nature portfolio

Peer Review File

AutoFocus: A hierarchical framework to explore multi-omic disease associations spanning multiple scales of biomolecular interaction

Reviewers' comments:

Reviewer #1 (Remarks to the Author):

General Comments:

The analysis of molecular perturbations associated with disease phenotypes at the multi-omic level is meaningful to a deeper understanding of pathophysiological mechanisms. In this article, the authors propose AutoFocus, a data-driven method that hierarchically organizes biomolecules and tests for phenotype enrichment. The evaluation results on two datasets show that AutoFocus can organize molecules as disease-associated modules at any scale. Overall, this work innovatively employs hierarchical clustering to organize molecules at different scales, which is of positive significance for the study of molecular perturbations related to diseases.

However, I have several concerns and suggestions, which I believe, if addressed, could enhance the quality of the research and its impact.

Major:

1.Comparison with other methods
In this article, the author solely presents the application results of AutoFocus on two datasets. As a methodological work, I suggest that the authors add the comparison with existing relevant studies (references 5, 9-13) or other clustering algorithms, and provide specific numerical comparisons to highlight the superiority of AutoFocus. Such as the number of identified modules, the proportion of inter-dataset modules, etc.

2.Enrichment threshold
A “majority vote” enrichment threshold of 0.5 were used to identify modules (enrichment peaks) associated with phenotypes. I wonder if 0.5 is an arbitrary setting without comparison. I believe that the selection of this threshold is crucial and can significantly impact the results. I suggest that the authors optimize the threshold by comparing the results obtained under different thresholds. Furthermore, it might be better to set different thresholds for different diseases or datasets. If the authors could provide a method for automatically selecting the threshold or offer some guidelines about threshold selection, it would enhance the applicability of this method.

3.Piggy-backer threshold
The threshold used for identifying piggy-backers can also be optimized by following the recommendations for enrichment threshold, rather than arbitrarily using a fixed value of 0.5.

Minor:

1.Making the code publicly available is helpful for other researchers to conduct further studies based on AutoFocus. However, the README in the author's GitHub repository is empty, which poses some difficulties for others in using AutoFocus. To enhance the applicability of this work, I recommend that the author supplement the README with necessary usage guidelines to make it more user-friendly.

2.The figures in this manuscript contain minor errors in terms of formatting and specific details. For instance: (1) The label on the y-axis of Figure 2b is slightly offset downward; (2) In Supplementary Figure 4b, the views marked with red boxes appear to be inconsistent. Specifically, the box before zooming should be shifted to the left to match the content displayed in the zoomed-in view.

There may still be other similar errors in this manuscript, so I suggest that the authors carefully check to avoid these minor errors that could have been prevented.

Reviewer #2 (Remarks to the Author):

This is an important tool to explore multi-omics disease associates.
There are a number of questions remaining:
1.There is no benchmarking of AutoFocus tool with other existing tool
2. Clear advantage of AutoFocus from WGCNA and other co-expression network tools misisng.
3. The clear advantage and unique feature of the tool is not clearly presented.
4. Performance/accuracy were not compared and assessed statistically.
5. Without statistical cutoff, it is not clear how false positives are controlled. However, they also implemented statistical cutoff to create networks.
6. Missing data- which is very common in multi-omics was not discussed and addressed
7. Platform, and data type and source differences were not discussed and addressed
8. Adjustment should go beyond age, sex and BMI for Type 2 diabetes.
9. Validation of results using independent dataset should be presented.
10. Multi-omics data are generated from different labs and sources. It is not clear how AutoFocus handle data coming from different sources.
11. The running time for two phenotypes in relatively short time is 4.5 hrs. This is extremely slow process and require additional development to accommodate more phenotype and several multi-Omics data.
12. Web-based tool for those not informatics savvy should be an option.

Response to Reviewers:

Reviewer comments

Responses

Note to both reviewers:

While performing the analyses in response to the comments below, a minor bug was found in the original AutoFocus code. After fixing the bug, the resulting clusters changed slightly for the ROS/MAP analysis, which explains the changes in the numbers of modules reported in Results section 2.4. The bug fix also changed the "Multi-phenotype associated module" in Figure 4c, and the exploration of that cluster's members. The bug fix did not affect the QMDiab modules, nor the overall message of any of the analyses in our paper. All code changes have been updated in Github.

