Knowledge Graphs and Explainable AI for Drug Repurposing on Rare Diseases

Pablo Perdomo-Quinteiro [a], Katy Wolstencroft[b], Marco Roos[c] and Núria Queralt-Rosinach [c,*]

[a] *Grupo de Aplicación de Telecomunicaciones Visuales, Escuela Técnica Superior de Ingenieros de Telecomunicación, Universidad Politécnica de Madrid, Avenida Complutense 30, 28040, Madrid, Spain*
[b] *Advanced Compute and Data Core, Amsterdam University Medical Centre, Meibergdreef 9, 1105 AZ, Amsterdam, The Netherlands*
[c] *Department of Human Genetics, Leiden University Medical Center, Albinusdreef 2, 2333 ZA, Leiden, The Netherlands*

Abstract. Artificial Intelligence (AI)-based drug repurposing is an emerging strategy to identify drug candidates to treat rare diseases. However, cutting-edge algorithms based on Deep Learning (DL) typically don't provide a human understandable explanation supporting their predictions. This is a problem because it hampers the biologists' ability to decide which predictions are the most plausible drug candidates to test in costly lab experiments. In this study, we propose *rd-explainer* a novel AI drug repurposing method for rare diseases which obtains possible drug candidates together with human understandable explanations. The method is based on Graph Neural Network (GNN) technology and explanations were generated as semantic graphs using state-of-the-art eXplainable AI (XAI). The model learns features from current background knowledge on the target rare disease structured as a Knowledge Graph (KG), which integrates curated facts and their evidence on different biomedical entities such as symptoms, drugs, genes and ortholog genes. Our experiments demonstrate that our method has excellent performance that is superior to state-of-the-art models. We investigated the application of XAI on drug repurposing for rare diseases and we prove our method is capable of discovering plausible drug candidates based on testable explanations.

Keywords: Rare Disease (RD), Knowledge Graph (KG), Drug Repurposing, Graph Neural Network (GNN), Explainable AI (XAI)

1. Highlights

- We demonstrated the use of graph-based explainable AI for drug repurposing on rare diseases to accelerate sound discovery of new therapies for this underrepresented group.
- We developed *rd-explainer* for rare disease specific drug research for faster translation. It predicts drugs to treat symptoms/phenotypes, it is highly performant and novel candidates are plausible according to evidence in the scientific literature and clinical trials. Key is that it learns a GNN model that is trained on a knowledge graph built specifically for a rare disease. We provide *rd-explainer* code freely available for the community.
- *rd-explainer* is researcher-centric interpretable ML for hypothesis generation and lab-in-the-loop drug research. Explanations of predictions are semantic graphs in line with human reasoning.
- We detected an effect of knowledge graph topology on explainability. This highlights the importance of knowledge representation for the drug repurposing task.

*Corresponding author. E-mail: n.queralt_rosinach@lumc.nl.

2. Introduction

Developing new drugs can be a challenging effort that often ends with the drug not being able to launch. Recent studies have shown that around 90% of drugs fail to be approved during their clinical development [1]. This leads to a fruitless expenditure of both time and money that will yield no financial returns. The situation is even worse in the case of rare diseases, as pharmaceutical companies may consider it risky to invest large amounts of resources into developing drugs that only a small percent of the population will need. Nonetheless, in total, human beings are affected by approximately 7,000 rare diseases, of which only 5% have an effective treatment [2]; and only in Europe between 27 and 36 million people suffer from rare diseases [3].

In this scenario, drug repurposing strategies have appeared as a possible approach to solve these issues. By reusing drugs that have already been approved, companies can avoid many of the costly and time-consuming steps of clinical trials. In this context, innovative approaches to drug repurposing, such as computational strategies and AI-driven methodologies, have emerged as promising solutions to address these challenges. Graph-based drug repurposing is another noticeable strategy that has gained attention in recent years. By constructing intricate networks of molecular interactions, genes, proteins, and diseases, this approach unveils hidden relationships and connections that might otherwise go unnoticed [4].

Still, many people remain skeptical about AI-driven decisions, specially Machine Learning (ML) and Deep Learning (DL), as many of them come with no explanation that can help to understand the reason why they should be trusted (also called black-box AI). This issue is especially significant in the healthcare field, where decisions may have an important impact on people's lives. Also, giving valid explanations can help researchers to point in the right direction in the generation of hypotheses that are testable in the lab and enable a solid knowledge discovery. Furthermore, the EU General Data Protection Regulation (GDPR) is requesting the AI industry to fulfill the 'right to explanation' [5]. This 'right to explanation' implies that when a decision is significantly affected by an automated process/algorithm, the individual can demand an explanation. In recent years, many different tools have appeared to try and cover this gap in the emerging explainable AI (XAI) research area [6–8].

In this study, we explore whether AI can be used to produce both predictions and explanations in computational drug repurposing for rare diseases and, if so, how helpful can these explanations be for hypothesis generation. The main objective of this work was to develop and implement a pipeline to find marketed drugs that can be used to treat the symptoms of a rare disease. Our approach is based on cutting-edge AI algorithms used in computational drug repurposing such as graph ML using knowledge graphs (KG) and graph neural networks (GNN), and XAI methodology to provide the explanations supporting the drug predictions made by the AI model. The approach was evaluated by selecting Duchenne muscular dystrophy (DMD) as a case study, a genetic disorder that is the most common form of muscular dystrophy [9]. We demonstrate the generalizability of our approach by applying the pipeline to different rare diseases.

3. Related work

3.1. Knowledge graph-based drug repurposing

The state-of-the-art of computational drug repurposing approaches make use of graph-based structures and AI techniques to find potential drug candidates. One of the main advantages of using graph structures is that they can easily incorporate information from different sources. This is especially important in the domain of rare diseases, where information is distributed and often scarce. The ability to integrate as much relevant data as possible can confer a significant advantage. An example of this would be the recent study of Al Saleem et al. [10], where a knowledge graph was used to discover drug candidates to treat COVID-19.

Different ML algorithms can be used to analyse knowledge graphs, including matrix factorization, random-walk approaches (node2vec [11]), geometric embeddings (DistMul [12]) and GNNs [13, 14], each one of them with its own advantages and disadvantages, see Table 1. In our study, we used a combination of random-walk approaches and GNNs as in contrast to other methods (like matrix factorization or geometric embeddings) they can easily

incorporate new information without the need of retraining the ML model. This is especially relevant in the field of drug repurposing where new information about drugs, genes and diseases is being published [15–17].

Table 1
Comparison of different graph-based machine learning methods in drug repurposing.

Method	Example	Advantages	Disadvantages	Applications
Matrix Factorization	ADA-GRMFC [18]	Captures global relationships between entities. Simple and interpretable. Effective for sparse graphs.	Computationally expensive for large graphs. Difficulty in incorporating new data without retraining.	Suitable for large-scale recommendation systems
Random-walks	node2vec [11]	Efficient for large graphs. Easy to implement. Can capture node proximity.	Limited to local information; misses long-range dependencies. Cannot utilize node features or graph structure.	Useful for tasks requiring efficient exploration of graph neighborhoods
Geometric Embeddings	DistMult [19]	Produces interpretable low-dimensional embeddings. Scalable and efficient for sparse graphs. Performs well on link prediction tasks.	Captures only local information, missing complex graph interactions. Cannot handle high-order relationships or complex structures.	Effective in link prediction or node classification tasks with relatively simple graph structures.
Graph Neural Networks (GNNs)	GraphSAGE [20]	Aggregates local and global node features. Inductive learning, generalizes to unseen nodes. Scalable and flexible.	Computationally intensive for large graphs. "Black-box" nature hinders interpretability. Sensitive to choice of aggregation function.	Ideal for large, dynamic graphs in drug repurposing, where new entities are constantly introduced.

3.2. Explainable AI on graph ML

One of the graph-based methods that can provide explanations of the predictions, also called local explanations, is (Graph)LIME [6], an adaptation of the popular and more general explainability method LIME [7]. The idea behind this method is the following: when trying to get an explanation for a given prediction, (Graph)LIME performs small perturbations to the features of nodes, and sees how the predictions vary with respect to the initial prediction. The more the prediction changes, the more the model is relying on that feature to obtain its prediction. This way, explanations in this model are given in the form of a set of node features. Among its drawbacks, this method can only be used in node classification tasks. Another explainability method is CRIAGE [8] where explanations are given as a set of rules.

Several other explainability methods have been proposed for Graph ML, including PGExplainer [21] and GRETEL [22]. PGExplainer generates explanations by learning a probabilistic mask over graph structures, making it more flexible in terms of capturing various graph features. GRETEL, on the other hand, is designed to provide global explanations, making it different from other methods that focus on local interpretability.

Finally, the method chosen in this work is GNNExplainer [23]. The insight of how this method works is the following: given an initial prediction (link prediction, node classification or graph classification) obtained through a GNN, GNNExplainer finds a subset of node features and edges that are responsible for the prediction. This subset is obtained by training an edge and node mask. This method was chosen as explanations are provided in the form of a subgraph that can be easily understandable. Additionally, it is a post-hoc XAI method, i.e., it is model-agnostic, which means that if more sophisticated GNNs are developed in the future, these new GNNs can be easily incorporated into the pipeline. Furthermore, as a post-hoc method, its explanations might not always be faithful to the model's decision-making process. If the GNN has been trained on noisy data, GNNExplainer may highlight irrelevant edges or nodes simply because they correlate with predictions. These features make it a popular method in the research community [24–26]. However, a major drawback is that it lacks consistency when obtaining explanations. This means that explanations on the same prediction can significantly change if running GNNExplainer several times.

Table 2
Summary of explainability methods in graph ML.

Method	Explanation Type	Main Drawback
GraphLIME [6]	Feature-based	Limited to node features
CRIAGE [8]	Rule-based	Requires rule extraction
GNNExplainer [23]	Subgraph-based	Inconsistent explanations
PGExplainer [21]	Probabilistic mask	Complexity in training
GRETEL [22]	Global explanation	Not applicable to local explanations

4. Methods

4.1. rd-explainer *method overview*

rd-explainer is the drug repurposing method we developed for rare diseases and its pipeline is illustrated in Figure 1. rd-explainer has three modules: the Knowledge Graph Construction module constructs a KG for the specific rare disease and drug repurposing task, the Prediction module trains a GNN model and predicts drug candidates for the rare disease symptoms, and the Explainer module computes the most important semantic subgraphs that explain the connection between the predicted drug and the symptom. Firstly, information related to the disease is gathered from different data sources: Monarch Initiative knowledge base [27] for disease pathology, and DrugCentral [28] and Therapeutic Target Database [29] for disease druggability. This information is then preprocessed and captured as a knowledge graph. Next, for each node in the graph a feature vector is obtained that will be used as input for the GNN model. This is done by making use of a method known as edge2vec [30] to consider the different edge semantics in the KG for node embedding learning. We used the version extracted from GitHub (accessed in 2021) [1]. The following step is to build and train the GNN model, which is done using the GraphSAGE framework for graph representation learning [20]. Next, link prediction is performed for each drug-symptom node embedding pair using the dot product as the scoring function. Finally, we produced prediction explanations as semantic graphs using GNNExplainer [23], a recent and, to our knowledge, one of the first XAI methods for obtaining explanations from GNN predictions.

4.2. Rare disease-specific drug repurposing knowledge graphs

4.2.1. Data sources

Data was obtained from three different sources: Monarch [27] (accessed in 2021), DrugCentral [28] (2021 version) and Therapeutic Target Database (TTD) [29] (November 8th, 2021 version). Monarch is a knowledge base built on semantic principles, unifying gene, variant, genotype, phenotype, and disease data across different species. Its primary aim is to establish links between genes and phenotypes, thereby facilitating computational exploration of human disease biology. Monarch was chosen as it contains curated information across different species. This way, because rare diseases are often less studied than common diseases, incorporating information from other species can maximize the amount of knowledge in the graph. However, Monarch does not specialize in drug information.

Drug information was incorporated from DrugCentral (drug-target information) and from Therapeutic Target Database (drug-disease information). DrugCentral is a comprehensive online database that provides information about approved drugs, active ingredients and other pharmaceutical products. One of its major features is that it is open source and its data is freely available to anyone. For this project, we made use only of the drug-target information (as it is the main piece of information that is not present in Monarch) downloaded as a tsv file from their site [29] [2]. Similarly, TTD is a database that specializes in drugs and their respective therapeutic targets. Once more, this database is freely accessible and its information can be easily downloaded in csv format (in this project,

[1] https://github.com/RoyZhengGao/edge2vec
[2] DrugCentral, *Download site*, accessed March 2022, https://unmtid-dbs.net/download/DrugCentral/2021_09_01/drug.target.interaction.tsv.gz

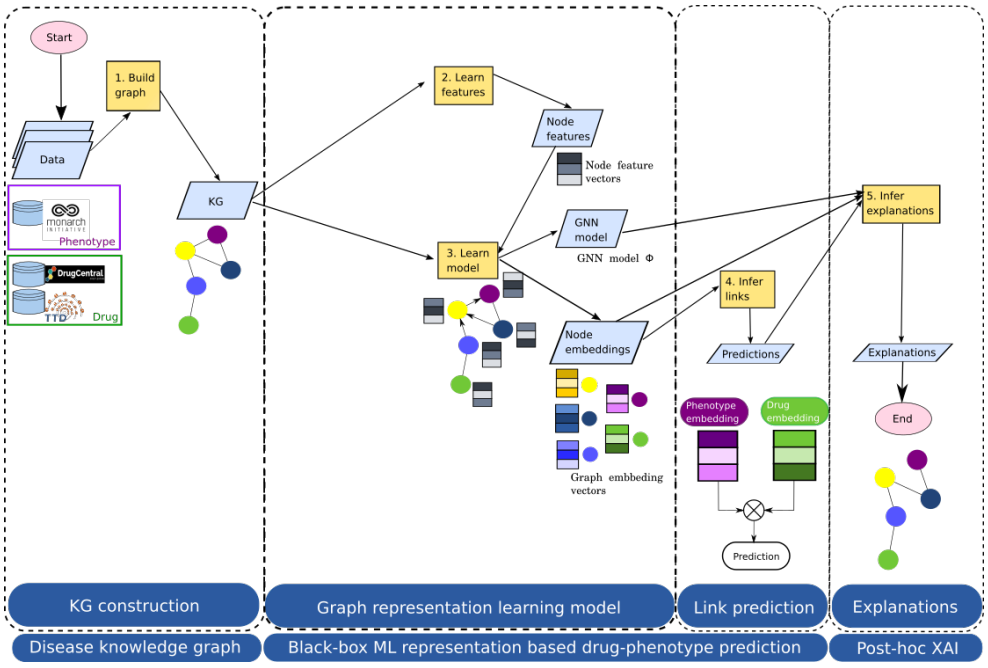

Fig. 1. *rd-explainer* drug repurposing method pipeline developed in this work.

we just made use of the drug-disease information [29] [3], once again because it is the information that is missing in Monarch).

4.2.2. Knowledge graph construction

To extract information from Monarch, the BioKnowledge Reviewer [31] tool was used. This tool was originally created to collect knowledge from several sources and create a knowledge graph that could be later used for hypothesis generation. It works by using several seeds (node identifiers (IDs)) as input to query the Monarch API and constructing the graph based on the neighborhood of those seeds. After introducing the seeds in the BioKnowledge Reviewer pipeline, the final output is the rare disease research question specific knowledge graph structured in two dataframes (stored as csv files). One of them contains a list of nodes with their respective name, IDs, semantic entity type, synonyms and description. The second file contains the list of edges, again containing the IDs of the entities participating in each link and other edge information such as type of edge, supporting evidence and reference date. Monarch was our main source of information, and so it served as a starting point to create the rest of the graph. This way, data from other data sources was modified to fit Monarch's standards by unifying the identifiers. Finally, the graphs were constructed using the networkx Python library [32]. With this library the dataframes extracted using BioKnowledge Reviewer were converted into a *Graph* object.

We integrated data in two different knowledge graphs to perform the experiments. Each one of them was constructed using different (number of) node seeds to extract information from Monarch. The first one (KG A) only uses two seeds: DMD seed (HGNC:2928), corresponding to the human gene that causes the disease; and DMD seed (MONDO:0010679), corresponding to the disease itself. The second graph (KG B), extends KG A by including as seeds all phenotypes of the rare disease (in total, 27 more seeds). The seeds used for the construction of each graph can be found in Tables S1 and S2. The idea of creating two different graphs is to find out if the performance of the model and the quality of the explanations increases by incorporating more (phenotypic) information.

