

DTIAM: A unified framework for predicting drug-target interactions, binding affinities and drug mechanisms

Corresponding Author: Professor Min Li

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)

The authors present a bioactivity prediction framework which includes quantitative target activity and functional effects.

The work conceptually makes sense, although it's more written in computer science-style, and it doesn't include prospective validation either which would be useful. The 'validations' given are more proof by example and based on existing literature, which I would see as quite weak a validation, and not really on the level I would like to see published in this journal. Also the datasets are too small to really be useful in practice, in particular for functional annotations (lines 562 ff.)

Some items that should be addressed before publication in any case if the editor decides it is appropriate are:

- 'binary' target prediction is obviously an arbitrary simplification of the task (lines 35 ff.), please point this out, also point out advantages (often better performance)
- Classification of methods in intro is a bit unusual; structure-; ligand-; and network-based methods might be more suitable
- General concept of DTIAM is similar to proteochemometrics approaches (although with learned features, and a different NN architecture etc); key PCM papers (Wikberg etc, <https://pubmed.ncbi.nlm.nih.gov/12851294/> is probably the first one but also many later ones) should also be cited in introduction and relationship between methods be pointed out
- Functional effects have been predicted before in target prediction and related papers should be cited (e.g. Mervin et al <https://www.frontiersin.org/articles/10.3389/fphar.2018.00613/full>)
- Performance doesn't pay any attention to chemistry and applicability domain, would suggest to not only treat as computer science exercise if the authors have practical applicability in mind (for this also validation should be performed)
- 305: CYPs are listed; what do the authors mean by 'activating' CYPs, is this upregulation? How are substrates treated?; Table 1 are not 'candidate drugs', please use appropriate terminology
- 2.5. is too small-scale and not a 'validation', same for 2.6., this is basically just proof-by-example

So overall the limitations of the dataset and the performance metrics used; and the lack of any discussion of chemistry and prospective validation makes this referee conclude that this manuscript isn't a very strong one, it would need to be extended along the above axes to get there.

Reviewer #2

(Remarks to the Author)

This paper presents a new approach for predicting DTIs, DTA and MoA which is entitled DTIAM. The approach leverages two transformer models, one for molecules and one for proteins, which have been pre-trained in a self-supervised manner on large volumes of unlabeled data. These two models are used to create feature vectors from drug and protein pairs which are fed into another model to predict the likelihood of a binary interaction between a given pair. The DTIAM approach is

shown to perform well compared to a series of competing approaches from the literature. The experiments consider both target and drug cold start problems, an inductive setup where proteins and drugs unseen during training are evaluated upon.

Overall I enjoyed the paper and found it overall easy to read. I did have some concerns with the current manuscript however:

- The contribution of the works seems quite low overall. Essentially this works takes two existing model types and combines them together. Could the authors articulate more clearly what is the contribution of their work?
- For the molecular transformer could the authors clarify if this model is created by them? If so, why did they chose to do this instead of using an existing approach? For this model the authors propose a multi-part objective function was an ablation study performed in order to assess the impact of each one of these terms?
- Did the authors consider other protein representation learning approaches than ESM? If so, why was ESM chosen?
- The actual model which predicts the DTIs is poorly explained and seemed to be created via some form of autoML package called AutoGluon. Why did the authors not use a neural network here? Did the authors consider updating the weights of the protein and drug transformer using the supervised signal from this model?
- Overall I would expect to see more experiments performed ablating different parts of the model to assess the impact each one makes.
- Whilst it is nice to see the code has been made available, it would be good if the authors could make the code more usable overall for example by have a more in depth readme to better explain how to use the code and how to replicate some of the key results.

Version 1:

Reviewer comments:

Reviewer #2

(Remarks to the Author)

Thanks for addressing my comment and improving the readme of the code.

I have been asked to comment on the concerns of reviewer #1.

- 1) Yes reviewer 1 has raised very valid technical questions – many of which overlapped with my own concerns. I do not disagree with any of reviewer 1's points.
- 2) The authors have taken steps to address many of the comments made from reviewer 1. I feel that point 7 could have been addressed more – the authors only provided 1 example, and I think more should be provided as this is the crucial result that validates their methodology.

(Remarks on code availability)

The code would run however it could be improved and brought up to to proper PEP8 standard.

Reviewer #3

(Remarks to the Author)

The authors present a prediction framework for drug-target interactions based on features of drug and protein pairs. Specific comments are listed below.

1. It is not clear to me how the sequence/residue information of the protein alone without any information of protein structure, together with some structural features of the compounds, can yield meaningful prediction. Does the prediction framework consider the structure of the protein targets? What kind of data was the model trained on and how does the model predict binding affinity, potency, and positive or negative efficacy?
2. What are the selection criteria for deciding whether the data are to be included or not in the training set? It is not uncommon that different values of potency and positive or negative efficacy are reported in the literature for a given drug and target pair, and the sensitivity to the same compound can be influenced by external factors beyond the residue information of the protein target. Please discuss these aspects in the text.
3. The authors stated that their framework can accurately predict activation/inhibition mechanisms. In what sense has this been accomplished? Based on the TMEM16A example, the framework did not result in any prediction of activation or inhibition mechanism.
4. The authors should justify their choice of TMEM16A as the only chosen target for experimentally testing their framework.
5. The authors should test more than just one class of proteins and more than just one compound for the experimental validation of their prediction framework. For instance, well-characterized kinases, enzymes, transporters, other ion channels, receptors etc. should be tested. In my opinion, one specific compound tested on one specific target is insufficient for generalization of the utility of the presented framework.
6. Please provide the experimental information on the construct (species, UniProt ID, details of the construct etc.) and cell

culture (cell lines, media, cultivation, transfection etc.). For electrophysiology, please provide information on series resistance and its compensation, liquid junction potential and whether this has been corrected for, the perfusion system, and whether the glass pipettes were fire-polished. Please move the methods to the 'methods' section.

7.. Please plot the 1) the current-voltage relations for each dehydrocostus lactone overlaid in the same graph and 2) the concentration-response relation of dehydrocostus lactone from the whole-cell experiments.

8. Please plot the scale bars, namely the current and time axes in Fig. 5.

9. Please perform outside-out patch-clamp experiments to verify that dehydrocostus lactone can act on excised membrane patches expressing TMEM16A.

10. What is the predicted potency and efficacy of dehydrocostus lactone and how do these compare to the experimentally measured potency and efficacy for TMEM16A? The authors should comment on this in the text.

11. The authors stated that the predicted affinity is 0.121 μ M for the TMEM16A inhibitor daidezein, while the previously reported potency is 1.39 μ M. While affinity and potency are different properties, a correspondence between the two is expected. Have the authors verified this previously reported potency? Please comment on this in the text.

12. Considering that TMEM16A was used as a specific example, the authors should omit data on TMEM16A and other TMEM16 proteins in the training set and quantify the accuracy and precision of the model in predicting TMEM16A modulators. In particular, the authors should state if the known blockers/inhibitors 1PBC, niclosamide, and benzbromarone can be successfully predicted in terms of their potency and efficacy. Please include this prediction data and address any discrepancies in the text.

13. Please comment on how quantitatively the described methods can predict experimental potency and efficacy in the text. The authors should additionally quantify the accuracy of the predictions using data on experimental potency and efficacy, for instance by investigating their correspondence, in a systematic manner.

(Remarks on code availability)

Version 2:

Reviewer comments:

Reviewer #2

(Remarks to the Author)

Thanks to the authors for addressing my comments.

(Remarks on code availability)

Code is much better thank you. I would also consider adding type hints for your code as well. Also why is PyTorch not in the requirements file?

Reviewer #3

(Remarks to the Author)

The authors should not claim that their model can predict mechanisms without any experimental support. Please remove this claim throughout the manuscript.

Unaddressed concerns:

Point 6 –

1. The authors should use a recombinant system such as HEK293T cells to test the effect of dehydrocostus lactone on TMEM16A; this is standard practice.

2. The authors should comment on the Hill coefficient of the current decrease caused by dehydrocostus lactone; the reported concentration dependence is unusually shallow, spanning 5 orders of magnitude. While multiple binding sites is common in symmetric or pseudosymmetric assemblies where each monomer or subunit interface contributes to the binding of one ligand molecule, this is mechanistically implausible in TMEM16A whose fundamental operating unit is a single subunit.

3. The method section for electrophysiology needs to be revised, including the reporting of pipette resistance and series resistance, errors associated with units, and the consistency of solution composition. The last sentence 'In a word, dose-dependent manner' does not belong to the methods.

Point 9 –

Outside-out patch-clamp data are needed to rigorously show that dehydrocostus lactone can inhibit TMEM16A in excised membranes. Satisfactorily addressing this issue is an absolute requirement for validating that the effect of the compound originates from directly binding to the channel. This is an important aspect of the authors' prediction model. Outside-out patches are formed after the establishment of the whole-cell configuration. There are no excuses for claiming that resources or expertise are not available to perform these experiments if the authors' have already succeeded in recording whole-cell currents. Please provide these data in the manuscript or completely remove TMEM16A from the manuscript if there is no inhibition in outside-out patches; the field does not need another false activator/inhibitor.

Point 12 –

Please directly address my concerns and provide the data on the prediction for the known inhibitors 1PBC, niclosamide, and benzbromarone. If the authors' model is generally applicable, it should be able to predict the effect of experimentally validated compounds. Please include these data in the manuscript and address any inconsistencies.

(Remarks on code availability)

Version 3:

Reviewer comments:

Reviewer #3

(Remarks to the Author)

The authors have addressed the major concerns that I raised; below are some minor comments.

Line 780: This is the adjustment of the voltage offset when the pipette is inserted into the bath solution but is not the correction of liquid junction potential. If junction potentials were not corrected for, please state the magnitude and that junction potentials were not corrected for. Please revise.

Lines 790 and 792: Inconsistency between the stated pH values of the pipette solutions.

Lines 800 and 801: The seal between the membrane and the glass pipette remains, giga-ohm or not.

(Remarks on code availability)

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.

The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

Response Letter

On behalf of all the contributing authors, we would like to express our sincere appreciation for reviewers' constructive comments and helpful suggestions regarding our article NCOMMS-23-43911 entitled "*DTIAM: A unified framework for predicting drug-target interactions, binding affinities and activation/inhibition mechanisms*". In response to the feedback from the reviewers, we have made substantial revisions to our manuscript and have added extra materials to strengthen our results. In the following pages, we provided our detailed point-by-point responses to the review comments.

Reviewer #1:

The authors present a bioactivity prediction framework which includes quantitative target activity and functional effects.

The work conceptually makes sense, although it's more written in computer science-style, and it doesn't include prospective validation either which would be useful. The 'validations' given are more proof by example and based on existing literature, which I would see as quite weak a validation, and not really on the level I would like to see published in this journal. Also the datasets are too small to really be useful in practice, in particular for functional annotations (lines 562 ff.).

Authors' response:

We highly appreciate the Reviewer's recognition on the validity of our computational method. We agree that experimental validation of our model predictions would strengthen the validity further. Consequently, we have conducted the new wet-lab experiments according to your suggestions. In terms of datasets, we used several benchmark datasets on each task and common evaluation metrics in order to make a fair comparison with previous methods. In practical applications, such as the screening of TMEM16A inhibitors, we compiled a larger dataset from the Bindingdb database for training and prediction, and successfully identified effective and novel inhibitors of TMEM16A. Please find the results in our point-by-point response below.

Some items that should be addressed before publication in any case if the editor decides it is appropriate are:

1. 'binary' target prediction is obviously an arbitrary simplification of the task (lines 35 ff.), please point this out, also point out advantages (often better performance)

Authors' response:

We agree with the Reviewer that binary classification-based drug-target interaction prediction is a simplification of the task. It only focuses on predicting whether there is an interaction

between the drug and the target, but ignores the interaction strength information. Such models tend to have better performance. The regression-based affinity prediction allows for the inference of finer binding affinity scores. However, errors from experimental measurements also introduce more data noise into such models. Both types of models have their own advantages and disadvantages. We have pointed these out in the introduction section of the revised manuscript, also shown below:

The binary classification-based DTI prediction models can successfully predict whether there is an interaction between a drug and a target and have strong performance. The drug-target interaction is not a simple binary on-off relationship, but a continuous value named binding affinity, which reflects how tightly the drug compound binds to a particular target protein¹.

In order to further predict the putative strengths of the interactions, various regression-based models have been proposed to infer the binding affinities between drugs and targets¹⁻³. Binding affinity reflects how tightly the drug binds to a particular target, which is quantified by measures such as inhibition constant (K_i), dissociation constant (K_d), and the half-maximal inhibitory concentration (IC_{50}).

