PAPER

A precise comparison of molecular target prediction methods

Tiantian He, Klaudia Caba and Pedro J. Ballester [ID] *

Small-molecule drug discovery has transitioned from traditional phenotypic screening to more precise target-based approaches, with an increased focus on understanding mechanisms of action (MoA) and target identification. With more research on off-target effects of approved drugs and the discovery of new therapeutic targets, revealing hidden polypharmacology can reduce both time and costs in drug discovery through off-target drug repurposing. However, despite the potential of *in silico* target prediction, its reliability and consistency remain a challenge across different methods. This project systematically compares seven target prediction methods, including stand-alone codes and web servers (MolTarPred, PPB2, RF-QSAR, TargetNet, ChEMBL, CMTNN and SuperPred), using a shared benchmark dataset of FDA-approved drugs. We also explore model optimization strategies, such as high-confidence filtering, which reduces recall, making it less ideal for drug repurposing. Furthermore, for MolTarPred, Morgan fingerprints with Tanimoto scores outperform MACCS fingerprints with Dice scores. This analysis shows that MolTarPred is the most effective method. For practical applications, we introduce a programmatic pipeline for target prediction and MoA hypothesis generation. A case study on fenofibric acid shows its potential for drug repurposing as a THRB modulator for thyroid cancer treatment.

Introduction

Over 90% of global pharmaceuticals are small-molecule drugs[1] for a wide range of diseases, from infectious diseases to cancer, due to their stability, accessibility, and cost-effectiveness.[2] From natural extracts in the early century, phenotypic screening was dominant in drug discovery.[3] In the last three decades, target-based approaches were developed with advances in molecular biology. While target-based screening offers efficiency, the complexity of biological systems has led to a resurgence of phenotypic methods as a complementary approach[4] and increasing research on polypharmacology.[5] Targeting multiple disease-related pathways, polypharmacology aims to address limitations of single-target strategies in treating complex diseases, such as Alzheimer's or cancer, and has also emerged as a promising solution to overcome drug tolerance, understand off-target effects and facilitate drug repurposing.[6,7] Off-target effects result in side effects and new indications. For example, nonsteroidal anti-inflammatory drugs (NSAIDs) primarily target cyclooxygenase (COX) enzymes to alleviate pain and inflammation, but they can also cause gastrointestinal damage due to COX-1 inhibition.[8,9] For positive outcomes of off-target effects, Gleevec and Viagra, originally developed for leukemia and hypertension, were repurposed to treat gastrointestinal stromal tumors[10] and erectile dysfunction,[11]

Department of Bioengineering, Imperial College London, London, UK. E-mail: p.ballester@imperial.ac.uk

respectively. These cases show the potential of polypharmacology to repurpose existing drugs, saving resources and reducing the need for extensive safety and pharmacokinetic testing.[12]

Precise identification and validation of drug–target interactions are essential. Experimental methods such as binding affinity assays, gene expression analyses, and proteomics are reliable,[13] but labour-intensive and complex. Recent advances in high-throughput techniques have significantly improved the efficiency of traditional wet-lab methods.[14] Moreover, *in silico* target fishing, including target-centric and ligand-centric approaches, can enhance the efficiency and accuracy of target prediction.

Target-centric methods build predictive models for each target to estimate whether a query molecule is likely to interact with those targets. They often use Quantitative Structure–Activity Relationship (QSAR) models built with various machine learning algorithms, such as random forest and the Naïve Bayes classifier. Some target-centric methods use molecular docking simulations based on 3D protein structures. For example, ponatinib, an FDA-approved tyrosine kinase inhibitor for leukemia, was repurposed as a PD-L1 inhibitor. After molecular docking and virtual screening using the ZINC database, *in vitro* experiments confirmed its binding to PD-L1, and *in vivo* studies demonstrated that ponatinib delayed tumor growth in mice, outperforming conventional anti-PD-L1 antibodies.[15] However, the application is limited by the availability of bioactivity data to train the QSAR models and accurate 3D protein structures.[16,17] Recent advances in cryo-electron microscopy[18] and
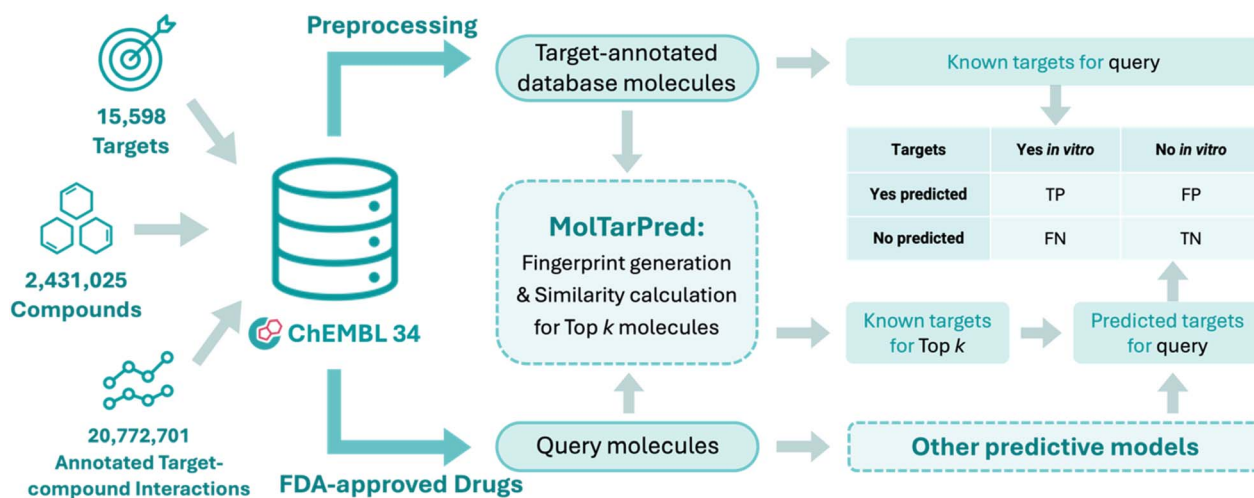
Fig. 1 Workflow of target fishing methods. Using MolTarPred, the most similar database molecules to the query are identified based on similarity scores derived from chemical structure fingerprints. Known targets for the query molecule are retrieved from annotated target–compound interactions with *in vitro* validation using the ChEMBL database. In other methods, prediction results crawled from websites or obtained from training models are compared with the known targets of the query and calculated to evaluate the predictive performance across all methods.

computational tools, such as AlphaFold,[19,20] have expanded the target coverage for protein structures. With these computational tools, high-quality structural models can be generated from amino acid sequences even without experimental determination, yet many protein targets still lack high-resolution ligand-bound structures.[21] There is also the issue that many structure-based scoring functions have low predictive accuracy, although this has been improving with the application of machine learning.[22] Even non-structure-based models based on sequences or features heavily rely on quality and comprehensiveness of existing datasets for training and validation.[23]

Ligand-centric methods, on the other hand, focus on the similarity between the query molecule and a large set of known molecules annotated with their targets.[24] Their effectiveness depends on the knowledge of known ligands. With data on proved interaction, several small-molecule drugs have been successfully repurposed. For example, MolTarPred discovered hMAPK14 as a potent target of mebendazole which was further proved by *in vitro* validation.[25,26] MolTarPred also predicted Carbonic Anhydrase II (CAII) as a new target of Actarit, suggesting

potential for repurposing this rheumatoid arthritis drug for conditions such as hypertension, epilepsy, and certain cancers.[27] Not all similarity-based methods are ligand-centric. For instance, the TAMOSIC method is also based on similarity but cannot be simply classified as either ligand-centric or target-centric because it learns the optimal similarity threshold for each target.[28]

In this study, we evaluate several target prediction methods and compare their performance on a shared dataset to identify optimal computational models for small-molecule drug repositioning (Fig. 1). Both available stand-alone codes and web servers are considered (see Table 1). Among them, MolTarPred is a primary focus of this study, and we further explore how model components such as fingerprints and similarity metrics influence its prediction accuracy and discuss the implications for model optimization.

