nature communications

Article

https://doi.org/10.1038/s41467-024-55129-6

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

Received: 30 November 2023

Accepted: 29 November 2024

Gina Bouchard[1], Weiruo Zhang[1], Ilayda Ilerten[1], Irene Li[1], Asmita Bhattacharya[2], Yuanyuan Li[1], Winston Trope[3], Joseph B. Shrager[3,4], Calvin Kuo[2], Michael G. Ozawa[5], Amato J. Giaccia[6,7], Lu Tian[1] & Sylvia K. Plevritis[1,8]

Spatial omics is enabling unprecedented tissue characterization, but the ability to adequately compare spatial features across samples under different conditions is lacking. We propose a quantitative framework that catalogs significant, normalized, colocalizations between pairs of cell subpopulations, enabling comparisons among a variety of biological samples. We perform cell-pair colocalization analysis on multiplexed immunofluorescence images of assembloids constructed with lung adenocarcinoma (LUAD) organoids and cancer-associated fibroblasts derived from human tumors. Our data show that assembloids recapitulate human LUAD tumor-stroma spatial organization, justifying their use as a tool for investigating the spatial biology of human disease. Intriguingly, drug-perturbation studies identify drug-induced spatial rearrangements that also appear in treatment-naïve human tumor samples, suggesting potential directions for characterizing spatial (re)-organization related to drug resistance. Moreover, our work provides an opportunity to quantify spatial data across different samples, with the common goal of building catalogs of spatial features associated with disease processes and drug response.

The emergence of highly multiplexed spatial omics technologies is advancing translational oncology research by enabling a more comprehensive understanding of the cellular organization within the tumor microenvironment (TME)[1]. Recent spatial biology technologies such as multiplexed immunofluorescence[2], imaging mass cytometry[3] and spatial transcriptomics[4] are revolutionizing our ability to interrogate the TME in solid tumors. While these diverse -omic technologies can capture spatially resolved molecular information, often at single cell resolution, analysis and comparison of spatial data across studies poses several challenges[5]. A standard spatial analysis pipeline for images at single cell resolution includes preprocessing raw imaging data, cell segmentation, cell type identification, and extraction of

---

[1]Department of Biomedical Data Science, Stanford University, Stanford, CA, USA. [2]Division of Hematology, Department of Medicine, Stanford University, Stanford, CA, USA. [3]Department of Cardiothoracic Surgery, Stanford University, Stanford, CA, USA. [4]Veterans Affairs Palo Alto Health Care System, Palo Alto, CA, USA. [5]Department of Pathology, Stanford University, Stanford, CA, USA. [6]Department of Radiation Oncology, Stanford University, Stanford, CA, USA. [7]Department of Oncology, University of Oxford, Oxford, UK. [8]Department of Radiology, Stanford University, Stanford, CA, USA.
e-mail: sylvia.plevritis@stanford.edu

statistically significant spatial features. Approaches such as spatial permutation are particularly important for evaluating the significance of spatial features[6,7]. As for identifying cell subpopulations, a common approach involves clustering cells with similar expression patterns. However, assigning cell clusters is subjective, time-consuming, and often requires manual assessment. More recent cell-identification methods that do not involve clustering are now available to overcome these limitations and are providing faster and more automated tools that can also be used to discover cell types and states[7,8]. With these tools, several studies have demonstrated the utility of cell-cell colocalization ensembles for quantifying spatial features within specific tissues and diseases[7,9,10].

Translational researchers are increasingly relying on patient-derived co-cultures as high-throughput experimental platforms to model human diseases[11,12]. Examples include organoids, tumor spheroids, assembloids, and various other models involving cell lines or primary cells derived from patients. While these in vitro models do not perfectly replicate human tissue architecture, cell heterogeneity, or the effects of environmental stimuli, their usage continues to rise due to their accessibility, ease of gene editing, cost advantages, and advocacy for alternatives to animal testing in research. For these reasons, a standardized, quantitative way to compare and analyze spatial data between in vitro patient-derived models and human tumor samples is an important step toward gaining clinically relevant information from in vitro models.

In this work, we build on these concepts and propose a quantitative framework, termed colocatome analysis, for comparing spatial features across samples, conditions, and studies. We demonstrate how the colocatome framework allows for the direct comparison of cell-cell colocalizations between three-dimensional in vitro models and patient specimens, enabling the identification of conserved spatial features between different types of samples. Through our colocatome framework, we combine pairwise cell-cell colocalization, spatial permutation, and normalization approaches, as a quantitative analysis for assessing spatial features that addresses variation across conditions. Our quantitative framework uses the colocation quotient (CLQ) spatial metric, which we introduced in prior work for identifying cell subpopulation pairs in close proximity (positive colocalization) versus those that are distant (negative colocalization)[13]. We apply spatial randomization to assess the significance of each colocalization compared to a null distribution established by permuting the cell types. Lastly, we apply normalization to the CLQs under a given condition to enable comparisons of statistically significant colocalizations to each other and across different conditions, including between in vitro assays and clinical samples.

We apply colocatome analysis to tumor-stroma assembloids generated with lung adenocarcinoma (LUAD) epithelial organoids and cancer-associated fibroblasts (CAFs) from spatially distinct tumor sites (edge vs. core) and correlate these spatial features with specific LUAD histopathological growth patterns in clinical samples. Further, we interrogate spatial rearrangements related to drug resistance in both assembloids and treatment-naïve human tumor samples. Using the colocatome framework, we identify which drug-resistant and drug-sensitive cell-pairs colocalizations from the in-vitro model are present in treatment-naïve clinical samples. In summary, our quantitative framework enables direct comparison of specimens across various conditions or assays with the potential to enable the spatial biology community to advance toward a common goal of cataloging and comparing cell-pairwise spatial features.

Results

Regionally distinct CAFs exhibit unique spatial organization in tumor–stroma assembloid models

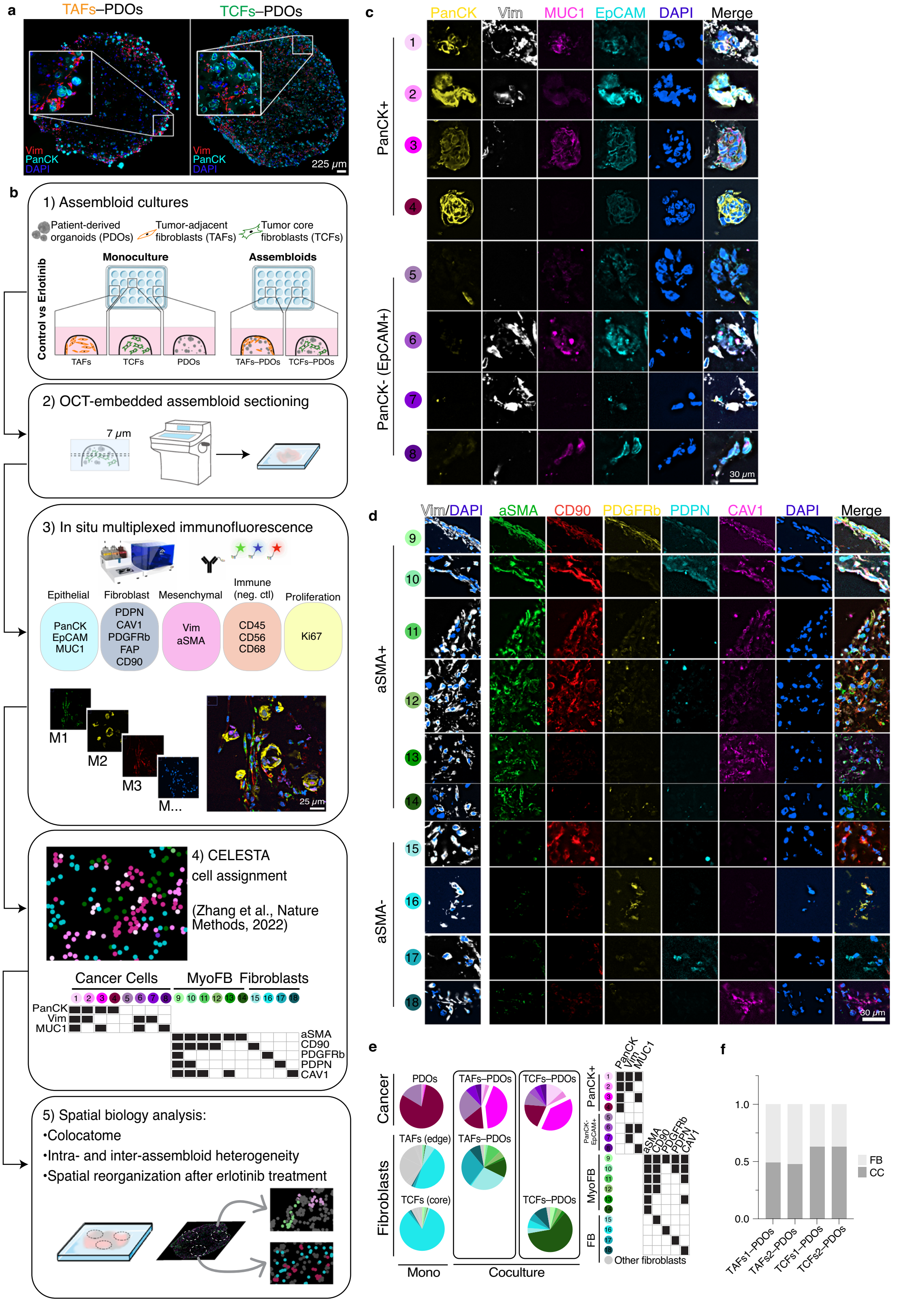
To study spatial organization between cancer cells and fibroblasts, we primarily established CAF–patient-derived organoid (PDO) assembloids. We investigated whether regionally distinct CAFs originating from a tumor leading edge vs. the tumor core influence cell heterogeneity and organization. In previous work, we observed that regionally distinct CAFs from the tumor edge, namely tumor-adjacent fibroblasts (TAFs), compared to tumor core fibroblasts (TCFs) promoted striking morphological and transcriptomic differences when cocultured with cancer cells, as well as demonstrated the increased promigratory role of TAFs vs. TCFs[8]. Here, we tested the hypothesis that regionally distinct CAFs uniquely influence tumor–stroma spatial organization using our quantitative spatial framework with LUAD assembloids. These assembloids were generated from single-cell suspensions of two cell types (cancer cells and fibroblasts), incubated for up to 10 days, embedded, sectioned, and stained with antibodies, as if they were pathological specimens. Immunofluorescence (IF) results showed that both assembloid types displayed inter-assembloid and intra-assembloid spatial heterogeneity (TAFs vs. TCFs and periphery vs. center, respectively). TAF-PDO assembloids were enriched in the periphery of the matrix dome and were less dense in the center (Fig. 1a). In contrast, TCF-PDO assembloids were more homogeneously organized throughout the matrix dome. Our results suggest that fibroblasts affect the spatial organization of cancer cells depending on their regional origin within the tumor. We quantitatively assessed these characteristics as described in the study workflow (Fig. 1b).

TAF–PDO and TCF–PDO assembloids reveal distinct fibroblast subpopulations but similar cancer cell subpopulations

To conduct a more comprehensive quantitative assessment of assembloid spatial organization, we used multiplexed immunofluorescence PhenoCycler technology (Akoya Biosciences). We acquired images using a panel of 15 markers (characterizing mainly epithelial and fibroblast), to which we applied the computational framework described below to resolve tumor-stroma spatial organization. Fibroblast and cancer cell markers were selected based on their known associations with lung cancer (Table 1). Fibrocytes[14,15] and negative control markers were also included (CD45: immune; CD68: pan-macrophage; CD56: bone marrow-derived mesenchymal stromal cell).