Reference

1. Pahikkala, Tapio, et al. "Toward more realistic drug–target interaction predictions." *Briefings in bioinformatics* 16.2 (2015): 325-337.
2. Nguyen, Thin, et al. "GraphDTA: predicting drug–target binding affinity with graph neural networks." *Bioinformatics* 37.8 (2021): 1140-1147.
3. Jiang, Mingjian, et al. "Drug–target affinity prediction using graph neural network and contact maps." *RSC advances* 10.35 (2020): 20701-20712.

2. Classification of methods in intro is a bit unusual; structure-; ligand-; and network-based methods might be more suitable

Authors' response:

Thanks for the suggestion. We have updated the classification of methods in the introduction section of the revised manuscript, also shown below:

Generally, the approaches for in silico DTI prediction can be divided into four major categories: structure-based approaches, ligand-based approaches, network-based approaches, and machine learning-based approaches.

3. General concept of DTIAM is similar to proteochemometrics approaches (although with learned features, and a different NN architecture etc); key PCM papers (Wikberg etc, <https://pubmed.ncbi.nlm.nih.gov/12851294/> is probably the first one but also many later ones) should also be cited in introduction and relationship between methods be pointed out

Authors' response:

Thanks for the suggestion. Wikberg et al.¹ provided a proteochemometric modeling approach to model the interaction of MSH peptides with the MCRs. In terms of feature representation, it used unlearnable physicochemical, structural, and binary descriptors to represent receptors and ligands. The relationship between ligand-receptor descriptors and affinities was modeled by applying the partial least squares latent structure projection (PLS) algorithm. These are fundamentally different from our method.

We have cited this work proposed by Wikberg et al.¹ in the introduction section of the revised manuscript, also shown below:

Recently, some proteochemometrics modeling and gene expression profiles-based approaches were proposed to predict the activation and inhibition relationships between drugs and targets¹⁻⁴. However, there is a manifest need for developing a more powerful deep learning model to predict and decipher MoA, which is crucial and challenging scientific research in drug development.

Reference

1. Wikberg, J. E. S., et al. "Melanocortin receptors: ligands and proteochemometrics modeling." *Annals of the New York Academy of Sciences* 994.1 (2003): 21-26.
2. Sawada, Ryusuke, et al. "Predicting inhibitory and activatory drug targets by chemically and genetically perturbed transcriptome signatures." *Scientific reports* 8.1 (2018): 156.
3. Mervin, Lewis H., et al. "Extending in silico protein target prediction models to include functional effects." *Frontiers in Pharmacology* 9 (2018): 613.
4. Lee, Won-Yung, Choong-Yeol Lee, and Chang-Eop Kim. "Predicting activatory and inhibitory drug–target interactions based on mol2vec and genetically perturbed transcriptomes." *bioRxiv* (2021): 2021-03.

4. Functional effects have been predicted before in target prediction and related papers should be cited (e.g. Mervin et al <https://www.frontiersin.org/articles/10.3389/fphar.2018.00613/full>)

Authors' response:

Thanks for the suggestion. We have cited these related papers in the revised manuscript.

5. Performance doesn't pay any attention to chemistry and applicability domain, would suggest to not only treat as computer science exercise if the authors have practical applicability in mind (for this also validation should be performed)

Authors' response:

Thanks for the suggestion. We have applied our model to predict and validate novel inhibitors of TMEM16A, which is a protein target involved in the growth and metastasis of several malignant tumors. Down-regulation of TMEM16A expression or functional activity has been shown to inhibit the growth of tumor cells¹. Specifically, we prioritized potential inhibitors for TMEM16A from a high-throughput screening molecular library using the DTI, DTA, and

MoA prediction model. First, the DTI prediction model of DTIAM was applied to score the compounds from a commercial high-throughput screening molecular library, which contains approximately 10 million compounds, and the top 50,000 molecules (~top 0.5%) were selected based on predicted interaction probabilities. Second, we filtered these compounds by the Lipinski rules and removed pan-assay interference compounds, and clustered these molecules based on their fingerprint similarities, obtaining approximately 3,500 clusters. Finally, 85 representative compounds were selected from the top-ranked clusters based on the predicted interaction, affinity, and inhibition scores.

The predicted compounds inhibiting the TMEM16A channel are shown in Supplemental Spreadsheet 1. We found that daidzein (ranked second in the predicted potential compounds) has been verified by Wang et al.² to be a high-potency TMEM16A inhibitor, with an IC_{50} of $1.39 \pm 0.59 \mu\text{M}$, while DTIAM inferred that its interaction probability with TMEM16A is 0.995, its inhibition probability is 0.999, and its affinity is $0.121 \mu\text{M}$.

Furthermore, we use the Auto-Docking Vina software to shortlist the predicted compounds for further experimental evaluation. We found that dehydrocostus lactone, ranked 12th in the predicted compounds, has a high Vina score with a relatively easier synthetic accessibility, and decided to test this compound experimentally. Specifically, whole-cell patch clamp was employed to study the inhibitory effect of different concentrations of dehydrocostus lactone on TMEM16A. The initial TMEM16A current is activated by 600 nM free Ca^{2+} in the pipette solution. As shown in **Figure R1**, when the concentration of dehydrocostus lactone increases, the current of TMEM16A is gradually inhibited. Furthermore, the high concentrations of dehydrocostus lactone can almost completely inhibit the current of TMEM16A. These results suggested that dehydrocostus lactone can inhibit TMEM16A channel in a dose-dependent manner. Taken together, the new validation results suggested that the DTIAM can be effectively used as a computational pipeline in combination with high-throughput screening to expedite the search for novel compounds targeting a specific target.

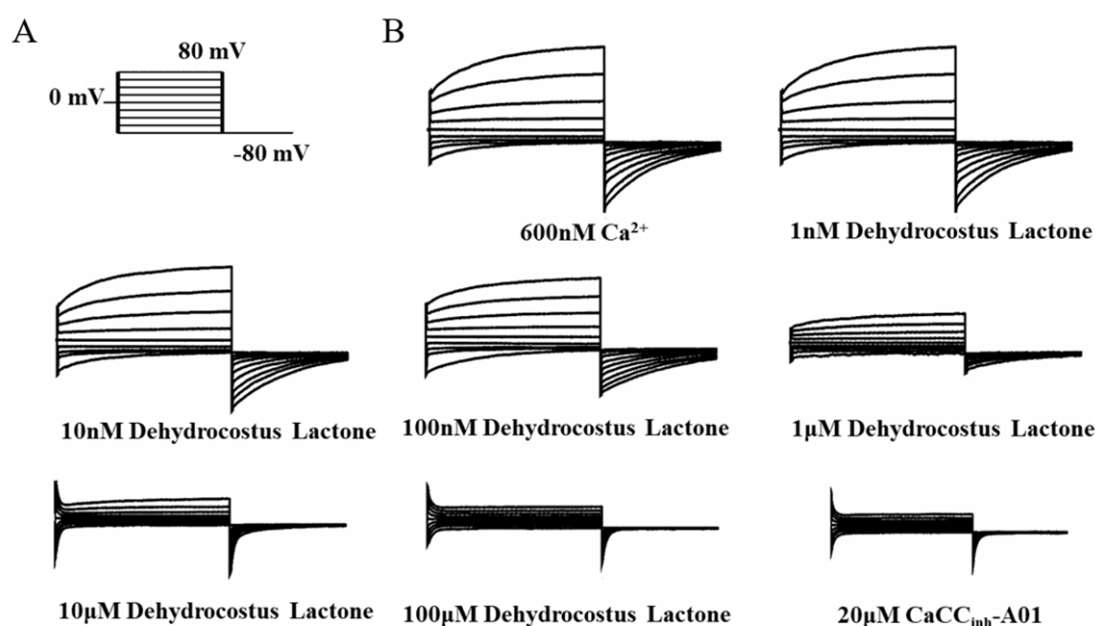


Figure R1 Dehydrocostus lactone inhibits TMEM16A current in a concentration-dependent manner. (A) Voltage stimulation waveform scheme. (B) Representative TMEM16A current is activated by 600 nM Ca^{2+} and inhibited by different concentrations of dehydrocostus lactone ($n = 4$), with 20 μM $\text{CaCC}_{\text{inh}}\text{-A01}$ as the inhibitor positive control.

Reference

1. Jia, Linghan, et al. "Inhibition of calcium-activated chloride channel ANO1/TMEM16A suppresses tumor growth and invasion in human lung cancer." *PloS one* 10.8 (2015): e0136584.
2. Wang, Xuzhao, et al. "Inhibitory effect of daidzein on the calcium-activated chloride channel TMEM16A and its anti-lung adenocarcinoma activity." *International Journal of Biological Macromolecules* 253 (2023): 127261.

6. 305: CYPs are listed; what do the authors mean by 'activating' CYPs, is this upregulation?

How are substrates treated?; Table 1 are not 'candidate drugs', please use appropriate terminology

Authors' response:

Cytochrome P450 (CYP) are a group of enzymes encoded by the P450 genes and responsible for the metabolism of most drugs seen in clinical practice. The expression of CYP enzymes varies between populations and will greatly influence drug metabolism and response. Activating CYP can upregulate corresponding genes to increase the expression of CYP enzymes, thereby accelerating drug metabolism.

In this work, we applied the trained model to infer the activating and inhibiting effects of a given compound on CYP, without considering what the corresponding substrate of the CYP enzyme is.

We have changed "candidate drugs " to "potential drugs " in the revised manuscript.

7. 2.5. is too small-scale and not a 'validation', same for 2.6., this is basically just proof-by-example

Authors' response:

We have modified these two subsections and also added the experimental validation of DTIAM predictions in the screening of TMEM16A inhibitors. As mentioned earlier, we applied the trained DTIAM model to infer potential inhibitors of TMEM16A from a high-throughput screening molecular library. We have successfully validated effective inhibitors of TMEM16A through a series of computational screening and molecular simulation scoring. These results indicate that DTIAM can provide a useful tool in drug screening.

So overall the limitations of the dataset and the performance metrics used; and the lack of any discussion of chemistry and prospective validation makes this referee conclude that this

manuscript isn't a very strong one, it would need to be extended along the above axes to get there.

Authors' response:

We understand the Reviewer's concern. In terms of datasets, we have used several benchmark datasets on each task and common evaluation metrics in order to make a fair comparison with previous methods. In terms of prospective validation, we have performed various validation experiments, including the validation of predicted potential drugs on EGFR and CDK 4/6, as well as experiments in which DTIAM distinguishes activation and inhibition on specific target proteins. In addition, we have added the wet-lab validation of DTIAM in the TMEM16A inhibitor screening. Based on the prediction scores of DTIAM and docking analysis, we have successfully identified effective and novel inhibitors of TMEM16A from a high-throughput molecular library (~10 million compounds). These results suggested that our method can provide a useful tool for efficient screening of potential molecules in practical drug discovery applications. We hope that with these new results we have addressed the concerns of the Reviewer.

Reviewer #2:

This paper presents a new approach for predicting DTIs, DTA and MoA which is entitled DTIAM. The approach leverages two transformer models, one for molecules and one for proteins, which have been pre-trained in a self-supervised manor on large volumes of unlabeled data. These two models are used to create features vectors from drug and protein pairs which are fed into another model to predict the likelihood of a binary interaction between a given pair. The DTIAM approach is shown to perform well compared to a series of competing approaches from the literature. The experiments consider both target and drug cold start problems, an inductive setup where proteins and drugs unseen during training are evaluated upon.

Overall I enjoyed the paper and found it overall easy to read. I did have some concerns with the current manuscript however:

Authors' response:

We thank the Reviewer for the overall positive comments. We have provided the point-by-point responses as below.

1. The contribution of the works seems quite low overall. Essentially this works takes two existing model types and combines them together. Could the authors articulate more clearly

what is the contribution of their work?

Authors' response:

In this paper, we have done a considerable amount of work. In summary, our contributions are: (1) We developed a molecular representation learning model based on self-supervised pre-training using massive unlabeled molecular data. The general knowledge of molecular substructure learned from the vast amount of existing molecules with self-supervised learning strategy have a greater ability for generalization to unseen drugs than existing methods. (2) For the first time, we provide a unified framework for accurate and robust prediction of drug-target interactions, binding affinities, and activation/inhibition mechanisms. Previous methods only focused on one of interaction prediction or affinity prediction, and were unable to distinguish the mechanism of action between the drug and the target. (3) We conduct comprehensive experiments on three tasks under several scenarios. Results show that our method consistently outperforms existing methods, particularly in the cold start scenario. Further practical application of drug screening demonstrates the reliability of DTIAM predictions and the candidate inhibitor of TMEM16A predicted by DTIAM was verified by wet-lab experiments.

2. For the molecular transformer could the authors clarify if this model is created by them? If so, why did they chose to do this instead of using an existing approach? For this model the authors propose a multi-part objective function was an ablation study performed in order to assess the impact of each one of these terms?

