Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Fibroblast-enriched senescence-associated transcription and measurement-context-dependent FAP–matrix covariation in colorectal cancer

Jinglin Wang
The Affiliated Hospital of Guizhou Medical University
Pan Sun
The Affiliated Hospital of Guizhou Medical University
Wei Zhu
The Affiliated Hospital of Guizhou Medical University
Yan Zhang
The Affiliated Hospital of Guizhou Medical University
Yao Tang
The Affiliated Hospital of Guizhou Medical University
Juan Yang
yjj63yjj63@163.com
The Affiliated Hospital of Guizhou Medical University

Research Article

Keywords: colorectal cancer, cancer-associated fibroblast, cellular senescence, fibroblast activation protein, extracellular matrix, single-cell RNA sequencing, spatial transcriptomics, composition adjustment

Posted Date: November 11th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11027032/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Fibroblast-enriched senescence-associated transcription and measurement-context-dependent FAP–matrix covariation in colorectal cancer

Jinglin Wang[1,2], Pan Sun[1], Wei Zhu[1], Yan Zhang[1], Yao Tang[1], Juan Yang[1,*]

1 Department of Geriatrics, The Affiliated Hospital of Guizhou Medical University, Guiyang, China

2 Department of Digestive Disease, Guizhou Provincial People's Hospital, Guiyang, China

* Corresponding author: Juan Yang, MD, PhD; Email: yj63yj63@163.com; ORCID iD: 0009-0001-2081-6950

Short title: Fibroblast senescence-associated transcription in CRC

Article type: Analysis

Abstract

Bulk tumour transcriptomes mix cellular composition, senescence-associated transcription and extracellular-matrix signals. We examined the compartmental distribution of senescence-associated transcription in colorectal cancer and its relationship to fibroblast activation protein (FAP)-linked matrix covariation. Patient-paired compartments were analysed in two single-cell RNA-sequencing cohorts, followed by FAP-specific models, expression-matched null signatures, target-purged bulk composition analyses, colon tumour proteomics and three-cohort spatial assessment. Source-available SenMayo scores were higher in fibroblast than epithelial compartments in GSE132465 (23 of 23 patients; median paired difference 0.466; exact $P = 2.38 \times 10^{-7}$) and GSE166555 (9 of 10; median difference 0.149; $P = 0.0195$). A 111-gene common-core sensitivity retained the direction in both cohorts. Within GSE132465 tumour fibroblasts, patient-cluster bootstrap estimates did not support additional enrichment among FAP-detected cells: $\beta = 0.0239$ (95% confidence interval [CI], −0.0175 to 0.0675) for SenMayo and $\beta = 0.0340$ (95% CI, −0.0374 to 0.1095) for the senescence-associated secretory phenotype. FAP–matrix correlations were positive in both single-cell cohorts, although direct contrasts did not establish that they were larger than FAP–SenMayo correlations. In 380 TCGA tumours, FAP13–matrix4 covariation remained positive after target-purged joint composition adjustment (partial $\rho = 0.466$; 95% CI, 0.363 to 0.556), but its magnitude varied across proxy definitions. Spatial common-core effects were positive in GSE280315 and the Valdeolivas cohort but mixed or negative in GSE334323. Together, the data localize senescence-associated transcription preferentially to fibroblast compartments while separating this finding from FAP detection and matrix abundance. Establishing durable cellular senescence or causal stromal signalling will require orthogonal validation.

Keywords: colorectal cancer; cancer-associated fibroblast; cellular senescence; fibroblast activation protein; extracellular matrix; single-cell RNA sequencing; spatial transcriptomics; composition adjustment

Introduction

Colorectal cancer contributes substantially to the global burden of cancer incidence and mortality [1]. Cancer-associated fibroblasts (CAFs) are heterogeneous components of the tumour microenvironment that contribute to extracellular-matrix organization, paracrine signalling and immune regulation [2]. Fibroblast activation protein (FAP) marks activated stromal populations in many malignancies, but FAP expression does not define a single functional CAF phenotype [3]. This measurement distinction is particularly relevant to studies of colorectal cancer. Stromal transcripts can dominate molecular classification and prognostic associations [4–6], and spatial studies have identified FAP-expressing fibroblasts in macrophage-rich niches [7]. FAP expression, fibroblast abundance, matrix transcription and inferred cell state are therefore related measurements, but they are not interchangeable.

SenMayo was developed as a cross-tissue panel of senescence-associated genes [8]. Cellular senescence itself is a durable growth-arrest phenotype that requires convergent molecular, structural and functional evidence [9–10]. Individual features such as lamin B1 loss can be

informative, but none is universal [11]. The senescence-associated secretory phenotype is similarly context dependent and overlaps inflammatory, stress and wound-response programmes [12]. Transcriptomic scores can thus localize senescence-associated expression, whereas stable arrest and a functional senescent phenotype require orthogonal tissue or perturbation assays.

Experimental studies show why this distinction matters. Senescent myofibroblasts can organize immunosuppressive tumour niches in pancreatic cancer, and primary human colon fibroblasts show inducer-dependent senescence and secretory phenotypes [13–14]. Using human colorectal cancer specimens and functional models, Ge et al. showed that senescent fibroblasts transferred lipids to CD8-positive T cells through a CD36-dependent process, increasing lipid peroxidation and impairing T-cell cytotoxic function [15]. Yang et al. identified macrophage-induced CAF senescence and senescence-associated secretory phenotype-mediated chemoresistance through IL1B–IL1R1 signalling [16]. A recent multi-omic study associated a COMP-positive CAF programme with a non-inflammatory senescence signature and CD8-positive T-cell exclusion while explicitly treating COMP–CD47 as a predicted, rather than established, interaction [17]. Together, these studies define experimentally supported senescent-fibroblast phenotypes while leaving a separate measurement problem for retrospective transcriptomic analyses. Buissant des Amorie et al. resolved telocyte-like and trophocyte-like CAF populations in early colorectal cancer and linked trophocyte-like CAFs to epithelial oncofetal plasticity [18]. This heterogeneity raises the additional question of whether CAF marker programmes align with senescence-associated transcription within the fibroblast compartment.

We addressed four related questions about the interpretation of these measurements. First, is senescence-associated transcription preferentially localized to fibroblast rather than epithelial tumour compartments at the patient level? Second, does FAP transcript detection identify a fibroblast subgroup with additional enrichment after sequencing depth, subtype and patient structure are considered? Third, how does FAP covary with a collagen/fibronectin matrix programme within fibroblasts and in bulk and protein measurements? Fourth, does the compartment contrast generalize across independent spatial platforms and annotation schemes? Non-supportive results were retained, and localization, covariation and generalizability were distinguished from claims about cellular identity or mechanism.

Materials and methods

Study design and evidence hierarchy

This retrospective study reanalysed publicly available, de-identified human colorectal cancer data. No participant was recruited and no new human or animal specimen was collected. The discovery estimand was the patient-paired difference in SenMayo score between tumour fibroblast-lineage and epithelial compartments in GSE132465. GSE166555 provided an external directional evaluation using patient-compartment pseudobulks. FAP-specific models, expression-matched nulls, within-fibroblast correlations, bulk composition analyses and proteomics were prespecified as supporting or boundary analyses. Three spatial cohorts were

evaluated as a generalizability stress test. Cells, bins and spots were treated as nested observations; the patient or independent tumour sample was the inferential unit wherever biological replication was available [19–20]. Analyses were exploratory and were not preregistered. Conclusions were limited to molecular localization and covariation.

Data sources and analysis populations

TCGA colon and rectal adenocarcinoma expression data were obtained through UCSC Xena, and 380 barcode-reviewed primary tumours were retained [21]. GSE39582 contributed 566 tumour microarrays after exclusion of 19 non-tumour mucosa samples [22]. GSE132465 comprised 63,689 annotated cells from 23 patients [23]; the primary analysis was restricted to 1,501 source-annotated tumour fibroblast-lineage cells (Myofibroblasts and Stromal 1–3) and 17,469 tumour epithelial cells. GSE166555 contained 68,702 annotated cells from 12 tumour patients [24]; patients required at least 20 tumour CAF-like cells and 100 tumour epithelial cells, leaving 10 eligible paired patients. Spatial analyses included three GSE280315 Visium HD colorectal cancer sections with author deconvolution labels [25], 12 eligible sections from six patients in the Valdeolivas Visium cohort with pathologist labels [26], and three tumour sections from the GSE334323 FFPE Visium CytAssist preprint cohort [27]. CPTAC colon tumour proteomics included 97 tumours with the required proteins, accessed through cBioPortal [28–30]. Dataset roles and independent units are summarized in Table 1.