[3] Therapeutic Target Database, *Download site*, accessed March 2022,https://idrblab.net/ttd/sites/default/files/ttd_database/P1-05-Drug_disease.txt

4.3. ML model and XAI

4.3.1. Node features

At this point, none of the nodes have any specific node features. It is possible to run a GNN relying only on graph information, i.e., network topology (this is done, for example, by using the node degree as graph feature); nonetheless, this resulted in a poor performance (results not shown). To increase the efficiency of the model, edge2vec was used to produce a specific embedding for each node that captures information about its neighborhood. edge2vec [30] is a tool that generates node embeddings based on the node neighborhood and types of edges connecting each node. After executing edge2vec, each node was given a unique feature vector. Since edge2vec is an unsupervised method that does not use task-specific labels, these embeddings serve as general-purpose representations of the graph structure rather than encoding direct knowledge of the downstream task. This approach ensures that the GNN still needs to learn task-relevant patterns, rather than relying only on the precomputed embeddings.

4.3.2. Data splitting

As any other machine learning task, data needs to be split into training, (validation) and test sets. However, when tackling a link prediction task, there are different ways to perform this split. In link or edge prediction tasks, edges can be divided into two groups: message passing edges and supervision edges. Message passing edges are the ones that will be used by our GNN to obtain the embeddings, while supervision edges are the ones that will be used to test the performance of our model [33, 34]. Additionally, when creating the supervision edges it is necessary to include negative examples by applying negative sampling. These negative sample edges are edges that are not present in our original graph, i.e., entities that it is known are not linked or there is no known link between them, and the idea is that the neural network is able to learn to distinguish true or positive edges from false or negative edges. In general, one negative edge is created for each true edge [33, 34].

In this work, we selected the all-graph transductive split [33, 34]. This method divides the data as follows: in the training set, the supervision edges and message-passing edges are the same. In the validation set, the message-passing edges are the same as those in the training set, while the supervision edges are different from the training supervision edges. Finally, in the test set, the message-passing edges consist of the validation edges, and the supervision edges are distinct from both the training and validation supervision edges.

This method is one of the standard settings when performing link prediction tasks, as the whole graph can be seen in all dataset splits [33]. The proportion used were 80% of edges used for training set, 10% for validation set and 10% for test set. The training set will be used to train the model, the validation set to select the best hyperparameters, and the test set to obtain the global performance of the model.

4.3.3. GNN model

We first utilized a GNN algorithm to learn vector representation embeddings for nodes in our knowledge graphs. Then, we applied these node embeddings for drug-phenotype link prediction. The GNN algorithm that we used in this work is called GraphSAGE [20]. GraphSAGE performs inductive graph representation learning by leveraging rich node attribute information. The main advantage that was brought by GraphSAGE is its scalability: instead of working with full batches (the whole graph is seen during the training) it works with mini-batches. Each mini-batch is a subset of computational graphs (a computational graph is the individual GNN that is built for each node) of N nodes. By applying this technique, the GNN can better manage larger graphs. The GraphSAGE model was created using the DeepSNAP library [34] to obtain the predictions. The hyperparameter optimization was performed using RayTune [35], as it is a model-agnostic library that allows to run multiple trials in parallel, reducing the training time. The list of hyperparameters that were needed to be tuned and the optimal values can be found in Table S5. In total, 30 models were created (each of them containing a random selection of parameters).

The final model consists in a GraphSAGE-based neural network that processes node embeddings through two graph convolutional layers using mean aggregation. The first SAGEConv layer transforms the input features into a 264-dimensional hidden representation, followed by batch normalization, LeakyReLU activation, and dropout (0.2) to prevent overfitting. The second SAGEConv layer maps the hidden representation to a 64-dimensional output space, which serves as the final node embeddings. Link prediction is performed by computing the dot product between the embeddings of node pairs. The model is trained for 150 epochs using the Binary Cross-Entropy with Logits Loss function and optimized with a learning rate of 0.07.

4.3.4. Drug-phenotype link predictions

The GNN model generates embeddings for individual nodes within the graph as its final output. By applying the dot product between distinct node pairs and applying a sigmoid function, we obtain a value that shows the likelihood of a link existing between those nodes. Consequently, we obtain dot products between each drug and every phenotype in the graph, and rank them in descending order. The top-ranked dot products are considered the most promising targets. Links that were already present in the graph were removed from the ranking.

4.3.5. Graph-based prediction explanations

We applied GNNExplainer to generate explanations for every drug-phenotype prediction. To do so, we adapted the pipeline code (from Pytorch geometric version 2.0.9) to generate explanations for the link prediction task, which was not implemented in authors' version [23] (see pseudocode in Algorithm 1 in the Supplementary material). However, this XAI algorithm has a problem of robustness in the explanations it produces [36] and, additionally, it may yield disconnected graphs affecting to the interpretability of explanations by domain-users. To solve this issue, we developed the following procedure. First, we make the assumption that a complete explanation is one that connects the two targeted nodes. If drug A can treat phenotype B, there must be some common pathway that allows A to interact with B. This way, the procedure starts by running GNNExplainer for several iterations. In each iteration, networkx is used to check if, in the subgraph generated by GNNExplainer, a path exists between both nodes. If no path is found, it continues with the next iteration; if it does exist, it stops iterating and that subgraph is considered to be the final explanation. If no subgraph is found that satisfies the 'pathway' condition, the last subgraph is returned as a possible explanation.

In total 7 phenotypes were selected to evaluate the explanations (Muscular Dystrophy (HP:0003560), Respiratory Insufficiency (HP:0002093), Arrhythmia (HP:0011675), Congestive Heart Failure (HP:0001635), Dilated Cardiomyopathy (HP:0001644), Cognitive Impairment (HP:0100543) and Progressive Muscle Weakness (HP:0003323)). These phenotypes were selected to cover all the main areas that are affected by the disease (muscular, respiratory, cardiac and intellectual symptoms). For each prediction obtained in these phenotypes (three drug predictions per phenotype), an explanation was obtained. This process was done for the predictions coming from KG A and for those coming from KG B. This makes a total of 42 explanations (21 for each graph).

Regarding the parameters of GNNExplainer, because the graphs are highly connected, explanations were generated by using the 1-hop neighborhood around the graph. Using a higher k-hop neighborhood is not recommended as the amount of nodes in the subgraph increases exponentially which can make it difficult to understand the explanation. This happens because both graphs are scale-free graphs, and thus, by increasing the number of hops there is a higher chance that a 'hub-node' is hit, and the number of nodes escalates exponentially (see Section 5.1 in the results).

Additionally, the maximum size of the explanations was set to 15 (this means that no more than 15 edges will be part of the explanation). This way, we will avoid obtaining too complex explanations with many edges that might be impossible to comprehend by researchers. This was done by selecting the edges whose contribution values are among the 15th highest values.

Finally, the maximum number of iterations was set to 10. In other words, if after 10 iterations GNNExplainer has not found an explanations that connects the drug candidate with the targeted phenotype it will conclude that no 'complete' explanation was found, and the last explanation produced by GNNExplainer will be the one that will serve as final answer. This parameter can be increased or reduced depending on the expectations of the researcher. A large number of iterations increases the chances of finding a complete explanation at the cost of more computational time. On the contrary, reducing the number of iterations reduces the computational time, which can be useful if a researcher wants to obtain explanations for a large number of predictions.

4.4. Evaluation and metrics

4.4.1. Evaluation of GNN model

Data. We used both graphs KG A and B. Data was split into three sets: training set, validation set and test set. **Baselines.** Our baselines include edge2vec [30], GraphSAGE [20], ComplEX [37], DistMult [19], and TransE [38]. **Evaluation metrics.** The Area Under the Precision-Recall curve (AUPRC) was used to validate and test the

performance of the model, as it has been shown that it leads to better precision when evaluating link prediction [39]. Additionally, we also computed the Area Under the ROC curve (AUROC), Precision, Recall and the F1-Score metrics - the harmonic mean of precision and recall.

Other evaluations were developed to further assess the performance of the model. These evaluations include the testing of different negative sampling sizes (n = 1, 5, 10 and 20) to determine the importance of keeping the data balanced. Additionally, both a regular 10-fold cross validation and a biased 7-fold cross validation were performed. The biased cross validation consists of the following: in each fold 4 phenotypes were removed from the training set, and it was observed how well the model was able to predict the links of the removed phenotypes.

4.4.2. Evaluation of explanations

The evaluation of explanations was done manually, following a two-step process. Firstly, they were classified as complete or incomplete explanations based on the appearance of connection between the drug and the phenotype. We developed a function to visualize the explanations as semantic graphs (see section 9.5 in the supplementary material for further details). This way, if the explanation contains a link between the drug and the phenotype it is considered to be a complete explanation. These explanations are considered the ones that are truly useful as they are the ones that can be easily understood and interpreted. On the other hand, explanations where there is no link between drug and phenotype (where there are two separate clusters) or where only one of the target elements (either the drug or the phenotype) is missing, are considered incomplete explanations. Several illustrative examples are provided in the supplementary material (See section 9.6).

During the second step, we evaluated explanations using an objective and a subjective approach. First, complete explanations were reviewed and a manual search was performed to check whether the explanation proposed by the model had been already described in the literature (objective evaluation). This process was only performed for those predictions that have supporting evidence in the literature and that were classified as complete explanations. The examination of the literature was performed using PubMed and Google Scholar during the first half of 2022. Finally, each explanation was evaluated with our own biological knowledge (subjective evaluation).

5. Results

5.1. Rare disease KG topology and representation for drug repurposing

We generated two different drug repurposing knowledge graphs for the Duchenne muscular dystrophy rare disease. KG A contains 10786 nodes, 93905 directed edges. The average node degree of the graph ($\frac{2\times numberofedges}{numberofnodes}$) is 10.83, being the node with the highest degree, the human DMD gene, with a total degree of 1683. The diameter of the graph was 6, meaning that the longest shortest path between two nodes is 6 (in other words, one can travel from one node to another in 6 steps or fewer). The final feature that was obtained is the clustering coefficient, which measures the extent to which a graph is clustered together. In a complete graph (where all nodes are connected to all nodes) this clustering coefficient is equal to 1, while in a tree-like graph this coefficient is equal to 0. In KG A this clustering coefficient is equal to 0.33. A summary of the features can be found in Table 3.

In the case of KG B (built from 29 nodes: KG A seeds extended by 27 phenotypes of DMD), the total number of nodes is 83665, with a total of 1984774 directed edges. The average degree in this case is of 34.43, being the node with the highest degree the physiological process 'Protein Binding' with a total degree of 4817. The diameter of the graph is of 7, which shows one of the features of scale-free networks: despite increasing the number of nodes 8 times and the number of edges 20 times, the diameter of graph B only increased one unit with respect to graph A. In this case, the clustering coefficient is equal to 0.48, showing that KG B is more clustered. Table 3 shows a summary of the features of both graphs.

Table 3
Table showing features of KG A and B.

Property	KG A	KG B
Number of Nodes	10786	83665
Number of Directed Edges	93855	1984774
Number of Undirected Edges	58435	1440418
Average Degree	10.83	34.43
Highest Degree	1683	4817
Diameter	6	7
Average Clustering Coefficient	0.33	0.48
Number of drugs	337	1565
Number of diseases	5419	25636
Number of drug-disease pairs	86	599

The schema of the knowledge graph, which is the same for KG A and KG B, can be seen in Figure 2 and shows how the 8 different node types interact with each other. The schema contains 24 and 29 different edge types for KG A and KG B respectively, which are not included in this figure for clarity, but are listed in the Supplementary material S3 and S4.

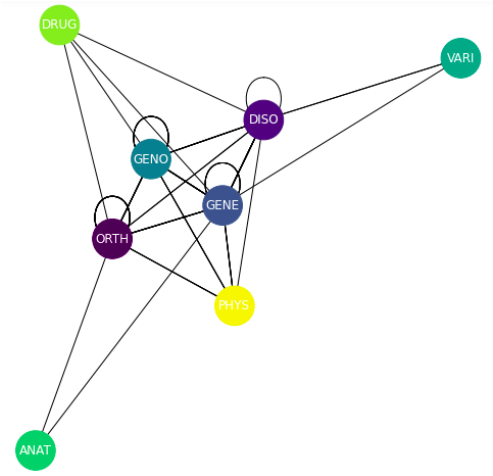

Fig. 2. Schema of the knowledge graph. Node types are: drugs or chemical compounds (DRUG), genes (GENE), symptoms/phenotypes or diseases (DISO), gene variants (VARI), genotypes (GENO), gene orthologs (ORTHO), anatomical structures (ANAT), and biological processes (PHYS).

5.2. *GNN model performance for rare disease specific drug repurposing*

In total, two GNNs were used, one trained on KG A and one trained on KG B. The hyperparameter optimization was developed using RayTune and the optimal values can be found in Table S5. These hyperparameters were obtained by training several GNN models (Random Search) on graph A; and were later used to train a GNN model on graph B.

Table 4
Precision, Recall and F1-Score obtained on each dataset, trained on each graph.

	Precision		**Recall**		**F1-Score**	
Dataset	**KG A**	**KG B**	**KG A**	**KG B**	**KG A**	**KG B**
Training	0.93	0.96	0.96	0.93	0.95	0.95
Validation	0.93	0.96	0.93	0.92	0.93	0.94
Test	0.93	0.96	0.93	0.92	0.93	0.94

To measure link prediction performance, the scores obtained were Precision, Recall and the F1-Score, and can be found in Table 4 (the threshold used was 0.8). We found that both models (the one trained with KG A and the one trained with KG B) yield to high performance (F1-Score = 0.93 and 0.94 in KG A and B, respectively). To visualise the performance of the link prediction task, the ROC curve of KG A and KG B obtained on the test set can be found in Figure 3 and Figure 4, repectively.

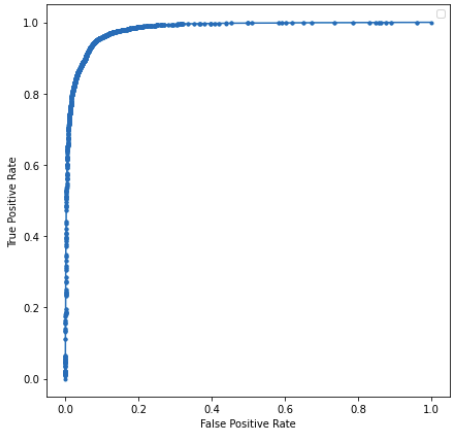

Fig. 3. AUROC on the test dataset using KG A.

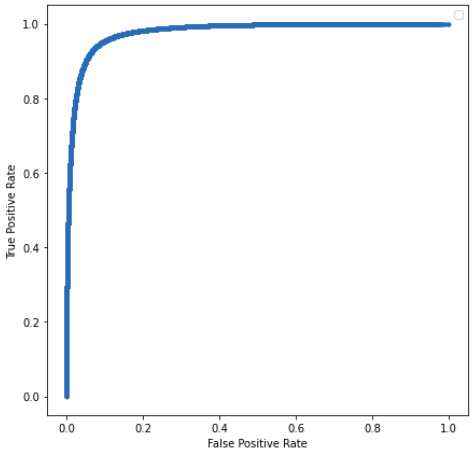

Fig. 4. AUROC on the test dataset using KG B.

5.3. Evaluating rd-explainer with state-of-the-art methods

Firstly, we evaluated our GNN model applying different strategies and compared its performance to the state-of-the-art graph embeddings used in drug repurposing methods. Then, we evaluated our approach based on its ability to predict drugs that are already reported in the literature for a new symptom or phenotype.

We performed a regular 10-fold cross-validation and a biased 7-fold cross-validation evaluation in KG A. The regular 10-fold cross-validation obtained an average AUPRC of 0.98 and an average AUROC of 0.98. For the biased 7-fold cross-validation, in each fold 4 symptoms (along with the edges connected to those symptoms) were removed from the training set. Then the performance of the model was tested on the removed symptoms. In this case, the average AUPRC was 0.75 and the AUROC was 0.8.

