

Tissue composition quantitatively reproduces bulk transcriptomic differences between paired primary gastric tumours and solid peritoneal metastases

Chao Huang

Second Affiliated Hospital of Nanchang University

Qin Yu

University of Science and Technology of China

Zheng-Ming Zhu

Second Affiliated Hospital of Nanchang University

Yinfei Kong

yikong@fullerton.edu

California State University, Fullerton

Research Article

Keywords: gastric cancer, peritoneal metastasis, tumour purity, tissue composition, transcriptional programmes, paired specimens

Posted Date: October 27th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11252276/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Bulk transcriptomic contrasts between primary gastric cancers and peritoneal metastases combine cancer-cell state with tissues present in each specimen. We tested whether tissue composition is quantitatively sufficient to explain such contrasts. Before site comparison, we froze a 100-gene malignant-epithelium-biased programme, control programmes, and non-overlapping host-compartment markers. The primary endpoint was the paired score difference in 17 patients. The programme was lower in 15 patients (mean difference -0.087 , 95% confidence interval -0.136 to -0.038 ; $P = 0.0017$). Fibroblast, adipocyte, endothelial and stromal scores increased, whereas gastric-epithelial markers decreased. Adjustment for fibroblast content reduced the remaining difference to -0.014 ($P = 0.457$). Pseudo-bulk calibration showed that the primary difference was reproduced by an epithelial-content difference of -0.305 ; 12 programmes, including epithelial–mesenchymal transition, implied a common median difference of -0.311 . Four scoring procedures retained the same direction, and an independent four-gene epithelial marker set also decreased. In a separate cohort comprising 66 specimens from 14 patients, the frozen score tracked published histological purity within patients (slope 0.126 ; patient-cluster bootstrap 95% confidence interval 0.052 to 0.250). The paired endpoint in five patients from that cohort was imprecise (-0.031 , 95% confidence interval -0.124 to 0.063). Thus, a single difference in epithelial content quantitatively reproduces the bulk programme changes. Cancer-cell-state claims from primary-versus-peritoneal bulk contrasts require cell-resolved confirmation or an explicit composition model.

Introduction

Gastric cancer remains one of the leading causes of cancer death worldwide (Bray et al. 2024). The peritoneum is among the most frequently involved metastatic sites, particularly in diffuse and poorly cohesive tumours: synchronous peritoneal metastasis is present in roughly one in seven patients at diagnosis and the proportion has risen over time, and when metachronous disease is included about one third of patients are affected (Thomassen et al. 2014; Riihimäki et al. 2016; Koemans et al. 2021; Rijken et al. 2024; Guchelaar et al. 2025). Survival is short even when systemic therapy is given (Koemans et al. 2021), and the trials that established current first-line treatment were not specific to peritoneal disease (Janjigian et al. 2021). Even free intraperitoneal tumour cells without macroscopic implants carry prognostic weight (Jamel et al. 2018), which is one reason the field distinguishes ascites, lavage and solid implants.

Because peritoneal metastases are rarely resected, paired specimens from the same patient are scarce, and the handful of public paired cohorts carry much of the molecular reasoning about the disease. Studies of such pairs have reported a gradient of epithelial–mesenchymal transition from normal tissue through the primary tumour to the metastasis, higher cancer-associated fibroblast and macrophage content, adipogenic signal, and remodelled ecotypes in peritoneal lesions (Kim et al. 2024; Zhao et al. 2024; Li et al. 2025; Charton et al. 2026). Single-cell work has added stromal–myeloid niches and route-specific programmes (Wang et al. 2023; Cheng et al. 2024; Hu et al. 2024; Li et al. 2025; Wang et al. 2025). A recurring conclusion is that the difference between a primary tumour and its peritoneal metastasis is carried largely by non-malignant compartments.

That conclusion sits on a measurement assumption that has not been tested in the paired setting. A peritoneal implant is not the same tissue as a gastric resection. It is surrounded by peritoneal and omental

host tissue — mesothelium, adipose, fibroblasts, vessels — while the matched primary specimen contains gastric mucosa and muscularis propria. Any score computed on bulk tissue therefore reflects both what the cancer cells are doing and what else was in the block. The direction of the problem is visible in published results: the genes reported as most increased in peritoneal lesions include mesothelial and adipocyte markers (Charton et al. 2026), and it is established more generally that most of the apparent mesenchymal signal in bulk tumours comes from fibroblasts rather than cancer cells (Tyler and Tirosh 2021). What has been missing is a quantitative answer: given a measured compositional shift, how much of a programme difference should we expect even if cancer cells were unchanged?

To quantify this contribution, we selected in advance a single functional programme that is defined independently of gastric cancer, is biased towards malignant epithelium, and has been shown not to depend on YAP/TEAD signalling: the EMP1 + high-relapse-cell programme derived from colorectal cancer, which marks the residual disseminating cells responsible for metastatic relapse (Cañellas-Socias et al. 2022). We paired it with four control programmes, a deliberately stroma-restricted negative control, and seven host-compartment marker sets built from an independent normal-tissue single-cell atlas (Karlsson et al. 2021). Gene membership, the scoring rule, the coverage threshold, eligibility, and the single primary endpoint were frozen with hashes before any site contrast was computed.

We first tested whether the frozen programme changed between paired primary and solid peritoneal specimens within the same patient. We then assessed whether compartment scores moved as tissue admixture would predict and estimated the difference in cell content required to reproduce the observed programme difference without any change of state inside cells.

Materials and Methods

Data sources and roles

All data were public. Roles were fixed before analysis and are listed in Table 1. The primary within-patient test used GSE314812 (Charton et al. 2026), comprising 55 bulk whole-transcriptome specimens from 26 patients with advanced gastric cancer. The independent paired bulk check used GSE237876 (Kim et al. 2024) (66 specimens, 14 patients, 6 peritoneal specimens). Single-cell sources were used for calibration and description only: GSE183904 (Kumar et al. 2022) and GSE163558 (Jiang et al. 2022) for cell identity and measurement calibration, using their gastric primary and normal donors; GSE308231 (Wang et al. 2025) and the peritoneal donors of GSE183904 for descriptive cell-level observation; and GSE239676 (Cheng et al. 2024) for a separately defined fluid-context contrast. GSE234129 was confirmed by cell-barcode overlap to contain the same libraries as GSE239676 and was counted once.

Table 1

Sources, specimens, patients and specimen types, with the role assigned to each source before analysis.

Source	Role	Specimens	Patients	Primary	Solid PM	Other material	Paired patients
GSE314812	Primary paired bulk	55	26	19	26	4 other metastasis; 6 normal	17
GSE237876	External paired bulk / purity	66	14	38	6	22 other metastasis	5
GSE183904	Single-cell calibration	40	29	26	3	11 normal	1
GSE163558	Single-cell calibration	10	6	3	1	5 other metastasis; 1 normal	0
GSE308231	Cell-level description	6	0	3	3	—	0
GSE239676	Fluid context	68	20	19	0	10 ascites; 9 other metastasis; 16 normal	0
GSE228598	Metadata/context	28	28	0	0	15 ascites; 13 lavage	0

Detailed specimen, programme, audit and result records are provided in Supplementary Tables S1–S14 (Online Resource 2); supplementary methods and item legends are provided in Online Resource 1.

Controlled-access data and data available only on request to authors were not used. Optional spatial and clinical-recurrence modules were switched off by protocol, with reasons recorded.

Specimen and patient definitions

One row per biological specimen was assembled from GEO records, paper supplementary tables and two small curated maps, each with a source locator. Solid peritoneal metastasis was distinguished from omental metastasis, ascites, peritoneal lavage and normal peritoneum; ascites and lavage were never merged with solid tissue. Analysis was at the patient level, with one primary value, one metastasis value and one difference per patient.

In GSE314812 one patient contributed two pairs, a pre-treatment pair and a pair taken at recurrence one year after treatment. The protocol assigned the earliest eligible pair to the primary analysis, so the 18 pair units reported by the original study correspond to 17 independent patients. Preservation method differed within 13 of the 18 pairs, in both directions; 13 of the GSE314812 specimens were formalin-fixed. Per-patient treatment before sampling was not reported for any patient except the one with two visits and was recorded as unknown rather than as untreated.

