HOKKAIDO UNIVERSITY

Title	Multi-Omics and Artificial Intelligence-Guided Drug Repositioning : Prospects, Challenges, and Lessons Learned from COVID-19
Author(s)	Cong, Yi; Endo, Toshinori
Citation	OMICS: A Journal of Integrative Biology, 26(7), 361-371 https://doi.org/10.1089/omi.2022.0068
Issue Date	2022
Doc URL	https://hdl.handle.net/2115/96103
Rights(URL)	https://creativecommons.org/licenses/by/4.0/
Type	journal article
File Information	OMICS J. Integr. Biol.26-7_361-413.pdf

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

OMICS A Journal of Integrative Biology
Volume 26, Number 7, 2022
Mary Ann Liebert, Inc.
DOI: 10.1089/omi.2022.0068

Review Article

Open camera or QR reader and scan code to access this article and other resources online.

Multi-Omics and Artificial Intelligence-Guided Drug Repositioning: Prospects, Challenges, and Lessons Learned from COVID-19

Yi Cong and Toshinori Endo

Abstract

Drug repurposing is of interest for therapeutics innovation in many human diseases including coronavirus disease 2019 (COVID-19). Methodological innovations in drug repurposing are currently being empowered by convergence of omics systems science and digital transformation of life sciences. This expert review article offers a systematic summary of the application of artificial intelligence (AI), particularly machine learning (ML), to drug repurposing and classifies and introduces the common clustering, dimensionality reduction, and other methods. We highlight, as a present-day high-profile example, the involvement of AI/ML-based drug discovery in the COVID-19 pandemic and discuss the collection and sharing of diverse data types, and the possible futures awaiting drug repurposing in an era of AI/ML and digital technologies. The article provides new insights on convergence of multi-omics and AI-based drug repurposing. We conclude with reflections on the various pathways to expedite innovation in drug development through drug repurposing for prompt responses to the current COVID-19 pandemic and future ecological crises in the 21st century.

Keywords: drug repositioning, Big Data, bioinformatics, machine learning, drug research and OMICS, COVID-19

Introduction

Drug repositioning is a method of developing new targets for existing drugs, which can significantly reduce time, cost, and other wastes as the targets are compounds that have already been tested for safety and pharmacokinetics (Jourdan et al, 2020; Rapicavoli et al, 2022). Drug repositioning relies on two main scientific bases that some diseases share common biological pathways and that a drug may have multiple targets that may be effective to different diseases (Jourdan et al, 2020). Within this context, and with the advent of the Big Data era, the generation of vast amounts of biological and chemical information has provided the scientific community with new opportunities to link drugs to diseases (March-Vila et al, 2017).

Since December 2019, coronavirus disease 2019 (COVID-19) has been recognized as a worldwide public health emergency (Hui et al, 2020) and declared as a global pandemic by the World Health Organization (WHO) in 2020. Although there is a massive vaccination campaign underway, emerging variants limit the efficacy of the campaign (Sibilio et al, 2021). Therefore, the search for new drugs that can treat patients remains important and urgent, and drug repositioning, with its time and cost advantages, certainly makes it highly visible.

This article highlights the emerging intersection with artificial intelligence (AI) and machine learning (ML) from a multi-omics perspective, and summarizes and discusses drug repositioning approaches that have been applied to COVID-19.

Laboratory of Information Biology, Information Science and Technology, Hokkaido University, Sapporo, Japan.

© Yi Cong and Toshinori Endo, 2022. Published by Mary Ann Liebert, Inc. This Open Access article is distributed under the terms of the Creative Commons License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited.

SARS Coronavirus-2: Pathogenesis and Focused Treatments

Coronaviruses are enveloped RNA viruses that are widely distributed in mammals, including humans, as well as birds, that cause respiratory, intestinal, hepatic, and neurological diseases (Zhu et al, 2020). Human coronaviruses (HCoVs) are positive-sense single-stranded RNA viruses and 30,000 bp long (Pirone et al, 2020). Before December 2019, six coronaviruses are known to cause human disease, of which, SARS coronavirus (SARS-CoV) and Middle East respiratory syndrome coronavirus (MERS-CoV) are associated with fatal disease (Cui et al, 2019; Zhu et al, 2020).

Infection with the novel coronavirus SARS-CoV-2 belongs to the sarbecovirus subgenus of the Coronaviridae family (Zhu et al, 2020), caused a cluster of severe respiratory illnesses, which can lead to acute respiratory failure and even death (Pirone et al, 2020). WHO reported that there had been 516,922,683 cumulative cases and 6,259,945 cumulative deaths as of May 13, 2022.

Clinical deterioration in patients with severe COVID-19 disease is usually rapid, and a large part of the severe disease course is due to a cytokine storm leading to a massive inflammatory response, which leads to multi-organ failure or even death, which is thought to be possibly related to immune checkpoint activation and immune system failure (Behrens and Koretzky, 2017; Kim et al, 2021; Sibilio et al, 2021). At the same time, this massive immune response has set the stage for testing several immunomodulatory agents simultaneously with antiviral drugs (Sibilio et al, 2021), and several specific immunomodulatory agents, anti-cytokines such as interleukin-1 (IL-1) and IL-6 receptor antagonists, are considered to have potential for the treatment of cytokine storms (Rizk et al, 2020).

Multi-Omics and AI in Drug Repositioning

Avalanche of omics data

With the continuous development of new technologies for the determination of multi-omics data, the possibility of obtaining high-dimensional histological data quickly and efficiently is offered. Because a disease often emerges as a complex interaction between multiple genetic variants (Hirschhorn and Daly, 2005), a single layer of ''omics'' usually provides only limited insight into the biological mechanisms of the disease, with DNA, RNA, proteins, and metabolites often acting in complementary roles to join some biological function (Sun and Hu, 2016). Therefore, the integrated analysis of multi-omics data is of great importance in the study of complex biological processes and disease mechanisms.

Figure 1 briefly represents the different layers of the multi-omics data. For the overview of omics modalities, background, and origin, please refer to the study by Manzoni et al (2018).

Through gene expression profiling, multi-omics data can be easily used for drug repurposing. The Connectivity Map (CMap) is a gene expression profiling database based on interventional gene expression proposed by Lamb et al (2006); it is mainly used to reveal functional associations between small-molecule compounds, genes, and diseases. CMap and Library of Integrated Network-Based Cellular Signatures (LINCS) as its extension are considered to be the key concepts behind various drug repurposing studies (Jarada et al, 2020). Specifically, the CMap database contains mainly gene expression profiles of different cell lines acted by a large number of perturbagens (small-molecule compounds, overexpressed genes, etc.).