Following cell segmentation, clustering cells with similar expression patterns is a common approach for identifying cell subpopulations. However, this approach is subjective, time-consuming, and requires manual assessment. Moreover, some clusters of mixed-cell subpopulations can be difficult to annotate. To overcome these challenges, we expanded the scope of our CELESTA algorithm to identify cell states. CELESTA is a semi-supervised machine learning tool that does not involve manual gating or clustering, but which uses prior (even partial) knowledge of cell expression profile[7]. Leveraging CELESTA's fast and iterative nature and its customable thresholds, we identified pre-defined cell types and cell states (e.g., subpopulations) based on their expression profiles and also generated hypotheses by systematically testing every marker combination (Supplementary Fig. 1a). In this manner, CELESTA allows us to identify cell types and subpopulations in a fast and automated way.

Identifying precise subpopulations in complex tissues of unknown cell composition is difficult, even with the assistance of an expert pathologist. Therefore, we used pre-defined fibroblasts and organoid cultures of known cell ratios to construct tumor-stroma assembloids, with the ultimate goal of transferring phenotype information from assembloid analysis as a guide to phenotype complex tissues. In total, we identified 18 cell subpopulations in the assembloids (Fig. 1c–f and Supplementary Data 1): 8 cancer cells and 10 fibroblasts, with additional fibroblast subpopulations specific to monocultures (Supplementary Fig. 1b). Cell subpopulations were segregated between PanCK+ and PanCK- (cancer cells), αSMA + (myofibroblasts) and αSMA- (fibroblasts), and then validated on the original images (Fig. 1c, d). Cancer cells displayed broad morphological differences,



ranging from adherent to mesenchymal cell bundles, and myofibroblasts appeared more elongated than αSMA- fibroblasts. Note that the assignment of PanCK- cells to cancer cells were confirmed by their EpCAM+ status. We found that PDO three-dimensional monocultures largely consisted of PanCK + /Vim-/MUC1- cancer cells (#4), whereas in the presence of TAFs and TCFs, the dominant cancer cell subpopulation was PanCK + /Vim-/MUC1+ (#3). We observed that cancer cells transformed fibroblasts into myofibroblasts (which were mainly absent from monocultures). Specifically, TAF–PDO assembloids were enriched for PDPN+ fibroblasts (#17), and TCF–PDO assembloids mainly contained aSMA+ fibroblasts (#14) (Fig. 1e). Interestingly, assembloids exhibited similar cancer cell heterogeneity. The initial one-to-one cancer cell:fibroblast ratio was mostly conserved in both assembloid models (TAF–PDO and TCF-PDO); neither TAFs nor TCFs

Fig. 1 | Cell composition of TAF–PDO and TCF–PDO LUAD assembloids. a IF images with two markers (PanCK, Vim) and one nuclei stain (DAPI) of OCT sections of TAF–PDO (left) and TCF–PDO (right) LUAD assembloids. Results were repeated in two independent experiments using biological replicates ($n = 2$) for each assembloid condition. **b** Study workflow and mIF analysis pipeline. Figure adapted from Servier Medical Art licensed under a Creative Commons Attribution license CC BY 4.0 and Phenocycler image used with permission from Akoya Biosciences, Marlborough, MA. **c** Representative PhenoCycler images of assembloids showing cancer cell and **d** fibroblast subpopulations identified with CELESTA[7]. PanCK- cancer cells (#5-8) were assigned as cancer cells based on their EpCAM+ status. Scale bar, 30 μm. **e** Representative example of subpopulation proportions in TAF–PDO and TCF–PDO LUAD assembloids. Additional biological replicate pie charts visualization can be found in Supplementary data Fig. 1b. **f** Stacked bar graph representing fibroblast and cancer cell ratios in TAF vs TCF–PDO assembloids from two independent experiment using biological replicates ($n = 2$). LUAD lung adenocarcinoma, CC cancer cell, FB fibroblast. OCT optimum cutting temperature, Vim vimentin, PanCK pan-cytokeratin, EpCAM epithelial cell adhesion molecule, MUC1 mucin 1, PDPN podoplanin, CAV1 caveolin 1, PDGFRb platelet derived growth factor receptor beta, FAP fibroblast activation protein, CD cluster of differentiation, aSMA smooth muscle actin alpha 2. Source data are provided as a Source Data file.

significantly induced cancer cell proliferation (Fig. 1f). Next, we quantified whether certain cell subpopulations were enriched in either the periphery or center of assembloids. TAFs subpopulations #9-11 (expressing three or more fibroblast markers) were enriched at the periphery, and subpopulations #14-18 (expressing only one marker) were enriched at the center. In contrast, TCFs and cancer cell subpopulations were not differentially enriched when comparing center to peripheral zones (Supplementary Figs. 1c, d, 2–5).

Quantitative spatial framework analysis demonstrates that TAF–PDO and TCF–PDO assembloids exhibit distinct tumor-stroma spatial organization

To quantitatively evaluate and compare the spatial organization of TAF–PDO and TCF–PDO assembloids, we introduce the general concept of the "colocatome", which refers to the complete ensemble of colocalizations. In the context of tumor-stroma assembloids, we catalog cell subpopulation pairs in close proximity (positive colocalization) versus those that are distant (negative colocalization). While other spatial metrics are available[6,16], our framework expands upon use of the CLQ, a metric designed to assess the spatial association between two distinct constituents within a complex population[7]. We compute the CLQ value for each pair of cell subpopulations, using the 20 nearest neighbors for each cell following the equation and parameters detailed in the Methods section. Then, we assess the significance of each CLQ value by randomly permuting the cell subpopulation labels 500 times within each sample while preserving subpopulation ratios to generate a CLQ null distribution for each cell pair. Observed CLQ values falling within the tail of their respective null distribution generation by permutation testing were considered as significant (non-random) (Fig. 2a, b, Supplementary Data 2). This permutation approach accounts for the prevalence of cell subpopulations when determining significance. For instance, when the CLQ is computed among two highly prevalent cell subpopulations, the null CLQ distribution tends to be narrow, whereas rare subpopulations lead to broader null CLQ distributions. In the latter case, observed CLQ values are more likely to be located within the null distribution and thereby considered non-significant; less significant CLQs represent rarer cell subpopulations. We then normalize CLQs (see Methods) and characterize positive or negative colocalizations as an attraction or avoidance of spatial states between cell pairs. Only significant CLQ values ($p < 0.05$) consistent across conditions were considered for downstream colocatome analysis (Fig. 2c). The normalization step enables us to consider the colocatome as a spatial omic given its generalizability across conditions and sample types.

Our results revealed that TAF-PDO, compared to TCF-PDO assembloids, exhibited a higher number of tumor-stroma colocalizations—with approximately twice as many significant spatial features (Fig. 2d–i). Surprisingly, more than 80% of tumor-stroma colocalizations observed in assembloids were negative, suggesting that cancer cells and fibroblasts are primarily segregated from each other rather than being proximally intermixed (Fig. 2d). Interestingly, TAF-PDO assembloids and TCF-PDO assembloids shared no positive cancer–fibroblast colocalizations, and cell subpopulations with higher cell counts did not necessarily exhibit a higher number of statistically significant colocalizations with other cell types. For instance, the larger cell subpopulations (#3 in the TAF-PDO and TCF-PDO assembloids and #14 in TCF-PDO assembloid) did not positively colocalize with any cancer cells (Fig. 2h, i). In summary, these results demonstrate that regionally distinct fibroblasts contribute to unique tumor–stroma spatial organization in tumor–stroma assembloids.

Colocatome analysis reveals that erlotinib alters spatial organization of tumor-stroma assembloids

Little is known about effects of oncologic treatments on the spatial architecture of the TME. Because CAFs protect cancer cells through various mechanisms, they have been proposed as a cancer therapeutic target[17,18]. We thus wondered whether fibroblasts can induce drug resistance through spatial rearrangement. Given that the LUAD PDOs used in this study harbor a mutation in the epidermal growth factor receptor (EGFR) (Supplementary Table 1), we used EGFR-targeting erlotinib as a means to disrupt assembloids. We found that erlotinib reduced cell density of PDO monocultures. However, we did not observe any significant decrease in cell density in erlotinib-treated TAF-PDO or TCF-PDO assembloids, compared to their respective treatment-naïve controls (Fig. 3a). Moreover, none of the assembloids cell subpopulations were affected considerably (either enriched or depleted) when comparing naïve and treatment conditions (Fig. 3b–g). While several changes were statistically significant using chi-square testing (Supplementary Data 4), effect sizes were small. For the vast majority of cell subpopulations, erlotinib did not induce changes greater than 5% across independent experiments. Interestingly, we did observe that erlotinib induced statistically significant spatial rearrangements, revealing numerous emergent, persistent and sensitive spatial colocalization features in both TAF–PDO and TCF–PDO assembloids. These spatial rearrangements were not homogeneous, enriched either at the periphery or at the center of the assembloids (Fig. 3h–k, Supplementary Figs. 6–7, Supplementary Data 5). Numerous cancer–fibroblast positive colocalizations, particularly in TAF–PDO assembloids, persisted in the presence of erlotinib, and emergent positive colocalizations arose in erlotinib-treated assembloids compared to naïve assembloid in both models, predominantly involving Vim+ cancer cells and CAV1+ or αSMA + /CAV1+ fibroblasts (Fig. 3h–k, Supplementary Fig. 8).

Next, we investigated the effect of erlotinib on phenotypic changes in cancer cells by assessing epithelial-to-mesenchymal transition (EMT) states of cancer cells as another possible cause of erlotinib resistance. To do so, we used a single cell tool developed in our lab called EMT-MET PHENOtypic STAte MaP (PHENOSTAMP)[19]. PHENOSTAMP employs a neural net algorithm to map cancer cells from different studies onto an EMT-MET spectrum. Although erlotinib induced a partial EMT in PDO monocultures after 72 h (which was associated with elevated levels of O-GlcNAc, MDR1, and MUC1, known mechanisms of erlotinib resistance[20–22]), this effect was modest (Supplementary Figs. 9–10) and unlikely to have considerably influenced the observed resistance. Overall, these intriguing findings suggest that TAFs and TCFs may protect cancer cells from erlotinib mainly through

Table 1 | Common CAF, mesenchymal, and epithelial cancer cell markers involved in LUAD and/or NSCLC included in the mIF antibody panel

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.	44
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.	45
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 PDGFRβ in NSCLC.	46,47
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.	48
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.	38
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.	49
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 progression and unfavorable survival outcome in NSCLC.	50
KRT	Pan-cytokeratin	Epithelial marker	Specific isoforms of individual cytokeratin may have utility as diagnostic or predictive markers in lung adenocarcinomas.	51,52
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.	22,53–55
EPCAM	EpCAM	Epithelial marker	Transmembrane glycoprotein involved in intercellular adhesion and cell adhesion. May have utility as diagnostic or predictive markers in lung adenocarcinomas.	56–58

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.

mechanisms associated with tumor-stroma spatial reorganization, rather than by selecting for specific cell subpopulations.

Colocatome analysis enables discovery of treatment-resistant tumor-stroma spatial features

To resolve the spatial organization of assembloids across treatment-naïve and erlotinib-treated conditions, we grouped all statistically significant cell–cell colocalizations into a matrix of three possibilities. Briefly, significant negative colocalizations were assigned -1, significant positive colocalizations were assigned 1, and insignificant spatial features were assigned 0. This schema allowed us to establish a simplified yet comprehensive reference of spatial features that we designate as the "composite tumor-stroma colocatome." Detailed methods describing the construction of the composite colocatome are in the Methods section. Using hierarchical clustering on heterotypic composite colocatome, we identified distinct colocation clusters (Fig. 4a–e, Supplementary Fig. 11) that highlight groups of cancer–fibroblast pairs that spatially behave in a similar manner. Prior work has shown that fibroblasts promote EMT in cancer cells, and cancer cells sustain the myofibroblast phenotype in fibroblasts[23–25]. We corroborated these findings by showing, for example, that Vim- cancer cells are unlikely to colocalize with myofibroblasts (aSMA +) (Fig. 4b–f), whereas myofibroblasts are generally proximal to Vim+ cancer cells. In addition to confirming established cancer–fibroblast colocalizations, our composite colocatome analysis provided insights into additional spatial configurations within the TME. Notably, we observed that MUC1+ cancer cells are unlikely to be proximal to fibroblasts. For instance, the cancer cell subpopulation #3 (PanCK + / Vim-/MUC1 +) was not proximal to any fibroblast regardless of the treatment condition. Moreover, we found that while CD90+ fibroblasts reside in the vicinity of cancer cells, cancer cells are not necessarily in the vicinity of CD90+ fibroblasts, showing that spatial relationships between cancer cells and fibroblasts can be asymmetrical (Fig. 4b–e). In summary, the composite colocatome analysis inferred from our assembloids experiments validated known cancer–fibroblast colocalizations and revealed additional tumor–stroma spatial features potentially associated with erlotinib resistance.

