Cell Systems

Transparent Peer Review Record

Deciphering tumor ecosystems at super-resolution from spatial transcriptomics with TESLA

Jian Hu, Kyle Coleman, Daiwei Zhang, Edward B. Lee, Humam Kadara, Linghua Wang, Mingyao Li

Summary

Initial Submission: Received Sept 28, 2022
Preprint: doi.org/10.1101/2022.11.05.515256
Scientific editor: Bernadett Gaal, DPhil

First round of review: Number of reviewers: 3
3 confidential, 0 signed
Revision invited Dec 20, 2022
Major changes anticipated
Revision received Feb 6, 2023

Second round of review: Number of reviewers: 2
2 original, 0 new
2 confidential, 0 signed
Accepted March 30, 2023

Cell Systems

Transparent Peer Review Record

Editorial decision letter with reviewers' comments, first round of review

Dear Mingyao,

I'm enclosing the comments that reviewers made on your paper, which I hope you will find useful and constructive. As you'll see, they express interest in the study, but they also have a number of criticisms and suggestions. Based on these comments, it seems premature to proceed with the paper in its current form; however, if it's possible to address the concerns raised with additional analysis, we'd be interested in considering a revised version of the manuscript.

As a matter of principle, I usually only invite a revision when I'm reasonably certain that the authors' work will align with the reviewers' concerns and produce a publishable manuscript. In the case of this manuscript, the reviewers and I feel that benchmarking of TESLA using more datasets, and forthright discussion of the limitations of TESLA, supported by analysis where reasonable, would be necessary to strengthen the manuscript and make it suitable for publication in Cell Systems.

In addition to the concerns I've detailed above, I've highlighted portions of the reviews that strike me as particularly critical. I'd also like to be explicitly clear about an almost philosophical stance that we take at Cell Systems. We believe that understanding how approaches fail is fundamentally interesting: it provides critical insight into understanding how they work. We also believe that all approaches do fail and that it's unreasonable, even misleading, to expect otherwise. Accordingly, when papers are transparent and forthright about the limitations and crucial contingencies of their approaches, we consider that to be a great strength, not a weakness. Please keep this in mind when addressing Reviewer 1's comments about additional benchmarking to BayesSpace and the reviewers' comments about potential limitations of TESLA more broadly.

As you address these concerns, it's important that you and I stay on the same page. I'm always happy to talk, either over email or over Zoom, if you'd like feedback about whether your efforts are moving the manuscript in a productive direction. Do note that we generally consider papers through only one major round of revision, so the revised manuscript would be either accepted or rejected based on the next round of comments we receive from the reviewers. If you have any questions or concerns, please let me know. More technical information and advice about resubmission can be found below my signature. Please read it carefully, as it can save substantial time and effort later.

I look forward to seeing your revised manuscript.

All the best,

Bernadett

Bernadett Gaal, DPhil
Editor-in-Chief, Cell Systems

Reviewers' comments:

Reviewer #1: TESLA is a machine learning method that enhances gene expression resolution of ST data beyond the spot-level, to resolutions that match histology slides, achieving pixel-level resolution. TESLA provides improvements over existing methods (two noted - BayesSpace, XFuse) by being more comprehensive (all tissue regions, including those not captured by spots) and by being faster. TESLA also incorporates histology information into the inference process and, specifically, into the imputation step. An interesting engineering advantage of TESLA is that it supports annotation, based on the ST data, of the histology slides at pixel-level resolution.

Major comments:

● Line 62: "but still leaves a large portion of the tissue unmeasured for gene expression." - requires:
1. Reference to Supplementary Note 1
2. How much? You mention this on line 147, but best to shift that explanation to this location, and change line 147 to the brief version you have here.

● Line 67:
"For many genes that are involved in innate and adaptive immune responses, their spatial patterns are weakly correlated with histology images…"
1. Requires reference and better quantification. "Many" is not an accurate scientific notion - how did you conclude this? At the very least, point the reader to these many examples or include them in the supplementary information (or reference the original paper's examples).
2. How was the correlation measured? Are you comparing your method to histology images in the same way? If not, why? (You should mention this also later when you present your results and describe how you measured them). If it is not accomplished using the same method - this needs to be justified.

● Regarding the comparison to BayesSpace ("To show…" in L152). First, the comparison is only using this one sample.
1. Can you provide more than one comparison? The more the better. This could strengthen your manuscript.
2. If not, or if the number of comparisons will remain small, you should indicate how these were selected.
3. Did you compare to BayesSpace using other tissue types? More samples? If not - why? If yes - why are they not included? (Even if they are in favour of BayesSpace, this is also interesting, and should be provided as a discussion of the pros and cons of the two methods - TESLA and BayesSpace, to allow future research to improve upon both approaches).

● Line 254 addresses an important point, indeed. Namely that methods using spot level resolution may not be able to obtain the same results. But - your claim regarding the number of cells per spot requires references as well as more accurate numbers (e.g. ±5 cells per spot). In any case,10x are fast approaching single-cell. There are also other technologies, like Stereo-seq (that you indeed mention and compare against) that are either close to or already provide single-cell ST. It's best that you research this further in order to provide: a. The actual facts with references; and b. Justification for using TESLA even if and when single-cell ST is already available (e.g. costs...? speed?).

● With regards to your comparison to Stereo-Seq (Supplementary Note 3) :

1. Why are you not also comparing using SSIM?
2. Why not also compare with BayesSpace?
This shouldn't be too different from your analysis using CD3ε staining.

Minor comments about the presentation and the writting:

Abstract:

The abstract feels a bit rushed and may need revision. For example:

● "In discrete spots…" - perhaps the word discrete is not the right word to use as a downside that you are addressing in your method, since your method will also be discrete. Perhaps "low resolution" is a better term to stick to. This may appear in the body as well.

● "…heterogeneous immune and tumor cells" - will help to be more precise about what you mean by this? Can it be quantified?

● "Differential transcriptome programs…" - what exactly does this mean?

Body

1. It only becomes clear what you mean when you say that your method uses histology information when the reader reaches the Methods section (specifically line 535). This is one of the advantages of the approach. Therefore - it would be beneficial for the reader to understand this when first mentioned in the text. This can be done by briefly mentioning (e.g. in L89) that you impute partially by using similarity in histology and not only using Euclidean distance on the ST image.
2. The manuscript would be more readable if you used sub-section-headings instead of many consecutive paragraphs that deal with different aspects of the manuscripts, all under: "Overview of TESLA and evaluation".
3. English needs to be revised. E.g. (there may be other places):
a. Line 206: "with pathologist's manual"
b. Line 264: "reach to single-cell resolution"

Reviewer #2: Hu, et al., present a novel method for constructing high-resolution data from low-resolution spatial transcriptomics data. This research is timely and may be useful for a wide range of readers from different fields. However, I still have a few questions that I hope the authors can help to answer them.

1. One of the major claims in the paper is that TESLA generated better results compared to BayesSpace. Is this conclusion based on one dataset or several datasets? Do they all generate similar results? Because from the published data of BayesSpace, it also seems to generate nice results.

2. TESLA can provide the gene expression patterns of the core and edge of a tumor. However, a few publications also show that sometimes there is a clear boundary between the basal/normal and the tumor in terms of gene expression, not such a gradient as shown in the paper. Does TESLA always generate such results from different tissue sections? Can this be an artifact due to the algorithm that the neighboring spots are used as reference that actually ignore the fact that cancer region and

normal region quite often has a clear boundary?

3. Sometimes, bioinformatic results do not agree with the wet lab results. Can you performed some wet lab experiments to confirm results such as that mentioned in the last point? Either by in situ hybridization or IHC is fine, alternatively other methods are acceptable.

Hopefully, these questions/suggestions can help to further demonstrate TESLA can provide accurate results.

