

<https://doi.org/10.1038/s42003-025-08799-x>

Cancer-associated fibroblasts shape the formation of budding cancer cells at the invasive front of human colorectal cancer



Juan L. García-Rodríguez¹✉, Ulrik Korsgaard^{2,3,4}, Stine M. Vissing¹, Thea Petersen Paasch¹, Mariana Semenova¹, Simon L. Vendelbo¹, Eva F. Jensby^{5,6}, Hannah L. Williams⁷, Paul Vinu Salachan^{5,6}, Camilla Blunk Brandt^{1,8}, Jacob Hanimann⁷, Lin Lin^{1,8}, Inti Zlobec⁷, Karina D. Sørensen^{5,6}, Jørgen Kjems^{9,10}, Henrik Hager^{4,5} & Lasse S. Kristensen¹✉

The formation of budding cancer cells at the invasive front of solid tumors is one of the first steps of metastasis. However, this process is still incompletely elucidated. Here, we used spatial molecular imaging to disentangle the complex interactions between cancer cells and the tumor microenvironment at the invasive front of colorectal tumors. Employing a 1000-plex gene panel, we defined all major cell types in tumors and adjacent normal tissue with accurate spatial information. Individual cancer cell clusters were located together, consistent with an expected mutation- and epigenetic-driven clonal evolution. However, cancer cell clusters encompassing budding cells exhibited a markedly different spatial distribution as they also contained cells that were scattered around the periphery of the main cancer cell masses. Moreover, these cells were frequently in contact with cancer-associated fibroblasts (CAFs) and underwent broad gene expression changes, mainly related to epithelial-mesenchymal transition (EMT), remodeling of the extracellular matrix (ECM), and migration. In addition, we defined an 11-gene signature (*TYK2*, *IL2RG*, *KRT17*, *HLA-B*, *NPPC*, *WIF1*, *IL32*, *B2M*, *CCND1*, *CRIP1*, *ITGB1*), which characterizes cancer cells en route to metastasis and is associated with inferior outcomes. Collectively, our findings suggest that CAFs induce pro-invasive gene expression changes involved in EMT, ECM remodeling, and migration.

The development of distant metastases accounts for most cancer-related deaths as surgical removal of metastatic lesions is often not possible or warranted due to a wider systemic spread of cancer cells in late-stage disease¹. The metastatic process consists of several discrete steps, including local invasion, entry into lymphatic and/or blood vessels, extravasation into another tissue and outgrowth at the distant site². In several solid tumors, including colorectal cancer, the sprouting of cancer cells at the invasive front (or tumor budding) of early-stage primary tumors is associated with disease progression³. Budding cells are usually defined as single cancer cells or small clusters of up to four cancer cells that are isolated from the main part of the tumor and their presence in patients with stage II colon cancer is increasingly being considered as a

poor prognostic marker that may, along with other high-risk factors, justify the consideration of adjuvant chemotherapy⁴.

Budding cells are hypothesized to undergo a process known as epithelial-mesenchymal transition (EMT), which is characterized by increased expression of mesenchymal markers like vimentin and decreased expression of epithelial markers. During EMT a loss or downregulation of cell-cell adhesion molecules like E-cadherin and EpCAM together with cytoskeletal rearrangements, leads to increased cell motility and tissue invasion^{5,6}. In addition, budding cells often have an increased expression of stem cell markers, gain the capacity to degrade the extracellular matrix, and become more resistant to apoptosis⁷. Cancer associated fibroblasts (CAFs) within the tumor microenvironment

¹Department of Biomedicine, Aarhus University, Aarhus, Denmark. ²Department of Clinical Pathology, Lillebaelt Hospital, University Hospital of Southern Denmark, Vejle, Denmark. ³Danish Colorectal Cancer Center South, Lillebaelt Hospital, University Hospital of Southern Denmark, Vejle, Denmark. ⁴Department of Pathology, Aarhus University Hospital, Aarhus, Denmark. ⁵Department of Molecular Medicine, Aarhus University Hospital (AUH), Aarhus, Denmark. ⁶Department of Clinical Medicine, Aarhus University, Aarhus, Denmark. ⁷Institute of Tissue Medicine and Pathology, University of Bern, Bern, Switzerland. ⁸Steno Diabetes Center Aarhus, Aarhus University Hospital, Aarhus, Denmark. ⁹Department of Molecular Biology and Genetics (MBG), Aarhus University, Aarhus, Denmark. ¹⁰Interdisciplinary Nanoscience Center (iNANO), Aarhus University, Aarhus, Denmark. ✉e-mail: juanlgr@biomed.au.dk; lasse@biomed.au.dk

(TME) are thought to promote tumor progression and metastasis^{8,9}, whereas CD68 positive macrophages and CD8 positive T cells may kill off the invading cancer cells¹⁰. The physical properties of the TME may also play a role in the formation of budding cells^{11–13}.

The signals from the CAFs and other cells within the TME that are hypothesized to promote budding as well as the proposed immune escape mechanisms of the budding cells are still incompletely understood and challenging to study. However, novel spatial proteomics and transcriptomics technologies have recently emerged and promise to revolutionize our understanding of tumor budding and cancer metastasis.

Here, we employed a technology called Spatial Molecular Imaging (SMI), which simultaneously measures a thousand different RNA molecules, to study the invasive front of colorectal adenocarcinomas. We demonstrate the robustness of the SMI technology for the profiling of transcriptomic changes in individual cells, including CAFs, immune cells, endothelial cells, and budding cancer cells, leading to novel discoveries of cancer biology. In particular, our study is focused on the spatial architecture of different cancer cell clusters and cells within the TME, emphasizing the crosstalk between cells in the TME and budding cancer cells at the invasive front.

Results

We employed four different spatial technologies to study the invasive front of colorectal adenocarcinomas. Our main focus was on SMI analyses, for which samples from four patients (four adenocarcinomas and one adjacent normal tissue) were included. Subsequently, validation studies were done using the 10X Visium platform (adenocarcinoma tissues from five patients), nanostring GeoMx (adenocarcinoma tissues from three patients), and RNAscope duplex analyses, including tissues from six patients with adenocarcinoma and three patients with adenoma (Supplementary Fig. 1).

Accurate spatial gene expression analysis of single cells in adjacent normal colon tissues

We initially assessed the performance of the SMI technology on colon tissues by examining structurally ordered adjacent normal tissue. In such tissues, it is well established that epithelial cells, including goblet cells, stem cells, and enterocytes, are situated in the crypts and surrounded by fibroblasts and immune cells. First, we visualized the crypts by immunofluorescent staining for pan-cytokeratin, which stains epithelial cells, and stained for T-cells using CD3 (Supplementary Fig. 2A). Then, from the same area within the same tissue slide, we identified a total of 3108 cells using SMI and performed Uniform Manifold Approximation and Projection (UMAP) analysis to visualize cell distribution patterns based on gene-expression, and unsupervised clustering using the Leiden algorithm¹⁴ to identify individual cell clusters. This analysis revealed nine different major cell clusters, including immune cells and fibroblasts as well as cells of epithelial origin, namely enterocytes, goblet, crypt, and tuft cells based on the PanglaoDB database¹⁵. Additionally, we identified a cluster comprising only 57 cells, some of which were located within the crypts, while others were found in the stromal area. This cluster was labeled as “unknown”. Overall, cells of epithelial origin clustered separately from other cell types (Supplementary Fig. 2B), and when inspecting their physical localization (Supplementary Fig. 2C), these cells were found within the crypts, strictly aligned with pan-cytokeratin staining (Supplementary Fig. 2D). The remaining cell types were located in the stroma between the crypts as expected, and T-cells were present in areas of positive CD3 staining (Supplementary Fig. 2E). Among the most differentially expressed genes between the different cell types, we identified many well-known canonical markers, e.g., *EPCAM*, *OLFM4*, and *KRT19* for epithelial cells, several collagens including *COL1A2* as well as *ACTA2* for fibroblasts, *JCHAIN*, *MZB1*, and *CD27* for B-cells and plasma cells, and *CD3E* for T-cells (Supplementary Fig. 2F). Visualization of individual epithelial marker genes illustrated a high specificity of the signal to pan-cytokeratin positive cells (Supplementary Fig. 3). Furthermore, we found that the *PIGR* gene was highly expressed and specific for the epithelial cells, in line with previous research¹⁶ (Supplementary Fig. 2G–J). To further

validate the specificity of the epithelial expression of this marker, we investigated a digital spatial profiling (DSP) dataset for which epithelial cells and immune cells from normal adjacent colon tissues were analyzed separately¹⁷. In accordance with the SMI data, we found a high expression of *PIGR* in the epithelial cell fractions (Supplementary Fig. 2K, L). We also used a single molecule fluorescent in situ hybridization (FISH) duplex assay to further validate the spatial expression patterns of *PIGR* and *OLFM4*. Here, it could be observed that *OLFM4* is mainly located at the base of the crypts as expected¹⁸, whereas *PIGR* is more abundant in the upper parts of the crypts (Supplementary Fig. 2M). Finally, further analysis of the “Unknown cluster” led to the identification of three subclusters of cells (Supplementary Fig. 4), with those located in the stroma showing the lowest expression of *PIGR*, some of which may represent pericryptal fibroblasts-like cells (*PDGFRA* and *CXCL12* high) (Supplementary Fig. 4). Together, based on these analyses, we conclude that SMI enables a highly accurate spatial gene expression analysis of single cells in adjacent normal colon tissue.

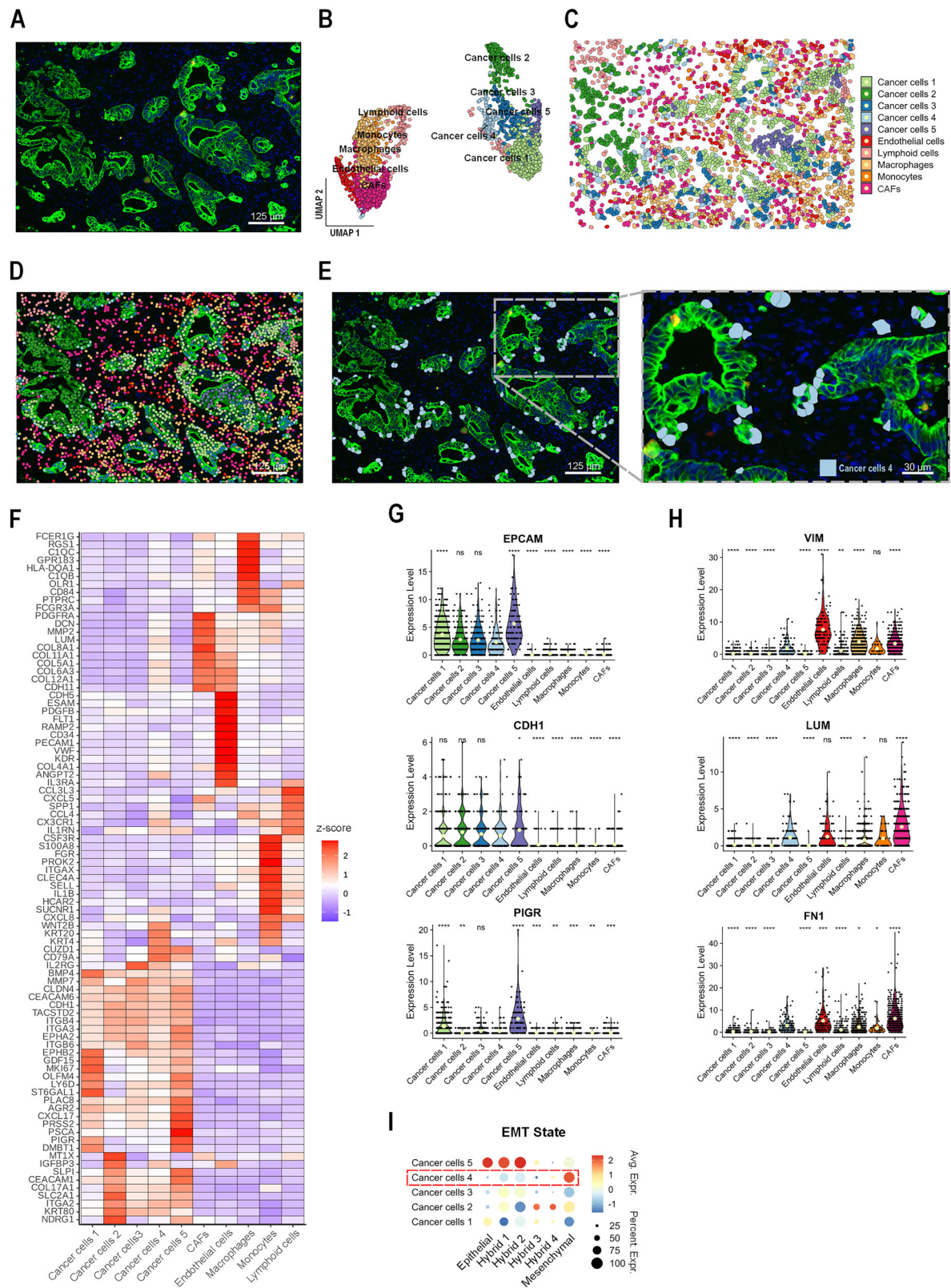
Spatial gene expression analyses of single cells in colorectal tumors reveal a unique spatial distribution of cell clusters that encompass budding cells

