Title

Preclinical side effect prediction through pathway engineering of protein interaction network models

Authors

Alidoost, Mohammadali
Wilson, Jennifer L

Publication Date

2024-05-12

DOI

10.1002/psp4.13150

Received: 14 November 2023 | Revised: 1 March 2024 | Accepted: 8 April 2024

DOI: 10.1002/psp4.13150

ARTICLE

Preclinical side effect prediction through pathway engineering of protein interaction network models

Mohammadali Alidoost | **Jennifer L. Wilson**

Department of Bioengineering, University of California, Los Angeles, California, USA

Correspondence
Jennifer L. Wilson, Department of Bioengineering, University of California, Los Angeles, California, USA.
Email: jenniferwilson@g.ucla.edu

Abstract

Modeling tools aim to predict potential drug side effects, although they suffer from imperfect performance. Specifically, protein–protein interaction models predict drug effects from proteins surrounding drug targets, but they tend to overpredict drug phenotypes and require well-defined pathway phenotypes. In this study, we used PathFX, a protein–protein interaction tool, to predict side effects for active ingredient-side effect pairs extracted from drug labels. We observed limited performance and defined new pathway phenotypes using pathway engineering strategies. We defined new pathway phenotypes using a network-based and gene expression-based approach. Overall, we discovered a trade-off between sensitivity and specificity values and demonstrated a way to limit overprediction for side effects with sufficient true positive examples. We compared our predictions to animal models and demonstrated similar performance metrics, suggesting that protein–protein interaction models do not need perfect evaluation metrics to be useful. Pathway engineering, through the inclusion of true positive examples and omics measurements, emerges as a promising approach to enhance the utility of protein interaction network models for drug effect prediction.

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?

Network-based methods in drug safety evaluation suffer from low prediction accuracy and overprediction due to reliance on accurate side effect pathways. Additionally, recent observations emphasize the significance of downstream proteins, in addition to drug targets, in predicting drug side effects.

WHAT QUESTION DID THIS STUDY ADDRESS?

This study addressed the impact of defining new gene pathways through pathway engineering, by including true positive examples and omics data, on the prediction performance of drug-induced side effects.

WHAT DOES THIS STUDY ADD TO OUR KNOWLEDGE?

The study demonstrates that refining pathways and incorporating gene expression data can reduce overprediction, enhancing preclinical prediction of drug-induced safety events.

HOW MIGHT THIS CHANGE DRUG DISCOVERY, DEVELOPMENT, AND/OR THERAPEUTICS?

The findings suggest that pathway engineering and the integration of omics data can improve the accuracy of predicting drug-induced side effects, offering valuable insights for drug discovery and development strategies.

INTRODUCTION

Many newly developed drugs are not approved despite lengthy and expensive research and development.[1,2] Even though drugs can save the lives of people, they sometimes result in serious and deadly adverse effects on humans. Pharmaceutical companies also suffer substantial losses when drugs must be pulled from the market due to these adverse drug reactions.[3] Before new drugs are tested on humans, regulatory organizations need a set of standard tests for pharmaceuticals to guarantee their safety and effectiveness. Preclinical prediction of drug-induced adverse events is crucial given that safety is a leading cause of drug attrition and has already proven useful for some adverse events. Preclinical animal research, as a gold standard for toxicity and efficacy evaluations, on rodents, non-rodents, and nonhuman primates, is a key component of this regulatory process.[4] Animal models have been a valuable technique for anticipating drug-induced side effects. However, other than the fact that some countries have been considering reducing and/or replacing animal use in their research studies,[5] animal studies are not uniformly sensitive to all side effects and can be costly compared with in silico approaches. It is also worth mentioning that in accordance with legislation approved in late December 2022, the US Food and Drug Administration (FDA) may rely less on animal testing of new drugs and may not require animal models when there is sufficient computation or non-animal studies.[6]

Many are applying in silico approaches to predict drug effects, understand potential side effects, and identify novel therapeutic targets.[7,8] These approaches include machine learning,[9] protein–protein interaction (PPI) network methods,[10,11] off-target drug-binding prediction,[12] ligand similarity metrics,[13] or side effect profile similarity.[14] Huang et al.[15] developed a systems pharmacology approach to predict adverse drug reactions, with a focus on cardiotoxicity. By combining clinical observation data, drug target information, PPI networks, and Gene Ontology (GO) annotations, their model achieved satisfactory predictive performance for cardiotoxic side effects. Importantly, the study highlighted the significance of incorporating prior knowledge networks to enhance side effect predictions, especially for drug targets in development, yet their study was limited to a single side effect. Li et al.[16] explored the significance of drug-induced liver injury in drug development, highlighting its role as a leading cause of drug failures and acute liver failure. They emphasized the absence of reliable clinical biomarkers and the potential of computational approaches, specifically artificial intelligence, to address this issue. Their study surveyed drug-induced liver injury predictive models based on various molecular representations. Song et al.[17] did a review that focused on the use of network-based computational methods in the field of drug repositioning to find new therapeutic uses for existing drugs. This paper acknowledged the rapid advancements made in this field, particularly with the advent of deep learning and the availability of large datasets. They aimed to provide an overview of various network-based methods used in drug repositioning, comparing, and discussing their development processes. Lin et al.[18] discussed that cancer drugs in clinical trials achieve efficacy not from the intended targets but through other, frequently unidentified pathway phenotypes. This finding underscores the broad-acting nature of drugs and the prevalence of off-target effects, which may be overlooked in traditional evaluation methods. This understanding emphasizes the need for employing advanced network modeling techniques to predict the complex interplay of drug interactions and their unintended effects more accurately, aiming to refine the predictive accuracy of drug side effects in preclinical safety evaluations. Lampa et al.[19] introduced a novel methodology for ligand-based target profiling in drug discovery, with a focus on providing confidence measures for predictions of off-target interactions. Their approach holds promise for integrating target predictions into drug discovery and safety assessments, with a focus on confidence and model reliability. In addition, it is worth mentioning that discovering drug targets typically initiates the prediction of drug side effects. A study by Campillos et al.[14] illustrated that when drug side effect profiles are similar, this can serve as a predictive factor for common drug targets.

A drug's primary target is often insufficient for understanding all drug-induced outcomes and PPI network approaches have the advantage of connecting drug targets to downstream adverse effect-associated proteins.[20] Furthermore, PPI networks improve the biological relevance of model predictions by identifying additional signaling molecules that have not been explicitly detected[11]

and they provide an interpretable understanding of drug responses. In our own work,[21] we developed the PathFX algorithm for anticipating drug effects from PPI networks. PathFX, like other network methods, assumes that proteins neighboring drug target proteins are candidates for a drug's signaling pathway. These methods use optimization or empirical analysis of network associations to identify the downstream proteins with the highest confidence, either based on experimental or database-derived data. Interestingly, in our previous pathway study of pharmacological side effects, we observed that in addition to drug targets, downstream proteins are significant in predicting drug side effects.[21,22] We also learned that true positive (TP) and false positive (FP) network predictions were associated with subsets of pathway phenotype-associated proteins. This encouraged us to consider new pathway phenotypes that may better predict drug-induced effects.

Furthermore, the integration of gene expression data into PPI networks that exclusively focus on human proteins, such as PathFX, can enhance our understanding of cellular processes and disease mechanisms. Gene expression data provide valuable information about the abundance and activity levels of individual proteins in specific tissues or under certain conditions. By incorporating this data into PPI networks, we can contextualize the interactions between proteins, revealing their functional relevance in specific biological processes.[23,24] Chen et al.[25] presented PharmOmics, a tool in drug development, aiming to combat the high failure rate in clinical trials due to insufficient knowledge about drug actions in diverse organs and species. It harnesses species- and tissue-specific transcriptome data from humans, mice, and rats to enable gene-network-based drug repositioning. PharmOmics identifies therapeutic drugs and potential tissue toxicity through computational evaluations, validated in a non-alcoholic fatty liver disease model. PharmOmics offers a valuable resource for network-based drug research, and relevant to our study, a comprehensive database of drug-induced gene expression changes.

While attractive, PPI network methods have some limitations, yet they may still be useful in development pipelines. Network models tend to overpredict drug phenotypes and there is insufficient evidence to validate these predictions. For many drug outcomes, we lack balanced sets of drugs that do and do not cause the outcome. Because we cannot validate all PPI predictions, many consider these predictions as FPs, and this assumption leads to low-performance metrics.[21,26,27] Yet, a meta-analysis of the predictive values of in vivo animal models for drug side effect prediction achieved sensitivity and specificity ranges of 0.00–0.74 and 0.33–1.00 depending on the organ categories and species.[4] This suggests that models can still be useful for decision-making without perfect prediction performance and sets a benchmark for improving PPI network methods to influence decision-making. Emphasizing an understanding of drug-induced effects instead of overall performance is consistent with other computational efforts that emphasize model utility over performance. This is a growing field of computational research that we[28] and others have emphasized.[29]

We also discovered that a per-pathway phenotype assessment enhanced prediction accuracy and reduced overprediction.[22] This analysis emphasized the importance of defining drug "pathway phenotypes"—or gene/protein lists—by calibrating these lists to drugs known to cause severe adverse reactions, as indicated on their labels, as opposed to optimizing statistical approaches for ranking drug network associations. However, there are many sources for defining pathway gene lists. In our original development of PathFX, we merged all pathway phenotype-associated genes from several data sources and a "pathway phenotype" can include diseases or side effects. Yet, these pathways, or gene lists, had not been calibrated to known drug-induced effects. Our objective was to engineer pathways by incorporating network genes associated with drugs that cause a side effect (distinct TP genes) and omics measurements, specifically, drug-induced gene expression changes, and assess PathFX sensitivity and specificity per side effect. We focused on our previously published set of side effects outcomes because it contained sufficient examples of drugs that do and do not cause side effects. Furthermore, we sought to understand general principles for pathway definition that could be applied when there are insufficient examples of positive and negative examples.

In this study, we aimed to address the shortcomings of network-based methods, which exhibit low prediction accuracy and a tendency to overpredict associations, to improve the prediction of drug-induced effects.

METHODS

Dataset

In this study, we utilized a dataset, called the "drug toxicity dataset," which comprises pairs of active ingredients and their corresponding side effects extracted from drug labels.[22,30] The dataset contained 1970 drugs and 34 side effects, emphasizing only severe drug-induced pathway phenotypes that could affect a drug development program. The entire dataset is provided with the original publication[30] and with this study (see *Data availability statement*). For the purpose of this study, we selected and highlighted a subset of side effects to demonstrate the effectiveness of our proposed approaches. Throughout, we

use "side effects" to refer to the drug's effects reflected on their labels.

Mapping drugs and side effects

We first mapped active ingredient names and side effects to DrugBank[31] drug names and PathFX pathway phenotypes using string matching in Python. In the drug mapping procedure, we retained all active ingredients that had a direct match to a drug name in the DrugBank version 5.1.6. Importantly, we retained all active ingredients in DrugBank and did not restrict our analysis to specific drug classes or targets.

