Response to Reviewer's Comments concerning BIB Submission BIB-21-1135

Thank you for reviewing our manuscript, and of course for assigning it a favorable verdict of *Major Revisions*. We are grateful for the detailed comments and suggestions which have been helpful to further improve the paper. Below, we have regrouped the sets of comments and have provided a detailed point-by-point replies. We have highlighted the updated contents in the revised manuscript with color red and sincerely hope that this satisfies the request for changes necessary to proceed with the publication of the updated manuscript.

Initial Comments by Reviewer 1

Comment 1:

There are 482 herbs chose to generate the TCM dataset. It is unclear what are the characters of these herbs. Why chose these 482 herbs?

$\gg$ **Reply**:

Than you so much for this comment. We downloaded the raw set of Chinese medicines from Traditional Chinese Medicine Systems Pharmacology Database and Analysis Platform[1] by using selenium, a python-based crawler. And then processed to a dataset that contains 480 herbal medicines and 13,448 related chemical compounds. And after conducting the filtering process of ADME and compound-protein docking steps, the number of compounds decreased to 12,735. And this update content is located in the revised manuscript line 867 $\sim$ line 874, page 13 to page 14.

Comment 2:

Are there any other approaches can be used to identify the drug combinations? It lacks detailed comparisons among different approaches.

$\gg$**Reply**:

Than you so much for this comment. In this work, we aim to signaling the candidate herbs and candidate herb combinations with the most synergistic therapeutic effects. There are some related works to signal re-purposing drugs, but they have the following limitations. Partial studies are based on the previous knowledge

[1] `https://old.tcmsp-e.com/browse.php?qc=herbs`

that the potential candidates had a strong inhibitory effect on MERS and SARS-CoV; it was not guaranteed that these candidates could strongly fight against SARS-CoV-2. Besides, most of current issued methods (1) only focused on chemical medicines not utilizing synthetic effects from multicomponent-based herbal medicines treating against COVID-19; (2) analyzed on single drug targeting single structural SARS-CoV-2 protein (eg. Spike protein), which is either hardly binding the virus' all protein's sites or own less therapeutic power; (3) adopted the "one-drug-fits-all" strategy by means of the same drug to treat different infected patients in various infected stages. In such case, infected patients cannot obtain the precision treatment; and (4) no experimental validation was provided to the results.

Therefore, in order to signal candidate drug combinations with more potential therapeutic effects, we conducted a mix-way of dry-lab and wet-lab research based on the herbs data. But restricted by the burden of time and cost consuming, we so far only tested one of the recommended candidate herb combinations without test other methods.

The updated contents are located in line (1) 129 $\sim$ line 161 page 3, (2) line 197 $\sim$ line 202 page 4.

***Comment 3**:*
The language of this manuscript needs to be improved.
$\gg$**Reply**:
Thank you so much for this comment and we apologize for the writing. The whole manuscript has been checked by someone with full professional proficiency in English, and we have revised the writing.

***Comment 4**: The result shows that Higher ranked herbs in a specific cluster have the potential to exhibit therapeutic effects. Can the authors use experiments to prove the result?*
$\gg$**Reply**: Thank you so much for this comment. The ranked herbs are generated by computing the similarity with the herbs issued by the National Health Commission of the People's Republic of China guidelines. These official issued suggested herbs have been adopted and some of them have under clinical trial tests. Therefore, we employed this issued set of herbs as the "Silver standard" to measure our ranking results.

In addition, since we aim to signal candidate herb combinations, we then only tested the herb combinations in a P2+ biosafety lab.

We do appreciate this comment, and we are planning to conduct estimating results of ranked herbs in our future work.

Initial Comments by Reviewer 2

In this manuscript, Yang and colleagues present an approach combining protein docking, heterogeneous graph construction, and node embedding to find repurposable traditional Chinese medicine herbs to target SARS-CoV-2 proteins. The problem is, obviously, highly critical, and the exact outcome the authors choose to

pursue – identify not just individual drugs but drug combinations – is valuable. The approach and subsequent analysis taken by the authors are both very interesting. However, there are a few major concerns that need to be addressed.