Table 1. Dataset roles, analysis populations and independent units

Dataset	Modality	Analysis population	Independent unit	Role
TCGA COAD/READ	Bulk RNA sequencing	380 primary tumours	Tumour/patient	Composition-proxy sensitivity
GSE39582	Bulk microarray	566 tumour samples	Tumour sample	Cross-platform marginal association
GSE132465	Single-cell RNA sequencing	1,501 tumour fibroblast-lineage and 17,469 tumour epithelial cells; 23 patients	Patient	Primary localization and FAP-specific analyses
GSE166555	Single-cell RNA sequencing	68,702 annotated cells; 10 eligible paired tumour patients	Patient	External directional evaluation
CPTAC colon	Proteomics	97 tumours with required proteins	Tumour	Protein-level FAP–matrix triangulation
GSE280315	Visium HD spatial transcriptomics	3 colorectal cancer sections	Patient/section	Spatial generalizability stress test
Valdeolivas	Visium spatial transcriptomics	12 sections from 6 patients	Patient	Spatial generalizability stress test
GSE334323	FFPE Visium CytAssist spatial transcriptomics	3 tumour sections	Patient/section	Spatial sensitivity and host-feature audit

Note. CAF, cancer-associated fibroblast; CPTAC, Clinical Proteomic Tumor Analysis Consortium; TCGA, The Cancer Genome Atlas

Gene sets, score construction and overlap control

FAP13 comprised FAP, POSTN, THY1, PDPN, TAGLN, ACTA2, MMP2, MMP9, CXCL12, TGFB1, INHBA, WNT2 and WNT5A. Matrix4 comprised COL1A1, COL1A2, COL3A1 and FN1. A 25-gene senescence-associated secretory phenotype (SASP25) set was retained as supportive rather than diagnostic evidence. The original 125-gene SenMayo list [8] was purged of CXCL12, MMP2, MMP9 and WNT2 because they overlap FAP13, leaving 121 candidates. Of these, 119 were represented with non-zero variance in GSE132465 and 114 in GSE166555. These source-available sets were the primary within-cohort definitions. The exact

intersection of 111 represented genes across both single-cell and all three spatial cohorts was frozen as a cross-dataset sensitivity; its effects were not pooled across platforms.

Unless otherwise specified, a gene-set score was the mean of gene-wise standardized expression within the relevant analysis universe. Single-gene FAP and MKI67 values used log-normalized expression. Gene membership, platform representation, zero-variance exclusions and overlap with composition-proxy features were recorded before each model. Transcript detection was labelled “FAP-detected” and was not treated as protein positivity.

GSE132465 public-data reconstruction and compartment analysis

The official GEO raw unique molecular identifier matrix and cell annotation were downloaded, and file checksums were recorded. Counts were normalized per cell as log1p(10,000 × count/library size). All 18,970 tumour fibroblast-lineage and epithelial cells used in the archived analysis matched the public annotations. Reconstructed SenMayo119 scores were identical to the archived scores within numerical precision (Pearson r = 1.000; root-mean-square error 2.74×10^{-16}).

For each patient, SenMayo119, SASP25 and MKI67 were summarized separately in tumour fibroblast-lineage and epithelial cells. Fibroblast-minus-epithelial differences were tested with exact two-sided Wilcoxon signed-rank tests. Percentile 95% confidence intervals used 10,000 patient-level bootstrap resamples where estimands and sample size permitted. A sensitivity required at least 20 cells in each compartment. Patient-level analysis avoided treating cells from the same patient as independent replicates.

External single-cell evaluation and common-core sensitivity

For GSE166555, raw counts were summed within each eligible patient-compartment group, transformed as log1p counts per million and standardized gene-wise across the eligible patient-compartment pseudobulks. The 114 represented genes from the frozen overlap-removed SenMayo candidates were scored without outcome-driven reselection. The same paired test and patient bootstrap procedures were applied. Because annotation and normalization differed from GSE132465, effect magnitudes were reported on separate scales and were not pooled. Both cohorts were also rescored with the exact 111-gene common core as a sensitivity analysis.

Expression-matched null signatures

To assess whether the GSE132465 compartment contrast merely reflected generic expression and detection characteristics, 5,000 random gene sets were matched to SenMayo119 on mean expression, standard deviation and detection rate. The median patient-level fibroblast-minus-epithelial score difference was recomputed for each draw; the corresponding mean and direction-count statistics were archived as secondary diagnostics. The upper-tail empirical P value was calculated as (1 + number of null statistics at least as large as the observed statistic)/5,001. This procedure tests the observed magnitude against the specified matched-gene background; it does not prove that the score is lineage neutral or specific for durable senescence. SASP25 underwent the same supportive analysis.

FAP-detected subgroup and within-fibroblast models

Within the 1,501 GSE132465 tumour fibroblast-lineage cells, FAP-detected cells had at least one FAP unique molecular identifier. The adjusted models included FAP detection, standardized MKI67, log1p total unique molecular identifiers, source fibroblast subtype and patient. Random-intercept mixed models with Satterthwaite inference were retained as conventional sensitivity analyses. The cluster-robust estimates used patient fixed effects and 5,000 bootstrap resamples of whole patient clusters; all cells belonging to a sampled patient were retained together. The same framework was applied to SenMayo119 and SASP25. A symmetric supporting model regressed matrix4 on continuous FAP expression with depth, subtype and patient adjustment. A confidence interval crossing zero was described as not supporting additional enrichment, not as equivalence.

Patient-level correlations and dependent-correlation contrasts

Within each single-cell cohort, we calculated patient-level Spearman correlations among tumour-fibroblast summaries of FAP, FAP13, matrix4 and SenMayo. Ten thousand paired patient bootstrap samples provided percentile confidence intervals. Within each resample, the differences $\rho(\text{FAP}, \text{matrix4}) - \rho(\text{FAP}, \text{SenMayo})$ and $\rho(\text{FAP}, \text{matrix4}) - \rho(\text{SenMayo}, \text{matrix4})$ were recomputed on the same patients. Two-sided bootstrap sign probabilities were reported. Formal contrasts, rather than separate significance labels, determined whether one correlation could be described as larger than another.

Bulk expression and target-purged composition sensitivity

TCGA COAD/READ primary-tumour expression matrices were restricted to one primary sample per patient and combined over common genes. FAP13, matrix4, the overlap-removed SenMayo score and a non-overlapping five-gene fibroblast lineage score (PDGFRA, PDGFRB, LUM, DCN and COL14A1) were computed as means of gene-wise standardized expression. GSE39582 provided a cross-platform marginal FAP13–matrix4 estimate. Spearman correlations described unadjusted tissue-level covariation.

Partial Spearman correlations in TCGA were obtained by rank-transforming each target and residualizing it against fibroblast-content covariates. Covariates were the five-gene lineage score, MCP-counter fibroblasts [31–32], EPIC CAFs [33], or joint MCP-counter and EPIC estimates. Because deconvolution results depend on reference and preprocessing choices [34–35], estimates from multiple proxy definitions were retained. The official MCP-counter fibroblast features overlapped FAP13 or matrix4 at COL1A1, COL3A1 and TAGLN; EPIC reference features overlapped at COL1A1 and COL3A1. We therefore recomputed pair-purged proxies after removing both target sets and globally disjoint proxies after removing all FAP13, matrix4 and SenMayo genes from proxy construction. Five thousand patient bootstrap resamples provided 95% confidence intervals. Full-proxy and target-purged results were displayed together to make incorporation bias visible.

CPTAC protein analysis

In 97 CPTAC colon tumours, FAP protein abundance was correlated with a FAP-excluded extracellular-matrix protein score comprising COL1A1, COL1A2 and FN1. Spearman ρ and a 5,000-sample tumour bootstrap confidence interval were reported. Benjamini–Hochberg adjustment covered the two prespecified protein comparisons in the archived analysis. This analysis triangulated FAP–matrix covariation at the protein level; it did not measure a coordinated senescence or SASP phenotype.

Spatial generalizability analysis

The spatial contract was frozen before target scores were inspected. Counts were normalized within section as log1p(10,000 × count/library size), and represented genes were standardized across retained in-tissue bins or spots. GSE280315 used author-provided singlet deconvolution labels: CAF, Fibroblast, Myofibroblast and Proliferating Fibroblast were the stromal group, and labels beginning with Tumor were the malignant epithelial group. The Valdeolivas cohort used pathologist-labelled Fibroblastic_stroma and Tumor spots, with replicate sections averaged within patient. GSE334323 lacked author cell-type labels; stromal and epithelial spots were therefore assigned by a frozen margin of 0.25 within-slide standard deviations between non-overlapping lineage-marker scores. Analyses required at least 30 assigned observations per compared compartment and retained patient, not spot, as the independent unit.

The primary spatial outcome was the within-patient stromal-minus-tumour SenMayo score difference. A secondary contrast compared Fisher-z-transformed partial correlations of FAP–matrix4 with the larger of FAP–SenMayo and SenMayo–matrix4 within stromal regions, adjusting for library size and the stromal-marker score. H1 and H2 were corrected by Holm's method within each prespecified cohort. MKI67 tumour-minus-stromal direction was a polarity control. The 111-gene common-core analysis was a cross-cohort sensitivity reported separately for each cohort. Five thousand expression- and detection-matched sets provided supportive null distributions. For GSE334323, a post-outcome technical audit identified 323 microbial, viral or control probes in the library-size denominator; the host-transcriptome sensitivity retained only ENSG-identified human features, and both the original and corrected outputs were preserved.

