Supplementary Methods

Patient Selection

Spatial transcriptomic (CosMx term) profiling was undertaken on FFPE fixed resection specimens post-neoadjuvant therapy and matched pre-treatment biopsies from 5 patients who had undergone neoadjuvant FLOT based chemotherapy for oesophageal/GOJ adenocarcinoma in the prior 2 years followed by curative oesophagectomy. All patients selected had experienced a partial response to therapy (TRG3) as defined by an NHS pathologist expert in gastro-oesophageal cancer. TRG3 was selected to provide the balance between areas of response and residual disease and allow investigation of mechanisms of treatment resistance and recurrence, including areas of minimally residual disease in fibrosis.

Computational analysis

Integrated single-cell RNA-seq reference

Reciprocal principal component analysis (RPCA) was used to integrate three published oesophageal adenocarcinoma (OAC) single-cell or single-nucleus RNA-seq datasets using Seurat v5.2.0. The datasets comprised 10x Genomics single-cell RNA-seq, Drop-seq single-cell RNA-seq, and single-nucleus RNA-seq/multiome. RPCA integration was used for cross-study co-embedding, harmonisation of cell-type annotations, and construction of a reference atlas.

Pre-processed, post-QC expression matrices and metadata were obtained from the original studies. Because the datasets were generated using different capture platforms and tissue-processing workflows, study-specific QC procedures were retained rather than applying a single uniform QC threshold across all datasets. These published QC pipelines included platform-appropriate filtering of low-complexity cells or nuclei, high mitochondrial-content barcodes, predicted doublets or empty droplets where applicable, and sample or library-quality metrics. In the 10x Genomics single-cell RNA-seq dataset, reported QC included processing of samples with high cell viability, filtering

of cells with low or excessively high detected gene counts, removal of cells with high mitochondrial RNA content, and exclusion of predicted doublets. In the single-nucleus/multiome dataset, published processing included filtering of low-complexity nuclei, high mitochondrial-content barcodes, ambient RNA correction, and doublet detection. For the Drop-seq dataset, post-QC cells from the published analysis were retained.

Where applicable, datasets were subset to retain primary oesophageal tumour and normal tissue-derived cells, with metastatic samples excluded. Metadata variables, including tissue source, treatment status, donor, sample, and study of origin, were harmonised across datasets. Cell-type annotations were standardised into nine broad lineages: normal epithelium, malignant epithelium, endothelial cells, fibroblasts, and five immune-cell lineages.

Each broad lineage was subsequently analysed independently. Within each lineage, cells underwent normalisation, principal component analysis, nearest-neighbour graph construction, clustering, and UMAP visualisation following the recommended Seurat workflow. Fine-level cell-type identities were assigned using Azimuth and SingleR, supplemented by manual curation based on established marker-gene expression, prior biological knowledge, and published literature. Azimuth mapping to the human PBMC reference was used to support immune-cell annotation and was not used to assign epithelial, malignant, endothelial, or fibroblast identities.

Pre-processing and quality control of CosMx spatial transcriptomics data

CosMx spatial transcriptomics data were analysed independently for each slide to minimise potential slide-specific batch effects. The dataset comprised five biopsy samples on one slide and five resection samples distributed across three slides. Pre-processing and quality control were performed following Bruker's recommended workflow described in the "Basics of CosMx Analysis in R" vignette.

Unique cell identifiers were generated using the prefix "c" followed by slide number, field of view, and cell number, for example c_1_23_452. These identifiers were applied consistently across the raw count matrix, metadata, polygon files, and field-of-view position files.

Negative control probes and system control probes were extracted from the raw count matrix and saved separately for downstream use, including supervised InSituType annotation. Cells containing fewer than 50 detected transcripts were removed to exclude low-complexity cells. Cells with a surface area greater than 30,000 μm^2 were also excluded to reduce the inclusion of likely mis-segmented, merged, or otherwise artefactual cells.

A Seurat object was created from the raw count matrix and associated metadata and cell centroid coordinates were extracted from the metadata and used to construct a centroid-based spatial feature layer. Coordinates were converted into a spatial field-of-view object, which was added to the Seurat object as the global spatial layer.

Cell-type annotation and consensus label generation for CosMx data

To generate robust and reproducible cell-type annotations for the OAC CosMx dataset, multiple independent annotation frameworks were applied. Their outputs were harmonised and combined into a consensus cell-type label. This approach was used to reduce dependence on any single classifier and to integrate complementary evidence from spatial-platform-aware models, reference mapping, marker-gene classifiers, and correlation-based annotation.

Individual annotation methods

Five annotation approaches were used.

InSituType was run in supervised mode using a pseudo-bulk reference derived from the integrated scRNA-seq reference. A reference gene-expression matrix was constructed from the scRNA-seq dataset using genes shared with the CosMx 6,000-plex panel. For each annotated reference cell type, the top differentially expressed genes were identified, and the subset overlapping the CosMx panel was retained. These genes were used to generate pseudo-bulk profiles representing each reference cell type. InSituType was then run using the raw CosMx counts together with the negative control count matrix. InSituType was treated as the primary spatial-platform-aware classifier because it is designed for in situ transcriptomic data, but its predictions were evaluated in combination with independent classifiers, marker expression, and manual review.

Latent transfer was performed using Seurat's anchor-based label transfer framework. The pre-processed CosMx Seurat object was used as the query and the integrated scRNA-seq object as the reference. Anchors were identified using 30 principal components, and reference labels were transferred to CosMx cells. This approach was used to capture global transcriptomic similarity between CosMx cells and the scRNA-seq reference while accounting for cross-platform differences.

Garnett was used as an interpretable marker-gene-based classifier. A hierarchical marker file was generated for 19 cell types known or expected to occur in OAC, using established positive and negative marker-gene patterns. Garnett predictions were included because marker-based classification is relatively interpretable and less dependent on global cross-platform alignment, although it is constrained by marker coverage and manual marker selection.

Azimuth was used for supervised reference mapping of immune populations. The human PBMC Azimuth reference was used with level 1 cell-type annotations and 30 principal components. Azimuth predictions were used only to support immune-cell annotation and were not used to classify epithelial, malignant, endothelial, or stromal cells.

SingleR was used as a correlation-based annotation method. CosMx expression profiles were compared with the Human Primary Cell Atlas reference provided through SingleR. SingleR was included as an independent broad cell-type annotation method because it performs well for abundant, transcriptionally distinct cell types.

Each method generated a predicted cell-type label and, where available, an associated confidence, probability, or prediction score.

