scientific reports

Check for updates

OPEN

Drug repurposing for Parkinson's disease by biological pathway based edge-weighted network proximity analysis

Manyoung Han, Seunghwan Jung & Doheon Lee✉

Parkinson's disease is the second most frequent neurodegenerative disease, and its severity is increasing with extended life expectancy. Most of current treatments provide symptomatic relief; however, disease progression is not inhibited. There are multiple trials for treatments that target the causes of the disease but they were flawed. The mechanisms underlying neurodegenerative diseases are intricate, and understanding the interplay among the biological elements involved is crucial. These relationships can be effectively analyzed through biological networks, and the application of network-based analyses in the context of neurodegenerative disease treatment has gained considerable attention. Moreover, considering the significance differences in interactions between biological elements within the network is important. Therefore, we introduce a novel biological pathway based edge-weighted network construction method for drug repurposing in Parkinson's disease. The interaction found in multiple Parkinson's disease-related pathways is more significant than other interactions, and this significance is reflected in the network edge weights. Using the edge-weighted network construction method, we found a significant difference in the efficacy between known and unknown Parkinson's disease drugs. The method predicts drug-disease interactions more accurately than approaches that do not consider the significance differences among interactions, and the paths between the drug and disease within the network correspond to the drug's mechanism of action. In summary, we propose a network-based drug repurposing method using the biological pathway based edge-weighted network. Using this methodology, researchers can find novel drug candidates for the parkinson's disease and their mechanism of actions.

Parkinson's disease (PD), a brain disorder characterized by uncontrollable or unintended movement, is caused by impaired dopamine production. Parkinson's disease is the 2nd most frequent neurodegenerative disease worldwide. PD affects over 1% of the population over the age of 60, accounting for 6.1 million affected individuals and 21,296 deaths in 2016. As a result, in 2016, there are 6.1 million Parkinson's disease patients and 211,296 deaths because of the disease worldwide. The number of individuals with PD is predicted to increase to 13 million in 2040[1].

PD is strongly associated with aging. There are currently no disease-modifying drugs for PD; the current treatments only provide symptomatic relief. Additionally, PD drugs can produce serious side effects such as hallucinations and confusion[2]. Therefore, the identification of novel PD drugs is crucial.

Despite technological improvements, drug discovery demands a significant amount of time and money. On average, it takes 13 to 15 years and costs around 2 to 3 billion dollars[3]. Additionally, according to Eroom's Law, the number of new drugs in the market per billion dollars has decreased. Therefore, there is a need for the development of new drug discovery strategies that offer low attrition rates, low costs, and shorter development times. One promising approach is drug repurposing, which suggests a new direction for drug discovery that satisfies the needs of the pharmaceutical industry. Drug repurposing involves the identification of new indications for already approved or investigational drugs[3]. For instance, sildenafil was originally indicated for hypertension, and has been repurposed as an effective treatment option for erectile dysfunction.

Most diseases result from the breakdown of functional gene groups rather than a single gene malfunction[4]. PD results from gene malfunctions such as alpha-synuclein, Parkin, PINK1, DJ-1, and LRRK2[5]. Because multiple genes cause the disease, interactions between the genes should be analyzed to elucidate the disease mechanism.

Korea Advanced Institute of Science and Technology, Daejeon KS015, Korea. ✉email: dhlee@kaist.ac.kr

Analyzing the drug-gene-disease interactions using the human gene network is a useful strategy, as it reveals the interactions between disease-related genes and aids in the discovery of candidate drugs[6]. Among network-based drug repurposing methods, the proximity-based approach identifies the shortest path between drug targets and disease genes in the network. The hypothesis behind this method is that drug targets that have high proximity with the disease genes within network tend to be more effective in targeting disease compared to those that are farther away. Proximity is measured with the calculation of shortest path length between two nodes. Statistically, in the human gene network, drug targets exhibit more proximal interactions with targeting disease genes[7]. By leveraging the proximity between drug targets and disease genes, we can predict the therapeutical interactions between drugs and diseases. Unknown drug-disease interactions that exhibit statistically high proximity become potential targets for drug repurposing.

However, the human gene network used for proximity calculations does not take into account the significant difference between gene interactions[8]. The interactions found in multiple biological pathways should have higher significance than the interactions found in only one biological pathway. Therefore, the number of pathways in which the interactions are involved, and the significant differences between interactions need to be considered[9]. As a result, focusing solely on the shortest path between drug targets and disease genes may prioritize interactions that are not relevant. To accurately infer drug-disease interactions, it is crucial to consider the significance of interactions within the disease context.

Since the human network does not consider the significance of differences between gene interactions, predicting drug-disease interactions in PD using the proximity-based approachs is associated with limitations. First, the proximity-based drug repurposing approachs using the human gene network exhibits low statistical performance in predicting drug-disease interactions in PD[4]. Second, this approach fails to explain how drugs affect the diseases. For instance, bromocriptine, a well-known PD drug, targets dopamine receptors to enhance dopamine absorption and alleviate symptoms[10]. Logically, the path between the bromocriptine targets and PD-associated genes should describe the underlying dopaminergic and neurodegenerative interactions. Although the mechanism of action of bromocriptine is known to involve interactions between DRD4, GNG13, and PLCB3 within the dopaminergic pathway, as well as interactions between PLCB3 and PRKN within the neurodegenerative pathway, the proximity-based drug repurposing method using the human gene network does not consider these pathways. Instead, interactions between DRD4, ARNT, and RREB1 are chosen based on the shorter path (distance=2), despite the fact that this path cannot accurately describe the drug's mechanism of action which would require a longer path (distance=3). The shortest path identified by the human gene network that does not consider the significant differences between interactions can not effectively capture the drug's mechanism of action within the biological pathways.

Therefore, in this study, we aimed to emphasize the importance of considering significant differences among gene interactions when constructing human gene network for PD drug repurposing, based on the following considerations: 1. The frequency of interactions within the pathway database and 2. The significance of PD-associated gene interactions within the pathological context. The research flow is presented in Fig.1.