Colocatome analysis demonstrates colocalizations derived from tumor-stroma assembloids that are recapitulated in LUAD clinical samples

Determining cell types or cell states in large and complex human tumor samples can be challenging because of shared markers between a multitude of cell types and cell states. 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 are not always associated with known distinctive markers in clinical samples. Importantly, we observed that rare features in clinical specimens were captured in the assembloids, given their inherent abundance of cell states. We applied our quantitative spatial framework to whole slide images from three different LUAD patients (Supplementary Table 1) imaged with a PhenoCycler with a comparable antibody panel (Supplementary Data 1). The samples were delineated into 13 histological regions, assigned as either lepidic, acinar, or solid by an expert pathologist. Of note, predominance of these histologic growth patterns have prognostic significance: lepidic predominant tumors are associated with a good prognosis due to their noninvasive nature, whereas acinar and solid predominant tumors are associated with worse prognosis, with solid worse than acinar[26,27]. We compared the CELESTA cell assignment using identical parameters, starting from either clinical specimens or assembloids (Supplementary Data 3). Identifying cell subpopulations directly from clinical specimens resulted in fewer identifiable cell

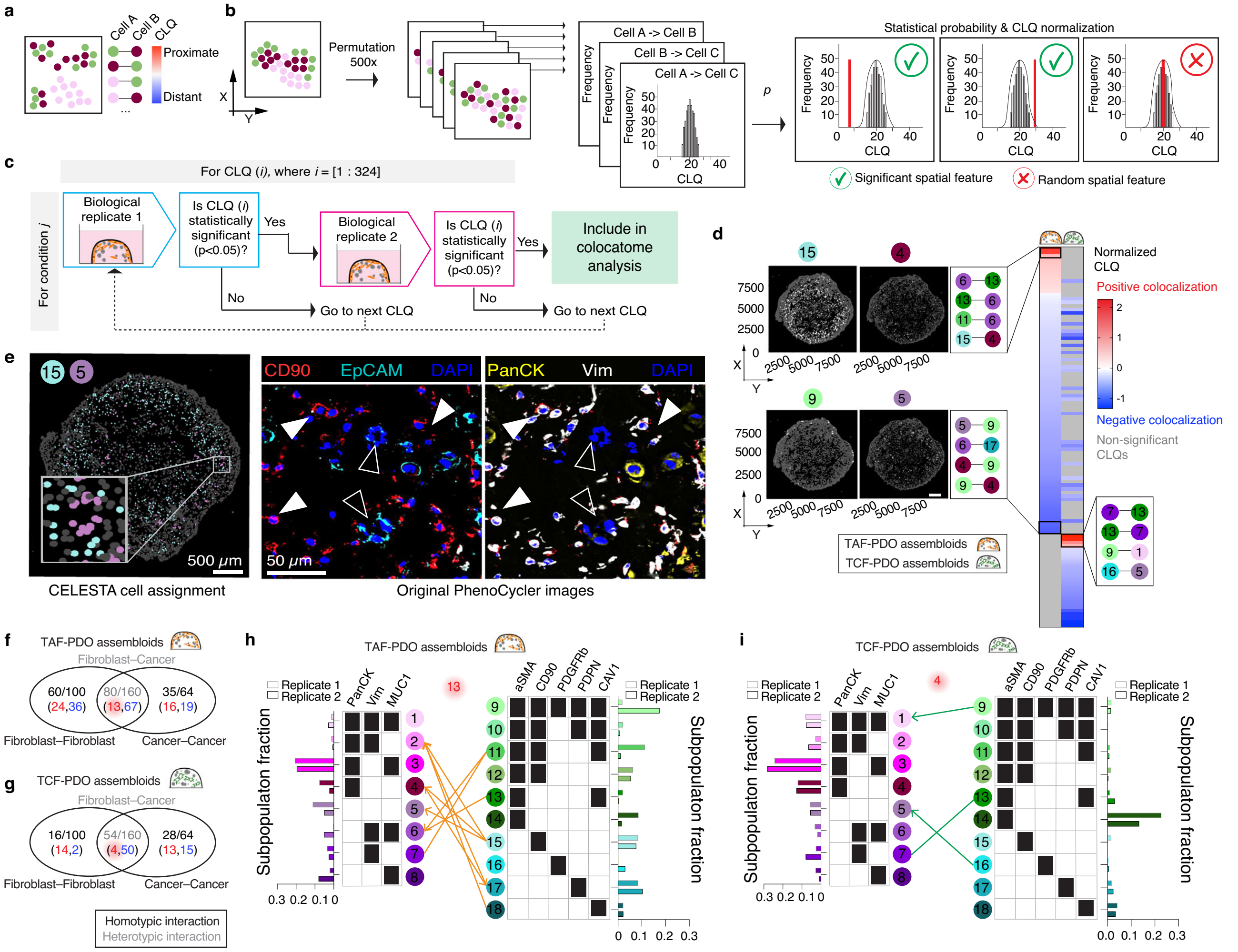
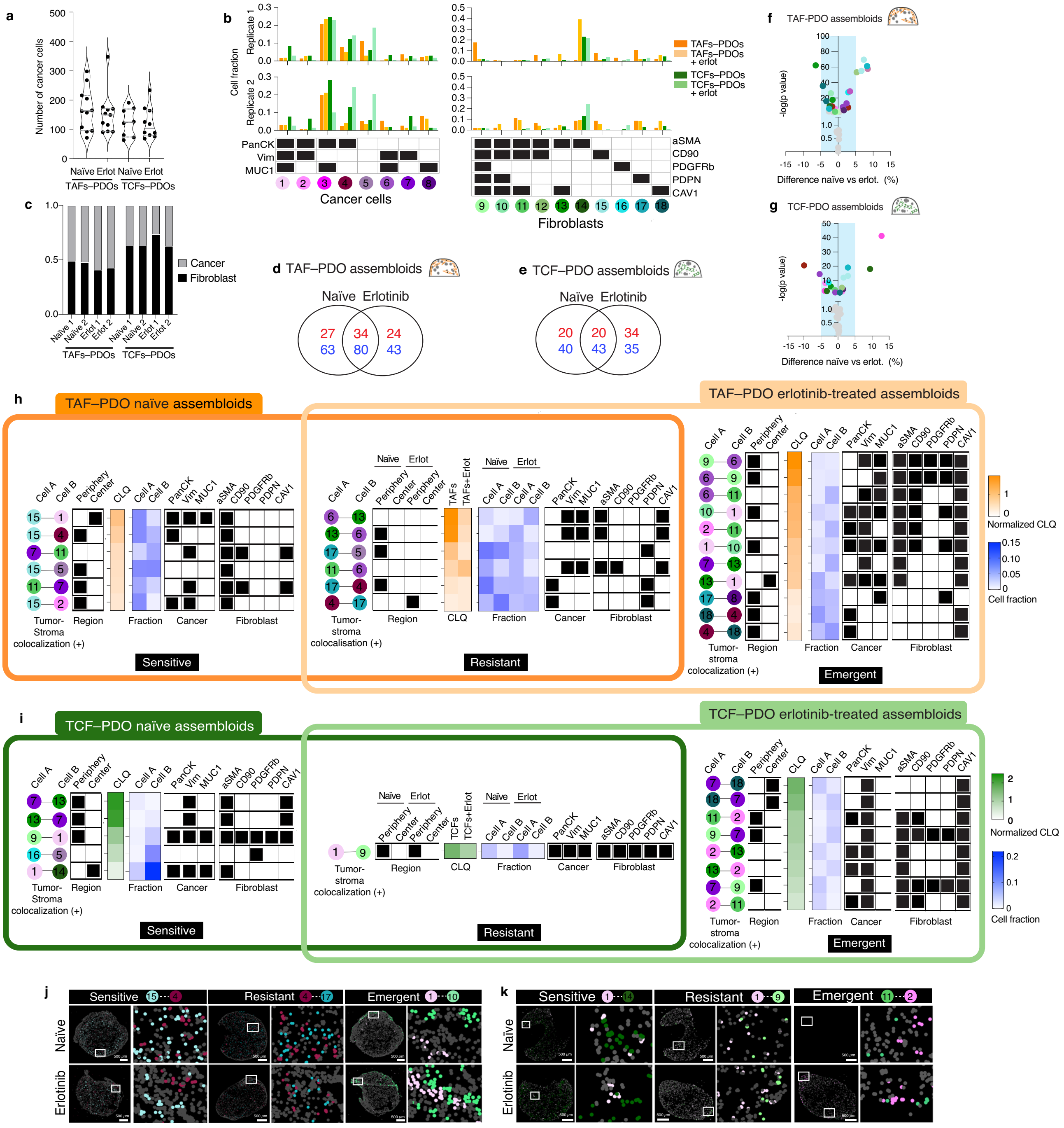


Fig. 2 | Colocatome analysis enables a comparisons of cell-cell colocalizations in TAF-PDO versus TCF-PDO assembloids. a Schematic representation of the colocation quotient (CLQ). **b** Workflow of the permutation analysis to determine statistically significant cell–cell colocalizations and, **c** selection of statistically-significant pairwise spatial features in each assembloid that are consistent across the two biological replicates are carried out for downstream colocatome analysis. More specifically, for each condition j, all CLQs are tested (for i = 1 to 324) and only CLQs that are statistically significant are included for downstream colocatome analysis. We generated the null distribution for each CLQ by randomly permuting 500 times the cell labels. Observed CLQ values falling within the 5% tails (two-sided test) of the null distribution were considered significant. **d** Heatmaps showing significant heterotypic negative (blue) and positive colocalizations (red) in TAF-PDO and TCF-PDO assembloids. For each assembloid condition, for each biological replicate (i.e., TAF-PDO and TCF-PDO assembloids, for two biological replicates), colocation analysis was conducted separately. The number of cells included in each colocation analysis are as follows: the TAF–PDO and TCF-PDO assembloids for the first biological replicate had 4592 cells and 1955 cells, respectively, and from the second biological replicate, had 9055 cells and 2440 cells, respectively. For the TAF–PDO assembloids, findings that were statistically significant ($p < 0.05$), through spatial permutation analysis, in each biological replicate and consistent across the two biological replicates are reported as the mean CLQ, highlighting positive (red) and negative (blue) colocalizations. A similar analysis was performed for the TCF–PDO assembloids. CLQs for all cell pairs, for each assembloid, can be found in Supplementary Data 2, with their statistical significance values. Scale bar = 500 μm. **e** Representative example of a significant colocalization validated on original PhenoCycler images. Results were obtained for each of two biological replicates for each assembloid condition, separately. White-solid arrowheads indicate fibroblasts, and clear-filled arrowheads indicate cancer cells. Scale bar = 500 μm (left panel) and 50 μm (right panel). Venn diagram showing the number of homotypic and heterotypic colocalizations in TAF–PDO (**f**) and TCF–PDO (**g**) assembloids resulting from the analysis in (**d**). Only findings that are statistically significant through spatial permutation analysis in each biological replicate ($p < 0.05$) and consistent across both biological replicates, for each assembloid condition, are reported here as the number of homo- and heterotypic colocalizations. The denominators sum to 324, which is the total number of pairwise cell-cell colocalizations. Red vs blue numbers indicate positive vs negative heterotypic colocalizations, respectively. Bipartite graphs showing the heterotypic positive colocalizations in TAF-PDO (**h**) and TCF-PDO (**i**) assembloids from the Venn diagrams illustrated in (**f**, **g**). Lines with (vs without) arrowheads represent unidirectional (vs bidirectional) colocalizations with corresponding bar graphs showing the subpopulation fractions from each biological replicate. PanCK- cancer cells were confirmed as epithelial cells using EpCAM+ status. Source data are provided as a Source Data file.

subpopulations (Supplementary Fig. 12a, b). Notably, assembloids adequately recapitulated all cell subpopulations found in clinical specimens, as well as improved detection of some cell subpopulations found rarely in tissues, but prevalent in assembloids (PanCK+/Vim-/MUC1+; #3).