Reviewer #1 (Remarks to the Author):

General Comments:

The analysis of molecular perturbations associated with disease phenotypes at the multi-omic level is meaningful to a deeper understanding of pathophysiological mechanisms. In this article, the authors propose AutoFocus, a data-driven method that hierarchically organizes biomolecules and tests for phenotype enrichment. The evaluation results on two datasets show that AutoFocus can organize molecules as disease-associated modules at any scale. Overall, this work innovatively employs hierarchical clustering to organize molecules at different scales, which is of positive significance for the study of molecular perturbations related to diseases.

However, I have several concerns and suggestions, which I believe, if addressed, could enhance the quality of the research and its impact.

Major:

1.Comparison with other methods
In this article, the author solely presents the application results of AutoFocus on two datasets. As a methodological work, I suggest that the authors add the comparison with existing relevant studies (references 5, 9-13) or other clustering algorithms, and provide specific numerical comparisons to highlight the superiority of AutoFocus. Such as the number of identified modules, the proportion of inter-dataset modules, etc.

We appreciate this comment and agree with the reviewer that a comparison of AutoFocus with existing relevant studies would serve to highlight the superiority of AutoFocus. As such, we have added to the Results section of the manuscript a section comprehensively comparing AutoFocus' module detection and testing to those of three methods, WGCNA, MEGENA, and MoDentify, which share key features in their method design against which it would be appropriate to compare with AutoFocus. This reasoning has similarly been added to the manuscript.

2.Enrichment threshold
A "majority vote" enrichment threshold of 0.5 were used to identify modules (enrichment peaks) associated with phenotypes. I wonder if 0.5 is an arbitrary setting without comparison. I believe that the selection of this threshold is crucial and can significantly impact the results. I suggest that the authors optimize the threshold by comparing the results obtained under different thresholds. Furthermore, it might be better to set different thresholds for different diseases or datasets. If the authors could provide a method for automatically selecting the threshold or offer some guidelines about threshold selection, it would enhance the applicability of this method.

A supplemental section has been added to discuss the effect of the enrichment threshold on various features of the clusters identified by AutoFocus, including the range of cluster heights and sizes, the number of clusters, the proportion of non-significant nodes found in clusters, and the number of significant nodes not in clusters. From this analysis, we can see that while there's substantial variability in the results based on threshold when the threshold is below 25% enrichment, both cluster height range and cluster size range level out quickly above 25%. This indicates that results are decently stable and not going to completely change with a small threshold change. In addition, the clusters themselves do not change positions in the hierarchical tree - the same underlying relationships will always be maintained regardless of threshold.

We observe two metrics that do change drastically with threshold: the proportion of non-significant nodes to total nodes returned in clusters decreases as the threshold is made more stringent, and the number of singleton significant hits (significant hits not found in clusters) increases as the threshold is made more stringent. If a user would like a guideline for threshold selection, they can use the global maximum of a score that combines these two metrics, indicating the threshold that incorporates the most significant hits in the returned clusters while reducing the proportion of non-significant hits in those clusters. This functionality has been added to the code in the GitHub, such that if a user calls the "threshold_analysis" function, they can see all these metrics and the score maximum. Moreover, this threshold selection guideline is now mentioned in the "Enrichment 'Peak' Calculation" in Methods section 4.4.

Finally, the threshold is a way of focusing on a specific region in the tree. If the user would like a zoomed-out view to analyze a greater process affected by a phenotype of interest, they can do that by choosing a lower threshold, and if they want to see smaller processes with a more concentrated signal, they can similarly achieve this with a higher threshold of enrichment. This is why the threshold was provided as a user input rather than an optimizable parameter. That being said, and as discussed above, we have added functions to the package that will allow the user to analyze key cluster metrics across a range of thresholds including cluster height range, cluster size range, number of clusters, number of singletons, and proportion of non-significant nodes so that they can make a more informed choice.

3.Piggy-backer threshold
The threshold used for identifying piggy-backers can also be optimized by following the recommendations for enrichment threshold, rather than arbitrarily using a fixed value of 0.5.

The threshold for identifying piggy-backers is the same as the enrichment threshold and is therefore not independent, a point that has been clarified in the supplementary text. We hope that our comment on the enrichment threshold above addresses this comment as well; users can examine the plots added above to examine how the threshold affects results.

Minor:

1.Making the code publicly available is helpful for other researchers to conduct further studies based on AutoFocus. However, the README in the author's GitHub repository is empty, which poses some difficulties for others in using AutoFocus. To enhance the applicability of this work, I recommend that the author supplement the README with necessary usage guidelines to make it more user-friendly.