Reviewer #3: This paper presented TESLA, a machine learning framework for multi-level tissue annotation in spatial transcriptome data. Although there are alternative tools, TESLA can integrate histological images with gene expression to annotate heterogeneous immune and tumor cells directly and efficiently. There are some novelties in the work, especially in how the image data and gene expression data are combined. It provides a useful tool for the research community. The tool is open source, with good documentation and tutorial. The method is compared with several other tools and it shows better performance. The overall results are rich in the paper. The writing is well polished. I think it is a valuable study. There are some limitations. Given extensive work has been done, more computational studies may not be needed, but the following issues should be discussed:

(1) The paper claims that "the framework is generic and can be applied to other medical conditions as long as high-resolution histology images are available." It is unclear how generalizable this method is. It is not an end-to-end framework. It appears various tuning and adjustment of some parameters are needed. The idea may be useful to study other similar data, but the software tool TESLA may not be readily usable in other cases.

(2) The image segmentation method is adapted from ref. [51]. It was designed for other types of images. The domain adaptation here may need more work.

(3) The paper did not address some specific issues in spatial transcriptome data. For example, the cell sizes may varies; spatial transcriptome spots may have significant variations from histological images due to measurement perturbation and cell diffusion.

(4) Superpixel gene expression imputation using top 10 nearest neighboring measured spots seems over-simplified. It could mix multiple cell types in many pixels.

(5) The benchmarks in terms of datasets and more tools to compare can be more.

Authors' response to the reviewers' first round comments

Attached.

Editorial decision letter with reviewers' comments, second round of review

Dear Mingyao,

I'm very pleased to let you know that the reviews of your revised manuscript are back, the peer-review process is complete, and only a few minor, editorially-guided changes are needed to move forward towards publication. Reviewer 1 was unfortunately not available to review the revised manuscript, but Reviewer 3 has kindly reviewed your revisions and replies in response to Reviewer 1, and we are happy to proceed.

In addition to the final comments from the reviewers, I've made some suggestions about your manuscript within the "Editorial Notes" section, below. Please consider my editorial suggestions carefully, ask any questions of me that you need, make all warranted changes, and then upload your final files into Editorial Manager.

I'm looking forward to going through these last steps with you. Although we ask that our editorially-guided changes be your primary focus for the moment, you may wish to consult our FAQ (final formatting checks tab) to make the final steps to publication go more smoothly. More technical information can be found below my signature, and please let me know if you have any questions.

All the best,

Bernadett

Bernadett Gaal, DPhil
Editor-in-Chief, Cell Systems

Editorial Notes

Transparent Peer Review: Thank you for electing to make your manuscript's peer review process transparent. As part of our approach to Transparent Peer Review, we ask that you add the following sentence to the end of your abstract: "A record of this paper's Transparent Peer Review process is included in the Supplemental Information." Note that this ***doesn't*** count towards your 150 word total!

Also, if you've deposited your work on a preprint server, that's great! Please drop me a quick email with your preprint's DOI and I'll make sure it's properly credited within your Transparent Peer Review record.

Manuscript Text:

- House style disallows editorializing within the text (e.g. strikingly, surprisingly, importantly, etc.), especially the Results section. These terms are a distraction and they aren't needed—

your excellent observations are certainly impactful enough to stand on their own. Please remove these words and others like them. “Notably” is suitably neutral to use once or twice if absolutely necessary.
- Please ensure that you use the word "significantly" in the statistical sense only.

Figures and Legends:

Please look over your figures keeping the following in mind:

- When color scales are used, please define them, noting units or indicating "arbitrary units," and specify whether the scale is linear or log.
- Please ensure that all figures included in your point-by-point response to the reviewers' comments are present within the final version of the paper, either within the main text or within the Supplemental Information.

STAR Methods: Note that Cell Press has recently changed the way it approaches "availability" statements for the sake of ease and clarity. Please revise the first section of your STAR Methods as follows, noting that the particular examples used might not pertain to your study. Please consult the STAR Methods guidelines for additional information.

RESOURCE AVAILABILITY

Lead Contact: Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Jane Doe (janedoe@qwerty.com).

Materials Availability: This study did not generate new materials. *-OR-* Plasmids generated in this study have been deposited at [Addgene, name and catalog number]. *-OR-* etc.

Data and Code Availability:

- **Source data statement** (described below)
- **Code statement** (described below)
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

Data and Code Availability statements **have three parts and each part must be present. Each part should be listed as a bullet point, as indicated above.**

Instructions for section 1: Data. The statements below may be used in any number or combination, but at least one must be present. They can be edited to suit your circumstance. Please ensure that all datatypes reported in your paper are represented in section 1. For more information, please consult this list of standardized datatypes and repositories recommended by Cell Press.

• [Standardized datatype] data have been deposited at [datatype-specific repository] and are publicly available as of the date of publication. Accession numbers are listed in the key resources table.

Instructions for section 2: Code. The statements below may be used in any number or combination, but at least one must be present. They can be edited to suit your circumstance. ***If you are using GitHub, please follow the instructions here to archive a "version of record" of your GitHub repo at Zenodo, then report the resulting DOI. Additionally, please note that the Cell Systems strongly recommends that you also include an explicit reference to any scripts you may have used throughout your analysis or to generate your figures within section 2.***

> • All original code has been deposited at [repository] and is publicly available as of the date of publication. DOIs are listed in the key resources table.

Instructions for section 3. Section 3 consists of the following statement: Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

In addition,

STAR Methods follows a standardized structure. Please reorganize your experimental procedures to include these specific headings in the following order: LEAD CONTACT AND MATERIALS AVAILABILITY (including the three statements detailed above); EXPERIMENTAL MODEL AND SUBJECT DETAILS (when appropriate); METHOD DETAILS (required); QUANTIFICATION AND STATISTICAL ANALYSIS (when appropriate); ADDITIONAL RESOURCES (when appropriate). We're happy to be flexible about how each section is organized and encourage useful subheadings, but the required sections need to be there, with their headings. They should also be in the order listed. Please see the STAR Methods guide for more information or contact me for help.

Please ensure that original code has been archived in a general purpose repository recommended by Cell Press and that its DOI is provided in the Software and Algorithms section of the Key Resources Table. If you've chosen to use GitHub, please follow the instructions here to archive a "version of record" of your GitHub repo at Zenodo, complete with a DOI. Thank you!

Currently, you don't have a **Key Resources Table** (KRT). Note that the key resources table is required for manuscripts with an experimental component, and if a purely computational manuscript links to any external datasets (previously published or new), code-containing websites (e.g. a GitHub repo, noting that DOIs are strongly preferred), or uses non-standard software, it needs to include a key resources table that details these aspects of the paper. Purely computational or theoretical papers that don't contain any external links and use standard software don't require a key resources table, although you're welcome to include one if you like. For details, please refer to the Table Template or feel free to ask me for help.

Thank you!

Reviewer comments:

Reviewer #2: The author has addressed my concerns. I think the paper is now ready for publication.

Reviewer #3: The authors addressed reviewers' comments well. The revised version is improved in quality. I have no further suggestions to make.

POINT-BY-POINT RESPONSE TO REVIEWERS' COMMENTS

We sincerely thank the reviewers for their constructive comments, which have helped improve the exposition of our manuscript. We have made substantial improvements to our manuscript and highlighted the revised/new parts in red for ease of reviewing. Below are our point-by-point responses to the reviewers' comments. The original reviewers' comments are in **bold** *italics* and our responses are in normal font colored in blue. The reviewers' comments that the editor noted are highlighted in yellow.

Reviewer #1:

TESLA is a machine learning method that enhances gene expression resolution of ST data beyond the spot-level, to resolutions that match histology slides, achieving pixel-level resolution. TESLA provides improvements over existing methods (two noted - BayesSpace, XFuse) by being more comprehensive (all tissue regions, including those not captured by spots) and by being faster. TESLA also incorporates histology information into the inference process and, specifically, into the imputation step. An interesting engineering advantage of TESLA is that it supports annotation, based on the ST data, of the histology slides at pixel-level resolution.

We thank the Reviewer for their encouraging remarks and thoughtful comments, each of which we address below.

