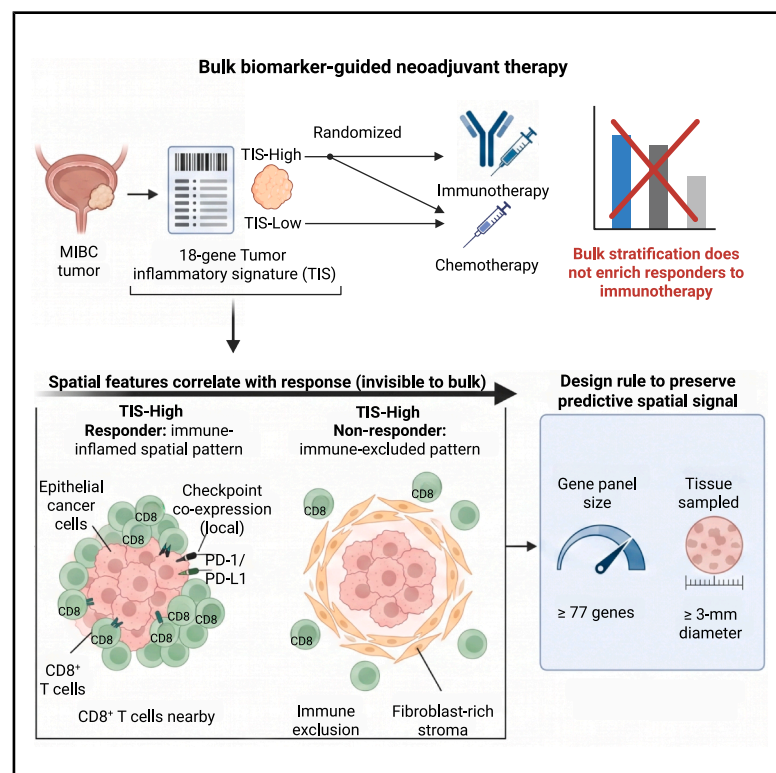


Spatial architecture contributes to failure of bulk biomarker-guided neoadjuvant immunotherapy selection in bladder cancer: The DUTRENEO study

Graphical abstract



Authors

Enrique Grande, Mustafa Sibai, Daniela Grases, ..., Ignacio Durán, Eduard Porta-Pardo, Francisco X. Real

Correspondence

egrande@oncomadrid.com (E.G.), eporta@carrerasresearch.org (E.P.-P.), freal@cnio.es (F.X.R.)

In brief

Grande et al. show that prospective selection of bladder cancer patients for neoadjuvant immunotherapy based on a bulk tumor inflammation score does not enrich for responders. Spatial transcriptomics points to CD8-tumor proximity, checkpoint niches, and fibroblast exclusion as spatial correlates of response, and signal is retained at ≥ 77 genes and ≥ 3 -mm sampling.

Highlights

- Patient selection using tumor inflammatory score (TIS) does not enrich for ICI responders
- A community resource of 5.4 million spatially resolved cells from 20 bladder cancer patients
- In TIS-high tumors, spatial architecture distinguishes responders from non-responders
- A framework for clinical translation of spatial transcriptomics biomarkers



Article

Spatial architecture contributes to failure of bulk biomarker-guided neoadjuvant immunotherapy selection in bladder cancer: The DUTRENEO study

Enrique Grande,^{1,30,*} Mustafa Sibai,^{2,28,30} Daniela Grases,² Elena Andrada,^{3,4} Oscar Reig,^{5,6,7} Marc Escobosa,² Aina Azueta,^{8,9} Daniel Castellano,¹⁰ Javier Puente,¹¹ Jaime Martínez de Villarreal,^{3,4} Albert Font,¹² Teresa Alonso-Gordoa,¹³ Raquel Benítez,^{4,14} Ane Moreno-Oya,^{4,14} Mario Álvarez-Maestro,¹⁵

(Author list continued on next page)

¹Medical Oncology Department, MD Anderson Cancer Center Madrid, Facultad de Medicina, Universidad Francisco de Vitoria, Madrid, Spain

