



A Study on Methodologies of Drug Repositioning Using Biomedical Big Data: A Focus on Diabetes Mellitus

Suehyun Lee^{1,2,*}, Seongwoo Jeon^{2,*}, Hun-Sung Kim^{3,4}

¹Department of Biomedical Informatics, Konyang University College of Medicine; ²Health Care Data Science Center, Konyang University Hospital, Daejeon; ³Department of Medical Informatics, College of Medicine, The Catholic University of Korea; ⁴Division of Endocrinology and Metabolism, Department of Internal Medicine, Seoul St. Mary's Hospital, College of Medicine, The Catholic University of Korea, Seoul, Korea

Drug repositioning is a strategy for identifying new applications of an existing drug that has been previously proven to be safe. Based on several examples of drug repositioning, we aimed to determine the methodologies and relevant steps associated with drug repositioning that should be pursued in the future. Reports on drug repositioning, retrieved from PubMed from January 2011 to December 2020, were classified based on an analysis of the methodology and reviewed by experts. Among various drug repositioning methods, the network-based approach was the most common (38.0%, 186/490 cases), followed by machine learning/deep learning-based (34.3%, 168/490 cases), text mining-based (7.1%, 35/490 cases), semantic-based (5.3%, 26/490 cases), and others (15.3%, 75/490 cases). Although drug repositioning offers several advantages, its implementation is curtailed by the need for prior, conclusive clinical proof. This approach requires the construction of various databases, and a deep understanding of the process underlying repositioning is quintessential. An in-depth understanding of drug repositioning could reduce the time, cost, and risks inherent to early drug development, providing reliable scientific evidence. Furthermore, regarding patient safety, drug repurposing might allow the discovery of new relationships between drugs and diseases.

Keywords: Drug repositioning; Semantics; Machine learning; Real-world data; Data science

INTRODUCTION

The coronavirus disease 19 (COVID-19) pandemic is bringing about socio-economic changes, inevitably affecting the overall healthcare system [1]. Effective strategies to curtail the spread of the virus and prevent virus-related increases in morbidity and mortality are urgently needed [2]. In an urgent scenario, drug and vaccine development processes are rapidly evolving and being updated [3], enabling the development of effective treatment

methods that are urgently needed. Although developing powerful therapeutic agents for virus control is paramount, in reality, the extended time, high cost, and low success rates associated with new drug development represent major obstacles [4,5]. Recently, several studies have reported that specific drugs previously approved for other purposes might be repurposed to treat COVID-19 [6]. Therefore, the importance of developing a treatment for COVID-19 through drug repositioning (or repurposing) is emphasized, and interest in drug repositioning is increasing.

Received: 11 January 2022, **Revised:** 4 March 2022, **Accepted:** 21 March 2022

Corresponding author: Hun-Sung Kim

Department of Medical Informatics, College of Medicine, The Catholic University of Korea, 222 Banpo-daero, Seocho-gu, Seoul 06591, Korea

Tel: +82-2-2258-8262, **Fax:** +82-2-2258-8297, **E-mail:** 01cadiz@hanmail.net

*These authors contributed equally to this work.

Copyright © 2022 Korean Endocrine Society

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Drug repositioning is the process of identifying new therapeutic applications for existing drugs and novel treatment methods for untreatable diseases [7]. It is a new drug development method allowing the use of drugs that have already been marketed or proven safe in clinical trials but have not been approved for efficacy reasons [7]. For instance, although sildenafil has been developed to treat angina pectoris through its vasodilator effect, it gained popularity as an erectile dysfunction drug [8]. Finasteride, initially used to treat benign prostatic hyperplasia, is currently used as a hair loss remedy after a dose adjustment [9]. Furthermore, for diabetes mellitus (DM) patients, sodium glucose cotransporter-2 inhibitor (SGLT2i) has been shown to decrease serum triglyceride levels and increase high-density lipoprotein cholesterol levels [10]; however, it cannot be used in hypercholesterolemia patients as it is only prescribed for diabetic patients. Although SGLT2i lowers blood pressure and has a diuretic effect [11], it cannot be used as an antihypertensive medication nor a first-line treatment for heart failure (HF). Glucagon-like peptide-1 receptor agonists (GLP1-RA) have been approved for diabetic patients while also being used in obese patients [12]. In the case of liraglutide, marketed under the name Victoza, it is used and approved for diabetic patients, whereas Saxenda is used and approved for obese individuals (Novo Nordisk, Bagsvaerd, Denmark). Pregabalin, a treatment for diabetic neuropathy, was originally used to treat epilepsy [13]. Overall, drug repositioning can reduce the risk of failures inherent to the early stages of drug development as it relies on drugs that have already been tested.

Early drug repositioning revealed the possibility of redirecting drugs based on serendipity; however, it recently opened up the opportunity to rationally reuse existing drugs [14]. Considering the COVID-19 outbreak as an opportunity, it is necessary to understand drug repositioning and its potential to rapidly unravel new drug uses. However, studies providing detailed explanations of the concept, methodology, and application of drug repositioning are lacking. In this study, we attempted to analyze various drug repositioning methods by retrieving related research reports, discussing the developmental potential of these methods. We also analyzed recent drug repositioning cases that have been implemented, with a particular focus on DM.

LITERATURE SEARCH OF DRUG REPOSITIONING-RELATED REPORTS

We retrieved articles on drug repositioning and analytical methodologies in the National Library of Medicine (NLM) PubMed

database using the “Drug repositioning” [MeSH] OR “Drug Repurposing” [All Fields] OR “Drug Repurposing” [All Fields] keywords (Fig. 1). Studies published from January 1, 2011, to December 31, 2020, were extracted. Of the 4,892 studies, 3,109 were available for download, of which 2,421 were included in MEDLINE. Following review, 494 studies classified as review or systemic review papers were excluded, and the authors conducted a manual review of the remaining 1,927 studies. The exclusion of studies unrelated to the topic yielded a collection of 490 studies. In this process, papers including terms such as ‘machine-learning,’ ‘deep-learning,’ ‘network-based’ were included, encompassing a significant number of methodological papers based on computational approaches. The following papers were excluded: (1) papers containing inappropriate keywords such as “Chinese” and “Herbal” in the title and abstract; (2) reports implementing methodologies such as “active learning” or “case study,” of which a few appeared among computational approaches. Two researchers reviewed abstracts and relevant topics of each report to evaluate the appropriateness of the subject. Of 139 papers for which the two experts had diverging opinions, 91 papers were maintained after discussion, and the remaining 48 were deleted due to a lack of agreement.