Next, we used SMI to study colorectal adenocarcinomas from four different patients, focusing on areas at the invasive fronts with different densities of tumor budding. Further details about the patients can be found in Supplementary Table 1. In the first FOV from patient 1, we visualized the cancer cells using immunofluorescent staining for pan-cytokeratin, revealing both larger areas of cancer cells and budding cells (Fig. 1A). We identified a total of 2541 cells and performed UMAP visualization and Leiden clustering analysis employing the same methodology as previously described for the adjacent normal tissue. This resulted in the identification of nine main cell clusters that were annotated based on enrichment of well-known canonical markers and confirmed with the PanglaoDB database. In the UMAP plot, all the different cancer cell clusters formed a separate supercluster, whereas the immune cells, CAFs, and endothelial cells clustered separately (Fig. 1B). When inspecting the localization of the cells, it was evident that cancer cells colocalized with cells being positive for the pan-cytokeratin immunofluorescent signal, as would be expected¹⁹ (Fig. 1C, D). Moreover, most individual cancer cells belonging to a distinct cluster were in close physical contact, indicating that gene expression patterns are a good predictor of spatial distribution. In contrast, the cells from one of the cancer cell clusters (denoted as Cancer cells 4) appeared to be randomly distributed throughout the tissue. Interestingly, this cluster was found to encompass most of the budding cells, and upon closer inspection, it became apparent that individual cells were always found at the outskirts of masses of cancer cells belonging to other clusters. Thus, all these cells were in close contact with cells within the TME (Fig. 1E). This unique spatial distribution of cancer cell clusters encompassing most of the budding cells was also observed for other FOVs with different densities of tumor budding (Supplementary Figs. 5 and 6). Together these analyses show that the distinct cancer cell clusters that encompass most of the budding cells have a unique spatial distribution.

Cancer cell clusters that encompass budding cells share changes related to EMT and migration

It has been hypothesized that budding cells undergo EMT, which is a cellular transdifferentiation process from an epithelial to a mesenchymal phenotype. These cells lose polarity and adhesion properties while acquiring migratory abilities. Indeed, the “Cancer cells 4” cluster from FOV1 possessed all these features, showing decreased expression of epithelial marker genes such as *EPCAM*, *CDH1*, and *PIGR* in parallel with an increased expression of mesenchymal marker genes such as *LUM*, *VIM*, and several collagen genes. In addition, an increased expression of *FN1* and *WNT2B* was observed (Fig. 1F, G, H).

EMT is known to be a continuous process, however, some studies have defined discrete intermediate hybrid states^{20–22}. Thus, we further investigated



the cancer clusters and confirmed the mesenchymal EMT state of the “Cancer cells 4” cluster. On the contrary, the other cancer cell clusters were in epithelial-like states. Specifically, “Cancer cells 1” scored highest for “Epithelial state”, and “Cancer cells 5” for “Epithelial state” as well as for Hybrid (H) stages 1 and 2. “Cancer cells 3” scored higher for H1 and H2 and

“Cancer cells 2” scored higher for H3 and H4 (Fig. 1I). This also applied to the cancer cell clusters encompassing most of the budding cells in other FOVs with different tumor budding densities (Supplementary Fig. 7). Together, these analyses show that the distinct cancer cell clusters that encompass most of the budding cells are undergoing EMT.

Fig. 1 | Spatial gene expression analyses of single cells in colorectal tumors reveal a unique spatial distribution of cell clusters that encompass budding cells.

A Multichannel immunofluorescent (IF) overlay of epithelial (PancK), immune (CD3) and nuclei (DAPI) immunofluorescent markers in green, yellow, and blue, respectively, from a field of view (FOV) containing colorectal adenocarcinoma tissue. Red colorings represent the fiducial beads. **B** UMAP projection showing the cell type composition of the tissue. **C** Spatial plot showing the polygonal shape, type, and location of each cell from the colorectal adenocarcinoma tissue. **D** Spatial plot showing the cellular overlay between the multichannel IF image and each cell detected, represented by a point at each cell centroid and color-labelled by cell type. **E** Spatial plot showing the overlay of the IF image with the cluster “Cancer cells 4” (light blue), showing a dispersed distribution across the tissue with cells located

mainly in the outskirts of the tumor masses and in tumor buds. **F** Heatmap of the top 10 gene markers specific for each cluster. **G** Violin plots showing the expression levels of the epithelial markers *EPCAM*, *CDH1*, and *PIGR* for each cluster. Each dot corresponds to a single cell. Significance was determined by unpaired Student’s T tests with “Cancer cells 4” as the reference cluster. **H** Violin plots showing the expression levels of mesenchymal markers *VIM*, *LUM*, and *FN1* for each cluster. Each dot corresponds to a single cell. Significance was determined by unpaired Student’s T tests with “Cancer cells 4” as the reference cluster. **I** Bubble plot showing the scored average expression (blue to red) and the percentage of expression (dot size) per cluster for each EMT state. The red scattered box highlights the “Cancer cells 4” cluster. ns (non-significant), * (*P*-value < 0.05), ** (*P*-value < 0.01), *** (*P*-value < 0.001), **** (*P*-value < 0.0001).

A gene signature of early budding is associated with disease free survival

To define the molecular features of early budding cells, we studied the “Cancer cells 4” cluster in FOV1 further. We extracted a list of genes that showed higher expression levels relative to all the other cancer cell clusters (61 genes) and relative to the CAF cluster (206 genes). From these two lists, 11 genes (*TYK2*, *IL2RG*, *KRT17*, *HLA-B*, *NPPC*, *WIF1*, *IL32*, *B2M*, *CCND1*, *CRIP1*, *ITGB1*) were shared (Fig. 2A). When plotting these 11 genes for all the individual cell clusters, we observed that most were highly enriched in “Cancer cells 4”, in terms of average expression and the percentage of cells expressing these genes (Fig. 2B and Supplementary Fig. 8). When inspecting other FOVs with varying tumor budding densities, we also observed a strong enrichment of this gene signature in the cancer cell clusters encompassing most budding cells (Supplementary Fig. 9). To further investigate the spatial expression patterns of these 11 genes, we explored a DSP dataset for which epithelial cells from adjacent normal tissues as well as cancer cells from the invasive front and central part of colon tumors, along with liver metastases were analyzed separately¹⁷. First, when comparing the global gene expression patterns of cancer cells from the invasive fronts with those from the central parts of the tumors, we detected 616 differentially expressed genes. Interestingly, *PIGR* and *OLFM4* were the most downregulated genes at the invasive front relative to the tumor mass (Supplementary Fig. 10A), consistent with these markers being lost in the cancer cell clusters that encompass most of the budding cells. We also found that many of the differentially expressed genes are involved in wound healing, angiogenesis, cell migration, and adhesion (Supplementary Fig. 10B, C). In this dataset, nine of the 11 signature genes were detected above background levels, and seven of these showed a higher expression at the invasive fronts relative to the central parts of the tumors. Additionally, in most cases, these genes also exhibited higher expression at the invasive fronts relative to the paired liver metastasis, which is expected since the cells no longer require a migratory phenotype at this point (Supplementary Fig. 10D). Interestingly, the average expression of these genes was higher in colon adenocarcinomas compared to adjacent normal tissues in a TCGA dataset (Fig. 2C) and significantly associated with a poor disease-free survival (Fig. 2D). When analyzing the genes individually, *CRIP1*, *IL2RG*, and *ITGB1* were significantly associated with poor disease-free survival (Supplementary Fig. 11). Next, to further substantiate the specificity of the 11-gene signature and its involvement in cancer progression, we explored a single-cell RNA sequencing dataset consisting of 371,223 cells from colorectal tumors and adjacent normal tissues²³ (Fig. 2E). Again, we found a higher expression of the gene-signature in cancer cells undergoing EMT and higher expression in the more advanced tumors (T3 and T4) compared to early-stage tumors (T1 and T2) and compared to epithelial cells from adjacent normal tissues (Fig. 2F, G). Thereafter, we analyzed the expression of these genes at the protein level using the human protein atlas (proteinatlas.org) and overall observed higher staining intensities at invasive front areas and in budding cells (Fig. 2H). Finally, to assess the role of one of the 11 genes in cell migration in a controlled setup, we established a knockdown of *ITGB1* in two different human colorectal cell lines (Caco-2 and HCT116) (Supplementary Fig. 12A, B). Using a wound healing assay, the Caco-2 cell line exhibited a clear

dependency of *ITGB1* for migratory potential (Supplementary Fig. 12C and D), whereas HCT116 cells were unaffected (Supplementary Fig. 12E and F). On the other hand, both cell lines showed a significant reduction in growth 72 h after *ITGB1* silencing (Supplementary Fig. 12G), whereas cell viability was unaffected (Supplementary Fig. 12H).

