Deep Learning Can Identify Explainable Reasoning Paths of Mechanism of Drug Action for Drug Repurposing from Multilayer Biological Network

Jiannan Yang

City University of Hong Kong

Zhen Li

HUST <https://orcid.org/0000-0001-8037-4245>

William Wu

Chinese University of Hong Kong <https://orcid.org/0000-0002-5662-5240>

Shi Yu

University of South California

Qian Chu

Tongji Hospital <https://orcid.org/0000-0001-8192-7630>

Qingpeng Zhang (✉ qingpeng.zhang@cityu.edu.hk)

City University of Hong Kong <https://orcid.org/0000-0002-6819-0686>

Article

Keywords:

Posted Date: February 11th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1269212/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Deep Learning Can Identify Explainable Reasoning Paths of Mechanism of Drug Action for Drug Repurposing from Multilayer Biological Network

Jiannan Yang¹, Zhen Li², William Ka Kei Wu³, Shi Yu^{4,5}, Qian Chu^{6,*}, and Qingpeng Zhang^{1,*}

¹School of Data Science, City University of Hong Kong, Hong Kong SAR, China

²Department of Radiology, Tongji Hospital, Huazhong University of Science and Technology, Wuhan, China

³Department of Anaesthesia and Intensive Care, Chinese University of Hong Kong, Hong Kong SAR, China

⁴The USC Norris Center for Cancer Drug Development, University of Southern California, Los Angeles, CA, USA

⁵Keck School of Medicine, University of Southern California, Los Angeles, CA, USA

⁶Department of Thoracic Oncology, Tongji Hospital, Huazhong University of Science and Technology, Wuhan, China

*correspondence: qingpeng.zhang@cityu.edu.hk (Q Zhang), qianchu@tjh.tjmu.edu.cn (Q Chu)

ABSTRACT

The discovery and repurposing of drugs require a deep understanding of the mechanism of drug action (MODA). Existing computational methods model these mechanisms with the help of protein-protein interaction (PPI) network. However, the molecular interactions of drugs in the human body are far beyond the PPI network, and the lack of interpretability of these models hinders their practical applications. In this study, we propose iDPath, an interpretable deep learning-based path-reasoning framework to identify potential drugs for the treatment of diseases by capturing the MODA based on simulating the paths from drugs to diseases in the human body. We create a multilayer biological network consisting of a gene regulatory layer, a PPI layer, a protein-chemical interaction layer, and a chemical-chemical interaction layer to comprehensively characterize the molecular interactions in the human body. Based on this multilayer biological network, iDPath utilizes a graph convolutional network module to capture the global connectivity information of the human molecular network; and a long short-term memory neural network module to capture the detailed mechanisms of drug action by modeling the sequential dependencies of the MODA-related biological paths between drugs and diseases. Moreover, iDPath applies two attention modules - node attention and path attention - to further enhance its interpretability. Experiments with drug screen data show that iDPath outperforms state-of-the-art machine learning methods on a general drug repurposing task featuring 1,993 drugs and 2,794 diseases. Further investigations demonstrate that iDPath can identify explicit critical paths that are consistent with clinical evidence. To demonstrate the practical value of iDPath, we apply it to identify potential drugs for the treatment of prostate cancer. Results show that iDPath can successfully discover new FDA-approved drugs. This research provides a novel interpretable artificial intelligence perspective on drug discovery.

Introduction

Artificial intelligence (AI) has recently shown the huge potential to subvert the typical drug discovery process¹. Scientists are now using deep learning technologies to discover candidate drugs for the treatment of COVID-19²⁻⁴, Alzheimer's disease⁵, cancers⁶, and so on. Among various applications, drug repurposing, the identification of the new use of approved or investigational drugs that are outside of the original medical indication, can shorten the time of drug development while ensuring safety, and thus attracts special attention from drug discovery communities and industries^{7,8}. Existing computational approaches aim to precisely model the effects of drugs in the human body. A common approach is to model the outcomes of drugs in a clinical environment, such as using electronic health records to discover the efficacy of drugs on a specific population⁹, or emulating clinical trials on real-world patient data¹⁰.

With the development of high-throughput omics technologies, the detailed characterization of the molecular interactions of drugs in the human body becomes possible¹¹. The protein-protein interaction (PPI) network serves as a “skeleton” for body's signaling circuitry¹² and shows tremendous power in guiding drug discovery¹³⁻¹⁶. A series of studies explored the potential of mining the network properties of drugs in the PPI network in synergistic drug combination identification¹³ and drug repurposing¹⁴. A recent study introduced advanced graph-based deep learning approaches to the identification of anti-cancer drug combinations by learning the graphic representations of the PPI network¹⁷. However, the molecular network in the human body is not limited to the PPI network, the gene regulatory mechanisms¹⁸, the binding work of the proteins and chemicals¹⁹, and the interactions of the chemicals^{20,21} also play a role in the mechanism of drug action (MODA). These processes rely on the drug's interactions with various proteins and chemicals in the human body^{22,23}. Usually, the MODA is described by biological pathways, a series of biochemical and molecular steps to achieve a specific function or to produce a certain

product²⁴. Such biological pathways can be naturally denoted as a series of paths in the biological network. Furthermore, instead of targeting specific proteins, some drugs need to take further chemical reactions to be effective²⁵. For example, cytarabine, an important drug in the treatment of acute myeloid leukemia²⁶, must be phosphorylated intracellularly to a nucleotide (cytarabine 5'-triphosphate, Ara-CTP) before it can exert its cytotoxic effect²⁷.

Recent studies have introduced novel machine learning models to drug discovery based on human biological networks²⁸⁻³¹. Most of these networks are combinations of PPI network and other networking information such as the drug interactions and drug similarities^{30,31}. The rich patterns hidden in the molecular interactions (such as chemical-chemical interactions) in the human body have not been fully used in drug discovery and repurposing tasks. Whether the addition of these molecular interactions can facilitate the accurate modeling of drug effects is under-researched. Furthermore, nearly all of these machine learning models are *black-box*. Users cannot explain why the model generates a certain prediction. The lack of interpretability hinders machine learning models' potential in practical drug discovery tasks. The need for explainable machine learning models led to the development of a series of novel neural network architectures, such as attribution methods^{32,33} and knowledge-graph-based models^{34,35}. These models have been further applied in healthcare³⁶⁻³⁹, such as using biological-informed neural networks to identify anti-cancer drug combinations³⁸ and predict disease prognostic based on comorbidity network³⁹. However, whether these explainable modules can accelerate drug discovery and further enhance the knowledge of the MODA is unknown.

