


RESEARCH

Open Access



Artificial intelligence-based computational framework for drug-target prioritization and inference of novel repositionable drugs for Alzheimer's disease

Shingo Tsuji^{1*†} , Takeshi Hase^{2,3,4,5†}, Ayako Yachie-Kinoshita^{2,4}, Taiko Nishino², Samik Ghosh², Masataka Kikuchi⁶, Kazuro Shimokawa⁷, Hiroyuki Aburatani¹, Hiroaki Kitano² and Hiroshi Tanaka³

Abstract

Background: Identifying novel therapeutic targets is crucial for the successful development of drugs. However, the cost to experimentally identify therapeutic targets is huge and only approximately 400 genes are targets for FDA-approved drugs. As a result, it is inevitable to develop powerful computational tools that can identify potential novel therapeutic targets. Fortunately, the human protein-protein interaction network (PIN) could be a useful resource to achieve this objective.

Methods: In this study, we developed a deep learning-based computational framework that extracts low-dimensional representations of high-dimensional PIN data. Our computational framework uses latent features and state-of-the-art machine learning techniques to infer potential drug target genes.

Results: We applied our computational framework to prioritize novel putative target genes for Alzheimer's disease and successfully identified key genes that may serve as novel therapeutic targets (e.g., DLG4, EGFR, RAC1, SYK, PTK2B, SOCS1). Furthermore, based on these putative targets, we could infer repositionable candidate-compounds for the disease (e.g., tamoxifen, bosutinib, and dasatinib).

Conclusions: Our deep learning-based computational framework could be a powerful tool to efficiently prioritize new therapeutic targets and enhance the drug repositioning strategy.

Keywords: Network embedding, Deep learning, Machine learning, Systems biology, Drug discovery, Protein interaction network

Background

Biomedical research, especially for the field of drug discovery, is currently experiencing a global paradigm shift with artificial intelligence (AI) technologies and their application to “Big Data” in the biomedical domain [1–3]. The complex, non-linear, multi-dimensional nature of big

data is accompanied by unique challenges and opportunities when employed for processing and analysis to derive actionable insights. In particular, existing statistical techniques, such as principle components analysis (PCA), are insufficient for capturing the complex interaction patterns that are hidden in multiple dimensions across the data spectrum [4]. Thus, a key challenge for future drug discovery research is the development of powerful AI-based computational tools that can capture multiple dimension of biomedical insights and obtain “value” in the form of actionable insights (e.g., insights toward to select and

*Correspondence: tsuji@genome.rcastu-tokyo.ac.jp

[†]Shingo Tsuji and Takeshi Hase contributed equally to this work.

¹Research Center for Advanced Science and Technology, The University of Tokyo, 4-6-1 Komaba, Meguro-ku, 153-8904, Tokyo, Japan

Full list of author information is available at the end of the article



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

prioritize candidate targets and repositionable drugs for candidate targets) from big data volumes.

“Big Data” in the biomedical domain are generally associated with high dimensionality. Their dimensionality should be reduced to avoid undesired properties of high-dimensional space, such as the curse of dimensionality [5]. Dimensionality reduction techniques facilitate classification, data visualization, and high-dimensional data compression [6]. However, classical dimensional reduction techniques (e.g., PCA) are generally linear techniques and thus insufficient to handle non-linear data [4, 6].

With the recent advancement in AI technologies, several dimensionality reduction techniques have become available for non-linear complex data [4, 6, 7]. Among the dimensionality reduction techniques, the multi-layer neural network-based technique, “deep autoencoder,” could serve as the most powerful technique for reducing the dimensionality of non-linear data [4, 6]. Deep autoencoders are composed of multilayer “encoder” and “decoder” networks. The multilayer “encoder” component transforms high-dimensionality data to a low-dimensional representation while multilayer “decoder” component recovers original high-dimensional data from the low-dimensional representation. Weights associated with the links that connect the layers are optimized by minimizing the discrepancy between the input and output of the network (i.e., in an ideal condition, the values for the nodes in the input layer is the same as those in the output layer). After the optimization steps, the middle-hidden encoder layer yields a low dimensional representation that preserves information that is considered original data as much as possible [6]. The values of nodes in the middle-hidden encoder layer would be useful features for classification, regression, and data visualization of high-dimensional data.

In drug discovery research, identifying novel drug-targets is critical for the successful development of a therapeutic drug [8–10]. However, the cost to experimentally predict drug targets is huge and only approximately 400 genes are used as targets of FDA-approved drugs [11]. Thus, it is inevitable to develop a powerful computational framework that can identify potentially novel drug-targets.

Drug repositioning is another promising approach for boosting new drug development. The advantage of drug repositioning is its established safety (i.e., toxicology studies have already been carried out with a target drug). Therefore, the development of computational methods to predict repositionable candidates could be a promising strategy to reduce the cost and time for drug development.

Different drug repositioning methods have been proposed in prior studies. Further, these methods can be classified into two different major categories: activity-

based drug repositioning and in silico drug repositioning. Several drugs for non-cancerous diseases have been discovered for cancer therapeutics using the former approach [12], and in recent years, the latter approach has become successful because of advancements of the protein-protein interaction database, protein structural database, and in-silico network analysis technology. Such types of applications for drug repositioning via the network theory have also been discussed. By verifying the similarity between CDK2 inhibitors and topoisomerase inhibitors, Iorio et al. [13] reported that Fasudil (a Rho-kinase inhibitor) might be applicable to several neurodegenerative disorders. Further, Cheng et al. [14] applied the inference method based on three similarities (drug-based, target-based, and network-based similarities) to predict the interactions between drugs and targets and finally confirmed that five old drugs could be repositioned.

PIN data could be a useful resource for computational investigations of potential novel drug-targets; that is because proteins derive their functions together with their interacting partners and a network of protein interaction captures downstream relationships between targets and proteins [8–10, 15]. With the recent advancement in network science, various network metrics are presently available and have been used to investigate the structure of molecular interaction networks and their relationship with drug-target genes [8–10, 15, 16]. For example, “degree,” which is the number of links to a protein, is a representative network metric for investigating the molecular interaction networks (i.e., almost all FDA-approved drug-targets are middle- or low-degree proteins; however, almost no therapeutic targets exist among high-degree proteins [10]). Such finding indicates that the key features for identifying potential drug target genes could be embedded in the complex architectures of the PIN [10].

Genome-wide PIN data are typical non-linear high-dimensional big-data in the biomedical domain that are composed of thousands of proteins as well as more than ten-thousand interactions among them [8, 9]. Mathematically, a PIN is represented as an adjacency matrix [17]. The adjacency matrices for PINs within rows and columns labeled by proteins and elements in the matrices are presented as a binary value (i.e., 1 or 0 in position (i, j) if protein i interacts with protein j or not). In the adjacency matrix, each row represents the interacting pattern for each protein and may be a useful feature for predicting potential drug target proteins.

Recently, researchers have developed “network embedding” methods that apply dimensional reduction techniques to extract low-dimensional representations of a large network from the high-dimensional adjacency matrix of the network [17, 18]. For example, several researchers have used singular value decomposition and

non-negative matrix factorization methods to map high-dimensional adjacency matrices of large-scale networks onto low-dimensional representations [19, 20]. However, the feature vector for a protein is high dimensional (e.g., several thousand dimensions) and sparse; this is because protein interaction network composed of thousands of proteins and the vast majority of proteins in the PIN have few interactions [17].

To address this issue, several researchers have employed network embedding methods based on deep learning techniques [21, 22]. Deep autoencoder-based network embedding methods would be especially useful for transforming non-linear large-scale networks into low-dimensional representations. Wang et al. applied a deep autoencoder-based network embedding method to large-scale social networks (e.g., arxiv-GrQc, blogcatalog, Flicker, and Youtube) and successfully mapped these networks onto low-dimensional representations [21].

Herein, to infer potentially novel target genes, we proposed a computational framework based on a representative network embedding method that employs a deep autoencoder to map a genome-wide protein interaction network onto low-dimensional representations. The framework builds a classifier based on state-of-the-art machine learning techniques to predict potentially novel drug-targets using the resultant low-dimensional representations. We applied the framework to predict potentially novel drug targets for Alzheimer's disease. Based on the list of predicted candidate novel drug targets, we further inferred potential repositionable drug candidates for Alzheimer's disease.

Methods

Overview

The first part of the method was preparing the PIN data and calculating the 100 dimension vector representation for each gene by using a deep autoencoder. To examine the performance of the deep autoencoder, we compared the 100 features with nine known network metrics. The second part was building a machine learning model which can predict whether a gene is a putative target of Alzheimer's therapeutic drug or not. In this step, we used Xgboost to build the model and SMOTE to mitigate the sample imbalance (i.e. only a few genes were known therapeutic targets).