Results

Data statistics

The human interactome information was obtained from the CODA (Context-Oriented Directed Association) database, resulting in a human gene network with 22,119 nodes and 530,761 edges[11]. CODA database is relational database aggregating biological relationship information within the human body by extracting and standardizing data from various sources, including international public databases (including BioGrid, CTD, DiseaseConnect, EndoNet, PhenoGO, RegNetwork and so on), literature databases, and experimental data. It is the world's largest biological relationship information database and enables high-level analysis that considers all levels from molecules to cellular functions and diseases. Biological pathway information was obtained from the KEGG (Kyoto Encyclopedia of Genes and Genomes) database, and 307 biological pathways were utilized in the research[12]. The gene expression data from Parkinson's disease patients were collected from the GEO (Gene Expression Omnibus) database, comprising three series of data: GSE20292 (26 samples: 15 controls, 11 Parkinson's disease patients), GSE20291 (35 samples: 20 controls, 15 Parkinson's disease patients), and GSE20168 (29 samples: 15 controls, 14 Parkinson's disease patients). Based on the biological pathway information, 34,702 edge weights were modified among the 530,709 edges (6.1%). Drug, disease and their relationship data were collected from the DrugBank and OMIM databases[4]. A total of 238 drugs were found to have more than 20 target genes, and among them, 11 were used for PD treatment. Additionally, 105 genes were identified as PD-related genes. PD-related genes were referenced from Menche et al.[13]. This paper collected relationships between diseases and genes from the OMIM database[14] and UniProtKB/Swiss-Prot[15]. Significant relationships between diseases and genes were filtered using GWAS data from PheGenI[16].

Proximity between drug targets and disease genes

According to the proximity-based drug repurposing approach, known drug-disease interactions tend to be proximal within the human gene network compared to unknown drug-disease interactions[7]. We performed calculations to determine the proximity of PD-associated genes with both known and unknown PD drugs using different versions of the gene network: the network with no edge weight modification, the network with modified edge weights considering the number of pathways (STEP 1) and the network with modified edge weights considering both the number of pathways and the correlation of the pathways with PD (STEP 1, 2) (Fig. 2). GSEA(Gene Set Enrichment Analysis) scores were used to quantify the correlation of the pathways with PD. Statistical analysis revealed no significant proximity difference between known and unknown Parkinson's disease drugs within the non-weighted human gene network (p-value> 0.05). In contrast, a statistically significant difference was observed

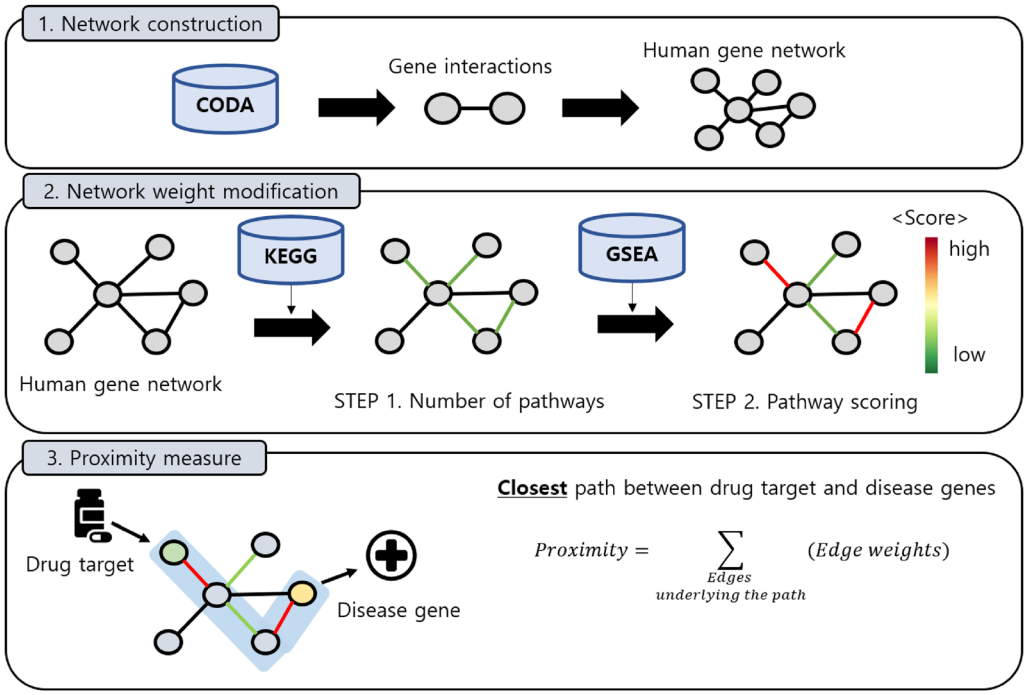


Fig. 1. Method overview for the biological pathway based edge-weighted human gene network construction **1**. Data on gene interactions were obtained from CODA(Context-Oriented Directed Association) database. **2**. Human gene network edge weights are modified based on the biological pathway information from KEGG(Kyoto Encyclopedia of Genes and Genomes) database and GSEA(Gene Set Enrichment Analysis). **3**. Proximity between drug and disease is the sum of edge weights underlying the closest path between drug target and disease genes.

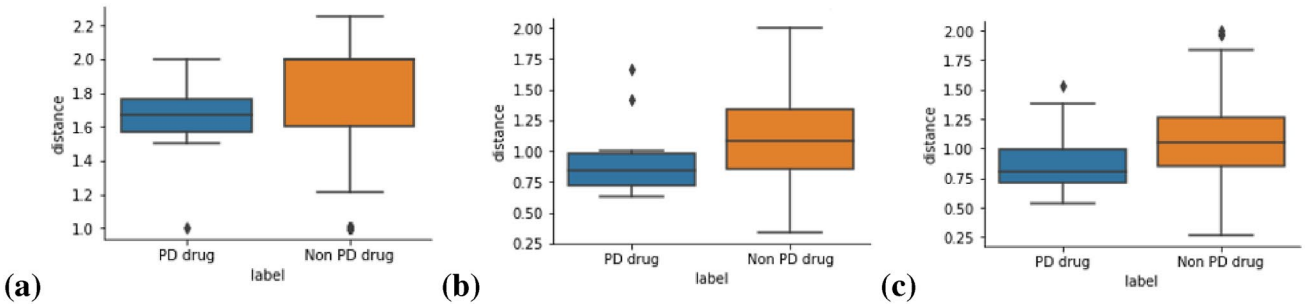


Fig. 2. Proximity difference between known and unknown Parkinson's disease drugs using (**a**) the non edge-weighted network(p-value = 0.13), (**b**) the edge-weighted network considering the number of pathways(p-value = 0.05), (**c**) the edge-weighted network considering the number of pathways and the pathways' correlation with the Parkinson's disease(p-value = 0.03).