Major comments:

● ***Line 62: "but still leaves a large portion of the tissue unmeasured for gene expression." - requires:***

1. Reference to Supplementary Note 1

We have added a reference to **Supplementary Note 1** in the revised manuscript.

2. How much? You mention this on line 147, but best to shift that explanation to this location, and change line 147 to the brief version you have here.

Thanks for this constructive comment. We have revised the text following your suggestion.

● ***Line 67:***

"For many genes that are involved in innate and adaptive immune responses, their spatial patterns are weakly correlated with histology images…"

1. Requires reference and better quantification. "Many" is not an accurate scientific notion - how did you conclude this? At the very least, point the reader to these many examples or include them in the supplementary information (or reference the original paper's examples).

We regret that our descriptions are not precise. To avoid confusion, we have deleted this sentence in the revised manuscript.

2. How was the correlation measured? Are you comparing your method to histology images in the same way? If not, why? (You should mention this also later when you present

your results and describe how you measured them). If it is not accomplished using the same method - this needs to be justified.

Thanks for raising these questions. We realize that our previous descriptions are not precise. Therefore, we have deleted this sentence in the revised manuscript.

● Regarding the comparison to BayesSpace ("To show…" in L152). First, the comparison is only using this one sample.

1. Can you provide more than one comparison? The more the better. This could strengthen your manuscript.

We thank the reviewer for this constructive comment. In this revision, we have added comparisons to BayesSpace for every spot-resolution dataset analyzed in the paper. In addition, we included a comparison with BayesSpace using a benchmark dataset generated from 10x Xenium breast cancer data.

First, to show that TESLA outperforms BayesSpace in retaining the original gene expression patterns for every spot-resolution dataset, we considered the top 2,000 highly variable genes selected by BayesSpace for each dataset and obtained the spot-level gene expression from the enhanced expression generated by both TESLA and BayesSpace. For TESLA, we obtained the spot-level gene expression from the super-resolution gene expression image by extracting expression from circles that exactly overlap with measured spots. For BayseSpace, we summed up the expression from all sub-spots within a spot to get the spot-level gene expression. As shown in revised **Fig. 2b**, TESLA's super-resolution gene expression derived spot-level expression yields significantly higher correlations with the original spot-level gene expression than BayesSpace across all five datasets.

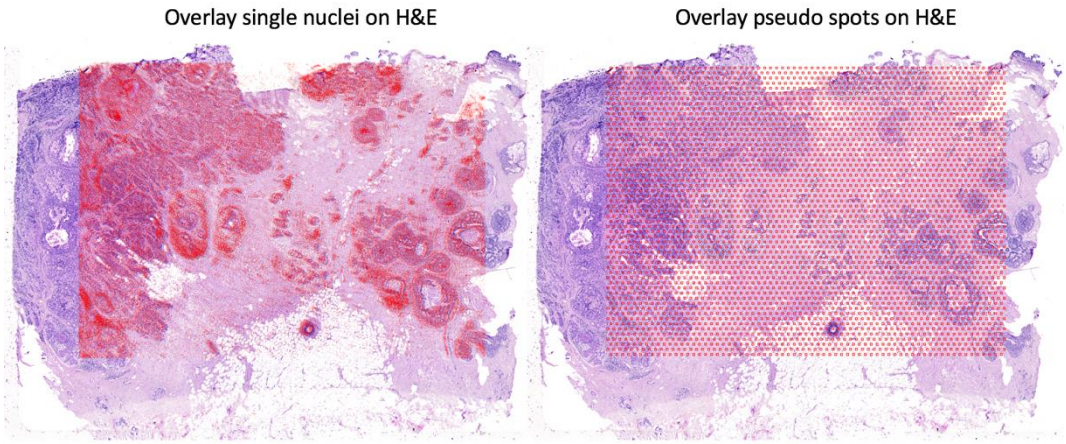
Revised **Fig. 2b**

For the IDC dataset, n=2000, median: 0.54 vs 0.30, two-sample t-test P < 2.2e-16.
For the CSCC dataset, n=1996, median: 0.67 vs 0.41, two-sample t-test P < 2.2e-16.

For the Melanoma dataset, n=,1985 median: 0.68 vs 0.41, two-sample t-test P < 2.2e-16.
For the Mouse Brain dataset, n=1976, median: 0.79 vs 0.55, two-sample t-test P < 2.2e-16.
For the Mouse Kidney dataset, n=1976, median: 0.60 vs 0.56, two-sample t-test P < 2.2e-16.

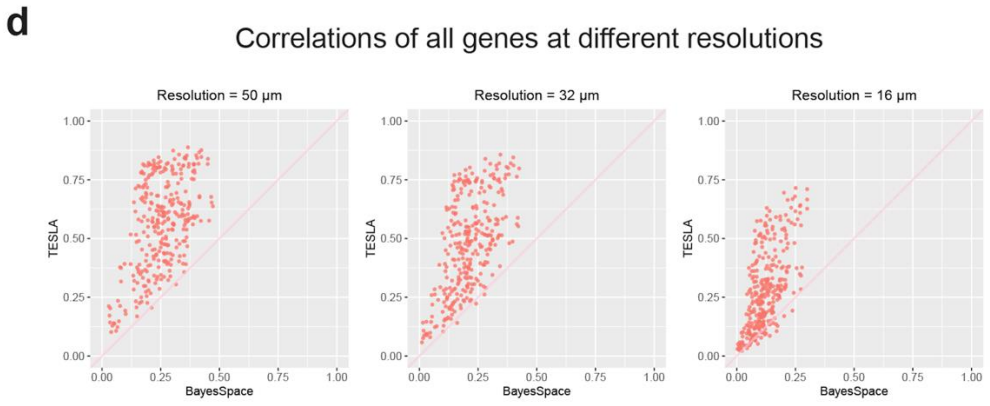
The above comparison shows that TESLA can retain the original expression pattern at the spot level better than BayesSpace.

To demonstrate TESLA can better recover gene expression pattern at a higher resolution than BayesSpace, we analyzed a newly released 10x Xenium dataset from human breast cancer tissue, which has single-cell resolution. This dataset includes 313 genes. We made a pseudo-Visium data from the Xenium data by extracting square regions with size equals 40 μm x 40 μm, which have the same area as the 55 μm circle spots in Visium. As shown in the Figure below, the center-to-center distances between adjacent pseudo spots is 100 μm, leaving the same amount of tissue gaps as the typical Visium data.



Revised **Fig. 2c**

After the pseudo-Visium data were generated, we then performed gene expression enhancement using TESLA and BayesSpace. To evaluate accuracy, we calculated Pearson correlations between the enhanced gene expression and the ground truth gene expression for all 313 genes measured in Xenium at different resolutions. As shown in the figure below, TESLA has higher correlations with the ground truth than BayesSpace across all resolutions:



Revised **Fig. 2d**

Resolution = 50 μm, median: 0.582 vs 0.253, two-sample t-test P < 2.2e-16.
Resolution = 32 μm, median: 0.486 vs 0.204, two-sample t-test P < 2.2e-16.
Resolution = 16 μm, median: 0.261 vs 0.115, two-sample t-test P < 2.2e-16.

We note that the Visium spot size is about 50 μm. As shown in **Supplementary Note 1**, there is ~70% of the tissue that is unmeasured in Visium due to the lack of spot coverage. With TESLA, we can fill in those tissue gaps and retain a relatively high correlation for the majority of genes. We also found that the correlations decrease as the spatial resolution increases. Neither TESLA nor BayesSpace can infer the single-cell level gene expression. We have acknowledged this limitation in the revised Discussion (lines 393-397). We have also added the above results in the revised manuscript.

2. If not, or if the number of comparisons will remain small, you should indicate how these were selected.

We agree. We hope the new results added in this revision demonstrate that TESLA is a more practical choice than BayesSpace.

3. Did you compare to BayesSpace using other tissue types? More samples? If not - why? If yes - why are they not included? (Even if they are in favour of BayesSpace, this is also interesting, and should be provided as a discussion of the pros and cons of the two methods - TESLA and BayesSpace, to allow future research to improve upon both approaches).