Agreed, this has been addressed and the README has been updated with necessary usage guidelines to make the code more user-friendly. Contact information has been included as well for further user questions.

2.The figures in this manuscript contain minor errors in terms of formatting and specific details. For instance: (1) The label on the y-axis of Figure 2b is slightly offset downward; (2) In Supplementary Figure 4b, the views marked with red boxes appear to be inconsistent. Specifically, the box before zooming should be shifted to the left to match the content displayed in the zoomed-in view.

These errors have been addressed in the figures.

There may still be other similar errors in this manuscript, so I suggest that the authors carefully check to avoid these minor errors that could have been prevented.

We have carefully combed through the manuscript for other minor errors in the figures and have addressed all found.

Reviewer #2 (Remarks to the Author):

This is an important tool to explore multi-omics disease associates.
There are a number of questions remaining:
1. There is no benchmarking of AutoFocus tool with other existing tool
2. Clear advantage of AutoFocus from WGCNA and other co-expression network tools misisng.
3. The clear advantage and unique feature of the tool is not clearly presented.

Please see the response to the first major point of Reviewer 1. We have added a new analysis to the Results section comparing the performance of AutoFocus to three other methods in response to this comment.

4. Performance/accuracy were not compared and assessed statistically.

In the realm of biomolecule clustering analysis, especially multi-omic clustering analysis, there does not exist a gold standard of clusters to be achieved by a method. For this reason, accuracy assessment is not possible in our opinion. Our tool is exploring existing associations between molecules and disease, associations that have been previously identified for the tested datasets. We have added clarification on this point to the Discussion.

5. Without statistical cutoff, it is not clear how false positives are controlled. However, they also implemented statistical cutoff to create networks.

False positives are controlled for in the original association step where individual analytes, or hierarchical tree “leaves”, are determined to be associated with a phenotype of interest using a rigorous statistical cutoff with multiple testing correction. The tool then allows us to explore those statistically significant hits in a structured way. To the second point, the mgm networks are purely for annotation purposes and are not used in cluster identification. In this manuscript, we assert that statistical cutoffs in cluster **identification** are the problem we are trying to address with AutoFocus, not in cluster **annotation.**

6. Missing data- which is very common in multi-omics was not discussed and addressed

Missing data can pose a problem in omics data analysis as sometimes missing data represents molecular concentrations that are too low to detect, and omitting missing data while only considering present data could introduce bias in results. To avoid this bias in our application of AutoFocus to the QMDiab and ROS/MAP datasets, we fully removed samples that are missing more than 20% of measured molecules, and molecules missing more than 10% of samples (25% in ROS/MAP). For the remaining missing values, we used k-nearest neighbors to impute the missing values in each of our multi-omics datasets, the reasoning for which can be found in Do, K. T. et al. *Metabolomics,* 2018. These steps are outlined in the Methods section of the manuscript.

While the various platforms had different rates of sample dropout due to missing measurements, the AutoFocus method computes the hierarchical structure using pairwise correlations, meaning the effect of missing samples will only impact the correlation power of the molecules for which they are missing; a missing sample on one platform will not affect the correlation power of molecules measured on different platforms. This is a positive improvement over existing clustering methods, such as those using partial correlations, as they require a complete matrix with no missing values for network calculation, and thus samples missing from one platform would have to be removed entirely.

Ultimately, preprocessing omics data is the user’s responsibility and addressing missing data in multi-omics datasets is not a goal of the AutoFocus method; AutoFocus expects missing data to be dealt with prior to cluster calculation. The steps we have outlined in the Methods section are variations we felt appropriate for our data but are not the only way to address missing values. We have added clarification to the text of the manuscript making it clearer that preprocessing is on the user’s side.

7. Platform, and data type and source differences were not discussed and addressed

Our current Methods section 4.2 includes all information about the technologies and sources of the platforms used in the QMDiab and ROSMAP datasets. In addition, we have now added more information about the different data types of the platforms used in both the QMDiab and ROSMAP analyses, specifically discussing the distributions of the data as is most relevant to our method and downstream analysis.

However, as in our response to comment 6, it is important to note that AutoFocus is not a preprocessing tool that *addresses* the differences between data types coming from different sources. Our tool takes as input preprocessed data whose analytes are ready to be correlated with one another, and our method provides flexible options for the correlation method used to perform this clustering analysis (parametric and non-parametric). We outline in Figure 2 and discuss in section 2.2 the effect of combining different platforms on correlation significance.