Authors' response:

Yes, the molecular transformer adopted in this work was created by us. We did consider existing approaches such as Mol2vec¹, ChemBERTa², MoLFormer³, which may perform better on molecular property prediction tasks, but their performance on drug-target interaction prediction tasks needs to be improved, especially for unseen molecules. We performed comparative experiments of these molecular representation learning models on the Yamanishi 08's dataset under three experimental settings. As shown in **Table R1**, DTIAM with BERMol model achieves substantial performance improvement over other models in all experimental scenarios, particularly in the drug cold start scenario. Therefore, we chose BERMol created by ourselves as the molecular representation learning module of our framework.

We have previously conducted ablation study to assess the impact of each self-supervised task, but we did not present the results in the paper. In the revised manuscript, we have added the ablation study and analyzed the experimental results. Specifically, to analyze the impact of using different combinations of self-supervised tasks for molecular representation learning pre-training, we conducted ablation experiments on seven combinations of three self-supervised tasks for our proposed molecular pre-training model. The three self-supervised tasks are: (1) Masked Language Modeling (MLM); (2) Molecular Descriptor

Prediction (MDP); and (3) Molecular Functional Group Prediction (MFGP). As shown in **Table R2**, we found that MFGP task has the highest impact on the performance, especially in the drug cold start scenario, with a significant leading margin of 50% in terms of AUPR when using the MFGP task alone (AUPR=0.305) versus using the MDP task alone (AUPR=0.203). This demonstrates that the local functional group information of the molecule is very helpful for DTI prediction. At the same time, our molecular representation learning model can accurately capture the key functional group information. Among the combinations of the two self-supervised tasks, the molecular model trained on both MLM and MFGP performed best. Although the best performing model is trained on the combination of MLM, MDP and MFGP, the additive gain from the MDP task is relatively minor.

Table R1 The experiment results of DTIAM using different molecule representation learning models.

Model	Warm Start	Drug Cold Start	Target Cold Start
Mol2vec	0.814	0.248	0.728
ChemBERTa	0.857	0.336	0.776
MoLFormer	0.903	0.377	0.81
BERMol	0.93	0.567	0.863

Note: We used different molecular representation learning models to conduct comparative experiments under the same division on the Yamanishi 08’s dataset, and obtained the AUPR results under three experimental settings. Bold numbers represent the best performance, higher values are better.

Table R2 The impact of pre-training BERMol with different self-supervised task combinations.

MLM	MDP	MFGP	Warm Start	Drug Cold Start	Target Cold Start
✓	×	×	0.697	0.247	0.543
×	✓	×	0.586	0.203	0.509
×	×	✓	0.783	0.305	0.594
✓	✓	×	0.82	0.418	0.673
✓	×	✓	0.902	0.527	0.802
×	✓	✓	0.851	0.489	0.769
✓	✓	✓	0.931	0.568	0.865

Note: We tested the impact of BERMol with different self-supervised task combinations on the Yamanishi 08’s dataset in terms of AUPR. Bold numbers represent the best performance, higher values are better.

Reference

1. Jaeger, Sabrina, Simone Fulle, and Samo Turk. "Mol2vec: unsupervised machine

learning approach with chemical intuition." Journal of chemical information and modeling 58.1 (2018): 27-35.

2. Chithrananda, Seyone, Gabriel Grand, and Bharath Ramsundar. "ChemBERTa: large-scale self-supervised pretraining for molecular property prediction." arXiv preprint arXiv:2010.09885 (2020).
3. Ross, Jerret, et al. "Large-scale chemical language representations capture molecular structure and properties." Nature Machine Intelligence 4.12 (2022): 1256-1264.

3. Did the authors consider other protein representation learning approaches than ESM? If so, why was ESM chosen?

Authors' response:

Yes, we did consider other protein representation learning approaches, such as TAPE¹, ESM-1b². Meanwhile, we also developed a protein representation learning model called BERProt to extract contextual information of protein residues from large-scale unlabeled data through multiple self-supervised pre-training tasks. However, the token size of protein sequences is much larger than molecular substructure tokens and requires more GPU resources. Limited by the GPU computing power and memory of our laboratory, it is difficult for us to train a model with the same parameter scale as ESM-2³ (created by Meta AI, FAIR Team). We performed comparative experiments of these protein representation learning models on the Yamanishi 08's dataset under three experimental settings. As shown in **Table R3**, DTIAM with ESM-2 model achieves better performance in various experimental scenarios, which can be attributed to ESM-2's larger parameter size and larger training data. Therefore, we chose ESM-2 as the protein representation learning module of our framework.

Table R3 The experiment results of DTIAM using different protein representation learning models.

Model	Parameters	Data size	Warm Start	Drug Cold Start	Target Cold Start
BERProt	56M	31M	0.913	0.531	0.812
TAPE	92M	31M	0.882	0.496	0.779
ESM-1b	650M	-	0.917	0.537	0.825
ESM-2	650M	60M	0.931	0.568	0.865

Note: We used different protein representation learning models to conduct comparative experiments under the same division on the Yamanishi 08's dataset, and obtained the AUPR results under three experimental settings. Bold numbers represent the best performance, higher values are better.

Reference

1. Rao, Roshan, et al. "Evaluating protein transfer learning with TAPE." Advances in neural information processing systems 32 (2019).
2. Rives, Alexander, et al. "Biological structure and function emerge from scaling

unsupervised learning to 250 million protein sequences." *Proceedings of the National Academy of Sciences* 118.15 (2021): e2016239118.

3. Lin, Zeming, et al. "Evolutionary-scale prediction of atomic-level protein structure with a language model." *Science* 379.6637 (2023): 1123-1130.

4. The actual model which predicts the DTIs is poorly explained and seemed to be created via some form of autoML package called AutoGluon. Why did the authors not use a neural network here? Did the authors consider updating the weights of the protein and drug transformer using the supervised signal from this model?

Authors' response:

In this work, our method mainly focuses on how to learn accurate representations of drugs and targets from massive unlabeled data through self-supervised pre-training, thereby improving the performance of various downstream prediction tasks. Our model is a unified framework designed with the concept that there is no need to tune the model architecture and parameters for different prediction tasks. AutoGluon¹ is an automated machine learning (AutoML) framework that can achieve strong predictive performance without the need for manual intervention, which exactly meets our needs. It is worth noting that AutoGluon integrates neural network models including NeuralNetFastAI, NeuralNetTorch, and various machine learning models. Limited by GPU resources and AutoGluon's model architecture, we have therefore not considered fine-tuning the representation learning model of drugs and targets during training.

Reference

1. Erickson, Nick, et al. "Autogluon-tabular: Robust and accurate automl for structured data." *arXiv preprint arXiv:2003.06505* (2020).

5. Overall I would expect to see more experiments performed ablating different parts of the model to assess the impact each one makes.

Authors' response:

Thanks for the suggestion. As mentioned before, our method mainly focuses on how to learn accurate representations of drugs and targets to improve the performance of various downstream prediction tasks. In this regard, we developed molecular representation learning and protein representation learning models based on self-supervised pre-training and compared them with different approaches, and then selected the best model as the representation learning module of the DTIAM framework. It is worth noting that by comparing **Tables R1** and **R3**, we found that training more accurate molecular representation models has a greater impact on downstream task performance. We also conducted ablation study to analyze the impact of using different combinations of self-supervised tasks for molecular representation learning pre-training. The results in **Tables R2** show that each self-supervised task has a different impact

on prediction performance, and the molecular pre-training model that combines three self-supervised tasks can achieve the best performance.

6. Whilst it is nice to see the code has been made available, it would be good if the authors could make the code more usable overall for example by have a more in depth readme to better explain how to use the code and how to replicate some of the key results.

Authors' response:

Thanks for the suggestion. We have already updated the README to provide more in-depth explanations on how to use the code and replicate key results. The file can be accessed at: <https://github.com/CSUBioGroup/DTIAM/blob/main/README.md>.

Response Letter

On behalf of all the contributing authors, we would like to express our sincere appreciation for reviewers' constructive comments and helpful suggestions regarding our article NCOMMS-23-43911 entitled "*DTIAM: A unified framework for predicting drug-target interactions, binding affinities and activation/inhibition mechanisms*". In response to the feedback from reviewers, we have made an additional revision to our manuscript and have added extra materials to strengthen our results. All changes can be seen in the revised manuscript with yellow highlighting. We hope that our revised version addresses the issues raised by the reviewers. In the following pages, we provided our detailed point-by-point responses to the review comments.

Reviewer #2 (Remarks to the Author):

Thanks for addressing my comment and improving the readme of the code.

I have been asked to comment on the concerns of reviewer #1.

1) Yes reviewer 1 has raised very valid technical questions – many of which overlapped with my own concerns. I do not disagree with any of reviewer 1's points.

2) The authors have taken steps to address many of the comments made from reviewer 1. I feel that point 7 could have been addressed more – the authors only provided 1 example, and I think more should be provided as this is the crucial result that validates their methodology.

Authors' response:

Thanks for the comment. In the revised manuscript, we conducted more in-depth and systematic experiments for the validation of TMEM16A inhibitors. Based on previous experiments, we further provided current-voltage curves and dose-response curves, confirming the concentration-dependent inhibitory effect of dehydrocostus lactone on TMEM16A. Ultimately, the IC₅₀ value was determined to be approximately 161 nM, which aligns closely with our predicted results (**Figure R1**). These results indicate that DTIAM can provide a useful tool in drug screening. In addition, we performed extensive computational validation experiments for drug-target interactions, affinity, and activation/inhibition mechanism predictions on multiple datasets. We also tested DTIAM's ability to distinguish between activation and inhibition on 10 specific targets and conducted independent validation of interaction predictions for EGFR and CDK 4/6 targets. All experimental results demonstrate the accuracy and effectiveness of DTIAM, which can provide an effective tool for drug discovery.

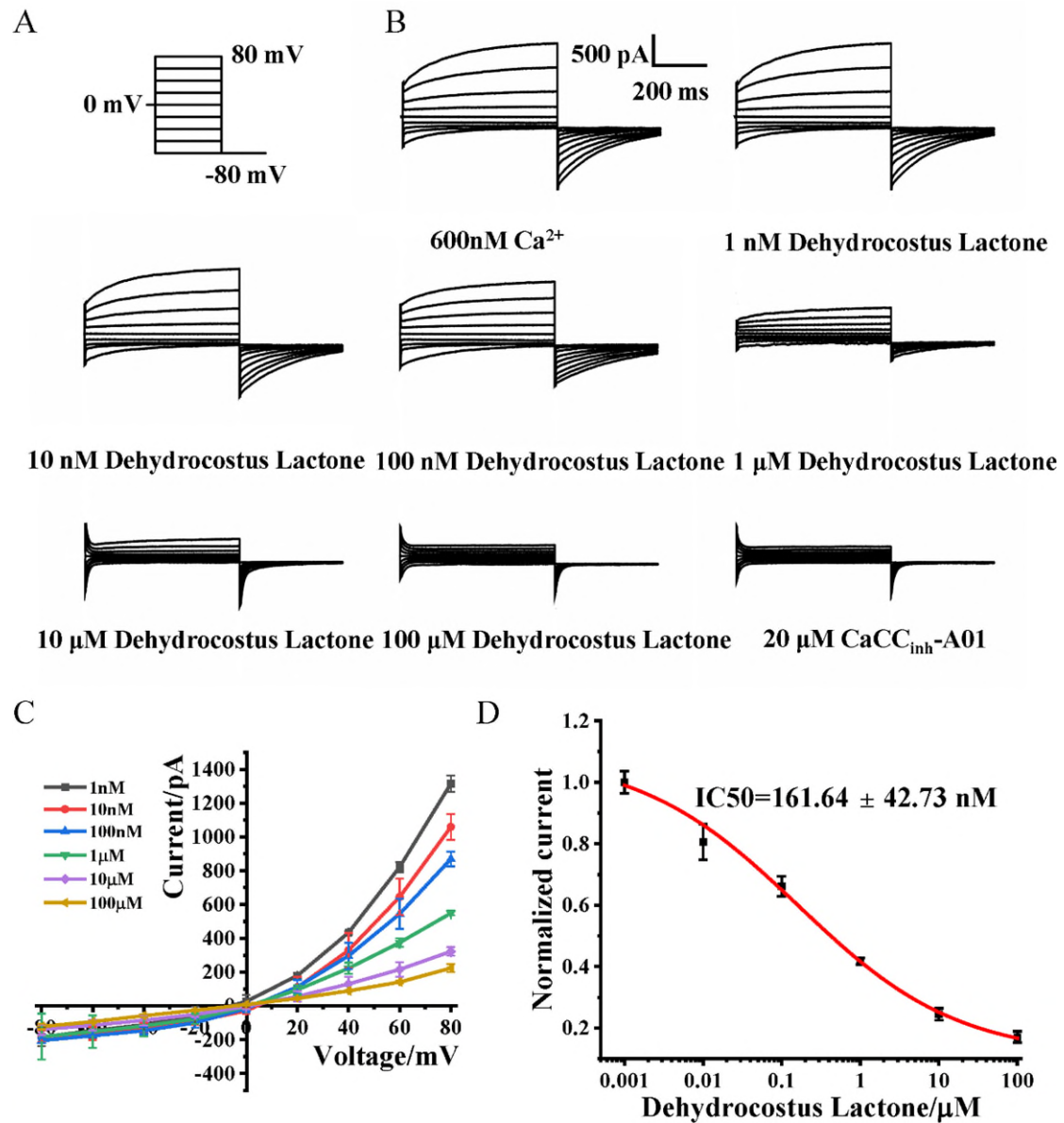


Figure R1 Dehydrocostus lactone inhibits TMEM16A current in a concentration-dependent manner. (A) Voltage stimulation waveform scheme. (B) Representative TMEM16A current is activated by 600 nM Ca^{2+} and inhibited by different concentrations of dehydrocostus lactone ($n=3$), with 20 μM $\text{CaCC}_{\text{inh}}\text{-A01}$ as the inhibitor positive control. (C) I-V curve of TMEM16A inhibited by various concentrations of dehydrocostus lactone. (D) Dose-response curve of TMEM16A currents at +80 mV incubated with indicated concentrations of dehydrocostus lactone. The IC_{50} of dehydrocostus lactone on TMEM16A channels was $161.64 \pm 42.73 \text{ nM}$.