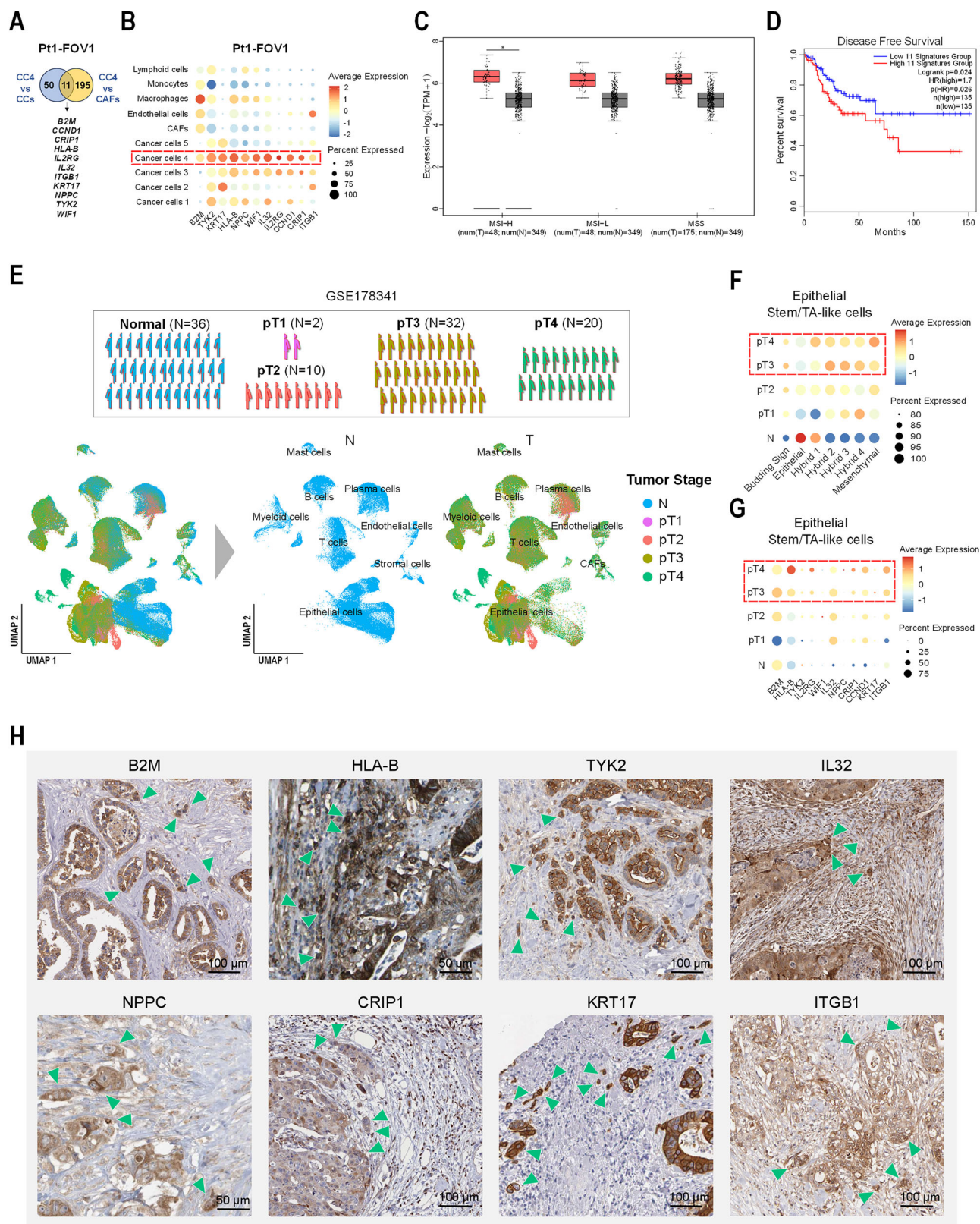
Individual FOVs from high, mid, and low-density budding areas share upregulated and downregulated genes

Next, to further investigate the gene expression changes taking place in the cancer cell clusters that encompass most of the budding cells, we performed differential gene expression analysis relative to all the other cancer cells within each individual FOV (Fig. 3). In two FOVs, which we annotated as high-density budding (HDB) (Fig. 3, upper panel), there was a high similarity in the transcriptomic signatures, with 129 genes shared among the significantly upregulated ones. Indeed, of the top 10 most upregulated, seven genes were shared (*TYK2*, *COL1A1*, *COL3A1*, *FN1*, *KRT8*, *MZT2A*, and *MT2A*), with *TYK2* and *COL1A1* being the two most upregulated in both FOVs.

In six FOVs, which we annotated as mid-density budding (MDB) (Fig. 3, middle panel), most of the commonly upregulated genes belonged to the collagen family. Specifically, *COL5A1*, *COL6A3*, *COL6A2*, *COL12A1* were significantly upregulated in all the FOVs, while *COL1A1*, *COL1A2*, *COL3A1*, *COL4A2*, *COL11A1*, *COL5A2* and *COL6A1* were found to be significantly upregulated in at least five of these FOVs. No genes were found to be significantly downregulated in all FOVs. However, *PIGR* and *OLFM4* were downregulated in four of the six FOVs. These findings were further substantiated by performing spatial whole transcriptome profiling (Visium 10X Genomics), which showed a stronger signal in the tumor masses, but also variation from patient to patient (Supplementary Fig. 13). Using the single molecule in situ hybridization duplex assay, we also found a large degree of inter-patient heterogeneity across colon adenocarcinoma samples, whereas colon adenoma samples stained positive for both genes (Supplementary Fig. 14). Finally, we explored previously published RNA-sequencing data²⁴, which confirmed the inter-patient heterogeneity across stage II and III colon adenocarcinomas (Supplementary Fig. 15).

In six FOVs, which we annotated as low-density budding (LDB) (Fig. 3, lower panel), we observed an upregulation of several genes related to the extracellular matrix (ECM) and migration. Specifically, *LGALS1*, *IGFBP7*, *VIM*, *COL9A2*, and *WIF1* were upregulated in all FOVs. Again, many collagen family genes were upregulated for most FOVs, along with several human leukocyte antigen (HLA) genes, including, *HLA-DPA1*, *HLA-DRB1*, *HLA-DRB5*, *HLA-B*, *HLA-C*. Regarding the early budding signature, seven of the 11 genes (*IL2RG*, *TYK2*, *B2M*, *HLA-B*, *IL32*, *NPPC* and *WIF1*) were upregulated in four of these FOVs, three genes (*NPPC*, *IL32* and *HLA-B*) in five FOVs and, finally, *WIF1* was found to be significantly upregulated in all six FOVs.

In summary, these comparisons revealed shared gene expression changes among the cancer cell clusters that contained most of the budding cells, marked by the downregulation of epithelial markers (including *OLFM4* and *PIGR*) and the upregulation of mesenchymal genes (collagens and migration-related genes).



Gene-set enrichment analysis of high, mid, and low-density budding areas revealed a progressive adaptation of the budding cells to the surrounding stroma

As described above, 129 genes were commonly upregulated in the HDB areas (Fig. 4A). Interestingly, enrichment analysis of these genes revealed molecular functions related to “extracellular matrix constituents that confer

tensile strength” and “integrin and collagen binding”, highlighting the active interactions of these cells with the surrounding stroma, likely affecting matrix stiffness (Fig. 4B). In parallel, we found many top hits related to “activation of developmental processes” which was also in line with the top significantly enriched pathways according to the WikiPathways²⁵ and Reactome²⁶ databases. For example, “TGF beta”, and “IL-4 and IL-13

Fig. 2 | A gene signature of early budding is associated with disease free survival. **A** Venn diagram showing the unique and common genes being upregulated when comparing the “Cancer cells 4” cluster (CC4) and all the other cancer cell clusters (CCs), and when comparing CC4 and the CAFs cluster. **B** Dot plot showing the average expression and percentage of expression of each of the 11 genes that were upregulated in both comparisons in A. The red scattered box highlights the “Cancer cells 4” cluster. **C** Box plots showing the interquartile range of the average expression data for the 11 genes identified in A, extracted from TCGA (COAD). The horizontal line in the box indicates median value and the whiskers indicate 10th and 90th percentiles. Colorectal adenocarcinoma (T), separated by Microsatellite instability status: MSI high (MSI-H), MSI low (MSI-L), and Microsatellite stable (MSS) were compared to adjacent normal tissue (N). Significance was determined by one-way ANOVA. **D** Kaplan-Meier plot showing the disease-free survival based on the average expression levels of the 11 genes identified in A. The two groups were defined based on the median. Significance was determined using a Mantel-Cox test. ns (non-significant), * (*P*-value < 0.05), ** (*P*-value < 0.01), *** (*P*-value < 0.001), **** (*P*-

value < 0.0001). **E** Schematic of the patient cohort derived from Pelka et al. single-cell RNA-seq dataset (GSE178341) (Upper panel). UMAP plots showing the integrated single-cell data, stratified by tissue type: adjacent normal colon (N) and tumor (T) tissues. Cells are colored by tumor stage (pT1-4) and annotated with major cell type labels (Lower panel). **F** Bubble plot showing the scored average expression (blue to red) and the percentage of expression (dot size) per cluster for the Budding Signature (11 genes found in A) and each EMT state for the epithelial subcluster Stem/transition amplifying (TA)-like cells. The red scattered box highlights pT3 and pT4 stages. **G** Bubble plot showing the scored average expression (blue to red) and the percentage of expression (dot size) per cluster for the each of the 11 genes found in A from the epithelial subcluster Stem/transition amplifying (TA)-like cells. The red scattered box highlights pT3 and pT4 stages. **H** IHC on human colorectal cancer tissue sections extracted from the Protein Atlas database showing the protein staining corresponding to 10 of the 11 genes identified in A. Green arrows highlight invasive front areas and budding cells.

signaling” were two of the top terms, both of which have been described to contribute to the generation of supportive niches for metastatic dissemination²⁷. Similarly, “TROP2” and “CKAP4 signaling pathways” were among the most upregulated pathways in these cells (Fig. 4B), which are also known to play pro-metastatic roles^{28,29}.

Similar analyses of 36 commonly upregulated genes in the MDB areas revealed many of the same molecular functions as found in the HDB areas (Fig. 4A). Additionally, the “PDGF pathway” was found, suggesting a key role of PDGF signaling in CAF activation and ECM remodeling towards the generation of a pro-migratory niche. The most significant pathway from the Reactome database was “collagen degradation” and the top biological processes were related to “endoderm formation”, “gastrulation”, and “ECM organization” (Fig. 4B).

Finally, 79 commonly upregulated genes in the LDB areas revealed activation of an immune defensive reaction as indicated by the molecular functions related to the MHC class II protein complex (Fig. 4A, B). The top biological processes were also related to immune system, and the most significant pathway from the Reactome database was “ECM proteoglycans”, a term that was also found for the HDB and MDB areas. Among the most enriched pathways according to WikiPathways were “burn wound healing”, which also appeared for the HDB and MDB areas.

Cancer cells from the budding clusters are frequently in contact with CAFs

Having observed that distinct cancer cell clusters that encompass most of the budding cells show a distinct spatial distribution with individual cells being in contact with cells within the TME, we investigated the nature of their neighboring cells. We found that these budding cancer cells were more often in contact with CAFs compared to the cells from any of the other cancer cell clusters (Fig. 5). Moreover, among the different cell types within the TME, the CAFs were the most frequent contact partner in nine of the 14 FOVs. In the remaining five FOVs, T-cells were the most frequent contact partner in two, macrophages in two, and basal cells in one (Fig. 5).

Cancer cells in contact with CAFs alter the ECM composition to facilitate migration and avoid apoptosis

To gain a better understanding of the role of the CAFs, we compared the gene expression profiles of cancer cells in contact with CAFs with those not in contact for each of the cancer cell clusters that encompass most of the budding cells. For these analyses, three FOVs with very few cancer cells being in contact with CAFs were omitted to ensure robust analyses.

Since the gene set enrichment analyses indicated extensive changes in the ECM organization, we hypothesized that this could be influenced by interactions between cancer cells and CAFs. Indeed, cancer cells in contact with CAFs exhibited higher scores for ECM organization compared to those not in contact in 10 out of 11 FOVs, with the highest levels observed for the HDB areas (Fig. 6A). Likewise, this was observed when analyzing the collagen ECM score (Fig. 6B). Additionally, we noted higher focal adhesion

levels in the cancer cells in contact with CAFs in 8 out of 11 FOVs, with the highest levels observed in the HDB areas (Fig. 6C). Moreover, we observed a downregulation of the apoptotic process in 9 out of 11 FOVs, specifically for cancer cells in contact with CAFs (Fig. 6D).

Finally, the previous gene set enrichment analysis highlighted “miRNA targets in ECM and membrane receptors” as a term commonly activated in the cancer cell clusters that encompass most of the budding cells compared to the other cancer cell clusters (Fig. 4C). Thus, we hypothesized that these genes were more active in cancer cells in contact with CAFs. Indeed, this was observed for all FOVs in the LDB areas, however, this effect was less pronounced in the MDB or HDB areas (Fig. 6E).

Discussion

Budding cancer cells at the invasive front of several different types of adenocarcinomas are prime suspects for being en route to metastasis, and a high degree of tumor budding is associated with poor patient prognosis³. Therefore, it is important to gain a better molecular understanding of tumor budding to enable the development of better treatments and to define new biomarkers to guide treatment decisions. For this purpose, we employed a state-of-the-art spatial transcriptomics method called SMI³⁰, which allowed us to study individual adenocarcinoma cells and their spatial context within the TME.

Initially, we analyzed adjacent normal tissue to better understand how effective the technology is for identifying different cell types in colon tissues. By performing UMAP analysis, we observed that the cells formed two superclusters: one containing all the epithelial cells and another containing all the stromal and immune cells. The epithelial cells were located within the crypts, as evidenced by a strong overlap with panCK immunofluorescent staining of the same tissue slide and T-cells were co-localized with the immunofluorescent staining for CD3. These findings suggest that the method can effectively define the boundaries and nature of individual colon cells based on nuclear and membrane staining as well as unsupervised clustering based on gene-expression patterns.