Experimental setup

Database selection

ChEMBL, PubChem, DrugBank and BindingDB are widely used databases of bioactive molecules, including chemical

Table 1 Publicly available methods for target prediction employed in this research

Methods		Source	Database	Algorithm	Fingerprints	Top similar ligand
Target-centric	RF-QSAR[32]	Web server	ChEMBL 20&21	Random forest	ECFP4	Top 4, 7, 11, 33, 66, 88 and 110
	TargetNet[33]		BindingDB	Naïve Bayes	FP2, Daylight-like, MACCS, E-state and ECFP2/4/6	Unclear
	ChEMBL[34]		ChEMBL 24	Random forest	Morgan	Unclear
	CMTNN[35]	Stand-alone	ChEMBL 34	ONNX runtime	Morgan	Unclear
Ligand-centric	MolTarPred[36]	code	ChEMBL 20	2D similarity	MACCS	Top 1, 5, 10 and 15
	PPB2 (ref. 37)	Web server	ChEMBL 22	Nearest neighbor–Naïve Bayes/deep neural network	MQN, Xfp and ECFP4	Top 2000
	SuperPred[38]		ChEMBL and BindingDB	2D/fragment/3D similarity	ECFP4	Unclear

structures, biological activities and ligand–target interactions.[29] In this research, ChEMBL was selected for its extensive and experimentally validated bioactivity data, including drug–target interactions, inhibitory concentrations, and binding affinities.[30] While DrugBank is ideal for predicting new drug indications against known targets due to its focus on drug-related information, ChEMBL is more suitable for novel protein targets because of its extensive chemogenomic data.[24]

Therefore, we used ChEMBL version 34, the most recent publicly available release, containing 15 598 targets, 2 431 025 compounds, and 20 772 701 interactions.[30] We hosted the PostgreSQL version of the ChEMBL 34 database locally and retrieved data from the molecule_dictionary and target_dictionary tables, including unique ChEMBL IDs for both compounds and targets, bioactivity interaction and canonical SMILES strings, by connecting *via* pgAdmin4 software.

Database preparation

The experimental bioactivity data provide binding affinity between targets and compounds. We retrieved data from the local PostgreSQL ChEMBL 34 database by querying the molecule_dictionary, target_dictionary, and activities tables and selected bioactivity records with standard values for IC50, Ki, or EC50 below 10 000 nM. The ChEMBL IDs and preferred names of each target were also extracted for further analysis.

To avoid confusion and ensure data quality, entries associated with non-specific or multi-protein targets were excluded by filtering out targets whose names contained keywords such as "multiple" or "complex." To prevent redundancy, duplicate compound–target pairs were removed, retaining only unique pairs. Finally, a total of 1 150 487 unique ligand–target interactions were retained.

To simplify the analysis and avoid redundant predictions, we consolidated data for a single ligand across multiple targets into one row, with the target IDs separated by colons. Finally, the ChEMBL IDs, canonical SMILES strings, and annotated targets were exported to a CSV file for further prediction and validation.

To enhance data quality for subsequent analyses, a filtered database was employed, only containing highly confident interactions with a minimum confidence score of 7. In the ChEMBL database,[30] the confidence score is defined from 0 (target unknown or has yet to be assigned) to 9 (direct single protein target assigned). A score of 7 means direct protein complex subunits assigned, which ensures that only well-validated interactions are included in the analysis. Additionally, a potentially improved fingerprint for similarity calculations, the Morgan hashed bit vector fingerprint with radius two and 2048 bits (Morgan), was tested for optimization effects.[31]

Dataset preparation for benchmark

We collected molecules with FDA approval years to prepare a benchmark dataset of FDA-approved drugs from the whole ChEMBL database. To ensure that the performance was not biased or overestimated, these molecules were excluded from the main database to prevent any overlap with any known drugs during prediction.

We randomly selected 100 samples from the FDA-approved drugs dataset to validate prediction methods. We removed the "CHEMBL" prefix from ChEMBL IDs and organized data into separate files: one containing the 100 random samples as query molecules and another containing the remaining molecules in the database to identify potential drug–target interaction candidates for these queries.

Target prediction *via* various methods

We considered seven target prediction methods: MolTarPred with published codes,[36] Polypharmacology Browser2 (PPB2),[37] RF-QSAR,[32] TargetNet,[33] ChEMBL,[34] ChEMBL Multitask Neural Network (abbreviated as CMTNN),[35] and SuperPred.[38] Among these, MolTarPred and CMTNN are run locally with stand-alone codes, while the others are web servers that require manual querying. To automate the extraction of data from these web servers, we implemented web crawlers respectively. These crawlers navigate the webpages, simulate form submissions and retrieve relevant target prediction results from responses. For PPB2, we used the "NN(ECfp4) + NB(ECfp4)" configuration favoured by the PPB2 webserver. This setting corresponds to using Extended Connectivity Fingerprint of radius 2 (ECFP4) with a combination of nearest neighbor (NN) similarity search and Naive Bayes (NB) classification. The Tanimoto coefficient was applied for similarity scoring.

The prediction result of drug–target interaction is binary: the model predicts that a drug will bind to a particular target (positive) or that there is no expected interaction between the drug and target (negative). MolTarPred uses ChEMBL data based on experimental evidence to determine interactions. Other methods, such as PPB2, RF-QSAR, SuperPred, and TargetNet, rank predicted targets by measuring similarity scores between the query molecule and known ligands of a target or the probabilities of a target interacting with the query.

For instance, SuperPred differentiates between positive and negative drug–target interactions using the Tanimoto score based on molecular fragment comparison, with a threshold of 0.45. The CMTNN generates probability scores ranging from 0 to 1 indicating the likelihood of targets interacting with the query molecules. A common threshold of 0.5 is used for classification, where scores above 0.5 indicate positive drug–target interaction and scores below 0.5 indicate negative drug–target interaction. For the ChEMBL prediction method, predictions were filtered to only retain targets consistently classified as positive (active in this context) across three confidence levels: 70%, 80%, and 90%, representing different degrees of the model's expected prediction certainty. These retained predictions were then ranked using confidence scores, with higher confidence predictions prioritized in the final analysis. To ensure consistency across methods, the Top 5 and Top 10 predictions will be chosen to be considered as positive predictions in the next step.

Predictions with low confidence were not uniformly interpreted across methods and were not treated as definitive negatives. For example, the ChEMBL model does not define any confidence thresholds for non-binding; low-confidence outputs are instead considered uncertain. Similarly, RF-QSAR and

TargetNet rank likely binders using confidence scores, but do not explicitly classify low-ranking targets as inactive. Only the CMTNN provides a clear probabilistic threshold, where values below 0.5 indicate predicted non-binding. Due to these differences in handling low-confidence predictions, our evaluation focused on reliable binding predictions and did not attempt to define a consistent threshold for negatives across methods.