PIN data and drug-target information

The PIN data was obtained elsewhere [23]. This network is composed of 6,338 genes and 34,814 non-redundant interactions among the genes.

We obtained information for drugs and their target genes from the DrugBank database [24, 25]. Thereafter, we investigated the "description" field for all the drugs in the DrugBank database and identified 61 therapeutic

drugs for Alzheimer's disease. The 61 targets for these drugs were regarded as the established drug targets for Alzheimer's disease. Among the 61 targets, 31 were mapped onto the PIN.

Feature extraction from PIN using a deep autoencoder

We build a deep autoencoder with a symmetric layer structure composed of 7 encoder layers and 7 decoder layers (e.g., 7 encoder layers (6338-3000-1500-500-250-150-100) and symmetric decoder layers (100-150-250-500-1500-3000-6338)). Layers are fully connected. In addition, layers, except output layer, use rectified linear unit (ReLU) [26] as an activation function while output layer uses sigmoid function to generate binary outputs. We optimized the deep autoencoder network by using "adam" [27] optimizer with a learning rate = 1.0×10^{-6} , number of epochs = 10,000, batch size = 10, and default values for the other parameters. In the optimization step, we minimized the binary cross-entropy loss between the values of nodes in the input layer and those in the output layer. We used a representative deep learning platform, "Keras" [28], with Tensorflow [29] backend to implement the deep autoencoder. To perform the deep autoencoder-based dimensionality reduction analysis of PIN, we used Tesla K80 GPU on the shirokane 5 super computer system [30].

Statistical and topological analysis of the PIN

To determine the statistical topological features in the PIN for each gene, we calculated the following representative network metrics: indegree, outdegree, betweenness, closeness, PageRank [31], cluster coefficient [32], nearest neighbor degree (NND) [33], bow-tie structures [34], and indispensable nodes [35, 36].

Indegree: Indegree for a given node represents the number of nodes connected to the node (i.e., upstream neighbors of the node).

Outdegree: Outdegree represents the number of links from the given node to other nodes (i.e., downstream neighbors of the nodes).

Betweenness: Betweenness for a given node i is the number of shortest paths between two nodes that pass through node i .

Closeness: The value of closeness for a given node i is the mean length of the shortest paths between node i and all other nodes in the network.

PageRank [31]: PageRank for a given node is a metric used to roughly estimate the importance of the node in the network. The PageRank score is calculated using the algorithm proposed by Google [37]. A given node has a higher PageRank if the nodes with a higher rank have links to the node.

Cluster coefficient [32]: Cluster coefficient of a node i (C_i) is calculated by using the following equation:

$C_i = \frac{2e_i}{k_i(k_i-1)}$, where k_i is the degree of node i and e_i is the number of links connecting the neighborhood of node i to one another.

Nearest neighbor degree (NND) [33]: The value of NND for a given node i is the average degree among nearest neighbor nodes of node i .

Bow-tie structure [34]: Biological networks often possess bow-tie structures that are composed of three components (i.e., input, core, and output layers). Yang et al. proposed a bow-tie decomposition method to classify nodes into three classes: the input layer, the core layer, and the output layer [34]. In the decomposition analysis, a strongly connected component composed of the largest number of nodes is defined as the nodes in the core layer. Nodes in the input layers can reach the core layer; however, those in the core layer cannot reach the input layer. Further, the nodes in the core layer can reach the nodes in the output layers but those in the output layer cannot reach the core layer. Herein, one-hot vector encoding was employed to represent the analysis results from bow-tie decomposition. For example, for a node assigned to the core layer, the value of the “core layer” of the node is equal to 1 while the value of the “input layer” and the “output layer” is equal to 0.

Indispensable nodes [35, 36]: Liu et al. developed a controllability analysis method to identify the minimum number of driver nodes (ND) that must be controlled to modulate the dynamics of the entire network [36] (i.e., they used the Hopcroft–Karp “maximum matching” algorithm [38] to identify the minimum set of driver nodes [36]). Indispensable nodes that are potential key player nodes and are sensitive to structural changes in a network are obtained from controllability analysis (i.e., removal of an indispensable node increases the ND in the network [35]). Vinayagam et al. reported that indispensable proteins in the human PIN tend to be targets of mutations associated with human diseases and human viruses [35]. One-hot vector encoding was also used to represent the analysis results of indispensable nodes. For example, for an indispensable node, the value of the binary variable of the node is equal to 1 while that for a non-indispensable node is equal to 0.

For network analysis, we employed the igraph R package [39].

Oversampling by the SMOTE algorithm

In order to prepare a class-balanced dataset for building binary classifier, we used a state-of-the-art sampling method, SMORT [40], to generate this class-balanced dataset to construct a binary classifier for drug target prediction. The SMOTE algorithm synthetically creates more cases in the minority class. Thus, the algorithm selects k nearest neighbours of a case in the minority class and randomly selects a point along the line that connects them.

The selected point is used as an additional case in the minority class. We used the Python module, imbalanced-learn[41], to perform oversampling based on the SMOTE algorithm. In addition, we used $k = 2$ to carry out SMOTE-based oversampling.

Binary classifier model based on Xgboost

To build a binary classifier for drug target prediction, we used Xgboost, which is the most efficient implementation of the gradient tree boosting algorithms [42]. The algorithm generates a large number of weak learners and builds a strong learner that exists as an ensemble of the weak learners. In the boosting step, the algorithm continues to update the weak learners by correcting the errors made by previous learners. Thereafter, the algorithm aggregates the predictions from the weak learners to make the final prediction by minimizing the loss with a gradient descent algorithm.

To build the Xgboost algorithm-based binary classifiers, we used the XGBClassifier and scikit-learn [43] python modules. The XGBClassifier has several parameters. Briefly, we employed the following values for each parameter (please see manual for XGBClassifier module [44] for details): learning_rate = (0.01, 0.1, 0.5), max_depth = (1, 2, 3, 5, 10), n_estimators = (100), gamma = (0, 0.3), booster = (“gblinear”), objective = (“binary:logistic”), reg_lambda = (0, 0.1, 1.0), and reg_alpha = (0, 0.1, 1). For the other parameters, we used a default value. To evaluate the binary classifier models and optimize the parameters of the models, we performed 5-fold cross validation.

Pathway enrichment analysis

To identify the pathways that are significantly associated with the putative targets inferred by our computational framework, we used WebGestalt web tool [45]. WebGestalt uses over-representation analysis (ORA) to statistically evaluate overlaps between the gene set of interest and a pathway [46]. In the analysis, the number of overlapped genes between the gene set of interest and a pathway is first counted. Thereafter, a hyper-geometric test is used to determine whether the pathway is over- or under-represented in the gene set of interest (for each pathway, the p value and FDR are calculated based on the overlap). Based on the ORA, we examined the pathways in Reactome, KEGG, and GO biological processes. The pathways with an $FDR < 0.05$ were regarded as significant pathways associated with the gene set of interest.

Results

Network embedding: deep autoencoder-based dimensional reduction of PIN

We obtained the directed human PIN from [23]; this PIN is composed of 6338 genes and 34,814 interactions (see the “Methods” section for details). Thereafter we gener-

ated an adjacency matrix for the human PIN. Elements in the matrix are represented as a binary value (i.e., 1 or 0 in position (i, j) denotes whether or not protein j is a downstream interacting partner of protein i). The resultant matrix is composed of 6,338 rows and 6,338 columns. Each row in the matrix presents the interacting pattern for each gene and used as features of the gene. Because there are 6,338 genes in the PIN, the features for each gene are of 6,338 dimensions (i.e., a gene is characterized by 6338 dimensional features based on the PIN data).