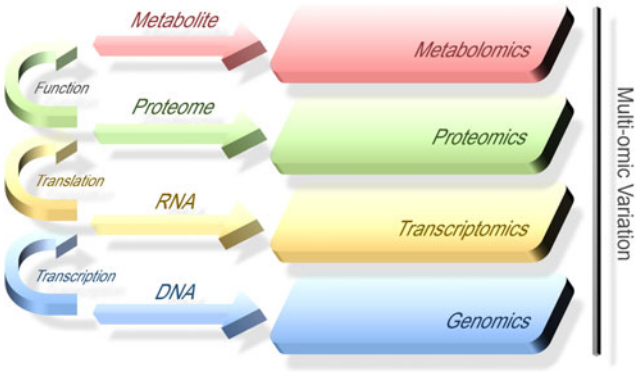


FIG. 1. Different layers of multi-omics data. The *straight arrows* in the *middle* indicate the organismal molecules that constitute the different layers of the so-called omics cascade, and the *curved arrows* on the *left side* imply the biological process between different types of molecules.

By comparing reference data, cellular gene expression profiles are found with high similarity and thus investigate whether cells have some kind of connection to different inductions. For use, the list of up/downregulated differentially expressed genes obtained from the experimental analysis is compared with the database reference data set by CMap; the enrichment direction and intensity of each reference is scored according to the enrichment of differentially expressed genes in the reference gene expression profile to determine the degree of enrichment with the query feature, that is, the similarity of differentially expressed genes to the reference gene expression profile, and ranked.

Amemiya et al (2019) proposed a computational drug repositioning approach to perform an integrated multi-omics analysis based on transcriptomic, proteomic, and interactomic data to detect drug candidates for dengue hemorrhagic fever. In this study, signature genes were identified by integrating the Gene Expression Omnibus (GEO) data set, drug candidates were identified by CMap search, disease specific pathways were detected using Gene Set Enrichment Analysis (GSEA) approach for transcriptomic and proteomic data, and finally, a human–dengue virus protein–protein interaction (PPI) network was constructed (Amemiya et al, 2019).

Convergence of multi-omics and AI/ML applications toward drug repurposing

The biggest challenge in drug repositioning is to customize or optimize methods to develop promising, affordable, and efficient drug repositioning pipelines for complex diseases (Jin and Wong, 2014; Zeng et al, 2020), so screening methods for drug reuse are particularly important. Depending on the classification, the approaches involved in drug repositioning can be classified as drug-oriented, target-oriented, and disease or therapy-oriented according to the information

available related to quality and quantity (Sahoo et al, 2021), or as network-based, ligand-based, chemogenomic and ML, ligand-based approaches, and so on, according to the means (March-Vila et al, 2017).

With the advent of high-throughput technologies, more and more data need to be explored and used by computational analysis and mining tools, and to achieve systematic or comprehensive repurposing, various computer-based approaches are gradually becoming mainstream (Rapicavoli et al, 2022). *In silico* drug repositioning is a hypothesis-driven approach that can translate available omics data through the collection of disease phenotypes and targets, such as genome-wide association analyses or gene expression response profiles, pathway mappings, compound structures, as well as data related to drug modes of action, into predictions of druggable targets, ideally, top provide a list of Food and Drug Administration (FDA)-approved drug candidates with potential modulatory/inhibitory functions (Mottini et al, 2021; Pushpakom et al, 2019).

Our focus here is on AI and ML methods that leverage publicly available databases and information sources. DSP-1181 was the first repurposed drug to enter clinical trials discovered through an AI approach, and the time from initial screening to the end of preclinical testing was reduced from the 4 years that would have been required to <12 months (Farghali et al, 2021). AI-based approaches have enabled a more nuanced and iterative process to rapidly identify potentially bioactive compounds from millions of drug candidates in a short period, which has revolutionized the drug development process and facilitated the realization of precision medicine (Boniolo et al, 2021; Nayarisseri et al, 2021).

AI has also led to the creation of many reverse vaccinology (RV) virtual frameworks that are often categorized as rule-based filtering models. ML enables the creation of models that can learn and generalize patterns within existing data and are able to make inferences from previously unseen data. With the advent of deep learning (DL), the learning process can also include the automatic extraction of features from raw data (Prasad and Kumar, 2021; Sarker, 2021).

As an example of fusing multi-omics and AI/ML applications, in our recent study, we proposed a new approach to drug repurposing involving two-stage prediction and ML with applications to inclusion body myositis, polymyositis, and dermatomyositis (Cong et al, 2022). First, diseases are clustered by gene expression, with the thought that similar patterns of altered gene expression imply critical pathways shared in different disease conditions. Second, drug efficacy is determined based on the ability to reverse altered gene expression, and the results are clustered to identify repurposing targets.

Since the number of clusters cannot be well determined by using Uniform Manifold Approximation and Projection for Dimension Reduction (UMAP) alone, we introduced the *k*-means method in combination with it to obtain effective grouping information while maintaining good clustering performance. As a result, disease-specific gene expression and 22 drugs for repurposing were identified. The details are presented in Figure 2. To date, many computational approaches to drug repositioning using ML technology continue to be proposed and improved, especially in the context of COVID-19 of global concern.

AI-Based Drug Repositioning as a Strategy for Identifying New Therapeutic Agents for COVID-19

Understanding the genetic regulatory code that controls gene expression is an important topic in the field of molecular biology and will provide us with the means to cure diseases. However, because the biological sequence space is too large to explore, experimental studies are limited to a single regulatory component in the context of single reporter genes (Zrimec et al, 2020). Therefore, various methods based on AI and DL have been introduced.

Classification and regression

Linear regression is the simplest model, usually used to solve continuous numerical prediction problems, which uses regression analysis of mathematical statistics to determine the interdependence between variables. Logistic regression adds a sigmoid function mapping to linear regression and is often used to solve classification problems for estimating the likelihood of something (Cox, 1958).

Buza et al, (2020) proposed a new method called MOLIERE for drug–target interaction (DTI) prediction and compared their results with the bipartite local model (BLM), previously proposed by Yamanishi et al, (2008), that is popular in DTI by predicting the target protein of a given drug to reason about the drug that targets a given protein (Bleakley and Yamanishi, 2009; Buza et al, 2020; Yamanishi et al, 2008). In this study, a framework called asymmetric loss models was instantiated using linear regression, adding the use of weighted profile (WP) to BLM, which was considered to have better performance relative to the original BLM and WP (Buza et al, 2020).

Gottlieb et al (2011) proposed an algorithm for predicting drug–disease associations using various data sources called PREdicting Drug IndiCaTion (PREDICT). The algorithm uses drug–drug and disease–disease similarity measures as classification features, and they apply a logistic regression classifier to distinguish between true and false drug–disease associations and ultimately predict new associations.

Neural network (NN)-based approaches are data-driven methods that can learn potential feature representations directly from labeled training data, and a review article addresses NN-based biomedical classification methods in detail (Jarada et al, 2020).

Deep neural networks (DNNs) increase the number of layers based on NN, optimizes the model structure and training methods, and improves the computational power to better cope with large amounts of data (Sze et al, 2017).

Unlike traditional ML, both convolutional neural networks (CNNs) and recurrent neural network (RNNs) can autonomously learn features and autonomously optimize the weights of each layer to obtain the set of values that best represent the features in the training network. CNNs are mostly used in computer vision and image processing applications, and CNN is a network model that does not consider sequential data. And RNNs are models that consider sequence data, mostly used for language model and video data processing applications (Krizhevsky et al, 2017; Lecun et al, 1998; Olurotimi, 1994).

