nature portfolio

Peer Review File

Evidential Deep Learning-based Drug-Target Interaction Prediction

Corresponding Author: Professor Song He

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
This manuscript presents an innovative DTI prediction framework, EviDTI, which addresses the important problem of the lack of confidence estimates in models for DTI predictions. EviDTI integrates multiple data dimensions, including 2D topology graphs and 3D spatial structures of drugs, as well as target sequence features, and utilizes evidential deep learning to quantify prediction uncertainty. Experimental results show that EviDTI is competitive with five state-of-the-art baseline models on three benchmark datasets. The authors demonstrate that EviDTI can calibrate prediction errors. The authors utilize the well-calibrated uncertainty information to prioritize high-confidence DTIs for experimental validation, which improves the efficiency of drug discovery. The manuscript performs rich experiments and discovers new potential inhibitors against FAK and FLT3 through uncertainty-based prediction. The manuscript is clearly described and easy to read. The reviewer believes that the EviDTI framework can support the decision-making process by providing a reliable measure of uncertainty, thus reducing the risk and cost associated with erroneous DTI predictions and pointing the way to experimental validation in the drug discovery process. My main recommendation focuses on the requirements for the developed model, in particular its generalization performance on the benchmark dataset. If the authors are able to address the following recommendations, the paper should be published.
points
1. The claim that uncertainty-based methods improve prediction reliability is well-supported through benchmarks and case studies. However, the relationship between uncertainty estimates and out-of-domain (OoD) sample performance could be explored in more depth. The reviewer recommended that the authors design a cold-start scenario to test the predictive power of the model on unseen targets and drugs.
2. The authors claim that EviDTI's UQ enhances prediction reliability. However, while the manuscript emphasizes uncertainty-guided ranking, it does not explore whether UQ improves overall predictive performance compared to non-uncertainty-based models. Recommendation: Provide ablation studies comparing prediction performance with and without UQ integration.
3. Although this paper focuses on deep learning-based methods, it is recommended to supplement the comparison with traditional machine learning models (e.g., Support Vector Machine SVM, Random Forest RF, etc.) in order to demonstrate the advantages of EviDTI more comprehensively.
4. Although EviDTI is compared against multiple baselines, methods like DeepDTA and GraphDTA are primarily designed for drug-target affinity prediction (regression tasks). Merely adapting them to classification might compromise their parameter optimization. Recommendation: Clarify how baseline models were adapted for classification. Include a broader comparison with other classification-specific methods, such as network-based DTI prediction models.
5. The manuscript lacks sufficient details on hyperparameter optimization. The authors should provide more information on the hyperparameter selection process and the training procedure, particularly how the parameters of different components of EviDTI (e.g., the drug and protein encoders) were optimized. This would help clarify the impact of hyperparameter tuning on the model's performance across different datasets.
6. To enhance the scholarly relevance of your paper, it is recommended to incorporate recent literature from 2022 to 2024, particularly focusing on uncertainty quantification and pre-trained models. This will ensure that your work reflects the latest advancements and situates it within the current research landscape.
7. In Tables 1, 2, and 3, the values in parentheses do not specify their meaning. Do these values represent variance? In order to ensure that the results are transparent and interpretable, authors are requested to clarify the exact meaning of these values in the text.
8. In Fig. 1, the authors use the symbol "⊕" to denote a vector collocation operation. However, in many cases, the symbol

"⊕" is commonly used to indicate a different-or operation.
9. In the EviDTI framework, drugs are represented using a fusion of 2D and 3D embeddings, while proteins are represented using only pre-trained sequence-based (1D) embeddings. Considering that tools such as AlphaFold2 (AF2) and AlphaFold3 (AF3) have been able to predict the 3D structure of proteins more accurately, what are the reasons and advantages of this choice?
10. Several grammatical errors were observed (e.g., inconsistent tenses in the abstract). A thorough proofreading is recommended.

(Remarks on code availability)
N/A

Reviewer #2

(Remarks to the Author)
In the manuscript, Zhou et al. developed an evidential deep learning tool to predict drug-target interactions (EviDTI). Their method quantifies uncertainty for its predictions. The EviDTI framework has components for a protein encoder which relies on primary sequence of the proteins, a component for a drug feature encoder which employs both 2D and 3D features of the molecules, and a component for evidence which includes uncertainty quantification. They then conducted a case study with EviDTI and tyrosine kinase inhibitors, followed by in vitro screening to test some of the predications.

By the authors' own admission, the performance of EviDTI is not impressive. Correspondingly, many of its high probability/low uncertainty ligands did not translate to these molecules affecting enzyme activity in vitro. Nevertheless, upon vadiation with DrugBank, Davis, and KIDA datasets, EviDTI outcompeted contemporary DTI models in several parameters, especially in accuracy, precision, and Matthews correlation coefficient. Having estimation of uncertainty in their framework is an advancement compared to other DTI models of drug-protein interaction, and this provides a measure of confidence.

The work generally supports the authors' conclusions. Inclusion of certain data and analysis would improve the story.

1. In line 188, the authors claim EviDTI improved accuracy, precision, etc. over the baseline model. This should likely be 'outperformed' baseline model or something similar. The methods don't describe combining EvitDTI with other models, just comparisons.

2. In Table S6, the authors seem to omit 'known targets' of the molecules and the relevant citations, except for the target and reference (target of Lyn, an SRC family tyrosine kinase) that nearly corroborate EviDTI's prediction (target of SRC).

a. Tyrphostin 9 has a target reported in the literature (PDFGR), and this is a peculiar omission from Table S6, when the authors cite Kandandapani et al. (Ref 52 in the main manuscript file). Ref 52 and citations within mention PDFGR.

b. In fact, the other three compounds in Table S6 have targets reported in the literature, or they wouldn't have been in the kinase library. These should be included, even if they do not match EviDTI's predictions.
c. In addition to inclusion in Table S6, it would be appropriate to also address whatever these predictions are in the Results text.

3. The authors say in Table S6 that the target for tolimidone in the literature is "SRC", when they cite a paper (Saporitio et al, Ref 51 in the main manuscript) that identifies the target as Lyn, a SRC family protein, but not SRC itself. If the predications from EviDTI are for kinase families, rather than specific kinase targets, this should be made clearer throughout the entire paper.

4. Table S6 should include probability and uncertainty values from EviDTI for the literature targets. I.e. How certain was EviDTI that tyrphostin 9 was or wasn't a ligand for PDFGR?

Additionally, there are some discussions/conclusions that would strengthen their report.

1. EviDTI predicted inhibition of SRC by tolimidone, when the literature citation the authors provide to support the predictions identifies tolimidone as an activator of an SRC family kinase, not an inhibitor. With this discrepancy, coupled with the in vitro activation they found in assay experiments, the authors should discuss this limitation of their model: that identifying ligands in an inhibitor prediction may retrieve ligands which may be activators in a biological system.

2. The real-world utility of the model would be enhanced if the authors spoke to abilities of EviDTI to predict for known oncoprotein sequences of the tyrosine kinases. E.g. there are two well-known mutations of FLT3 in AML: Is EviDTI aware of these mutations? If so, is there data for this (does EviDTI predict binding of vodobatinib to FLT3-ITD or FLT3-TKD)? If not, do the authors anticipate EviDTI to yield distinguished probabilities and uncertainty for compounds against FLT3 vs. FLT3-ITD

vs. FLT3-TKD? Perhaps not, since protein data is 1D, but this should be addressed in the discussion. In an ideal world, an antineoplastic agent would inhibit the oncoproteins more than protooncogenic proteins.

Some concerns with data analysis and interpretation include the confusion around some of the data presented.

1. The effect of Tyrphostin 9 and Vodobatinib on FLT3 is indecipherable and seems to have conflicting explanations between the text and figures.

a. Line 397 suggests Tyrphostin 9 and Vodobatinib inhibited FLT3, though the term “EC50” is used instead of “IC50”, and the dose response has an opposite trend and low doses gave the most inhibition.

b. If there was little inhibition at high concentration of Tryphostin 9 or Vodobatinib, but high inhibition at low concentration, the authors should address this in the text.

c. If Tryphostin 9 and Vodobatinib were found to agonize, the text should reflect this, and the x-axes on Fig. 5e and 5f should reflect activation (or ‘relative activity to control assays’ or something similar to normalized activity from a negative enzymatic control) and not inhibition.

d. If these compounds are all inhibiting in a dose-dependent fashion as the text suggests, Figure 5 should show this, and I would suggest abandoning “EC50” if all trends are inhibitory, for simplicity for the reader.

2. Further the effect of Tyrphostin 9 and Vodobatinib is also unclear in Table S6. It may be helpful to include in reach row in Table S6, whether the molecule’s effect was activating or inhibitory.

3. Throughout page 23, IC50 values are reported in ‘micromoles’. Besides not being reported as concentrations, if these values were intended to be ‘micromolar’ rather than ‘micromoles’, the values are off by an order of magnitude from the nanomolar values displayed in Figure 2b-d and Table S7.

4. Figure 5b-f implies the authors only achieved a maximum of 1% inhibition at the highest inhibitor concentrations.

a. If only 1% inhibition was achieved, that makes the results less impressive than the authors already admit.

b. Assuming 1 represents normalization or 100%, the graphs should indicate this, either by labeling the axis relative inhibition, or by keeping % on the axis and changing the values from 0-100 instead of 0-1.

5. In Tables 1-3, are the underlined values the second highest? If so, it would be helpful if this was explained in the table legend. And if so, Table 2 Accuracy is missing an underlined value. If parenthetical values in Tables 1-3 are error values, this should be denoted, at least in Comparison methods in Methods, if not also in the table.

6. In Figure 2, the nature of the error bars should be denoted.

7. A replicate statement exists in Table S7, but there should be one in Figure 2 legend, and IC50 and/or EC50 figures should have error bars, and corresponding explanation in the figure legend.

I will combine the soundness of the methodology with the detail in the methods required to reproduce the work, since some of my suggestions to methodology may have been done, but not conveyed in detail in the Methods section. The following comments all relate to the authors’ enzymatic activity assay, which in their current state, would be difficult to reproduce without additional details.

1. Inhibition assays should be run alongside negative controls. It would seem the authors have done this, since their IC50s are normalized to presumably the negative control enzymatic activity. However inhibition assays should also be run with at least single-point positive controls for their system. A positive control of known inhibitor(s) of FAK and Flt3, like sorafenib (or alternatively two different positive controls for each enzyme) assayed alongside unknowns would strengthen the methodology.

2. Inclusion of IC50 values for their chosen positive controls within the Methods section would allow comparison to literature values and validate their enzymatic system.

3. Inclusion of typical metabolism for negative control kinase assays for each FAK and Flt3 assays within the Methods section would also help validate their enzymatic system. I.e. What enzymatic activity corresponds to 0% inhibition, in concentration of substrate consumed or product generated (or concentration per reaction time per enzyme concentration). Representative values for relative light units at these activities can also be helpful to the reader.

4. There are several components of reaction conditions omitted, including:

a. The concentration of tyrosine kinase, either FAK or Flt3

b. The identity of the protein or peptide substrate used (or a statement that autophosphorylation was measured in the absence of protein or peptide substrate), and the concentration of that substrate if applicable.

c. The concentration of ATP used

d. Reaction volume

e. Reaction vessel/plate

f. Reaction duration

g. Reaction temperature

h. Buffer conditions, especially pH and magnesium concentration

i. The instrument used to measure light generated by ADP-Glo

5. What vehicle was used for inhibitor dilutions, as well as the final concentration of this vehicle in enzyme assays and a statement that this concentration was used in negative control assays are omitted.

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)
Dear Authors,
I am writing regarding the manuscript "Evidential Deep Learning for Guided Drug-Target Interaction Prediction," which presents EviDTI, a novel method for drug-target interaction prediction. The authors have developed an innovative approach that integrates multiple drug and protein feature representations - specifically, 1D protein features alongside 2D and 3D drug features. The model architecture employs a protein feature encoder for sequence processing, while utilizing MG-BERT and GeoGNN models for drug feature extraction. These features are then concatenated to generate comprehensive drug representations, which, combined with an evidential layer, produce both interaction probability predictions and associated uncertainty estimates.
While the authors have demonstrated their model's performance through comparisons on established benchmarks (DrugBank, KIBA, and Davis) and conducted ablation studies, I recommend the following improvements:
1. The comparative analysis would benefit from including additional state-of-the-art models, particularly those employing different architectural approaches, to better contextualize EviDTI's performance.
2. The manuscript's structure and formatting should adhere more closely to journal guidelines to enhance readability and accessibility.
3. Figure 1 would be more effective if redesigned to better highlight the model's key architectural components and their interactions, particularly the feature integration process.
4. The literature review should be expanded to incorporate recent advances in drug-target interaction prediction, specifically those published within the last 1-2 years, to better establish the work's novelty and contextual relevance.
5. The inclusion of GraphDTA as a baseline model requires further justification, given its primary design for binding affinity prediction rather than interaction prediction. The authors should elaborate on how this choice relates to their work's objectives.
6. The manuscript would benefit from a clearer articulation of its motivations and contributions, particularly highlighting how EviDTI advances the field beyond existing approaches.

(Remarks on code availability)
The authors provided comprehensive code accompanied by well-documented instructions, which offers readers a good opportunity for reproducibility of their work.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
The authors have addressed my previous comments.

(Remarks on code availability)
The codes run correctly.

Reviewer #2

(Remarks to the Author)
Zhao et al. describe a new evidential deep learning tool for predicting drug target interactions that they developed called EviDTI. Their method quantifies uncertainty and incorporates protein sequence with 2D and 3D models of ligands. To validate EviDTI, they performed a case study on its predictions with tyrosine kinase inhibitors, followed by in vitro screening. The impetus for this study was to begin to address probabilities and uncertainties in current DTI methods, and this work contributes to the field.

The authors have addressed most of the concerns of reviewers, concerning both comparative learning models and in vitro validation. The addition of data, references, and discussion since the original submission have strengthened the manuscript.

My only concern that remains are the in vitro activities reported for tyrphostin 9 and vodobatinib against FLT3, highlighted in Figure S7. The authors are still claiming activation, when their graphical representation indicates dose-dependent inhibition in the opposite direction as one would expect. Having changed the y-axis labels to relative activity was good. Normally dose-dependent inhibition would be measured as activity relative to a negative control (kinase assay with DMSO alone). All of the experiments where the authors note inhibition look normal, and look to be normalized to a negative control. e.g. their tyrphostin 9 and FAK (Fig. S71) indicates that around 0.1 to 1 nM of tyrphostin 9, activity of FAK is the same as it was for a data point which contained DMSO and no tyrphostin 9; and at, 10 μM tyrphostin 9, there was zero activity of tyrphostin 9. That all looks normal.

However, in the tryphostin 9 and FLT3 plot, at 0.1 - 1 nM tryphostin 9, according to their plot, FLT3 is completely inhibited and has 0% of negative control activity. The plot also suggests that at high concentrations of tryphostin 9 above 1 μM, FLT3 activity is approaching 100% of control activity. Meaning that low concentrations of tyrphostin 9 inhibit FLT and high concentrations do not. This doesn't make sense, unless the authors were inhibiting FLT with another molecule and tryphostin 9 recovered activity, which does not seem what their describing. Hypothetically, if a molecule at low concentrations had no effect (like tryphostin 9 on FAK), one would expect the enzyme activity relative to a non-inhibited data point to be close to 100% (as it is for tryphostin 9 on FAK). If the hypothetical compound was activating at high concentrations, one would expect the enzyme activity relative to a non-inhibited data point to be *above* 100%. It seems as though these plots might be normalized to themselves, that the highest activity within an EC50 experiment was called 100 and the lowest was called 0? These should be normalized to enzyme activity without compound (but the same DMSO concentration as with compound). Or if my interpretation is off, the authors should explain how they’re normalizing their relative enzyme activities.

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)
The authors have solved all my concerns. I have no other suggestions.

(Remarks on code availability)

Version 2:

Reviewer comments:

Reviewer #2

(Remarks to the Author)
The authors have effectively clarified their biochemical results.

(Remarks on code availability)

Response to Referees Letter

Title: Evidential Deep Learning-based Drug-Target Interaction Prediction

Manuscript ID: NCOMMS-24-71365A

Authors: Yanpeng Zhao [a, b †], Yuting Xing [c †], Yixin Zhang [a †], Yifei Wang [a], Mengxuan Wan [a, b], Duoyun Yi [a, d], Chengkun Wu [e, f], Shangze Li [a], Huiyan Xu [a, d], Hongyang Zhang [a, d], Ziyi Liu [a, c], Guowei Zhou [a, g], Mengfan Li [a], Xuanze Wang [a], Zhengshan Chen [a], Ruijiang Li [a], Lianlian Wu [a, g], Dongsheng Zhao [a], Peng Zan [d] *, Song He [a] *, Xiaochen Bo [a] *

[a] Academy of Military Medical Sciences, Beijing 100850, China

[b] School of Medicine, Shanghai University, Shanghai 200444, China

[c] Defense Innovation Institute, Beijing 100071, China

[d] Shanghai Key Laboratory of Power Station Automation Technology, School of Mechatronics Engineering and Automation, Shanghai University, Shanghai 200444, China

[e] Laboratory of Digitizing Software for Frontier Equipment, National University of Defense Technology, Changsha 410073, China

[f] National Key Laboratory of Parallel and Distributed Computing, National University of Defense Technology, Changsha 410073, China

[g] Academy of Medical Engineering and Translational Medicine, Tianjin University, Tianjin 300072, China

---

* Corresponding authors.

E-mail addresses: boxc@bmi.ac.cn (Xiaochen Bo), hes1224@163.com (Song He), zanpeng@shu.edu.cn (Peng Zan)

† These authors contributed equally to this work.

Point-by-point Response to Reviewers' Comments

Reviewer #1:

This manuscript presents an innovative DTI prediction framework, EviDTI, which addresses the important problem of the lack of confidence estimates in models for DTI predictions. EviDTI integrates multiple data dimensions, including 2D topology graphs and 3D spatial structures of drugs, as well as target sequence features, and utilizes evidential deep learning to quantify prediction uncertainty. Experimental results show that EviDTI is competitive with five state-of-the-art baseline models on three benchmark datasets. The authors demonstrate that EviDTI can calibrate prediction errors. The authors utilize the well-calibrated uncertainty information to prioritize high-confidence DTIs for experimental validation, which improves the efficiency of drug discovery. The manuscript performs rich experiments and discovers new potential inhibitors against FAK and FLT3 through uncertainty-based prediction. The manuscript is clearly described and easy to read. The reviewer believes that the EviDTI framework can support the decision-making process by providing a reliable measure of uncertainty, thus reducing the risk and cost associated with erroneous DTI predictions and pointing the way to experimental validation in the drug discovery process. My main recommendation focuses on the requirements for the developed model, in particular its generalization performance on the benchmark dataset. If the authors are able to address the following recommendations, the paper should be published.

Response:

Thank you for the efforts you have put into reviewing our manuscript. In accordance with your suggestions, we have revised the manuscript from six aspects (**Part 1-6** below).