It is worth mentioning that the PathFX v2[32] (version 2, released with Wilson et al.[32]) database contained 29,831 pathway phenotypes merged from multiple sources of pathway phenotype associations,[33–37] including ClinVar, OMIM, DisGeNet, and PhenotypeGenotype Integrator; thus, there were many PathFX pathway phenotypes that could be relevant to any side effect. Throughout, we refer to a PathFX-predicted drug effect as a "pathway phenotype." We developed a string-matching approach, a two-step process, to find the PathFX pathway phenotypes relevant to the labeled side effects. For the first step, we wrote an ensemble function of three distinct techniques capable of detecting similar matches between two lists of words. The first component of the string-matching search checks similar letters with the same positions in both strings and sees how many similar letters they have. Eventually, the algorithm takes the one with the highest number of similar letters and returns one match. This algorithm can be changed to return more similar matches too. The second method uses a Python library called Jaro-Winkler[38] that computes the similarity between two strings and the returned value lies in the interval of 0.0 and 1.0. When we have no similarity, the score will be 0.0. The last component applies a Python module, difflib.get_close_matches,[39] by giving back a list of good enough matches. For the second step, after applying the string-matching search, we further validated their relevance by manual review of the literature.

Running PathFX for all mapped drugs

We generated drug networks for all active ingredients mapped to drug names (and DrugBank identifiers) in DrugBank. PathFX required drug-binding proteins to seed drug networks. Thus, PathFX was unable to generate networks for drugs without documented targets in DrugBank or targets that were not connected to the PathFX interaction network. For drugs with sufficient data, PathFX made multiple files, two of which were important for our analysis: a network file that included the prioritized PPIs (a tab-delimited file containing edges between two proteins) and an association file that contained network-associated pathway phenotypes (a tab-delimited file containing the pathway phenotype name and drug network proteins associated to a pathway phenotype). Network proteins associated with a pathway phenotype are subsets of the full list of pathway phenotype genes in PathFX.

Measuring PathFX baseline predictions per side effect

To evaluate the predictive performance of PathFX for each side effect, we searched PathFX association tables of mapped drug names for side effect-related pathway phenotypes obtained previously. Because multiple PathFX pathway phenotypes were used to discover a single side effect, we ensured that synonymous pathway phenotypes were not double-counted. To illustrate whether the predicted pathway phenotypes "Proteinuria" and "Mild Proteinuria" were both indicative of the same side effect, "Proteinuria," we considered them as a single instance. We also devised an end-to-end method, SEPred, which identified pathway proteins specific to drugs that were accurately associated with their corresponding labeled side effects and shared with drugs' networks that contained a FP prediction. Eventually, we computed the confusion matrix along with key evaluation metrics, including sensitivity and specificity to quantitatively assess the performance of the prediction methodology. In assessing the predictive performance of PathFX for each side effect, we define TP as instances where the model predicts a pathway phenotype that is corroborated by the drug toxicity dataset. FP occurs when a predicted pathway phenotype is not supported by this dataset. True negative (TN) is identified when PathFX does not predict a pathway phenotype, and this is consistent with its absence in the drug toxicity dataset. Conversely, a false negative (FN) arises when the model fails to predict a pathway phenotype that is present in the dataset. Our approach involves comparing the drug names associated with each pathway phenotype, as identified by PathFX, against the labeled side effects in the drug toxicity dataset.

GO enrichment analysis to discover further biological relevance

We performed GO enrichment using the GOrilla tool,[40] to compare pathway phenotype proteins to the entire interaction network. For the foreground gene list, we used

all shared network proteins for pathway phenotypes identified previously, and for the background gene list, we used all network proteins in the interaction network as published in Wilson et al.[21] We repeated this process for all pathway phenotypes, where there were sufficient examples of unique or shared network proteins and we merged network proteins for pathway phenotypes that represented similar side effects (e.g., we performed GO enrichment for the union of all genes associated with the 12 pathway phenotypes representative of the side effect "hypertension").

Bias reduction for incorporation of new pathways into PathFX

When creating the original PathFX v1 (version 1, released with Wilson et al.[21]), we used two procedures to minimize biases inherent in network prediction: interaction specificity analysis and expected association analysis. The former minimized biases due to connectivity in the interactome and the latter minimized biases due to pathway phenotypes having a different number of associated genes using Fisher's exact test. To incorporate new pathway phenotypes into PathFX, we repeated the latter bias reduction technique (because we were not changing the interactome) with several new pathways (described below). This procedure included generating 100 random networks with a range of random input targets (we assessed networks with 1 to 40 random targets) and measuring their association with the newly generated pathway phenotypes. When assessing real drug networks, we used these expected association scores to determine whether a pathway phenotype would be retained in the final network. We required that the pathway phenotype have an association that was below the median association for that pathway phenotype within 100 random networks with the same number of input targets. This is the same process as described in Wilson et al.[21] Every new pathway phenotype was subject to this bias reduction technique before we incorporated them into a new version of PathFX that would assess real drug network associations relative to these randomizations (new version released with this publication, see *Data availability statement*).

Defining novel pathway phenotypes

We made a pipeline, called "DefPath," implementing pathway engineering, that defined novel custom pathway phenotypes using diverse input data: counting network proteins from drugs known to cause the side effect in the drug toxicity dataset, as "distinct TP baseline" and counting network proteins from drugs not known to cause the side effect in the drug toxicity dataset, as "distinct FP baseline." Through this pipeline, we systematically assessed the performance of new pathway phenotypes for each side effect. Moreover, to account for scenarios with insufficient distinct TP example networks, we further devised novel pathway phenotypes using an omics dataset, called the "drug signature dataset," previously compiled by Chen et al.[25] This dataset contains comprehensive gene expression changes in the human, mouse, and rat tissues, obtained from various databases, along with a curated list of crucial drugs utilized for treating diseases. Specifically, they curated 13,530 human, mouse, and rat transcriptome datasets published in Gene Expression Omnibus (GEO) and summarized top and bottom differentially expressed genes collected from more than 20 tissues and 941 drugs. For our analysis, we exclusively focused on the differential expression of human genes to ensure the relevance of the findings. Additionally, because we wanted to broadly explore all possible gene expression changes, we did not restrict our analysis to any specific tissue and instead used any differentially expressed gene to understand the drug's effects. We did not have patient-level or sample numbers in processing human gene expression data. The dataset included "top" genes, which are the top 500 differentially expressed genes (or less if there are <500 significant genes), as assessed by differential gene expression analysis (LIMMA).[25] Using these gene lists, we initially searched for top genes shared among drugs that cause the same side effect. However, we discovered this to be a null set. Instead, we took the union of top genes from drugs known to cause the side effect in the drug toxicity dataset, as "TP signature," and for top genes from drugs not known to cause the side effect in the drug toxicity dataset, as "FP signature." This process yielded four pathway categories for each side effect, TP baseline, FP baseline, TP signature, and FP signature. Afterward, we took various intersections between these categories, that is, the pathway engineering process, generated new PathFX pathway phenotypes, and measured the ability to recover associations to drugs known to cause the side effect. Notably, we had insufficient data for multiple side effects.

Assessing gene pathway robustness through randomization techniques in PPI networks

To assess the robustness of our pathways, we measured the predictive utility of pathway phenotypes assembled with random genes. Firstly, we implemented a gene knockout process, where we randomly excluded 15% of the distinct TP genes linked to each pathway phenotype.

Subsequently, to evaluate the impact of noisy input data, we augmented the TP distinct genes list by including 15% of the shared TP and FP genes associated with pathway phenotype. By varying the inputs in this manner, we systematically investigated the effects of different gene sets on defining novel pathway phenotypes, enabling a comprehensive exploration of the method's robustness and reliability in predicting drug-induced safety events.

Generating new PathFX predictions and evaluation metrics for novel custom pathways

Having established novel custom pathway phenotypes, utilized distinct TP and deferentially expressed genes, and completed bias reduction, we executed PathFX for all drugs mapped from the drug toxicity dataset. PathFX generated a table of network-associated pathway phenotypes, ranked by their multiple-hypothesis-corrected *p*-values using the "Benjamini–Hochberg" method.[41] Furthermore, we tracked the newly defined pathway phenotypes and quantified their ability to recover associations to drugs known to cause the side effects in the toxicity dataset. Subsequently, we retrieved the drugs associated with specific pathways, considering them as new PathFX predictions, and compared these with the drug names present in the drug toxicity dataset. This pipeline, DefPath, enabled us to calculate evaluation metrics per side effect and for each pathway phenotype, providing a comprehensive assessment of the predictive performance of our approach.

RESULTS

Baseline performance analysis: Preliminary results and evaluations

Integration of the drug toxicity dataset: Mapping drugs and side effects to the PathFX database

To test PathFX predictions, we needed to map side effects from drug labels ("labeled side effects") to pathway phenotypes in the PathFX database ("predicted side effects") and active ingredients to DrugBank identifiers (to obtain drug targets which are required inputs to PathFX). In the original PathFX v2, we had 8238 drugs (with DrugBank identifiers) and 29,831 pathway phenotypes. In the drug toxicity dataset, we successfully mapped 1132/1970 active ingredients to DrugBank identifiers and within this set, drugs were associated with an average of 6.5 (median = 5.0, SD = 4.7) side effects. For instance, the drug name "Atropine" was mapped to the drug name "Atropine." In case, we could not map to the drug name, we mapped to the DrugBank identifier. For example, the drug name "Valproic acid" was mapped to DrugBank Identifier "DB00313." Importantly, we did not restrict our analysis to any therapeutic use classes or specific drug targets. Indeed, looking at the ATC codes of all mapped drugs, there were 1600 unique level-3 ATC codes in the dataset.

Of these 1132 drugs, we generated PathFX networks for 890 drugs with sufficient binding targets and protein interactions (Figure 1). Of the 34 side effects in the drug toxicity dataset, we found sufficient PathFX pathway phenotypes for 32: 4 side effects had a direct match to a PathFX pathway phenotype (e.g., the side effect, "Delirium," matched to pathway phenotype, "Delirium"), 8 had a synonymous match only (e.g., the side effect, "Sleep apnea syndrome," matched to the pathway phenotype, "Sleep apnea syndromes"), 20 side effects had both a synonymous and direct match (e.g., the side effect, "Gastric ulcer," matched to the pathway phenotypes, "Gastric ulcer" and "Peptic ulcer"), and 2 side effects did not match any pathway phenotypes tracked in PathFX (Figure 1, Table 1).