Prediction method validation

Model validation varies across methods and is categorized into two groups: virtual screening and target prediction. Virtual screening methods, such as SuperPred, RF-QSAR, and Target-Net, are evaluated based on metrics per target, assessing how well a method can rank or identify potential targets for a given molecule across a large set of targets. In contrast, target prediction methods, including MolTarPred, PPB2, ChEMBL and CMTNN, are evaluated based on metrics per query molecule, focusing on how well the method predicts specific drug–target interactions for each query molecule. Such ligand-centric assessment provides a much more precise evaluation of target prediction performance, as hit rates and how many true targets are missed are calculated per molecule.[36] Therefore, our work focused on comparing methods from a target prediction perspective, evaluating how well each method predicts drug–target interactions on a per-molecule basis. We evaluated their performance on a common standard to identify the most effective target prediction method for comprehensively predicting more known targets and discovering potential new targets.

To evaluate the performance of the tested methods, we classified them as true positives (TPs), false positives (FPs), false negatives (FNs), or true negatives (TNs) by comparing predicted drug–target interactions against experimental evidence (Table 2).

True Positives (TPs): correctly identified targets of a query molecule verified experimentally.

False Positives (FPs): predicted targets that are not considered identified targets for a given query molecule. FPs can be categorized into two groups: confirmed FPs are targets which have been tested *in vitro* with the query molecule and show negative results, proving that they are not true targets. Unconfirmed FPs are targets that have not yet been tested *in vitro* with the query molecule and, therefore, remain unannotated, which could potentially represent true targets if confirmed by future experiments.

False Negatives (FNs): targets incorrectly predicted to not interact with a given query molecule although experimental evidence suggests otherwise.

True Negatives (TNs): correctly identified non-interacting targets for a given query molecule.

Whether the target IDs were obtained from the top hits in the annotated database or from the top-ranked targets, they were extracted and compiled in a standardized format. The prediction results included the ligand ID, SMILES, and predicted target IDs. After categorizing the predicted targets as TPs, FPs, TNs, or FNs, various performance metrics, such as accuracy, recall, specificity, and precision, can be calculated for validation.

Accuracy is the proportion of correct target predictions:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \quad (1)$$

Recall, also known as sensitivity or the true positive rate, is the proportion of correctly predicted targets out of all the known targets:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (2)$$

Specificity, or the true negative rate, is the proportion of predicted non-interacting targets that are correctly identified:

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (3)$$

Precision measures the proportion of predicted targets that are experimentally confirmed:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (4)$$

As one of the most robust measures in binary classification, the Matthews Correlation Coefficient (MCC) is a balanced measure of prediction quality, particularly useful for imbalanced datasets, and regarded as a more reliable evaluation criterion.[36,39] In target prediction methods, the number of non-interacting targets far exceeds the number of true targets, leading to significant imbalance. The MCC takes all four metrics into consideration, reflecting the correlation between predictions and true values, making it a more appropriate measure of performance.[40]

$$\text{MCC} = \frac{\text{TP} \cdot \text{TN} - \text{FP} \cdot \text{FN}}{\sqrt{(\text{TP} + \text{FP})(\text{TP} + \text{FN})(\text{TN} + \text{FP})(\text{TN} + \text{FN})}} \quad (5)$$

MoA hypothesis generation pipeline

In this study, we first chose the optimal target prediction method, focusing on FPs and correcting misclassifications caused by targets with identical names but different target IDs. Next, we generated a MoA hypothesis to help understand the predicted drug–target interactions and disease pathways. Hetionet is a heterogeneous network that connects biomedical
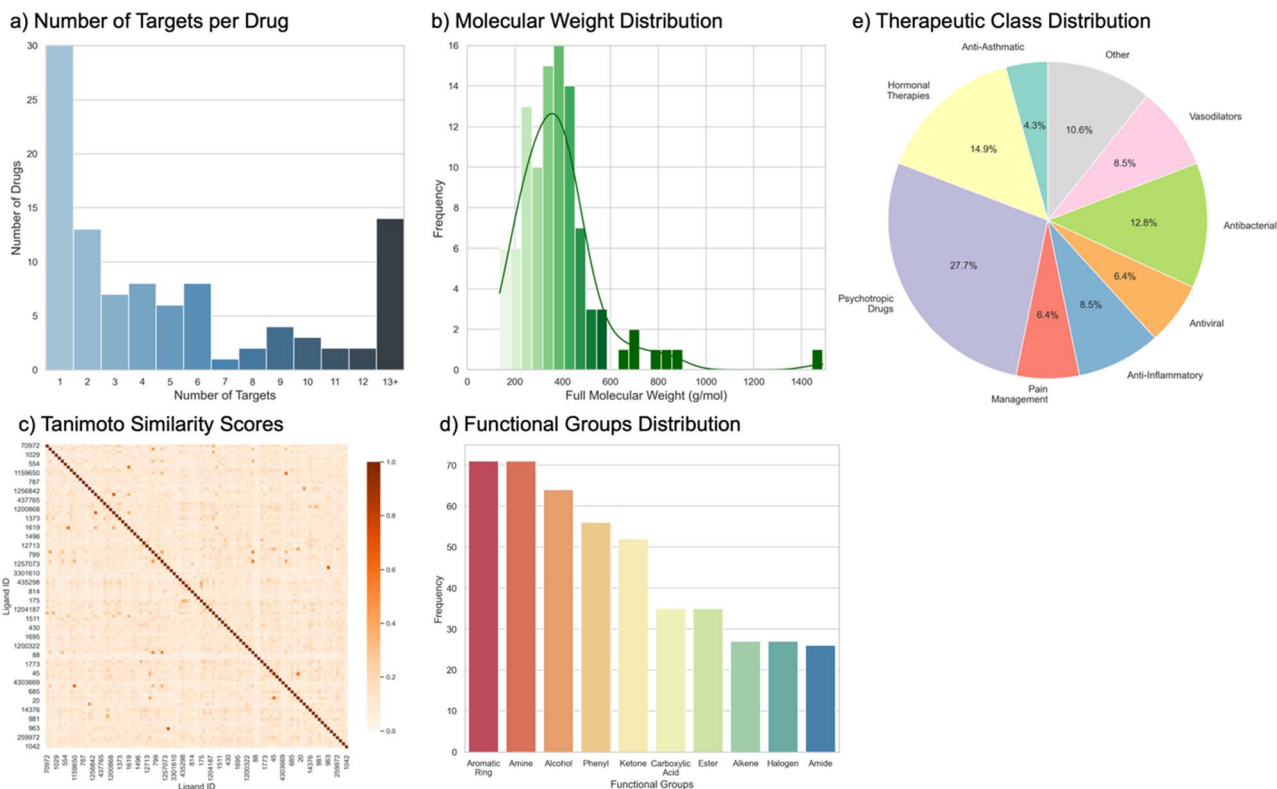
Table 2 Confusion matrix for a query molecule

Targets	Yes predicted	No predicted
Yes *in vitro*	TP	FN
No *in vitro*	FP[a]	TN

[a] FPs include predicted targets for the query molecule that have not been tested *in vitro*. Thus, we evaluated the worst-case scenario by assuming that all these are true FPs. However, these predictions could be true targets for drug repositioning after *in vitro* validation in the future.

Fig. 2 Diversity analysis of drug datasets from 100 randomly selected FDA-approved drugs. (a) Number of targets per drug, (b) molecular weight distribution, (c) pairwise Tanimoto similarity scores, (d) functional group distribution, and (e) therapeutic class distribution.

entities (such as genes, compounds, and diseases) from 29 databases. *Via* the Rephetio[41] platform (**https://het.io/repurpose/**), we investigate potential links between drugs, new targets, and related indications. Degree-Weighted Path Counts (DWPCs) were used to quantify the strength of these connections, showing the possibility of a predicted target playing a role in the MoA for a given disease. As a case study, we applied the pipeline to a small molecule drug and analysed these results with published studies to explore possible MoA hypotheses and discover new therapeutic applications for drug repurposing.