Exploratory CAF marker programme analysis

We examined three CAF marker programmes derived from Supplementary Table 5 of Buissant des Amorie et al. [18]: trophocyte-like matrix metalloproteinase (MMP)-positive, trophocyte-like MMP-negative and telocyte-like CAFs. An initial top-100 fold-change extraction did not meet the prespecified minimum of 50 represented genes per programme. Before testing programme–SenMayo associations, we therefore restricted source markers to adjusted $P < 0.05$, positive log2 fold change, at least 25% detection within the source cluster and a detection-frequency difference of at least 10 percentage points, then selected the top 100 fold changes. Genes shared with SenMayo, FAP13, matrix4, the five-gene fibroblast proxy or

MKI67 were removed. These source-detectability-adapted programmes were analysed as continuous scores, not used to assign CAF subtypes.

This analysis included 1,435 tumour fibroblasts from 15 GSE132465 patients with at least 20 fibroblasts and 1,315 tumour fibroblasts from the ten eligible GSE166555 patients. Log1p counts per 10,000 were standardized gene-wise within each cohort's eligible fibroblasts. The three exposure programmes shared 82, 86 and 85 genes across cohorts; the outcome used 114 shared SenMayo genes. This within-fibroblast definition was separate from the 111-gene cross-platform compartment sensitivity. Within each patient, ranked programme and SenMayo scores were residualized against ranked library size and MKI67 and correlated. Patient coefficients were averaged with equal weights on the Fisher-z scale and back-transformed. Two-sided one-sample t tests of patient Fisher z were adjusted by the Benjamini–Hochberg method across three programmes within each cohort; q denotes this adjusted P value. Percentile confidence intervals used 5,000 patient bootstraps. Subtype-adjusted models and 1,000 approximately expression- and detection-matched random programmes were sensitivity analyses. In TCGA, the same programme definitions were evaluated with cohort-specific gene coverage before and after adjustment for the disjoint five-gene fibroblast expression proxy (fib5). Full extraction, scoring and inferential details are provided in the supporting methods, Fig. S7, Table S8 and S1 Data.

Statistical analysis and software

All tests were two-sided unless an upper-tail matched-null test was prespecified. Exact P values, effect estimates and confidence intervals were prioritized over cell- or spot-level tests. No cross-platform meta-analysis was performed. R 4.6.1 was invoked explicitly with the R 4.6 user library. Key packages included Matrix 1.7-5, lme4 2.0-6, lmerTest 3.2-1, ggplot2 4.0.3, MCPcounter and EPIC. Random seeds, session information, consumed-file manifests and checksums were retained with the analysis outputs. The exploratory CAF module used Python 3.12, NumPy, pandas and SciPy, with session versions and seeds recorded in S1 Data.

Results

Cohorts and inferential hierarchy

We first compared paired tumour compartments in two single-cell cohorts. Subsequent analyses examined FAP-specific and matrix associations, bulk composition sensitivity, protein triangulation and spatial generalizability. GSE132465 contributed 23 paired patients; GSE166555 contributed 10 eligible paired patients. The bulk, protein and spatial modules contributed 380 TCGA tumours, 566 GSE39582 tumours, 97 CPTAC tumours and 12 independent spatial patients or tumour sections across three cohorts. Cell and spot counts described measurement density but did not increase the biological replicate count.

Senescence-associated transcription was higher in fibroblast compartments in two single-cell cohorts

In GSE132465, the source-available 119-gene SenMayo score was higher in tumour fibroblast-lineage than epithelial compartments in all 23 patients. The median paired difference was 0.466 (95% patient-bootstrap confidence interval [CI], 0.418 to 0.553; exact $P = 2.38 \times 10^{-7}$). Requiring at least 20 cells in each compartment retained 15 patients, with the same median difference of 0.466 (95% CI, 0.410 to 0.551; $P = 6.10 \times 10^{-5}$). SASP25 was higher in fibroblasts in all 23 patients (median fibroblast-minus-epithelial difference, 0.757; 95% CI, 0.612 to 0.863; $P = 2.38 \times 10^{-7}$), whereas MKI67 was lower in 21 of 23 patients (median difference, −0.118; 95% CI, −0.137 to −0.074; $P = 1.81 \times 10^{-4}$). This combination localizes a senescence-associated transcriptional pattern and provides a directionally concordant proliferation control; stable arrest remains an orthogonal phenotype to be tested (Fig. 1a–c and Table 2).

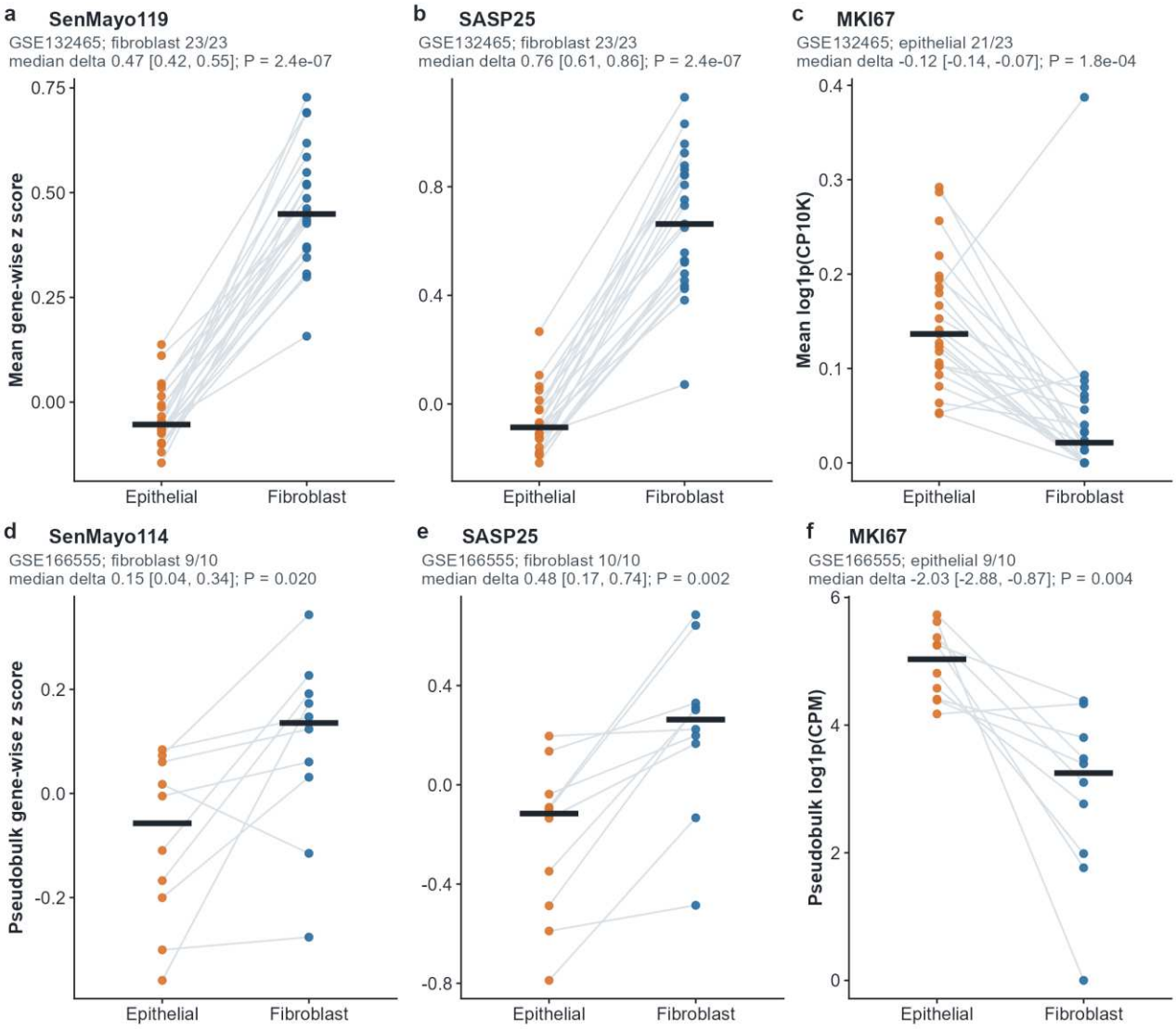


Fig. 1 Patient-paired localization of senescence-associated transcription in two colorectal cancer single-cell cohorts. (a–c) GSE132465 patient-level tumour epithelial and fibroblast-lineage summaries for source-available SenMayo119, SASP25 and MKI67. (d–f) GSE166555 patient-compartment pseudobulk summaries for source-available SenMayo114, SASP25 and MKI67. Grey segments join compartments from the same patient; black crossbars show medians. Displayed differences are fibroblast minus epithelial, whereas the MKI67 direction annotation counts patients with higher epithelial values. Exact two-sided Wilcoxon signed-rank P values and patient-bootstrap confidence intervals are reported. Cohort-specific score scales were not pooled

Table 2. Patient-level fibroblast-minus-epithelial contrasts in two single-cell cohorts