²Josep Carreras Leukaemia Research Institute, Badalona, Spain

³Epithelial Carcinogenesis Group, Spanish National Cancer Research Centre-CNIO, Madrid, Spain

⁴Centro de Investigación Biomédica en Red Cáncer, Madrid, Spain

⁵Medical Oncology Department, Hospital Clinic, Barcelona, Spain

⁶Translational Genomics and Targeted Therapies in Solid Tumours, IDIBAPS, Barcelona, Spain

⁷Department of Medicine, University of Barcelona, Barcelona, Spain

⁸Pathology Department, Hospital Universitario Basurto, Bilbao, Spain

⁹Translational Prostate Cancer Research Lab, Biobizkaia Health Research Institute, Barakaldo, Spain

¹⁰Servicio de Oncología Médica, Hospital 12 de Octubre, Madrid, Spain

¹¹Medical Oncology Department, Hospital Clínico San Carlos, Instituto de Investigación Sanitaria del Hospital Clínico San Carlos (IdISSC), Madrid, Spain

¹²Medical Oncology Department, Institut Català d'Oncologia, Badalona Applied Research Group in Oncology (B-ARGO Group), Institut Investigació Germans Trias i Pujol (IGTP), Badalona, Spain

¹³Medical Oncology Department, Hospital Universitario Ramón y Cajal, Madrid, Spain

¹⁴Genetic and Molecular Epidemiology Group, Spanish National Cancer Centre-CNIO, Madrid, Spain

¹⁵Hospital La Paz, Madrid, Spain

¹⁶Urology Department, Hospital Ramón y Cajal, Alcalá University, IRYCIS, Madrid, Spain

¹⁷Servicio de Oncología Médica, Fundación Instituto Valenciano de Oncología, Valencia, Spain

¹⁸Servicio de Urología, Hospital Marqués de Valdecilla, Santander, Spain

¹⁹Cancer Institute and Blood Disorders, Hospital Clinic, Barcelona, Spain

(Affiliations continued on next page)

SUMMARY

Predictive biomarkers for immune checkpoint inhibitors (ICIs) are largely identified retrospectively, but their prospective clinical utility remains unproven. Here, we report DUTRENEO, a prospective randomized phase 2 trial testing whether a retrospectively validated 18-gene bulk tumor inflammation signature (TIS) can guide neoadjuvant ICI therapy in muscle-invasive bladder cancer. The trial does not meet its primary endpoint, and this demonstrates that bulk gene-expression stratification does not sufficiently enrich for responders. To define the biology underlying this failure, we generate single-cell spatial transcriptomic profiles of 377 genes in ~5.4 million cells across large tissue areas. Response is governed by spatial architectures invisible to bulk assays, including CD8⁺ T cell proximity to cancer cells, localized checkpoint co-expression within epithelial cancer-rich neighborhoods, and fibroblast-rich immune-excluded communities in non-responders. We provide a quantitative framework showing that ≥77 genes and ≥3-mm tissue diameter regions preserve predictive spatial signal at scalable throughput. The registration details of the trial are EudraCT 2017-002246-68.

INTRODUCTION

More than 570,000 new cases of bladder cancer are diagnosed every year worldwide, causing approximately 200,000 yearly deaths.¹ While most of them are non-muscle invasive tumors

(75%), a subset of patients (25%) present with muscle-invasive bladder cancer (MIBC), an aggressive disease. Radical cystectomy remains the most common therapeutic approach, although cisplatin-based neoadjuvant chemotherapy (NAC) provides an absolute 5%–8% increase in 5-year overall survival (OS) and a



Javier Burgos,¹⁶ M. Angel Climent,¹⁷ Mario Domínguez,¹⁸ Patricia Galván,^{6,19} Isabel Galante,²⁰ Juan F. García,²¹ Elena Perez,² Xavier García del Muro,²² Félix Guerrero-Ramos,²³ Miriam Marqués,^{3,4} Pablo Maroto,²⁴ Jesús M. Paramio,²⁵ Alvaro Pinto,²⁶ Aleix Prat,^{6,7,19} Núria Malats,^{4,14} Ignacio Durán,²⁷ Eduard Porta-Pardo,^{2,28,*} and Francisco X. Real^{3,4,29,31,*}

