

A quantitative spatial cell-cell colocalizations framework enabling comparisons between in vitro assembloids and pathological specimens

Corresponding Author: Professor Sylvia Plevritis

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 study by Bouchard et al. introduces a novel approach to characterise spatial relationships between cell subpopulations and states within the tumour microenvironment. This analysis approach is an interesting concept and these types of approaches are increasingly in demand as large spatial datasets are being generated with greater frequency and size. Therefore, the study has the potential to have a significant impact on the field. I think the principle of this approach is exciting and novel. However, I do have some concerns regarding the statistical rigour of some aspects of the analysis pipeline and question the biological relevance of the findings presented to human disease as detailed specifically below.

Introduction:

This section doesn't provide sufficient detail on current literature, including alternate approaches for spatial analysis of the TME and their limitations; or known mechanisms of interactions and/or colocalization between fibroblast subpopulations and cancer cell phenotypes. The material that is covered in this section is also largely repeated in the results, where I would argue is where it should be covered.

Results:

Figure 1: It is not clear what the subpopulations analysed in this study represent in terms of known functional roles in the tumour microenvironment. This is a key limitation to how much can be learnt/interpreted from this study in the context of lung cancer progression and the underlying biological mechanisms. In my opinion the rationale (line 153) for using this assembloid model system to determine the subpopulations that should be investigated is misguided (153). I would expect this approach to be more informative if the subpopulations analysed for colocalization were based on those found in human tumour samples, rather in a model system where non-physiologically relevant phenotypes may have emerged. Additionally, The labelling of cancer cell subpopulations is odd. If I have understood correctly and based on the matrix provided in figure 1. Cell populations 5-8 should be shown as EPCAM+. Including this information would make it easier to understand why these cells are classed as cancer cells. For example, in the current figure population is shown as PanCK-VIM+ and classed as cancer cells, which on its own is not correct. But in fact these cells are PanCK-VIM+EPCAM+.

Figure 2: There is no clear description of how many biological replicates were used to establish the assembloid culture colocalome. It is shown that many "Statistically significant" colocalizations were observed, but whether these are within a single assembloid culture or consistently observed across cultures formed from different PDOs or fibroblast cultures (i.e. from different patients) is not described. If the former, the biological relevance of these "statistically significant" colocalisations is limited and it should be demonstrated that these colocalizations are consistently observed across biological replicates. If the latter, then the use of biological replicates in the calculation of significant colocalisations needs to be more clearly described.

Figure 3: Similar to above the use of biological replicates for these analyses needs to be clearly explained. For panel A, the legend states "Quantification of 7 to 12 representative areas" were used to assess statistical significance, this suggests that the sample size is inflated by sampling "representative regions" as opposed to biological replicates. It is also not clear why the analysis is performed on representative regions and not run across the complete scan of each assembloid culture, presumably this region sampling has the potential for introducing bias and it should be clearly described how these biases

were mitigated.

Panel b, does refer to biological replicates but only shows two and then uses these data to determine that there are no significant changes in cell population enrichments following erlotinib. This experimental design is likely to be insufficiently powered to assess whether cell population abundance is changed by erlotinib treatment. Further biological replicates are required to conclusively determine this. This should similarly be assessed in mono-culture as phenotypic shifts in cancer cell or fibroblast phenotypes could be an explanation for the emergence or preservation of certain colocalization pairings. The results section here should also be worded more carefully, as currently written it describes the data as if the measurements are made longitudinally, following the response to erlotinib treatment in time, as opposed to being measured as endpoints from control and treated cultures. This distinction also further emphasises the importance of understanding the degree of variation in colocalization pairings between biological replicates (or even merely between technical replicates of the same assembloid culture) as the conclusions drawn are heavily reliant on the assumption that these colocalization scores would be consistent across two independent assembloid cultures, which has either not been clearly demonstrated here or not clearly described.

Figure 4: The significance of these findings are dependent on the response or revisions made in response to the comments above.

Figure 5: As described for figure 1. I think this analysis would provide greater insight if the cell subpopulations were based on measurements made from human tumours (presented in figure 5), i.e. generating the CELESTA matrix from analysing human LUAD samples, as opposed to projecting the cell populations found in the model system onto the tumour samples.

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)

Evaluation of Bouchard et al:

In this manuscript, Bouchard et al. present a colocalization quotient (CLQ) based computational approach, which the authors term colocatome, as a spatial-omics feature for consideration in spatial analysis. The authors demonstrate the utilization of the “colocatome” in human lung adenocarcinoma (LUAD) assembloids using CODEX, and attempted to demonstrate the physiological relevance of tumor-stroma assembloids in LUAD tissues, guided by the “colocatome”.

Which substantial effort has been put into this study, the CLQ colocatome based method proposed here can be seen as incremental, based upon a number of spatial permutation tests already developed by multiple other groups, including as part of the Histocat package (PMID 28783155). The material & methods lacks sufficient details in multiple regards to reproduce the work, and there are various other confusing aspects that hinder the delivery in this study.

Major concerns:

1. The term “cytoarchitecture” is defined for use in the context of the brain and CNS, and is incredibly confusing to see here. The authors should refrain from redefining a well documented terminology
2. The feasibility of “colocatome”, essentially a pairwise cell-cell interaction probability and extent, has been demonstrated using orthogonal methods by multiple others groups in tissues (eg PMID 30193111, 31959985). Here, this is done only in an assembloid situation with less than ideal situations.
3. Why does the analysis of the colocatome only consider colocalization between two cell types? Will colocalization involving three or more cell types be considered, and if not, could you explain why? Furthermore, if this method is applied in a more complex tumor microenvironment, how can we ensure that it will uncover more insightful cell-cell colocalizations? This would signify a more impactful methodology beyond what has been done currently.
4. Given that assembloids were generated from only 2 cell types (cancer cells and fibroblasts), how do the following impact the physiological relevance of assembloids and downstream analysis?
 - a. Range of size of the assembloids can impact the spatial statistics
 - b. Why are immune markers (including CD45, CD56, CD68) included in the CODEXstain? (See Figure 1b panel 3)
5. Related to above, it is not clear how the authors selected the size of the field of views (FOVs). Will the size of FOV will affect prediction? As the permutation test is an empirical method and would permute all cells in a given size of FOV, I am concerned increasing the size of FOV will lead to more false positives. The author may need a simulation test to confirm the FOV size.
6. It is unclear how the CODEX antibody stainings were titrated, and annotations from CELESTA verified?
 - a. These also include how were signals normalized between sections of tissues, and how were the “customable” thresholds defined? This appears highly qualitative
7. It is unclear (line 184) how CLQ scores were cutoff to determine “high” vs “low” values.
8. Lines 200-212 needs to be substantially expanded to describe the CLQ-based methodology and analysis, given it's the backbone of this methods paper
9. Fig 3a shows number of cancer cells while Fig 3b shows fractions. Are total proportions of cancer and CAFs consistent across treatment?
10. Lines 247-249 is overstating the claim for extrapolating some of the findings.
11. Fig 5a, how the whole slide mIF LUAD clinical samples (using what technologies? Markers? Validation of annotations? Material & Methods unclear and a non working accession number for Enable Medicine is linked.

12. Additionally for Figs 5a and 5b: How much variability do we see between the patients, regions, and sections? How were the venn diagram constructed?

13. The authors point out astutely that cancer cells display “broad morphological differences”, and this may be a better area of focus for an unique spatial feature, than just a relatively incremental modification of the general cell-cell permutation analysis.

Minor concerns:

1. Given the frequent use of the term “spatial-omics”, the authors should consider other modalities to demonstrate the applicability of CLQ/colocatome, besides CODEX spatial proteomics

2. It is unclear how 20 was selected as the default parameter for determining the number of neighbors, 100 as the default Euclidean bandwidth parameter, and 500 iterations for the permutation test? The authors should demonstrate an empirical approach of arriving at such parameters.

3. In line 144, the author claims, “We expanded the capabilities of the CELESTA algorithm.” However, the new innovation of expanded CELESTA is unclear. Please explain in depth.

(Remarks on code availability)

Not available

Reviewer #3

(Remarks to the Author)

The manuscript proposes the development of a tool, the "Colocatome," for systematically recording colocalizations of cell subpopulations or states in histology slides. The authors claim that the data on colocalizations, obtained by their previously published annotating approach 'CELESTA' and their new method 'Colocatome', can serve as a reference for categorizing tumor-stroma assembloids and cancer samples. The study demonstrates the efficacy of assembloids in replicating the cellular structure of LUAD and the spatial relationships observed in clinical samples. Additionally, the authors show disparities in cell composition, colocalization, and medication responsiveness between TAF-PDO and TCF-PDO.

Overall, the authors introduce colocatome analysis as a means to classify samples and provide a functional spatial readout for clinical prediction. While the method shows potential value, there are noteworthy limitations and issues on both technical and conceptual fronts.

Major Comments

1. Sample size: The authors claim that tumor-stroma assembloids, generated using LUAD patient-derived cancer organoids (PDO) and fibroblasts isolated from cancers, exhibit distinct cellular composition and colocalization pattern. However, the overall conclusions may be compromised due to the limited sample size. The cell-cell colocalizations depicted in Fig. 2 cannot be considered reliable based on a single sample. Furthermore, in Fig. 3, the authors showed the biological replicates of the subpopulation ratio in Fig. 3b, however they did not include any repetitions of colocalizations in Fig. 3e and 3f.

2. Significance: The authors argue that the utilization of colocatome and cytoarchitecture enables the classification of tumor samples and the prediction of medication response through the use of assembloid models. However, colocatome has limitations in identifying biological characteristics, making it less valuable compared to other platforms. With the advancement of single cell RNA-seq and spatial genomics, numerous platforms have emerged that can analyze spatial transcriptome or proteome data with single-cell resolution and over 100 features. In this field, a new approach is needed to utilize the abundant data available, including not only spatial information but also other important variables. Similar to Xenium, CosMx, and MERSCOPE, Phenocycler can detect over 100 proteins in a single sample slide. The new versions of Phenocycler are not limited to detecting subpopulations like the previous version was when it was released alongside the development of 'CELESTA'. The new versions can also handle expression data matrix, such as scRNA-seq. Therefore, the usefulness of colocatome appears to be limited.

3. Statistical significance and data analysis: The authors assert in Fig. 5 that distinct colocation patterns can be identified among various types of cancer samples or fibroblasts, and they link an in vitro model to patient samples. However, out of the 262 spatial features identified in assembloids, only 69 features are present in the LUAD specimens. Furthermore, the authors claim that the solid regions of LUAD samples predominantly clustered with TAF-PDO when analyzing the colocalization patterns using hierarchical clustering. The heatmap presented in Fig. 5d does not appear to effectively differentiate between the sample groups, and the use of hierarchical clustering does not ensure their difference with statistical significance. The authors should provide statistical significance about this matter. Comparing the distinguishability of cell fractions is not meaningful because it is not even possible to categorize samples with subpopulations using scRNA-seq, which is significantly superior in cell annotation.

Minor Comments

1. The authors should carefully check the misspellings.

2. Scale bars should be added to Fig. 2c, 2d, 3g, Extended Fig. 2-5, and 7b.

3. In Fig. 3a, the mean or median line should be with violin plots.

(Remarks on code availability)

Due to the limited time, I was unable to review the details of the code.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

I would like to thank the authors for their efforts in addressing my original comments by editing the manuscript and providing a very clear response to each point. I think these changes have significantly improved the manuscript and only have minor comments that should be addressed prior to publication.

Minor comments:

I accept the points made by the authors regarding challenges associated with reliably identifying fibroblasts in human tissue sections and how this can present a significant obstacle to performing their discovery analysis; and in addition, providing the opportunity to examine later stage phenotypes where clinical samples are difficult (or not possible) to access. However, the statement in lines 326-328 (copied below) of the revised version should be re-worded to not over-state the pros of this approach and accurately convey the compromise/limitations associated with it being founded on a potentially non-physiological system.

- “To overcome these difficulties, we leveraged the cancer cell and fibroblast subpopulations identified in the assembloids as a guide to phenotype the human samples which we demonstrated is a more robust approach compared to identifying subpopulations directly from the clinical specimen.”

- o How is it demonstrated that this approach is more robust, and what is meant by more robust in this context?

- o I don't think using this method “overcome[s] these difficulties” it simply enables a comparative analysis to be performed between in vitro and human tissues. In a different context, where the cell types under investigation are identifiable in human tissues (e.g. for well defined cell populations or with the development of more accurate phenotyping and segmentation tools) this would not be a more robust approach. This caveat should be reflected in the statement above and perhaps in the discussion.

It should be noted that the fibroblast markers selected will likely fail to differentiate between mural cells and CAFs (as ACTA2, CD90 and PDGFRB are all likely to be expressed by both cell types). Evidently the assembloid analyses are not affected by this but it should be noted that this is a limitation to how comprehensively the CAF populations in human tissue sections can be delineated with this approach and the conclusions drawn from analysing the human tumour sections in the manner described.

There are also a number of grammatical errors that need to be addressed with further proof reading.

(Remarks on code availability)

The code provided gives details of the novel functions generated and used to perform the analysis described in the manuscript, which are clearly annotated.

Reviewer #2

(Remarks to the Author)

We appreciate the substantial work undertaken by the authors, including various responses that address some of our initial concerns. There are still concerns regarding various definitions and scope of analysis, which we would request the authors address appropriately for publication:

Major concerns:

1. Re 2.2, on the Feasibility of “colocatome” which is essentially a pairwise cell-cell interaction that has been demonstrated in multiple prior publications.

