nature portfolio

Peer Review File

A deep learning-based multiscale integration of spatial omics with tumor morphology

Corresponding Author: Dr Benoit Schmauch

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

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
This manuscript introduces MISO, a multiscale deep learning model that integrates spatial transcriptomics with tissue morphology to predict gene expression patterns at near single-cell resolution. The authors leverage a novel dataset and demonstrate the model's capacity to capture spatial gene expression in colorectal cancer, with promising applications in other cancer types as well. MISO's ability to improve spatial resolution and enhance performance in downstream tasks. The study effectively employs both supervised and weakly supervised learning, with thorough validation across TCGA dataset and tasks.
While the manuscript presents notable advancements, there are several areas where further clarification and analysis would strengthen the work:
1. The authors effectively demonstrate the MISO model's performance on genes with high spatial autocorrelation, particularly through its performance across varying neighborhood sizes (k). However, the manuscript does not thoroughly address how the model handles genes with low spatial autocorrelation. I recommend assessing the model's performance on distinct sets of genes categorized by their spatial autocorrelation levels (high versus low). This would provide a more comprehensive evaluation of the model's effectiveness across a broader range of spatial gene expression patterns, potentially identifying limitations or opportunities for refinement.
2. The MISO model has demonstrated potential in colorectal cancer and other cancer types. I suggest expanding the discussion to explore its ability to predict gene expression within the tumor microenvironment. This could enhance its utility for more tasks, such as molecular subtypes, staging, and prognosis. Expanding on these aspects would give readers a sense of how the model could be applied to real-world clinical scenarios.
3. The manuscript indicates that the training dataset consists of only 48 samples. Given the availability of additional spatial transcriptomics data for colorectal cancer in public repositories, has the team considered fine-tuning the model on external datasets? If not, it would be useful to discuss whether batch effects have been tested and if they significantly impact model performance. This would provide valuable insight into the model's generalization capacity across datasets.
4. The discussion on limitations focuses primarily on dataset size. I believe a more nuanced discussion should include not only data volume but also challenges related to the quality and variability of spatial transcriptomics data, potential biases, and assumptions made in the model. A broader discussion would help contextualize the model's current limitations and highlight avenues for future improvement.

Points of Curiosity:
1. In the vision transformer model, input images are divided into 196 patches of 16x16 pixels, with patch-level descriptors trained through the self-supervised "Masked Image Modeling" mechanism to learn useful features without labeled data. I am curious how this process translates into high-resolution gene expression predictions. Specifically, what are the model's inputs and outputs? How does this approach compare to multiple-instance learning (MIL), where the final prediction typically reflects the overall expression level of a tile rather than a finer resolution prediction? It would be helpful if the authors could elaborate on how gene expression is predicted at the patch level for each tile in this high-resolution setting.
2. The authors discuss the limitations of mean squared error (MSE) as a loss function when significant patient effects are present. While cosine similarity helps mitigate some of these issues by focusing on directional accuracy rather than magnitude—thus reducing the impact of patient effects and outliers—it's still unclear how well this addresses sample variability. Could the authors quantify the advantages of using cosine similarity in their specific context and provide empirical data to show how it outperforms MSE in handling patient heterogeneity?

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)
Schmauch et al. presented a deep learning framework, MISO, trained on in-house Visium spatial transcriptomics (ST) data and designed to predict gene expression from H&E images alone. The authors also illustrated the wide variety of applications of the proposed in-silico ST method, including super-resolution reconstruction for the gap between Visium spots and histology image feature embedding for molecular and clinical machine learning tasks. Such in-silico spatial transcriptomics promises cost-effectiveness and, therefore, wider clinical adoption. MISO was trained and validated on an unprecedentedly large, unpublished dataset with 72 COAD Visium samples using a state-of-the-art vision transformer backbone. The authors provided a novel contribution to the field by outlining the maximally expected performance and gene coverage in a gene expression-predicting histology deep learning model.