Results and discussion

Diversity of the drug dataset from 100 random samples

To evaluate the effectiveness of various prediction methods, we selected a set of 100 FDA-approved drugs from the ChEMBL database. These molecules were chosen randomly, aiming to ensure diversity and comprehensiveness of the dataset across key parameters, such as the number of targets per drug, molecular weight distribution, structural features, and therapeutic coverage (Fig. 2).
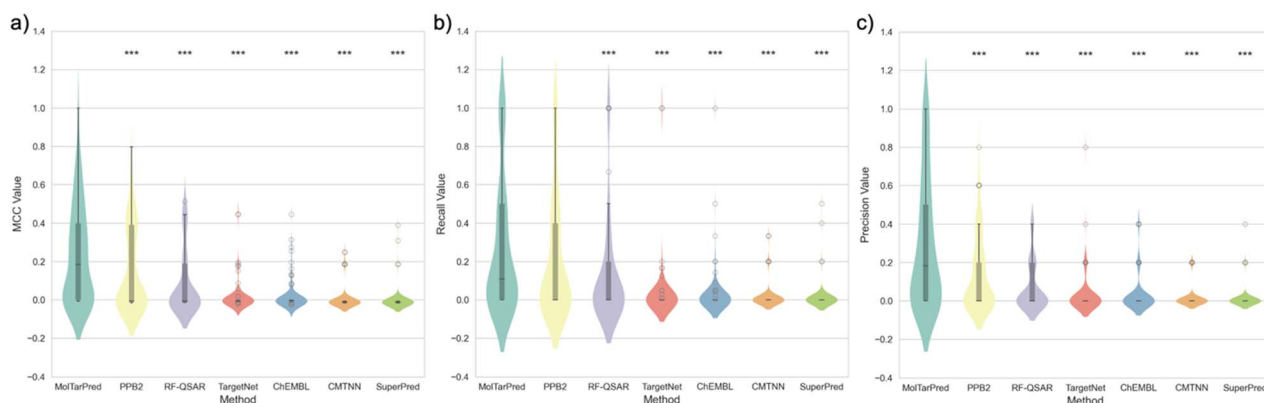

Fig. 3 Performance metrics across different methods on 100 query molecules (Top $k = 5$), (a) MCC, (b) recall, and (c) precision (*$0.01 <= p < 0.05$, **$0.001 <= p < 0.01$, and ***$p < 0.001$). Pairwise statistical significance calculated between MolTarPred and each of the other methods.

Table 3 Example of prediction results for a drug. Targets IDs are separated by ":" without the CHEMBL prefix. Information of target 1994: https://www.ebi.ac.uk/chembl/web_components/explore/target/CHEMBL1994

Drug	Finerenone (CHEMBL2181927)
Known targets	1994:2034:1871:208
MolTarPred	**208**:220:239:247:364:378:**1871:1994:2034**:2459:3979:4158
PPB2	**1994:1871**:5023:2564:**208**:279:4302:2276:4792:1936
RF-QSAR	5785:3286:3032:2148:**1994:1871**:244:220:**208**:202
TargetNet	4282:2219:252:4427:302:226:4566:4068:4552:4616
ChEMBL	4086:2073:4601:2283:4828:2966:2781:3321:1849:1929
CMTNN	3815:5023:2 146 302:2525:3922:2781:2083:5263:2954:6164
SuperPred	3251:1 293 237:4203:5409:2535:3 137 262:3060:4793:241:1 293 249

The presence of both single-target and polypharmacological drugs with varying numbers of targets in the dataset tests whether the models' predictive ability is influenced by the number of known targets. Meanwhile, the molecular weights varying from small molecules ($\sim$200 g mol^{-1}) to larger compounds ($\sim$1500 g mol^{-1}) ensure that both small and large molecules are well-represented in this dataset. The pairwise Tanimoto similarity score visualized in a heatmap (Fig. 2c) presents low similarity between molecules and their high chemical diversity. The variety of functional groups further proves this, and distinct therapeutic classes offer an opportunity for a wide range of therapeutic indication predictions. This diversity ensures that the prediction methods are tested on an unbiased and representative benchmark.

Comparative performance of prediction methods in Top 5

Using a dataset of 100 random samples from FDA-approved drugs as query molecules, we evaluated the performance of seven target prediction methods: MolTarPred, PPB2, RF-QSAR, TargetNet, ChEMBL, CMTNN and SuperPred. The definition of "Top k" varies across methods, with MolTarPred referring to the number of most similar ligands identified for each query molecule. Predictions are made from targets associated with these top similar ligands. For other methods, "Top k" generally refers to the top-ranked targets.

We select "Top 5" and "Top 10" as evaluation standards based on their promising results balanced between sensitivity and precision, according to previous study.[36] "Top 5" minimizes

false positives, making it ideal for precision, while "Top 10" provides broader target coverage, useful for identifying potential off-target effects but with a slightly higher risk of false positives. For each method, we computed the number of predicted targets (NPTs), True Positives (TPs), True Negatives (TNs), False Positives (FPs), and False Negatives (FNs), and calculated performance metrics, including recall, precision, and MCC for each query molecule.

The violin plots (Fig. 3) illustrate the distribution of main metrics across the methods for the 100 query molecules under the Top 5 predictions condition. Using the Kruskal–Wallis test, a non-parametric statistical significance test that accounts for non-normal distribution and outliers, we compared the performance of MolTarPred with that of each of the other methods. MolTarPred was identified as the optimal method, evidenced by its highest median precision, recall, and MCC score, indicating its promising performance in target prediction.

Comparative performance of prediction methods in Top 10

When evaluating the performance under the Top 10 condition (Table 3), the violin plots (Fig. 4) reveal that MolTarPred continues its notable performance. The recall significantly increases, indicating that expanding prediction results to the Top 10 enhances the breadth of target identification. Despite a slight decrease in accuracy and precision, which can be expected, MolTarPred's performance still surpasses that of the other methods according to the statistical significance tests between MolTarPred and each of the other methods.
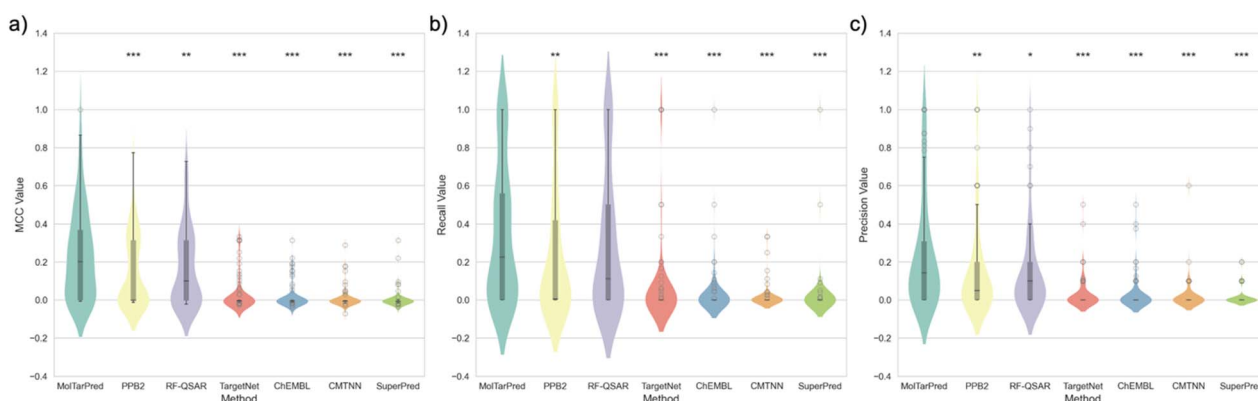

Fig. 4 Performance metrics across different methods on 100 query molecules (Top k = 10), (a) MCC, (b) recall, and (c) precision (*$0.01 <= p < 0.05$, **$0.001 <= p < 0.01$, and ***$p < 0.001$). Pairwise statistical significance was calculated between MolTarPred and each of the other methods.