Expression matrices

The GSE314812 supplementary matrix contains gene-level RSEM expected counts for 57,773 Ensembl GRCh37 identifiers (Li and Dewey 2011); the authors' TPM matrix is not public, so we did not attempt to reproduce it. GSE237876 provides transcript-level Salmon estimates per specimen (Patro et al. 2017), which we summed to gene level (Soneson et al. 2015). Both were mapped to Ensembl GRCh37 release 75 annotation and to approved HGNC symbols, retaining original identifiers and the mapping route. Single-cell sources were processed per sample: symbol harmonisation, per-sample robust quality control, doublet removal (Germain et al. 2021), graph-based clustering (Hao et al. 2021), marker-based lineage assignment using fixed lineage markers that do not overlap any scored programme, and expression-based copy-number scoring against same-sample non-epithelial reference cells to identify malignant epithelium (Tirosh et al. 2016).

Frozen programmes and scoring

The primary programme was the 100-gene core high-relapse-cell programme (Cañellas-Socias et al. 2022). Control programmes were hallmark epithelial–mesenchymal transition and hallmark G2/M checkpoint (Liberzon et al. 2015), the ESTIMATE stromal signature (Yoshihara et al. 2013), and a stress/immediate-early meta-programme (Gavish et al. 2023). The negative control comprised consensus YAP/TAZ–TEAD target genes restricted to those that are not epithelium-dominant in the Human Protein Atlas, chosen so that it should move if and only if stromal admixture changes. Seven host-compartment sets (gastric epithelium, mesothelium, adipocyte, fibroblast, smooth muscle, endothelium, immune) were built from Human Protein Atlas single-cell-type expression (Karlsson et al. 2021) by requiring at least 10 nCPM in that compartment and at least three times the maximum of every other compartment and excluding every gene belonging to any scored programme, so that compartment covariates could not be circular. Two pre-specified alternative definitions of the primary programme were also frozen.

Scores were within-sample equal-weight mean gene ranks on a common gene universe defined by expression level alone, without reference to site or group (15,888 genes in GSE314812). A programme was declared not evaluable on a platform when fewer than 80% of its defined genes, or fewer than ten genes, were in that universe. The primary programme had 99 of 100 genes measured in the primary cohort. The mesothelium set reached only 31 of 40 genes and was therefore declared not evaluable under the frozen rule; we did not relax the rule after seeing results, and the consequence is recorded as a limitation.

Post-hoc scoring robustness used the identical expression universe and gene set but replaced the primary mean-rank statistic with three alternatives: the rank-based single-sample *singscore* method (Foroutan et al. 2018), mean log2 counts per million, and mean expression after standardising each programme gene across samples. A separate epithelial score used EPCAM, KRT8, KRT18 and KRT19, an established marker panel (Gao et al. 2021) with no genes shared with the primary programme; both mean-rank and mean-log-expression versions were evaluated.

Statistical analysis

The single primary endpoint was the mean of the patient-level differences, tested with a two-sided paired t-test, with a patient-level bootstrap (5,000 resamples), median, sign and signed-rank statistics as pre-specified robustness. Secondary families (control programmes; host compartments) were corrected within family by the Benjamini–Hochberg procedure. Pre-specified adjusted models added at most one composition term. Pre-specified sensitivity analyses were leave-one-patient-out, same-preservation pairs only, the two alternative programme definitions, and substitution of the post-treatment pair for the pre-treatment pair.

After the primary result was known we added analyses labelled post hoc throughout. For measurement calibration, pseudo-bulk mixtures of 300 cells from calibration donors were constructed at known malignant- or total-epithelial-cell fractions. The slope of each frozen score on known fraction expressed the observed tissue-level difference as the cell-content difference that would reproduce it. We also tested scoring robustness as described above. Finally, the 66 GSE237876 specimens were used to relate the frozen score to histological tumour purity reported by the source study. A patient fixed-effects model estimated the within-patient score change per unit purity; its confidence interval was obtained from 5,000 patient-cluster bootstrap resamples. Region-level specimens were additionally averaged within patient and lesion role. The five eligible primary–peritoneal pairs were reported separately and were not treated as evidence for the across-lesion purity association. Analyses used R 4.5.2.

Generative AI tools (Anthropic Claude and OpenAI Codex) were used only to assist with refinement of selected code and with manuscript writing and language. They were not used to generate data or to independently determine the study design, statistical analyses, interpretation, or conclusions. All analyses were performed by executing the archived code against the cited public data. The authors take full responsibility for the code, numerical results, accuracy, originality, interpretation, and conclusions of the submitted work.

Results

Which patients and specimens carry evidence

Of 26 patients in the primary cohort, 17 had both a primary tumour and a solid peritoneal specimen and entered the primary analysis, one pair each (Fig. 1a). In the independent cohort, 6 peritoneal specimens came from five patients, all women with diffuse or mixed histology, three sampled after chemotherapy, and with low reported tumour purity in most peritoneal specimens (Fig. 1b). Specimen types across all sources are shown in Fig. 1c. No patient in either cohort had the combination of a primary tumour, a solid peritoneal metastasis and another metastatic site needed to estimate site specificity, so that pre-specified comparison was not estimable and no peritoneum-specific claim is made.

The frozen programme is lower in solid peritoneal metastasis

The high-relapse-cell programme was lower in the peritoneal specimen than in the paired primary tumour in 15 of 17 patients (Fig. 2). The mean within-patient difference was -0.087 (95% CI -0.136 to -0.038 ; $P = 0.0017$; standardised effect -0.910), with a patient-level bootstrap interval of -0.133 to -0.045 , a sign-test P of 0.0023 and a signed-rank P of 0.0008 . Leaving out any single patient moved the mean only between -0.094 and $-$

0.075 with a largest P of 0.0040. Both pre-specified alternative definitions agreed in direction: the epithelium-restricted subset gave -0.165 (95% CI -0.253 to -0.077) and the alternative source definition gave -0.068 (95% CI -0.114 to -0.023). Substituting the post-treatment pair for the single patient with two visits gave -0.093 ($P = 0.0008$). Restricting to the five pairs with identical preservation gave -0.110 (95% CI -0.250 to 0.031 ; $P = 0.0966$), in the same direction but uninformative at that size.

Every programme moves in the direction tissue admixture predicts

If the paired specimens differ mainly in what tissue surrounds the cancer, then programmes should move together according to which compartment carries their genes, and this is what we observed (Fig. 3; Table 2). Stromal abundance rose (0.112 , 95% CI 0.048 to 0.176 , $q = 0.0091$) and so did hallmark epithelial–mesenchymal transition (0.065 , 95% CI 0.021 to 0.108 , $q = 0.0153$), reproducing the direction reported by earlier paired analyses. Among host compartments, fibroblast (0.118 , $q = 0.0035$), adipocyte (0.094 , $q = 0.0182$) and endothelium (0.078 , $q = 0.0035$) increased, while gastric-epithelial markers fell sharply (-0.172 , 95% CI -0.245 to -0.100 , $q = 0.0007$), as expected for tissue that contains no gastric mucosa. Immune (0.028 , $q = 0.1646$) and smooth muscle (0.019 , $q = 0.4056$) did not change appreciably. Proliferation moved slightly downwards (-0.033 , $q = 0.0707$) and the stress control was not significant (0.044 , $q = 0.0938$).

Table 2

Within-patient differences for the primary endpoint, alternative definitions, control programmes and host compartments, with confidence intervals and family-corrected q values.