We continue to be confused by the definition of “colocatome”, which is now presented by the authors as a new “omics feature”, which again is essentially a cell-cell pairwise permutation test, and seems inappropriately termed as “omics”. This reviewer continues to be hesitant in this exaggerated definition and novelty of a cell-cell interaction test.

2. Re 2.4.1, on the physiological relevance of assembloids and downstream analysis.

The authors claim that CLQs in given samples form a Gaussian distribution independent of region size. This is an important point that should be substantiated with data in the manuscript.

We also further point out that the approach here favors more prevalent cell subpopulations, leading to a narrower null CLQ distribution and potentially overstating their significance due to less variability in spatial permutations. In contrast, rare subpopulations will reflect a broader null distribution, which increases the likelihood of their observed CLQ values being deemed non-significant. This discrepancy suggests a reduced sensitivity in detecting significant interactions among rarer subpopulations. The author should discuss the sensitivity issue in the discussion section.

3. Re 2.14, on the term “spatial-omics” and how the authors only demonstrate spatial proteomics applicability
The authors did not address our concerns, and further highlight the addition of TNBC IMC data. The authors should either include other “-omics” application, or tone down the claims in this study. The addition of the “validation” dataset here also raises more questions on the differences between HistoCAT and CLQ results, which should warrant a further discussion given the context of prior studies, and claims of superiority of CLQ.

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)

The authors have made commendable efforts to address the reviewers' comments by providing additional analyses and explanations. However, a major concern regarding sample size persists: the authors have used only one PDO sample for all experiments and only two biological replicates for each assembloid condition. These limitations raise concerns about the robustness and reproducibility of their colocalization utility, despite its potential as an intriguing characteristic in the context of spatial omics data.

(Remarks on code availability)

Version 2:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

I would like to thank the authors for their efforts in revising the manuscript following previous comments. I have no further concerns to raise regarding this manuscript proceeding to publication.

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)

The authors have addressed my comments. Thank you.

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)

I would like to thank the authors for addressing the concerns raised by the Reviewers. I don't have any further comments.

(Remarks on code availability)

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

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

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

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

June 2nd, 2024

Subject: Request for Reconsideration to Address Reviewers' Suggestions and Resubmit

Re: NCOMMS-23-58305-T

We thank the reviewers for their time and effort in reviewing our manuscript entitled “A spatial –omic approach for quantitative analysis of tumor-stroma assembloids and architectural mapping onto pathological specimens using the colocatome”. Their helpful feedback and constructive criticism guided us on how to better present the novelty and statistical rigor of our study and prompted us to incorporate new analyses, in an effort to improve the overall clarity and demonstrate the significance of our work. In particular, we have addressed concerns regarding the statistical rigor of certain aspects of our analytical pipeline, demonstrated the biological significance of utilizing *in vitro* assembloid models alongside human clinical specimens, and provided new evidence regarding the reproducibility of our findings. Lastly, we are pleased to present new results comparing the colocatome with other spatial metrics and alternative spatial biology platform, as suggested by the reviewers. Below is a point-by-point response addressing each comment with corresponding changes highlighted in red in the main manuscript:

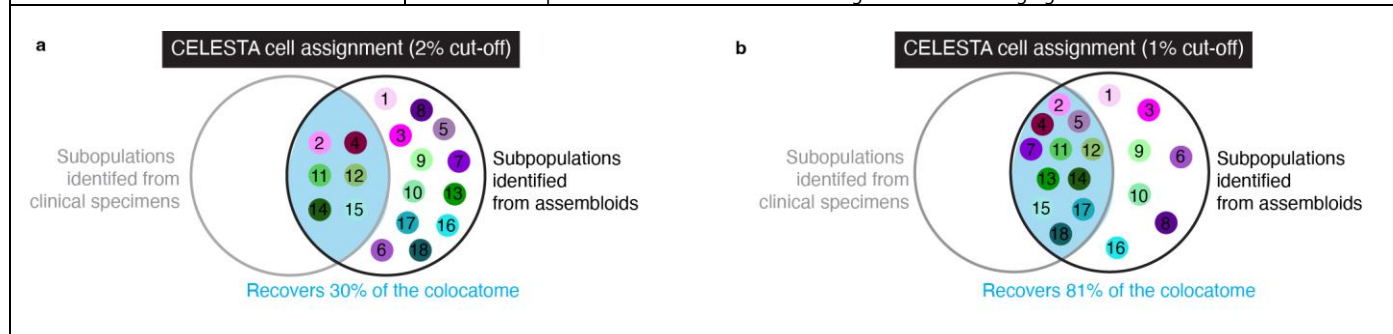
Reviewer #1 – Comprehensive response	
We thank Reviewer #1 for providing their insightful reviews and suggestions. We are delighted that Reviewer #1 describes our work as an “interesting concept” and acknowledges “the applicability of these types of approaches, given the increase in demand for analyzing spatial datasets”, and recognizes “the potential of our work to have a significant impact on the field”. We also appreciate that Reviewer #1 describes our approach as “exciting and novel”. Below, we address Reviewer #1’s concerns regarding the statistical rigor of some aspects of our analysis pipeline, as well as clarify our biological findings and demonstrate additional examples emphasizing the relevance of using <i>in vitro</i> assembloid models alongside human clinical specimens.	
Reviewer #1 – Major comments	
Comments	Answers
<p>1.1. Introduction: This section doesn’t provide sufficient detail on current literature, including alternate approaches for spatial analysis of the TME and their limitations; or known mechanisms of interactions and/or colocalization between fibroblast subpopulations and cancer cell phenotypes.</p> <p>The material that is covered in this section is also largely repeated in the results, where I would argue is where it should be covered.</p>	<p>We appreciate Reviewer #1 for highlighting missing elements of our manuscript, including alternate approaches for spatial analysis of the tumor microenvironment (TME) and their limitations, as well as known interactions and/or colocalization between fibroblast subpopulations and cancer cell phenotypes.</p> <p>We have addressed these points by adding limitations regarding common steps of data analysis related to cell identification, spatial analysis, and statistical approaches used for spatial biology data in the Introduction and strengthen our rationale for employing <i>in vitro</i> 3D patient-derived models to study complex microenvironments (lines 49 to 63). Additionally, we have incorporated additional citations about known colocalizations between fibroblast subpopulations and cancer cell phenotypes, the main cell subtypes used in our study, along with the challenges associated with studying these cell types in tissues directly (lines 98 to 102).</p> <p>We removed the duplicated content between the Introduction and Results sections.</p>
<p>1.2. Results: Figure 1: It is not clear what the subpopulations analysed in this study represent in terms of known functional roles in the tumour microenvironment. This is a key limitation to how much can be learnt/interpreted from this study in the context of lung cancer progression and the underlying biological mechanisms.</p>	<p>We appreciate the feedback from Reviewer #1 regarding the need for additional details on the functional roles of the subpopulations analyzed in our study.</p> <p>In response, we have provided more details regarding the choices of the fibroblast and epithelial markers included in our antibody panel in the result section (lines 158-162) and their association with survival outcomes in non-small-cell lung cancer or lung adenocarcinoma specifically, when available (Table 1). Our selection of markers focuses on established canonical CAF markers previously investigated in LUAD studies, most of them showing inconsistencies in association with outcomes across various studies as cited in Table 1. This inconsistency highlights the necessity for new tools and frameworks, such as ours, to study heterogenous cell subpopulations or cell states, that can be otherwise difficult to identify with confidence <i>in situ</i> because of their lack of exclusive targets and well-established morphologies, as is the case for CAFs. We address these limitations through the combinatorics of multiplexed molecular image targets at single cell resolution.</p> <p>We also agree that every spatial biology platform comes with a trade-off between single-cell information vs number of targets, which may limit what can be learnt. However, a primary objective of our study is to establish a standardized and quantitative framework for investigating the spatial organization, which can be applied to compare spatial features between different conditions, assays, and human specimens. While this study does not aim to work out specific biological mechanisms of findings that may be associated with lung cancer progression; this framework lays the groundwork for identifying cell-cell interactions worthy of subsequent research to address deep mechanisms. For example, our framework facilitates the statistically rigorous exploration of tumor-stroma reorganization in 3D assembloids following drug perturbations to address resistant, emergent and persistent spatial features, which could then be assessed by direct comparison with imaging data from patients’ specimens. Essentially, we are proposing a data mining approach to identify tumor-stroma spatial features potentially associated with lung cancer progression and drug resistance. In our case, mechanistic work to validate the biological mechanism is underway for subsequent studies designed for such purposes.</p>

Gene name	Protein name	Category	Biological role in LUAD/NSCLC	Citation
ACTA2	α -SMA	Myofibroblast and EMT marker	Well-accepted marker of myofibroblast differentiation, role in the production of contractile force during wound healing and fibrotic diseases. Highly expressed in LUAD but no direct association with survival.	1
FAP	FAP	CAF marker	Cell surface antigen expressed on the reactive stromal fibroblasts of epithelial cancers. Involved in cellular functions such as migration, extracellular matrix remodeling and immunomodulation. Highly expressed in LUAD but no direct association with survival.	2
PDGFRB	PDGFR- β	Myofibroblast marker	Cell surface tyrosine kinase receptor for members of the platelet-derived growth factor family. Involved in cellular functions such as migration, extracellular matrix production and tissue assembly. No final conclusion on the prognostic impact of PDGFR β in NSCLC.	3,4
THY1	CD90	CAF marker	Heavily N-glycosylated cell surface protein that is expressed on a variety of tumor and normal cell types including fibroblasts. Involved in cell adhesion and cell communication. Associated with worse outcome in LUAD.	5
CAV1	Caveolin-1	CAF marker	CAV1 is the main component of caveolae, which are complex plasma membrane structures with important role in cellular processes such as transport and signaling, Pro and anti-tumor functions in lung cancer.	6
PDPN	Podoplanin	CAF marker	Cell-surface mucin-like glycoprotein that plays a critical role in tumor development and normal development of the lung, kidney, and lymphatic vascular systems. Associated with worse outcome in LUAD.	7
VIM	Vimentin	EMT and pan-fibroblast marker	EMT marker, which is pivotal in tumorigenesis, metastasis, and invasion in NSCLC. An overexpression of vimentin may predict the progression and unfavorable survival outcome in NSCLC.	8
KRT	Pan-cytokeratin	Epithelial marker	Specific isoforms of individual cytokeratin may have utility as diagnostic or predictive markers in lung adenocarcinomas.	9,10
MUC1	Mucin-1	Epithelial marker	Glycoprotein present in normal epithelial tissue and in various cancers that can act as a lubricant, moisturizer, and physical barrier in cells. Overexpression predicts worse survival in NSCLC patients. Associated with EGFR-mediated resistance.	11–14
EPCAM	EpCAM	Epithelial marker	Transmembrane glycoprotein involved in intercellular adhesion and cell adhesion. May have utility as diagnostic or predictive markers in lung adenocarcinomas.	15–17
α -SMA: α -smooth muscle actin; FAP: fibroblast activation protein; PDGFR- β : Platelet-derived growth factor receptor beta; CD90: cluster of differentiation 90; LUAD: lung adenocarcinoma; EMT: epithelial-to-mesenchymal transition; NSCLC: non-small-cell lung cancer.				

Table 1 | Summary of common cancer-associated fibroblast (CAFs) and epithelial markers involved in Lung Adenocarcinoma (LUAD).

<p>1.3. In my opinion the rationale (line 153) for using this assembloid model system to determine the subpopulations that should be investigated is misguided (153). I would expect this approach to be more informative if the subpopulations analysed for colocalization were based on those found in human tumour samples, rather in a model system where non-physiologically relevant phenotypes may have emerged.</p>	<p>We understand that it may seem counterintuitive to identify the cell subpopulations from the <i>in vitro</i> assembloids rather than directly from the tumor samples. However, determining the subpopulations or cell state directly from tumor tissues poses numerous challenges including, but not limited to, the complexity of the human tumor microenvironment, shared markers between a multitude of cell types and cell states, density of tissues, computational resources required to analyzing whole-slide images and, autofluorescence from the extracellular matrix when using immunofluorescence platforms. Additional details explaining some of these challenges have been added to the Introduction (lines 61 to 63) and the Results sections (lines 343 to 348 and lines 354-359).</p> <p>Additional common cell type identification challenge in human tumors arise from the complete loss of cytokeratins in mesenchymal-like cancer cells, which often cause them to be mislabeled as fibroblasts. Moreover, subpopulations of fibroblasts pose equal difficulty in identification, as vimentin, the most widely accepted fibroblast marker, may also be expressed in immune cells, endothelial cells, and mesenchymal-like cancer cells. Of note, there isn't a unique marker to identify fibroblasts. Utilizing an assembloid model of known composition of malignant and stroma cells offers a higher level of certainty when identifying cell subpopulations or states. This serves as a guide for phenotyping the malignant and stromal compartments of complex specimens, significantly reducing the uncertainties incurred when labeling more cell types (ex: immune and endothelial, which can exhibit similar morphology or expression patterns with cancer or fibroblast subtypes) while enhancing the ability to transfer information between our in-vitro system and the human tissue samples.</p> <p>We also agree with Reviewer #1 that <i>in vitro</i> models may exhibit greater heterogeneity or less physiologically relevant phenotypes. To mitigate this issue the downstream analyses of our <i>in vitro</i> models are exclusively focused on spatial features shared between the assembloids and the clinical specimens, thereby excluding likely non-relevant biologically spatial information. We demonstrate that the vast majority of spatial features in the clinical samples are observed in the assembloids. In this regard, the abundance of cell states across the assembloids provides robustness by increasing our confidence in asserting the prevalence of certain rare features observed in clinical specimens but common <i>in vitro</i>. To further support our approach, we repeated the analysis as suggested by Reviewer #1 by first identifying the cancer cell and fibroblast subpopulations from the clinical specimens, rather than from the assembloids as we have presented in this work. Employing the exact same methodology, we applied CELESTA to identify as many cell subpopulations as possible, and utilizing the same combination of markers described in Fig. 1. The outcomes of all the tested scenarios have been added in the Supplementary Table 7 under the tab “CELESTA from clinical specimens”. When we used the same cut-off as in the original analysis and filtered out the subpopulations that accounted for less than 2% across all clinical specimens, we identified only 6 subpopulations (#2, #4, #11, #12, #14, and #15) out of 18</p>
---	--