Cohort and outcome	Genes	Patients	Effect estimate (95% CI)	Direction	Exact P
GSE132465 SenMayo	119	23	Median 0.466 (0.418 to 0.553)	23/23 higher	2.38×10^{-7}
GSE132465 SASP	25	23	Median 0.757 (0.612 to 0.863)	23/23 higher	2.38×10^{-7}
GSE132465 MKI67	1	23	Median −0.118 (−0.137 to −0.074)	21/23 lower	1.81×10^{-4}
GSE166555 SenMayo	114	10	Median 0.149 (0.044 to 0.336)	9/10 higher	0.0195
GSE166555 SASP	25	10	Median 0.475 (0.170 to 0.740)	10/10 higher	0.00195
GSE166555 MKI67	1	10	Median −2.033 (−2.878 to −0.874)	9/10 lower	0.00391

Note. CI, confidence interval; SASP, senescence-associated secretory phenotype. Direction is relative to the epithelial compartment. P values are from exact two-sided paired Wilcoxon signed-rank tests; confidence intervals are patient-level percentile bootstrap intervals

The GSE166555 analysis reproduced the direction of the compartment contrast on its cohort-specific pseudobulk scale. SenMayo was higher in CAF-like than epithelial compartments in 9 of 10 patients (median difference, 0.149; 95% CI, 0.044 to 0.336; exact P = 0.0195). SASP25 was higher in 10 of 10 patients (median difference, 0.475; 95% CI, 0.170 to 0.740; P = 0.00195), and MKI67 was lower in 9 of 10 (median difference, −2.033; 95% CI, −2.878 to −0.874; P = 0.00391). The exact 111-gene common core retained the compartment direction in GSE132465 (mean difference, 0.517; 95% CI, 0.463 to 0.571; 23 of 23 positive) and GSE166555 (0.189; 0.078 to 0.306; 9 of 10 positive). Source-available patient pairs are shown in Fig. 1d–f, and definition sensitivity is shown in Fig. S2; cohort-specific primary effects were not pooled.

The GSE132465 SenMayo contrast exceeded its matched-gene background

The observed median patient-level SenMayo fibroblast-minus-epithelial standardized difference was 0.466. The median of 5,000 expression-, variability- and detection-matched random sets was 0.216, and the 97.5th percentile was 0.339, giving an upper-tail empirical P = 2.00×10^{-4}. The corresponding SASP25 statistic was 0.757, compared with a null median of 0.403 and 97.5th percentile of 0.759 (empirical P = 0.0264; Fig. 2a and Fig. S4, panels a and b). Many matched signatures were nevertheless higher in fibroblasts, so the null analysis supports the relative magnitude of the observed contrast under the tested matching scheme rather than proving a senescence-specific cell identity.

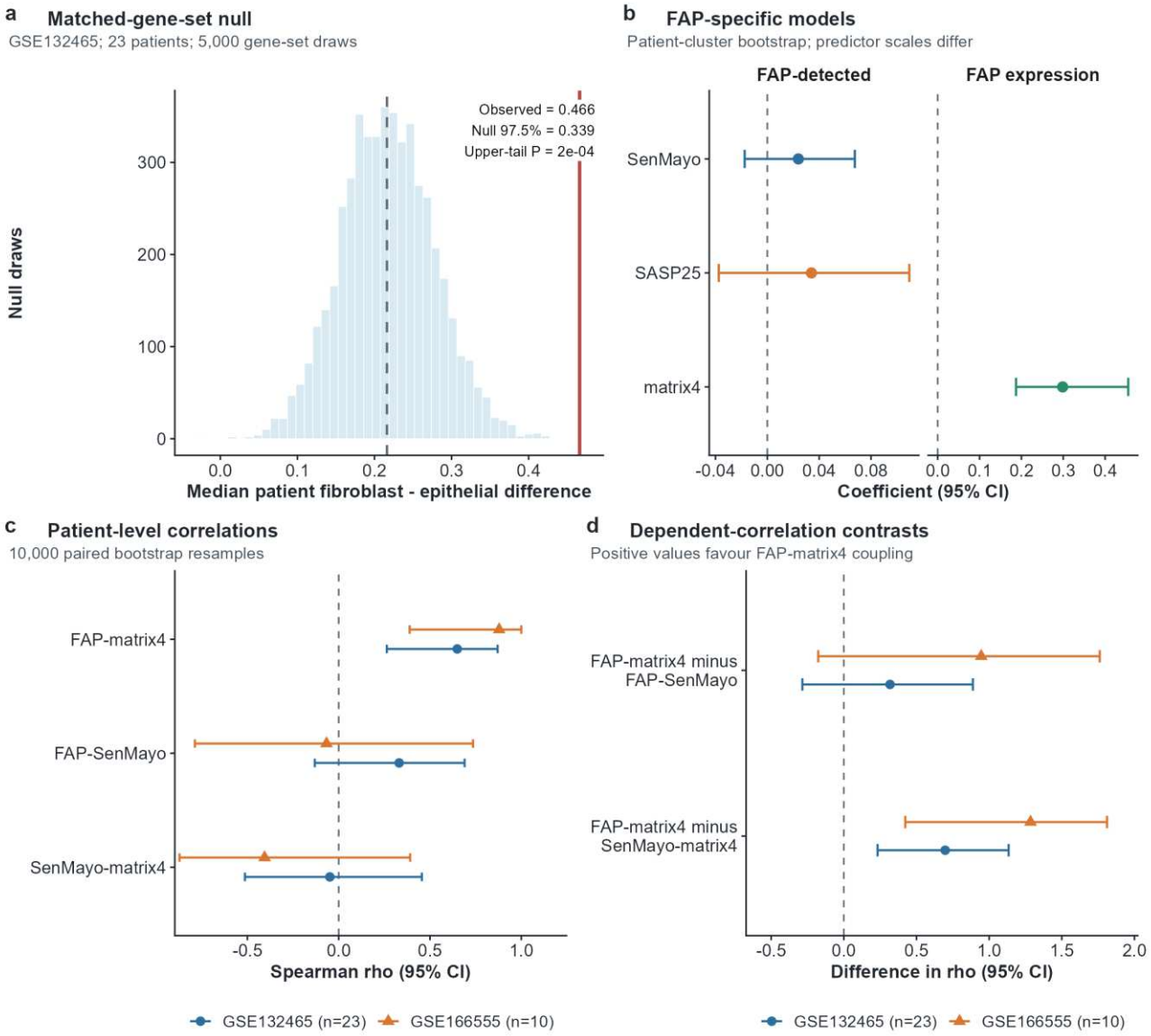


Fig. 2 Matched-null, FAP-specific and dependent-correlation analyses. (a) Distribution of 5,000 expression-, variability- and detection-matched GSE132465 gene-set statistics for the median patient SenMayo fibroblast-minus-epithelial contrast; the dashed line marks the null median and the red line the observed statistic. (b) Patient-cluster-bootstrap coefficients for binary FAP detection with SenMayo and SASP25, and for continuous FAP expression with matrix4, after the prespecified covariate adjustments. Predictor scales differ across facets. (c) Patient-level Spearman correlations in GSE132465 and GSE166555 with 10,000 paired-bootstrap confidence intervals. (d) Dependent-correlation contrasts recomputed within the same patient-bootstrap samples. Confidence intervals crossing zero are not interpreted as evidence of a difference

FAP detection did not identify additional SenMayo or SASP enrichment after patient-cluster resampling

Within GSE132465 tumour fibroblasts, cluster-bootstrap estimates for FAP-detected cells crossed zero for both SenMayo and SASP25 after depth, subtype, MKI67 and patient

adjustment. The SenMayo coefficient was 0.0239 (95% CI, −0.0175 to 0.0675; P = 0.274), and the SASP25 coefficient was 0.0340 (95% CI, −0.0374 to 0.1095; P = 0.363; Fig. 2b). Conventional random-intercept estimates were similar in magnitude but narrower. Thus, the cluster-aware analysis did not identify additional enrichment among FAP-detected fibroblasts, while smaller effects and protein-defined subgroups remain unresolved because the models were not designed as equivalence tests.

FAP covaried with matrix4, whereas comparative correlation estimates remained imprecise

The symmetric GSE132465 model showed positive within-fibroblast covariation between continuous FAP expression and matrix4 after depth, subtype and patient adjustment (cluster-bootstrap $\beta = 0.299$; 95% CI, 0.187 to 0.455; $P = 4.00 \times 10^{-4}$). At the patient level, FAP–matrix4 correlation was 0.649 (95% CI, 0.264 to 0.870) in GSE132465 and 0.879 (95% CI, 0.389 to 1.000) in GSE166555.

Direct bootstrap contrasts assessed whether the observed correlations differed in magnitude (Fig. 2c,d). FAP–matrix4 exceeded SenMayo–matrix4 in GSE132465 (difference in $\rho = 0.698$; 95% CI, 0.233 to 1.134) and GSE166555 (1.285; 95% CI, 0.423 to 1.811). The comparison of FAP–matrix4 with FAP–SenMayo was less precise in both cohorts: 0.318 (95% CI, −0.285 to 0.888) and 0.945 (95% CI, −0.176 to 1.761), respectively. This pattern supports FAP–matrix covariation while leaving its relative magnitude versus FAP–SenMayo unresolved.