Programme	Role	Mean difference	95% CI	P	q
relapse program (primary endpoint)	Primary endpoint	-0.087	-0.136 to -0.038	0.002	—
relapse program, alternative definition	Alternative HRC definition	-0.068	-0.114 to -0.023	0.006	—
relapse program, epithelium-restricted	Alternative HRC definition	-0.165	-0.253 to -0.077	0.001	—
fibroblast	Host compartment	0.118	0.052 to 0.184	0.002	0.004
adipocyte	Host compartment	0.094	0.024 to 0.164	0.012	0.018
endothelium	Host compartment	0.078	0.034 to 0.122	0.002	0.004
immune	Host compartment	0.028	-0.010 to 0.067	0.137	0.165
smooth muscle	Host compartment	0.019	-0.028 to 0.066	0.406	0.406
gastric epithelium	Host compartment	-0.172	-0.245 to -0.100	< 0.001	< 0.001
stroma	Control	0.112	0.048 to 0.176	0.002	0.009
EMT	Control	0.065	0.021 to 0.108	0.006	0.015
immediate-early stress	Control	0.044	-0.008 to 0.096	0.094	0.094
stroma-restricted YAP targets (negative control)	Control	0.037	-0.000 to 0.075	0.051	0.071
proliferation	Control	-0.033	-0.067 to 0.001	0.057	0.071

The pre-specified negative control behaved as designed. Stroma-restricted YAP/TAZ–TEAD target genes, which carry no epithelial information, increased (0.037, 95% CI 0.000 to 0.075, q = 0.0707) – the pattern expected from a shift in stromal content rather than from epithelial signalling.

Adjusting the primary endpoint for a single composition term reduced it to within the range of chance: with fibroblast content in the model the remaining difference was – 0.014 (95% CI -0.053 to 0.025, P = 0.4573), and with stromal abundance it was – 0.029 (95% CI -0.080 to 0.023, P = 0.2558). Adjusting for preservation change left – 0.031 (P = 0.3592), and adjusting for proliferation left – 0.060 (P = 0.0164). These adjusted models are

descriptive: the composition scores are themselves expression-based and cannot separate cell number from cell state.

The complete set of one-covariate adjustment estimates is shown in Supplementary Figure S1 (Online Resource 1).

One difference in epithelial content reproduces every programme change

In calibration donors the programme was, in practice, a readout of epithelial identity: median scores were 0.789 to 0.837 in epithelial classes, whether malignant or not, against 0.435 to 0.517 in every non-epithelial lineage (Fig. 4a). In pseudo-bulk mixtures with known composition from 8 donors, the score rose monotonically with epithelial content (Fig. 4b), with a slope of 0.285 per unit epithelial fraction; against malignant fraction alone the slope was shallower (0.156), because the cells replaced in that mixture still included non-malignant epithelium.

Converting the observed tissue-level difference through this slope, the primary endpoint corresponds to an epithelial-content difference of -0.305 (interval - 0.478 to -0.133) between paired specimens — that is, peritoneal specimens containing roughly thirty percentage points less epithelium than their matched primaries would reproduce the entire observed difference with no change of state inside any cell. Applying the same conversion to every evaluable programme gave a cluster of implied values (median - 0.311, interquartile range - 0.406 to -0.260 across 12 programmes; Fig. 4c; Table 3), including the hallmark epithelial-mesenchymal transition increase. The gastric-mucosa marker set was the expected outlier, shifting further than dilution predicts because peritoneal tissue contains none of it; the widest implied intervals belong to the programmes whose observed change was itself not significant, such as smooth muscle and immune. A single compositional difference is therefore a sufficient explanation for the entire pattern, including results previously interpreted as programme-level biology.

Table 3

For each evaluable programme, the observed tissue-level difference, its slope against known epithelial fraction in calibration mixtures, and the implied difference in epithelial content.

Programme	Observed difference	Slope vs epithelial fraction	Implied epithelial-fraction difference	95% CI
gastric epithelium	-0.1724	0.152	-1.133	-1.607 to -0.658
fibroblast	0.1180	-0.197	-0.598	-0.934 to -0.263
proliferation	-0.0330	0.068	-0.487	-0.990 to 0.015
EMT	0.0646	-0.171	-0.379	-0.633 to -0.124
relapse program, epithelium-restricted	-0.1651	0.450	-0.367	-0.562 to -0.172
stroma	0.1120	-0.354	-0.316	-0.496 to -0.136
relapse program (primary endpoint)	-0.0870	0.285	-0.305	-0.478 to -0.133
stroma-restricted YAP targets (negative control)	0.0374	-0.133	-0.282	-0.565 to 0.001
endothelium	0.0780	-0.286	-0.273	-0.427 to -0.118
relapse program, alternative definition	-0.0684	0.308	-0.222	-0.371 to -0.073
immune	0.0284	-0.201	-0.141	-0.331 to 0.050
smooth muscle	0.0188	-0.206	-0.092	-0.319 to 0.136

The result is independent of the scoring algorithm and marker definition

Returning to the primary endpoint, every pre-specified sensitivity analysis reported above retained its direction (Fig. 5a), and the paired decrease did not depend on the mean-rank scoring rule (Fig. 5b). The same 99 measured programme genes gave a mean difference of -0.088 with *singscore* ($P = 0.0017$; Cohen's $d_z = -0.910$), -0.631 by mean log2 counts per million ($P = 0.0103$; $d_z = -0.705$), and -0.435 by mean gene-standardised expression ($P = 0.0280$; $d_z = -0.586$). All four algorithms therefore retained the direction of the frozen endpoint despite using different scales and normalisations.

The result also extended beyond the primary programme. EPCAM, KRT8, KRT18 and KRT19 shared no genes with it and were selected as an independent canonical epithelial panel. Their mean-rank score was lower in 14

of 17 peritoneal specimens (mean difference -0.061 , 95% CI -0.114 to -0.009 ; $P = 0.0245$), and their mean log-expression was likewise lower (-1.203 , 95% CI -2.146 to -0.259 ; $P = 0.0157$). The paired decrease is therefore not an artefact of one scoring implementation or one imported gene set.

External histological purity supports the composition interpretation

GSE237876 provided a composition measurement that did not depend on any expression signature: histological tumour purity reported by the source study. Across all 66 specimens from 14 patients, a patient fixed-effects model showed that higher reported purity was associated with a higher frozen programme score (slope 0.126 per unit purity, parametric 95% CI 0.035 to 0.217 , $P = 0.0075$; Fig. 5c). A 5,000-resample patient-cluster bootstrap gave a 95% interval of 0.052 to 0.250 . Averaging multiple regions within patient and lesion role retained the association (32 patient-site values; slope 0.120 , 95% CI 0.016 to 0.224 ; $P = 0.0257$).

This association does not imply that the small external paired subset had a consistent site-level purity difference. Among its five eligible primary–peritoneal pairs, the mean reported purity difference was 0.022 (95% CI -0.548 to 0.592 ; $P = 0.921$). The external purity analysis therefore supports what the score measures across lesions, while the five-patient contrast remains too imprecise to establish how purity differs between the two sites.

Independent paired cohort and the limits of cell-level resolution

Run once in the independent paired cohort, the same frozen endpoint gave -0.031 (95% CI -0.124 to 0.063) in 5 patients, with 3 of 5 in the same direction (Fig. 5d). With five patients the smallest attainable two-sided exact P is 0.0625 , so this cohort can indicate direction only and cannot support or refute the primary result; excluding the one patient whose metastatic site is recorded inconsistently across that study's own figures left 4 patients and -0.013 (95% CI -0.129 to 0.104). We therefore treat this estimate as limited directional support rather than replication.

The decisive observation would be the programme measured inside malignant cells from paired specimens. The public single-cell data cannot currently provide it (Fig. 6). Across three sources, solid peritoneal libraries yielded at most 28 malignant cells each, with 3 peritoneal libraries reaching the pre-specified twenty-cell threshold at all, against thousands of malignant cells in primary libraries (Fig. 6b). In those few libraries the programme value inside malignant cells was 0.781 to 0.837 in peritoneal specimens against 0.813 to 0.860 in primary specimens — overlapping, near the top of the tissue-level range, and far above the values seen in non-epithelial cells (Fig. 6a). This is consistent with the compositional explanation, but the numbers are too small, and donor identity in one source too uncertain, for inference. The fluid-context contrast in malignant cells from ascites could not be evaluated at all, because the public matrix for that study retains only 8,630 genes and the programme fell below the frozen coverage threshold.