Although the large Visium dataset provides a solid foundation for evaluating the model's performance, the manuscript lacks comprehensive benchmarking to convincingly demonstrate its superiority over current models. Furthermore, the biological interpretation of the relationship between gene expression and morphology is underexplored, which reduced the overall depth and rigor of the work. As it stands, the individual objectives are not sufficiently developed to support a standalone publication.

Major comments

Novelty: Many similar methodologies exist for integrating H&E and Visium data for gene expression prediction and other applications. It was not explained why MISO's performance is only compared to MLP and iStar, rather than to other top-performing methods of a similar nature, such as TESLA, DeepSpaCE, BLEEP, ST-Net, HisToGene, Hist2ST, or XFuse (or simpler methods such as BayesSpace). In line 196, iStar was the only method for which the authors specifically stated its shortcomings. MISO should be compared to more methods, rather than only 1-2, to demonstrate sufficient advances in this rapidly progressing field. Since the dataset is not intended to be published as part of the paper, it cannot be used as a merit for publication of the current paper. Even as a method-paper, there was insufficient elaboration on patient characteristics or clinical and biological insights from the large ST cohort.

Comparison: The authors promised multiple applications of MISO's 1024-dimensional feature vectors but also admitted that the difference in performance was not significant compared to the baseline model, iBOTViTBase.

Explainability: The authors seem to have missed an opportunity to link morphology with molecular data and further explain the performance of MISO. The observation that MYC and other oncogenic genes/pathways were among the top genes is interesting and could potentially be tied to the morphological features MISO used for prediction. Exploring this connection would strengthen the biological relevance and provide a more comprehensive understanding of MISO's predictive capabilities.

Generalizability: Table 1 shows a significant decrease in performance after increasing the number of genes, indicating that gene selection is a crucial step in ensuring model performance. However, for applications where Visium is not available, it is difficult to accurately perform gene selection (100 out of ~20k protein-coding genes) without strong prior assumptions. This makes the model unstable in cases where there is a lack of pilot Visium data. As the authors mentioned, the impact of the study is to enable routine clinical use. It would be ideal if the authors could provide a method to recommend gene selection based solely on H&E images and clinical information. If this is not feasible, the claim of clinical routine application should be limited to the provided COAD gene list.

Minor comments

P-hacking: The use of the Stouffer test to combine p-values is inappropriate, as this test assumes that all experiments are independent of each other. In this case, that assumption may not hold. Therefore, relying on this test, as presented in Table 3 and the paragraph at line 313, should not be used as a basis to claim that the performance comparison was "significantly in favor of MISO." A different statistical approach that accounts for the potential dependence between experiments would be more suitable.

Reproducibility:

The Visium data was processed without providing code for crucial preparation steps, such as the 'in-house matter detector' or the code for selecting the most variable genes. This will make it difficult for other users to reproduce the results and apply the model to their own Visium datasets.

It is also unclear whether the authors successfully removed the batch effect. It is not explained how the model will handle an

unseen sample in terms of batch correction. Providing clarification on these points would improve the reproducibility and robustness of the study.

There are insufficient details about the Visium data generation, including which version of Visium was used, the size of the covered area, and whether the data was generated from formalin-fixed paraffin-embedded (FFPE) or fresh-frozen blocks.

Coding environment:

Authors should state which versions of python it was based on (python 3.9) in code readme, and authors should encourage users to create a virtual environment for MISO setup.

Dependency file generated by Poetry contains unnecessary packages, e.g. sqlalchemy, alembic, tinycss2, etc. Authors should refine the dependency document before publishing.

About the histopathology slides: The training data consisted of Haematoxylin, Eosin & Saffron (HES)-stained slides. However, there is a lack of details to justify how a model trained on HES slides could be effectively adapted to H&E slides. Clarifying this adaptation process would be beneficial.

When reporting the standard deviation of correlations, it should be ‘std’ instead of ‘st’, line 140 and line 141.

(Remarks on code availability)

Reviewer #4

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

(Remarks on code availability)
Please refer to the review report.