Bulk and protein estimates were informative but measurement dependent

Marginal FAP13–matrix4 Spearman correlations were 0.930 in TCGA COAD/READ and 0.913 in GSE39582 (Fig. 3a). In TCGA, the partial correlation was 0.650 (95% CI, 0.571 to 0.714) after adjustment for the non-overlapping five-gene fibroblast score. Joint MCP-counter and EPIC adjustment using the full proxy definitions gave $\rho = 0.150$ (95% CI, 0.036 to 0.261). Because the full proxies incorporated several target genes, this attenuation could not be interpreted without the purge analysis. After target-pair purging, the joint partial correlation was 0.466 (95% CI, 0.363 to 0.556); the globally disjoint result was 0.466 (95% CI, 0.364 to 0.555; Fig. 3b,c and Table 3). FAP13–matrix4 covariation therefore remained positive, although its magnitude changed with proxy definition. Under globally disjoint joint adjustment, SenMayo–matrix4 was −0.016 (95% CI, −0.134 to 0.109; P = 0.761), leaving the residual monotonic association unresolved rather than establishing equivalence.

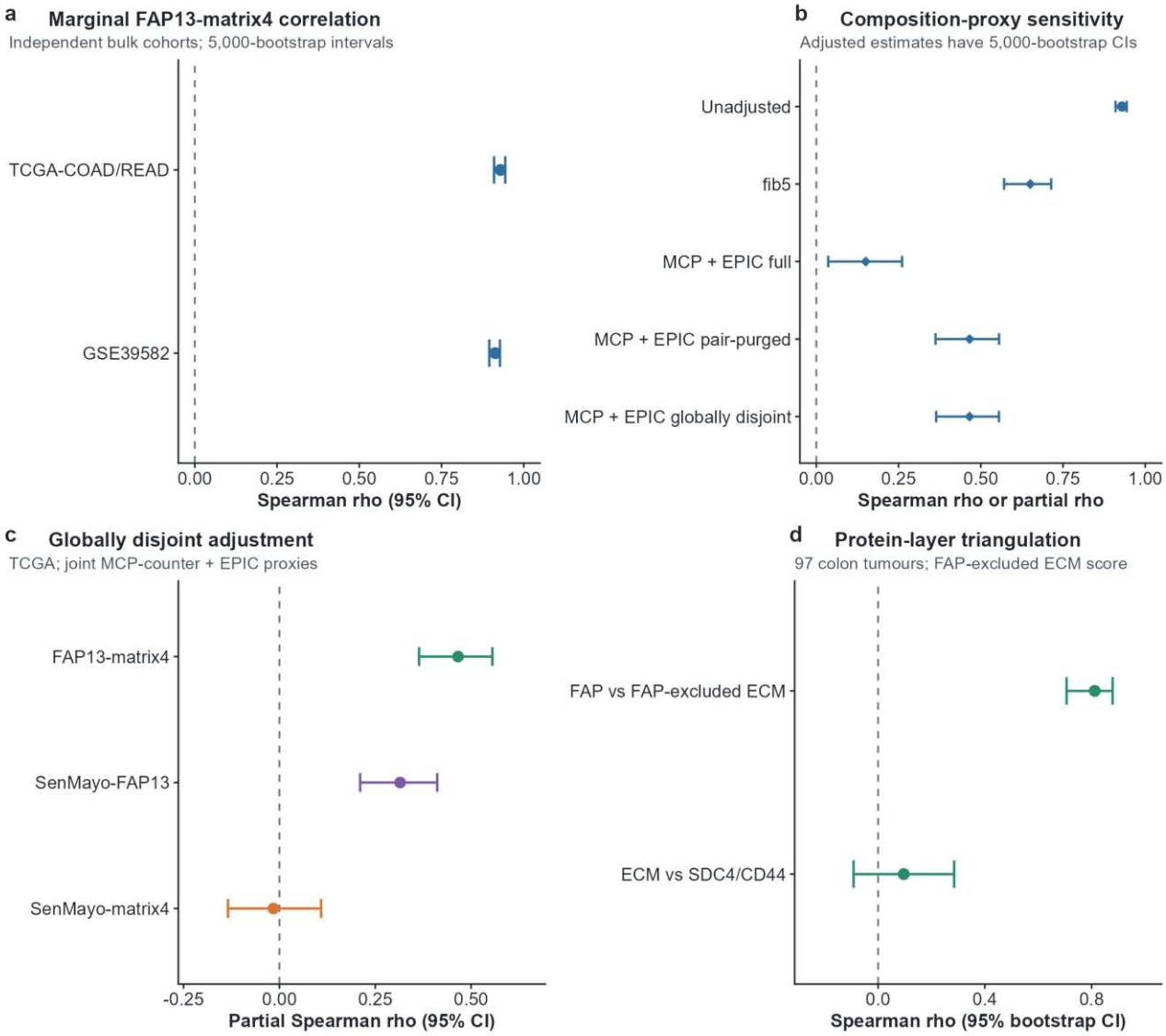


Fig. 3 Bulk composition-proxy sensitivity and protein-layer triangulation. (a) Marginal FAP13–matrix4 Spearman correlations in TCGA COAD/READ and GSE39582 with 5,000-bootstrap confidence intervals. (b) TCGA FAP13–matrix4 estimates before adjustment and after fib5, full joint MCP-counter plus EPIC, pair-purged joint, and globally disjoint joint adjustment. (c) TCGA globally disjoint joint-adjustment estimates for FAP13–matrix4, SenMayo–FAP13 and SenMayo–matrix4. (d) CPTAC colon-tumour protein correlations for FAP with a FAP-excluded extracellular-matrix score and for that matrix score with SDC4/CD44. Estimates are associative and were not pooled across measurement layers

Table 3. TCGA FAP13–matrix4 correlations under alternative fibroblast-content adjustments

Adjustment	Target-feature status	Spearman ρ	95% CI	P
None	Not applicable	0.930	0.910 to 0.944	6.38×10^{-166}
Five-gene fibroblast score	Non-overlapping	0.650	0.571 to 0.714	5.79×10^{-47}
MCP-counter + EPIC (full)	Target overlaps retained	0.150	0.036 to 0.261	0.00342
MCP-counter + EPIC (pair-purged)	FAP13 and matrix4 features removed	0.466	0.363 to 0.556	8.70×10^{-22}
MCP-counter + EPIC (globally disjoint)	FAP13, matrix4 and SenMayo features removed	0.466	0.364 to 0.555	8.95×10^{-22}

Note. CI, patient-bootstrap confidence interval. The full MCP-counter and EPIC definitions contained three and two FAP13/matrix4 target-feature overlaps, respectively. The unadjusted confidence interval uses the 5,000 patient-bootstrap resamples shown in Fig. 3a

CPTAC provided orthogonal protein-level support for FAP–matrix covariation. FAP protein correlated with the FAP-excluded COL1A1/COL1A2/FN1 score (ρ = 0.812; 95% CI, 0.706 to 0.879; false-discovery rate = 1.33×10^{-23}; n = 97). This result extends FAP–matrix covariance to the tumour-protein layer (Fig. 3d), while senescent phenotype and pathway direction remain outside the measurement scope.

Spatial localization was heterogeneous across cohorts

The spatial cohorts did not show uniform localization under the prespecified integration rule. Under the original cohort-specific definitions, GSE280315 showed a positive SenMayo stromal-minus-tumour effect in 2 of 3 patients (median, 0.00255; 95% CI, −0.00995 to 0.05366; raw and Holm P = 0.75), the Valdeolivas cohort was positive in 6 of 6 patients (median, 0.0961; 95% CI, 0.0405 to 0.1687; raw P = 0.03125; Holm P = 0.0625), and corrected GSE334323 was positive in 1 of 3 sections. The 111-gene common-core sensitivity produced mean patient effects of 0.0313 in GSE280315 (3 of 3 positive), 0.1196 in the Valdeolivas cohort (6 of 6 positive) and −0.0600 in GSE334323 (1 of 3 positive; Fig. 4a,b and Table 4). These cohort-specific results were not pooled.