Initial assessment of PathFX predictions discovers low performance, yet identifies biologically relevant pathway connections

We next investigated PathFX networks for the 890 drugs with complete information and assessed PathFX v2's ability to correctly predict pathway phenotypes related to the drugs' labeled side effects. PathFX identifies high-priority downstream proteins using a drug's target proteins as inputs and then assesses pathway phenotypes associated with this network using a bias-controlled Fisher's exact test (see 'Section 2'). For example, PathFX created a network for the drug atropine, which has 9 drug-binding proteins in our interactome. The entire network included 348 proteins and 604 pathway phenotypes. Within the subnetwork, atropine is associated with 4 pathway phenotypes relevant to our side effect prediction: "essential hypertension," "genetic hypertension," "idiopathic pulmonary arterial hypertension," and "pulmonary hypertension" (Figure 2a). Atropine is connected to these pathway phenotypes through five drug-binding proteins (CHRM-1-5) and three downstream proteins (KNG1, AGT, and ADN1). Also, PathFX analysis of caffeine, which has 19 drug-binding proteins in our interactome, generated a network of 552 proteins and 1040 pathway phenotypes. Caffeine is associated with a pathway phenotype, "edema," related to the side effect in that subnetwork (Figure 2b), through seven drug-binding targets (in red) and eight downstream

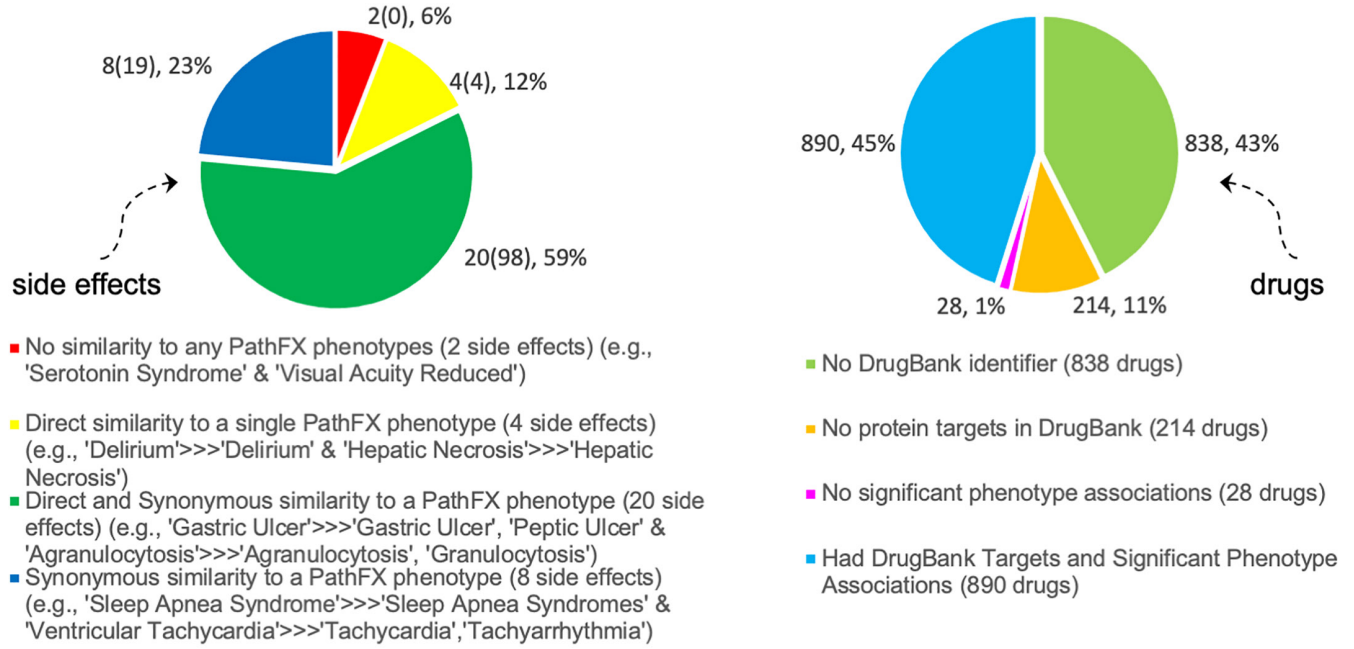

FIGURE 1 Side effects and drugs of the study. On the left, 34 side effects mapped to 121 PathFX pathway phenotypes: the pie chart indicates the fraction of total PathFX pathway phenotypes where red, yellow, blue, and green indicate whether the side effect matched to no, a direct, a synonymous, or both a synonymous and direct PathFX pathway phenotype, respectively. On the right, 890 drugs had sufficient data to generate network predictions: the pie chart indicates the active ingredients where blue, green, pink, and orange represent active ingredients that had drug bank targets and sufficient interactions, had no DrugBank identifier, had drug-binding targets but empty networks, or lacked drug-binding proteins, respectively.

TABLE 1 Detailed mapping information of drug toxicity side effects to PathFX pathway phenotypes.

Side effects to pathway phenotypes	Example
Four side effects only had 4 direct matches	'Delirium': 'Delirium'
20 side effects had 98 direct or synonymous and/or related matches	'Gastric Ulcer': 'Gastric Ulcer', 'Peptic Ulcer'
Eight side effects had no direct matches but 19 synonymous and/or related matches	'Ventricular Tachycardia': 'Tachycardia', 'Tachyarrhythmia'

proteins (in gray). In these examples, atropine is considered a "true positive" drug because PathFX correctly predicted multiple pathway phenotypes related to the drug's labeled side effect, "hypertension." Further, atropine's targets and network proteins – EDN1, KING1, and AGT – are considered true positive proteins because they were used in a "correct" PathFX prediction. Similarly, caffeine is a true positive drug for edema because PathFX-predicted edema, a pathway phenotype relevant to caffeine's labeled side effect, "edema." In this network, the protein, ERBB3, is considered "distinct" because it is not used in any PathFX edema predictions for drugs that are not labeled for the side effect. All baseline PathFX predictions, pathway phenotypes, and network proteins are available on GitHub.

Using our side effect prediction method, SEPred, which assessed all networks for drugs associated with one side effect, we observed average sensitivity and specificity values of 0.23 (Median = 0.20, SD = 0.21) and 0.83 (Median = 0.83, SD = 0.15), respectively (two examples highlighted in Table 2). Additionally, in Table S1, we provide detailed PathFX predictions for drugs associated with all 32 side effects, providing a comprehensive overview of the predictive outcomes for each individual side effect. As anticipated, PathFX v2 prediction performance demonstrated a relatively low and variable accuracy across different side effects.

Despite the modest overall performance, our approach discovered network genes associated with either or both TP and FP pathways, as depicted in Figure 2. Example networks for two drugs are shown in Figure 2 and demonstrate PathFX correctly identifying pathway phenotypes relevant to the drug's labeled side effects. For instance, atropine is associated with the side effect, "hypertension," and PathFX predicted four relevant pathway phenotypes: "essential hypertension," "genetic hypertension," "idiopathic pulmonary arterial hypertension," and "pulmonary hypertension." PathFX also used five drug-binding proteins and three downstream proteins to support these predictions. Interestingly, some proteins are shared between

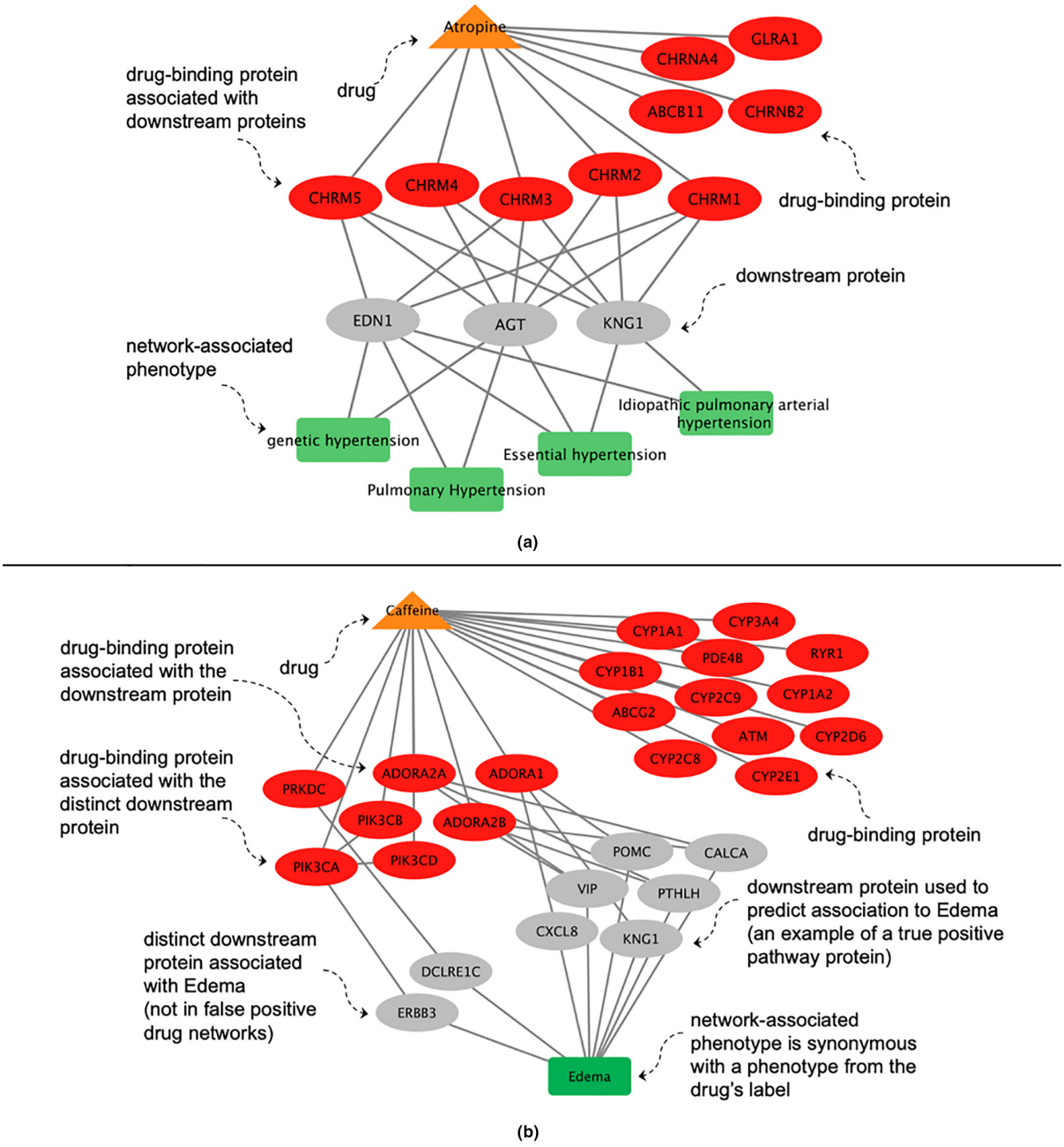

FIGURE 2 PathFX identified hypertension (a) and edema (b) genes in subnetworks for atropine and caffeine, respectively.

true and false positive drug predictions, including EDN1 which is used in true positive networks like atropine, and in additional false positive drug networks. Moreover, PathFX correctly predicts an association between caffeine and the side effect, "edema." PathFX used the network protein ERBB3, and because this protein was not used in any false positive drug networks, it is considered distinct from TP edema predictions (further illustrated in Figure 3). In this case, genes such as EDNI would be considered a TP gene for the hypertension side effect, in this subnetwork, because it connects a drug to a relevant side effect pathway phenotype (Figure 2a). It should be emphasized that EDN1 is considered a TP and FP pathway gene for hypertension in our interactome due to its connection to relevant and irrelevant side effect pathway phenotypes. We assessed these proteins for all side effects

TABLE 2 PathFX prediction results for two side effects (hypertension and edema).

Side effect	Sensitivity	Specificity	# of associated drugs in the toxicity dataset	# of drugs predicted by PathFX	# of shared TP & FP genes	# of distinct TP genes	# of distinct FP genes
Hypertension	0.54	0.63	448	406	255	108	29
Edema	0.15	0.85	442	134	21	4	21

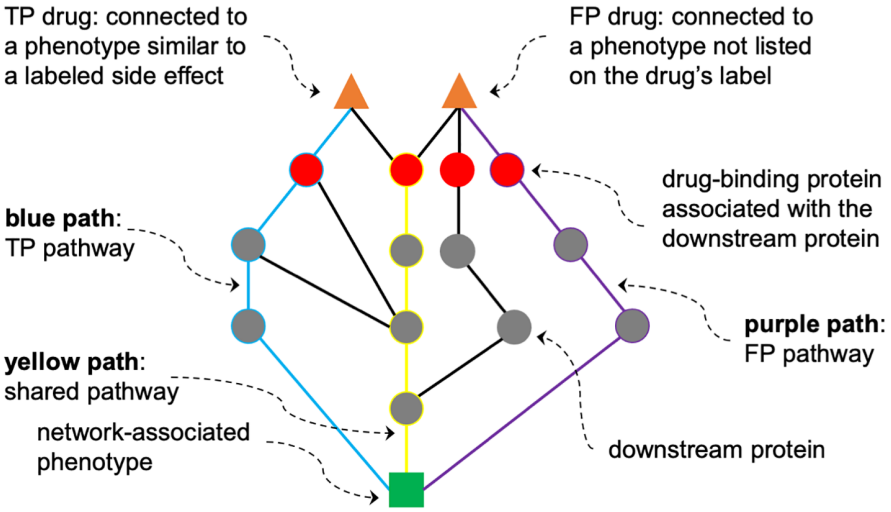

FIGURE 3 Comparison of true positive (TP) (in blue), false positive (FP) (in purple), and shared (in yellow) pathways in drug-induced side effect prediction using SEPred.

and assessed their frequencies in drug-pathway phenotype predictions. The most frequent genes in the TP and FP PathFX network associations to hypertension included as follows: [TP: ("AGT," 666), ("EDN1," 449), ("KNG,1" 425)] and [FP: ("AGT," 363), ("EDN1," 245), ("KNG1," 186)] (Figure 2a). In addition, "ERBB3" and "DCLRE1C" are downstream distinct edema genes discovered by PathFX (Figure 2b). In total, we found PathFX network associations to hypertension connected with 406 genes; 108 and 29 were uniquely associated with TP and FP associations, respectively. Similarly, for edema, we discovered 134 genes; 4 and 21 were uniquely associated with TP and FP associations, respectively (Table 2). We collected distinct TP gene lists to use as new pathway phenotypes and wanted to measure the ability to reduce overprediction for these side effects using these new pathway genes.