1) Newly added cold start dataset. We introduced a cold-start dataset to evaluate the model's ability to predict novel DTIs. The performance of EviDTI was compared with baseline models in the cold-start dataset (Page 10 Line 185-191 in the revised manuscript). Additionally, the capability of EviDTI to provide a reliable measure of uncertainty in the cold-start scenario was verified. (Page 17 Line 275-279

in the revised manuscript)

2) Newly added models and performance comparative study. We introduced additional widely used models for comparison with EviDTI. These models include **machine learning algorithms** (e.g., Random Forest, Support Vector Machines) as well as the **latest deep learning algorithms** (e.g., GraphformerDTI, DLMDTI, AIGO-DTI)**. Specifically, AIGO-DTI is a network-based DTI prediction model.** (Page 9-10 Line 163-184 in the revised manuscript) In addition, we explained in detail why GraphDTA was chosen as the comparison model and how GraphDTA can be applied to classification tasks. (Page 43 Line 757-762 in the revised manuscript)

3) **Newly added ablation study to compare prediction performance with and without UQ integration.** In order to assess the impact of uncertainty quantification on the overall prediction performance, we performed an ablation study comparing prediction performance between model with and without UQ integration. (Page 22 Line 362-369 in the revised manuscript)

4) **Expanded description of the hyperparameter selection process.** We provided detailed information on the hyperparameter selection process and training procedures. Specifically, for the different components of EviDTI (e.g., drug encoder and protein encoder), we described the process of finding the optimal parameter configurations. (Page 18 Line 197-215 in the revised Supplementary Information)

5) **Incorporation of the latest literature.** In the '*Introduction*' section, we included the latest research from 2022 to 2025, focusing on advances in the areas of uncertainty quantification and pre-trained models. (Page 4 Line 74 and Page 5 Line 94 in the revised manuscript)

6) **Made correction of inaccurate expressions, grammar, and representations of figures and tables.**

- Clearly stating in Tables 1-3 that the values in parentheses represent standard deviation (SD). (Page 11 Line 193, Page 12 Line 198 and Page 13 Line 204 in the revised manuscript)
- Modifying symbols in the graphs, e.g., replacing "⊕" with a more intuitive identifier. (Page 8 Line 149 in the revised manuscript)

- Correcting grammar and tenses throughout the manuscript.

Below we provide a point-by-point response to each suggestion. All modifications in the revised manuscript are highlighted in red text.

Points

1. The claim that uncertainty-based methods improve prediction reliability is well-supported through benchmarks and case studies. However, the relationship between uncertainty estimates and out-of-domain (OoD) sample performance could be explored in more depth. The reviewer recommended that the authors design a cold-start scenario to test the predictive power of the model on unseen targets and drugs.

Response:

Thank you for your insightful feedback and constructive suggestion. We added a cold-start dataset to evaluate the performance of EviDTI **(Part 1** below). In addition, we assessed whether EviDTI can provide a reliable measure of uncertainty under cold-start scenarios (**Part 2** below).

(1) Statement of performance comparative study on the cold-start dataset

In order to assess the ability of EviDTI to predict novel DTIs, we implemented a cold-start scenario following the practice established by Wang et al[1]. Specifically, we selected 10% of drugs from DrugBank dataset, and all DTIs associated with these drugs were used as test set. This resulted in a total of 3,170 cold-start test samples and 29,894 training samples. (The corresponding revision can be seen Line 626-630 on Page 36 of the revised manuscript)

Table R1 (corresponding to Table S2 of the revised supplemental information) demonstrates the performance of EviDTI under the cold-start scenario. EviDTI outperforms other models in several evaluation metrics, especially in accuracy (79.96%), recall (81.20%), F1 score (79.61%) and MCC value (59.97%). Its AUC value (86.69%) is slightly lower than TransformerCPI's 86.93%. These results demonstrate that EviDTI is a competitive model in cold start scenarios. (The corresponding revision can be seen Line 185-191 on Page 10 of the revised manuscript)

Table R1 Comparison results of EviDTI and baselines on the cold start dataset. The best results are highlighted in bold, while the second-best results are underlined.

Cold start	Accuracy	Recall	Precision	MCC	F1 score	AUC	AUPR
DeepConv-DTI	78.56	78.32	76.32	57.75	77.31	86.43	82.12
GraphDTA	76.23	74.76	75.34	52.54	75.05	84.23	82.54
MolTrans	78.70	78.59	**86.64**	57.37	78.05	85.87	<u>86.64</u>
HyperAttention	73.66	<u>79.61</u>	70.10	48.06	72.14	82.40	82.93
TransformerCPI	78.94	76.05	79.40	57.82	77.67	**86.93**	**87.78**
GraphormerDTI	76.16	68.88	<u>80.87</u>	53.19	74.14	83.15	85.27
AIGO-DTI	73.17	61.60	78.57	47.21	68.64	80.16	79.05
DLM-DTI	<u>79.16</u>	76.56	81.26	<u>59.33</u>	<u>78.80</u>	86.43	86.10
EVIDTI	**79.96**	**81.20**	78.09	**59.97**	**79.61**	<u>86.69</u>	85.43

(2) Extended analysis of uncertainty quantification under cold-start datasets

Next, we analyzed whether the model could provide a reliable measure of uncertainty in the cold-start dataset. Here, we verified the hypothesis that samples with correct predictions should have low uncertainty, while samples with incorrect predictions should have high uncertainty.

Figure R1 (<u>corresponding to **Supplementary Figure S5** of the revised supplemental information</u>) illustrates the relationship between sample prediction results and uncertainty values in cold-start dataset. The horizontal axis divides the sample into true positives (TP, true value 1, predicted value 1), false positives (FP, true value 0, predicted value 1), false negatives (FN, true value 1, predicted value 0), and true negatives (TN, true value 0, predicted value 0). The vertical axis represents the distribution of uncertainty values for each group. In the cold-start dataset, samples with incorrect predictions (FP and FN) typically exhibit higher uncertainty than samples with correct predictions (TP and TN). These findings suggest that EviDTI provides a reliable measure of uncertainty under cold-start dataset. <u>(The corresponding revision can be seen Line 275-279 on Page 17 of the revised manuscript)</u>

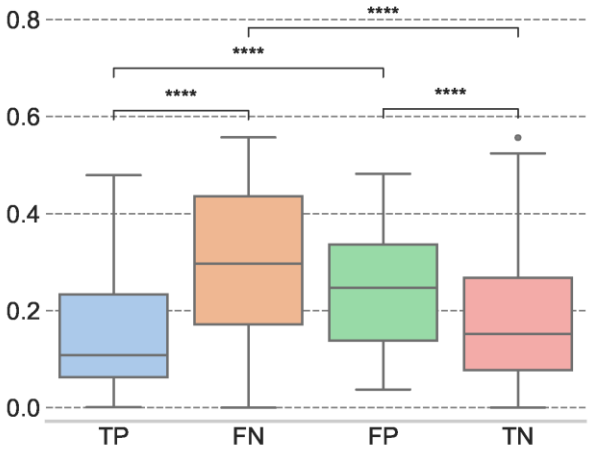


Figure R1 A Mann-Whitney test was performed on the error distribution of uncertainty in samples classified as TP, FP, FN, TN, respectively. All p-values <0.0001. The box plots show the median as the center lines. The black dots represent outliers.

2. The authors claim that EviDTI's UQ enhances prediction reliability. However, while the manuscript emphasizes uncertainty-guided ranking, it does not explore whether UQ improves overall predictive performance compared to non-uncertainty-based models. Recommendation: Provide ablation studies comparing prediction performance with and without UQ integration.

Response:

We sincerely appreciate the reviewer's insightful critique regarding the role of uncertainty quantification (UQ) in predictive performance. We compared model with UQ integration and the model without UQ integration across three benchmark datasets (DrugBank, KIBA and Davis). The model without UQ integration has the same architecture but does not use the evidential layer. As shown in **Figure R2** (corresponding to **Supplementary Figure S6** of the revised supplemental information), the model with UQ integration performs comparably to the model without UQ integration respect to the standard evaluation metrics.

However, uncertainty quantification remains important for DTI predictions. The primary objective of DTI identification is to maximize the chances of discovering true DTI by prioritizing high-confidence predictions rather than exhaustively testing all possibilities. The incorporation of UQ into the DTI prediction can increase hit rates,

reduce false positives and increase the efficiency of drug discovery by focusing experimental validation on the most reliable predictions[2]. (The corresponding revision can be seen Line 362-369 on Page 22 of the revised manuscript)

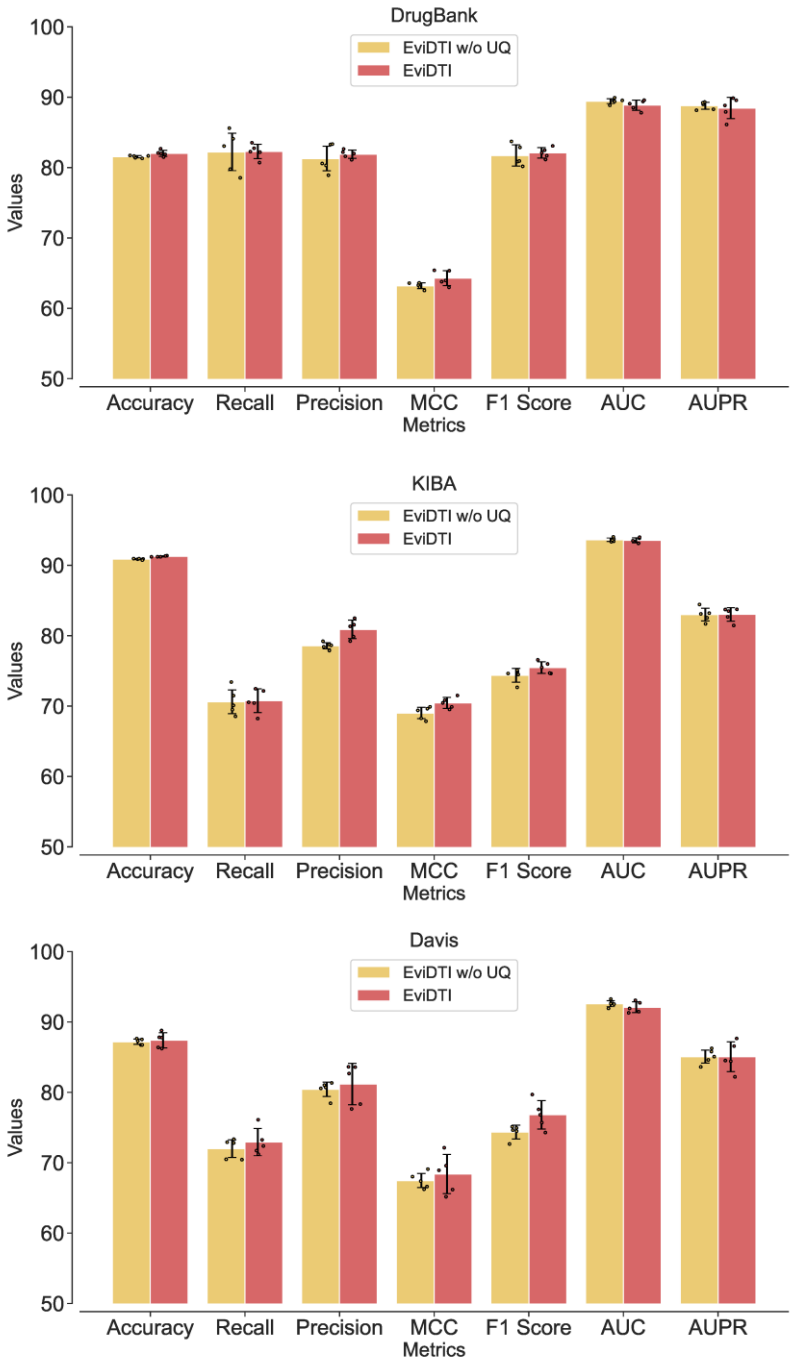


Figure R2 Performance comparison on the DrugBank, KIBA and Davis datasets using uncertainty quantification and using non-uncertainty quantification strategies. Five independent replications of each method were performed (n=5). Data are expressed as means ± std.

3. Although this paper focuses on deep learning-based methods, it is recommended to supplement the comparison with traditional machine learning models (e.g., Support Vector Machine SVM, Random Forest RF, etc.) in order to demonstrate the advantages of EviDTI more comprehensively.

Response:

Thank you for this valuable suggestion. We agree that comparing EviDTI with traditional machine learning (ML) models could further demonstrate its strengths.

We systematically evaluated EviDTI against RF, SVM, and NB across three datasets (DrugBank, KIBA, Davis). The results on the DrugBank dataset are shown in **Table R2**, EviDTI significantly outperforms all machine learning models on all evaluation metrics except Recall. The greatest enhancement in terms of MCC (64.29%) and F1 score (82.09%) was observed in EviDTI, surpassing the second-ranked RF model by 13.06 and 7.89, respectively. The results on the KIBA dataset are shown in **Table R3**, EviDTI demonstrated superior performance in all evaluation metrics with the exception of Recall. The metric that improved the most was MCC (7.23 higher than RF), followed by AUPR (6.98 higher than RF). The results on the Davis dataset are shown in **Table R4**, EviDTI outperforms the machine learning models on all metrics.

Overall, on all three benchmark datasets, EviDTI outperforms traditional machine learning models on most of metrics such as accuracy, MCC, F1 score, AUC and AUPR.

Table R2 Comparison results of the EviDTI and ML models on the DrugBank dataset. The best results are highlighted in bold, while the second-best results are underlined.

DrugBank	Accuracy (std)	Recall (std)	Precision (std)	MCC (std)	F1 score (std)	AUC (std)	AUPR (std)
RF	71.07 (1.34)	73.08 (0.95)	76.92 (1.42)	51.23 (1.54)	74.94 (0.92)	83.03 (0.65)	84.78 (0.67)
NB	52.90 (1.13)	**87.99 (1.77)**	51.71 (1.01)	42.34 (3.24)	65.13 (0.89)	60.76 (1.24)	59.37 (1.77)
SVM	78.02 (1.92)	75.04 (2.93)	73.86 (2.14)	53.87 (1.29)	74.39 (1.40)	82.02 (1.62)	82.85 (1.49)
EviDTI	**82.02 (0.4)**	82.28 (1.0)	**81.90 (0.5)**	**64.29 (1.0)**	**82.09 (0.7)**	**88.87 (0.7)**	**88.45 (0.2)**

Note: Five independent replications of each method were performed (n=5). Data are expressed as means (std).

Table R3 Comparison results of the EviDTI and ML models on the KIBA dataset. The best results are highlighted in bold, while the second-best results are underlined.

KIBA	Accuracy (std)	Recall (std)	Precision (std)	MCC (std)	F1 score (std)	AUC (std)	AUPR (std)
RF	88.41 (0.34)	64.84 (2.48)	70.57 (2.30)	63.21 (1.67)	67.59 (1.37)	91.05 (0.38)	76.05 (1.04)
NB	52.92 (0.83)	**71.72 (1.77)**	24.22 (1.40)	24.56 (3.24)	36.21 (1.47)	65.92 (0.78)	35.51 (0.94)
SVM	81.72 (0.46)	3.87 (0.76)	66.40 (2.91)	5.24 (1.34)	7.32 (0.92)	73.27 (0.56)	43.38 (1.60)
EviDTI	**91.27 (0.7)**	70.77 (1.6)	**80.90 (1.3)**	**70.44 (0.7)**	**75.48 (0.8)**	**93.57 (0.3)**	**83.03 (0.9)**

Note: Five independent replications of each method were performed (n=5). Data are expressed as means (std).

Table R4 Comparison results of the EviDTI and ML models on the Davis dataset. The best results are highlighted in bold, while the second-best results are underlined.

Davis	Accuracy (std)	Recall (std)	Precision (std)	MCC (std)	F1 score (std)	AUC (std)	AUPR (std)
RF	<u>84.48 (0.60)</u>	<u>65.01 (1.47)</u>	<u>80.70 (1.87)</u>	<u>64.12 (2.11)</u>	<u>72.01 (1.49)</u>	<u>89.64 (0.68)</u>	<u>83.04 (1.68)</u>
NB	67.53 (0.92)	61.70 (1.82)	47.78 (1.95)	45.21 (1.68)	53.85 (1.72)	70.11 (1.39)	58.43 (2.09)
SVM	82.13 (0.59)	59.70 (1.57)	76.96 (2.19)	60.23 (1.45)	67.24 (1.61)	87.85 (0.78)	80.16 (1.65)
EviDTI	**87.40 (0.7)**	**72.93 (1.6)**	**81.17 (1.3)**	**68.40 (0.7)**	**76.81 (0.8)**	**92.10 (0.3)**	**85.07 (0.9)**

Note: Five independent replications of each method were performed (n=5). Data are expressed as means (std).

4. Although EviDTI is compared against multiple baselines, methods like DeepDTA and GraphDTA are primarily designed for drug-target affinity prediction (regression tasks). Merely adapting them to classification might compromise their parameter optimization. Recommendation: Clarify how baseline models were adapted for classification. Include a broader comparison with other classification-specific methods, such as network-based DTI prediction models.

Response:

Thank you for this thoughtful critique. We appreciate the opportunity to clarify our baseline adaptation process and expand the comparison to classification-specific methods.

GraphDTA is primarily designed for drug-target affinity prediction. However, GraphDTA has been widely adopted as a benchmark model by many other state-of-the-art DTI[3-5] methods due to its advantages in processing molecular structure information. These studies demonstrate that GraphDTA serves as a robust baseline model, providing an important reference standard for evaluating novel DTI prediction models. To clarify

how GraphDTA can be adjusted for fair comparisons, we provided detailed explanations in **Part 1** below. (The corresponding revision can be seen Line 757-762 on Page 43 of the revised manuscript)

To further enhance the comparative analyses, we have added three methods covering two categories of methods: network-based models and Proteochemometrics -based models (**Table R5**). Details of these methods are provided in **Part 2**. (The corresponding revision can be seen Line 780-804 on Page 44-45 of the revised manuscript)

Furthermore, we conducted a comprehensive performance comparison between the baseline methods and EviDTI on three benchmark datasets **(Part3**). (The corresponding revision can be seen Line 163-184 on Page 9-10 of the revised manuscript)

(1) Description of how to adjust GraphDTA for fair comparison

In order to facilitate fair comparisons, GraphDTA was adapted according to the following steps: (1) A Sigmoid activation function was added after the last layer to convert the model output to binary probabilities. (2) A binary cross-entropy was used as the loss function instead of the original mean square error (MSE) loss. (3) A grid search was performed to optimize key hyperparameters such as learning rate and batch size. (The corresponding revision can be seen Line 757-762 on Page 43 of the revised manuscript)

(2) Newly added of three comparison methods

Table R5 Categories of newly added DTI prediction methods

Category of methods	Subcategories of methods	Model
Network-based		AIGO-DTI
Proteochemometrics-based	Based on gated cross-attention mechanism	GraphormerDTI
	Based on protein and small molecule pre-training	DLM-DTI

In order to further enrich our comparative analyses, we have incorporated three methods, which are divided into two categories: network-based DTI prediction models and protein chemometrics-based DTI prediction models. Below are details of these additional methods:

Category 1: Network-Based Models (The corresponding revision can be seen Line 780-787 on Page 44 of the revised manuscript)

AIGO-DTI: AIGO-DTI proposes an Adaptive Iterative Graph Optimization (AIGO)-DTI prediction framework. This framework integrates atomic cluster information and enhances molecular features through the design of functional group prompts and graph encoders, thereby optimizing the construction of DTI association networks. Furthermore, the optimization of graph structure is transformed into a node similarity learning problem, utilizing multi-head similarity metric functions to iteratively update the network structure to improve the quality of DTI information.

Category 2: Proteochemometrics-Based Models (The corresponding revision can be seen Line 788-804 on Page 45 of the revised manuscript)

GraphormerDTI: GraphormerDTI integrates the Graph Transformer neural network with a 1D-CNN to extract the representations of drug and target, respectively. Specifically, GraphormerDTI embeds molecular graphs into vector representations through iterative Transformer-based message passing. The structural representations of the molecules are encoded through node centrality encoding, node spatial encoding and edge encoding. The attentional mechanisms are used to model interactions.

DLM-DTI: DLM-DTI was comprised of three primary components: the drug encoder, the target encoder, and the interaction prediction head. The target encoder incorporates both the teacher and student models of language models for protein sequences. The teacher model employed for target sequence encoding was the ProtBERT model. The student model was designed to be consistent with the original teacher model with fewer layers to conduct knowledge distillation. The function of the drug encoder is to convert SMILES sequences to molecular representations, and this is achieved by employing the ChemBERTa encoder. The interaction prediction header predicts interaction probabilities by concatenating drug and target representations.