The performance of the pipeline was evaluated for a different number of negative edges. This evaluation was only performed in KG A due to the large increase in the number of edges in the evaluation tests (and the consequential increase in the computational time). The results can be seen in Table 5. It is seen that as the number of negative edges increases, the PR curve is affected while the ROC curve remains mostly intact, a result that has previously been reported [40].

Table 5
Performance as the number of negative edge samples increases. This results were obtained using KG A.

Number of negative edges	**Precision**	**Recall**	**F1-Score**	**AUROC**	**AUPRC**
1	0.95	0.95	0.95	0.99	0.99
5	0.94	0.90	0.92	0.94	0.94
10	0.93	0.91	0.86	0.86	0.85
20	0.93	0.57	0.62	0.82	0.70

Finally, the performance of rd-explainer (tested in KG A) was also compared to other state-of-the-art methods, including edge2vec, GraphSAGE, ComplEX, DistMult, and TransE. Our results can be seen in Table 6 and they revealed that rd-explainer outperformed all the other methods based on the different evaluation metrics measured.

Table 6
Prediction performance metrics comparing rd-explainer with other state-of-the-art graph embedding methods including edge2vec, GraphSAGE, CompleEX, DistMult and TransE. The best results are **highlighted**. In the headings, **P** stands for *Precision*, **R** for *Recall*, and **F1** for *F1-Score*.

Method	**P**	**R**	**F1**	**AUROC**	**AUPRC**
edge2vec	0.90	0.90	0.90	0.98	0.97
GraphSage	0.71	0.65	0.62	0.64	0.87
ComplEX	0.84	0.76	0.74	0.95	0.99
DistMult	0.93	0.93	0.92	0.95	0.98
TransE	0.88	0.87	0.87	0.95	0.95
rd-explainer	**0.95**	**0.95**	**0.95**	**0.99**	**0.99**

5.4. Drug predictions validation based on the scientific literature

We also evaluated the prediction performance based on the capacity of our method to discover marketed drugs already reported being used for a new phenotype. First, we listed for each of the 7 selected phenotypes the three drugs with the highest scores. Because the objective is to find new indications for drugs; if any of the reported drugs already appears in the graph as a treatment for the targeted symptom, this drug will be skipped and the next one with the highest score will be selected. For example, if aprindine is selected as the drug with the highest score to

treat arrhythmia, but the relation 'aprindine is a substance that treats arrhythmia' is already present in our graph, aprindine won't be reported as a possible drug candidate.

For each possible drug candidate, a literature search was carried out to find preliminary evidence if that drug had already been used to treat the symptom. If the drug was contraindicated to treat the symptom (or if it could cause the symptom) it was also annotated. Results regarding each drug candidate obtained using KG A can be found in Table S9. Additionally, Table 7 summarizes the amount of drugs (in percentage) that contained supporting evidence, contraindication evidence or no evidence at all. We found that only a fifth of the drug candidates had supporting evidence in the literature, and that the vast majority of the candidates (65.43%) did not have any evidence at all. There is a small percentage of them that are actually contraindicated to treat the targeted symptom/phenotype. Finally, the amount of supporting/contraindicating evidence can be found summarized in Table S7.

Table 7
Percentage of drugs containing supporting evidence, contraindication evidence or no evidence at all for both Graph A and B.

Property	**KG A**	**KG B**
Supporting Evidence	20.99 %	27.16 %
Contraindication Evidence	13.58 %	14.82 %
No Evidence	65.43 %	58.02 %

The same approach was followed in the case of KG B. Information regarding the drug candidates for each symptom (as well as the supporting evidence) can be found also in Table S10. Additionally, the percentage of drugs with supporting evidence, contraindication evidence or no evidence at all can be seen in Table 7. In this case, the number of drug candidates with evidence has increased with respect to the drug candidates obtained with KG A (27% in B vs 21% in A), and the number of drug candidates with no evidence has been reduced (58% in B vs 65% in A). The number of drug candidates with contraindications remains almost the same (13% in A vs 14% in B).

5.5. Evaluating drug repurposing explanations as semantic graphs

Evaluating an explanation is a tough task and many different benchmarks are recently appearing to evaluate them [41]. In this work, we followed two different approaches to evaluate the explanations: a more subjective one, where the explanation was evaluated with our own biological knowledge; and a more objective one, where a manual literature search and curation was performed to check if the suggested explanation has already been reported. We selected 7 phenotypes (muscular dystrophy, respiratory insufficiency, arrhythmia, dilated cardiomyopathy, congestive heart failure, progressive muscle weakness and cognitive impairment) and their top 3 predictions, then explanations were produced from the models trained on both KGs. The selection of these phenotypes aimed to cover the diverse systems affected by the disease. Each explanation was analyzed and, if possible, compared to the one that was found in the literature.

Explanations were classified into complete and incomplete explanations. Complete explanations are those that show a connection (path) between the drug candidate and the targeted symptom/phenotype (Figure S3). They are considered complete as they allow for an easy human-understandable interpretation. On the other hand, incomplete explanations are those where the explanation is composed of two separated clusters (one for the drug and one for the phenotype) (Figure S4) or by a unique cluster where either the drug or the phenotype is missing (Figure S5).

Table 8
Number and percentage of complete and incomplete explanations in each evidence type.

	Complete Explanations	Percentage Complete Explanations	Incomplete Explanations	Percentage Incomplete Explanations
Supporting Evidence	13	68 %	6	32 %
Contraindication Evidence	3	30 %	7	70 %
No Evidence	5	38 %	8	62 %
Total	21	50 %	21	50 %

The global analysis of the completeness of explanations generated can be seen in Table 8 (amount of complete and incomplete explanations in each type of supporting evidence) and Table 9 (amount of supporting evidence in each type of explanations). This analysis was performed taking into account the explanations from both graphs. As it can be seen in Table 8, in total the same number of complete and incomplete explanations was obtained (21 each). However, when looking at each category separately, it is seen that when there is evidence GNNExplainer tends to produce complete explanations (68 %), and conversely when there is no supporting evidence or when the drug is contraindicated the resulting explanation is usually incomplete (62 % and 70 %, respectively). As it can be seen in Table 9, when a complete explanation is created, almost 2/3 of the time the explanation contains supporting evidence (62 %); while when the explanation is incomplete, only 1/4 of the times it contains supporting evidence (28 %).

Table 9
Number and percentage of explanations with no evidence, with supporting evidence and with contraindications in each type of explanation.

	Supporting Evidence	Percentage with Evidence	Contraindication Evidence	Percentage with Contraindications	No Evidence	Percentage No Evidence
Complete Explanations	13	62 %	3	14 %	5	24 %
Incomplete Explanations	6	28%	7	33 %	8	38%

An additional analysis was performed, this time considering each graph separately. This can be seen in Table 10 and Table S7. There is a clear difference between the explanations obtained in graph A and B. Firstly, KG A explanations are more likely to be complete (72 % in A vs 28 % in B), while KG B produces more incomplete explanations (72 % in B vs 28 % in A) (Table 10).

Table 10
Number and percentage of complete and incomplete explanations in each evidence type and in each graph.

	Evidence Type	**Complete Explanations**		**Incomplete Explanations**	
		Number	**Percentage**	**Number**	**Percentage**
KG A	Supporting Evidence	9	100%	0	0%
	Contraindication Evidence	1	17%	5	83%
	No Evidence	5	83%	1	17%
	Total	**15**	**72%**	**6**	**28%**
KG B	Supporting Evidence	4	40%	6	60%
	Contraindication Evidence	2	50%	2	50%
	No Evidence	0	0%	7	100%
	Total	**6**	**28%**	**15**	**72%**

An example of an explanation produced by rd-explainer can be seen in Figure 5. This explanation is classified into complete and suggests why Doxorubicin should be considered for treating respiratory insufficiency; as it is a drug that targets CHRM1 a gene that interacts with DAG1, which causes the disease. Throughout this section explanations have been classified into complete and incomplete. However, an explanation being complete does not make it a good explanation. This way, for example, an explanation of the type 'Drug A targets Gene B, Gene B interacts with Gene C, and Gene C causes Disease D' can make biological sense such as in Figure 5. On the other hand, an explanation of the type 'Drug A treats Disease B, Disease B is caused by Gene C, Gene C causes Disease D' does not make full biological sense (Drug A could treat Disease B by targeting a gene other than Gene C; this way, the same treatment could not be applied for Disease D). This is in fact what is observed in Figure S6, where disopyramide is said to treat muscular dystrophy following the next explanation: disopyramide treats urinary incontinence, affectation in DMD gene can cause urinary incontinence, and DMD gene has as phenotype muscular dystrophy. In this case, a person may have urinary incontinence for several reasons, and disopyramide may be able to treat one of them, but not necessarily the one caused by affectation in DMD gene.

The objective evaluation is undoubtedly more unbiased and equitable. Nonetheless, subjective evaluations are also significant since there are drug-phenotype interactions that are not fully understood (specially when a certain drug is producing an undesired side effect), and so they are not well established in the literature. But, analyzing the proposed explanations based on expert domain knowledge might shed light on the interaction and help to formulate a hypothesis that can be clearly designed to be tested in the wet laboratory.

After applying the objective evaluation only one explanation (levosimendan - progressive muscle weakness) was found to have supporting evidence (where levosimendan treats the disease by increasing the troponin C affinity for calcium), and two links' explanations (doxorubicin - respiratory insufficiency and sorafenib - respiratory insufficiency) contained unclear interactions (both were of type contraindications). The results after applying this evaluation can be found summarised in the Table S6. Regarding the subjective evaluations, 17 out of 21 explanations were found to be good explanations (they were in accordance with biological reasoning) such as the one illustrated by doxorubicin - respiratory insufficiency in Figure 5; and 4 were considered bad explanations (they made no biological sense), the previously mentioned disopyramide - muscular dystrophy in Figure S6, and the explanations in Figures S7, S8 and S9.

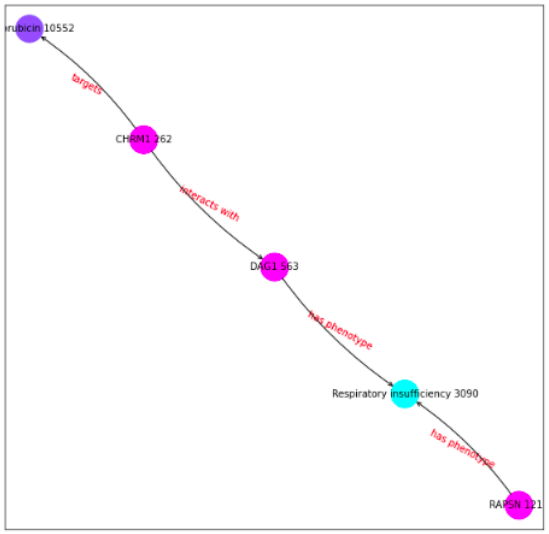

Fig. 5. Explanation of drug candidate Doxorubicin as possible treatment for Respiratory Insufficiency. Classified as complete explanation.

5.6. Generalizability of rd-explainer tested on other case studies

To show that this method can be extended to other rare diseases it was also tested in Alzheimer's Disease (AD) and Amyotrophic Lateral Sclerosis (ALS) type 1. Despite Alzheimer Disease not being a rare disease, there are different types of Alzheimer with very little prevalence. This way, for the Alzheimer's knowledge graph we used the general disease (MONDO:0004975) and all its causal genes that were present in Monarch (APP (HGNC:620), APOE (HGNC:613), PSEN1 (HGNC:9501) and PSEN2 (HGNC:9509)) as seeds. The final result would be a knowledge graph that specializes in Alzheimer diseases and that we can use to focus on the symptoms of the rare types of the disease. For the ALS type 1 knowledge graph we used the seed for the disease (MONDO:0007103) and the causal gene according to Monarch (SOD1 (HGNC:11179)). Table 11 shows the GNN performance in both diseases, showing once more a high AUROC and AUPRC for these diseases.

Table 11
Table showing different performance metrics tested in AD and ALS.

	Precision	**Recall**	**F1-score**	**AUROC**	**AUPRC**
AD	0.95	0.95	0.95	0.98	0.97
ALS	0.94	0.93	0.94	0.97	0.97

Next, the same approach that was followed for DMD was followed for both diseases: for each symptom we analysed the three drug candidates with the highest score (this drug candidates should not appear in the knowledge graph); then a literature search was performed to check if the drug candidates had been reported by the scientific community. The complete list of phenotypes as well as the drug candidates and scores for each phenotype can be found in Tables S11 and S12. This tables also contain whether the drug candidates had supporting evidence in the literature.

Among the predictions, it is worth mentioning Pexidartinib, a drug candidate that was proposed by the model to treat memory impairment in AD and that is currently undergoing a clinical trial as a drug that could be potentially beneficial to treat the disease [42].

6. Discussion

We integrated disease-specific knowledge graphs in combination with GNN and XAI for interpretable drug repurposing. We found that state-of-the-art XAI methods based on GNNs support *in silico* predictions of candidate repurposable drugs for rare diseases by providing interpretable reasoning paths of mechanism of action. We developed *rd-explainer*, a method to perform computational drug repurposing specifically for rare diseases. It utilizes cutting-edge deep learning methods such as edge2vec and GNNs and provides drug-symptom/phenotype predictions with high performance scores, and utilizes a modified version of GNNExplainer to provide explanations as semantic graphs for the interpretability of the results. We also found that these explanations have different levels of usefulness to generate testable hypotheses: paths linking drug and phenotype nodes are more understandable versus isolated clusters since they are similar to human reasoning; adding semantics to relations adds biological meaning to help to formulate a hypothesis and design the experiment in the laboratory; and providing clear semantic graphs by removing relations that are not contributors in the learning process. We tested the generalizability of our method executing it on two additional diseases: ALS and AD. ALS type 1 was selected to test the pipeline in another monogenic disease with fewer information available. AD was selected as it is a common disease with rare subtypes that can be caused by several genes, and we wanted to test the pipeline in a polygenic and multifactorial disease. We demonstrated that our pipeline performs well on mono- and polygenic rare diseases.

rd-explainer is a researcher-centered drug repurposing method that has been demonstrated as an innovative AI based method for rare disease drug research. rd-explainer's main advantage is its interpretability. The main motivation of this study was to provide explanations underlying AI predictions. rd-explainer provides explanations as semantic graphs, a type of explanation that resembles to human reasoning. This is in line with current research

on user-centric XAI [43]. Not only does this have the high value to support rare disease researchers to formulate evidence-based hypotheses testable in the wet laboratory (and reduce cost, time and risk), but to gain new disease knowledge and speed up robust drug research. Our approach was to use state-of-the-art AI and XAI methods used in drug repurposing such as knowledge graphs to naturally represent known associations among biological entities with expressive semantics and supporting curated evidence, graph learning, and graph based XAI methods. The advance in the rare disease field is that we provide interpretable predictions thanks to a pipeline that it seamlessly integrates a graph learning model with an explainer, combining results of both model performance and explanation accuracy to mitigate the black-box problem and foster XAI adoption in the field [44]. BioKnowledge Reviewer tool provides rare disease specific knowledge graphs for disease biology data collection by means of the Monarch knowledge base API [31]. We argue that a tool or approach that can collect associations from a virtual, federated knowledge graph via APIs could extend this feature to any biomedical associations such as for drug data collection, and improve data and knowledge driven research. Another great advantage of the rd-explainer method is its modular implementation; this means that different parts of the workflow (data, features, GNN and explanations) can be independently modified and the pipeline can still be run. For example, if one is interested in using another node feature embedding algorithm instead of edge2vec, one can just modify that component of the pipeline and still run the rest of the workflow.