Furthermore, to understand the biological functions of TP genes/proteins, we used GO enrichment (Table S2, the other GO terms in Table S3). Generally, GO enrichment uncovered associations to multiple cellular processes, though, a few terms stood out for their connections to cardiovascular disease. We chose five GO terms from the most statistically significant GO terms list and found additional evidence in the literature to show their biological relevance (see 'Section 2'). For instance, the TP genes were associated with a "regulation of endothelial cell migration" (GO:0010594, false discovery rate (FDR) q-value of 0.0168). Research has elucidated that disturbances in the regulation of endothelial cell migration are intricately linked with hypertension—a condition where elevated blood pressure can compromise endothelial integrity and disrupt the normal flow regulation within blood vessels. As blood pressure surges, it can precipitate endothelial dysfunction, potentially exacerbating the severity of hypertension.[42] The interplay between hypertension and endothelial dysfunction is also multifaceted, influencing a spectrum of cellular activities that include the migration of endothelial cells.[43] Reviews in the field have further detailed how endothelial dysfunction, as a central element of cardiovascular pathologies such as hypertension, correlates with the regulatory mechanisms of endothelial cell migration.[44] These results affirmed the relevance of PathFX-prioritized proteins per side effect.

Through meticulous examination of TP, FP, and shared pathways, (as illustrated in example cases in Figure 3), we gained a holistic understanding of the predictive performance of our model, identified frequent molecular interactions, and generated new pathways to test PathFX predictions. Taken together, these results suggest that prediction performance is weak but biologically relevant. We hypothesized that meta-analysis of PathFX proteins, or the integration of new omics data may improve the ability to predict drug-induced side effects. These findings contribute to our overall goal of improving preclinical safety evaluation and drug development strategies.

Pathway engineering and performance assessment of newly defined pathways

Pathway engineering can reduce overprediction of drug-induced side effects and increase sensitivity when side effect pathways are poorly defined

We hypothesized that using downstream proteins correctly associating a drug to its side effect to define new pathway phenotypes could improve our ability to predict drug-induced effects. In a previous study,[22] we made an interesting discovery that edema-associated downstream proteins were highly predictive of whether a drug caused edema, even when all edema-causing drugs did not share direct drug targets. Furthermore, our baseline analysis confirmed that there were distinct sets of genes associated with TP and FP-predicted pathway phenotypes (e.g., case study side effects in Table 3) and that they were functionally relevant to drug side effects as texted by GO enrichment (Tables S2 and S3). These findings suggest that by gaining a better understanding of the pathway phenotypes and identifying genes that are linked with both TP and FP drugs, we could potentially improve the accuracy of our predictions. However, we used PathFX v1 in our previous analysis and we applied the logistic regression on all network proteins, instead of just pathway phenotype-associated proteins. In this work, we initially examined all drug network proteins rather than just the pathway phenotype-associated proteins to define new pathway lists. Nevertheless, reviewing the lists of network proteins associated with TP and FP results, we found no significant distinctions between them. We were primarily interested in defining gene lists with distinct associations; thus, our focus shifted toward utilizing proteins with direct associations to pathway phenotypes.

In our study, we leveraged a list of distinct TP genes associated with side effects, identified in the previous section, and made new gene-to-phenotype associations. To address potential biases in the data, we performed random exclusions, removing 15% of distinct TP genes to create knockout gene lists and adding 15% of shared TP and FP genes to the distinct TP gene list to simulate the effects of noisy input data. Subsequently, we included these new pathways with the genes associated with the previously mentioned 121 mapped pathway phenotypes. After integrating these pathways into PathFX, we again generated networks for all TP and FP drugs for each selected side effect and assessed whether our new pathway phenotypes were recovered in positive (drugs cause the side effect in the toxicity dataset) or negative (drugs did not cause the side effect in our toxicity dataset) drug networks, using the DefPath method. This analysis resulted in tables of pathway phenotypes associated with the respective drug networks.

We next assessed which of these newly defined pathway genes were recovered in TP and FP drug networks (Table 3). The first three rows of Table 3 provide evidence of a reduced overprediction. We highlighted four selected side effects, that had the highest number of TP genes among all 32 side effects (all other side effects shown in Table S1). For instance, with hypertension, our baseline results recovered 255 shared hypertension-associated genes found in TP and FP drug networks; 108 and 29 genes were also unique to TP and FP drug networks, respectively. When we redefined, new pathway phenotypes using the 108 genes associated with only TP drugs and ran the new version of PathFX with 241 TP and 165 FP drugs identified in the baseline analysis, we found fewer genes recovered in PathFX networks. Specifically, for 241 TP drugs associated with hypertension, 21 out of 108 distinct TP genes were discovered ("TP defined genes-In TP drug networks" column in Table 3). When using a 15% knockout pathway, we observed that all TP predictions were connected through 22 pathway genes, a similar recovery to the TP distinct pathway. Importantly, for both pathways, the number of FP predictions was zero. This aligns with our initial hypothesis, affirming the reduction in overprediction. These results further suggest that smaller, curated pathways may be valuable for reducing overprediction. For myocardial infarction, the number of TP genes falls below the PathFX detection threshold (PathFX requires at least 25 genes in a pathway; otherwise, the pathway is skipped entirely based on findings from Menche et al.[45]). However, for side effects with more than 25 genes, that is, hypertension, pancreatitis, and thrombocytopenia, our analysis demonstrates compelling findings, as we have not captured any TP-defined pathways in FP drug networks.

Subsequently, we evaluated the recovery of pathway genes in noisy pathways. We reasoned that it is difficult to precisely define a side effect pathway phenotype, and investigating noisy inputs would quantify the impact of having erroneous genes in the pathway gene lists. The number of noisy pathway genes discovered for all four side effects is relatively high and not surprisingly, noisy pathways were identified for both TP and FP drugs. In the case of hypertension, 183 and 90 genes from the noisy pathway were discovered in the TP and FP drug networks (Table 3). The relative increase in FP predictions is likely due to the relatively large number of genes randomly added to the pathways, and because these genes were added from the shared gene list. This further supports our finding that a smaller, less noisy list (i.e., the knockout list) is preferable for reducing overprediction compared with a larger, noisier list. In the case of myocardial infarction, we had insufficient genes to define TP and knockout pathways (discussed above), but generating a noisy pathway enabled prediction of the side effects when we

TABLE 3 Number of discovered baseline and defined pathways per side effect.

Side effect	Baseline results			New results					
				TP-defined genes		Knockout-defined genes		Noise-defined genes	
	Shared genes	Distinct TP genes	Distinct FP genes	In TP drug networks	In FP drug networks	In TP drug networks	In FP drug networks	In TP drug networks	In FP drug networks
Hypertension	255	108	29	21	0	22	0	183	90
Pancreatitis	282	43	128	11	0	9	0	92	168
Thrombocytopenia	62	31	20	21	0	16	0	0	0
Myocardial infarction	219	20	104	0	0	0	0	147	174

Note: Shared genes are downstream proteins associated with a relevant PathFX pathway phenotype found in TP and FP drug networks. TP and FP genes are distinct pathway phenotype proteins found only in TP or FP drug networks, respectively. The TP-defined genes are recovered proteins in TP and FP drug networks. We removed 15% of distinct TP genes and added 15% of shared TP and FP genes to the distinct TP gene list to make the knockout and noise genes.

previously had insufficient data. The discovery of FP associations indicates that the process is not robust to noise but suggests that noisy pathways may be necessary for increasing sensitivity.

Integration of gene expression data increases pathway identification

It is important to note that, while we initially considered a broader range of pathway phenotypes in our analysis, a majority of them, except for four, exhibited fewer than 25 distinct TP genes in their pathways. Consequently, the PathFX tool excluded these pathways from our study due to their limited gene lists. In order to address this challenge and extend the applicability of our approach to cases with inadequate distinct TP example networks, we undertook the generation of novel pathway phenotypes using an omics dataset, the "drug signature data" published with the PharmOmics dataset.[25] Furthermore, gene expression data are often widely available, especially in cases of novel drugs under development, and we were eager to understand the utility of this data for pathway-based prediction of drug side effects. By leveraging this omics dataset, we conducted experiments to assess the effectiveness of these newly defined pathway phenotypes in mitigating overprediction and enhancing the precision of our predictive model. Although all side effects are not driven by gene expression changes, pathway engineering could illuminate whether gene expression changes had any long-range connections to drug targets.

Briefly, the drug signature dataset included lists of top genes affected by exposure to 941 drugs in three species. PathFX uses only human protein data; therefore, we prioritized differential gene lists tested in human cells and drug names listed in our toxicity dataset. We originally hypothesized that a pathway defined by all differentially expressed genes shared by positive drugs, or each drug's "signature" would be the most sensitive for predicting a side effect (e.g., all top genes from hypertension-causing drugs would constitute a new "hypertension signature" pathway). However, we discovered that the intersection of top gene lists for all side effect drugs was zero for all side effects. Given this limitation, we instead used gene sets discovered from the union of top gene lists associated with both positive and negative drugs and repeated this process per side effect. This yielded TP and FP "signature" gene lists. This scenario is most like our previous definition of noisy pathways, and we anticipated a trade-off between sensitivity and specificity using this approach.

We incorporated four gene sets, namely TP and FP baseline genes, along with TP and FP signature genes, to establish new pathways as described previously (see

'Section 2'). For example, hypertension was associated with 51 TP and 35 FP drugs, and the gene lists were as follows: TP baseline: 363, FP baseline: 284, TP signature: 5502, and FP signature: 3117 drug-gene pairs, respectively. Initially, we took the union of all top TP signature genes for drugs associated with eight side effects that had the highest number of distinct TP genes, in the toxicity dataset. This resulted in side effect pathways encompassing 78,435 gene–phenotype associations, with individual pathway phenotypes ranging from 250–6787 genes. Table S4 shows all pathway phenotypes and their corresponding pathways in Supplementary Materials S1. Surprisingly, after incorporating these into PathFX, we were unable to detect drug associations to the pathways, likely because the pathways were too large and not uniquely associated with any drug pathway. This is consistent with earlier findings that PathFX is unable to detect pathway associations when the pathway sizes are large relative to the size of the interactome.[21] Subsequently, we repeated the process, but this time, we took the top 500 ranked differentially expressed signature genes (per side effect) to define new TP signature pathways for four side effects, yet still failed to recover associations to these pathways in PathFX networks, likely since gene expression does not always align with protein signaling pathways. To address this issue, we created a process called "pathway engineering," aiming to discover more performant sets of intersecting genes that could yield defined proteins (signals) in PathFX results.