When analyzing the colorectal cancer specimens, we also observed a distinct clustering of the epithelial cancer cells and the various cells within the TME across the majority of the FOVs. Most of the cancer cells within individual clusters were predominantly located physically together, in line with the anticipated spatial distribution if their clonal evolution is guided by mutations and epigenetic alterations influenced by natural selection. However, clusters of cancer cells containing most of the budding cells displayed a notably distinct spatial distribution, with individual cells scattered along the periphery of masses of other cancer cells. This spatial expression pattern cannot be explained solely by Darwinian evolution³¹. Instead, it suggests that interactions with cells in the TME are responsible for the major gene expression changes that occur during the formation of budding cells and their migration away from the main part of the tumor. This is in line with a recent study of genetic profiles of primary and metastatic tumors across many different cancer types from thousands of patients, suggesting

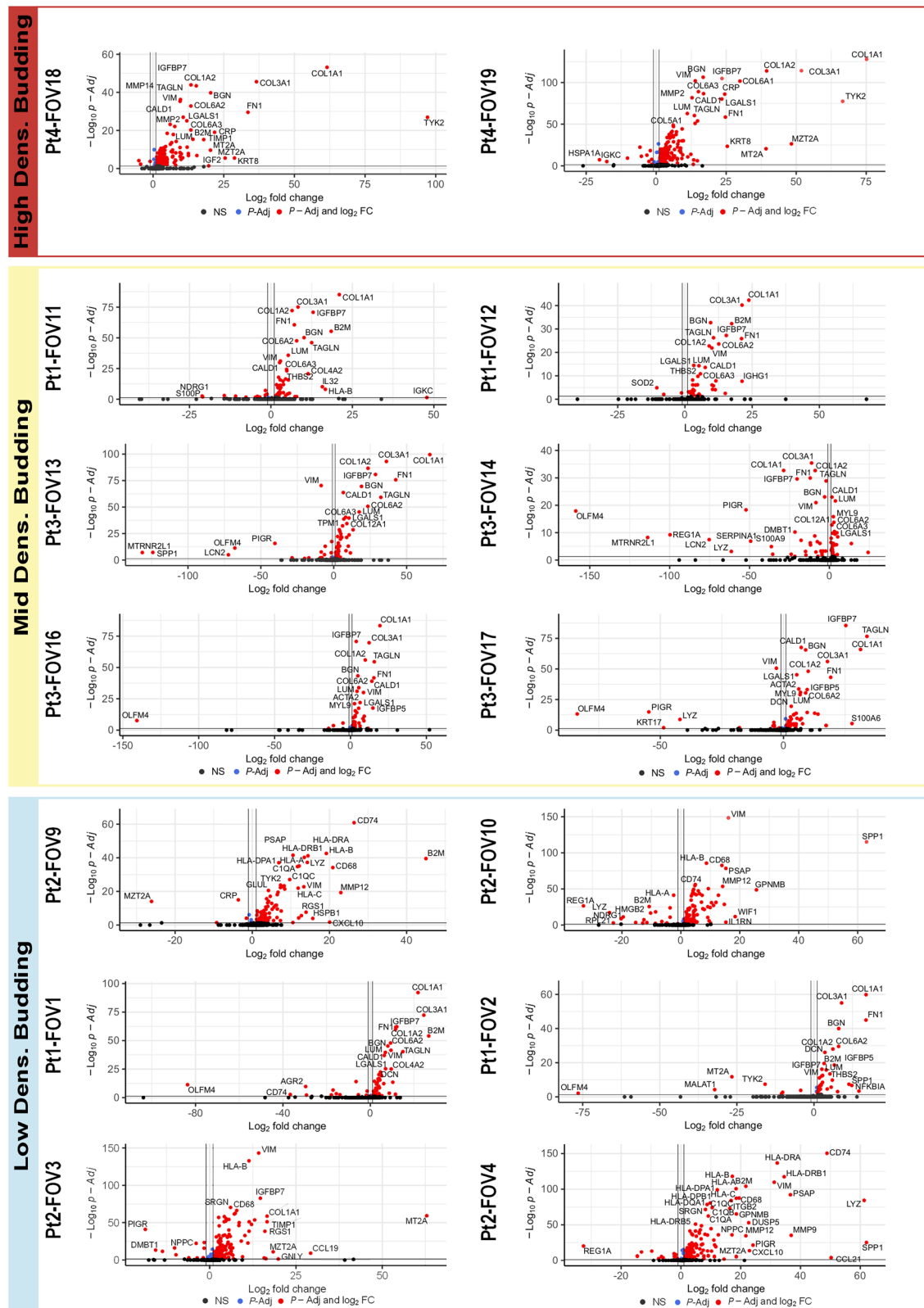


Fig. 3 | Individual FOVs from high, mid, and low-density budding areas share upregulated and downregulated genes. Volcano plots depicting P -adjusted values relative to log2 fold changes (FCs) for comparisons between the cancer clusters that encompass most budding cells and all other cancer clusters for each field of view (FOV). The top panel in red shows the two FOVs annotated as high-density budding areas. The mid panel in light yellow shows the six

FOVs annotated as mid-density budding areas. The bottom panel in light blue shows the six FOVs annotated as low-density budding areas. Red dots depict genes with a P -Adjusted (Benjamini-Hochberg correction) value < 0.05 and a \log_2 FC > 1 . Blue dots depict genes with a P -Adjusted (Benjamini-Hochberg correction) value < 0.05 and a \log_2 FC < 1 . ns: non-significant.

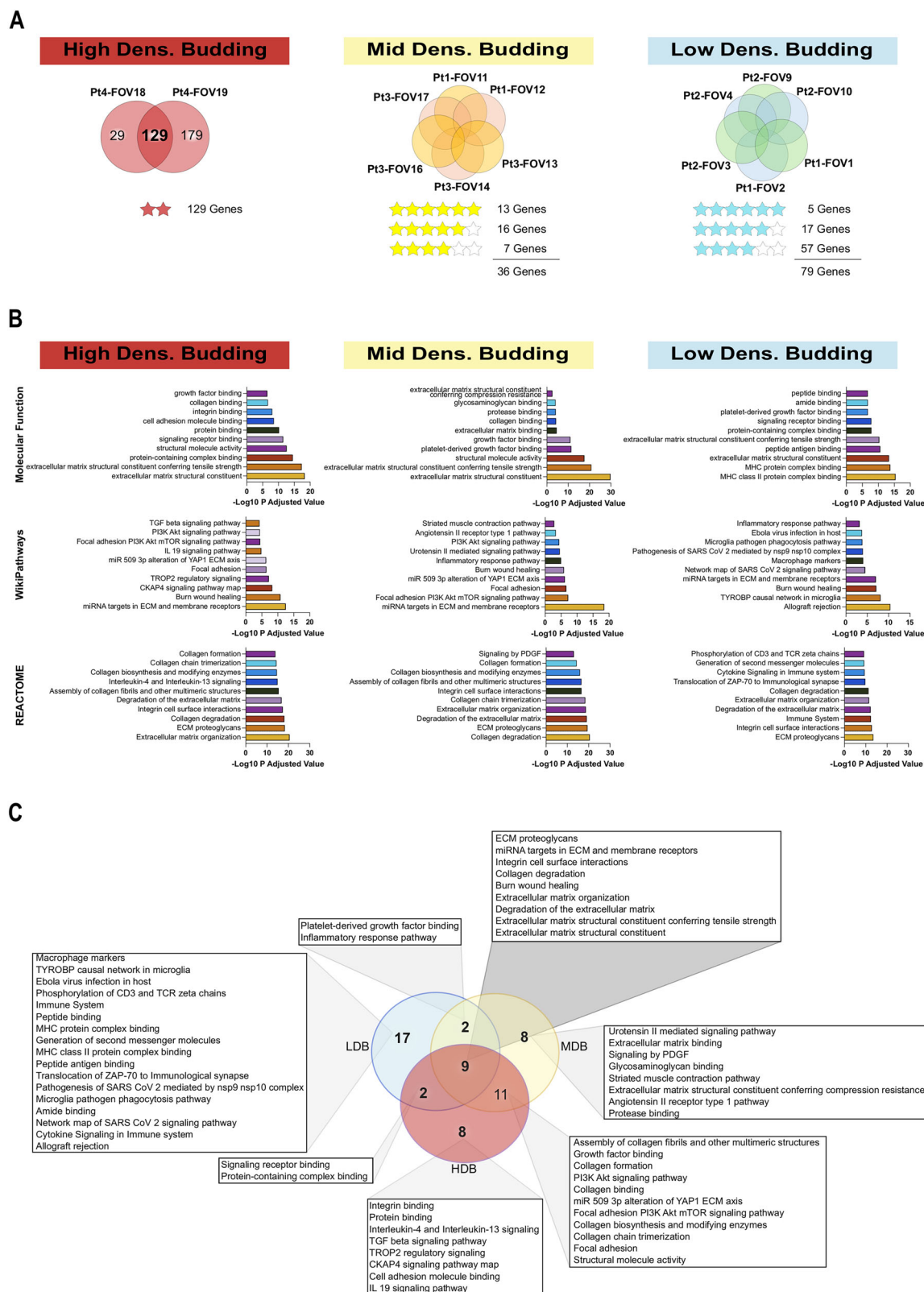


Fig. 4 | Enrichment analysis on high, mid, and low-density budding areas revealed a progressive adaptation of the budding cells to the surrounding stroma. **A** Venn diagrams showing the overlap of upregulated genes in cancer cell clusters containing most budding cells relative to all other cancer cell clusters for the two high-density budding (HDB) field of views (FOVs) (left panel), the six mid-density budding (MDB) FOVs (middle panel), and the six low-density budding (LDB) FOVs (right panel). For the two HDB FOVs, 129 genes were commonly upregulated. For the six MDB FOVs, 13 genes were commonly upregulated, 16 genes were

upregulated in five FOVs, and seven genes were upregulated in four FOVs. For the six LDB FOVs, five genes were commonly upregulated, 17 genes were upregulated in five FOVs, and 57 genes were upregulated in four FOVs. **B** Bar plots showing pathway enrichment analyses performed on the 129, 36, and 79 genes identified in the HDB, MDB, and LDB areas, respectively. The Top 10 terms ranked by *P*-Adjusted values for “Molecular Function” from Gene Ontology, “WikiPathways”, and “REACTOME” databases are shown. **C** Venn diagram highlighting the shared and unique pathways from the enrichment analyses shown in **B**.