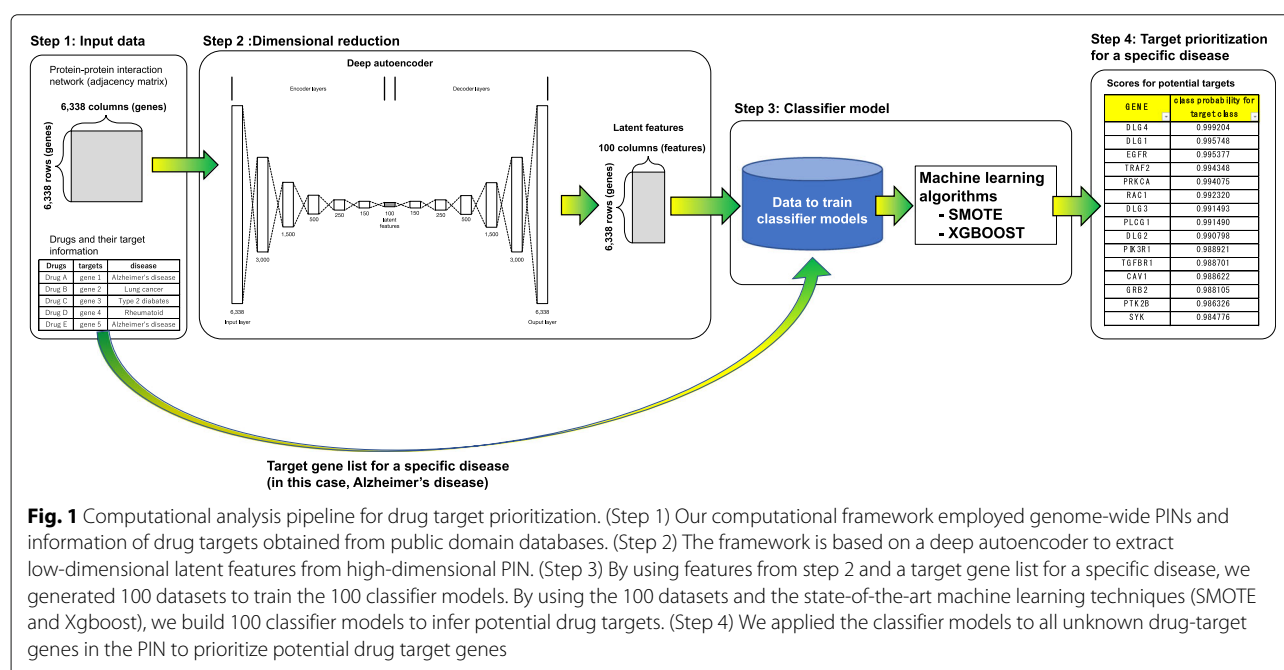
As shown in Fig. 1, to map the high dimensionality of the features (6338 dimensions) for each gene onto low dimensional features, we built and used a deep autoencoder. The deep autoencoder is composed of 7 encoder layers (6338-3000-1500-500-250-150-100) and symmetric decoder layers (100-150-250-500-1500-3000-6338) (see Fig. 1). In the deep autoencoder, layers are fully connected and weights of links connecting layers are optimized by minimizing binary cross-entropy loss between values of nodes in the input layer and those in the output layer (for details, see the “Methods” section). Following optimization, for each gene, we used the optimized deep autoencoder to map the high dimensionality of the original features (6,338 dimensional features) into low dimensionality (100 dimensional features) through the middle layer (layer with 100 nodes) in the network. Accordingly the resultant features for each gene are of 100-dimensional features.

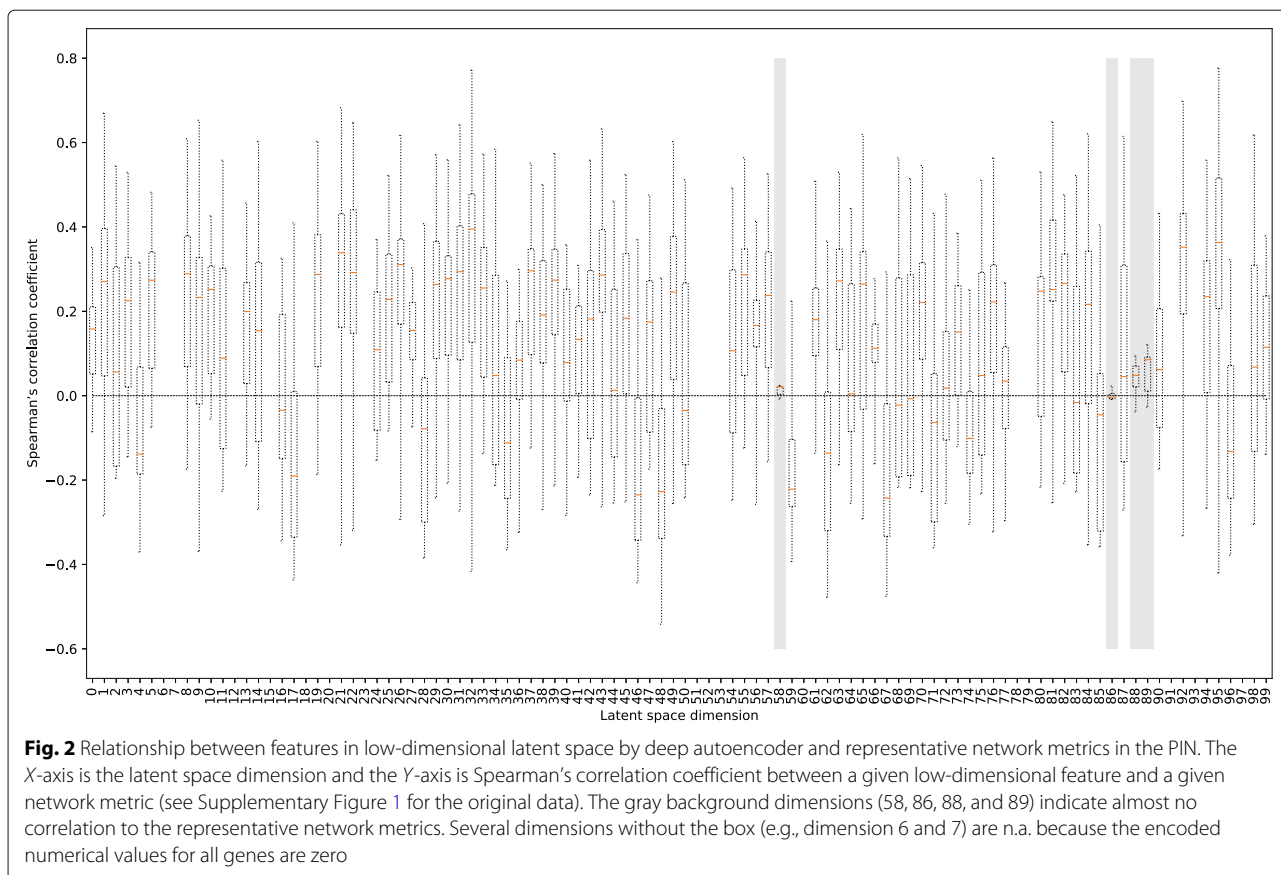
The low-dimensional latent space contains enough information to represent original high-dimensional human PIN. However, it is still unclear whether the low-dimensional features in the latent space can explain the topological and statistical properties obtained from the

representative network metrics. To examine this issue, we calculated nine representative network metrics for each gene in the PIN (e.g., indegree, outdegree, betweenness, closeness, PageRank, cluster coefficient, nearest neighbor degree (NND), bow-tie structure, and node dispensability, see the “Methods” section for details) and compared the metrics to the 100-dimensional features for the gene from the network embedding analysis (see Fig. 2 and the original data for Supplementary Figure 1). As shown in the figure, among the 100-dimensional feature, most of the features were correlated with the representative network metrics. Interestingly, several features (e.g., dimensions 58, 86, 88, and 89) did not correlate with the nine representative network-metrics (shown in gray background). Such findings indicate that the low-dimensional features from the network embedding analysis can capture not only the topological and statistical properties of network metrics but also information that cannot be obtained from analysis using representative network metrics.

Machine learning-based drug target prediction using the extracted feature from PIN

In this study, we treated the issue of drug-target prediction as a binary classification model. To construct a binary classifier for drug-target prediction, we generated a training dataset using the low-dimensional features extracted from PIN and public domain drug-target information. From the public domain drug-target database, we obtained known drug-target genes for Alzheimer's disease. Among the known targets, we could map 31 onto PIN. These 31 genes were further regarded as positive cases and the negative cases were selected from the





remaining 6,307 genes. We randomly selected 500 negative cases (genes) from the 6307 genes 100 times to build 100 datasets composed of 500 negative and 31 positive cases (genes). In the 100 datasets, each gene had 100 dimensional features that were obtained from deep autoencoder. Further, we employed the 100 datasets to build 100 binary classifier models to predict novel candidate targets for Alzheimer's disease.