After classifying the studies based on the analysis of the drug repositioning methodology, 186 papers were categorized as network-based approaches, 35 were based on text mining, 26 on semantics, and 168 on machine learning/deep learning. In addition to the network-based, machine learning, and deep learning-based approaches, a few studies combined text mining and semantic approaches. Furthermore, it was confirmed that high-throughput screening, virtual screening, and clinical trials are being used in drug repositioning research.

APPROACHES TO DRUG REPOSITIONING ANALYSIS

The main purpose of drug repositioning is to detect new relationships between drugs and diseases [15]. General studies screen for pharmacological actions against new targets and investigate the general properties of drug compounds, such as chemical structure and side effects. In addition, drug repositioning focuses on revealing the similarity between drug effects and modes of action by discovering the relationship between drugs and diseases [3,16,17]. Various approaches have been developed to analyze drug repositioning. Although text-mining and semantic-based approaches are both categorized as data-mining strategies [3], we divided these categories as both fields have

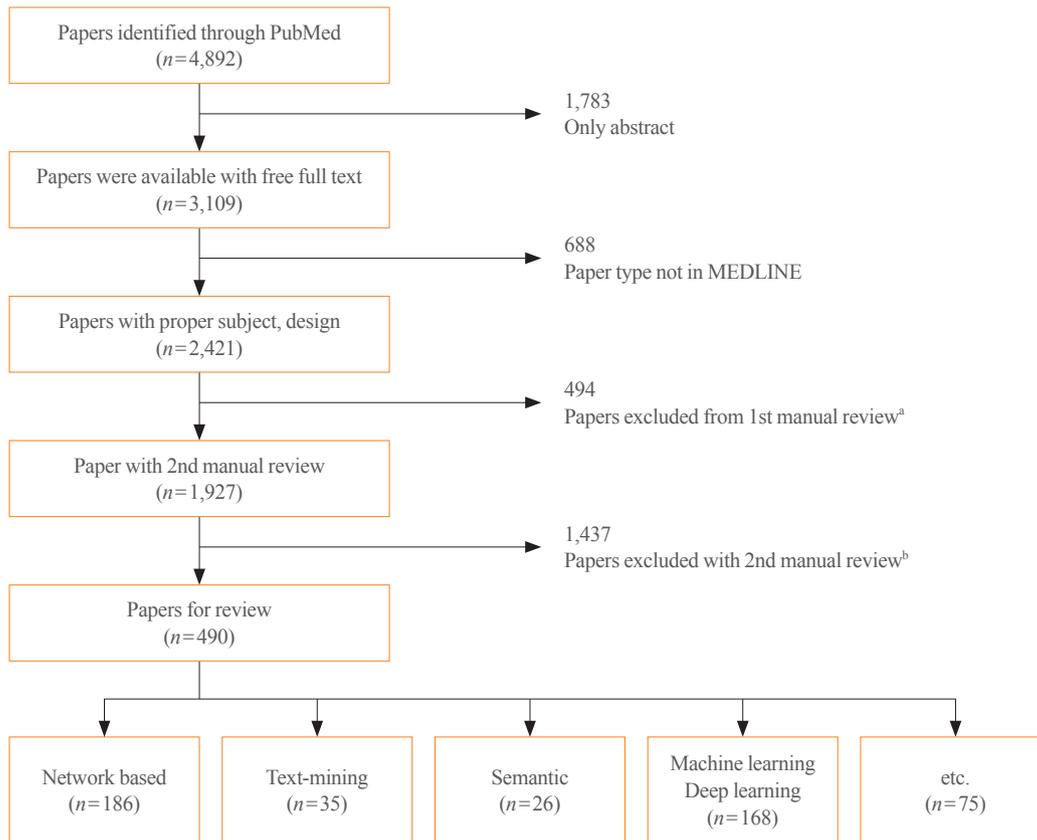


Fig. 1. Flowchart depicting the study selection process. ^aThe 1st manual review by article type: Excluding review and systemic review articles; ^bThe 2nd manual review by article topic in the abstract.

recently gained independent importance. Besides, deep learning is a machine learning-based approach that recently grew along with the increased availability of datasets [15]. Therefore, in this study, we classified network-based, text-mining-based, semantic-based, and machine learning/deep learning-based approaches with reference to previous cases [3,15] and presented the corresponding methodologies (Table 1).

Network-based approaches

In the past, the focus has been on exploring the shared properties of drug compounds, such as their chemical structures and side effects. However, recently, to explore the relationship between drugs and diseases, pharmacological, genetic, and clinical data are first considered to explore the relationship between drug compounds [18]. This is based on the hypothesis that similar drugs are usually associated with similar diseases and *vice versa*. Algorithms are typically implemented to detect structural or network similarities between distinct networks, such as drugs, diseases, proteins, and genes (Supplemental Fig. S1A) [19]. Wu et al. [20] constructed a heterogeneous network of dis-

ease-gene and drug-target relationships and weighted diseases and drugs using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database [21,22]. Furthermore, all possible drug-disease pairs (drug re-creation candidates) were assembled to validate predictions. For instance, hydroxychloroquine has been proposed to exert potentially beneficial effects in coronary artery disease due to evidence from a protein-protein interaction network and large-scale patient data [23].

Looking at the research on DM using these network-based approaches, cyclooxygenase-2 (COX2) represents a potential repositioning candidate for type 1 DM treatment [24]. In studies based on electronic medical records (EMR), calcium channel blockers were safely prescribed during pregnancy to effectively treat and prevent gestational DM [25]. Lastly, metformin has shown promise as a therapeutic agent for neurodegenerative diseases [26].

Text mining-based approaches

Text mining is the process of acquiring meaningful knowledge from unstructured documents. Keywords for a specific drug and

Table 1. Computational Drug Repositioning Approaches

Network-based approaches [20,23]
Assume that two drugs with structurally similar components perform similar roles
Integrate information regarding drugs and diseases from large-scale biological datasets
Use gene, protein, molecular, phenotypic, biological, or biomedical interactions
Text mining-based approaches [29]
Estimate information and knowledge from the literature
Identify drug functions, drug metabolic pathways, and diseases using specific keywords
Effective in predicting associations between drugs and diseases
Semantics-based approaches [31]
Add the existing ontology of network-based prior information to the existing semantic information extracted from a large-scale medical database
Combine multiple sources for predictive indications and therapeutic potential of existing drugs
Improve the accuracy of predicting biological entity relationships
Machine learning/deep learning-based approaches [34]
Identify new indications using computational approaches for extracting features from biological data
Train a model based of disease and drug characteristics obtained from various biological and biomedical datasets
Predict new uses based on the trained model

its targets, pathways, associated disease, and function are used. Based on such keywords, it is possible to reveal the overall research direction regarding the target keywords and obtain new knowledge regarding other keywords associated with the target keyword. The text-mining-based approach extracts and preprocesses text data of interest from literature by recognizing entity terms. A knowledge graph was constructed by identifying the relationships between recognized entity terms [27]. In drug repositioning, the text mining approach recognizes the information and properties of the linguistic context of each biological concept to predict associations and detect new indications (Supplemental Fig. S1B) [27]. Such an approach is effective in predicting the association between drugs and diseases as well as enabling the detection of new indications and side effects of existing drugs [18]. With the development of natural language processing technology, an increasing number of text mining tools are being developed and used to aid drug development [28]. Kostoff et al. [29] derived potential treatments by prioritizing them according to the prevalence and clinical relevance of inflammatory bowel disease in the literature using a text mining approach.