(3) Comparative evaluation of EviDTI and baseline models (The corresponding revision can be seen Line 163-184 on Page 9-10 of the revised manuscript)

The performance results on the DrugBank dataset are shown in **Table R6** (corresponding to **Table 1** of the revised manuscript). EviDTI demonstrates robust overall performance on all metrics, especially in terms of precision (81.90%), and competitive values for Accuracy (82.02%), MCC (64.29) and F1 score (82.09%). Besides, we evaluated EviDTI on the Davis and KIBA datasets, which are particularly challenging due to class imbalance. The performance results on the Davis and KIBA datasets are presented in **Tables R7 and R8** (corresponding to **Table 2 and 3** of the revised manuscript), respectively. Specifically, on the KIBA dataset, EviDTI outperformed the best baseline model by 0.6% in accuracy, 0.4% in precision, 0.3% in MCC, 0.4% in F1 score, and 0.1% in AUC. On the Davis dataset, EviDTI exceeds 0.8% in accuracy, 0.6% in precision, 0.9% in MCC, 2% in F1 score, 0.1% in AUC, and 0.3% in AUPR. These results demonstrate the robustness and superior performance of EviDTI when dealing with complex and unbalanced datasets. The findings underscore the efficacy and competitiveness of EviDTI.

Table R6 Comparison results of EviDTI and baselines on the DrugBank dataset. The best results are highlighted in bold, while the second-best results are underlined.

DrugBank	Accuracy (std)	Recall (std)	Precision (std)	MCC (std)	F1 score (std)	AUC (std)	AUPR (std)
RF	71.07 (1.34)	73.08 (0.95)	76.92 (1.42)	51.23 (1.54)	74.94 (0.92)	83.03 (0.65)	84.78 (0.67)
NB	52.90 (1.13)	87.99 (1.77)	51.71 (1.01)	42.34 (3.24)	65.13 (0.89)	60.76 (1.24)	59.37 (1.77)
SVM	78.02 (1.92)	75.04 (2.93)	73.86 (2.14)	53.87 (1.29)	74.39 (1.40)	82.02 (1.62)	82.85 (1.49)
DeepConv-DTI	79.50 (0.2)	80.45 (0.2)	79.30 (0.6)	57.91 (0.6)	79.87 (0.6)	86.69 (0.2)	86.28 (0.1)
GraphDTA	77.69 (0.1)	78.77 (1.1)	76.92 (1.0)	55.41 (0.6)	77.83 (1.1)	83.62 (0.2)	83.26 (0.1)
MolTrans	78.48 (0.1)	81.97 (0.2)	77.01 (0.3)	55.65 (0.6)	79.41 (0.7)	85.98 (0.6)	86.67 (0.1)
HyperAttention	80.85 (0.2)	**88.16 (0.1)**	77.09 (0.1)	62.32 (0.4)	**82.26 (0.3)**	89.94 (0.1)	88.95 (0.1)
TransformerCPI	79.01 (0.2)	76.26 (0.3)	80.47 (0.3)	58.09 (0.2)	78.31 (0.3)	84.62 (0.3)	84.23 (0.2)
GraphormerDTI	80.22 (1.07)	82.05 (1.23)	79.11 (0.91)	60.48 (2.16)	80.55 (1.02)	88.63 (1.02)	88.36 (0.99)
AIGO-DTI	81.88 (1.37)	82.28 (2.69)	81.26 (0.36)	63.80 (2.78)	81.74 (1.40)	88.62 (1.02)	88.44 (1.65)
DLM-DTI	**82.13 (0.27)**	82.54 (1.91)	81.68 (1.56)	**66.32 (0.55)**	81.08 (0.56)	**90.48 (0.48)**	**90.20 (0.60)**
EviDTI	82.02 (0.4)	82.28 (1.0)	**81.90 (0.5)**	64.29 (1.0)	82.09 (0.7)	88.87 (0.7)	88.45 (0.2)

Note: Five independent replications of each method were performed (n=5). Data are expressed as means (std)

Table R7 Comparison results of EviDTI and baselines on the KIBA dataset. The best results are highlighted in bold, while the second-best results are underlined.

KIBA	Accuracy (std)	Recall (std)	Precision (std)	MCC (std)	F1 score (std)	AUC (std)	AUPR (std)
RF	88.41 (0.34)	64.84 (2.48)	70.57 (2.30)	63.21 (1.67)	67.59 (1.37)	91.05 (0.38)	76.05 (1.04)
NB	52.92 (0.83)	71.72 (1.77)	24.22 (1.40)	24.56 (3.24)	36.21 (1.47)	65.92 (0.78)	35.51 (0.94)
SVM	81.72 (0.46)	3.87 (0.76)	66.40 (2.91)	5.24 (1.34)	7.32 (0.92)	73.27 (0.56)	43.38 (1.60)
DeepConv-DTI	89.05 (0.3)	68.87 (0.4)	80.52 (0.3)	70.21 (0.2)	74.24 (0.6)	90.67 (0.1)	80.42 (0.6)
GraphDTA	88.92 (0.1)	59.42 (3.2)	77.53 (2.0)	67.51 (0.5)	71.24 (0.8)	91.41 (0.1)	77.61 (0.7)
MolTrans	88.91 (0.1)	73.53 (0.2)	70.42 (0.3)	67.62 (0.2)	71.94 (0.1)	92.32 (0.3)	79.49 (0.6)
HyperAttention	89.33 (0.1)	79.63 (0.2)	68.93 (0.1)	70.12 (0.2)	74.21 (0.3)	93.51 (0.1)	81.41 (0.2)
TransformerCPI	87.01 (0.1)	63.12 (0.3)	66.93 (0.3)	66.2 (0.2)	69.12 (0.2)	88.81 (0.1)	70.81 (0.1)
GraphormerDTI	86.54 (0.85)	**85.40 (1.68)**	60.63 (1.84)	64.03 (1.10)	70.87 (0.89)	92.58 (0.25)	81.95 (0.91)
AIGO-DTI	90.68 (0.32)	73.65 (0.77)	76.77 (1.12)	69.46 (0.97)	75.17 (0.77)	93.06 (0.26)	**83.91 (0.77)**
DLM-DTI	88.25 (0.25)	68.10 (2.15)	69.86 (0.90)	61.72 (0.78)	68.94 (0.87)	91.14 (0.39)	75.63 (0.41)
EviDTI	**91.27 (0.7)**	70.77 (1.6)	**80.90 (1.3)**	**70.44 (0.7)**	**75.48 (0.8)**	**93.57 (0.3)**	83.03 (0.9)

Note: Five independent replications of each method were performed (n=5). Data are expressed as means (std).

Table R8 Comparison results of the EviDTI and baselines on the Davis dataset. The best results are highlighted in bold, while the second-best results are underlined.

Davis	Accuracy (std)	Recall (std)	Precision (std)	MCC (std)	F1 score (std)	AUC (std)	AUPR (std)
RF	84.48 (0.60)	65.01 (1.47)	80.70 (1.87)	64.12 (2.11)	72.01 (1.49)	89.64 (0.68)	83.04 (1.68)
NB	67.53 (0.92)	61.70 (1.82)	47.78 (1.95)	45.21 (1.68)	53.85 (1.72)	70.11 (1.39)	58.43 (2.09)
SVM	82.13 (0.59)	59.70 (1.57)	76.96 (2.19)	60.23 (1.45)	67.24 (1.61)	87.85 (0.78)	80.16 (1.65)
DeepConv-DTI	86.41 (0.2)	72.14 (0.3)	77.85 (0.4)	64.53 (0.3)	74.89 (0.6)	91.83 (0.2)	83.87 (0.6)
GraphDTA	81.72 (0.1)	53.01 (1.7)	74.33 (1.4)	62.42 (0.6)	71.83 (0.8)	85.92 (0.4)	74.31 (0.7)
MolTrans	83.61 (0.1)	76.31 (0.3)	68.87 (0.2)	62.91 (0.1)	72.40 (0.1)	89.84 (0.2)	80.26 (0.6)
HyperAttention	86.61 (0.1)	78.01 (0.2)	75.41 (0.1)	67.54 (0.2)	74.21 (0.2)	92.01 (0.1)	83.92 (0.1)
TransformerCPI	82.22 (0.1)	68.82 (0.3)	68.81 (0.3)	64.21 (0.2)	72.41 (0.2)	87.72 (0.1)	76.71 (0.1)
GraphormerDTI	84.81 (0.84)	**83.78 (2.52)**	68.43 (2.18)	65.22 (1.84)	75.28 (1.43)	91.24 (0.74)	84.31 (1.48)
AIGO-DTI	86.19 (0.79)	72.55 (2.94)	78.01 (2.11)	65.72 (1.96)	75.12 (1.52)	92.01 (0.97)	84.81 (1.10)
DLM-DTI	85.66 (0.47)	74.08 (2.50)	75.26 (1.92)	64.68 (1.27)	74.62 (1.03)	91.30 (0.35)	82.48 (0.86)
EviDTI	**87.40 (0.7)**	72.93 (1.6)	**81.17 (1.3)**	**68.40 (0.7)**	**76.81 (0.8)**	**92.10 (0.3)**	**85.07 (0.9)**

Note: Five independent replications of each method were performed (n=5). Data are expressed as means (std).

5. The manuscript lacks sufficient details on hyperparameter optimization. The authors should provide more information on the hyperparameter selection process and the training procedure, particularly how the parameters of different components of EviDTI (e.g., the drug and protein encoders) were optimized. This would help clarify the impact of hyperparameter tuning on the model's performance across different datasets.

Response:

Thank you for your valuable comments on our manuscript. We provided detailed additions and explanations on the missing details of hyperparameter optimization.

For the key hyperparameters of each major component (e.g., learning rate, batch size, hidden layer dimension, etc.), we adopted a grid search approach to explore different hyperparameter combinations. First, we conducted a series of experiments to ascertain the impact of varying learning rates and batch sizes on the outcomes. Next, we explored the impact of different feature dimensions of the target and the drug on the model performance, with the objective of identifying the optimal set of dimensions that would adequately capture the features of the input data without resulting in an overly complex model structure. Third, we further explored the effect of the number of neurons in the fully connected layer on model performance. As shown in **Figure R3** (corresponding to **Supplementary Figure S9** of the revised Supplementary Information)**,** the final selected hyperparameter combination demonstrates strong performance in terms of AUC.

The above results have been added to the supplementary information. (The corresponding revision can be seen Line 198-215 on Page 18 of the revised Supplementary Information)

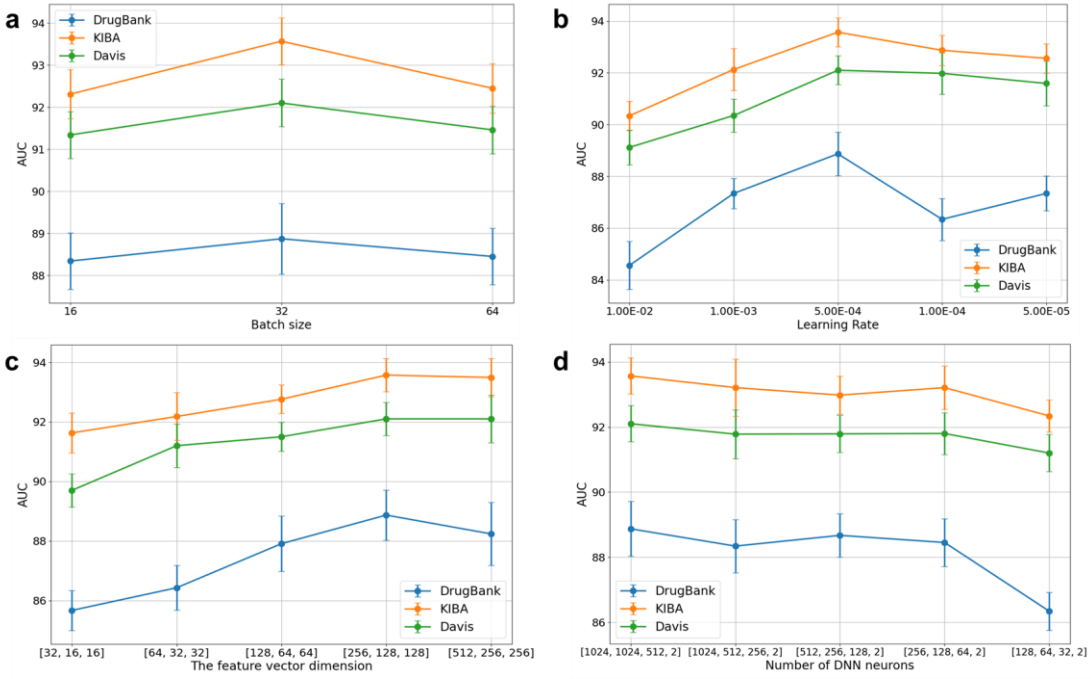


Figure R3 Impact of hyperparameter on AUC.

6. To enhance the scholarly relevance of your paper, it is recommended to incorporate recent literature from 2022 to 2024, particularly focusing on uncertainty quantification and pre-trained models. This will ensure that your work reflects the latest advancements and situates it within the current research landscape.

Response:

Thank you for your valuable suggestions for our research. In the revised manuscript, we have incorporated recent literature from 2022 to 2025, particularly focusing on uncertainty quantification and pre-trained models.

Uncertainty Quantification (Page 5 Line 94 in the revised manuscript)

1. Tan, H. S., Wang, K. & McBeth, R. Deep evidential learning for radiotherapy dose prediction. *Computers in Biology and Medicine* **182**, 109172 (2024).
2. Lin, Y. *et al.* Uncertainty quantification via a memristor Bayesian deep neural network for risk-sensitive reinforcement learning. *Nature Machine Intelligence* **5**, 714–723 (2023).

Pre-trained models (Page 4 Line 74 in the revised manuscript)

3 Hayes, T. *et al.* Simulating 500 million years of evolution with a language model. *Science* **387**, 850–858 (2025).

4. Lin, Z. *et al.* Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science* **379**, 1123–1130 (2023).
5. Fang, Y. *et al.* Knowledge graph-enhanced molecular contrastive learning with functional prompt. *Nature Machine Intelligence* **5**, 542–553 (2023).
6. Li, H. *et al.* A knowledge-guided pre-training framework for improving molecular representation learning. *Nature Communications* **14**, 7568 (2023).

7. In Tables 1, 2, and 3, the values in parentheses do not specify their meaning. Do these values represent variance? In order to ensure that the results are transparent and interpretable, authors are requested to clarify the exact meaning of these values in the text.

Response:

Thank you for your valuable comments. The parenthetical values in Table 1-3 are

standard deviation. We further explain the meaning of the parenthetical values in the table legend: "*Five independent replications of each method were performed (n=5). Data are expressed as means (std)*". (Page 11 Line 193, Page 12 Line 198 and Page 13 Line 204 in the revised manuscript) In addition, in the "Comparative Methods" section, we have further explained the experimental design and the way the results are reported: "*All methods were conducted with five repeated experiments, and the results are reported as the mean and variance.*" (Page 42 Line 746 in the revised manuscript)

8. In Fig. 1, the authors use the symbol "⊕" to denote a vector collocation operation. However, in many cases, the symbol "⊕" is commonly used to indicate a different-or operation.

Response：

Thank you for your comment. As shown in **Figure R4** (corresponding to Figure 1 of the revised manuscript), we have replaced ⊕ with the text "Concat". (The corresponding revision can be seen Line 149 on Page 8 of the revised manuscript)

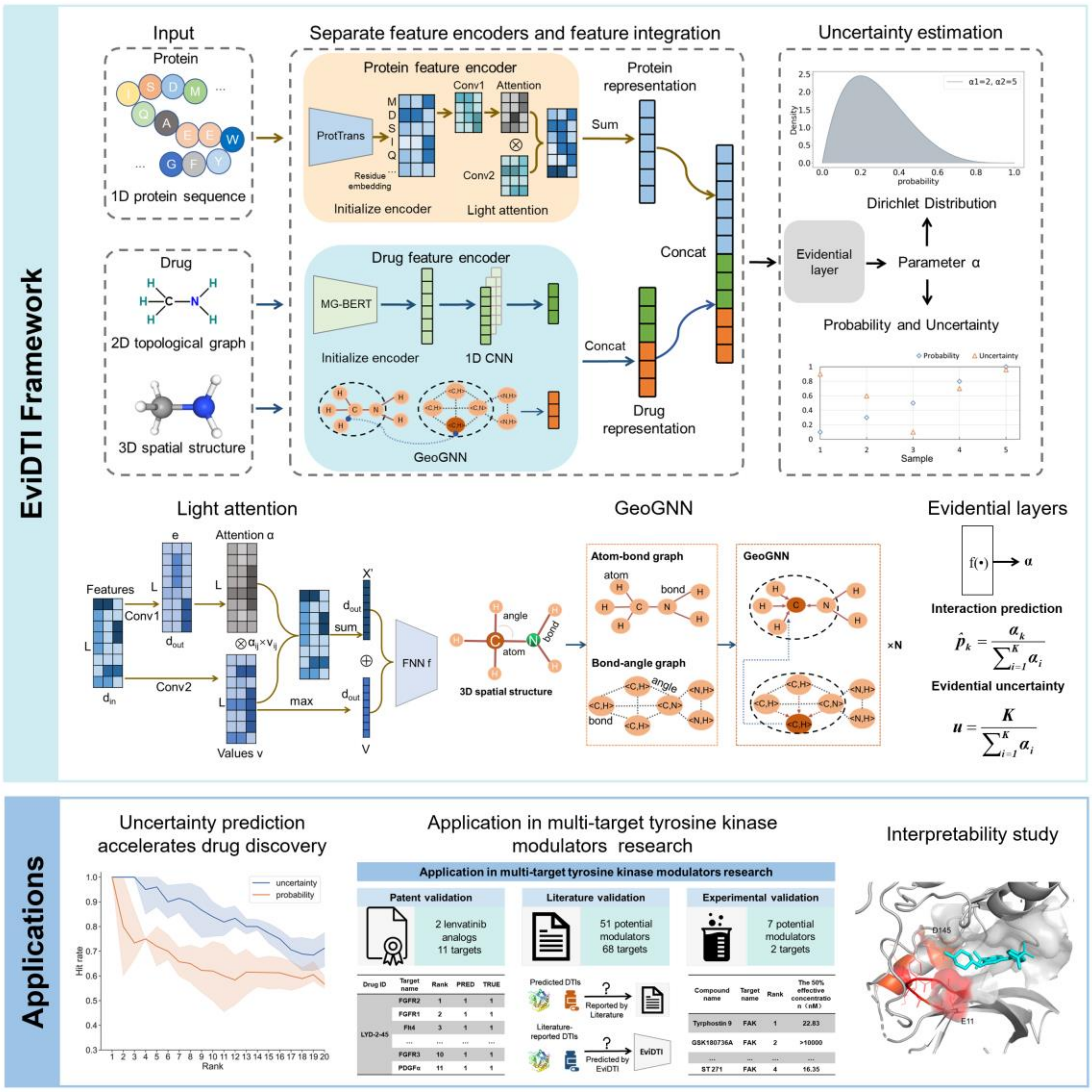


Figure R4 The flowchart of EviDTI.

9. In the EviDTI framework, drugs are represented using a fusion of 2D and 3D embeddings, while proteins are represented using only pre-trained sequence-based (1D) embeddings. Considering that tools such as AlphaFold2 (AF2) and AlphaFold3 (AF3) have been able to predict the 3D structure of proteins more accurately, what are the reasons and advantages of this choice?

Response:

Thank you for raising this important point regarding protein representation. While AlphaFold2/3 (AF2/AF3) have revolutionized protein structure prediction, we intentionally chose sequence-based (1D) embeddings for proteins in EviDTI for the

following reasons:

(1) Structural Hallucinations of disordered regions

While AF2/AF3 predicts globular proteins very accurately, the prediction of intrinsically disordered regions (IDRs) or membrane proteins remains error-prone[6,7]. These errors propagate into the 3D embedding, compromising the reliability of the model.

(2) Efficiency limitations of mass screening

Converting protein sequences to 3D coordinates via AF2 requires~10 GPU-minutes per protein[8], which is prohibitive for large-scale virtual screening (e.g., 1M+ pairs in our experiments). EviDTI's 1D embeddings can be computed in ~1 GPU-second per protein using pre-trained ProtTrans.

(3) Sufficient Information of Pretrained 1D Embeddings

Recent studies show that protein language models (pLMs) encode functional site information (e.g., binding pockets) in their 1D embeddings, even without explicit 3D supervision.