Label harmonisation

Because the five classifiers produced labels with different naming conventions and levels of granularity, predicted labels were harmonised to a unified ontology of approximately 20 biologically interpretable cell classes. These included malignant epithelial cells, non-malignant epithelial cells, fibroblasts or stromal cells, endothelial cells, CD4 T cells, CD8 T cells, regulatory T cells, B cells, plasma cells or plasmablasts, NK cells, myeloid cells, dendritic cells, mast cells, and other or unknown cells.

Harmonisation was performed using string matching and regular-expression-based rules, followed by manual review. Particular attention was given to ambiguous labels, including plasmacytoid dendritic cells versus plasma cells, broad lymphoid labels versus specific T- or B-cell annotations, and epithelial versus malignant epithelial assignments.

Weighted consensus annotation

To combine predictions from the five classifiers, a weighted-voting scheme was implemented. For each cell, each classifier contributed support to its predicted harmonised label. The contribution of classifier *i* to label *l* was calculated as:

$$W_{label} = \sum_{i} (w_i \times s_i)$$

where w_i is the fixed method weight and s_i is the classifier-specific confidence score. For methods that did not provide a continuous confidence score, a score of 1 was assigned to the predicted label and 0 to all other labels.

InSituType predictions were assigned the highest weight, $w = 3$, because they were generated using a CosMx-aware spatial transcriptomic framework. Latent transfer and Azimuth were assigned intermediate weights, $w = 2$ and $w = 1.5$, respectively, reflecting their utility for reference-based mapping. SingleR was assigned $w = 1$, reflecting its broad but less spatially specific correlation-based annotation. Garnett was assigned $w = 0.5$ because it relies on manually curated marker sets and does not provide continuous confidence scores.

For each cell, the label with the highest total weighted support was retained as the initial consensus label. A weighted consensus confidence score was calculated as the support for the winning label divided by the total support across all candidate labels. In parallel, an unweighted agreement score was calculated as the proportion of classifiers assigning the final consensus label. These scores were retained as measures of annotation confidence.

Manual annotation refinement

Initial review with a pathologist indicated that most cells annotated as malignant epithelial cells by the automated pipeline corresponded to malignant regions on

matched H&E images. However, some discrepancies were observed. In particular, subsets of cells initially labelled as lymphoid, epithelial, other, or unknown were located in regions morphologically consistent with tumour, whereas some cells initially labelled as malignant epithelial cells appeared inconsistent with malignancy on H&E review.

To refine malignant-cell annotations, a sample-wise manual refinement procedure was applied. For each sample, cells annotated as malignant epithelial, epithelial, other, unknown, and any additional cell types suspected to include misclassified malignant cells were subset. These cells were reprocessed using Seurat, including normalisation, highly variable feature selection, scaling, PCA, nearest-neighbour graph construction, clustering, and UMAP visualisation. Cluster marker genes were identified using FindAllMarkers(). Marker-gene expression, spatial localisation, H&E morphology, and prior biological knowledge were used to assign each cluster to a broad lineage. Cell labels were revised when cluster-level evidence supported a different lineage assignment from the initial consensus annotation.

Cell-type annotation and iterative lineage refinement of CosMx data

Cell-type annotation of the CosMx spatial transcriptomics dataset was performed using a two-stage strategy. First, cells were assigned to broad biological lineages and iteratively refined based on unsupervised clustering and marker-gene expression. Second, fine cell-type annotations were generated within each broad lineage using lineage-restricted supervised InSituType classification.

Broad lineage annotation and iterative refinement

CosMx cells were initially assigned to six broad biological compartments: cancer, epithelial, endothelial and mural, stromal, lymphoid, and myeloid. These broad compartments were selected to capture the major cellular components of OAC tissue while minimising over-classification at the initial annotation stage.

To refine broad lineage assignments, cells from each compartment were analysed iteratively. For each broad lineage, cells were subset and reprocessed independently in Seurat. This included normalisation, identification of variable features, scaling, principal component analysis, nearest-neighbour graph construction, clustering, and UMAP

visualisation. Cluster marker genes were identified using FindAllMarkers() and examined to determine whether clusters were consistent with their assigned broad lineage.

Clusters showing marker-gene profiles inconsistent with their initial lineage assignment were reassigned to the appropriate compartment. For example, clusters initially assigned to lymphoid or other categories but expressing epithelial markers were moved to the epithelial compartment, whereas clusters expressing canonical fibroblast, endothelial, mural, myeloid, or lymphoid markers were reassigned accordingly. This process was repeated iteratively until the major clusters within each broad compartment showed coherent lineage-specific marker expression.

Broad lineage refinement was guided by canonical marker-gene expression, spatial localisation, inspection of H&E morphology where available, and prior biological knowledge of OAC tissue composition. The final broad lineage labels were used as the basis for subsequent lineage-restricted fine annotation.

Lineage-restricted InSituType annotation

Fine cell-type annotation was performed using InSituType in supervised mode. Rather than applying a single reference containing all cell types to the full CosMx dataset, InSituType was run separately within each refined broad lineage. For each CosMx lineage, the corresponding subset of cell types from the integrated scRNA-seq reference was used to construct a lineage-matched reference. This approach reduced competition between unrelated cell types and improved the biological specificity of fine annotation within each compartment.

For each lineage-specific InSituType run, the scRNA-seq reference was subset to include only cell types relevant to the corresponding CosMx broad compartment. For example, CosMx stromal cells were annotated using only fibroblast reference populations, including F1–F7 fibroblast states, from the scRNA-seq reference. Similarly, epithelial CosMx cells were annotated using epithelial reference populations; endothelial and mural cells were annotated using endothelial, pericyte, and smooth muscle or mural reference populations; lymphoid cells were annotated using T-cell, B-cell, plasma-cell, NK-cell, and related lymphoid reference populations; and myeloid cells were annotated

using monocyte, macrophage, dendritic-cell, and mast-cell reference populations where present.

For each lineage-specific reference, genes were restricted to those shared between the scRNA-seq reference and the CosMx 6,000-plex panel. Differentially expressed genes defining each reference cell type or state were identified from the scRNA-seq reference, and genes overlapping the CosMx panel were retained. These genes were used to generate pseudo-bulk expression profiles for each reference cell type or cell state. The resulting lineage-specific pseudo-bulk matrices were supplied to InSituType together with the CosMx raw count matrix and negative control count matrix.