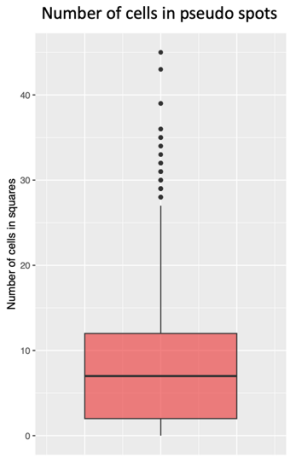
In the revised manuscript, we have expanded the comparison with BayesSpace on five more datasets, including cutaneous squamous cell carcinoma, melanoma tumor, mouse posterior brain, and mouse kidney. In all datasets, we have shown that TESLA can better retain the original gene expression patterns than BayesSpace. In addition, we added a comparison with BayesSpace on pseudo-Visium data generated from a 10x Xenium dataset to show that TESLA can recover high-resolution gene expression better than BayesSpace.

● ***Line 254 addresses an important point, indeed. Namely that methods using spot level resolution may not be able to obtain the same results. But - your claim regarding the number of cells per spot requires references as well as more accurate numbers (e.g. ±5 cells per spot). In any case,10x are fast approaching single-cell. There are also other technologies, like Stereo-seq (that you indeed mention and compare against) that are either close to or already provide single-cell ST. It's best that you research this further in order to provide: a. The actual facts with references; and b. Justification for using TESLA even if and when single-cell ST is already available (e.g. costs...? speed?).***

Thanks for these insightful comments. The number of cells within a spot and their transcriptional output depends on their type, state, and local morphology of the corresponding tissue. As reported by 10x Genomics (https://kb.10xgenomics.com/hc/en-us/articles/360035487952-How-many-cells-are-captured-in-a-single-spot-), the number of cells per spot is between 1 to 10. But in another publication (PMC8763026), it is reported that the number of cells per spot may range from 1 to 30. We believe the discrepancies are due to variations in tissue types and the state of the tissues that were studied. For the Spatial Transcriptomics platform, the diameter size of each spot is 100 μm, which is much larger than a single cell. Based on a recently published study that utilized the Spatial Transcriptomics platform (Moncada et al. 2020 Nature Biotech,

https://www.nature.com/articles/s41587-019-0392-8), it is estimated that the number of cells per spot is ~40. We have added this reference in the revised manuscript (lines 288-289).

In addition, we calculated how many cells fall into each pseudo spot in the pseudo-Visium data generated from the 10x Xenium data. As shown in the boxplot below, the number of cells in each square region ranges from 0 to 45, with a median of 7, which is similar to the number of cells reported by 10x Genomics.



Thanks for the comments about the recent advances in ST technologies. We are aware of the new ST technologies, e.g., 10x Xenium, NanoString CosMx, MERSCOPE, and Stereo-seq etc. Although these platforms provide sub-cellular resolution, none of them are perfect. For example, Xenium, CosMx and MERSCOPE are *in situ* hybridization or imaging-based, and these platforms can only measure a few hundred genes. The lack of full transcriptome coverage will limit their applicability when the goal is to do biological discoveries. Stereo-seq, on the other hand, is next-generation sequencing-based, thus it measures gene expression for the entire transcriptome. However, Stereo-seq is not easily accessible by a typical research lab due to the lack of commercialization. In addition, the sequencing cost will be extremely high because the number of spots to be sequenced is much larger than Visium. Given these constraints, we think Stereo-seq is not a practical choice for most of the research labs. Since Visium covers the entire transcriptome, its cost is reasonable, and is the most popular ST platform based on a recent survey (https://www.nature.com/articles/s41592-022-01409-2), we believe TESLA would offer a practical solution to existing Visium users. We have added a discussion about these points in the revised Discussion (lines 446-456; text also pasted below).

> *"We are aware of the new ST technologies, e.g., 10x Xenium [25] , NanoString CosMx [50], MERSCOPE, and Stereo-seq [51] etc. Although these platforms provide sub-cellular resolution, none of them are perfect. For example, Xenium, CosMx and MERSCOPE are in situ hybridization-based, and these platforms can only measure a few hundred genes. The lack of full transcriptome coverage will limit their applicability when the goal is to do biological discoveries. Stereo-seq, on the other hand, is next-generation sequencing-based, thus it measures gene expression for the entire transcriptome. However, Stereo-seq is not easily accessible by a typical research lab due to the lack of commercialization. In addition, the sequencing cost will be extremely high because the number of spots to be sequenced is much larger than Visium. Given these constraints, we think Stereo-seq is not a practical choice for most of the research labs. Since Visium*

covers the entire transcriptome, its cost is reasonable, and is the most popular ST platform based on a recent survey [52], we believe TESLA would offer a practical solution to existing Visium users."

● ***With regards to your comparison to Stereo-Seq (Supplementary Note 3):***

1. Why are you not also comparing using SSIM?

We have decided to delete the Stereo-seq example in the revised manuscript. Please see our response to your next question.

2. Why not also compare with BayesSpace? This shouldn't be too different from your analysis using CD3ε staining.

Thanks for this suggestion. The Stereo-seq data do not have companion image data, i.e., neither H&E histology image nor immunofluorescence image is available. Although TESLA can run without images, it is a method that was developed to utilize high-resolution morphology information provided by images. The performance of TESLA is suboptimal in the absence of companion image data. We analyzed the Stereo-seq data in our original submission simply because it was the only single-cell resolution spatial transcriptomics data that captures a relatively large tissue area at that time. However, after our manuscript was submitted, 10x Genomics released their Xenium dataset, which has companion high-resolution H&E histology images. Since the Xenium data fit better for the goal of TESLA, we have decided to delete the Stereo-seq application in the revision. As shown in our response to your previous comments, TESLA outperformed BayesSpace by a significant margin based on the Xenium derived benchmarking dataset.

Minor comments about the presentation and the writting:

Abstract:

The abstract feels a bit rushed and may need revision. For example:

● ***"In discrete spots…" - perhaps the word discrete is not the right word to use as a downside that you are addressing in your method, since your method will also be discrete. Perhaps "low resolution" is a better term to stick to. This may appear in the body as well.***

Thanks for this thoughtful comment. We have changed "discrete" to "low resolution" both in the Abstract and other parts of the paper.

● ***"…heterogeneous immune and tumor cells" - will help to be more precise about what you mean by this? Can it be quantified?***

We use "heterogeneous" to describe the situation where the immune and tumor cells include multiple cell subtypes and states. This is common for cancer tissues. Please see the explanation of "tumor heterogeneity" from NCI's website:
https://www.cancer.gov/publications/dictionaries/cancer-terms/def/tumor-heterogeneity

"Tumor heterogeneity: A term that describes the differences between tumors of the same type in different patients, the differences between cancer cells within a single tumor, or the differences

between a primary (original) tumor and a secondary tumor. These differences may involve the tumor's genes and proteins. For example, some cancer cells in a tumor may have genetic mutations (changes) that aren't present in other cancer cells in that tumor. Tumor heterogeneity can play an important role in how cancer is diagnosed and treated and how it responds to treatment." Due to word limit for the Abstract, we are not able to fully explain this term.

● ***"Differential transcriptome programs…" - what exactly does this mean?***

We realize that the use of "transcriptome programs" is confusing. As such, we have deleted it in the revised manuscript.

We regret that the Abstract was not clearly rewritten and thank you for the constructive comments. Following your suggestions, we have rewritten the Abstract. We hope the new version is now clearer. For your convenience, we also pasted the updated Abstract below.