To fill the aforementioned research gaps, based on our previous research on interpretable machine learning models^{4,17,39,40}, we propose the interpretable Deep learning-based **Path**-reasoning framework for drug repurposing (**iDPath**), which captures the mechanisms of drug actions by identifying the paths from drugs to diseases in the human body. To accurately characterize the MODA, we build a multilayer

biological network instead of using the PPI network alone. The multilayer biological network is the integration of a gene regulatory layer, a protein-protein interaction layer, a protein-chemical interaction layer, and a chemical-chemical interaction layer. Starting from this multilayer biological network, iDPath utilizes a graph convolutional network module to capture the global connectivity information of the human molecular network; and a long-short term memory neural network module to capture the detailed mechanisms of drug action based on the shortest paths between drugs and diseases. Furthermore, iDPath introduces two attention modules, namely the path attention and the node attention, to enhance model interpretability. Experiments with drug screen data demonstrate the superior performance of iDPath in a general drug repurposing task featuring 1,993 drugs and 2,794 diseases. Further investigations demonstrate that iDPath can identify explicit critical paths that are consistent with clinical evidence. To demonstrate the practical value of iDPath, we apply it to identify potential drugs for the treatment of prostate cancer. Results show that iDPath can successfully discover new FDA-approved drugs. These results indicate that iDPath can facilitate drug discovery and repurposing, and has the potential to address other computational chemistry and biology tasks involving the understanding of the molecular interactions in the human body.

Methods

In this section, we describe the datasets and the proposed iDPath model, as well as baseline models for drug repurposing, including DeepWalk, Graph Convolutional Networks (GCN), Long Short-Term Memory (LSTM) networks, and Knowledge-aware Path Recurrent Network (KPRN).

Data

Multilayer Biological Network

Gene regulatory network (GRN) layer. The gene regulatory network is adopted from RegNetwork⁴¹, which collects the experimentally validated and the predicted regulations based on the transcription factor

binding sites. The edges in RegNetwork start from transcription factor (TF) and microRNA (miRNA) and target the regulated genes. In total, RegNetwork provides us with 369,277 gene regulations between 1,456 TFs, 1,904 miRNAs and 19,719 genes.

Protein-protein interaction (PPI) layer. The PPI network consists of information from two sources. The first dataset, STRING dataset⁴², is the most comprehensive database of known and predicted PPIs till now, with more than 1,380 million PPIs among over 9 million proteins. We only keep the PPIs in the human body (homo sapiens) and at high confidence or better (confidence level > 0.7). Another PPI dataset is the human interactome built by Cheng et al.¹⁴. This dataset is harnessed from multiple databases with experimental evidence. After pre-processing, our PPI network contains 614,970 interactions connected by 13,758 proteins.

Protein-chemical interaction (PCI) layer. We obtain a PCI network by curating from STITCH database⁴³, which is the most comprehensive database of known and predicted interactions between chemicals and proteins till now. We select PCIs in the human body (homo sapiens) at high confidence or better (confidence level > 0.7). The processed PCI network consists of 203,551 interactions among 9,393 proteins and 73,199 chemicals.

Chemical-chemical interaction (CCI) layer. CCI network is curated from STITCH⁴³ database and further processed by selecting CCIs in the human body (homo sapiens) at high confidence or better (confidence level > 0.7). The processed CCI network has 396,284 interactions among 107,055 chemicals.

Constructing the multilayer biological network. We construct a multilayer biological network by mapping all the entities in GRN, PPI, PCI and CCI to same nomenclature. The proteins are named by their encoded genes and the miRNAs are mapped to their corresponding genes by BioMart⁴⁴. All the genes are encoded to their Entrez IDs⁴⁵. All the chemicals are denoted by their PubChem CIDs (Compound ID

number)⁴⁶.

Therapeutic drug-disease pairs

For the drug repurposing task, we collect therapeutic drug-disease pairs from Therapeutic Target Database (TTD)⁴⁷, which provides the known and explored therapeutic protein and nucleic acid targets, the targeted disease, and the pathway information of tens of thousands of drugs. We only keep the FDA-approved drugs in TTD and map them to PubChem CID in order to be consistent with the chemicals in our multilayer biological network. The diseases in TTD are in the ICD-11 coding system and are mapped to their corresponding ICD-10 codes. The cleaned dataset of therapeutic drug-disease pairs includes 1,993 drugs and 2,794 diseases and constitutes 19,500 pairs.

Drug-protein associations and drug-chemical associations

We collect drug-protein associations from 4 datasets: the PCI network from STITCH⁴³, the drug-protein associations built by Cheng et al.¹⁴, therapeutic target database (TTD)⁴⁷, and DrugBank⁴⁸. The drug-chemical associations are extracted from STITCH by selecting the compounds that are drugs. DrugBank is a commonly used database containing comprehensive molecular information about drugs, their mechanisms, interactions and targets. The aggregated dataset contains 85,305 drug-protein associations between 20,405 drugs and 7,796 proteins, 83,271 drug-chemical associations between 4,630 drugs and 12,042 chemicals. All the drugs and chemicals are denoted by their PubChem CIDs, and all the proteins are represented by their encoded genes using Entrez ID.

Disease-gene associations and disease-miRNA associations

The disease-gene associations include genes and variants associated with human diseases, curated from DisGeNET⁴⁹ by selecting expert-curated repositories. The miRNAs associated with human diseases come from the Human microRNA Disease Database (HMDD)⁵⁰, which is a database about curated experiment-

supported evidence for human microRNA (miRNA) and disease associations. All the genes, variants, and miRNAs are mapped to Entrez IDs, and diseases are mapped to ICD-10 codes. After processing, we have 230,837 associations among 7,559 genes, 6,830 variants, 705 miRNAs with 5,602 diseases.

Overall architecture of iDPath

The iDPath framework for drug repurposing is presented in Figure 1. The MODA-related biological paths are identified by the shortest path between drugs and diseases (Figure 1b) in the multilayer biological network (Figure 1a). To learn the global connectivity information of the multilayer biological network, iDPath first utilizes a three-layer GCN to learn the embeddings of associated nodes in the multilayer biological network. Then, to capture the detailed MODA patterns, the embeddings of the nodes along the shortest path between a drug and a disease are fed into an LSTM module to model their sequential dependencies. iDPath also introduces two attention modules to aggregate the embeddings of nodes and paths - path attention and node attention. These two attention modules are capable of discriminating the contribution of different nodes to one MODA-related biological path, as well as the contribution of different paths to the final prediction.