using the weighted network considering the number of pathways (STEP 1; p-value = 0.05). Finally, the weighted network considering both the number of pathways and the correlation of the pathways with PD (STEP 1, 2) demonstrated the most statistically significant difference (p-value = 0.03). Consequently, it was observed that the proximity-based PD drug interaction inference was not appropriate using the non-weighted human gene network, as there was no statistical distinction between known and unknown PD drug proximity. In contrast, the human gene network with edge weight modification based on the biological pathway information showed statistical distinction between known and unknown PD drug proximity and the network yielded more accurate inference of PD drug interactions.

Performance measurement using the biological pathway based edge-weighted human gene network

The statistical significance of the biological pathway based edge-weighted human gene network in inferring drug-disease interactions was evaluated using Receiver Operating Characteristic (ROC) curves. The tendency of

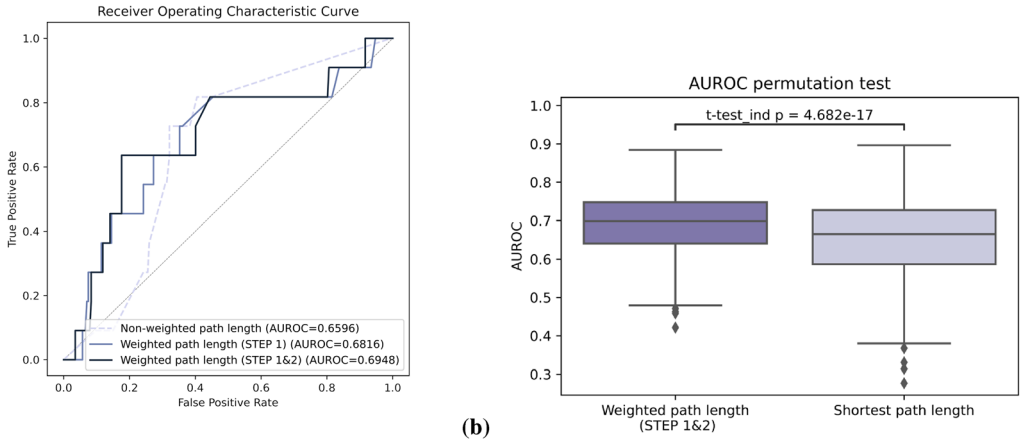
known drug-disease interactions to exhibit higher proximity rankings than unknown drug-disease interactions was assessed by comparing the Area Under the ROC Curve (AUROC) values (Fig.3(a)). The non-weighted path length analysis, calculated using the non edge-weighted human gene network, resulted in an AUROC value of 0.66. The weighted path length analysis considering only the number of pathways (STEP 1) showed a higher AUROC value (AUROC=0.68). Finally, the weighted path length analysis utilizing both the number of pathways and the pathway relationship with PD (STEP 1, 2) demonstrated the highest performance, with an AUROC value of 0.69. In conclusion, when employing the proximity-based method, incorporating biological pathway knowledge into the human gene network yielded superior performance in inferring Parkinson's disease-drug relationships. To demonstrate the statistically significant difference in AUROC between proximity calculations using non-weighted and weighted path length, we employed a permutation test. We calculated AUROC based on proximity values after randomly extracting the same number of drugs for PD drugs from the remaining drug set. Fig.3(b) demonstrates that the AUROC obtained from proximity calculations using weighted path length is significantly improved compared to calculations using non-weighted path length (p-value < 0.05).

In order to demonstrate the statistical reliability of drug-disease interaction inference using the biological pathway based edge-weighted human gene network, we utilized the recall measure among the top 50 drugs with high proximity values. That is, we calculated the ratio of known PD drugs among these top 50 drugs. Using the non edge-weighted human gene network, the recall value was 0.27, and only three PD drugs were identified among the top 50 drugs. However, when the number of pathways was incorporated into the network edge weights, the recall improved to 0.45, with five PD drugs appearing in the top 50 drugs. Finally, when the pathway relationship with PD was included, the recall further increased to 0.64, with seven PD drugs found among the top 50 drugs. These findings indicate that the inclusion of pathway information and its relationship with PD in the biological pathway based edge-weighted human gene network increases the number of known PD drugs within the top 50 drugs with high proximity values (Fig. 4).

To demonstrate the superiority of the relationship prediction performance between PD and drugs using the biological pathway based edge-weighted network construction method, we compared it with disease-drug relationship prediction models based on gene expression and biological network. The comparative models were SAveRUNNER, MNBDR, and OCTAD. SAveRUNNER utilizes proximity and similarity between drugs and diseases module on biological networks, MNBDR leverages gene connectivity in the network and gene expression in the context of diseases and drugs, and OCTAD predicts relationships between drugs and diseases by calculating the rank of gene expression for each gene in the context of diseases and drugs, thus identifying drugs that can reverse gene expression changes caused by diseases. For the construction of MNBDR and OCTAD models using gene expression, changes in gene expression due to PD were based on the GEO datasets (GSE20292, GSE20291, GSE20168) already utilized in the research, while changes in gene expression due to drugs were obtained from CMap LINCS 2020 Data an expansion upon the previous 2017 data (GSE92742) and contains $\tilde{3}$M gene expression profiles. As a result, in both AUROC and recall evaluation metrics, the biological pathway based edge-weighted network construction method demonstrated superior performance in predicting relationships between PD and drugs. Fig.5

Parkinson's disease drugs MoA analysis with biological pathway based edge-weighted human gene network

Despite the importance of understanding the mechanism of action(MoA), the proximity-based drug MoA analysis using the non edge-weighted human gene interaction network fails to explain how drugs affect diseases. In order to address this limitation, we conducted an analysis of PD drugs that exhibited the most significant increase in proximity compared to the non edge-weighted human gene network. We focused on the three drugs



(a) (b)

Fig. 3. (**a**) ROC curve using the non edge-weighted network, the edge-weighted network considering the number of pathways (STEP 1) and the edge-weighted network considering the number of pathways and the relationship of the pathways with the Parkinson's disease (STEP 1, 2). (**b**) Permutation test showing significantly improved AUROC using weighted path length.