Recently, we used DTIAM to participate in the Shanghai International Computational Biology Innovation Competition - Drug Screening AI Algorithm "Lingyue" Challenge, which

used N-methyl-D-aspartate (NMDA) receptors as targets for virtual drug screening and wet experiments for verification. NMDA receptors are associated with many diseases, such as stroke, depression, epilepsy, Alzheimer's disease, and pain. The development of drugs targeting NMDA receptors has important clinical significance and market value. The organizers selected the top-ranked candidate molecules we submitted for experimental validation. The results showed that 3 candidates had an IC₅₀ of less than 100 μM, and 2 had an IC₅₀ of less than 10 μM (3.49 μM and 4.64 μM), as illustrated in **Figure R2A**. These findings further demonstrate the effectiveness of DTIAM. However, due to intellectual property constraints, we were not authorized by the organizers to include this experimental data in our paper. Nonetheless, these results clearly demonstrate the robust utility of DTIAM, achieving excellent performance across various application scenarios.

In summary, we believe that the experiments in the revised manuscript are sufficient to demonstrate the accuracy and effectiveness of DTIAM, and the wet lab validation on TMEM16A and NMDA are sufficient for generalization of the utility of our framework.

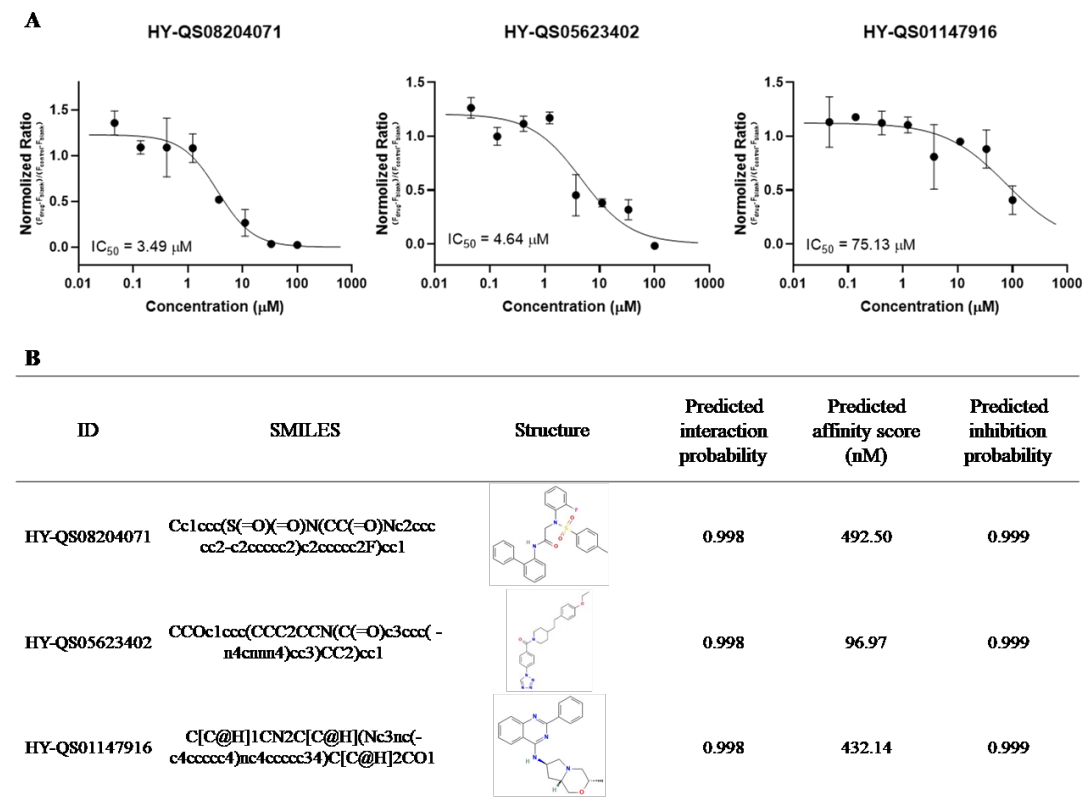


Figure R2 The experimental results of NMDA candidate molecules predicted by DTIAM. (A) Dose-response curves of 3 candidate molecules with NMDA inhibitory effects predicted by

DTIAM. The organizer used the FDSS/ μ Cell high-throughput screening equipment to perform multi-concentration fluorescence screening on the top-ranked candidate molecules and fit the dose-response curve to obtain the maximum half-inhibitory concentration (IC₅₀). We found that among the top-ranked candidate molecules, 3 candidates had an IC₅₀ of less than 100 μ M, and 2 had an IC₅₀ of less than 10 μ M (3.49 μ M and 4.64 μ M for HY-QS08204071 and HY-QS05623402). (B) SMILES, structure and DTIAM prediction of candidate molecules with NMDA inhibitory effects.

Reviewer #2 (Remarks on code availability):

The code would run however it could be improved and brought up to proper PEP8 standard.

Authors' response:

Thanks for the professional suggestion. We have brought up the code to the PEP8 standard. Specifically, we have made improvements in areas such as indentation, naming conventions, line length limits, spacing, and import order to comply with PEP8 standards. Additionally, we have added comments to important functions, classes, and variables.

Reviewer #3 (Remarks to the Author):

The authors present a prediction framework for drug-target interactions based on features of drug and protein pairs. Specific comments are listed below.

1. It is not clear to me how the sequence/residue information of the protein alone without any information of protein structure, together with some structural features of the compounds, can yield meaningful prediction. Does the prediction framework consider the structure of the protein targets? What kind of data was the model trained on and how does the model predict binding affinity, potency, and positive or negative efficacy?

Authors' response:

Thanks for this careful comment. In this work, we employ a protein language model ESM-2¹ to extract the structural representations of the protein targets. Similar to AlphaFold², ESM-2 is able to predict the atomic-level protein structure based solely on the amino acid sequence of proteins. As language models of protein sequences are scaled up, an atomic-resolution picture of protein structure emerges in the learned representations. For drug compounds, we pre-trained a

molecular representation learning model BERMol, which learns structural features of molecular substructures from large-scale unlabeled data with the language model and domain-relevant auxiliary tasks. So, we did consider the structure of the protein targets and the compounds in our prediction framework.

The protein language model is pre-trained on the protein sequence and structure data, and the molecular representation learning model is pre-trained on the molecular substructure sequence, descriptor and functional group data. Based on the learned representations of proteins and compounds, we used interaction, affinity, and mechanism of action data to train three models for predicting drug-target interactions, binding affinities and activation/inhibition mechanisms, respectively.

Reference

1. Lin, Zeming, et al. "Evolutionary-scale prediction of atomic-level protein structure with a language model." *Science* 379.6637 (2023): 1123-1130.
2. Jumper, John, et al. "Highly accurate protein structure prediction with AlphaFold." *nature* 596.7873 (2021): 583-589.

2. What are the selection criteria for deciding whether the data are to be included or not in the training set? It is not uncommon that different values of potency and positive or negative efficacy are reported in the literature for a given drug and target pair, and the sensitivity to the same compound can be influenced by external factors beyond the residue information of the protein target. Please discuss these aspects in the text.

Authors' response:

Thanks for the suggestion. In the cross-validation experiments, we use six benchmark datasets to evaluate the performance of our model. Each dataset is randomly split into k folds, and the model is trained k times, where each time a different fold is held out as the test set, and the remaining k-1 folds are used as the training set. In the screening of TMEM16A inhibitors, we selected the BindingDB¹ database to construct a universal dataset to train DTIAM. We clean the data using the following procedure:

(1) The “target type” attribute was set to “single protein”, and the “standard type” attribute constants Ki, Kd, IC50 and EC50;

(2) The “molecule type” was set to “Small molecule” and can be handled by RDKit²;

(3) Duplicate data points and drug-target pairs with different labels are excluded.

We agree with the reviewer that there are different affinity values and mechanisms of action for a given drug and target pair in the database and literature, which is also called noise data. In this work, these data points are excluded from the training set. And we also agree that the sensitivity of a protein target to a specific compound is influenced by many factors beyond the protein’s residue information. These factors, including the protein dynamics, protein mutations, cellular and membrane environment, etc., can alter the compound's binding efficiency and efficacy. In future work, we will consider the dynamics of proteins in the model to improve the robustness and interpretability of the model, and integrate protein mutation data to improve the sensitivity of the model at key mutation sites.

We have discussed these aspects in the discussion section of the revised manuscript, also shown below:

DTIAM uses only molecular SMILES and protein sequences as input, and it effectively improves the performance of downstream prediction tasks by using massive amounts of label-free data for self-supervised pre-training. However, the sensitivity of a protein target to a specific compound is influenced by many factors beyond the protein sequences information. These factors, including the protein dynamics, protein mutations, cellular and membrane environment, etc., can alter the compound's binding efficiency and efficacy. In future work, we will consider the dynamics of proteins in the model to improve the robustness and interpretability of the model, and integrate protein mutation information to improve the sensitivity of the model at key mutation sites.

Reference

1. Gilson, Michael K., et al. "BindingDB in 2015: a public database for medicinal chemistry, computational chemistry and systems pharmacology." *Nucleic acids research* 44.D1 (2016): D1045-D1053.
2. Landrum, Greg. "RDKit: A software suite for cheminformatics, computational chemistry, and predictive modeling." *Greg Landrum* 8.31.10 (2013): 5281.
3. The authors stated that their framework can accurately predict activation/inhibition mechanisms. In what sense has this been accomplished? Based on the TMEM16A example,

the framework did not result in any prediction of activation or inhibition mechanism.

Authors' response:

Our framework can accurately predict activation/inhibition mechanisms in cross-validation on benchmark datasets, leave-one-out validation on 10 specific targets, and screening of TMEM16A inhibitors. In the cross-validation experiments, we evaluated the performance of DTIAM and compared it with AI-DTA using 5-fold cross validation under three experimental settings. As shown in **Figure R1A-D**, DTIAM significantly outperformed AI-DTI in all three experiment settings, on both Activation and Inhibition datasets. Specifically, on the Activation dataset, DTIAM outperformed AI-DTI with 16.1%, 17.9%, and 26.8% improvement in terms of AUPR in the scenario of the warm start, the drug cold start, and the target cold start, respectively. In the leave-one-out validation experiments, we evaluated the reliability of DTIAM in distinguishing activation and inhibition interactions on 10 specific targets (including ADRA1A, ADRA1B, ADRA1D, CYP3A43, CYP3A5, CYP3A7, DRD1, HTR1B, HTR1D, and HTR2A). The experimental results are shown in **Figure R1E-H**, including accuracy, recall of activation, recall of inhibition, and AUROC. We found that the prediction accuracy exceeded 93% for all targets, including 100% for HTR1D, and 9 of the targets had AUROC values above 0.96. These results suggest that DTIAM can accurately distinguish the activation and inhibition relationships between drugs and targets.