InSituType was then run independently for each broad CosMx lineage, producing fine cell-type or cell-state predictions within that lineage only. The final fine annotation for each CosMx cell was therefore constrained by its manually refined broad lineage. This prevented, for example, stromal cells from being assigned epithelial or immune labels and ensured that fine labels were selected only from biologically plausible candidates within each compartment.

Final CosMx annotation curation

Following lineage-restricted InSituType annotation, predicted fine labels were reviewed within each broad compartment. Annotation quality was assessed using marker-gene expression, spatial distribution, cluster structure, and consistency with the scRNA-seq reference. Where InSituType predictions were inconsistent with marker expression or spatial context, annotations were manually reviewed and, where appropriate, adjusted at the cluster level rather than the individual-cell level.

The final CosMx annotation therefore consisted of two linked levels: a refined broad lineage label and a lineage-restricted fine cell-type or cell-state label. The broad lineage label was used for compartment-level analyses and tumour-cell selection, whereas the fine annotation was used to investigate cell-state heterogeneity within epithelial, stromal, endothelial and mural, lymphoid, and myeloid compartments. Where fine-grained subtype resolution was limited by low cell numbers, restricted CosMx gene-panel coverage, or transcriptional similarity between closely related populations, annotations were collapsed to broader biologically interpretable classes; consequently,

not all 34 scRNA-seq reference cell types were resolved in the CosMx data, and closely related populations such as T-cell subtypes were grouped as “T cells”.

Inference of copy-number alteration signal from CosMx expression data

Large-scale copy-number alteration (CNA)-like signal was inferred from cell-level CosMx spatial transcriptomic profiles using infercnvpy, the Python implementation of the inferCNV approach. Because this analysis infers CNV signal indirectly from transcript abundance, results were interpreted as *putative large-scale CNA signal* rather than direct DNA copy-number calls. infercnvpy estimates CNV-like signal by comparing expression across genes ordered by genomic position to a reference population, followed by genomic smoothing, cell-wise centering, noise filtering and median filtering. Analyses were performed independently for each of the 10 CosMx samples, comprising five biopsy specimens and five matched post-treatment resections from OAC TRG3 cases. For each sample, genes from the CosMx 6,000-gene panel were annotated with chromosome and genomic position, restricted to autosomal genes with available coordinates, and ordered by genomic location. Cell annotations in Mid_confidence_fine were used to define malignant and non-malignant compartments. Cells annotated as Lymphoid**,** Myeloid**,** Endothelial and Stromal were used as the non-malignant reference population, while cells annotated as Cancer were evaluated for tumour-associated CNV signal.

infercnvpy was run using reference_key="Mid_confidence_fine" and non-malignant reference categories as above. A running genomic window of 100 genes was used (window_size=100), with the default step size of 10 genes unless otherwise specified. Sex chromosomes were excluded, retaining autosomal CNV profiles. The resulting smoothed CNV matrix was stored for each sample and visualised as chromosome-ordered heatmaps grouped by cell annotation. The infercnvpy documentation lists window_size=100, step=10, dynamic_threshold=1.5 and exclusion of chrX/chrY as defaults, so this wording is appropriate if you did not override those parameters.

For downstream summarisation, a per-cell CNV amplitude score was calculated as the mean absolute CNV signal across all inferred genomic windows:

$$\text{CNV amplitude}_i = \frac{1}{W}\sum_{w=1}^{W} | X_{iw} |$$

where $X_{iw}$is the inferred CNV signal for cell $i$in genomic window w, and $W$is the total number of inferred windows. This score was used to compare CNV burden across annotated cell types and to visualise the distribution of CNV signal in cancer versus non-malignant populations.

Identification and spatial mapping of malignant transcriptional programmes

Selection of malignant cells from the scRNA-seq reference

Malignant epithelial cells were selected from the harmonised single-cell RNA-seq reference using the consensus cancer-cell annotation. The RPCA-integrated object was used only to define harmonised cell identities and metadata. All transcriptional programme discovery was performed using the uncorrected RNA assay count matrix.

To minimise the influence of low-quality or poorly represented donors, only donors with at least 50 malignant epithelial cells were retained. To prevent donors with very large malignant-cell numbers from dominating downstream analyses, a maximum of 500 malignant cells per donor was randomly sampled using a fixed random seed. Donor, dataset, sequencing platform, treatment status, and tissue-source metadata were retained for downstream assessment of potential biological and technical associations with inferred programmes.

Protein-coding genes were retained based on Ensembl annotations obtained using biomaRt. Mitochondrial genes and ribosomal genes were excluded to reduce technical and housekeeping-driven variation. Raw UMI counts were extracted from the RNA assay and library-size normalised to counts per million (CPM). CPM values were log1p-transformed for variance estimation. Highly variable genes were identified by gene-wise variance across the log1p-transformed CPM matrix, and the top 3,000 genes were retained for NMF. For NMF input, the selected log1p-transformed CPM matrix was transformed back to linear scale using expm1, yielding a non-negative CPM-scale gene-by-cell matrix.

Consensus non-negative matrix factorisation

Non-negative matrix factorisation was performed using the Brunet multiplicative-update algorithm implemented in the NMF R package. The expression matrix, with genes as rows and cells as columns, was factorised into a gene-by-programme matrix W and a programme-by-cell matrix H. Factorisation was performed for ranks $k = 4$ to $k = 10$. For each rank, 50 random initialisations were performed, with a maximum of 400 iterations per run.

For each run, the W matrix was extracted and column-normalised so that gene weights for each programme summed to one. To identify stable consensus programmes, W matrices from all runs at a given rank were concatenated, and pairwise cosine similarity was calculated between programmes. Programmes were clustered using hierarchical clustering on cosine distance, defined as $1 -$ cosine similarity, with average linkage. The dendrogram was cut into k clusters, and consensus programmes were defined by averaging gene weights across programmes within each cluster, followed by column normalisation.

Selection of optimal factorisation rank

Programme stability was evaluated across values of k by calculating the mean within-cluster cosine similarity of programmes across independent NMF runs. This metric reflects the reproducibility of inferred transcriptional programmes across random initialisations. Candidate ranks were also assessed for biological interpretability, pathway enrichment, top-weighted marker genes, and distribution of programme usage across donors, datasets, and sequencing platforms.

Based on the balance between programme stability, interpretability, and avoidance of donor- or platform-dominated factors, $k = 5$ was selected for downstream analyses.