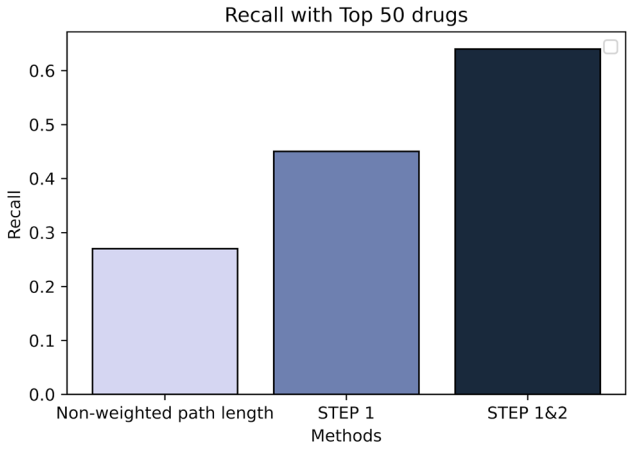


Fig. 4. Recall with top 50 highest proximity drugs using the non edge-weighted network, the edge-weighted network considering the number of pathways (STEP 1) and the edge-weighted network considering the number of pathways and the relationship of the pathways with the Parkinson's disease (STEP 1, 2).

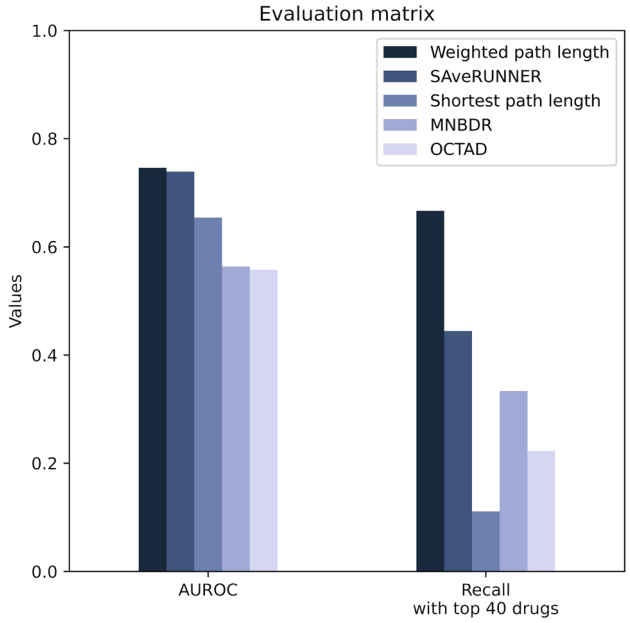


Fig. 5. AUROC and recall performance between PD and drugs using the biological pathway based edge-weighted network construction method(Weighted path length) and other prediction methods. The method demonstrated superior performance in predicting relationships between PD and drugs.

with the most significant increase in proximity and examined the shortest paths between the drug targets and PD-associated genes using the biological pathway based edge-weighted human gene network. Remarkably, the biological pathway based edge-weighted human gene network provided logical explanations for the mechanism of actions of the three drugs. We observed a remarkable coincidence between the improved proximity of these drugs and the shortest paths derived from the biological pathway based edge-weighted human gene network. These findings suggest that by utilizing the biological pathway based edge-weighted human gene network, we gain valuable insights into the mechanisms of action of drugs, providing a more comprehensive understanding of how the drugs impact PD.

Trihexyphenidyl exhibits the most improved proximity when utilizing the biological pathway based edge-weighted human gene network. The mechanism of action of trihexyphenidyl involves the inhibition of efferent impulses by affecting dopamine and M1 muscarinic (acetylcholine) receptors[17]. In contrast to the non edge-weighted human gene network, which suggested a path between the drug targets and disease genes (CHRM4 - RREB1), the biological pathway based edge-weighted human gene network provides a more accurate representation. The biological pathway based edge-weighted network suggests the path; CHRM4 - GNG12 - PLCB3 - PRKN(Fig. 6). The interaction between CHRM4 and GNG12 is found in the cholinergic synapse, while the