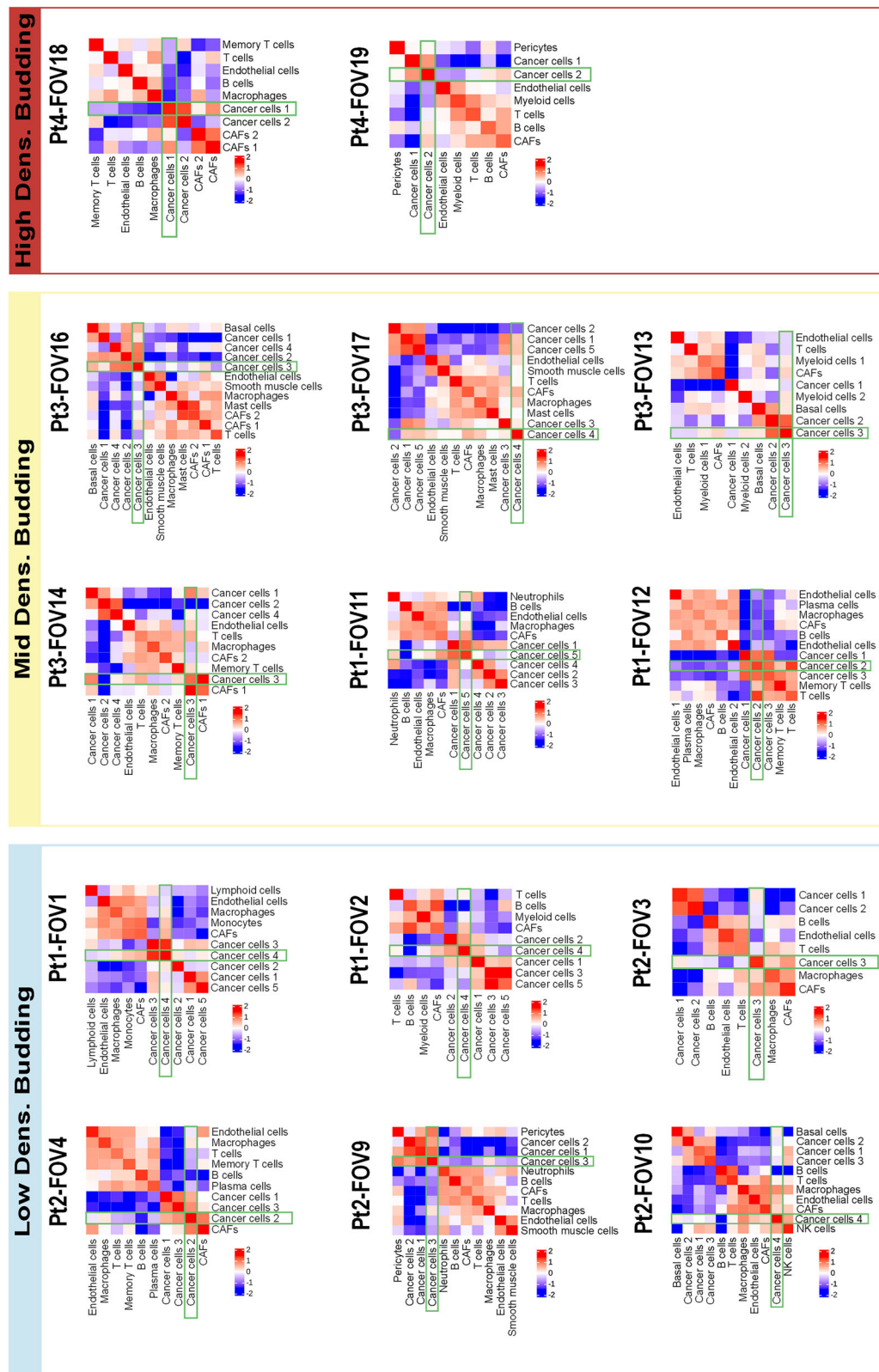


Fig. 5 | Cancer cells from the budding clusters are frequently in contact with CAFs. Heatmaps showing z-scores of the cell-cell interaction values for each of the cell types per field of view (FOV). The FOVs are divided into three sub-groups

based on their budding density, high, mid, and low-density budding areas. Green boxes highlight the budding clusters.

that metastasis is not primarily driven by genetic mutations³². Indeed, cancer cells can exhibit substantial phenotypic plasticity, allowing them to adopt states or phenotypes not directly inherited from their predecessors^{33,34}. However, it is challenging to distinguish cancer cell plasticity from heritable

changes and the influence of the TME³⁵. In support of non-genetic events being the main drivers of metastasis, we found that most of the gene expression changes occurring at the invasive front and in the budding cells are lost again in liver metastasis, suggesting that these changes result from

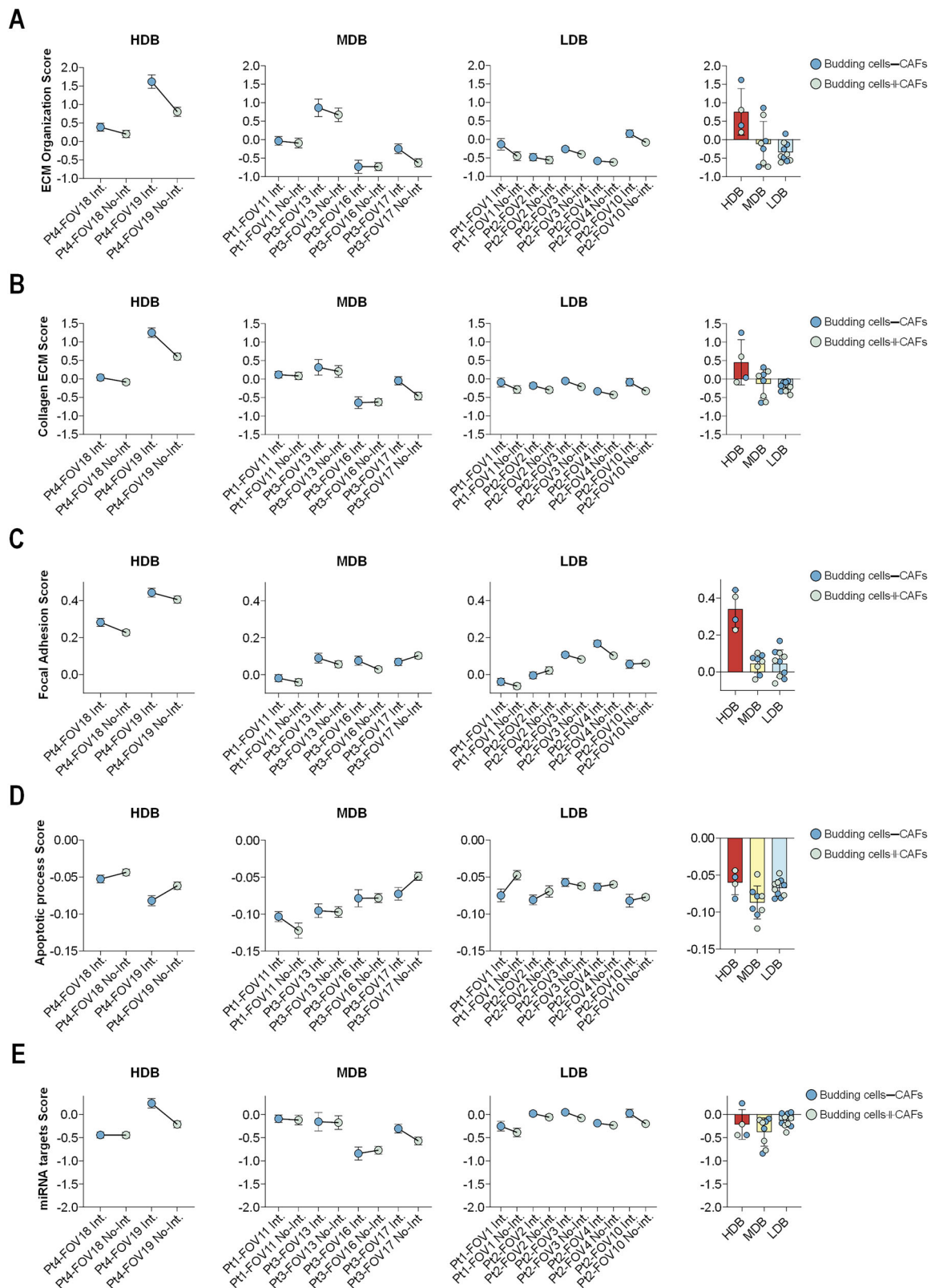


Fig. 6 | Cancer cells in contact with CAFs change the ECM composition to facilitate migration and avoid apoptosis. Dot plots showing the “ECM Organization Score” (A), “Collagen ECM Score” (B), “Focal Adhesion Score” (C), “Apoptotic process Score” (D), and “miRNA Targets Score” (E). The dots represent mean values, and the error bars indicate standard deviation. FOVs are separated into

high-density budding (HDB), mid-density budding (MDB), and low-density budding (LDB) areas. Blue dots correspond to the mean score of the budding cells that are in contact with CAFs, whereas light green dots correspond to the mean score of budding cells not in contact with CAFs. The box plots to the right show the mean values for the HDB, MDB, and LDB areas. Error bars indicate standard deviations.

phenotypic plasticity and interactions with the TME rather than from the acquisition of mutations and stable epigenetic changes.

It became immediately clear that many genes involved in EMT were changed in the cancer cell clusters containing most of the budding cells, including downregulation of epithelial markers like *EPCAM* and *CDH1* as well as *PIGR* and *OLFM4*, alongside a marked upregulation of mesenchymal markers such as *LUM*, *VIM*, and several collagen genes, including *COL1A1* for which cancer-cell intrinsic expression has recently been shown to promote metastatic colonization of breast cancer cells³⁶. Moreover, we found an upregulation of *WNT2B*, a gene recently shown to promote the EMT process in colon cancer³⁷. In addition to individual genes, we also observed a high degree of EMT based on a broader previously published gene expression signature^{20–22}.

To further elucidate the molecular characteristics associated with tumor budding, we identified a set of 11 genes (*TYK2*, *IL2RG*, *KRT17*, *HLA-B*, *NPPC*, *WIF1*, *IL32*, *B2M*, *CCND1*, *CRIP1*, *ITGB1*) that exhibited elevated expression levels in the cluster containing most budding cells in FOV1, compared to all other cancer cell clusters as well as to the CAF cluster. When examining other FOVs, we consistently observed a strong enrichment of this gene signature within the cancer cell clusters containing budding cells. Interestingly, most of these 11 genes have previously been shown to promote various aspects of cancer metastasis. For instance, *TYK2*, is a pro-metastatic gene known to promote cell death resistance³⁸ and *IL2RG* expression has been linked to compromised immune surveillance and more advanced tumor stages and poor prognosis in colorectal cancer³⁹. *IL32* is a potent pro-inflammatory gene capable of inducing the production of other cytokines and chemokines that contribute to the establishment of a TME favoring invasion and metastasis⁴⁰. *CCND1*, *CRIP1*, and *KRT17* overexpression has been observed in several types of cancers and has been linked to cancer cell migration, invasion, and metastasis in experimental studies^{41–45}. *ITGB1* encodes an integrin subunit, which is important for adhesion to the ECM and migration of breast cancer cells^{46,47}, in line with our in vitro data where the migratory capacity of Caco-2 cells was decreased upon *ITGB1* silencing. This effect was not observed in HCT116 cells. However, these cell lines have distinct molecular backgrounds: Caco-2 cells harbor mutations in *APC* and *CTNNB1*, while HCT116 cells possess a *KRAS* mutation and mismatch repair defects. These differences may explain their divergent responses to *ITGB1* silencing. Elevated levels of *B2M* have been associated with poor prognosis in several cancers and the absence of *B2M* protein expression indicates that no metastases are present and a favorable clinical outcome among mismatch repair-deficient colorectal cancer⁴⁸. However, it has also been suggested that low *B2M* expression levels are associated with lymph node metastasis and poor survival of colorectal cancer patients⁴⁹. On the other hand, *WIF1* encodes a secreted protein inhibiting the Wnt signaling pathway, which is associated with increased cell survival and metastasis⁵⁰. Finally, the role of *NPPC* in metastasis has not been elucidated. In a TCGA dataset, these genes showed higher expression in colon adenocarcinomas than in adjacent normal tissues and were significantly associated with poor disease-free survival, despite the dataset being generated by bulk RNA sequencing.

Several other genes were upregulated in cancer cell clusters containing most of the budding cells across the majority of FOVs, including *FN1*, a key gene in migration and invasion⁵¹, and *MT2A*, which encode a metallothionein involved in metal ion homeostasis and protection against oxidative stress. Notably, *MT2A* can exhibit both tumor-suppressive and oncogenic properties depending on the cancer type and context^{52–55}.

When dividing the FOVs according to the density of budding cells, not to be confused with the ITBCC guidelines for scoring of tumor budding⁵⁶, and performing gene set enrichment analyses on commonly upregulated genes, an interesting picture emerged as these analyses suggested a progressive adaptation of the budding cells to the new environment and surrounding cells. First, the activation of a defense response and the upregulation MHC class II molecules, which are known to be expressed on the surface of epithelial cells of the gastrointestinal tract, especially in inflamed tissues⁵⁷, alongside initial changes in the ECM organization, as observed in the LDB areas, may represent the earliest events in the budding process. Whereas in

the MDB areas, an increase in protease binding, matrix contraction, and an upregulation of pathways related to tissue development were seen. Finally, as observed in the HDB areas, further changes in the ECM configuration coincided with the activation of PDGF and TGF beta signaling pathways.