²⁰Servicio de Urología, Hospital Clínico San Carlos, Madrid, Spain

²¹Pathology Department, MD Anderson Cancer Center, Madrid, Spain

²²Institut Català d'Oncologia. IDIBELL, University of Barcelona, Barcelona, Spain

²³Urology Department, Hospital 12 de Octubre, Madrid, Spain

²⁴Oncología Médica, Hospital Sant Pau, Barcelona, Spain

²⁵Molecular and Translational Oncology, CIEMAT, Biomedical Research Institute, University Hospital "12 de Octubre", CIBERONC, Madrid, Spain

²⁶Medical Oncology Service, Hospital Universitario La Paz - IdiPAZ, Madrid, Spain

²⁷Hospital Universitario Marqués de Valdecilla, IDIVAL, Santander, Spain

²⁸Barcelona Supercomputing Center, Barcelona, Spain

²⁹Departament de Medicina i Ciències de la Vida, Universitat Pompeu Fabra, Barcelona, Spain

³⁰These authors contributed equally

³¹Lead contact

*Correspondence: egrande@oncomadrid.com (E.G.), eporta@carrerasresearch.org (E.P.-P.), freal@cnio.es (F.X.R.)

<https://doi.org/10.1016/j.xcrm.2026.102878>

16% relative reduction in the risk of cancer-related death.^{2–6} Recently, the NIAGARA study has shown that adding the anti-PDL1 immune checkpoint inhibitor (ICI) durvalumab to the standard regimen of cisplatin and gemcitabine as neoadjuvant therapy followed by durvalumab for 1 year after cystectomy increases the OS of patients. This regimen has been recently incorporated into international guidelines for the management of these tumors.⁷ Yet, as the therapeutic landscape shifts toward these immune-based combinations, the fundamental spatial barriers to immune infiltration (such as stromal exclusion) remain critical determinants of efficacy that are likely to impact response across regimens. Despite this evidence, real-world data reveal that there is a significant underuse of NAC, possibly related to concerns about the relatively modest absolute survival benefit, treatment-related toxicity, and the fact that 40%–50% of patients are medically ineligible to receive cisplatin-based NAC.⁸

Small single-arm phase 2 trials using ICI doublets have revealed pathological complete response (pCR) rates in the range of 37%–46%.^{9–12} Retrospective analyses in several tumor types, including urothelial bladder cancer, have identified gene expression profiles associated with T cell infiltration, high interferon gamma (IFN- γ), and low TGF β signaling^{13–15} as predictive biomarkers of response to ICI. Particularly, a tumor inflammation signature (TIS) score based on 18 genes related to IFN- γ signaling has been reported.¹⁵ In these studies, the TIS score is one of the most robust biomarkers predictive of response to ICI, and its ability to discriminate is preserved across tumor types. The TIS has also been associated with prognosis at the PanCancer level and has been proposed as an indirect measure of an inflamed tumor phenotype that might guide immunotherapy across cancer types.^{16,17}

A major problem in translational research is that few of the promising markers identified in retrospective studies, including the TIS, are applied prospectively to demonstrate clinical utility to stratify and optimize patient treatment. To our knowledge, the DOMINI trial in patients with stage III melanoma is the only study that randomized patients to allocate treatment according to a signature predictive of response to ICI.¹⁸ In this trial, patients whose tumors had a high IFN- γ score showed a higher response

rate than those with a low score.¹⁸ However, this study did not include a control arm of standard therapy.

Here, we present the results of the DUTRENEO trial, a prospective, biomarker-driven, randomized study testing the 18-gene TIS to guide neoadjuvant therapy selection in MIBC. While the trial provides a definitive prospective test of TIS utility, the primary endpoint was not met, demonstrating that bulk inflammation profiling alone is insufficient for robust patient stratification in this setting. However, this clinical outcome provided a unique opportunity to decipher the reasons for biomarker failure. We generated the largest single-cell spatial transcriptomics dataset in bladder cancer to date (377 genes analyzed in ~5.4 million cells), revealing that response is governed by specific architectural constraints, such as fibroblast-mediated exclusion and immune-tumor distance, that are invisible to bulk assays. Leveraging this dataset, we not only provide a mechanistic explanation for the limitations of bulk profiling but also establish a quantitative framework for the design of future spatial biomarkers, illustrating how gene panel size and tissue sampling area influence the ability to capture predictive spatial features in MIBC.