Our results showed that rd-explainer is a highly performant graph ML based drug repurposing method. Our method builds rare disease-specific models trained on newly generated KG for the disease of focus and enriched with data for the prediction task. In comparison with state-of-the-art AI-based drug repurposing approaches, rd-explainer demonstrates outstanding performance. Throughout this paper, we have compared rd-explainer with various AI methods that employ different techniques for their predictions, including GNNs such as GraphSAGE, random walk embeddings like edge2vec, and geometric embeddings using models like ComplEX, DistMult, and TransE. By combining random walk models (edge2vec) with GNNs (GraphSAGE), rd-explainer achieves superior results in the link prediction task. Notably, edge2vec outperforms GraphSAGE, suggesting that the exceptional performance of rd-explainer is primarily attributed to the random walk model, with the GNN providing an additional performance boost. This level of performance rivals other models developed for drug repurposing, such as deepDR (AUROC = 0.908) [45]. Although there are benchmarks and frameworks to evaluate the performance of GNNs [46–51] to the best of our knowledge there is no a standard for drug repurposing, and makes it challenging to directly compare rd-explainer to other methods due to one of its key features: the creation of high-quality disease specific knowledge graphs. These knowledge graphs are enriched with data from a wide array of sources including domain expert knowledge via the seed nodes, and curated known relations among genes, anatomical structures, biological processes and diseases not only from humans, but also importantly numerous other species to fill the lack of molecular knowledge. This comprehensive approach significantly boosts the graph's richness and diversity, making it a valuable resource for tackling rare diseases, which often suffer from limited research attention. By maximizing the information available, rd-explainer enhances our ability to identify potential treatments for these understudied conditions and, ultimately, enable more effective and faster translation. Conversely, Huang et al. recently proposed a clinician-centered drug repurposing foundation model pre-trained on a medical KG composed of 17.000 diseases and transfer learning by disease mechanism similarity [52]. It would be interesting to combine both approaches and investigate the effect of extending our KGs with similar disease networks from well-known diseases.

Our new predictions are valid drug candidates since they are consistent with recent findings in the literature. We demonstrated that rd-explainer can provide new interesting drug-phenotype predictions. For instance, Sunitinib, one of the drugs that appear to be a good candidate to treat the symptoms of the disease according to both models (using KG A and KG B), has been considered as a good drug candidate to treat DMD and in 2019 appeared to be in preclinical trials [53]. This drug belongs to the group of tyrosine kinase inhibitors, and many other drugs that belong to this category have been proposed by our model (Fedratinib, Sorafenib, Bosutinib, Ruxolitinib and Midostaurin). Similarly, Mezlocillin, an antibiotic used to treat gram-negative bacterial infections, has also been proposed by our model; while Gentamicin, another gram-negative antibiotic, was in 2019 in clinical trials to treat DMD [53]. This way, despite not producing drugs candidates that are undergoing a clinical trial or treating the disease, it produces drug candidates that participate in similar biological processes (i.e., tyrosine kinases inhibitor, gram negative antibiotics)

Importantly, explanations for hypothesis generation may enable to move towards *lab-in-the-loop* framework. With respect to the interpretability and utility of explanations, one of the 21 examined explanations was supported by evidence in the literature. Nonetheless, this does not mean that the explanations are useless. A good example of this would be the explanation for the Methylprednisolone-Muscular Dystrophy link (Figure S10). The explanation is simple: 'Methylprednisolone treats DMD, DMD has Muscular Dystrophy as phenotype; thus methylprednisolone can treat Muscular Dystrophy'. In this case the explanation does not contain supporting evidence but the explanation still makes sense. In the literature, methylprednisolone is said to be a good candidate to treat muscular dystrophies because it interacts with the glucocorticoid receptor and this leads to the activation of anti-inflammatory signaling and the inhibition of proinflammatory signaling [54]. The explanation proposed by rd-explainer doesn't provide the underlying causative mechanism that relates methylprednisolone and muscular dystrophy, but a researcher can still be able to see that muscular dystrophies and methylprednisolone are interrelated. This illustrates how even though an explanation may lack comprehensive supporting evidence, it can still provide valuable directional cues for further more precise investigation. Another important aspect is that rare disease findings in the lab can be introduced back in the knowledge graph to update and improve the disease specific AI model for continual learning and enabling precise experimental design. Besides, this synergy fosters collaboration between computational and wet lab researchers to increase efficiency for disease specific drug research [31].

Finally, we found that knowledge graph topology has an impact on explainability. It was also seen that KG A usually produces more complete explanations, while in KG B incomplete explanations appear to be more numerous. This could happen due to the difference in the graph structure itself: graph A has a smaller clustering coefficient than graph B (see Section 5.1), which leads to more edges being present in the subgraphs produced by GNNExplainer. This way, because the 15th edges with the highest scores are selected, it is more likely to find a path between drug and phenotype in KG A than in B. Another interesting difference is that explanations generated with KG A tend to have a higher 'sensitivity', while explanations generated with KG B tend to have a higher 'specificity'. When an incomplete explanation is produced using graph A it is very unlikely that the explanation will contain supporting evidence (0 explanations were found to have evidence if the explanation was incomplete in KG A). Similarly, when a complete explanation is produced in KG B, it is very likely that the explanations have supporting evidence or contraindication evidence (67% of complete explanations had supporting evidence and 33% of complete explanations had contraindication evidence). For this reason, if one remains skeptical about the explanations themselves, this quality of the explanations might be used as filter/validation. For example, if an incomplete explanation is obtained with KG A, it is unlikely that it is trustworthy (none of the incomplete explanations had supporting evidence). Similarly, if a complete explanation is obtained using KG B, it is likely that there is some interaction between the drug and the phenotype (all of the complete explanations generated with graph B had either supporting or contraindication evidence). Our findings are aligned with recent studies where the influence of clustering coefficient and topology has been observed in embedding-based predictions [55, 56], here we extend these observations to its impact on graph-based explanations.

Limitations and future directions

An important limitation of this study is that we only utilize one XAI method, which is not model agnostic. XAI is a hot research topic in the AI field, where new and more sophisticated methods are frequently published [57]. It would be good to extend our study to other XAI types to check how applicable they are given the unique characteristics of rare diseases, including limited annotated data, lack of knowledge of pertinent entity relations, and lack of a gold drug-phenotype standard. Another important limitation is the lack of standard benchmarking and metrics to systematically evaluate explainers and explanations. Currently, there are some initial efforts going in this direction [36, 58–62], but there is still a lack of a common standard [63]. The known reproducibility issue of our explainer [36] that may imply that the explanations are different each time it is used, may reduce the confidence and reliance on the explanations. We did several experiments to try and bring consistency to explanations; for example, executing GNNExplainer several times and using the mean mask as the final mask or increasing the number of epochs. However, this still did not solve the issue. This experience makes us strongly recommend to work on the standard evaluation of explanations by the XAI community to foster trust on the application of AI in bioinformatics and biomedicine. Additionally, many times the explanation would consist in a subgraph where the two targeted nodes would be disconnected from each other, which might bring confusion and could be seen as a 'bad' explanation.

Therefore, work towards methods that prioritize or focus on providing just connecting paths such as metapath based ones [64–68] and on improving path visualisation for user interpretation [69–71] is arguably recommended. Finally, while we focused primarily on integrating a graph ML model with an explainer, a clear line of research will be to work on interpretability and reproducibility of explanations in the context of the drug repurposing task. The reproducibility/incosistency could be affected by the size and complexity of our data. This inconsistency could make the users of this pipeline skeptical about its explanations and for this reason more investigation should be done in this element of the pipeline to make it a more robust model. To improve this, ontologies could be incorporated into the knowledge graph to increase the quality and interpretability of our data. Ontologies help to standardize data into the shared meaning by a community enhancing thus interpretability by domain users. Importantly, the formal description of knowledge embedded in ontologies can be leveraged for data consistency checking, and for inference to add implicit knowledge into the graph [72]. Nonetheless, knowledge graph and ontology changes pose a great interoperability challenge to the community to keep up downstream bioinformatics and data science workflows and analyses [73, 74]. Finally, it would make our work more 'FAIR' [75], i.e., not only understandable by humans, but also by machines, by providing our drug repurposing for DMD KG from a FAIR data point [76], and rd-explainer from workflowHub [77].

7. Conclusion

We present the application of explainable AI on state-of-the-art computational drug repurposing for rare diseases. Our knowledge graph based deep learning method provides human understandable explanations for the phenotype-drug link prediction and we demonstrated that graph XAI can be applied to rare diseases. The *rd-explainer* method provides an innovative approach that can maximize the available disease-specific knowledge and generate valuable predictions with its explanations. Our model has proven to obtain high evaluation scores, providing drug candidates that are often supported by evidence. The key contribution of our study is that our pipeline gives possible explanations in the form of semantic graphs that may help rare disease researchers to make informed decisions to experimentally validate deep learning model predictions. However, we detected that data topology affects explanations, highlighting the importance of investigating further how best represent graphical knowledge for model performance and explanation accuracy. *rd-explainer* can be extended to other rare diseases and provide computer-aided guidance for biologists and accelerate translational research. Finally, future studies should advance our understanding of the necessary standard mechanism to evaluate explainability to foster adoption from domain experts and to mitigate the black-box problem of trust on AI, especially for biomedicine where decisions can have an important impact on people's lives.

8. Code availability

The code is freely accessible with an open license at https://github.com/PPerdomoQ/rare-disease-explainer.

References

[1] Duxin Sun, Wei Gao, Hongxiang Hu, and Simon Zhou. Why 90*Acta Pharmaceutica Sinica. B*, 12:3049, 7 2022. ISSN 22113843. . URL /pmc/articles/PMC9293739//pmc/articles/PMC9293739/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC9293739/.

[2] Melissa Haendel, Nicole Vasilevsky, Deepak Unni, Cristian Bologa, Nomi Harris, Heidi Rehm, Ada Hamosh, Gareth Baynam, Tudor Groza, Julie McMurry, Hugh Dawkins, Ana Rath, Courtney Thaxon, Giovanni Bocci, Marcin P. Joachimiak, Sebastian Köhler, Peter N. Robinson, Chris Mungall, and Tudor I. Oprea. How many rare diseases are there? 19(2):77–78. ISSN 1474-1776. . URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7771654/.

[3] Rare diseases, . URL https://ec.europa.eu/health/non-communicable-diseases/steering-group/rare-diseases_en.

[4] Emre Guney, Jörg Menche, Marc Vidal, and Albert-László Barábasi. Network-based in silico drug efficacy screening. 7(1):10331. ISSN 2041-1723. . URL https://www.nature.com/articles/ncomms10331. Number: 1 Publisher: Nature Publishing Group.

[5] Bryce Goodman and Seth Flaxman. European Union Regulations on Algorithmic Decision-Making and a 'Right to Explanation'. *AI Magazine*, 38(3):50–57, sep 2017. ISSN 07384602. . URL https://api.semanticscholar.org/CorpusID:7373959#id-name=S2CID.

[6] Qiang Huang, Makoto Yamada, Yuan Tian, Dinesh Singh, Dawei Yin, and Yi Chang. GraphLIME: Local interpretable model explanations for graph neural networks. (arXiv:2001.06216). . URL http://arxiv.org/abs/2001.06216. type: article.

[7] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. "why should i trust you?": Explaining the predictions of any classifier. (arXiv:1602.04938). . URL http://arxiv.org/abs/1602.04938. type: article.

[8] Pouya Pezeshkpour, Yifan Tian, and Sameer Singh. Investigating robustness and interpretability of link prediction via adversarial modifications. (arXiv:1905.00563). . URL http://arxiv.org/abs/1905.00563. type: article.

[9] Cristina Al-Khalili Szigyarto and Pietro Spitali. Biomarkers of duchenne muscular dystrophy: current findings. *Degenerative Neurological and Neuromuscular Disease*, 8:1–13, 1 2018. . URL https://www.dovepress.com/biomarkers-of-duchenne-muscular-dystrophy-current-findings-peer-reviewed-fulltext-article-DNND.

[10] Jacob Al-Saleem, Roger Granet, Srinivasan Ramakrishnan, Natalie A. Ciancetta, Catherine Saveson, Chris Gessner, and Qiongqiong Zhou. Knowledge graph-based approaches to drug repurposing for COVID-19. page acs.jcim.1c00642. ISSN 1549-9596. . URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8340579/.

[11] Aditya Grover and Jure Leskovec. node2vec: Scalable feature learning for networks. (arXiv:1607.00653). URL http://arxiv.org/abs/1607.00653. type: article.

[12] Bishan Yang, Wen-tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. Embedding entities and relations for learning and inference in knowledge bases. (arXiv:1412.6575). . URL http://arxiv.org/abs/1412.6575. type: article.

[13] Xiang Yue, Zhen Wang, Jingong Huang, Srinivasan Parthasarathy, Soheil Moosavinasab, Yungui Huang, Simon M. Lin, Wen Zhang, Ping Zhang, and Huan Sun. Graph embedding on biomedical networks: methods, applications and evaluations. *Bioinformatics*, 36:1241–1251, 2 2020. ISSN 1367-4803. . URL https://dx.doi.org/10.1093/bioinformatics/btz718.

[14] Ilaria Ferrari, Giacomo Frisoni, Paolo Italiani, Gianluca Moro, and Claudio Sartori. Comprehensive analysis of knowledge graph embedding techniques benchmarked on link prediction. *Electronics 2022, Vol. 11, Page 3866*, 11:3866, 11 2022. ISSN 2079-9292. . URL https://www.mdpi.com/2079-9292/11/23/3866/htmhttps://www.mdpi.com/2079-9292/11/23/3866.

[15] Shaghayegh Sadeghi, Jianguo Lu, and Alioune Ngom. An integrative heterogeneous graph neural network–based method for multi-labeled drug repurposing. *Frontiers in Pharmacology*, 13:908549, 7 2022. ISSN 16639812. .

[16] Yuchen Zhang, Xiujuan Lei, Yi Pan, and Fang Xiang Wu. Drug repositioning with graphsage and clustering constraints based on drug and disease networks. *Frontiers in Pharmacology*, 13:872785, 5 2022. ISSN 16639812. .

[17] Kanglin Hsieh, Yinyin Wang, Luyao Chen, Zhongming Zhao, Sean Savitz, Xiaoqian Jiang, Jing Tang, and Yejin Kim. Drug repurposing for covid-19 using graph neural network and harmonizing multiple evidence. *Scientific Reports 2021 11:1*, 11:1–13, 11 2021. ISSN 2045-2322. . URL https://www.nature.com/articles/s41598-021-02353-5.

[18] Junjun Zhang and Minzhu Xie. Graph regularized non-negative matrix factorization with prior knowledge consistency constraint for drug–target interactions prediction. *BMC Bioinformatics*, 23, 12 2022. ISSN 14712105. . URL /pmc/articles/PMC9798666//pmc/articles/PMC9798666/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC9798666/.

[19] Bishan Yang, Wen tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. Embedding entities and relations for learning and inference in knowledge bases. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 12 2014. URL https://arxiv.org/abs/1412.6575v4.

[20] William L. Hamilton, Rex Ying, and Jure Leskovec. Inductive Representation Learning on Large Graphs. *Advances in Neural Information Processing Systems*, 2017-December:1025–1035, jun 2017. ISSN 10495258. . URL https://arxiv.org/abs/1706.02216v4.

[21] Dongsheng Luo, Tianxiang Zhao, Wei Cheng, Dongkuan Xu, Feng Han, Wenchao Yu, Xiao Liu, Haifeng Chen, and Xiang Zhang. Towards inductive and efficient explanations for graph neural networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.

[22] Qianqian Xie, Jimin Huang, Tulika Saha, and Sophia Ananiadou. GRETEL: Graph contrastive topic enhanced language model for long document extractive summarization. In Nicoletta Calzolari, Chu-Ren Huang, Hansaem Kim, James Pustejovsky, Leo Wanner, Key-Sun Choi, Pum-Mo Ryu, Hsin-Hsi Chen, Lucia Donatelli, Heng Ji, Sadao Kurohashi, Patrizia Paggio, Nianwen Xue, Seokhwan Kim, Younggyun Hahm, Zhong He, Tony Kyungil Lee, Enrico Santus, Francis Bond, and Seung-Hoon Na, editors, *Proceedings of the 29th International Conference on Computational Linguistics*, pages 6259–6269, Gyeongju, Republic of Korea, October 2022. International Committee on Computational Linguistics. URL https://aclanthology.org/2022.coling-1.546/.

[23] Rex Ying, Dylan Bourgeois, Jiaxuan You, Marinka Zitnik, and Jure Leskovec. GNNExplainer: Generating explanations for graph neural networks. (arXiv:1903.03894). . URL http://arxiv.org/abs/1903.03894. type: article.

[24] Bastian Pfeifer, Anna Saranti, and Andreas Holzinger. Gnn-subnet: disease subnetwork detection with explainable graph neural networks. *Bioinformatics*, 38:ii120–ii126, 9 2022. ISSN 1367-4803. . URL https://dx.doi.org/10.1093/bioinformatics/btac478.