The 100 datasets are class-imbalanced (e.g., 31 positive and 500 negative cases, respectively). Furthermore, classification using class-imbalanced data is biased toward the majority class. In the datasets, the number of "positive" cases was very small (i.e., only 31 positive cases were found in the datasets). These problems can be mitigated by using over-samplings that are often used to produce class-balanced training datasets from class-imbalance data. To generate class-balanced training datasets for binary classifiers, we used a state-of-the-art sampling method, SMOTE (Synthetic Minority Oversampling TEchnique) [40] that synthetically creates new cases in the minority class (in this study, "positive" case) (see the "Methods" section in details).

By using the class-balanced training datasets from SMOTE, we trained binary classifiers for drug target pre-

diction. The binary classifier models are based on the Xgboost algorithm which is the most efficient implementation of the gradient boosting algorithm [42]. The trained binary classifier models calculate two class probabilities for each gene based on 100 dimensional features (e.g., probability of "positive" and that of "negative"). Accordingly, a gene with a higher class probability of "positive" is more likely to be a member of the "positive" class.

To optimize the binary classifiers based on Xgboost for drug target prediction, we performed a grid search with 5-fold cross validations. Notably to avoid data leakage, we conducted data splits for cross validations before SMOTE-based over-sampling to generate class balancing training datasets. To evaluate the predictive performance of each parameter combination, we calculated area under the receiver operator characteristic curve (AUC ROC). The mean value of AUC ROC for the 100 binary classifiers with the optimal parameters was 0.661. Such result indicates that the 100 binary classifiers tend to assign a high class probability of "positive" for known drug-target genes of Alzheimer's disease. Therefore, unknown drug-target genes with a high probability of "positive" could serve as novel drug-targets for Alzheimer's disease.

Further, to infer the putative therapeutic targets for Alzheimer's disease, we used the mean value of the class probability of "positive" from the 100 binary classifier to prioritize the 6,307 genes (see Table 1 and Supplementary Table 1 for details); i.e., the unknown targets with a higher mean value of "positive" for the class probability (e.g., *DLG4* in Table 1 and Supplementary Table 1) are more likely potential novel drug targets. A total of 187 unknown drug-target genes had a mean value greater than 0.75 for a class probability of "positive" (see Supplementary Table 1). These 187 genes were thus regarded as putative novel target genes for Alzheimer's disease.

Pathway enrichment analysis of putative target genes

To deduce the potential target pathways for Alzheimer's disease, we determined the significant pathways that are associated with the 187 putative targets inferred using our computational framework (see Figs. 3, 4, and 5). The 187 putative targets were significantly associated with the pathways that control Alzheimer's disease mechanisms (e.g., cytokine-related signaling pathways and EGF receptor signaling pathway), especially those associated with

inflammatory mechanisms and the immune system. The innate immune system is a key component of Alzheimer's disease pathology [47]. In fact, continuous amyloid- β formation and deposition chronically activate the immune system, causing disruption of the microglial clearance systems [47]. Accordingly, the progression of Alzheimer's disease could be suppressed by modulating these pathways, especially the immune system and inflammation-related pathways, by targeting these putative target genes.

Inference of repositionable drug candidates

Networks connecting drugs, targets, and diseases could serve as useful resources for investigating novel indications for FDA-approved drugs, i.e., if target gene P is a putative target for disease A and is a known target gene of drug R for disease B, disease A may be a novel target disease for drug R (see Fig. 6). Thus, to infer the putative repositionable drugs and their potential target disease, we further examined the list of 187 predicted putative target genes (genes with a class probability of target class > 0.75 in Supplementary Table 1) from our computational framework and drug-target information across different

Table 1 Top 20 genes with the highest mean probability value for the "positive (drug target)" class

Gene	Mean probability
<i>DLG4</i>	0.99859
<i>PLCG1</i>	0.99775
<i>EGFR</i>	0.99758
<i>SYK</i>	0.99752
<i>PTK2B</i>	0.99617
<i>RAC1</i>	0.99585
<i>CAV1</i>	0.99579
<i>DLG1</i>	0.99512
<i>PIK3R1</i>	0.99500
<i>PRKCA</i>	0.99292
<i>KIT</i>	0.99224
<i>JAK1</i>	0.99154
<i>PTPN6</i>	0.98968
<i>CRKL</i>	0.98918
<i>SHC1</i>	0.98840
<i>NCK1</i>	0.98760
<i>ZAP70</i>	0.98750
<i>PTPN11</i>	0.98630
<i>DLG3</i>	0.98551
<i>PTK2</i>	0.98537
<i>DLG2</i>	0.98471
<i>IL2RB</i>	0.98328
<i>JAK2</i>	0.98299
<i>GRB2</i>	0.98278

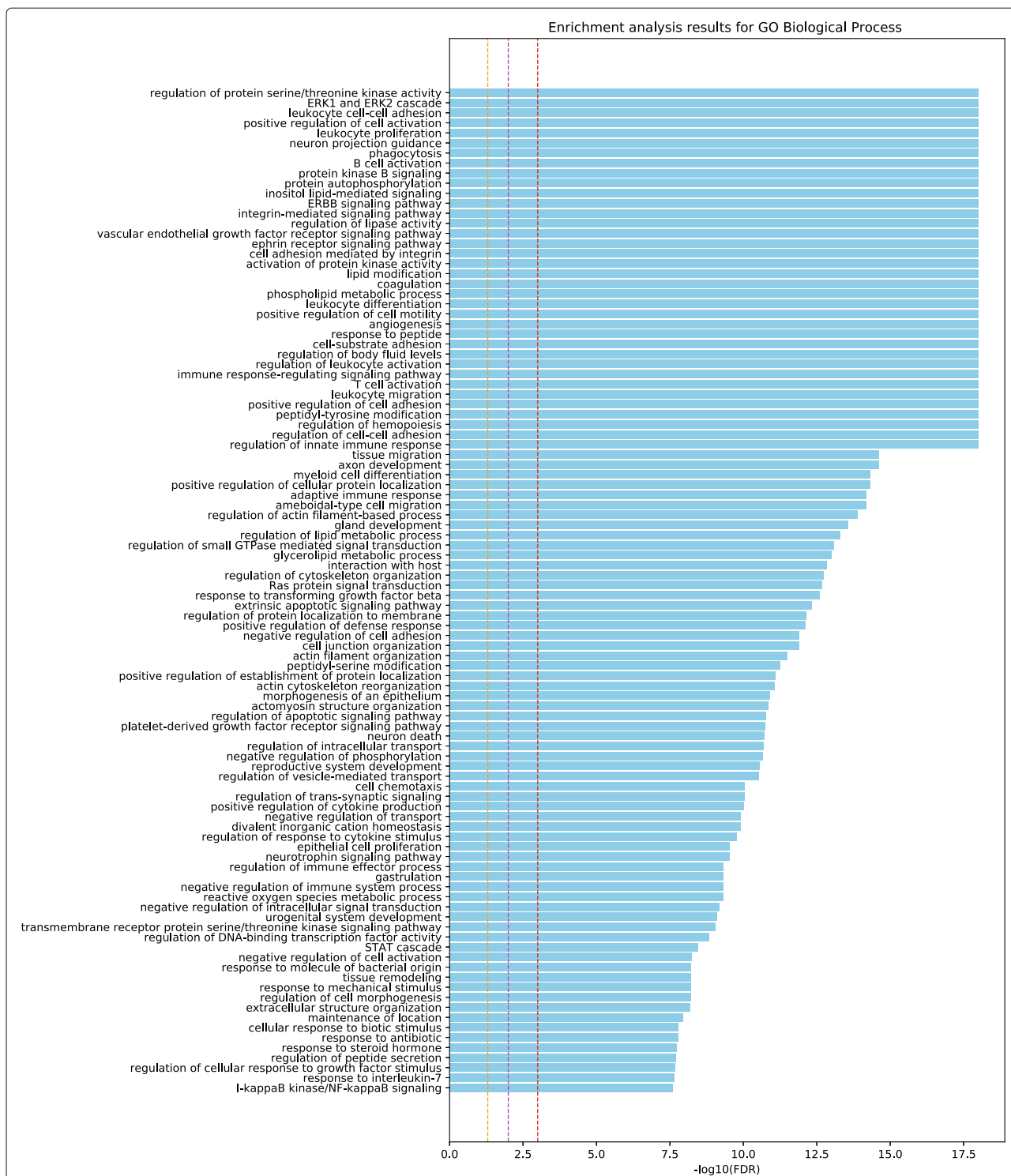


Fig. 3 Pathway enrichment analysis using GO biological database for the 187 putative targets from our computational pipeline for Alzheimer's disease. The names of the pathways are shown on the vertical axis, and the bars on the horizontal axis represent the $-\log_{10}(p \text{ value})$ of the corresponding pathway. Dashed lines in orange, magenta, and red indicate $p \text{ value} < 0.05, 0.01,$ and $0.001,$ respectively

diseases. If at least one target of a known drug is included among the 187 putative targets, the drug was regarded as a potential repositionable drug. As shown in Supple-

mentary Table 2, we inferred 244 candidate repositionable drugs for Alzheimer's disease. For each candidate repositionable drug, we calculated the number of overlapping



genes between the known targets of the drug and the 187 putative targets. Thereafter, we ranked the candidate repositionable drugs based on the number of overlapped genes. Among the predicted repositionable drug candidate, the top ranked candidates may be effective for the target disease. Table 2 lists the 20 highest ranked candidate compounds.

