repositioning

Genes play a critical role in gene signature-based drug repositioning. Especially, the targets of drugs are of importance in traditional drug development. In General, the targets of drugs are human or viral proteins, which are druggable [12] and associated with a particular disease or multi diseases. So far, there are about 900 biomolecules targeted by about 1500 US FDA-approved drugs as curated by Rita et al. [13]. Obtaining this information will facilitate the process of gene signature-based drug repositioning. Some databases and web servers have gene information, which are useful in drug development [14].

GeneCards (https://www.genecards.org/) is an integrative knowledge base and web server with comprehensive information on all human genes, scratching more than 150 high-quality web sources, from genotype to phenotypes and functional information [15]. Though it is a general database, which is not centric on drug development, it provides comprehensive knowledge about a gene of interest. It is highly recommended to browse this website at the beginning of a study of a target.

DGIdb (drug-gene interaction database, www.dgidb.org) is a webserver with drug-gene interaction and druggable genes information, collected from more than thirty high-quality web sources [16]. If biomarkers or therapeutic targets are identified, then researchers could search which drugs could target the biomarker or therapeutic target using DGIdb, achieving a quick translational opportunity.

The Open Targets database (https://www.opentargets.org/) aims to identify and prioritize promising therapeutic targets of drugs by analyzing human genetics, genomics and functional genomics data [17, 18]. The database emphasizes the importance of genetics of diseases via genome-wide association studies to approach gene causal inference, which is beneficial to drug development [19, 20].

The Clue.io webserver (https://clue.io/) includes the updated CMap LINCS gene expression resource perturbed by CRISPR gene over-expression, RNAi gene knock-down and CRISPR gene knockout generating loss-of-function mutants [9, 21]. This webserver has abundant data about the gene perturbation, providing a great resource to study the effect of a target, mimicking the targets affected by drugs [22–24]. Meanwhile, it also supplies a drug repositioning hub for researchers, a curated library of drugs with a companion knowledge resource [25].

Pathways, besides gene level, could also be a key resource in drug repositioning. Pathway, consisting of a set of genes, could be the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway, gene ontology (GO), Reactome Pathway Database (https://reactome.org/) and other gene sets. As genes in a pathway are not randomly selected, a generalized pathway concept is the gene set, substantially enlarging the function aspects of pathways. A good resource of the gene sets is the Molecular Signatures Database (MSigDB, https://www.gsea-msigdb.org/gsea/msigdb/) as it supplied a downloadable gmt-formatted gene set dataset, facilitating its use in the bioinformatic analysis [26]. Several reasons highlight the importance of the pathway. Firstly, it could be used to illuminate the mode of action of drugs by connecting the genes and drugs [27]. Secondly, it could be a feature summarizing the gene-signature at a higher level, which is useful in machine learning-based modeling. It is different from the gene level as it captures different information about drugs or diseases [28–30]. Thirdly, the pathway analysis could enhance the confidence of the prediction of the candidate drugs [31].

The information about drugs is an invaluable resource to drug repositioning and an evaluation dataset of drug repositioning. The repoDB database is a standard dataset to benchmark various computational repositioning methods, which consist of 6677 approved and 4123 failed drug-indication pairs [32]. The Experimental Knowledge-Based Drug Repositioning Database (EK-DRD, http://www.idruglab.com/drd/index.php) curated 1861 FDA-approved and 102 withdrawn drugs with validated drug repositioning annotations [33]. These datasets will facilitate the training and testing of the machine-learning-based models.

3. Gene signature databases related to drugs

The gene signature databases of drugs and compounds are fundamental resources determining the searching space for drug repositioning. For a long time, researchers have been pursuing the enlargement of the gene signature library of drugs and compounds. For example, researchers have explored a bunch of bioactive compounds and ligands, such as growth factors and cytokines, which are not drugs but with known functions [8–10]. There are lots of data resources related to drugs. The sources of these data are mainly from two aspects. One is the public data, such as GEO, which is scattered in the database. A manual curation by professional researchers is necessary to make a usable dataset for drug repositioning. There is a trend for advanced metadata curation from the GEO [34]. The other one is from large projects, such as CMap, aiming to create a reference dataset of gene signatures for drug development.

NCBI GEO [35], EMBL-EBI ArrayExpress [36] and NGDC Gene Expression Nebulas [37] store massive omics data, including many transcriptome data of drugs and other compounds. But researchers need to search, collect and tidy them before their use for drug repositioning. Fortunately, several groups have collected multi-gene expression signatures related to the drugs.