[25] Yu Sun, Xiang Xu, Lin Lin, Kang Xu, Yang Zheng, Chao Ren, Huan Tao, Xu Wang, Huan Zhao, Weiwei Tu, Xuemei Bai, Junting Wang, Qiya Huang, Yaru Li, Hebing Chen, Hao Li, and Xiaochen Bo. A graph neural network-based interpretable framework reveals a novel dna fragility–associated chromatin structural unit. *Genome Biology*, 24, 12 2023. ISSN 1474760X. . URL /pmc/articles/PMC10124043//pmc/articles/PMC10124043/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC10124043/.

[26] So Yeon Kim. Personalized explanations for early diagnosis of alzheimer's disease using explainable graph neural networks with population graphs. *Bioengineering*, 10, 6 2023. ISSN 23065354. . URL /pmc/articles/PMC10295378//pmc/articles/PMC10295378/?report=abstract https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10295378/.

[27] Monarch initiative explorer, . URL https://monarchinitiative.org/.

[28] Sorin Avram, Cristian G Bologa, Jayme Holmes, Giovanni Bocci, Thomas B Wilson, Dac-Trung Nguyen, Ramona Curpan, Liliana Halip, Alina Bora, Jeremy J Yang, Jeffrey Knockel, Suman Sirimulla, Oleg Ursu, and Tudor I Oprea. DrugCentral 2021 supports drug discovery and repositioning. 49:D1160–D1169. ISSN 0305-1048, 1362-4962. . URL https://academic.oup.com/nar/article/49/D1/D1160/5957163.

[29] Ying Zhou, Yintao Zhang, Xichen Lian, Fengcheng Li, Chaoxin Wang, Feng Zhu, Yunqing Qiu, and Yuzong Chen. Therapeutic target database update 2022: facilitating drug discovery with enriched comparative data of targeted agents. 50:D1398–D1407. ISSN 0305-1048, 1362-4962. . URL https://academic.oup.com/nar/article/50/D1/D1398/6413598.

[30] Zheng Gao, Gang Fu, Chunping Ouyang, Satoshi Tsutsui, Xiaozhong Liu, Jeremy Yang, Christopher Gessner, Brian Foote, David Wild, Ying Ding, and Qi Yu. edge2vec: Representation learning using edge semantics for biomedical knowledge discovery. 20(1):306. ISSN 1471-2105. . URL https://doi.org/10.1186/s12859-019-2914-2.

[31] Núria Queralt-Rosinach, Gregory S. Stupp, Tong Shu Li, Michael Mayers, Maureen E. Hoatlin, Matthew Might, Benjamin M. Good, and Andrew I. Su. Structured reviews for data and knowledge-driven research. 2020:baaa015. ISSN 1758-0463. .

[32] NetworkX — NetworkX documentation, . URL https://networkx.org/.

[33] CS224W Machine Learning with Graphl Home. URL http://web.stanford.edu/class/cs224w/.

[34] DeepSNAP documentation — DeepSNAP 0.2.0 documentation, . URL https://snap.stanford.edu/deepsnap/.

[35] Richard Liaw, Eric Liang, Robert Nishihara, Philipp Moritz, Joseph E. Gonzalez, and Ion Stoica. Tune: A research platform for distributed model selection and training. 7 2018. URL http://arxiv.org/abs/1807.05118.

[36] Chirag Agarwal, Owen Queen, Himabindu Lakkaraju, and Marinka Zitnik. Evaluating explainability for graph neural networks. *Scientific Data 2023 10:1*, 10:1–18, 3 2023. ISSN 2052-4463. . URL https://www.nature.com/articles/s41597-023-01974-x.

[37] Théo Trouillon, Johannes Welbl, Sebastian Riedel, Eric Gaussier, and Guillaume Bouchard. Complex embeddings for simple link prediction. In Maria Florina Balcan and Kilian Q. Weinberger, editors, *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pages 2071–2080, New York, New York, USA, 20–22 Jun 2016. PMLR. URL https://proceedings.mlr.press/v48/trouillon16.html.

[38] Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. Translating embeddings for modeling multi-relational data. In *Advances in Neural Information Processing Systems*, volume 26. Curran Associates, Inc. URL https://proceedings.neurips.cc/paper/2013/hash/1cecc7a77928ca8133fa24680a88d2f9-Abstract.html.

[39] Yang Yang, Ryan N. Lichtenwalter, and Nitesh V. Chawla. Evaluating link prediction methods. *Knowledge and Information Systems*, 45:751–782, 5 2015. ISSN 02193116. . URL https://arxiv.org/abs/1505.04094v1.

[40] Ruthwik R. Junuthula, Kevin S. Xu, and Vijay K. Devabhaktuni. Evaluating link prediction accuracy on dynamic networks with added and removed edges. *Proceedings - 2016 IEEE International Conferences on Big Data and Cloud Computing, BDCloud 2016, Social Computing and Networking, SocialCom 2016 and Sustainable Computing and Communications, SustainCom 2016*, pages 377–384, 7 2016. . URL https://arxiv.org/abs/1607.07330v1.

[41] Aniek F. Markus, Jan A. Kors, and Peter R. Rijnbeek. The role of explainability in creating trustworthy artificial intelligence for health care: A comprehensive survey of the terminology, design choices, and evaluation strategies. *Journal of Biomedical Informatics*, 113:103655, 1 2021. ISSN 1532-0464. .

[42] Antonio Ancidoni, Ilaria Bacigalupo, Giulia Remoli, Eleonora Lacorte, Paola Piscopo, Giulia Sarti, Massimo Corbo, Nicola Vanacore, and Marco Canevelli. Anticancer drugs repurposed for alzheimer's disease: a systematic review. *Alzheimer's Research & Therapy*, 13, 12 2021. ISSN 17589193. . URL /pmc/articles/PMC8101105//pmc/articles/PMC8101105/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC8101105/.

[43] Qianwen Wang, Kexin Huang, Payal Chandak, Marinka Zitnik, and Nils Gehlenborg. Extending the nested model for user-centric xai: A design study on gnn-based drug repurposing. *IEEE Transactions on Visualization and Computer Graphics*, 29(1):1266–1276, 2023. .

[44] Claudio Borile, Alan Perotti, and André Panisson. Evaluating link prediction explanations for graph neural networks. *CoRR*, abs/2308.01682, 2023. . URL https://doi.org/10.48550/arXiv.2308.01682.

[45] Xiangxiang Zeng, Siyi Zhu, Xiangrong Liu, Yadi Zhou, Ruth Nussinov, and Feixiong Cheng. deepdr: a network-based deep learning approach to in silico drug repositioning. *Bioinformatics*, 35:5191–5198, 12 2019. ISSN 1367-4803. . URL https://dx.doi.org/10.1093/bioinformatics/btz418.

[46] Weihua Hu, Matthias Fey, Marinka Zitnik, Yuxiao Dong, Hongyu Ren, Bowen Liu, Michele Catasta, and Jure Leskovec. Open graph benchmark: datasets for machine learning on graphs. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS '20, Red Hook, NY, USA, 2020. Curran Associates Inc. ISBN 9781713829546.

[47] Wentao Zhang, Zeang Sheng, Yuezihan Jiang, Yikuan Xia, Jun Gao, Zhi Yang, and Bin Cui. Evaluating deep graph neural networks. 2021.

[48] Vijay Prakash Dwivedi, Chaitanya K. Joshi, Anh Tuan Luu, Thomas Laurent, Yoshua Bengio, and Xavier Bresson. Benchmarking graph neural networks. *Journal of Machine Learning Research*, 24(43):1–48, 2023. URL http://jmlr.org/papers/v24/22-0567.html.

[49] Zhiyao Zhou, Sheng Zhou, Bochao Mao, Xuanyi Zhou, Jiawei Chen, Qiaoyu Tan, Daochen Zha, Yan Feng, Chun Chen, and Can Wang. Opengsl: a comprehensive benchmark for graph structure learning. In *Proceedings of the 37th International Conference on Neural Information Processing Systems*, NIPS '23, Red Hook, NY, USA, 2024. Curran Associates Inc.

[50] Juanhui Li, Harry Shomer, Haitao Mao, Shenglai Zeng, Yao Ma, Neil Shah, Jiliang Tang, and Dawei Yin. Evaluating graph neural networks for link prediction: current pitfalls and new benchmarking. In *Proceedings of the 37th International Conference on Neural Information Processing Systems*, NIPS '23, Red Hook, NY, USA, 2024. Curran Associates Inc.

[51] Xin Zheng, Miao Zhang, Chunyang Chen, Soheila Molaei, Chuan Zhou, and Shirui Pan. Gnnevaluator: evaluating gnn performance on unseen graphs without labels. In *Proceedings of the 37th International Conference on Neural Information Processing Systems*, NIPS '23, Red Hook, NY, USA, 2024. Curran Associates Inc.

[52] Kexin Huang, Payal Chandak, Qianwen Wang, Shreyas Havaldar, Akhil Vaid, Jure Leskovec, Girish N Nadkarni, Benjamin S Glicksberg, Nils Gehlenborg, and Marinka Zitnik. A foundation model for clinician-centered drug repurposing. *Nat. Med.*, September 2024.

[53] Libero Vitiello, Lucia Tibaudo, Elena Pegoraro, Luca Bello, and Marcella Canton. Teaching an old molecule new tricks: Drug repositioning for duchenne muscular dystrophy. 20(23):6053. ISSN 1422-0067. . URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6929176/.

[54] Mattia Quattrocelli, Aaron S. Zelikovich, Isabella M. Salamone, Julie A. Fischer, and Elizabeth M. McNally. Mechanisms and clinical applications of glucocorticoid steroids in muscular dystrophy. *Journal of Neuromuscular Diseases*, 8:39, 2021. ISSN 22143602. . URL /pmc/articles/PMC7902991//pmc/articles/PMC7902991/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC7902991/.

[55] Anand Kumar Gupta and Neetu Sardana. Significance of clustering coefficient over jaccard index. In *2015 Eighth International Conference on Contemporary Computing (IC3)*. IEEE, August 2015.

[56] Omar F Robledo, Xiu-Xiu Zhan, Alan Hanjalic, and Huijuan Wang. Influence of clustering coefficient on network embedding in link prediction. *Appl. Netw. Sci.*, 7(1), December 2022.

[57] A. Saranya and R. Subhashini. A systematic review of explainable artificial intelligence models and applications: Recent developments and future trends. *Decision Analytics Journal*, 7:100230, 6 2023. ISSN 2772-6622. .

[58] Chirag Agarwal, Satyapriya Krishna, Eshika Saxena, Martin Pawelczyk, Nari Johnson, Isha Puri, Marinka Zitnik, and Himabindu Lakkaraju. Openxai: Towards a transparent evaluation of model explanations. *36th Conference on Neural Information Processing Systems (NeurIPS 2022) Track on Datasets and Benchmarks*, 6 2022. URL https://arxiv.org/abs/2206.11104v3.

[59] Thomas Fel, Lucas Hervier, David Vigouroux, Antonin Poche, Justin Plakoo, Remi Cadene, Mathieu Chalvidal, Julien Colin, Thibaut Boissin, Louis Bethune, Agustin Picard, Claire Nicodeme, Laurent Gardes, Gregory Flandin, and Thomas Serre. Xplique: A deep learning explainability toolbox. 2022.

[60] Hao Yuan, Haiyang Yu, Shurui Gui, and Shuiwang Ji. Explainability in graph neural networks: A taxonomic survey. *IEEE Trans. Pattern Anal. Mach. Intell.*, 45(5):5782–5799, May 2023.

[61] Claudio Borile, Alan Perotti, and André Panisson. Evaluating link prediction explanations for graph neural networks. August 2023.

[62] Daniel Daza, Cuong Xuan Chu, Trung-Kien Tran, Daria Stepanova, Michael Cochez, and Paul Groth. Explaining graph neural networks for node similarity on graphs. July 2024.

[63] Mehwish Alam, Frank van Harmelen, and Maribel Acosta. Towards semantically enriched embeddings for knowledge graph completion. July 2023.

[64] Daniel Scott Himmelstein, Antoine Lizee, Christine Hessler, Leo Brueggeman, Sabrina L Chen, Dexter Hadley, Ari Green, Pouya Khankhanian, and Sergio E Baranzini. Systematic integration of biomedical knowledge prioritizes drugs for repurposing. *Elife*, 6, September 2017.

[65] Xinyu Fu, Jiani Zhang, Ziqiao Meng, and Irwin King. Magnn: Metapath aggregated graph neural network for heterogeneous graph embedding. *The Web Conference 2020 - Proceedings of the World Wide Web Conference, WWW 2020*, pages 2331–2341, 2 2020. . URL http://arxiv.org/abs/2002.01680http://dx.doi.org/10.1145/3366423.3380297.

[66] Michael Mayers, Roger Tu, Dylan Steinecke, Tong Shu Li, Núria Queralt-Rosinach, and Andrew I Su. Design and application of a knowledge network for automatic prioritization of drug mechanisms. *Bioinformatics*, 38(10):2880–2891, April 2022.

[67] Ayush Noori, Michelle M Li, Amelia L M Tan, and Marinka Zitnik. Metapaths: similarity search in heterogeneous knowledge graphs via meta-paths. *Bioinformatics*, 39(5), May 2023.

[68] Ana Jiménez, María José Merino, Juan Parras, and Santiago Zazo. Explainable drug repurposing via path based knowledge graph completion. *Sci. Rep.*, 14(1):16587, July 2024.

[69] Qianwen Wang, Kexin Huang, P Chandak, Nils Gehlenborg, and M Zitnik. Interactive visual explanations for deep drug repurposing. https://icml.cc/virtual/2021/workshop/8358, 2021. Accessed: 2024-10-8.

[70] Daniel S Himmelstein, Michael Zietz, Vincent Rubinetti, Kyle Kloster, Benjamin J Heil, Faisal Alquaddoomi, Dongbo Hu, David N Nicholson, Yun Hao, Blair D Sullivan, Michael W Nagle, and Casey S Greene. Hetnet connectivity search provides rapid insights into how biomedical entities are related. *Gigascience*, 12, December 2022.

[71] Qianwen Wang, Kexin Huang, Payal Chandak, Marinka Zitnik, and Nils Gehlenborg. Extending the nested model for user-centric xai: A design study on gnn-based drug repurposing. *IEEE Transactions on Visualization and Computer Graphics*, 29(1):1266–1276, 2023. .

[72] Mona Alshahrani, Mohammad Asif Khan, Omar Maddouri, Akira R Kinjo, Núria Queralt-Rosinach, and Robert Hoehndorf. Neuro-symbolic representation learning on biological knowledge graphs. *Bioinformatics*, 33(17):2723–2730, 04 2017. ISSN 1367-4803. . URL https://doi.org/10.1093/bioinformatics/btx275.

[73] Deepak Unni, Vasundra Touré, Philip Krauss, Katrin Crameri, and Sabine Österle. SPHN strategy to unravel the semantic drift between versions of standard terminologies. December 2023.

[74] Harshad Hegde, Jennifer Vendetti, Damien Goutte-Gattat, J Harry Caufield, John B Graybeal, Nomi L Harris, Naouel Karam, Christian Kindermann, Nicolas Matentzoglu, James A Overton, Mark A Musen, and Christopher J Mungall. A change language for ontologies and knowledge graphs. 2024.

[75] Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan Willem Boiten, Luiz Bonino da Silva Santos, Philip E. Bourne, Jildau Bouwman, Anthony J. Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T. Evelo, Richard Finkers, Alejandra Gonzalez-Beltran, Alasdair J.G. Gray, Paul Groth, Carole Goble, Jeffrey S. Grethe, Jaap Heringa, Peter A.C. t Hoen, Rob Hooft, Tobias Kuhn, Ruben Kok, Joost Kok, Scott J. Lusher, Maryann E. Martone, Albert Mons, Abel L. Packer, Bengt Persson, Philippe Rocca-Serra, Marco Roos, Rene van Schaik, Susanna Assunta Sansone, Erik Schultes, Thierry Sengstag, Ted Slater, George Strawn, Morris A. Swertz, Mark Thompson, Johan Van Der Lei, Erik Van Mulligen, Jan Velterop, Andra Waagmeester, Peter Wittenburg, Katherine Wolstencroft, Jun Zhao, and Barend Mons. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data 2016 3:1*, 3(1):1–9, mar 2016. ISSN 2052-4463. . URL https://www.nature.com/articles/sdata201618.