Discussion

Paired primary and peritoneal specimens from the same patient are the strongest design available for studying peritoneal metastasis in humans, and they are rare enough that a small number of public cohorts carry a large share of the field's molecular conclusions. Our result is that in these specimens the apparent behaviour of a transcriptional programme is governed by what tissue the block contained. A programme chosen for its malignant-epithelial bias appeared clearly lower in peritoneal lesions, in 15 of 17 patients and with a stable interval; yet the same data show that peritoneal specimens carry more fibroblast, adipose and endothelial signal and no gastric mucosa, that adjusting for a single composition term reduces the programme difference to within the range of chance, and that one epithelial-content difference of about thirty percentage points reproduces the observed change in every programme we scored, including the epithelial–mesenchymal transition increase that paired studies have reported as a feature of peritoneal disease.

The main advance is quantitative. A single epithelial-content difference reproduced the observed magnitudes across 12 biologically distinct programmes. The result persisted under three alternative scoring algorithms and in a four-gene epithelial marker panel with no overlap with the primary programme. In a separate multi-lesion cohort, the frozen score also rose with histological tumour purity within the same patient, with a patient-cluster bootstrap interval excluding zero. Together, these checks make an idiosyncratic gene-set definition or rank statistic an unlikely explanation for the main finding.

This does not overturn the biology those studies describe. Increased fibroblast and myeloid content in peritoneal lesions has been shown with cell-resolved data and spatial measurement (Zhao et al. 2024; Li et al. 2025; Wang et al. 2025), and stromal–myeloid niches in peritoneal metastasis are supported by evidence that does not rest on bulk deconvolution. What our results constrain is the inference step from a paired bulk contrast to a statement about cancer-cell state. The observation that metastasis-associated mesenchymal signal derives largely from fibroblasts rather than carcinoma cells has been made before in gastric cancer (Kim et al. 2024) and across tumour types (Tyler and Tirosh 2021); we extend it by showing that in the paired peritoneal setting the compositional component is large enough to account for programme differences in either direction, including a decrease in an epithelial programme, and by providing the conversion that makes the size of that component explicit.

The primary programme was defined outside gastric cancer, in a colorectal study of the cells that survive resection and seed relapse, and was reported there to be independent of YAP/TEAD (Cañellas-Socias et al. 2022). We chose it partly for that independence because canonical YAP target genes are expressed strongly by fibroblasts, endothelium and mesothelium and are therefore poorly identifiable in peritoneal tissue, a consideration that also applies to interpreting YAP-focused work on peritoneal carcinomatosis (Ajani et al. 2021). The compartment marker sets came from a normal-tissue atlas (Karlsson et al. 2021), with every programme gene excluded so that the composition covariates could not be circular with the endpoint. The negative control was constructed to move only with stromal content, and it did.

The limitations are specific. The mesothelium marker set did not meet the frozen coverage threshold in the primary cohort, missing it by one gene; we kept the rule rather than relax it after seeing the result, so the compartment most characteristic of peritoneal tissue is represented only indirectly, through adipose, fibroblast and endothelial signal. Preservation method differs within most pairs and cannot be fully separated from site, although it does so in both directions and adjustment did not change the qualitative picture. Per-

patient treatment before sampling is unknown in the primary cohort and was not assumed to be absent. The independent paired cohort contains only five women with diffuse or mixed histology, most sampled after chemotherapy, and therefore cannot provide a precise replication estimate. Its reported purity data are valuable across all 66 lesions, but the five paired patients do not show a stable primary–peritoneal purity difference. The calibration mixtures come from primary tumours rather than peritoneal implants, and the score–composition relationship saturates at high epithelial content, so the implied differences are approximations rather than direct purity measurements. The chosen programme also separates epithelial from non-epithelial cells more strongly than malignant from non-malignant epithelium; a programme with sharper within-epithelium discrimination might behave differently. Finally, we cannot determine what the programme does inside cancer cells in solid peritoneal implants, because the available single-cell data recover too few malignant cells from those lesions.

Studies that aim to characterise cancer-cell state in peritoneal metastasis need either cell-resolved measurement with adequate malignant-cell recovery from solid implants, or an explicit composition model with independently defined compartment markers and a stated equivalence between compositional and state explanations. Paired bulk specimens remain valuable for describing what the lesion as a whole contains — a biological question in its own right, since the peritoneal niche is built from mesothelium, adipose and fibroblasts (Fidler 2003; Sandoval et al. 2013; Kanda and Kodera 2016) — but a difference in a bulk programme score between a gastric resection and a peritoneal implant should not by itself be read as a change in what the cancer cells are doing.

Declarations

Competing interests.

The authors have no relevant financial or non-financial interests to disclose.

Consent to participate.

Not applicable to this secondary analysis of public de-identified data.

Consent for publication.

Not applicable.

Funding.

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Author Contribution

C.H., Z.-M.Z. and Y.K. contributed to all stages of the study, including study conception and design, data curation, methodology, formal analysis, interpretation, visualisation, and manuscript preparation and revision. Q.Y. contributed to data analysis and manuscript preparation and revision. All authors reviewed and approved the final manuscript.

Data availability.

This study analysed only public data: GSE314812, GSE237876, GSE183904, GSE163558, GSE308231, GSE239676, GSE234129 and GSE228598 from the Gene Expression Omnibus; Ensembl GRCh37 release 75 annotation; HGNC approved symbols; Human Protein Atlas single-cell-type expression; and gene sets from the Molecular Signatures Database and the publications cited. Accession numbers, URLs, file sizes, download dates and SHA-256 checksums for every input are provided with the analysis code.

Code availability.

Analysis code, the protocol and protocol lock file, and a public-input manifest with accessions and checksums are available at <https://github.com/kingterkong/gastric-pm-tissue-composition> (version v1.0.1).

Figures 1–6 were generated from the cited public data and archived analysis outputs by the archived code.

The limited assistance with selected code refinement is described in the Methods section.

References

1. Ajani J, Xu Y, Huo L et al (2021) YAP1 mediates gastric adenocarcinoma peritoneal metastases that are attenuated by YAP1 inhibition. *Gut* 70:55–66. <https://doi.org/10.1136/gutjnl-2019-319748>
2. Bray F, Laversanne M, Sung H et al (2024) Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *Cancer J Clin* 74:229–263. <https://doi.org/10.3322/caac.21834>
3. Cañellas-Socias A, Cortina C, Hernando-Momblona X et al (2022) Metastatic recurrence in colorectal cancer arises from residual EMP1 + cells. *Nature* 611:603–613. <https://doi.org/10.1038/s41586-022-05402-9>
4. Charton C, Kang S, Kim S et al (2026) Divergent clonal evolution and tumor microenvironment remodeling shape gastric cancer peritoneal metastasis. *Commun biology* 9:666. <https://doi.org/10.1038/s42003-026-09906-2>
5. Cheng X, Dai E, Wu J et al (2024) Atlas of metastatic gastric cancer links ferroptosis to disease progression and immunotherapy response. *Gastroenterology* 167:1345–1357. <https://doi.org/10.1053/j.gastro.2024.07.038>
6. Fidler I (2003) The pathogenesis of cancer metastasis: The 'seed and soil' hypothesis revisited. *Nat Rev Cancer* 3:453–458. <https://doi.org/10.1038/nrc1098>
7. Foroutan M, Bhuva D, Lyu R et al (2018) Single sample scoring of molecular phenotypes. *BMC Bioinformatics* 19:404. <https://doi.org/10.1186/s12859-018-2435-4>