The Venn diagram in Figure 4a visually depicts all possible intersections among our gene sets. After considering various intersection cases, we found that taking the intersection of the TP baseline and signature gene sets, while removing the FP baseline and signature gene sets (Figure 4b), could remove the FP signals in PathFX results. However, for all side effects, the final intersection contained fewer than 25 genes, leading to PathFX disregarding those networks. To overcome this limitation, we explored two additional cases. In one instance, we included the intersection of all gene sets and excluded FP baseline genes (Figure 4c), and in the other case, we incorporated the intersection of all gene sets and excluded the FP signature genes (Figure 4d). Like before, we next assessed which pathway genes were recovered in PathFX networks (Table 4) for various cases. Zero values in the table represent cases with fewer than 25 genes in the network. Moreover, the non-zero values in the fourth ("# of TP signature (not included FP distinct) genes") and fifth ("# of TP signature (not included FP signature) genes") columns confirm that signature genes are recovered in PathFX networks and validate their utility in defining new side effect pathways. For myocardial infarction, no pathway genes were recovered in PathFX networks when only including TP distinct genes or TP signature genes without FP distinct genes. However, when incorporating TP signature genes without FP signature genes (a similar case to Figure 4d), we recovered 113 genes, highlighting the significance of utilizing drug signature data in our analysis.

Furthermore, we considered that the inclusion of the FP signature data may most resemble our previous analysis of noisy pathways and hypothesized that these pathways may increase sensitivity (TP predictions) for predicting drug side effects at the expense of reduced specificity. We next sought to quantify the utility of these new pathways for correctly predicting drug-induced side effects.

Novel pathways increase specificity at the cost of reduced sensitivity

Our primary objective was to minimize FPs and overprediction in our model. Unlike our initial analysis, where we focused on TP and FP drugs for specific side effects, here, we studied all 890 drugs in our dataset. Subsequently, we calculated sensitivity and specificity by comparing the PathFX drug predictions with the drugs listed in the drug toxicity dataset (refer to the 'Section 2' section for more details). In Figure 5, we present the evaluation metrics for four selected side effects and the performance of our novel side effect pathways. All engineered pathways had improved specificity compared with the baseline analysis. For all side effects, the TP distinct pathway (side effect gene list generated from networks correctly associated with a side effect) had the highest specificity. However, these pathways often had relatively low sensitivity, likely due to having relatively few genes. Similarly, the "TP dist&sig/noFP dist" pathway had high specificity and relatively low sensitivity. Removing FP-network genes from the pathway reduced overprediction, but the addition of signature genes did not increase sensitivity for all side effects; in pancreatitis, the addition of signature genes increased sensitivity, suggesting that pathway engineering could perform differently across pathway phenotypes. This endeavor resulted in a trade-off between specificity and sensitivity values. Consequently, we achieved higher specificity (fewer FPs) but lower sensitivity in our predictions. We were able to increase sensitivity, but only at the cost of reduced specificity. Improved sensitivity required either noisy pathways or the inclusion of genes associated with FP drugs; this suggests that there is not a single pathway for identifying side effects or that it is impossible to sufficiently discern the networks of drugs that do or do not cause a side effect.

To assess the utility of our newly defined pathways for side effect prediction, we conducted a comparison with an animal testing study,[4] where we observed variations in the

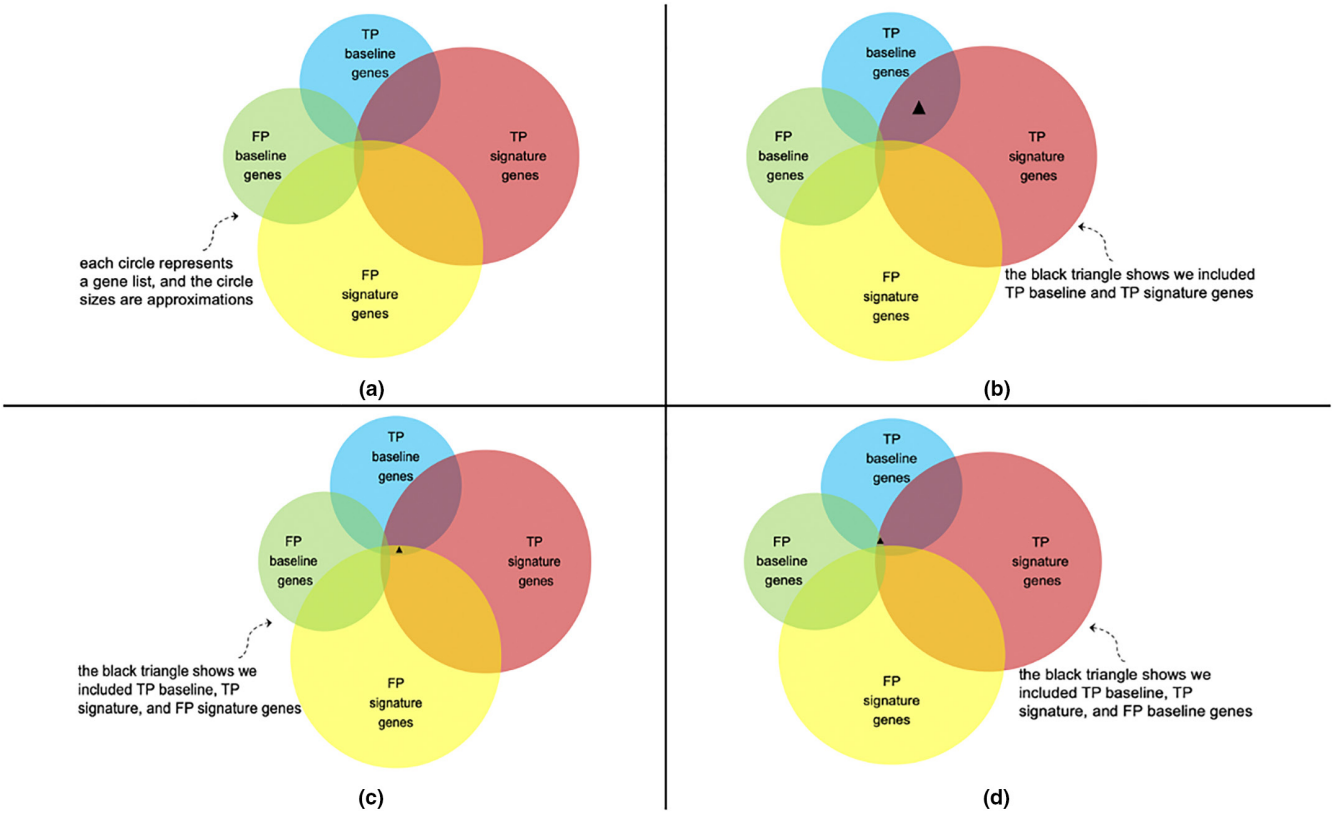

FIGURE 4 Venn diagram illustrating all gene sets intersections in our pathway engineering process. (a) True positive (TP) and false poitive (FP) baseline genes are represented by blue and green circles, respectively, and include the network-identified gene lists described previously in our preliminary analysis. The TP and FP signature gene lists are represented by red and yellow circles, respectively, and include the lists of genes identified from the PharmOmics dataset. As an example, for hypertension, the number of genes in these lists were as follows: TP baseline: 363, FP baseline: 2 84, TP signature: 5502, and FP signature: 3117. (b) The black triangle shows that we included genes intersecting the TP baseline and TP signature sets and removed the FP baseline and FP signature sets. (c) The black triangle shows that we included genes intersecting TP baseline, TP signature, and FP signature sets and removed FP baseline sets. (d) The black triangle shows that we included genes intersecting TP baseline, TP signature, and FP baseline sets and removed FP signature sets.

TABLE 4 New pathway genes recovered in updated PathFX networks.

Side effect/New pathway	# of TP distinct baseline genes	# of TP distinct (not included FP distinct) genes	# of TP signature (not included FP distinct) genes	# of TP signature (not included FP signature) genes
Hypertension	108	21	21	195
Pancreatitis	43	11	12	95
Thrombocytopenia	31	21	0	32
Myocardial infarction	20	0	0	113

Note: TP distinct genes: Network proteins from TP drug networks not contained in FP drug networks. FP distinct genes: Network proteins from FP drug networks not contained in TP drug networks. TP signature genes: Differentially expressed genes from PharmOmics TP drug networks not contained in FP drug networks. FP signature genes: Differentially expressed genes from PharmOmics TP drug networks not contained in FP drug networks.

performance depending on the side effect and the organ system category. We compared our results to Monticello et al.[4] which collected preclinical and clinical rates of side effects by organ system (e.g., "cardiovascular" or "gastrointestinal") and reported sensitivity and specificity of three animal models—"rodent," "dogs," and "non-human primates"—for their ability to anticipate clinical side effects (Figure 5). They observed that while animal testing can predict some side effects in humans, their sensitivity is limited, and sensitivity varies depending on the organ system. Comparing our results with the animal study findings (we matched side effect pathways to the most relevant organ system—for example, hypertension, thrombocytopenia, and myocardial infarction were compared