[76] Luiz Olavo Bonino da Silva Santos, Kees Burger, Rajaram Kaliyaperumal, and Mark D. Wilkinson. Fair data point: A fair-oriented approach for metadata publication. *Data Intelligence*, 5:163–183, 12 2023. ISSN 2641435X. . URL https://dx.doi.org/10.1162/dint_a_00160.

[77] Rafael Ferreira da Silva, Loïc Pottier, Tainã Coleman, Ewa Deelman, and Henri Casanova. Workflowhub: Community framework for enabling scientific workflow research and development – technical report. *Proceedings of 15th Workshop on Workflows in Support of Large-Scale Science, WORKS 2020 - Held in conjunction with SC 2020: The International Conference for High Performance Computing, Networking, Storage and Analysis*, pages 49–56, 9 2020. . URL http://arxiv.org/abs/2009.00250http://dx.doi.org/10.1109/WORKS51914.2020.00012.

[78] Visualization function - pytorch geometric. URL https://pytorch-geometric.readthedocs.io/en/latest/_modules/torch_geometric/nn/models/explainer.html#Explainer.visualize_subgraph.

9. Supplementary section

9.1. Knowledge Graph seeds

Table S1
Table containing seeds used to build KG A.

Seed Name	**Seed ID**
DMD	HGNC:2928
DMD	MONDO:0010679

Table S2
Table containing seeds used to build KG B.

Seed Name	**Seed ID**
DMD	HGNC:2928
DMD	MONDO:0010679
Hypotonia	HP:0001252
Specific learning disability	HP:0001328
Arrhythmia	HP:0011675
Congestive heart failure	HP:0001635
Dilated cardiomyopathy	HP:0001644
Calf muscle hypertrophy	HP:0008981
Motor delay	HP:0001270
Muscular dystrophy	HP:0003560
Delayed speech and language development	HP:0000750
Hypoventilation	HP:0002791
Intellectual disability, mild	HP:0001256
Hyporeflexia	HP:0001265
Cognitive impairment	HP:0100543
Proximal muscle weakness	HP:0003701
Abnormal EKG	HP:0003115
Calf muscle pseudohypertrophy	HP:0003707
Cardiomyopathy	HP:0001638
Flexion contracture	HP:0001371
Elevated circulating creatine kinase concentration	HP:0003236
Global developmental delay	HP:0001263
Skeletal muscle atrophy	HP:0003202
Respiratory insufficiency	HP:0002093
Waddling gait	HP:0002515
Gowers sign	HP:0003391
Generalized hypotonia	HP:0001290
Progressive muscle weakness	HP:0003323
Scoliosis	HP:0002650
Hyperlordosis	HP:0003307

9.2. Number of edge types

Table S3
Number and percentage of edge types in KG A.

Edge Type	**Count**	**Percentage**
in 1 to 1 orthology relationship with	35650	37.96%
in orthology relationship with	25242	26.88%
has phenotype	15730	16.75%
interacts with	9824	10.46%
is part of	1465	1.56%
has affected feature	1101	1.17%
expressed in	1079	1.14%
enables	983	1.04%
pathogenic for condition	976	1.03%
targets	518	0.55%
involved in	432	0.46%
likely pathogenic for condition	182	0.19%
contributes to condition	171	0.18%
has role in modeling	134	0.14%
is allele of	96	0.10%
is substance that treats	86	0.09%
colocalizes with	84	0.09%
source	29	0.03%
is causal germline mutation in	16	0.02%
has genotype	7	0.01%
contributes to	5	0.01%
causes condition	3	0.003%
is marker for	1	0.001%
is causal germline mutation partially giving rise to	1	0.001%

Table S4
Number and percentage of edge types in the KG B.

Edge Type	**Count**	**Percentage**
has phenotype	836138	42.13%
in 1 to 1 orthology relationship with	520547	23.23%
in orthology relationship with	333288	16.79%
interacts with	226174	11.40%
expressed in	14589	0.74%
is part of	9427	0.47%
colocalizes with	8112	0.41%
involved in	7790	0.39%
enables	7053	0.36%
targets	5070	0.26%
has role in modeling	3449	0.17%
causes condition	2479	0.12%
contributes to condition	2203	0.11%
is allele of	1167	0.06%
has affected feature	1137	0.06%
pathogenic for condition	1024	0.05%
is causal germline mutation in	900	0.04%
is substance that treats	599	0.03%
contributes to	198	0.01%
likely pathogenic for condition	185	0.01%
is causal loss of function germline mutation of in	179	0.01%
is reference allele of	130	0.01%
is marker for	97	0.005%
has genotype	67	0.003%
is causal susceptibility factor for	42	0.002%
source	32	0.002%
is causal somatic mutation in	16	0.001%
is causal gain of function germline mutation of in	15	0.001%
is causal germline mutation partially giving rise to	12	0.001%

9.3. GNNExplainer algorithm

$GNN, NodeIdx1, NodeIdx2, G\ G_{s,m}$, Mask $Emb = GNN(G)$ // Obtain embeddings
$InitialPred = Emb[NodeIdx1] \cdot Emb[NodeIdx2]$ // Get initial prediction
$G_s = Subgraph(G, NodeIndex1, NodeIndex2)$ // Obtain subgraph
$Mask = InitializeMask(G_s)$ // Initialize Mask
for *Epoch in Epochs* **do**
    $G_{s,m} = ApplyMask(G_s, Mask)$ // Apply Mask to subgraph
    $NewEmb = GNN(G_{s,m})$ // Get new embeddings
    $NewPred = NewEmb[NodeIdx1] \cdot NewEmb[NodeIde2]$ // Get new prediction
    $Loss = GetLoss(InitialPred, NewPred)$ // Calculate loss
    $Mask = Backpropagate(Mask, Loss)$ // Backpropagate loss
end
return $G_{s,m}$, *Mask*

Algorithm 1: GNNExplainer Link Prediction Pseudocode. *GNN* stands for the trained GNN model. *G* stands for the Graph.

9.4. List of hyperparameters

Table S5
Table showing the different options of hyperparameters that were tested as well as their optimal values.

Process	**Hyperparameter**	**Options**	**Optimal Value**
edge2vec	Number of walks	2, 4, 6	2
	Walk Length	3, 5, 7	7
	Embedding Dimension	32, 64, 128	32
	Edge Direction	Undirected, Directed	Directed
	p	0.5, 0.7, 1	0.7
	q	0.5, 0.7, 1	1
	Epochs	5, 10	10
GNN	Hidden Dimension	64, 128, 256	256
	Output Dimension	64, 128, 256	64
	Layers	2, 4, 6	2
	Aggregation Function	mean, sum	mean
	Dropout	0, 0.1, 0.2	0.2
	Learning Rate	0.001 - 0.1	0.07
	Epochs	100, 150, 200	150

9.5. Visualization of explanations

To visualize the resulting explanations, a custom visualization function was developed to represent explanations as more human readable and semantic graphs and, thus, improving the one provided by Pytorch Geometric [78]. In the first place, the possibility of visualizing the edge types has been incorporated. Additionally, in this new formula several customizable parameters have been added. Now, it is possible to only visualize the active edges of the explanation, removing non-important edges. This will allow for clearer visualization of the subgraph. Figure S1 shows how an explanation is modified after applying this option. Finally, it is also possible to remove unconnected clusters from the explanations. This way, if an explanation is formed by several clusters, there is the possibility of just viewing the ones that contain the drug candidate and the targeted phenotype. Figure S2 shows how the explanation is modified after applying this filter.

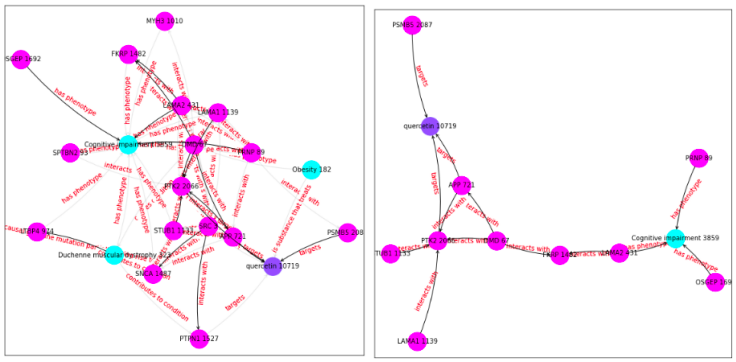

Fig. S1. Explanation after removing non-important edges. Left: Explanation keeping all the edges. Right: Explanation removing non-important edges.

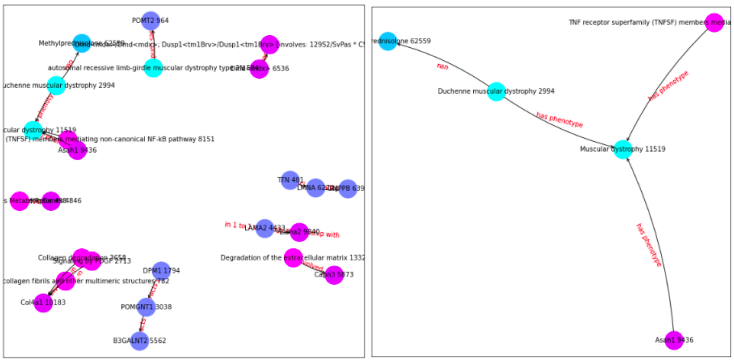

Fig. S2. Explanation after removing unconnected clusters. Left: Explanation keeping all the clusters. Right: Explanation removing additional clusters.

9.6. Complete/Incomplete explanation Example

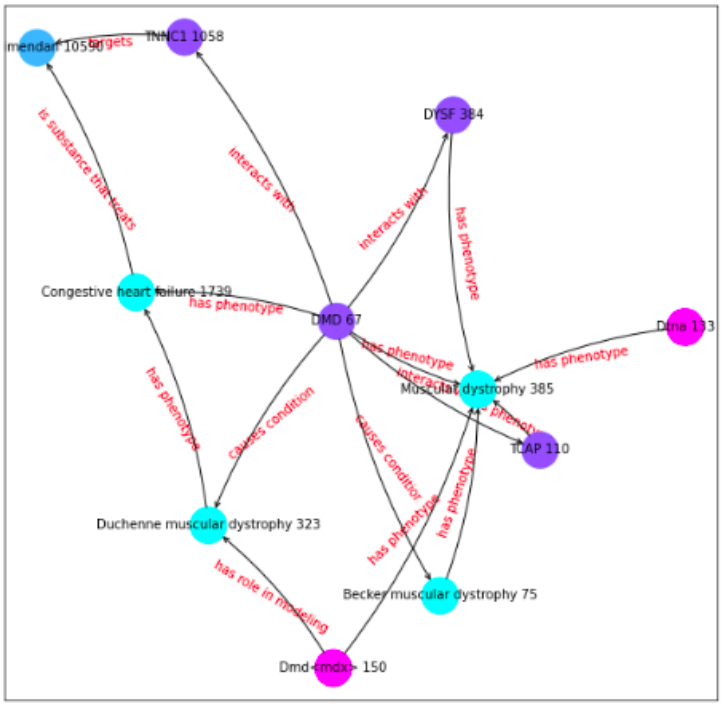

Fig. S3. Explanation of drug candidate Levosimendan as possible treatment for Muscular Dystrophy. Classified as complete explanation.

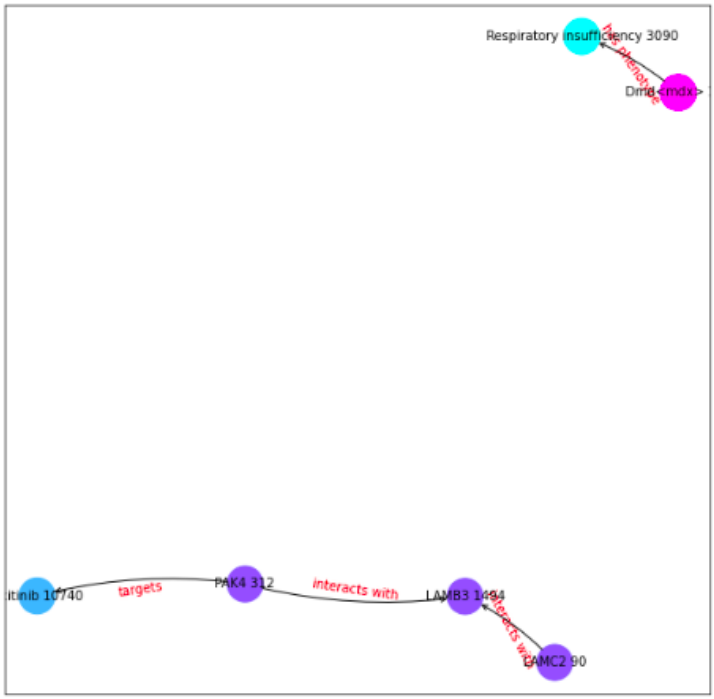

Fig. S4. Explanation of drug candidate Axitinib as possible treatment for Respiratory Insufficiency. Classified as incomplete explanation.

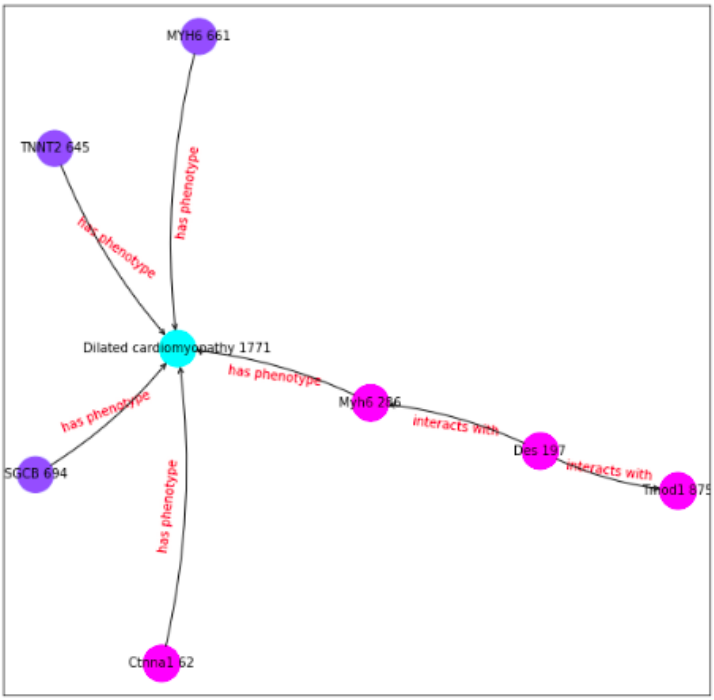

Fig. S5. Explanation of drug candidate Entrectinib as possible treatment for Dilated cardiomyopathy. Classified as incomplete explanation.

9.7. Evaluation of explanations

Table S6

Table showing the amount of times each drug appears as one of the top 3 drug candidates with highest score treat one of the 27 symptoms. It is also shown the amount of supporting evidence and contraindication evidence for each drug. This information was obtained using Graph A.

Drug	Appearances	Percentage	With Evidence	With Contraindications
Entrectinib	25	92.59 %	0	8
Axitinib	19	70.37 %	1	1
Nintedanib	12	44.44 %	2	0
Levosimendan	7	25.92 %	6	0
Disopyramide	6	22.22 %	2	0
Doxorubicin	2	7.40 %	0	2
Aprindine	2	7.40 %	2	0
Amiodarone	1	3.70 %	1	0
Acepromazine	1	3.70 %	0	0
Mezlocillin	1	3.70 %	0	0
Sunitinib	1	3.70 %	0	0
Fedratinib	1	3.70 %	0	0
Carvedilol	1	3.70 %	1	0
Queracetin	1	3.70 %	1	0

Table S7

Table showing the number and percentage of explanations with no evidence, with supporting evidence, and with contraindications for each type of explanation and each graph.