Reviewer #5

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

(Remarks on code availability)

Version 2:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

(Remarks on code availability)
The authors have adequately resolved my previous concerns and clarified the issues I raised. I appreciate their efforts.

Reviewer #6

(Remarks to the Author)
In this manuscript, the Schmauch et al. introduced MISO, a deep learning model designed to predict spatial gene expression from pathology slides. While this work to addresses a “hot topic”, aiming to overcome the cost of spatial sequencing, the

current manuscript might not yet be mature enough for publication in Nature Comm.

My major concern is the lack of downstream analyses. Without such analyses, no definitive conclusions can be drawn about the meaningful benefits of this model. Although the author demonstrate that MISO improves the correlation between inferred and measure gene expression, compared with existing methods, the improvement is marginal. It is therefore difficult to argue for model's potential in clinical applications.

Additionally, I concur with the Reviewer #3 that the manuscript lacks a clear linkage between morphological features and predictions. This issue remains unaddressed in the revised version.

(Remarks on code availability)
The training data is not available.

Version 3:

Reviewer comments:

Reviewer #6

(Remarks to the Author)
I appreciate the authors' thorough responses and find that all concerns have been satisfactorily addressed. I have no further comments.

(Remarks on code availability)

Reviewers' comments:

Reviewer #1 (Remarks to the Author): Expert in machine learning and deep learning for digital pathology and omics, and computational cancer omics

This manuscript introduces MISO, a multiscale deep learning model that integrates spatial transcriptomics with tissue morphology to predict gene expression patterns at near single-cell resolution. The authors leverage a novel dataset and demonstrate the model's capacity to capture spatial gene expression in colorectal cancer, with promising applications in other cancer types as well. MISO's ability to improve spatial resolution and enhance performance in downstream tasks. The study effectively employs both supervised and weakly supervised learning, with thorough validation across TCGA dataset and tasks.
While the manuscript presents notable advancements, there are several areas where further clarification and analysis would strengthen the work:
1. The authors effectively demonstrate the MISO model's performance on genes with high spatial autocorrelation, particularly through its performance across varying neighborhood sizes (□). However, the manuscript does not thoroughly address how the model handles genes with low spatial autocorrelation. I recommend assessing the model's performance on distinct sets of genes categorized by their spatial autocorrelation levels (high versus low). This would provide a more comprehensive evaluation of the model's effectiveness across a broader range of spatial gene expression patterns, potentially identifying limitations or opportunities for refinement.

We thank the reviewer for this insightful comment. In section "MISO can robustly predict spatial gene expression from histology images", and Fig. S8 and S9, we illustrate the fact that spatial autocorrelation, as measured by Moran's index, is very highly correlated with average gene expression and to our new metric Rmax, making these criterion redundant. As a result, it is difficult to assess independently the association between these metrics and the model performance. Figure S2 and S3 further show how the model performance varies with Rmax and confirm that performances tend to diminish when including genes with low spatial autocorrelation.

2. The MISO model has demonstrated potential in colorectal cancer and other cancer types. I suggest expanding the discussion to explore its ability to predict gene expression within the tumor microenvironment. This could enhance its utility for more tasks, such as molecular subtypes, staging, and prognosis. Expanding on these aspects would give readers a sense of how the model could be applied to real-world clinical scenarios.

We appreciate the reviewer's insightful suggestion regarding the potential application of our AI model to predict gene expression within the tumor microenvironment (TME), which indeed could broaden its clinical relevance in areas such as molecular subtyping, staging, and prognosis.
In the current study, our focus was on establishing and validating the methodological framework of the model, rather than tailoring it to a specific biological context. As a result, we did not enrich the cohort for immune-rich samples, which would be more appropriate for a targeted analysis of TME-associated genes. Additionally, the majority of well-predicted genes in our current dataset tend to be those with higher expression levels, which are often cancer-related and thus overrepresented in large cohorts.