> *"**Abstract***
> *Cell populations in the tumor microenvironment (TME), including their abundance, composition, and spatial location are critical determinants of patient response to therapy. Recent advances in spatial transcriptomics (ST) have enabled the comprehensive characterization of gene expression in the TME. However, popular ST platforms such as Visium only measure expression in low-resolution spots and have large tissue areas not covered by any spots, which limits their usefulness in studying the detailed structure of TME. Here we present TESLA, a machine learning framework for tissue annotation with pixel-level resolution in ST. TESLA integrates histological information with gene expression to annotate heterogeneous immune and tumor cells directly on the histology image. TESLA further detects unique TME features such as tertiary lymphoid structures, which represents a promising avenue for understanding the spatial architecture of the TME. Although we mainly illustrated the applications in cancer, TESLA can also be applied to other diseases."*

Body

1. It only becomes clear what you mean when you say that your method uses histology information when the reader reaches the Methods section (specifically line 535). This is one of the advantages of the approach. Therefore - it would be beneficial for the reader to understand this when first mentioned in the text. This can be done by briefly mentioning (e.g. in L89) that you impute partially by using similarity in histology and not only using Euclidean distance on the ST image.

Thanks for this suggestion. We have revised the manuscript to make it clear that TESLA utilizes the similarity in histology for super-resolution gene expression imputation (lines 88-97). The revised text is pasted below:

> *"The most important step in TESLA is to enhance gene expression resolution by leveraging information provided by the companion high-resolution histology image generated from the same tissue section (**Fig. 1a**). This is based on the intuition that tissue regions that share similar histological features are also likely to share similar gene expression, implying that histology similarity can help impute gene expression. To impute super-resolution gene expression and fill in tissue gaps, TESLA will first generate superpixels from the histology image. At each superpixel,*

TESLA will identify its nearest neighboring spots based on physical location and histology similarity. TESLA then impute the superpixel's gene expression through weighted aggregation of spot-level gene expression values from these neighboring spots. By the end of this imputation, TESLA will return a super-resolution gene expression image for each gene."

2. The manuscript would be more readable if you used sub-section-headings instead of many consecutive paragraphs that deal with different aspects of the manuscripts, all under: "Overview of TESLA and evaluation".

Thanks for this suggestion. We have added sub-section-headings in the revised manuscript.

3. English needs to be revised. E.g. (there may be other places):
a. Line 206: "with pathologist's manual"
b. Line 264: "reach to single-cell resolution"

Thanks for catching these problems. We have revised the corresponding text in the revised manuscript.

Reviewer #2:

Hu, et al., present a novel method for constructing high-resolution data from low-resolution spatial transcriptomics data. This research is timely and may be useful for a wide range of readers from different fields. However, I still have a few questions that I hope the authors can help to answer them.

We thank the reviewer for their positive appraisal of our paper and thoughtful comments, each of which we address below.

1. One of the major claims in the paper is that TESLA generated better results compared to BayesSpace. Is this conclusion based on one dataset or several datasets? Do they all generate similar results? Because from the published data of BayesSpace, it also seems to generate nice results.

Thanks for this comment. Reviewer #1 has raised a similar concern. In this revision, we have added comparisons to BayesSpace for every spot-resolution dataset analyzed in the paper. In addition, we included a comparison with BayesSpace using a benchmark dataset generated from 10x Xenium breast cancer data.

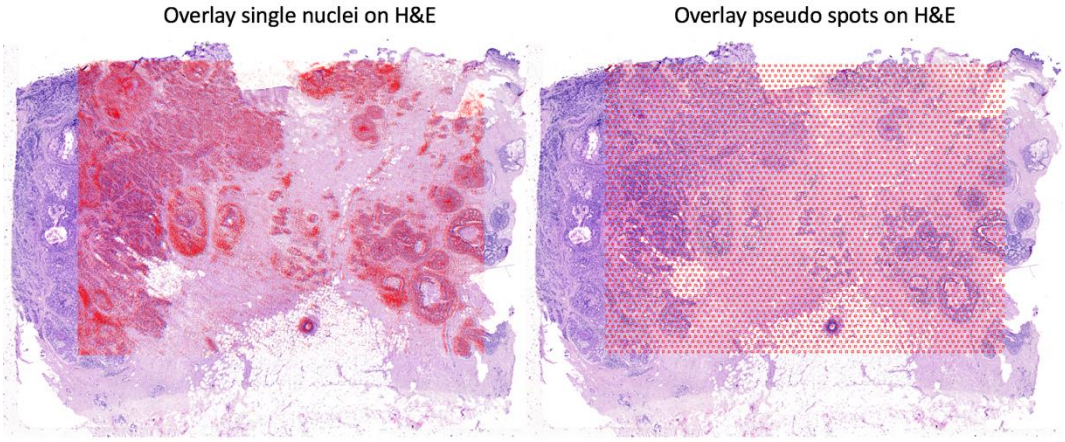
First, to show that TESLA outperforms BayesSpace in retaining the original gene expression patterns for every spot-resolution dataset, we considered the top 2,000 highly variable genes selected by BayesSpace for each dataset and obtained the spot-level gene expression from the enhanced expression generated by both TESLA and BayesSpace. For TESLA, we obtained the spot-level gene expression from the super-resolution gene expression image by extracting expression from circles that exactly overlap with measured spots. For BayseSpace, we summed up the expression from all sub-spots within a spot to get the spot-level gene expression. As shown in revised **Fig. 2b**, TESLA's super-resolution gene expression derived spot-level expression yields significantly higher correlations with the original spot-level gene expression than BayesSpace across all five datasets.

Revised **Fig. 2b**

For the IDC dataset, n=2000, median: 0.54 vs 0.30, two-sample t-test P < 2.2e-16.
For the CSCC dataset, n=1996, median: 0.67 vs 0.41, two-sample t-test P < 2.2e-16.
For the Melanoma dataset, n=,1985 median: 0.68 vs 0.41, two-sample t-test P < 2.2e-16.
For the Brain Posterior dataset, n=1976, median: 0.79 vs 0.55, two-sample t-test P < 2.2e-16.
For the Mouse Kidney dataset, n=1976, median: 0.60 vs 0.56, two-sample t-test P < 2.2e-16.

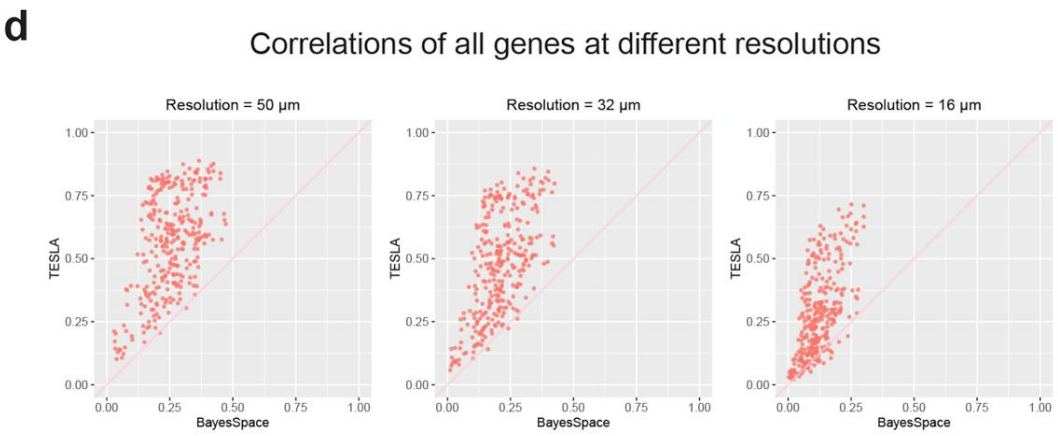
The above comparison shows that TESLA can retain the original expression pattern at the spot level better than BayesSpace.

To demonstrate TESLA can better recover gene expression pattern at a higher resolution than BayesSpace, we analyzed a newly released 10x Xenium dataset from human breast cancer tissue, which has single-cell resolution. This dataset includes 313 genes. We made a pseudo-Visium data from the Xenium data by extracting square regions with size equals 40 μm x 40 μm, which have the same area as the 55 μm circle spots in Visium. As shown in the Figure below, the center-to-center distances between adjacent pseudo spots is 100 μm, leaving the same amount of tissue gaps as the typical Visium data.