		With Evidence	Percentage With Evidence	With Contraindications	Percentage With Contraindications	No Evidence	Percentage No Evidence
KG A	*Complete Explanations*	9	60%	1	7%	5	33%
	Incomplete Explanations	0	0%	5	83%	1	17%
KG B (Large)	*Complete Explanations*	4	67%	2	33%	0	0%
	Incomplete Explanations	6	40%	2	13%	7	47%

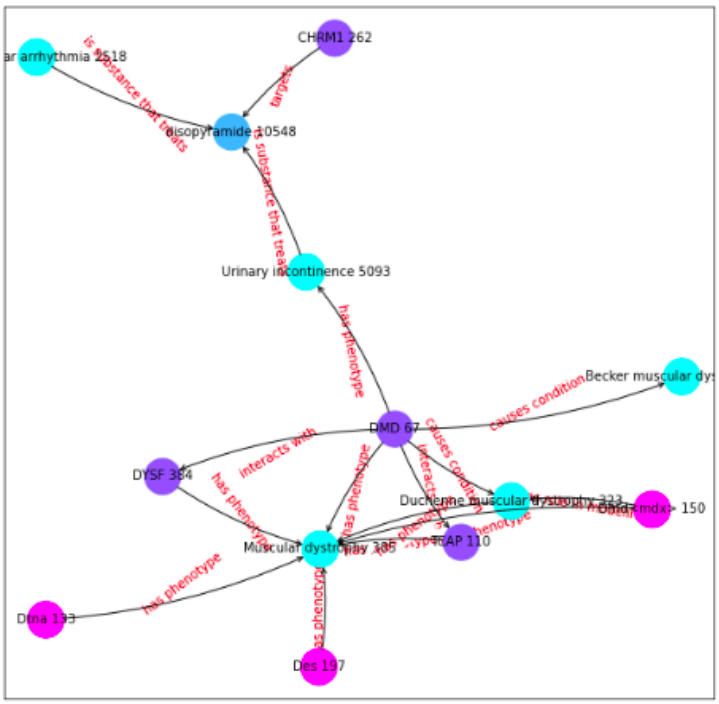

Fig. S6. Explanation of drug candidate Disopyramide as possible treatment for Muscular Dystrophy. Classified as complete explanation.

Table S8

Table showing the analysis of the explanation. The Good/Bad column shows the subjective evaluation. The Supporting Evidence Link shows if the Drug-Disease link contains supporting evidence. The Supporting Evidence Explanation shows if the explanation itself has supporting evidence.

Graph	**Drug**	**Disease**	**Good/Bad**	**Supporting Evidence Link**	**Supporting Evidence Explanation**
KG A	Levosimendan	*Muscular Dystrophy*	Good	Yes	-
	Disopyramide	*Muscular Dystrophy*	Bad	Yes	-
	Entrectinib	*Muscular Dystrophy*	Good	No	-
	Entrectinib	*Respiratory Insufficiency*	Good	No	-
	Doxorubicin	*Respiratory Insufficiency*	Good	Contraindication	Unclear: https://grantome.com/grant/NIH/R01-HL146443-01
	Levosimendan	*Arrhythmia*	Bad	Yes	-
	Amiodarone	*Arrhythmia*	Good	Yes	-
	Isradipine	*Arrhythmia*	Good	Yes	-
	Levosimendan	*Dilated Cardiomyopathy*	Bad	Yes	-
	Aprindine	*Congestive Heart Failure*	Bad	Yes	-
	Nintedanib	*Congestive Heart Failure*	Good	No	-
	Levosimendan	*Progressive Muscle Weakness*	Good	Yes	https://www.frontiersin.org/articles/10.3389/fphys.2021.786895/full
	Entrectinib	*Cognitive Impairment*	Good	Contraindication	-
	Axitinib	*Cognitive Impairment*	Good	Yes	-
	Quercetin	*Cognitive Impairment*	Good	Yes	-
KG B (Large)	Methylprednisolone	*Muscular Dystrophy*	Good	Yes	-
	Methylprednisolone	*Respiratory Insufficiency*	Good	Yes	-
	Sorafenib	*Respiratory Insufficiency*	Good	Contraindication	Unclear: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3961597/
	Methylprednisolone	*Progressive Muscle Weakness*	Good	Contraindication	-
	Resveratrol	*Progressive Muscle Weakness*	Good	Yes	-
	Sorafenib	*Cognitive Impairment*	Good	Contraindication	-

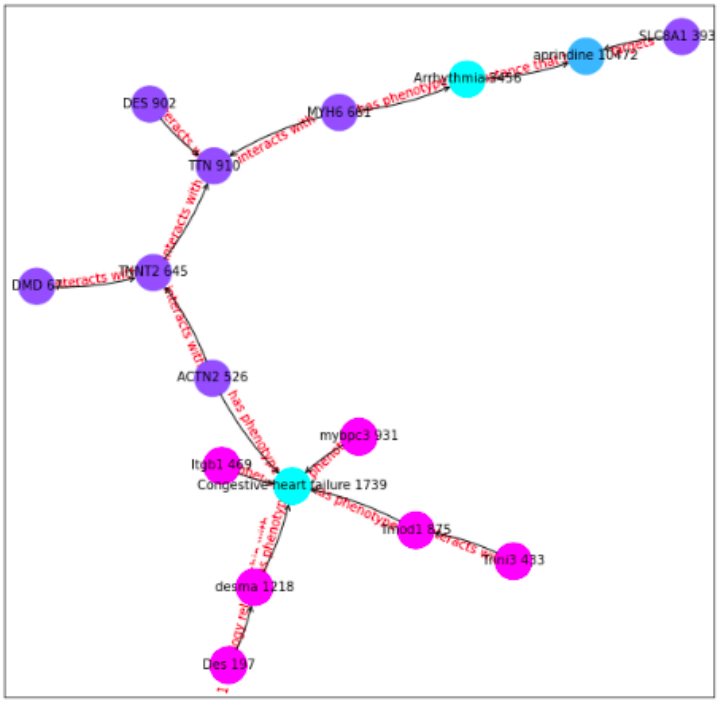

Fig. S7. Explanation of drug candidate Aprindine as possible treatment for Congestive heart failure. Classified as complete explanation.

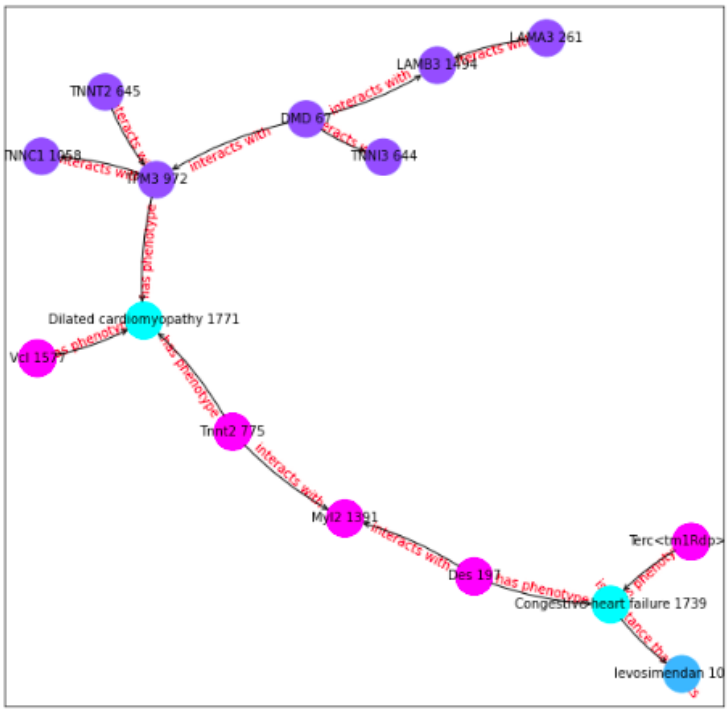

Fig. S8. Explanation of drug candidate Levosimendan as possible treatment for Dilated cardiomyopathy. Classified as complete explanation.

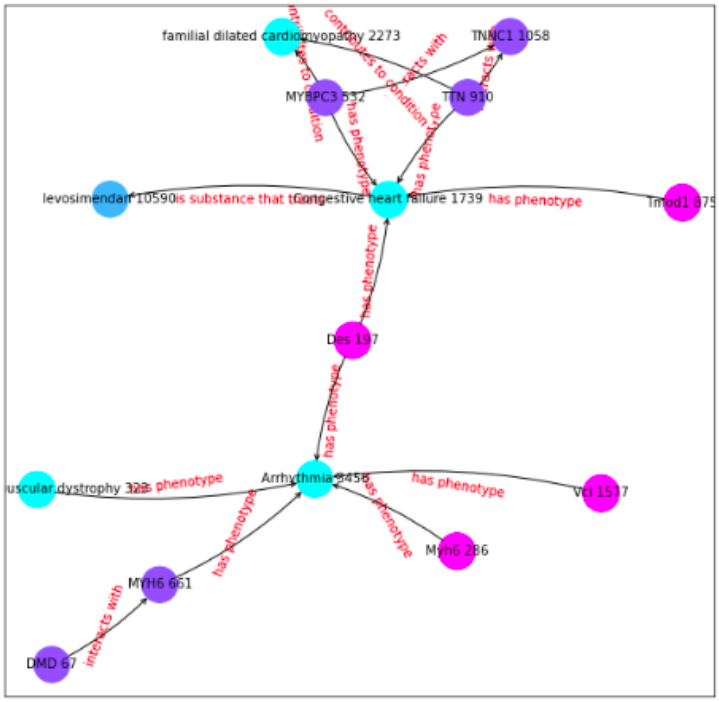

Fig. S9. Explanation of drug candidate Levosimendan as possible treatment for Arrhythmia. Classified as complete explanation.

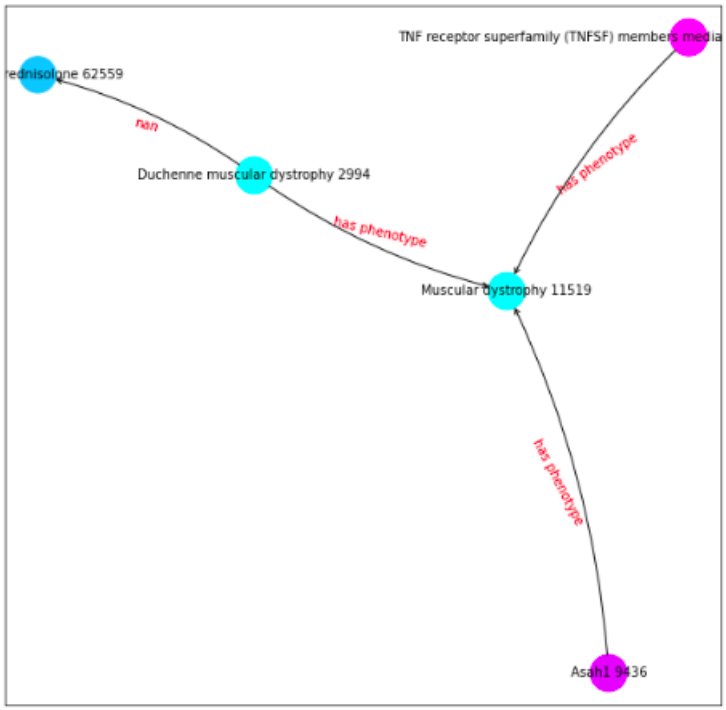

Fig. S10. Explanation of drug candidate Methylprednisolone as possible treatment for Muscular dystrophy. Classified as complete explanation.

9.8. Drug Candidates on KG A

Table S9: Table showing the drug candidates with the highest scores for each symptom/phenotype obtained in Graph A. Any evidence that supports the prediction will be shown in the *Supporting Evidence* column. If the drug is contraindicated for the given symptom/phenotype it will also be shown in this column.

Symptom	**ID**	**Drug Candidate**	**Score**	**Supporting Evidence**
Muscular dystrophy	HP:0003560	Levosimendan	0.849	https://pubmed.ncbi.nlm.nih.gov/30796500/
		Disopyramide	0.848	https://pubmed.ncbi.nlm.nih.gov/7045292/
		Entrectinib	0.845	None
Respiratory insufficiency	HP:0002093	Entrectinib	0.954	None
		Axitinib	0.925	None
		Doxorubicin	0.915	May produce respiratory dysfunction: https://grantome.com/grant/NIH/R01-HL146443-01
Gowers sign	HP:0003391	Entrectinib	0.963	None
		Axitinib	0.945	None
		Nintedanib	0.932	None
Global developmental delay	HP:0001263	Entrectinib	0.985	Can produce developmental delay: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8341080/
		Axitinib	0.974	None
		Nintedanib	0.968	None
Hyporeflexia	HP:0001265	Entrectinib	0.923	None
		Axitinib	0.905	None
		Nintedanib	0.872	None
Proximal muscle weakness	HP:0003701	Entrectinib	0.961	Can produce muscle weakness: https://www.drugs.com/sfx/entrectinib-side-effects.html
		Axitinib	0.944	None
		Nintedanib	0.925	https://pubmed.ncbi.nlm.nih.gov/29991677/
Intellectual disability	HP:0001256	Entrectinib	0.947	None
		Axitinib	0.921	None
		Doxorubicin	0.884	Can produce cognitive impairment: https://pubmed.ncbi.nlm.nih.gov/34055643
Calf muscle pseudohypertrophy	HP:0003707	Disopyramide	0.813	None
		Entrectinib	0.784	None
		Axitinib	0.776	None
Elevated serum creatine kinase	HP:0003236	Entrectinib	0.929	Can increase more: https://www.oncolink.org/cancer-treatment/oncolink-rx/entrectinib-rozlytrek
		Levosimendan	0.920	None
		Disopyramide	0.915	None

Abnormal EKG	HP:0003115	Levosimendan	0.777	https://pubmed.ncbi.nlm.nih.gov/20814559/
		Aprindine	0.747	https://pubmed.ncbi.nlm.nih.gov/10068848/
		Disopyramide	0.713	https://pubmed.ncbi.nlm.nih.gov/9141608/
Arrhythmia	HP:0011675	Levosimendan	0.890	https://ccforum.biomedcentral.com/articles/10.1186/cc1595#:$\sim$:text=Effects%20of%20levosimendan%20on%20cardiac%20arrhythmia%20in%20patients%20with%20severe%20heart%20failure,-J%20Lilleberg%20%26&text=Levosimendan%20(LS)%20is%20a%20novel,oxygen%20consumption%2C%20and%20induces%20vasodilation.
		Amiodarone	0.792	https://www.aafp.org/pubs/afp/issues/2003/1201/p2189.html#:$\sim$:text=Amiodarone%20is%20a%20potent%20antiarrhythmic,deaths%20in%20high%2Drisk%20patients.
		Isradipine	0.953	https://pubmed.ncbi.nlm.nih.gov/8480504/
Waddling gait	HP:0002515	Entrectinib	0.976	None
		Axitinib	0.964	None
		Nintedanib	0.947	None
Dilated cardiomyopathy	HP:0001644	Entrectinib	0.967	Can produce heart disease: https://www.drugs.com/cons/entrectinib.html
		Levosimendan	0.950	https://pubmed.ncbi.nlm.nih.gov/25863426/#:$\sim$:text=Conclusions%3A%20Levosimendan%20seems%20to%20improve,support%20while%20awaiting%20heart%20transplantation.
		Nintedanib	0.933	None
Flexion contracture	HP:0001371	Entrectinib	0.980	None
		Axitinib	0.975	None
		Nintedanib	0.958	None
Specific learning disability	HP:0001328	Entrectinib	0.871	None
		Axitinib	0.862	None
		Acepromazine	0.830	None
Skeletal muscle atrophy	HP:0003202	Entrectinib	0.962	None
		Axitinib	0.946	None

		Nintedanib	0.925	https://pubmed.ncbi.nlm.nih.gov/29991677/
Hypoventilation	HP:0002791	Axitinib	0.781	None
		Entrectinib	0.769	None
		Mezlocillin	0.759	None
Calf muscle hypertrophy	HP:0008981	Entrectinib	0.978	None
		Axitinib	0.977	None
		Disopyramide	0.976	None
Motor delay	HP:0001270	Entrectinib	0.991	None
		Sunitinib	0.985	None
		Fedratinib	0.978	None
Generalized hypotonia	HP:0001290	Entrectinib	0.995	None
		Axitinib	0.988	None
		Nintedanib	0.983	None
Cardiomyopathy	HP:0001638	Levosimendan	0.899	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6588712/
		Entrectinib	0.848	Can produce myocarditis: https://pubmed.ncbi.nlm.nih.gov/34315748/
		Carvedilol	0.837	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4055878/#:$\sim$:text=Pathways%20through%20which%20carvedilol%20exert,for%20beneficial%20effects%20in%20cardiomyopathy.
Hyperlordosis	HP:0003307	Entrectinib	0.970	None
		Axitinib	0.959	None
		Disopyramide	0.932	None
Congestive heart failure	HP:0001635	Entrectinib	0.863	Can produce heart failure: https://www.rozlytrek.com/ntrk/how-rozlytrek-may-help/possible-side-effects.html
		Aprindine	0.857	https://pubmed.ncbi.nlm.nih.gov/6871919/
		Nintedanib	0.835	None
Delayed speech and language development	HP:0000750	Entrectinib	0.986	None
		Axitinib	0.977	None
		Nintedanib	0.969	None
Scoliosis	HP:0002650	Entrectinib	0.994	None
		Axitinib	0.989	None
		Nintedanib	0.981	None
Progressive muscle weakness	HP:0003323	Levosimendan	0.864	https://www.frontiersin.org/articles/10.3389/fphys.2021.786895/full
		Entrectinib	0.985	Can cause weakness: https://www.drugs.com/sfx/entrectinib-side-effects.html

		Axitinib	0.960	Can cause weakness: https://www.mayoclinic.org/drugs-supplements/axitinib-oral-route/side-effects/drg-20075455?p=1#:$\sim$:text=This%20medicine%20may%20cause%20serious,trouble%20talking%2C%20or%20vision%20changes.
Cognitive impairment	HP:0100543	Entrectinib	0.952	Can induce cognitive disorders: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8149347/#:$\sim$:text=Cognitive%20disorders%20included%20events%20reported,(0.2%25)%20%5B20%5D.
		Axitinib	0.931	https://www.neuro-central.com/reversing-alzheimers-symptoms-in-mice-with-axitinib-treatment/
		Quercetin	0.991	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3736941/#:$\sim$:text=In%20vitro%20research%20also%20suggests,similar%20to%20that%20of%20caffeine.