Interestingly, the cancer cells from clusters that included most of the budding cells were most frequently in contact with CAFs. Our analyses, conducted on cancer cells in contact with CAFs compared to cancer cells not in contact with CAFs, indicated that CAFs trigger gene expression changes related to ECM organization and focal adhesion, which facilitate cellular migration and play a protective role against mechanical stress and apoptosis^{58,59}. Indeed, the CAFs seemed to induce a downregulation of apoptotic processes. These findings suggest a key role for CAFs in protecting the budding cancer cells and facilitating their migration into the TME. Unfortunately, the employed 1000-plex panel hindered us to further subclassify CAFs due to the lack of the key genes required for their identification⁶⁰. Thus, future research should focus on defining the roles of specific CAF subtypes in metastasis.

While several different methods have been developed for spatial transcriptomics, our study represents the first to profile the budding cancer cells at the invasive front of colon adenocarcinomas with single cell resolution. In a study involving seven colorectal (CRC) patients, Valdeolivas and colleagues used the 10x Genomics Visium platform to map intratumoral heterogeneity⁶¹ with respect to the consensus molecular subtypes (CMS)⁶². Although this method covers more genes, it is based on spots that contain more than a single cell and do not cover the entire tissue. In addition, DSP has been employed in a study involving 10 CRC patients showing that tumor deposits, which are discrete nodules of cancerous tissue within the lymphatic drainage area with prognostic value⁶³, have distinctive biological profiles characterized by more matrix remodeling, EMT, and cell motility⁶⁴. Similar to the Visium platform, DSP also does not provide data from single cells and, therefore, cannot offer the insights into the budding cells and their neighboring cells that we have provided here. However, while DSP and Visium covers the entire transcriptome, our study was limited to analyzing close to 1,000 genes, and our analyses would also have benefited from being able to analyze samples in 3D. This is particularly important when comparing cancer cells in contact with CAFs to those not in contact, as it is possible that these cells might be in contact with other CAFs in the z-plane and cells that appear to be budding cells may also be in contact with other cancer cells when visualized in 3D⁶⁵. Finally, it will also be interesting to focus on other immune cells within the TME and their interactions with cancer cells, such as T-cells and macrophages in future spatial transcriptomics studies^{66,67}.

Further investigations, using advanced models, such as co-cultures or organoids, are warranted to further explore the metastasis-related gene expression changes and cellular crosstalk uncovered in this study.

In conclusion, our study provides novel insights into the biology of tumor budding and early metastasis. Notably, by using an unsupervised approach, we were able to identify cancer cell clusters that encompass most budding cells, which displayed a distinct spatial distribution. Our analyses further suggest that CAFs are driving many of the gene-expression changes behind this phenomenon, which enable cancer cells to migrate into the TME. While we were able to pinpoint many individual genes and pathways that may drive the earliest events in cancer metastasis, additional advancements of spatial transcriptomics methodologies will further enhance our understanding of these complex processes.

Materials and methods

Patient samples

This study mainly focused on CosMx analyses performed on material from four patients (four adenocarcinomas and one adjacent normal tissue). Extensive information regarding these patients can be found in Supplementary Table 1. Additional patient samples were included for the various validation studies, five for the 10X Visium analyses (all adenocarcinomas) and nine for the RNAscope duplex analyses (six adenocarcinomas and three adenomas). All patients had their tumors surgically resected and were diagnosed and treated according to the guidelines from The Danish

Colorectal Cancer Group. All tumors were classified according to the latest WHO classification of digestive tumors, 5th edition. None of the patients received chemotherapy prior to surgery and none were suspected of having hereditary cancer.

We only used excess surgical material from the patients for this study, which was conducted in accordance with the principles of the Declaration of Helsinki. The requirement for patient consent to the specific analyses was waived and The Danish National Committee on Health Research Ethics approved the study (approval number 2016120).

Spatial molecular imaging

CosMx SMI technology using the Human Universal Cell Characterization RNA Panel (960 genes and 20 negative control probes targeting non-human sequences) was used to analyze 20 FOVs from the formalin-fixed paraffin-embedded (FFPE) samples described above. The 20 FOVs included invasive front, tumor mass and adjacent normal areas. The same tissue sections were stained by immunofluorescent antibodies against B2M/CD298 (membrane markers), DAPI (nuclear marker) (CosMx Universal Segmentation kit, CMX-H-UCS-12-P), PanCK (epithelial markers), CD45 (CosMx Human IO PanCK/CD45 Morphology marker kit: CMX-H-IO-PCKCD45-MM34-P) and CD3 (immune markers) (CosMx Human CD3 A la Carte marker kit: CMX-H-CD3-MM5-P).

FFPE tissues were deparaffinized following the guidelines described elsewhere³⁰. In brief, 5-micron tissue sections were incubated overnight at 60 °C, followed by in-situ hybridization (ISH) on the Leica Bond RXm with a deparaffinization step and epitope retrieval at 100 °C for 15 min using epitope retrieval buffer (Leica Biosystems, Tris/EDTA, pH 9.0). This was followed by a tissue digestion step with 3 µg/ml Proteinase K diluted in ACD Protease Plus at 40 °C for 30 min.

Then, following a washing with diethyl pyrocarbonate (DEPC) water, the samples were incubated with fiducial beads (Bangs laboratories) in 2X saline sodium citrate and Tween-20 solution (SSCT solution) at room temperature (RT) for 5 min. This incubation was followed by three consecutive rinsing steps of 5 min; the first two with Tris-glycine buffer (0.1 M glycine, 0.1 M Tris-base) and the last one with 1X PBS. Then, the fixed samples were blocked with 100 mM N-succinimidyl S-acetylthioacetate (SATA) (Thermo Fisher Scientific) buffer for 15 min at RT. Sections were rinsed with 2X saline sodium citrate (SSC) for another 5 min and sealed with Adhesive SecureSeal Hybridization Chamber (Grace Bio-Labs).

The ISH probes were incubated at 95 °C for 2 min and placed on ice, then prepared the ISH probe mix (1 nM 960-plex ISH probe, 10 nM Attenuation probes, 1X Buffer R, 0.1 U/uL SUPERase InTM (Thermo Fisher Scientific)) to be pipetted into the hybridization chamber. The hybridization was performed overnight at 37 °C. Thereafter, the tissue sections were washed with 50% formamide (VWR) in 2X SSC at 37 °C for 25 min twice, followed by another 2 washes with 2X SSC for 2 min at RT and blocked with 100 mM SATA for 15 min protected from light.

The RNA target readout was performed in the CosMx SMI instrument³⁰. Briefly, the loaded flow cells were washed with the Reporter Wash Buffer to remove air bubbles. Then, a preview scan was performed followed by the FOV selection. As a reference, in parallel, an adjacent serial slide from the same tissue was hematoxylin & eosin (H&E)-stained to identify the regions of interest to be selected at this step inside the SMI. Then, the RNA readouts were started, where 100 µL of the corresponding Reporter pool flowed into the flow cell and was incubated for 15 min. The Reporter Wash Buffer (1 mL) washed unbound reporter probes; the Imaging buffer was then added to the flow cell to acquire an image per round (a total of 16 times). Three Z-stack images were acquired per FOV, and the photo-cleavable linkers on the reporter probes fluorophores released by UV light and washed with Strip Wash buffer.

The end of the 16 cycles of RNA readout was followed by the 4-fluorophore-conjugated antibody cocktail and DAPI incubation for 2 h. Reporter Wash buffer was added, and Imaging buffer to each flow cell before making Z-stack images on the five corresponding channels (488 nm for CD298/B2M, 532 nm for PanCK, 594 nm for CD45, 647 nm for CD3 and

blue channel for DAPI). Next, segmentation workflow was implemented using Cellpose algorithm⁶⁸ and the SMI nuclear-based signal to delineate the cell boundaries. Finally, the cell count matrix and all transcript information were retrieved for downstream analyses.

SMI data analyses

Initially five FOVs were removed from the analyses, one due to detachment of tissue and four due to an overall low gene count (Supplementary Fig. 1). From the remaining 15 FOVs, a total of 50,727 cells were detected by the Cellpose algorithm⁶⁸ and of these 50,457 passed a conservative quality filter of having a minimum of 20 genes detected per cells and the mean count of the negative control probes were close to 1 (Supplementary Fig. 16).

We then computed the filtered, normalized and scaled gene expression counts (log base 2 with a scale factor of 6000 was applied) using the Giotto toolkit (v 3.3.0)⁶⁹, excluding the negative probes. Detection of high-variable features was performed using *cov_groups* algorithm with a threshold of 0.5 and a difference of variance of 0.1. The shared nearest neighbor method was performed using 30 principal components, 15 k neighbors and a minimum of five shared edges. Dimensional reduction using Uniform manifold approximation and projection algorithm (UMAP) was then performed selecting the 30 principal components. With this setup applied to each of the FOVs of the study, unsupervised Leiden algorithm¹⁴ (0.5 resolution) was used for clustering annotation. Cell type annotations were performed using canonical markers and validated using PanglaoDB database¹⁵. The differential expression analysis was performed using Gini-coefficient method⁷⁰. The working Giotto objects were interconverted to AnnData or Seurat objects for further analysis. Enrichment analysis was performed using g:Profiler⁷¹. The “EMT states”, “11-gene signature”, and the scores for “ECM Organization” (REAC:R-HSA-1474244), “Collagen ECM” (GO:0062023), “Focal adhesion” (WP:WP306), “Apoptotic process” (GO:0006915), and “miRNA targets” (WP:WP2911) were calculated using AddmoduleScore function from Seurat package. The gene lists from each discrete EMT state were extracted from Simeonov KP et al.²⁰.

The cell-cell interaction analysis was performed using the Giotto function cellProximityEnrichment. Briefly, a 2D network is constructed using the Delaunay triangulation method⁷². Then, spatial proximity enrichment between pairs of cell types was calculated based on the Delaunay data, where observed over expected frequency of cell-cell proximity interactions are annotated. Expected interactions were derived from the average frequency of 1000 spatial network simulations, where each cell type was serially reshuffled from each node in the spatial network. These values were then converted to z-scores and represented in the heatmap. Cell types derived from Leiden clustering were used for the annotations.

Digital spatial profiling

GeoMx (NanoString Technologies) digital spatial profiling (DSP) data were from a previous study¹⁷ focusing on circular RNAs. This study was conducted on tissues from three patients diagnosed with colon cancer and included both primary tumors (10 regions of interest (ROIs) in total) as well as paired adjacent normal tissues (6 ROIs in total) and paired liver metastasis (9 ROIs in total). The experiments were performed according to the manufacturer's recommendations and the processing of the tissues and next-generation sequencing procedures¹⁷. Smooth muscle antibody, pan-cytokeratin (PanCK) antibody, and CD45 antibody (GeoMx NanoString morphology kit: GMX-PRO-MORPH-HST-12) were used as morphology markers for visualization of smooth muscle cells, epithelial and cancer cells as well as immune cells, respectively, and nuclei were visualized using SYTO-83 (S11364, Thermo Fisher/Invitrogen). The immunofluorescence images were used for the selection of ROIs, which were further segmented into a total of 42 areas of illumination (AOIs) based on cell type (SMA + , PanCK + , CD45 + , and PanCK-/CD45-) and tissue architecture (invasive front or tumor mass). At least 200 cells were present within each AOI.

The GeoMx DSP Analysis Suite (v. 2.5.1.143) was used for quality control (QC) and normalization of the data according to the manufacturer's recommendations. All 42 AOIs passed the QC thresholds where a raw read

limit of 1000 was applied. Trimmed, stitched, and aligned reads were limited to a minimum length of 80 bp, a sequencing saturation of 50%, and a limit of quantification of 10%. Of the total of 18,681 genes that were detected, 11,741 passed the QC filtering. Finally, prior to gene expression analyses, a Q3 normalization was applied.