The CREEDS (CRowd Extracted Expression of Differential Signatures) extracted and analyzed the signatures of 875 drugs and 828 diseases from GEO

via a crowdsourcing project, setting in a massive open online course on Coursera [38]. The dataset could be downloaded from the website, https://maayanlab.cloud/CREEDS/.

HERB (http://herb.ac.cn) is a high-throughput experiment database of traditional Chinese medicine, consisting of 7263 herbs and 49,258 ingredients, from 472 high-throughput GEO datasets, providing complementary and valuable drug resources [39].

The CMap version 1 (https://portals.broadinstitute.org/cmap/) consists of Affymetrix-based 6100 gene signatures of 1309 compounds perturbing five different cell lines (such as PC3, MCF7, HL60) with varying doses (mainly 10 μM). Notably, there were 164 distinct perturbagens, including approved drugs and non-drug bioactive compounds, in the original article published in the *Science* journal [8]. Indeed, this dataset stimulates the rapid development of drug repositioning, indicated by the high citations (more than 1800 times). It suggests the great value and success of a large-scale community Connectivity Map project.

The CMap version 2 (https://clue.io/cmap), belonging to NIH's Library of Integrated Network-Based Cellular Signatures (LINCS) program, includes 1.3 million L1000 profiles and 25,200 unique perturbations on variable cell lines [9]. They used L1000 technology due to the cost and argued that about 1000 landmark genes could recover 82% of the information in the full transcriptome based on a comprehensive comparison [9]. As expected, the updated dataset also motivated the continual development of drug repositioning. It should be noted that the consistency between the two versions of CMap is not high with a low recall [40]. It suggests that drug repositioning based on the CMap should consider other evidence to filter false positives in the computational drug repositioning.

In summary, the availability of huge gene signatures of drugs makes the gene signature-based drug repositioning possible as a big data basis. Meanwhile, researchers are still developing new transcriptome technology to make the large-scale transcriptome sequencing of millions of drugs treating different cell lines with various doses possible at a relatively low cost. In addition, with the cost of conventional RNASeq lower, it is also possible to use the RNASeq directly soon.

4. Gene signature databases of various diseases

The gene signature databases of various diseases are a complementary resource to drug repositioning. Importantly, the gene signatures of diseases are robust across different tissues and experiments to some extent (Dudley et al. 2009). As mentioned in the introduction section, it is difficult to apply a high-throughput way to model various diseases in parallel. Researchers have collected some gene signature datasets related to numerous diseases. However, in practice, biologists usually focus on a specific disease, which means that they could obtain the gene signature of the disease by themselves. Once they have the gene signature of the disease, they could directly query the gene signature library of drugs to get the candidate drugs for this disease.

The gene signatures of diseases were mainly collected from the GEO. ADEPTUS (Annotated Disease Expression Profiles Transformed into a Unified Suite) supplied about 14,000 ready-to-use gene signature profiles, annotated with Disease Ontology terms [41]. ADEPTUS built a classic way to form a gene signature of various diseases. The STARGEO (Search Tag Analyze Resource for GEO) project generated annotations of disease-related samples in GEO to identify robust signatures of disease by meta-analysis via a crowdsourcing approach [42]. It covered about 250 types of diseases and could be improved via the webserver. The DrugVsDiseasedata

(Drug versus Disease data) package defined 45 gene signatures of diseases, such as Breast with Small-cell Lung, Cervical, Bladder and Prostate cancer, collected from GEO [43]. Recently, Porcu et al. reported that differentially expressed genes reflect disease-induced rather than disease-causing changes in the transcriptome via the Mendelian randomization method. Thus, identifying the upstream genes, which cause the diseases, would be a promising direction in the transcriptome data of diseases.

Although, there are several gene signature datasets of diseases, more efforts are necessary to enlarge the library of the types of diseases. The disease ontology is a fruitful resource for reference when searching for a disease. With the scale of gene signatures of diseases increasing, there will be more possibility of connecting drugs and diseases as the searching space for the algorithm is expanded.

5. Gene signature-based methods and tools for drug repositioning

Once the gene signatures of drugs and diseases, as well as other useful information (such as the structure of drugs), are ready, we could make a computational drug repositioning analysis. In the end, it is to find a method to connect the drug and disease. This connecting method could be a similarity metric [44], community discovery, matrix factorization and completion, machine learning-based models and so on. A good method should significantly enrich true positive results and deplete false-positive results.

There are several biologist-friendly web servers, convenient to use without the need for programming. The CMap version 1 website is one of the most popular websites in the field of drug repositioning. The CMap version 2 website supplies a more fruitful website. The enrichr website (https://maayanlab.cloud/Enrichr/) also provides the drug repositioning module with the drug and disease libraries (for example, Drug_Perturbations_from_GEO_down gene set) [45, 46]. Biologists could easily use these websites for drug repositioning without programming.