Based on CNNs, the computational framework named convolutional neural network for coexpression (CNNC) was proposed, which provides the ability to perform genetic relationship

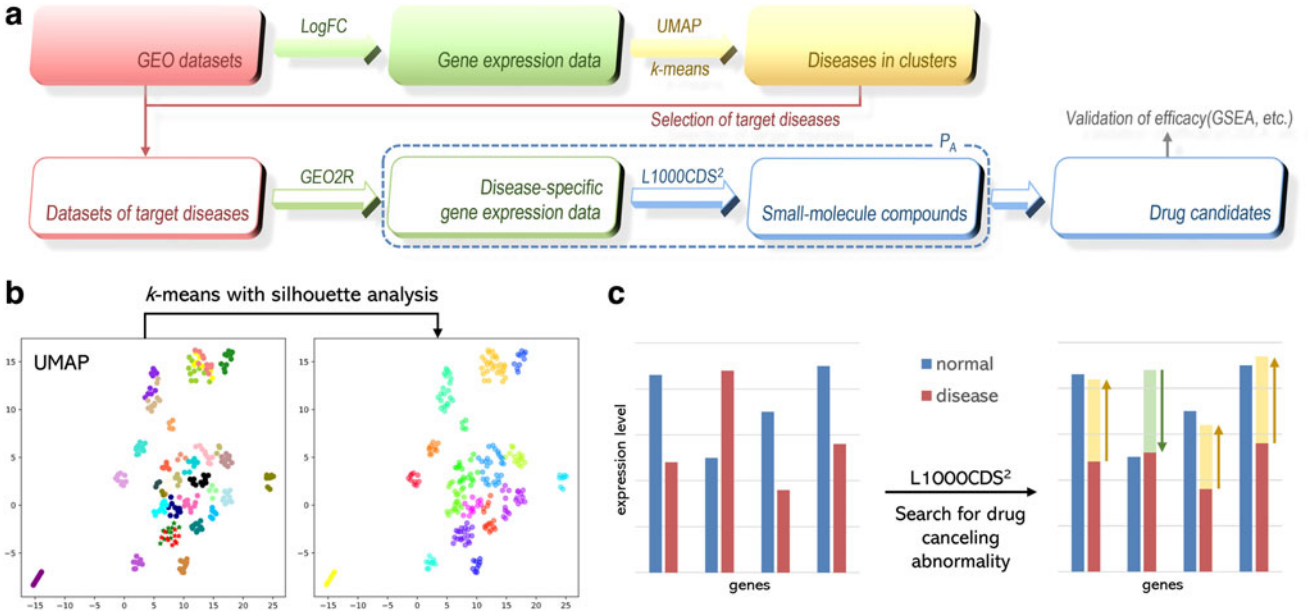


FIG. 2. A new approach to drug repurposing with two-stage prediction, ML, and unsupervised clustering of gene expression. (**a**) The flow of the method. First, gene expression data from 262 cases of 31 diseases and 268 controls were transformed into log fold-change (logFC) values. Then, to cluster similar diseases, the transformed gene expression data were analyzed by UMAP, followed by *k*-means to optimize the number of clusters. For evaluation, we examined disease-specific gene expression data for the three target diseases, inclusion body myositis, polymyositis, and dermatomyositis, and used L1000CDS^2 to obtain lists of small-molecule compounds that reversed the expression patterns of these specifically altered genes as candidates for drug repurposing. Finally, the functions of affected genes were analyzed by GSEA to examine consistency with expected drug efficacy. (**b**) Clustering using UMAP and *k*-means method. The *left figure* shows the visualization results obtained by UMAP, where *different colors* represent samples of different diseases, and it can be seen that samples belonging to the same disease are basically well clustered together. However, the UMAP results do not show how to group the different diseases together, so we introduced the *k*-means method, and the results are shown on the *right*, where the *same colors* represent the samples that are grouped in the same cluster. By comparing the *left* and *right figures*, it is easy to get the clustering information of different kinds of diseases. (**c**) Exploration of small-molecule compounds at P_A step that recover gene expression patterns affected by diseases. Using data for genes with variable expression. These small-molecule compounds can reduce/increase the expression of genes that are over/underexpressed in disease samples relative to healthy samples. GSEA, Gene Set Enrichment Analysis; L1000CDS^2, LINCS L1000 characteristic direction signatures search engine; LINCS, Library of Integrated Network-Based Cellular Signatures; logFC, log fold-change; ML, machine learning; UMAP, Uniform Manifold Approximation and Projection for Dimension Reduction.

inference in a supervised manner and was shown to be superior to previous approaches in inferring interactions, causality, and function assignment (Yuan and Bar-Joseph, 2019).

It is worth mentioning that different kinds of NNs can be combined to build hybrid models for classification. As synthetic screening experiments are considered to be limited in their ability to study the overall relationship between different parts of gene regulatory structures and co-regulation, deep convolutional neural network (DCNN) was used to predict gene expression levels from natural DNA sequences (Agarwal and Shendure, 2020; Zrimec et al, 2020).

Support-vector network (SVM) is a two-group classification model, whose basic model is defined as a linear classifier with maximum interval on the feature space, and whose learning strategy is interval maximization, which can eventually be translated into the solution of a convex quadratic programming problem (Cortes et al, 1995).

Lin et al (2020) proposed a new ML pipeline using the support-vector regression (SVR) algorithm that is a regression version of SVM. Using models trained from AA2AR, DHI1, and AL5AP, respectively, they demonstrated its ability in predicting the binding affinity of drugs to specific proteins and assessed the similarity of drug binding profiles between proteins by the Spearman correlation coefficient (SCC) of coefficients between models, which suggested that the models could be applied for screening of candidate drugs.

The goal of generative adversarial network (GAN) is to train two NNs, where one generative model tries to generate images similar to the real training samples by replicating the distribution of the data, and the other discriminative network predicts the probability that the generated images are from the real training set. These two models compete with each other so as to ultimately produce image results similar to the real training samples (Fanny and Cenggoro, 2018; Goodfellow et al, 2014).

Fanny and Cenggoro (2018) proposed a deep learning approach for imbalance data classification using class expert generative adversarial network (CE-GAN), which attaches a generator of GAN to the classifier instead of sharing a single architecture for differentiation and classification, and effectively increases the size of the data set and improves the classification.

Gene expression clustering

The goal of ML is to extract feature patterns from data that can represent the relationship between input data features and

the output targets to be predicted and to use these patterns to make predictions, and the representation of these input features directly affects the nature and quality of the acquired patterns, and to some extent, it can be said that the selection of features determines the upper limit of ML. Therefore, it is particularly important to extract meaningful features from large and noisy data sets, and clustering and dimensionality reduction methods are introduced to assist in feature value selection.

Eisen et al (1998) used hierarchical clustering to classify the generated dendrograms according to predefined criteria and is considered the most commonly used clustering method, which results in side-by-side display of dendrograms and genetic heat maps (D'haeseleer, 2005). However, since it is based on the principle of creating a hierarchical nested tree by calculating the similarity between different classes of data points, in bioinformatics, gene expression information tends to be somewhat correlated, usually leading to the appearance of larger errors.