We acknowledge that 3D structural information can improve specific application scenarios. Therefore, in the discussion section, we provide an outlook on how 3D structural information can be incorporated into models. (The corresponding revision can be seen Line 574-583 on Page 33-34 of the revised manuscript)

10. Several grammatical errors were observed (e.g., inconsistent tenses in the abstract). A thorough proofreading is recommended.

Response:

Thank you for your meticulous attention to detail and for identifying the grammatical inconsistencies in our manuscript. We sincerely apologize for these oversights and have taken comprehensive steps to ensure the clarity and correctness of the text.

Reference：

1. Wang, X. *et al.* Multitask joint strategies of self-supervised representation learning on biomedical networks for drug discovery. *Nature Machine Intelligence* **5**, 445–456 (2023).
2. Lin, Y. *et al.* Uncertainty quantification via a memristor Bayesian deep neural network for risk-sensitive reinforcement learning. *Nature Machine Intelligence* **5**, 714–723 (2023).
3. Zhao, Q., Zhao, H., Zheng, K. & Wang, J. HyperAttentionDTI: improving drug–protein interaction prediction by sequence-based deep learning with attention mechanism. *Bioinformatics* **38**, 655–662 (2022).
4. Zeng, X., Chen, W. & Lei, B. CAT-DTI: cross-attention and Transformer network with domain adaptation for drug-target interaction prediction. *BMC Bioinformatics* **25**, 141 (2024).
5. Zhang, S. *et al.* AIGO-DTI: Predicting Drug–Target Interactions Based on Improved Drug Properties Combined with Adaptive Iterative Algorithms. *J. Chem. Inf. Model.* **64**, 4373–4384 (2024).
6. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
7. Ahdritz, G. *et al.* OpenFold: retraining AlphaFold2 yields new insights into its learning mechanisms and capacity for generalization. *Nature Methods* **21**, 1514–1524 (2024).
8. Mirdita, M. *et al.* ColabFold: making protein folding accessible to all. *Nature Methods* **19**, 679–682 (2022).

Reviewer #2:

In the manuscript, Zhou et al. developed an evidential deep learning tool to predict drug-target interactions (EviDTI). Their method quantifies uncertainty for its predictions. The EviDTI framework has components for a protein encoder which relies on primary sequence of the proteins, a component for a drug feature encoder which employs both 2D and 3D features of the molecules, and a component for evidence which includes uncertainty quantification. They then conducted a case study with EviDTI and tyrosine kinase inhibitors, followed by in vitro screening to test some of the predications.

By the authors' own admission, the performance of EviDTI is not impressive. Correspondingly, many of its high probability/low uncertainty ligands did not translate to these molecules affecting enzyme activity in vitro. Nevertheless, upon vadiation with DrugBank, Davis, and KIDA datasets, EviDTI outcompeted contemporary DTI models in several parameters, especially in accuracy, precision, and Matthews correlation coefficient. Having estimation of uncertainty in their framework is an advancement compared to other DTI models of drug-protein interaction, and this provides a measure of confidence.

The work generally supports the authors' conclusions. Inclusion of certain data and analysis would improve the story.

We are deeply grateful for your comprehensive and detailed review of our manuscript. All of your suggestions are valuable in improving our work and making the results more convincing.

Our motivation for developing EviDTI as follows. Although existing DTI prediction methods have demonstrated high accuracy on benchmark datasets, the results in practical applications are often unsatisfactory. A significant challenge is the lack of reliability of results with high predictive probabilities, which often leads to excessive false positives and failure of the experiment. This study proposes a novel DTI prediction framework, EviDTI, which innovatively integrates uncertainty quantification and aims to improve the utility and reliability of computational models.

In a comprehensive comparison with 11 DTI prediction models, EviDTI outperformed these methods on most metrics. More importantly, EviDTI is able to effectively identify and prioritize high-confidence DTIs for experimental validation by utilizing well-calibrated uncertainty information. The experimental validation of interactions between two tyrosine kinase targets and seven potential tyrosine kinase modulators (totally 14 potential DTIs) was performed. Among these 14 pairs of predicted DTIs, 5 pairs showed binding affinity, and 4 of them reached the nanomolar level of binding affinity. These results demonstrate that EviDTI is a powerful tool that can accelerate the translation process from theoretical models to actual drug development.

In accordance with your suggestions, we have revised the manuscript from four aspects (**Part 1-4 below**).

1) Redesign of the literature validation section.

The original design of Table S6 was for the validation of DTIs with low uncertainty through the literature. For those DTIs where we could not find evidence in the literature, we have not listed the targets reported in the literature for small molecules in those DTIs. However, the column name "Targets Reported in the Literature" in the original table may cause confusion. Therefore, we redesigned the literature-based validation to verify the performance of EviDTI from 3 aspects, and reorganized Table S6 (expanded to **Supplementary Table S7-S9** in the revised Supplementary Information, Page 10-12).

a. Validation of EviDTI's performance based on the literature in 3 aspects. First, we validated how many predicted DTIs with the lowest uncertainty score have been reported in literature (Table R1, corresponding to **Supplementary Table S7** in the revised Supplementary Information Page 10). Second, we validated how many literature-reported DTIs between tyrosine kinase modulators and tyrosine kinase targets are predicted by EviDTI (Table R2, corresponding to **Supplementary Table S8** in the revised Supplementary Information Page 11). Third, we validated how many DTIs between drugs in the top 10 predicted DTIs (with the lowest uncertainty score) and their literature-reported targets could be predicted by EviDTI. (Table R3, corresponding to **Supplementary Table S9** in the revised Supplementary Information Page 12).

b. Correction of the target of Tolimidone. A literature review confirmed that the

reported target of Tolimidone should be Lyn kinase (originally mislabeled as SRC), and the relevant data have been corrected in the table in parallel with the main text. (**Supplementary Table S7-9** in the revised Supplementary Information, Page 10-12)

2) Incorporation of new discussions on limitations you proposed.

In the Discussion section, we explore in detail the existing limitations of EviDTI, including:

a. The inability to distinguish whether a small molecule is an inhibitor or an activator (Page 33 Line 559-564 in the revised manuscript)

b. The difficulty of yield distinguished probabilities and uncertainty for interactions between compounds and a target and its mutants. (Page 33 Line 565-573 in the revised manuscript)

3) Correction of inaccurate description, data analyses and interpretations

a. Harmonization of terminology. For consistency of terminology, we unified **'50 percent effective concentration'** instead of **'EC_{50}'** and **'IC_{50}'**. Given that the model was unable to determine whether a drug was acting as an activator or an inhibitor, it would therefore be an inaccuracy to refer to all drugs as inhibitors. In addition, in vitro experiments have shown that two of the five drugs identified by the model are actually activators and three are inhibitors. The term '50% effective concentration' refers to the concentration of a compound required to achieve 50% of the maximum effect in a bioassay. This term is more appropriate for describing the effect of a compound on enzyme activity, either activation or inhibition.

b. Modification of the y-axis of the Figure5b-f. We modified the labels for the y-axis of Figures 5b-f to "**Relative activity to control assay**" and modified the range of values from **0-1 to 0-100%.** In addition, to improve the readability of the main manuscript, we moved Figures 5b-f to the Supplementary Information and replaced these figures with two tables (corresponding to **Figure 5b-c** in the revised manuscript Page 28). These tables present the '50% effective concentration' of the drug on the two targets, as well as the mean and standard deviation of the three replicate experiments. (corresponding to **Supplementary Figure S7** in the revised Supplementary Information Page 14)

c. In **Table S7** (corresponding to **Supplementary Table S10** in the revised Supplementary Information Page 13), we added information about whether compounds are activators or inhibitors of the enzyme based on the results of the ADP-Glo™ kinase assay.

d. Harmonization of units. We harmonized the units of all the 50% effective concentration in the manuscript to nM.

e. Newly added error bars of concentration-activity curve data. We added error bars representing the standard deviation for the ADP-Glo assay and explained them accordingly in the legend. (**Supplementary Figure S7** in the revised manuscript Supplementary Information Page 14)

f. Statement of values in Table 1-3. We specified that the underlined values in Table 1-3 are the second highest values and state that the values in parentheses are standard deviations. (Page 11 Line 193, Page 12 Line 198 and Page 13 Line 204 in the revised manuscript)

g. Statement of error bars in Figure 2. We stated the nature of the error bars as standard deviations in legend of Figure 2. (Page 15 Line 242 in the revised manuscript)

h. Correction of descriptions in performance comparison section. The incorrect use of descriptions in performance comparison has been corrected by replacing "improved" with "outperformed". (Page 9 Line 177-180 in the revised manuscript)

4) Enhancement of methodological components.

a. To validate the accuracy of the enzyme system, we supplemented the positive control results with the 50% effective concentrations of selected positive controls. These values were compared with literature-reported values. (Page 27 Line 453-466 in the revised manuscript)

b. In the Methods section, we added comprehensive details about all reaction conditions. This includes the carrier used for inhibitor dilution, the final concentration of the carrier in the enzyme assay, and instructions for its application in the negative control assay (Page 47-48 Line 838-862 in the revised manuscript).

Points

1. In line 188, the authors claim EviDTI improved accuracy, precision, etc. over the baseline model. This should likely be 'outperformed' baseline model or something similar. The methods don't describe combining EvitDTI with other models, just comparisons.

Response:

Thank you for your valuable comments on our manuscript. We have made changes to the manuscript to use the term "outperformed" rather than "improved". The manuscript has been revised as follows: "Specifically, on the KIBA dataset, EviDTI outperformed the best baseline model by 0.6% in accuracy, 0.4% in precision, 0.3% in MCC, 0.4% in F1 score, and 0.1% in AUC. On the Davis dataset, EviDTI exceeds 0.8% in accuracy, 0.6% in precision, 0.9% in MCC, 2% in F1 score, 0.1% in AUC, and 0.3% in AUPR." (The corresponding revision can be seen Line 177-180 on Page 9-10 of the revised manuscript)

2. In Table S6, the authors seem to omit 'known targets' of the molecules and the relevant citations, except for the target and reference (target of Lyn, an SRC family tyrosine kinase) that nearly corroborate EviDTI's prediction (target of SRC).

a. Tyrphostin 9 has a target reported in the literature (PDFGR), and this is a peculiar omission from Table S6, when the authors cite Kandandapani et al. (Ref 52 in the main manuscript file). Ref 52 and citations within mention PDFGR.

b. In fact, the other three compounds in Table S6 have targets reported in the literature, or they wouldn't have been in the kinase library. These should be included, even if they do not match EviDTI's predictions.

c. In addition to inclusion in Table S6, it would be appropriate to also address whatever these predictions are in the Results text.

Response:

We sincerely appreciate your critical review of data integrity. We sincerely apologize for the misrepresentation in Table S6. The original Table S6 was dedicated to

verifying whether DTIs with the lowest uncertainty are reported in the literature. However, the column name "Targets Reported in the Literature" in the original table may cause confusion.

Therefore, we redesigned the literature-based validation to verify the performance of EviDTI from 3 aspects. We reorganized Table S6 (**Part 0 below, expanded to Table S7-S9**), and addressed these predictions in the results section.

First, we validated how many predicted DTIs with the lowest uncertainty score have been reported in literature. (**Part 1** below, **Table R1**, corresponding to the **Supplementary Table S7** in revised supplementary information, Page 10).

Second, we validated how many literature-reported DTIs between tyrosine kinase modulators and tyrosine kinase targets are predicted by EviDTI. (**Part 2 below, Table R2，** corresponding to the **Supplementary Table S8** in revised supplementary information, Page 11).

Third, we validated how many DTIs between drugs in the top 10 predicted DTIs and their literature-reported targets could be predicted by EviDTI. (**Part 3 below, Table R3,** corresponding to the **Supplementary Table S9** in revised supplementary information, Page 12).

(0) Reorganized Table S6. In the new version of Table S6 (as shown in **Table R1**, corresponding to the **Supplementary Table S7** in revised supplementary information, Page 10), we removed "Literature-reported targets" column and added a new column named "Consistent with Literature" to minimize confusion. In addition, we have updated Table S6 to focus on the top 10 DTIs predicted by EviDTI.

(1) Validated how many predicted DTIs with lowest uncertainty score have been reported in literature. As shown in **Table R1 (**corresponding to the **Supplementary Table S7** in revised supplementary information, Page 10), among the top 10 predicted DTIs, 2 DTIs were validated in the literature (Interaction between Flumatinib mesylate and c-Kit [Rank 6, Uncertainty value=0.0419]; Interaction between Flumatinib mesylate and Bcr-Abl [Rank 10, Uncertainty value=0.0429]). (The corresponding revision can be seen Line 426-431 on Page 25 of the revised manuscript)

Table R1 Top 10 predicted interactions with the lowest uncertainty between tyrosine kinase modulators and tyrosine kinase targets

Compound	Target	Rank	Probability (Uncertainty)	Consistency with literature
RG 14620	SRC	1	98.37 (0.0327)	-
Tyrphostin 9	KIT/ SCFR	2	98.14 (0.0373)	-
GSK180736A	KIT/ SCFR	3	98.08 (0.0384)	-
NS309	SRC	4	98.03 (0.0393)	-
Tolimidone	SRC	5	97.94 (0.0412)	-
Flumatinib mesylate	**c-Kit**	**6**	**97.91 (0.0419)**	Zhao, J. *et al.*[1]
BIX02188	KIT/ SCFR	7	97.90 (0.0421)	-
Tyrphostin 9	VEGFR3	8	97.89 (0.0421)	-
GSK180736A	FLT3	9	97.85 (0.0429)	-
Flumatinib mesylate	**Bcr-Abl**	**10**	**97.85 (0.0429)**	Zhao, J. *et al.*[1]

(2) Validated how many literature-reported DTIs between tyrosine kinase modulators and tyrosine kinase targets are predicted by EviDTI. To conduct this validation, a total of 27 DTIs were collected from literatures between 68 tyrosine kinase targets and 51 tyrosine kinase modulators. The 27 DTIs and 68 tyrosine kinase targets have been obtained from multiple literature sources. The 51 tyrosine kinase modulators were obtained from the Targetmol-Tyrosine kinase modulators library.

The results of the predictions are shown in **Table R2 (**corresponding to the **Supplementary Table S8** in revised supplementary information Page 11). Among the 27 known DTIs, 21 DTIs were predicted by EviDTI. 10 out of 21 predicted DTIs have high confidence (the uncertainty score less than 0.1). This finding suggests EviDTI's potential in the discovery of tyrosine kinase modulators. (The corresponding revision

can be seen Line 431-438 on Page 26 of the revised manuscript)

Table R2 Predicted probabilities and uncertainties of literature-reported DTIs between tyrosine kinase modulators and tyrosine kinase targets

Compound	Target	Probability	Uncertainty	Predicted results	Reference
Flumatinib mesylate	c-Kit	97.91	0.0419	1	Zhao, J. *et al.*[1]
Flumatinib mesylate	Bcr-Abl	97.85	0.0429	1	Zhao, J. *et al.*[1]
5'-Fluoroindirubinoxime	FLT3	97.70	0.0461	1	Choi, S. J. *et al.*[2]
MNS	Src	97.46	0.0508	1	Wang *et al.*[3]
SB1317 hydrochloride	FLT3	97.39	0.0523	1	William, A. D. *et al.*[4]
Tyrphostin 9	PDGFR	96.29	0.0742	1	Kandandapani, S. *et al.*[5]
Edicotinib	CSF-1R	96.16	0.0769	1	Mancuso, R. *et al.*[6]
Edicotinib	FLT3	95.74	0.0852	1	Genovese, M. C. *et al.*[7]
XL228	Lyn	95.37	0.0926	1	Clary, D. O. *et al.* [8]
Ilginatinib maleate	JAK2	95.25	0.0949	1	Nakaya, Y. *et al.*[9]
MNS	Syk	94.12	0.1176	1	Wang *et al.*[3]
SB1317 hydrochloride	JAK2	94.11	0.1178	1	William, A. D. *et al.*[4]
SSR128129E	FGFR	93.71	0.1258	1	Bono, F. *et al.*[10]
MCB613	SRC	93.66	0.1268	1	Wang, L. *et al.*[11]
Su1498	VEGFR2	93.14	0.1372	1	Boguslawski, G. *et al.*[12]
(E/Z)-Zotiraciclib	JAK2	93.12	0.1375	1	Khalid Pasha *et al.*[13]
Flumatinib mesylate	PDGFRβ	92.29	0.1541	1	Zhao, J. *et al.* [1]
FLLL32	JAK2	89.75	0.2049	1	Bill, M. A. *et al.*[14]
(E/Z)-Zotiraciclib	FLT3	89.10	0.2181	1	Khalid Pasha. *et al.*[13]
MNS	FAK	83.61	0.3278	1	Wang *et al.*[3]
Degrasyn	JAK2	74.90	0.5019	1	Bartholomeusz. *et al.*[15]
Tyrphostin AG 879	TrKA	7.29	0.1457	0	Levitzki, A. *et al.*[16]
Tolimidone	LYN	5.73	0.1145	0	Saporito, M. S. *et al.*[17]
AG1024	IGF-1R	3.13	0.0627	0	Párrizas, M. *et al.*[18]
Tanshinone IIA	VEGFR2	2.93	0.0585	0	Xie, J. *et al.*[19]
ZM 39923 HCl	JAK3	2.64	0.0528	0	Zhang, Z. *et al.*[20]
Pralsetinib	RET	1.17	0.0233	0	Subbiah, V. *et al.*[21]

Note: A predicted result of 1 indicates that the model considers the drug and target to interact, while a predicted result of 0 indicates that the model considers the drug and target to have no interaction.

(3) Validated how many DTIs between drugs in top 10 predicted DTIs and their literature-reported targets could be predicted by EviDTI. We collected drugs in the top 10 predicted DTIs and their all literature-reported targets, totally 10 DTIs between these drugs and their all targets. The results of these predictions are displayed in **Table R3 (**corresponding to the **Supplementary Table S9** in revised supplementary information Page 12).

Among these 10 DTIs, 7 DTIs were predicted by EviDTI, 4 out of 7 predicted DTIs have high confidence (the uncertainty score less than 0.1). It suggests EviDTI's potential in the discovery of tyrosine kinase modulators. (The corresponding revision can be seen Line 438-444 on Page 26 of the revised manuscript)

Table R3 Predicted probabilities and uncertainties of interactions between drugs in the top 10 DTIs and their literature-reported targets

Compound	Targets reported in the literature	Probability (Uncertainty)	Predicted results
RG 14620	EGFR[22]	96.52 (0.0696)	1
Tyrphostin 9	PDGFR[23]	96.29 (0.0741)	1
GSK180736A	GRK[24]	2.08 (0.0415)	0
	ROCK1[25]	89.19 (0.2163)	1
NS309	Potassium Channel[26]	93.89(0.1222)	1
Tolimidone	LYN[27]	5.73 (0.1145)	0
Flumatinib mesylate	c-Kit[1]	97.91 (0.0419)	1
	Bcr-Abl[1]	97.85 (0.0429)	1
	PDGFR[1]	92.29 (0.1541)	1
BIX02188	MEK5[28]	12.26 (0.0245)	0

Note: A predicted result of 1 indicates that the model considers the drug and target to interact, while a predicted result of 0 indicates that the model considers the drug and target to have no interaction.

It is hoped that these modifications will provide a more comprehensive picture of

the predictive power and limitations of EviDTI, and that readers will gain a clearer understanding of the experimental validation and data analysis process. We would like to reiterate our gratitude for the constructive comments that have been provided, as they play a pivotal role in enhancing the quality and robustness of our work.

3. The authors say in Table S6 that the target for tolimidone in the literature is "SRC", when they cite a paper (Saporitio et al, Ref 51 in the main manuscript) that identifies the target as Lyn, a SRC family protein, but not SRC itself. If the predications from EviDTI are for kinase families, rather than specific kinase targets, this should be made clearer throughout the entire paper.