	<p>from our original analysis (Extended Fig. 12 a and b). (Of note, the 2% threshold was used in our original analysis.) These resulting cell subpopulations recapitulated only 30% of the LUAD colocalome from the original analysis (Supplementary Table 7 under the tab “Recovered colocalome”). As an example, rare cancer cell subpopulations expressing MUC1+, an interesting glycoprotein associated with cancer resistance¹⁴, has been filtered out, missing out on interesting statistically significant spatial features associated with erlotinib resistance in the clinical specimens. Of note, no subpopulation was exclusively found in the clinical specimens, supporting the relevance of patient-derived <i>in vitro</i> models to help phenotyping complex microenvironment, particularly for rare cell subpopulations.</p> <p>Using an even stricter cut-off threshold (1% as opposed to 2%) to include rarer subpopulations, we were only able to recover 80% of the spatial features that were identified from our original analysis. Subpopulations as rare as 0.09% would have needed to be included to fully recover the LUAD colocalome from our original analysis. Incorporating such rare subpopulations raise concerns and would have resulted in the identification of several low-certainty (likely false positive) cell subpopulations. Examples of such populations are highlighted in yellow in the Supplementary Table 7 under the “CELESTA from specimens” tab. The Supplementary Table 7 results are presented for individual clinical specimens but were also assessed on the individual histological regions (lepidic, acinar, solid) to account for fraction’s variations across regions, which led to similar results (results not shown). These new results are briefly described in the revised Result section of the manuscript at lines 353 to 358, with more details in the Supplementary Results section.</p> <p>In summary, these additional results reinforce our rationale that <i>in vitro</i> assembloids help identify a greater number of cell subpopulation or states, faster, with a higher degree of certitude, without demanding the extensive computational resources needed for large whole-slide imaging datasets.</p>
--	---



<p>Extended Data Fig. 12a, b CELESTA assignments made from clinical specimens. Identification of cancer cell and fibroblast subpopulations when applying CELESTA to clinical samples rather than the assembloids using (a) a 2% cut-off and (b) a 1% cut-off for allowed cell types.</p>	
<p>1.4. Additionally, the labelling of cancer cell subpopulations is odd. If I have understood correctly and based on the matrix provided in figure 1. Cell populations 5-8 should be shown as EPCAM+. Including this information would make it easier to understand why these cells are classed as cancer cells. For example, in the current figure population is shown as PanCK- VIM+ and classed as cancer cells, which on its own is not correct. But in fact these cells are PanCK- VIM+EPCAM+.</p>	<p>We thank reviewer #1 for highlighting the confusion regarding the EpCAM status labeling. The following statement: “Note that Vim+PanCK-EpCAM- cells were assumed as fibroblasts, and that we used the expression of EpCAM to confirm that PanCK- cancer cells were not fibroblasts” is included in the Methods section at lines 591 to 592. In addition, we followed Reviewer 1’s recommendation and revised the Results section (lines 187 to 188) and modified Fig.1 and figure captions throughout the text accordingly.</p>
<p>1.5. Figure 2: There is no clear description of how many biological replicates were used to establish the assembloid culture colocalome. It is shown that many “Statistically significant” colocalizations were observed, but whether these are within a single assembloid culture or consistently observed across cultures formed from different PDOs or fibroblast cultures (I.e. from different patients) is not described. If the former, the biological relevance of these “statistically significant” colocalisations is limited and it should be demonstrated that these colocalizations are consistently observed across biological replicates. If the latter, then the use of biological replicates in the calculation of significant colocalisations needs to be more clearly described.</p>	<p>We appreciate Reviewer #1’s comment and agree that only colocalizations that are consistently observed across biological replicates should be considered. Indeed, each of the four assembloid conditions (namely: 1. TAFs–PDOs untreated, 2. TAFs–PDOs + erlotinib, 3. TCFs–PDOs untreated and, 4. TCFs–PDOs + erlotinib) has been repeated to produce two biological replicates by using fibroblast from different patients (n = 2 for each of the 4 conditions). As for cancer cells, the same PDOs were conserved across conditions to reduce variability.</p> <p>In addition to repeating every condition independently using fibroblasts from different patients, we leveraged the single cell nature of per-sample. Within each sample we performed spatial permutation analysis to identify the spatial features that display statistically significant organization. We describe the spatial permutation analysis in the Results section at lines 215 to 220 and in the Methods section at lines 616 to 624. The entirety of the CLQs with their significance are listed in Supplementary Tables 2, 5 and 7, and color-coded by replicates.</p> <p>Briefly, the statistical significance of the CLQ values were obtained through spatial permutations by randomly permuting 500 times the cell labels while preserving the spatial locations of the cells; this has the desired property of preserving the subpopulation proportions. The spatial permutation analysis allows us to generate a null distribution for each pairwise CLQ. Only actual CLQ values falling within tails of the null distribution were considered significant (P values < 0.05).</p> <p>Of note, only spatial features that were regarded as the statistically significant across the biological replicates were carried out for downstream colocalome analysis. To emphasize these points further, we added a new</p>

panel to Fig. 2 (see panel c) to clearly show that only colocalizations that are consistently observed across biological replicates are considered for colocatome analysis. Moreover, we edited the caption of Fig. 2 to clarify that the heatmap at Fig. 2d shows the mean CLQ between colocalizations that are consistently regarded as statistically significant for each biological replicate.

Additionally, our co-author, Professor Lu Tian is a statistician and helped develop the statistical framework presented here to robustly estimate cell-cell colocalizations from different conditions and assays. Dr. Tian is a well published world expert in resampling methods^{18–24}.

In summary, we employed a rigorous approach to ensure the reliability of our results that include: 1) utilizing biological fibroblast replicates from different patients, 2) conducting spatial permutation analysis to identify statistically significant colocalizations unlikely to occur by random chance, and 3) focusing solely on colocalizations that were consistently regarded as statistically significant across biological replicates.

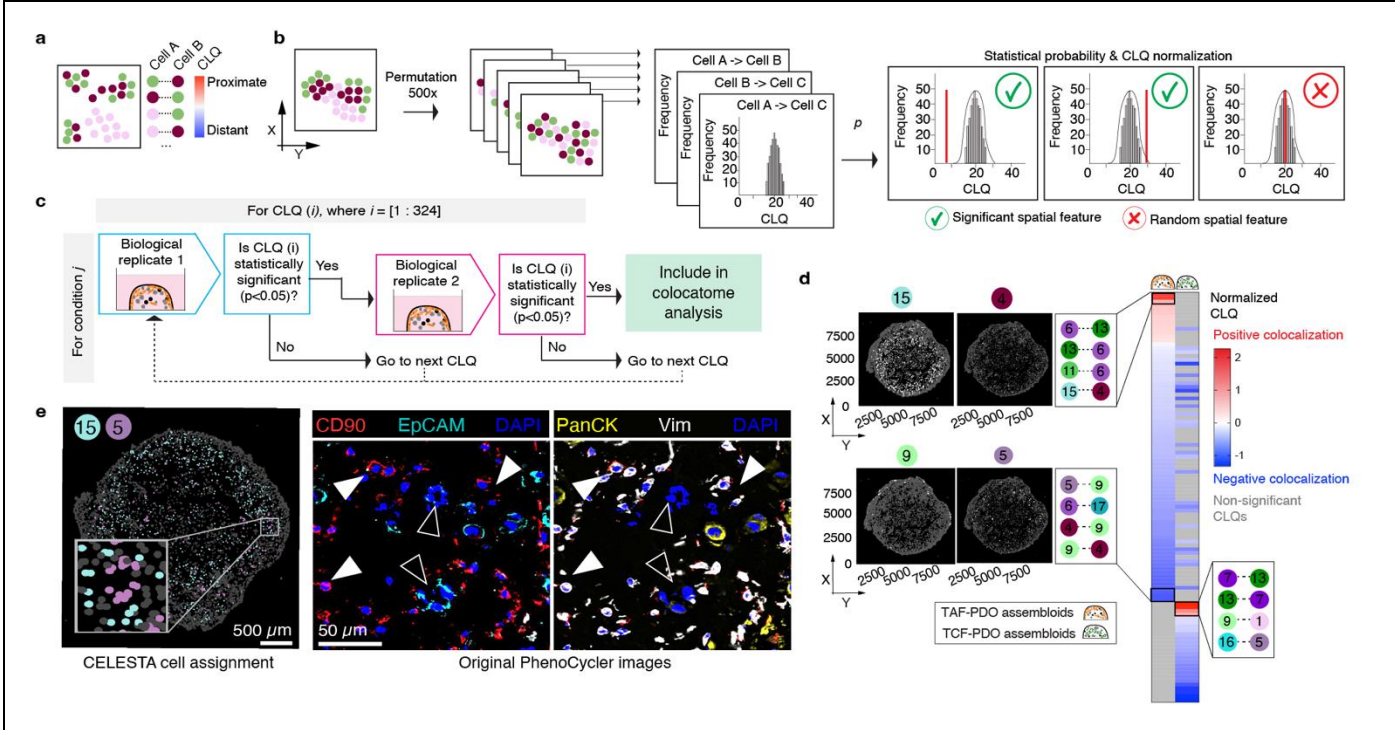


Fig. 2 a-e | Colocatome analysis enables a comparison of cell-cell colocalizations in TAF-PDO versus TCF-PDO assembloids. **a**, Schematic representation of the colocation quotient (CLQ) and **(b)** workflow of the permutation analysis to determine statistically significant cell-cell colocalizations and **(c)** selection of statistically significant cell-cell colocalizations that are consistent across biological replicates that are carried out for downstream colocatome analysis. **d**, Heatmaps showing statistically significant heterotypic negative (blue) and positive colocalizations (red) in TAF-PDO and TCF-PDO assembloids with example dot plots highlighting examples. CLQs and their corresponding p values can be found in [Supplementary Table 2](#). Only statistically significant anti- and colocalizations ($p < 0.05$) across two independent biological replicates for each condition were investigated in downstream colocatome.

1.6. Figure 3: Similar to above the use of biological replicates for these analyses needs to be clearly explained. For panel A, the legend states “Quantification of 7 to 12 representative areas” were used to assess statistical significance, this suggests that the sample size is inflated by sampling “representative regions” as opposed to biological replicates. It is also not clear why the analysis is performed on representative regions and not run across the complete scan of each assembloid culture, presumably this region sampling has the potential for introducing bias and it should be clearly described how these biases were mitigated.

1.7. Panel b, does refer to biological replicates but only shows two and then uses these data to determine that there are no significant changes in cell population enrichments following erlotinib. This experimental design is likely to be insufficiently powered to assess whether cell population abundance is changed by erlotinib

We thank Reviewer #1 for pointing out this confusing aspect of our study. In our revisions, we clarified that the spatial analyses were conducted on the complete scans of the assembloids, not on representative areas (see [lines 600 to 601](#)).

Also, we would like to emphasize that the analysis presented in [Fig. 3a](#) is independent of the CLQ calculation, spatial permutation, and colocatome analysis. Instead, the regions mentioned in the caption of [Fig. 3a](#) were exclusively designated for measuring the effect of erlotinib on cancer cell density. We included between 7 and 12 regions, each with identical dimensions, to ensure comprehensive coverage across assembloid sections of varying sizes.