In the validation experiment of TMEM16A, our framework predicts the interactions, affinities, and activation/inhibition scores of TMEM16A with the molecules from the high-throughput screening molecular library. Since we are screening for inhibitors of TMEM16A, we ranked the candidate molecules based on the predicted interactions, affinity, and inhibition scores. We found that daidzein (ranked second in the predicted potential compounds) has been verified to be a high-efficiency TMEM16A inhibitor, with an IC₅₀ of 1.39 ± 0.59 μ M, while DTIAM inferred that its interaction probability with TMEM16A is 0.995, its inhibition probability is 0.999, and its affinity is 0.121 μ M. Furthermore, we found that dehydrocostus lactone (ranked 12th in the predicted potential compounds) was verified its inhibitory effect with TMEM16A by wet-lab experiments, and DTIAM inferred that its interaction probability with TMEM16A is 0.976, its inhibition probability is 0.989, and its affinity is 205.84 nM. The whole-cell patch clamp experiments indicated that the IC₅₀ value for dehydrocostus lactone

in inhibiting TMEM16A-mediated currents was 161.64 ± 42.73 nM, which is broadly consistent with our predictions.

Details about these validation experiments can be found in section 2.4 and 2.5 of the manuscript.

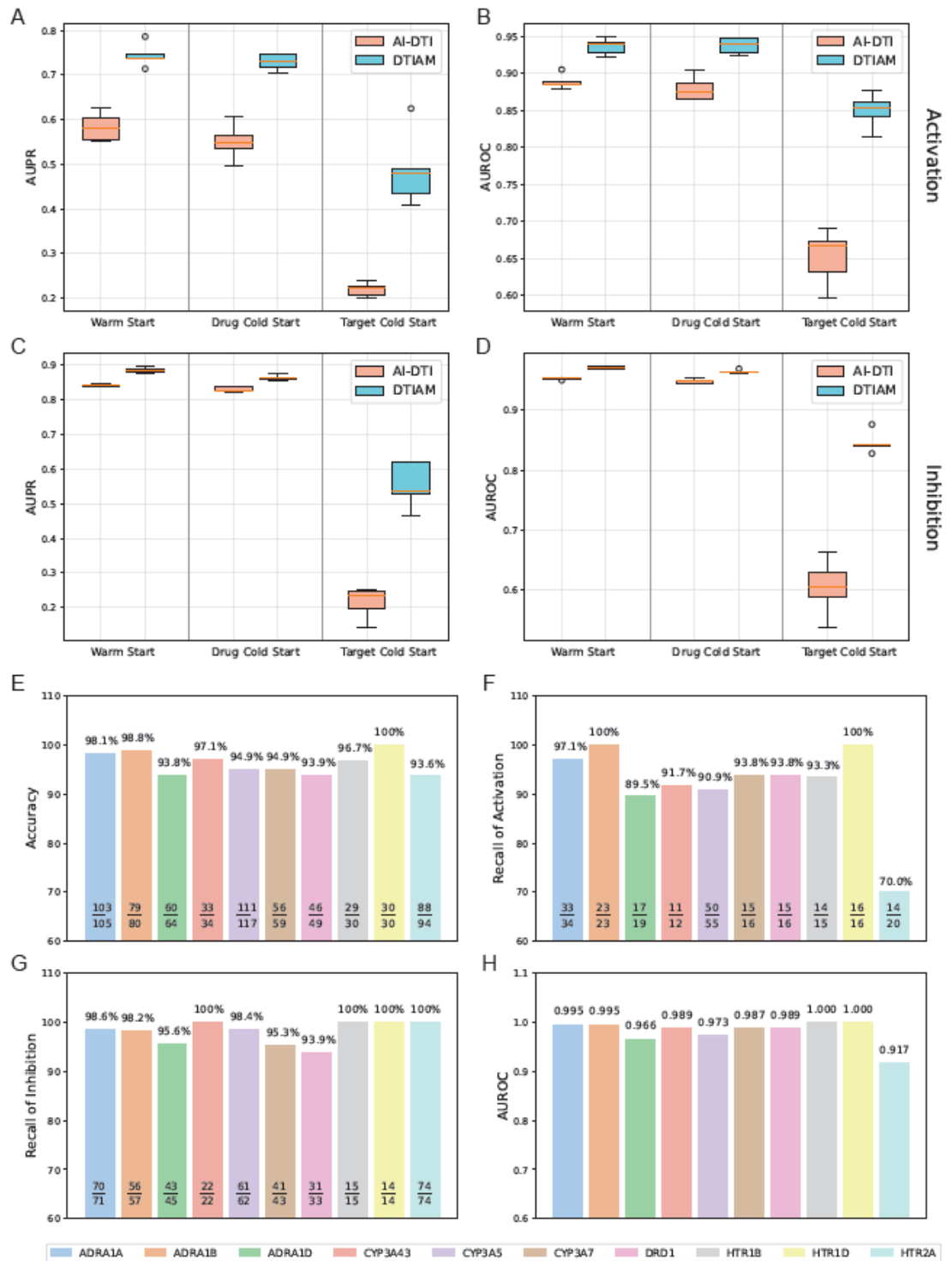


Figure R1 Performance evaluation on the MoA prediction task. (A-D) Performance comparison of DTIAM with AI-DTI on the Activation and Inhibition datasets under three

experiment settings in terms of AUPR and AUROC. All results were obtained by 5-fold cross-validation. The ratio between the positive and negative samples is 1:10. (E-H) Validating DTIAM on 10 specific targets. (E) Prediction accuracy for each target, with the score at the top of the bar and the ratio inside indicating the number of correctly predicted drugs out of the total number. (F-G) The recall of activation and inhibition, respectively, with the top value indicating recall and the ratio inside indicating the number of correctly predicted activators or inhibitors out of the total number. (H) AUROC score for each target.

4. The authors should justify their choice of TMEM16A as the only chosen target for experimentally testing their framework.

Authors' response:

Thanks for the suggestion. TMEM16A is widely expressed in many cells including epithelia, airway and vascular smooth muscle cells, vascular endothelium, cardiac muscles, olfactory sensory neurons, somatosensory neurons interstitial cells of Cajal (ICC), and nociceptive neurons¹. TMEM16A regulates many cellular functions, such as fluid secretion in secretory epithelia, smooth muscle contraction, gut mobility, cell volume regulation, apoptosis, and pain². In addition, the dysfunction of TMEM16A can lead to many diseases, including various cancers, gastrointestinal (GI) disorders, hypertension, and cystic fibrosis³. TMEM16A inhibitors have potential therapeutic uses in conditions such as hypertension (pulmonary and systemic), stroke, and overactive bladder while activators could be used in the treatment of disorders associated with epithelial dysfunction such as CF, chronic obstructive pulmonary disease (COPD), Sjögren's syndrome, and dry eye syndrome⁴. TMEM16A is expressed in more than a single organ or tissue; this might imply that drugs acting on TMEM16A may lead to effects on more than one site. Consequently, TMEM16A represents a significant pharmacological target. Simultaneously, DTIAM serves as a comprehensive framework for predicting drug-target interactions, binding affinities, and the mechanisms of activation or inhibition for targets such as TMEM16A. We utilized DTIAM to predict dehydrocostus lactone, a potential inhibitor of TMEM16A, and subsequently validated this prediction through the whole-cell patch clamp experiments. In summary, there are exciting times ahead for the development of new specific TMEM16A modulators with the appropriate potency,

selectivity, pharmacokinetic, and safety profile for the therapeutic exploitation of the TMEM16A ion channel target.

We have modified the text in section 2.5 of the manuscript to incorporate the justification of TMEM16A, also shown below:

TMEM16A is widely expressed in various cells, including epithelial, smooth muscle, endothelial, and neuronal cells, and regulates key functions such as fluid secretion, smooth muscle contraction, gut motility, cell volume regulation, apoptosis, and pain^{1,2}. Its dysfunction can lead to many diseases, including various cancers, gastrointestinal disorders, hypertension, and cystic fibrosis³. TMEM16A inhibitors have potential therapeutic uses in hypertension, stroke, and overactive bladder, while activators may treat epithelial dysfunction disorders like cystic fibrosis, and dry eye syndrome⁴. Consequently, TMEM16A represents a significant pharmacological target. Identifying potential drugs targeting TMEM16A through computational models is a challenging task with important implications for the treatment of various diseases.

Reference

1. Wang, Hui, et al. "Cell-specific mechanisms of TMEM16A Ca²⁺-activated chloride channel in cancer." *Molecular Cancer* 16.1 (2017): 152.
2. Liu, Yani, Liu, Zongtao, & Wang, Kewei. "The Ca²⁺-activated chloride channel ANO1/TMEM16A: An emerging therapeutic target for epithelium-originated diseases?" *Acta Pharmaceutica Sinica B* 11.6 (2021): 1412-1433.
3. Bai, Weiliang, Liu, Mei & Xiao, Qinghuan. "The diverse roles of TMEM16A Ca²⁺-activated Cl⁻ channels in inflammation." *Journal of Advanced Research* 4.33 (2021): 53-68.
4. Al-Hosni, Rumaitha, Ilkan, Zeki, Agostinelli, Emilio & Tammam, P. "The pharmacology of the TMEM16A channel: therapeutic opportunities." *Trends in Pharmacological Sciences* 43.9 (2022): 712-725.

5. The authors should test more than just one class of proteins and more than just one compound for the experimental validation of their prediction framework. For instance, well-characterized kinases, enzymes, transporters, other ion channels, receptors etc. should be tested. In my opinion, one specific compound tested on one specific target is insufficient for generalization of the utility of the presented framework.

Authors' response:

Thanks for the comment. In the revised manuscript, we conducted more in-depth and systematic experiments for the validation of TMEM16A inhibitors. Based on previous experiments, we further provided current-voltage curves and dose-response curves, confirming the concentration-dependent inhibitory effect of dehydrocostus lactone on TMEM16A. Ultimately, the IC₅₀ value was determined to be approximately 161 nM, which aligns closely with our predicted results. These results indicate that DTIAM can provide a useful tool in drug screening. In addition, we performed extensive computational validation experiments for drug-target interactions, affinity, and activation/inhibition mechanism predictions on multiple datasets. We also tested DTIAM's ability to distinguish between activation and inhibition on 10 specific targets and conducted independent validation of interaction predictions for EGFR and CDK 4/6 targets. All experimental results demonstrate the accuracy and effectiveness of DTIAM, which can provide an effective tool for drug discovery.

Recently, we used DTIAM to participate in the Shanghai International Computational Biology Innovation Competition - Drug Screening AI Algorithm "Lingyue" Challenge, which used N-methyl-D-aspartate (NMDA) receptors as targets for virtual drug screening and wet experiments for verification. NMDA receptors are associated with many diseases, such as stroke, depression, epilepsy, Alzheimer's disease, and pain. The development of drugs targeting NMDA receptors has important clinical significance and market value. The organizers selected the top-ranked candidate molecules we submitted for experimental validation. The results showed that 3 candidates had an IC₅₀ of less than 100 μ M, and 2 had an IC₅₀ of less than 10 μ M (3.49 μ M and 4.64 μ M), as illustrated in **Figure R2A**. These findings further demonstrate the effectiveness of DTIAM. However, due to intellectual property constraints, we were not authorized by the organizers to include this experimental data in our paper. Nonetheless, these results clearly demonstrate the robust utility of DTIAM, achieving excellent performance across various application scenarios.

In summary, we believe that the experiments in the revised manuscript are sufficient to demonstrate the accuracy and effectiveness of DTIAM, and the wet lab validation on TMEM16A and NMDA are sufficient for generalization of the utility of our framework.