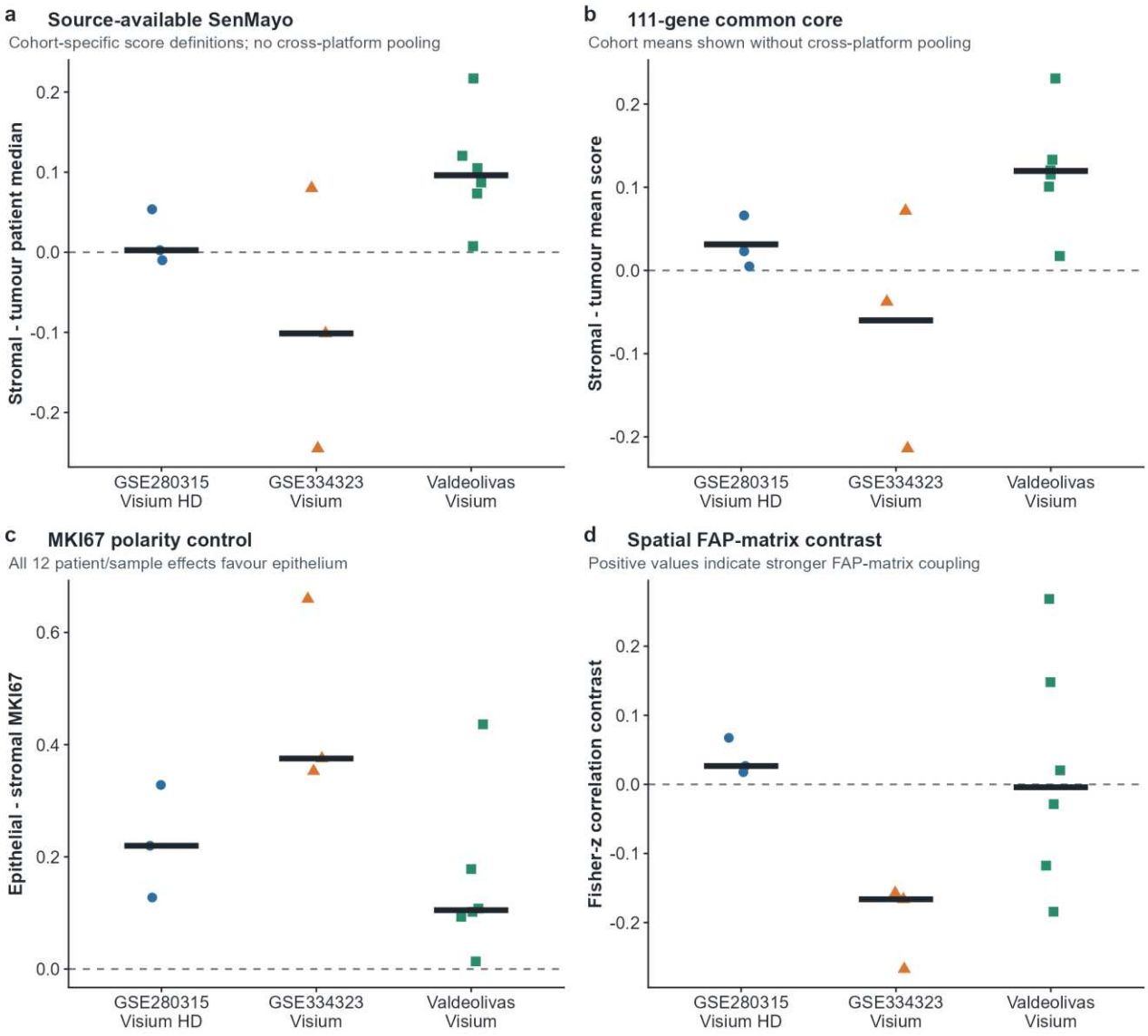


Fig. 4 Spatial generalizability across three colorectal cancer cohorts. (a) Patient or independent-section stromal-minus-tumour effects for each cohort's source-available SenMayo definition. (b) Effects after rescoring with the exact 111-gene common core. (c) Epithelial-minus-stromal MKI67 effects as a polarity control. (d) Fisher-z contrasts comparing stromal FAP–matrix4 correlation with the larger of FAP–SenMayo and SenMayo–matrix4 correlations. Points are patient or independent-section effects. Black crossbars show cohort medians in panels a, c and d and the prespecified cohort means in panel b. Platforms and annotation rules differ; no cross-platform pooled estimate or P value was calculated

Table 4. Cohort-specific spatial effects and directional controls

Cohort	Original H1 effect	111-gene common-core H1	H2 direction	MKI67 polarity
GSE280315	Median 0.00255 (−0.00995 to 0.05366); 2/3 positive; Holm P = 0.75	Mean 0.0313; 3/3 positive; P = 0.25	3/3 positive	Tumour > stromal in 3/3
Valdeolivas	Median 0.0961 (0.0405 to 0.1687); 6/6 positive; raw P = 0.03125; Holm P = 0.0625	Mean 0.1196; 6/6 positive; P = 0.03125	3/6 positive	Tumour > stromal in 6/6
GSE334323	Median −0.1013; 1/3 positive; descriptive	Mean −0.0600; 1/3 positive; P = 0.75	0/3 positive	Tumour > stromal in 3/3

Note. H1 is the stromal-minus-tumour SenMayo contrast. H2 is the prespecified within-stromal dependent-correlation contrast. Effects and P values are cohort specific and were not pooled. Positive MKI67 polarity denotes higher tumour/epithelial than stromal expression

The supportive centered matched-null tests placed the GSE280315 and Valdeolivas source-available SenMayo statistics above their respective expression- and detection-matched null centres. The observed-minus-null-centre differences were 0.128 and 0.181, respectively, with centered absolute empirical P = 2.00×10^{-4} for each (Fig. S4, panels c and d). However, the H2 direction counts were 3 of 3, 3 of 6 and 0 of 3 for GSE280315, Valdeolivas and GSE334323, respectively, and no cross-platform pooled inference was performed (Fig. 4d). MKI67 epithelial-minus-stromal direction was positive in all 12 spatial patients or sections (Fig. 4c). Taken together, the spatial data show context-dependent localization and a stable proliferative polarity control rather than universal spatial replication of the SenMayo contrast.

CAF marker programme associations with SenMayo were not consistently reproduced across cohorts

After library-size and MKI67 adjustment, the three adapted CAF programmes showed small, non-significant mean patient associations with SenMayo in GSE132465 (back-transformed correlations: MMP-positive trophocyte-like, 0.026; MMP-negative trophocyte-like, −0.068; telocyte-like, 0.063; all $q \geq 0.337$). In GSE166555, the corresponding estimates were 0.206 (95% patient-bootstrap CI, 0.068 to 0.331; $q = 0.0313$), 0.158 (95% CI, 0.028 to 0.289; $q = 0.0534$) and 0.409 (95% CI, 0.339 to 0.482; $q = 1.66 \times 10^{-5}$). No programme met the positive-association criterion at $q < 0.05$ in both cohorts (Fig. S7, panel a; Table S8). The bootstrap intervals and t-test-based q values use different inferential procedures; a percentile interval excluding zero did not override the multiplicity-adjusted test.

Only the GSE166555 telocyte-like programme lay clearly above its approximately matched random-programme reference; the GSE132465 MMP-negative programme instead lay below

its reference (empirical q = 0.0030 for each; Fig. S7, panel b). In 380 TCGA tumours, marginal programme–SenMayo correlations ranged from 0.738 to 0.814. After fib5 adjustment, the MMP-positive estimate remained 0.343 (95% CI, 0.238 to 0.440), whereas the MMP-negative and telocyte-like estimates were 0.088 (95% CI, −0.014 to 0.195) and 0.067 (95% CI, −0.044 to 0.176), respectively (Fig. S7, panel c). These tissue-level associations did not establish a CAF subtype with consistently elevated senescence-associated transcription across cohorts.

Discussion

Principal inference

Across two colorectal cancer single-cell cohorts, patient-level analyses localized senescence-associated transcription preferentially to fibroblast compartments. The result was present for each cohort's source-available SenMayo genes and persisted when both cohorts were rescored with an exact 111-gene common core. The discovery contrast also exceeded the tested matched-gene background. At the same time, cluster-aware estimates did not support additional SenMayo or SASP enrichment among FAP-detected tumour fibroblasts. These results support preferential fibroblast-compartment localization rather than the identification of a uniquely senescent FAP-defined subgroup.

This distinction is biologically consequential. A transcriptomic panel aggregates genes that can participate in senescence, inflammation, tissue repair and stress responses. The lower fibroblast MKI67 distribution is directionally compatible with reduced proliferation, but neither it nor SenMayo demonstrates durable arrest. Current consensus recommends multiple markers adapted to tissue and cell context [9–11]. The present study identifies where such orthogonal assays should be concentrated; it does not replace them.

FAP–matrix covariation is robust as an association but sensitive to measurement context

FAP expression covaried with the collagen/fibronectin matrix score within tumour fibroblasts in both single-cell cohorts, in TCGA after target-purged adjustment and at the protein level in CPTAC. This convergence supports FAP–matrix covariation across measurement layers. It does not establish that FAP causes matrix deposition. Direct bootstrap contrasts were important here: FAP–matrix correlation exceeded SenMayo–matrix correlation, but its difference from FAP–SenMayo correlation remained uncertain. Interpreting separate significance labels as preferential FAP–matrix linkage would have overstated the direct comparison.

Bulk estimates varied substantially with the definition of the composition proxy. The full MCP-counter and EPIC proxies contained genes from FAP13 or matrix4, so residualization partly adjusted for the targets themselves. Removing those overlaps increased the joint residual estimate from 0.150 to 0.466. Conversely, the non-overlapping five-gene proxy retained an estimate of 0.650. Each coefficient remains conditional on the proxy used to represent fibroblast content. The tissue-level association remains positive, but its magnitude changes materially with the composition model.

Relation to experimental and recent computational colorectal cancer studies

Recent experimental work provides the functional context that retrospective scores cannot. Ge et al. linked single-cell classification to tissue staining, fibroblast–T-cell–organoid co-culture, lipid tracing and animal experiments, tracing a CD36-dependent lipid-transfer route from senescent fibroblasts to impaired T-cell function [15]. Yang et al. used spatial, organoid and genetic models to show that macrophage-derived IL1B can induce CAF senescence and SASP-mediated chemotherapy resistance [16]. By directly measuring phenotype or perturbation, those studies address a functional question that complements the present patient-level localization analysis.