Scoring scRNA-seq malignant cells for consensus programmes using AUCell

To map cNMF-derived tumour transcriptional programmes onto malignant cells from the scRNA-seq reference, a rank-based enrichment strategy was implemented using AUCell. The consensus W matrix containing five tumour programmes was used to define

programme-specific gene sets. For each programme, genes were ranked by consensus weight, and the top 50 genes were selected as the initial programme gene set.

To reduce redundancy between programmes and improve specificity of programme scoring, gene overlap across the five programme gene sets was quantified. Genes assigned to more than one programme were removed, retaining only genes uniquely associated with a single programme. These refined non-overlapping gene sets were used for AUCell scoring.

Raw RNA assay counts from malignant scRNA-seq cells were library-size normalised to CPM. Gene-expression rankings were generated for each cell using AUCell_buildRankings(), with block-wise ranking enabled to improve memory efficiency. Programme enrichment was quantified using AUCell_calcAUC(), which calculates the area under the curve for each programme gene set within each cell's ranked gene list. The maximum rank considered for AUC calculation was restricted to the top 100 genes per cell to prioritise highly expressed genes and reduce the influence of low-level noise. The resulting AUCell scores were stored as cell-by-programme metadata in the Seurat object.

Because AUCell scores are relative to each cell's ranked expression distribution and may vary between datasets, AUC values were z-score normalised within cohort prior to categorical programme assignment. For each cell, the programme with the highest z-scored AUCell value was designated as the candidate dominant programme. To avoid forced assignment of weak or ambiguous programme states, two additional criteria were applied: a minimum z-score of 1.0 for the top programme and a minimum margin of 0.25 between the highest and second-highest z-scored programme scores. Cells failing the z-score threshold and margin criterion were labelled ambiguous. Cells satisfying both criteria were assigned a final AUCell-based tumour programme label.

Hallmark enrichment of consensus cNMF programmes (fgsea)

Hallmark pathway enrichment was performed directly on the consensus cNMF gene-programme matrix (consensus W), treating each programme as a ranked gene list. Consensus W (genes × programmes) was loaded from the cNMF output and gene identifiers were verified to be HGNC gene symbols matching MSigDB naming. Hallmark

gene sets were obtained using msigdbr (species = Homo sapiens, category = “H”) and converted into a list of pathways (each pathway containing its member gene symbols). For each cNMF programme, genes were ranked by their consensus W weight in that programme (decreasing order) and used as the statistic vector for enrichment testing. Enrichment testing was performed using fgseaMultilevel (from the fgsea package), which estimates enrichment without requiring a fixed permutation count and is stable for long ranked lists. For each programme, we retained enrichment results with associated NES and Benjamini-Hochberg adjusted P values (FDR).

Transcription factor activity analysis of cNMF tumour programmes

Transcription factor activity was inferred for the five malignant transcriptional programmes using DoRothEA regulons and the decoupleR framework. Malignant cells assigned to one of the five cNMF tumour programmes by AUCell were selected from the scRNA-seq reference. Cells labelled as ambiguous were excluded from this analysis.

For each programme, RNA expression was aggregated across all malignant cells assigned to that programme using Seurat’s AggregateExpression() function with the RNA assay and log-normalised expression values from the data slot. This produced a gene-by-programme expression matrix representing the average transcriptional profile of each cNMF programme.

Transcription factor activities were inferred using run_mlm() from decoupleR with the human DoRothEA regulon database. DoRothEA confidence levels A, B, and C were retained to balance regulon confidence and coverage. The regulon network was restricted to target genes present in the aggregated scRNA-seq expression matrix. Transcription factors with fewer than 10 measured target genes were excluded. For each transcription factor, decoupleR’s multivariate linear model estimated an activity score for each cNMF programme based on the expression of its signed target-gene regulon.

To identify transcription factors preferentially active in each tumour programme, a programme-specificity score was calculated for each transcription factor and programme. For a given transcription factor t and programme p, specificity was defined as:

$$Specificity(t,p) = Activity(t,p) - mean(Activity(t,q))$$

where q represents all other cNMF programmes. For each programme, the top 15 transcription factors with the highest specificity scores were selected. The union of these programme-specific transcription factors was retained for visualisation.

For the heatmap shown in panel f, transcription factor activity scores for the selected transcription factors were extracted across the five cNMF programmes and z-score scaled across programmes for each transcription factor. This row-wise scaling was used to visualise the relative activity pattern of each transcription factor across tumour programmes. Heatmaps were generated using ComplexHeatmap, with cNMF programmes displayed as columns and transcription factors as rows.

Selection of representative programme genes

To provide a compact and interpretable summary of gene-level drivers for each programme, representative genes were selected from the consensus W matrix using a combined weight-and-specificity criterion. For each gene g and programme p, a specificity ratio was calculated as:

$$Specificity(g,p) = \frac{W_{g,p}}{\max_{q \neq p}(W_{g,q}) + \epsilon}$$

where $W_{g,p}$ is the consensus weight of gene g in programme p, q indexes all other programmes, and ϵ is a small constant to avoid division by zero. Genes were ranked within each programme using the combined score:

$$Score(g,p) = W_{g,p} \times Specificity(g,p)$$

For each programme, representative genes were selected by prioritising genes with high combined scores, programme-specific weighting, and biological relevance. Where genes were highlighted in the Results section and were present in the W matrix, they were preferentially included, and remaining slots were filled using the highest-ranked genes by the combined weight-specificity score. Four to six representative genes were selected per programme.

Summarising programme phenotypes across mechanistic axes

To summarise each tumour programme's biological phenotype, a set of mechanistic axes was defined a priori. These axes included proliferation, interferon response, MHC class I antigen presentation, stress or AP-1 activation, Notch-associated plasticity, polarity or junctional organisation, secretory differentiation, and regenerative repair. Each axis was represented by a compact gene set selected based on pathway biology, marker coherence, and relevance to epithelial tumour-cell states.

Axis scores were calculated per cell using Seurat's AddModuleScore(), with each mechanistic gene set scored against matched control features. Axis scores were summarised across cells assigned to each tumour programme using the median. For comparability across programmes, median axis scores were scaled to the unit interval from 0 to 1 across programmes for each axis:

$$x' = \frac{x - min(x)}{max(x) - min(x) + \epsilon}$$

where x is the median axis score and ϵ is a small constant to avoid division by zero. Radar plots were generated using the scaled axis values, with axes and scaling held constant across all programmes to enable direct comparison of phenotypic profiles.