Another popular unsupervised learning method, *k*-means, which aims to partition the data set and update the center of mass by minimizing the sum of squares within clusters, is characterized by good scalability as the sample size increases (Chaudhuri and Chaudhuri, 1997; D'haeseleer, 2005; Hozumi et al, 2021). However, since it relies on computing the distance between randomly given cluster centers and each sample, it requires a large high-dimensional feature space that can lead to expensive computations, large memory requirements, and poor clustering performance (Hozumi et al, 2021) and is therefore often used together with dimensionality reduction methods in genomics research, as we will present in the next section.

Considering that the effectiveness of the *k*-means method is largely influenced by the *k*-value and the selection of the initial value of cluster centers and is more sensitive to noise and outliers, it extends to variants such as *k*-means++, intelligent *k*-means, genetic *k*-means, *k*-medians, and so on. It is worth to be mentioned that a completely unsupervised kernel-based clustering algorithm called Intelligent Kernel *K*-Means (IKKM) has been proposed to cluster the kernel matrix without any information. It has been used in experiments targeting gene expression clustering in human colorectal cancer and positive results have been obtained (Handhayani and Hiryanto, 2015).

The self-organizing map (SOM) method starts from a predetermined number of place names, and each iteration moves a place name closest to the selected gene to that gene, eventually forming a grid of clusters, which neighboring clusters show related expression patterns (D'haeseleer, 2005; Tamayo et al, 1999).

Clustering methods based on multi-omics data

A large number of multi-omics approaches for data integration exist for different objectives. These methods can be classified into three types according to their biological objectives and the way they process the data: regression/association-based methods, clustering-based methods, and network-based methods (Vahabi and Michailidis, 2022). Here, we focus on the clustering-based methods that are of great importance in precision medicine as well as drug research; for specific analytical evaluation of integration methods, please refer to the studies by Chauvel et al (2020) and Vahabi and Michailidis (2022).

According to the algorithmic approach, clustering methods for multi-omics are classified into three categories: (1) early integration by joining multi-omics data to form a single matrix with multiple omics features on which a single histological clustering algorithm is applied, (2) intermediate integration by building a model containing all omics data and thus achieving clustering, and (3) late integration by clustering each omics data individually and then integrating the resulting clustering solution. The specific review methods are described in detail in Rappoport and Shamir (2018).

Early integrations, such as the LDACluster model proposed by Wu et al (2015), assume that the features of different omics data are random variables obeying a hidden parameter distribution, and clustering is performed on this benchmark by obtaining the parameter matrix and decomposing it into a lower order representation of the original data (Rappoport and Shamir, 2018).

The Cancer Genome Atlas (TCGA) team proposed Cluster-Of-Cluster-Assignments (C-of-C/COCA) as one of the late integrations to deepen the understanding of cancer at the molecular level, which uses data from five different genomic/proteomics platforms to cluster tumors from different tissues to define breast cancer subtypes (The Cancer Genome Atlas Network, 2012). After that, COCA was used again to perform a comprehensive analysis of samples from 12 cancer types through 5 genome-wide platforms and 1 proteomics platform and successfully revealed a uniform classification of 11 major subtypes (Hoadley et al, 2014).

There are more types of methods included in intermediate integration. Similarity network fusion (SNF), which is recognized as a proven classical algorithm, is a similarity-based method for intermediate integration. SNF considers the complementarity contained in different omics data and first builds a fully connected network on each histology with samples as nodes and similarity as weights, and subsequently updates the network using iterative methods (Rappoport and Shamir, 2018; Wang et al, 2014).

In 2018, a robust and adaptive to noise clustering algorithm, the RSC-OTRI, was established to better identify noisy high-dimensional gene expression clusters with different histological features. It aims to maximize the separation in terms of survival curves by building a gene data matrix and computing eigenvalues and eigenvectors. In the article, it was compared with SNF and *t*-mixture (TMIX: model-based clustering based on Student's *t*-distribution), respectively, and the results showed that it was able to combine good and sparse gene correlation estimates and performed robustly to noise and survival analysis (Coretto et al, 2018).

Dimensionality reduction assisted clustering of large-scale data sets

In the previous section, we mentioned that for commonly used clustering methods such as *k*-means, high-latitude gene expression data can lead to expensive costs with poor performance. In fact, most clustering methods are prone to performance degradation because the feature space becomes sparse in high-dimensional space (Coretto et al, 2018), and dimensionality reduction should effectively avoid this problem.

Principal component analysis (PCA) is one of the oldest and most widely used. It is based on the principle of

projecting data matrices onto a low-dimensional space, and being a descriptive tool that does not require assumptions makes it well suited for various types of adaptive exploration methods but has limitations for data sets that are relatively more important to maintain local distances, such as genome sequences (Hozumi et al, 2021; Jolliffe and Cadima, 2016).

UMAP is a relatively new dimensionality reduction method based on multibody theory for topological data analysis techniques (McInnes et al, 2018). The principle lies in optimizing the spectral layout of the data in the lower dimensional space so that the error between two topological spaces will be minimized (McInnes et al, 2018). Since it can capture the global and topological structure while maintaining the local distance and its fast and effective operation, it has advantages for large data sets and is better for capturing recognition interactions and analyzing transcriptome data as well as visualize genetic interactions with biomolecular spatial relationships (Dorrity et al, 2020).

In addition, it is considered to be the most suitable dimensionality reduction method for use with *k*-means clustering methods that the combination has been used in genome phylogenetic analysis (Hozumi et al, 2021) and drug repositioning (Cong et al, 2022). It has been shown to have