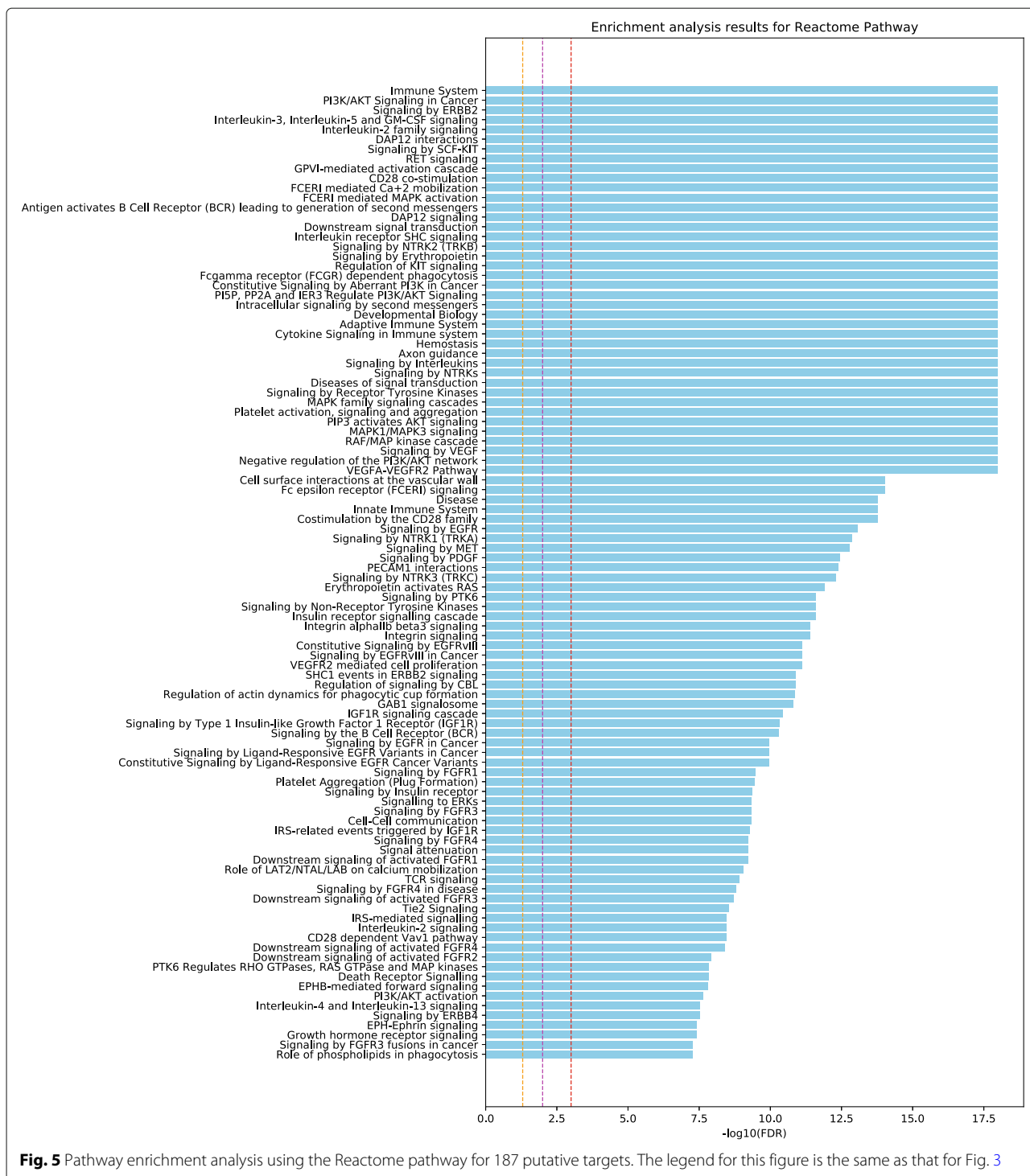
Discussion

Putative targets from our computational framework

Among the 187 putative targets from our analysis (see Supplementary Table 1), we investigated the top ranked

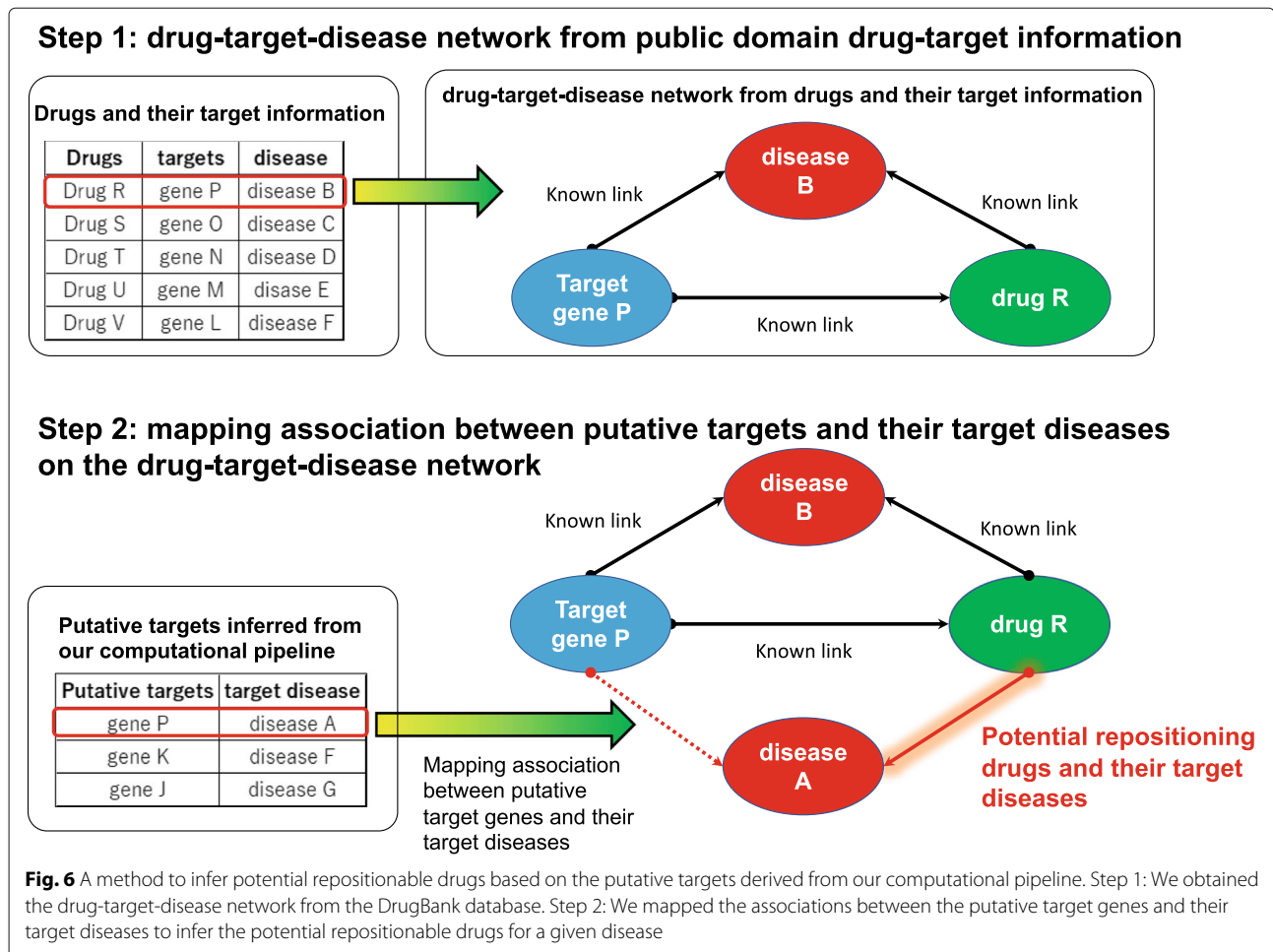
genes and found that several of these genes play an important role in the mechanism of Alzheimer's disease.