Diflunisal, nabumetone, niflumic acid, and valdecoxib targeting COX2 have been repositioned as therapeutic agents for type 1 DM. In addition, phenoxybenzamine and idazoxan, targeting alpha 2A adrenergic receptor (ADRA2A), have been reported to exhibit therapeutic effects in type 2 DM [30].

Semantics-based approaches

The semantic-based approach is widely used in information and image retrieval due to its effectiveness in predicting drug-disease associations when combined with text mining approaches [18]. A semantic network is built by adding prior information based on the existing ontology network to the information extracted from a large-scale medical database. In this network, mining algorithms are designed to predict new relationships. Although the semantic-based approach has improved the accuracy of predicting biological entity relationships by maximizing the semantic information contained in the vast literature, it is still difficult to integrate and construct various data sources [15]. Zhang et al. [31] discovered potential drugs for Parkinson's disease by mining the semantic relationships between genes and molecular sequences, chemicals, and drugs. This method can improve the detection of potential relationships between drugs and disorders such as Alzheimer's disease and cancer. For instance, bromocriptine, with neurotransmitter action, is known to improve blood sugar control [32].

Machine learning/deep learning-based approaches

The literature search revealed that a drug repositioning study based on machine learning was grafted in various fields. A drug is expressed as a vector derived from characteristics such as drug fingerprints, chemical structures, and side effects. Diseases are trained according to various characteristics of drugs and disorders using a machine learning model, and the relationship be-

tween them can be predicted using the trained model [18]. In particular, Napolitano et al. [33] suggested three approaches to predict drug repositioning based on machine learning algorithms and showed an accuracy of 78%. Distance of drugs according to the degree of chemical structural similarity, integration of multiple layers of information regarding the proximity of a target within the protein-protein interaction network, and the correlation of gene expression patterns following treatment are some of the characteristics used by the algorithms. Menden et al. [34] developed a machine learning model to predict the response of cancer cells to drug treatment using a combination of cell line genomics and drug chemical structure. This could be useful for personalized medicine by correlating cell line genomics with drug hypersensitivity.

An alpha 1-adrenoceptor antagonist, known to treat benign prostate hyperplasia, was reportedly beneficial for blood sugar control [35]. Furthermore, dipeptidyl peptidase-4 inhibitor (DPP-4i) showed promising results in the prognosis of colorectal cancer [36].

CONSIDERATIONS FOR THE INTRODUCTION OF DRUG REPOSITIONING

This study examined the concept and clinical application of efficient drug repositioning. However, various challenges need to be overcome before achieving clinical use. Aspects to be considered when introducing drug repositioning using these methodologies and data are as follows.

Excluding optimism about clinical applicability

Various studies have reported the effects of drug repositioning; however, it remains difficult to be optimistic about the effect of this method in clinical settings. In one meta-analysis, chloroquine, usually used for systemic lupus erythematosus and rheumatoid arthritis, suppressed the maximum effective virus concentration in a laboratory study; however, high-quality evidence was not confirmed in patients infected with the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) [37]. Despite particular cases that benefited from the use of chloroquine, it is difficult to be confident regarding the efficacy of a drug before similar effects are observed in large-scale studies [38,39]. Taking DM as an example, it appears that various antihyperglycemic therapies have other than hypoglycemic effects (Table 2) [10-12,26,36,40-51]. In addition, other types of drugs are likely to help lower blood sugar levels in diabetic patients (Table 3)

[13,24,25,30,32,35,52]. In one study, caution was required, as there is no guarantee that the preclinical findings on the neuroprotective effects of DPP4i in acute stroke are identical in clinical practice [53].

Approach utilization based on diverse data sources

In recent years, an increasing number of researchers have attempted to combine computational and experimental approaches to identify new drug uses. By combining various approaches, new indications for drugs have been explored, and the results have been verified through biological experiments and clinical trials. This is supported by the vast amount of data generated by advances in genomics and proteomics. In drug repositioning, it is difficult to evaluate the performance of this method using a one-dimensional approach due to low reliability. Therefore, it is necessary to increase the reliability of the repurposed drugs by approaching and verifying them from multiple perspectives. Furthermore, it is important to efficiently interpret and use large datasets [54]. By combining different types of data sources to discover the potential value of drugs, the discovery time can be shortened, and the efficiency, as well as reliability, can be improved. In particular, large-scale clinical data stored in EMRs and personal health records represent a promising and inexhaustible data source for discovering new drug indications [55].

Intellectual property issues

Intellectual property (IP) is an important problem associated with introducing repositioned drugs. Since IP protection is limited for non-patented drugs, the economic advantages associated with drug repositioning are counteracted by developing new indications for unprotected drugs. Regardless of the brand of the drug, if prescribed in clinical practice, the patient use can be expanded, and tangible benefits can be obtained. However, without a patent, drugs that prove novel indications are often protected by regulatory agencies, so their use might be limited. Furthermore, the researcher or research institution does not hold the license for that specific drug repositioning. For example, new associations between drug and disease targets discovered by researchers have been identified in publications or online databases; however, these are not protected by IP rights. In addition, some drug repositioning projects may be forcibly stopped; thus, wasting time and resources [56]. Because the existing drug commercial model is not discontinued and investment problems can overlap, the development of a new commercial model is sometimes necessary [57]. In such situations, academia, corporations, and regulatory agencies must cooperate to potentially benefit patients.