The nonparametric Kolmogorov–Smirnov statistic, formalized in Gene Set Enrichment Analysis (GSEA), was used in the original CMap article, indicating its power [8, 47]. It tests whether the empirical distribution of data (a set of genes) is different from a reference distribution (such as a ranked gene list related to a drug). The nonparametric test simplifies the statistical test process, making it feasible to multi situations.

PAGE (parametric analysis of gene set enrichment) was more sensitive and less-computational than GSEA [48], which could be used to evaluate the similarity between two gene expression signatures. Dr. Insight used the concordantly expressed genes in a frame-breaking statistical model to connect the drug and disease [49]. The eXtreme Sum (XSum) was a similarity scoring algorithm, which was developed by Jie et al. It showed a better performance than the KS statistic based on the area under the curve using 890 drug-indication pairs with 496 compounds and 238 disease signatures [50].

Network-based community discovery could exploit the similarity in gene expression signatures of drugs and identify the similar drugs, which should be clustered together [51]. They also implemented a tool, MANTRA (Mode of Action by NeTwoRk Analysis), which was accessible and biologist-friendly at http://mantra.tigem.it [52]. GPSnet (Genome-wide Positioning Systems network) associated the drug and the gene signature-based disease modules in the protein–protein interactome network [53]. DeMAND (detecting mechanism of action by network dysregulation) developed a regulatory network-based approach to elucidate the MoA using gene expression signatures [54]. Chemical Checker integrated five-level

data of drugs, such as targets, morphology and gene expression signatures, to evaluate the similarity of the drugs via the dimensionality reduction and network embedding algorithm [55].

Cogena, co-expressed gene-set enrichment analysis, focused on the idea of targeting co-expressed genes instead of all the differentially expressed genes for drug repositioning [27]. It empowered simultaneous, gene set knowledgebase-driven drug repositioning analysis and illustrated the mode of action of the predicted drug and disease pairs. Cogena has been widely used in drug repositioning for several diseases, including psoriasis, Coronavirus Disease 2019 (COVID-19) [56, 57], Crohn's disease [58], periodontitis [59].

Machine learning, especially deep learning algorithms, are suitable to the gene expression signatures inherently. The low-rank matrix approximation and randomized algorithms were used in drug repositioning by filling out the unknown connection in the drug-disease pairs [60] The iDrug could reposition drugs via a cross-network embedding and transferring knowledge from the drug target information [61]. DLEPS (deep learning-based efficacy prediction system) used one-dimensional convolutional neural networks to learn the relationship between the structure of drugs and gene expression signatures to predict drug efficacy [62]. Clearly, with the advances of deep learning, especially the graph neural network, lots of innovative algorithms will be continually applied in the drug repositioning field to improve performance.

6. New high-throughput technology for drug repositioning

Researchers try to develop new high-throughput RNASeq technology to improve the precision of transcriptome with the constraint of cost. For example, the microarray was used in the first version of CMap, while the L1000 technology was used in the second version of CMap, that is LINCS with a more than 1000-fold scale-up of the CMap. Via a Luminex bead-based probe hybridization, the L1000 only measured the mRNA abundance of 978 "landmark" genes with the expression of the remaining gene inferred by a machine learning algorithm [9]. This selection largely resulted from lowering the cost of obtaining the transcriptome of a huge scale of drugs and compounds.

RNA-Seq via Next-Generation Sequencing is a relatively new emerging technology in the drug repositioning field. Due to the higher cost, researchers tried to maintain the transcriptome performance when lowering the cost in several ways. For example, a subset of genes with a reduced representation of the transcriptome could be sequenced instead of all the mRNA. The L1000 technology used the most informative genes, named "landmark" genes [9]. Deepak et al. argued that a knowledge-driven subset of 1500 sentinel genes could precisely predict pathway perturbations [63]. RASL-seq (RNA-mediated oligonucleotide annealing, selection, and ligation) only measured hundreds of pre-defined genes in response to a set of 350 chemicals and their mixtures, which provided a cost-effective approach to quantify gene expression signature with a panel of marker genes [64]. TempO-Seq, Templated Oligo assay with Sequencing readout, could determine the whole transcriptome via a targeted way, requiring less sequencing depth [65].