8. Gao R, Bai S, Henderson Y et al (2021) Delineating copy number and clonal substructure in human tumors from single-cell transcriptomes. *Nat Biotechnol* 39:599–608. <https://doi.org/10.1038/s41587-020-00795-2>
9. Gavish A, Tyler M, Greenwald A et al (2023) Hallmarks of transcriptional intratumour heterogeneity across a thousand tumours. *Nature* 618:598–606. <https://doi.org/10.1038/s41586-023-06130-4>
10. Germain P, Lun A, Garcia Meixide C et al (2021) Doublet identification in single-cell sequencing data using scDblFinder. <https://doi.org/10.12688/f1000research.73600.2>. *F1000Research* 10:979
11. Guchelaar N, de Neijis M, Noordman B et al (2025) The prognostic value of peritoneal metastases in patients with gastric cancer: A nationwide population-based study. *EClinicalMedicine* 81:103109. <https://doi.org/10.1016/j.eclinm.2025.103109>
12. Hao Y, Hao S, Andersen-Nissen E et al (2021) Integrated analysis of multimodal single-cell data. *Cell* 184:3573–3587e29. <https://doi.org/10.1016/j.cell.2021.04.048>
13. Hu S, Hu C, Xu J et al (2024) The estrogen response in fibroblasts promotes ovarian metastases of gastric cancer. *Nat Commun* 15:8447. <https://doi.org/10.1038/s41467-024-52615-9>
14. Jamel S, Markar S, Malietzis G et al (2018) Prognostic significance of peritoneal lavage cytology in staging gastric cancer: Systematic review and meta-analysis. *Gastric cancer: official J Int Gastric Cancer Association Japanese Gastric Cancer Association* 21:10–18. <https://doi.org/10.1007/s10120-017-0749-y>
15. Janjigian Y, Shitara K, Moehler M et al (2021) First-line nivolumab plus chemotherapy versus chemotherapy alone for advanced gastric, gastro-oesophageal junction, and oesophageal adenocarcinoma (CheckMate 649): A randomised, open-label, phase 3 trial. *Lancet (London England)* 398:27–40. [https://doi.org/10.1016/S0140-6736\(21\)00797-2](https://doi.org/10.1016/S0140-6736(21)00797-2)
16. Jiang H, Yu D, Yang P et al (2022) Revealing the transcriptional heterogeneity of organ-specific metastasis in human gastric cancer using single-cell RNA sequencing. *Clin translational Med* 12:e730. <https://doi.org/10.1002/ctm2.730>
17. Kanda M, Kodera Y (2016) Molecular mechanisms of peritoneal dissemination in gastric cancer. *World J Gastroenterol* 22:6829–6840. <https://doi.org/10.3748/wjg.v22.i30.6829>
18. Karlsson M, Zhang C, Méar L et al (2021) A single-cell type transcriptomics map of human tissues. *Sci Adv* 7:eabh2169. <https://doi.org/10.1126/sciadv.abh2169>
19. Kim K, Lee J, Cheong J et al (2024) Deciphering metastatic route-specific signals and their microenvironment interactions in peritoneal metastasis of gastric cancer. *Cancer Commun (London England)* 44:514–517. <https://doi.org/10.1002/cac2.12533>
20. Koemans W, Lurvink R, Grootsholten C et al (2021) Synchronous peritoneal metastases of gastric cancer origin: Incidence, treatment and survival of a nationwide dutch cohort. *Gastric cancer: official J Int Gastric Cancer Association Japanese Gastric Cancer Association* 24:800–809. <https://doi.org/10.1007/s10120-021-01160-1>
21. Kumar V, Ramnarayanan K, Sundar R et al (2022) Single-cell atlas of lineage states, tumor microenvironment, and subtype-specific expression programs in gastric cancer. *Cancer Discov* 12:670–691. <https://doi.org/10.1158/2159-8290.CD-21-0683>
22. Li B, Dewey C (2011) RSEM: Accurate transcript quantification from RNA-seq data with or without a reference genome. *BMC Bioinformatics* 12:323. <https://doi.org/10.1186/1471-2105-12-323>

23. Li Y, Zheng Y, Huang J et al (2025) CAF-macrophage crosstalk in tumour microenvironments governs the response to immune checkpoint blockade in gastric cancer peritoneal metastases. *Gut* 74:350–363. <https://doi.org/10.1136/gutjnl-2024-333617>
24. Liberzon A, Birger C, Thorvaldsdóttir H et al (2015) The molecular signatures database (MSigDB) hallmark gene set collection. *Cell Syst* 1:417–425. <https://doi.org/10.1016/j.cels.2015.12.004>
25. Patro R, Duggal G, Love M et al (2017) Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods* 14:417–419. <https://doi.org/10.1038/nmeth.4197>
26. Riihimäki M, Hemminki A, Sundquist K et al (2016) Metastatic spread in patients with gastric cancer. *Oncotarget* 7:52307–52316. <https://doi.org/10.18632/oncotarget.10740>
27. Rijken A, Pape M, Simkens G et al (2024) Peritoneal metastases from gastric cancer in a nationwide cohort: Incidence, treatment and survival. *Int J Cancer* 154:992–1002. <https://doi.org/10.1002/ijc.34780>
28. Sandoval P, Jiménez-Heffernan J, Rynne-Vidal Á et al (2013) Carcinoma-associated fibroblasts derive from mesothelial cells via mesothelial-to-mesenchymal transition in peritoneal metastasis. *J Pathol* 231:517–531. <https://doi.org/10.1002/path.4281>
29. Sonesson C, Love M, Robinson M (2015) Differential analyses for RNA-seq: Transcript-level estimates improve gene-level inferences. <https://doi.org/10.12688/f1000research.7563.2>. *F1000Research* 4:1521
30. Thomassen I, van Gestel Y, van Ramshorst B et al (2014) Peritoneal carcinomatosis of gastric origin: A population-based study on incidence, survival and risk factors. *Int J Cancer* 134:622–628. <https://doi.org/10.1002/ijc.28373>
31. Tirosh I, Izar B, Prakadan S et al (2016) Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Sci (New York NY)* 352:189–196. <https://doi.org/10.1126/science.aad0501>
32. Tyler M, Tirosh I (2021) Decoupling epithelial-mesenchymal transitions from stromal profiles by integrative expression analysis. *Nat Commun* 12:2592. <https://doi.org/10.1038/s41467-021-22800-1>
33. Wang Q, Ni Y, Lu S et al (2025) Multi-dimensional omics integrated machine learning framework identifies macrophage-fibroblast-tumor co-infiltration patterns to predict prognosis in gastric cancer. *NPJ Digit Med* 9:10. <https://doi.org/10.1038/s41746-025-02179-9>
34. Wang R, Song S, Qin J et al (2023) Evolution of immune and stromal cell states and ecotypes during gastric adenocarcinoma progression. *Cancer Cell* 41:1407–1426e9. <https://doi.org/10.1016/j.ccell.2023.06.005>
35. Yoshihara K, Shahmoradgoli M, Martínez E et al (2013) Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun* 4:2612. <https://doi.org/10.1038/ncomms3612>
36. Zhao J, Ong C, Srivastava S et al (2024) Spatially resolved niche and tumor microenvironmental alterations in gastric cancer peritoneal metastases. *Gastroenterology* 167:1384–1398e4. <https://doi.org/10.1053/j.gastro.2024.08.007>