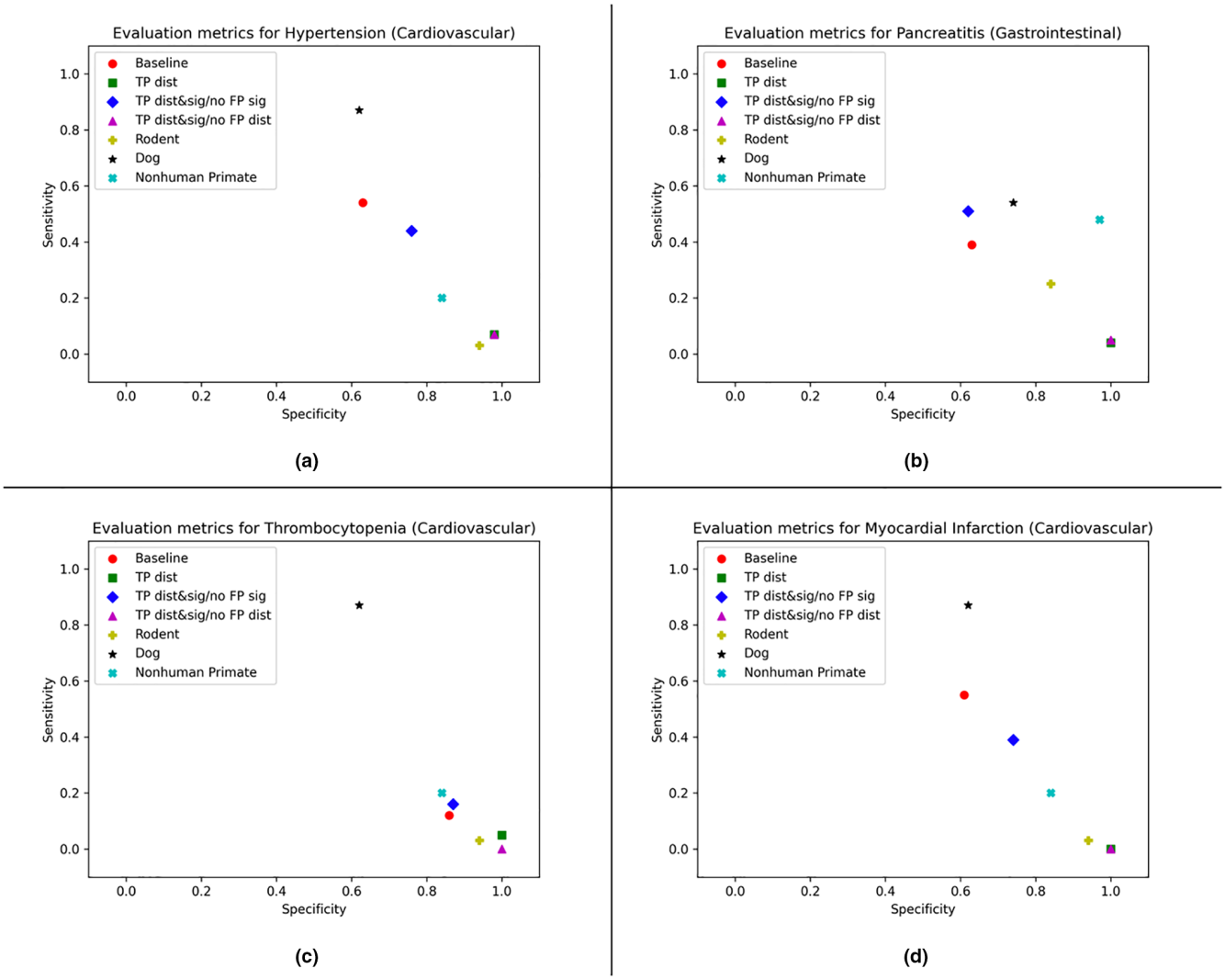

FIGURE 5 Comparison of human in silico and animal testing[4] evaluation metrics for four side effects in cardiovascular and gastrointestinal organ system categories. We plotted the sensitivity (y-axis) against the specificity (x-axis) for 4 PathFX pathway phenotypes, hypertension (upper left), pancreatitis (upper right), thrombocytopenia (lower left), and myocardial infarction (lower right), to compare evaluation metrics for relevant organ systems published in Monticello et al.[4] In all cases, the "baseline," "TP distinct," "intersection of TP distinct, TP signature, and FP distinct, w/o FP signature," and "intersection of TP distinct, TP signature, and FP signature, w/o FP distinct" genes are represented by a red circle, green square, blue diamond, and purple triangle, respectively. Rodent, dog, and nonhuman primate data are represented by a yellow plus, black star, and cyan x, respectively, and are as published in Monticello et al.[4]

with the cardiovascular organ system and pancreatitis was compared with the gastrointestinal organ system), it is crucial to consider that animals and humans differ in terms of physiology, anatomy, and genetics, leading to varying responses to drugs and treatments. However, our results are within the same range as those obtained from the animal study, depending on the organ system. For cardiovascular side effects, regarding sensitivity, the dog outperformed other animal models and our pathway predictions. Our engineered pathways had similar sensitivity to the dog for gastrointestinal side effects and greater sensitivity than the rodent or nonhuman primate for this organ class. By conducting this comparative analysis, we demonstrate the potential of our pathway engineering approach to enhance the accuracy of drug-induced side effect predictions, considering the differences between animal and human responses.

As previously discussed, each side effect can be associated with several pathway phenotypes. For example, hypertension encompasses 12 pathway phenotypes, which are synonymous or related terms. Table 5 presents the evaluation results for each pathway phenotype associated with the hypertension side effect. Notably, our findings align with the outcomes of animal testing studies, demonstrating the potential utility of our predictions. Through our pathway engineering approach, we succeeded in reducing overpredictions by minimizing the number of FPs. However, achieving a specificity

of 1 remained unattainable, primarily due to limitations within our dataset (not having enough distinct TP genes to incorporate and the inclusion of a few distinct FP genes in our analysis, as illustrated in Figure 4d). It is also important to note that "TP distinct" does not rule out the fact that these genes could show up in a FP drug pathway. Because we defined the pathway from genes enriched in a drug's pathway in the association table, another drug network may have the pathway genes in the network but will not be enriched because of Fisher's exact test. Once we change the pathway length, that second drug network may become enriched in the pathway phenotype, leading to the discovery of the TP distinct genes in the FP pathway.

More importantly, pathway engineering enhanced signal detection and uncovered previously unidentified effects. For example, after pathway engineering, caffeine's network included 59 new pathway phenotypes, and increased support for predicting the drug's association with hypertension (Figure 6a). This process revealed newly defined pathways, such as "phen_sig_tp_noFP_Sig_hypertension" and "phen_sig_tp_noFP_Dis_hypertension." Moreover, we identified new pathway phenotypes that support a drug's labeled side effect. Specifically, we observed the new pathway phenotype, "phen_sig_tp_noFP_Sig_myocardial infarction," in the network for acebutolol, which previously had no network associations with myocardial infarction, demonstrating the improved detection capabilities of the PathFX network after pathway engineering (Figure 6b). We measured the general utility of pathway engineering through changes in sensitivity for predicting additional side effects: hypertension (sensitivity: 0.44 (TP:198 and FP:104) vs 0.54 (TP:241 and FP:165), Table 5, Table S1, and Figure 7, Figure S1), myocardial infarction (sensitivity: 0.39 (TP:126 and FP:145) vs. 0.55 (TP:178 and FP:218), Tables S1, S6 and Figure S2), thrombocytopenia (sensitivity: 0.16 (TP:72 and FP:57) vs. 0.12 (TP:57 and FP:61), Tables S1, S7, and Figure S3), and pancreatitis (sensitivity: 0.51 (TP:142 and FP:235) vs. 0.39 (TP:109 and FP:227), Tables S1, S8, and Figure S4). This shows that the impact of pathway engineering can change depending on the pathway phenotype.

After pathway engineering, we also discovered new drug predictions, and distinct predictions when comparing the results with and without the integration of gene expression data (Figure 4b,d, respectively). For instance, as shown in Table S5, in the case of hypertension, among 198 drugs associated with "phen_sig_tp_noFP_Sig_hypertension" pathway phenotype, a gene expression-derived pathway (Table 5), only 16 were shared with the "phen_tp_hypertension" pathway phenotype, a non-gene expression-derived pathway, which was associated with a total of 31 drugs (Table 5). This highlights the utility of using the drug signature dataset to identify new and unique drugs.

To further display our tool's performance and enhance clarity, we incorporated visual confusion matrices with

TABLE 5 Per-pathway phenotype performance metrics and comparison with animal testing results for hypertension.

Side effect	Pathway phenotype	TP	TN	FP	FN	Sensitivity	Specificity
Hypertension	Idiopathic pulmonary arterial hypertension	0	442	0	448	0.00	1.00
Hypertension	Hypertension	22	438	4	426	0.05	0.99
Hypertension	Phen_tp_hypertension	31	434	8	417	0.07	0.98
Hypertension	Phen_sig_tp_noFP_Dis_hypertension	30	434	8	418	0.07	0.98
Hypertension	Essential hypertension	26	423	19	422	0.06	0.96
Hypertension	Hypertension, portal	13	419	23	435	0.03	0.95
Cardiovascular organ category (Rodent)						0.03	0.94
Hypertension	Renal hypertension	19	412	30	429	0.04	0.93
Hypertension	Ocular hypertension	121	396	46	327	0.27	0.90
Hypertension	Idiopathic pulmonary hypertension	50	393	49	398	0.11	0.89
Cardiovascular organ category (Nonhuman primate)						0.20	0.84
Hypertension	Hypertension, renovascular	121	368	74	327	0.27	0.83
Hypertension	Hypertensive disease	156	362	80	292	0.35	0.82
Hypertension	Genetic hypertension	98	350	92	350	0.22	0.79
Hypertension	Phen_sig_tp_noFP_Sig_hypertension	198	338	104	250	0.44	0.76
Hypertension	Pulmonary hypertension	191	315	127	257	0.43	0.71
Hypertension	Prehypertension	225	298	144	223	0.50	0.67
Cardiovascular organ category (Dog)						0.87	0.62

Note: TP, TN, FP, and FN represent the true positive, true negative, false positive, and false negative drugs identified by each pathway phenotype.

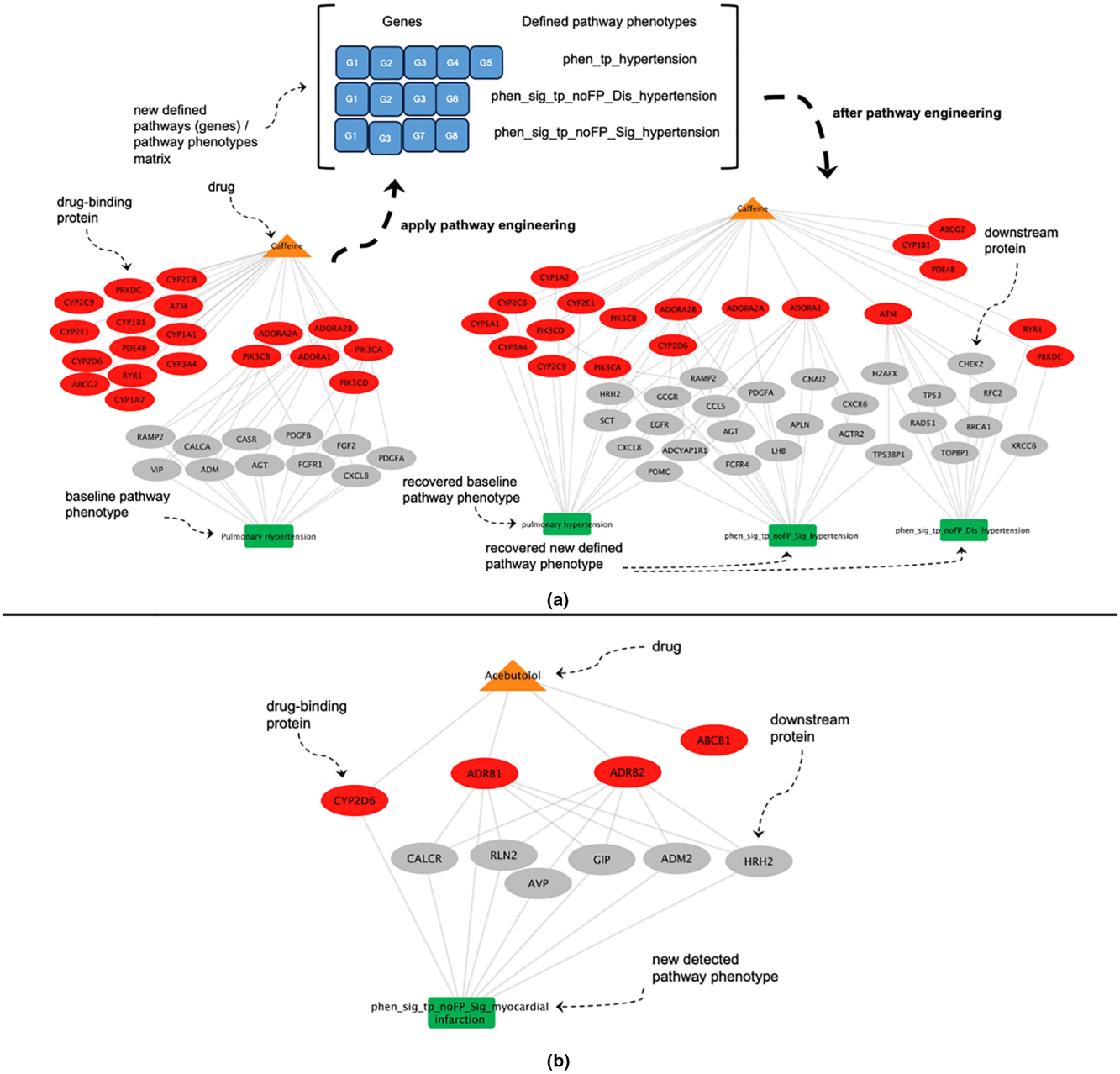

FIGURE 6 Enhanced detection of pathway phenotypes following pathway engineering, highlighting the emergence of (a) the newly defined and (b) previously undetected pathway phenotypes associated with caffeine and acebutolol, respectively, exemplifying the improved efficacy of the PathFX network in characterizing drug-related side effects.

an initial emphasis on hypertension for the baseline case as well as cases in which we considered distinct TP genes and omics data (Figure 7). In the case of hypertension, the confusion matrix shows that the engineered pathways have decreased overprediction compared with the baseline. We have also provided confusion matrix visualizations for all four side effects under investigation in our study in Figures S1–S4. These additions serve to offer a clearer and more accessible illustration of our methodology's effectiveness across various scenarios.