9.9. Drug Candidates on KG B

Table S10: Table showing the drug candidates with the highest scores for each symptom/phenotype obtained with Graph B. Any evidence that supports the prediction will be shown in the *Supporting Evidence* column. If the drug is contraindicated for the given symptom/phenotype it will also be shown in this column.

Symptom	**ID**	**Drug Candidate**	**Score**	**Reference**
Muscular dystrophy	HP:0003560	Methylprednisolone	0.993	https://pubmed.ncbi.nlm.nih.gov/17541998/
		Resveratrol	0.963	https://www.nature.com/articles/s41598-020-77197-6
		Tofisopam	0.919	https://extrapharmacy.ru/grandaxin-tofisopam-50mg-60tabs
Respiratory insufficiency	HP:0002093	Methylprednisolone	0.984	https://jintensivecare.biomedcentral.com/articles/10.1186/s40560-018-0321-9
		Fedratinib	0.981	None
		Sorafenib	0.975	Can cause pneumonia: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3961597/
Gowers sign	HP:0003391	Fedratinib	0.994	None
		Bosutinib	0.991	None
		Nintedanib	0.990	None
Global developmental delay	HP:0001263	Fedratinib	0.995	None
		Sorafenib	0.994	None
		Bosutinib	0.994	None
Hyporeflexia	HP:0001265	Fedratinib	0.996	None
		Sunitinib	0.994	None
		Bosutinib	0.994	None
Proximal muscle weakness	HP:0003701	Fedratinib	0.997	Can produce muscle weakness: https://medlineplus.gov/druginfo/meds/a619058.html
		Bosutinib	0.995	None
		Methylprednisolone	0.995	Can produce weakness: https://erj.ersjournals.com/content/21/2/377.2#:$\sim$:text=Methylprednisolone%20is%20often%20given%20in,weakness%20following%20high%2Ddose%20steroids.
Intellectual disability	HP:0001256	Fedratinib	0.996	None
		Sorafenib	0.995	None
		Bosutinib	0.995	None
Calf muscle pseudohypertrophy	HP:0003707	Methylprednisolone	0.970	https://www.britannica.com/science/pseudohypertrophy
		Ruxolitinib	0.967	https://www.sciencedirect.com/science/article/pii/S147148921630100X

		Fedratinib	0.948	None
Elevated serum creatine kinase	HP:0003236	Methylprednisolone	0.994	Can increase creatinine: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4275145/
		Fedratinib	0.989	Can increase more: https://jamanetwork.com/journals/jamaoncology/fullarticle/2330618
		Bosutinib	0.982	Can increase more: https://www.sciencedirect.com/science/article/pii/S2152265017305840
Abnormal EKG	HP:0003115	Methylprednisolone	0.982	Can affect EKG: https://pubmed.ncbi.nlm.nih.gov/29668335/
		Patisiran	0.879	None
		Silodosin	0.878	None
Arrhythmia	HP:0011675	Methylprednisolone	0.989	Can produce arrhythmia: http://www.ijps.ir/article_2090.html#:$\sim$:text=Cardiac%20dysrhythmias%20have%20been%20reported,turn%2C%20may%20initiate%20cardiac%20dysrhythmias.
		Fedratinib	0.980	None
		Sorafenib	0.979	None
Waddling gait	HP:0002515	Fedratinib	0.991	Can produce gait: https://www.accessdata.fda.gov/drugsatfda_docs/nda/2019/212327Orig1s000MultidisciplineR.pdf
		Sorafenib	0.990	Can produce gait: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4094497/
		Midostaurin	0.990	None
Dilated cardiomyopathy	HP:0001644	Methylprednisolone	0.993	https://pubmed.ncbi.nlm.nih.gov/25614863/
		Adefovir dipivoxil	0.980	None
		Milrinone	0.966	https://pubmed.ncbi.nlm.nih.gov/10488574/#:$\sim$:text=Conclusion%3A%20Milrinone%20lactate%20is%20an,and%20IV%20of%20heart%20failure.
Flexion contracture	HP:0001371	Fedratinib	0.997	None
		Sorafenib	0.996	https://pubmed.ncbi.nlm.nih.gov/35274715/
		Bosutinib	0.995	None
Specific learning disability	HP:0001328	Fedratinib	0.984	None
		Sorafenib	0.978	None
		Sunitinib	0.977	https://pubmed.ncbi.nlm.nih.gov/27046396/
Skeletal muscle atrophy	HP:0003202	Fedratinib	0.995	None
		Ruxolitinib	0.994	None

		Sunitinib	0.993	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4413636/
Hypoventilation	HP:0002791	Methylprednisolone	0.990	https://jintensivecare.biomedcentral.com/articles/10.1186/s40560-018-0321-9
		Resveratrol	0.966	None
		Fedratinib	0.993	None
Calf muscle hypertrophy	HP:0008981	Methylprednisolone	0.978	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2879072/
		Fedratinib	0.977	None
		Resveratrol	0.976	https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0083518
Motor delay	HP:0001270	Fedratinib	0.995	None
		Sunitinib	0.994	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6586148/
		Vincristine	0.993	None
Generalized hypotonia	HP:0001290	Fedratinib	0.982	None
		Sorafenib	0.980	None
		Primidone	0.980	None
Cardiomyopathy	HP:0001638	Methylprednisolone	0.995	https://pubmed.ncbi.nlm.nih.gov/7971647/
		Resveratrol	0.974	https://onlinelibrary.wiley.com/doi/full/10.1002/fsn3.92
		Adefovir dipivoxil	0.971	None
Hyperlordosis	HP:0003307	Methylprednisolone	0.986	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4897302/
		Fedratinib	0.982	None
		Sorafenib	0.980	None
Congestive heart failure	HP:0001635	Methylprednisolone	0.979	https://www.sciencedirect.com/science/article/pii/S1071916414005843#:$\sim$:text=Methylprednisolone%20improved%20HF%20outcomes.,of%20patients%20from%20the%20study.
		Daunorubicinol	0.957	Can produce cardiotoxicity: https://www.sciencedirect.com/topics/medicine-and-dentistry/daunorubicinol
		Adefovir dipivoxil	0.946	None
Delayed speech and language development	HP:0000750	Fedratinib	0.994	None
		Midostaurin	0.993	None
		Sunitinib	0.993	None
Scoliosis	HP:0002650	Sorafenib	0.995	None
		Fedratinib	0.995	None
		Midostaurin	0.994	None
Progressive muscle weakness	HP:0003323	Methylprednisolone	0.999	Can cause weakness: https://pubmed.ncbi.nlm.nih.gov/14629908/

		Resveratrol	0.985	https://pubmed.ncbi.nlm.nih.gov/33239684/
		Patisiran	0.960	None
Cognitive impairment	HP:0100543	Sunitinib	0.997	None
		Ruxolitinib	0.997	Can produce cognitive impairment: https://pubmed.ncbi.nlm.nih.gov/24661373/
		Bosutinib	0.997	https://pubmed.ncbi.nlm.nih.gov/34484904/

9.10. Drug Candidates AD

Table S11: Table showing the drug candidates with the highest scores for each symptom/phenotype obtained in the AD KG. Any evidence that supports the prediction will be shown in the *Supporting Evidence* column. If the drug is contraindicated for the given symptom/phenotype it will also be shown in this column.

Symptom	Symptom ID	Candidate	Score	Evidence?
Personality changes	HP:0000751	flortaucipir F 18	0.986	https://www.sciencedirect.com/science/article/abs/pii/S0006322321015663
		fedratinib	0.979	May cause: https://medlineplus.gov/druginfo/meds/a619058.html
		lansoprazole	0.978	None
Dysphagia	HP:0002015	fedratinib	0.998	None
		midostaurin	0.998	Causes no dysphagia ? https://www.ons.org/cjon/23/6/midostaurin-nursing-perspectives-managing-treatment-and-adverse-events-patients-flt3
		nintedanib	0.997	None
Alzheimer disease	HP:0002511	Resveratrol	0.983	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5664214/
		pexidartinib	0.980	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8101105/
		memantine	0.980	https://pubmed.ncbi.nlm.nih.gov/16906789/
Cerebral cortical atrophy	HP:0002120	midostaurin	0.998	None
		fedratinib	0.998	None
		sunitinib	0.998	treats Brain Cancer: https://clinicaltrials.gov/ct2/show/NCT00923117
Abnormality of extrapyramidal motor function	HP:0002071	midostaurin	0.991	None
		fedratinib	0.990	None
		bosutinib	0.989	None
Dementia	HP:0000726	midostaurin	0.995	https://www.sciencedirect.com/topics/chemistry/midostaurin
		fedratinib	0.995	None
		pazopanib	0.994	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5757517/
Babinski sign	HP:0003487	midostaurin	0.998	None
		fedratinib	0.997	None
		sunitinib	0.997	None
Lower limb hyperreflexia	HP:0002395	flortaucipir F 18	0.964	None
		Donepezil	0.956	None
		Clioquinol	0.956	None
Dysarthria	HP:0001260	midostaurin	0.999	None
		fedratinib	0.999	None

		sunitinib	0.998	None
Memory impairment	HP:0002354	flortaucipir F 18	0.982	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8175307/
		pexidartinib	0.977	https://www.alzdiscovery.org/uploads/cognitive_vitality_media/Pexidartinib-Cognitive-Vitality-For-Researchers.pdf
		sorafenib	0.974	None
Dystonia	HP:0001332	fedratinib	0.997	None
		midostaurin	0.997	None
		bosutinib	0.995	None
Optic ataxia	HP:0031868	Clioquinol	0.986	None
		Donepezil	0.986	None
		Memantine	0.970	Optic nerve atrophy: https://pubmed.ncbi.nlm.nih.gov/26666888/
Myoclonus	HP:0001336	fedratinib	0.996	None
		midostaurin	0.996	None
		bosutinib	0.996	None
Apraxia	HP:0002186	midostaurin	0.989	None
		fedratinib	0.988	None
		nintedanib	0.988	None
Seizure	HP:0001250	fedratinib	0.999	Can cause: https://www.mskcc.org/cancer-care/patient-education/medications/fedratinib
		midostaurin	0.999	None
		bosutinib	0.998	Can cause: https://www.ema.europa.eu/en/documents/product-information/bosulif-epar-product-information_en.pdf
Gait disturbance	HP:0001288	fedratinib	0.986	None
		bosutinib	0.977	None
		midostaurin	0.973	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8301989/
Neurofibrillary tangles	0.973	flortaucipir F 18	0.974	https://pubchem.ncbi.nlm.nih.gov/compound/70957463
		cycloserine	0.962	https://pubmed.ncbi.nlm.nih.gov/36159454/
		lansoprazole	0.961	https://pubmed.ncbi.nlm.nih.gov/24900410/
Spastic tetraparesis	HP:0001285	duloxetine	0.959	None
		flortaucipir F 18	0.952	None
		metformin	0.951	None
Agnosia	HP:0010524	Donepezil	0.980	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3504981/
		Clioquinol	0.980	None
		Memantine	0.967	https://pubmed.ncbi.nlm.nih.gov/19898670/

9.11. Drug Candidates ALS

Table S12: Table showing the drug candidates with the highest scores for each symptom/phenotype obtained in the ALS KG. Any evidence that supports the prediction will be shown in the *Supporting Evidence* column. If the drug is contraindicated for the given symptom/phenotype it will also be shown in this column.

Symptom	Symptom ID	Candidate	Score	Evidence?
Sleep apnea	HP:0010535	Riluzole	0.808	https://pubmed.ncbi.nlm.nih.gov/11732759/
		Gabapentin	0.767	Can cause: https://pubmed.ncbi.nlm.nih.gov/28116804/
		Vitamin E	0.756	https://pubmed.ncbi.nlm.nih.gov/23389837/
Degeneration of anterior horn cells	HP:0002398	Riluzole	0.821	Spinal muscular atrophy: https://pubmed.ncbi.nlm.nih.gov/14623733/
		tacrolimus	0.785	Not significant: https://www.nature.com/articles/sc2015172
		brilliant Blue G	0.768	Can help ELA: https://peerj.com/articles/3064/
Dysarthria	HP:0001260	hexachlorophene	0.976	None
		dabrafenib	0.972	None
		dichlorophen	0.954	None
Skeletal muscle atrophy	HP:0003202	hexachlorophene	0.953	None
		dabrafenib	0.935	Can cause: https://pubmed.ncbi.nlm.nih.gov/32898388/
		quercetin	0.907	https://pubmed.ncbi.nlm.nih.gov/25614714/#:$\sim$:text=Together%2C%20these%20findings%20suggest%20that,induced%20muscle%20inflammation%20and%20sarcopenia.
Muscle weakness	HP:0001324	hexachlorophene	0.951	None
		dabrafenib	0.944	None
		quercetin	0.989	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6356612/#:$\sim$:text=Taken%20together%2C%20the%20findings%20from,sarcolemmal%20action%20potential%20propagation%20impairment.
Muscle spasm	HP:0003394	hexachlorophene	0.975	None
		dabrafenib	0.958	Can cause: https://www.macmillan.org.uk/cancer-information-and-support/treatments-and-drugs/dabrafenib-and-trametinib
		dichlorophen	0.945	None
Amyotrophic lateral sclerosis	HP:0007354	hexachlorophene	0.911	https://pubmed.ncbi.nlm.nih.gov/25987361/

		oleic acid	0.884	https://pubmed.ncbi.nlm.nih.gov/29760648/
		dabrafenib	0.881	None
Dysphagia	HP:0002015	hexachlorophene	0.987	None
		dabrafenib	0.980	None
		dichlorophen	0.968	None
Fasciculations	HP:0002380	hexachlorophene	0.920	None
		oleic acid	0.891	Can increase: https://www.sciencedirect.com/science/article/pii/S0006899314005861?via%3Dihub
		dabrafenib	0.998	None
Degeneration of the lateral corticospinal tracts	HP:0002314	hexachlorophene	0.829	None
		dabrafenib	0.787	None
		celecoxib	0.787	None
Pseudobulbar paralysis	HP:0007024	Riluzole	0.788	https://www.nejm.org/doi/full/10.1056/NEJM199403033300901
		Gabapentin	0.727	None
		celecoxib	0.720	None
Hyperreflexia	HP:0001347	hexachlorophene	0.983	Can cause: https://pubchem.ncbi.nlm.nih.gov/compound/Hexachlorophene#section=Human-Toxicity-Excerpts
		dabrafenib	0.977	None
		dichlorophen	0.963	None
Spasticity	HP:0001257	hexachlorophene	0.989	None
		dabrafenib	0.986	None
		sotorasib	0.970	None