Projection of tumour transcriptional programmes into CosMx spatial data

CosMx malignant-cell processing for programme projection

CosMx spatial transcriptomics data were subset to include high-confidence malignant epithelial cells based on the final consensus and manually refined annotations. Raw UMI count matrices were extracted from the RNA assay with genes as rows and cells as columns. To ensure compatibility with the scRNA-seq-derived cNMF programmes, CosMx counts were normalised to CPM without log-transformation, preserving non-negativity for downstream NNLS projection.

Only genes shared between the CosMx panel and the consensus cNMF W matrix were retained. The consensus programme matrix was subset to this shared gene set, and programme columns were re-normalised after subsetting to account for differences in gene coverage between whole-transcriptome scRNA-seq and targeted CosMx profiling.

Projection of cNMF programmes onto CosMx data using non-negative least squares

Consensus cNMF programmes derived from scRNA-seq malignant cells were projected onto individual CosMx malignant cells using non-negative least squares regression. NNLS was used rather than retraining NMF on the spatial data, allowing the scRNA-seq-derived tumour programmes to be transferred into spatial coordinates while avoiding the introduction of additional latent structure from the lower-complexity CosMx panel.

For each CosMx cell, its CPM-normalised expression vector x was modelled as a non-negative linear combination of the consensus programmes:

$$x \approx Wh$$

where W is the shared-gene consensus programme matrix and h is the vector of non-negative programme usage coefficients for that cell. NNLS regression was performed independently for each cell using the nnls R package. The resulting coefficients were retained as continuous programme-usage scores. Where relative programme composition was required, coefficients were normalised per cell to sum to one.

Unlike simple correlation or matrix multiplication, NNLS constrains programme contributions to be non-negative and additive. This makes it suitable for estimating the contribution of multiple transcriptional programmes within each CosMx cell while preserving interpretability of programme usage scores.

AUCell scoring of tumour programmes in CosMx data

As a complementary approach to NNLS-based programme projection, CosMx malignant cells were scored using AUCell. The same refined, non-overlapping programme gene sets used for scRNA-seq scoring were applied to the CosMx data, restricted to genes present in the CosMx panel.

AUCell rankings were generated for each CosMx cell, and programme enrichment was quantified using AUCell_calcAUC()with the same maximum-rank parameter used for scRNA-seq scoring. AUCell scores were z-score normalised within each sample to account for sample-specific variation in transcript detection, segmentation, and local RNA density. Dominant programme labels were assigned using the same z-score and margin criteria applied to the scRNA-seq data.

AUCell was used as a rank-based measure of programme identity, whereas NNLS was used to estimate continuous programme usage. Using both approaches enabled categorical assignment of dominant tumour-cell state together with quantitative assessment of programme composition.

Spatial visualisation of tumour transcriptional programmes

Programme usage scores, dominant programme assignments, assignment confidence, and ambiguity status were visualised in tissue coordinates across CosMx sections. Spatial maps were generated using cell centroid coordinates and overlaid with tumour programme labels or continuous programme scores. These visualisations were used to assess spatial organisation of malignant transcriptional states, intra-tumoural heterogeneity, co-localisation of programmes, and boundary regions between transcriptionally distinct tumour compartments.

Where relevant, spatial programme distributions were compared with H&E morphology, manually refined malignant annotations, and local tissue architecture to support biological interpretation of tumour-cell states in situ.

Spatial autocorrelation analysis using Moran's I

Spatial autocorrelation of malignant cNMF programme activity was assessed in CosMx spatial transcriptomics data using global Moran's I. Cancer cells were identified using the final refined cell-type annotation and were defined as cells labelled "Cancer". Cell centroid coordinates were extracted from the CosMx metadata and converted from pixels to micrometres using a scaling factor of 0.12 µm per pixel. Only cells from retained fields of view were included in the analysis.

Analyses were performed on continuous cNMF programme scores rather than thresholded programme labels. This preserved the full range of programme activity and allowed Moran's I to test whether neighbouring cancer cells had similar levels of each transcriptional programme.

For each analysis group, cancer-cell centroid coordinates were used to construct a spatial neighbour graph. Each cancer cell was connected to its eight nearest neighbouring cancer cells using a k-nearest-neighbour approach. The neighbour graph

was converted to a row-standardised spatial weights matrix. Global Moran's I was then calculated separately for each continuous cNMF programme score using the spdep R package. Positive Moran's I values were interpreted as evidence that neighbouring cancer cells had similar programme activity, consistent with spatial clustering of that malignant transcriptional state. Values close to zero indicated limited spatial autocorrelation.

Statistical significance was assessed using both the analytical Moran's I test and a Monte Carlo permutation-based Moran's I test. Analytical tests were performed using spdep::moran.test(). For the Monte Carlo test, spdep::moran.mc() was used with 1,000 permutations. In each permutation, cNMF programme scores were randomly reassigned across the fixed cancer-cell spatial graph, preserving the cell coordinates and neighbour relationships while disrupting the association between spatial position and programme activity. The observed Moran's I statistic was compared with this permutation-derived null distribution to obtain an empirical P value. P values were adjusted across the five cNMF programmes using the Benjamini-Hochberg false discovery rate procedure.

Comparison of cNMF tumour programme activity between paired biopsies and resections

To compare malignant transcriptional programme activity between diagnostic biopsies and matched resections, CosMx objects from five paired OAC cases were analysed. High-confidence malignant epithelial cells were used for the comparison. For each cell, cNMF programme activity scores were extracted from the metadata and converted to long format, producing a cell-by-programme table containing patient, condition, sample identifier, programme, and programme score.

For visualisation of the cell-level distribution of programme activity, programme scores were plotted by condition using violin plots with overlaid boxplots. These plots show the distribution of single-cell cNMF programme scores across biopsy and resection samples for each programme. Mean differences between resection and biopsy scores were calculated for each programme and displayed as Δ values, where positive values indicate higher scores in resection samples.

To account for the paired patient structure, per-patient summaries were calculated for each condition and programme. For each patient, condition, and cNMF programme, the median programme score across malignant cells was calculated. These paired patient-level medians were visualised using connected line plots, with each line representing one patient and showing the change from biopsy to matched resection. Paired Wilcoxon signed-rank tests were used to compare biopsy and resection median programme scores for each cNMF programme. P values were adjusted across the five programmes using the Benjamini-Hochberg false discovery rate procedure.