Table 2. Candidate Diabetes Drugs for Diseases Other than Diabetes Mellitus

Drug	Other candidate for indication	Reference
Sulfonylurea	Treatment of Alzheimer's disease	[40]
Metformin	Cancer treatment as it reduces cancer incidence and mortality	[41]
	Therapeutic agent for neurodegenerative diseases	[26]
	Cerebroprotective potential for ischemic stroke	[42]
	Suitable candidate in aging-related CNS disorders	
	Improves depressive symptoms	[43]
Sulfonylurea+Metformin	Decrease affective disorder	[44]
DPP4i	Good prognosis of colorectal cancer	[36]
	Antiviral properties, suggesting the broad-spectrum antiviral agents	[45]
	Potential agents to treat SARS-CoV-2 infection	[45]
SGLT2i	Protective role in the occurrence of AF	[46]
	Decrease triglyceride and increase HDL-C	[10]
	Lowers blood pressure and exhibits a diuretic effect	[11]
DPP4i+SGLT2i	Neuroprotection in the obese-insulin resistance	[47]
Thiazolidinedione	Improves depressive symptoms	[43]
GLP1-RA	Neuroprotection, substance against neurodegeneration	[48]
	Prevent heart failure was obtained	[49]
	Treatment options in Parkinson's disease	[50]
	Treatment of metabolic syndrome	[51]
	Weight loss	[12]

CNS, central nervous system; DPP4i, dipeptidyl peptidase-4 inhibitor; SARS-CoV-2, severe acute respiratory syndrome coronavirus-2; SGLT2i, sodium glucose cotransporter-2 inhibitor; AF, atrial fibrillation; HDL-C, high-density lipoprotein cholesterol; GLP1-RA, glucagon-like peptide-1 receptor agonist.

Table 3. Potential Drugs Repositioned as Diabetes Medication

Drug	Original indication	Potential as an anti-diabetic drug	Reference
Alpha 1 (α 1)-adrenoceptor antagonist	Benign prostate hyperplasia	Increases the success rate of blood sugar control	[30,35]
Bromocriptine	Parkinson's disease	Treatment of type 2 diabetes mellitus	[32]
Calcium channel blockers	Anti-hypertensive drug	Effective in treating or preventing GDM	[25]
Colesevelam	Hyperlipidemia	Management of prediabetes and type 2 diabetes mellitus	[52]
Cyclooxygenase-2 inhibitor	Non-steroidal anti-inflammatory drug	Can be used as a treatment for type 1 diabetes mellitus	[24,30]
Pregabalin	Epilepsy	Treatment for diabetic neuropathy	[13]

GDM, gestational diabetes mellitus.

DATASET FOR DRUG REPOSITIONING

The amount of publicly available biomedical, pharmaceutical, and genomic data is increasing exponentially (Table 4) [21,22, 57-80]. Using data from various sources from specific fields (genes, compounds, proteins, drugs, diseases, etc.) reveals associations between field-specific entities. Several computational drug repositioning approaches have been developed as multiple

data sources are integrated and used to repurpose existing drugs. To implement each approach effectively, a reliable dataset must first be built [18]. For effective drug repositioning, a database must exist to connect information such as the interaction, similarity, and relevance of various elements. Extensive research efforts on building such databases for drug repositioning are currently in progress. Recently, data-based drug repositioning research has been performed, and electronic health records

Table 4. Database for Drug Repositioning

Database	Comment
ChEMBL [58,59]	http://chembank.broad.harvard.edu/ A public web-based information technology environment Freely available data and resources
ChEMBL [57]	https://www.ebi.ac.uk/chembl/ Database of bioactive drug-like small molecules Open database of EMBL-EBI with ADMET information Additional data on clinical progress of compounds has been integrated.
ClinicalTrials.gov [60]	https://clinicaltrials.gov/ Comprehensive clinical trial data representing the US, EU, and Japan
DailyMed [61]	https://dailymed.nlm.nih.gov/dailymed/about-dailymed.cfm Database containing labels for products submitted to FDA
DrugBank [62]	https://go.drugbank.com/ Comprehensive, open access, online database containing information on drugs and drug targets
FAERS [63,64]	https://fis.fda.gov/sense/app/95239e26-e0be-42d9-a960-9a5f7f1c25ec/sheet/7a47a261-d58b-4203-a8aa-6d3021737452/state/analysis New user-friendly search tool that improves access to real-world adverse event data
Gene Ontology [65]	http://www.geneontology.org The world's largest source of information on the functions of genes A community-based bioinformatics resource Uses ontology to represent biological knowledge
KEGG [21,22]	http://www.kegg.jp/ or http://www.genome.jp/kegg/ Large-scale molecular datasets generated by genome sequencing and other high-throughput experimental technologies
MedHelp [66]	https://www.medhelp.org/ Source of medical, health and wellness information created by users
MEDLINE [67]	https://www.nlm.nih.gov/medline/medline_overview.html Database of the NLM that contains more than 28 million documents in the life sciences
MedlinePlus [68]	https://medlineplus.gov Health information website for the general public with NLM's consumer-focused health information
MeSH [69]	https://www.ncbi.nlm.nih.gov/mesh/ Comprehensive vocabulary for genes, diseases, and drugs that co-occur in the literature
OMIM [70]	https://omim.org/ Comprehensive knowledge base of human genes and genetic phenotypes
PharmGKB [71,72]	https://www.pharmgkb.org Interactive tool for researchers investigating how genetic variation affects drug response
PreMedKB [73]	http://www.fudan-pgx.org/premedkb/index.html#/home A knowledge base that integrates the four basic components of precision medicine: disease, genes, variants and drugs
PubChem [74]	https://pubchem.ncbi.nlm.nih.gov/ Open chemistry database at the NIH
PubMed [75]	https://pubmed.ncbi.nlm.nih.gov/about/ MEDLINE's database of biomedical literature, life science journals, online books
RepoDB [76]	http://apps.chiragjgroup.org/repoDB/ Standard set of drug repositioning successes and failures
SemMedDB [77,78]	https://lhncbc.nlm.nih.gov/ii/tools/SemRep_SemMedDB_SKR/SemMed.html Summarizes MEDLINE citations returned by a PubMed search Extract semantic predications from titles and abstracts by natural language processing
SIDER [79,80]	http://sideeffects.embl.de/ Computer-readable side effect resource linking drug and side effects terms

ChEMBL, Chemical database of the European Molecular Biology Laboratory; ADMET, Absorption, Distribution, Metabolism, Excretion, Toxicity; EMBL, European Molecular Biology Laboratory; EBI, European Bioinformatics Institute; US, United States; EU, European Union; FDA, US Food and Drug Administration; FAERS, FDA Adverse Event Reporting System; KEGG, Kyoto Encyclopedia of Genes and Genomes; MEDLINE, Medical Literature Analysis and Retrieval System Online; NLM, National Library of Medicine; MeSH, Medical Subject Headings; OMIM, Online Mendelian Inheritance in Man; PharmGKB, Pharmacogenomics Knowledge Base; PreMedKB, Precision Medicine KnowledgeBase; NIH, National Institutes of Health; RepoDB, Drug Repositioning Database; SemMed DB, Semantic Medical Literature Analysis and Retrieval System Online (MEDLINE) database; SIDER, Side Effect Resource.