RESULTS

Study design, patient characteristics, and safety

The DUTRENEO trial was a multicenter, open-label phase 2 study conducted in 11 Spanish tertiary care hospitals (Figure 1A). Patients were eligible following a transurethral resection of bladder tumor (TURBT) demonstrating MIBC (T2–T4a), regardless of the presence of visible residual disease. They were also eligible for cisplatin therapy, had a planned cystectomy as part of their treatment plan, were Eastern Cooperative Oncology Group (ECOG) 0–1 performance status, and had tumor tissue available for TIS quantification at baseline recruitment. A total of 101 patients were screened for study eligibility. Of them, 28 had screening failures (Figure S1A; STAR Methods). Upon establishment of the TIS score, patients whose tumors were TIS-high (i.e., hot) were randomized to receive ICI or standard cisplatin-based NAC. In contrast, patients whose tumors

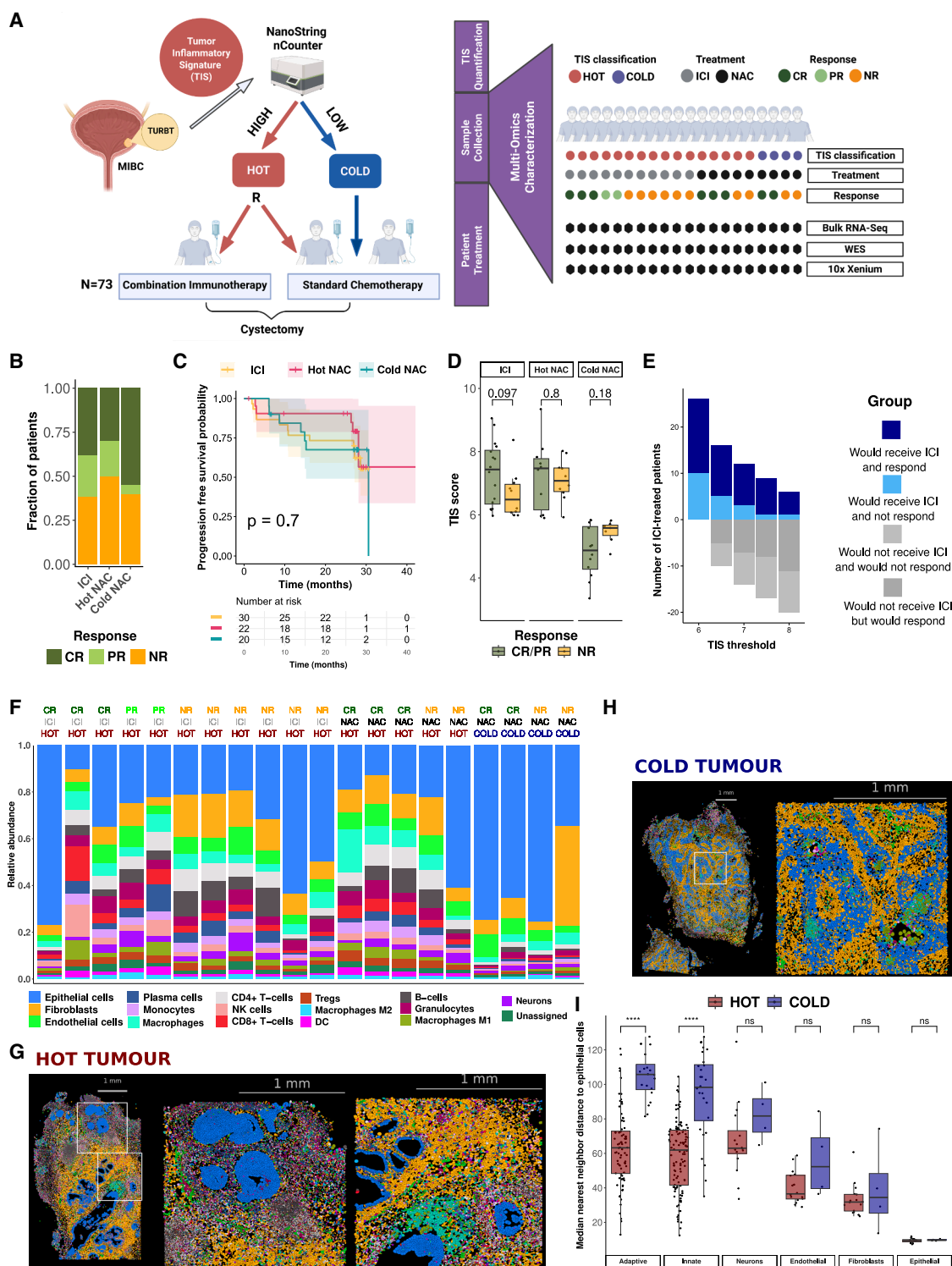


Figure 1. Study design, outcome, and multiomics analysis

(A) DUTRENEO study design. Patients' tumors were classified as immune-hot (hot, for simplicity) or immune-cold (cold, for simplicity) based on the score of the TIS signature. Patients with hot tumors were randomized to receive ICI or NAC, whereas patients with cold tumors received standard NAC. Multiomics analysis of pre-treatment tumor samples was performed using WES ($n = 44$), bulk RNA sequencing ($n = 66$), as well as spatial transcriptomics with Xenium ($n = 20$). (B) Barplot showing the fraction of patients who had complete (CR), partial (PR), or no (NR) pathological response in each of the treatment arms.