Table 1. Summary of Important Artificial Intelligence-Based Methods and Studies

Name	*Tasks*	*Methods*	*References*
Linear regression	Regression	Linear regression	Cox (1958)
Logistic regression	Regression	Logistic regression	Cox (1958)
BLM	DTI	BLM	Bleakley and Yamanishi (2009)
MOLIERE	DTI	Linear regression; BLM	Buza et al (2020)
PREDICT	Predicting drug-disease	Logistic regression	Gottlieb et al (2011)
NN	Classification	NN	Jarada et al (2020)
DNN	Classification	DNN	Sze et al (2017)
CNN	Classification	CNN	Krizhevsky et al (2017); Lecun et al (1998)
RNN	Classification	RNN	Olurotimi (1994)
CNNC	Inferring gene relationship	CNN	Yuan and Bar-Joseph (2019)
DCNN	Predicting mRNA abundance	CNN; DNN	Agarwal and Shendure (2020)
SVM	Classification	SVM	Cortes et al (1995)
n.a.	Protein prediction	SVM (SVR)	Lin et al (2020)
GAN	Classification	GAN	Goodfellow et al (2014)
CE-GAN	Classification	GAN	Fanny and Cenggoro (2018)
Hierarchical clustering	Clustering	Hierarchical clustering	Eisen et al (1998)
k-Means	Clustering	*k*-Means	Chaudhuri and Chaudhuri (1997); Hozumi et al (2021)
IKKM	Clustering	IKKM	Handhayani and Hiryanto (2015)
SOM	Clustering	SOM	Tamayo et al (1999)
LDACluster	Clustering (multi-omics)	LDACluster	Wu et al (2015)
COCA	Clustering (multi-omics)	COCA	Hoadley et al (2014); The Cancer Genome Atlas Network (2012)
SNF	Clustering (multi-omics)	SNF	Wang et al (2014)
RSC-OTRI	Clustering (multi-omics)	RSC-OTRI	Coretto et al (2018)
PCA	Dimensionality reduction	PCA	Jolliffe and Cadima (2016)
UMAP	Dimensionality reduction	UMAP	Dorrity et al (2020); McInnes et al (2018)
t-SNE	Dimensionality reduction	t-SNE	van der Maaten and Hinton (2008)
LDA	Dimensionality reduction	LDA	Fisher (1936)
MathDL	Predicting the interaction	MathDL	Nguyen et al (2020a); Nguyen et al (2020b)
SBDR	Drug repositioning	Ligand	Gao et al (2020)
n.a.	Phylogenetic analysis	*k*-Means; UMAP	Hozumi et al (2021)
n.a.	Drug repositioning	*k*-Means; UMAP	Cong et al (2022)
n.a.	Drug repositioning	BenevolentAI knowledge graph	Richardson et al (2020)
n.a.	Drug repositioning	SVM	Kowalewski and Ray (2020)
n.a.	Drug repositioning	DNN	Ke et al (2020)
MT-DTI	Drug repositioning	NLP	Beck et al (2020)
n.a.	Drug repositioning	NN; naive bayes	Kadioglu et al (2021)

AI, artificial intelligence; BLM, bipartite local model; CE-GAN, class expert generative adversarial network; CNN, convolutional neural network; CNNC, convolutional neural network for coexpression; COCA, Cluster-Of-Cluster-Assignments; DCNN, deep convolutional neural network; DNN, deep neural network; DTI, drug–target interaction; GAN, generative adversarial network; IKKM, Intelligent Kernel *K*-Means; LDA, linear discriminant analysis; MathDL, mathematical deep learning; mRNA, messenger RNA; MT-DTI, Molecule Transformer-Drug–Target Interaction; n.a., not available; NLP, natural language process; NN, neural network; PCA, principal component analysis; PREDICT, PREdicting Drug IndiCaTion; RNN, recurrent neural network; SBDR, structure-based drug repositioning; SNF, similarity network fusion; SOM, self-organizing map; SVM, support-vector network; SVR, support vector regression; t-SNE, t-distributed Stochastic Neighbor Embedding; UMAP, Uniform Manifold Approximation and Projection for Dimension Reduction.

significantly improved clustering accuracy and is gradually becoming one of the popular dimensionality reduction methods recently in genomics and other areas.

t-distributed Stochastic Neighbor Embedding (t-SNE) is a nonlinear dimensional reduction algorithm, as an improvement of the SNE, it uses a symmetric probability formula and does not resort to a normal distribution such as the SNE when calculating the probabilities between sample points in low-dimensional space, but instead uses a t-distribution. In measuring the scatter distance between two probability distributions, t-SNE uses the Kullback–Leibler (KL) divergence as SNE does (Hozumi et al, 2021; van der Maaten and Hinton, 2008).

The goal of linear discriminant analysis, which also uses projection, aims to minimize intra-class variation and maximize directional projection of inter-class variation (Fisher, 1936).

AI approaches in COVID-19 drug discovery

AI methods are widely used in public health, disease prediction, and drug development. With the advent of DL, automatic feature extraction from raw data has led to improved performance compared with other computer-aided models. Different DL algorithms are used to fight the COVID-19 pandemic, including artificial neural network, CNN, and long short-term memory (Prasad and Kumar, 2021). Next, we discuss some representative examples of AI-driven drug development that used in COVID-19.

Generative models were proposed for antiviral drug discovery in the early days of the pandemic. Nguyen et al (2020b) proposed a mathematical deep learning (MathDL) model for generating a low-dimensional representation of high-dimensional chemical/physical interactions. They integrated this representation into different DL models such as CNN and GAN for predicting the pose and energy of the interaction. They applied this model for finding inhibitors for 3CLpro of SARS-CoV-2 (Nguyen et al, 2020a).

3CLpro is an important potential drug target that is critical for the inhibition of viral replication. On this basis, Gao et al (2020) evaluated the binding affinity of drugs to SARS-CoV-2 3CLpro using a structure-based drug repositioning (SBDR) ML model, and 314 SARS-CoV-2/SARS-CoV-3CL inhibitors were trained on a two-dimensional fingerprint-based DL gradient-enhanced decision tree model from 8565 drugs were evaluated and scored, and finally, the top 20 drugs approved by FDA and the top 20 non-marketed drugs in the study were selected as effective inhibitors of SARS-CoV-2 3CL protease.

Hozumi et al (2021) successfully performed a phylogenetic analysis of large-scale SARS-CoV-2 genomic sequence data sets by UMAP-assisted *k*-means method and also compared the performance of various dimensionality reduction methods used for assisted clustering to further analyze the effectiveness of different methods in terms of speed and scalability, which has positive implications in analyzing mutation patterns of viruses and predicting transmission routes for effective drug discovery and vaccine production. Meanwhile, UMAP, being validated as the most suitable dimensionality reduction method for auxiliary clustering, could be used in the future for more drug repurposing as well as other fields.

Richardson et al (2020) integrated biomedical data from structured and unstructured sources through the BenevolentAI knowledge graph to propose a list of potentially effective drugs including the anti-HIV lopinavir plus ritonavir combination. Although some of the compounds obtained in this study result in serious side effects and cannot be useful as effective treatments in this infection today, it is undeniable that the BenevolentAI knowledge graph, which can respond quickly at an early stage, has potential in areas such as future prevention of infectious diseases.

Kowalewski and Ray (2020) presented a ML drug discovery pipeline that identifies at least 6 potential lead drugs that may be effective against COVID-19 from FDA-registered chemicals and approved drugs as well as $\sim$14 million purchasable chemicals.

Ke et al (2020) developed an AI-based model using DNN to identify 80 promising marketed drugs and tested the activity of all AI-predicted drugs against feline coronavirus (FCoV) in a cell-based *in vitro* assay, 8 of which inhibited proliferation of feline infectious peritonitis virus (Mottaqi et al, 2021).