GCN to capture the global connectivity information of the multilayer biological network

With the uniform nomenclature of the nodes, we aggregate the drug-related information (drug-protein associations, drug-chemical associations), the disease-related information (disease-gene associations and disease-miRNA associations), and the multilayer biological network to the network $G = (V, E)$, where V and E are the node set and edge set, respectively, iDPath introduces a three-layer GCN, following a spatial-based GCN architecture⁵¹, to encode the global connectivity topological information of the multilayer biological network. Suppose there are N nodes in total, the initial embeddings of these nodes are $E \in \mathbb{R}^{N \times d}$ (d is the dimension of the embedding), and the adjacency matrix of network G is A , the

Figure 1. The framework of iDPath on drug repurposing tasks. **a** The multilayer biological network consists of four layers: gene regulatory network layer (one-way red arrows), protein-protein interaction layer (two-way black arrows), protein-chemical interaction layer (two-way purple arrows), and chemical-chemical interaction layer (two-way orange arrows). The blue two-way dashed arrows represent that the two corresponding nodes in different layers are identical. The nodes associated with the drugs and diseases are marked by green dashed lines. **b** The MODA-related biological paths are identified by the shortest paths between drug and disease generated in the multilayer biological network. **c** The schematic representation of the algorithm: the multilayer biological network is fed into three-layer GCN to learn the embeddings of all nodes. The GCN embeddings of nodes along the shortest path between one drug and one disease are fed into an LSTM module to learn the sequential dependencies. Node attention and path attention modules are introduced to aggregate the embeddings of the nodes and paths. The final prediction is the probability that one drug is effective for one disease.

computation formula of layer l of GCN is shown as below:

$$E^{l+1} = \sigma(D^{-1}(A+I)D^{-1}E^lW^l), \quad (1)$$

where D is the diagonal node degree matrix, I is the identity matrix.

MODA-related biological paths

The MODA is dependent on the interactions of drugs with molecules in the human body, which can be represented as a series of paths in our multilayer biological network²⁴. To accurately model the effects of drugs, we need to identify informative paths to represent the MODA in an efficient way.

Among various approaches to generate these paths, the shortest paths are the most informative ones, which have already been applied in many drug-related studies^{13,14,52}. Considering the huge amount of nodes and edges of our multilayer biological network, finding the shortest paths between all drug-disease pairs would be time-consuming. To overcome this challenge, we introduced GPU-accelerated graph algorithms implemented by cuGraph⁵³. For a drug and a disease, the shortest path between them is connected by their associated nodes in the multilayer biological network. Given a drug and a disease, the shortest paths between this pair form a set $PATH = \{path_1, path_2, \dots, path_L\}$, where $path_i = \{node_{drug} \rightarrow node_{m1} \rightarrow node_{m2} \rightarrow \dots \rightarrow node_{disease}\}$, $node_{m1}$ and $node_{m2}$ denote the middle nodes of one path, and L is the number of shortest paths. Since L differs among drug-disease pairs, we choose a fixed value for L by randomly sampling from the shortest paths set $PATH$.

LSTM layer

Given a drug-disease pair, the embeddings E_{GCN} generated by the GCN, and the shortest path set between this pair $PATH$, we employ LSTM⁵⁴ to encode both long-term and short-term dependencies in a MODA-related biological path. Such sequential dependencies are crucial to the model intelligibility. Meanwhile, we introduce the type of these nodes to strengthen the model's capability of identifying different nodes. Here, we consider four types: protein (gene), chemical, drug, and disease. Their embeddings $E_{TYPE} \in \mathbb{R}^{4 \times d}$ are randomly initialized. Therefore, given one node $node_j$ of one path $path_p$, the input to LSTM is the concatenation of its GCN embedding $e_j = E_{GCN}[node_j]$ and type embedding $e'_j = E_{TYPE}[node_j]$, that is:

$$x_j = e_j \oplus e'_j, \quad (2)$$

where $\oplus$ denotes the concatenation operation. The hidden state h_{j-1} generated by the previous node $node_{j-1}$ in the same path and x_j are used to learn the hidden state of the input of the next node $node_{j+1}$,

which is defined as follows:

$$\begin{aligned}
i_j &= \sigma(W_i x_j + W_h h_{j-1} + b_i), \\
f_j &= \sigma(W_f x_j + W_h h_{j-1} + b_f), \\
g_j &= \tanh(W_g x_j + W_h h_{j-1} + b_g), \\
o_j &= \sigma(W_o x_j + W_h h_{j-1} + b_o), \\
c_j &= f_j \odot c_{j-1} + i_j \odot g_j, \\
h_j &= o_j \tanh(c_j),
\end{aligned} \tag{3}$$

where i_j , f_j , g_j , and o_j are the input, forget, cell, and output gates, respectively; $c_j \in \mathbb{R}^{d'}$ and $h_j \in \mathbb{R}^{d'}$ are the cell state and hidden state at path step j , and d' is the dimension of the output; W_i , W_f , W_g , W_o , and W_h are learnable weights, and b_i , b_f , b_g , and b_o are bias; σ denotes the sigmoid function and $\odot$ is the Hadamard product. The hidden states h_j for each path step are aggregated to the attention modules for the representation of the whole path and final prediction. Since the shortest paths are not of equal length, we borrow the padding method as follows⁵⁵. Suppose the maximum length of one path is set to l_{max} , for the paths that are shorter than l_{max} , we use a padding value pad (such as 0) to fill the path, and the following processing will ignore these padding positions not to affect the performance.

Node attention and path attention

Attention mechanism⁵⁶ is widely used in various deep learning tasks to enhance model intelligibility¹⁷. In this study, we introduced two attention modules to separately learn the importance of different nodes to one MODA-related biological path, as well as the importance of different paths to the final prediction.