As an additional cell-level analysis accounting for repeated sampling within patients, linear mixed-effects models were fitted separately for each cNMF programme. Programme score was modelled as a function of condition, with biopsy as the reference level and patient included as a random intercept:

$$Score_{ij} = \beta_0 + \beta_1 Condition_{ij} + u_{patient} + \epsilon_{ij}$$

where $\beta_1$estimates the change in programme score in resection relative to biopsy. Models were fitted using the lme4 R package. For each programme, the condition effect was reported as the estimated resection-minus-biopsy change with an approximate 95% confidence interval calculated as estimate ± 1.96 × standard error. Statistical significance was assessed by comparing the full model with a null model lacking the condition term using a likelihood ratio test. P values were adjusted across programmes using the Benjamini-Hochberg method.

The violin plots were used to display the single-cell distribution of cNMF programme activity, the paired line plots were used to show patient-level biopsy–resection changes, and the mixed-effects model was used to estimate the overall condition effect while accounting for inter-patient variation.

Spatial niche identification and cancer-centred neighbourhood analysis

Spatial niche analysis was performed using the final annotated CosMx Seurat objects. For the analyses shown in panels c–f, resection samples were used. Cell centroid coordinates were extracted from the CosMx metadata using the global x and y pixel

coordinates and converted to micrometres using a scaling factor of 0.12 µm per pixel. Cell identities were taken from the final refined annotation.

Unsupervised spatial niche identification

To identify recurrent multicellular neighbourhoods within individual tissue sections, unsupervised niche analysis was performed using Seurat's BuildNicheAssay() function. For each selected CosMx section, a niche assay was constructed using final refined cell-type annotations as the categorical input. For each cell, the local cellular neighbourhood was defined using its 30 nearest spatial neighbours. The composition of these local neighbourhoods was encoded as a cell-type count matrix and clustered into discrete spatial niches.

Spatial niche labels were mapped back onto the tissue coordinates and visualised using ImageDimPlot(). To interpret the cellular composition of each niche, the niche count matrix was extracted and converted to neighbourhood cell-type proportions. For each niche, the mean local cell-type composition was compared with the background cell-type frequency across the section. Cell-type enrichment within each niche was quantified using Fisher's exact tests on aggregated neighbourhood cell-type counts. P values were adjusted using the Benjamini-Hochberg method, and cell types with FDR < 0.05 and log2 fold enrichment greater than log2(1.5) were considered enriched. Enriched cell types were used to annotate the dominant cellular constituents of each niche.

Selection of spatial niche number

The number of spatial niche clusters was selected using a data-driven assessment of clustering performance across candidate values of k. For each CosMx section, spatial niche features were first generated using Seurat's BuildNicheAssay() function with final refined cell-type annotations as input. Local neighbourhoods were defined using the 30 nearest neighbours of each cell. The resulting niche assay encoded, for each cell, the composition of its local cellular neighbourhood.

To evaluate the optimal number of niche clusters, candidate values from $k = 2$ to $k = 10$ were tested. For each candidate k, k-means clustering was applied to the niche

composition matrix using 50 random starts. Prior to clustering, the niche matrix was filtered to remove cells with zero neighbourhood counts and was scaled across features. Clustering performance was assessed using total within-cluster sum of squares, between-cluster sum of squares, the ratio of between-cluster sum of squares to total sum of squares, average silhouette width, and minimum and maximum cluster size.

To assess clustering robustness, repeated k-means runs were performed for each candidate k, and pairwise adjusted Rand indices were calculated between replicate cluster assignments. The mean adjusted Rand index was used as a stability metric. Candidate niche numbers were ranked by average silhouette width, clustering stability, and parsimony, with a small penalty applied to larger values of k. The selected niche number was then used to rebuild the final niche assay using BuildNicheAssay(), and the resulting niche labels were mapped back to spatial coordinates for visualisation.

Elbow plots of total within-cluster sum of squares and line plots of average silhouette width and clustering stability were inspected to support the final selection of niche number. Final niche numbers were chosen to balance statistical separation, clustering stability, cluster size, and biological interpretability of the resulting spatial patterns.

Cancer-centred radius-based neighbourhood composition

To quantify the cellular composition surrounding malignant epithelial cells, a cancer-centred neighbourhood analysis was performed across the resection samples. Cancer cells were defined as cells labelled “Cancer” in the final refined annotation. For each sample, Euclidean distances were calculated from each cancer-cell centroid to neighbouring cells using the global micrometre coordinates.

For each cancer cell, fixed-radius neighbourhoods were defined at radii of 10, 25, 50, 75, 100, and 200 µm using dbscan::frNN(). The central cancer cell itself was excluded from its own neighbourhood. For each radius, the proportion of neighbouring cells belonging to each fibroblast or immune subtype was calculated as:

$$\%_{celltype} = \frac{n_{celltype\ within\ radius}}{n_{all\ neighbours\ within\ radius}} \times 100$$

where the denominator included all neighbouring cells within the radius, excluding the index cancer cell. This generated a per-cancer-cell neighbourhood composition profile for each radius.

For each sample, percentages were averaged across all cancer cells to obtain a sample-level mean percentage for each cell type and radius. Sample-level values were then averaged across resection samples to generate curves.

Fibroblast subtypes included quiescent fibroblast 1, quiescent fibroblast 2, universal fibroblast, transitionary fibroblast, stress responsive CAF, MyoCAF, and desmoplastic CAF. Immune subtypes included T cells, Tregs, NK cells, B cells, plasma cells, M1 macrophages, M2 macrophages, monocytes, dendritic cells, and granulocytes. Curves were plotted using ggplot2, with radius in micrometres on the x-axis and the mean percentage among neighbours on the y-axis.

Distance-to-cancer analysis of non-malignant cell types

To assess the spatial proximity of stromal and immune populations to malignant epithelial cells, nearest-neighbour distances to cancer cells were calculated for all non-cancer cell types in resection samples. Cell centroid coordinates were extracted from the global CosMx coordinate metadata and converted from pixels to micrometres using a scaling factor of 0.12 µm per pixel.

For each resection sample, the Euclidean distance from every cell to its nearest cancer cell was calculated using the get.knnx() function from the FNN R package. Distances were calculated independently within each sample to avoid measuring distances across separate tissue sections. For each non-cancer cell type, empirical cumulative distribution functions were generated from the distribution of nearest-cancer distances within each sample. ECDF values were evaluated over a regular distance grid from 0 to 500 µm in 5 µm increments.

To summarise distance distributions across samples, ECDF values were averaged across resection samples for each cell type and distance. The resulting curves show, for each distance threshold, the mean fraction of cells of a given type located within that distance of the nearest cancer cell. MyoCAFs and M2 macrophages were highlighted, while other non-cancer cell types were shown in grey. This analysis was used to compare

the relative proximity of MyoCAFs, M2 macrophages, and other stromal or immune cell types to malignant epithelial regions.