Revised **Fig. 2c**

After the pseudo-Visium data were generated, we then performed gene expression enhancement using TESLA and BayesSpace. To evaluate accuracy, we calculated Pearson correlations between the enhanced gene expression and the ground truth gene expression for all 313 genes measured in Xenium at different resolutions. As shown in the figure below, TESLA has higher correlations with the ground truth than BayesSpace across all resolutions:



Revised **Fig. 2d**

Resolution = 50 μm, median: 0.582 vs 0.253, two-sample t-test P < 2.2e-16.
Resolution = 32 μm, median: 0.486 vs 0.204, two-sample t-test P < 2.2e-16.
Resolution = 16 μm, median: 0.261 vs 0.115, two-sample t-test P < 2.2e-16.

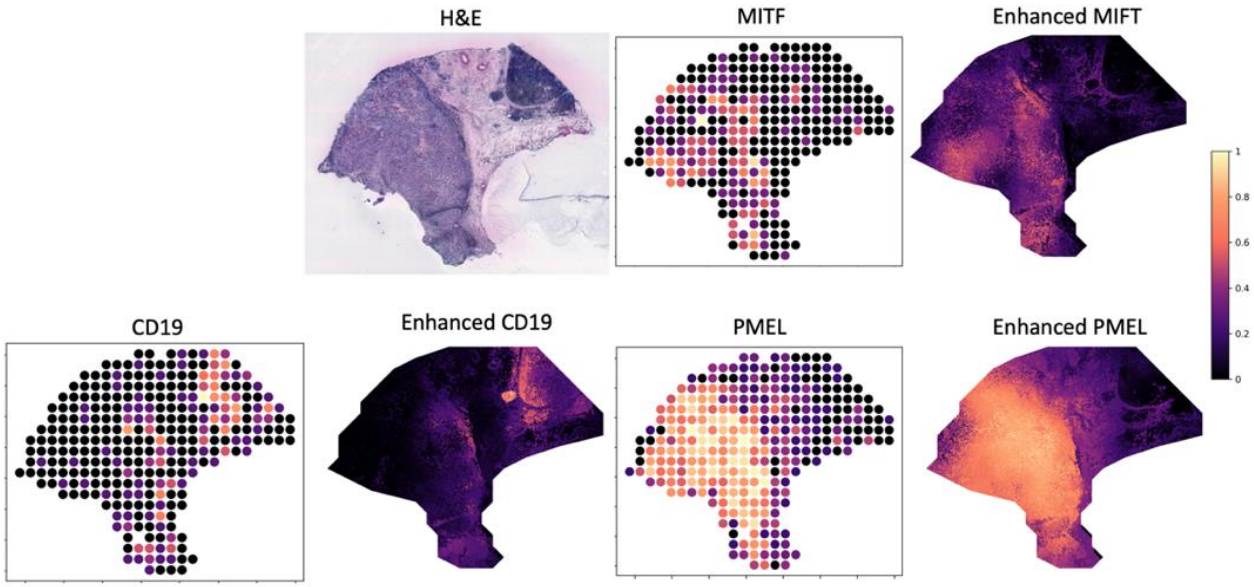
We note that the Visium spot size is about 50 μm. As shown in **Supplementary Note 1**, there is ~70% of the tissue that is unmeasured in Visium due to the lack of spot coverage. With TESLA, we can fill in those tissue gaps and retain a relatively high correlation for the majority of genes. We also found that the correlations decrease as the spatial resolution increases. Neither TESLA nor BayesSpace can infer the single-cell level gene expression. We have acknowledged this limitation in the revised Discussion (lines 393-397). We have also added the above results in the revised manuscript.

2. TESLA can provide the gene expression patterns of the core and edge of a tumor. However, a few publications also show that sometimes there is a clear boundary between the basal/normal and the tumor in terms of gene expression, not such a gradient as shown in the paper. Does TESLA always generate such results from different tissue sections? Can this be an artifact due to the algorithm that the neighboring spots are used as reference that actually ignore the fact that cancer region and normal region quite often has a clear boundary?

We thank the reviewer for raising this concern. The neighboring spots detection step takes the pixel color intensities in the H&E image into consideration. If the tumor and the adjacent normal region have distinct colors, and spots in these two regions will not be considered as each other's neighbor. Therefore, TESLA's gene expression enhancement will not blur their boundary.

Below, we provide some additional examples to show how the original gene expression and H&E image affect the region boundary in TESLA's enhanced gene expression. As shown in the figure below, in the melanoma dataset, the tumor (left), normal (middle), and lymphoid aggregate (upper-right) have very different morphology features in the H&E image. *MIFT* is a marker gene for tumor, and after TESLA's enhancement, the tumor-normal boundary is still clear. Similarly, *CD19* is a marker gene for B cells, and its enhanced gene expression still reveals the detailed structure of the lymphoid aggregate. An opposite example is *PMEL*, which is also a marker gene for tumor. However, this gene has blurred normal-tumor boundary after TESLA's enhancement. By checking the original spot-level data, we found that the boundary in the original data is not as clear as *MITF*, which led to the blurred boundary after enhancement. Based on these results, we

think TESLA will not blur tissue boundaries or add artificial boundaries that are not present in the original gene expression.



3. Sometimes, bioinformatic results do not agree with the wet lab results. Can you performed some wet lab experiments to confirm results such as that mentioned in the last point? Either by in situ hybridization or IHC is fine, alternatively other methods are acceptable.

We agree with the reviewer that more evaluation of the super-resolution gene expression with ground truth will strength this paper. Therefore, we have added evaluations using a 10X Xenium dataset generated from human breast cancer tissue. The single-cell level spatial gene expression in Xenium was treated as ground truth to compare the performance of TESLA and BayesSpace. The results are shown in our response to your previous comments. We hope these results addressed your comment.

Hopefully, these questions/suggestions can help to further demonstrate TESLA can provide accurate results.

We sincerely appreciate the reviewer's constructive comments, which have helped us to improve the quality of the paper.

Reviewer #3:

This paper presented TESLA, a machine learning framework for multi-level tissue annotation in spatial transcriptome data. Although there are alternative tools, TESLA can integrate histological images with gene expression to annotate heterogeneous immune and tumor cells directly and efficiently. There are some novelties in the work, especially in how the image data and gene expression data are combined. It provides a useful tool for the research community. The tool is open source, with good documentation and tutorial. The method is compared with several other tools and it shows better performance. The overall results are rich in the paper. The writing is well polished. I think it is a valuable study. There are some limitations. Given extensive work has been done, more computational studies may not be needed, but the following issues should be discussed:

We thank the reviewer for their positive appraisal of our paper and thoughtful comments, each of which we address below.

(1) The paper claims that "the framework is generic and can be applied to other medical conditions as long as high-resolution histology images are available." It is unclear how generalizable this method is. It is not an end-to-end framework. It appears various tuning and adjustment of some parameters are needed. The idea may be useful to study other similar data, but the software tool TESLA may not be readily usable in other cases.

We thank the reviewer for this comment. We agree that more applications to tissue types other than tumor will support the strength of TESLA. In this revision, we have added comparisons to BayesSpace for every spot-resolution dataset analyzed in the paper, which includes a mouse brain dataset and a mouse kidney dataset that are not generated from tumor.

To show that TESLA outperforms BayesSpace in retaining the original gene expression patterns for every spot-resolution dataset, we considered the top 2,000 highly variable genes selected by BayesSpace for each dataset and obtained the spot-level gene expression from the enhanced expression generated by both TESLA and BayesSpace. For TESLA, we obtained the spot-level gene expression from the super-resolution gene expression image by extracting expression from circles that exactly overlap with measured spots. For BayseSpace, we summed up the expression from all sub-spots within a spot to get the spot-level gene expression. As shown in revised **Fig. 2b**, TESLA's super-resolution gene expression derived spot-level expression yields significantly higher correlations with the original spot-level gene expression than BayesSpace across all five datasets.