Comment 1*: The biggest concern is how are the approach and the predictions evaluated?*

Comment 1.1*: As it is impossible to evaluate performance for SARS-CoV-2 given the limited knowledge of potential drugs, the approach could be applied to a disease with well-known TCM drugs to appropriately assess performance.*

≫**Reply**: Thank you so much for this comment and the suggestion is very valuable. We concerned that if we applied the proposed approaches to some diseases (eg. MERS and SARS-CoV) with already known herbs, which was not guaranteed that these candidates could strongly fight against SARS-CoV-2.

To estimate our recommended candidate herb combination, we conducted the enzyme active test to estimate one of our recommended herb combination in a P2+ biosafety lab.

We appreciate this data-driven estimation strategy which can save huge time and cost consuming. Due to the unaffordable burden conducting the wet-lab experiment, there was only one of recommended herb combination to be estimated by the wet-lab experiment test in our study. Therefore, we will adopt the suggestion to modify the experimental results estimation in our future work.

Comment 1.2*: The authors mention in the last point in the 'Key Points' section at the end that their recommended drugs all "belong to the treatment of" COVID-19 drugs proposed by the government, because of which they state that their model achieves high performance.*

≫**Reply**: Thank you so much for this comment. We aware that there need more clinical trial evidences to test the results. There are some herbs have been conducted clinical trials though, they still not completed. In such case, we adopted the herbs issued by the National Health Commission of the People's Republic of China. These official issued herbs have already been employed to most of people in China, which recognized as the experiential "Silver standard".

We appreciate this valuable suggestion for making more reasonable experimental design, and we will keep following up the clinical trial results for related herbs.

Comment 2*: At the outset, the authors wish to find drug combinations, where each 'combination' is a small number of drugs that have complementary actions that together achieve more than any of the individual drugs. However, at the end of the analysis, they identify ranked lists of drugs for each SARS-CoV-2 protein (or protein group: structural, non-structural, and auxiliary) but there is no indication that these drug lists are any more than a collection of potential independent drugs. In other words, there is no part of the underlying approach or the subsequent analysis of the prioritized drugs that identifies a subset of 'complementary' drugs that could work as a combination.*

≫**Reply**: Thank you so much for this comment. We apologized for the missing part of recommending herb

combinations. We have added the experiments of signaling precision candidate herb combinations for the specific SARS-CoV-2 infected stage.

The updated content located in: (1) line 156 $\sim$ line 161 page 3, (2) line 177 $\sim$ line 196 page 4, (3) line 796 $\sim$ line 863 page 13, (4) line 902 $\sim$ line 909 page 14. And the updated codes and related supplements can be accessed online `https://github.com/fanyang-AI/TCM-COVID19`.

Comment 3*: The authors point out two other studies that take different approaches to identify TCM herbs for SARS-CoV-2 (references 18 and 19). How do the predictions made in the current study compare to the drugs identified in those two studies?*

$\gg$**Reply**: Thank you so much for this comment. These two work have some contributions to treat against COVID-19. But they are limited in: (1) employing the "one-drug-fits-all" without considering patients in different infected stage who need personalized treatment; (2) lacking of analysis on herb combinations that make less power to binding all sites of SARS-CoV-2 proteins; (3) not to conduct a wet-lab test to estimate their results.

In comparison, we (1) recommended precision herb combinations considering specific infected stage which contribute to conducting personalized treatment; (2) we not only signal candidate herbs but also compute the combination of herbs that take advantage of synthetic effects of herb combination; (3) we conducted the enzyme activity test for one of our recommended herb combinations targeting two nonstructural proteins 3CLpro (nsp 5) and PLpro (nsp 3) in a P2+ biosafety lab. And the results are described in the Supplementary file S6.

Comment 4*: Though the authors have made a number of Python scripts available at* `https://github.com/fanyang-AI/TCM-COVID19`*, there is no documentation on how to use the code and there are no input datasets that the code can be run on to reproduce the approach/analysis described in this manuscript. These two need to be made available.*

$\gg$**Reply**: Thank you so much for this comment. We have uploaded the modified codes and data. And we also revised the ReadME file to present a detailed experimental steps.

MINOR CORRECTIONS*: The following three typos are noted: - Page 6, Column 1, Line 21/22: "The method –has– is..."; remove 'has' - Page 6, Column 1, Line 48/49: "... a compound can –be– effectively..."; remove 'be' - Page 6, Column 2, Line 23/24: "... and the jth –chemical compound–"; change 'chemical compound' to 'protein'?*

$\gg$ **Reply**: Thank you so much for this comment. We do appreciate these suggestions, and we have revised these typos and mistakes.