Beck et al (2020) used a DL-based DTI prediction model called Molecule Transformer-Drug–Target Interaction (MT-DTI), which they proposed in 2019, to identify a list of commercially available antiviral drugs that could disrupt the components of the SARS-CoV-2 virus. Prediction using

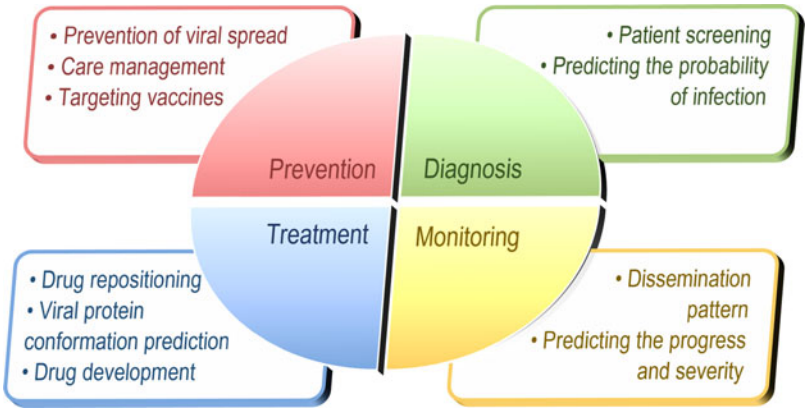


FIG. 3. Current applications of AI and ML in response to COVID-19. AI, artificial intelligence; COVID-19, coronavirus disease 2019.

established drug repurposing methods is a fast and efficient approach, which can help to provide a timely response in the face of an unexpected and complex situation.

By constructing a compound database with virtual drug screening, molecular docking, and supervised ML algorithm identification, Kadioglu et al (2021) obtained a list of the best compounds targeting spike protein, nucleocapsid protein, and 2′-o-ribose-methyltransferase and identified the top nine compounds with the highest protein–drug interactions.

Although some of these candidates, such as the antiviral Remdesivir, have been shown to be effective in adaptive platform trials (Beigel et al, 2020), most have not been proven in clinical trials or have been found to be ineffective suggesting that efficient and accurate identification of drug candidates is still a problem that needs to be addressed.

Conclusions and Future Perspectives

Through this sudden explosion of COVID-19, drug repositioning is back in the spotlight with its advantages in terms of time and cost in the drug development process. For ease of viewing, the methods we discuss this time are briefly summarized in Table 1. Although the most of these studies cannot yet be considered clinically successful, empirical facilitation of standardization through computational methods and predictive capabilities can lead to faster and more accurate responses in the face of possible future outbreaks of epidemics.

Figure 3 briefly summarizes the current applications of AI and ML in response to COVID-19; a more comprehensive summary of 146 articles is presented by Comito and Pizzuti (2022), and we can find that the majority of AI methods are used for detection/monitoring, prevention/treatment, and pathogenesis analysis, but only few are used for drug repositioning (Mottaqi et al, 2021). In other words, while AI approaches have been more fully researched in areas such as medical image processing diagnosis and its prediction of propagation patterns through ML calculations, there is still much room for progress in areas such as patient treatment and drug development.

In these areas, the potential of AI is undoubtedly huge, and if more advanced and targeted AI-based drug repositioning methods can be developed, it will undoubtedly be a great help for future efforts to fight infections similar to COVID-19.

AI/ML methods open the door not only drug repurposing with combining existed multi-omics data but also development of new drugs under complex conditions as well. For this sake, establishing relevant databases should be helpful. For emerging viruses, data collection activities are of great importance, as almost all computational methods need to be based on large high-quality data sets.

During this COVID-19 pandemic, the data sharing mechanism of the Global Initiative on Sharing All Influenza Data (GISAID), that the world's largest repository for SARS-CoV-2 sequences, the Centers for Disease Control and Prevention (CDC), WHO, and other databases greatly facilitated the progress of related research, while most of the data resources that have been established before the pandemic and expanded with COVID-19, such as molecular data resources such as GenBank, UniProt, PPI networks resources, or various drug databases, networks resources, and databases for various drugs or compounds have been utilized (Galindez et al, 2021).

It would be prospected that expansion of multi-omics data will be concerted for the drug repurposing as well as drug design by the conductor of AI/ML on the harmony of appropriately organized database in the future.

Acknowledgments

The authors wish to thank the anonymous reviewers as well as the editor for their valuable and constructive suggestions to improve the article.

Author Disclosure Statement

The authors declare they have no conflicting financial interests.

Funding Information

No funding was received for this article.

References

Agarwal V, Shendure J. Predicting mRNA abundance directly from genomic sequence using deep convolutional neural networks. Cell Rep 2020;31(7):107663; doi: 10.1016/j.celrep.2020.107663

Amemiya T, Gromiha MM, Horimoto K, et al. Drug repositioning for dengue haemorrhagic fever by integrating multiple omics analyses. Sci Rep 2019;9(1):523; doi: 10.1038/s41598-018-36636-1

Beck BR, Shin B, Choi Y, et al. Predicting commercially available antiviral drugs that may act on the novel coronavirus (SARS-CoV-2) through a drug-target interaction deep learning model. Comput Struct Biotechnol J 2020;18:784–790; doi: 10.1016/j.csbj.2020.03.025

Behrens EM, Koretzky GA. Review: Cytokine storm syndrome: looking toward the precision medicine era. Arthritis Rheumatol 2017;69(6):1135–1143; doi: 10.1002/art.40071

Beigel JH, Tomashek KM, Dodd LE, et al. Remdesivir for the treatment of Covid-19—Final report. N Engl J Med 2020;383(19):1813–1826; doi: 10.1056/NEJMoa2007764

Bleakley K, Yamanishi Y. Supervised prediction of drug–target interactions using bipartite local models. Bioinformatics 2009;25(18):2397–2403; doi: 10.1093/bioinformatics/btp433

Boniolo F, Dorigatti E, Ohnmacht AJ, et al. Artificial intelligence in early drug discovery enabling precision medicine. Expert Opin Drug Discov 2021;16(9):991–1007; doi: 10.1080/17460441.2021.1918096

Buza K, Peška L, Koller J. Modified linear regression predicts drug-target interactions accurately. PLoS One 2020;15(4):e0230726; doi: 10.1371/journal.pone.0230726

Chaudhuri D, Chaudhuri BB. A novel multiseed nonhierarchical data clustering technique. IEEE Trans Syst Man Cybern B Cybern 1997;27(5):871–876; doi: 10.1109/3477.623240

Chauvel C, Novoloaca A, Veyre P, et al. Evaluation of integrative clustering methods for the analysis of multi-omics data. Brief Bioinform 2020;21(2):541–552; doi: 10.1093/bib/bbz015

Comito C, Pizzuti C. Artificial intelligence for forecasting and diagnosing COVID-19 pandemic: A focused review. Artif Intell Med 2022;128:102286; doi: 10.1016/j.artmed.2022.102286

Cong Y, Shintani M, Imanari F, et al. A new approach to drug repurposing with two-stage prediction, machine learning, and unsupervised clustering of gene expression. OMICS: A Journal of Integrative Biology 2022;26(6):339–347; doi: 10.1089/omi.2022.0026.