(legend continued on next page)

Table 1. Baseline characteristics of patients recruited to the DUTRENEO study

	Hot ICI N = 30	Hot NAC N = 22	Cold NAC N = 21	p value
Age (years), median (IQR)	66.0 (60.2;71.8)	66.0 (61.0;72.8)	66.0 (62.0;72.0)	0.904
Gender, n (%)	–	–	–	0.471
Female	3 (10.0%)	1 (4.55%)	3 (14.3%)	–
Male	27 (90.0%)	21 (95.5%)	18 (85.7%)	–
ECOG, n (%)	–	–	–	0.192
0	19 (63.3%)	18 (81.8%)	12 (57.1%)	–
1	11 (36.7%)	4 (18.2%)	9 (42.9%)	–
Clinical stage, n (%)	–	–	–	>0.999
T2	23 (76.7%)	17 (77.3%)	17 (81.0%)	–
T3	7 (23.3%)	5 (22.7%)	4 (19.0%)	–
Lymph node invasion, n (%)	–	–	–	0.474
N0	24 (80.0%)	21 (95.5%)	19 (90.5%)	–
N1	5 (16.7%)	1 (4.55%)	2 (9.52%)	–
Nx	1 (3.33%)	0 (0.00%)	0 (0.00%)	–
Histologic class, n (%)	–	–	–	0.519
Adenocarcinoma	0 (0.00%)	0 (0.00%)	1 (4.76%)	–
Infiltrating ductal carcinoma	1 (3.33%)	1 (4.55%)	0 (0.00%)	–
Glandular differentiation	2 (6.67%)	1 (4.55%)	1 (4.76%)	–
Micropapillary	1 (3.33%)	0 (0.00%)	4 (19.0%)	–
Squamous differentiation	5 (16.7%)	2 (9.09%)	3 (14.3%)	–
Urothelial carcinoma (pure)	21 (70%)	17 (81.9%)	11 (57.2%)	–
Prior BCG, n (%)	–	–	–	0.452
Yes	4 (13.3%)	4 (18.2%)	1 (4.76%)	–

BCG, Bacillus Calmette-Guérin; ECOG, Eastern Cooperative Oncology Group; IQR, interquartile range.