Next, we applied our quantitative spatial framework to characterize the architecture of the clinical specimens (Supplementary Fig. 12c–e and Supplementary Data 6). While numerous statistically significant colocalizations were identified in assembloids, only significant colocalizations shared between clinical specimens and assembloids were included in downstream analyses. In other words, we excluded non-clinically relevant spatial features that may have emerged artifactually from in vitro culture conditions. See the Methods section for specific details related to generation of the LUAD



colocatome and statistical analysis. Of 133 colocalizations identified across 13 histological regions, most colocalizations (107/133) were recapitulated in the assembloids (Fig. 5a–c, Supplementary Data 6). Note that 19 interactions were categorized as avoidance (negative) in the assembloids, and attraction (positive) in the specimens, or vice-versa (Fig. 5a), which are indicated as inconsistencies in the colocatome composite summary (Supplementary Data 6).

Hierarchical clustering of shared colocatomes between all assembloid conditions and clinical specimens showed that solid regions mostly grouped with TAF-PDO assembloids (mauve cluster 3, Fig. 5d), whereas lepidic and acinar regions grouped predominantly with TCF-PDO assembloids (pink cluster 2, Fig. 5d). These results align with our previous results, which suggest a more tumor-promoting role of TAFs compared to TCFs[28] by associating TAF-PDO spatial features with worse prognostic growth patterns (solid) and TCF-PDO spatial features with better prognostic growth patterns (lepidic and acinar). The statistical significance of these clusters were independently tested using multiscale bootstrap resampling to assess the uncertainty in hierarchical cluster analysis using pvclust[29] (Supplementary Fig. 13). We also found that acinar regions from different clusters displayed distinct spatial features (Supplementary Fig. 14). Of note, these results were not supported by hierarchical clustering based on cell composition alone, further demonstrating the added value of the colocatome as an informative spatial omic dimension.

We also compared our quantitative spatial framework to a published imaging mass cytometry breast cancer dataset that also uses spatial permutation to identify significant spatial features (Schapiro et al.[6]). Our analysis replicated most of the colocalizations from that original study but also highlighted additional findings. For instance, our approach suggests colocalization between macrophages (CD68 +)

Fig. 3 | Effect of erlotinib treatment on tumor-stroma assembloids. a Violin plot representing cancer cell density in treatment-naïve and erlotinib-treated assembloids. Results are presented as cell counts from representative areas for each assembloid condition for one biological replicate (i.e., specimen 2 in Supplementary Table 1). The number of areas included in each analysis is as follows: treatment-naïve TAF–PDO ($n = 11$) and TCF–PDO ($n = 8$) assembloids, and erlotinib-treated TAF–PDO ($n = 11$) and TCF–PDO ($n = 10$) assembloids. Gray solid lines represent the mean and gray dash lines represent quartiles. Results were repeated for a second replicate in Supplementary Fig. 6a. **b** Bar graphs showing the fraction of each cell subpopulation in the TAF–PDO and TCF–PDO assembloids, in the treatment-naive and erlotinib-treated conditions, for each of two biological replicates. **c** Bar graphs showing the fraction of cancer cells and fibroblasts in the treatment-naïve vs erlotinib-treated assembloids, separately for each of two biological replicates, described in Supplementary Table 1. **d**, **e** Venn diagram showing the number of positive (red) and negative (blue) colocalizations following the workflow described in Fig. 2a–c. For each assembloid condition, colocation analysis was conducted separately for each biological replicate. The number of cells included in the colocation analysis is as follows: treatment-naïve TAF–PDO and TCF-PDO assembloids for the first biological replicate had 4592 cells and 1955 cells, and for the second biological replicate had 9055 cells and 2440 cells, respectively. For the erlotinib-treated TAF–PDO and TCF-PDO assembloids, the first biological replicate had 4809 cells and 3403 cells, and the second biological replicate had 7401 cells and 3643 cells, respectively. For each assembloid condition (e.g., treatment-naïve TAF–PDO, erlotinib-treated TAF–PDO, treatment-naïve TCF–PDO, erlotinib-treated TCF–PDO) only findings that were statistically significant and consistent across the two biological replicates are presented, as positive or negative colocalizations. Significance was assessed through spatial permutation testing to evaluate whether the CLQ scores were statistically significant. **f**, **g** Volcano plots representing the differences in cell proportions between treatment-naïve and erlotinib-treated TAF–PDO and TCF–PDO assembloids, pooled for the two biological replicates for visualization purposes. *P* values are calculated using chi-square ($\chi 2$) tests (two-sided) with $p < 0.05$ considered as significant (Supplementary Data 4), for each biological replicate separately. Color coding refers to subpopulations in **b**, and throughout study, and total number of cells analyzed per assembloid condition are detailed in (**d**, **e**). **h**, **i** Heatmaps representing normalized pairwise CLQ values for significant heterotypic positive colocalizations and subpopulation fractions in treatment-naïve and erlotinib-treated assembloids as described in (**d**, **e**). CLQs and cell fraction heatmaps are presented as mean values across the two biological replicates. Note that colocation analysis was conducted individually for each sample, as detailed in (**d**, **e**). The normalized CLQs and their corresponding significance are listed in the Supplementary Data 2 and 5. CELESTA cell assignment plots showing representative examples of sensitive, resistant, and emergent heterotypic colocalizations in (**j**) TAF-PDO and (**k**) TCF-PDO assembloids from the analysis described in (**d**, **e**). Scale bar = 500 μm. Source data are provided as a Source Data file.

and fibroblasts (spindle-shaped Vim +), neither identified in the original study, but which we confirmed on the original mass cytometry images (Supplementary Fig. 15). Colocalization between macrophages and fibroblasts in breast cancer has been predicted in various models[30,31], as well as in recent breast cancer spatial transcriptomics studies using spot-based technologies[32,33]. In a contrasting example, we found that cell phenotypes #2 and #3 (Supplementary 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 breast dataset (Supplementary Fig. 15c). These examples highlight the added value of applying colocatome analysis using the CLQ alongside other tools such as the histoCAT toolbox to increase the discovery of potentially relevant biological colocalizations and architectural features.

Taken together, these results suggest that colocatome analysis based on use of the CLQ spatial metric, among other spatial metrics, offers added value over analyses based solely on cell composition and may help reveal spatial heterogeneity among histologically similar regions when applied to larger datasets.

Colocatome analysis informs on suitable in vitro validation platforms for downstream functional validation

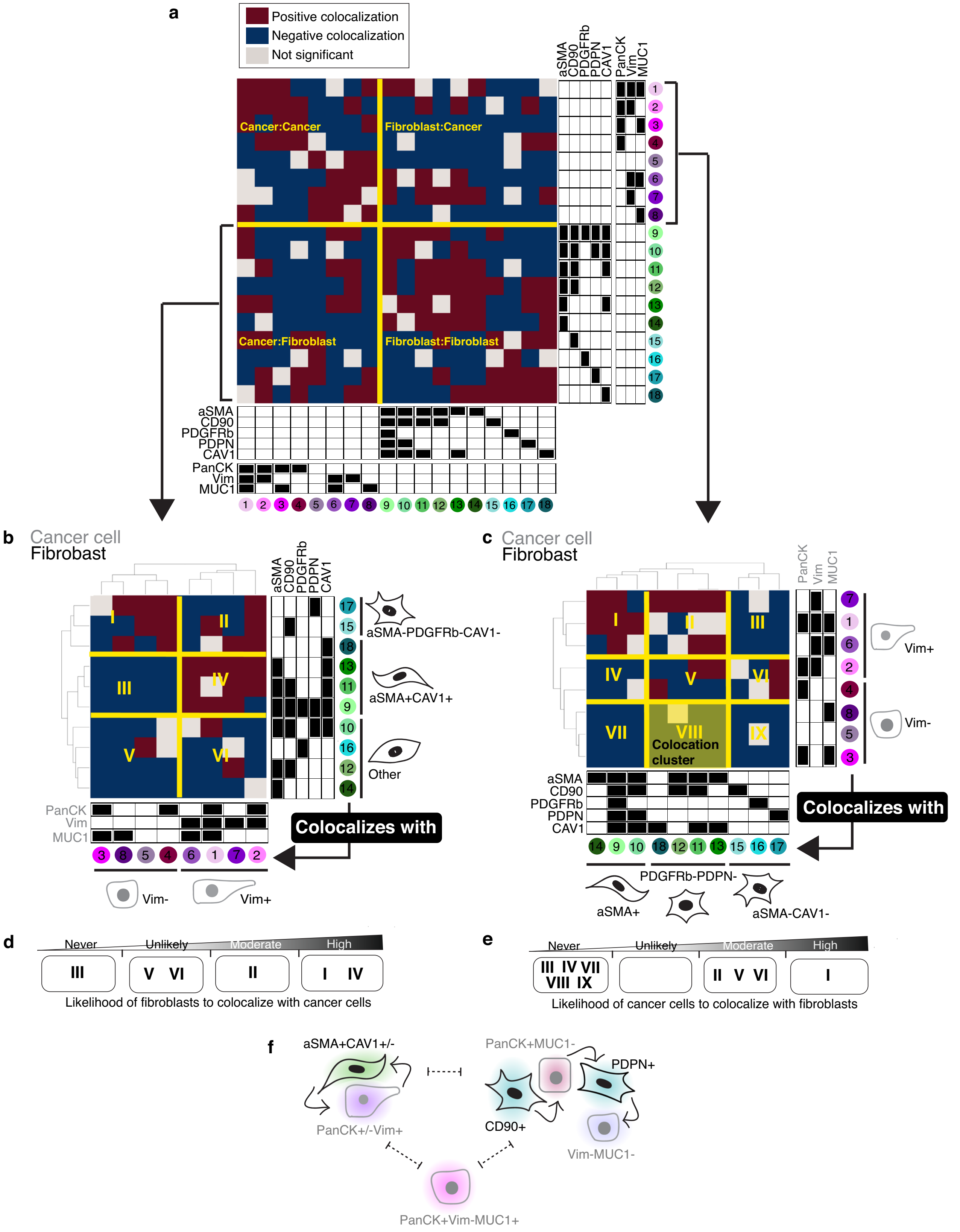
Other advantages of integrating the spatial architecture of clinical specimens with in vitro patient-derived models include the ability to study spatial re-organization during treatment and to select appropriate models for downstream functional analysis based on spatial features associated with resistance. To illustrate this, we compared the shared spatial features between treatment-naïve clinical specimens and the persistent and emergent colocalizations identified in erlotinib-treated assembloids.

We found that treatment-persistent cellular colocalizations vary significantly depending on the origin of CAFs in the TME, and that many spatial features associated erlotinib resistance appear in treatment-naïve tumor samples (Fig. 5f). Our observations thus suggest that certain cellular colocalizations may persist after treatment, contributing to drug resistance. As an example, we identified a cancer–fibroblast colocalization exclusively recapitulated in TAF-PDO assembloids, but not in TCF-PDO assembloids. Specifically, PanCK + / Vim-/MUC1+ cancer cells (#3) negatively colocalize with CD90+ fibroblasts (#15), which appear as segregated compartments in LUAD specimens of solid histology and TAF–PDO assembloids, but not in TCF–PDO assembloids (Fig. 5g). These findings suggest that spatial re-organization accompanies drug resistance. We also showed that some cellular colocalizations in acinar regions are only recapitulated in either TAF–PDO or TCF–PDO assembloids (Supplementary Fig. 16). Overall, our findings suggest that using quantitative colocatome analysis to compare spatial features in assembloids and clinical samples can (i) guide the use of in vitro validation platforms for subsequent functional analysis and (ii) unearth unknown biology of potential clinical relevance.