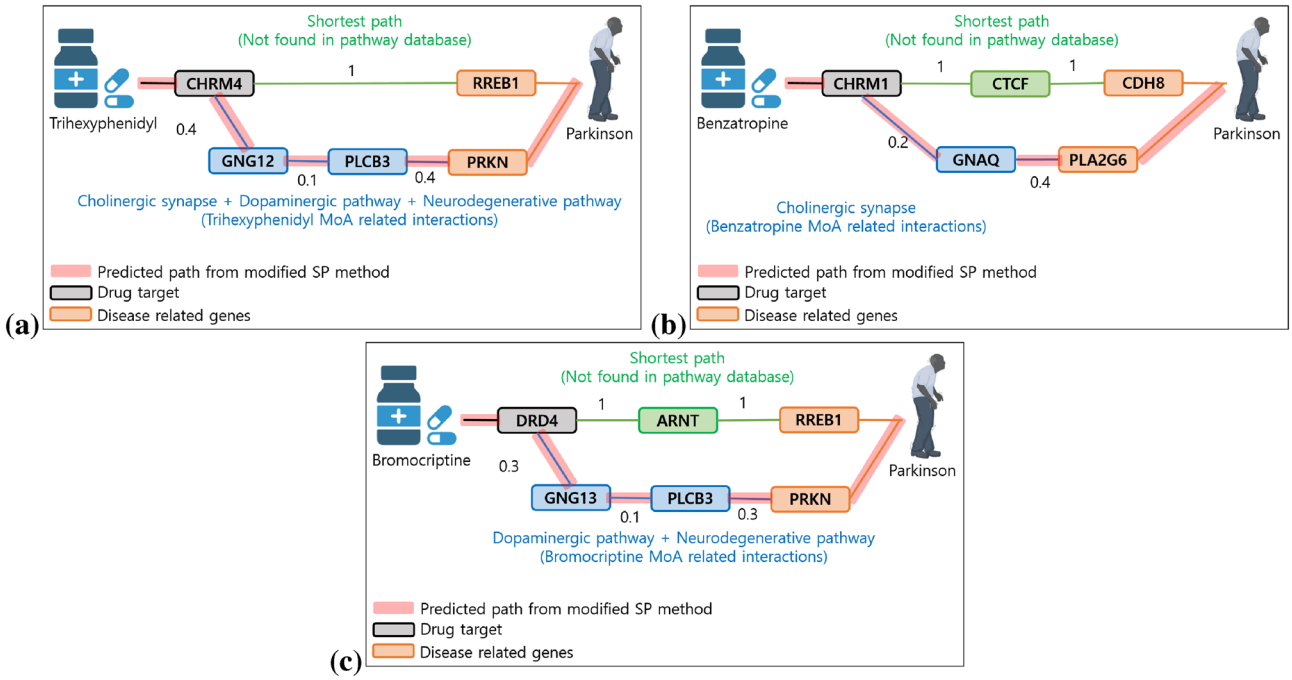


Fig. 6. Shortest path between (**a**) trihexyphenidyl, (**b**) benzatropine, (**c**) bromocriptine and the Parkinson's disease using the pathway-weighted network corresponded to the each drug mechanism of actions.

interaction between GNG12 and PLCB3 is found in the dopaminergic pathway. Finally, the interaction between PLCB3 and PRKN is found in the neurodegenerative pathway. These interactions align with the mechanism of action of trihexyphenidyl, providing a more comprehensive and accurate depiction of its mode of action.

Benzatropine exhibits the second most improved proximity when using the biological pathway based edge-weighted human gene network. Benzatropine acts as an antagonist of acetylcholine and muscarinic receptors, effectively prolonging the action of dopamine[18]. In contrast to the non edge-weighted human gene network, which suggested the path CHRM1 - CTCF - CDH8, these interactions were not found in the KEGG pathways, thus failing to accurately describe the mechanism of action for benzatropine (Fig. 6). However, with the biological pathway based edge-weighted human gene network, the suggested path CHRM1 - GNAQ - PLA2G6 partially explains the drug's mechanism of action; the interaction between CHRM1 and GNAQ is found in the cholinergic synapse pathway, providing a more comprehensive understanding of benzatropine's effects in PD.

Bromocriptine exhibits the third most improved proximity when utilizing the biological pathway based edge-weighted human gene network. Bromocriptine functions as a dopamine receptor agonist[10]. However, the non edge-weighted human gene network suggests a shortest path consisting of DRD4 - ARNT - RREB1. None of these interactions were found in the pathway database, therefore this path could not sufficiently describe the mechanism of action for bromocriptine. Conversely, the biological pathway based edge-weighted human gene network proposes an alternative path: DRD4 - GNG13 - PLCB3 - PRKN. Notably, the interaction between DRD4 and GNG13, as well as the interaction between GNG13 and PLCB3, were identified within the dopaminergic pathway. Additionally, the interaction between PLCB3 and PRKN was found within the neurodegenerative pathway (Fig. 6). These pathways are directly relevant to the mechanism of action of bromocriptine. Thus, the biological pathway based edge-weighted human gene network suggests a more accurate path between the drug and the disease, effectively describing the drug's biological functions.

Drug repurposing with biological pathway based edge-weighted human gene network

We utilized the distance between drug targets and PD-associated genes to identify unknown drug-disease interactions with significantly high proximity. Through this approach, we filtered and identified unknown drug-PD interactions with improved and high proximity, indicating their potential for repurposing. We suggested 10 drugs with the most improved proximity and 10 drugs with the highest proximity as potential repurposable drugs for PD, along with supporting literature evidence (Table. 1). Improved proximity means the difference between the distance between drug and PD-associated genes measured from non-weighted and biological pathway based edge-weighted human gene network. For example, if proximity between drug and PD-associated genes in the non-weighted human gene network is 1 and in the biological pathway based edge-weighted human gene network is 0.4, improved proximity is 0.6. We analyzed highly improved 10 drugs and among them, 9 drugs showed statistically significant improvement (p-value < 0.05); Baclofen showed not significant improvement (p-value=0.05). First, previous evidence supports the use of drugs with highly improved proximity when utilizing the biological pathway based edge-weighted network for PD treatment. Losartan, originally an hypertension drug, was found to reduce neurodegeneration and behavioral symptoms in a PD rat model[19]. Tetracycline, an antibiotic used for susceptible infections, was proposed as a treatment for PD[20]. Gabapentin, primarily used for peripheral neuropathic

Drug	Proximity	Proximity difference	Parkinson's disease relation
Rufinamide	0.45	1.55	-
Losartan	0.46	1.54	Moradganjeh et al.[19]
Tetracycline	0.54	1.46	Mariza et al.[20]
Pregabalin	0.63	1.34	-
Tiotropium	0.63	1.34	-
Pilocarpine	0.63	1.34	-
Diphenhydramine	0.72	1.28	-
Gabapentin	0.73	1.27	Walter et al.[21]
Insulin aspart	0.77	1.23	-
Baclofen	0.77	1.23	Rodolphe et al.[22]

Table 1. Non-Parkinson's disease drugs with highly improved proximity.

pains, postherpetic neuralgia, and partial-onset seizures, showed improvements in rigidity, bradykinesia, and tremor associated with the PD[21]. Baclofen, a GABA-ergic agonist used for spasticity treatment, was suggested for PD management when used in conjunction with acamprosate, which is utilized for alcohol abstinence[22].