Initial Comments by Reviewer 3

***Comment 1**: The manuscript lacks the literatures of network embedding, drug Combinations, and drug repositioning for COVID-19. Some important references are missing.*

≫**Reply**: Thank you so much for this comment. We have followed the comments to add the contents of network embedding, drug Combinations, and drug repositioning for COVID-19 in the Section of Related Works.

The updated content located in: (1) line 221 $\sim$ line 225 page 4, (2) line 281 $\sim$ line 349 page 5 $\sim$ page 6, (3) line 369 $\sim$ line 418 page 6.

***Comment 2**: In section "Related work", the author write "The three most commonly used categories...". However, I only find "Matrix decomposition" and "Graph embedding". In addtion, the author maybe fail to understand the network embedding. To the best of my knowledge, many studies(e.g., `https://www.sciencedirect.com/science/article/pii/S0950705118301540`, and `https://academic.oup.com/bioinformatics/article/36/4/1241/5581350`) have suggested that network embedding can be are categorized into three groups: matrix factorization based, random walk-based, and neural network-based.*

≫**Reply**: Thank you so much for this comment and we appreciate these two suggested references. We have enriched the Related Works by adding contents of Matrix decomposition and Graph embedding following the suggested references.

The updated content located in Section 2.1, Section 2.2, and Section 2.3.

***Comment 3**: The author should validate the repurposable therapeutic drug combinations targeting SARS-CoV-2 Proteins via data from PubMed publications, ongoing clinical trials.*

≫**Reply**: Thank you so much for this comment and this suggestion is valuable. We conducted research on herb combinations treating against SARS-CoV-2 proteins based on the characteristic of herbs' multicomponent targeting virus' multitargets. Since most of current works focused on mining single chemical drug binding single or partial SARS-CoV-2 protein sites, which caused there exist a few clinical trial results about herb combinations from PubMed[2] and ClinicalTrials[3]. In such case, we can only to design and conduct new wet-lab test.

We appreciate this valuable suggestion. And we will continue to focus on the related clinical trial experiments of herbs.

***Comment 4**: The parameters of the embedding algorithms. What are the parameters used in the proposed system and how their values are set? Also, how the parameter values can affect the proposed system?*

≫**Reply**: Thank you so much for this comment. We apologized for the missing part and we have added

[2] `https://pubmed.ncbi.nlm.nih.gov/`
[3] `https://clinicaltrials.gov/`

the content in the ReadMe file that located in the upload code package. And the code can be available from `https://github.com/fanyang-AI/TCM-COVID19`.

We also revised the content w.r.t. the training process of the model, which located from line 782 $\sim$ line 795 page 11 $\sim$ page 12.

***Comment 5**: The text in Figure 4 is blurred, which affects the aesthetics of the manuscript. The format of Eq. (8) is terrible. The author should adjust it.*

$\gg$**Reply**: Thank you so much for this comment. We do apologize the blurred figure and unsuitable equation. We have updated the figure that shown as Fig. 6 in page 12. And we also re-organized the equation that termed as Eq. 18 located in page 10.

***Comment 6**: The source code is public, but the authors only provided their source code without any description of how to use the program. The current version of the GitHub repository doesn't help to reproduce the entire experiments in the manuscript.*

$\gg$**Reply**: Thank you so much for this comment. We have uploaded the modified codes and data. And we also revised the ReadME file to present a detail experimental steps. The code and data can be accessed at Github `https://github.com/fanyang-AI/TCM-COVID19`.

***Comment 7**: The experiments is poor. The author should add more experiments to analyse the results and model perforances.*

$\gg$**Reply**: Thank you so much for this comment, we apologize the weak experiment design.

To enrich the experiment, we have added experiment of recommending precision herb combinations targeting specific SARS-CoV-2 infected stage. To be more specific, we employed the variational graph autoencoder model to compute the similarity between the 20 candidate herb combinations and six official issued herb combinations. And we recommend the 20 candidate herb combinations with certain ranking order for various infected stage.

In order to estimate the recommended results, we also conducted the enzyme activity test for one recommended herb combinations targeting two SARS-CoV-2 nonstructural proteins, 3CLpro and PLpro, in a P2+ biosafety lab.

The details are shown in the supplementary file S3, S4, S5, S6, S7, and S8.