Nevertheless, we agree that the model holds promise for applications in the TME. As a preliminary demonstration of this potential, we highlight our analysis of IGKC, an immunoglobulin coding gene, which showed strong predictive performance. This supports the feasibility of extending our model to immune-related genes and TME-specific tasks in future work. We have also added a brief discussion of this potential application in the revised manuscript.

3. The manuscript indicates that the training dataset consists of only 48 samples. Given the availability of additional spatial transcriptomics data for colorectal cancer in public repositories, has the team considered fine-tuning the model on external datasets? If not, it would be useful to discuss whether batch effects have been tested and if they significantly impact model performance. This would provide valuable insight into the model's generalization capacity across datasets.

We thank the reviewer for this important remark. In the updated version, we applied the model trained on PETACC8-Visium (colorectal cancer) on external data (293 samples) from the HEST1K database. This data comes from giving multiple organs and therapeutic areas (cancer and non-cancer), allowing to distinguish the impact of different factors on the performance.
We also retrained the model from scratch on data from the MOSAIC cohort, encompassing 348 samples from 5 different diseases to investigate the impact of quantity and diversity in the training set. To the best of our knowledge, MISO has now been trained on the largest homogeneous 10x Visium dataset.

4. The discussion on limitations focuses primarily on dataset size. I believe a more nuanced discussion should include not only data volume but also challenges related to the quality and variability of spatial transcriptomics data, potential biases, and assumptions made in the model. A broader discussion would help contextualize the model's current limitations and highlight avenues for future improvement.

We have updated the discussion to reflect limitations linked to resolution and overall dataset quality. Regarding the impact of data quality, our initial evaluations on high-quality datasets such as MOSAIC and PETACC-8 were complemented by a benchmark on the HER2ST dataset which uses an older sequencing technology, therefore affecting its quality. We showed that performances tended to be indeed lower on this dataset , yet MISO still outperformed the rest of the SoTA (best competing method R=0.168). Our experiments on HEST1K also show the impact of the resolution, as the model tends to perform better on slides with the same resolution (0.5 microns per pixel) as the one encountered in the training dataset.

Points of Curiosity:
1. In the vision transformer model, input images are divided into 196 patches of 16x16 pixels, with patch-level descriptors trained through the self-supervised "Masked Image Modeling" mechanism to learn useful features without labeled data. I am curious how this process translates into high-resolution gene expression predictions. Specifically, what are the model's inputs and outputs? How does this approach compare to multiple-instance learning (MIL), where the final prediction typically reflects the overall expression level of a tile

rather than a finer resolution prediction? It would be helpful if the authors could elaborate on how gene expression is predicted at the patch level for each tile in this high-resolution setting.

We use a new vision transformer which divides images in 256 patches of 14x14 pixels. Each patch is associated with a feature vector of size 768, and the model also outputs a class token feature vector with the same dimension. Prediction at the tile level is performed from the class token feature vector, while prediction at higher resolution is performed using the 256 patch-level feature vectors as input, precisely in a MIL setting. This is detailed in the Methods section, in Preprocessing of histology slides:
In practice, H0-mini output class tokens are used for the tile-level prediction task, while patch tokens are used for the weakly supervised approach as they provide a more fine-grained representation of the input slides.

2. The authors discuss the limitations of mean squared error (MSE) as a loss function when significant patient effects are present. While cosine similarity helps mitigate some of these issues by focusing on directional accuracy rather than magnitude—thus reducing the impact of patient effects and outliers—it's still unclear how well this addresses sample variability. Could the authors quantify the advantages of using cosine similarity in their specific context and provide empirical data to show how it outperforms MSE in handling patient heterogeneity?

The improvement brought by the loss function based on cosine similarity was assessed in the Methods section, in *Loss function* and in supplementary Figure 1:
Experiments with the baseline architecture (MLP) demonstrated superior performance of our cosine similarity-based loss compared to MSE (Average Spearman correlation of 0.346 and 0.293 respectively, p-value < 10-4, Fig. S1).