Node attention. For one shortest path $path_p$ between one drug-disease pair, the hidden states of each node generated by LSTM layers are $H_p \in \mathbb{R}^{l_{max} \times d'}$ where $H_p = \{h_1, h_2, \dots, h_{pad}, h_{pad}\}$ and h_{pad} are the hidden states of the padding step. We first transfer all the hidden values at the padding positions to negative

infinity for the following Softmax transformation. Then, we applied a linear layer with the Softmax activation to aggregate the embeddings to one numeric value denoting the importance (weight). That is

$$\Omega_p = H_p W_n, \quad (4)$$

where $W_n \in \mathbb{R}^{d' \times 1}$ is the learnable parameter, $\Omega_p \in \mathbb{R}^{l_{max} \times 1}$ denotes the weights of each node in path $path_p$, where $\Omega_p = \{\omega_1, \omega_2, \dots, \omega_{l_{max}}\}$. For one node j in $path_p$, its weight is computed as follows:

$$\hat{\omega}_j = \frac{e^{\omega_j}}{\sum_{k=1}^{l_{max}} e^{\omega_k}}. \quad (5)$$

Then, we aggregate the hidden states of these nodes weighted by $\hat{\omega}_i$ to get the embedding of $path_p$:

$$e_{path_p} = \sum_{k=1}^{l_{max}} \hat{\omega}_k h_k. \quad (6)$$

Path attention. After the aggregation operation, for one drug-disease pair, the embeddings of their shortest paths are: $E_{PATH} = \{e_{path_1}, e_{path_2}, \dots, e_{path_L}\}$, the path attention layer is similar to the node attention:

$$\begin{aligned} \Omega_{PATH} &= e_{PATH} W_p, \\ \hat{\omega}_{path_p} &= \frac{e^{\omega_{path_k}}}{\sum_{k=1}^L e^{\omega_{path_k}}}, \\ \hat{y} &= \sigma_p \left(\sum_{k=1}^L e_{path_k} \hat{\omega}_{path_k} \right), \end{aligned} \quad (7)$$

where σ_p is the sigmoid function. The final prediction $\hat{y}$ is the aggregation of the weighted embedding of each path, indicating the probability that the drug is effective in treating the disease.

Objective function

In this study, we treat the training of iDPath as a binary classification task following common practice. That is, besides all the therapeutic drug-disease pairs (marked as 1), we introduce the negative sampling to get an equal number of non-therapeutic drug-disease pairs (marked as 0). The objective function used by iDPath is the binary-cross-entropy loss with l_2 regularization:

$$\mathbb{L} = \sum -y \log \hat{y} - (1 - y) \log (1 - \hat{y}) + \lambda ||\Theta||_2^2, \quad (8)$$

where Θ is the set of parameters to be learnt in iDPath, λ is the l_2 regularizer to prevent over-fitting.

Baselines

In this section, we describe a set of baseline models to compare with, including DeepWalk, GCN, LSTM, and Knowledge aware Path Recurrent Network (KPRN).

DeepWalk. DeepWalk⁵⁷ is a widely used graph embedding approach by modeling a stream of short random walks and has already been introduced to several drug-related tasks, such as drug-target identification⁵⁸.

GCN. GCN⁵¹ is widely applied to many drug-related tasks, such as drug discovery using the drug’s Smiles features⁵⁹, and anti-cancer drug combination identification¹⁷. As a component of iDPath, we apply GCN individually to test its performance on the drug repurposing task. Specifically, besides the basic graph convolutional layer, we introduce a fully connected layer to combine the embeddings of drug and disease for the final prediction.

LSTM. LSTM has been applied to drug discovery^{60,61}. We apply a vanilla LSTM network and use the last hidden states of each path as its representation. A two-layer fully connected layer followed by the

LSTM layer is employed to generate the final prediction.

KPRN. Knowledge aware Path Recurrent Network (KPRN)³⁵ is an advanced path-based model for a reasonable recommendation based on a knowledge graph. We apply KPRN to the drug repurposing task by feeding it the same input as iDPath.

Performance evaluation

All models are trained on a binary classification task, where we utilize commonly used metrics to evaluate and fine-tune the models. The metrics used in the binary classification task include accuracy, recall, the area under the receiver operating characteristic curve (AUROC), and the area under the precision-recall curve (AUPRC). The drug repurposing task can be viewed as a recommendation task, that is, for each disease in our dataset, we use the model to calculate the probability that one drug is effective in treating one disease and then rank all the drugs based on the probability. We introduce two commonly used metrics in the recommendation system, $NDCG@K$ and $Hit@K$. Details of these metrics and the experiment setup can be found in the supplementary information.

Results and Discussions

iDPath consistently outperforms baselines in drug repurposing.

In general, baseline models can be classified into two categories: graph-embedding-based models (GCN and DeepWalk), and path-based models (LSTM and KPRN). iDPath presents a modeling framework that combines graph-embedding-based and path-based approaches. We compare the performance of iDPath with baselines in the drug repurposing problem. As shown in Figure 2a and c, iDPath outperforms all the baselines with an AUPRC of 0.97. In detail, iDPath achieves 91.51% true-negative rate (TN) and 91.23% true-positive rate (TP) in the test dataset, indicating that only less than 10.00% of drug-disease pairs have

not been correctly classified (Figure 2b). The utilization of the shortest paths can significantly improve the performance, as demonstrated by the superior performance of iDPath and path-based models over graph-embedding-based models (Figure 2a and c). These results indicate that the extracted MODA-related biological paths have pharmacological relevance.

Incorporating multiple biological network layers improves the prediction performance.

We further investigate the performance of iDPath with different biological network layers. Existing studies mainly make predictions using the PPI network alone^{13,14}. Here, we evaluate the performance of iDPath with only one layer (PPI, GRN or PCI). Note that CCI cannot be directly linked to diseases, so we do not evaluate the model with only the CCI layer. As shown in Figure 2d, the full multilayer biological network can improve the performance of iDPath. Comparing individual networks, PPI performs the best, followed by GRN, and finally PCI. Note that the iDPath variants combining two or three layers cannot compete with the model with all four layers. Then, we investigate the proportion of nodes and interactions at each network layer in the identified MODA-related biological paths (Figure 3), to examine the impact of different network layers on iDPath performance. We find that nodes and interactions in the PPI layer are not the most prevalent in the identified paths. Instead, GRN nodes and PCI interactions are more common. Combining these results with the dominating role of the PPI network in prediction performance, we find that although the connectivity in the PPI network can capture the key relationships between drugs and diseases, it requires much more information at the GRN, PCI, and CCI layers to further reveal the hidden biological paths related to MODA. By revealing these hidden paths, iDPath achieves higher prediction accuracy.