Discussion

In this study, we present the concept of the colocatome as a spatial omic dimension that can be used to quantitatively characterize and compare spatial organization between in vitro patient-derived models and human tumor tissues under various conditions (Fig. 6). Our spatial framework uses a quantitative approach to measure pairwise CLQs and assess their statistical significance, followed by a normalization step to enable further comparisons between conditions and across studies. Our results confirm that this spatial framework is useful for both for validating in vitro patient models as a means to interrogate human cancer and as a source for generating and testing hypotheses (e.g., the role of spatial reorganization in drug-resistance). Although we specifically applied colocatome analysis to spatial proteomics data, this general framework is applicable to other spatial -omics platforms with single cell resolution.

We show that distinct cellular colocalizations TAF–PDO and TCF–PDO assembloids correspond to histologic growth patterns in clinical samples. Of note, acinar-specific spatial features were predominantly recapitulated in TCF–PDO assembloids, whereas solid-specific features were recapitulated in TAF–PDO assembloids. These results align with our previous work aligning TAF-PDO spatial features[28] with worse prognostic growth patterns (solid). Importantly, our analysis reproduced known findings but also generated original hypotheses. For example, mesenchymal cancer cells expressing vimentin are generally found in proximity to myofibroblasts, and vice-versa, a known co-occurrence[34]. Through our colocatome composite analysis, we also identified additional cancer–fibroblast spatial features potentially associated with drug resistance. We observed, for example, that cancer cells and fibroblasts tend to be separate rather than intermixed. We also found that cancer cells expressing MUC1 (PanCK + /Vim-/MUC1 + , #3) were unlikely to be proximal to any



fibroblasts: an observation confirmed with analysis of human LUAD samples. MUC1 is a glycoprotein lubricant that protects epithelial cells. Its aberrant expression has been involved in cancer development, invasion, metastasis, and treatment resistance, including to erlotinib[22,35]. Interestingly, we observed that both TAFs and TCFs significantly increased the number of MUC1+ cancer cells within assembloids compared to PDO monocultures. However, MUC1+ cancer cells were not found in the vicinity of any fibroblasts. These findings suggest that MUC1+ cancer cells may not need fibroblasts in close proximity to survive erlotinib treatment. They also suggest more

Fig. 4 | The tumor–stroma colocatome. a Heatmap of the statistically significant colocalizations across conditions generated in this study establish a reference tumor–stroma colocatome (colocatome composite). Results are presented as the summarized positive and negative colocalizations resulting from the CLQ analysis of a total of 37,298 cells across four assembloid conditions, each repeated under two biological replications. The number of cells included in the analysis per is as follows: treatment-naïve TAF–PDO and TCF-PDO assembloids from the first biological replicate had 4592 cells and 1955 cells, and from the second biological replicate had 9055 cells and 2440 cells, respectively. For erlotinib-treated TAF–PDO and TCF-PDO assembloids, the first biological replicate had 4809 cells and 3403 cells, while the second biological replicate had 7401 cells and 3643 cells, respectively. Colocalization analysis was performed independently for each assembloid condition, for each biological replicate. For each assembloid condition, only findings that were statistically significant and consistent across the two replicates are presented here as binary values, either positive or negative colocalizations. If inconsistent, results are presented as insignificant. Blue boxes represent negative colocalizations, red boxes represent positive colocalizations, and gray boxes represent insignificant colocalizations. **b**–**e** Hierarchical clustering of the tumor-stroma colocatome. Colocation clusters highlight paired cell-cell colocalizations and indicate the likelihood of cancer cells to colocalize with fibroblasts or vice-versa. **f** Graphical summary of the cancer–fibroblast spatial organization inferred from the colocatome. Solid arrows represent positive colocalization with arrowhead identifying direction of colocalization (i.e., CD90+ fibroblasts colocalize PanCK+MUC1- cancer cells) and dotted lines indicate segregated spatial features that negatively colocalize. Source data are provided as a Source Data file.

complex tumor-stroma spatial configurations, such as fibroblasts surrounding cancer cell nests, or other yet-to-be-revealed structures that may be relevant for understanding and mitigating treatment resistance.

To explore this further, we investigated spatial features involved in treatment resistance by observing distinct spatial organization between TAF–PDO and TCF–PDO assembloids under treatment-naïve and erlotinib-treatment conditions. Our findings suggest that TAFs and TCFs may protect cancer cells from erlotinib through reorganization of the tumor-stroma environment rather than by promoting cell heterogeneity. Moreover, our findings show an enrichment of positive heterotypic colocalizations predominantly involving CAV1+ fibroblasts in erlotinib-treated assembloids. Several reports, including results from our group in which we computationally reconstructed the LUAD interactome[36], conclude that CAV1 is associated with poor outcome in LUAD. However, some studies have suggested the opposite[37], based upon high expression of CAV1 in normal lung tissue compared to lung tumors[38–40]. These conflicting findings raise further questions about whether erlotinib potentially restore a normal-like lung spatial architecture or if CAV1's association with poor outcomes is dependent on specific tumor-stroma interactions. These hypotheses warrant further study in larger cohorts to better understand the roles and interactions of erlotinib and CAV1 expression in the TME.

While assembloids effectively recapitulated most spatial features measured in human LUAD specimens, a wider range of cell subpopulations were observed in vitro. We attribute this observation to lower cell density and partially recapitulated environmental cues in assembloids compared to LUAD specimens, which may affect cell plasticity and overall phenotype. Moreover, while in vitro models did not perfectly replicate tissue architecture, we demonstrated how identifying cell subpopulations within assembloids of known composition can enhance the detection of rare cell subpopulations. As such, our quantitative framework can also help mitigate sensitivity issues that arise when analyzing rarer subpopulations using permutation analysis. Permutation analysis tends to reflect a broader null distribution for smaller subpopulations, thereby increasing the likelihood of their observed CLQ values being deemed non-significant. For example, the presence of the PanCK + /Vim-/MUC1+ cellular subpopulation (which is not commonly found in surgically resected early-stage LUAD specimens) would likely have been overlooked if we had directly identified cell subpopulations from clinical specimens. Our method is particularly useful for identifying highly heterogeneous cell types (e.g., cancer cells and fibroblasts), which often lack known unique protein markers for identification. However, this may not apply to well-defined lineages that have distinctive markers that easily allows their identification in situ with high confidence, such as immune cells. Lastly, future studies should investigate the role of cell density and other cell types (such as immune and endothelial cells) or include additional markers. For instance, the antibody panel used in this study may not be sufficient to distinguish fibroblasts from pericytes or vascular smooth muscle cells. Therefore, the conclusions from our analysis of CAF subpopulations may require further validation, as we cannot exclude the possibility that some CAF subpopulations identified in human tumors may be mural cells. Future work should also compare spatial findings captured with alternate colocalization metrics and statistics beyond the CLQ and spatial permutation testing.

In conclusion, our quantitative spatial framework for colocatome analysis provides insights into spatial configurations between cancer cells and fibroblasts in LUAD. It can serve as a spatial readout for characterizing, cataloging, and comparing cell-cell colocalization ensembles in both in vitro patient-derived models and clinical samples. Furthermore, this approach may also guide analysis of other diseases and complex tissues, toward identifying pathological mechanisms and improving treatment response.

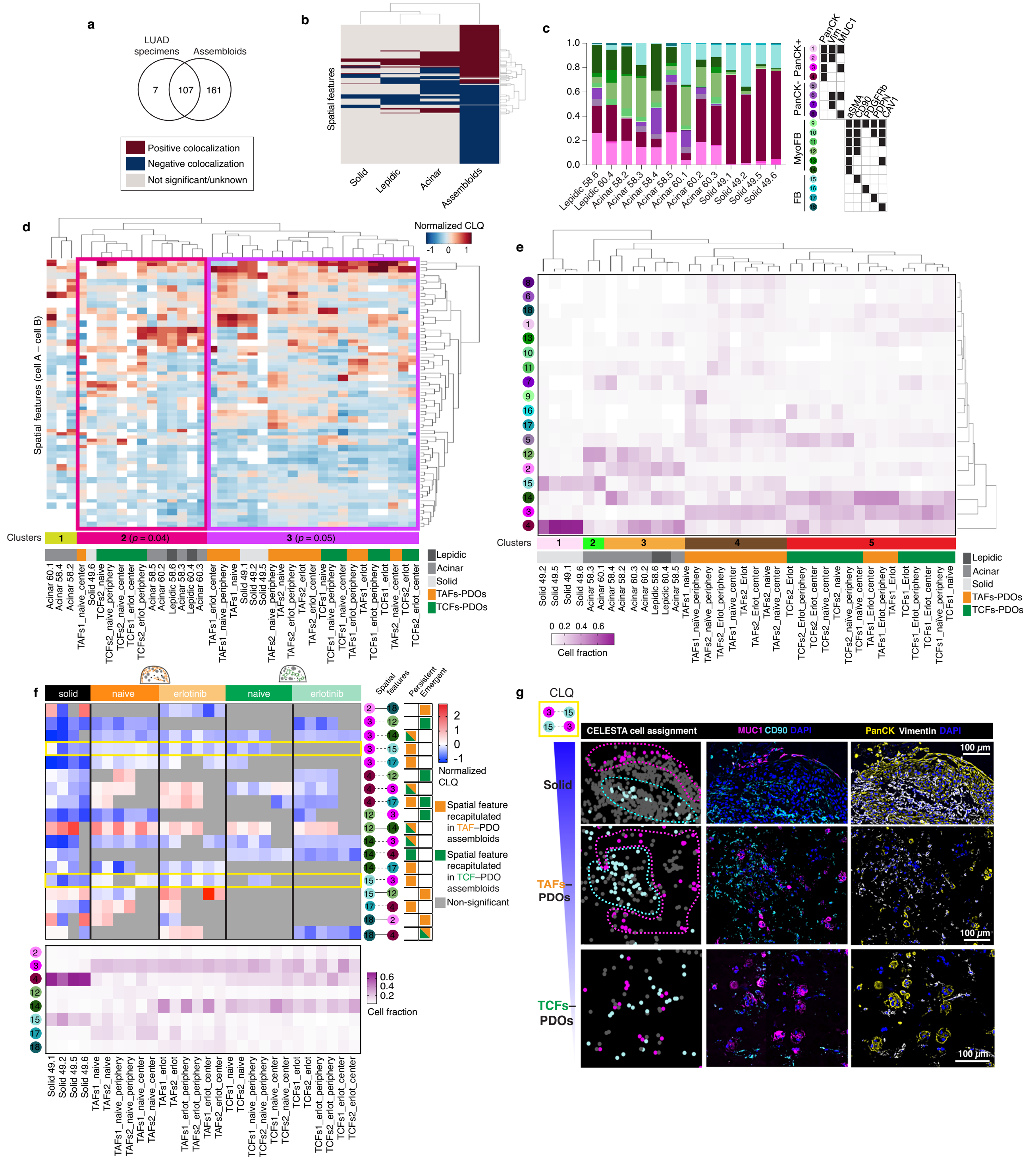
Methods

Human studies

Clinical aspects of this study were approved by the Stanford Institutional Review Board (IRB) in accordance with the Declaration of Helsinki guidelines for the ethical conduct of research. All patients involved provided a written informed consent. Collection and use of human tissues were approved and in compliance with data protection regulations regarding patient confidentiality (IRB protocol #15166). Following surgical resection of primary tumors, lung adenocarcinoma specimens were immediately processed to establish primary cell cultures as previously described[28].