We are showing that the cell type distributions of naïve vs erlotinib-treated assembloids (under both the PDO-TCF and PDO-TAF conditions) do not differ between biological replicates using ratio paired t-tests calculated with GraphPad Prism version 10.2.2. We acknowledge that increasing the number of replicates would enhance the power to demonstrate equivalence of the distributions. However, we are justifying our claims given the single cell nature of the data underlying the ratios when comparing two independent experiments as biological replicates. Each replicate represents an average number of 4,516 individual cells per sample ranging from 2,023 to 7,822 cells. This substantial sample size of single cells ensures adequate statistical power to assess whether the distributions of each cell subpopulation change between naïve and erlotinib-treated assembloids using chi-square (χ^2) tests. Given the very large sample size, extremely small differences in distributions can

<p>treatment. Further biological replicates are required to conclusively determine this.</p> <p>This should similarly be assessed in mono-culture as phenotypic shifts in cancer cell or fibroblast phenotypes could be an explanation for the emergence or preservation of certain colocalization pairings.</p>	<p>be statistically significant, as indicated by colored dots in the volcano plots (Fig. 3f and g). However, none of these changes exceeded a 5% increase or decrease across both biological replicates for each assembloid condition. Hence, we find that within a 5% change, the distributions are equivalent. This further reinforces the robustness of our analysis stating that erlotinib does not promote heterogeneity or cell subpopulation selection in the assembloids, given very small differences were observed across 36,129 cells. We included this chi-square analysis and edited the Results section accordingly at lines 265 to 267. Chi-square calculations and p-values are detailed in Supplementary Table 4.</p> <p>As for the effect of erlotinib on monocultures, erlotinib-naïve and erlotinib-treated PDO and CAF monocultures grown in 3D were also analyzed using multiplexed immunofluorescence. We agree with Reviewer #2 that some phenotypic shifts could be an explanation for the emergence or preservation of certain colocalization pairings. These controls are presented in Fig. 1i, Extended Data 6, and Extended Data Fig. 10.</p>
<div data-bbox="526 453 1096 730"> </div>	
<p>Fig. 3. f,g Volcano plots representing the differences in cell proportions between naïve and erlotinib-treated assembloids. P values are calculated using chi-square (χ^2) tests and $p < 0.05$ was considered as significant.</p>	
<p>1.8. The results section here should also be worded more carefully, as currently written it describes the data as if the measurements are made longitudinally, following the response to erlotinib treatment in time, as opposed to being measured as endpoints from control and treated cultures.</p>	<p>We carefully revised the text and refer the samples as treatment-naïve or controls, and erlotinib-treated samples to avoid confusion.</p>
<p>1.9. This distinction also further emphasizes the importance of understanding the degree of variation in colocalization pairings between biological replicates (or even merely between technical replicates of the same assembloid culture) as the conclusions drawn are heavily reliant on the assumption that these colocalization scores would be consistent across two independent assembloid cultures, which has either not been clearly demonstrated here or not clearly described.</p>	<p>Please refer to section 1.5 for clarification regarding the consistency of colocalization across biological replicates.</p> <p>Reviewer #1 is correct that there is variation in colocalization pairings between biological replicates. To clarify this point, we added the number of consistent colocalizations for each condition in Supplementary Table 2, under the "Colocalization consistency" tab. We find that approximately 54% of the statistically significant colocalizations were consistent across both TAF-PDO assembloid replicates, whereas 30% were consistent in TCF-PDO assembloids. Our downstream analysis is focused on colocalizations that are statistically significant in both biological replicates.</p>
<p>1.10. Figure 4: The significance of these findings are dependent on the response or revisions made in response to the comments above.</p>	<p>We confirm that this figure has been generated following the recommendation above and that only the colocalization consistent across both independent experiments using fibroblast biological replicates were carried on for colocalization analysis.</p>
<p>1.11. Figure 5: As described for figure 1. I think this analysis would provide greater insight if the cell subpopulations were based on measurements made from human tumours (presented in figure 5), i.e. generating the CELESTA matrix from analysing human LUAD samples, as opposed to projecting the cell populations found in the model system onto the tumour samples.</p>	<p>We reproduced our analysis using Reviewer 1's recommendation to start with the cell types identified from the clinical samples, and detailed the results in section 1.3. As discussed above, although it may seem counterintuitive to identify cell subpopulations from <i>in vitro</i> patient-derived models rather than clinical specimens, our findings highlight the added value of this approach in several ways including but not limited to: 1) increases the number of identified cell subpopulations or states, 2) significantly reduces analysis time, 3) enhances the certainty of cell assignment since the assembloids are of known constitution, and 4) alleviates the extensive computational resources required for large whole-slide imaging datasets.</p>
<p>Reviewer #2 – Comprehensive response</p>	
<p>We appreciate Reviewer #2 for recognizing the substantial effort that we invested in this study. Additionally, we thank Reviewer #2 for their time and thorough review of our paper, particularly in highlighting the missing elements to enhance the impact of our work. We carefully addressed all of them. While we acknowledge Reviewer #2's point that other groups have utilized spatial metrics and permutation analysis to identify statistically significant cell-cell colocalizations, our methods differ, and the additional novelty of our study lies in applying such methods to study <i>in vitro</i> patient-derived models and clinical specimens together. With the increasing popularity of the patient-derived models in translational research, our study provides the scientific community with a framework to assess their spatial features in the context of clinical samples. In particular, we introduce the colocalization as a spatial -omic</p>	

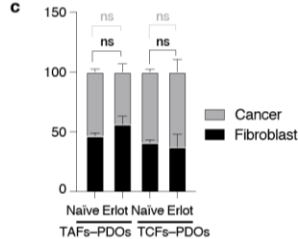
that facilitates direct comparison of spatial organization between *in vitro* and clinical specimens, as well as the comparison of different treatment conditions. This addresses the crucial need to bridge the gap between *in vitro* observations and human histological specimens.

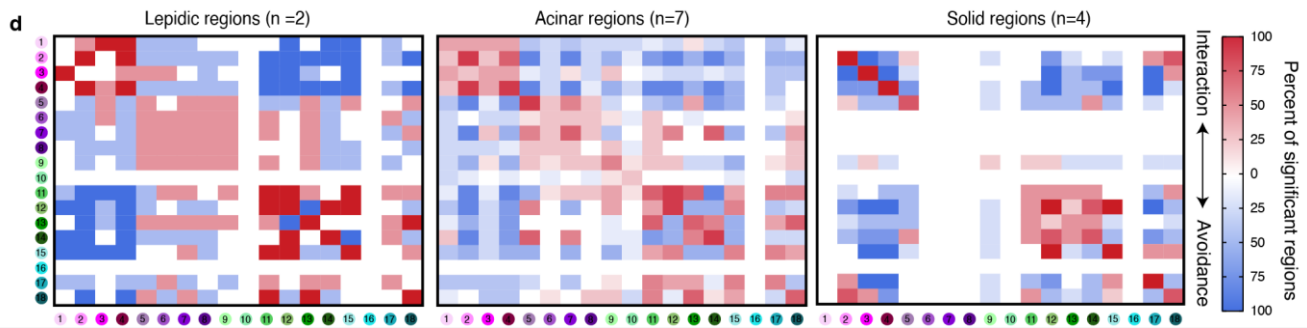
Furthermore, to fully address Reviewer 2's comments, we added a new analysis to our study to demonstrate the added value of utilizing the Colocation Quotient (CLQ) compared to other spatial metrics. The CLQ operates effectively with categorical variables and has the capability to capture asymmetric colocalization. As a new analysis we applied CLQ to the histoCAT breast cancer dataset, which was cited as a comparable analysis to our work by Reviewer #2. In doing so, we uncovered new spatial relationships that were not identified using the histoCAT toolbox, which we were able to validate on the breast cancer dataset used in the original study. We achieved this in collaboration with the first author of the histoCAT manuscript Dr. Denis Schapiro, who kindly shared raw data from their original manuscript with us, and have recognized Dr. Schapiro's help in our Acknowledgements.

Lastly, we have taken several steps to thoroughly describe the novelty and added value of conducting colocatome analysis on patient-derived *in vitro* models. These models can improve complex tissues phenotyping, as we detailed in [section 1.3](#) of this rebuttal and in the main manuscript.

Reviewer #2 – Major comments	
Comments	Answers
2.1. The term "cytoarchitecture" is defined for use in the context of the brain and CNS, and is incredibly confusing to see here. The authors should refrain from redefining a well documented terminology.	While the concept of cytoarchitecture can apply to any tissue with organizational structure at single cell resolution, including cancer, we understand Reviewer #2's point that its usage outside of neurosciences may lead to confusion. Therefore, we have removed this terminology in our manuscript and instead refer to single-cell-resolved spatial organization or tissue architecture instead.
2.2. The feasibility of "colocatome", essentially a pairwise cell-cell interaction probability and extent, has been demonstrated using orthogonal methods by multiple others groups in tissues (eg PMID 30193111, 31959985). Here, this is done only in an assembloid situation with less than ideal situations.	<p>We agree with reviewer #2 that cell-cell co-occurrence or interaction probability networks have been applied in different contexts, including the studies cited by Reviewer #2. However, our main objective is to provide a framework tailored for 3D patient-derived models, whose popularity is steadily growing. We agree that assembloid models have limitations, like other models including cell lines and animal models, but our study shows that they provide valuable information. While <i>in vitro</i> models can't perfectly replicate human tissue architecture, cell heterogeneity, or environmental cues, their usage will likely continue to rise due to their accessibility, cost advantages especially in drug perturbation and gene editing studies, and advocacy groups who promote alternatives to animal testing in research. Therefore highly-multiplexed spatial data and quantitative frameworks to assess the spatial organization of 3D patient-derived models, their statistical significance, and a standardized way to integrate the information established from such models that allow direct comparison with clinical histopathological specimens are needed.</p> <p>By rigorously quantifying the spatial information of these 3D models, we can utilize them appropriately while being aware of what aspects they accurately reproduce from tissues. Our framework facilitates direct comparison between <i>in vitro</i> and clinical tissue data. By employing the CLQ value and assessing its significance using spatial permutation null distribution, then normalizing the CLQ values within a sample, we adjust for size and prevalence variations among samples, thereby enabling comparison between different types of samples or treatment conditions.</p> <p>This is important because it adds a new -omic dimension (the colocatome) to the existing repertoire (proteome, transcriptome, etc.) for comparing <i>in vitro</i> models and tumors. Comparing the spatial organization between organoids and tumor specimens offers a novel and valuable perspective worthy of pursuit that allows for spatial -omic comparisons across different conditions and assays.</p> <p>Furthermore, we understand Reviewer #2's concerns regarding the application of colocatome analysis solely to assembloids. However, our study extends beyond assembloids as we apply the colocatome analysis to clinical specimens, namely human LUAD histopathological specimens. Additionally, in our revised manuscript, we present additional results obtained from applying colocatome analysis to a published breast cancer imaging mass cytometry dataset. These additional findings are presented in section 2.14 of this rebuttal, where Reviewer #2 suggests the application of the CLQ/colocatome to modalities other than PhenoCycler multiplexed immunofluorescence data.</p>
2.3.1. Why does the analysis of the colocatome only consider colocalization between two cell types? Will colocalization involving three or more cell types be considered, and if not, could you explain why?	We thank Reviewer #2 for bringing up this point and agree that conducting colocatome analysis on assembloid models comprising more than two cell types is meaningful. We are in the process of assessing colocalizations among more than 2 cell pairs while trying to control the bandwidth parameter, which defines the region of colocation enrichment. Also, in related ongoing research, we are extending the colocatome analysis to a network framework to identify spatial motifs comprised of 3 or more cell types. This work is underway, and the statistical properties are still being evaluated and therefore beyond the scope of the current study.
2.3.2. Furthermore, if this method is applied in a more complex tumor microenvironment, how can we ensure that it will uncover more insightful cell-cell colocalizations? This would signify a more impactful methodology beyond what has been done currently.	<p>The main goal of this study was to introduce a framework to investigate the spatial information of patient-derived <i>in vitro</i> models and empower researchers to utilize them effectively and extrapolate their findings to human tissues. While 3D patient-derived models have limitations, we demonstrate how they can accelerate our understanding of complex tumor microenvironment.</p> <p>That being said, we acknowledge that the tumor microenvironment involves more than just pairwise interactions. Currently, our lab is exploring cell-cell colocalizations that can involve more than two cell types using network-based methods. While this will undoubtedly capture more complex spatial phenotypes, we view the current method has having value in its ease of use and transparency to aid discovery process between cell type pairs, as demonstrated throughout the manuscript.</p>
2.4.1 Given that assembloids were generated from only 2 cell types (cancer	We thank Reviewer #2 for bringing up this point, and we acknowledge that the size of the assembloids can indeed impact spatial statistics. We have taken this into consideration in several ways, as summarized below.