Furthermore, external validation on HEST1K and on Her2ST demonstrate the robustness of our approach when applied on samples from different sources, organs and diseases

Reviewer #3 (Remarks to the Author): Expert in machine learning, deep learning, cancer digital pathology, and spatial omics

Schmauch et al. presented a deep learning framework, MISO, trained on in-house Visium spatial transcriptomics (ST) data and designed to predict gene expression from H&E images alone. The authors also illustrated the wide variety of applications of the proposed in-silico ST method, including super-resolution reconstruction for the gap between Visium spots and histology image feature embedding for molecular and clinical machine learning tasks. Such in-silico spatial transcriptomics promises cost-effectiveness and, therefore, wider clinical adoption. MISO was trained and validated on an unprecedentedly large, unpublished dataset with 72 COAD Visium samples using a state-of-the-art vision transformer backbone. The authors provided a novel contribution to the field by outlining the maximally expected performance and gene coverage in a gene expression-predicting histology deep learning model.

Although the large Visium dataset provides a solid foundation for evaluating the model's performance, the manuscript lacks comprehensive benchmarking to convincingly

demonstrate its superiority over current models. Furthermore, the biological interpretation of the relationship between gene expression and morphology is underexplored, which reduced the overall depth and rigor of the work. As it stands, the individual objectives are not sufficiently developed to support a standalone publication.

Major comments

Novelty: Many similar methodologies exist for integrating H&E and Visium data for gene expression prediction and other applications. It was not explained why MISO's performance is only compared to MLP and iStar, rather than to other top-performing methods of a similar nature, such as TESLA, DeepSpaCE, BLEEP, ST-Net, HisToGene, Hist2ST, or XFuse (or simpler methods such as BayesSpace). In line 196, iStar was the only method for which the authors specifically stated its shortcomings. MISO should be compared to more methods, rather than only 1-2, to demonstrate sufficient advances in this rapidly progressing field. Since the dataset is not intended to be published as part of the paper, it cannot be used as a merit for publication of the current paper. Even as a method-paper, there was insufficient elaboration on patient characteristics or clinical and biological insights from the large ST cohort.

We agree with this important remark and have updated the manuscript accordingly. Besides baselines (MLP and transformer-based) exploiting the same features as MISO and iStar, we reimplemented and retrained 2 additional models, Hist2ST and HisToGene. Furthermore, Wang *et al* recently published a comprehensive benchmark of methods aiming to predict spatial transcriptomic data from histology slides. Here, we demonstrate that our model, trained in the same conditions, outperformed all proposed methods on the same benchmark on the Her2ST dataset. Even when trained in a different dataset (OOD setting), MISO remained the best model when applied on this dataset. We believe that our updates rigorously demonstrate that MISO outperforms every other method of similar nature.

Wang, C. et al. Benchmarking the translational potential of spatial gene expression prediction from histology. Nat. Commun. 16, 1544 (2025)

Patient and side characteristics were further detailed in Methods:
HES slides were scanned at a resolution of 0.5 micron per pixel. Among those 48 patients, 23 were males and 25 were females. Age ranged from 26 to 74, with a median of 60 years old. 11 patients were MMR deficient and 37 were MMR proficient. All patients were stage III according to the American Joint Commission on Cancer (AJCC) staging system.24 new samples were made available after model development and were used as an external validation cohort.

Comparison: The authors promised multiple applications of MISO's 1024-dimensional feature vectors but also admitted that the difference in performance was not significant compared to the baseline model, iBOTViTBase.

We thank the reviewer for this comment. In the new version, this part was removed as it was not conclusive enough and misaligned from the main messages of the manuscript.

Explainability: The authors seem to have missed an opportunity to link morphology with molecular data and further explain the performance of MISO. The observation that MYC and other oncogenic genes/pathways were among the top genes is interesting and could potentially be tied to the morphological features MISO used for prediction. Exploring this connection would strengthen the biological relevance and provide a more comprehensive understanding of MISO's predictive capabilities.