Figure 2. Performance of iDPath. **a** The Precision-Recall (PR) curves of all the models on the testing set, the values in the bracket denotes the AUPRC. **b** The true-negative rate (TN), false-negative rate (FN), false-positive rate (FP), and true-positive rate (TP) of iDPath on the testing set. **c** The performance ($NDCG@K$) of all the models on the drug recommendation task on the testing set. These models are trained on the binary classification task and used to generate the repurposing probabilities of all the drugs on different diseases in the testing set. **d** The performance of iDPath with different biological network layers. Here GRN-PPI-PCI-CCI denotes the multilayer biological network generated by these four networks, GRN denotes that using gene regulatory network (GRN) alone, and the same goes for PPI, PCI. The K values in **c** and **d** denote the top K drugs used to compute for $NDCG$.

Figure 3. The proportion of nodes (a) and interactions (b) at each network layer in the identified MODA-related biological paths.

iDPath identified the critical paths related to MODA.

We visualize the critical paths of correctly classified drugs for prostate cancer and show one example of drug abiraterone in Figure 4. Here we define the critical paths as the top 50 paths ranked by their weights identified by the path attention module. The top 15 paths are presented in Figure 4. Among 256 shortest paths between abiraterone and prostate cancer, iDPath prefers the paths traversing through the gene targeted by both abiraterone and prostate cancer, such as abiraterone $\rightarrow$ CYP3A4 $\rightarrow$ prostate cancer and abiraterone $\rightarrow$ AR $\rightarrow$ prostate cancer. Previous studies have shown that abiraterone is a moderate inhibitor of CYP3A4⁶², and CYP3A4 is associated with oxidative deactivation of testosterone, which is the etiology of prostate cancer⁶³. Androgen receptor (AR) is highly relevant to the growth and differentiation of prostate cancer⁶⁴, and abiraterone inhibits androgen biosynthesis to control the progression of prostate cancer⁶⁵. Abiraterone is found to be an inhibitor of CYP17A1,⁶⁶ which has also been identified by iDPath. Specifically, the path abiraterone $\rightarrow$ CYP17A1 $\rightarrow$ Hydrogen $\rightarrow$ DL-Pyroglutamic Acid $\rightarrow$ MSLN $\rightarrow$

prostate cancer contributes to the prediction the most among all the CYP17A1 related paths, which is also consistent with previous biological studies⁶⁷.

To further validate that the paths with higher weights are more relevant to the progression of prostate cancer, we perform the KEGG Pathway enrichment analysis⁶⁸ on the proteins of the top 50 paths and bottom 50 paths for the abiraterone-prostate cancer pair. As shown in Figure 4b and 4c, the paths with higher weights focus on the P13K-Akt signaling pathway⁶⁹, regulation of actin cytoskeleton⁷⁰, prostate cancer pathway, and so on, which are highly related to the progression of prostate cancer. For example, the activation of P13K-Akt signaling pathway appears to be characteristic of many aggressive prostate cancers and is more frequently observed as prostate cancer progresses toward a resistant and metastatic disease⁶⁹. In contrast, the paths with lower weights (Figure 4c) are more enriched in the pathways related to other cancers or more general cancer progression, not specific to prostate cancer.

iDPath identified the critical paths to uncover the synergistic effect of drug combinations.

iDPath represents a multilayer network approach to understanding the MODA. The interactions between proteins and chemicals and the interactions among chemicals can reveal more detailed therapeutic effects of individual drugs and potential drug combinations. Among the 16 interacted chemicals of abiraterone in our dataset, docetaxel and cabozantinib are identified as the most relevant contributors to the positive treatment effects of abiraterone to prostate cancer (Figure 4a). We notice that both docetaxel and cabozantinib show synergistic effects on prostate cancer while being used together with abiraterone. The combination of docetaxel and abiraterone can improve radiographic progression-free survival for patients with metastatic castration-sensitive prostate cancer⁷¹. Cabozantinib enhances the anti-prostate cancer activity of abiraterone by inhibiting abiraterone's upregulation of IGFIR Phosphorylation⁷². The identification of these combinations shows that iDPath has the capability to herald the synergistic drug combinations, even iDPath is not explicitly trained to perform this task.

Figure 4. Interpretation of the actions of abiraterone to prostate cancer. **a** The Sankey diagram of the critical paths between abiraterone and prostate cancer identified by iDPath. The colors of edges are determined by the path attention module, and edges with darker colors are more important. The colors of nodes are determined by the node paths generated by the node attention module and nodes with darker colors are more important. **b** and **c** are the dotplots of the KEGG pathway enrichment analysis for the proteins that existed in the top-50 paths and bottom-50 paths between abiraterone and prostate cancer, respectively.

Drug repurposing for prostate cancer.

To demonstrate iDPath's utility in the real-world setting, we apply it to the discovery of potential drugs for treating prostate cancer among 1,572 FDA-approved drugs that have not been labeled as therapeutic drugs for prostate cancer in our dataset. The ten drugs with the highest score, together with their top-3

critical paths and top-3 KEGG pathways, are shown in Table 1. Among the ten drugs identified for prostate cancer, six drugs have already been proved effective in previous studies, including dutasteride⁷³, aspirin⁷⁴, erlotinib⁷⁵, midostaurin⁷⁶, apalutamide⁷⁷, and atorvastatin⁷⁸. Among them, apalutamide has recently been approved for the treatment of prostate cancer⁷⁹, but has not been labeled in our dataset. Specifically, iDPath finds that the most relevant paths for the efficacy of Apalutamide are through Enzalutamide or Abiraterone (both are FDA-approved drugs for the treatment of prostate cancer and labeled as therapeutic in our dataset), where the combination with Abiraterone has already been proved synergistic in a recent study⁸⁰. For other drugs identified as therapeutic but not officially approved, the KEGG pathway enrichment analysis shows that the proteins that existed in their critical paths enriched in prostate cancer-related pathways, such as PI3K-Akt signaling pathway⁸¹ and prostate cancer pathway.

Conclusion

In this study, we propose iDPath, an advanced deep learning framework to identify explainable biological paths to characterize the mechanism of drug actions and predict the drugs that can be repurposed for treating certain diseases. iDPath is built on a multilayer biological network consisting of GRN, PPI, PCI, and CCI networks. The proposed model achieves superior prediction performance compared to state-of-the-art models on a general drug repurposing task. Furthermore, we find that extending the PPI network to a multilayer biological network of the human body can significantly improve the prediction performance in drug repurposing. We investigate the identified critical paths of drugs to prostate cancer and find that the critical paths are consistent with the known mechanism of the drug action. At last, we apply iDPath to the challenging problem of identifying potential drugs for the treatment of prostate cancer. Results show that iDPath can effectively identify the newly approved drugs not recorded by the database. We believe iDPath can bring revelation to the explainable deep learning technologies to drug discovery.