In addition, we extended our analysis to three additional side effects, and the results are available in Tables S6–S8. This comprehensive evaluation allows us to gain further insights into the performance and effectiveness of our pathway engineering method across various side effects.

DISCUSSION

Network-based methods suffer from low prediction accuracy; their prediction relies on accurate side effect pathways, but there are many ways to define pathways without a full understanding of pathway definition on network

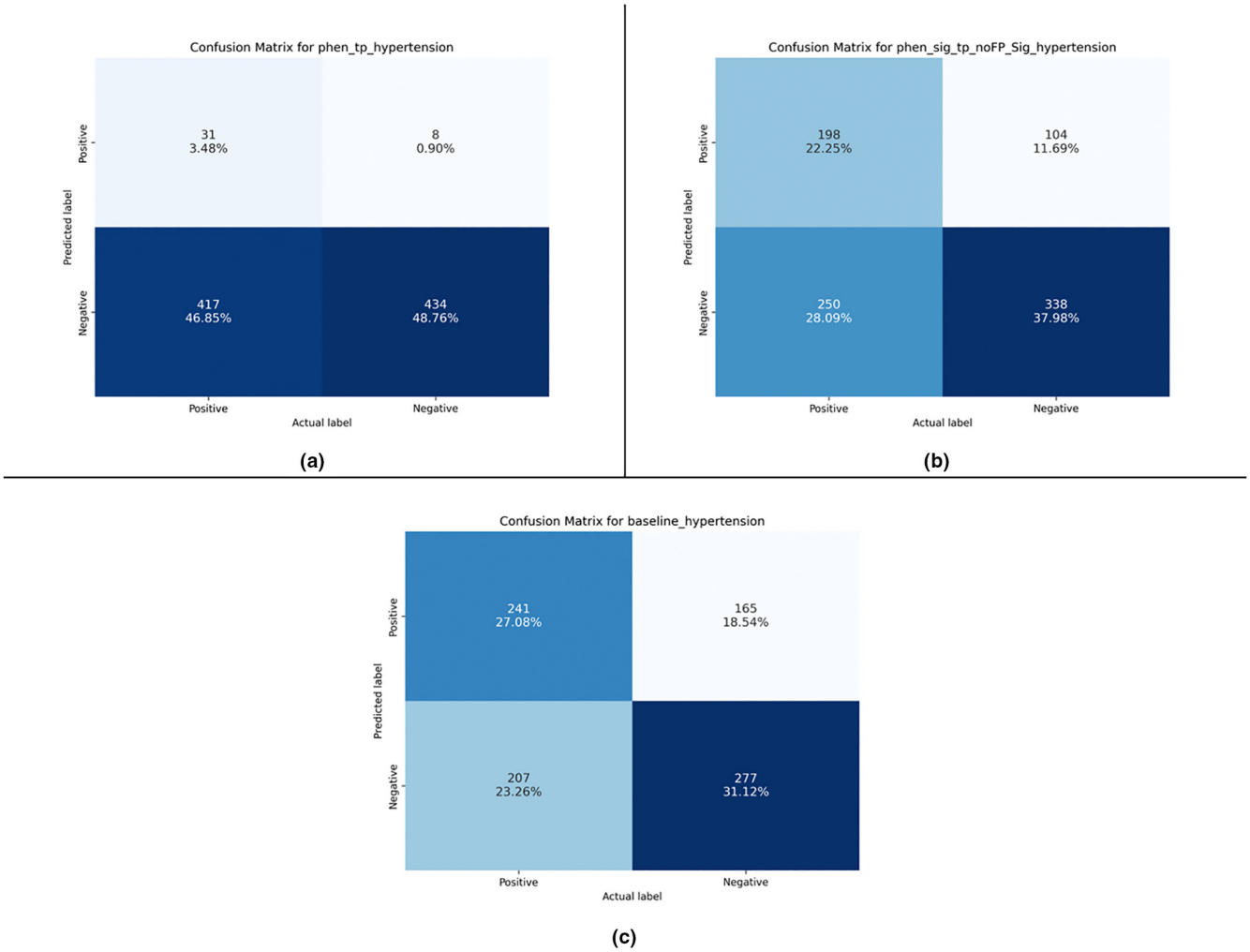

FIGURE 7 Confusion matrices for hypertension represent the number of drugs predicted by various engineered pathway phenotypes. (a) The case includes only genes from the true positive (TP) baseline set. (b) The case includes genes intersecting TP baseline, TP signature, and FP baseline sets and removes false positive (FP) signature sets. (c) The case includes genes from the baseline analysis.

prediction performance. Here we investigated the effect of defining new gene pathways on prediction performance of one network method for anticipating several drug-induced side effects. We generated pathways that were able to reduce overprediction when there were sufficient examples of TP drugs. We also demonstrated that the incorporation of gene expression data can increase sensitivity at the expense of reduced specificity. The results of our study demonstrate the potential of protein interaction network models and pathway engineering strategies to enhance preclinical prediction of drug-induced safety events. By leveraging our PPI tool, PathFX, we studied the limitations of network-based methods, that is, low prediction accuracy and overprediction.

By refining the pathways in our algorithm and developing pathway phenotypes that reduced overprediction, we made progress in improving preclinical safety evaluation and drug development strategies. The gene-to-phenotype associations derived from distinct TP genes yielded the greatest reduction in overprediction, but at a great cost to sensitivity. We observed this trend for specific side effects, such as hypertension, pancreatitis, and thrombocytopenia. Additionally, we incorporated omics data from a drug signature dataset to generate novel side effect pathway phenotypes. These pathways generally improved our sensitivity to detect pathway phenotypes related to drug side effects and specifically enabled us to predict side effects missed by our baseline mode (we specifically highlighted an example for myocardial infarction). These results highlighted the potential of integrating gene expression data into pathway phenotypes to improve network model predictions. Generally, pathways with omics data increased the sensitivity of our network model for prediction pathway phenotypes related to drugs' labeled side effects, but at the cost of reduced specificity (these pathways were also associated with false positive drug predictions). These results highlight the difficulty in defining highly performant pathway

phenotypes and demonstrate a possible limit of pathway engineering. We also compared our model predictions to a published animal testing study and observed that network predictions were better, worse, or similar to animal models, depending on the side effect. Essentially, animal models also had a trade-off in sensitivity and specificity, yet they are still widely used in drug development and provide valuable insights despite imperfect prediction performance. Taken together, these results highlight the utility of our pathway engineering approach to enhance drug-induced side effect predictions. We specifically envision network methods to be a complementary analysis to other methods of preclinical side effect prediction.

Many have modeled side effect prediction, albeit with different datasets, and different model types. We can qualitatively compare their performance, but they are generally different from our approach in the emphasis on different side effects and different types of prediction. Liang et al.[46] proposed a binary classification model incorporating a refined negative sample selection strategy to predict drug side effects. This study utilized three classification algorithms: Random Forest (RF), Support Vector Machine (SVM), and Artificial Neural Network (ANN), with RF demonstrating superior performance, evidenced by a sensitivity of 0.923 and specificity of 0.999. However, their approach only predicted a binary classification and is distinct from our efforts to predict specific side effects. Galeano et al.[47] introduced geometric self-expressive model (GSEM) to predict unknown side effects for drugs with a small number of side effects identified in clinical trials. While they reported the area under the receiver operating characteristic (AUROC) values above 0.9, their goals are again different from ours in that clinical side effects are exploratory, collected from small patient samples, and may include less severe outcomes, such as nausea and dizziness. In contrast, we emphasized severe adverse outcomes that warranted inclusion on drug labels, further highlighting our contribution to the understanding of severe adverse events. Lastly, Huang et al.[15] developed an in silico model that achieved a satisfactory cardiotoxic side effect prediction performance (median AUC = 0.771, Accuracy = 0.675, Sensitivity = 0.632, and Specificity = 0.789). Although these performance metrics are higher, they only represent prediction for a single side effect, and we aimed to apply pathway engineering to multiple severe adverse reactions. Taken together, these examples underscore the challenges and variations in methods for predicting side effects. They also highlight that our contribution—leveraging PPI networks to understand severe side effects—complements existing approaches.

However, it is important to acknowledge several limitations of the study, largely related to available data. It should be noted that we were unable to define PathFX pathway phenotypes for all labeled side effects, due to insufficient pathway data and we were unable to create networks for all drugs due to insufficient drug target data. Future work may consider the inclusion of additional, or predicted, drug targets to better connect drugs to side effect pathway phenotypes. We originally considered the full TP and FP drug network, and we observed no substantial differences between the associated network genes; drugs with and without the labeled side effects shared similar networks. We next investigated statistically significant pathway phenotype associations because these were associated with subsets of drug network proteins. However, using these network proteins to define new pathway phenotypes reduced sensitivity, likely because the pathway gene lists became too small. Considering these results and our prior results suggested that searching for distinct pathway proteins would not yield highly predictive pathway phenotypes. Instead, drugs may converge on similar network proteins, but proteins may need to be weighted or directed (to mimic the activation or deactivation of a pathway) to distinguish between true and false positive effects. Future work could consider the discovery of weighted or directed pathways to better predict drug outcomes. Moreover, many side effects had insufficient distinct TP genes, leading to exclusion from the study. This limitation restricted the applicability of the pathway engineering approach to all side effects and emphasized that our approach relied on verified positive examples. Our analysis also indicates that the pathway engineering approach may not be entirely robust to noise, as numerous noisy pathways were recovered in TP and FP drug networks. While the comparison with animal testing is informative, it is essential to recognize that animal responses may not fully reflect human reactions to drugs due to physiological and genetic differences. Further research and validation are necessary to strengthen the reliability and applicability of pathway engineering in drug development and safety evaluation processes.

The incorporation of drug signature genes had similar trade-offs to our noisy pathways in terms of increased sensitivity at the expense of reduced specificity. This suggests that there may not be a single pathway to define a side effect or that TP and FP drug networks may be too similar to accurately predict their side effects from pathways alone. However, the comparison to published animal studies underscores the utility of these pathways, despite perfect performance. As highlighted, a limitation in our work is the requirement of positive and negative examples for defining pathways. Nevertheless, for pathway phenotypes without these example cases, it may still be possible

to generate sufficiently performant pathways using omics datasets. This suggests that pathway engineering could be extended to pathway phenotypes other than side effects. Another limitation in our approach is the use of differentially expressed genes and the inclusion of effects in any tissues. Using all tissues gave us a larger pool of differentially expressed genes to consider but missed tissue-specific effects. Gene expression changes alone may insufficiently capture all drug-induced effects, and other omics data, such as proteomic changes, could enhance future efforts for pathway engineering.