Response:

Thank you for your important corrections to our study. Regarding the inaccurate description of the Tolimidone target in Table S6 (original manuscript), we acknowledge that this was an oversight on our part. Therefore, we have corrected this description in the revised **Supplementary Table S7-9** **(Table R1-3,** Page 10-12 in revised supplementary information**)** to clearly state that the target of Tolimidone is Lyn.

4. Table S6 should include probability and uncertainty values from EviDTI for the literature targets. I.e. How certain was EviDTI that tyrphostin 9 was or wasn't a ligand for PDFGR?

Response:

We have provided probability and uncertainty values from EviDTI for the DTIs reported in the literature in 3 aspects, as discussed in the second point.

First, we validated how many predicted DTIs with the lowest uncertainty score have been reported in literature. (**Part 1** below, **Table R1**, corresponding to the **Supplementary Table S7** in revised supplementary information, Page 10).

Second, we validated how many literature-reported DTIs between tyrosine kinase modulators and tyrosine kinase targets are predicted by EviDTI. (**Part 2** below, Table **R2,** corresponding to the **Supplementary Table S8** in revised supplementary information, Page 11). In particular, as shown in **Table R2**, the probability of DTI

between Tyrphostin 9 and PDFGR is 96.29%, with an uncertainty of 0.0742, indicating a high confidence in the prediction.

Third, we validated how many DTIs between drugs in the top 10 predicted DTIs and their literature-reported targets could be predicted by EviDTI. (**Part 3** below, **Table R3,** corresponding to the **Supplementary Table S9** in revised supplementary information, Page 12).

(1) Validated how many predicted DTIs with lowest uncertainty score have been reported in literature. As shown in **Table R1 (**corresponding to the **Supplementary Table S7** in revised supplementary information, Page 10), among the top 10 predicted DTIs, 2 DTIs were validated in the literature (Interaction between Flumatinib mesylate and c-Kit [Rank 6, Uncertainty value=0.0419]; Interaction between Flumatinib mesylate and Bcr-Abl [Rank 10, Uncertainty value=0.0429]). (The corresponding revision can be seen Line 426-431 on Page 25 of the revised manuscript)

(2) Validated how many literature-reported DTIs between tyrosine kinase modulators and tyrosine kinase targets are predicted by EviDTI. To conduct this validation, a total of 27 DTIs were collected from literatures between 68 tyrosine kinase targets and 51 tyrosine kinase modulators. The 27 DTIs and 68 tyrosine kinase targets have been obtained from multiple literature sources. The 51 tyrosine kinase modulators were obtained from the Targetmol-Tyrosine kinase modulators library.

The results of the predictions are shown in **Table R2 (**corresponding to the **Supplementary Table S8** in revised supplementary information Page 11). Among the 27 known DTIs, 21 DTIs were predicted by EviDTI. 10 out of 21 predicted DTIs have high confidence (the uncertainty score less than 0.1). This finding suggests EviDTI's potential in the discovery of tyrosine kinase modulators. (The corresponding revision can be seen Line 431-438 on Page 26 of the revised manuscript)

(3) Validated how many DTIs between drugs in top 10 predicted DTIs and their literature-reported targets could be predicted by EviDTI. We collected drugs in the top 10 predicted DTIs and their all literature-reported targets, totally 10 DTIs between these drugs and their all targets. The results of these predictions are displayed in **Table R3 (**corresponding to the **Supplementary Table S9** in revised supplementary

information Page 12).

Among these 10 DTIs, 7 DTIs were predicted by EviDTI, 4 out of 7 predicted DTIs have high confidence (the uncertainty score less than 0.1). It suggests EviDTI's potential in the discovery of tyrosine kinase modulators. (The corresponding revision can be seen Line 438-444 on Page 26 of the revised manuscript)

Additionally, there are some discussions/conclusions that would strengthen their report. 1. EviDTI predicted inhibition of SRC by tolimidone, when the literature citation the authors provide to support the predictions identifies tolimidone as an activator of an SRC family kinase, not an inhibitor. With this discrepancy, coupled with the in vitro activation they found in assay experiments, the authors should discuss this limitation of their model: that identifying ligands in an inhibitor prediction may retrieve ligands which may be activators in a biological system.

Response:

Thank you very much for your important feedback on our study, we fully agree with you and feel the need to discuss this issue in detail in the manuscript. We added a new paragraph to the Discussion section that discuss this limitation.

Although EviDTI was designed with the objective of identifying potential DTIs, it is currently cannot determine whether the molecule has an activating or inhibitory effect on the target. This limitation underscores the need for further enhancements. To address this, we propose the following directions for future improvements, including introducing more molecular biochemical properties into the model, developing a multi-task learning framework capable of predicting activation and inhibition simultaneously. (The corresponding revision can be seen Line 559-564 on Page 33 of the revised manuscript)

2. The real-world utility of the model would be enhanced if the authors spoke to abilities of EviDTI to predict for known oncoprotein sequences of the tyrosine kinases. E.g. there are two well-known mutations of FLT3 in AML: Is EviDTI aware of these mutations? If so, is there data for this (does EviDTI predict binding of vodobatinib to

FLT3-ITD or FLT3-TKD)? If not, do the authors anticipate EviDTI to yield distinguished probabilities and uncertainty for compounds against FLT3 vs. FLT3-ITD vs. FLT3-TKD? Perhaps not, since protein data is 1D, but this should be addressed in the discussion. In an ideal world, an antineoplastic agent would inhibit the oncoproteins more than protooncogenic proteins.

Response:

We appreciate the insightful comments on our research. We validated that EviDTI cannot yield distinguished probabilities and uncertainty for interactions between compounds and FLT3 and its mutants (**Part 1 below**). Consequently, in the **Part 2 below**, we explore the limitations of the model and propose potential improvements.

(1) Validated whether EviDTI can give distinguished probabilities and uncertainties on the interactions between FLT3 wild-type and its mutants with compounds. We used EviDTI to obtain probability and uncertainty scores for interactions between FLT3 wild-type (FLT3-WT) and two common FLT3 mutants (FLT3-ITD (627E) and FLT3-TKD (D835E)) with two FLT3-targeted drugs (Vodobatenib and Tyrphostin 9).

Table R4 Predicted probabilities and uncertainties between Vodobatinib and Tyrphostin 9 and FLT3 and its mutants

Compound	Targets	Probability (Uncertainty)
Vodobatinib	FLT3-WT	97.67 (0.0461)
	FLT3-ITD[29]	97.61 (0.0472)
	FLT3-TKD[30]	97.71 (0.0453)
Tyrphostin 9	FLT3-WT	97.91 (0.0419)
	FLT3-ITD[29]	97.81 (0.0403)
	FLT3-TKD[30]	97.85 (0.0408)

As demonstrated in **Table R4**, the probability and to uncertainty score of interactions between Vodobatinib and Tyrphostin 9 and FLT3 and its mutants is very

close. The predictions given by EviDTI may not possess the capability to effectively distinguishing oncoproteins with protooncogenic proteins.

(2) Extended discussion of directions for future improvement

Since EviDTI relies on one-dimensional protein sequence information, its ability to capture the nuanced effects of mutations may be limited. We believe in incorporating more informative representations such as protein 3D structural information and pharmacologically perturbed transcriptomics data can more effectively address these limitations. (The corresponding revision can be seen Line 565-573 on Page 33 of the revised manuscript)

It is hoped that the analyses and discussions presented herein will facilitate a more comprehensive understanding of the current status and potential of EviDTI. Furthermore, it is anticipated that they will provide a clear direction and suggestions for improvement in subsequent studies.

Some concerns with data analysis and interpretation include the confusion around some of the data presented.

1. The effect of Tyrphostin 9 and Vodobatinib on FLT3 is indecipherable and seems to have conflicting explanations between the text and figures.

a. Line 397 suggests Tyrphostin 9 and Vodobatinib inhibited FLT3, though the term "EC50" is used instead of "IC50", and the dose response has an opposite trend and low doses gave the most inhibition.

b. If there was little inhibition at high concentration of Tryphostin 9 or Vodobatinib, but high inhibition at low concentration, the authors should address this in the text. （

c. If Tryphostin 9 and Vodobatinib were found to agonize, the text should reflect this, and the x-axes on Fig. 5e and 5f should reflect activation (or 'relative activity to control assays' or something similar to normalized activity from a negative enzymatic control) and not inhibition.

d. If these compounds are all inhibiting in a dose-dependent fashion as the text suggests,

Figure 5 should show this, and I would suggest abandoning “EC50” if all trends are inhibitory, for simplicity for the reader.

Response:

We are deeply grateful for your comprehensive and careful review of our manuscript. To address the inconsistencies between the figures and the text regarding the effects of Tyrphostin 9 and Vodobatinib on FLT3 kinase, we have made the following revisions based on the ADP-Glo™ Kinase Assay results.

(1) Correction of inaccurate descriptions

We hereby express our sincere apologies for this inaccuracy. The statement in Line 397 of the original manuscript—"Tyrphostin 9 and Vodobatinib inhibited FLT3"—was inaccurate. According to the dose-response trends, the upward trend in the dose-response curve suggested potential activation of FLT3 kinase activity. (**Figure R1**).

In the revised version, we have corrected Line 397 of the original manuscript to “Tyrphostin 9 and Vodobatinib activated FLT3 activity with the 50% effective concentration 1265.9 ± 244.6 nM and 406.8 ± 74.5 nM, respectively” (Line 463 page 27 in the revised manuscript).

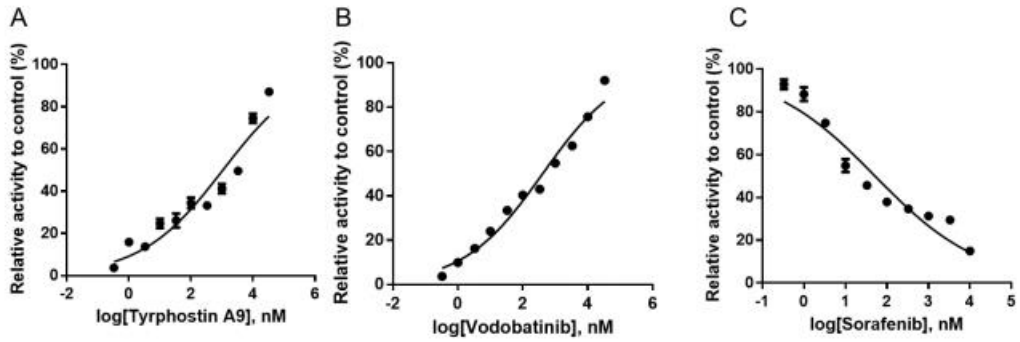


Figure R1 A-C Concentration-activity curves of Tyrphostin 9, Vodobatinib and sorafenib in the FLT3 kinase ADP-Glo assays, respectively. Sorafenib as positive control. Mean ± SEM of three independent experiments is shown (n=3).

(2) Correction of inaccurate Figure coordinates

Based on the ADP-Glo™ Kinase Assay results, Tyrphostin 9 and Vodobatinib exert an activating effect on FLT3 kinase. Consequently, the Y-axis of the

concentration-activity curves should not be labeled as inhibition.

In order to provide a more accurate depiction of the interaction between the compounds and the kinase, the y-axis label in all concentration-activity curves has been updated to "**relative activity to control assays**" (**Figure R2**). (corresponding to **Supplementary Figure S7** in the revised Supplementary Information, Page 14).

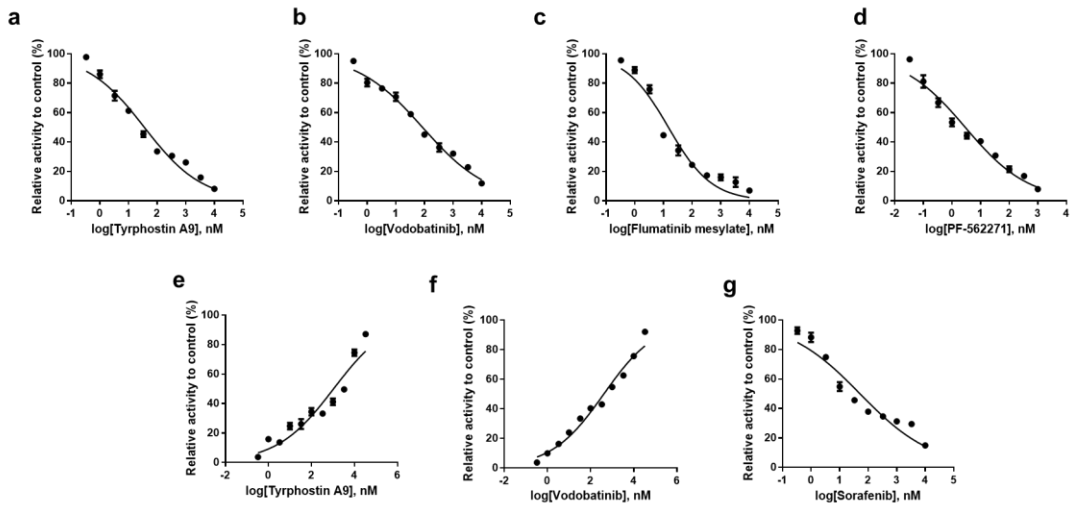


Figure R2 a-d Concentration-activity curves of Tyrphostin 9, Vodobatinib, Flumatinib and PF-562271 in the FAK kinase ADP-Glo assays, respectively. PF-562271 as positive control. Mean ± SEM of three independent experiments is shown(n=3). **e-g** Concentration-activity curves of Vodobatinib, Tyrphostin 9 and sorafenib in the FLT3 kinase ADP-Glo assays, respectively. Sorafenib as positive control. Mean ± SEM of three independent experiments is shown(n=3).

(3) Unification of description indicators for concentration-activity curves

In the ADP-Glo™ Kinase Assay, we found that certain compounds were able to activate the kinase, while others were able to inhibit it. To enhance the clarity of the results, we use the term "the 50% effective concentration" to describe the effect of compounds on the kinase, regardless of whether compounds are activating or inhibiting.

The 50% effective concentration (EC_{50}) is defined as the concentration of a compound required to achieve 50% of the maximum effect in a biological assay. In the literature, EC_{50} is commonly used to indicate the effect of compounds on enzyme activity, including both activation and inhibition. For example, '*Viramidine, the 3-*

carboxamidine derivative of ribavirin, was effective against a spectrum of influenza A (H1N1, H3N2 and H5N1) and B viruses in vitro, with the 50% effective concentration (EC_{50}) ranging from 2 to 32 μg/ml'[31]; '*AG7088 inhibited the replication of all HRV serotypes (48 of 48) tested with a mean 50% effective concentration (EC_{50}) of 0.023 μM*'[32]. In accordance with this principle, the "50% effective concentration" has been adopted as a means of describing the effect of the compounds on kinase activity.

The manuscript has been revised as follows (The corresponding revision can be seen Line 453-466 on Page 27 of the revised manuscript):

As shown in Fig. 5b, Tyrphostin 9 and Vodobatinib exhibited an inhibition of FAK activity with the 50% effective concentration of 35.7±3.4 nM and 85.7±8.2 nM, respectively. Also, Flumatinib mesylate inhibited FAK activity with a 50% effective concentration of 14.9± 2.1 nM (Supplementary Table S10a and Figure S7 a-d).

As shown in Fig. 5c, Tyrphostin 9 and Vodobatinib activated FLT3 activity with the 50% effective concentration 1265.9 ± 244.6 nM and 406.8± 74.5 nM, respectively (Supplementary Table S10b and Figure S7 e-g).

2. Further the effect of Tyrphostin 9 and Vodobatinib is also unclear in Table S6. It may be helpful to include in reach row in Table S6, whether the molecule's effect was activating or inhibitory.

Response:

Thank you for raising this important suggestion regarding the clarity of the effects of Tyrphostin 9 and Vodobatinib. In our previous response, we addressed the issue with Table S6. We infer that you'd like us to include the effect of molecules on kinases in Table S7. We added information on the activator or inhibitor status of the compounds in **Table R5a-b** based on the results of the ADP-Glo™ Kinase Assay. (Corresponding to **Supplementary Table S10** in the revised supplementary information Page 13)

Table R5a The 50% effective concentration of compounds with the lowest uncertainty against FAK

Compound	FAK			
	Rank	Probability (Uncertainty)	The 50% effective concentration (nM)[a]	The molecule's effect on FAK
Tyrphostin 9	1	97.69 (0.0461)	35.7±3.4	inhibitor
GSK180736A	2	97.65 (0.0469)	>10000	-
Vodobatinib	7	97.04 (0.0591)	85.7 ± 8.2	inhibitor
BIX02188	3	97.58 (0.0484)	>10000	-
PKG drug G1	5	97.22 (0.0555)	>10000	-
ST 271	4	97.31 (0.0537)	>10000	-
Flumatinib	6	97.18 (0.0564)	14.9 ± 2.1	inhibitor

[a] Values are the means of three replicates

Table R5b The 50% effective concentration of compounds with the lowest uncertainty against FLT3

Compound	FLT3			
	Rank	Probability (Uncertainty)	The 50% effective concentration (nM)[a]	The molecule's effect on FLT3
Tyrphostin 9	1	97.90 (0.0419)	1265.9±244.6	activator
GSK180736A	2	97.85 (0.0429)	>10000	-
Vodobatinib	3	97.69 (0.0461)	406.8±74.5	activator
BIX02188	4	97.60 (0.0480)	>10000	-
PKG drug G1	5	97.60 (0.0481)	>10000	-
ST 271	6	97.53 (0.0494)	>10000	-
Flumatinib	9	97.13 (0.0573)	>10000	-

[a] Values are the means of three replicates

3. Throughout page 23, IC50 values are reported in 'micromoles'. Besides not being reported as concentrations, if these values were intended to be 'micromolar' rather than 'micromoles', the values are off by an order of magnitude from the nanomolar values displayed in Figure 5b-d and Table S7.

Response:

Thank you for your valuable comments. We acknowledge the inaccuracies in our description of the drug concentration units. To correct these inaccurate descriptions, we have adopted the term "micromolar (nM)" as the uniform unit. We have refined the experimental methods and incorporated the updated version into the revised manuscript, as detailed below.

The manuscript has been revised as follows (The corresponding revision can be seen Line 453-466 on Page 27 of the revised manuscript):

As shown in Fig. 5b, Tyrphostin 9 and Vodobatinib exhibited an inhibition of FAK activity with the 50% effective concentration of 35.7±3.4 nM and 85.7±8.2 nM, respectively. Also, Flumatinib mesylate inhibited FAK activity with a 50% effective concentration of 14.9± 2.1 nM (Supplementary Table S10a and Figure S7 a-d).

As shown in Fig. 5c, Tyrphostin 9 and Vodobatinib activated FLT3 activity with the 50% effective concentration 1265.9 ± 244.6 nM and 406.8± 74.5 nM, respectively (Supplementary Table S10b and Figure S7 e-g).

4. Figure 5b-f implies the authors only achieved a maximum of 1% inhibition at the highest inhibitor concentrations.

a. If only 1% inhibition was achieved, that makes the results less impressive than the authors already admit.

b. Assuming 1 represents normalization or 100%, the graphs should indicate this, either by labeling the axis relative inhibition, or by keeping % on the axis and changing the values from 0-100 instead of 0-1.

Response:

Thank you so much for your careful check and your expert guidance. We feel sorry for our mistake of wrong y-axis of "inhibition (%) in Figure 5b-5f (original manuscript).