10X Visium Spatial Gene Expression analysis

Visium Spatial Gene Expression (10x Genomics) analysis using FFPE tissues from five human colon adenocarcinomas were performed following the manufacturer's instructions. Briefly, 5 µm thick tissue sections were cut from five FFPE blocks corresponding to five patients and placed on a glass slide. The tissues were deparaffined and subsequently stained using H&E. Bright-field images were captured using a NanoZoomer 2.0 HT (Hamamatsu) at 20X magnification. The areas of interest (6.5 × 6.5 mm) were marked by an experienced pathologist for each slide, and the tissues were placed in a Visium CytAssist Tissue Slide Cassette (10x Genomics). The tissues were destained and de-crosslinked prior to hybridization with the 18,000 gene target-specific probe pairs. This was followed by a ligation step, which ensured that specific hybridized probes were captured. Next, probe ligation, RNA digestion, and tissue removal was performed, together with transfer of probes from each of the 5000 barcoded spatial spots containing a poly-A sequence complementary to the probes. The spatially barcoded probe products were extended and amplified with indexes during the library construction. The number of cycles required for library amplification was determined by qPCR using the KAPA SYBR® FAST qPCR kit (Sigma Aldrich). The samples were sequenced with NovaSeq 6000 system (Illumina). The secondary analysis was performed using 10X Loupe Browser (v. 7).

Chromogenic in situ hybridization

The localization of *OLFM4* and *PIGR* within colon adenocarcinomas and adjacent normal tissues was investigated by chromogenic in situ hybridization using the chromogenic RNAscope® 2.5 HD Duplex Detection kit (ACD, Biotechnique). Specific probes (Hs-PIGR-C1 and Hs-OLFM4-C2) targeting *PIGR* and *OLFM4* were used for the assay (ACD, Biotechnique). In short, a 5 µm thick FFPE tissue biopsies were deparaffinized and treated according to the manufacturer's protocol. This was followed by probe hybridization with a 1:50 ratio *OLFM4*-C2 to *PIGR*-C1 probes during the hybridization step and overnight storage in 5x SSC buffer diluted from a 20x SSC stock solution (Invitrogen, Thermo Fisher Scientific). Ten subsequent amplification steps allowed the signals to be detected as red (*OLFM4*-C2) or blue (*PIGR*-C1) chromogenic signal. Counterstaining was done using 50% hematoxylin (Sigma Aldrich). Data were collected using a Slide Scanner (Upright Wide-field Fluorescence, Olympus VS120, Olympus Life Science) and analyzed using OlivIA (v. 2.9.1, Olympus Life Science) and QuPath (v. 0.3.1)⁷³.

Single-cell RNA-seq analysis

The scRNA-seq data were obtained from the NCBI Gene Expression Omnibus (GEO) repository with accession number GSE178341. Raw sequencing data were processed using the Seurat R package (v5.1.0) following standard quality control and normalization workflows. Briefly, cells with fewer than 100 detected genes or >50% mitochondrial reads were excluded. Count data were log-normalized, and highly variable genes were selected for downstream analysis. Dimensionality reduction was performed using PCA, followed by UMAP for visualization using 30 dimensions. Patient information, including tumor stage and cell type labels, were transferred from the reference cluster list and meta tables provided at the GEO repository.

RNA-sequencing data

Sample details, alignment, and processing details can be found at the original article from Korsgaard U. et al.²⁴. Secondary analysis and normalization were performed using DESeq2 (v 1.38.3)⁷⁴.

Wound healing assay, cell counting, and viability measurements

Caco-2 and HCT116 human adenocarcinoma cell lines were maintained at 37 °C in humidified air containing 5% CO₂. Caco-2 cells were cultured in

EMEM (Sigma-Aldrich) with 1% non-essential amino acids (Sigma-Aldrich). HCT116 cells were cultured in McCoy's 5 A medium (Sigma-Aldrich). Both media contained 10% FBS, 2 mmol/L L-glutamine, penicillin (100 ug/mL) and streptomycin (100 ug/mL). Knockdown was performed using siPOOLS (siTOOLS Biotech) and Lipofectamine RNAiMax (Thermo Fisher Scientific), following manufacturer's guidelines. Twenty-four hours after transfection (siControl and si*ITGB1*), the wound area was introduced by removing a 2 mm width adherent sealing film present in each well before the seeding step (150,000 cells/well). The images were captured using a Leica DMIL microscope and the cell-free scratch wound area measured with ImageJ/Fiji software using the wound healing size tool⁷⁵. The cell numbers and viability were measured using Trypan Blue (Thermo Fisher Scientific) and the automated cell counter Luna II (Logos Biosystems).

Reverse transcriptase quantitative PCR analysis

cDNA from Caco-2 and HCT116 cells for RT-qPCR expression analyses of *ITGB1* was synthesized from total RNA using M-MLV-RT (Invitrogen) according to the manufacturer's protocol. RT-qPCR was performed using SYBR Green PCR Master Mix (Thermo Fisher Scientific) on a LightCycler 480 (Roche). The following primers were employed: *ITGB1* forward: 5'-GCCGCGCGGAAAAGATGAA -3' and *ITGB1* reverse: 5'-TGAGCAAACACACAGCAAAC -3' and the housekeeping *MRLP19* forward: 5'-GTTCTATGTTGGAAGTATTCTTCGT-3' and *MRLP19* reverse: 5'-TCTCTGAATGCAAATCCCCAGA-3'.

Statistics and reproducibility

Nonparametric t tests were performed to compute the *P* values between groups as described in the figure legends. For TCGA derived data, one-way ANOVA was used for statistical significance determination. Adjusted *P* values were calculated using FDR correction or Benjamini-Hochberg methods as described in the figure legends and methods section. For Disease-Free survival plots, the Mantel-Cox statistical test was used. Gene-specific violin plots were generated using Seurat (v. 5.0) and statistical analyses were performed by T tests using the ggpvr package (v. 0.6.0). For generation of box plots, GraphPad Prism software (v. 10) was used. Bubble plots were generated using Seurat (v. 5.0). Volcano plots were performed using Enhanced volcano package (v 1.16.0). CNET plot and GeoMx enrichment dot plots were generated using the R package clusterProfiler (v 4.6.2) and plotted using the R package ggplot2 (v 3.5.0).

Reporting summary

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

Data availability

All ethical regulations relevant to human research participants were followed. The raw genomic data related to individual patients are not publicly available due to the waived requirement of consent from the patients for this study. This is because the Danish law and the European Union General Data Protection Regulation (GDPR) classify all genomic data as sensitive personal information, and therefore, it cannot be shared with third parties without prior approval. Source data from all graphs in this study can be found in Supplementary Data File.

Received: 7 October 2024; Accepted: 27 August 2025;

Published online: 16 September 2025

References

1. Cervantes, A. et al. Metastatic colorectal cancer: ESMO Clinical Practice Guideline for diagnosis, treatment and follow-up. *Ann. Oncol.* **34**, 10–32 (2023).
2. Lambert, A. W., Pattabiraman, D. R. & Weinberg, R. A. Emerging Biological Principles of Metastasis. *Cell* **168**, 670–691 (2017).
3. Lugli, A., Zlobec, I., Berger, M. D., Kirsch, R. & Nagtegaal, I. D. Tumour budding in solid cancers. *Nat. Rev. Clin. Oncol.* **18**, 101–115 (2021).

4. Baxter, N. N. et al. Adjuvant Therapy for Stage II Colon Cancer: ASCO Guideline Update. *J. Clin. Oncol.* **40**, 892–910 (2022).
5. Kim, S. A. et al. Loss of CDH1 (E-cadherin) expression is associated with infiltrative tumour growth and lymph node metastasis. *Br. J. Cancer* **114**, 199–206 (2016).
6. Gosens, M. J., van Kempen, L. C., van de Velde, C. J., van Krieken, J. H. & Nagtegaal, I. D. Loss of membranous Ep-CAM in budding colorectal carcinoma cells. *Mod. Pathol.* **20**, 221–232 (2007).
7. Hostettler, L., Zlobec, I., Terracciano, L. & Lugli, A. ABCG5-positivity in tumor buds is an indicator of poor prognosis in node-negative colorectal cancer patients. *World J. Gastroenterol.* **16**, 732–739 (2010).
8. Peng, Z., Ye, M., Ding, H., Feng, Z. & Hu, K. Spatial transcriptomics atlas reveals the crosstalk between cancer-associated fibroblasts and tumor microenvironment components in colorectal cancer. *J. Transl. Med.* **20**, 302 (2022).
9. Hu, J. L. et al. CAFs secreted exosomes promote metastasis and chemotherapy resistance by enhancing cell stemness and epithelial-mesenchymal transition in colorectal cancer. *Mol. cancer* **18**, 91 (2019).
10. Koelzer, V. H. et al. Phenotyping of tumor-associated macrophages in colorectal cancer: Impact on single cell invasion (tumor budding) and clinicopathological outcome. *Oncoimmunology* **5**, e1106677 (2016).
11. Ueno, H. et al. Desmoplastic Pattern at the Tumor Front Defines Poor-prognosis Subtypes of Colorectal Cancer. *Am. J. Surg. Pathol.* **41**, 1506–1512 (2017).
12. Ueno, H. et al. Histologic categorization of desmoplastic reaction: its relevance to the colorectal cancer microenvironment and prognosis. *Ann. Surg. Oncol.* **22**, 1504–1512 (2015).
13. Righi, A. et al. Tumour budding is associated with hypoxia at the advancing front of colorectal cancer. *Histopathology* **66**, 982–990 (2015).
14. Traag, V. A., Waltman, L. & van Eck, N. J. From Louvain to Leiden: guaranteeing well-connected communities. *Sci. Rep.* **9**, 5233 (2019).
15. Franzén, O., Gan, L. M. & Björkegren, J. L. M. PanglaoDB: a web server for exploration of mouse and human single-cell RNA sequencing data. *Database*. <https://doi.org/10.1093/database/baz046>. (2019).
16. Nie, Y. et al. The downregulation of tight junction proteins and plgR in the colonic epithelium causes the susceptibility of EpCAM(+/-) mice to colitis and gut microbiota dysbiosis. *Front. Mol. Biosci.* **11**, 1442611 (2024).
17. García-Rodríguez, J. L. et al. Spatial Profiling of Circular RNAs in Cancer Reveals High Expression in Muscle and Stromal Cells. *Cancer Res.* **83**, 3340–3353 (2023).
18. van der Flier, L. G., Haegebarth, A., Stange, D. E., van de Wetering, M. & Clevers, H. OLFM4 is a robust marker for stem cells in human intestine and marks a subset of colorectal cancer cells. *Gastroenterology* **137**, 15–17 (2009).
19. Menz, A. et al. Pan-keratin immunostaining in human tumors: a tissue microarray study of 15,940 tumors. *Int. J. Surg. Pathol.* **31**, 927–938 (2023).
20. Simeonov, K. P. et al. Single-cell lineage tracing of metastatic cancer reveals selection of hybrid EMT states. *Cancer Cell* **39**, 1150–1162.e1159 (2021).
21. Pastushenko, I. & Blanpain, C. EMT Transition States during Tumor Progression and Metastasis. *Trends Cell Biol.* **29**, 212–226 (2019).
22. McFaline-Figueroa, J. L. et al. A pooled single-cell genetic screen identifies regulatory checkpoints in the continuum of the epithelial-to-mesenchymal transition. *Nat. Genet.* **51**, 1389–1398 (2019).
23. Pelka, K. et al. Spatially organized multicellular immune hubs in human colorectal cancer. *Cell* **184**, 4734–4752.e4720 (2021).
24. Korsgaard, U. et al. The Transcriptional Landscape of Coding and Noncoding RNAs in Recurrent and Nonrecurrent Colon Cancer. *Am. J. Pathol.* **194**, 1424–1442 (2024).
25. Slenter, D. N. et al. WikiPathways: a multifaceted pathway database bridging metabolomics to other omics research. *Nucleic Acids Res.* **46**, D661–d667 (2018).
26. Croft, D. et al. Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Res.* **39**, D691–D697 (2011).
27. Hallett, M. A., Venmar, K. T. & Fingleton, B. Cytokine stimulation of epithelial cancer cells: the similar and divergent functions of IL-4 and IL-13. *Cancer Res.* **72**, 6338–6343 (2012).
28. Huebner, K. et al. ATF2 loss promotes tumor invasion in colorectal cancer cells via upregulation of cancer driver TROP2. *Cell Mol. Life Sci.* **79**, 423 (2022).
29. Sun, X. et al. Elucidation of CKAP4-remodeled cell mechanics in driving metastasis of bladder cancer through aptamer-based target discovery. *Proc. Natl Acad. Sci. USA* **119**, e2110500119 (2022).
30. 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).
31. Ciriello, G. et al. Cancer Evolution: A Multifaceted Affair. *Cancer Discov.* **14**, 36–48 (2024).
32. Nguyen, B. et al. Genomic characterization of metastatic patterns from prospective clinical sequencing of 25,000 patients. *Cell* **185**, 563–575.e511 (2022).
33. Fennell, K. A. et al. Non-genetic determinants of malignant clonal fitness at single-cell resolution. *Nature* **601**, 125–131 (2022).
34. Yang, D. et al. Lineage tracing reveals the phylogenetics, plasticity, and paths of tumor evolution. *Cell* **185**, 1905–1923.e1925 (2022).
35. Karras, P., Black, J. R. M., McGranahan, N. & Marine, J. C. Decoding the interplay between genetic and non-genetic drivers of metastasis. *Nature* **629**, 543–554 (2024).
36. Couto, J. P. et al. Nicotinamide N-methyltransferase sustains a core epigenetic program that promotes metastatic colonization in breast cancer. *EMBO J.* **42**, e112559 (2023).
37. Ortiz-Masià, D. et al. WNT2b Activates Epithelial-mesenchymal Transition Through FZD4: Relevance in Penetrating Crohn's Disease. *J. Crohns Colitis* **14**, 230–239 (2020).
38. Sanda, T. et al. TYK2-STAT1-BCL2 pathway dependence in T-cell acute lymphoblastic leukemia. *Cancer Discov.* **3**, 564–577 (2013).
39. Gharib, E. et al. IL-2RG as a possible immunotherapeutic target in CRC predicting poor prognosis and regulated by miR-7-5p and miR-26b-5p. *J. Transl. Med.* **22**, 439 (2024).
40. Tsai, C. Y. et al. Interleukin-32 increases human gastric cancer cell invasion associated with tumor progression and metastasis. *Clin. Cancer Res.* **20**, 2276–2288 (2014).
41. Ji, R. et al. Keratin 17 upregulation promotes cell metastasis and angiogenesis in colon adenocarcinoma. *Bioengineered* **12**, 12598–12611 (2021).
42. Wang, Z. et al. Overexpression of KRT17 promotes proliferation and invasion of non-small cell lung cancer and indicates poor prognosis. *Cancer Manag Res* **11**, 7485–7497 (2019).
43. Kolapalli, S. P. et al. RNA-Binding RING E3-Ligase DZIP3/hRUL138 Stabilizes Cyclin D1 to Drive Cell-Cycle and Cancer Progression. *Cancer Res.* **81**, 315–331 (2021).
44. Zhang, L. Z., Huang, L. Y., Huang, A. L., Liu, J. X. & Yang, F. CRIP1 promotes cell migration, invasion and epithelial-mesenchymal transition of cervical cancer by activating the Wnt/ β -catenin signaling pathway. *Life Sci.* **207**, 420–427 (2018).
45. Wu, Z. et al. CRIP1 reshapes the gastric cancer microenvironment to facilitate development of lymphatic metastasis. *Adv. Sci.* **10**, e2303246 (2023).
46. Barnawi, R. et al. β 1 Integrin is essential for fascin-mediated breast cancer stem cell function and disease progression. *Int. J. Cancer* **145**, 830–841 (2019).
47. Zhang, L. et al. EZH2 engages TGF β signaling to promote breast cancer bone metastasis via integrin β 1-FAK activation. *Nat. Commun.* **13**, 2543 (2022).
48. Koelzer, V. H., Baker, K., Kassahn, D., Baumhoer, D. & Zlobec, I. Prognostic impact of β -2-microglobulin expression in colorectal