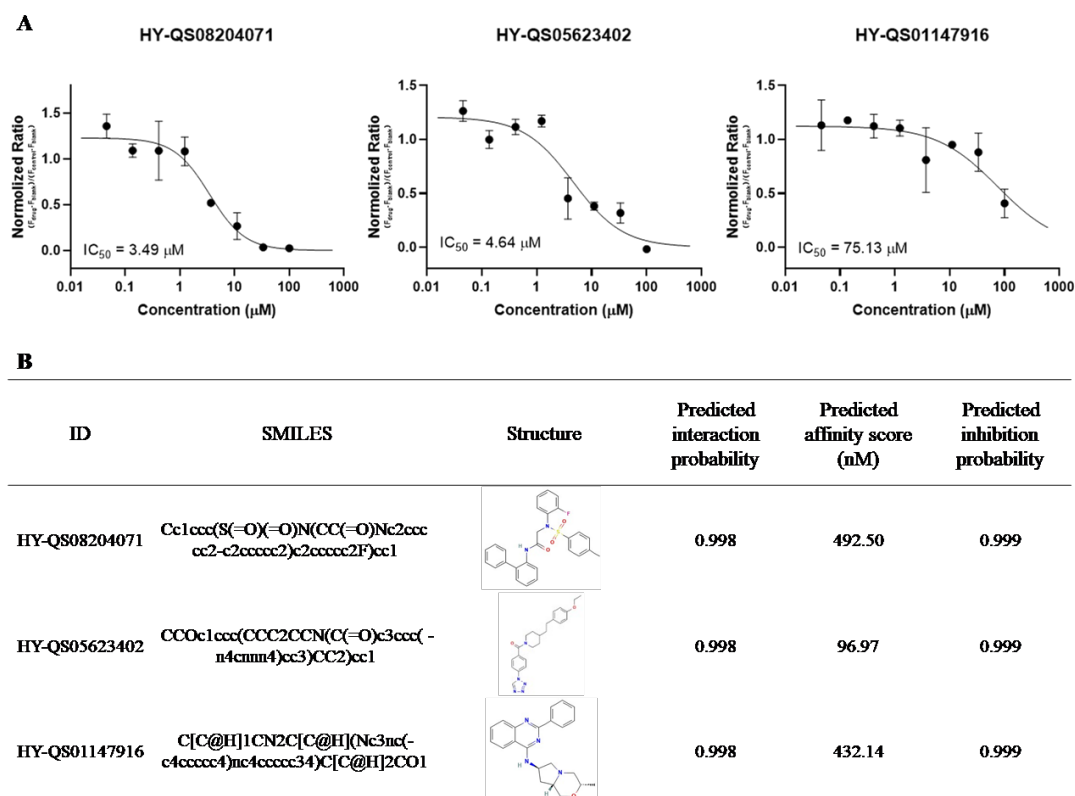


Figure R2 The experimental results of NMDA candidate molecules predicted by DTIAM. (A) Dose-response curves of 3 candidate molecules with NMDA inhibitory effects predicted by DTIAM. The organizer used the FDSS/ μ Cell high-throughput screening equipment to perform multi-concentration fluorescence screening on the top-ranked candidate molecules and fit the dose-response curve to obtain the maximum half-inhibitory concentration (IC₅₀). We found that among the top-ranked candidate molecules, 3 candidates had an IC₅₀ of less than 100 μ M, and 2 had an IC₅₀ of less than 10 μ M (3.49 μ M and 4.64 μ M for HY-QS08204071 and HY-QS05623402). (B) SMILES, structure and DTIAM prediction of candidate molecules with NMDA inhibitory effects.

6. Please provide the experimental information on the construct (species, UniProt ID, details of the construct etc.) and cell culture (cell lines, media, cultivation, transfection etc.). For electrophysiology, please provide information on series resistance and its compensation, liquid junction potential and whether this has been corrected for, the perfusion system, and whether the glass pipettes were fire-polished. Please move the methods to the ‘methods’ section.

Authors' response:

Thanks for the suggestion. In our experiment, lung adenocarcinoma cells with high expression of TMEM16A were used, and HEK cells were not used for transfection, so there was no construct information. For cell culture, the LA795 cells are mouse-derived lung adenocarcinoma cells, cultured in 1640 medium containing 10% fetal bovine serum in a 5% CO₂ incubator at 37 °C. For electrophysiology, the series resistance has been compensated, the junction potential has been corrected, the perfusion system has been cleaned before each experiment, and the glass pipette has been fire-polished.

The experimental details have been supplemented in the method section of the revised manuscript, also shown below:

DTIAM was applied to screen potential inhibitors of TMEM16A from a high-throughput molecular library, and the dehydrocostus lactone was selected for further wet lab validation. Specifically, whole-cell patch clamp was employed to study the inhibitory effect of different concentrations of dehydrocostus lactone on TMEM16A. Before performing the whole-cell patch clamp experiment, the LA795 cells were cultured on cover glass slides. The LA795 cells are mouse-derived lung adenocarcinoma cells, cultured in 1640 medium containing 10% fetal bovine serum in a 5% CO₂ incubator at 37 °C. Patch software was adopted for patch-clamp recording at room temperature using Pulse software with Digi LIH1600 interface (HEKA) to control EPC10 amplifiers. The borosilicate glass patch pipettes with a pipette resistance of 3-5 M in the bath solution (Sutter Instruments, Novato, USA) were drawn by a P-97 puller (Sutter Instruments, Novato, USA). The recording was done with an EPC10 amplifier controlled by Pulse software and a Digi LIH1600 interfaced (HEKA, Lambrecht, Germany). The data was sampled at 10 kHz after being low pass filtered at 2.9 kHz. The stimulation procedures included 150 ms voltage steps from a holding potential of 0 mV and membrane voltage (mV) clamping in 20 mV increments from -80 to +80 mV, with a -80 mV hold at room temperature (22-25 °C). The pipette solution of whole-cell patch clamp experiments contained CsCl (130 mM), ethylene glycol-bis(B-aminoethyl ether)-N,N,N',N'-tetraacetic acid (EGTA) (10 mM), MgATP (1 mM), MgCl₂·6H₂O (1 mM), and 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) (10 mM) and adjusted to pH 7.3 using CsOH. Whole-cell patch clamps were performed with a bath solution containing 150

mM NaCl, 1 mM MgCl₂·6H₂O, 10 mM HEPES, 10 mM glucose, and 10 mM mannitol, adjusted to pH 7.4 using NaOH. A 600 nM free Ca²⁺ bath was prepared by adding a standard CaCl₂ solution (1 M) to a final concentration of 8.69 mM and adjusting the pH to 7.4 with CsOH. The osmotic pressure was determined using the OM815 osmometer. The osmotic pressure of the pipette and the bath is in the range of 290-300 mOsm/L and 300-310 mOsm/L, respectively. Finally, we found that as the concentration of dehydrocostus lactone increases, the current of TMEM16A is gradually inhibited, while high concentrations of dehydrocostus lactone can almost completely inhibit the current of TMEM16A. In a word, whole-cell patch clamp data indicated that dehydrocostus lactone inhibits TMEM16A channel in a dose-dependent manner.

7.. Please plot the 1) the current-voltage relations for each dehydrocostus lactone overlaid in the same graph and 2) the concentration-response relation of dehydrocostus lactone from the whole-cell experiments.

Authors' response:

Thanks for the suggestion. We have plotted these two results in Fig. 5CD of the revised manuscript, also shown in Figure R3CD.

Whole-cell patch clamp experiments were conducted to assess the dose-dependent influence of dehydrocostus lactone on TMEM16A. A pipette solution containing 600 nM Ca was used to initiate TMEM16A current in LA795 cells, and dehydrocostus lactone was introduced to the bath solution to establish a concentration gradient within the perfusion system. The typical outward rectification and the slow activation kinetics of the TMEM16A current remained largely unaffected, as illustrated in **Figure R3BC**. The dose-response relationship was determined by plotting the normalized current values following exposure to various concentrations of dehydrocostus lactone at a holding potential of +80 mV. The curve was then fitted using the Hill equation. The findings indicated that the IC₅₀ value for dehydrocostus lactone in inhibiting TMEM16A-mediated currents was 161.64 ± 42.73 nM (**Figure R3D**), which is broadly consistent with our predictions. These results from patch clamp experiments demonstrated that dehydrocostus lactone was a putative inhibitor of TMEM16A.

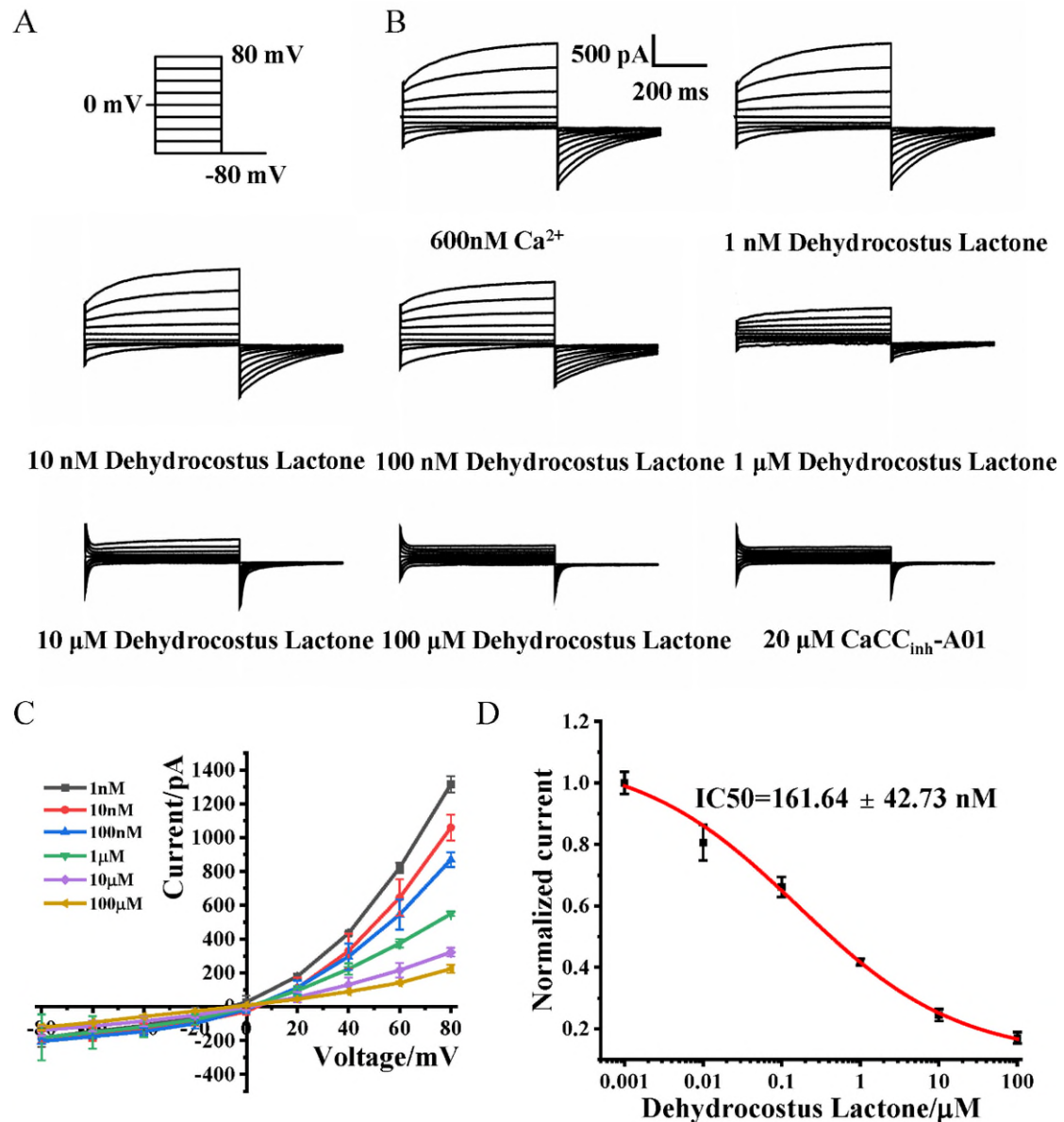


Figure R3 Dehydrocostus lactone inhibits TMEM16A current in a concentration-dependent manner. (A) Voltage stimulation waveform scheme. (B) Representative TMEM16A current is activated by 600 nM Ca^{2+} and inhibited by different concentrations of dehydrocostus lactone ($n=3$), with 20 μM $\text{CaCC}_{\text{inh}}\text{-A01}$ as the inhibitor positive control. (C) I-V curve of TMEM16A inhibited by various concentrations of dehydrocostus lactone. (D) Dose-response curve of TMEM16A currents at +80 mV incubated with indicated concentrations of dehydrocostus lactone. The IC_{50} of dehydrocostus lactone on TMEM16A channels was 161.64 ± 42.73 nM.

8. Please plot the scale bars, namely the current and time axes in Fig. 5.

Authors' response:

Thanks for pointing this out. We have plotted the scale bars of the current and time axes in Fig. 5B of the revised manuscript, also shown in **Figure R3B**.

9. Please perform outside-out patch-clamp experiments to verify that dehydrocostus lactone can act on excised membrane patches expressing TMEM16A.

Authors' response:

Thanks for the constructive suggestion. In response to your request for outside-out patch-clamp experiments to verify that dehydrocostus lactone can act on excised membrane patches expressing TMEM16A, we have given this considerable thought. While we agree that such experiments would provide valuable additional data, we regret to inform you that our laboratory currently does not have the resources or expertise to perform these specific experiments. We believe that the whole-cell patch clamp data in our manuscript are sufficient to support our conclusions.

10. What is the predicted potency and efficacy of dehydrocostus lactone and how do these compare to the experimentally measured potency and efficacy for TMEM16A? The authors should comment on this in the text.

Authors' response:

Our framework predicts an affinity value of 205.84 nM between dehydrocostus lactone and TMEM16A, with the inhibition probability at 0.989 and the interaction probability at 0.976. These predictions serve as a guide for further experimental validation. The whole-cell patch clamp experiments indicated that the IC₅₀ value for dehydrocostus lactone in inhibiting TMEM16A-mediated currents was 161.64 ± 42.73 nM, which is broadly consistent with our predictions. We have commented on this in the revised manuscript.