(EHR), EMR, insurance claim data, genomic data, and drug databases represent good data source examples [81].

Electronic medical records/electronic health records, insurance claim data

EMR data contains disease- and phenotype-related information that can be used as raw data for drug discovery [25,81]. The ability to conduct large-scale follow-up studies related to patient outcomes collected from EMR is an important advantage [82-84]. Medical databases such as EMR and claim data provide health records of millions of individuals, rendering them suitable for discovering new indications for available drugs [85].

Recently, an algorithm was developed to identify drug candidates effective for DM and dyslipidemia by analyzing large amounts of EMR data and clinical trial results [86]. This algorithm can be used to monitor the post-marketing safety of drugs and re-evaluate their effectiveness in clinical trials to ultimately discover new indications [86]. Metformin has been suggested as a repositioning candidate for cancer treatment because it reduces cancer incidence and mortality [41]. GLP1-RA has shown neuroprotective effects, rendering it a candidate therapeutic substance against neurodegeneration [48]. Thus, drug repositioning based on EMR data is most suitable due to the advantage of obtaining large-scale, sophisticated, and structured medical data in a short time [87]. However, sample size reduction problems due to missing data or exclusion of data from multiple patients might still represent a problem for this approach [88]. Furthermore, data quality is often unreliable, and in most cases, preprocessing is required [87,88], ultimately implying a critical privacy issue. To compensate for this, all information that can identify individuals should be removed from the accumulated data, and the extracted information should be stored in an encrypted file and made accessible only to designated persons [83,89].

Genomic data

If the expression of a certain gene changes from baseline in association with a specific disease, a drug that can alleviate this change in gene expression may have a therapeutic effect. Network biology using genomic data is an efficient and high-potential next-generation approach for drug repositioning or drug-to-drug combination of existing drugs. From this pharmacological perspective, drug repositioning for genetically rare diseases can be promoted by combining human genetics or genome-wide studies with network biology. In addition, such an approach has been proposed to solve the challenges of personalized medicine along with machine learning approaches [90]. Denny et al. [91]

confirmed the association between single nucleotide polymorphism (SNP) and diseases using the diagnostic code of the EMR dataset. A method of confirming the association between a target SNP and disease is called a phenome-wide association study (PheWAS). The PheWAS framework enable taking advantage of the genetic diversity between populations, ultimately making it possible to understand the functional role of specific genes. Therefore, PheWASs might be useful for prioritizing candidate drug targets [90].

For instance, PheWAS provided genetic evidence that GLP1-RA can prevent HF [49]. Furthermore, DPP4i, such as gemigliptin, linagliptin, and evogliptin, reportedly exert antiviral properties, suggesting their potential as broad-spectrum antiviral agents [45]. Another study indicated that SGLT2i play a protective role in the occurrence of AF [46]. Thus, by analyzing the change in gene expression according to the drug, a new point of action or indication of the drug can be identified. Such gene expression information can be obtained for almost any compound or disease regardless of whether the drug is approved or not. Additionally, it is a popular method because even small changes in each drug and disease can be obtained in an objective and detailed manner.

Drug database

Adverse drug reaction (ADR) information can be utilized in a network-based approach using pharmaceutical databases such as the Pharmacogenomics Knowledge Base (PharmGKB) [92], Side Effect Resource (SIDER) [79], and MedHelp [66]. It is possible to research health-related networks, including drugs and diseases, or use drug and disease names as well as ADRs as keywords to reach the relevant social network service. Numerous studies on drug repositioning, including studies investigating biological relationships, have been published. It is sometimes presented as a supporting basis in the method of acquiring new knowledge, such as a text-mining approach and/or semantic approach, ultimately exploring and predicting the relationship between biological concepts or entities. Researchers built a disease-adverse event relationship database using drug-adverse event data extracted from SIDER and drug-disease relationship data extracted from PharmGKB, and identified drug repositioning candidates by measuring the strength of the disease-adverse event relationship [93].

CONCLUSIONS

Drug repositioning is a method of identifying new drug indica-

tions by detecting new relationships between diseases and clinically proven drugs in an economical and efficient manner. In addition, repositioned drugs can be released to the market relatively quickly by applying an appropriate approach and analyzing a large amount of diverse data. With respect to drug safety and pharmacokinetics, it has a higher success rate than the traditional drug development method [94], and the resulting indications can be used to treat infectious diseases or cancers. Chen et al. [95]. demonstrated the effectiveness of pyrvinium in liver cancer patients by modeling the inverse correlation coefficient in gene expression and response in cancer patients. According to Paik et al. [96], clinical signatures extracted from EHR show that terbutaline sulfate, a known bronchodilator, can be repurposed to treat amyotrophic lateral sclerosis. Furthermore, an active movement encourages the development of therapeutic drugs for leukemia, Alzheimer's disease, Parkinson's disease, and diabetes via drug repositioning. As such, the potential demand and necessity for drug repositioning are expected to increase, and new indications for drugs in various fields, including intractable diseases, are expected to be identified.

Despite various limitations, drug repositioning is a field that will inevitably receive a spotlight in the drug development arena. The process of identifying new therapeutic drugs and treatments for untreated diseases will continue to accelerate. Learning the concept, pros, and cons of various methods used in drug repositioning will pose the basis for implementing new innovative technologies based on scientific evidence, taking into consideration patient safety.

CONFLICTS OF INTEREST

No potential conflict of interest relevant to this article was reported.

ACKNOWLEDGMENTS

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. NRF-2021R1G1A1091471).