We sincerely thank the reviewer for this thoughtful comment highlighting the opportunity to further explore the connection between morphological features and molecular data to enhance MISO's explainability.
Our method enables gene expression prediction at near single-cell resolution, which we believe provides a strong foundation for interpreting associations between tissue morphology and gene activity. We fully agree that linking specific morphological patterns to the prediction of oncogenic genes such as *MYC* is an important direction, particularly for reinforcing the biological and clinical relevance of the model.
In this study, our primary objective was to present and validate the general framework and methodological capabilities of MISO across multiple datasets and cancer types. As such, we did not perform an in-depth analysis of morphology-driven predictions for individual genes, which we recognize would be a valuable extension. We see this as a promising avenue for future work and have added a note to this effect in the revised manuscript.

Generalizability: Table 1 shows a significant decrease in performance after increasing the number of genes, indicating that gene selection is a crucial step in ensuring model performance. However, for applications where Visium is not available, it is difficult to accurately perform gene selection (100 out of ~20k protein-coding genes) without strong prior assumptions. This makes the model unstable in cases where there is a lack of pilot Visium data. As the authors mentioned, the impact of the study is to enable routine clinical use. It would be ideal if the authors could provide a method to recommend gene selection based solely on H&E images and clinical information. If this is not feasible, the claim of clinical routine application should be limited to the provided COAD gene list.

We thank the reviewer for this comment. The aim of the present work is to deliver a rigorously validated and robust pre-trained model. In order to increase the level of validation with respect to the previous version, we extensively validated the model on external data. For instance, inference on colorectal cancer samples from HEST1K shows little degradation of performance with respect to in-domain evaluation (0.415 vs 0.457 of average Pearson correlation). Thus, the model should be used as-is and gene selection should not be performed again on new, unseen data.

Minor comments

P-hacking: The use of the Stouffer test to combine p-values is inappropriate, as this test assumes that all experiments are independent of each other. In this case, that assumption may not hold. Therefore, relying on this test, as presented in Table 3 and the paragraph at line 313, should not be used as a basis to claim that the performance comparison was

"significantly in favor of MISO." A different statistical approach that accounts for the potential dependence between experiments would be more suitable.

We thank the reviewer for this comment. As mentioned above, in the new version, this part was removed as it was not conclusive enough.

Reproducibility:

The Visium data was processed without providing code for crucial preparation steps, such as the 'in-house matter detector' or the code for selecting the most variable genes. This will make it difficult for other users to reproduce the results and apply the model to their own Visium datasets.

It is also unclear whether the authors successfully removed the batch effect. It is not explained how the model will handle an unseen sample in terms of batch correction. Providing clarification on these points would improve the reproducibility and robustness of the study.

We thank the reviewer for this comment. In this updated version, the demonstration of the model's robustness with respect to batch effects is made through external validation on samples from different hospitals, organs and diseases for instance in HEST1K.

There are insufficient details about the Visium data generation, including which version of Visium was used, the size of the covered area, and whether the data was generated from formalin-fixed paraffin-embedded (FFPE) or fresh-frozen blocks.

We thank the reviewer for this remark, we added this information in the Methods section, for instance for PETACC8-Visium:

PETACC8-Visium. *The cohort used to train the model contained 48 slides of distinct Colon Adenocarcinoma patients with 10X Genomics Visium V1 spatial transcriptomics from a subset of the PETACC-8 cohort. Spatial transcriptomics was obtained from Formalin-Fixed Paraffin-Embedded (FFPE) samples and the captured area covered 6.5 mm x 6.5 mm.*

Coding environment:

Authors should state which versions of python it was based on (python 3.9) in code readme, and authors should encourage users to create a virtual environment for MISO setup.

Dependency file generated by Poetry contains unnecessary packages, e.g. sqlalchemy, alembic, tinycss2, etc. Authors should refine the dependency document before publishing.