Figures

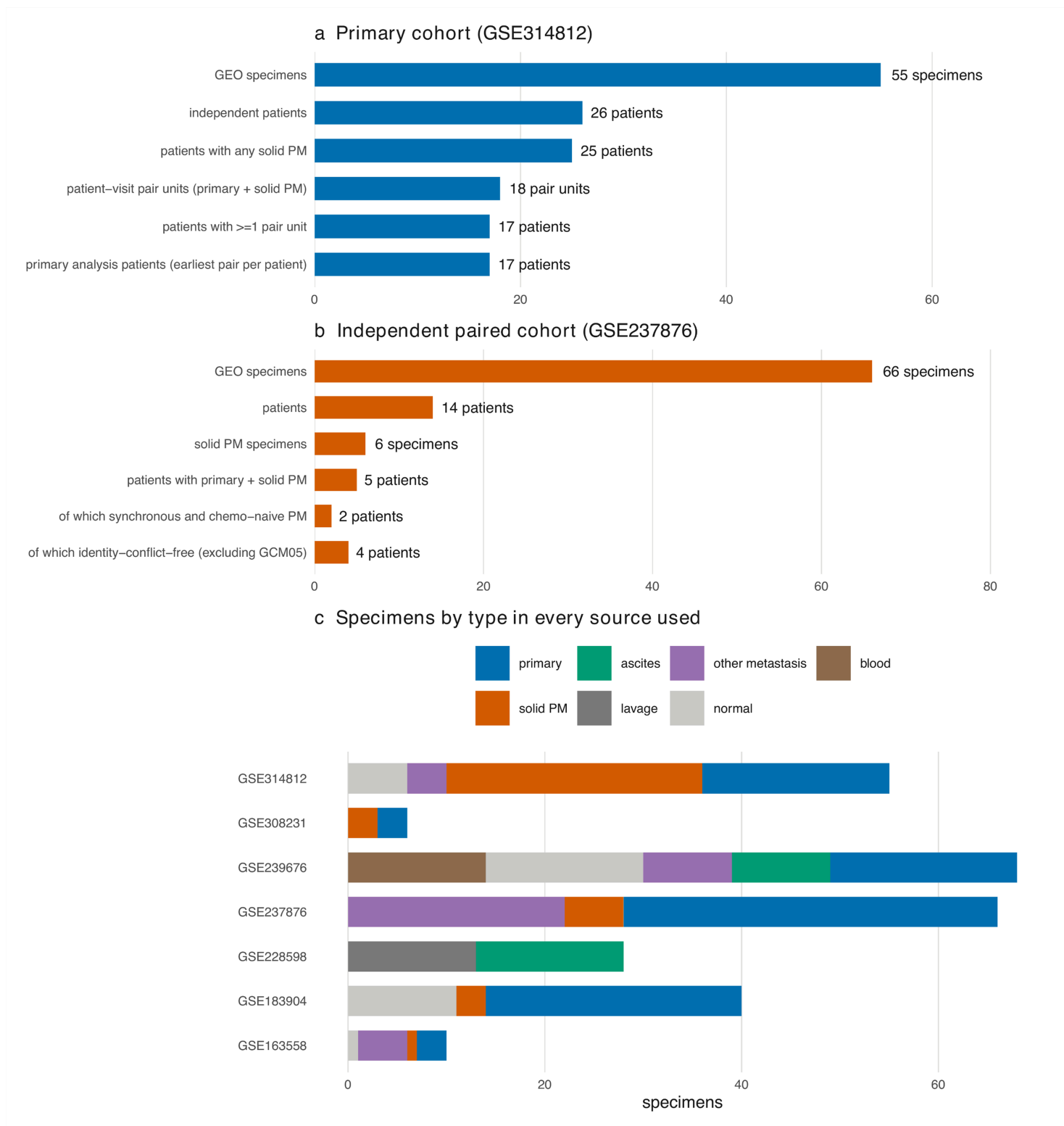


Figure 1

Which patients and specimens carry which evidence. (a) Specimen and patient flow in the primary cohort (GSE314812), from deposited specimens to the patients contributing one pair each to the primary analysis. (b) Corresponding flow in the independent paired cohort (GSE237876), including the subsets defined by metastasis timing, chemotherapy exposure and specimen identity consistency. (c) Specimen types in every

source used, counted from the sample crosswalk; ascites and lavage are shown separately from solid tissue throughout

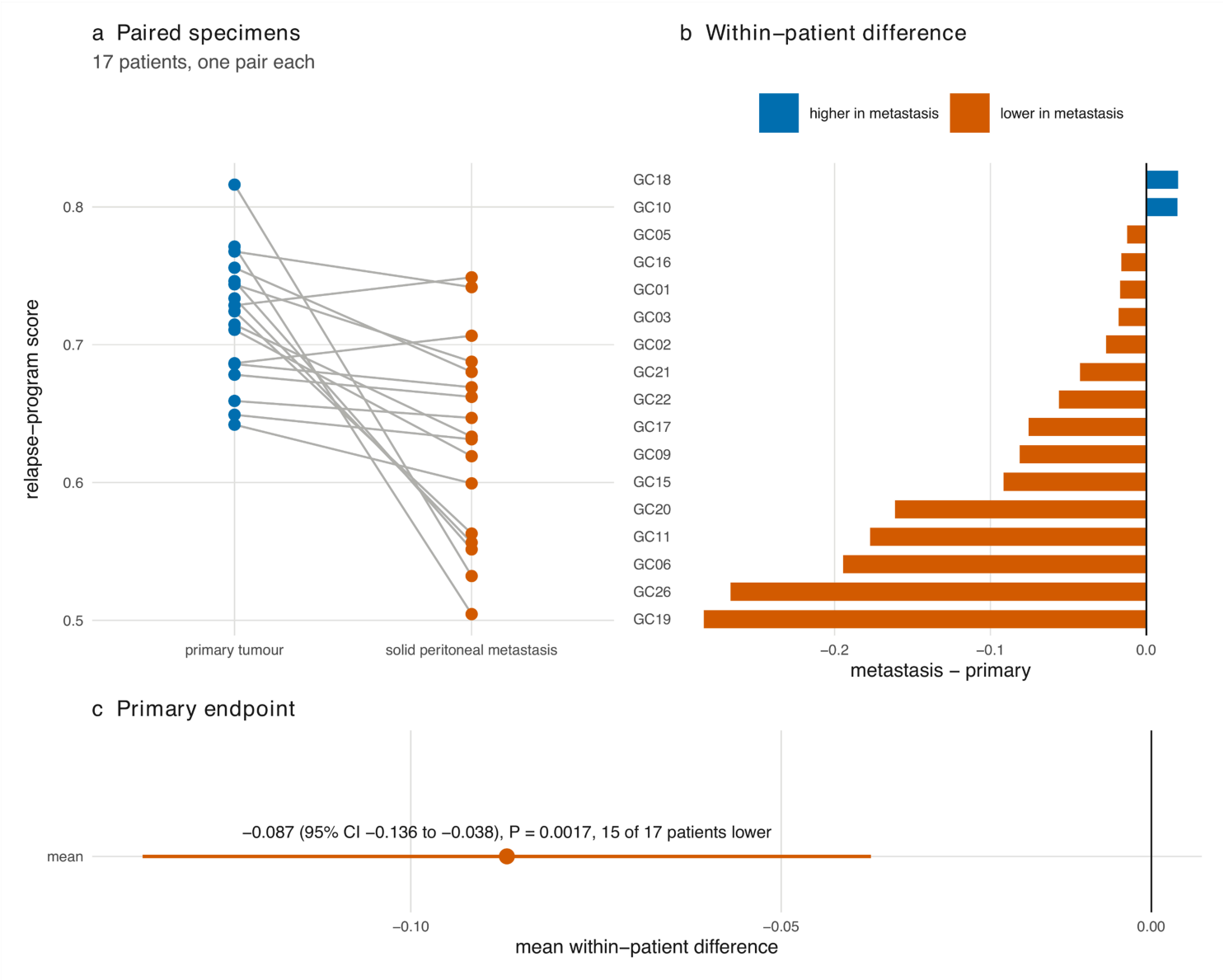


Figure 2

The frozen programme in paired specimens. (a) High-relapse-cell programme score in each patient’s primary tumour and solid peritoneal metastasis, one pair per patient, 17 patients. (b) Within-patient difference for each patient, ordered by magnitude. (c) Primary endpoint: mean within-patient difference with its 95% confidence interval. Statistics are a two-sided paired t-test on patient-level differences; each patient contributes one difference

Every program moves as tissue composition predicts
 mean within-patient difference, 17 patients; bars are 95% confidence intervals