The pooled and low-depth Next-Generation Sequencing is another approach to lower the cost but maintain the performance. PLATE-seq (pooled library amplification for transcriptome expression) introduced the sample-specific barcodes, allowing pooled library construction in 96 wells and low-depth sequencing, which is about 15-fold less expensive than canonical RNA-Seq [66]. DRUG-seq efficiently captured transcriptional changes with low-depth reads by importing cell barcode

and Unique Molecular Index (UMI) in 384- and 1536-well format with fewer steps, compared with PLATE-seq [67]. Notably, DRUG-seq also supplied an open-source R program analysis pipeline at Github recently [68]. BRB-seq (Bulk RNA Barcoding and sequencing) used early-stage multiplexing to produce 3′ cDNA libraries for multi-samples, while with a lower cost [69]. 3'Pool-seq was an optimized cost-efficient method of transcriptome profiling, which was also adapted for a 96-well plate format and ERCC spike-ins. Collectively, researchers have developed multi new transcriptome technologies while lowering the cost of sequencing to implement the RNASeq for large-scale samples, which could be due to the different doses, different treatments, and different periods of treatment.

Other types of gene signatures, such as the proteome and metabolome, could also be used in drug repositioning. Zhao et al. created a systematic map of protein-drug connectivity that compiled 210 clinically relevant protein signatures based on antibody-based proteomics technology in more than 12,000 cell-line samples in response to about 150 drugs [70]. ProTargetMiner was a proteome signature library of 56 molecules in A549 cancer cell lines, forming a valuable tool in drug discovery [71]. Benjamin et al. profiled the proteomes of five lung cancer cell lines (such as A549, Calu6 and Calu1) perturbed by more than 50 drugs based on the label-free proteomics platform [72]. Moreover, an atlas (http://bbmri.researchlumc.nl/atlas/) of 87 drugs and 150 clinically relevant plasma-based metabolite associations will contribute to the drug development as well [73]. Other omics data, besides transcriptome, related to drugs and diseases will promote the drug repositioning flourishing. In summary, new omics technology will precisely quantify the signatures related to drugs and diseases with a low cost, permitting the large-scale omics project, enlarging the searching library for drug repositioning.

7. Drug repositioning examples with reproducible code

Due to the pandemic of COVID-19 and no effective drugs for this disease, drug repositioning is a great way to combat this disease. Several researchers have used cogena for drug repositioning to fight the COVID-19 [56, 74].

We used the metatranscriptome data of the bronchoalveolar lavage fluid from 8 severe COVID-19 patients and 20 healthy controls to obtain the gene expression signature of COVID-19 [75]. The co-expression analysis, pathway analysis and drug repositioning analysis were done using the cogena pipeline [56]. We identified several drugs which were associated with COVID-19 reported before. For example, Saquinavir, a protease inhibitor, is a drug for human immunodeficiency virus infection. This drug was also identified by several docking methods [76]. Dexamethasone is a "major development" in the fight against COVID-19 in the RECOVERY trial [77]. Ribavirin can be used to treat SARS-CoV and MERS-CoV infections [78]. Importantly, it is a recommended drug in the diagnosis and treatment protocol for COVID pneumonia (trial version 5–latest) published by the National Health Commission of the P.R. of China. It was also identified by several docking methods [79]. Furthermore, we identified several other candidate drugs for COVID-19, for example, dinoprost, a smooth muscle activator, and (−)-isoprenaline, a bronchodilator for obstructive lung diseases. These candidate drugs could be tested in vitro and in vivo to validate their possibility.

The whole pipeline of this gene-signature-based drug repositioning for COVID-19 using cogena is accessible at https://github.com/zhilongjia/COVID-19 with data and code, forming a good resource for drug repositioning and reproducible study.

There are also other examples of drug repositioning using cogena with reproducible codes. For instance, the code of the drug repositioning for psoriasis is

available at https://github.com/zhilongjia/psoriasis and the code of drug repositioning for periodontitis is available at https://github.com/zhilongjia/Fn_HGFcell. These examples will enhance our understanding of how drug repositioning works and how to implement drug repositioning.

8. Future perspectives and conclusion

The future of gene signature-based drug repositioning is bright. The booming biotechnology and pharmaceutical industry, especially the emerging sequencing and MS field, supplies an important motivation to sequence more omics data related to drugs and diseases. The artificial intelligence industry, particularly the deep learning algorithm, will also promote the rapid development of the drug repositioning field as it will improve the rate of the true positives and lower the rate of false positives. The omics data of drugs and diseases is like electricity, while the algorithm is like a machine. The seamless combinations of them will produce new opportunities for gene signature-based drug repositioning. More data means a larger searchable space to identify the new relationship between drugs and diseases. Additionally, the signatures-based combination of drugs could also be investigated to deal with intractable diseases. Meanwhile, more evidence from different aspects of the drug-disease pairs will improve the quality of perdition.

In the end, we highlight the key points of this chapter.