ORCID

Suehyun Lee <https://orcid.org/0000-0003-0651-6481>

Seongwoo Jeon <https://orcid.org/0000-0002-2090-0706>

Hun-Sung Kim <https://orcid.org/0000-0002-7002-7300>

REFERENCES

1. Chakravarty K, Antontsev VG, Khotimchenko M, Gupta N, Jagarapu A, Bunday Y, et al. Accelerated repurposing and drug development of pulmonary hypertension therapies for COVID-19 treatment using an AI-integrated biosimulation platform. *Molecules* 2021;26:1912.
2. Spinelli A, Pellino G. COVID-19 pandemic: perspectives on an unfolding crisis. *Br J Surg* 2020;107:785-7.
3. Jarada TN, Rokne JG, Alhaji R. A review of computational drug repositioning: strategies, approaches, opportunities, challenges, and directions. *J Cheminform* 2020;12:46.
4. Yeu Y, Yoon Y, Park S. Protein localization vector propagation: a method for improving the accuracy of drug repositioning. *Mol Biosyst* 2015;11:2096-102.
5. Saberian N, Peyvandipour A, Donato M, Ansari S, Draghici S. A new computational drug repurposing method using established disease-drug pair knowledge. *Bioinformatics* 2019;35:3672-8.
6. Mohapatra S, Nath P, Chatterjee M, Das N, Kalita D, Roy P, et al. Repurposing therapeutics for COVID-19: rapid prediction of commercially available drugs through machine learning and docking. *PLoS One* 2020;15:e0241543.
7. Mickael ME, Pajares M, Enache I, Manda G, Cuadrado A. Nrf2 drug repurposing using a question-answer artificial intelligence system. *bioRxiv* 2019 Apr 4 [Preprint]. <https://www.biorxiv.org/content/10.1101/594622v1>.
8. Ghofrani HA, Osterloh IH, Grimminger F. Sildenafil: from angina to erectile dysfunction to pulmonary hypertension and beyond. *Nat Rev Drug Discov* 2006;5:689-702.
9. Ekman P. Finasteride in the treatment of benign prostatic hypertrophy: an update. New indications for finasteride therapy. *Scand J Urol Nephrol Suppl* 1999;203:15-20.
10. Sanchez-Garcia A, Simental-Mendia M, Millan-Alanis JM, Simental-Mendia LE. Effect of sodium-glucose co-transporter 2 inhibitors on lipid profile: a systematic review and meta-analysis of 48 randomized controlled trials. *Pharmacol Res* 2020;160:105068.
11. Filippatos TD, Tsimihodimos V, Elisaf MS. Mechanisms of blood pressure reduction with sodium-glucose co-transporter 2 (SGLT2) inhibitors. *Expert Opin Pharmacother* 2016;17:1581-3.
12. Ryan D, Acosta A. GLP-1 receptor agonists: nonglycemic clinical effects in weight loss and beyond. *Obesity (Silver Spring)* 2015;23:1119-29.
13. Blommel ML, Blommel AL. Pregabalin: an antiepileptic

- agent useful for neuropathic pain. *Am J Health Syst Pharm* 2007;64:1475-82.
14. Paranjpe MD, Taubes A, Sirota M. Insights into computational drug repurposing for neurodegenerative disease. *Trends Pharmacol Sci* 2019;40:565-76.
 15. Xue H, Li J, Xie H, Wang Y. Review of drug repositioning approaches and resources. *Int J Biol Sci* 2018;14:1232-44.
 16. Keiser MJ, Setola V, Irwin JJ, Lagner C, Abbas AI, Hu-feisen SJ, et al. Predicting new molecular targets for known drugs. *Nature* 2009;462:175-81.
 17. 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. *Proc Natl Acad Sci U S A* 2010;107:14621-6.
 18. Tian Z, Teng Z, Cheng S, Guo M. Computational drug repositioning using meta-path-based semantic network analysis. *BMC Syst Biol* 2018;12(Suppl 9):134.
 19. Wu Z, Wang Y, Chen L. Network-based drug repositioning. *Mol Biosyst* 2013;9:1268-81.
 20. Wu C, Gudivada RC, Aronow BJ, Jegga AG. Computational drug repositioning through heterogeneous network clustering. *BMC Syst Biol* 2013;7 Suppl 5:S6.
 21. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res* 2016;44:D457-62.
 22. Kyoto Encyclopedia of Genes and Genomes. KEGG: Kyoto Encyclopedia of genes and genomes [Internet]. Kyoto: Kanehisa Laboratories; 2021 [cited 2022 Mar 26]. Available from: <https://www.genome.jp/kegg/>.
 23. Cheng F, Desai RJ, Handy DE, Wang R, Schneeweiss S, Barabasi AL, et al. Network-based approach to prediction and population-based validation of in silico drug repurposing. *Nat Commun* 2018;9:2691.
 24. Emig D, Ivliev A, Pustovalova O, Lancashire L, Bureeva S, Nikolsky Y, et al. Drug target prediction and repositioning using an integrated network-based approach. *PLoS One* 2013;8:e60618.
 25. Goldstein JA, Bastarache LA, Denny JC, Roden DM, Pulley JM, Aronoff DM. Calcium channel blockers as drug repurposing candidates for gestational diabetes: mining large scale genomic and electronic health records data to repurpose medications. *Pharmacol Res* 2018;130:44-51.
 26. Rotermund C, Machtetanz G, Fitzgerald JC. The therapeutic potential of metformin in neurodegenerative diseases. *Front Endocrinol (Lausanne)* 2018;9:400.
 27. Datamet. Development of new drugs using AI (3): Using text mining to streamline new drug development [Internet]. Seoul: Network TIMES; 2021 [updated 2021 Jun 3; cited 2022 Mar 26]. Available from: <https://www.datamet.co.kr/news/articleView.html?idxno=160227>.
 28. Fleuren WW, Alkema W. Application of text mining in the biomedical domain. *Methods* 2015;74:97-106.
 29. Kostoff RN, Briggs MB, Shores DR. Treatment repurposing for inflammatory bowel disease using literature-related discovery and innovation. *World J Gastroenterol* 2020;26:4889-99.
 30. Zhang M, Luo H, Xi Z, Rogaeva E. Drug repositioning for diabetes based on 'omics' data mining. *PLoS One* 2015;10:e0126082.
 31. Zhang L, Hu J, Xu Q, Li F, Rao G, Tao C. A semantic relationship mining method among disorders, genes, and drugs from different biomedical datasets. *BMC Med Inform Decis Mak* 2020;20(Suppl 4):283.
 32. Pijl H, Ohashi S, Matsuda M, Miyazaki Y, Mahankali A, Kumar V, et al. Bromocriptine: a novel approach to the treatment of type 2 diabetes. *Diabetes Care* 2000;23:1154-61.
 33. 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 Cheminform* 2013;5:30.
 34. Menden MP, Iorio F, Garnett M, McDermott U, Benes CH, Ballester PJ, et al. Machine learning prediction of cancer cell sensitivity to drugs based on genomic and chemical properties. *PLoS One* 2013;8:e61318.
 35. Koren G, Nordon G, Radinsky K, Shalev V. Identification of repurposable drugs with beneficial effects on glucose control in type 2 diabetes using machine learning. *Pharmacol Res Perspect* 2019;7:e00529.
 36. Ng L, Foo DC, Wong CK, Man AT, Lo OS, Law WL. Repurposing DPP-4 inhibitors for colorectal cancer: a retrospective and single center study. *Cancers (Basel)* 2021;13:3588.
 37. Sanders JM, Monogue ML, Jodlowski TZ, Cutrell JB. Pharmacologic treatments for coronavirus disease 2019 (COVID-19): a review. *JAMA* 2020;323:1824-36.
 38. Gao J, Tian Z, Yang X. Breakthrough: chloroquine phosphate has shown apparent efficacy in treatment of COVID-19 associated pneumonia in clinical studies. *Biosci Trends* 2020;14:72-3.
 39. Gautret P, Lagier JC, Parola P, Hoang VT, Meddeb L, Mailhe M, et al. Hydroxychloroquine and azithromycin as a treatment of COVID-19: results of an open-label non-randomized clinical trial. *Int J Antimicrob Agents* 2020;56:105949.
 40. Baruah P, Das A, Paul D, Chakrabarty S, Aguan K, Mitra S.