Overall, our study contributes to the advancement of network-based prediction methods for drug safety evaluation. The integration of pathway engineering, distinct TP genes, and omics data offers valuable insights into the predictive performance of protein interaction network models for drug-induced side effects. By understanding the pathways connecting TP and FP predictions, we can further improve prediction accuracy. Our findings provide essential guidance for future drug development endeavors and highlight the potential of our approach in enhancing the utility of protein interaction network models for preclinical side effect prediction.

CONCLUSION

Our study aimed to enhance the preclinical prediction of drug-induced safety events through the application of protein interaction network models and pathway engineering strategies. By leveraging the PathFX algorithm and a drug toxicity dataset, we focused on downstream proteins. We defined new pathways to reduce overprediction and improve prediction accuracy. Our findings demonstrated that the identification of key genes associated with TP and FP pathways is essential for reliable predictions. Incorporating omics data from a drug-gene expression signature dataset further enhanced our network predictions. Notably, we achieved a trade-off between sensitivity and specificity values, while predicting the drug effects. The comparison of our results with an animal modeling study provided insights into the differences between animal and human responses to drugs, emphasizing the need for accurate side effect pathways for effective predictions. Our pathway engineering approach shows promise in preclinical safety evaluation and drug development strategies, providing valuable insights into drug-induced side effect predictions and pathway prioritization. Our comprehensive evaluation of various side effects further supports the effectiveness of our pathway engineering method. Overall, our study contributes to advancing the field of drug safety evaluation and offers valuable guidance for future drug development endeavors.

AUTHOR CONTRIBUTIONS

M.A. and J.L.W. wrote the manuscript. J.L.W. designed the research. M.A. performed the research. M.A. and J.L.W. analyzed the data.

ACKNOWLEDGMENTS

We would like to thank Dr. Xia Yang, Professor in the Department of Integrative Biology & Physiology at UCLA, and Dr. Jessica Ding for sharing invaluable omics data that significantly enriched our research.

FUNDING INFORMATION

No funding was received for this work.

CONFLICT OF INTEREST STATEMENT

The authors declared no competing interests for this work.

DATA AVAILABILITY STATEMENT

The code and data utilized in this study have been made accessible through a GitHub repository, "ToxPathEngine": https://github.com/jenwilson521/ToxPathEngine. The repository contains instructions for downloading and formatting the Drug Toxicity Dataset and PharmOmics data. The repository further contains scripts for rerunning key analyses (in the "scripts" directory) and the full release of the updated version of PathFX (in the "pathfx" folder).

ORCID

Mohammadali Alidoost https://orcid.org/0000-0003-2042-8736
Jennifer L. Wilson https://orcid.org/0000-0002-2328-2018

REFERENCES

1. Hay M, Thomas DW, Craighead JL, Economides C, Rosenthal J. Clinical development success rates for investigational drugs. *Nat Biotechnol*. 2014;32:40-51.
2. Wong CH, Siah KW, Lo AW. Estimation of clinical trial success rates and related parameters. *Biostat (Oxf, Engl)*. 2019;20:273-286.
3. Dirven H, Vist GE, Bandhakavi S, et al. Performance of preclinical models in predicting drug-induced liver injury in humans: a systematic review. *Sci Rep*. 2021;11:6403.
4. Monticello TM, Jones TW, Dambach DM, et al. Current nonclinical testing paradigm enables safe entry to first-in-human clinical trials: the IQ consortium nonclinical to clinical translational database. *Toxicol Appl Pharmacol*. 2017;334:100-109.
5. Fröhlich E. Replacement strategies for animal studies in inhalation testing. *Forensic Sci*. 2021;3:45.
6. Wadman M. FDA no longer has to require animal testing for new drugs. *Science*. 2023;379:127-128.
7. Ekins S, Mestres J, Testa B. In silico pharmacology for drug discovery: methods for virtual ligand screening and profiling: in silico pharmacology for drug discovery. *Br J Pharmacol*. 2007;152:9-20.

8. Gross SM, Mohammadi F, Sanchez-Aguila C, et al. Analysis and modeling of cancer drug responses using cell cycle phase-specific rate effects. *Nat Commun*. 2023;14:3450.
9. Hao Y, Romano JD, Moore JH. Knowledge-guided deep learning models of drug toxicity improve interpretation. *Patterns*. 2022;3:100565.
10. Hur J, Guo A, Loh W, Feldman E, Bai J. Integrated systems pharmacology analysis of clinical drug-induced peripheral neuropathy. *CPT Pharmacometrics Syst Pharmacol*. 2014;3:1-11.
11. Melas IN, Sakellaropoulos T, Iorio F, et al. Identification of drug-specific pathways based on gene expression data: application to drug induced lung injury. *Integr Biol*. 2015;7:904-920.
12. Zhou H, Gao M, Skolnick J. Comprehensive prediction of drug-protein interactions and side effects for the human proteome. *Sci Rep*. 2015;5:11090.
13. Keiser MJ, Roth BL, Armbruster BN, Ernsberger P, Irwin JJ, Shoichet BK. Relating protein pharmacology by ligand chemistry. *Nat Biotechnol*. 2007;25:197-206.
14. Campillos M, Kuhn M, Gavin A-C, Jensen LJ, Bork P. Drug target identification using side-effect similarity. *Science*. 2008;321:263-266.
15. Huang L-C, Wu X, Chen JY. Predicting adverse side effects of drugs. *BMC Genomics*. 2011;12:S11.
16. Li T, Kusko R, Thakkar S, Liu Z, Tong W. Artificial intelligence, machine learning, and deep learning in precision medicine in liver diseases. *Sect 3: Artif Intell Appl Specif Dis Liver*. 2023;233-251. doi:10.1016/b978-0-323-99136-0.00012-x
17. Song T, Wang G, Ding M, Rodriguez-Paton A, Wang X, Wang S. Network-based approaches for drug repositioning. *Mol Inform*. 2022;41.
18. Lin A, Giuliano CJ, Palladino A, et al. Off-target toxicity is a common mechanism of action of cancer drugs undergoing clinical trials. *Sci Transl Med*. 2019;11.
19. Lampa S, Alvarsson J, Arvidsson Mc Shane S, Berg A, Ahlberg E, Spjuth O. Predicting off-target binding profiles with confidence using conformal prediction. *Front Pharmacol*. 2018;9:1256.
20. Zhang C, Hong H, Mendrick DL, Tang Y, Cheng F. Biomarker-based drug safety assessment in the age of systems pharmacology: from foundational to regulatory science. *Biomark Med*. 2015;9:1241-1252.
21. Wilson JL, Racz R, Liu T, et al. PathFX provides mechanistic insights into drug efficacy and safety for regulatory review and therapeutic development. *PLoS Comput Biol*. 2018;14:e1006614.
22. Wilson JL, Gravina A, Grimes K. From random to predictive: a context-specific interaction framework improves selection of drug protein-protein interactions for unknown drug pathways. *Integr Biol: Quant Biosci Nano Macro*. 2021;14:13-24.
23. Wang X, Zhu H, Jiang Y, et al. PRODeepSyn: predicting anticancer synergistic drug combinations by embedding cell lines with protein–protein interaction network. *Brief Bioinform*. 2022;23:bbab587.
24. Meng X, Li W, Xiang J, et al. Temporal-spatial analysis of the essentiality of hub proteins in protein-protein interaction networks. *IEEE Trans Netw Sci Eng*. 2022;9:3504-3514.
25. Chen Y-W, Diamante G, Ding J, et al. PharmOmics: a species- and tissue-specific drug signature database and gene-network-based drug repositioning tool. *iScience*. 2022;25:104052.
26. Harpaz R, Odgers D, Gaskin G, et al. A time-indexed reference standard of adverse drug reactions. *Sci Data*. 2014;1:140043.
27. Ruiz C, Zitnik M, Leskovec J. Identification of disease treatment mechanisms through the multiscale interactome. *Nat Commun*. 2021;12:1796.
28. Alidoost M, Ghodrati V, Ahmadian A, et al. Model utility of a deep learning-based segmentation is not dice coefficient dependent: a case study in volumetric brain blood vessel segmentation. *Intell-Based Med*. 2023;7:100092.
29. Jung K, Kashyap S, Avati A, et al. A framework for making predictive models useful in practice. *J Am Méd Inform Assoc*. 2020;28:1149-1158.
30. Wilson JL, Steinberg E, Racz R, et al. A network paradigm predicts drug synergistic effects using downstream protein–protein interactions. *CPT Pharmacometrics Syst Pharmacol*. 2022;11:1527-1538.
31. Wishart DS, Knox C, Guo AC, et al. DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res*. 2006;34:D668-D672.
32. Wilson JL, Wong M, Stepanov N, Petkovic D, Altman R. PhenClust, a standalone tool for identifying trends within sets of biological phenotypes using semantic similarity and the unified medical language system metathesaurus. *JAMIA Open*. 2021;4:ooab079.
33. Landrum MJ, Lee JM, Riley GR, et al. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res*. 2014;42:D980-D985.
34. Amberger J, Bocchini CA, Scott AF, Hamosh A. McKusick's online Mendelian inheritance in man (OMIM®). *Nucleic Acids Res*. 2009;37:D793-D796.
35. Piñero J, Queralt-Rosinach N, Bravo A, et al. DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes. *J Biol Databases Curation*. 2015;2015:bav028.
36. Piñero J, Bravo À, Queralt-Rosinach N, et al. DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Res*. 2017;45:D833-D839.
37. Ramos EM, Hoffman D, Junkins HA, et al. Phenotype–genotype integrator (PheGenI): synthesizing genome-wide association study (GWAS) data with existing genomic resources. *Eur J Hum Genet*. 2014;22:144-147.
38. The pyjarowinkler module in Python. https://pypi.org/project/pyjarowinkler/
39. The difflib module in Python. https://docs.python.org/3/library/difflib.html
40. Gene Ontology enRIchment anaLysis and visuaLizAtion tool (GOrilla). https://cbl-gorilla.cs.technion.ac.il/
41. Wilson JL, Wong M, Chalke A, Stepanov N, Petkovic D, Altman RB. PathFXweb: a web application for identifying drug safety and efficacy phenotypes. *Bioinformatics*. 2019;35:4504-4506.
42. Higashi Y, Kihara Y, Noma K. Endothelial dysfunction and hypertension in aging. *Hypertens Res*. 2012;35:1039-1047.
43. Gallo G, Volpe M, Savoia C. Endothelial dysfunction in hypertension: current concepts and clinical implications. *Front Med*. 2022;8:798958.
44. Sun H-J, Wu Z-Y, Nie X-W, Bian J-S. Role of endothelial dysfunction in cardiovascular diseases: the link between inflammation and hydrogen sulfide. *Front Pharmacol*. 2020;10:1568.
45. Menche J, Sharma A, Kitsak M, et al. Uncovering disease-disease relationships through the incomplete interactome. *Science*. 2015;347:1257601.

46. Liang H, Chen L, Zhao X, Zhang X. Prediction of drug side effects with a refined negative sample selection strategy. *Comput Math Methods Med*. 2020;2020:1573543.
47. Galeano D, Paccanaro A. Machine learning prediction of side effects for drugs in clinical trials. *Cell Rep Methods*. 2022;2:100358.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Alidoost M, Wilson JL. Preclinical side effect prediction through pathway engineering of protein interaction network models. *CPT Pharmacometrics Syst Pharmacol*. 2024;00:1-21. doi:10.1002/psp4.13150