Tri-cell cancer–fibroblast–immune niche analysis

To quantify whether cancer cells were embedded in multicellular niches containing both fibroblast and immune populations, a tri-cell neighbourhood analysis was performed. Cancer cells were used as anchor cells. For each cancer cell, fixed-radius neighbourhoods were queried to determine whether at least one fibroblast subtype and at least one immune-cell subtype were present within the specified radius.

Fibroblast subtypes included quiescent fibroblast 1, quiescent fibroblast 2, universal fibroblast, transitionary fibroblast, stress responsive CAF, MyoCAF, and desmoplastic CAF. Immune subtypes included T cells, Tregs, NK cells, B cells, plasma cells, M1 macrophages, M2 macrophages, monocytes, dendritic cells, and granulocytes. Cancer cells were also stratified by cNMF tumour programme where relevant.

For each sample, radius-based neighbour searches were performed using dbscan::frNN() on cell centroid coordinates in micrometres. For every cancer cell, binary indicators were recorded for whether a given fibroblast subtype was present within the radius and whether a given immune subtype was present within the same radius. A tri-cell niche was defined as a cancer cell for which both conditions were true:

$$TriCellNiche = hasFibroblastSubtype \cap hasImmuneSubtype$$

For each sample, patient, condition, cancer programme, fibroblast subtype, immune subtype, and radius, the percentage of cancer cells in a tri-cell niche was calculated as:

$$\%TriCellNiche = \frac{n_{\,cancer\ cells\ with\ both\ fibroblast\ and\ immune\ neighbour}}{n_{cancer\ cells}} \times 100$$

For the analysis shown in panel h, the model focused on MyoCAF-containing tri-cell niches. The outcome was the percentage of cancer cells with both a MyoCAF and a given immune-cell subtype nearby. To determine whether M2 macrophages were preferentially associated with MyoCAF-containing cancer niches after accounting for

differences in cell-type abundance, a linear mixed-effects model was fitted using lme4. The model included immune-cell subtype, fibroblast subtype, cancer programme, immune-cell abundance, fibroblast abundance, and cancer-programme abundance as fixed effects, with patient included as a random intercept:

$$\%TriCellNiche \sim ImmuneSubtype + FibroblastSubtype + CancerProgram + scale(\%Immune) + scale(\%Fibroblast) + scale(\%CancerProgram) + (1 \mid Patient)$$

M2 macrophage was used as the reference immune subtype. Sample-level immune, fibroblast, and cancer-programme abundances were calculated as the percentage of all cells in each sample assigned to the corresponding class and were included as scaled covariates. Estimated marginal means were calculated using emmeans to obtain abundance-adjusted tri-cell niche frequencies for each immune subtype, allowing comparison of M2 macrophages with other immune populations while accounting for sample-level cell-type abundance and patient structure.

CellPhoneDB ligand–receptor analysis of the cancer–MyoCAF–M2 macrophage niche

Cell–cell communication analysis was performed on the integrated single-cell RNA-seq reference using CellPhoneDB. Cancer cells were stratified by their dominant cNMF tumour programme, generating five malignant sender/receiver groups corresponding to the Immune-primed, Proliferative, Invasive, Stress repair, and Stress adapted programmes. Stromal and immune populations of interest were defined from the fine cell-type annotations.

CellPhoneDB was run using the single-cell expression matrix and corresponding cell-group metadata. The resulting ligand–receptor interaction table was exported in long format, retaining source cell type, target cell type, ligand gene, receptor gene, interacting pair, CellPhoneDB classification, significant mean, and interaction score. Downstream analysis focused on interactions involving the tri-cell niche of interest: cNMF-defined cancer programmes, MyoCAFs, and M2 macrophages.

Interactions were filtered to retain directed edges between cancer programmes and MyoCAFs, cancer programmes and M2 macrophages, and reciprocal MyoCAF–M2 macrophage interactions. Self-interactions within the same stromal or immune population were excluded..

To aid biological interpretation, ligand–receptor interactions were collapsed into manually curated pathway categories using the CellPhoneDB classification field and ligand/receptor gene identities. Interactions were assigned to the following pathway classes: WNT/RSPO, BMP, TGF-beta, FGF, NOTCH, Myeloid-conditioning, Immune evasion, HGF/MET, EGFR/ERBB, and ECM/integrin. Interactions that did not match these predefined categories were retained as “Other” where appropriate.

For global visualisation of the niche communication network, directed chord diagrams were generated using circlize. Cell sectors represented cNMF cancer programmes, MyoCAFs, and M2 macrophages. Links represented directed ligand–receptor interactions between source and target populations and were coloured by pathway class. Chord diagrams were generated either using the number of interactions per cell-pair/pathway or the summed CellPhoneDB interaction score.

To summarise pathway-level communication patterns, interactions were aggregated by directed cell pair and pathway. For each source–target pair, the number of interactions and the summed CellPhoneDB interaction score were calculated for each pathway. Summed interaction scores were log1p-transformed for heatmap visualisation. Heatmaps were generated using pheatmap, with rows representing directed cell pairs and columns representing pathway classes. Directed cell-pair ordering was fixed to show cancer-to-MyoCAF, cancer-to-M2 macrophage, MyoCAF-to-cancer, MyoCAF-to-M2 macrophage, M2 macrophage-to-cancer, and M2 macrophage-to-MyoCAF interactions.

To compare the overall number of interactions sent or received by each tumour programme, CellPhoneDB interactions were summarised by source and target group. For each cNMF cancer programme, the number of outgoing interactions to MyoCAFs and M2 macrophages was calculated, as was the number of incoming interactions from MyoCAFs and M2 macrophages. These counts were visualised as column-scaled heatmaps to highlight relative sending and receiving patterns across tumour programmes.

Curated ligand–receptor interaction plots

Curated ligand–receptor dot plots were generated to visualise selected CellPhoneDB interactions within the cancer–MyoCAF–M2 macrophage niche. For these plots, each point represents a directed ligand–receptor interaction between a source and target cell population. Point size represents the CellPhoneDB interaction score, and point colour represents the CellPhoneDB significant mean.