Coretto P, Serra A, Tagliaferri R. Robust clustering of noisy high-dimensional gene expression data for patients subtyping. Bioinformatics 2018;34(23):4064–4072. doi: 10.1093/bioinformatics/bty502

Cortes C, Vapnik V, Saitta L. Support-vector networks editor. Mach Learn 1995;20:273–297; doi: 10.1023/A:1022627411411

Cox DR. The regression analysis of binary sequences. J R Stat Soc Series B Methodol 1958;20(2):215–232; doi: 10.1111/j.2517-6161.1958.tb00292.x

Cui J, Li F, Shi Z-L. Origin and evolution of pathogenic coronaviruses. Nat Rev Microbiol 2019;17(3):181–192; doi: 10.1038/s41579-018-0118-9

D'haeseleer P. How does gene expression clustering work? Nat Biotechnol 2005;23(12):1499–1501; doi: 10.1038/nbt1205-1499

Dorrity MW, Saunders LM, Queitsch C, et al. Dimensionality reduction by UMAP to visualize physical and genetic interactions. Nat Commun 2020;11(1):1537; doi: 10.1038/s41467-020-15351-4

Eisen MB, Spellman PT, Brown PO, et al. Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A 1998;95(25):14863–14868; doi: 10.1073/pnas.95.25.14863

Fanny, Cenggoro TW. Deep learning for imbalance data classification using class expert generative adversarial network. Procedia Comput Sci 2018;135:60–67; doi: 10.1016/j.procs.2018.08.150

Farghali H, Kutinová Canová N, Arora M. The potential applications of artificial intelligence in drug discovery and development. Physiol Res 2021;70(Suppl 4):S715–S722; doi: 10.33549/physiolres.934765

Fisher RA. The use of multiple measurements in taxonomic problems. Ann Eugenics 1936;7(2):179–188; doi: 10.1111/j.1469-1809.1936.tb02137.x

Galindez G, Matschinske J, Rose TD, et al. Lessons from the COVID-19 pandemic for advancing computational drug repurposing strategies. Nat Comput Sci 2021;1(1):33–41; doi: 10.1038/s43588-020-00007-6

Gao K, Nguyen DD, Chen J, et al. Repositioning of 8565 existing drugs for COVID-19. J Phys Chem Lett 2020;11(13):5373–5382; doi: 10.1021/acs.jpclett.0c01579

Goodfellow IJ, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks. arXiv:1406.2661. 2014; doi: 10.48550/arXiv.1406.2661

Gottlieb A, Stein GY, Ruppin E, et al. PREDICT: A method for inferring novel drug indications with application to personalized medicine. Mol Syst Biol 2011;7(1):496; doi: 10.1038/msb.2011.26

Handhayani T, Hiryanto L. Intelligent Kernel K-means for clustering gene expression. Procedia Comput Sci 2015;59:171–177; doi: 10.1016/j.procs.2015.07.544

Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. Nat Rev Genet 2005;6(2):95–108; doi: 10.1038/nrg1521

Hoadley KA, Yau C, Wolf DM, et al. Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. Cell 2014;158(4):929–944; doi: 10.1016/j.cell.2014.06.049

Hozumi Y, Wang R, Yin C, et al. UMAP-assisted K-means clustering of large-scale SARS-CoV-2 mutation datasets. Comput Biol Med 2021;131:104264; doi: 10.1016/j.compbiomed.2021.104264

Hui DS, I Azhar E, Madani TA, et al. The continuing 2019-NCoV epidemic threat of novel coronaviruses to global health—The latest 2019 novel coronavirus outbreak in Wuhan, China. Int J Infect Dis 2020;91:264–266; doi: 10.1016/j.ijid.2020.01.009

Jarada TN, Rokne JG, Alhajj R. A review of computational drug repositioning: Strategies, approaches, opportunities, challenges, and directions. J Cheminform 2020;12(1):46; doi: 10.1186/s13321-020-00450-7

Jin G, Wong STC. Toward better drug repositioning: prioritizing and integrating existing methods into efficient pipelines. Drug Discov Today 2014;19(5):637–644; doi: 10.1016/j.drudis.2013.11.005

Jolliffe IT, Cadima J. Principal component analysis: A review and recent developments. Philos Trans Royal Soc A Math Phys Eng Sci 2016;374(2065):20150202; doi: 10.1098/rsta.2015.0202

Jourdan J-P, Bureau R, Rochais C, et al. Drug repositioning: A brief overview. J Pharm Pharmacol 2020;72(9):1145–1151; doi: 10.1111/jphp.13273

Kadioglu O, Saeed M, Greten HJ, et al. Identification of novel compounds against three targets of SARS CoV-2 coronavirus by combined virtual screening and supervised machine learning. Comput Biol Med 2021;133:104359; doi: 10.1016/j.compbiomed.2021.104359

Ke Y-Y, Peng T-T, Yeh T-K, et al. Artificial intelligence approach fighting COVID-19 with repurposing drugs. Biomed J 2020;43(4):355–362; doi: 10.1016/j.bj.2020.05.001

Kim JS, Lee JY, Yang JW, et al. Immunopathogenesis and treatment of cytokine storm in COVID-19. Theranostics 2021;11(1):316–329; doi: 10.7150/thno.49713

Kowalewski J, Ray A. Predicting novel drugs for SARS-CoV-2 using machine learning from a >10 million chemical space. Heliyon 2020;6(8):e04639; doi: 10.1016/j.heliyon.2020.e04639

Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Commun ACM 2017;60(6):84–90; doi: 10.1145/3065386

Lamb J, Crawford ED, Peck D, et al. The Connectivity Map: Using gene-expression signatures to connect small molecules, genes, and disease. Science 2006;313(5795):1929–1935; doi: 10.1126/science.1132939

Lecun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition. Proc IEEE 1998;86(11):2278–2324; doi: 10.1109/5.726791

Lin Y-T, Sheu S-Y, Lin C-C. Prediction of drug-protein interaction and drug repositioning using machine 1 learning model 2. 2020; doi: https://www.bioXiv.org/content/10.1101/2020.07.29.218826v1

Manzoni C, Kia DA, Vandrovcova J, et al. Genome, transcriptome and proteome: the rise of omics data and their integration in biomedical sciences. Brief Bioinform 2018;19(2):286–302; doi: 10.1093/bib/bbw114

March-Vila E, Pinzi L, Sturm N, et al. On the integration of in silico drug design methods for drug repurposing. Front Pharmacol 2017;8:298; doi: 10.3389/fphar.2017.00298

McInnes L, Healy J, Melville J. UMAP: Uniform manifold approximation and projection for dimension reduction. 2018; doi: https://arXiv.org/abs/1802.03426

Mottaqi MS, Mohammadipanah F, Sajedi H. Contribution of machine learning approaches in response to SARS-CoV-2 infection. Inform Med Unlocked 2021;23:100526; doi: 10.1016/j.imu.2021.100526

Mottini C, Napolitano F, Li Z, et al. Computer-aided drug repurposing for cancer therapy: Approaches and opportunities to challenge anticancer targets. Semin Cancer Biol 2021;68:59–74; doi: 10.1016/j.semcancer.2019.09.023

Nayarisseri A, Khandelwal R, Tanwar P, et al. Artificial intelligence, big data and machine learning approaches in precision medicine & drug discovery. Curr Drug Targets 2021; 22(6):631–655; doi: 10.2174/1389450122999210104205732

Nguyen DD, Gao K, Chen J, et al. Potentially highly potent drugs for 2019-NCoV. 2020a; doi: https://www.bioxiv.org/content/10.1101/2020.02.05.936013v1

Nguyen DD, Gao K, Wang M, et al. MathDL: mathematical deep learning for D3R grand challenge 4. J Comput Aided Mol Des 2020b;34(2):131–147; doi: 10.1007/s10822-019-00237-5