11. The authors stated that the predicted affinity is 0.121 uM for the TMEM16A inhibitor daidezein, while the previously reported potency is 1.39 uM. While affinity and potency are different properties, a correspondence between the two is expected. Have the authors verified this previously reported potency? Please comment on this in the text.

Authors' response:

Drug-target affinity (DTA) prediction is a regression task, and it is extremely challenging to accurately predict the affinity between drugs and targets. The real affinity training data follows a long-tail distribution, and it is difficult for the model to directly fit the training data. Therefore, we use a log transformation to make the data distribution more closely to the normal distribution to reduce the complexity of model training. However, this introduces another issue: while the prediction error on the log scale might be small, once transformed back to the original scale, the error could be larger. Although there is a certain discrepancy between 0.121 uM and 1.39 uM, the error on the log scale is acceptable. Additionally, the lower value of affinity indicates higher strength binding. The predicted affinity value between daidezein and TMEM16A is lower than the actual value suggests that the model has higher confidence in daidezein. We have commented on this in the revised manuscript.

12. Considering that TMEM16A was used as a specific example, the authors should omit data on TMEM16A and other TMEM16 proteins in the training set and quantify the accuracy and precision of the model in predicting TMEM16A modulators. In particular, the authors should state if the known blockers/inhibitors 1PBC, niclosamide, and benzbromarone can be successfully predicted in terms of their potency and efficacy. Please include this prediction data and address any discrepancies in the text.

Authors' response:

We agree with the Reviewer that the data associated with the specific target should be omitted in the training set for validation of the specific example. In the leave-one-out validation of 10 specific targets (including ADRA1A, ADRA1B, ADRA1D, CYP3A43, CYP3A5, CYP3A7, DRD1, HTR1B, HTR1D, and HTR2A), we did remove the data associated with each target in the training set, and evaluated the prediction accuracy of DTIAM on each target. As shown in **Figure R1E-H**, the prediction accuracy exceeded 93% for all targets, including 100% for HTR1D, and 9 of the targets had AUROC values above 0.96, details can be found in section 2.4 of the manuscript. While in the case of TMEM16A, we aim to screen the novel inhibitors of TMEM16A from molecular library by learning known relationships in the training data. So we did not omit TMEM16A-related data in the training set. In fact, we successfully identified

effective and novel inhibitors of TMEM16A through known relationships.

13. Please comment on how quantitatively the described methods can predict experimental potency and efficacy in the text. The authors should additionally quantify the accuracy of the predictions using data on experimental potency and efficacy, for instance by investigating their correspondence, in a systematic manner.

Authors' response:

Thanks for the suggestion. In our proposed method, we quantitatively predict drug-target interactions, binding affinity, and activation/inhibition mechanisms. The prediction of drug-target interaction is a binary classification task where our model infers the probability of interaction between a given drug and target. A predicted probability closer to 1 indicates a higher likelihood of interaction, while a probability closer to 0 suggests a lower likelihood. Similarly, the prediction of activation/inhibition mechanisms is also a classification task: a predicted activation probability closer to 1 implies a higher chance of activation efficacy, while a predicted inhibition probability closer to 1 indicates a higher likelihood of inhibition efficacy. The prediction of binding affinity is a regression task, where our model quantitatively predicts the binding affinity value between a drug and its target.

We did quantify the accuracy of the predictions using various benchmark datasets consisting of experimentally validated data. For classification tasks (interaction and activation/inhibition mechanism), we used the area under the precision-recall curve (AUPR) and the area under the receiver operating characteristics curve (AUROC) as evaluation metrics, where higher values of AUPR and AUC indicate better predictive performance. For regression tasks (binding affinity prediction), we employed the mean squared error (MSE) and the Pearson correlation as evaluation metrics, with smaller MSE values indicating less deviation from experimental values and higher Pearson correlation indicating stronger correlations between predicted and experimental results. Detailed quantitative evaluation results can be found in sections 2.2, 2.3, and 2.4 of the manuscript. We hope that our response addresses the concerns raised by the reviewer.

Response Letter

On behalf of all the contributing authors, we would like to express our sincere appreciation for reviewers' constructive comments and helpful suggestions regarding our article NCOMMS-23-43911 entitled "*DTIAM: A unified framework for predicting drug-target interactions, binding affinities and activation/inhibition mechanisms*". In response to the feedback from reviewers, we have made an additional revision to our manuscript and carefully addressed all the comments and suggestions. All changes can be seen in the revised manuscript with yellow highlighting. We hope that our revised version addresses the issues raised by the reviewers. In the following pages, we provided our detailed point-by-point responses to the review comments.

Reviewer #2 (Remarks to the Author):

Thanks to the authors for addressing my comments.

Authors' response:

We are deeply thankful for the comments and suggestions that have greatly improved our manuscript.

Reviewer #2 (Remarks on code availability):

Code is much better thank you. I would also consider adding type hints for your code as well.

Also why is PyTorch not in the requirements file?

Authors' response:

Thanks for the professional suggestion. We have added type hints for the code, which significantly improves its readability.

For the requirements file, we have provided a separate command line for PyTorch: "conda install pytorch==1.12.1 torchvision==0.13.1 torchaudio==0.12.1 cudatoolkit=11.3 -c pytorch -y" in the README.md file to enhance the success rate of setting up the environment. At the same time, we have included these packages in the new version of the requirements file.

Reviewer #3 (Remarks to the Author):

The authors should not claim that their model can predict mechanisms without any experimental support. Please remove this claim throughout the manuscript.

Authors' response:

Thanks for the comment. We have double-checked to ensure that the revised manuscript does not include this claim.

Unaddressed concerns:

Point 6 –

1. The authors should use a recombinant system such as HEK293T cells to test the effect of dehydrocostus lactone on TMEM16A; this is standard practice.

Authors' response:

Thank you for your constructive comments. We totally agree that HEK293T recombinant cells is more suitable to test the effect of dehydrocostus lactone on TMEM16A. Therefore, we conducted a series of experiments utilizing TMEM16A recombinant HEK293T cells and obtained similar results with those in LA795 cells. HEK293T is a recombinant system that exclude the influence of other components in LA795 cells, and these two similar sets of results demonstrate that dehydrocostus lactone is a highly effective inhibitor of TMEM16A. After careful consideration, we decide to put the experimental results of the those in HEK293T transfected with TMEM16A in the main text and move the experimental results in LA795 cells to the supplementary materials. The description and details of the experiments have been updated in Sections 2.6 and 4.6 of the revised manuscript, also shown below:

Sections 2.6

Specifically, to investigate the inhibitory effect of dehydrocostus lactone on TMEM16A channels, we performed whole-cell patch clamp recordings in HEK293T cells transfected with TMEM16A plasmids. The pipette solution containing 600 nM Ca^{2+} was adopted to elicit TMEM16A currents in the transfected HEK293T cells. As illustrated in Fig. 5BC, the characteristic outward rectification and slow activation kinetics of the TMEM16A current were obvious. The dose-response relationship was established by plotting normalized current

values at a holding potential of +80 mV following exposure to various concentrations of dehydrocostus lactone. Our findings revealed an IC_{50} value of 111.97 ± 22.96 nM for dehydrocostus lactone in inhibiting TMEM16A transfected HEK293T cells (Fig. 5D). We further verified the inhibition effect of dehydrocostus lactone on TMEM16A with an outside-out patch clamp. It was demonstrated that 100 μ M of dehydrocostus lactone can completely inhibit the current of TMEM16A (Fig. 5EF), indicating that the inhibition of dehydrocostus lactone on TMEM16A was directly exerted. In addition, we performed experiments on LA795 cell with high endogenous expression of TMEM16A, which also proved the inhibitory effect of dehydrocostus lactone on TMEM16A (Supplementary Materials Fig. 1). The IC_{50} of dehydrocostus lactone inhibiting TMEM16A in LA795 cell was 161.64 ± 61.96 nM. All these experimental results demonstrate that dehydrocostus lactone predicted by DTIAM is a highly effective inhibitor of TMEM16A.

Sections 4.6

4.6.1 Cell culture and transfection

HEK293T cells were maintained in Dulbecco's modified eagle medium (DMEM) (Gibco), supplemented with 10% fetal calf serum (FBS) (Gibco), and 1% penicillin–streptomycin solution (100 U/ml penicillin and 100 μ g/ml streptomycin) in a humidified incubator at 37°C with 5% CO₂. The cells were transiently transfected with a cDNA for mouse TMEM16A (mTMEM16A), using Lipofectamine 3000 Transfection Reagent (Thermo Fisher Scientific) with a ratio of 1:3. The mouse cDNA clone mTMEM16A (ANO1, Accession Number: NM_178642.5) was a kind gift from Prof. Young Duk Yang (Seoul National University, Korea) and was subcloned into expression vector pEGFPN1. Following transfection, the cells were maintained at 37°C for 24 h before patch clamp recording. The UniProt ID of TMEM16A is A0A0A0MQF2. LA795 cells were cultured using RPMI 1640 medium (Gibco) with 10% fetal bovine serum (FBS) (Gibco) and 1% penicillin–streptomycin solution at 37°C and 5% CO₂. LA795 cell is a mouse lung adenocarcinoma cell line with high endogenous expression of TMEM16A.

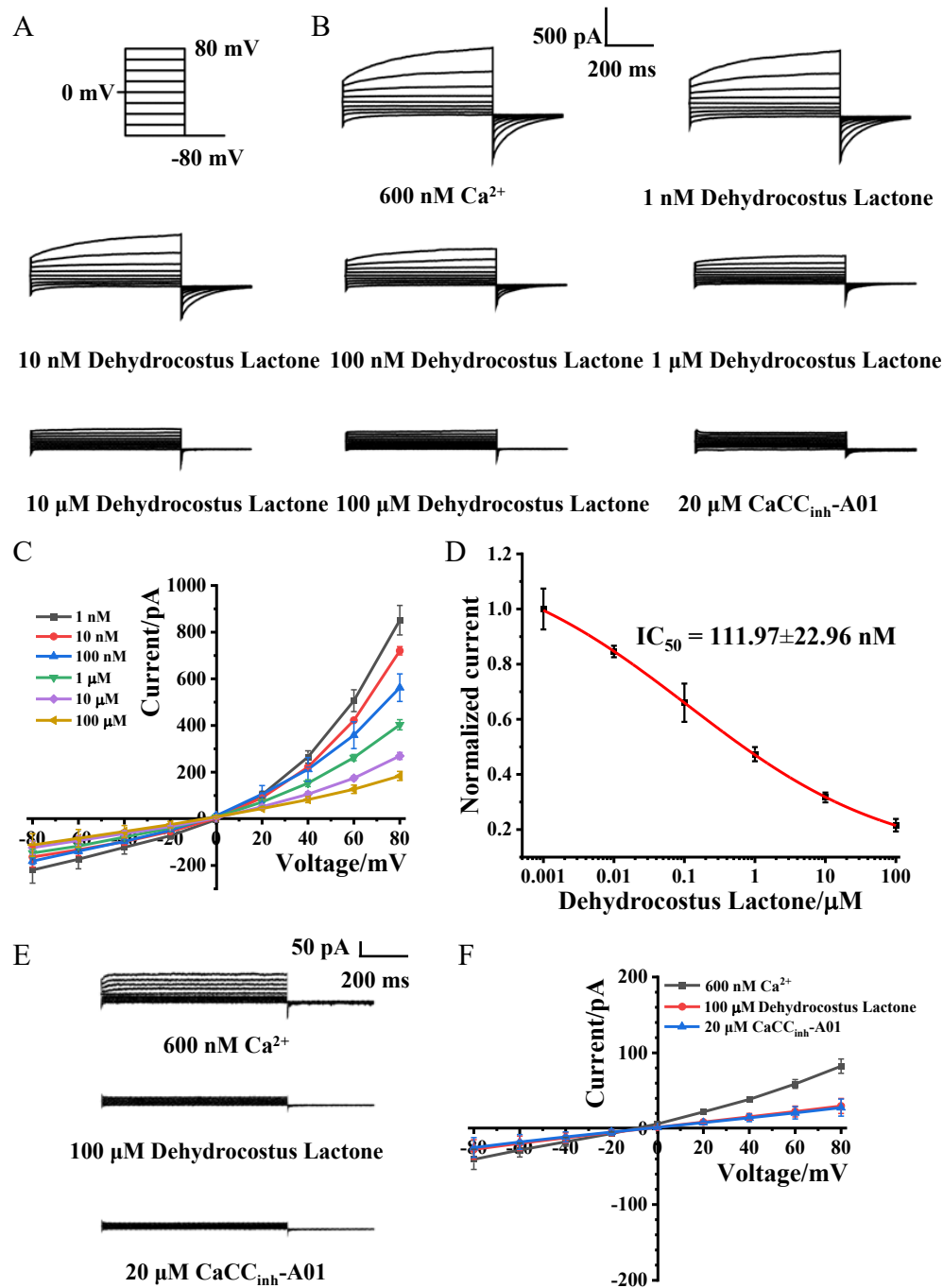


Figure R1 (Fig. 5) Dehydrocostus Lactone inhibits TMEM16A current in the TMEM16A recombinant HEK293T cells in a concentration-dependent manner. (A) Voltage stimulation waveform scheme. (B) Representative TMEM16A current in whole cell patch clamp in HEK293T recombination system is activated by 600 nM Ca^{2+} and inhibited by different concentrations of Dehydrocostus Lactone ($n=3$), with 20 μM $\text{CaCC}_{\text{inh}}\text{-A01}$ as the positive control. (C) I-V curve of TMEM16A inhibited by various concentrations of dehydrocostus lactone. (D) Dose-response curve of TMEM16A currents at +80 mV incubated with indicated

concentrations of dehydrocostus lactone. The IC_{50} of dehydrocostus lactone on TMEM16A channels was 111.97 ± 22.96 nM. (E) Representative TMEM16A current in outside-out patch-clamp in HEK293T recombination system, which is activated by 600 nM Ca^{2+} and inhibited by 100 μ M dehydrocostus lactone ($n=3$), with 20 μ M $CaCC_{inh}$ -A01 as the positive control. (F) I-V curve of TMEM16A inhibited by 100 μ M dehydrocostus lactone ($n=3$), with 20 μ M $CaCC_{inh}$ -A01.