- cancers stratified by mismatch repair status. *J. Clin. Pathol.* **65**, 996–1002 (2012).
49. Shrout, J. et al. beta(2)microglobulin mRNA expression levels are prognostic for lymph node metastasis in colorectal cancer patients. *Br. J. Cancer* **98**, 1999–2005 (2008).
 50. Ji, D. et al. MicroRNA-181a promotes tumor growth and liver metastasis in colorectal cancer by targeting the tumor suppressor WIF-1. *Mol. Cancer* **13**, 86 (2014).
 51. Spada, S., Tocci, A., Di Modugno, F. & Nisticò, P. Fibronectin as a multiregulatory molecule crucial in tumor matrisome: from structural and functional features to clinical practice in oncology. *J. Exp. Clin. Cancer Res* **40**, 102 (2021).
 52. Lin, S. et al. Transcription Factor Myeloid Zinc-Finger 1 Suppresses Human Gastric Carcinogenesis by Interacting with Metallothionein 2A. *Clin. Cancer Res.* **25**, 1050–1062 (2019).
 53. Dias, A. M. et al. Downregulation of metallothionein 2A reduces migration, invasion and proliferation activities in human squamous cell carcinoma cells. *Mol. Biol. Rep.* **49**, 3665–3674 (2022).
 54. Shimizu, M. et al. Metallothionein 2A Expression in cancer-associated fibroblasts and cancer cells promotes esophageal squamous cell carcinoma progression. *Cancers* **13**. <https://doi.org/10.3390/cancers13184552>. (2021).
 55. Kim, H. G. et al. Metallothionein-2A overexpression increases the expression of matrix metalloproteinase-9 and invasion of breast cancer cells. *FEBS Lett.* **585**, 421–428 (2011).
 56. Lugli, A. et al. Recommendations for reporting tumor budding in colorectal cancer based on the International Tumor Budding Consensus Conference (ITBCC) 2016. *Mod. Pathol.* **30**, 1299–1311 (2017).
 57. Wosen, J. E., Mukhopadhyay, D., Macaubas, C. & Mellins, E. D. Epithelial MHC Class II Expression and Its Role in Antigen Presentation in the Gastrointestinal and Respiratory Tracts. *Front. Immunol.* **9**, 2144 (2018).
 58. Seetharaman, S. & Etienne-Manneville, S. Cytoskeletal crosstalk in cell migration. *Trends Cell Biol.* **30**, 720–735 (2020).
 59. Cooper, J. & Giancotti, F. G. Integrin Signaling in Cancer: Mechanotransduction, Stemness, Epithelial Plasticity, and Therapeutic Resistance. *Cancer Cell* **35**, 347–367 (2019).
 60. Khaliq, A. M. et al. Refining colorectal cancer classification and clinical stratification through a single-cell atlas. *Genome Biol.* **23**, 113 (2022).
 61. Valdeolivas, A. et al. Profiling the heterogeneity of colorectal cancer consensus molecular subtypes using spatial transcriptomics. *NPJ Precis Oncol.* **8**, 10 (2024).
 62. Guinney, J. et al. The consensus molecular subtypes of colorectal cancer. *Nat. Med.* **21**, 1350–1356 (2015).
 63. Nagtegaal, I. D. et al. Tumor deposits in colorectal cancer: improving the value of modern staging—a systematic review and meta-analysis. *J. Clin. Oncol.* **35**, 1119–1127 (2017).
 64. Brouwer, N. P. et al. Transcriptomics and proteomics reveal distinct biology for lymph node metastases and tumour deposits in colorectal cancer. *J. Pathol.* **261**, 401–412 (2023).
 65. Lin, J. R. et al. Multiplexed 3D atlas of state transitions and immune interaction in colorectal cancer. *Cell* **186**, 363–381.e319 (2023).
 66. Zhu, H. et al. Location matters: spatial dynamics of tumor-infiltrating T cell subsets is prognostic in colon cancer. *Front. Immunol.* **15**, 1293618 (2024).
 67. van Vlerken-Ysla, L., Tyurina, Y. Y., Kagan, V. E. & Gabrilovich, D. I. Functional states of myeloid cells in cancer. *Cancer Cell* **41**, 490–504 (2023).
 68. Stringer, C., Wang, T., Michaelos, M. & Pachitariu, M. Cellpose: a generalist algorithm for cellular segmentation. *Nat. Methods* **18**, 100–106 (2021).
 69. Del Rossi, N., Chen, J. G., Yuan, G. C. & Dries, R. Analyzing Spatial Transcriptomics Data Using Giotto. *Curr. Protoc.* **2**, e405 (2022).
 70. Jiang, L., Chen, H., Pinello, L. & Yuan, G. C. GiniClust: detecting rare cell types from single-cell gene expression data with Gini index. *Genome Biol.* **17**, 144 (2016).
 71. Kolberg, L. et al. g:Profiler-interoperable web service for functional enrichment analysis and gene identifier mapping (2023 update). *Nucleic Acids Res.* **51**, W207–w212 (2023).
 72. Kim, J. & Cho, J. Delaunay Triangulation-Based Spatial Clustering Technique for Enhanced Adjacent Boundary Detection and Segmentation of LiDAR 3D Point Clouds. *Sensors* **19**. <https://doi.org/10.3390/s19183926>. (2019).
 73. Bankhead, P. et al. QuPath: Open source software for digital pathology image analysis. *Sci. Rep.* **7**, 16878 (2017).
 74. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
 75. Suarez-Arnedo, A. et al. An image J plugin for the high throughput image analysis of in vitro scratch wound healing assays. *PLoS One* **15**, e0232565 (2020).

Acknowledgements

We acknowledge the MOMA Core Center, Department of Molecular Medicine, Aarhus University Hospital, Denmark, and the Danish Single-Cell Examination Platform (CellX) for Next generation sequencing and providing access to CosMx and GeoMx instruments that supported the research reported in this paper. CellX was established with support from the Danish Research Agency through the Danish national research infrastructure program (5229-0009B). This work was supported by the Lundbeck Foundation (R307-2018-3433), Riisfort Fonden, Novo Nordisk Foundation (Open Discovery Innovation Network (ODIN) program, grant NNF20SA0061466), the Danish Cancer Society (R304-A17698) (L.S.K.), and Købmand i Odense Johann og Hanne Weimann Født Seedorffs Legat (J.L.G.-R.).

Author contributions

Conceptualization, J.L.G.-R., H.H., and L.S.K.; Data curation, J.L.G.-R., U.K., E.F.J., P.V.S., C.B.B., Formal analysis, J.L.G.-R., and L.S.K.; Investigation, J.L.G.-R., L.S.K., H.L.W., J.H., I.Z., J.K., and H.H.; Validation, J.L.G.-R., S.M.V., T.P.P., M.S., and S.L.V.; Resources, L.L., J.K., and L.S.K.; Writing original draft, J.L.G.-R., and L.S.K.; Writing review and editing, J.L.G.-R., U.K., S.M.V., T.P.P., M.S., S.L.V., E.F.J., H.L.W., P.V.S., C.B.B., J.H., L.L., I.Z., K.D.S., J.K., H.H., L.S.K.; Visualization, J.L.G.-R., S.M.V., and T.P.P.; Supervision, L.S.K., I.Z., and K.D.S.; Funding acquisition, J.L.G.-R., and L.S.K.

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/s42003-025-08799-x>.

Correspondence and requests for materials should be addressed to Juan L. García-Rodríguez or Lasse S. Kristensen.

Peer review information *Communications Biology* thanks Helmut Dolznig and the other, anonymous, reviewers for their contribution to the peer review of this work. Primary Handling Editor: Johannes Stortz. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025