1. A systematic introduction to gene signature-based drug repositioning and the core principle of gene signature-based drug repositioning;

2. Gene signature could be achieved based on molecular phenotypes, such as transcriptome and proteome;

3. Basic databases of gene, pathway and drug for drug repositioning;

4. Gene signature databases of drugs and diseases

5. Gene signature-based methods and tools for drug repositioning;

6. New high-throughput technology for drug repositioning;

7. Drug repositioning examples with reproducible code;

8. The future direction of gene signature-based drug repositioning.

Acknowledgements

Conflict of interest

The authors declare no conflict of interest.

Author details

Zhilong Jia[1,2]*, Xinyu Song[1,2], Jinlong Shi[2], Weidong Wang[2] and Kunlun He[2]

1 Center for Artificial Intelligence in Medicine, Medical Innovation Research Division of Chinese PLA General Hospital, Beijing, China

2 Key Laboratory of Biomedical Engineering and Translational Medicine, Ministry of Industry and Information Technology, Medical Innovation Research Division of Chinese PLA General Hospital, Beijing, China

*Address all correspondence to: jiazhilong@plagh.org

IntechOpen

References

[1] Raju TN. The Nobel chronicles. 1988: James Whyte Black, (b 1924), Gertrude Elion (1918-99), and George H Hitchings (1905-98). Lancet. 2000;**355**(9208):1022

[2] Pillaiyar T, Meenakshisundaram S, Manickam M, Sankaranarayanan M. A medicinal chemistry perspective of drug repositioning: Recent advances and challenges in drug discovery. European Journal of Medicinal Chemistry. 2020;**1**(195):112275

[3] Kettle JG, Wilson DM. Standing on the shoulders of giants: a retrospective analysis of kinase drug discovery at AstraZeneca. Drug Discovery Today. 2016;**21**(10):1596-1608

[4] Hughes TR, Marton MJ, Jones AR, Roberts CJ, Stoughton R, Armour CD, et al. Functional discovery via a compendium of expression profiles. Cell. 2000;**102**(1):109-126

[5] Arakelyan A, Nersisyan L, Nikoghosyan M, Hakobyan S, Simonyan A, Hopp L, et al. Transcriptome-guided drug repositioning. Pharmaceutics. 2019;**11**(12):677. DOI: 10.3390/pharmaceutics11120677

[6] Schuster A, Erasimus H, Fritah S, Nazarov PV, van Dyck E, Niclou SP, et al. RNAi/CRISPR screens: From a pool to a valid hit. Trends in Biotechnology. 2019;**37**(1):38-55

[7] Zarei A, Razban V, Hosseini SE, Tabei SMB. Creating cell and animal models of human disease by genome editing using CRISPR/Cas9. The Journal of Gene Medicine. 2019;**21**(4):e3082

[8] 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**:1929-1935

[9] Subramanian A, Narayan R, Corsello SM, Peck DD, Natoli TE, Lu X, et al. A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. Cell. 2017;**171**(6): 1437-52.e17

[10] 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 Systems. 2018;**6**(1):13-24, 24

[11] Aronson JK, Green AR. Me-too pharmaceutical products: History, definitions, examples, and relevance to drug shortages and essential medicines lists. British Journal of Clinical Pharmacology. 2020;**86**(11):2114-2122

[12] 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. Science Translational Medicine. 2017;**9**(383):eaag1166. DOI: 10.1126/scitranslmed.aag1166

[13] Santos R, Ursu O, Gaulton A, Bento AP, Donadi RS, Bologa CG, et al. A comprehensive map of molecular drug targets. Nature Reviews. Drug Discovery. 2017;**16**(1):19-34

[14] Tanoli Z, Seemab U, Scherer A, Wennerberg K, Tang J, Vähä-Koskela M. Exploration of databases and methods supporting drug repurposing: a comprehensive survey. Briefings in Bioinformatics. 2021;**22**(2):1656-1678

[15] Fishilevich S, Nudel R, Rappaport N, Hadar R, Plaschkes I, Iny Stein T, et al. GeneHancer: Genome-wide integration of enhancers and target genes in GeneCards. Database. 2017;**2017**:bax028. DOI: 10.1093/database/bax028

[16] Freshour SL, Kiwala S, Cotto KC, Coffman AC, McMichael JF, Song JJ,

et al. Integration of the drug-gene interaction database (DGIdb 4.0) with open crowdsource efforts. Nucleic Acids Research. 2021;**49**(D1):D1144-D1151

[17] Ochoa D, Hercules A, Carmona M, Suveges D, Gonzalez-Uriarte A, Malangone C, et al. Open targets platform: Supporting systematic drug-target identification and prioritisation. Nucleic Acids Research. 2021;**49**(D1):D1302-D1310