The COMP-positive CAF study is closer in data architecture because it integrated bulk, single-cell, spatial and tissue evidence [17]. Its predicted COMP–CD47 interaction was appropriately framed as a candidate mechanism. The contribution here is complementary: source-gene representation was separated from a common-core sensitivity, patient rather than cell or spot was used for inference, FAP specificity was directly tested, composition proxies were audited for target incorporation, dependent correlations were compared within the same bootstrap samples and a non-supportive spatial cohort was retained. These choices narrow the claim but improve traceability of what each modality can establish.

The exploratory CAF analysis places a further limit on subtype interpretation. Source-detectability-adapted marker programmes did not show a reproducible positive relationship with SenMayo across the two single-cell cohorts, despite strong marginal bulk correlations. Sequencing depth, fibroblast subtype composition and the choice of expression proxy affected these comparisons. The organoid experiments of Buissant des Amorie et al. demonstrate that trophocyte-like CAFs can induce epithelial oncofetal plasticity [18]; they do not establish that this activity is a senescence phenotype. Their fibroblast lines were generated using hTERT and BMI1, further separating those functional models from a test of persistent fibroblast arrest. Our results support keeping CAF marker identity, senescence-associated transcription and epithelial plasticity as separate questions for future tissue and co-culture experiments.

Why the spatial results differ

The spatial cohorts differed in platform, physical resolution, feature design and compartment assignment. GSE280315 used high-definition bins and author deconvolution labels; the Valdeolivas cohort used lower-resolution Visium spots and pathologist annotations; GSE334323 used a custom FFPE probe reference and marker-based fallback labels. Spatial spots or bins also mix cell state, local abundance and technical capture. A shared gene set cannot remove these differences. Common-core effects were positive in GSE280315 and Valdeolivas but mixed or negative in GSE334323. This variation indicates cohort dependence and limits generalization across spatial settings.

The matched-null results place the observed statistics in two cohorts outside the tested expression- and detection-matched background, while leaving platform and annotation differences unchanged. The common-core contrast remained mixed or negative in

GSE334323. Consistent MKI67 polarity provides a useful compartment control but does not identify senescence.

Strengths

Strengths include reconstruction of GSE132465 from official public counts, patient-level inferential units, an external single-cell cohort, direct resampling of dependent correlations, target-purged composition adjustment, transparent preservation of a post-outcome technical correction and complete presentation of spatial heterogeneity. The core analysis code, source data and diagnostic outputs are publicly archived, and the accompanying CAF data document the exploratory extension.

Limitations

Several limitations remain. All analyses were retrospective and exploratory. FAP transcript detection is affected by dropout and is not equivalent to protein-defined FAP positivity. SenMayo and SASP are transcriptional summaries rather than assays of persistent arrest. The FAP-specific model used one single-cell cohort with 23 patient clusters, limiting random-slope estimation and precision for small subgroup effects; the external paired single-cell analysis included 10 patients. Bulk deconvolution proxies remained model dependent even after target purging. The three spatial cohorts included 3, 6 and 3 independent patients or sections, used different platforms and labels, and GSE334323 required a post-outcome host-feature correction. The CAF marker analysis required a source-detectability adaptation and approximate random-programme matching; different significance labels across cohorts were not a formal test of between-cohort heterogeneity. No multiplex protein localization, senescence-associated β-galactosidase, DNA-damage response, longitudinal arrest assay or perturbation was available. The study therefore cannot support therapeutic claims or a causal immune-communication mechanism.

Future validation

Organoid dependency maps may help select epithelial partners for CAF co-culture. Herranz-Ors et al. linked molecular profiles to CRISPR fitness effects in a tumour-derived organoid biobank [36]. Such resources can support model selection based on genotype and epithelial fitness requirements. Testing a CAF-derived signal would still require a defined CAF or ligand perturbation and an appropriate rescue condition, because endogenous gene knockout and exposure to an external stromal signal address different biological questions.

A decisive next study would combine independently segmented colorectal cancer tissue with multiplex FAP protein, p16, p21, lamin B1 and MKI67 measurement, followed by isolation or induction of CAFs to test persistent arrest and secreted factors. A prespecified CAF–tumour-organoid–CD8 T-cell co-culture could then compare FAP-defined and senescence-marker-defined CAF conditions while measuring lipid transfer, lipid peroxidation and tumour-cell killing. This staged design converts the present localization result into falsifiable tissue and functional tests without treating transcript correlation as mechanism.

Conclusion

Two colorectal cancer single-cell cohorts localized senescence-associated transcription preferentially to fibroblast rather than epithelial tumour compartments. FAP–matrix covariation persisted across single-cell, target-purged bulk and protein measurements, with magnitudes that depended on measurement context. FAP detection did not define a subgroup with additional SenMayo/SASP enrichment. Heterogeneous spatial effects limit universal generalization and identify multiplex tissue phenotyping and functional co-culture as the next tests.

Declarations

Funding

This work was supported by the National Natural Science Foundation of China (No. 82360467) and the Science and Technology Fund Project of Guizhou Provincial Science and Technology Program (Nos. QKH JC-ZK [2022]-358 and [2022]-348). The funders had no role in study design, data collection and analysis, the decision to publish or preparation of the manuscript.

Competing interests

The authors declare no competing financial or non-financial interests.

Ethics approval and consent to participate

Ethics approval and consent procedures for each contributing cohort are described in its source publication. Our analyses were limited to publicly released, de-identified records; we did not contact participants or collect additional specimens.

Clinical trial number: not applicable.

Consent for publication

Not applicable. This manuscript contains no identifiable individual participant information.

Data and code availability

TCGA COAD/READ data are available through the Genomic Data Commons and UCSC Xena. GSE39582, GSE132465, GSE166555, GSE280315 and GSE334323 are available through the Gene Expression Omnibus. The Valdeolivas spatial data are available from Zenodo (version DOI 10.5281/zenodo.7760264; concept DOI 10.5281/zenodo.7551712), and CPTAC colon proteomics are available through cBioPortal. The analysis scripts, result files, figure source data, supplementary tables and SHA-256 checksums, including the exploratory CAF module, are available in GitHub release v2.1.1 (https://github.com/wangjinglin0905-gif/fap-ecm-cldn-crc-reproducibility/releases/tag/v2.1.1). The same release is archived in Zenodo (version DOI 10.5281/zenodo.22668754; concept DOI 10.5281/zenodo.21441731). The source CAF marker table is available with the Nature publication (doi: 10.1038/s41586-026-10344-7). Adapted marker definitions, patient-level estimates, full sensitivity results and figure source data for the exploratory CAF analysis accompany this manuscript in Table S8 and S1 Data.

AI-assisted analysis and manuscript preparation

OpenAI Codex, including GPT-6 Astra for the final manuscript audit, was used for code review and document formatting. The authors reviewed the analysis contracts, source tables, code, references and manuscript text; the tool did not generate biological measurements or qualify for authorship, and the authors retain responsibility for the work.

Author contributions

J.W. and J.Y. conceived the review. J.W. coordinated the literature search and drafted the manuscript. P.S., W.Z., Y.Z. and Y.T. contributed to literature screening, evidence checking, figure development and critical revision. J.Y. acquired funding and supervised the work. All authors critically reviewed and approved the final manuscript.

Acknowledgments

The authors thank the investigators and participants who generated and shared the public datasets analysed in this study.

References

1. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2024;74:229–263. https://doi.org/10.3322/caac.21834

2. Sahai E, Astsaturov I, Cukierman E, DeNardo DG, Egeblad M, Evans RM, et al. A framework for advancing our understanding of cancer-associated fibroblasts. Nat Rev Cancer. 2020;20:174–186. https://doi.org/10.1038/s41568-019-0238-1

3. Fitzgerald AA, Weiner LM. The role of fibroblast activation protein in health and malignancy. Cancer Metastasis Rev. 2020;39:783–803. https://doi.org/10.1007/s10555-020-09909-3

4. Isella C, Terrasi A, Bellomo SE, Petti C, Galatola G, Muratore A, et al. Stromal contribution to the colorectal cancer transcriptome. Nat Genet. 2015;47:312–319. https://doi.org/10.1038/ng.3224

5. Isella C, Terrasi A, Bellomo SE, Petti C, Galatola G, Muratore A, et al. Erratum: Corrigendum: Stromal contribution to the colorectal cancer transcriptome. Nat Genet. 2016;48:1296. https://doi.org/10.1038/ng1016-1296d

6. Calon A, Lonardo E, Berenguer-Llergo A, Espinet E, Hernando-Momblona X, Iglesias M, et al. Stromal gene expression defines poor-prognosis subtypes in colorectal cancer. Nat Genet. 2015;47:320–329. https://doi.org/10.1038/ng.3225

7. Qi J, Sun H, Zhang Y, Wang Z, Xun Z, Li Z, et al. Single-cell and spatial analysis reveal interaction of FAP+ fibroblasts and SPP1+ macrophages in colorectal cancer. Nat Commun. 2022;13:1742. https://doi.org/10.1038/s41467-022-29366-6

8. Saul D, Kosinsky RL, Atkinson EJ, Doolittle ML, Zhang X, LeBrasseur NK, et al. A new gene set identifies senescent cells and predicts senescence-associated pathways across tissues. Nat Commun. 2022;13:4827. https://doi.org/10.1038/s41467-022-32552-1