Assembloid cultures

The PDOs originated from LUAD fresh specimens were kindly provided by Dr. Calvin Kuo, Stanford University and were grown in DMEM:F12 (1:1) media supplemented with 20 ng/ml human FGF-basic, 1X N-2 supplement, 1X B-27 supplement, 10 μM rock inhibitor, 50 ng/ml human EGF and 1X Normocin at 5% CO_2 and 37 °C in a Cultrex reduced growth factor basement membrane extract solid matrix. The LUAD primary fibroblasts were established by our group as previously described[28] and grown in RPMI-1640 with L-Glutamine, supplemented with 10% fetal bovine serum, and 5% antibiotic solution (penicillin/streptomycin), at 5% CO_2 and 37 °C. Fresh, paired LUAD specimens were collected from the tumor leading edge and tumor core of two patients. For the first patient, TAFs1 (edge) and TCFs1 (core) were harvested; for the second patient, TAFs2 (edge) and TCFs2 (core) were harvested. Experiments involving the specimens from the two patients are referred as biological replicates. Samples were immersed and transferred from Stanford Hospital to the laboratory in MACS Tissue Storage Solution (Miltenyi). Then, tumors were cut into small pieces with dissecting scissors. The dissociation was performed using the MACS Tumor Tissue Dissociation Kit (Miltenyi iotec) for the tumor core samples or the Lung Dissociation Kit (Miltenyi) for the adjacent tissues as per the manufacturer's protocol. Cells were then washed, resuspended in RPMI1640, and plated in a 35-mm dish for fibroblast expansion. At confluency, all cells were trypsinized and transferred to a 25 cm^2 flask until confluency, then all cells were split in two 75 cm^2



flasks for further expansion. At confluency of the 75 cm^2 flasks, primary fibroblasts were divided equally and frozen in 10 vials, designated here as passage 1. All assembloid coculture experiments were realized using primary fibroblasts at passage 2. More precisely, one vial (passage 1) was thawed in a 75 cm^2 flask, expanded until confluency, then split at a maximum dilution ratio of 1:10 or lower into subsequent flasks, designated here as passage 2. Passage 2 cells were collected at confluency of 80% to 95% to perform assembloid cocultures. The primary fibroblasts, PDOs and fibroblast cultures used in this study were subjected to mycoplasma testing through DAPI staining. Cells were authenticated by their expression of epithelial markers for PDOs and vimentin for fibroblasts, as well as their morphology. Clinical annotations and histopathological information can be found in Supplementary Table 1. Race and sex information for human fibroblast lines, organoids and LUAD tissue samples were not considered in our analysis because of the limited sample size. Gender information was not available.

PDOs and regionally distinct fibroblasts from the tumor edge (TAFs) and core (TCFs) were used to generate the assembloids. Briefly, confluent PDOs grown in a 3D solid dome of Cultrex matrix were rinsed with PBS 1X, and detached from the bottom of the plate with a tip. Then, the domes containing the PDOs were collected in a tube and incubated with TRYPLE at 37 °C for 15 min under agitation. After the incubation, the domes were gently dissociated using

Fig. 5 | The colocatome from tumor-stroma assembloids recapitulate LUAD spatial features. **a** Venn Diagram and **b** heatmap of the positive and negative colocalizations found in the LUAD specimens and assembloids. All colocalizations and corresponding p values are provided in Supplementary Data 6. CLQ results are from the independent analysis of 13 histological regions with the number of cells per individual region as follows: Solid 49.1: 10,765 cells, Solid 49.2: 22,858 cells, Solid 49.5: 5145 cells, Solid 49.6: 1506 cells, Lepidic 58.6: 8504 cells, Lepidic 60.4: 1673 cells, Acinar 58.2: 2590 cells, Acinar 58.3: 23,121 cells, Acinar 58.4: 172 cells, Acinar 58.5: 403 cells, Acinar 60.1: 2472 cells, Acinar 60.2: 31,269 cells, and Acinar 60.3: 9626 cells. Statistical significance was determined through spatial permutation testing for each region individually ($p < 0.05$) and CLQ values consistent across more than half of the regions of a same growth pattern were included in the heatmap. Results were integrated with the assembloids results described at Fig. 4a in which treatment-naïve TAF–PDO and TCF-PDO assembloids the first biological replicate had 4592 cells and 1955 cells, and from the second biological replicate had 9055 cells and 2440 cells, respectively. For erlotinib-treated TAF–PDO and TCF-PDO assembloids, the first biological replicate had 4809 cells and 3403 cells, and the second biological replicate had 7401 cells and 3643 cells, respectively. **c** Cell proportions identified with CELESTA for each histological region (lepidic, acinar, solid) analyzed from three different LUAD samples from patients #49, #58, and #60. **d** Hierarchical clustering of the normalized CLQ values shared between assembloids and clinical specimens. **e** Hierarchical clustering of the LUAD specimens and assembloids according to their cell composition **f**, Heatmap of the resistant spatial features that overlapped between erlotinib-treated assembloids, and solid growth patterns. Non-significant CLQ values and self-colocalizations were excluded for visualization purposes. Yellow boxes highlight a negative colocalization between CD90+ fibroblasts and MUC1+ malignant cells from a representative LUAD solid region and recapitulated in the TAF–PDO, but not in TCF–PDO assembloids, and (**g**) validated on the original Phenocycler images. CLQs are from the independent colocalization analysis performed on four solid regions and assembloid samples from two biological replicates per condition as described in (**b**). Source data are provided as a Source Data file.

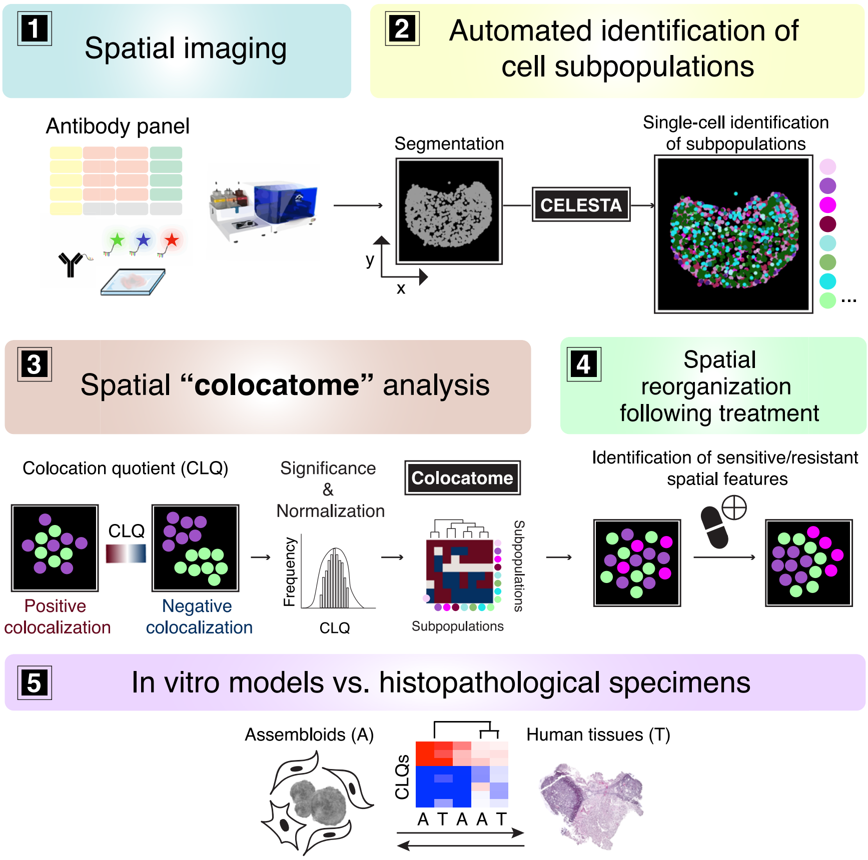


Fig. 6 | Schematic overview of the colocatome analysis spatial framework. The main steps of the colocatome framework are summarized as follows: (1) Multiplexed immunofluorescence imaging; (2) Cell subpopulations and states annotation using the CELESTA tool[7]; (3) Quantification of positive and negative colocation quotients (CLQ), followed by a spatial permutation test to establish significance and normalization for comparing cell pairs and conditions; (4) Evaluation of treatment effects on pairwise CLQs; and (5) Comparison of cell-cell colocalization patterns between in vitro models and patient histopathological specimens. Phenocycler image used with permission from Akoya Biosciences, Marlborough, MA.

repeated pipetting. In parallel, primary fibroblasts cultures at passage 2 were grown in a monolayer until reaching 75–80% confluency. Next, cells were rinsed with PBS 1X, incubated 5 min with TRYPLE, collected and spun down to remove debris. A 1:1 fibroblast: PDO mixture totalizing 300k cells was prepared, spun down to remove the remaining liquid, then resuspended in cold liquid Cultrex. Next, 40 µl of the cell mixture was added at the bottom of a 24-well plate pre-warmed at 37 °C. Cell mixtures were incubated at 37 °C for up to 30 min or until the domes solidified. Then, DMEM:F12 (1:1) media formulation supplemented as described above was added to the wells to cover the solidified assembloids. Assembloids were cultivated for a period of 7 days, followed by treatment with a 2 µM erlotinib solution, which was replenished every 24 h. The LUAD PDOs employed in this study carry an EGFR del19 mutation that is responsive to erlotinib. Naïve assembloids did not receive any treatment and were used as controls. After 72 h, the whole assembloids were rinsed with PBS 1X and embedded in OCT. OCT blocks were kept at -80 °C until sectioning at 7 µm with a cryostat on slides or coverslips (22 × 22 mm) pre-treated with poly-L-lysine overnight, washed 5x with double-distilled water (ddH2O), and dried.

In total, four assembloid conditions (1. TAF–PDO naïve, 2.TCF–PDO naïve, 3. TAF–PDO erlotinib-treated and 4.TCF–PDO erlotinib-treated) were performed with patient-matched primary fibroblasts (edge vs core), across two biological replicates; in other words, each assembloid condition was analyzed using fibroblasts established from two different patients.

Immunofluorescence

TAF–PDO and TCF–PDO assembloids sections on slides were rinsed 3x with a cold solution of 5% BSA (diluted in PBS 1x) and cells were fixed for 10 min in PFA 4% (diluted in PBS 1x). Then, slides were incubated in 0.1% Triton X-100 for 10 min, rinsed with PBS 1x, blocked with 10% goat serum diluted in PBS 1x for 30 min and incubated with the PanCK and vimentin antibodies in blocking solution in a humidified chamber overnight at 4 °C. After 3 washes in PBS 1x, the slides were incubated with the Cy3-conjugated goat anti-mouse IgG and Cy5-conjugated goat anti-rabbit IgG in blocking solution for 1 hat RT in the dark. The slides were intensively washed with PBS 1x and mounted on slides with a drop of mounting medium containing DAPI. The sections were observed under a BZ-X800 fluorescence microscope. All antibodies and reagents are listed in the Supplementary Data 7.

PhenoCycler® image acquisition and data processing

Phenocycler imaging requires the antibodies to be conjugated to DNA barcodes available from Akoya Biosciences according to the manufacturer protocol or can be purchased directly from the manufacturer if available. All antibodies were manually verified via standard immunofluorescence staining on appropriate tissues using the staining protocol from Akoya Biosciences to determine protein abundance and specificity. Results were compared to current literature and The Human Protein Atlas online database. The list of reporters and antibodies with target (clone), barcode, vendor and catalog number can be found in the Supplementary Data 7.

All imaging acquisition were performed with a PhenoCycler® (formerly CODEX) connected to a Keyence BZ-X800 fluorescence microscope equipped with a 20x objective (Nikon CFI Plan Apo 20x/0.75). Before imaging, the region and the z-stack were configurated on the Keyence software and the number of cycles with targets were designated in the Akoya's PhenoCycler® Instrument Manager (version 1.29.3.6) as described in the CODEX user manual (Akoya Biosciences, 2021a). Acquired images were processed, stitched, and segmented using the CODEX Processor set at default manufacturer values with a radius of 10 for segmentation (Akoya Biosciences, 2021b). Final data were then viewed using the CODEX Multiplex Analysis Viewer (MAV) plugin (Akoya Biosciences, 2021c) for FIJI (ImageJ)[41] and the Enable-Medicine Visualizer tool (beta version).