[18] Ghoussaini M, Mountjoy E, Carmona M, Peat G, Schmidt EM, Hercules A, et al. Open targets genetics: Systematic identification of trait-associated genes using large-scale genetics and functional genomics. Nucleic Acids Research. 2021;**49**(D1): D1311-D1320

[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. Nature Genetics. 2015;**47**(8):856-860

[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. PLoS Genetics. 2019;**15**(12):e1008489

[21] Smith I, Greenside PG, Natoli T, Lahr DL, Wadden D, Tirosh I, et al. Evaluation of RNAi and CRISPR technologies by large-scale gene expression profiling in the Connectivity Map. PLoS Biology. 2017;**15**(11):e2003213

[22] Housden BE, Valvezan AJ, Kelley C, Sopko R, Hu Y, Roesel C, et al. Identification of potential drug targets for tuberous sclerosis complex by synthetic screens combining CRISPR-based knockouts with RNAi. Science Signaling. 2015;**8**(393):rs9

[23] Behan FM, Iorio F, Picco G, Gonçalves E, Beaver CM, Migliardi G,

et al. Prioritization of cancer therapeutic targets using CRISPR-Cas9 screens. Nature. 2019;**568**(7753):511-516

[24] Szlachta K, Kuscu C, Tufan T, Adair SJ, Shang S, Michaels AD, et al. CRISPR knockout screening identifies combinatorial drug targets in pancreatic cancer and models cellular drug response. Nature Communications. 2018;**9**(1):4275

[25] Corsello SM, Bittker JA, Liu Z, Gould J, McCarren P, Hirschman JE, et al. The drug repurposing hub: A next-generation drug library and information resource. Nature Medicine. 2017;**23**(4): 405-408

[26] Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The molecular signatures database (MSigDB) hallmark gene set collection. Cell Systems. 2015;**1**(6):417-425

[27] Jia Z, Liu Y, Guan N, Bo X, Luo Z, Barnes MR. Cogena, a novel tool for co-expressed gene-set enrichment analysis, applied to drug repositioning and drug mode of action discovery. BMC Genomics. 2016;**27**(17):414

[28] 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**(9):1498-1505

[29] Hernández-Lemus E, Martínez-García M. Pathway-based drug-repurposing schemes in Cancer: The role of translational bioinformatics. Frontiers in Oncology. 2020;**10**:605680

[30] Li J, Lu Z. Pathway-based drug repositioning using causal inference. BMC Bioinformatics. 2013;**14**(Suppl. 16):S3

[31] Jahchan NS, Dudley JT, Mazur PK, Flores N, Yang D, Palmerton A, et al. A drug repositioning approach identifies

tricyclic antidepressants as inhibitors of small cell lung cancer and other neuroendocrine tumors. Cancer Discovery. 2013;**3**(12):1364-1377

[32] Brown AS, Patel CJ. A standard database for drug repositioning. Sci Data. 2017;**14**(4):170029

[33] Zhao C, Dai X, Li Y, Guo Q, Zhang J, Zhang X, et al. EK-DRD: A comprehensive database for drug repositioning inspired by experimental knowledge. Journal of Chemical Information and Modeling. 2019;**59**(9): 3619-3624

[34] Wang Z, Lachmann A, Ma'ayan A. Mining data and metadata from the gene expression omnibus. Biophysical Reviews. 2019;**11**(1):103-110

[35] Clough E, Barrett T. The gene expression omnibus database. Methods in Molecular Biology. 2016;**1418**:93-110

[36] Athar A, Füllgrabe A, George N, Iqbal H, Huerta L, Ali A, et al. ArrayExpress update - from bulk to single-cell expression data. Nucleic Acids Research. 2019;**47**(D1):D711-D715

[37] CNCB-NGDC Members and Partners. Database resources of the national genomics data center, China national center for bioinformation in 2021. Nucleic Acids Research. 2021;**49**(D1):D18-D28

[38] Wang Z, Monteiro CD, Jagodnik KM, Fernandez NF, Gundersen GW, Rouillard AD, et al. Extraction and analysis of signatures from the gene expression omnibus by the crowd. Nature Communications. 2016;**26**(7):12846

[39] Fang S, Dong L, Liu L, Guo J, Zhao L, Zhang J, et al. HERB: a high-throughput experiment- and reference-guided database of traditional Chinese medicine. Nucleic Acids Research. 2021;**49**(D1):D1197-D1206

[40] Lim N, Pavlidis P. Evaluation of connectivity map shows limited reproducibility in drug repositioning. Scientific Reports. 2021;**11**(1):17624

[41] Amar D, Hait T, Izraeli S, Shamir R. Integrated analysis of numerous heterogeneous gene expression profiles for detecting robust disease-specific biomarkers and proposing drug targets. Nucleic Acids Research. 2015;**43**(16): 7779-7789