Second, drugs with the highest proximity based on the biological pathway based edge-weighted network were linked to PD (Table. 2). Colchicine, used to relieve gout pain, was shown to protect dopaminergic neurons in a rat model, suggesting its potential efficacy as a PD therapeutic[23]. Vincristine, used to treat various conditions such as acute leukemia and malignant lymphoma, was inversely associated with PD. When administered with adriamycin to acute leukemia patients, vincristine was found to induce parkinsonian-like symptoms as a side effect[24]. It is worth noting that the biological pathway based edge-weighted network did not consider the improvement or aggravation of PD symptoms, hence suggesting drugs with PD as a side effect. Finally, felbamate, used for epilepsy treatment, exhibited anti-parkinsonian potential in a rat model[25]. In summary, the proximity-based drug repurposing approach utilizing the biological pathway based edge-weighted network proposed potential PD drugs supported by previous findings offering reliable options for further investigation and potential treatment options.

Discussion

PD is a prevalent neurodegenerative disease, particularly affecting the elderly population. Given the increasing lifespan of individuals, it is crucial to focus disease prevention and treatment. However, currently available PD drugs only provide symptomatic relief and do not impede disease progression. Moreover, these drugs are associated with numerous side effects. Therefore, there is a pressing need for the discovery of new PD drugs. However, drug discovery is a challenging process, with high attrition rates.

To overcome these limitations, drug repurposing has gained attention as it offers reduced time and cost compared to traditional drug discovery approaches. Network-based approaches, which consider the interplay between multiple genes, are frequently employed in drug repurposing. Among these approaches, we utilized a proximity-based method that analyzed the distance between disease-related genes and drug targets on the human gene network. However, the non edge-weighted network did not adequately account for the significance of gene interactions within biological pathways. Additionally, interactions specific to PD should carry more significance compared to non-related interactions. These limitations resulted in poor performance of the proximity-based method using the non edge-weighted network, and the paths between drug targets and disease genes did not align with the drug's mechanism of action.

Drug	Proximity	Proximity difference	Parkinson's disease relation
Colchicine	0.26	0.74	Salama et al.[23]
Vinorelbine	0.26	0.74	-
Vincristine	0.26	0.74	Boranic et al.[24]
Docetaxel	0.42	0.58	-
Rufinamide	0.45	1.55	-
Losartan	0.46	1.54	Moradganjeh et al.[19]
Irbesartan	0.49	1.01	Sarah et al.[26]
Vinblastine	0.51	0.49	-
Tetracycline	0.54	1.46	Mariza et al.[20]
Felbamate	0.60	1.06	Kretschmer et al.[25]

Table 2. Non-Parkinson's disease drugs with highest proximity.

To address these limitations, we proposed the use of a biological pathway based edge-weighted human gene network that assigns significance to interactions based on biological pathway and disease context knowledge. This method prioritized interactions found in multiple PD related pathways. By utilizing the biological pathway based edge-weighted human gene network, we overcame the limitations of the non edge-weighted network.

The drug-PD interaction inference using the biological pathway based edge-weighted human gene network demonstrated a significant proximity difference between PD drugs and other drugs. Moreover, the biological pathway based edge-weighted human gene network yielded more accurate interaction inference compared to the non edge-weighted network, as evidenced by higher AUROC and recall values. Importantly, the paths between drug targets and disease genes identified using the human gene network corresponded to the drug's mechanism of action. Furthermore, non-PD drugs with high proximity using the modified network were supported by literature evidence showcasing their potential for PD treatment.

The methodology proposed in this paper is applicable to all diseases if there is a data of the differences in gene expression between disease and control samples and if disease-related genes are known. In other words, it has scalability in terms of applicability to other neurological or rare diseases besides PD. Moreover, in addition to FDA-approved drugs, the methodology could serve as a stepping stone for establishing a large-scale drug repositioning atlas knowing the target genes of new compounds, enabling the assessment of the therapeutic potential of the compounds in treating diseases.

This study is associated with limitations. We did not consider the direction and type of interactions, such as activation or inhibition, and whether the predicted repurposable drugs would improve or aggravate PD. For instance, vincristine, which exhibited high proximity to PD-associated genes, induced a parkinson-like syndrome when used in a child with acute leukemia[24]. Future research could involve filtering the types of interactions between repurposable drugs and PD based on additional biological knowledge, including side effects, to provide a more comprehensive understanding. The methodology utilized the human gene network as a backbone network for drug repurposing, which is a homogeneous network containing information for only one type of node. However, interactions in humans occur not only between genes but also among various biological entities such as proteins, miRNAs, and others. Biological functions arise from interactions among diverse entities, and abnormalities in these functions can lead to diseases. Previous researches employed heterogeneous networks that consider various biological entities including drugs-proteins[27] and diseases-miRNAs[28] to address this complexity. Expanding the scope of the research beyond a homogeneous human gene network to incorporate interactions among diverse biological entities at the molecule level (genes, metabolites), function level (biological processes, molecular functions), and phenotype level (diseases) through a heterogeneous network is expected to improve drug-disease interaction prediction ability.

Overall, the use of the biological pathway based edge-weighted human gene network in proximity-based drug repurposing for PD addressed the limitations of the non edge-weighted network and demonstrated promising results. This approach offers potential avenues for the discovery of new PD treatments.

Methods

Pathway-weighted human gene network construction

Pathway-weighted human gene network construction To incorporate biological pathway information into the human gene network, we formulated a two-step hypothesis. The first step involved assessing the frequency of interactions between genes using KEGG pathway analysis. We hypothesized that interactions that are frequently observed in these pathways would hold greater significance in the inference of drug-disease interactions compared to less frequently observed interactions. By considering the prevalence of these interactions within the pathways, we aimed to prioritize their relevance in the context of drug-disease interactions. In the second step, we focused on interactions that occur within pathways associated with PD. We hypothesized that biological pathways containing differentially expressed genes between control and PD samples are associated with PD. Gene expression data in control and PD samples are collected from three individual GEO studies; GSE20292, GSE20291, GSE20168. Considering batch effects, we normalized expression values of each sample using z-score normalization. Z-score normalization can help mitigate batch effects to some extent by adjusting the data distribution across batches. The expression differences of genes in these biological pathways were measured using GSEA, and pathways with differentially expressed genes exhibited high absolute GSEA scores. For example, dopaminergic synapse pathway, well-known PD-associated pathways, has high absolute GSEA scores compared with others. Therefore, in our research, PD-associated pathways refer to those pathways with high GSEA scores. By utilizing GSEA scores in edge weight modifications, the association of PD with biological pathways is reflected in the inference of PD drug interactions. We posited that interactions found within these specific pathways would increase the inference of PD drug interactions. By highlighting interactions occurring in pathways relevant to PD, we aimed to emphasize their significance and potential relevance in the context of drug interactions within the disease. By incorporating these two steps into our analysis, we sought to enhance the accuracy and relevance of drug-disease interaction inference, particularly in PD. This approach allowed us to leverage the wealth of biological pathway information available and provided a more comprehensive understanding of the underlying mechanisms and interactions involved in drug-disease relationships.