Cell assignment with CELESTA

CELESTA is an an unsupervised machine learning algorithm for individual cell identification using the cell's marker expression profile and, when needed, its spatial information[7]. The CELESTA package was downloaded from GitHub under https://github.com/plevritis-lab/CELESTA. As an initial step, a cell type signature matrix that relies on prior knowledge of markers was designed and used as CELESTA input, along with the segmentation data as previously described[7]. Our initial cell type signature matrix included all the possible combinations of a pre-selected set of markers from our antibody panel to identify cancer cells (panCK, VIM, MUC1) and fibroblasts (αSMA, CD90, PDGFRb, PDPN, CAV1), along with other general markers to identify subpopulations. Every cell subpopulation (designated by unique combination of markers) with a cell fraction smaller than 2% across all samples were removed from the cell type signature matrix before a subsequent iteration of CELESTA, until reaching a final cell type assignment. 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. The cell type signature matrix and the CELESTA thresholds can be found in Supplementary Data 1. All CELESTA annotations were manually verified by comparing the cell assignment dot 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, as described in the CELESTA user guide[7].

Flow cytometry analysis

After TRYPLE incubation, PDOs were dissociated into single-cell suspension. Cells were counted and aliquots of 1×10^6 cells per condition were prepared. The samples were incubated for 5 minutes with 1 μL of Zombie Aqua™ fixable viability dye in PBS 1X, then washed with flow cytometry buffer (FCB, 0.5% BSA, 0.02% NaN_3 and 2 mM EDTA in PBS 1X) and centrifuged (500 × *g*, 5 min) before adding PFA at a final concentration of 1.6% for 10 min at room temperature in FCB. Cells were then centrifuged at 500 × *g* for 5 min at 4 °C to pellet cells, PFA was removed, and cells were washed again with FCB. Cells were either long-term stored at −80 °C in 500 μL of FCB or permeabilised with 100 μL of eBioscience™ Permeabilisation Buffer diluted at 1X concentration for 30 min on ice with a master mix of primary antibodies (see Supplementary Data 7). After the incubation with primary antibodies, cells were washed with FCB and spun down at 500 × *g* for 5 min at 4 °C (2X). Then, cells were resuspended in 500 µl of FCB, strained and analysed using a BD LSRFortessa™ X-20. Results were analysed using Cytobank single-cell analysis software with the gating strategies described in Supplementary Fig. 9.

EMT analysis with PhenoSTAMP

PHENOSTAMP was downloaded from GitHub under https://github.com/anchangben/PHENOSTAMP. Briefly, FCS files previously gated in Cytobank according to singlets and Live/Dead (Supplementary Fig. 9), were uploaded into R, and the PHENOSTAMP algorithm was used to project the PDOs on the 2D EMT–MET state map, as previously described here[19].

Colocatome analysis

To quantify the tumor–stroma spatial organization within the assembloids, we used the CLQ and identified positive (proximate) and negative (distant) cell–cell spatial features between each cell subpopulation as previously described[7,42]. CLQ calculations were performed on the complete scans of the assembloids. Briefly, we use the CLQ to quantify how a cell subpopulation co-locates spatially with another cell subpopulation among a set of nearest neighbors, defined here as 20. 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 and reflecting a realistic number of cells for a biological neighborhood (Supplementary Fig. 11a). CLQs were calculated for pairwise cell types identified under naïve and treatment conditions using the following equation: $CLQ_{b \to a} = (C_{b \to a}/N_a) / (N_b/(N-1))$ where $C_{b \to a}$ is the number of cells of cell type b among the defined nearest neighbors of cell type a, N is the total number of cells and N_a and N_b are the numbers of cells for cell type a and cell type b, respectively. We use a Euclidean distance bandwidth parameter, which is determined based on the segmented X and Y coordinates, with a default value set to 100 according to the recommendations from the authors of the original CELESTA manuscript to always ensure 20 neighboring cells in the vicinity[7]. This CLQ is sometimes referred to be as the "observed CLQ" to distinguish it from the normalized CLQ, defined below.

We generated the null distribution for each CLQ by randomly permuting 500 times the cell labels, thus preserving the subpopulation proportions. The number of permutations was set to 500 as we found it generated an adequate CLQ distribution within a reasonable compute time and larger numbers of permutations did not change the findings. Observed CLQ values falling within the 5% tails of the null distribution were considered significant, associating statistical significance with a $p < 0.05$. Of note, rare subpopulations were more likely to yield a broader null distribution of CLQ values. A broader distribution increases the likelihood of categorizing the observed CLQ value as statistically insignificant. Only statistically significant colocalizations consistent between biological replicates per condition were included in downstream colocatome analysis (Fig. 2c).

The normalized CLQ is reported throughout the manuscript. We normalize the observed CLQ (defined above), in order to compare spatial features across different conditions and assays (ex: assembloids vs histopathological specimens). We observed that the distribution of CLQs under each condition generally followed a Gaussian-shaped distribution independent of region size (Supplementary Fig. 12d). Hence, for a given condition, we computed the normalized CLQ as the Z-score, calculated as the observed CLQ values minus (−) the mean of the observed CLQs under the given condition, divided by the standard deviation of the observed CLQ distribution for that condition. When the observed CLQ is 0, the normalized CLQ is -1; when the observed CLQ is the maximum value of the null distribution, the normalized CLQ is 1; and when the observed CLQ is the mean value of the null distribution (set here to the number of neighbors; 20), the normalized CLQ is near 0. Note that when the observed CLQs is greater (or smaller) than the maximum (or minimum) value of the CLQ distribution, it is in the 5% tail of the distribution and significant, and the observed CLQ is set to the maximum (or minimum) value of the CLQ distribution, so that the upper (or lower) bound of the normalized CLQ is 1 (or -1). Normalized CLQ values were calculated only for cell counts greater than 5 in any given sample, excluding irrelevant cell subpopulations with low counts. This was necessary because certain samples or regions did not express all cell types in significant amounts.

Additionally, CLQ calculations were repeated in the periphery and center regions of the same assembloid to evaluate whether specific spatial features were enriched in a particular region. To determine each region, the assembloid sections were considered as circular, and periphery and center zones were established with a consistent ratio to the diameter, ensuring uniform zone ratios regardless of the total area that can vary between samples (Supplementary Fig. 7). All CLQ values with their statistical significance are listed in Supplementary Data 2 and 4.

Composite colocatome of assembloids

To establish a comprehensive reference colocatome, we identified colocalizations consistent across biological replicates as described in Fig. 2 for each of the four assembloid conditions. Briefly, all negative colocalizations (cells repelling each other or segregated) were assigned to -1, and positive colocalizations (cells found in proximity) were assigned to 1. Unsignificant spatial features were assigned to 0. The final colocatome composite include all assembloid conditions (treatment-naïve TAF–PDO and TCF–PDO assembloids and erlotinib-treated TAF–PDO and TCF-PDO assembloids) and regions (whole assembloid, periphery and center). Cell pairs colocalized under one condition, but negatively colocalized under another were assigned to 1 in the final composite colocatome, as the reference colocatome intended to generate a matrix of possibilities of cells that can be found in proximity in some, but not necessarily all instances. Next, we used hierarchical clustering on the full (homo- and heterotypic spatial features) or partial (heterotypic spatial features) colocatome to group cell subpopulations into subclusters with similar colocalization partners. Then, we ordered the subclusters, namely colocation clusters, according to their likelihood of being found in proximity or segregated from each other.

LUAD clinical samples

To validate the clinical significance of our assembloid-derived colocatome, we transposed the cell subpopulations identified in the assembloids to a small LUAD cohort generated in our lab. Briefly, the same panel of markers for cancer cell and fibroblasts were used, but additional immune and endothelial markers were added. The full list of antibodies and reporter used are listed in Supplementary Data 7. The antibody conjugation and the staining were done according to the manufacturer's protocol (Akoya Biosciences) at the exception of an additional photobleaching step before the incubation with the primary antibody cocktail solution. Briefly, tissues fixed on coverslips were incubated in 4.5% H_2O_2 and 20 mM NaOH solution diluted in 1X PBS (5 ml/well) in a translucid vessel in between two white LED light panels (White light; Light intensity: 20,000 Lux) for 90 min at RT. Details regarding primary antibody, vendors, catalog numbers and barcode assignments are listed in the Supplementary Data 7. Data acquisition, software configuration and data were done as described above under the PhenoCycler image acquisition and data processing section.

The cell signature matrix developed for the assembloids was used to guide the identification of fibroblast and cancer cell subpopulations in LUAD specimens (Supplementary Data 1 and 3) of unknown cell composition and architecture. Lepidic, acinar and solid regions were defined by a pathologist, overlayed on the Phenocycler images using the EnableMedicine Beta visualizer tool, and cells were assigned to a histological growth pattern (lepidic, acinar or solid). The LUAD samples were then broken into histopathological regions associated with a growth pattern, on which we repeated the pipeline described above to reconstruct the cancer-fibroblast lepidic, acinar and solid colocatomes. Significant CLQ values consistent across more than half of the regions of a same growth pattern were considered as significant. If the majority of the normalized CLQ values for pairwise colocalization were negative, the interaction was considered avoidance, or vice versa for attraction. Pairwise interactions with equal number of regions with positive and negative CLQ values, which did not allow us to determine if cells were being attracted or repelled, were excluded. Hierarchical clustering was performed on the assembloid colocatome composite vs clinical specimens using R (version 4.3.1).

Hierarchical clustering on spatial colocalizations was also performed 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 to improve visualization. The significance of the clusters obtained was assessed using the pvclust R package[29], a tool for assessing the uncertainty in hierarchical cluster analysis using multiscale bootstrap resampling with the complete method, Euclidean distance, and 10,000 iterations for bootstrap resampling as input parameters. The approximate unbiased (AU) p-values were used as recommended in the user guide, and a p value of 0.05 was considered significant (Supplementary Fig. 13).

Breast cancer imaging mass cytometry dataset

The imaging mass cytometry dataset[6] comprising 49 breast cancer and 3 normal breast images was generously provided by Dr. Denis Schapiro from the Heidelberg University Hospital. The centroids of each cell were extracted from the segmentation masks provided with the images using MATLAB (R2023b). The colocalization analysis was repeated as described above, utilizing CLQ calculation and permutation analysis. Cell annotations were preserved from the original manuscript, and the number of neighbors was set to 5, consistent with the average number of neighbors per cell used in the original study. CLQ values with a $p < 0.05$ were considered as significant.

Statistical analysis

The significance of the CLQ values were obtained by randomly permuting 500 times the cell labels while preserving the subpopulation proportions. CLQ values falling outside or at the tail of the distribution generated by the permutation analysis were considered significant, whereas values within the distribution were deemed non-significant, as they can be reproduced after spatial randomization. In assembloids, CLQs consistently significant between biological replicates of a same condition were included, and for clinical specimens, spatial features significant across more than half of the regions of a same histology were considered as significant. The significance of the hierarchical clustering was assessed using multiscale bootstrap resampling with the pvclust R package[29]. $P < 0.05$ were considered as significant. See methods, results and figure captions for additional details about individual statistical analyses used in this study.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The assembloids imaging data generated in this study have been deposited on the EnableMedicine platform at https://app.enablemedicine.com/portal/atlas-library/studies/2e12f499-027b-4f06-813f-e16491ab34c1?sid=496, which allows both direct visualization and local download of the images. Due to the extremely large number of files and their substantial size, segmentation data, CELESTA cell annotations, and raw unprocessed images for all samples will be made available upon request. The publicly available mass cytometry breast data[6] used in this study is available through the Bodenmiller group Github at https://github.com/BodenmillerGroup/histoCAT?tab=readme-ov-file. Reagents and antibodies used are available in Supplementary Data 7 and Reporting Summary. Further information and requests for resources and reagents should be directed to, and will be fulfilled by, the lead contact, S.K.P.. This study did not generate new unique reagents or cell lines. The remaining data are available within the Article, Supplementary Information or Source Data file. Source data are provided with this paper.

Code availability

The R scripts used in this study are deposited on the Plevritis lab Github at https://github.com/plevritis-lab/Spatial_Permutation_and_Normalization.git and a permanent version is available at https://doi.org/10.5281/zenodo.13909851[43].