[42] Hadley D, Pan J, El-Sayed O, Aljabban J, Aljabban I, Azad TD, et al. Precision annotation of digital samples in NCBI's gene expression omnibus. Scientific Data. 2017;**19**(4):170125

[43] Pacini C, Iorio F, Gonçalves E, Iskar M, Klabunde T, Bork P, et al. DvD: An R/Cytoscape pipeline for drug repurposing using public repositories of gene expression data. Bioinformatics. 2013;**29**(1):132-134

[44] Struckmann S, Ernst M, Fischer S, Mah N, Fuellen G, Möller S. Scoring functions for drug-effect similarity. Briefings in Bioinformatics. 2021;**22**(3):bbaa072. DOI: 10.1093/ bib/bbaa072

[45] Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Research. 2016;**44**(W1):W90-W97

[46] Xie Z, Bailey A, Kuleshov MV, Clarke DJB, Evangelista JE, Jenkins SL, et al. Gene set knowledge discovery with Enrichr. Current Protocols. 2021;**1**(3):e90

[47] Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences of the

United States of America. 2005;**102**: 15545-15550

[48] Kim S-Y, Volsky DJ. PAGE: parametric analysis of gene set enrichment. BMC Bioinformatics. 2005;**8**(6):144

[49] Chan J, Wang X, Turner JA, Baldwin NE, Gu J. Breaking the paradigm: Dr Insight empowers signature-free, enhanced drug repurposing. Bioinformatics. 2019;**35**(16):2818-2826

[50] Cheng J, Yang L, Kumar V, Agarwal P. Systematic evaluation of connectivity map for disease indications. Genome Medicine. 2014;**6**:540

[51] Iorio F, Bosotti R, Scacheri E, Belcastro V, Mithbaokar P, Ferriero R, et al. Discovery of drug mode of action and drug repositioning from transcriptional responses. Proceedings of the National Academy of Sciences of the United States of America. 2010;**107**:14621-14626

[52] Carrella D, Napolitano F, Rispoli R, Miglietta M, Carissimo A, Cutillo L, et al. Mantra 2.0: An online collaborative resource for drug mode of action and repurposing by network analysis. Bioinformatics. 2014;**30**(12):1787-1788

[53] Cheng F, Lu W, Liu C, Fang J, Hou Y, Handy DE, et al. A genome-wide positioning systems network algorithm for in silico drug repurposing. Nature Communications. 2019;**10**(1):3476

[54] Woo JH, Shimoni Y, Yang WS, Subramaniam P, Iyer A, Nicoletti P, et al. Elucidating compound mechanism of action by network perturbation analysis. Cell. 2015;**162**(2):441-451

[55] Duran-Frigola M, Pauls E, Guitart-Pla O, Bertoni M, Alcalde V, Amat D, et al. Extending the small-molecule similarity principle to all levels of biology with the Chemical Checker. Nature Biotechnology. 2020;**38**(9):1087-1096

[56] Jia Z, Song X, Shi J, Wang W, He K. Transcriptome-based drug repositioning for coronavirus disease 2019 (COVID-19). Pathogens and Disease. 2020;**78**(4): ftaa036. DOI: 10.1093/femspd/ftaa036

[57] Li F, Michelson AP, Foraker R, Zhan M, Payne PRO. Computational analysis to repurpose drugs for COVID-19 based on transcriptional response of host cells to SARS-CoV-2. BMC Medical Informatics and Decision Making. 2021;**21**(1):15

[58] Kwak MS, Lee HH, Cha JM, Shin HP, Jeon JW, Yoon JY. Novel candidate drugs in anti-tumor necrosis factor refractory Crohn's diseases: In silico study for drug repositioning. Scientific Reports. 2020;**10**(1):10708

[59] Kang W, Jia Z, Tang D, Zhao X, Shi J, Jia Q, et al. Time-course transcriptome analysis for drug repositioning in fusobacterium nucleatum-infected human gingival fibroblasts. Frontiers in Cell and Development Biology. 2019;**20**(7):204

[60] Luo H, Li M, Wang S, Liu Q, Li Y, Wang J. Computational drug repositioning using low-rank matrix approximation and randomized algorithms. Bioinformatics. 2018;**34**(11):1904-1912

[61] Chen H, Cheng F, Li J. iDrug: Integration of drug repositioning and drug-target prediction via cross-network embedding. PLoS Computational Biology. 2020;**16**(7):e1008040

[62] Zhu J, Wang J, Wang X, Gao M, Guo B, Gao M, et al. Prediction of drug efficacy from transcriptional profiles with deep learning. Nature Biotechnology. 2021;**39**(11):1444-1452. DOI: 10.1038/s41587-021-00946-z