The CellPhoneDB significant mean is an expression-based summary of interaction strength for statistically significant ligand–receptor pairs. It reflects the mean expression of the ligand in the source cell group and the receptor in the target cell group, with multi-subunit complexes represented by the limiting subunit. In the CellPhoneDB significant-means output, values are retained for interactions passing the permutation-based significance threshold, whereas non-significant interactions are not assigned a positive significant-mean value. Higher significant-mean values therefore indicate stronger expression support for a statistically significant interaction in a given source–target direction, but should be interpreted as a relative communication score rather than a direct biochemical measure of ligand–receptor binding.

Interactions were selected to highlight cancer-programme-specific communication with MyoCAFs and M2 macrophages. For each ligand–receptor pair, interaction strength was compared across the five cNMF cancer programmes. Interactions were retained if they were unique to a single cancer programme or if their interaction strength in one cancer programme was at least 1.5-fold higher than in the other cancer programmes. This selection strategy was used to identify ligand–receptor interactions preferentially associated with individual malignant transcriptional states.

MyoCAF–M2 macrophage interactions were summarised at the pathway level. Within each curated pathway class, interactions were ranked by CellPhoneDB significant mean, and up to the top 12 interactions per pathway were retained for visualisation. Additional biologically relevant interactions were manually retained where appropriate to aid interpretation of the MyoCAF–M2 macrophage niche. Dot plots were faceted by pathway and sender–receiver direction.

Spatial-context differential expression of cancer cells in the MyoCAF–M2 macrophage niche

To identify cancer-cell transcriptional changes associated with localisation in a MyoCAF–M2 macrophage niche, spatial-context differential expression analysis was performed on CosMx resection samples. Resection samples were analysed independently and stored as sample-specific Seurat objects. Final refined cell-type annotations and cNMF cancer-programme annotations were used as input. Cancer cells were defined as cells annotated as one of the five malignant cNMF programmes: Immune-primed, Proliferative, Invasive, Stress repair, or Stress adapted. Cells annotated as MyoCAF and M2 macrophage were used to define the niche of interest.

Cell centroid coordinates were extracted from the CosMx metadata using global x and y pixel coordinates. Distances were converted to micrometres using a pixel size of 0.12µm. For each sample, radius-based neighbourhoods were constructed using dbscan::frNN(). For each cell, all neighbouring cells within the specified radius were identified, excluding the index cell itself. For cancer cells, the number of neighbouring MyoCAFs and M2 macrophages within the radius was counted.

Each cancer cell was assigned to one of four spatial-context categories:

$$\text{MyoCAF\_M2\_niche} = n_{\text{MyoCAF}} > 0 \ \cap \ n_{\text{M2 macrophage}} > 0$$

$$\text{MyoCAF\_only} = n_{\text{MyoCAF}} > 0 \ \cap \ n_{\text{M2 macrophage}} = 0$$

$$\text{M2\_only} = n_{\text{MyoCAF}} = 0 \ \cap \ n_{\text{M2 macrophage}} > 0$$

$$\text{Neither} = n_{\text{MyoCAF}} = 0 \ \cap \ n_{\text{M2 macrophage}} = 0$$

The primary contrast for differential expression was cancer cells in the "MyoCAF_M2_niche" compared with cancer cells in the "Neither" context. This contrast was performed separately within each cNMF cancer programme.

For each cancer programme and spatial context, raw RNA counts were aggregated into sample-level pseudobulk profiles. Pseudobulk counts were generated by summing raw counts across cancer cells from the same sample, cancer programme, and spatial-context category. Contexts containing fewer than 10 cells in a given sample were excluded. Only samples with paired pseudobulk profiles for both "MyoCAF_M2_niche"

and “Neither” contexts were retained for the corresponding programme-level comparison. Programmes with fewer than three paired samples were not tested.

Differential expression was performed using edgeR. Genes were initially filtered to retain genes with a total count of at least 10 and expression in at least two pseudobulk samples, followed by filterByExpr(). Library-size normalisation was performed using trimmed mean of M-values normalisation with calcNormFactors(). For each cancer programme, a quasi-likelihood negative binomial generalized linear model was fitted with sample identity and spatial context as covariates:

$$\text{Expression} \sim \text{sample} + \text{spatial context}$$

The spatial-context coefficient estimated the log2 fold change for MyoCAF_M2_niche versus Neither while accounting for paired sample structure. Dispersion was estimated using robust estimation, and differential expression was assessed using glmQLFit() and glmQLFTest(). P values were adjusted for multiple testing using the Benjamini-Hochberg false discovery rate procedure. Genes with FDR < 0.1 and absolute log2 fold change ≥ 0.25 were considered significant for exploratory interpretation.

Niche-radius sensitivity analysis

To assess the effect of neighbourhood radius on niche-associated differential expression, the spatial-context assignment and pseudobulk differential expression workflow were repeated across a series of candidate niche radii (15, 20, 25, 30, 35, 40, 45, 50, 55, 60uM). For each radius, cancer cells were reassigned to “MyoCAF_M2_niche”, “MyoCAF_only”, “M2_only”, or “Neither” based on the presence of MyoCAFs and M2 macrophages within that distance. Programme-specific differential expression was then repeated for the “MyoCAF_M2_niche” versus “Neither” contrast.

For each radius and cancer programme, the number of significant genes was counted using the threshold FDR < 0.1 and absolute log2 fold change ≥ 0.25. These counts were plotted against niche radius to evaluate the stability of the transcriptional signal across neighbourhood definitions. The main downstream visualisations used the 25 μm niche definition.

Gene-category summarisation and heatmap visualisation

To summarise biological patterns in the spatial-context differential expression results, genes were grouped into manually curated categories. The primary categories shown were epithelial/glandular genes and stress/inflammatory genes. For each cancer programme and gene category, the number of significant genes was calculated using FDR < 0.1 and absolute log2 fold change ≥ 0.25. The mean log2 fold change among significant genes in each category was also calculated. These values were visualised using a dot plot, where dot size represents the number of significant genes and colour represents the mean log2 fold change among significant genes.

For heatmap visualisation, selected epithelial/glandular and stress/inflammatory genes were extracted from the differential expression results. Genes were retained for plotting if they were significant in at least one cancer programme. Heatmap values represent the edgeR log2 fold change for “MyoCAF_M2_niche” versus “Neither” within each cancer programme. Positive values indicate higher expression in cancer cells located in the MyoCAF–M2 macrophage niche, whereas negative values indicate higher expression in cancer cells lacking nearby MyoCAFs and M2 macrophages. Values were capped for colour-scale visualisation. Significance symbols indicate genes passing the log2 fold-change threshold of ≥0.25 together with FDR thresholds of <0.1, <0.01, or <0.001.