References

1. Moffitt, J. R., Lundberg, E. & Heyn, H. The emerging landscape of spatial profiling technologies. *Nat. Rev. Genet*. **23**, 741–759 (2022).
2. Schürch, C. M. et al. Coordinated cellular neighborhoods orchestrate antitumoral immunity at the colorectal cancer invasive front. *Cell* **182**, 1341–1359 (2020).
3. Giesen, C. et al. Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry. *Nat. Methods* **11**, 417–422 (2014).
4. He, S. et al. High-plex imaging of RNA and proteins at subcellular resolution in fixed tissue by spatial molecular imaging. *Nat. Biotechnol*. **40**, 1794–1806 (2022).
5. Palla, G., Fischer, D. S., Regev, A. & Theis, F. J. Spatial components of molecular tissue biology. *Nat. Biotechnol*. **40**, 308–318 (2022).
6. Schapiro, D. et al. HistoCAT: analysis of cell phenotypes and interactions in multiplex image cytometry data. *Nat. Methods* **14**, 873–876 (2017).
7. Zhang, W. et al. Identification of cell types in multiplexed in situ images by combining protein expression and spatial information using CELESTA. *Nat. Methods* **19**, 759–769 (2022).
8. Brbić, M. et al. Annotation of spatially resolved single-cell data with STELLAR. *Nat. Methods* **19**, 1411–1418 (2022).
9. Risom, T. et al. Transition to invasive breast cancer is associated with progressive changes in the structure and composition of tumor stroma. *Cell* **185**, 299–310.e18 (2022).
10. Yeh, C.Y. et al. Mapping spatial organization and genetic cell-state regulators to target immune evasion in ovarian cancer. *Nat. Immunol*. **25**, 1953–1958. https://doi.org/10.1038/s41590-024-01943-5 (2024).
11. Bose, S., Clevers, H. & Shen, X. Promises and challenges of organoid-guided precision medicine. *Med* **2**, 1011–1026 (2021).
12. Corsini, N. S. & Knoblich, J. A. Leading edge human organoids: new strategies and methods for analyzing human development and disease. *Cell* **185**, 2756–2769 (2022).
13. Liu, B., Li, Y. & Zhang, L. Analysis and visualization of spatial transcriptomic data. *Front. Genet*. **12**, 1–15 (2022).
14. Galligan, C. L. & Fish, E. N. The role of circulating fibrocytes in inflammation and autoimmunity. *J. Leukoc. Biol*. **93**, 45–50 (2013).
15. Reilkoff, R. A., Bucala, R. & Herzog, E. L. Fibrocytes: emerging effector cells in chronic inflammation. *Nat. Rev. Immunol*. **11**, 427–435 (2011).
16. Rees, M. & Cressie, N. *Statistics for Spatial Data*. 156. https://doi.org/10.2307/2982871 (1993).
17. Valkenburg, K. C., de Groot, A. E. & Pienta, K. J. Targeting the tumour stroma to improve cancer therapy. *Nat. Rev. Clin. Oncol*. 15, 336–381. https://doi.org/10.1038/s41571-018-0007-1 (2018).
18. De Jaeghere, E. A., Denys, H. G. & De Wever, O. Fibroblasts fuel immune escape in the tumor microenvironment. *Trends Cancer* **5**, 704–723 (2019).
19. Karacosta, L. G. et al. Mapping lung cancer epithelial-mesenchymal transition states and trajectories with single-cell resolution. *Nat. Commun*. **10**, 5587 (2019).
20. Ganthala, P. D. et al. Co-encapsulated nanoparticles of Erlotinib and Quercetin for targeting lung cancer through nuclear EGFR and PI3K/AKT inhibition. *Colloids Surf. B Biointerfaces* **211**, 112305 (2022).
21. Lin, L., Li, L., Chen, X., Zeng, B. & Lin, T. Preliminary evaluation of the potential role of β-elemene in reversing erlotinib-resistant human NSCLC A549/ER cells. *Oncol. Lett*. **16**, 3380–3388 (2018).
22. Jin, W. et al. Muc1 induces acquired chemoresistance by upregulating abcb1 in egfr-dependent manner. *Cell Death Dis*. **8**, e2980–13 (2017).
23. Shoucair, I., Mello, F. W., Jabalee, J., Maleki, S. & Garnis, C. The role of cancer-associated fibroblasts and extracellular vesicles in tumorigenesis. *Int J. Mol. Sci*. **21**, 1–37 (2020).
24. Gascard, P. & Tlsty, T. D. Carcinoma-associated fibroblasts: Orchestrating the composition of malignancy. *Genes Dev*. **30**, 1002–1019 (2016).
25. Pradhan, R. N., Krishnamurty, A. T., Fletcher, A. L., Turley, S. J. & Müller, S. A bird's eye view of fibroblast heterogeneity: A pan-disease, pan-cancer perspective. *Immunol. Rev*. **302**, 1–22 (2021).
26. Russell, P. A. et al. Does lung adenocarcinoma subtype predict patient survival?: A clinicopathologic study based on the new international association for the study of lung cancer/American thoracic society/European respiratory society international multidisciplinary lung adeno. *J. Thorac. Oncol*. **6**, 1496–1504 (2011).
27. Solis, L. M. et al. Histologic patterns and molecular characteristics of lung adenocarcinoma associated with clinical outcome. *Cancer* **118**, 2889–2899 (2012).
28. Bouchard, G. 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*. **82**, 648–664 (2021).
29. Suzuki, R. & Shimodaira, H. Pvclust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* **22**, 1540–1542 (2006).

30. Zhou, X. et al. Circuit design features of a stable two-cell system. *Cell* **172**, 744–757.e17 (2018).
31. Buechler, M. B., Fu, W. & Turley, S. J. Fibroblast-macrophage reciprocal interactions in health, fibrosis, and cancer. *Immunity* **54**, 903–915 (2021).
32. Angarola, B. L. et al. Comprehensive single-cell aging atlas of healthy mammary tissues reveals shared epigenomic and transcriptomic signatures of aging and cancer. *Nat. Aging* https://doi.org/10.1038/s43587-024-00751-8 (2024).
33. Ben-chetrit, N. et al. Breast cancer macrophage heterogeneity and self-renewal are determined by spatial localization. Preprint at *bioRxiv* https://doi.org/10.1101/2023.10.24.563749 (2023).
34. Chen, Y. I. et al. Homophilic ATP1A1 binding induces activin A secretion to promote EMT of tumor cells and myofibroblast activation. *Nat. Commun*. **13**. https://doi.org/10.1038/s41467-022-30638-4 (2022).
35. Kharbanda, A. et al. Targeting the oncogenic MUC1-C protein inhibits mutant EGFR-Mediated signaling and survival in non-small cell lung cancer cells. https://doi.org/10.1158/1078-0432.CCR-13-3168.
36. Yu, A. et al. Reconstructing codependent cellular cross-talk in lung adenocarcinoma using REMI. *Sci. Adv*. **8**. https://doi.org/10.1126/sciadv.abi4757 (2022).
37. Yan, Y. et al. Identification of CAV1 and DCN as potential predictive biomarkers for lung adenocarcinoma. *Am. J. Physiol. Lung Cell Mol. Physiol.* **316**, L630–L643 (2019).
38. Shi, Y. B. et al. Multifaceted roles of caveolin-1 in lung cancer: a new investigation focused on tumor occurrence, development and therapy. *Cancers* **12**, 1–19 (2020).
39. Zhan, P. et al. Expression of caveolin-1 is correlated with disease stage and survival in lung adenocarcinomas. *Oncol. Rep.* **27**, 1072–1078 (2012).
40. Sunaga, N. et al. Different roles for caveolin-1 in the development of non-small cell lung cancer versus small cell lung cancer. *Cancer Res.* **64**, 4277–4285 (2004).
41. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
42. Leslie, T. F. & Kronenfeld, B. J. The colocation quotient: a new measure of spatial association between categorical subsets of points. *Geogr. Anal.* **43**, 306–326 (2011).
43. Bouchard, G. et al. A quantitative spatial cell-cell colocalizations framework enabling comparisons between in vitro assembloids and pathological specimens. Zenodo https://doi.org/10.5281/zenodo.13909851 (2024).
44. Kilvaer, T. K. et al. Cancer associated fibroblasts in stage I-IIIA NSCLC: Prognostic impact and their correlations with tumor molecular markers. *PLoS ONE* **10**, 1–15 (2015).
45. Fitzgerald, A. A. & Weiner, L. M. The role of fibroblast activation protein in health and malignancy. 39. https://doi.org/10.1007/s10555-020-09909-3 (2020).
46. Kilvaer, T. K. et al. Differential prognostic impact of platelet-derived growth factor receptor expression in NSCLC. *Sci. Rep.* **9**, 1–11 (2019).
47. Hewitt, K. J. et al. PDGFRβ expression and function in fibroblasts derived from pluripotent cells is linked to DNA demethylation. *J. Cell Sci.* **125**, 2276–2287 (2012).
48. Schliekelman, M. J. et al. Thy-1+ cancer-associated fibroblasts adversely impact lung cancer prognosis. *Sci. Rep.* **7**, 1–12 (2017).
49. 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.* **11**, 5607–5619 (2018).
50. Ye, Z. 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* **11**, 1–17 (2016).
51. Luo, H. T., Liang, C. X., Luo, R. C. & Gu, W. G. Identification of relevant prognostic values of cytokeratin 20 and cytokeratin 7 expressions in lung cancer. *Biosci. Rep.* **37**, 1–12 (2017).
52. Gharib, T. G. et al. Proteomic analysis of cytokeratin isoforms uncovers association with survival in lung adenocarcinoma. *Neoplasia* **4**, 440–448 (2002).
53. Huang, X. et al. MUC1 overexpression predicts worse survival in patients with non-small cell lung cancer: Evidence from an updated meta-analysis. *Oncotarget* **8**, 90315–90326 (2017).
54. Lan, Y., Ni, W. & Tai, G. Expression of MUC1 in different tumours and its clinical significance (Review). *Mol. Clin. Oncol.* **17**, 1–10 (2022).
55. Xu, X. et al. Muc1 knockout potentiates murine lung carcinogenesis involving an epiregulin-mediated EGFR activation feedback loop. *Carcinogenesis* **38**, 604–614 (2017).
56. Baeuerle, P. A. & Gires, O. EpCAM (CD326) finding its role in cancer. *Br. J. Cancer* **96**, 417–423 (2007).
57. Kim, Y. et al. Clinicopathological implications of EpCAM expression in adenocarcinoma of the lung. *Anticancer Res.* **29**, 1817–1822 (2009).
58. Pak, M. G., Shin, D. H., Lee, C. H. & Lee, M. K. Significance of EpCAM and TROP2 expression in non-small cell lung cancer. *World J. Surg. Oncol.* **10**, 1–8 (2012).

Acknowledgements

S.K.P. discloses support for the publication of this study from National Institute of Health, National Cancer Institute [R25CA180993, U54CA274511]. G.B. discloses support for the research described in this study from the National Institute of Health, National Cancer Institute [K99CA255586] and Les Fonds de Recherche du Québec – Santé [35603 and 267646]. The authors thank Enable Medicine for providing visualization and analytical tools, as well as technical support for mIF images. The authors also thank Dr. Schapiro for graciously providing assistance with the human breast cancer mass cytometry imaging dataset. Lastly, the authors would like to acknowledge editorial contributions from Alison F. Davis.

Author contributions

Study Conception & Design: G.B. and S.K.P.; Performed Experiment and Data Collection: G.B., I.L., A.B., and Y.L.; Performed Data Analysis: G.B., W.Z., I.L, I.I.; Interpretation of data analysis: G.B., W.Z. I.L., I.I, L.T, M.G.O., A.J.G., and S.K.P.; Writing the first draft: G.B. and S.K.P.; Figures Design: G.B; Patient Sample Management: A.B., C.K., W.T., and J.B.S; and Supervision: J.B.S., A.J.G., L.T. and S.K.P. All authors contributed to manuscript editing and revision.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-024-55129-6.

Correspondence and requests for materials should be addressed to Sylvia K. Plevritis.

Peer review information *Nature Communications* thanks Christopher Hanley, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.