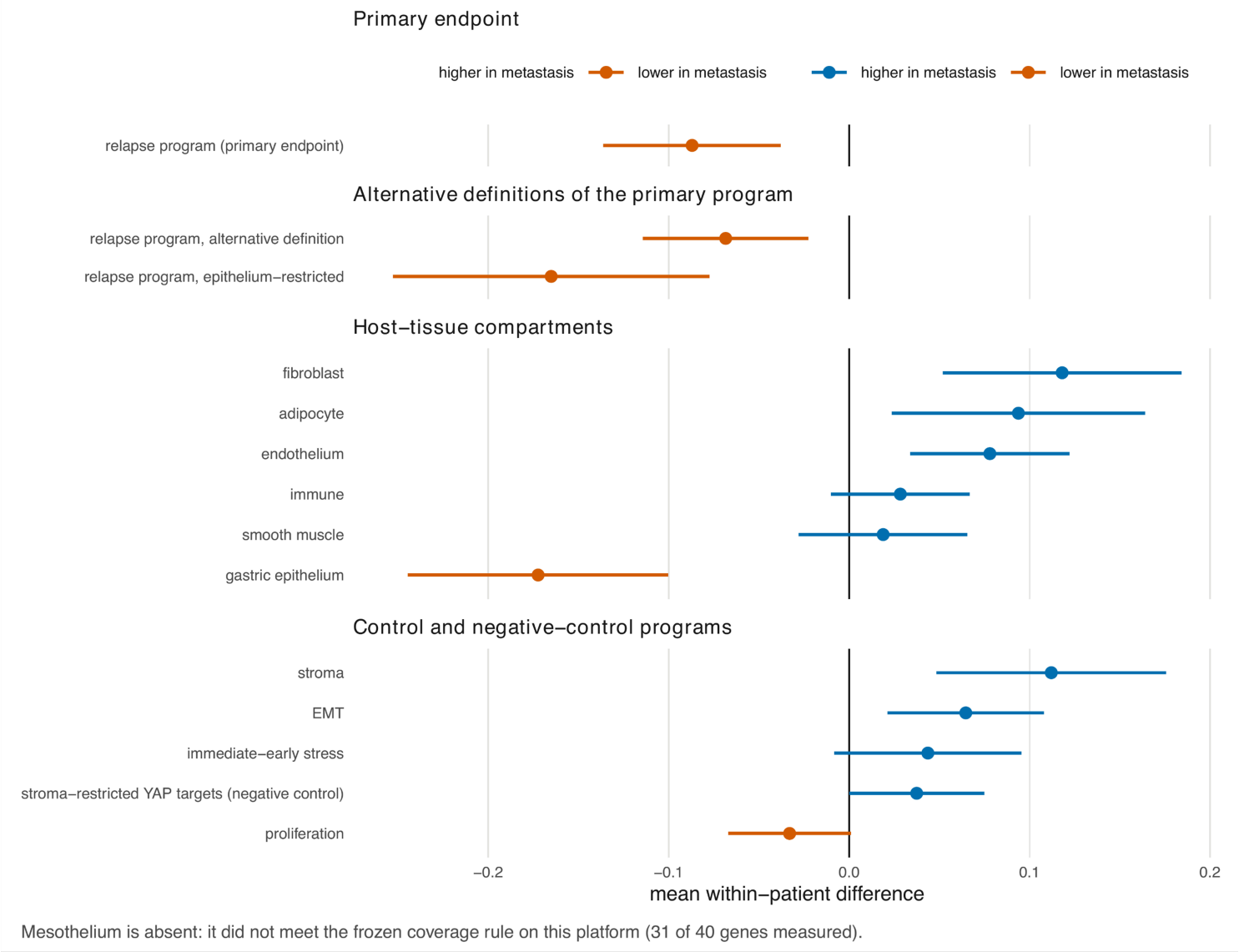


Figure 3

Programme and compartment changes. Mean within-patient difference (metastasis minus primary) for the primary endpoint, its two pre-specified alternative definitions, the four control programmes with the stroma-restricted negative control, and the host-compartment marker sets. Bars are 95% confidence intervals; 17 patients. The mesothelium set did not meet the frozen coverage rule on this platform (31 of 40 genes measured) and is therefore absent

a The program scores epithelial identity

GSE183904 gastric primary tumours, development side

b Known mixtures track epithelial content

300-cell pseudo-bulk mixtures, 8 calibration donors

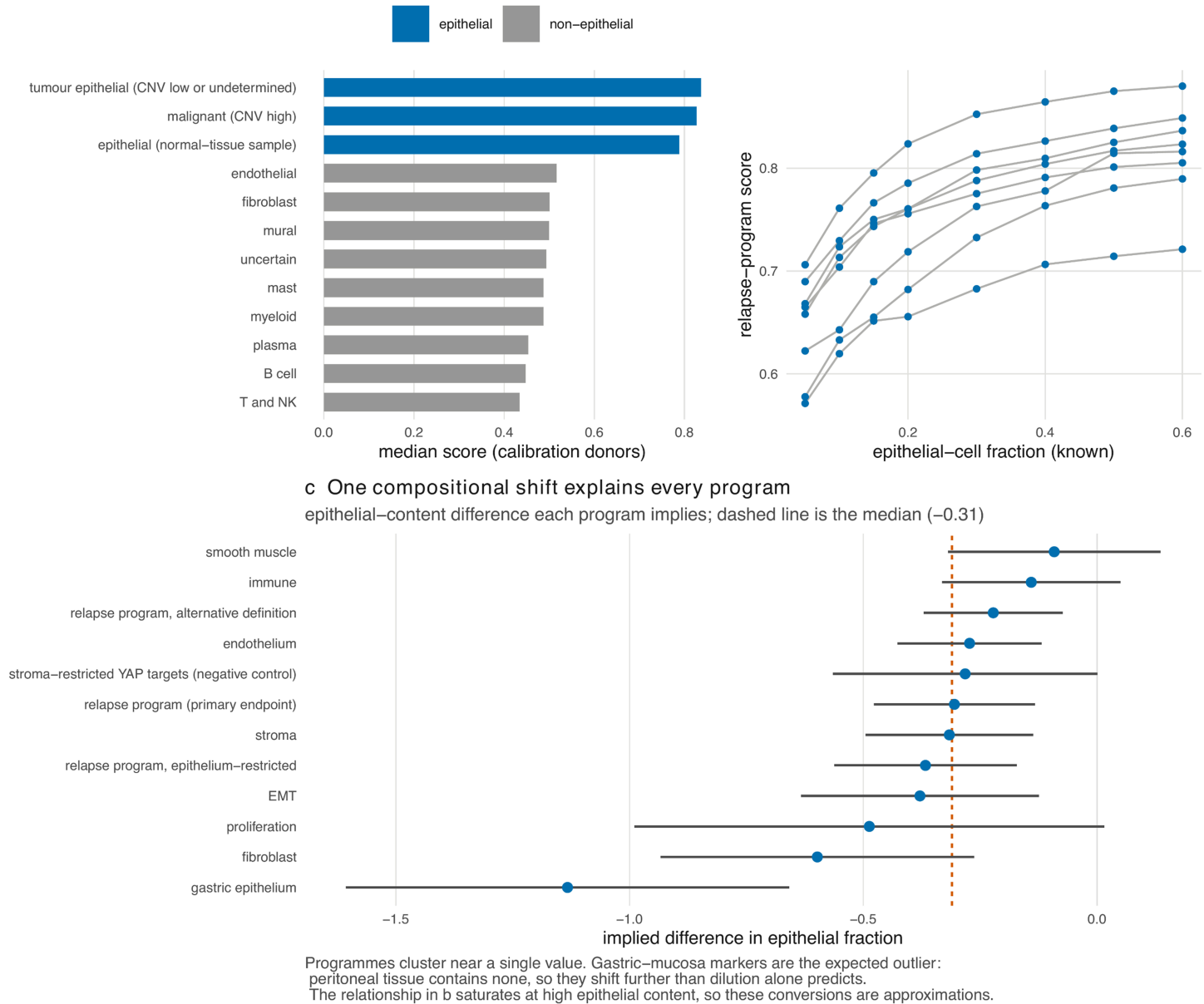


Figure 4

A single compositional difference accounts for the pattern. (a) Median programme score by cell class in calibration donors (GSE183904 gastric primary tumours, development side). (b) Programme score in 300-cell pseudo-bulk mixtures with known epithelial fraction, 8 calibration donors, ten replicates per fraction. (c) For each evaluable programme, the epithelial-content difference that would reproduce its observed tissue-level change; the dashed line is the median across programmes. Post-hoc measurement calibration