Drug	Critical Paths (Top-3)	KEGG Pathways (Top-3)
Dutasteride	SRD5A2	PI3K-Akt signaling pathway
	ORM1 → TMSB4X	Lipid and atherosclerosis
	SRD5A1 → UGT2B17	Hepatitis B
Aspirin	PLAUR	Proteoglycans in cancer
	FASLG	Hepatitis B
	TGFB1	Lipid and atherosclerosis
Erlotinib	STAT3	EGFR tyrosine kinase inhibitor resistance
	CYP3A5	Chemical carcinogenesis - receptor activation
	STAT3 → NT5C2	Prostate cancer
Nicergoline	ADRA1A	Steroid hormone biosynthesis
	ADRA1A → Hydrogen → Azelaic Acid → SRD5A2	Prostate cancer
	ADRA1A → Triphosadenine → Diethylstilbestrol → SLC30A3	Cysteine and methionine metabolism
Acetohydroxamic Acid	MMP13	Human T-cell leukemia virus 1 infection
	MMP8 → KLK2	Prostate cancer
	MMP13 → MMP7	Human cytomegalovirus infection
Midostaurin	RET	PI3K-Akt signaling pathway
	AURKB	Chemical carcinogenesis - receptor activation
	CYP3A5	Chemical carcinogenesis - DNA adducts
Apalutamide	Enzalutamide → CYP3A5	Prostate cancer
	ABCB1	Chemical carcinogenesis - DNA adducts
	Abiraterone → SULT2A1	Metabolism of xenobiotics by cytochrome P450
Atorvastatin	ABCC4	Chemical carcinogenesis - DNA adducts
	CYP3A5	Drug metabolism - cytochrome P450
	CYP2C19	Metabolism of xenobiotics by cytochrome P450
Carisoprodol	CYP2C19	Arachidonic acid metabolism
	CYP2C19 → Hydrogen → Glycerol → FERMT2	Chemical carcinogenesis - DNA adducts
	Oxicone → Chloride ion → ITGAV	Drug metabolism - cytochrome P450
Oxcarbazepine	AKR1C3	Steroid hormone biosynthesis
	CYP2C19	Chemical carcinogenesis - DNA adducts
	ABCB1	Chemical carcinogenesis - reactive oxygen species

Table 1. The top-3 critical paths and top-3 KEGG pathways of the potential drugs for the treatment of prostate cancer. These drugs are ranked top 10 by iDPath among all the FDA-Approved drugs used in this study. The head (drug) and tail (prostate cancer) of these critical paths are ignored due to the limit of space. The top-3 critical paths are determined by the weights generated by the path attention module. The KEGG pathways are identified by KEGG enrichment analysis on the proteins existed in the top-50 critical paths and ranked by p-adjust values.

As a deep learning approach, iDPath is limited to *in silico* study, which can be extended by *in vitro* and *in vivo* experiments to further validate its practical value and consistency with clinical evidence in future studies. In addition, the identified paths may contain rich biological knowledge beyond this study, such as some popular paths may be associated with common mechanisms of action in a class of diseases, which is worth further study.

Code availability

iDPath source code and its usage instructions are available in Github (<https://github.com/JasonJYang/iDPath>).

Acknowledgements

This work was supported by National Natural Science Foundation of China (NSFC): 71972164; Health and Medical Research Fund of the Food and Health Bureau of Hong Kong: 16171991; Innovation and Technology Fund of Innovation and Technology Commission of Hong Kong: MHP/081/19; National Key Research and Development Program of China, Ministry of Science and Technology of China: 2019YFE0198600.

Author contributions statement

J.N.Y. conceived the study, searched the literature, designed the study, did the coding, and drafted and reviewed the article. Z.L., S.Y., and Q.C. reviewed the article and provided domain comments. W.K.K.W. reviewed the article. Q.C. and Q.P.Z designed the study. Q.P.Z. drafted and reviewed the article.

References

1. Fleming, N. How artificial intelligence is changing drug discovery. *Nature* **557**, S55–S55 (2018).
2. Pham, T.-H., Qiu, Y., Zeng, J., Xie, L. & Zhang, P. A deep learning framework for high-throughput mechanism-driven phenotype compound screening and its application to covid-19 drug repurposing. *Nat. Mach. Intell.* **3**, 247–257 (2021).
3. Jin, W. *et al.* Deep learning identifies synergistic drug combinations for treating covid-19. *Proc. Natl. Acad. Sci.* **118** (2021).
4. Yan, V. K. *et al.* Drug repurposing for the treatment of covid-19: A knowledge graph approach. *Adv. Ther.* 2100055 (2021).
5. Rodriguez, S. *et al.* Machine learning identifies candidates for drug repurposing in alzheimer's disease. *Nat. communications* **12**, 1–13 (2021).
6. Baptista, D., Ferreira, P. G. & Rocha, M. Deep learning for drug response prediction in cancer. *Briefings bioinformatics* **22**, 360–379 (2021).
7. Zhou, Y., Wang, F., Tang, J., Nussinov, R. & Cheng, F. Artificial intelligence in covid-19 drug repurposing. *The Lancet Digit. Heal.* (2020).
8. Sanseau, P. & Koehler, J. Computational methods for drug repurposing (2011).
9. Xu, H. *et al.* Validating drug repurposing signals using electronic health records: a case study of metformin associated with reduced cancer mortality. *J. Am. Med. Informatics Assoc.* **22**, 179–191 (2015).
10. Liu, R., Wei, L. & Zhang, P. A deep learning framework for drug repurposing via emulating clinical trials on real-world patient data. *Nat. Mach. Intell.* **3**, 68–75 (2021).
11. Yang, X. *et al.* High-throughput transcriptome profiling in drug and biomarker discovery. *Front. genetics* **11**, 19 (2020).
12. Silverbush, D. & Sharan, R. A systematic approach to orient the human protein–protein interaction network. *Nat. communications* **10**, 1–9 (2019).
13. Cheng, F. *et al.* Network-based approach to prediction and population-based validation of in silico drug repurposing. *Nat. communications* **9**, 1–12 (2018).
14. Cheng, F., Kovács, I. A. & Barabási, A.-L. Network-based prediction of drug combinations. *Nat. communications* **10**, 1–11 (2019).
15. Gysi, D. M. *et al.* Network medicine framework for identifying drug-repurposing opportunities for covid-19. *Proc. Natl. Acad. Sci.* **118** (2021).
16. Zhou, Y. *et al.* Network-based drug repurposing for novel coronavirus 2019-ncov/sars-cov-2. *Cell discovery* **6**, 1–18 (2020).
17. Yang, J., Xu, Z., Wu, W. K. K., Chu, Q. & Zhang, Q. GraphSynergy: a network-inspired deep learning model for anticancer drug combination prediction. *J. Am. Med. Informatics Assoc.* **28**, 2336–2345, DOI: [10.1093/jamia/ocab162](https://doi.org/10.1093/jamia/ocab162) (2021). <https://academic.oup.com/jamia/article-pdf/28/11/2336/40576607/ocab162.pdf>.
18. Lopes-Ramos, C. M. *et al.* Gene regulatory network analysis identifies sex-linked differences in colon cancer drug metabolism. *Cancer research* **78**, 5538–5547 (2018).