Compared to the metrics for Top 5, an increase in recall is observed in PPB2 and RF-QSAR, while the CMTNN shows an improvement in precision. However, compared to the Top 5, MolTarPred's lower precision and recall in the Top 10 suggest that the model may require further refinement. Optimizing the model or integrating it with other methods could potentially enhance its performance.

Evaluation with highly confident target–ligand interactions

To assess whether filtering for high-confidence interactions enhances MolTarPred's performance, we applied a new filtered database containing only drug–target interactions with a confidence score of at least 7, based on ChEMBL 34. This high-confidence filter was used to test whether MolTarPred's performance would improve with more reliable interaction data. In the post-filtered model, metrics including the MCC, precision, recall and NPT decreased, indicating a more conservative prediction model (Fig. 5). The high-confidence filter led to a marginal increase in accuracy but common declines in the NPT, precision, recall, and MCC due to the loss of true targets. As a result, this conservative model is not suitable for drug repositioning or polypharmacology research. For such purposes, identifying a wide range of potential targets, including those with lower confidence scores, might still be meaningful.

Fig. 5 Performance metrics on 100 query molecules before (pink) and after filtering for high-confidence interactions (blue).

Evaluation with different fingerprints and similarity scores

We assessed the impact of different fingerprints on MolTarPred's performance by comparing Morgan hashed bit vector fingerprints with a radius of two and 2048 bits (Morgan)[31] to MACCS fingerprints, using both the Dice score and Tanimoto score. These fingerprints represent contrasting approaches: Morgan offers a more detailed, comprehensive representation of molecular structures, while MACCS provides a simpler, more generalized view. Using a dataset of 100 randomly selected query molecules, the model considered the Top 10 nearest neighbors ($k = 10$) from a pre-filtered database.

As we can see in Fig. 6, Morgan generally outperformed MACCS across several performance metrics. Under the Dice score, Morgan showed a higher MCC and precision, though it had a slight reduction in recall. The average number of predicted targets (NPTs) for Morgan was 30% lower than that of MACCS, indicating that while Morgan predicted fewer targets,
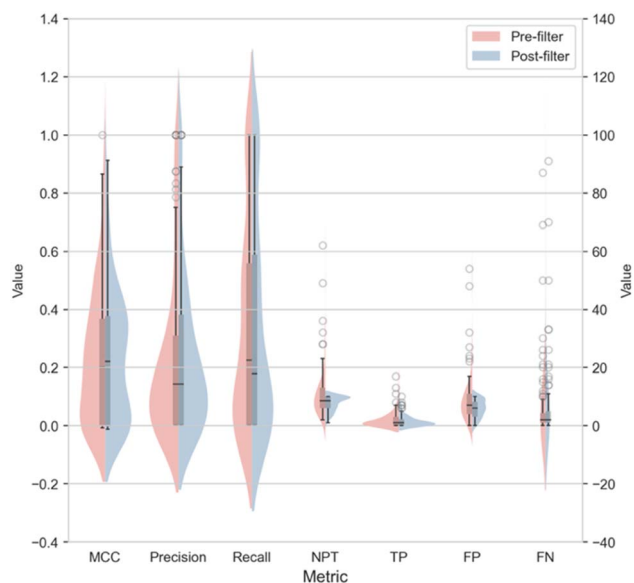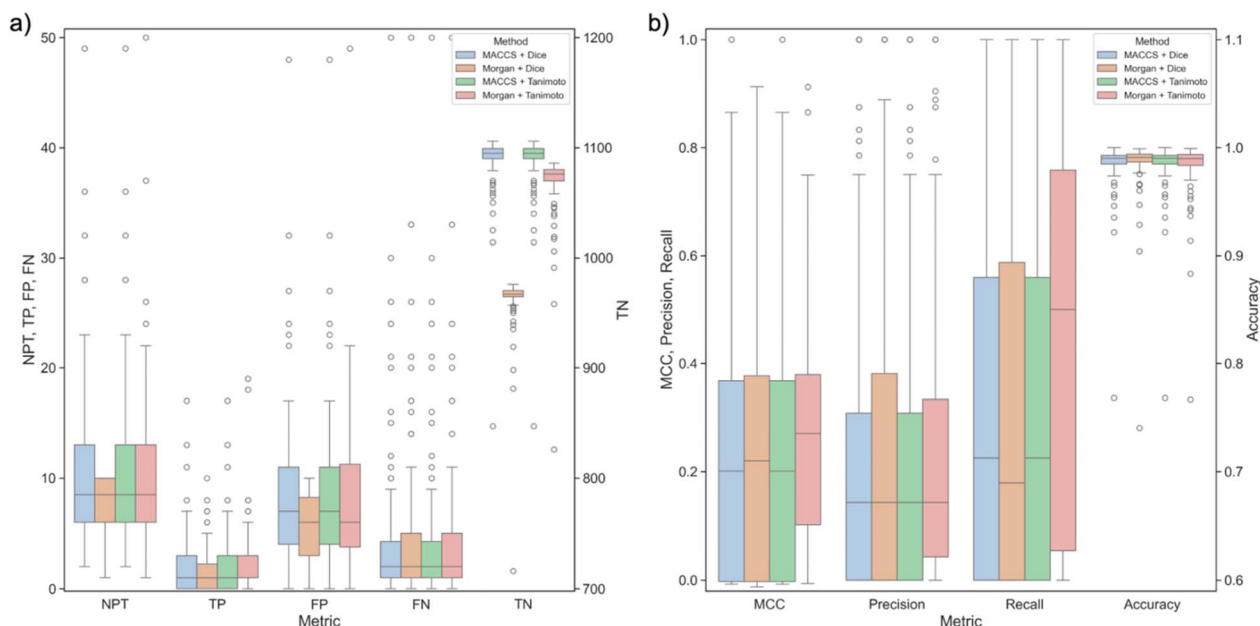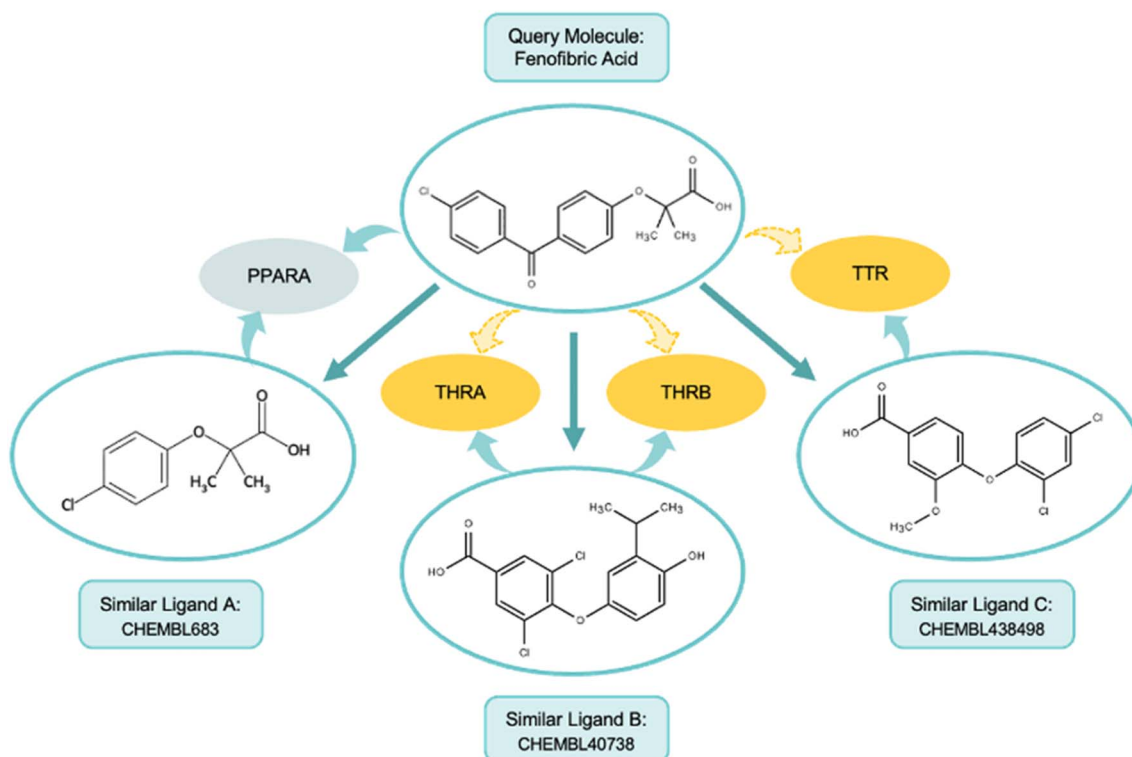