9. González-Gualda E, Baker AG, Fruk L, Muñoz-Espín D. A guide to assessing cellular senescence in vitro and in vivo. FEBS J. 2021;288:56–80. https://doi.org/10.1111/febs.15570

10. Suryadevara V, Hudgins AD, Rajesh A, Pappalardo A, Karpova A, Dey AK, et al. SenNet recommendations for detecting senescent cells in different tissues. Nat Rev Mol Cell Biol. 2024;25:1001–1023. https://doi.org/10.1038/s41580-024-00738-8

11. Freund A, Laberge RM, Demaria M, Campisi J. Lamin B1 loss is a senescence-associated biomarker. Mol Biol Cell. 2012;23:2066–2075. https://doi.org/10.1091/mbc.E11-10-0884

12. Faget DV, Ren Q, Stewart SA. Unmasking senescence: context-dependent effects of SASP in cancer. Nat Rev Cancer. 2019;19:439–453. https://doi.org/10.1038/s41568-019-0156-2

13. Belle JI, Sen D, Baer JM, Liu X, Lander VE, Ye J, et al. Senescence defines a distinct subset of myofibroblasts that orchestrates immunosuppression in pancreatic cancer. Cancer Discov. 2024;14:1324–1355. https://doi.org/10.1158/2159-8290.CD-23-0428

14. Hattangady NG, Carter K, Maroni-Rana B, Wang T, Ayers JL, Yu M, et al. Mapping the core senescence phenotype of primary human colon fibroblasts. Aging (Albany NY). 2024;16:3068–3087. https://doi.org/10.18632/aging.205577

15. Ge M, Sun S, Chen W, Xu Z, Kang D, Wang Z, et al. Senescent fibroblasts drive CD8+ T cell dysfunction in colorectal cancer via CD36-mediated lipid transfer and peroxidation. J Transl Med. 2026;24:310. https://doi.org/10.1186/s12967-025-07636-3

16. Yang S, Chen Y, Liu S, Zuo A, Ba Y, Zhang Y, et al. Macrophage-induced senescent cancer-associated fibroblasts promote SASP-mediated chemoresistance in colorectal cancer. Cancer Res. 2026 Jul 8 [Epub ahead of print]. https://doi.org/10.1158/0008-5472.CAN-25-4870

17. Chen Y, Zhang Y, Lv C. Targeting a candidate stromal immune checkpoint: a cellular senescence signature identifies COMP+ cancer-associated fibroblasts associated with CD8+ T-cell exclusion via a predicted COMP–CD47 axis in colorectal cancer. J Adv Res. 2026 Aug 20 [Epub ahead of print]. https://doi.org/10.1016/j.jare.2026.08.047

18. Buissant des Amorie JR, Hageman JH, Brunner SR, van der Horst SEM, Puschhof MC, van Hoeck A, et al. Emergence of oncofetal plasticity is ubiquitous in early colorectal cancers. Nature. 2026;654:229–239. https://doi.org/10.1038/s41586-026-10344-7

19. Zimmerman KD, Espeland MA, Langefeld CD. A practical solution to pseudoreplication bias in single-cell studies. Nat Commun. 2021;12:738. https://doi.org/10.1038/s41467-021-21038-1

20. Squair JW, Gautier M, Kathe C, Anderson MA, James ND, Hutson TH, et al. Confronting false discoveries in single-cell differential expression. Nat Commun. 2021;12:5692. https://doi.org/10.1038/s41467-021-25960-2

21. Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. Nature. 2012;487:330–337. https://doi.org/10.1038/nature11252

22. Marisa L, de Reyniès A, Duval A, Selves J, Gaub MP, Vescovo L, et al. Gene expression classification of colon cancer into molecular subtypes: characterization, validation, and prognostic value. PLoS Med. 2013;10:e1001453. https://doi.org/10.1371/journal.pmed.1001453

23. Lee HO, Hong Y, Etlioglu HE, Cho YB, Pomella V, Van den Bosch B, et al. Lineage-dependent gene expression programs influence the immune landscape of colorectal cancer. Nat Genet. 2020;52:594–603. https://doi.org/10.1038/s41588-020-0636-z

24. Uhlitz F, Bischoff P, Peidli S, Sieber A, Trinks A, Lüthen M, et al. Mitogen-activated protein kinase activity drives cell trajectories in colorectal cancer. EMBO Mol Med. 2021;13:e14123. https://doi.org/10.15252/emmm.202114123

25. Oliveira MF de, Romero JP, Chung M, Williams SR, Gottscho AD, Gupta A, et al. High-definition spatial transcriptomic profiling of immune cell populations in colorectal cancer. Nat Genet. 2025;57:1512–1523. https://doi.org/10.1038/s41588-025-02193-3

26. Valdeolivas A, Amberg B, Giroud N, Richardson M, Gálvez EJC, Badillo S, et al. Profiling the heterogeneity of colorectal cancer consensus molecular subtypes using spatial transcriptomics. NPJ Precis Oncol. 2024;8:10. https://doi.org/10.1038/s41698-023-00488-4

27. Aburajab R, Karmouch JL, Jenq RR, Craig DW. Custom probe-based spatial transcriptomics enables microbiome detection in FFPE colorectal cancer tissue. bioRxiv [Preprint]. 2026 [posted 2026 Jul 30]. https://doi.org/10.64898/2026.07.29.741338

28. Vasaikar S, Huang C, Wang X, Petyuk VA, Savage SR, Wen B, et al. Proteogenomic analysis of human colon cancer reveals new therapeutic opportunities. Cell. 2019;177:1035–1049.e19. https://doi.org/10.1016/j.cell.2019.03.030

29. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov. 2012;2:401–404. https://doi.org/10.1158/2159-8290.CD-12-0095

30. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal. 2013;6:pl1. https://doi.org/10.1126/scisignal.2004088

31. Becht E, Giraldo NA, Lacroix L, Buttard B, Elarouci N, Petitprez F, et al. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. Genome Biol. 2016;17:218. https://doi.org/10.1186/s13059-016-1070-5

32. Becht E, Giraldo NA, Lacroix L, Buttard B, Elarouci N, Petitprez F, et al. Erratum to: Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. Genome Biol. 2016;17:249. https://doi.org/10.1186/s13059-016-1113-y

33. Racle J, de Jonge K, Baumgaertner P, Speiser DE, Gfeller D. Simultaneous enumeration of cancer and immune cell types from bulk tumor gene expression data. eLife. 2017;6:e26476. https://doi.org/10.7554/eLife.26476

34. Avila Cobos F, Alquicira-Hernandez J, Powell JE, Mestdagh P, De Preter K. Benchmarking of cell type deconvolution pipelines for transcriptomics data. Nat Commun. 2020;11:5650. https://doi.org/10.1038/s41467-020-19015-1

35. Avila Cobos F, Alquicira-Hernandez J, Powell JE, Mestdagh P, De Preter K. Author Correction: Benchmarking of cell type deconvolution pipelines for transcriptomics data. Nat Commun. 2020;11:6291. https://doi.org/10.1038/s41467-020-20288-9

36. Herranz-Ors C, Bhosle SG, Beck AE, Gilbert JGR, Picco G, Espejo Valle-Inclan J, et al. A tumour-derived organoid biobank maps cancer gene dependencies. Nature. 2026 Aug 5 [Epub ahead of print]. https://doi.org/10.1038/s41586-026-10830-y

Supplementary Information captions

Supplementary_Information.pdf contains Figs. S1–S7, supplementary methods and the table index. Supplementary_Tables.zip contains Tables S1–S8 and four companion CSV files. S1_Data_CAF_programme_source_data.zip provides the exploratory CAF source data (S1 Data).

Fig. S1 Dataset flow, eligibility criteria and independent analysis units

Fig. S2 GSE132465 cell-count and 111-gene common-core sensitivities

Fig. S3 FAP-detected single-gene models and model diagnostics

Fig. S4 Expression-matched null diagnostics and corrected spatial empirical statistics

Fig. S5 Composition-proxy feature overlap and target-purged sensitivity

Fig. S6 Spatial compartment definitions, quality control and definition sensitivities

Fig. S7 Exploratory associations of adapted CAF marker programmes with senescence-associated transcription

Table S1. Dataset accessions, consumed files, sample eligibility and checksums

Table S2. Signature membership, overlap removal, represented genes and the 111-gene common core

Table S3. Full single-cell compartment statistics and bootstrap intervals

Table S4. Full coefficients, standard errors, confidence intervals, P values and diagnostics for GSE132465 models

Table S5. Dependent-correlation bootstrap results in both single-cell cohorts

Table S6. Full TCGA composition-proxy overlap audit and partial-correlation sensitivity analyses

Table S7. Cohort-specific spatial effects, multiplicity adjustment and GSE334323 pre/post-correction provenance

Table S8. Full CAF programme association estimates, inferential procedures and matched-null results

S1 Data. CAF marker definitions, gene coverage, patient-level coefficients, sensitivity results and figure source data

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryInformation.pdf
- SupplementaryTables.zip
- S1DataCAFprogrammesourcedata.zip