Revised **Fig. 2b**

For the IDC dataset, n=2000, median: 0.54 vs 0.30, two-sample t-test P < 2.2e-16.
For the CSCC dataset, n=1996, median: 0.67 vs 0.41, two-sample t-test P < 2.2e-16.
For the Melanoma dataset, n=,1985 median: 0.68 vs 0.41, two-sample t-test P < 2.2e-16.
For the Brain Posterior dataset, n=1976, median: 0.79 vs 0.55, two-sample t-test P < 2.2e-16.
For the Mouse Kidney dataset, n=1976, median: 0.60 vs 0.56, two-sample t-test P < 2.2e-16.

We also performed spatial domain annotation for a 10x Visium mouse brain dataset using TESLA. We first analyzed this dataset using our previous developed method, SpaGCN (Hu et al. 2021 Nature Methods, DOI:10.1038/s41592-021-01255-8), to detect Spatially Variable Genes (SVGs) for different brain regions. Next, we used these SVGs to identify brain regions at super-resolution using TESLA. As shown in the figure below, TESLA can successfully identify distinct neuroanatomic subregions in the brain with “granular cell layer of the cerebellum” shown on the left and “molecular layer of the cerebellum” shown on the right. The clear boundaries between these two neuroanatomic subregions indicate the high accuracy of TESLA in spatial domain annotation in non-cancer tissues.

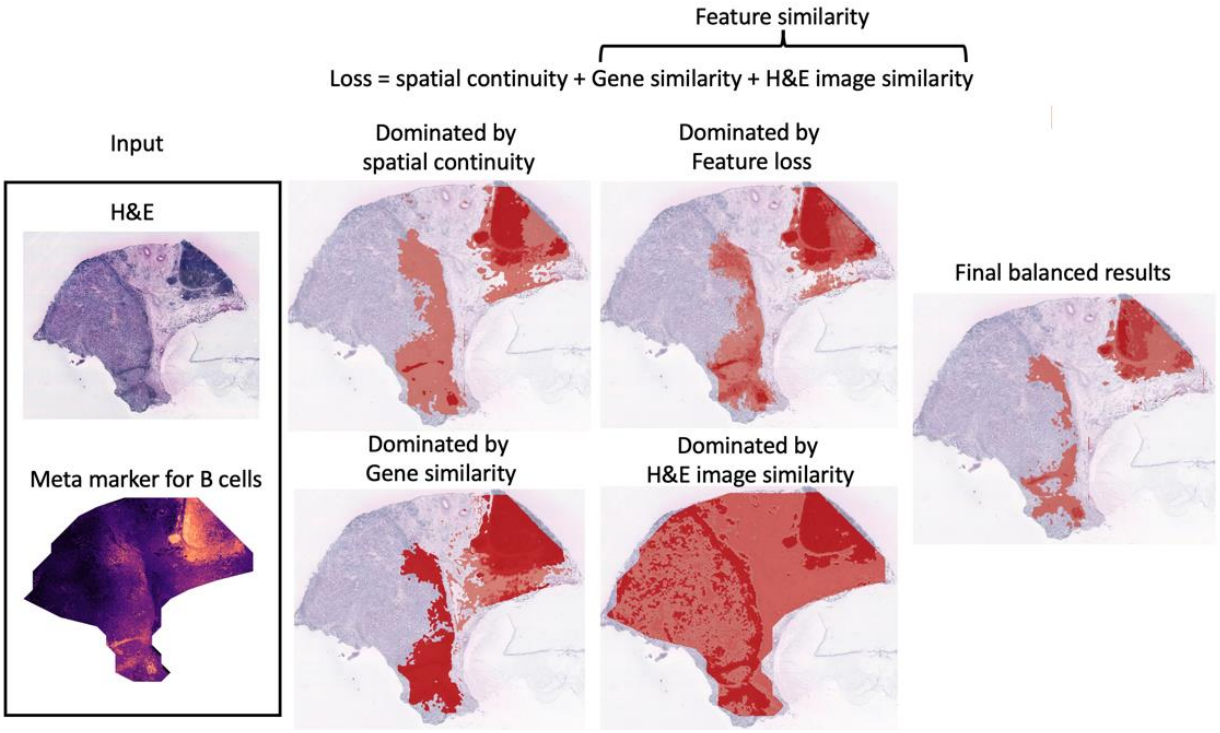
We have added the above results in **Supplementary Note 3**.

We acknowledge that the use of TESLA requires some domain knowledge from users. In particular, we require the users to specify a list of marker genes for cell type or spatial domain annotation. While we could predetermine a list of marker genes, we decide not to do so because the marker genes will vary from tissue to tissue. It is difficult to provide a comprehensive list of marker genes for all tissue types. We also think that by asking users to specify marker genes themselves, we can give them more flexibility because the users may have study-specific marker genes in their analysis.

(2) The image segmentation method is adapted from ref. [51]. It was designed for other types of images. The domain adaptation here may need more work.

Thanks for this insightful question. During the development stage of TESLA, we tried multiple image segmentation methods and the current one performed the best. An advantage of this method is that it can be directly used to analyze high-dimensional images. However, in order to make it suitable for segmenting gene expression image and H&E image, we also did several modifications.

In the original method, the loss function has 2 parts with one part for the spatial continuity loss and the other part for feature similarity loss for RGB channels in the input image. In TESLA, the input image for segmentation has 2 channels in which one channel corresponds to the gray color in the histology image and the other channel corresponds to meta gene expression in the gene image. We have modified the loss function so that it can take the two-channel image as input. As shown in the figure below,



the modified loss function in TESLA has three parts: spatial continuity, gene similarity, and H&E image similarity. Different weights are assigned to each part to make the relative contribution of the three parts similar, so that the final segmentation results will not be dominated by a single

component. We found our modified model is robust to the balance between spatial continuity and feature similarity loss. However, it is important to keep the balance between the gene similarity loss and the H&E image similarity loss; for example, if the total loss function is dominated by gene similarity, the results will miss some tissue boundaries revealed by the H&E image. On the other hand, if the total loss is dominated by the H&E image loss, the final results will fail to identify the target region since the H&E image by itself may not provide enough information about the target region. In TESLA, we normalized the gray color in the H&E image and meta-gene expression value to the same range so that the H&E image and gene expression contribute equally to the final segmentation.

We would also like to note that the original method only performs image segmentation, but does not have the functionality to do target cell type or region annotation. In TESLA, we are able to annotate each segment with a cell type or region label based on information provided by marker genes.

We have added the above explanations in the revised Methods.

(3) The paper did not address some specific issues in spatial transcriptome data. For example, the cell sizes may varies; spatial transcriptome spots may have significant variations from histological images due to measurement perturbation and cell diffusion.

We thank the reviewer for this insightful comment. We agree with the reviewer that there are some specific issues in spatial transcriptomics data that TESLA does not account for. Although it has shown promising performance, it might be overly simplistic for complex tissue types due to reasons that the reviewer pointed out. We are currently exploring a more sophisticated approach to further improve the performance of TESLA. The main idea of this extension is to extract histology image features that account for both local and long-range dependencies of pixel intensities and use a weakly supervised machine learning method to impute super-resolution gene expression by imposing constraints provided by spot-level gene expression. Fully implementing and testing this idea would require a substantial amount of work. Given the promising performance of our current version of TESLA over existing methods, we believe that implementing and testing more sophisticated approaches is beyond the scope of the current study. Nevertheless, it is an exciting future direction and we have indicated this in the Discussion section of the revised manuscript (lines 458-467; text also pasted below).