Fig. 6 Box plots of performance on 100 query molecules using different fingerprints and similarity scores ((a): target counts and (b): metrics).

Paper

Digital Discovery

these predictions were more reliable. This trend reflects Morgan's higher specificity, focusing on a narrower range of true interactions with improved overall prediction reliability.

Using the Tanimoto score makes no significant difference in MACCS's performance. In contrast, Morgan's performance improved, showing an increased NPT, higher than both its Dice
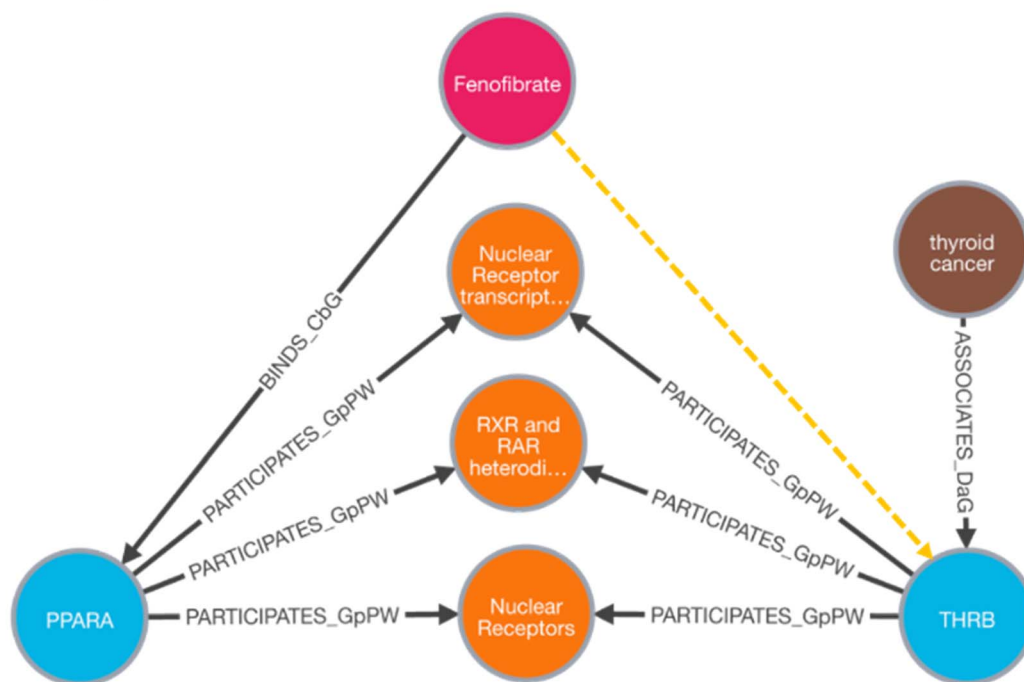
Fig. 7 An example of the MoA hypothesis generation pipeline (a) targets from similar ligands predicted by MolTarPred (b) MoA hypothesis of fenofibric acid in treating thyroid cancer *via* THRB generated by Rephetio. The analysis is based on network relationships, where GpPW refers to gene-participates-pathway, CbG indicates compound-binds-gene, and DaG denotes disease-associates-gene.

score result and MACCS result. Accuracy is a flattering metric, with a median of around 0.99 while the median MCC is around 0.2. With the Tanimoto score, Morgan achieved the highest MCC and recall among all evaluated methods, indicating an enhanced ability to identify a broader range of true targets. Also, the Tanimoto score with Morgan results in more unconfirmed FPs, beneficial for identifying a broader range of new potential targets.

The consistent performance of MACCS fingerprints across both Dice and Tanimoto scores is attributed to their simple structure feature representation, making them less sensitive to different similarity measures.[42] In contrast, Morgan fingerprints, which capture more detailed structural information, particularly from atoms' local environments, performed better with the Tanimoto score. The Dice score is more sensitive to smaller overlaps and less effective with Morgan's dense representation, while the Tanimoto score normalizes the similarity score based on the total number of bits set to 1 by subtracting the overlapping portion between two fingerprints to avoid double-counting common bits, proving to be better suited for Morgan's complexity.[43]

Overall, while MACCS is robust to different similarity metrics due to its simplicity, Morgan fingerprints are more sensitive and perform better under the Tanimoto score, especially for identifying a wider range of true drug–target interactions.

Case study

To balance reliability and breadth of predictions, we selected MolTarPred with Morgan fingerprints using the Tanimoto score, according to the findings of the previous analysis. This approach was chosen due to MolTarPred's high MCC score, which shows its effectiveness and reliability in identifying more promising new targets among those categorized as FPs that have not been experimentally validated. In some cases, TPs were initially misclassified as FPs due to having different target IDs while sharing the same name as a known target. These targets were discarded to avoid the misleading outcomes in the discovery of new targets and indications.

We applied a MoA hypothesis generation pipeline to explore the predicted drug–target interactions and disease pathways. Rephetio leverages Hetionet, a heterogeneous network, to establish connections between drugs, targets, and diseases. By applying advanced algorithms to Hetionet's data, Rephetio evaluates the strength and relevance of these connections, supporting MoA hypothesis generation, drug repurposing and the identification of new therapeutic applications. Fig. 7 shows an example.