- Sulfonylurea class of antidiabetic drugs inhibit acetylcholinesterase activity: unexplored auxiliary pharmacological benefit toward Alzheimer's disease. *ACS Pharmacol Transl Sci* 2021;4:193-205.
41. Xu H, Aldrich MC, Chen Q, Liu H, Peterson NB, Dai Q, et al. Validating drug repurposing signals using electronic health records: a case study of metformin associated with reduced cancer mortality. *J Am Med Inform Assoc* 2015;22:179-91.
 42. Sharma S, Nozohouri S, Vaidya B, Abbruscato T. Repurposing metformin to treat age-related neurodegenerative disorders and ischemic stroke. *Life Sci* 2021;274:119343.
 43. Moulton CD, Hopkins C, Ismail K, Stahl D. Repositioning of diabetes treatments for depressive symptoms: a systematic review and meta-analysis of clinical trials. *Psychoneuroendocrinology* 2018;94:91-103.
 44. Wahlqvist ML, Lee MS, Chuang SY, Hsu CC, Tsai HN, Yu SH, et al. Increased risk of affective disorders in type 2 diabetes is minimized by sulfonylurea and metformin combination: a population-based cohort study. *BMC Med* 2012;10:150.
 45. Rao PPN, Pham AT, Shakeri A, El Shatshat A, Zhao Y, Karuturi RC, et al. Drug repurposing: dipeptidyl peptidase IV (DPP4) inhibitors as potential agents to treat SARS-CoV-2 (2019-nCoV) infection. *Pharmaceuticals (Basel)* 2021;14:44.
 46. Bonora BM, Raschi E, Avogaro A, Fadini GP. SGLT-2 inhibitors and atrial fibrillation in the Food and Drug Administration adverse event reporting system. *Cardiovasc Diabetol* 2021;20:39.
 47. Pintana H, Apaijai N, Chattipakorn N, Chattipakorn SC. DPP-4 inhibitors improve cognition and brain mitochondrial function of insulin-resistant rats. *J Endocrinol* 2013;218:1-11.
 48. Markaki I, Winther K, Catrina SB, Svenningsson P. Repurposing GLP1 agonists for neurodegenerative diseases. *Int Rev Neurobiol* 2020;155:91-112.
 49. Daghlas I, Karhunen V, Ray D, Zuber V, Burgess S, Tsao PS, et al. Genetic evidence for repurposing of GLP1R (glucagon-like peptide-1 receptor) agonists to prevent heart failure. *J Am Heart Assoc* 2021;10:e20331.
 50. Foltyniec T, Athauda D. Repurposing anti-diabetic drugs for the treatment of Parkinson's disease: rationale and clinical experience. *Prog Brain Res* 2020;252:493-523.
 51. Rameshrad M, Razavi BM, Lalau JD, De Broe ME, Hosseinzadeh H. An overview of glucagon-like peptide-1 receptor agonists for the treatment of metabolic syndrome: a drug repositioning. *Iran J Basic Med Sci* 2020;23:556-68.
 52. Levy P, Jellinger PS. The potential role of colesivelam in the management of prediabetes and type 2 diabetes mellitus. *Postgrad Med* 2010;122(3 Suppl):1-8.
 53. Magkou D, Tziomalos K. Antidiabetic treatment, stroke severity and outcome. *World J Diabetes* 2014;5:84-8.
 54. Kobayashi Y, Banno K, Kunitomi H, Tominaga E, Aoki D. Current state and outlook for drug repositioning anticipated in the field of ovarian cancer. *J Gynecol Oncol* 2019;30:e10.
 55. Xu H, Li J, Jiang X, Chen Q. Electronic health records for drug repurposing: current status, challenges, and future directions. *Clin Pharmacol Ther* 2020;107:712-4.
 56. Oprea TI, Bauman JE, Bologna CG, Buranda T, Chigaev A, Edwards BS, et al. Drug repurposing from an academic perspective. *Drug Discov Today Ther Strateg* 2011;8:61-9.
 57. ChEMBL. ChEMBL database [Internet]. Hinxton: The European Bioinformatics Institute (EMBL-EBI); 2018 [cited 2022 Mar 26]. Available from: <https://www.ebi.ac.uk/chembl/>.
 58. Broad Institute. ChemBank [Internet]. Cambridge: Broad Institute [cited 2022 Mar 26]. Available from: <http://chembank.broad.harvard.edu/>.
 59. Seiler KP, George GA, Happ MP, Bodycombe NE, Carrinski HA, Norton S, et al. ChemBank: a small-molecule screening and cheminformatics resource database. *Nucleic Acids Res* 2008;36:D351-9.
 60. U.S. National Library of Medicine. clinicaltrials.gov [Internet]. Bethesda: National Library of Medicine; 2021 [cited 2022 Mar 26]. Available from: <https://clinicaltrials.gov/>.
 61. DailyMed. Dailymed overview [Internet]. Bethesda: National Library of Medicine; 2021 [cited 2022 Mar 26]. Available from: <https://dailymed.nlm.nih.gov/dailymed/about-dailymed.cfm>.
 62. DrugBank. DrugBank online: database for drug and drug target info [Internet]. Alberta: DrugBank; 2017 [cited 2022 Mar 26]. Available from: <https://go.drugbank.com/>.
 63. Xu R, Wang Q. Large-scale combining signals from both biomedical literature and the FDA Adverse Event Reporting System (FAERS) to improve post-marketing drug safety signal detection. *BMC Bioinformatics* 2014;15:17.
 64. U.S. Food and Drug Administration. FDA adverse event reporting System (FAERS) Public Dashboard [Internet]. Silver Spring: FDA; 2021 [cited 2022 Mar 26]. Available from: <https://www.fda.gov/drugs/questions-and-answers-fdas-adverse-event-reporting-system-faers/fda-adverse-event-reporting-system-faers-public-dashboard>.
 65. Gene ontology. The gene ontology resource [Internet]. Bar Harbor: The Gene Ontology Consortium; 2021 [cited 2022 Mar 26]. Available from: <http://www.geneontology.org/>.