We have made revisions in two main aspects:

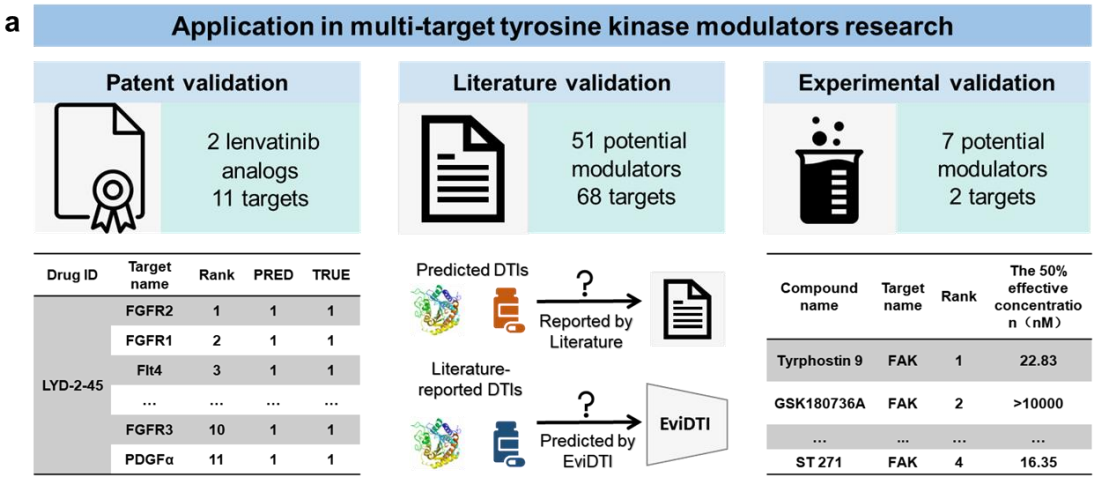
A. Correction of inaccurate Figure coordinates

For your suggestion, the appropriate value for the y-axis is "**relative activity to control (%)**", which represents the relative FAK or Flt3 kinase activity of the drug treatment compared with control group (DMSO). In the revised manuscript, the y-axis labelling and values have been corrected, and the data are shown in **Figure R2** (Corresponding to **Supplementary Figure S7** in the revised Supplementary Information Page 14). As shown in the revised figures, the effects of the compounds on the enzyme are dose-dependent.

B Modification of the y-axis value

To present the data more clearly, we changed the y-axis label to "**relative activity to control (%)**" and modified the value range from 0-1 to 0-100%, to more intuitively show the changes in effects.

In the revised manuscript, we have moved the revised figures to the **Supplementary Figure S7** (Page 14 in the revised Supplementary Information) and presented the effects of compounds on enzymes from triplicate independent experiments in **Figure 5** (As shown in **Figure R3**, Line 471 page 28 in the revised manuscript) within the main text.



b

FAK

Compounds	The 50% effective concentration (nM)
Tyrphostin A9	35.7 ± 3.4
Vodobatinib	85.7 ± 8.2
Flumatinib mesylate	14.9 ± 2.1
PF-562271	2.91 ± 0.47

c

FLT3

Compounds	The 50% effective concentration (nM)
Tyrphostin A9	1265.9 ± 244.6
Vodobatinib	406.8 ± 74.5
Sorafenib	51.7 ± 8.47

Fig. 5 Application of EviDTI in the discovery of multi-target tyrosine kinase modulators. a The validation framework on multi-target tyrosine kinase modulators. Initially, the validation was carried out using the data reported in the patents. Two lenvatinib analogs with 11 known targets were collected from the patents to conduct this validation. Subsequently, the validation was carried out using the data reported in the literature. The uncertainty score and probability of interactions between the 68 tyrosine kinase targets and 51 tyrosine kinase modulators were predicted via EviDTI. Finally, two targets of interest were selected from the 68 targets collected above. Based on the uncertainty for these two targets and 51 modulators, the interactions with lowest uncertainty between these two targets and the seven modulators were validated experimentally. b The 50% effective concentration of Tyrphostin 9, Vodobatinib, Flumatinib and PF-562271 in the FAK kinase ADP-Glo assays, respectively. PF-562271 as positive control. Mean ± SEM of three independent experiments is shown (n=3). c The 50% effective concentration of Vodobatinib, Tyrphostin 9 and sorafenib in the FLT3 kinase ADP-Glo assays, respectively. Sorafenib as positive control. Mean ± SEM of three independent experiments is shown (n=3).

5. In Tables 1-3, are the underlined values the second highest? If so, it would be helpful

if this was explained in the table legend. And if so, Table 2 Accuracy is missing an underlined value. If parenthetical values in Tables 1-3 are error values, this should be denoted, at least in Comparison methods in Methods, if not also in the table.

Thank you for your keen insight into the details of table data presentation. To response to your suggestions, we have made the following modifications:

(1) Statement of underlined values in Table 1-3. The underlined values in Tables 1-3 indicate the second-highest value for each metric. To ensure readability, we have added the following text to the table legend "*The best results are highlighted in bold, while the second-best results are underlined.*". (Page 11 Line 193, Page 12 Line 198 and Page 13 Line 204 in the revised manuscript)

(2) Completion of missing underlined values. We have added the missing underline in Table 2 and verified similar cases in other tables to ensure that all second highest values are correctly labelled. (Page 11 Line 193, Page 12 Line 198 and Page 13 Line 204 in the revised manuscript)

(3) **Statement of parenthetical values in Table 1-3.** The parenthetical values in Table 1-3 denote standard deviation (std). We have explained the meaning of the parenthetical values in the table note: "*Five independent replications of each method were performed (n=5). Data are expressed as means (std)*". (Page 11 Line 193, Page 12 Line 198 and Page 13 Line 204 in the revised manuscript)

In addition, we have provided further explanation in "*Comparison methods*" section: "*All methods were conducted with five repeated experiments, and the results are reported as the mean and standard deviations.*" (Page 42 Line 746 in the revised manuscript)

6. In Figure 2, the nature of the error bars should be denoted.

Thank you for your valuable comments on our study. The error bars in Figure 2 represent the standard deviation (std). We have added a statement to the legend of Figure 2 to the following text: "*Five independent replications of each method were performed (n=5). Data are expressed as means ± std.*" (The corresponding revision can be seen Line 242 on Page 15 of the revised manuscript)

7. A replicate statement exists in Table S7, but there should be one in Figure 2 legend, and IC50 and/or EC50 figures should have error bars, and corresponding explanation in the figure legend.

Response:

Thank you for your suggestion. We highly appreciate your valuable suggestions regarding the inclusion of replicate statements and error bars in the figures. To response to your suggestions, we have made the following modifications:

A. Newly Added of Error Bars in ADP-Glo Assay Figures: We included error bars in the ADP-Glo assays to visually represent the variability of the data (**Figure R2**).

B. Newly Added of Replicate Statement in Figure 5 Legends: We added a replicate statement to the legend of Figure 5, and added "Mean ± SEM of three independent experiments is shown (n=3)"in revised vision. (The corresponding revision can be seen Line 471 on Page 28 of the revised manuscript)

C. Statement of Error Bars in Figure 2: We added a statement to the legend of Figure 2: "*Five independent replications of each method were performed (n=5). Data are expressed as means ± std.*" (The corresponding revision can be seen Line 242 on Page 15 of the revised manuscript)

I will combine the soundness of the methodology with the detail in the methods required to reproduce the work, since some of my suggestions to methodology may have been done, but not conveyed in detail in the Methods section. The following comments all relate to the authors' enzymatic activity assay, which in their current state, would be difficult to reproduce without additional details.

1. Inhibition assays should be run alongside negative controls. It would seem the authors have done this, since their IC50s are normalized to presumably the negative control enzymatic activity. However inhibition assays should also be run with at least single-point positive controls for their system. A positive control of known inhibitor(s) of FAK and Flt3, like sorafenib (or alternatively two different positive controls for each enzyme) assayed alongside unknowns would strengthen the methodology.

Response:

Thank you for your valuable comments. To demonstrate the feasibility and operability of our experiment system, we used positive control drugs for the evaluation assays between FAK /FLT3 kinases and chemicals.

For the FAK kinase experiment, we selected PF-562271 as the positive control. In our experimental system, the EC_{50} value of this compound was 2.91±0.47 nM, which is consistent with the results reported in the literature (FAK :1.5 nM)[33]. **(Figure R4-A)** (The corresponding revision can be seen Line 453-456 on Page 27 of the revised manuscript)

For the FLT3 kinase experiment, we used sorafenib as the positive control. The EC50 value of sorafenib in our system was 51.7±8.47 nM, which is in line with the literature values (FLT3: 58 ± 20 nM)[34]. (**Figure R4-B**) (The corresponding revision can be seen Line 460-463 on Page 27 of the revised manuscript)

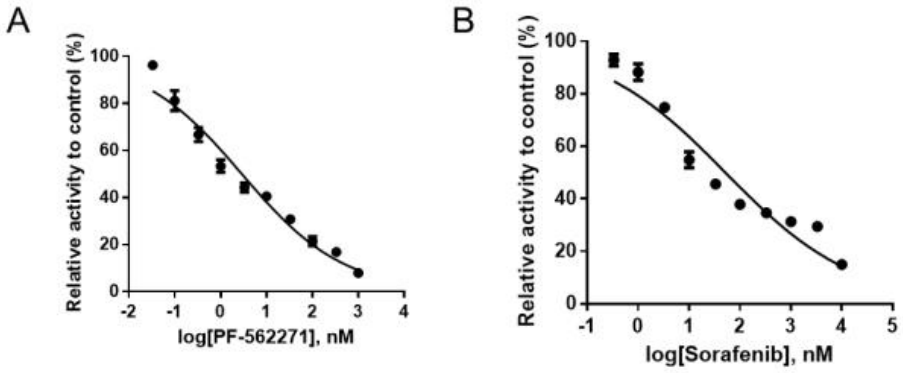


Figure R4 A The concentration-activity curves of PF-562271 in the FAK kinase ADP-Glo assays. Mean ± SEM of three independent experiments is shown(n=3). **B** Concentration-activity curves of Sorafenib in the FLT3 kinase ADP-Glo assays, respectively. Mean ± SEM of three independent experiments is shown(n=3).

2. Inclusion of IC50 values for their chosen positive controls within the Methods section would allow comparison to literature values and validate their enzymatic system.

Response:

Thank you for your valuable comments. We added positive controls to the revised manuscript.

For the FAK kinase experiment, we selected PF-562271 as the positive control. In our experimental system, the EC_{50} value of this compound was 2.91±0.47 nM, which is consistent with the results reported in the literature (FAK :1.5 nM)[33]. (The corresponding revision can be seen Line 453-456 on Page 27 of the revised manuscript)

For the FLT3 kinase experiment, we used sorafenib as the positive control. The 50% effective concentration of sorafenib in our system was 51.7±8.47 nM, which is in line with the literature values (FLT3: 58 ± 20 nM)[34]. (The corresponding revision can be seen Line 460-463 on Page 27 of the revised manuscript)

3. Inclusion of typical metabolism for negative control kinase assays for each FAK and Flt3 assays within the Methods section would also help validate their enzymatic system. I.e. What enzymatic activity corresponds to 0% inhibition, in concentration of substrate consumed or product generated (or concentration per reaction time per enzyme concentration). Representative values for relative light units at these activities can also be helpful to the reader.

Response:

Thank you for your comments. The reliability of enzyme activity assay is demonstrated through the use of positive and negative controls. To response your questions and suggestions, we respond from the following aspects.

(1) Description of positive and negative controls

For each experiment, 5% dimethyl sulfoxide (DMSO) was used as negative control, representing the substrate consumed by kinase without drug. For the FAK kinase experiment, we selected PF-562271 as the positive control, which has been reported to act as a reversible, potent, ATP-competitive FAK kinase inhibitor of (the 50% effective concentration =1.5 nM)[33]. For the FLT3 kinase experiment, we used sorafenib as the positive control, which has been shown to be a multi-kinase inhibitor with potent FLT3 kinase inhibition (the 50% effective concentration 50 = 57nM) [34].

In our enzyme activity assay, we found that PF-562271 significantly inhibited FAK kinase activity in a dose-dependent manner with the 50% effective concentration of 2.91 ± 0.47 nM. Additionally, we identified that sorafenib exhibited dose-dependent

FLT3 kinase inhibition (The 50% effective concentration = 51.7 ± 8.47 nM). The positive control results are consistent with those reported in the literature. These results demonstrated the feasibility and operability of our experiment system.

(2) Newly added Description of Method

To facilitate reading, we have revised the Methods section as follows (The corresponding revision can be seen Line 838-862 on Page 47-48 of the revised manuscript): In vitro kinase activity assays were conducted through ADP-Glo assays (#V9101) provided by Promega. The protocol for the FAK and FLT3 assay is described as follows: Enzyme, substrate, ATP, and compounds were diluted in 1X reaction buffer composed of 40 mM Tris (pH 7.5), 2 mM MnCl, 100 μM sodium vanadate. In a 384-well low-volume plate, 1 μl of the compound at indicated doses or 5% dimethyl sulfoxide (DMSO), 2 μl of FAK or FLT3 enzyme (15 ng/well), and 2 ul of substrate/ATP mix (final concentration: 20 μM ATP) were added to each well. The plate was then incubated at 25℃ for 60 min to allow for kinase activity. Following the enzymatic reaction, 5 μl of ADP-GloT Reagent was added to each well, and the plate was incubated at 25℃ for an additional 40 min. Subsequently, 10 μl of Kinase Detection Reagent was added to convert ADP to ATP and introduce luciferase and luciferin to detect ATP, and the plate was incubated for a final 30 min at 25°C. Luminescence was recorded with an integration time of 0.5 s using the SpectraMax iD3 instrument (Molecular Devices).

The 50% effective concentration was calculated using Prism 8 by fitting the following equation:

$$Y = Bottom + (Top - Bottom)/(1 + 10^{(logEC50-X)} \times HillSlope)$$

where X is a log of concentration, Y is a response, and top and bottom are the responses of controls, each assay was repeated at least three times, and we computed the mean and standard deviation for the values.

(3) Newly added the Raw Experimental Values for Compounds

The values for each compound with indicated dosed were suppled in Supplement

excel, which could can be helpful to the reader.

4. There are several components of reaction conditions omitted, including:

Response:

Thank you for your valuable comments. We have refined the experimental methods and incorporated the updated version into the latest manuscript. (The corresponding revision can be seen Line 838-862 on Page 47-48 of the revised manuscript)

a. The concentration of tyrosine kinase, either FAK or Flt3: FAK or FLT3 enzyme (15 ng/well)

b. The identity of the protein or peptide substrate used (or a statement that autophosphorylation was measured in the absence of protein or peptide substrate), and the concentration of that substrate if applicable.

2 µl of FAK or FLT3 enzyme (15 ng/well), and 2 ul of substrate/ATP mix (final concentration: 20 µM ATP)

c. The concentration of ATP used: ATP mix (final concentration: 20 µM ATP)

d. Reaction volume: 20µl

e. Reaction vessel/plate: a 384-well low-volume plate

f. Reaction duration: the plate was incubated at 25°C for an additional 40 min.

g. Reaction temperature: the plate was incubated at 25°C for an additional 40 min.

h. Buffer conditions, especially pH and magnesium concentration: Enzyme, substrate, ATP, and compounds were diluted in 1X reaction buffer composed of 40 mM Tris (pH 7.5), 2 mM MnCl, 100 µM sodium vanadate.

i. The instrument used to measure light generated by ADP-Glo: Luminescence was recorded with an integration time of 0.5 s using the SpectraMax iD3 instrument (Molecular Devices)

5. What vehicle was used for inhibitor dilutions, as well as the final concentration of this vehicle in enzyme assays and a statement that this concentration was used in negative control assays are omitted.

Response:

Thank you for your suggestion. We acknowledge that we omitted some experimental details that are crucial for reproducing the experiments. For the ADP-Glo assays, we used DMSO as the solvent, and the concentration of this vehicle in the enzyme assays was 15 ng/well. These contents are reflected in the Methods section of the revised version. (The corresponding revision can be seen Line 838-862 on Page 47-48 of the revised manuscript)

Reference

1. Zhao, J. *et al.* Flumatinib, a selective inhibitor of BCR-ABL/PDGFR/KIT, effectively overcomes drug resistance of certain KIT mutants. *Cancer Science* **105**, 117–125 (2014).
2. Choi, S. J. *et al.* Indirubin derivatives as potent FLT3 inhibitors with anti-proliferative activity of acute myeloid leukemic cells. *Bioorganic & Medicinal Chemistry Letters* **20**, 2033–2037 (2010).
3. Wang, W.-Y., Wu, Y.-C. & Wu, C.-C. Prevention of Platelet Glycoprotein IIb/IIIa Activation by 3,4-Methylenedioxy-β-Nitrostyrene, A Novel Tyrosine Kinase Inhibitor. *Molecular Pharmacology* **70**, 1380–1389 (2006).
4. William, A. D. *et al.* Discovery of kinase spectrum selective macrocycle (16E)-14-methyl-20-oxa-5,7,14,26-tetraazatetracyclo[19.3.1.1(2,6).1(8,12)]heptacosa-1(25),2(26),3,5,8(27),9,11,16,21,23-decaene (SB1317/TG02), a potent inhibitor of cyclin dependent kinases (CDKs), Janus kinase 2 (JAK2), and fms-like tyrosine kinase-3 (FLT3) for the treatment of cancer. *J Med Chem* **55**, 169–196 (2012).
5. Kandandapani, S., Ridzwan ,Nor Farrah Wahidah, Mohamad ,Saharuddin B. & and Tayyab, S. Exploring the interaction between tyrphostin 9 and human serum albumin using biophysical and computational methods. *Journal of Biomolecular Structure and Dynamics* **38**, 4134–4142 (2020).
6. Mancuso, R. *et al.* CSF1R inhibitor JNJ-40346527 attenuates microglial proliferation and neurodegeneration in P301S mice. *Brain* **142**, 3243–3264 (2019).
7. Genovese, M. C. *et al.* Results from a Phase IIA Parallel Group Study of JNJ-40346527, an Oral CSF-1R Inhibitor, in Patients with Active Rheumatoid Arthritis despite Disease-modifying Antirheumatic Drug Therapy. *J Rheumatol* **42**, 1752 (2015).
8. Clary, D. O. *et al.* Abstract C192: Characterization of the target profile of XL228, a multi-targeted protein kinase inhibitor in phase 1 clinical development. *Molecular Cancer Therapeutics* **8**, C192–C192 (2009).
9. Nakaya, Y. *et al.* Efficacy of NS-018, a potent and selective JAK2/Src inhibitor, in primary cells and mouse models of myeloproliferative neoplasms. *Blood Cancer*

Journal **1**, e29–e29 (2011).

10. Bono, F. *et al.* Inhibition of Tumor Angiogenesis and Growth by a Small-Molecule Multi-FGF Receptor Blocker with Allosteric Properties. *Cancer Cell* **23**, 477–488 (2013).
11. Wang, L. *et al.* Characterization of a Steroid Receptor Coactivator Small Molecule Stimulator that Overstimulates Cancer Cells and Leads to Cell Stress and Death. *Cancer Cell* **28**, 240–252 (2015).
12. Boguslawski, G., McGlynn, P. W., Harvey, K. A. & Kovala, A. T. SU1498, an Inhibitor of Vascular Endothelial Growth Factor Receptor 2, Causes Accumulation of Phosphorylated ERK Kinases and Inhibits Their Activity in Vivo and in Vitro*. *Journal of Biological Chemistry* **279**, 5716–5724 (2004).
13. Khalid Pasha, M. *et al.* Preclinical Metabolism and Pharmacokinetics of SB1317 (TG02), a Potent CDK/JAK2/FLT3 Inhibitor. *Drug Metabolism Letters* **6**, 33–42 (2012).
14. Bill, M. A. *et al.* The small molecule curcumin analog FLLL32 induces apoptosis in melanoma cells via STAT3 inhibition and retains the cellular response to cytokines with anti-tumor activity. *Molecular Cancer* **9**, 165 (2010).
15. Bartholomeusz, G. A., Donato, N., Estrov, Z., Priebe, W. & Talpaz, M. Activation of a Novel Proteasomal Independent Bcr/Abl Degradation Pathway by WP1130 Induces Apoptosis in CML Cells. *Blood* **106**, 2862–2862 (2005).
16. Levitzki, A. & Gazit, A. Tyrosine Kinase Inhibition: An Approach to Drug Development. *Science* **267**, 1782–1788 (1995).
17. Saporito, M. S., Ochman, A. R., Lipinski, C. A., Handler, J. A. & Reaume, A. G. MLR-1023 Is a Potent and Selective Allosteric Activator of Lyn Kinase In Vitro That Improves Glucose Tolerance In Vivo. *The Journal of Pharmacology and Experimental Therapeutics* **342**, 15–22 (2012).
18. Párrizas, M., Gazit, A., Levitzki, A., Wertheimer, E. & LeRoith, D. Specific Inhibition of Insulin-Like Growth Factor-1 and Insulin Receptor Tyrosine Kinase Activity and Biological Function by Tyrphostins. *Endocrinology* **138**, 1427–1433 (1997).