Fenofibric acid, a fibric acid derivative and the active metabolite of fenofibrate (a prodrug), is primarily used to reduce triglyceride levels and increase HDL cholesterol levels by activating peroxisome proliferator-activated receptor alpha (PPARA), a key regulator in lipid metabolism.[44] Fenofibrate is converted into Fenofibric acid in the body, which then exerts the therapeutic effects. Among the Top 10 similar ligands predicted by MolTarPred, some share the same known target, PPARA, while others suggest new potential targets, such as THRB (Fig. 7a). Thyroid Hormone Receptor β (THRB) is also

a member of the nuclear receptor superfamily as is Peroxisome Proliferator-Activated Receptor α (PPARA). They both function as ligand-activated transcription factors regulating metabolism.[45] THRB is a key regulator in thyroid hormone signaling. In thyroid cancer, mutations or loss of function in THRB are often observed, leading to uncontrolled cell proliferation and metastasis.[46] To investigate the potential mechanism of action (MoA) for fenofibric acid targeting thyroid hormone receptor beta (THRB), we utilized the Rephetio platform to generate hypotheses (Fig. 7b). The results show that fenofibrate binds the known target PPARA and predicted target THRB (compound-binds-gene, CbG), which participate in the gene-participates-pathway (GpPW) involving nuclear receptor transcription, RXR and RAR heterodimers, and nuclear receptor signaling. THRB is linked to thyroid cancer through the Disease-associates-Gene (DaG) relationship. If the drug molecule targets THRB in a similar manner to PPARA agonists, restoring or mimicking the normal function of THRB, it could potentially reduce tumorigenic properties and inhibit metastasis, especially in tumor types with THRB gene loss or mutations, such as thyroid cancer.

Conclusions

In this project, we evaluated seven target prediction methods, including MolTarPred, PPB2, RF-QSAR, TargetNet, ChEMBL, CMTNN and SuperPred, using FDA-approved drugs as query molecules. MolTarPred consistently outperformed the other methods in terms of median precision, recall, and MCC for both Top 5 and Top 10 predictions. Filtering for high-confidence interactions significantly reduced precision and recall, making MolTarPred less suitable for uncovering broad polypharmacology. Morgan fingerprints, particularly when combined with the Tanimoto score, showed better performance compared to MACCS fingerprints. A case study of fenofibric acid showed its potential for drug repurposing as a THRB modulator for thyroid cancer.

Overall, we recommend MolTarPred for its effectiveness in target prediction and suggest further enhancing its performance by integrating it with other methods or optimizing the model. Consistent names for targets and comprehensive databases for drug–target interactions are essential for successful drug repurposing.

Data availability

The bioactivity data used for the comparative analysis and programmatic pipeline were retrieved from the ChEMBL database (version 34), available at **https://www.ebi.ac.uk/chembl/** (DOI: **https://doi.org/10.6019/chembl.database.34**). The code to reproduce this study is available on GitHub at **https://github.com/the614/Compare-target-prediction-methods/** (DOI: **https://doi.org/10.5281/zenodo.16102111**).

Author contributions

P. J. B. conceived the idea and designed the experiments. T. H. performed the experiments and wrote the manuscript with the assistance of K. C. and P. J. B. All authors analyzed the results.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

Notes and references

1 M. W. Y. Southey and M. Brunavs, Introduction to small molecule drug discovery and preclinical development, *Front. Drug Discovery*, 2023, **3**, 1314077.

2 F. D. Makurvet, Biologics *vs.* small molecules: Drug costs and patient access, *Med. Drug Discovery*, 2021, **9**, 100075.

3 S. Wang, Z. Wang, L. Fang, Y. Lv and G. Du, Advances of the Target-Based and Phenotypic Screenings and Strategies in Drug Discovery, *Int J Drug Discov. Pharmacol.*, 2022, **1**, 2.

4 G. E. Croston, The utility of target-based discovery, *Expert Opin. Drug Discovery*, 2017, **12**, 427–429.

5 A. Anighoro, J. Bajorath and G. Rastelli, Polypharmacology: Challenges and Opportunities in Drug Discovery, *J. Med. Chem.*, 2014, **57**, 7874–7887.

6 J. L. Medina-Franco, M. A. Giulianotti, G. S. Welmaker and R. A. Houghten, Shifting from the single to the multitarget paradigm in drug discovery, *Drug Discovery Today*, 2013, **18**, 495–501.

7 A. Kabir and A. Muth, Polypharmacology: The science of multi-targeting molecules, *Pharmacol. Res.*, 2022, **176**, 106055.

8 I. G. V. Gerriets, *Nonsteroidal Anti-Inflammatory Drugs (NSAIDs)*, StatPearls Publishing, Treasure Island (FL), 2023.

9 C. Sostres, C. J. Gargallo, M. T. Arroyo and A. Lanas, Adverse effects of non-steroidal anti-inflammatory drugs (NSAIDs, aspirin and coxibs) on upper gastrointestinal tract, *Best Pract. Res., Clin. Gastroenterol.*, 2010, **24**, 121–132.

10 A. Frolov, S. Chahwan, M. Ochs, J. P. Arnoletti, Z. Z. Pan, O. Favorova, J. Fletcher, M. von Mehren, B. Eisenberg and A. K. Godwin, Response markers and the molecular mechanisms of action of Gleevec in gastrointestinal stromal tumors, *Mol. Cancer Ther.*, 2003, **2**, 699–709.

11 R. F. DeBusk, C. J. Pepine, D. B. Glasser, A. Shpilsky, H. DeRiesthal and M. Sweeney, Efficacy and safety of sildenafil citrate in men with erectile dysfunction and stable coronary artery disease, *Am. J. Cardiol.*, 2004, **93**, 147–153.

12 K. Sachdev and M. K. Gupta, A comprehensive review of feature based methods for drug target interaction prediction, *J. Biomed. Inf.*, 2019, **93**, 103159.

13 A. Cichonska, J. Rousu and T. Aittokallio, Identification of drug candidates and repurposing opportunities through compound–target interaction networks, *Expert Opin. Drug Discovery*, 2015, **10**, 1333–1345.

14 B. K. Wagner and S. L. Schreiber, The Power of Sophisticated Phenotypic Screening and Modern Mechanism-of-Action Methods, *Cell Chem. Biol.*, 2016, **23**, 3–9.

15 A. Barnwal, S. Das and J. Bhattacharyya, Repurposing Ponatinib as a PD-L1 Inhibitor Revealed by Drug Repurposing Screening and Validation by *In Vitro* and *In Vivo* Experiments, *ACS Pharmacol. Transl. Sci.*, 2023, **6**, 281–289.

16 A. Ezzat, M. Wu, X.-L. Li and C.-K. Kwoh, Computational prediction of drug–target interactions using chemogenomic approaches: an empirical survey, *Briefings Bioinf.*, 2018, **20**, 1337–1357.

17 F. Shaikh, H. K. Tai, N. Desai and S. W. I. Siu, LigTMap: ligand and structure-based target identification and activity prediction for small molecular compounds, *J. Cheminf.*, 2021, **13**, 44.

18 A. V. Sadybekov and V. Katritch, Computational approaches streamlining drug discovery, *Nature*, 2023, **616**, 673–685.

19 J. Abramson, J. Adler, J. Dunger, R. Evans, T. Green, A. Pritzel, O. Ronneberger, L. Willmore, A. J. Ballard, J. Bambrick, S. W. Bodenstein, D. A. Evans, C.-C. Hung, M. O'Neill, D. Reiman, K. Tunyasuvunakool, Z. Wu, A. Žemgulytė, E. Arvaniti, C. Beattie, O. Bertolli, A. Bridgland, A. Cherepanov, M. Congreve, A. I. Cowen-Rivers, A. Cowie, M. Figurnov, F. B. Fuchs, H. Gladman, R. Jain, Y. A. Khan, C. M. R. Low, K. Perlin, A. Potapenko, P. Savy, S. Singh, A. Stecula, A. Thillaisundaram, C. Tong, S. Yakneen, E. D. Zhong, M. Zielinski, A. Žídek, V. Bapst, P. Kohli, M. Jaderberg, D. Hassabis and J. M. Jumper, Accurate structure prediction of biomolecular interactions with AlphaFold 3, *Nature*, 2024, **630**, 493–500.