19. Kalinina, O. V., Wichmann, O., Apic, G. & Russell, R. B. Combinations of protein-chemical complex structures reveal new targets for established drugs. *PLOS computational biology* **7**, e1002043 (2011).
20. Hu, L.-L., Chen, C., Huang, T., Cai, Y.-D. & Chou, K.-C. Predicting biological functions of compounds based on chemical-chemical interactions. *PloS one* **6**, e29491 (2011).
21. Liu, X. & Pan, L. Detection of driver metabolites in the human liver metabolic network using structural controllability analysis. *BMC systems biology* **8**, 1–17 (2014).
22. Zhou, W., Wang, Y., Lu, A. & Zhang, G. Systems pharmacology in small molecular drug discovery. *Int. journal molecular sciences* **17**, 246 (2016).
23. T Issa, N., Wathieu, H., Ojo, A., W Byers, S. & Dakshanamurthy, S. Drug metabolism in preclinical drug development: a survey of the discovery process, toxicology, and computational tools. *Curr. drug metabolism* **18**, 556–565 (2017).
24. Sun, Y. V. Integration of biological networks and pathways with genetic association studies. *Hum. genetics* **131**, 1677–1686 (2012).
25. Snape, T. J., Astles, A. M. & Davies, J. Understanding the chemical basis of drug stability and degradation. *Pharm. J.* **285**, 416–417 (2010).
26. Löwenberg, B. *et al.* Cytarabine dose for acute myeloid leukemia. *New Engl. J. Medicine* **364**, 1027–1036 (2011).
27. Hamada, A., Kawaguchi, T. & Nakano, M. Clinical pharmacokinetics of cytarabine formulations. *Clin. pharmacokinetics* **41**, 705–718 (2002).
28. Wan, F., Hong, L., Xiao, A., Jiang, T. & Zeng, J. Neodti: neural integration of neighbor information from a heterogeneous network for discovering new drug–target interactions. *Bioinformatics* **35**, 104–111 (2019).
29. Berenstein, A. J., Magariños, M. P., Chernomoretz, A. & Agueero, F. A multilayer network approach for guiding drug repositioning in neglected diseases. *PLoS neglected tropical diseases* **10**, e0004300 (2016).
30. Wang, W., Yang, S., Zhang, X. & Li, J. Drug repositioning by integrating target information through a heterogeneous network model. *Bioinformatics* **30**, 2923–2930 (2014).
31. Zhang, F., Wang, M., Xi, J., Yang, J. & Li, A. A novel heterogeneous network-based method for drug response prediction in cancer cell lines. *Sci. reports* **8**, 1–9 (2018).
32. Shrikumar, A., Greenside, P. & Kundaje, A. Learning important features through propagating activation differences. In *International Conference on Machine Learning*, 3145–3153 (PMLR, 2017).
33. Ribeiro, M. T., Singh, S. & Guestrin, C. " why should i trust you?" explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, 1135–1144 (2016).
34. Wang, H. *et al.* Ripplenet: Propagating user preferences on the knowledge graph for recommender systems. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, 417–426 (2018).
35. Wang, X. *et al.* Explainable reasoning over knowledge graphs for recommendation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, 5329–5336 (2019).
36. Elmarakeby, H. A. *et al.* Biologically informed deep neural network for prostate cancer discovery. *Nature* **598**, 348–352 (2021).

37. Ma, J. *et al.* Using deep learning to model the hierarchical structure and function of a cell. *Nat. methods* **15**, 290–298 (2018).
38. Kuenzi, B. M. *et al.* Predicting drug response and synergy using a deep learning model of human cancer cells. *Cancer cell* **38**, 672–684 (2020).
39. Xu, Z., Zhang, J., Zhang, Q., Xuan, Q. & Yip, P. S. F. A comorbidity knowledge-aware model for disease prognostic prediction. *IEEE Transactions on Cybern.* (2021).
40. Guo, M., Xu, Z., Zhang, Q., Liao, X. & Liu, J. Deciphering feature effects on decision-making in ordinal regression problems: An explainable ordinal factorization model. *ACM Transactions on Knowl. Discov. from Data (TKDD)* **16**, 1–26 (2021).
41. Liu, Z.-P., Wu, C., Miao, H. & Wu, H. Regnetwork: an integrated database of transcriptional and post-transcriptional regulatory networks in human and mouse. *Database* **2015** (2015).
42. Szklarczyk, D. *et al.* The string database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic acids research* **49**, D605–D612 (2021).
43. Szklarczyk, D. *et al.* Stitch 5: augmenting protein–chemical interaction networks with tissue and affinity data. *Nucleic acids research* **44**, D380–D384 (2016).
44. Kasprzyk, A. Biomart: driving a paradigm change in biological data management. *Database* **2011** (2011).
45. Maglott, D., Ostell, J., Pruitt, K. D. & Tatusova, T. Entrez gene: gene-centered information at ncbi. *Nucleic acids research* **35**, D26–D31 (2007).
46. Kim, S. *et al.* Pubchem 2019 update: improved access to chemical data. *Nucleic acids research* **47**, D1102–D1109 (2019).
47. Zhou, Y. *et al.* Therapeutic target database update 2022: facilitating drug discovery with enriched comparative data of targeted agents. *Nucleic Acids Res.* (2021).
48. Wishart, D. S. *et al.* Drugbank 5.0: a major update to the drugbank database for 2018. *Nucleic acids research* **46**, D1074–D1082 (2018).
49. Piñero, J. *et al.* Disgenet: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic acids research* gkw943 (2016).
50. Huang, Z. *et al.* Hmdd v3. 0: a database for experimentally supported human microrna–disease associations. *Nucleic acids research* **47**, D1013–D1017 (2019).
51. Kipf, T. N. & Welling, M. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907* (2016).
52. Ren, Y., Ay, A. & Kahveci, T. Shortest path counting in probabilistic biological networks. *BMC bioinformatics* **19**, 1–19 (2018).
53. Hricik, T., Bader, D. & Green, O. Using rapids ai to accelerate graph data science workflows. In *2020 IEEE High Performance Extreme Computing Conference (HPEC)*, 1–4 (IEEE, 2020).
54. Hochreiter, S. & Schmidhuber, J. Long short-term memory. *Neural computation* **9**, 1735–1780 (1997).
55. Dwarampudi, M. & Reddy, N. Effects of padding on lstms and cnns. *arXiv preprint arXiv:1903.07288* (2019).