Olurotimi O. Recurrent neural network training with feedforward complexity. IEEE Trans Neural Netw 1994;5(2):185–197; doi: 10.1109/72.279184

Pirone L, del Gatto A, di Gaetano S, et al. A multi-targeting approach to fight SARS-CoV-2 attachment. Front Mol Biosci 2020;7:186; doi: 10.3389/fmolb.2020.00186

Prasad K, Kumar V. Artificial intelligence-driven drug repurposing and structural biology for SARS-CoV-2. Curr Res Pharmacol Drug Discov 2021;2:100042; doi: 10.1016/j.crphar.2021.100042

Pushpakom S, Iorio F, Eyers PA, et al. Drug repurposing: Progress, challenges and recommendations. Nat Rev Drug Discov 2019;18(1):41–58; doi: 10.1038/nrd.2018.168

Rapicavoli RV, Alaimo S, Ferro A, et al. Computational methods for drug repurposing. Adv Exp Med Biol 2022; 1361:119–141; doi: 10.1007/978-3-030-91836-1_7

Rappoport N, Shamir R. Multi-omic and multi-view clustering algorithms: Review and cancer benchmark. Nucleic Acids Res 2018;46(20):10546–10562; doi: 10.1093/nar/gky889

Richardson P, Griffin I, Tucker C, et al. Baricitinib as potential treatment for 2019-NCoV acute respiratory disease. Lancet 2020;395(10223):e30–e31; doi: 10.1016/S0140-6736(20)30304-4

Rizk JG, Kalantar-Zadeh K, Mehra MR, et al. Pharmacoimmunomodulatory therapy in COVID-19. Drugs 2020; 80(13):1267–1292; doi: 10.1007/s40265-020-01367-z

Sahoo BM, Ravi Kumar BVV, Sruti J, et al. Drug repurposing strategy (DRS): Emerging approach to identify potential therapeutics for treatment of novel coronavirus infection. Front Mol Biosci 2021;8:628144; doi: 10.3389/fmolb.2021.628144

Sarker IH. Deep learning: A comprehensive overview on techniques, taxonomy, applications and research directions. SN Comput Sci 2021;2(6):420; doi: 10.1007/s42979-021-00815-1

Sibilio P, Bini S, Fiscon G, et al. In silico drug repurposing in COVID-19: A network-based analysis. Biomed Pharmacother 2021;142:111954; doi: 10.1016/j.biopha.2021.111954

Sun YV, Hu Y-J. Integrative analysis of multi-omics data for discovery and functional studies of complex human diseases. Adv Genet 2016;93:147–190; doi: 10.1016/bs.adgen.2015.11.004

Sze V, Chen Y-H, Yang T-J, et al. Efficient processing of deep neural networks: A tutorial and survey. Proc IEEE 2017; 105(12):2295–2329; doi: 10.1109/JPROC.2017.2761740

Tamayo P, Slonim D, Mesirov J, et al. Interpreting patterns of gene expression with self-organizing maps: Methods and application to hematopoietic differentiation. Proc Natl Acad Sci U S A 1999;96(6):2907–2912; doi: 10.1073/pnas.96.6.2907

The Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. Nature 2012;490(7418): 61–70; doi: 10.1038/nature11412

Vahabi N, Michailidis G. Unsupervised multi-omics data integration methods: A comprehensive review. Front Genet 2022;13:854752; doi: 10.3389/fgene.2022.854752

van der Maaten L, Hinton G. Visualizing data using t-SNE. Journal of Machine Learning Research 2008; 9(86):2579–2605. https://jmlr.org/papers/v9/vandermaaten08a.html

Wang B, Mezlini AM, Demir F, et al. Similarity network fusion for aggregating data types on a genomic scale. Nat Methods 2014;11(3):333–337; doi: 10.1038/nmeth.2810

Wu D, Wang D, Zhang MQ, et al. Fast dimension reduction and integrative clustering of multi-omics data using low-rank approximation: Application to cancer molecular classification. BMC Genomics 2015;16(1):1022; doi: 10.1186/s12864-015-2223-8

Yamanishi Y, Araki M, Gutteridge A, et al. Prediction of drug-target interaction networks from the integration of chemical and genomic species. Bioinformatics 2008;24(3):i232–i240; doi: 10.1093/bioinformatics/btn162

Yuan Y, Bar-Joseph Z. Deep learning for inferring gene relationships from single-cell expression data. Proc Natl Acad Sci U S A 2019;116(52):27151–27158; doi: 10.1073/pnas.1911536116

Zeng X, Zhu S, Lu W, et al. Target identification among known drugs by deep learning from heterogeneous networks. Chem Sci 2020;11(7):1775–1797; doi: 10.1039/C9SC04336E

Zhu N, Zhang D, Wang W, et al. A novel coronavirus from patients with pneumonia in China, 2019. N Engl J Med 2020;382(8):727–733; doi: 10.1056/NEJMoa2001017

Zrimec J, Börlin CS, Buric F, et al. Deep learning suggests that gene expression is encoded in all parts of a co-evolving interacting gene regulatory structure. Nat Commun 2020; 11(1):6141; doi: 10.1038/s41467-020-19921-4

Address correspondence to:
Toshinori Endo, PhD
Laboratory of Information Biology
Research Group of Bioinformatics
Faculty of Information Science and Technology
Hokkaido University
N14 W9, Sapporo, 060-0814
Japan

E-mail: endo@ist.hokudai.ac.jp

Abbreviations Used

AI = artificial intelligence
BLM = bipartite local model
CE-GAN = class expert generative adversarial network
CMap = Connectivity Map
CNN = convolutional neural network
CNNC = convolutional neural network for coexpression
COCA = Cluster-Of-Cluster-Assignments
COVID-19 = coronavirus disease 2019
DCNN = deep convolutional neural network
DL = deep learning
DNN = deep neural network
DTI = drug–target interaction
FDA = Food and Drug Administration
GAN = generative adversarial network
GSEA = Gene Set Enrichment Analysis
IKKM = Intelligent Kernel *K*-Means
IL = interleukin
L1000CDS^2 = LINCS L1000 characteristic direction signatures search engine
LDA = linear discriminant analysis
LINCS = Library of Integrated Network-Based Cellular Signatures

(*continued*)

Abbreviations Used (Cont.)

logFC = log fold-change
MathDL = mathematical deep learning
ML = machine learning
MT-DTI = Molecule Transformer-Drug–Target Interaction
n.a. = not available
NLP = natural language process
NN = neural network
PCA = principal component analysis
PPI = protein–protein interaction
PREDICT = PREdicting Drug IndiCaTion
RNN = recurrent neural network
SARS-CoV = SARS coronavirus
SBDR = structure-based drug repositioning
SNF = similarity network fusion
SOM = self-organizing map
SVM = support-vector network
SVR = support-vector regression
t-SNE = t-distributed Stochastic Neighbor Embedding
UMAP = Uniform Manifold Approximation and Projection for Dimension Reduction
WHO = World Health Organization
WP = weighted profile