For example, the first ranked putative target, DLG4, encodes PSD95, which is a key protein for synaptic plasticity that is downregulated in under aged patients as well as patients with Alzheimer's disease. Recently, Bustos et al. demonstrated that epigenetic editing of DLG4/PSD95 ameliorates cognitions in model mice with Alzheimer's disease [48]. Thus, epigenetic editing of DLG4 may serve as a novel therapy for rescuing cognitive impairment induced by Alzheimer's disease.



EGFR is the third ranked putative target and is frequently upregulated in certain cancers. By employing an amyloid- β -expressing fruit fly model, Wang et al. demonstrated that the upregulation of EGFR causes memory impairment [49]. Furthermore, they administered several EGFR inhibitors (e.g., erlotinib and gefitinib) to

transgenic fly and a mouse model of Alzheimer's disease and found that the inhibitors prevented memory loss in both animal models. Based on these findings, they suggested that EGFR may be a therapeutic target for the treatment of amyloid- β -induced memory impairment.



RAC1, the sixth ranked putative target, is a small signaling GTPase, that controls different cellular processes, including cell growth, cellular plasticity, and inflammatory responses. Inhibition of RAC1 downregulates amyloid precursor protein (APP) and amyloid- β through regulation of the APP gene in hippocampal primary neurons [50]. RAC1 inhibitors can prevent cell death caused by amyloid- β 42 in primary neurons of the hippocampus and those of the entorhinal cortex [51]. Furthermore, based on an analysis of the protein-domain interaction network and experiments using drosophila genetic models, Kikuchi et al. demonstrated that RAC1 is a hub gene in the network and thus causes age-related alterations in behavior and neuronal degenerations [52]. The RAC1 gene could be a potential therapeutic target for preventing amyloid- β -induced neuronal cell death in Alzheimer's disease.

Spleen tyrosine kinase (SYK), the fourth ranked potential target, could modulate the accumulation of amyloid- β and hyperphosphorylation of Tau protein, which is associated with Alzheimer's disease [53]. Nilvadipine, an antagonist of the L-type calcium channel (LCC), inhibits the

accumulation of amyloid- β ; however, this does not occur because of LCC inhibition, but rather other mechanisms. Paris et al. demonstrated that the down-regulation of SYK exerts an effect that is similar to an enantiomer of Nilvadipine ((-)-nilvadipine) for the clearance of amyloid- β and reduction of Tau hyperphosphorylation [53]. Schweig et al. demonstrated that in mice with overexpressing amyloid- β , SYK activation occurred in the microglia. Further, neurite degeneration was found to increase because of the association between amyloid- β plaques and aging [54]. These researchers also demonstrated that in mice overexpressing Tau, SKY was activated in the microglia while misfolded and hyperphosphorylated Tau was accumulated in the hippocampus and cortex. Schweig et al. demonstrated that SYK inhibition induces Tau reduction in an autophagic manner [55]. Moreover, they demonstrated that SYK acts as an upstream target in the mTOR pathway and its inhibition induces Tau degradation by decreasing the activation of mTOR pathway.

The 5th ranked putative target, PTK2B, is a key gene in the mediation of synaptic dysfunction induced

Table 2 Top 20 candidate repositioning drugs for Alzheimer's disease

DRUG	Overlaps between known targets and predicted targets	# of overlaps
Regorafenib	RET; FLT1; KDR; KIT; PDGFRA; PDGFRB; FGFR1; TEK; NTRK1; EPHA2; ABL1	11
Tamoxifen	ESR1; ESR2; PRKCA; PRKCB; PRKCD; PRKCE; PRKCG; PRKCQ; PRKCZ; ESRRG	10
Ponatinib	ABL1; KIT; RET; TEK; FGFR1; LCK; SRC; LYN; KDR; PDGFRA	10
Dasatinib	ABL1; SRC; FYN; LCK; KIT; PDGFRB; EPHA2; BTK; FGR; LYN	10
Imatinib	PDGFRB; ABL1; KIT; RET; NTRK1; CSF1R; PDGFRA	7
Brigatinib	EGFR; ABL1; IGF1R; INSR; MET; ERBB2	6
Sorafenib	PDGFRB; KIT; KDR; FGFR1; RET; FLT1	6
Sunitinib	PDGFRB; FLT1; KDR; KIT; CSF1R; PDGFRA	6
Nintedanib	FLT1; KDR; FGFR1; LCK; LYN; SRC	6
Pazopanib	FLT1; KDR; PDGFRA; PDGFRB; KIT	5
Midostaurin	PRKCA; KDR; KIT; PDGFRA; PDGFRB	5
Resveratrol	ITGA5; ITGB3; SNCA; ESR1; AKT1	5
Diethylstilbestrol	ESR1; ESRRG; ESR2; ESRRA	4
Tofacitinib	TYK2; JAK2; JAK1; JAK3	4
Lenvatinib	FLT1; KDR; FGFR1; KIT	4
Foreskin fibroblast (neonatal)	FLT1; CSF2RA; PDGFRB; TGFB1	4
Baricitinib	JAK1; JAK2; PTK2B; JAK3	4
Foreskin keratinocyte (neonatal)	EGFR; CSF2RA; PDGFRA; TGFB1	4
Bosutinib	ABL1; LYN; SRC	3
Estradiol valerate	ESR1; ESR2; ESRRG	3

by amyloid- β in Alzheimer's disease [56]. Salazar et al. demonstrated that in a transgenic mice model of Alzheimer's disease, PTK2B deletion improves deficits in memory and learning functions as well as synaptic loss [56].

Although SOCS1 is the 78th ranked putative target, it modulates cytokine responses by suppressing JAK/STAT signaling to control inflammation in the CNS (central nerve system) [57]. Thus, SOCS1 may be a key therapeutic modulator in Alzheimer's disease.

GWAS and other sequencing technologies have identified over 20 genes that modify Alzheimer's disease risk. We obtained 29 genes listed in [58] and compared them with our 187 genes. PTK2B and INPP5D were listed as the overlap between the two gene sets. While as mentioned above, PTK2B is the 5th ranked strong candidate gene, INPP5D was the 68th ranked putative gene in the set of our 187 genes. INPP5D (Inositol Polyphosphate-5-Phosphatase D) is selectively expressed in brain microglia and likely a crucial player in Alzheimer's disease pathophysiology. Tsai et al. reported that INPP5D expression was upregulated in late-onset Alzheimer's disease and positively correlated with amyloid plaque density [59].

Collectively, these findings indicate that our computational framework could successfully identify key genes

that may be novel target candidates for Alzheimer's disease.

Promising repositionable drugs for Alzheimer's disease

In our computational drug repositioning analysis, our method predicted that tamoxifen (the second ranked candidate, see Table 2), an FDA-approved estrogen receptor modulator for the treatment of hormone-receptor-positive breast cancer patients, could serve as a potential drug target for Alzheimer's disease. As mentioned in Wise PM [60], estrogen therapy could protect neuronal cells from cell death by modulating the expression of key genes that inhibit the apoptotic cell death pathway. Based on a nation-wide cohort study in Taiwan, Sun et al. reported that patients with long-term use of tamoxifen exhibited a reduced risk of dementia [61].

Our method also predicted that bosutinib (the nineteenth ranked target), an FDA-approved tyrosine-kinase-inhibitor (TKI) drug (Bcr-Abl kinase inhibitor) for the treatment of Philadelphia chromosome-positive (Ph+) chronic myelogenous leukemia, may be a repositionable drug for Alzheimer's disease (see Table 2). Lonskaya et al. reported that Bosutinib combined with nilotinib systematically modulates with immune system in the CNS by inhibiting the non-receptor tyrosine kinase, Abl, to remove amyloid and decrease neuroinflammation [62].

Such findings indicates that TKIs, especially bosutinib, could be potential repositionable drugs for the treatment of early stage Alzheimer's disease.

Among the predicted repositionable candidates, 19 are immunosuppressive agents. These 19 candidates may include promising repositionable drugs for Alzheimer's disease; this is because of the important role played by inflammation in the mechanisms of Alzheimer's disease. Among the 19 candidates, dasatinib (the fourth ranked compound) may be the most promising candidate. Recently, Zhang et al. reported that senolytic therapy (a combination of dasatinib and quercetin) could reduce the production of proinflammatory cytokine and alleviate deficits of cognitive functions in Alzheimer's disease mouse models, via the selective removal of senescent oligodendrocyte progenitor cells [63, 64]. Furthermore, the combined therapy of dasatinib and quercetin is now registered in a clinical trial (ClinicalTrials.gov Identifier: NCT04063124).