> *"There are some specific issues in ST data that TESLA does not account for. Although it has shown promising performance, it might be overly simplistic for complex tissue types. We are currently exploring a more sophisticated approach to further improve the performance of TESLA. The main idea of this extension is to extract histology image features that account for both local and long-range dependencies of pixel intensities and use a weakly supervised machine learning method to impute super-resolution gene expression by imposing constraints provided by spot-level gene expression. Fully implementing and testing this idea would require a substantial amount of work. Given the promising performance of our current version of TESLA over existing methods, we believe that implementing and testing more sophisticated approaches is beyond the scope of the current study. Nevertheless, it is an exciting future direction that we will pursue."*

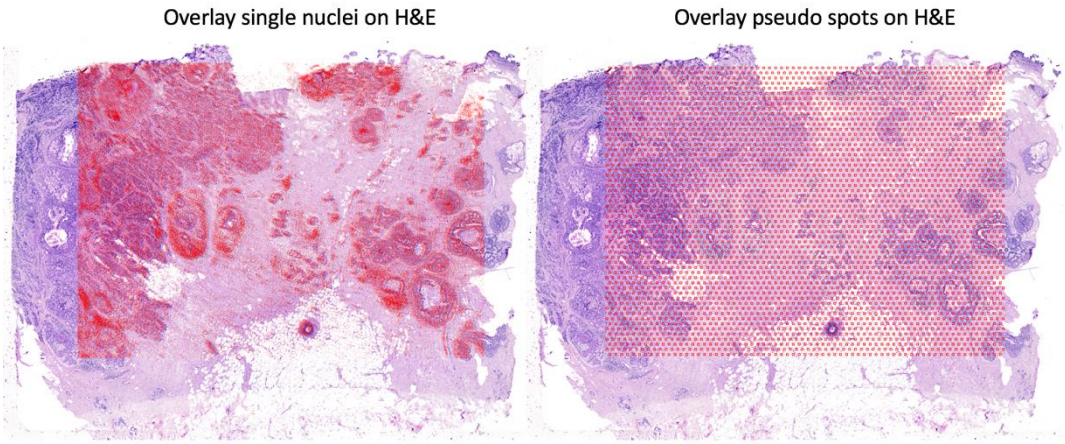
(4) Superpixel gene expression imputation using top 10 nearest neighboring measured spots seems over-simplified. It could mix multiple cell types in many pixels.

We agree with the reviewer that the current implementation of TESLA is over-simplified. In this paper, our main purpose is to demonstrate that high-resolution histology images provide valuable information for gene expression resolution enhancement. While the current results are encouraging, it is possible that the accuracy of the imputed super-resolution gene expression can be improved further with more sophisticated methods. As we responded to your previous comment, a better approach is to use a more sophisticated approach to extract histology image features that account for both local and global dependencies and use a weakly supervised machine learning method to add constraints provided spot-level gene expression. We are currently exploring this new approach and plan to report our new findings in a separate paper.

(5) The benchmarks in terms of datasets and more tools to compare can be more.

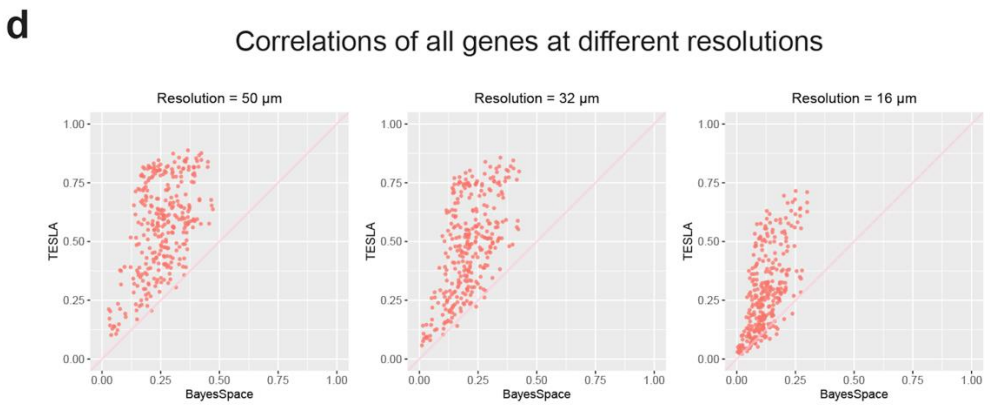
We appreciate the reviewer's constructive comment. In this revision, we have added benchmark comparisons with BayesSpace for every dataset analyzed in the paper. The results are shown in our response to your previous comments.

To demonstrate TESLA can better recover gene expression pattern at a higher resolution than BayesSpace, we also analyzed a newly released 10x Xenium dataset from human breast cancer tissue, which has single-cell resolution. This dataset includes 313 genes. We made a pseudo-Visium data from the Xenium data by extracting square regions with size equals 40 μm x 40 μm, which have the same area as the 55 μm circle spots in Visium. As shown in the Figure below, the center-to-center distances between adjacent pseudo spots is 100 μm, leaving the same amount of tissue gaps as the typical Visium data.



Revised **Fig. 2c**

After the pseudo-Visium data were generated, we then performed gene expression enhancement using TESLA and BayesSpace. To evaluate accuracy, we calculated Pearson correlations between the enhanced gene expression and the ground truth gene expression for all 313 genes measured in Xenium at different resolutions. As shown in the figure below, TESLA has higher correlations with the ground truth than BayesSpace across all resolutions:



Revised **Fig. 2d**

Resolution = 50 μm, median: 0.582 vs 0.253, two-sample t-test P < 2.2e-16.
Resolution = 32 μm, median: 0.486 vs 0.204, two-sample t-test P < 2.2e-16.
Resolution = 16 μm, median: 0.261 vs 0.115, two-sample t-test P < 2.2e-16.

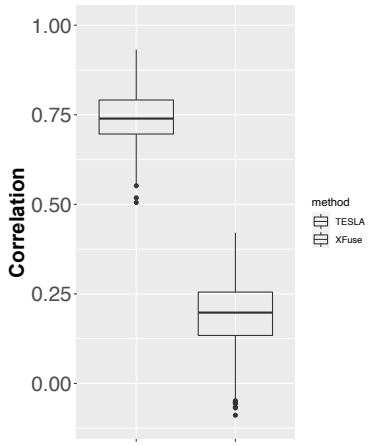
We note that the Visium spot size is about 50 μm. As shown in Supplementary Note 1, there is ~70% of the tissue that is unmeasured in Visium due to the lack of spot coverage. With TESLA, we can fill in those tissue gaps and retain a relatively high correlation for the majority of genes. We also found that the correlations decrease as the spatial resolution increases. Neither TESLA nor BayesSpace can infer the single-cell level gene expression. We have acknowledged this limitation in the revised Discussion (lines 393-397). We have also added the above results in the revised manuscript.

In addition to BayesSpace, we also compared TESLA with XFuse using the CSCC data in two aspects:

1) Run time: The gene expression enhancement step in TESLA only took less than 5 minutes on a CPU machine, while it took XFuse 17 days (408.5 hours) on the same machine.

2) The Pearson correlation between the original spot level and the enhanced gene expression for tissue region that overlaps with spots. This comparison is to ensure that the enhanced gene expression retains the original expression pattern and does not generate artificial patterns.

XFuse automatically filtered out 7376 genes. Among the remaining 9968 genes that it predicted, we selected the top 2000 highly expressed genes for evaluation. Similar to **Figure 2b**, we calculated the Pearson correlation between the spot-level enhanced gene expression and the original gene expression for TESLA and XFuse. As shown in the figure below, the median Pearson correlation for TESLA is 0.74 while the median Pearson correlation for XFuse is only 0.20. As admitted by the XFuse authors, their method may perform well only for a limited number of genes. This is not surprising as XFuse's performance is highly dependent on the histology image. For genes whose expression patterns are not similar to histology image, XFuse does not perform well.



Given how slow XFuse is, its lack of flexibility in generating super-resolution gene expression in a transparent manner, and its poor performance in our evaluations, we think it is not necessary to run XFuse for the remaining datasets included in this paper.

Finally, we would like to note that gene expression resolution enhancement is just one part of TESLA. We consider our most important contribution to be treating super-resolution gene expression data as images and integrating with histology information for high-resolution tissue annotation. To the best of our knowledge, we are the first that integrates gene expression and histology for high-resolution tissue annotation. Our tissue annotation procedure can take any super-resolution gene expression data, e.g., those generated from XFuse, as input. Since gene expression and histology provide complementary information, such integrated tissue annotation can reveal subtle tissue structures that cannot be revealed by gene expression or histology alone.