8. Adjustment should go beyond age, sex and BMI for Type 2 diabetes.

Unfortunately, these are the applicable covariates present in our dataset, and have adjusted for them accordingly. Further applicable adjustment data such as medications, is not available in the QMDiab dataset. A user of our tool can choose the appropriate confounders for their own datasets.

9. Validation of results using independent dataset should be presented.

While we agree that validation of our results using an independent dataset would be a wonderful addition to this analysis, the QMDiab dataset for Type 2 diabetes and the ROS/MAP dataset for Alzheimer's brain tissue are one of a kind. Therefore, we do not have independent datasets with which to perform validation.

10. Multi-omics data are generated from different labs and sources. It is not clear how AutoFocus handle data coming from different sources.

AutoFocus is a method that takes as input pre-processed datasets from different omic sources, each molecule of which is then associated with an input phenotype and pairwise correlated with each other molecule. AutoFocus is not a method to process data from different sources. We see the blocklike correlation biases that are a natural feature of this multi-omics data from different sources, which is extensively discussed in the paper. Please refer to our response in point 7 for further discussion.

11. The running time for two phenotypes in relatively short time is 4.5 hrs. This is extremely slow process and require additional development to accommodate more phenotype and several multi-Omics data.

While we agree that 4.5 hours is quite a long time, there are a few things to note. First, the ROS/MAP dataset has over 8,000 analytes and we are performing the same calculation twice on two different phenotypes. Therefore, the 4.5 hours is an extreme by our standards. Second, as mentioned in Methods section 4.5, the majority of this time is spent calculating MGM networks for each cluster.

We have added functionality to our method that can disable the MGM calculation, thus reducing the algorithm runtime on the same dataset to just over 6 minutes. This time analysis has been added to the manuscript, and the function update to toggle off MGM calculation has been added to our GitHub codebase, along with a message to the user indicating the option to turn off MGM calculation for improved runtime.

12. Web-based tool for those not informatics savvy should be an option.

A web-based, hosted version of AutoFocus could be possible with a cloud-hosted solution. However, due to the high memory and CPU demands of AutoFocus, this would be substantially too expensive to implement. We have provided all code used as an R package so labs can use their own resources instead, which may exceed the computing power of our own. In addition, we have provided contact information on the GitHub page for this package for those not informatics savvy to use in the case of an issue or bug.

Reviewers' comments:

Reviewer #1 (Remarks to the Author):

I thank the authors for carefully addressing my comments and modifying the manuscript correspondingly. Therefore, I would recommend it for publication.

Reviewer #3 (Remarks to the Author):

1. The authors acknowledged the correlation bias within the datasets, with intra-dataset correlations being higher than inter-dataset correlations. Could the authors provide further analysis or methods to assess and correct for the impact of this bias on the study's results?

2. The authors mentioned the difficulty of validation with independent datasets. Can the authors provide strategies for enhancing the reproducibility of the results through cross-validation or other statistical methods?

3. The authors mentioned improvements to reduce computational time. Can the authors further discuss the computational complexity of the algorithm and its scalability and resource consumption on datasets of different sizes?

4. In terms of multi-omics data integration, can the authors provide more details on the data processing workflow and the strategies for choosing and optimizing the integration of different data types?

5. For the modules identified by AutoFocus, can the authors provide more biological interpretations, especially how these modules relate to known biological processes or disease mechanisms?

6.Can the authors provide more information on the rationale behind the choice of statistical methods and thresholds, and how these choices affect the interpretation of the final results?

7. In the comparison with other methods, the author only mentioned the comparison with WGCNA, MEGENA, and MoDentify. Are there any other relevant methods that can be compared? For example, comparisons with deep learning-based methods.

Response to Reviewers:

Reviewer comments
Responses

Reviewer #3 (Remarks to the Author):

1. The authors acknowledged the correlation bias within the datasets, with intra-dataset correlations being higher than inter-dataset correlations. Could the authors provide further analysis or methods to assess and correct for the impact of this bias on the study's results?

We thank the reviewer for their comment regarding the correlation bias within the datasets. We acknowledge that intra-dataset correlations are higher than inter-dataset correlations, a known omics effect that is also present in our data. We believe that Figure 2 adequately assesses and summarizes the impact of this bias within the QMDiab dataset on which we tested our method.