66. MedHelp. Be your healthiest [Internet]. San Francisco: MedHelp; 2021 [cited 2022 Mar 26]. Available from: <https://www.medhelp.org/>
67. National Library of Medicine. Medline: overview [Internet]. Bethesda: National Library of Medicine; 2021 [cited 2022 Mar 26]. Available from: https://www.nlm.nih.gov/medline/medline_overview.html.
68. MedlinePlus. Health information from the National library of medicine [Internet]. Bethesda: National Library of Medicine; 2021 [cited 2022 Mar 26]. Available from: <https://medlineplus.gov/>.
69. Lowe HJ, Barnett GO. Understanding and using the medical subject headings (MeSH) vocabulary to perform literature searches. *JAMA* 1994;271:1103-8.
70. Hamosh A, Scott AF, Amberger JS, Bocchini CA, McKusick VA. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res* 2005;33:D514-7.
71. PharmGKB. Search pharmGKB [Internet]. Stanford: PharmGKB; 2021 [cited 2022 Mar 26]. Available from: <https://www.pharmgkb.org/>.
72. Thorn CF, Klein TE, Altman RB. PharmGKB: the Pharmacogenomics Knowledge Base. *Methods Mol Biol* 2013; 1015:311-20.
73. PreMedKB. Precision Medicine Knowledge Base [Internet]. Shanghai: PreMedKB; 2018 [cited 2022 Mar 26]. Available from: <http://www.fudan-pgx.org/premedkb/index.html#/home>.
74. PubChem. Explore chemistry [Internet]. Bethesda: National Library of Medicine; 2021 [cited 2022 Mar 26]. Available from: <https://pubchem.ncbi.nlm.nih.gov/>.
75. National Library of Medicine. PubMed [Internet]. Bethesda: National Library of Medicine; 2021 [cited 2022 Mar 26]. Available from: <https://pubmed.ncbi.nlm.nih.gov/>.
76. Malas TB, Vlietstra WJ, Kudrin R, Starikov S, Charrouf M, Roos M, et al. Drug prioritization using the semantic properties of a knowledge graph. *Sci Rep* 2019;9:6281.
77. Rindfleisch TC, Kilicoglu H, Fiszman M, Roseblat G, Shin D. Semantic medline: an advanced information management application for biomedicine. *Inf Serv Use* 2011;31:15-21.
78. National Library of Medicine. Access to SemRep/SemMed-DB/SKR Resources [Internet]. Bethesda: National Library of Medicine; 2021 [cited 2022 Mar 26]. Available from: <https://skr3.nlm.nih.gov/SemMedDB/>.
79. Kuhn M, Campillos M, Letunic I, Jensen LJ, Bork P. A side effect resource to capture phenotypic effects of drugs. *Mol Syst Biol* 2010;6:343.
80. SIDER 4.1. SIDER 4.1 Side effect resource [Internet]. SIDER 4.1; 2015 [cited 2022 Mar 26]. Available from: <http://sideeffects.embl.de/>.
81. Yao L, Zhang Y, Li Y, Sanseau P, Agarwal P. Electronic health records: implications for drug discovery. *Drug Discov Today* 2011;16:594-9.
82. Bate A, Juniper J, Lawton AM, Thwaites RM. Designing and incorporating a real world data approach to international drug development and use: what the UK offers. *Drug Discov Today* 2016;21:400-5.
83. Kim HS, Lee S, Kim JH. Real-world evidence versus randomized controlled trial: clinical research based on electronic medical records. *J Korean Med Sci* 2018;33:e213.
84. Lee S, Kim HS. Prospect of artificial intelligence based on electronic medical record. *J Lipid Atheroscler* 2021;10:282-90.
85. Ozery-Flato M, Goldschmidt Y, Shaham O, Ravid S, Yanover C. Framework for identifying drug repurposing candidates from observational healthcare data. *JAMIA Open* 2020;3:536-44.
86. Kim DH, Lee JE, Kim YG, Lee Y, Seo DW, Lee KH, et al. High-throughput algorithm for discovering new drug indications by utilizing large-scale electronic medical record data. *Clin Pharmacol Ther* 2020;108:1299-307.
87. Kim HS, Kim DJ, Yoon KH. Medical big data is not yet available: why we need realism rather than exaggeration. *Endocrinol Metab (Seoul)* 2019;34:349-54.
88. Kim HS, Kim JH. Proceed with caution when using real world data and real world evidence. *J Korean Med Sci* 2019; 34:e28.
89. Shin SY, Kim HS. Data pseudonymization in a range that does not affect data quality: correlation with the degree of participation of clinicians. *J Korean Med Sci* 2021;36:e299.
90. Nabirotkin 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;51:78-92.
91. Denny JC, Ritchie MD, Basford MA, Pulley JM, Bastarache L, Brown-Gentry K, et al. PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. *Bioinformatics* 2010;26:1205-10.
92. Whirl-Carrillo M, McDonagh EM, Hebert JM, Gong L, Sangkuhl K, Thorn CF, et al. Pharmacogenomics knowledge for personalized medicine. *Clin Pharmacol Ther* 2012;92: 414-7.

93. Yang L, Agarwal P. Systematic drug repositioning based on clinical side-effects. *PLoS One* 2011;6:e28025.
94. Lotfi Shahreza M, Ghadiri N, Mousavi SR, Varshosaz J, Green JR. A review of network-based approaches to drug repositioning. *Brief Bioinform* 2018;19:878-892.
95. Chen B, Ma L, Paik H, Sirota M, Wei W, Chua MS, et al. Reversal of cancer gene expression correlates with drug efficacy and reveals therapeutic targets. *Nat Commun* 2017;8:16022.
96. Paik H, Chen B, Sirota M, Hadley D, Butte AJ. Integrating clinical phenotype and gene expression data to prioritize novel drug uses. *CPT Pharmacometrics Syst Pharmacol* 2016;5:599-607.