[63] Mav D, Shah RR, Howard BE, Auerbach SS, Bushel PR, Collins JB, et al. A hybrid gene selection approach to create the S1500+ targeted gene sets for use in high-throughput

transcriptomics. PLoS One. 2018;**13**(2):e0191105

[64] Simon JM, Paranjape SR, Wolter JM, Salazar G, Zylka MJ. High-throughput screening and classification of chemicals and their effects on neuronal gene expression using RASL-seq. Scientific Reports. 2019;**9**(1):4529

[65] Yeakley JM, Shepard PJ, Goyena DE, VanSteenhouse HC, McComb JD, Seligmann BE. A trichostatin a expression signature identified by TempO-Seq targeted whole transcriptome profiling. PLoS One. 2017;**12**(5):e0178302

[66] Bush EC, Ray F, Alvarez MJ, Realubit R, Li H, Karan C, et al. PLATE-Seq for genome-wide regulatory network analysis of high-throughput screens. Nature Communications. 2017;**8**(1):105

[67] Ye C, Ho DJ, Neri M, Yang C, Kulkarni T, Randhawa R, et al. DRUG-seq for miniaturized high-throughput transcriptome profiling in drug discovery. Nature Communications. 2018;**9**(1):4307

[68] Li J, Ho DJ, Henault M, Yang C, Neri M, Ge R, et al. DRUG-seq provides unbiased biological activity readouts for drug discovery. bioRxiv. 2021:2021.06.07.447456 Available from: https://www.biorxiv.org/content/10.1101/2021.06.07.447456v1. [cited 2021 Sep 6]

[69] Alpern D, Gardeux V, Russeil J, Mangeat B, Meireles-Filho ACA, Breysse R, et al. BRB-seq: ultra-affordable high-throughput transcriptomics enabled by bulk RNA barcoding and sequencing. Genome Biology. 2019;**20**(1):71

[70] Zhao W, Li J, Chen M-JM, Luo Y, Ju Z, Nesser NK, et al. Large-scale characterization of drug responses of clinically relevant proteins in cancer cell lines. Cancer Cell. 2020;**38**(6):829-43.e4

[71] Saei AA, Beusch CM, Chernobrovkin A, Sabatier P, Zhang B, Tokat ÜG, et al. ProTargetMiner as a proteome signature library of anticancer molecules for functional discovery. Nature Communications. 2019;**10**(1):5715

[72] Ruprecht B, Di Bernardo J, Wang Z, Mo X, Ursu O, Christopher M, et al. Publisher correction: A mass spectrometry-based proteome map of drug action in lung cancer cell lines. Nature Chemical Biology. 2020;**16**(10):1149

[73] Liu J, Lahousse L, Nivard MG, Bot M, Chen L, van Klinken JB, et al. Integration of epidemiologic, pharmacologic, genetic and gut microbiome data in a drug-metabolite atlas. Nature Medicine. 2020;**26**(1):110-117

[74] Krishnamoorthy P, Raj AS, Roy S, Kumar NS, Kumar H. Comparative transcriptome analysis of SARS-CoV, MERS-CoV, and SARS-CoV-2 to identify potential pathways for drug repurposing. Computers in Biology and Medicine. 2021;**128**:104123

[75] Zhou Z, Ren L, Zhang L, Zhong J, Xiao Y, Jia Z, et al. Heightened innate immune responses in the respiratory tract of COVID-19 patients. Cell Host & Microbe. 2020;**27**(6):883-90.e2

[76] Bello M, Martínez-Muñoz A, Balbuena-Rebolledo I. Identification of saquinavir as a potent inhibitor of dimeric SARS-CoV2 main protease through MM/GBSA. Journal of Molecular Modeling. 2020;**26**(12):340

[77] RECOVERY Collaborative Group, Horby P, Lim WS, Emberson JR, Mafham M, Bell JL, et al. Dexamethasone in Hospitalized Patients with Covid-19. The New England Journal of Medicine. 2021;**384**(8):693-704

[78] Wang Y, Li W, Jiang Z, Xi X, Zhu Y. Assessment of the efficacy and safety of Ribavirin in treatment of coronavirus-related pneumonia (SARS, MERS and COVID-19): A protocol for systematic review and meta-analysis. Medicine. 2020;**99**(38):e22379

[79] Elfiky AA. Ribavirin, remdesivir, sofosbuvir, galidesivir, and tenofovir against SARS-CoV-2 RNA dependent RNA polymerase (RdRp): A molecular docking study. Life Sciences. 2020;**15**(253):117592