We direct the reviewer to the hierarchical trees in Figures 3 and 4, where we believe the visualization of platform distribution across the hierarchical trees clearly illustrates the impact of the bias. The text accompanying these figures outlines the number of clusters that are multi-omic and multi-platform, and the supplemental tables provide detailed distribution of platforms across all clusters.

This bias is fundamentally complex and remains an unsolved issue in the field of multi-omics research. Correction methods, such as varying cutoff thresholds between inter- and intra-dataset correlations, will always introduce some form of significant bias. As such, we chose not to correct for the bias, and designed AutoFocus with this bias in mind, utilizing the data structure as it is.

We have revised the language in the discussion section to emphasize that this bias is a characteristic of the multi-omics data field rather than a limitation of our specific method.

2. The authors mentioned the difficulty of validation with independent datasets. Can the authors provide strategies for enhancing the reproducibility of the results through cross-validation or other statistical methods?

We thank the reviewers for this comment, as reproducibility and result stability are important aspects in assessing the validity of our findings. While we are unable to validate using an independent dataset, we can use other statistical methods, namely bootstrapping, to test the robustness of our results from different sample populations.

The modules returned by the AutoFocus algorithm mainly rely on the underlying hierarchical structure. To test the stability of this structure, we performed 100 bootstraps on the original QMDiab data, sampling with replacement from the original 388 samples. We then performed hierarchical clustering on each bootstrap and compared the resulting structure to the hierarchical structure from the original data via cophenetic correlation using the dendextend R package (details can be found here: https://www.rdocumentation.org/packages/dendextend/versions/1.17.1/topics/cor_cophenetic). From this analysis, we found that all bootstrapped hierarchies had a mean cophenetic correlation of 0.924 with the original hierarchy (standard deviation of 0.006), indicating that the hierarchical structure is very stable even on different subsets of data.

We see a different result in the ROS/MAP data a mean cophenetic correlation of 0.62 (standard deviation of 0.02), indicating that the hierarchy is less stable for the dataset.

The bootstrapping methodology has been added to the Methods section and we have added a section to the paper in both the QMDiab and ROS/MAP results sections called "Stability Analysis" to discuss these results.

3. The authors mentioned improvements to reduce computational time. Can the authors further discuss the computational complexity of the algorithm and its scalability and resource consumption on datasets of different sizes?

The following is a breakdown of the **computational** complexity of AutoFocus on a dataset of N samples and p features:

1. Pairwise correlation of p features over N samples has a computational complexity of $O(Np^2)$
2. Hierarchical clustering of p features using the hclust algorithm and average linkage has a complexity of $O(p^2)$. Details of the hclust backend can be found here: https://classification-society.org/csna/mda-sw/
3. Calculating the association of p features with a phenotype over N samples has a computational complexity of $O(Np)$.
4. Finally, top-down traversal of the hierarchy to identify enriched clusters has a computational complexity $O(p)$.

The dominant term of all these calculations is that of the pairwise correlation operation, meaning the computational complexity of this algorithm is $O(Np^2)$. However, in a high-throughput multi-omic setting, we see $p \gg N$ as the amount of data generated surpasses the number of samples collected, leading to a time complexity of $O(p^2)$.

The following is a breakdown of the **space** complexity of AutoFocus on a dataset of N samples and p features:

The input to the AutoFocus algorithm is the original matrix which has p features and N samples. Therefore, the input data storage requires $O(Np)$ space

Intermediate calculations:

1. For pairwise correlation, the output is the pairwise correlation matrix of $p \times p$ features. Therefore, this operation requires $O(p^2)$ space
2. For hierarchical clustering, the output is a dendrogram object requiring $O(p)$ space. However, the algorithm stores each cluster created by an internal node. The space complexity of this operation is $O(p \log p)$.
3. For association of p features with a phenotype over N samples, each association is calculated sequentially (one feature at a time), indicating that the storage requirement is $O(N)$.

The output to the AutoFocus algorithm is the hclust object ($O(p)$), a list of all clusters within the hierarchy ($O(p^2)$ in worst case scenario, $O(p \log p)$ on average), along with annotations on the samples ($O(N)$) and the features ($O(p)$).

The dominant terms of all these calculations are the input matrix and the intermediate pairwise correlation matrix, meaning the space complexity of this algorithm is $O(Np + p^2)$. As mentioned above, as p becomes much greater than N in practice, the space complexity of this algorithm becomes $O(p^2)$.

The details concerning computational and space complexity of AutoFocus have been added to Methods section 4.5 (Runtime performance and complexity).