About the histopathology slides: The training data consisted of Haematoxylin, Eosin & Saffron (HES)-stained slides. However, there is a lack of details to justify how a model trained on HES slides could be effectively adapted to H&E slides. Clarifying this adaptation process would be beneficial.

When reporting the standard deviation of correlations, it should be ‘std’ instead of ‘st’, line 140 and line 141.

We thank the reviewer for their useful comments on the code. The readme has been clarified accordingly to explicitly state the python version and give instructions to create a virtual environment.
No specific adaptation was used to adapt the model from HES to H&E, as previous works (e.g. Saillard et al., 2019 and Schmauch et al., 2025) have shown that deep learning models can transfer robustly from HES to H&E and conversely. Furthermore, the foundation model that we use in the new version of the manuscript has been proven robust to data acquisition conditions in a recent paper (Filiot et al., 2025).

Saillard, C., Schmauch, B., Laifa, O., Moarii, M., Toldo, S., Zaslavskiy, M., ... & Calderaro, J. (2020). *Predicting survival after hepatocellular carcinoma resection using deep learning on histological slides*. Hepatology, 72(6), 2000-2013.
Schmauch, B., Cabeli, V., Domingues, O. D., Le Douget, J. E., Hardy, A., Belbahri, R., ... & Cesaroni, M. (2025). *Deep learning uncovers histological patterns of YAP1/TEAD activity related to disease aggressiveness in cancer patients*. iScience, 28(1).
Filiot, A., Dop, N., Tchita, O., Riou, A., Dubois, R., Peeters, T., ... & Olivier, A. (2025). *Distilling foundation models for robust and efficient models in digital pathology.* arXiv preprint arXiv:2501.16239.

Reviewer #4 (Remarks to the Author): Early-Career Researcher co-reviewer

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Reviewer #5 (Remarks to the Author): Early-Career Researcher co-reviewer

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

REVIEWER COMMENTS

Reviewer #6 (Remarks to the Author): Expert in machine learning, deep learning, digital pathology, and spatial omics; replaces Reviewer #3

In this manuscript, the Schmauch et al. introduced MISO, a deep learning model designed to predict spatial gene expression from pathology slides. While this work to addresses a “hot topic”, aiming to overcome the cost of spatial sequencing, the current manuscript might not yet be mature enough for publication in Nature Comm.

My major concern is the lack of downstream analyses. Without such analyses, no definitive conclusions can be drawn about the meaningful benefits of this model. Although the author demonstrate that MISO improves the correlation between inferred and measure gene expression, compared with existing methods, the improvement is marginal. It is therefore difficult to argue for model’s potential in clinical applications.

Additionally, I concur with the Reviewer #3 that the manuscript lacks a clear linkage between morphological features and predictions. This issue remains unaddressed in the revised version.

Response to Reviewer #6

We thank the reviewer for their insightful feedback. We have substantially revised the manuscript to address the major concerns regarding the need for a definitive downstream analysis and a clearer link between morphology and MISO's predictions.

1. On the concern regarding the lack of downstream analyses:

Our new section, “**MISO identifies spatial prognostic patterns of gene expression”,** moves beyond model performance to demonstrate a powerful clinical application: the discovery of spatial transcriptomics prognostic biomarkers in a large cohort where only H&E slides are available. This analysis, which stratifies patients by molecular subtype and clinical risk, is a significant downstream use case that showcases MISO's ability to generate novel, clinically relevant hypotheses. It demonstrates that MISO’s value is not just in recapitulating spatial transcriptomics data, but in enabling large-scale spatial biology studies that would otherwise be infeasible.

2. On the concern regarding the linkage between morphological features and predictions:

We have revised the text (**From spot resolution to spatial super-resolution)** to make the connection between MISO's predictions and specific, observable histopathological features more explicit.

Overall, the new analyses demonstrates that MISO learns to identify prognostically relevant tissue architectures and associate them with underlying molecular states, directly addressing the reviewer's concern and illustrating the model's utility in bridging the gap between morphology and molecular biology.