4.6.2 Electrophysiology

Currents were recorded using an EPC10 amplifier and Pulse software with a Digi LIH1600 interface (HEKA, Lambrecht, Germany). The borosilicate glass patch pipettes with a pipette resistance of 3-5 M Ω (Sutter Instruments, Novato, USA) were drawn by a P-97 puller (Sutter Instruments, Novato, USA). The series resistance is less than 20 M Ω , with a compensation rate ranging from 60% to 80% in the whole cell patch-clamp. The liquid connection potentials were corrected before patch formation and the glass pipette has been fire-polished. The data was sampled at 10 kHz after being low pass filtered at 2.9 kHz. The stimulation procedures included 150 ms voltage steps from a holding potential of 0 mV and membrane voltage (mV) clamping in 20 mV increments from -80 to +80 mV, with a -80 mV hold at room temperature (22-25 °C). The bath solution used to record TMEM16A currents contained the following (in mM): NaCl 150 mM, MgCl₂·6H₂O 1 mM, HEPES 10mM, glucose 10 mM, and mannitol 10 mM adjusted to pH 7.4 with NaOH. The pipette solution contained the following (in mM): CsCl 130 mM, EGTA 10 mM, MgATP 1 mM, MgCl₂·6H₂O 1 mM, and HEPES 10 mM adjusted to pH 7.3 with CsOH. The 600 nM free Ca^{2+} bath solution was prepared by adding standard $CaCl_2$ solution (1 M, Sigma Aldrich, St. Louis, MO, USA) to a final concentration of 8.69 mM, and adjusted to pH 7.4 with CsOH. The osmotic pressure was determined using the OM815 osmometer. The osmotic pressure of the pipette and the bath solutions was in the range of 290-300 mOsm/L and 300-310 mOsm/L, respectively. Outside-out membrane patches are formed by gradually retreating the patch pipette from the cell after a whole cell recording has been established. This has the effect of drawing out a strand of membrane that eventually breaks and often reseals such that its extracellular surface faces into the recording chamber. The membrane attached to the pipette follows the

retreating pipette, and a giga-ohm seal should form again when the membrane breaks and an outside-out patch is excised. The single channel current of TMEM16A is recorded when the drug is administered through the perfusion system. The bath and pipette solutions of the outside-out patch-clamp were the same as those in the whole cell patch-clamp recording. Different concentrations of dehydrocostus lactone were prepared in the bath solution and administered through the perfusion system. Specifically, whole-cell patch clamp and outside-out patch-clamp were employed to study the inhibitory effect of dehydrocostus lactone on TMEM16A.

2. The authors should comment on the Hill coefficient of the current decrease caused by dehydrocostus lactone; the reported concentration dependence is unusually shallow, spanning 5 orders of magnitude. While multiple binding sites is common in symmetric or pseudosymmetric assemblies where each monomer or subunit interface contributes to the binding of one ligand molecule, this is mechanistically implausible in TMEM16A whose fundamental operating unit is a single subunit.

Authors' response:

Thank you for your insightful suggestion. We completely agree that the operating unit of TMEM16A is a single subunit. During our data processing, we founded that the Hill coefficient of the stoichiometric response curve made by LA795 cells and TMEM16A recombinant HEK293T cells was 0.4273 and 0.2993, respectively, both of which were less than 1, indicating that the binding of the drug with the target (TMEM16A) had a negative synergistic effect. TMEM16A is a domain with two subunits, each of which is believed to contain at least one binding site for drug molecules. Notably, there might exists a negative synergistic effect between the binding sites of the two subunits. It is noted that the concentration dependence of drug-induced current drops is very flat, spanning five orders of magnitude. This corresponds to the case where the Hill coefficient is less than 1. Due to the existence of negative synergistic effect, with the increase of drug concentration, the binding of drug and TMEM16A does not show an obvious acceleration stage, but a relatively gentle process.

3. The method section for electrophysiology needs to be revised, including the reporting of pipette resistance and series resistance, errors associated with units, and the consistency of solution composition. The last sentence 'In a word,dose-dependent manner' does not belong to the methods.

Authors' response:

Thank you very much for your precious advice. We have modified the method section for electrophysiology in the section 4.6 of the manuscript. The resistance of the pipettes was ranged from 3 to 5 MΩ. The series resistance is less than 20 MΩ, with a compensation rate ranging from 60% to 80% in the whole cell patch-clamp. We also carefully checked all the units. The bath and pipette solutions of the outside-out patch-clamp were the same as those in the whole cell patch-clamp recording. Different concentrations of dehydrocostus lactone were prepared in the bath solution and administered through the perfusion system. The last sentence 'In a word, dose-dependent manner' has been removed.

Point 9 –

Outside-out patch-clamp data are needed to rigorously show that dehydrocostus lactone can inhibit TMEM16A in excised membranes. Satisfactorily addressing this issue is an absolute requirement for validating that the effect of the compound originates from directly binding to the channel. This is an important aspect of the authors' prediction model. Outside-out patches are formed after the establishment of the whole-cell configuration. There are no excuses for claiming that resources or expertise are not available to perform these experiments if the authors' have already succeeded in recording whole-cell currents. Please provide these data in the manuscript or completely remove TMEM16A from the manuscript if there is no inhibition in outside-out patches; the field does not need another false activator/inhibitor.

Authors' response:

Thank you very much for your helpful suggestion. We agree with you that outside-out patch clamp can directly demonstrate that dehydrocostus lactone is a bona fide inhibitor of TMEM16A. We studied the outside-out patch clamp technology, and were lucky enough to grasp the technique quickly and obtained the outside-out patch experimental results. We confirmed that dehydrocostus lactone can directly inhibit TMEM16A in outside-out

patch-clamp recording in the TMEM16A transfected HEK293T cells (**Figure R1EF**). The description and details of the experiments have been provided in Sections 2.6 and 4.6 of the revised manuscript.

Point 12 –

Please directly address my concerns and provide the data on the prediction for the known inhibitors 1PBC, niclosamide, and benzbromarone. If the authors' model is generally applicable, it should be able to predict the effect of experimentally validated compounds. Please include these data in the manuscript and address any inconsistencies.

Authors' response:

Thanks for the suggestions. We have provided the prediction results between TMEM16A and five known inhibitors, including 1PBC¹, niclosamide², benzbromarone³, crofelemer⁴ and CaCC_{inh}-A01⁵. As shown in the **Table R1**, our framework achieved consistently high scores in predicting the interactions, binding affinities, and inhibitory potential for all five inhibitors. Specifically, the prediction results show that all five inhibitors were predicted to have a high likelihood of inhibiting TMEM16A (predicted inhibition probability>0.99), and the interaction probability of four inhibitors exceeded 0.95. These results demonstrate the robust predictive capability of DTIAM in screening TMEM16A inhibitors.

We have provided these data in the revised manuscript, also shown below:

To evaluate the performance of our framework on known inhibitors of TMEM16A, we also applied DTIAM to infer the interactions, affinities, and inhibition scores between TMEM16A and five known inhibitors, including 1PBC¹, niclosamide², benzbromarone³, crofelemer⁴ and CaCC_{inh}-A01⁵. The prediction results are shown in Supplemental Spreadsheet 2. We found that all five inhibitors were predicted to have a high likelihood of inhibiting TMEM16A, and the interaction probability of four inhibitors exceeded 0.95. Moreover, the predicted binding affinity scores for the five inhibitors were closely aligned with experimental measurements, demonstrating the robust predictive capability of DTIAM in screening TMEM16A inhibitors.

Table R1 The prediction results of the interactions, affinities, and inhibition scores between

TMEM16A and five known inhibitors.

Name	SMILES	Interaction pred_prob	Affinity scores (nM)	Inhibition pred_prob
1PBC	<chem>C1=CC=C2C(=C1)NC3=C(C(=CC(=O)N23)C(F)(F)F)C#N</chem>	0.9544	352.47	0.9997
Niclosamide	<chem>C1=CC(=C(C=C1[N+](=O)[O-])Cl)NC(=O)C2=C(C=CC(=C2)Cl)O</chem>	0.9544	989.44	0.9997
Benzbromarone	<chem>CCC1=C(C2=CC=CC=C2O1)C(=O)C3=CC(=C(C(=C3)Br)O)Br</chem>	0.9599	348.39	0.9999
Crofelemer	<chem>COC1=CC=CC=C1CNC(=O)C2=NOCC(=C2)C3=CC=CC=C3</chem>	0.8022	1053.52	0.9999
CaCC _{inh} -A01	<chem>CC(C)(C)C1CCC2=C(C1)SC(=C2C(=O)O)NC(=O)C3=CC=CO3</chem>	0.9631	478.75	0.9992

Reference

1. Lam, Andy KM, Sonja Rutz, and Raimund Dutzler. "Inhibition mechanism of the chloride channel TMEM16A by the pore blocker 1PBC." *Nature Communications* 13.1 (2022): 2798.
2. Miner, Kent, et al. "Drug repurposing: the anthelmintics niclosamide and nitazoxanide are potent TMEM16A antagonists that fully bronchodilate airways." *Frontiers in pharmacology* 10 (2019): 51.
3. Huang, Fen, et al. "Calcium-activated chloride channel TMEM16A modulates mucin secretion and airway smooth muscle contraction." *Proceedings of the National Academy of Sciences* 109.40 (2012): 16354-16359.
4. Tradtrantip, Lukmanee, Wan Namkung, and Alan S. Verkman. "Crofelemer, an antisecretory antidiarrheal proanthocyanidin oligomer extracted from *Croton lechleri*, targets two distinct intestinal chloride channels." *Molecular pharmacology* 77.1 (2010): 69-78.
5. Tradtrantip, Lukmanee, Wan Namkung, and Alan S. Verkman. "Crofelemer, an

antisecretory antidiarrheal proanthocyanidin oligomer extracted from *Croton lechleri*, targets two distinct intestinal chloride channels." *Molecular pharmacology* 77.1 (2010): 69-78.

Response Letter

We sincerely appreciate the time and effort spent by the reviewers in assessing our article NCOMMS-23-43911 entitled “*DTIAM: A unified framework for predicting drug-target interactions, binding affinities and activation/inhibition mechanisms*”. We have carefully addressed all the comments and suggestions provided by the reviewers, which have improved the quality of our manuscript. In the following pages, we provided our detailed point-by-point responses to the review comments.

Reviewer #3 (Remarks to the Author):

The authors have addressed the major concerns that I raised; below are some minor comments.

Line 780: This is the adjustment of the voltage offset when the pipette is inserted into the bath solution but is not the correction of liquid junction potential. If junction potentials were not corrected for, please state the magnitude and that junction potentials were not corrected for. Please revise.

Authors' response:

Thank you very much for your helpful suggestion. The liquid junction potentials were not corrected and its magnitude was about 10 mV. We have revised it in the manuscript.

Lines 790 and 792: Inconsistency between the stated pH values of the pipette solutions.

Authors' response:

Thank you for your careful review and the valuable comment. The inconsistency in the stated pH values of the pipette solutions is attributable to a clerical error on our part. We have corrected it throughout the manuscript.

Lines 800 and 801: The seal between the membrane and the glass pipette remains, giga-ohm or not.

Authors' response:

After the establishment of the whole cell patch-clamp configuration, a gigaseal persists.

Remarkably, this gigaseal is maintained not only throughout the whole-cell state but also during the subsequent formation of the outside-out patch clamp configuration.