19. Xie, J. *et al.* The antitumor effect of tanshinone IIA on anti-proliferation and decreasing VEGF/VEGFR2 expression on the human non-small cell lung cancer A549 cell line. *Acta Pharmaceutica Sinica B* **5**, 554–563 (2015).
20. Zhang, Z. *et al.* Jak3 is involved in CCR7-dependent migration and invasion in metastatic squamous cell carcinoma of the head and neck Corrigendum in /10.3892/ol.2024.14712. *Oncol Lett* **13**, 3191–3197 (2017).
21. Subbiah, V. *et al.* Precision Targeted Therapy with BLU-667 for RET-Driven Cancers. *Cancer Discovery* **8**, 836–849 (2018).
22. Yoneda, T. *et al.* The Antiproliferative Effects of Tyrosine Kinase Inhibitors Tyrphostins on a Human Squamous Cell Carcinoma in Vitro and in Nude Mice1. *Cancer Research* **51**, 4430–4435 (1991).
23. Gazit, A., Yaish, P., Gilon, C. & Levitzki, A. Tyrphostins I: synthesis and biological activity of protein tyrosine kinase inhibitors. *J. Med. Chem.* **32**, 2344–2352 (1989).
24. Inagaki, S. *et al.* G Protein-Coupled Receptor Kinase 2 (GRK2) and 5 (GRK5) Exhibit Selective Phosphorylation of the Neurotensin Receptor in Vitro. *Biochemistry* **54**, 4320–4329 (2015).
25. Homan, K. T. *et al.* Identification and Structure–Function Analysis of Subfamily Selective G Protein-Coupled Receptor Kinase Inhibitors. *ACS Chem. Biol.* **10**, 310–319 (2015).
26. Stankevičius, E. *et al.* Opening of Small and Intermediate Calcium-Activated Potassium Channels Induces Relaxation Mainly Mediated by Nitric-Oxide Release in Large Arteries and Endothelium-Derived Hyperpolarizing Factor in Small Arteries from Rat. *The Journal of Pharmacology and Experimental Therapeutics* **339**, 842–850 (2011).
27. Saporito, M. S., Ochman, A. R., Lipinski, C. A., Handler, J. A. & Reaume, A. G. MLR-1023 Is a Potent and Selective Allosteric Activator of Lyn Kinase In Vitro That Improves Glucose Tolerance In Vivo. *The Journal of Pharmacology and Experimental Therapeutics* **342**, 15–22 (2012).
28. Tatake, R. J. *et al.* Identification of pharmacological inhibitors of the MEK5/ERK5 pathway. *Biochemical and Biophysical Research Communications* **377**, 120–125

(2008).

29. Breitenbuecher, F. *et al.* Identification of a novel type of ITD mutations located in nonjuxtamembrane domains of the FLT3 tyrosine kinase receptor. *Blood* **113**, 4074–4077 (2009).
30. Cheng, W., Yang ,Xiaofei, Ping ,Nana, Chen ,Suning & and Sun, A. A case of acute erythroleukemia with concurrence of BCR–ABL and FLT3-TKD (D835E) mutation. *Leukemia & Lymphoma* **56**, 2463–2464 (2015).
31. Sidwell, R. W., Bailey, K. W., Wong, M.-H., Barnard, D. L. & Smee, D. F. In vitro and in vivo influenza virus-inhibitory effects of viramidine. *Antiviral Research* **68**, 10–17 (2005).
32. Patick A. K. *et al.* In Vitro Antiviral Activity of AG7088, a Potent Inhibitor of Human Rhinovirus 3C Protease. *Antimicrobial Agents and Chemotherapy* **43**, 2444–2450 (1999).
33. Roberts, W. G. *et al.* Antitumor Activity and Pharmacology of a Selective Focal Adhesion Kinase Inhibitor, PF-562,271. *Cancer Research* **68**, 1935–1944 (2008).
34. Wilhelm, S. M. *et al.* BAY 43-9006 Exhibits Broad Spectrum Oral Antitumor Activity and Targets the RAF/MEK/ERK Pathway and Receptor Tyrosine Kinases Involved in Tumor Progression and Angiogenesis. *Cancer Research* **64**, 7099–7109 (2004).

Reviewer #3:

Dear Authors,

I am writing regarding the manuscript "Evidential Deep Learning for Guided Drug-Target Interaction Prediction," which presents EviDTI, a novel method for drug-target interaction prediction. The authors have developed an innovative approach that integrates multiple drug and protein feature representations - specifically, 1D protein features alongside 2D and 3D drug features. The model architecture employs a protein feature encoder for sequence processing, while utilizing MG-BERT and GeoGNN models for drug feature extraction. These features are then concatenated to generate comprehensive drug representations, which, combined with an evidential layer, produce both interaction probability predictions and associated uncertainty estimates.

While the authors have demonstrated their model's performance through comparisons on established benchmarks (DrugBank, KIBA, and Davis) and conducted ablation studies, I recommend the following improvements:

Thank you for reviewing in our manuscript. We have carefully revised the manuscript according to your comments. All of your suggestions are helpful to improve our work and make the results more convinced. According to your suggestions, we mainly answer your comments and revise the manuscript from five aspects (**Part 1-5** below):

1) **Newly added models and performance comparative study.** We introduced additional widely used models for comparison with EviDTI. These models cover machine learning algorithms (e.g., Random Forest, Support Vector Machines) as well as the **latest deep learning algorithms** (e.g., GraphformerDTI, DLMDTI, AIGO-DTI) that **use different architectural approaches**. (Page 9-10 Line 163-184 in the revised manuscript) In addition, we explained in detail why GraphDTA was chosen as the comparison model and how GraphDTA can be applied to classification tasks. (Page 43 Line 757-762 in the revised manuscript)

2) **Revision of the structure and formatting of the manuscript** We have critically adjusted the structure and format of the manuscript according to the journal

guidelines to ensure compliance with the journal's requirements.

3) **Redesign of Figure 1.** We have redesigned Figure 1 to better highlight the key architectural components of EviDTI and their interactions, particularly the functional integration process. (Page 8 Line 149 in the revised manuscript)

4) **Extended literature review.** In the '*Introduction*' section, we have included the recent advances from 2022 to 2025, focusing on advances in the areas of drug-target interactions. (Page 3 Line 60-76 in the revised manuscript)

5) **Clear articulation of motivation and contribution.** In the introduction and discussion section, we have clearly articulated EviDTI's motivation and main contributions. (Page 4-5 Line 77-97, Page 5-6 Line 107-113 and Page 32 Line 535-558 in the revised manuscript)

Below we provide a point-by-point response to each suggestion. All modifications in the revised manuscript are highlighted in red text.

Points:

1. The comparative analysis would benefit from including additional state-of-the-art models, particularly those employing different architectural approaches, to better contextualize EviDTI's performance.

Response:

Thank you for your valuable feedback on the model performance evaluation. We expanded our comparisons with other state-of-the-art models, with a particular focus on those approaches that use different architectural designs.

To further enhance the comparative analyses, we have added three methods covering two categories of methods: network-based models and Proteochemometrics - based models (**Table R1**). Details of these methods are provided in **Part 2**. (The corresponding revision can be seen Line 780-804 on Page 44-46 of the revised manuscript)

Furthermore, we conducted a comprehensive performance comparison between the baseline methods and EviDTI on three benchmark datasets, as shown in **Part 2**

below. (The corresponding revision can be seen Line 163-184 on Page 9-10 of the revised manuscript)

Table R1 Categories of newly added DTI prediction methods

Category of methods	Subcategories of methods	Model
Network-based		AIGO-DTI
Proteochemometrics-based	Based on gated cross-attention mechanism	GraphormerDTI
	Based on protein and small molecule pre-training	DLM-DTI

(1) Newly added of three comparison methods

In order to further enrich our comparative analyses, we have incorporated three methods, which are divided into two categories: network-based DTI prediction models and protein chemometrics-based DTI prediction models. For the network-based DTI prediction models, AIGO-DTI has been selected. AIGO-DTI, with its unique advantages at the network level, is able to provide a different perspective from that of EviDTI. For proteochemometrics-based DTI prediction models, two approaches were selected: the GraphormerDTI, a model that utilizes a gated cross-attention mechanism; and the DLM-DTI, a model based on pre-training of proteins and small molecules.

Category 1: Network-Based Models (The corresponding revision can be seen Line 780-787 on Page 44 of the revised manuscript)

AIGO-DTI: AIGO-DTI proposes an Adaptive Iterative Graph Optimization (AIGO)-DTI prediction framework. This framework integrates atomic cluster information and enhances molecular features through the design of functional group prompts and graph encoders, thereby optimizing the construction of DTI association networks. Furthermore, the optimization of graph structure is transformed into a node similarity learning problem, utilizing multi-head similarity metric functions to iteratively update the network structure to improve the quality of DTI information.

Category 2: Proteochemometrics-Based Models (The corresponding revision can be seen Line 788-804 on Page 44-45 of the revised manuscript)

GraphormerDTI: GraphormerDTI integrates the Graph Transformer neural network with a 1D-CNN to extract the representations of drug and target, respectively. Specifically, GraphormerDTI embeds molecular graphs into vector representations through iterative Transformer-based message passing. The structural representations of the molecules are encoded through node centrality encoding, node spatial encoding and edge encoding. The attentional mechanisms are used to model interactions.

DLM-DTI: DLM-DTI was comprised of three primary components: the drug encoder, the target encoder, and the interaction prediction head. The target encoder incorporates both the teacher and student models of language models for protein sequences. The teacher model employed for target sequence encoding was the ProtBERT model. The student model was designed to be consistent with the original teacher model with fewer layers to conduct knowledge distillation. The function of the drug encoder is to convert SMILES sequences to molecular representations, and this is achieved by employing the ChemBERTa encoder. The interaction prediction header predicts interaction probabilities by concatenating drug and target representations.

(2) Comparative evaluation of EviDTI and baseline models (The corresponding revision can be seen Line 163-184 on Page 9-10 of the revised manuscript)

The performance results on the DrugBank dataset are shown in **Table R2** (corresponding to **Table 1** of the revised manuscript). EviDTI demonstrates robust overall performance on all metrics, especially in terms of precision (81.90%), and competitive values for Accuracy (82.02%), MCC (64.29) and F1 score (82.09%). Besides, we evaluated EviDTI on the Davis and KIBA datasets, which are particularly challenging due to class imbalance. The performance results on the Davis and KIBA datasets are presented in **Tables R3 and R4** (corresponding to **Table 2 and 3** of the revised manuscript), respectively. Specifically, on the KIBA dataset, EviDTI outperformed the best baseline model by 0.6% in accuracy, 0.4% in precision, 0.3% in MCC, 0.4% in F1 score, and 0.1% in AUC. On the Davis dataset, EviDTI exceeds 0.8%

in accuracy, 0.6% in precision, 0.9% in MCC, 2% in F1 score, 0.1% in AUC, and 0.3% in AUPR. These results demonstrate the robustness and superior performance of EviDTI when dealing with complex and unbalanced datasets. The findings underscore the efficacy and competitiveness of EviDTI.

Table R2 Comparison results of EviDTI and baselines on the DrugBank dataset. The best results are highlighted in bold, while the second-best results are underlined.

DrugBank	Accuracy (std)	Recall (std)	Precision (std)	MCC (std)	F1 score (std)	AUC (std)	AUPR (std)
RF	71.07 (1.34)	73.08 (0.95)	76.92 (1.42)	51.23 (1.54)	74.94 (0.92)	83.03 (0.65)	84.78 (0.67)
NB	52.90 (1.13)	87.99 (1.77)	51.71 (1.01)	42.34 (3.24)	65.13 (0.89)	60.76 (1.24)	59.37 (1.77)
SVM	78.02 (1.92)	75.04 (2.93)	73.86 (2.14)	53.87 (1.29)	74.39 (1.40)	82.02 (1.62)	82.85 (1.49)
DeepConv-DTI	79.50 (0.2)	80.45 (0.2)	79.30 (0.6)	57.91 (0.6)	79.87 (0.6)	86.69 (0.2)	86.28 (0.1)
GraphDTA	77.69 (0.1)	78.77 (1.1)	76.92 (1.0)	55.41 (0.6)	77.83 (1.1)	83.62 (0.2)	83.26 (0.1)
MolTrans	78.48 (0.1)	81.97 (0.2)	77.01 (0.3)	55.65 (0.6)	79.41 (0.7)	85.98 (0.6)	86.67 (0.1)
HyperAttention	80.85 (0.2)	**88.16 (0.1)**	77.09 (0.1)	62.32 (0.4)	**82.26 (0.3)**	89.94 (0.1)	88.95 (0.1)
TransformerCPI	79.01 (0.2)	76.26 (0.3)	80.47 (0.3)	58.09 (0.2)	78.31 (0.3)	84.62 (0.3)	84.23 (0.2)
GraphormerDTI	80.22 (1.07)	82.05 (1.23)	79.11 (0.91)	60.48 (2.16)	80.55 (1.02)	88.63 (1.02)	88.36 (0.99)
AIGO-DTI	81.88 (1.37)	82.28 (2.69)	81.26 (0.36)	63.80 (2.78)	81.74 (1.40)	88.62 (1.02)	88.44 (1.65)
DLM-DTI	**82.13 (0.27)**	82.54 (1.91)	81.68 (1.56)	**66.32 (0.55)**	81.08 (0.56)	**90.48 (0.48)**	**90.20 (0.60)**
EviDTI	82.02 (0.4)	82.28 (1.0)	**81.90 (0.5)**	64.29 (1.0)	82.09 (0.7)	88.87 (0.7)	88.45 (0.2)

Note: Five independent replications of each method were performed (n=5). Data are expressed as means (std).

Table R3 Comparison results of EviDTI and baselines on the KIBA dataset. The best results are highlighted in bold, while the second-best results are underlined.

KIBA	Accuracy (std)	Recall (std)	Precision (std)	MCC (std)	F1 score (std)	AUC (std)	AUPR (std)
RF	88.41 (0.34)	64.84 (2.48)	70.57 (2.30)	63.21 (1.67)	67.59 (1.37)	91.05 (0.38)	76.05 (1.04)
NB	52.92 (0.83)	71.72 (1.77)	24.22 (1.40)	24.56 (3.24)	36.21 (1.47)	65.92 (0.78)	35.51 (0.94)
SVM	81.72 (0.46)	3.87 (0.76)	66.40 (2.91)	5.24 (1.34)	7.32 (0.92)	73.27(0.56)	43.38 (1.60)
DeepConv-DTI	89.05 (0.3)	68.87 (0.4)	80.52 (0.3)	70.21 (0.2)	74.24 (0.6)	90.67 (0.1)	80.42 (0.6)
GraphDTA	88.92 (0.1)	59.42 (3.2)	77.53 (2.0)	67.51 (0.5)	71.24 (0.8)	91.41 (0.1)	77.61 (0.7)
MolTrans	88.91 (0.1)	73.53 (0.2)	70.42 (0.3)	67.62 (0.2)	71.94 (0.1)	92.32 (0.3)	79.49 (0.6)
HyperAttention	89.33 (0.1)	79.63 (0.2)	68.93 (0.1)	70.12 (0.2)	74.21 (0.3)	93.51 (0.1)	81.41 (0.2)
TransformerCPI	87.01 (0.1)	63.12 (0.3)	66.93 (0.3)	66.2 (0.2)	69.12 (0.2)	88.81 (0.1)	70.81 (0.1)
GraphormerDTI	86.54 (0.85)	**85.40 (1.68)**	60.63 (1.84)	64.03 (1.10)	70.87 (0.89)	92.58 (0.25)	81.95 (0.91)
AIGO-DTI	90.68 (0.32)	73.65 (0.77)	76.77 (1.12)	69.46 (0.97)	75.17 (0.77)	93.06 (0.26)	**83.91 (0.77)**
DLM-DTI	88.25 (0.25)	68.10 (2.15)	69.86 (0.90)	61.72 (0.78)	68.94 (0.87)	91.14 (0.39)	75.63 (0.41)
EviDTI	**91.27 (0.7)**	70.77 (1.6)	**80.90 (1.3)**	**70.44 (0.7)**	**75.48 (0.8)**	**93.57 (0.3)**	83.03 (0.9)

Note: Five independent replications of each method were performed (n=5). Data are expressed as means (std).

Table R4 Comparison results of the EviDTI and baselines on the Davis dataset. The best results are highlighted in bold, while the second-best results are underlined.

Davis	Accuracy (std)	Recall (std)	Precision (std)	MCC (std)	F1 score (std)	AUC (std)	AUPR (std)
RF	84.48 (0.60)	65.01 (1.47)	80.70 (1.87)	64.12 (2.11)	72.01 (1.49)	89.64 (0.68)	83.04 (1.68)
NB	67.53 (0.92)	61.70 (1.82)	47.78 (1.95)	45.21 (1.68)	53.85 (1.72)	70.11 (1.39)	58.43 (2.09)
SVM	82.13 (0.59)	59.70 (1.57)	76.96 (2.19)	60.23 (1.45)	67.24 (1.61)	87.85 (0.78)	80.16 (1.65)
DeepConv-DTI	86.41 (0.2)	72.14 (0.3)	77.85 (0.4)	64.53 (0.3)	74.89 (0.6)	91.83 (0.2)	83.87 (0.6)
GraphDTA	81.72 (0.1)	53.01 (1.7)	74.33 (1.4)	62.42 (0.6)	71.83 (0.8)	85.92 (0.4)	74.31 (0.7)
MolTrans	83.61 (0.1)	76.31 (0.3)	68.87 (0.2)	62.91 (0.1)	72.40 (0.1)	89.84 (0.2)	80.26 (0.6)
HyperAttention	86.61 (0.1)	78.01 (0.2)	75.41 (0.1)	67.54 (0.2)	74.21 (0.2)	92.01 (0.1)	83.92 (0.1)
TransformerCPI	82.22 (0.1)	68.82 (0.3)	68.81 (0.3)	64.21 (0.2)	72.41 (0.2)	87.72 (0.1)	76.71 (0.1)
GraphormerDTI	84.81 (0.84)	**83.78 (2.52)**	68.43 (2.18)	65.22 (1.84)	75.28 (1.43)	91.24 (0.74)	84.31 (1.48)
AIGO-DTI	86.19 (0.79)	72.55 (2.94)	78.01 (2.11)	65.72 (1.96)	75.12 (1.52)	92.01 (0.97)	84.81 (1.10)
DLM-DTI	85.66 (0.47)	74.08 (2.50)	75.26 (1.92)	64.68 (1.27)	74.62 (1.03)	91.30 (0.35)	82.48 (0.86)
EviDTI	**87.40 (0.7)**	72.93 (1.6)	**81.17 (1.3)**	**68.40 (0.7)**	**76.81 (0.8)**	**92.10 (0.3)**	**85.07 (0.9)**

Note: Five independent replications of each method were performed (n=5). Data are expressed as means (std).

2. The manuscript's structure and formatting should adhere more closely to journal guidelines to enhance readability and accessibility.

Response:

Thank you for your comments. In order to enhance the readability and accessibility of the manuscript, the structure and formatting of the manuscript will be adapted in strict accordance with the journal guidelines.

First, we checked the details including font, font size, line spacing and margins in accordance with the journal format guidelines.

Second, we adjusted the formatting of headings at all levels (e.g., bold, italic, or specific font sizes) according to journal requirements to ensure clarity of hierarchy.

Third, we displayed all data points on all charts with a sample size of less than 10.

Forth, we included error bars to represent standard deviation for all repeated experiments and indicate the number of repeated experiments.

3. Figure 1 would be more effective if redesigned to better highlight the model's key architectural components and their interactions, particularly the feature integration process.