Edge weight modification based on the number of contained pathways

The first step of the edge weight modification combines the network approach with biological knowledge. Biological pathway knowledge is reflected by considering the number of pathways in which the gene interactions are involved. The significant interactions that are involved in many pathways have shorter distances than less significant interactions. To reflect this hypothesis, we set the weight of each edge between genes with the reciprocal of the number of pathways that the interaction between the genes is involved.

$$edge\ weight\ (gene\ A,\ gene\ B) = \frac{1}{Number\ of\ pathways\ in\ PW} \tag{1}$$

(where PW is biological pathways set the gene A-B interaction is involved)

Frequent occurrences in biological pathways result in assigning importance to edge weights. When there are numerous biological pathways, higher importance is assigned to edge weights, meaning shorter distances. If the path between drug and disease in the network contains relationships between genes found in biological pathways with many occurrences, the path has short distance and high proximity, indicating a significant relationship between the drug and disease. To achieve this, equation (1) is formulated to have smaller edge weights as the relationship between gene A and B is more frequently found in biological pathways. The collection of biological pathways containing relationships between gene A and B is defined as PW. If the number of pathways is 0, the denominator of the edge weight is zero and we set the denominator as 1+(Number of pathways in PW).

Edge weight modification using the correlation between pathways and the Parkinson's disease

In the first step of our edge weight modification, we aimed to combine the network approach with biological knowledge by incorporating information from biological pathways. We hypothesized that gene interactions that are involved in a greater number of pathways would be more significant and therefore have shorter distances in the network. To reflect this hypothesis, we assigned edge weights to the network based on the reciprocal number of pathways in which the edge's interacting genes are involved. Specifically, for interactions that are involved in multiple pathways, the edge weight would be higher, indicating a shorter distance between the genes. On the other hand, interactions involved in fewer pathways would have lower edge weights, reflecting a longer distance. In cases where an edge is not involved in any pathways (i.e., the number of pathways is zero), the denominator of the edge weight was set to 1 plus the number of pathways. This adjustment ensures that the edge weight is non-zero and allows for the inclusion of interactions that may not be explicitly captured in known pathways. By modifying the edge weights in this manner, we aimed to incorporate the significance of gene interactions based on their involvement in biological pathways. This approach allowed us to prioritize interactions that are more likely to be biologically relevant and improve the accuracy of drug-disease interaction inference within the network.

$$edge\ weight\ (gene\ A,\ gene\ B) = \frac{1}{\sum_{PW_i \in PW} GSEA\ score\ of\ PW_i} \tag{2}$$

(where PW is biological pathways set the gene A-B interaction is involved)

The collection of biological pathways containing gene A-B interaction is defined as *PW*, where

$$PW = PW_1, PW_2, \ldots PW_n$$

; n is the number of biological pathways the gene A-B interaction is involved. Each biological pathway has a GSEA score, which represents its relevance to PD and the higher the GSEA score, the more closely related the pathway is to the disease. As mentioned earlier, short distance (high proximity) indicates a significant relationship between the drug and disease. Accordingly, edge weight between gene A and B is reciprocal of the GSEA score sum of *PW*.

Proximity between drug and the Parkinson's disease

In the proximity-based analysis using the pathway-weighted protein-protein interaction network, the proximity between a drug and PD was determined by calculating the shortest path length between the drug targets and PD-associated genes. It should be noted that there may be multiple paths between a drug target and a disease gene. To select the shortest path, we consider the sum of the edge weights along each path. The edge weights represent the significance of the interactions based on their involvement in biological pathways, as discussed earlier. By summing the edge weights along each path, we identified the path with the smallest cumulative weight, indicating the shortest distance between the drug target and disease gene. Furthermore, multiple genes are involved in the pathogenesis of PD. To determine the closest path length, we selected the gene that has the shortest path length to the drug target. Compared to using the average path length across all disease genes, this approach has been shown to yield better performance. Several drugs have multiple targets; therefore, we calculated the path length with the smallest sum of edge weights for each target. Then, the proximity between the drug and PD was determined by the average of these path lengths across all drug targets. By considering the shortest path length between drug targets and the closest disease gene, and averaging the path lengths for drugs with multiple targets, we aimed to capture the most relevant and significant interactions in the pathway-weighted network and a more accurate measure of proximity between drugs and PD.

$$proximity = \frac{1}{n} \sum_{t \in T} \min_{g \in G} \left(\sum_{e_i \in e} e_i\ weight \right) \tag{3}$$

(where n=number of targets of drug, T=target set of drug, G=parkinson's disease gene set, e=edge set under shortest path between t and g)

Data availability

The datasets analysed during the current study are available in the GEO (Gene Expression Omnibus) database repository, GSE20292 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE20292), GSE20291 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE20291), GSE20168 (https://www.ncbi.nlm.nih.gov/geo/query/

acc.cgi?acc=GSE20168). CMap LINCS 2020 Data is available in https://console.cloud.google.com/bigquery?p=cmap-big-table&d=cmap_lincs_public_views&page=dataset.