4. In terms of multi-omics data integration, can the authors provide more details on the data processing workflow and the strategies for choosing and optimizing the integration of different data types?

We believe the paper provides sufficient detail on the data processing workflow. Data types were not “chosen” as all available data was used. Integration of data types is done through simple concatenation, correlation, and hierarchical clustering, all of which is outlined already in the manuscript. As was mentioned in the response to reviewer 2, question 10, AutoFocus is a method that takes as input pre-processed datasets from different omic sources, each molecule of which is then associated with an input phenotype and pairwise correlated with each other molecule. AutoFocus is not a method to process data from different sources.

5. For the modules identified by AutoFocus, can the authors provide more biological interpretations, especially how these modules relate to known biological processes or disease mechanisms?

We believe the manuscript already provides enough biological information regarding the modules discussed and their relationships to the diseases studied. Specifically, the *Type 2 Diabetes Modules* paragraphs in Results section 2.3 outline the various molecules in the largest module identified by AutoFocus, their prior implication in the disease, and functional interpretation of their role in the disease. The same biological analysis was done for three modules of note in Results section 2.4 as it pertained to the modules associated with tau neurofibrillary tangles and cognitive decline.

Given the number of modules identified by the method, providing the biological interpretations for each would be out of scope of this paper.

6. Can the authors provide more information on the rationale behind the choice of statistical methods and thresholds, and how these choices affect the interpretation of the final results?

The only statistical test that was used in this method was the association of features with phenotypes using a linear model. This was used for its simplicity and its ability to integrate covariates in the calculation to account for confounding effects. This is a standard statistical evaluation, as is the p<0.05 threshold used to determine significance. For p-value adjustment, the Bonferroni method was used on the QMDiab dataset while FDR was used on the ROS/MAP dataset, which is outlined in the manuscript. The Bonferroni p-adjustment method led to sparser significant hits and thus fewer modules.

Pearson correlation was the statistical method used in the manuscript to find the distance between molecules to be passed into the `hclust` algorithm, but the AutoFocus function allows for other correlation methods to be used, such as Spearman correlation. Even though these are two different correlation methods, they lead to similar hierarchical trees. The hierarchical tree made on the QMDiab dataset using Pearson correlation had a cophenetic correlation of 0.93 with the hierarchical tree using Spearman correlation, indicating that the final results would be incredibly similar between the two methods. However, this dropped to 0.7 for the ROS/MAP dataset, indicating the choice of distance metric could have a greater impact on those results. Ultimately, the extent of the impact of the choice of statistical methods is dataset dependent.

We have added to the manuscript that the user has the option to choose between the two p-value adjustment methods and the two correlation methods when running the AutoFocus algorithm (section 4.4).

7. In the comparison with other methods, the author only mentioned the comparison with WGCNA, MEGENA, and MoDentify. Are there any other relevant methods that can be compared? For example, comparisons with deep learning-based methods.

The choice of the three comparison methods was for very specific criteria: clustering algorithms that clustered biomolecular features (instead of samples) and had infrastructure for multi-scale examination of identified clusters. WGCNA was developed for gene expression and has a hierarchical structure from which it identifies clusters; MEGENA was also developed for gene expression and is a multiscale algorithm; MoDentify was designed for metabolomics and claims to identify modules at different resolutions. Therefore, these three methods met the criteria set forth. We have updated the text of Results section 2.5 to clarify the reasoning behind the three methods we chose for comparison to AutoFocus.

Deep learning-based methods were not considered for comparison to AutoFocus for many reasons. First and foremost, deep learning models require vast amounts of data to train a multitude of parameters (i.e. N >> p). However, in our application there are many more features than samples, which could lead to massive overfitting of deep learning models. Second, the few deep-learning clustering methods that cluster features rather than samples do not have an implementation that allows for multi-scale analysis (see https://www.sciencedirect.com/science/article/pii/S1046202318303591 and https://ojs.aaai.org/index.php/AAAI/article/view/8916) as they generally represent molecules in a learned latent space and then perform k-means clustering on the resulting distances.

Deep learning could be incorporated into the AutoFocus method by using learned latent representations of molecules to approximate molecular distances that could be fed into the hierarchical clustering step. However, our datasets do not have big enough sample sizes for these methods and this distance metric would drastically increase computational complexity. Therefore, the comparison is outside the scope of this manuscript.

REVIEWERS' COMMENTS:

Reviewer #3 (Remarks to the Author):

No more comments