cells and fibroblasts), how do the following impact the physiological relevance of assembloids and downstream analysis? a. Range of size of the assembloids can impact the spatial statistics.	<p>1. The CLQ, as a ratio, has desirable properties for controlling differences in the size of the region of interest. The numerator accounts for the enrichment of cell type "b" in the local neighborhood of cell type "a," while the denominator accounts for the prevalence of all cell types. After pooling the CLQs in given samples, they form a Gaussian distribution independent of the region size. Therefore, we renormalize the CLQ as a z-score and this allow us to compare samples of different sizes.</p> <p>2. When assessing the statistical significance of a given CLQ, we apply a spatial permutation method which inherently takes into account population size. As described at lines 215 to 220, this approach considers the prevalence of each cell subpopulation when determining significance. For instance, a small cell subpopulation may be sparsely redistributed within the assembloid during spatial permutation, compared to a highly prevalent cell subpopulation, resulting in less radical redistribution. Consequently, when computing the Colocation Quotient (CLQ) between two highly prevalent cell subpopulations, the null CLQ distribution tends to narrow. In contrast, rare subpopulations lead to broader null CLQ distributions due to the high variation, resulting in a broad range of CLQ values after each permutation. In the latter case, observed CLQ values are more likely to fall within the null distribution and be considered non-significant, attributing a reduced ability to assess statistical significance for the rarer cell subpopulations.</p> <p>3. We recognize that the size of the assembloid can also influence microenvironmental factors such as nutrient accessibility in the core, oxygen gradients, mechanical forces, cellular heterogeneity, etc. Therefore, we computed the CLQ values for the entire assembloid as well as for independent regions (center vs periphery of the assembloids) to evaluate if specific spatial patterns were enriched in either of these regions. Indeed, we observed numerous colocalizations that were significantly enriched at the periphery or the center. These results are included in Fig. 3 and Fig. 5 and are fully listed in Supplementary Table 5.</p>
2.4.2. b. Why are immune markers (including CD45, CD56, CD68) included in the CODEX stain? (See Figure 1b panel 3).	While we did not have immune cells in our assembloid models, we have been exploring the possibility of new "immune-like" stromal states. In previous work, we dedicated significant effort to characterizing the lung primary fibroblasts used in this study, with particular focus on the Tumor-Adjacent-Fibroblasts (TAFs) ²⁵ because of their unique morphology and transcriptomic profile. We sought to better understand their origin and hypothesized that TAFs may be fibrocytes. Fibrocytes have been described as antigen-presenting leukocytes derived from peripheral blood mononuclear cells, implicated in wound healing. Previous studies have indicated that fibrocytes are bone-marrow-derived mesenchymal cells present in injured organs and exhibit both inflammatory features of macrophages and tissue remodeling properties of fibroblasts. Therefore, we included known fibrocyte markers such as CD45 (immune marker), CD68 (pan-macrophage marker), and CD56 ²⁶ (bone marrow-derived mesenchymal stromal marker). While we detected several CD68+ TAFs, the expression was not consistent across samples to confirm or infirm our hypothesis, so we only used them as negative controls for cancer cells. We clarified the use of these markers in Fig. 1b and at lines 158 to 162 and in the Discussion at lines 449 to 454 .
2.5. Related to above, it is not clear how the authors selected the size of the field of views (FOVs). Will the size of FOV will affect prediction? As the permutation test is an empirical method and would permute all cells in a given size of FOV, I am concerned increasing the size of FOV will lead to more false positives. The author may need a simulation test to confirm the FOV size.	We agree with Reviewer #2 that field of view sizes would affect the spatial permutation analysis. As stated in 1.6 , we confirm that the spatial permutation was performed on the entire assembloids, and not on representative field of views.
2.6. It is unclear how the CODEX antibody stainings were titrated, and annotations from CELESTA verified? These also include how were signals normalized between sections of tissues, and how were the "customable" thresholds defined? This appears highly qualitative.	<p>We described the antibody validation in the Reporting Summary section reserved to this purpose. However, we realized that the Reporting Summary document may not be easily accessible, so we moved back this section to the main Methods section at lines 562 to 568.</p> <p>The signals were not normalized between sections as every assembloid section was analyzed individually to identify the cell subpopulations individually. We did not pool all the cells from the different conditions to determine the cell types; this method of cell type identification is a feature of our CELESTA algorithm as described in Zhang et al., Nature Methods 2022. Our spatial analysis is based on spatial organization between cell subpopulation, reducing the marker expression variation biases across samples. The sections are compared together through normalization of the CLQ values based on Z-score calculations, which allows comparison <i>in vitro</i> and clinical specimens together (see lines 625 to 639).</p> <p>The thresholds were chosen following the author recommendation of the original CELESTA manuscript²⁷ and the README document available at https://rdrr.io/github/plevritis-lab/CELESTA/f/README.md. A dedicated section outlining how to define thresholds was available and served as a guide. All CELESTA annotations were manually verified by comparing the cell assignment plots with the original images. Thresholds were adjusted by increasing to impose stricter conditions or decreasing to relax the conditions accordingly if cells were erroneously assigned and are included in Supplementary Table 1. This statement was added at lines 593-595.</p>
2.7. It is unclear (line 184) how CLQ scores were cutoff to determine "high" vs "low" values.	CLQ values were categorized as either positive (indicating attraction) or negative (indicating avoidance) colocalization between pairs of cell subpopulations. We revised the terms "high" and "low" to "positive" and "negative" to prevent confusion. Note that all statistically significant CLQ values were included on the continuous spectrum and that no CLQ above or below a cutoff were excluded. We appreciate Reviewer #2 for bringing this inconsistency to our attention.

2.8. Lines 200-212 needs to be substantially expanded to describe the CLQ-based methodology and analysis, given it's the backbone of this methods paper.	We have provided a more detailed explanation of the CLQ-based methodology in Methods section describing the colocalization analysis (lines 598-647). We also compare colocalization analysis using the CLQ-based method to the method presented in the HistoCat manuscript by Schapiro et al.. These results are presented in the Supplementary Results section.
2.9. Fig 3a shows number of cancer cells while Fig 3b shows fractions. Are total proportions of cancer and CAFs consistent across treatment?	We thank Reviewer #2 for raising this important point. No significant change is observed between the ratio of cancer cells and fibroblasts between naïve and erlotinib-treated assembloids. We added the summarized results as Fig. 3c and the individual values at Extended Data 6e .
 <p>Fig. 3c Bar graphs showing the fraction of cancer cells and fibroblasts in naïve vs erlotinib-treated assembloids from two independent biological replicates for each condition. Error bars indicate standard deviation of the mean. Statistical significance was determined using two-way ANOVA showing no significant difference.</p>	
2.10. Lines 247-249 is overstating the claim for extrapolating some of the findings.	We edited our statements to avoid any perception of overstating our findings. The new statement noted by the reviewer now reads as following “ <i>Overall, our findings suggest that TAFs and TCFs may protect cancer cells from erlotinib mainly through mechanisms associated with tumor-stroma spatial reorganization, rather than selecting for specific cell subpopulations.</i> ” (lines 304 to 306).
2.11. Fig 5a, how the whole slide mIF LUAD clinical samples (using what technologies? Markers? Validation of annotations? Material & Methods unclear and a non-working accession number for Enable Medicine is linked.	<p>The whole-slide multiplexed immunofluorescence (mIF) LUAD clinical samples were imaged using a Phenocycler, and the annotations were validated by a certified pathologist, Dr. Michael Osawa, who is a co-author on this manuscript. This information is available in the method section under “LUAD clinical specimens”.</p> <p>The markers used are provided in the Supplementary Table 1 with the CELESTA prior-knowledge matrix information, under the tab “LUAD matrix”.</p> <p>We apologize for the “place holder accession number” (https://www.enablemedicine.com/XX12345678XX). We deleted it to avoid confusion. As mentioned in the Data availability section, the imaging data will be made available at the moment of publication. More details were added to the LUAD clinical samples and Data availability method sections.</p>
2.12. Additionally for Figs 5a and 5b: How much variability do we see between the patients, regions, and sections? How were the venn diagram constructed?	<p>The variability of cell subpopulations between patients can be observed in Fig. 5c and Extended Fig. 12c, and all the CLQ values and significance are listed in Supplementary Table 6. Additionally, we generated heatmaps illustrating the number of significant regions for each pairwise colocalization to enable the reader to visualize variability. These new results have been added to Extended Data Fig.12d.</p> <p>More details regarding the construction of the Venn Diagram were included in the Results section (lines 360 to 370) and in the Methods section at lines 682 to 688. Briefly, CLQ values that were statistically significant in more than half of the LUAD regions of the same growth pattern (lepidic, acinar or solid) were included. If a majority of CLQ values were negative, the spatial relationship is regarded as avoidance, or vice versa for a spatial pattern that is regarded as attraction. Pairwise interactions with equal positive and negative CLQ values, which did not allow us to determine if cell pairs were being attracted or repelled, were excluded. For the assembloids, interactions consistent across conditions were included as described above in section 1.5. If all CLQs were negative across the assembloid conditions, the spatial colocalization appears as avoidance on the heatmap (Fig. 5b); otherwise, it appears as attraction.</p> <p>The left circle of the Venn Diagram (LUAD samples) represents the number of unique statistically significant colocalizations found in the clinical specimens, while the right circle represents the unique statistically significant colocalizations found in the assembloids. We found 126 shared colocalizations between assembloids and LUAD specimens, with 75% of the tumor-stroma LUAD colocalization being recapitulated <i>in vitro</i>. Note that of these 126 shared colocalizations, only 19 (roughly 10%) were identified as inconsistencies (avoidance in assembloids but attraction in the specimens, or vice-versa), and are indicated on the Venn Diagram at Fig. 5a.</p>



Extended Data Fig. 12d | Heatmaps showing the number of regions that display statistical significance for each pairwise colocalization across histological regions in the LUAD samples.

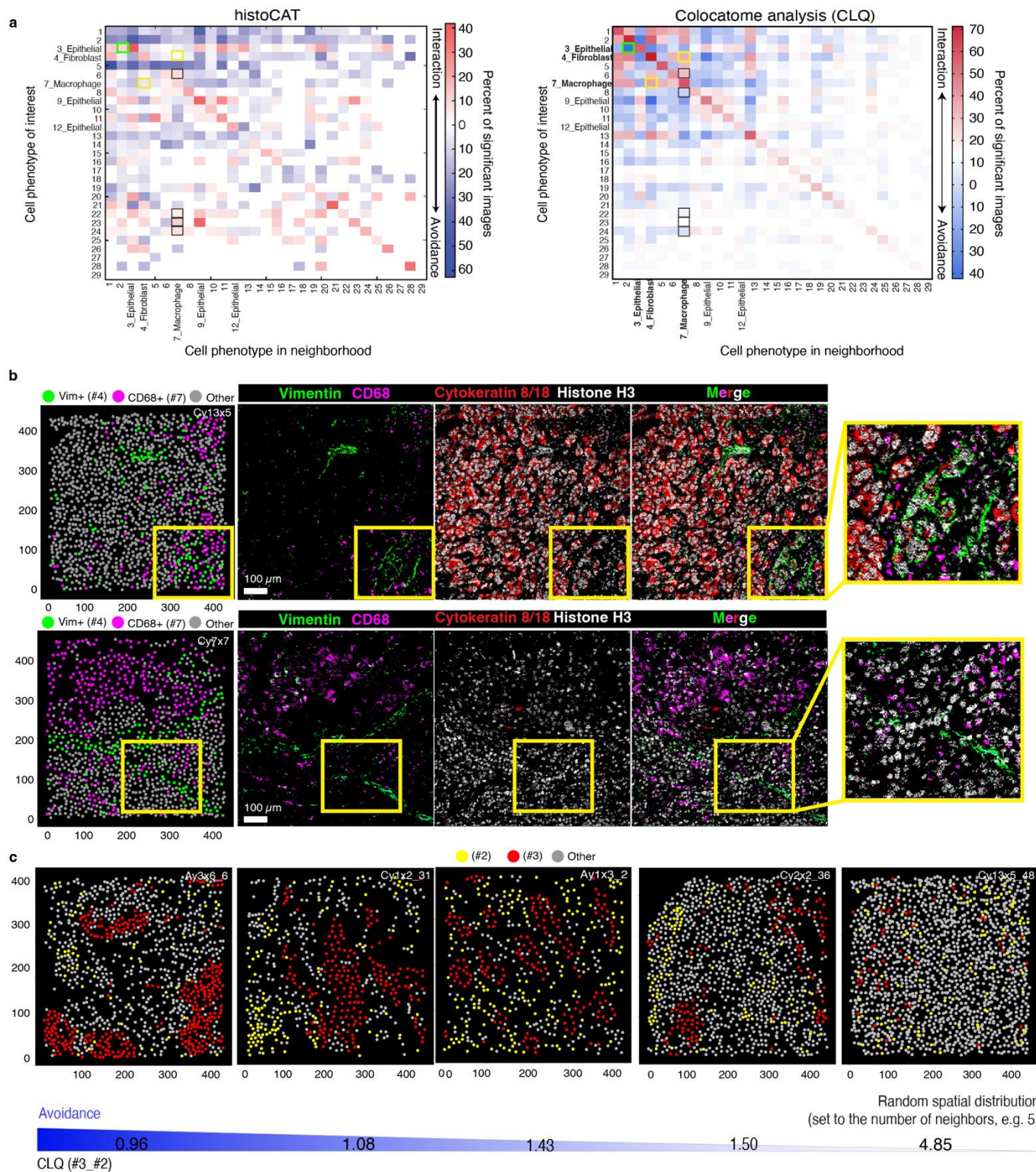
<p>2.13. The authors point out astutely that cancer cells display “broad morphological differences”, and this may be a better area of focus for a unique spatial feature, than just a relatively incremental modification of the general cell-cell permutation analysis.</p>	<p>We appreciate reviewer #2’s comment as we are actively investigating the morphological differences of cancer cells in this study, and in another ongoing study from our group.</p> <p>In this study, we measured the effect of erlotinib on the epithelial-to-mesenchymal transition (EMT) of cancer cells using a single-cell tool developed in our lab called EMT-MET PHENotypic STATE MaP (PHENOSTAMP)²⁸. PHENOSTAMP employs a neural net algorithm to characterize the phenotypic profile by mapping cancer cells onto the EMT-MET spectrum. Briefly, we assessed expression changes that are associated with morphological changes of the dissociated PDOs over the course of erlotinib using flow cytometry. We observed only a modest increase of the partial-EMT (subpopulations #1, #2, #5, and #6) or mesenchymal cell states (subpopulations #7 and #8) over time. Therefore, we did not characterize further the morphological changes of cancer cells following erlotinib in the context of this work. This data is described in the Result section (lines 297 to 306) and included in Extended Data Fig. 10e-h.</p> <p>However, we are currently expanding this work for developing a spatio-temporal analysis pipeline for multiplexed immunofluorescence data using an extensive time series dataset that takes into account EMT changes associated with drug resistance.</p>
--	---