56. Vaswani, A. *et al.* Attention is all you need. In *Advances in neural information processing systems*, 5998–6008 (2017).
57. Perozzi, B., Al-Rfou, R. & Skiena, S. Deepwalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, 701–710 (2014).
58. Chen, Z.-H. *et al.* Prediction of drug–target interactions from multi-molecular network based on deep walk embedding model. *Front. Bioeng. Biotechnol.* **8**, 338 (2020).
59. Sakai, M. *et al.* Prediction of pharmacological activities from chemical structures with graph convolutional neural networks. *Sci. reports* **11**, 1–14 (2021).
60. Xu, Z., Wang, S., Zhu, F. & Huang, J. Seq2seq fingerprint: An unsupervised deep molecular embedding for drug discovery. In *Proceedings of the 8th ACM international conference on bioinformatics, computational biology, and health informatics*, 285–294 (2017).
61. Santiso, S., Perez, A. & Casillas, A. Exploring joint ab-lstm with embedded lemmas for adverse drug reaction discovery. *IEEE journal biomedical health informatics* **23**, 2148–2155 (2018).
62. Benoist, G. E. *et al.* Pharmacokinetic aspects of the two novel oral drugs used for metastatic castration-resistant prostate cancer: abiraterone acetate and enzalutamide. *Clin. pharmacokinetics* **55**, 1369–1380 (2016).
63. Zeigler-Johnson, C. Cyp3a4: a potential prostate cancer risk factor for high-risk groups. *Clin. journal oncology nursing* **5** (2001).
64. Fujita, K. & Nonomura, N. Role of androgen receptor in prostate cancer: a review. *The world journal men's health* **37**, 288–295 (2019).
65. Jentzmik, F., Azoitei, A., Zengerling, F., Damjanoski, I. & Cronauer, M. V. Androgen receptor aberrations in the era of abiraterone and enzalutamide. *World journal urology* **34**, 297–303 (2016).
66. Malikova, J. *et al.* Cyp17a1 inhibitor abiraterone, an anti-prostate cancer drug, also inhibits the 21-hydroxylase activity of cyp21a2. *The J. steroid biochemistry molecular biology* **174**, 192–200 (2017).
67. DeVore, N. M. & Scott, E. E. Structures of cytochrome p450 17a1 with prostate cancer drugs abiraterone and tok-001. *Nature* **482**, 116–119 (2012).
68. Ogata, H., Goto, S., Fujibuchi, W. & Kanehisa, M. Computation with the kegg pathway database. *Biosystems* **47**, 119–128 (1998).
69. Toren, P. & Zoubeidi, A. Targeting the pi3k/akt pathway in prostate cancer: challenges and opportunities. *Int. journal oncology* **45**, 1793–1801 (2014).
70. Yamaguchi, H. & Condeelis, J. Regulation of the actin cytoskeleton in cancer cell migration and invasion. *Biochimica et Biophys. Acta (BBA)-Molecular Cell Res.* **1773**, 642–652 (2007).
71. Fizazi, K. *et al.* A phase 3 trial with a 2x2 factorial design of abiraterone acetate plus prednisone and/or local radiotherapy in men with de novo metastatic castration-sensitive prostate cancer (mcspc): First results of peace-1. (2021).
72. Wang, X. *et al.* Cabozantinib inhibits abiraterone's upregulation of igfir phosphorylation and enhances its anti-prostate cancer activity. *Clin. Cancer Res.* **21**, 5578–5587 (2015).
73. Andriole, G. L. *et al.* Effect of dutasteride on the risk of prostate cancer. *N Engl J Med* **362**, 1192–1202 (2010).

74. Joshi, S., Murphy, E., Olaniyi, P. & Bryant, R. The multiple effects of aspirin in prostate cancer patients. *Cancer Treat. Res. Commun.* **26**, 100267 (2021).
75. Gravis, G. *et al.* Monocentric evaluation of erlotinib in advanced prostate cancer. *J. Clin. Oncol.* **25**, 15569–15569 (2007).
76. Krishnappa, K., Mallesh, N. K., Sharma, S. C. & Manjulakumari, D. Midostaurin inhibits hormone-refractory prostate cancer pc-3 cells by modulating npkcs and ap-1 transcription factors and their target genes involved in cell cycle. *Front. biology* **12**, 421–429 (2017).
77. Smith, M. R. *et al.* Apalutamide treatment and metastasis-free survival in prostate cancer. *New Engl. J. Medicine* **378**, 1408–1418 (2018).
78. Khosropanah, I. *et al.* Assessment of atorvastatin effectiveness on serum psa level in hypercholesterolemic males. *Acta Medica Iran.* 789–794 (2011).
79. Al-Salama, Z. T. Apalutamide: first global approval. *Drugs* **78**, 699–705 (2018).
80. Saad, F. *et al.* Apalutamide plus abiraterone acetate and prednisone versus placebo plus abiraterone and prednisone in metastatic, castration-resistant prostate cancer (acis): a randomised, placebo-controlled, double-blind, multinational, phase 3 study. *The Lancet Oncol.* **22**, 1541–1559 (2021).
81. Shorning, B. Y., Dass, M. S., Smalley, M. J. & Pearson, H. B. The pi3k-akt-mtor pathway and prostate cancer: at the crossroads of ar, mapk, and wnt signaling. *Int. journal molecular sciences* **21**, 4507 (2020).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SI.pdf](#)