The composition result is robust to scoring choices and external purity data

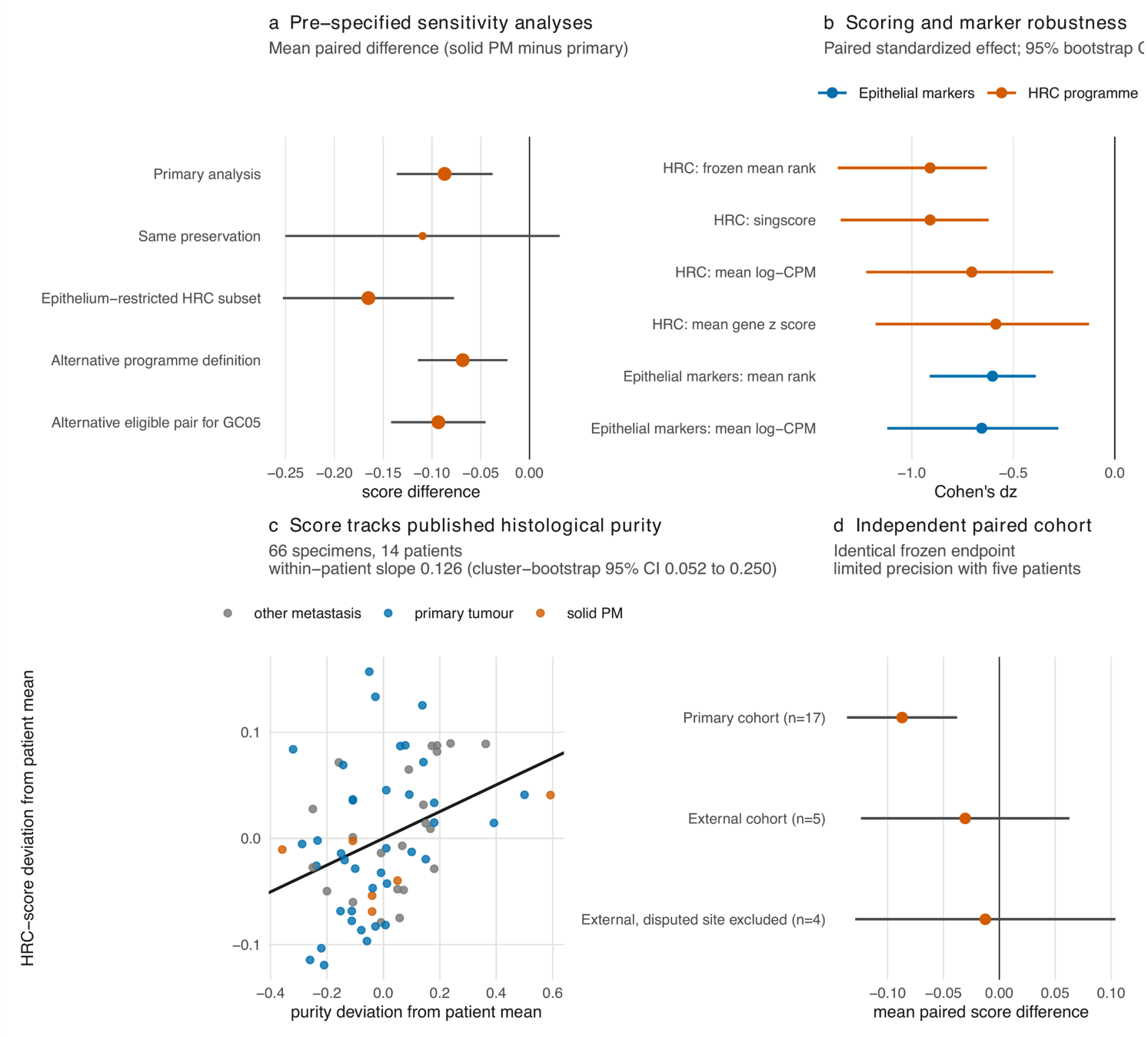


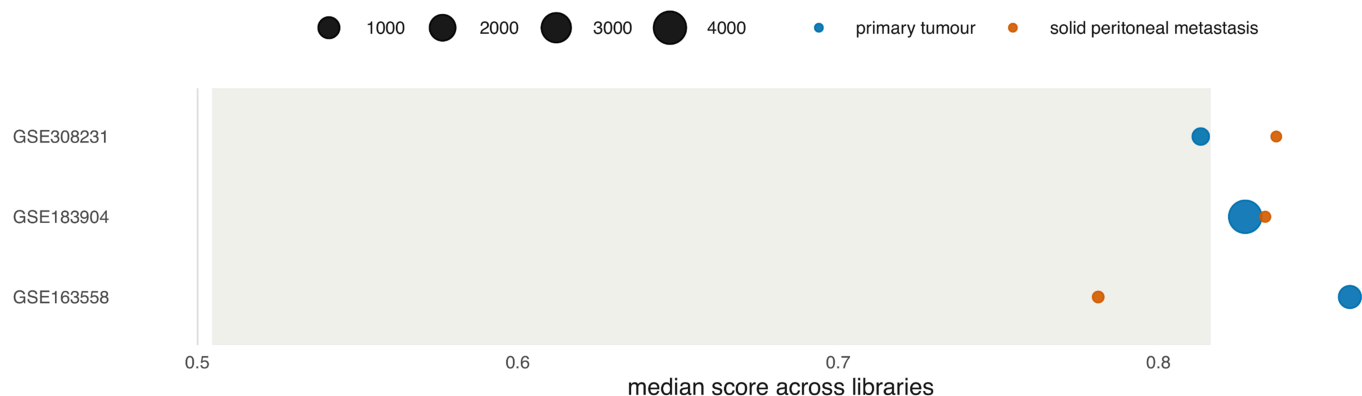
Figure 5

Robustness to scoring choices and external histological purity. (a) Pre-specified sensitivity analyses of the primary endpoint; point size is the number of patients. (b) Post-hoc scoring robustness, expressed as paired standardised mean differences with patient-bootstrap 95% confidence intervals. The first four rows score the same HRC genes by different algorithms; the last two score the independent EPCAM/KRT8/KRT18/KRT19 panel. (c) Frozen HRC score and published histological purity after centring both within patient in GSE237876; the line is the patient fixed-effects slope, and the interval was obtained by patient-cluster bootstrap. (d) The identical frozen paired endpoint across sources; the five-patient external estimate is shown with its 95% confidence interval

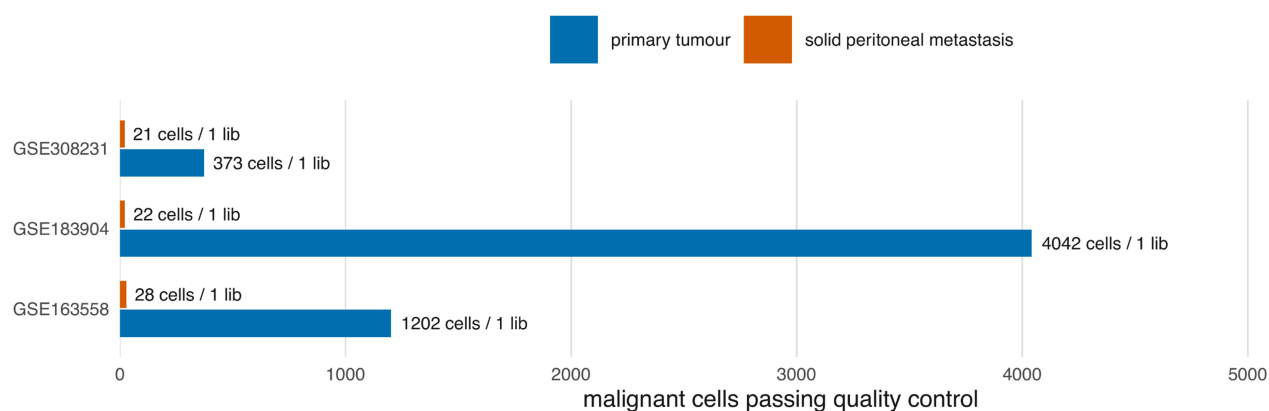
a Program value inside malignant cells

libraries reaching 20 malignant cells; donors unverified in GSE308231.

Shaded band is the full range of paired tissue-level scores in the primary cohort.



b How little malignant material the peritoneal libraries contain



This is why the cell-intrinsic question cannot be settled with the public data available.

Figure 6

What the available single-cell data can and cannot resolve. (a) Median programme score inside malignant cells for libraries reaching twenty malignant cells; the shaded band is the full range of paired tissue-level scores in the primary cohort. (b) Malignant cells passing quality control per library and specimen type. Donor identity is not verifiable in GSE308231, so libraries rather than patients are the unit and no paired inference is possible

Supplementary Files

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

- [ESM1SupplementaryInformation.pdf](#)
- [FigS1.tif](#)
- [ESM2SupplementaryTables.xlsx](#)