One limitation of our method was that the process of identifying the putative target genes was dependent on the drug target gene database (i.e., the DrugBank in this research). This means that there is a possibility of bias in the known target genes because the DrugBank contains the existing therapeutic drugs and compounds which may have failed the clinical trials. However, we could overcome this limitation by adding new drug and target relationships, such as tau targeting compounds.

Conclusions

In this study, we developed a deep autoencoder-based computational framework and applied it to prioritize putative target genes for Alzheimer's disease. The method identified key genes (e.g., DLG4, EGFR, RAC1, SYK, PTK2B, SOCS1) associated with the disease mechanisms. Furthermore, by using the putative targets, we successfully inferred promising repositionable candidate-compounds (e.g., tamoxifen, bosutinib, dasatinib) for Alzheimer's disease. Our method could be a powerful tool for inferring potential repositionable drugs, especially those that could be used to treat Alzheimer's disease. Notably, our computational framework can be easily applied to the investigation of novel potential therapeutic targets and repositioning compounds for any disease. Accordingly, we anticipate that our method will be used by large pharmaceutical companies that house large volumes of their own non-public data.

Abbreviations

PIN: Protein-protein interaction network; AI: Artificial intelligence; PCA: Principle component analysis; ReLU: Rectified linear unit; NND: Nearest neighbor degree; ORA: Over-representation analysis

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13195-021-00826-3>.

Additional file 1: The original data of Fig. 2. Rows and columns represent the names of features in the low-dimensional latent space and names of the network metrics, respectively. The numeric value in a cell represents Spearman's correlation coefficient between a given low-dimensional feature and a given network metric (i.e., the correlation coefficient between the feature "Dimension 1" and the network metric "outdegree" is 0.67). Darker red (blue) indicates a higher (lower) correlation coefficient. Dimensions that are zero for all genes are denoted as n.a.

Additional file 2: A list of potential therapeutic targets for Alzheimer's disease.

Additional file 3: A list of all candidate repositionable compounds for Alzheimer's disease.

Acknowledgements

We would like to thank Editage (www.editage.com) for English language editing.

Authors' contributions

Conceived the experiments: ST, TH, AY, TN, SG, MK, SK, and HK. Designed the experiments and analyses: ST, TH, and AY. Performed the experiments: ST and TH. Analyzed the data: ST, TH, AY, and TN. Wrote the paper: ST, TH, AY, TN, SG, and MK. Supervised the research: ST, TH, HK, HA, and HT. The authors read and approved the final manuscript.

Funding

Not applicable.

Availability of data and materials

Documentation and source code are available on the author's Github site[65].

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Research Center for Advanced Science and Technology, The University of Tokyo, 4-6-1 Komaba, Meguro-ku, 153-8904, Tokyo, Japan. ²The Systems Biology Institute, Saisei Ikedayama Bldg. 5-10-25 Higashi Gotanda Shinagawa, 141-0022, Tokyo, Japan. ³Institute of Education, Tokyo Medical and Dental University, 20F, M&D Tower, 1-5-45 Yushima, Bunkyo-ku, 113-8510 Tokyo, Japan. ⁴SBX BioSciences, Inc. 1600 - 925 West Georgia Street, Vancouver BC V6C 3L2, Canada. ⁵Faculty of Pharmacy, Keio University, 1-5-30 Shibakoen, Minato-ku, 105-8512 Tokyo, Japan. ⁶Department of Genome Informatics, Graduate School of Medicine, Osaka University, 2-2 Yamadaoka, Suita, 565-0871, Osaka, Japan. ⁷Center for Mathematical Modeling and Data Science, Osaka University, 1-3 Machikaneyama-cho, Toyonaka City, 560-8531 Osaka, Japan.

Received: 23 November 2020 Accepted: 12 April 2021

Published online: 03 May 2021

References

1. Fleming N. How artificial intelligence is changing drug discovery. *Nature*. 2018;557(7706):55.
2. Rossi RL, Grifantini RM. Big data: challenge and opportunity for translational and industrial research in healthcare. *Front Digit Humanit*. 2018;5:13.