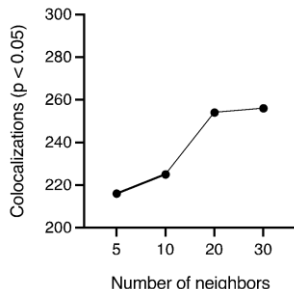
Reviewer #2 – Minor comments

Comments	Answers
<p>2.14. Given the frequent use of the term “spatial-omics”, the authors should consider other modalities to demonstrate the applicability of CLQ/colocatome, besides CODEX spatial proteomics.</p>	<p>Reviewer #2 referred to the histoCAT toolbox from Schapiro et al.²⁹, as being a comparable method to our work. This comment provided an opportunity for us to compare our colocatome analysis using the CLQ and apply our framework to another modality besides multiplexed immunofluorescence imaging data, as suggested by Reviewer #2. We reached out to Dr. Schapiro, the first author of the histoCAT manuscript, who kindly shared the breast cancer images and annotations of the cell phenotypes with us. We applied colocatome analysis and reproduced the cell-cell colocalization analysis from the histoCAT original manuscript using the CLQ and described these results at lines 397 to 404, and in the Supplementary Results. Additionally, the methodology is explained in the Methods section at lines 700-706.</p> <p>Briefly, our colocatome analysis reproduces approximately 80% of the interactions or avoidances from the original histoCAT manuscript by Schapiro et al. However, our method also reveals differences of relevance to breast cancer not identified in the histoCAT manuscript (Extended Data Fig.15a). In the original study, Schapiro et al. identified colocalization between macrophages (phenotype #7) and cell phenotypes #6, #8, #22, #23, and #24 (highlighted in black boxes). Through colocatome analysis, we observed macrophages in proximity to cell subpopulation #6, while the associations with #8, #22, #23, and #24 were noted as avoidance with overall modest significance across the images.</p> <p>Notably, we identified a positive colocalization between CD68+ cells (subpopulation #7) and subpopulations #4, which was not identified as a proximate colocalization in the original histoCAT study (yellow boxes). Based on their high vimentin expression and negative status for epithelial markers, we suggest that cell subpopulation #4 may represent fibroblasts, which we demonstrate on the original images and show that CD68+ cells colocalize with spindle-shaped vim+ cells, providing further evidence for the colocalization between macrophages and fibroblasts (Extended Data Fig.15b). Colocalization between macrophages and fibroblasts in breast cancer has been predicted in various models^{30,31}, or in recent breast cancer spatial transcriptomics studies using spot-based technologies^{32,33}, which we confirm here using a single-cell resolution approach.</p> <p>In an additional counter example, we show that cell phenotypes #2 and #3 (Extended Data Fig.15a, green boxes) were positively colocalized with histoCAT and negatively colocalized with colocatome analysis based on the CLQ. Interestingly, this colocalization is displayed as segregated compartments with adjoining sections, as shown on representative breast images from the histoCAT dataset (Extended Data Fig.15c). These differences between HistoCAT and the CLQ spatial metrics may be attributed to the “connectivity” parameter in the HistoCAT tool. HistoCAT utilizes an average connectivity approach, which appears to spread connections evenly among neighboring cells. This can result in a lower count for a specific phenotype compared to the “accumulated numbers” method used in the CLQ calculation. In essence, while HistoCAT</p>

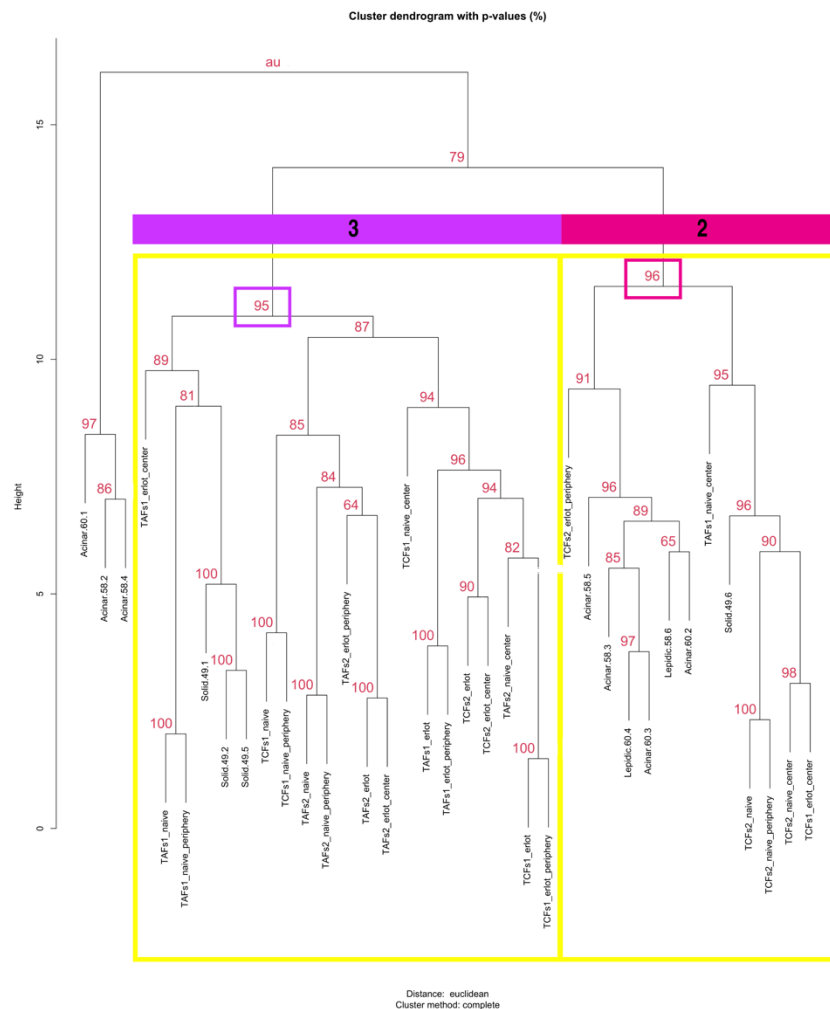
averages connections, CLQ counts each connection individually, potentially leading to variations in the reported numbers.

These examples highlight the added value of applying colcatome analysis using the CLQ alongside other tools such as histoCAT to identify new relevant biological colocalizations and architectural features. Future work that highlights the differences captured by different colocalization statistics is warranted but beyond the scope of our current study.



<p>Extended Data Fig. 15 Colocatome analysis applied to a breast cancer imaging mass cytometry dataset. a, Interactions from 49 breast tumor images and three matched normal tissues represented as a heatmap from the original histoCAT article from Schapiro et al., (left) compared to colocatome analysis using the CLQ (right). Significance was determined by permutation test ($p < 0.05$). Black boxes represent examples of colocalizations highlighted in the original manuscript, and yellow boxes represent an example of a unique finding to colocatome analysis. The histoCAT figure has been adapted from the original manuscript. Rights and permissions were obtained from Springer Nature. b, Representative examples of a breast metastasis (Cy13x5; upper panel) and breast tumor (Cy7x7; lower panel) showing colocalization between CD68+ and vim+ spindle-shaped cells on original mass cytometry images, suggesting colocalization between macrophages and fibroblasts. Left images represent a dot plot where each dot represents a cell, located at the cell centroid, and is color-coded by its cell annotation from the original histoCAT manuscript. c, Representative examples of anti-colocalizations and CLQ values.</p>	
<p>2.15. It is unclear how 20 was selected as the default parameter for determining the number of neighbors, 100 as the default Euclidean bandwidth parameter, and 500 iterations for the permutation test? The authors should demonstrate an empirical approach of arriving at such parameters.</p>	<p>We thank Reviewer #2 for highlighting these points regarding our analysis. We carefully addressed them in the Colocatome Analysis Method section at lines 610-614.</p> <p>Briefly, the number of neighbors was set to 20 after evaluating the statistically significant colocalizations obtained with spatial permutation testing and varying the number of neighbors. This metric was chosen because it represented the elbow point where the curve started to plateau, as shown in Extended Data Fig. 10a, and repeated below.</p> <p>The default Euclidean bandwidth parameter of 100 was chosen according to recommendations from the original CELESTA manuscript²⁷. Zhang et al. tested the average number of neighbors found with various bandwidth parameters. A bandwidth of 100 was selected to always ensure 20 neighboring cells in the vicinity. These results are presented at Extended Data Fig. 3 of the original CELESTA manuscript²⁷.</p> <p>Lastly, 500 permutations were selected as this number was sufficient to generate each CLQ distribution while considerably reducing the analysis processing time compared to employing 1000+ iterations.</p>
	
<p>Extended Data Fig. 10a Line graph illustrating a representative example of the number of statistically significant colocalizations identified with a spatial permutation test ($n = 500$) within the assembloids, as the number of neighbors tested in the CLQ analysis varies.</p>	
<p>2.16. In line 144, the author claims, "We expanded the capabilities of the CELESTA algorithm." However, the new innovation of expanded CELESTA is unclear. Please explain in depth.</p>	<p>CELESTA has been designed to analyze each cell individually, determining pre-defined cell type based on marker expression profiles and spatial information when required. Here, we discuss a broadening of scope rather than the introduction of new functions. We utilized CELESTA's fast and iterative nature not only to identify predefined cell types based on their expression profile and spatial information, but also to uncover new cell subpopulations by systematically testing every marker combination and assessing its prevalence. This allows us to quickly assess which cell states exist, or not, in a system or tissue. We have revised our statement accordingly (lines 170-176).</p>
<p>Reviewer #3 – Comprehensive response</p>	
<p>Reviewer #3 recognizes the potential value of our work. We thank them for their time and thorough comments. We have carefully addressed Reviewer #3's concerns regarding the repetitions of colocalizations observed across samples, the significance, and implications of the colocatome, as well as provided additional statistics to support the hierarchical clustering integrating LUAD histopathological specimens and <i>in vitro</i> assembloid colocatome data.</p>	
<p>Reviewer #3 – Major Comments</p>	
<p>Comments</p> <p>3.1. Sample size: The authors claim that tumor-stroma assembloids, generated using LUAD patient-derived cancer organoids (PDO) and fibroblasts isolated from cancers, exhibit distinct cellular composition and colocalization pattern. However, the overall conclusions may be compromised due to the limited sample size. The cell-cell colocalizations depicted in Fig. 2 cannot be considered reliable based on a single sample.</p> <p>Furthermore, in Fig. 3, the authors showed the biological replicates of the subpopulation ratio in Fig. 3b, however they did not include any repetitions of colocalizations in Fig. 3e and 3f.</p> <p>3.2. Significance: The authors argue that the utilization of colocatome and cytoarchitecture enables the classification of tumor samples and the prediction of medication response through the use of assembloid models.</p>	<p>Answers</p> <p>We appreciate Reviewer #3's comment and agree that cell-cell colocalizations based on a single sample cannot be considered as generalizable to a broader population. We want to reassure Reviewer #3 that only colocalizations consistently observed across biological replicates were considered in our study. We address these points in comment 1.5 to Reviewer 1.</p> <p>As for Fig. 3e and f, (now h and i) these panels represent repeated data as described in 1.5, and heatmaps display the average normalized CLQ values or cell fractions for each pairwise colocalization displayed. These details were clarified in the caption.</p> <p>We appreciate Reviewer #3's comments on the numerous exciting spatial biology platforms that are currently available, as well as the importance of combining different -omics dimensions (ex: protein and RNA).</p> <p>Reviewer #3 states that "colocatome has limitations in identifying biological characteristics, making it less valuable compared to other platforms." We would like to emphasize that the colocatome is not an imaging platform, but a computational approach to analyze spatial information at the single-cell resolution from any</p>

<p>However, colocatome has limitations in identifying biological characteristics, making it less valuable compared to other platforms.</p> <p>With the advancement of single cell RNA-seq and spatial genomics, numerous platforms have emerged that can analyze spatial transcriptome or proteome data with single-cell resolution and over 100 features. In this field, a new approach is needed to utilize the abundant data available, including not only spatial information but also other important variables. Similar to Xenium, CosMx, and MERSCOPE, Phenocycler can detect over 100 proteins in a single sample slide. The new versions of Phenocycler are not limited to detecting subpopulations like the previous version was when it was released alongside the development of 'CELESTA'. The new versions can also handle expression data matrix, such as scRNA-seq. Therefore, the usefulness of colocatome appears to be limited.</p>	<p>platform that allows single cell measurement. The colocatome defines the ensemble of colocalizations present in a system, similar to other -omic dimensions (Ex: RNA-transcriptome, protein-proteome, and in our case colocalization-colocatome). Here, the colocatome represents a spatial -omic dimension agnostic to a platform. We could equally establish the colocatome between different categorical variables (which don't need to be cell subpopulations) using MERSCOPE, Phenocycler, CosMx, or even metabolomics, as long as single-cell spatial information is provided. The significance and implications of the colocatome are detailed in the introduction at lines 89 to 94.</p> <p>Reviewer #3 also states that "In this field, a new approach is needed to utilize the abundant data available, including not only spatial information but also other important variables." We agree with Reviewer #3 that integrating spatial data with other types of data is important and are actively working in this area, as stated in comment 2.3.2.</p>
<p>3.3.1 Statistical significance and data analysis: The authors assert in Fig. 5 that distinct colocation patterns can be identified among various types of cancer samples or fibroblasts, and they link an <i>in vitro</i> model to patient samples. However, out of the 262 spatial features identified in assembloids, only 69 features are present in the LUAD specimens.</p>	<p>We appreciate Reviewer #3's comments and acknowledge that the assembloids exhibit numerous spatial features that are not reproduced in the LUAD samples.</p> <p>When cells are cultured <i>in vitro</i>, factors such as nutrient availability, oxygen levels, growth factors, and physical cues can influence cellular behavior and phenotypes. Therefore, the greater number of colocalizations found in the assembloids is not unexpected. We find it remarkable that 75% of the cancer-fibroblast spatial features identified in the clinical specimens are replicated in the assembloids, which supports the notion that <i>in vitro</i> assembloids can replicate most LUAD cancer-fibroblast spatial features found in LUAD specimens. If the opposite scenario were observed, wherein the in-vitro models were limited to their phenotypic properties and the clinical specimen colocalizations were not represented in the assembloids, that would be problematic, as it would suggest that the assembloids are insufficient to adequately recapitulate the clinical specimen's tumor-stroma architecture. As it stands, our downstream analyses exclusively focus on spatial features shared between the assembloids and the clinical specimens, excluding non-clinically relevant spatial information.</p> <p>With the increasing popularity of 3D patient-derived <i>in vitro</i> models in biology, our study provides the spatial biology research community with a framework to understand the reliability and limitations of these widely used <i>in vitro</i> validation platforms. More importantly, we address the important need to bridge the gap between <i>in vitro</i> observations and human histological specimens.</p>
<p>3.3.2 Furthermore, the authors claim that the solid regions of LUAD samples predominantly clustered with TAF-PDO when analyzing the colocalization patterns using hierarchical clustering. The heatmap presented in Fig. 5d does not appear to effectively differentiate between the sample groups, and the use of hierarchical clustering does not ensure their difference with statistical significance. The authors should provide statistical significance about this matter. Comparing the distinguishability of cell fractions is not meaningful because it is not even possible to categorize samples with subpopulations using scRNA-seq, which is significantly superior in cell annotation.</p>	<p>We support Reviewer #3's comment that categorizing samples solely based on cell fractions, as we demonstrated in Fig. 5e, is insufficient. Had we obtained similar clustering results using only cell fractions, similar to those in Fig. 5d based on CLQ values, we would have concluded that the colocatome, as a standalone spatial -omic, does not offer additional information. However, our findings demonstrate that the colocatome does indeed provide added value by revealing that the TAF-PDO assembloids exhibit a colocatome more similar to the colocatome of solid regions, whereas TCF-PDO assembloids better recapitulate the colocatome of lepidic or acinar-like regions.</p> <p>To further support results presented in Fig. 5d, we repeated hierarchical clustering using the pvclust R package³⁴, a tool for assessing the uncertainty in hierarchical cluster analysis using multiscale bootstrap resampling. Briefly, we repeated the clustering on all the significant spatial features shared between assembloids and clinical specimens. Pairwise colocalizations with more than 20% missing values across all conditions and self-colocalizations were excluded. The "complete" clustering method, Euclidean distance, and 10,000 iterations for bootstrap resampling were used as input parameters. As recommended by the authors of the pvclust R package, we used the approximate unbiased (AU) p-values (red values on the dendrogram), and a p-value of 0.05 was considered significant. The significance of the clusters (p values) was added to Fig. 5d and the dendrogram with annotations is included in Extended Data Fig 13 and repeated below.</p>