Response:

Thank you for your comments. We have redesigned Figure 1. As illustrated by **Figure R1**, the different modules of the model are clearly distinguished by different dashed boxes. Furthermore, new modules have been incorporated to demonstrate the application of the model in the field of drug discovery. (The corresponding revision can be seen Line 149 on Page 8 of the revised manuscript)

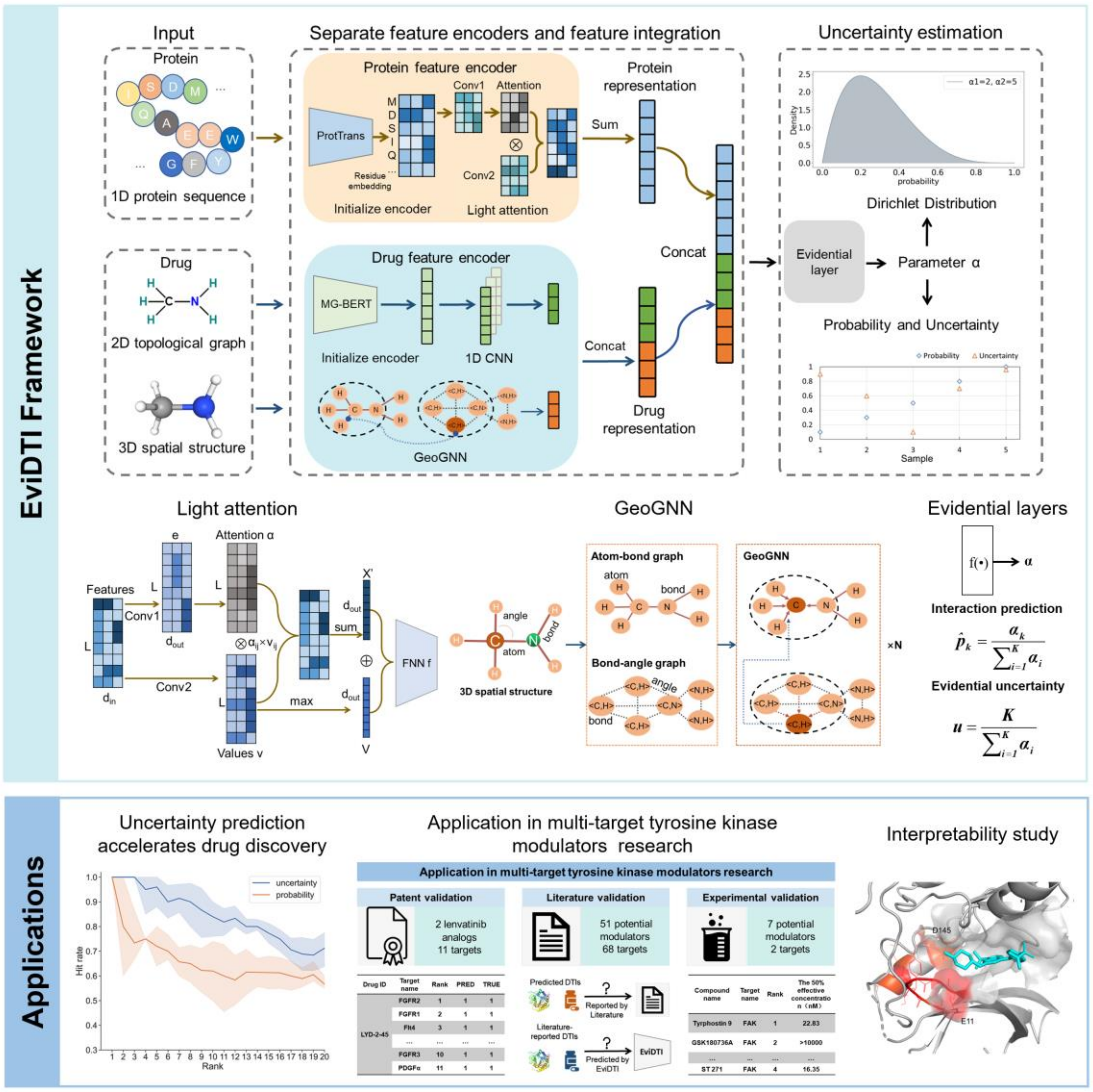


Figure R1 The flowchart of EviDTI.

4. The literature review should be expanded to incorporate recent advances in drug-target interaction prediction, specifically those published within the last 1-2 years, to better establish the work's novelty and contextual relevance.

Response:

Thank you for your valuable suggestions for our research. In the revised manuscript, we added 10 references and expanded on recent advances in the field of DTI prediction in the introduction. (The corresponding revision can be seen Line 67-76 on Page 3-4 of the revised manuscript)

The expanded literature review in the revised manuscript is shown below: "A great deal of innovative work has been undertaken to predict interactions more effectively[1].

To improve model interpretability and capture local interactions of drug-target, the gated cross-attention mechanisms have received more attention[2-5]. To address the problem of small dataset sizes and incomplete representations in DTI prediction, pre-training models are emerging as a promising solution[6-8]. These models demonstrate excellent scalability and generalization capabilities across a wide range of prediction tasks. To achieve more comprehensive and nuanced representations, many methods integrate multimodal techniques that combine different types of data[9-10].

The new additions to the literature are as follows, corresponding to the citations above:

1. de la Fuente, J. et al. Towards a more inductive world for drug repurposing approaches. Nature Machine Intelligence 7, 495–508 (2025).
2. Bian, J., Zhang, X., Zhang, X., Xu, D. & Wang, G. MCANet: shared-weight-based MultiheadCrossAttention network for drug–target interaction prediction. Briefings in Bioinformatics 24, bbad082 (2023).
3. He, Y. et al. Flexible drug-target interaction prediction with interactive information extraction and trade-off. Expert Systems with Applications 249, 123821 (2024).
4. Hadipour, H. et al. GraphBAN: An inductive graph-based approach for enhanced prediction of compound-protein interactions. Nature Communications 16, 2541 (2025).
5. Gao, M. et al. GraphormerDTI: A graph transformer-based approach for drug-target interaction prediction. Computers in Biology and Medicine 173, 108339 (2024).
6. Sun, Y., Li, Y. Y., Leung, C. K. & Hu, P. iNGNN-DTI: prediction of drug–target interaction with interpretable nested graph neural network and pretrained molecule models. Bioinformatics 40, btae135 (2024).
7. Lu, Z. et al. DTIAM: a unified framework for predicting drug-target interactions, binding affinities and drug mechanisms. Nature Communications 16, 2548 (2025).
8. Lee, J., Jun, D. W., Song, I. & Kim, Y. DLM-DTI: a dual language model for the prediction of drug-target interaction with hint-based learning. J Cheminform 16, 14 (2024).

9. He, C., Yang, C., Zhang, H., Long, Y. & Zhao, X. DTI-MPFM: A multi-perspective fusion model for predicting potential drug–target interactions. Expert Systems with Applications 264, 125740 (2025).
10. Zhao, W. et al. MSI-DTI: predicting drug-target interaction based on multi-source information and multi-head self-attention. Briefings in Bioinformatics 25, bbae238 (2024).

5. The inclusion of GraphDTA as a baseline model requires further justification, given its primary design for binding affinity prediction rather than interaction prediction. The authors should elaborate on how this choice relates to their work's objectives.

Response:

Thank you for your valuable suggestions for our research. GraphDTA is primarily designed for drug-target affinity prediction. However, GraphDTA has been widely adopted as a benchmark model by many other state-of-the-art DTI[11-13] methods due to its advantages in processing molecular structure information. These studies demonstrate that GraphDTA serves as a robust baseline model, providing an important reference standard for evaluating novel DTI prediction models. To clarify how GraphDTA can be adjusted for fair comparisons, we have provided detailed explanations. (The corresponding revision can be seen Line 757-762 on Page 43 of the revised manuscript)

In order to facilitate fair comparisons, GraphDTA was adapted according to the following steps:(1) A Sigmoid activation function was added after the last layer to convert the model output to binary probabilities. (2) A binary cross-entropy was used as the loss function instead of the original mean square error (MSE) loss. (3) A grid search was performed to optimize key hyperparameters such as learning rate and batch size.

6. The manuscript would benefit from a clearer articulation of its motivations and contributions, particularly highlighting how EviDTI advances the field beyond existing approaches.

Response:

Thank you for your suggestion to clearly articulate the contribution and motivation of EviDTI. In the Introduction and Discussion sections, we provided a clearer articulation of the manuscript's motivations and contributions.

The newly added description of EviDTI's motivation and contribution in the revised version are described below:

(1) **Newly Added description of EviDTI's motivation and contribution to the "Introduction"**:

Motivation (The corresponding revision can be seen Line 77-97 on Page 4-5 of the revised manuscript)**:** Despite the significant advances achieved by DL in DTI prediction, its practical application still faces a major challenge: high probability predictions do not necessarily correspond to high confidence. The root of this problem lies in the fundamental difference between DL models and human cognitive models. Humans are able to dynamically adjust the confidence level according to the knowledge boundary, giving certain answers to familiar questions and explicitly expressing 'uncertainty' about unknown domains. In contrast, traditional DL models generate predictions for all inputs, including out-of-distribution and noisy samples. More critically, traditional DL models lack probability calibration ability and may produce high prediction probabilities even in low confidence situations. This phenomenon of "overconfidence" has the tendency to introduce unreliable predictions into downstream processes, including the pushing of false positives into experimental validation, the omission of potentially active compounds in virtual screening, and even the designing of clinical trial protocols based on false predictions. These situations not only lead to inefficient use of resources, but also have the potential to delay the drug discovery process.

Uncertainty quantification (UQ) methods can address above challenge and thus improve the robustness of neural models in scientific applications, especially in the field of drug discovery. The core value of UQ is to provide a reliable basis for decision-making by distinguishing between plausible predictions and high-risk predictions.

Contribution (The corresponding revision can be seen Line 107-113 on Page 5-6

of the revised manuscript)**:** In order to enhance the reliability and applicability of DTI prediction methods, we propose an EDL-based DTI prediction framework (EviDTI). The framework utilizes pre-trained knowledge as well as multi-dimensional representations to enhance the model performance. More importantly, EviDTI provides prediction confidence estimation by introducing EDL to help identify drug candidates that are most likely to be successful, thus reducing the risk and cost associated with false positives.

(2) **Newly Added description of EviDTI's motivation and contribution to the "Disscussion"**

Motivation (The corresponding revision can be seen Line 535-545 on Page 32 of the revised manuscript):

Although current deep learning models have made significant progress in DTI prediction, they typically lack the ability to provide confidence estimates for predictions. This limitation seriously affects the effectiveness of these models in real-world applications, where decision making often requires a clear understanding of the reliability of the predicted results. To bridge the gap between predictive models and practical applications, we present EviDTI, a novel framework that integrates evidential-based deep learning. This method not only predicts DTI with high accuracy, but also provides reliable quantification of uncertainty, which enhances the practicality and reliability of the model. This is important for advancing the field of drug discovery as it is directly related to the success and cost-effectiveness of experiments.

Contribution (The corresponding revision can be seen Line 548-558 on Page 32-33 of the revised manuscript)**:**

The performance comparison across three datasets demonstrates that EviDTI achieves competitive performance compared to 11 DTI prediction models. This study demonstrates that EDL can provide a reliable uncertainty quantification for DTIs prediction. Importantly, we showed that uncertainty-guided predictions can prioritize experiments in drug discovery and drug repositioning studies. The experimental validation of interactions between two tyrosine kinase targets and seven potential tyrosine kinase modulators (totally 14 potential DTIs) was performed. Among those

predicted DTIs, 5 DTIs showed binding affinity, and 4 of them reached the nanomolar level of binding affinity. These results demonstrate that EviDTI is a powerful tool that can accelerate the translation process from theoretical models to actual drug development.

Reviewer #3 (Remarks on code availability):

The authors provided comprehensive code accompanied by well-documented instructions, which offers readers a good opportunity for reproducibility of their work.

Response:

Thank you for your positive comments about the code we provide and its documentation. We are very pleased to learn that you think that the code and detailed instructions we have provided provide good opportunities for readers to reproduce our research work.

Reference:

1. de la Fuente, J. *et al.* Towards a more inductive world for drug repurposing approaches. *Nature Machine Intelligence* **7**, 495–508 (2025).
2. Bian, J., Zhang, X., Zhang, X., Xu, D. & Wang, G. MCANet: shared-weight-based MultiheadCrossAttention network for drug–target interaction prediction. *Briefings in Bioinformatics* **24**, bbad082 (2023).
3. He, Y. *et al.* Flexible drug-target interaction prediction with interactive information extraction and trade-off. *Expert Systems with Applications* **249**, 123821 (2024).
4. Hadipour, H. *et al.* GraphBAN: An inductive graph-based approach for enhanced prediction of compound-protein interactions. *Nature Communications* **16**, 2541 (2025).
5. Gao, M. *et al.* GraphormerDTI: A graph transformer-based approach for drug-target interaction prediction. *Computers in Biology and Medicine* **173**, 108339 (2024).
6. Sun, Y., Li, Y. Y., Leung, C. K. & Hu, P. iNGNN-DTI: prediction of drug–target interaction with interpretable nested graph neural network and pretrained molecule models. *Bioinformatics* **40**, btae135 (2024).
7. Lu, Z. *et al.* DTIAM: a unified framework for predicting drug-target interactions, binding affinities and drug mechanisms. *Nature Communications* **16**, 2548 (2025).
8. Lee, J., Jun, D. W., Song, I. & Kim, Y. DLM-DTI: a dual language model for the prediction of drug-target interaction with hint-based learning. *J Cheminform* 16, 14 (2024).
9. He, C., Yang, C., Zhang, H., Long, Y. & Zhao, X. DTI-MPFM: A multi-perspective fusion model for predicting potential drug–target interactions. *Expert Systems with Applications* **264**, 125740 (2025).
10. Zhao, W. *et al.* MSI-DTI: predicting drug-target interaction based on multi-source information and multi-head self-attention. *Briefings in Bioinformatics* **25**, bbae238 (2024).

11. Zhao, Q., Zhao, H., Zheng, K. & Wang, J. HyperAttentionDTI: improving drug–protein interaction prediction by sequence-based deep learning with attention mechanism. *Bioinformatics* **38**, 655–662 (2022).
12. Zeng, X., Chen, W. & Lei, B. CAT-DTI: cross-attention and Transformer network with domain adaptation for drug-target interaction prediction. *BMC Bioinformatics* **25**, 141 (2024).
13. Zhang, S. *et al.* AIGO-DTI: Predicting Drug–Target Interactions Based on Improved Drug Properties Combined with Adaptive Iterative Algorithms. *J. Chem. Inf. Model.* **64**, 4373–4384 (2024).

Point-by-point Response to Reviewers' Comments

Reviewer #2:

Zhao et al. describe a new evidential deep learning tool for predicting drug target interactions that they developed called EviDTI. Their method quantifies uncertainty and incorporates protein sequence with 2D and 3D models of ligands. To validate EviDTI, they performed a case study on its predictions with tyrosine kinase inhibitors, followed by in vitro screening. The impetus for this study was to begin to address probabilities and uncertainties in current DTI methods, and this work contributes to the field.

The authors have addressed most of the concerns of reviewers, concerning both comparative learning models and in vitro validation. The addition of data, references, and discussion since the original submission have strengthened the manuscript.

Response:

Thank you for the efforts you have put into reviewing our manuscript. Below we provide a point-by-point response to your suggestion. All modifications in the revised manuscript are highlighted in red text.

Points

My only concern that remains are the in vitro activities reported for tyrphostin 9 and vodobatinib against FLT3, highlighted in Figure S7. The authors are still claiming activation, when their graphical representation indicates dose-dependent inhibition in the opposite direction as one would expect. Having changed the y-axis labels to relative activity was good. Normally dose-dependent inhibition would be measured as activity relative to a negative control (kinase assay with DMSO alone). All of the experiments where the authors note inhibition look normal, and look to be normalized to a negative control. e.g. their tyrphostin 9 and FAK (Fig. S71) indicates that around 0.1 to 1 nM of tyrphostin 9, activity of FAK is the same as it was for a data point which contained DMSO and no tyrphostin 9; and at, 10 μM tyrphostin 9, there was zero activity of

tyrphostin 9. That all looks normal.

However, in the tryphostin 9 and FLT3 plot, at 0.1 - 1 nM tryphostin 9, according to their plot, FLT3 is completely inhibited and has 0% of negative control activity. The plot also suggests that at high concentrations of tryphostin 9 above 1 μM, FLT3 activity is approaching 100% of control activity. Meaning that low concentrations of tyrphostin 9 inhibit FLT and high concentrations do not. This doesn't make sense, unless the authors were inhibiting FLT with another molecule and tryphostin 9 recovered activity, which does not seem what their describing. Hypothetically, if a molecule at low concentrations had no effect (like tryphostin 9 on FAK), one would expect the enzyme activity relative to a non-inhibited data point to be close to 100% (as it is for tryphostin 9 on FAK). If the hypothetical compound was activating at high concentrations, one would expect the enzyme activity relative to a non-inhibited data point to be *above* 100%. It seems as though these plots might be normalized to themselves, that the highest activity within an EC50 experiment was called 100 and the lowest was called 0? These should be normalized to enzyme activity without compound (but the same DMSO concentration as with compound). Or if my interpretation is off, the authors should explain how they're normalizing their relative enzyme activities.

Response:

Thank you for your insightful feedback and constructive suggestions. We apologize for the confusing y-axis labeling in Figure S7 e-f. In the revised manuscript, we have refined Figure S7 e-f. The y-axis values range from 100% to 200%, normalizing enzyme activity to the initial value with the same DMSO concentration without compound as a negative control (Line 119, page 14 in the revised Supplementary Information). As shown in Figure S7 e-f, the activation of FLT3 by Tyrphostin 9 and Vodobatinib ranged from 0.33 nM to 10000 nM, respectively, and both compounds exhibited increased FLT3 activation effects, reaching up to 180% of control activity at high concentrations above 10 μM. The source data of Figure S7 are provided through Figshare (https://doi.org/10.6084/m9.figshare.28816634.v1).

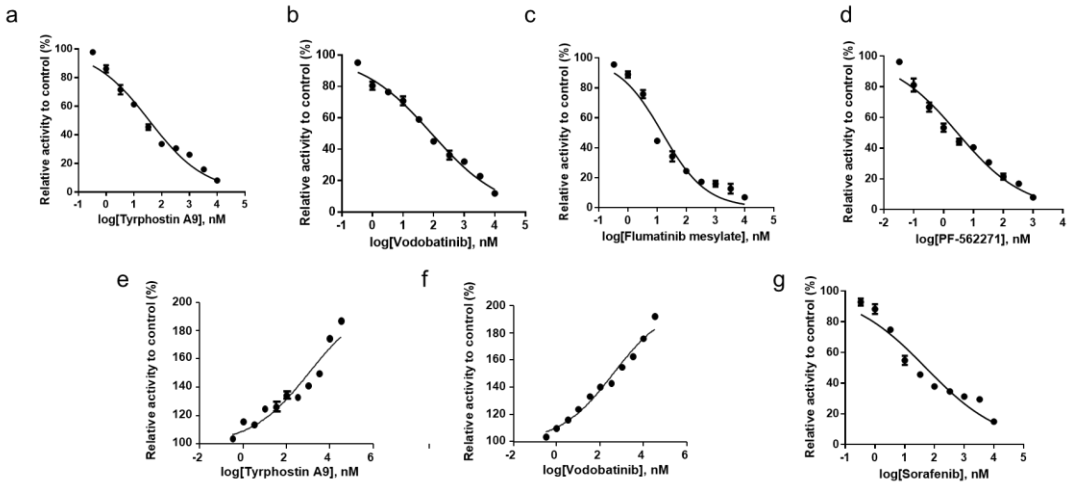


Figure S7 a-d Concentration-activity curves of Tyrphostin 9, Vodobatinib, Flumatinib and PF-562271 in the FAK kinase ADP-Glo assays, respectively. PF-562271 as positive control. Mean ± SEM of three independent experiments is shown (n=3). **e-g** Concentration-activity curves of Vodobatinib, Tyrphostin 9 and sorafenib in the FLT3 kinase ADP-Glo assays, respectively. Sorafenib as positive control. Mean ± SEM of three independent experiments is shown (n=3).