3. Luo J, Wu M, Gopukumar D, Zhao Y. Big data application in biomedical research and health care: a literature review. *Biomed Inform Insights*. 2016;8:31559.
4. Van Der Maaten L, Postma E, Van den Herik J. Dimensionality reduction: a comparative. *J Mach Learn Res*. 2009;10(66-71):13.
5. Ramlee R, Muda AK, Ahmad SSS. PCA and LDA as dimension reduction for individuality of handwriting in writer verification. In: 2013 13th International Conference on Intelligent Systems Design and Applications. IEEE; 2013. p. 104–8. <https://doi.org/10.1109/isd.2013.6920716>.
6. Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. *Science*. 2006;313(5786):504–7.
7. Sorzano COS, Vargas J, Montano AP. A survey of dimensionality reduction techniques. 2014. arXiv preprint arXiv:1403.2877.
8. Barabási A-L, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. *Nat Rev Genet*. 2011;12(1):56–68.
9. Hase T, Niimura Y. Protein-protein interaction networks: structures, evolution, and application to drug design. *Protein-Protein Interactions—Computational and Exp Tools*. 2012:2405–26. <https://doi.org/10.5772/36665>.
10. Hase T, Tanaka H, Suzuki Y, Nakagawa S, Kitano H. Structure of protein interaction networks and their implications on drug design. *PLoS Comput Biol*. 2009;5(10):1000550.
11. Rask-Andersen M, Almén MS, Schiöth HB. Trends in the exploitation of novel drug targets. *Nat Rev Drug Discov*. 2011;10(8):579–90.
12. Shim JS, Liu JO. Recent advances in drug repositioning for the discovery of new anticancer drugs. *Int J Biol Sci*. 2014;10(7):654.
13. Iorio F, Bosotti R, Scacheri E, Belcastro V, Mithbaokar P, Ferriero R, Murino L, Tagliaferri R, Brunetti-Pierri N, Isacchi A, et al. Discovery of drug mode of action and drug repositioning from transcriptional responses. *Proc Natl Acad Sci*. 2010;107(33):14621–6.
14. Cheng F, Liu C, Jiang J, Lu W, Li W, Liu G, Zhou W, Huang J, Tang Y. Prediction of drug-target interactions and drug repositioning via network-based inference. *PLoS Comput Biol*. 2012;8(5):1002503.
15. Hase T, Ghosh S, Palaniappan SK, Kitano H. Cancer network medicine. *Netw Med*. 2017:294–323.
16. Hase T, Kikuchi K, Ghosh S, Kitano H, Tanaka H. Identification of drug-target modules in the human protein–protein interaction network. *Artif Life Robot*. 2014;19(4):406–13.
17. Cui P, Wang X, Pei J, Zhu W. A survey on network embedding. *IEEE Trans Knowl Data Eng*. 2018;31(5):833–52.
18. Hamilton WL, Ying R, Leskovec J. Representation learning on graphs: methods and applications. 2017. arXiv preprint arXiv:1709.05584.
19. Ou M, Cui P, Pei J, Zhang Z, Zhu W. Asymmetric transitivity preserving graph embedding. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2016. p. 1105–14. <https://doi.org/10.1145/2939672.2939751>.
20. Wang X, Cui P, Wang J, Pei J, Zhu W, Yang S. Community preserving network embedding. In: Thirty-first AAAI Conference on Artificial Intelligence. 2017. <https://ojs.aaai.org/index.php/AAAI/article/view/10488>.
21. Wang D, Cui P, Zhu W. Structural deep network embedding. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2016. p. 1225–34. <https://doi.org/10.1145/2939672.2939753>.
22. Cao S, Lu W, Xu Q. Deep neural networks for learning graph representations. In: Thirtieth AAAI Conference on Artificial Intelligence. 2016. <https://ojs.aaai.org/index.php/AAAI/article/view/10179>.
23. Vinayagam A, Stelzl U, Fouille R, Plassmann S, Zenkner M, Timm J, Assmus HE, Andrade-Navarro MA, Wanker EE. A directed protein interaction network for investigating intracellular signal transduction. *Sci Signal*. 2011;4(189):8.
24. Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, Sajed T, Johnson D, Li C, Sayeeda Z, et al. Drugbank 5.0: a major update to the drugbank database for 2018. *Nucleic Acids Res*. 2018;46(D1):1074–82.
25. DrugBank. Detailed drug and drug target information. <https://go.drugbank.com/>. Accessed 8 Nov 2020.
26. Dahl GE, Sainath TN, Hinton GE. Improving deep neural networks for LVCSR using rectified linear units and dropout. In: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE; 2013. p. 8609–13. <https://doi.org/10.1109/icassp.2013.6639346>.
27. Kingma DP, Ba J. Adam: a method for stochastic optimization. 2014. arXiv preprint arXiv:1412.6980.
28. Chollet F, et al. Keras. 2015. <https://keras.io>. Accessed 23 Apr 2021.
29. Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, Devin M, Ghemawat S, Irving G, Isard M, et al. Tensorflow: a system for large-scale machine learning. In: 12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16); 2016. p. 265–283.
30. Human Genome Center. Supercomputer. <https://supcom.hgc.jp/english/>. Accessed 8 Nov 2020.
31. Brin S, Page L. The anatomy of a large-scale hypertextual web search engine. 1998. [https://doi.org/10.1016/s0169-7552\(98\)00110-x](https://doi.org/10.1016/s0169-7552(98)00110-x).
32. Watts DJ, Strogatz SH. Collective dynamics of ‘small-world’ networks. *Nature*. 1998;393(6684):440–2.
33. Newman ME. Assortative mixing in networks. *Phys Rev Lett*. 2002;89(20):208701.
34. Yang R, Zhuhadar L, Nasraoui O. Bow-tie decomposition in directed graphs. In: 14th International Conference on Information Fusion. IEEE; 2011. p. 1–5. <https://ieeexplore.ieee.org/document/5977625>.
35. Vinayagam A, Gibson TE, Lee H-J, Yilmazel B, Roesel C, Hu Y, Kwon Y, Sharma A, Liu Y-Y, Perrimon N, et al. Controllability analysis of the directed human protein interaction network identifies disease genes and drug targets. *Proc Natl Acad Sci*. 2016;113(18):4976–81.
36. Liu Y-Y, Slotine J-J, Barabási A-L. Controllability of complex networks. *Nature*. 2011;473(7346):167–73.
37. Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual web search engine. <http://infolab.stanford.edu/~backrub/google.html>. Accessed 8 Nov 2020.
38. Hopcroft JE, Karp RM. An $n^5/2$ algorithm for maximum matchings in bipartite graphs. *SIAM J Comput*. 1973;2(4):225–31.
39. Csardi G, Nepusz T. The igraph software package for complex network research. *InterJournal*. 2006;Complex Systems:1695.
40. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. Smote: synthetic minority over-sampling technique. *J Artif Intell Res*. 2002;16:321–57.
41. Lemaître G, Nogueira F, Aridas CK. Imbalanced-learn: a python toolbox to tackle the curse of imbalanced datasets in machine learning. *J Mach Learn Res*. 2017;18(17):1–5.
42. Chen T, Guestrin C. Xgboost: a scalable tree boosting system. In: Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining. 2016. p. 785–794. <https://dl.acm.org/doi/abs/10.1145/2939672.2939785>.
43. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, et al. Scikit-learn: machine learning in python. *J Mach Learn Res*. 2011;12:2825–30.
44. XGBoost. Python API Reference. <https://xgboost.readthedocs.io/en/latest/python>. Accessed 8 Nov 2020.
45. Wang J, Vasaikar S, Shi Z, Greer M, Zhang B. Webgestalt 2017: a more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. *Nucleic Acids Res*. 2017;45(W1):130–7.
46. Khatri P, Sirota M, Butte AJ. Ten years of pathway analysis: current approaches and outstanding challenges. *PLoS Comput Biol*. 2012;8(2):1002375.
47. Heneka MT, Golenbock DT, Latz E. Innate immunity in Alzheimer’s disease. *Nat Immunol*. 2015;16(3):229–36.
48. Bustos FJ, Ampuero E, Jury N, Aguilar R, Falahi F, Toledo J, Ahumada J, Lata J, Cubillos P, Henríquez B, et al. Epigenetic editing of the Dlg4/PSD95 gene improves cognition in aged and Alzheimer’s disease mice. *Brain*. 2017;140(12):3252–68.
49. Wang L, Chiang H-C, Wu W, Liang B, Xie Z, Yao X, Ma W, Du S, Zhong Y. Epidermal growth factor receptor is a preferred target for treating amyloid- β -induced memory loss. *Proc Natl Acad Sci*. 2012;109(41):16743–8.
50. Wang P-L, Niidome T, Akaike A, Kihara T, Sugimoto H. Rac1 inhibition negatively regulates transcriptional activity of the amyloid precursor protein gene. *J Neurosci Res*. 2009;87(9):2105–14.
51. Manterola L, Hernando-Rodríguez M, Ruiz A, Apraiz A, Arrizabalaga O, Vellón L, Alberdi E, Cavaliere F, Lacerda HM, Jimenez S, et al. 1–42 β -amyloid peptide requires PDK1/nPKC/Rac 1 pathway to induce neuronal death. *Transl Psychiatry*. 2013;3(1):219–219.
52. Kikuchi M, Sekiya M, Hara N, Miyashita A, Kuwano R, Ikeuchi T, Iijima KM, Nakaya A. Disruption of a Rac1-centred network is associated with

- Alzheimer's disease pathology and causes age-dependent neurodegeneration. *Human Mol Genet.* 2020;29(5):817–33.
53. Paris D, Ait-Ghezala G, Bachmeier C, Laco G, Beaulieu-Abdelahad D, Lin Y, Jin C, Crawford F, Mullan M. The spleen tyrosine kinase (Syk) regulates Alzheimer amyloid- β production and tau hyperphosphorylation. *J Biol Chem.* 2014;289(49):33927–44.
 54. Schweig JE, Yao H, Beaulieu-Abdelahad D, Ait-Ghezala G, Mouzon B, Crawford F, Mullan M, Paris D. Alzheimer's disease pathological lesions activate the spleen tyrosine kinase. *Acta Neuropathol Commun.* 2017;5(1):1–25.
 55. Schweig JE, Yao H, Coppola K, Jin C, Crawford F, Mullan M, Paris D. Spleen tyrosine kinase (Syk) blocks autophagic tau degradation in vitro and in vivo. *J Biol Chem.* 2019;294(36):13378–95.
 56. Salazar SV, Cox TO, Lee S, Brody AH, Chyung AS, Haas LT, Strittmatter SM. Alzheimer's disease risk factor Pyk2 mediates amyloid- β -induced synaptic dysfunction and loss. *J Neurosci.* 2019;39(4):758–72.
 57. Baker BJ, Akhtar LN, Benveniste EN. SOCS1 and SOCS3 in the control of CNS immunity. *Trends Immunol.* 2009;30(8):392–400.
 58. Naj AC, Schellenberg GD, (ADGC) ADGC. Genomic variants, genes, and pathways of Alzheimer's disease: an overview. *Am J Med Genet Part B Neuropsychiatr Genet.* 2017;174(1):5–26.
 59. Tsai AP, Lin PB-C, Dong C, Moutinho M, Casali BT, Liu Y, Lamb BT, Landreth GE, Oblak AL, Nho K. INPP5D expression is associated with risk for Alzheimer's disease and induced by plaque-associated microglia. *Neurobiol Dis.* 2021105303. <https://doi.org/10.1016/j.nbd.2021.105303>.
 60. Wise PM. Estrogen therapy: does it help or hurt the adult and aging brain? Insights derived from animal models. *Neuroscience.* 2006;138(3):831–5.
 61. Sun L-M, Chen H-J, Liang J-A, Kao C-H. Long-term use of tamoxifen reduces the risk of dementia: a nationwide population-based cohort study. *QJM Int J Med.* 2015;109(2):103–9.
 62. Lonskaya I, Hebron M, Selby S, Turner R, Moussa C-H. Nilotinib and bosutinib modulate pre-plaque alterations of blood immune markers and neuro-inflammation in Alzheimer's disease models. *Neuroscience.* 2015;304:316–27.
 63. Zhang P, Kishimoto Y, Grammatikakis I, Gottimukkala K, Cutler RG, Zhang S, Abdelmohsen K, Bohr VA, Sen JM, Gorospe M, et al. Senolytic therapy alleviates $\alpha\beta$ -associated oligodendrocyte progenitor cell senescence and cognitive deficits in an Alzheimer's disease model. *Nat Neurosci.* 2019;22(5):719–28.
 64. Curtis A. Targeting senescence within the Alzheimer's plaque. *Sci Transl Med.* 2019;11(488):4869.
 65. Github. AI based computational framework for drug development. <https://github.com/tsjshg/ai-drug-dev>. Accessed 8 Nov 2020.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