Received: 31 January 2024; Accepted: 2 September 2024
Published online: 11 September 2024

References

1. Hernández-Parra, H. *et al.* Repositioning of drugs for Parkinson's disease and pharmaceutical nanotechnology tools for their optimization. *J. Nanobiotechnol.* **20**(1), 413 (2022).
2. Stott, S. R., Wyse, R. K. & Brundin, P. Drug repurposing for Parkinson's disease: the international linked clinical trials experience. *Front. Neurosci.* **15**, 653377 (2021).
3. Nosengo, N. Can you teach old drugs new tricks?. *Nature* **534**, 314–316. https://doi.org/10.1038/534314a (2016).
4. Guney, E., Menche, J., Vidal, M. & Barábasi, A.-L. Network-based in silico drug efficacy screening. *Nat. Commun.* **7**, 1–13. https://doi.org/10.1038/ncomms10331 (2016).
5. Sai, Y., Zou, Z., Peng, K. & Dong, Z. The parkinson's disease-related genes act in mitochondrial homeostasis. *Neurosci. Biobehav. Rev.* **36**, 2034–2043. https://doi.org/10.1016/j.neubiorev.2012.06.007 (2012).
6. Hwang, S. *et al.* Humannet v2: human gene networks for disease research. *Nucleic Acids Res.* **47**, D573–D580. https://doi.org/10.1093/nar/gky1126 (2019).
7. Yıldırım, M. A., Goh, K.-I., Cusick, M. E., Barabási, A.-L. & Vidal, M. Drug–target network. *Nat. Biotechnol.* **25**, 1119–1126. https://doi.org/10.1038/nbt1338 (2007).
8. Pham, M. & Lichtarge, O. Graph-based information diffusion method for prioritizing functionally related genes in protein-protein interaction networks. *Pac. Symp. Biocomput.* **439–450**, 2019. https://doi.org/10.1142/9789811215636_0039 (2020).
9. Hu, J. B. *et al.* Characteristic analysis of the pathway-based weighted network of hypertension-related genes. *Phys. A: Stat. Mech. Appl.* **533**, 122069 (2019).
10. Parkes, D. Bromocriptine. *N. Engl. J. Med.* **301**, 873–878. https://doi.org/10.1056/NEJM197910183011606 (1979).
11. Yu, H. *et al.* Coda: Integrating multi-level context-oriented directed associations for analysis of drug effects. *Sci. Rep.*[SPACE]https://doi.org/10.1038/s41598-017-07448-6 (2017).
12. Kanehisa, M. & Goto, S. Kegg: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.*[SPACE]https://doi.org/10.1093/nar/28.1.27 (2000).
13. Menche, J. *et al.* Uncovering disease-disease relationships through the incomplete interactome. *Science* **347**(6224), 1257601 (2015).
14. Hamosh, A., Scott, A. F., Amberger, J. S., Bocchini, C. A. & McKusick, V. A. Online mendelian inheritance in man (omim), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res.* **33**, D514–D517. https://doi.org/10.1093/nar/gki033 (2005).
15. Mottaz, A., Yip, Y. L., Ruch, P. & Veuthey, A.-L. Mapping proteins to disease terminologies: From uniprot to mesh. *BMC Bioinf.* **9**, S3. https://doi.org/10.1186/1471-2105-9-S5-S3 (2008).
16. Ramos, E. M. *et al.* Phenotype-genotype integrator (phegeni): Synthesizing genome-wide association study (gwas) data with existing genomic resources. *Eur. J. Hum. Genet.* **22**, 144–147. https://doi.org/10.1038/ejhg.2013.96 (2014).
17. Jilani, T. N., Sabir, S. & Sharma, S. Trihexyphenidyl. *StatPearls [Internet]* (2022).
18. Ahuja, A. & Abdijadid, S. Benztropine. *StatPearls [Internet]* (2020).
19. Ali, M., Ziai, S. A. & Roghani, M. Losartan pretreatment reduces neurodegeneration and behavioural symptoms in 6-hydroxydopamine induced unilateral rat model of parkinson's disease. *Pathophysiology* **20**, 243–248. https://doi.org/10.1016/j.pathophys.2013.10.001 (2013).
20. Bortolanza, M. *et al.* Tetracycline repurposing in neurodegeneration: focus on parkinson's disease. *J. Neural Transm.* **125**, 1403–1415. https://doi.org/10.1007/s00702-018-1913-1 (2018).
21. Gruenthal, W. L. O. M., Mueller, M. E. & Olson, W. H. Gabapentin for parkinsonism: A double-blind, placebo-controlled, crossover trial. *Am. J. Med.* **102**, 60–66. https://doi.org/10.1016/S0002-9343(96)00381-6 (1997).
22. Hajj, R. *et al.* Combination of acamprosate and baclofen as a promising therapeutic approach for parkinson's disease. *Sci. Rep.* **5**, 1–13. https://doi.org/10.1038/srep16084 (2015).
23. Salama, M. *et al.* Colchicine protects dopaminergic neurons in a rat model of Parkinson's disease. *CNS Neurol. Disorders-Drug Targets* **11**(7), 836–43 (2012).
24. Boranic, M. & Raci, F. A parkinson-like syndrome as side effect of chemotherapy with vincristine and adriamycin in a child with acute leukaemia. *Biomedicine/[publiee Pour l'AAICIG]* **31**, 124–125 (1979).
25. Kretschmer, B. D. Felbamate, an anti-convulsive drug, has anti-parkinsonian potential in rats. *Neurosci. Lett.* **179**, 115–118. https://doi.org/10.1016/0304-3940(94)90948-2 (1994).
26. Kamal, S. J. & Khadhim, H. M. Effects of irbesartan in induced parkinson's disease in mice. *Int. J. Pharmaceut. Qual. Assur.* **12**, 31–39. https://doi.org/10.25258/ijpqa.12.1.5 (2021).
27. Zhao, B.-W. *et al.* igrldti: an improved graph representation learning method for predicting drug-target interactions over heterogeneous biological information network. *Bioinformatics*[SPACE]https://doi.org/10.1093/bioinformatics/btad451 (2023).
28. Zhao, B.-W. *et al.* Motif-aware mirna-disease association prediction via hierarchical attention network. *IEEE J. Biomed. Health Inform.*[SPACE]https://doi.org/10.1109/JBHI.2024.3383591 (2024).

Acknowledgements

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT)(RS-2023-00262747).

Author contributions

Manyoung Han developed the method, performed the experiments, and analyzed the results. Seunghwan Jung reviewed the method and experiments. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-71922-1.

Correspondence and requests for materials should be addressed to D.L.