20 J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli and D. Hassabis, Highly accurate protein structure prediction with AlphaFold, *Nature*, 2021, **596**, 583–589.

21 A. E. Wakefield, D. Kozakov and S. Vajda, Mapping the binding sites of challenging drug targets, *Curr. Opin. Struct. Biol.*, 2022, **75**, 102396.

22 V.-K. Tran-Nguyen, M. Junaid, S. Simeon and P. J. Ballester, A practical guide to machine-learning scoring for structure-based virtual screening, *Nat. Protoc.*, 2023, **18**, 3460–3511.

23 Y. Li, Z. Fan, J. Rao, Z. Chen, Q. Chu, M. Zheng and X. Li, An overview of recent advances and challenges in predicting compound-protein interaction (CPI), *Med. Rev.*, 2023, (3), 465–486.

24 S.-Q. Yang, Q. Ye, J.-J. Ding, M.-Z. Yin, A.-P. Lu, X. Chen, T.-J. Hou and D.-S. Cao, Current advances in ligand-based target prediction, *Wiley Interdiscip. Rev.:Comput. Mol. Sci.*, 2021, **11**, e1504.

25 A. Peón, H. Li, G. Ghislat, K.-S. Leung, M.-H. Wong, G. Lu and P. J. Ballester, *MolTarPred: A web tool for comprehensive target prediction with reliability estimation*, Chemical Biology & Drug Design, 2019, vol. 94, pp. 1390–1401.

26 J. Ariey-Bonnet, K. Carrasco, M. Le Grand, L. Hoffer, S. Betzi, M. Feracci, P. Tsvetkov, F. Devred, Y. Collette, X. Morelli, P. Ballester and E. Pasquier, In silico molecular target prediction unveils mebendazole as a potent MAPK14 inhibitor, *Mol. Oncol.*, 2020, **14**, 3083–3099.

27 G. Ghislat, T. Rahman and P. J. Ballester, Identification and Validation of Carbonic Anhydrase II as the First Target of the Anti-Inflammatory Drug Actarit, *Biomolecules*, 2020, **10**, 1570.

28 L. Wang, C. Ma, P. Wipf, H. Liu, W. Su and X. Q. Xie, TargetHunter: an *in silico* target identification tool for predicting therapeutic potential of small organic molecules based on chemogenomic database, *AAPS J.*, 2013, **15**, 395–406.

29 H. Wu, J. Liu, R. Zhang, Y. Lu, G. Cui, Z. Cui and Y. Ding, A review of deep learning methods for ligand based drug virtual screening, *Fundam. Res.*, 2024, **4**, 715–737.

30 B. Zdrazil, E. Felix, F. Hunter, E. J. Manners, J. Blackshaw, S. Corbett, M. de Veij, H. Ioannidis, D. M. Lopez, J. F. Mosquera, M. P. Magarinos, N. Bosc, R. Arcila, T. Kizilören, A. Gaulton, A. P. Bento, M. F. Adasme, P. Monecke, G. A. Landrum and A. R. Leach, The ChEMBL Database in 2023: a drug discovery platform spanning multiple bioactivity data types and time periods, *Nucleic Acids Res.*, 2024, **52**, D1180–D1192.

31 A. Peón, S. Naulaerts and P. J. Ballester, Predicting the Reliability of Drug–target Interaction Predictions with Maximum Coverage of Target Space, *Sci. Rep.*, 2017, **7**, 3820.

32 K. Lee, M. Lee and D. Kim, Utilizing random Forest QSAR models with optimized parameters for target identification and its application to target-fishing server, *BMC Bioinf.*, 2017, **18**, 567.

33 Z.-J. Yao, J. Dong, Y.-J. Che, M.-F. Zhu, M. Wen, N.-N. Wang, S. Wang, A.-P. Lu and D.-S. Cao, TargetNet: a web service for predicting potential drug–target interaction profiling *via* multi-target SAR models, *J. Comput.-Aided Mol. Des.*, 2016, **30**, 413–424.

34 N. Bosc, F. Atkinson, E. Felix, A. Gaulton, A. Hersey and A. R. Leach, Large scale comparison of QSAR and conformal prediction methods and their applications in drug discovery, *J. Cheminf.*, 2019, **11**, 4.

35 Eloy, *ChEMBL Multitask Nerual Network model*, GitHub repository, 2019.

36 A. Peón, C. C. Dang and P. J. Ballester, How Reliable Are Ligand-Centric Methods for Target Fishing?, *Front. Chem.*, 2016, **4**, 15.

37 M. Awale and J. L. Reymond, Polypharmacology Browser PPB2: Target Prediction Combining Nearest Neighbors with Machine Learning, *J. Chem. Inf. Model.*, 2019, **59**, 10–17.

38 J. Nickel, B.-O. Gohlke, J. Erehman, P. Banerjee, W. W. Rong, A. Goede, M. Dunkel and R. Preissner, SuperPred: update on drug classification and target prediction, *Nucleic Acids Res.*, 2014, **42**, W26–W31.

39 D. Chicco and G. Jurman, The Matthews correlation coefficient (MCC) should replace the ROC AUC as the standard metric for assessing binary classification, *BioData Min.*, 2023, **16**, 4.

40 A. S. Rifaioglu, E. Nalbat, V. Atalay, M. J. Martin, R. Cetin-Atalay and T. Doğan, DEEPScreen: high performance drug-target interaction prediction with convolutional neural networks using 2-D structural compound representations, *Chem. Sci.*, 2020, **11**, 2531–2557.

41 D. S. Himmelstein, A. Lizee, C. Hessler, L. Brueggeman, S. L. Chen, D. Hadley, A. Green, P. Khankhanian and S. E. Baranzini, Systematic integration of biomedical knowledge prioritizes drugs for repurposing, *eLife*, 2017, **6**, e26726.

42 C. L. Mellor, R. L. Marchese Robinson, R. Benigni, D. Ebbrell, S. J. Enoch, J. W. Firman, J. C. Madden, G. Pawar, C. Yang and M. T. D. Cronin, Molecular fingerprint-derived similarity measures for toxicological read-across: Recommendations for optimal use, *Regul. Toxicol. Pharmacol.*, 2019, **101**, 121–134.

43 H. Kuwahara and X. Gao, Analysis of the effects of related fingerprints on molecular similarity using an eigenvalue entropy approach, *J. Cheminf.*, 2021, **13**, 27.

44 R. Arakawa, N. Tamehiro, T. Nishimaki-Mogami, K. Ueda and S. Yokoyama, Fenofibric acid, an active form of fenofibrate, increases apolipoprotein A-I-mediated high-density lipoprotein biogenesis by enhancing transcription of ATP-binding cassette transporter A1 gene in a liver X receptor-dependent manner, *Arterioscler., Thromb., Vasc. Biol.*, 2005, **25**, 1193–1197.

45 Y. Xiao, M. Kim and M. A. Lazar, Nuclear receptors and transcriptional regulation in non-alcoholic fatty liver disease, *Mol. Metab.*, 2021, **50**, 101119.

46 W. G. Kim and S. Y. Cheng, Thyroid hormone receptors and cancer, *Biochim. Biophys. Acta*, 2013, **1830**, 3928–3936.