Extended Data Fig. 13 | Uncertainty calculation in hierarchical cluster analysis using pvclust. Cluster dendrogram of LUAD tumor-stroma assembloids and histopathological specimens. 10,000 iterations were used for bootstrap resampling. The approximate unbiased (AU) p-values were used as recommended in the user guide. P-values (red) of a cluster is a value between 0 and 100, which indicates how strong the cluster is supported by data in percentage. P-values of 0.05 was considered significant corresponding to AU values ≥ 95 on the dendrogram. Yellow boxes delimitate the clusters described in Fig. 5d.

Reviewer #3 – Minor comments	
Comments	Answers
3.4. The authors should carefully check the misspellings.	The manuscript was reviewed for typos.
3.5. Scale bars should be added to Fig. 2c, 2d, 3g, Extended Fig. 2-5, and 7b.	Scale bars were added as an approximation of the size of each assembloid. Note that the dots do not represent a true measure of each cell but only a point designated in the centroid.
3.6. In Fig. 3a, the mean or median line should be with violin plots.	The lines were edited to be more apparent and described in the figure legend.

REFERENCES

- Kilvaer TK, Khanekhenari MR, Hellevik T, et al. Cancer associated fibroblasts in stage I-III NSCLC: Prognostic impact and their correlations with tumor molecular markers. *PLoS One*. 2015;10(8):1-15. doi:10.1371/journal.pone.0134965
- Fitzgerald AA, Weiner LM. *The Role of Fibroblast Activation Protein in Health and Malignancy*. Vol 39; 2020. doi:10.1007/s10555-020-09909-3
- Kilvaer TK, Rakaee M, Hellevik T, et al. Differential prognostic impact of platelet-derived growth factor receptor expression in NSCLC. *Sci Rep*. 2019;9(1):1-11. doi:10.1038/s41598-019-46510-3
- Hewitt KJ, Shamis Y, Knight E, et al. PDGFR β expression and function in fibroblasts derived from pluripotent cells is linked to DNA demethylation. *J Cell Sci*. 2012;125(9):2276-2287. doi:10.1242/jcs.099192
- Schliekelman MJ, Creighton CJ, Baird BN, et al. Thy-1+ Cancer-associated Fibroblasts Adversely Impact Lung Cancer Prognosis. *Sci Rep*. 2017;7(1):1-12. doi:10.1038/s41598-017-06922-5
- Shi YB, Li J, Lai XN, Jiang R, Zhao RC, Xiong LX. Multifaceted roles of caveolin-1 in lung cancer: A new investigation focused on tumor occurrence, development and therapy. *Cancers (Basel)*. 2020;12(2):1-19. doi:10.3390/cancers12020291
- Hu G, Zhong K, Chen W, Wang S, Huang L. Podoplanin-positive cancer-associated fibroblasts predict poor prognosis in lung cancer patients. *Onco Targets Ther*. 2018;11:5607-5619. doi:10.2147/OTT.S175566
- Ye Z, Zhang X, Luo Y, et al. Prognostic values of vimentin expression and its clinicopathological significance in non-small cell lung cancer: A meta-analysis of observational studies with 4118 cases. *PLoS One*. 2016;11(9):1-17. doi:10.1371/journal.pone.0163162

9. Luo HT, Liang CX, Luo RC, Gu WG. Identification of relevant prognostic values of cytokeratin 20 and cytokeratin 7 expressions in lung cancer. *Biosci Rep*. 2017;37(6):1-12. doi:10.1042/BSR20171086
10. Gharib TG, Chen G, Wang H, et al. Proteomic analysis of cytokeratin isoforms uncovers association with survival in lung adenocarcinoma. *Neoplasia*. 2002;4(5):440-448. doi:10.1038/sj.neo.7900257
11. Huang X, Sun Q, Chen C, et al. MUC1 overexpression predicts worse survival in patients with non-small cell lung cancer: Evidence from an updated meta-analysis. *Oncotarget*. 2017;8(52):90315-90326. doi:10.18632/oncotarget.19861
12. Lan Y, Ni W, Tai G. Expression of MUC1 in different tumours and its clinical significance (Review). *Mol Clin Oncol*. 2022;17(6):1-10. doi:10.3892/mco.2022.2594
13. Xu X, Chen W, Leng S, et al. Muc1 knockout potentiates murine lung carcinogenesis involving an epiregulin-mediated EGFR activation feedback loop. *Carcinogenesis*. 2017;38(6):604-614. doi:10.1093/carcin/bgx039
14. Jin W, Liao X, Lv Y, et al. Muc1 induces acquired chemoresistance by upregulating abcb1 in egfr-dependent manner. *Cell Death Dis*. 2017;8(8):e2980-13. doi:10.1038/cddis.2017.378
15. Baeuerle PA, Gires O. EpCAM (CD326) finding its role in cancer. *Br J Cancer*. 2007;96(3):417-423. doi:10.1038/sj.bjc.6603494
16. Kim Y, Hyo SK, Zheng YC, et al. Clinicopathological implications of EpCAM expression in adenocarcinoma of the lung. *Anticancer Res*. 2009;29(5):1817-1822.
17. Pak MG, Shin DH, Lee CH, Lee MK. Significance of EpCAM and TROP2 expression in non-small cell lung cancer. *World J Surg Oncol*. 2012;10:1-8. doi:10.1186/1477-7819-10-53
18. Tian L, Tibshirani R. Adaptive index models for marker-based risk stratification. *Biostatistics*. 2011;12(1):68-86. doi:10.1093/biostatistics/kxq047
19. Tian L, Liu JS, Wei LJ. Implementation of estimating function-based inference procedures with markov chain Monte Carlo samplers. *J Am Stat Assoc*. 2007;102(479):881-888. doi:10.1198/016214506000000122
20. Gronsbell J, Hong C, Nie L, Lu Y, Tian L. Exact inference for the random-effect model for meta-analyses with rare events. *Stat Med*. 2020;39(3):252-264. doi:10.1159/000444169
21. Parast L, Cai T, Tian L. Testing for heterogeneity in the utility of a surrogate marker. *Biometrics*. 2023;79(2):799-810. doi:10.1111/biom.13600
22. Zhao XG, Dai W, Li Y, Tian L. AUC-based biomarker ensemble with an application on gene scores predicting low bone mineral density. *Bioinformatics*. 2011;27(21):3050-3055. doi:10.1093/bioinformatics/btr516
23. Huang B, Tian L. Utilizing restricted mean duration of response for efficacy evaluation of cancer treatments. *Pharm Stat*. 2022;21(5):865-878. doi:10.1002/pst.2198
24. Tian LU, Cai T, Goetghebuer E, Wei LJ. Model evaluation based on the sampling distribution of estimated absolute prediction error. *Biometrika*. 2007;94(2):297-311. doi:10.1093/biomet/asmo36
25. Bouchard G, Garcia-Marques FJ, Karacosta LG, et al. Multi-omics analysis of spatially distinct stromal cells reveals tumor-induced O-glycosylation of the CDK4-pRB axis in fibroblasts at the invasive tumor edge. *Cancer Res*. 2021;82(4):648-664. doi:10.1158/0008-5472.CAN-21-1705
26. Skog MS, Nystedt J, Korhonen M, et al. Expression of neural cell adhesion molecule and polysialic acid in human bone marrow-derived mesenchymal stromal cells. *Stem Cell Res Ther*. 2016;7(1):1-12. doi:10.1186/s13287-016-0373-5
27. Zhang W, Li I, Reticker-flynn NE, et al. Identification of cell types in multiplexed in situ images by combining protein expression and spatial information using CELESTA. *Nat Methods*. 2022;19(7):759-769.
28. Karacosta LG, Anchang B, Ignatiadis N, et al. Mapping Lung Cancer Epithelial-Mesenchymal Transition States and Trajectories with Single-Cell Resolution. *Nat Commun*. 2019;10:5587.
29. Schapiro D, Jackson HW, Raghuraman S, et al. HistoCAT: Analysis of cell phenotypes and interactions in multiplex image cytometry data. *Nat Methods*. 2017;14(9):873-876. doi:10.1038/nmeth.4391
30. Zhou X, Franklin RA, Adler M, et al. Circuit Design Features of a Stable Two-Cell System. *Cell*. 2018;172(4):744-757.e17. doi:10.1016/j.cell.2018.01.015
31. Buechler MB, Fu W, Turley SJ. Fibroblast-macrophage reciprocal interactions in health, fibrosis, and cancer. *Immunity*. 2021;54(5):903-915. doi:10.1016/j.immuni.2021.04.021
32. Angarola BL, Sharma S, Katiyar N, et al. Comprehensive single cell aging atlas of mammary tissues reveals shared epigenomic and transcriptomic signatures of aging and cancer. *bioRxiv Prepr Serv Biol*. Published online 2023.
33. Ben-chetrit N, Niu X, Sotelo J, et al. Breast Cancer Macrophage Heterogeneity and Self-renewal are Determined by Spatial Localization. Published online 2023.
34. Suzuki R, Shimodaira H. Pvcust: An R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics*. 2006;22(12):1540-1542. doi:10.1093/bioinformatics/btl117

September 2nd, 2024

Subject: Request for Reconsideration to Address Reviewers' Suggestions and Resubmit

Re: NCOMMS-23-58305-T

We thank the reviewers for their time and effort in reviewing our revised manuscript entitled “A spatial-omic approach for quantitative analysis of tumor-stroma assembloids and architectural mapping onto pathological specimens using the colocalome.” Based on the feedback, we have now entitled our manuscript as “**A quantitative spatial cell-cell colocalizations framework enabling comparisons between in vitro assembloids and pathological specimens**”. We appreciate the reviewer’s detailed evaluation of our extensive revisions and are pleased that all three reviewers found our efforts to be “very clear,” “substantial,” and “commendable.” The additional feedback and constructive criticism were invaluable in guiding our final revisions. We have addressed the remaining five comments by revising relevant statements, incorporating the recommended discussion points, and providing the last piece of requested data (e.g., evidence that co-location quotients (CLQ) in given samples form a Gaussian-shaped distribution independent of region size). Additionally, we collaborated closely with a scientific writer to polish our manuscript and ensure clear and concise message delivery. Below is a point-by-point response addressing each comment, with corresponding changes highlighted in red in the main manuscript:

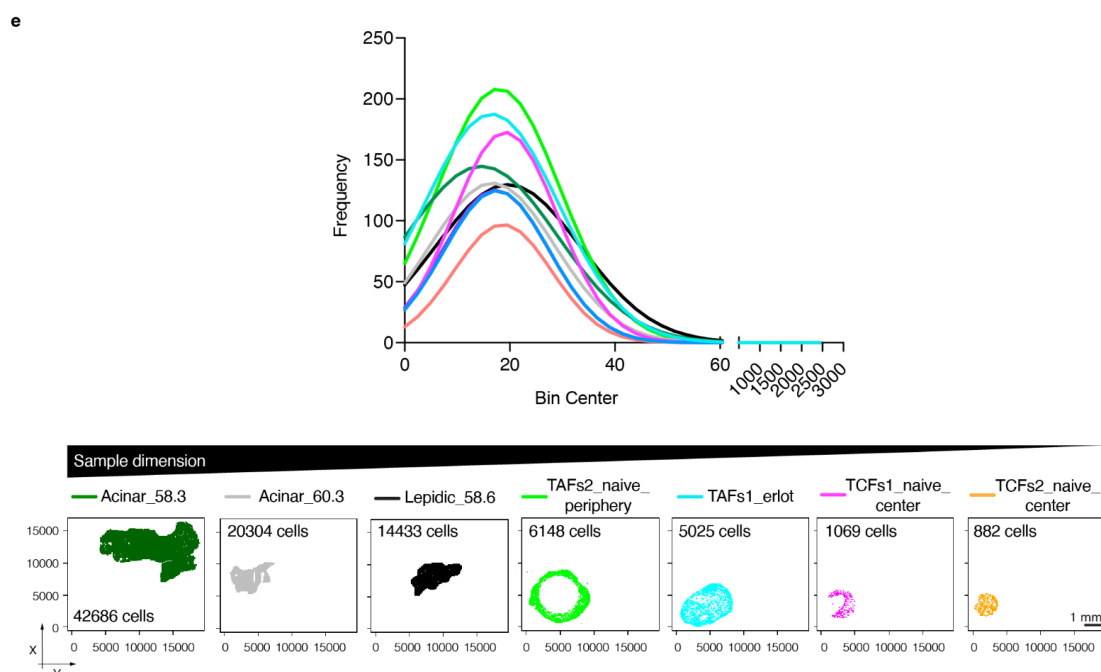
Reviewer #1 – Comprehensive response	
We thank Reviewer #1 for their time and effort in reviewing our revised manuscript. We appreciate their acknowledgment of our efforts in addressing their original comments and are pleased that our responses were clear. We have revised the statement highlighted by Reviewer #1 and added relevant discussion points accordingly. Additionally, a scientific writer has proofread the manuscript for grammar, typographical errors, and concision.	
Comments	Answers
<p>1. I accept the points made by the authors regarding challenges associated with reliably identifying fibroblasts in human tissue sections and how this can present a significant obstacle to performing their discovery analysis; and in addition, providing the opportunity to examine later stage phenotypes where clinical samples are difficult (or not possible) to access. However, the statement in lines 326-328 (copied below) of the revised version should be re-worded to not over-state the pros of this approach and accurately convey the compromise/limitations associated with it being founded on a potentially non-physiological system.</p> <p>“To overcome these difficulties, we leveraged the cancer cell and fibroblast subpopulations identified in the assembloids as a guide to phenotype the human samples which we demonstrated is a more robust approach compared to identifying subpopulations directly from the clinical specimen.”</p> <p>How is it demonstrated that this approach is more robust, and what is meant by more robust in this context?</p> <p>I don’t think using this method “overcome[s] these difficulties” it simply enables a comparative analysis to be performed between in vitro and human tissues. In a different context, where the cell types under investigation are identifiable in human tissues (e.g. for well defined cell populations or with the development of more accurate phenotyping and segmentation tools) this would not be a more robust approach. This caveat</p>	<p>We thank Reviewer #1 for their suggestion and the referred statement now reads “To address these difficulties, we leveraged the designated subpopulations identified in assembloids as a guide to phenotype human samples, as opposed to identifying cell subpopulations directly from clinical specimens. This approach was found to be robust for highly heterogeneous cell types, such as fibroblasts and cancer cells, which do not always express specific, distinctive markers” (see lines 321-325).</p> <p>Here, by “a more robust approach”, we refer to using predefined subpopulations from assembloids as a guide, which can be more reliable for identifying poorly defined cell subpopulations in human samples. This was demonstrated by comparing the identification of cell subpopulations starting with the assembloids vs. starting with the clinical specimens. Using the assembloids as a starting point led to the identification of more subpopulations in the clinical specimens given the increased confidence of identifying diverse cell states in the assembloids. This experiment, as well as the meaning of robustness in this context have been emphasized in the Supplementary Results at lines 7 to 11.</p> <p>We agree with Reviewer #1 that in some instances, using in vitro patient-derived models for phenotyping may not be a more robust approach (e.g., well-defined cell populations such as immune cells). Therefore, we modified the statement referred to by Reviewer #1 to specify that this approach was found to be robust for highly heterogeneous cell types, such as fibroblasts and cancer cells, which do not always express specific,</p>

<p>should be reflected in the statement above and perhaps in the discussion.</p> <p>It should be noted that the fibroblast markers selected will likely fail to differentiate between mural cells and CAFs (as ACTA2, CD90 and PDGFRB are all likely to be expressed by both cell types). Evidently the assembloid analyses are not affected by this but it should be noted that this is a limitation to how comprehensively the CAF populations in human tissue sections can be delineated with this approach and the conclusions drawn from analysing the human tumour sections in the manner described.</p>	<p>distinctive markers (lines 323 to 325) and added a statement in the Discussion at lines 462 to 465.</p> <p>Lastly, Reviewer #1 highlighted an important point regarding the fact that the number of fibroblast markers used in this study may not be sufficient to distinguish fibroblasts from pericytes and vascular smooth muscle cells. Therefore, we added a statement in the discussion at lines 467 to 470 noting that the conclusions drawn from analyzing the human tumor sections should be further validated and that we cannot exclude the possibility that some CAF subpopulations identified may be mural cells.</p>
<p>1.2. There are also a number of grammatical errors that need to be addressed with further proof reading.</p>	<p>We apologize for the grammatical errors and have worked closely with a scientific writer to proofread the manuscript and improve the message delivery.</p>
<p>Reviewer #2 – Comprehensive response</p> <p>We appreciate Reviewer #2 for recognizing the substantial effort we invested in revising this study. We also thank Reviewer #2 for pointing out key terms in our study that were not clearly defined. In response, we revised the definition of the colocatome and clarified its use as a spatial -omic. Additionally, we provided further data to substantiate that CLQs in given samples form a Gaussian-shaped distribution independent of region size, and we integrated the new discussion points suggested. Lastly, we worked closely with a scientific writer to clarify the definitions, the novelty of our approach, and the overall message, which we believe has significantly strengthened our manuscript and supports its suitability for publication."</p>	
<p>Comments</p> <p>1. Re 2.2, on the Feasibility of "colocatome" which is essentially a pairwise cell-cell interaction that has been demonstrated in multiple prior publications. We continue to be confused by the definition of "colocatome", which is now presented by the authors as a new "omics feature", which again is essentially a cell-cell pairwise permutation test, and seems inappropriately termed as "omics". This reviewer continues to be hesitant in this exaggerated definition and novelty of a cell-cell interaction test.</p>	<p>Answers</p> <p>We appreciate Reviewer #2's highlighting this important point. We would like to emphasize that the colocatome itself is a comprehensive catalog of all cell-cell colocalizations. Reviewer #2 is absolutely correct in stating that the colocatome framework includes statistically significant pairwise cell-cell colocalizations, which has been demonstrated in prior publications that we have cited¹⁻³. In this study, we conducted colocatome analysis based on the CLQ, but other cell-cell colocalization spatial metrics could be employed depending on their properties. Because we found that the CLQs for a sample form a bell-shaped distribution, we implemented a z-score-based normalization step allowing for direct comparisons of CLQs between in vitro models and patient specimens. By performing this normalization, we propose the colocatome as a spatial omic that is generalizable across conditions and sample types and an important step toward an ultimate goal of cataloguing cell-cell localizations for comparison across different studies.</p> <p>We clarified the definition of the colocatome and the purpose of the analytical framework throughout the manuscript (see Introduction at lines 55 to 58, and lines 68 to 70, and Results at lines 177 to 198.</p>

2. Re 2.4.1, on the physiological relevance of assembloids and downstream analysis. The authors claim that CLQs in given samples form a Gaussian distribution independent of region size. This is an important point that should be substantiated with data in the manuscript.

We thank Reviewer #2 for their thorough review of our manuscript and highlighting this key feature of the CLQ. To address this important point, we have added a new figure that illustrates representative examples of CLQ distributions across various samples of different shapes and sizes, demonstrating that the CLQs in these samples form a Gaussian-shaped distribution, independent of region size (see [Methods lines 598 to 599](#) and [Extended Data Fig. 12e](#)).

Briefly, the CLQ distributions were calculated with the function “Frequency distribution” in GraphPad Prism Version 10.2.3 with a bin size of 20. The distributions were then analyzed using the “Gaussian” model parameter. As shown in the left panel, the original CLQs form a Gaussian-shaped distribution independent of the sample size, with sample sizes ranging from 882 to 42686 cells.



Extended Data Fig. 12e | The colocalome in LUAD histopathological specimens. e, CLQ distributions for LUAD and assembloid regions of different sizes and shapes. Frequency distributions were plotted, and a non-linear curve fit with the Gaussian parameter was calculated in GraphPad. Bins of 20 were used.

We also further point out that the approach here favors more prevalent cell subpopulations, leading to a narrower null CLQ distribution and potentially overstating their significance due to less variability in spatial permutations. In contrast, rare subpopulations will reflect a broader null distribution, which increases the likelihood of their observed CLQ values being deemed non-significant. This discrepancy suggests a reduced sensitivity in detecting significant interactions among rarer subpopulations. The author should discuss the sensitivity issue in the discussion section.

We agree with Reviewer #2 about this important point on the permutation analysis, which we explain in the [Method](#) section at [lines 590 to 592](#). In addition, a statement addressing sensitivity issues for rarer cell subpopulations has been included in the [Discussion](#) section, as suggested by Reviewer #2 (see [lines 455 to 461](#)).

<p>Re 2.14 on the term “spatial-omics” and how the authors only demonstrate spatial proteomics applicability. The authors did not address our concerns, and further highlight the addition of TNBC IMC data. The authors should either include other “-omics” application, or tone down the claims in this study.</p> <p>The addition of the “validation” dataset here also raises more questions on the differences between HistoCAT and CLQ results, which should warrant a further discussion given the context of prior studies, and claims of superiority of CLQ.</p>	<p>In their prior review, Reviewer #2 suggested “considering other modalities to demonstrate the applicability of CLQ/colocator, aside from CODEX spatial proteomics”, as a “minor concern”. In response, we provided an extensive new analysis using an external imaging mass cytometry breast dataset. We realize that Reviewer #2 was referring to an alternative -omic data type, such as single-cell transcriptomics, rather than a different modality besides CODEX spatial proteomics.</p> <p>We apologize for the confusion and consequently have toned down our claims regarding other data -omics applications, as recommended by Reviewer #2 (see lines 414 to 415).</p> <p>In addition, we clarified the concept of using the colocator as a spatial -omic, describing it as a spatial readout that can be expanded to work towards the common goal of building a catalog and enabling comparisons across studies. This approach aims to advance the field of spatial biology with sets of features that can be compared across studies. Similar to how gene sets can be associated with biological mechanisms or diseases, it is possible that spatial features could also be linked to biological mechanisms and outcomes (see Results, lines 198 to 199, and Conclusion, lines 474 to 479).</p> <p>Lastly, we agree with Reviewer #2 that it is crucial to understand the main differences between spatial metrics used in various cell-cell interaction or cell-cell colocalization studies, particularly with the emergence of various spatial biology platforms. We recognize the value of all colocalization analysis methods and would like to emphasize that we do not claim that CLQ is superior to the analytical framework presented in the HistoCAT toolbox³. Instead, we emphasize the added value of the CLQ over analyses based on cell composition only (see lines 370 to 372). We also emphasize certain differences between CLQ and HistoCAT, which we suggest may be partly attributed to the "connectivity" parameter in the HistoCAT tool, which is computed differently than for CLQ analysis.</p> <p>While we agree that further work evaluating different spatial metrics and their performance will be crucial to explore in the future, which we highlight in the Discussion at lines 470 to 472, it is beyond the scope of this work.</p>
Reviewer #3 – Comprehensive response	
<p>We would like to thank Reviewer #3 for their time and valuable feedback on our revised manuscript. We greatly appreciate the recognition of the efforts made to address the reviewers' comments and the potential of the colocator as a spatial -omic. We hope that the emphasis on the 13 LUAD histological regions and the breast cohort of 52 samples will address the concerns on sample size raised by Reviewer#3, which were initially shared across all reviewers, and substantiate the potential of our approach.</p>	
Comments	Answers

The authors have made commendable efforts to address the reviewers' comments by providing additional analyses and explanations. However, a major concern regarding sample size persists: the authors have used only one PDO sample for all experiments and only two biological replicates for each assembloid condition. These limitations raise concerns about the robustness and reproducibility of their colocatome utility, despite its potential as an intriguing characteristic in the context of spatial omics data.

We appreciate Reviewer #3's comment regarding the sample size for the assembloid experiments and agree that experimental findings should be robust and reproducible. We would like to emphasize that we view the "colocatome" as comprehensive catalog of all cell-cell colocalizations that can be compared across conditions. The innovation of our work lies in the quantitative framework which combines cell-cell colocalization quantification, spatial permutation and a normalization process, enabling us to catalog the colocatome and perform cross-sample comparisons. This approach facilitates direct comparison between in vitro models and patient specimens, ensuring robust and meaningful results.

In addition to the assembloids used in this study, we performed colocatome analyses on 13 distinct LUAD regions and 52 breast samples from external validation datasets. We demonstrated the robustness and reproducibility of our analytical framework through these comparisons. For instance, the colocatome derived from the assembloids was either cross-referenced with literature (ex: [lines 422 to 423](#)) or validated using the human LUAD dataset cited above (ex: [Fig. 5](#) exclusively focuses on shared spatial features between the assembloids and LUAD specimens). As for hypotheses generated from the in vitro samples that were not corroborated by clinical specimens, we indicated in the [Discussion](#) that these hypotheses warrant further study in larger cohorts (see [lines 448-449](#)). Lastly, we compared the colocatome resulting from our analytical framework vs the HistoCAT toolbox³ on the 52 breast samples mentioned above. Therefore, our study encompasses more than 70 samples in total, and our results were reproducible across datasets which we hope will alleviate Reviewer #3's concerns about sample size, robustness and reproducibility.