were TIS-low (i.e., cold) were assigned to receive standard NAC (STAR Methods). Overall, 73 patients were eligible and received treatment, as follows: 21 patients with cold tumors received cisplatin-based NAC according to the investigator's choice. Fifty-two patients had hot tumors. The first 6 patients were treated with durvalumab plus tremelimumab under a safety run-in period for at least two cycles, and the remaining 46 patients were randomized to ICI ($n = 24$) or NAC ($n = 22$), representing the full intention-to-treat population of the trial (Figure S1A). Cystectomy, post-surgery follow-up, and therapy were administered according to standard practice. Clinical and pathological baseline characteristics of patients in the three

arms were not significantly different (Table 1). Adverse effects and toxicities were in line with those previously described for NAC and ICI (Table 1).

Efficacy outcomes

Among patients with hot tumors, a pCR was observed in 10/26 patients treated with ICI (39%) and in 6/20 patients treated with NAC (30%). Among patients with cold tumors, 11/20 had a pCR (55%) (Table 2; Figure 1B). The primary endpoint was not met; pCR rates in hot tumors treated with ICI (39%) were not significantly superior to those treated with NAC (30%) (Fisher's exact test, $p = 0.51$, estimated odds ratio of 1.64,

(C) Kaplan-Meier plot showing progression-free survival for patients included in the three treatment arms. The p value was calculated using the log rank test. Shaded areas indicate 95% confidence intervals.

(D) Boxplot showing the TIS score by treatment arm and response. Boxplots show the median, interquartile range, and $1.5 \times$ interquartile range whiskers; points represent individual patients. p values were calculated using two-sided Wilcoxon rank-sum tests.

(E) Barplot showing the retrospective analysis of DUTRENEO outcomes. The y axis shows the fraction of patients that would have been included (positive values) or excluded (negative values) to receive neoadjuvant immunotherapy at different TIS thresholds (x axis). Bars are colored depending on the outcome of the patient.

(F) Relative cell-type abundance in the samples analyzed with the Xenium platform.

(G and H) Overlay of the major cell types in the whole sample analyzed and in selected regions of exemplary hot (G) and cold tumors (H). Scale bars, 1 mm.

(I) Boxplot showing the median nearest-neighbor distance of the different cell types to epithelial cells, calculated within a 200- μ m radius, in samples analyzed with the Xenium platform and split by TIS category. Median distances for individual adaptive and innate immune cell types were pooled under the "adaptive" and "innate" categories, respectively; each point represents the patient-level median distance for one constituent cell type. For non-grouped categories, each point represents the patient-level median distance for the indicated cell type. Boxplots show the median, interquartile range, and $1.5 \times$ interquartile range whiskers. Statistical significance was assessed using two-sided Wilcoxon rank-sum tests. ns: $p > 0.05$; * $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$; **** $p \leq 0.0001$.

Table 2. Pathological response and downstaging rates

	Hot ICI N = 26 ^a	Hot NAC N = 20 ^a	Cold NAC N = 20 ^a	p value
Pathological response, n (%)	–	–	–	0.349
Complete response	10 (38.5%)	6 (30.0%)	11 (55.0%)	–
Partial response	6 (23.1%)	4 (20.0%)	1 (5.00%)	–
No response	10 (38.5%)	10 (50.0%)	8 (40.0%)	–
Downstaging, n (%) ^b	–	–	–	0.710
Yes	16 (61.5%)	10 (50.0%)	12 (60.0%)	–

^aCystectomy was not performed in 4 patients from the hot ICI group (two patients progressed while receiving treatment, one patient refused cystectomy, one developed myocarditis) and 1 patient from the cold group who was incorrectly enrolled in this group. One patient from the hot chemotherapy group withdrew consent to use information on cystectomy outcome; a second patient developed grade 3 rash leading to a delay in cystectomy beyond protocol rules.

^bDownstaging was defined as any improvement in T or N stage from the time of randomization.

95% confidence interval [CI]: 0.36–7.89). Downstaging was achieved in 10/20 (50%) and 12/20 (60%) patients with hot or cold tumors, respectively, and in 16/26 (62%) of patients treated with ICI. Notably, patients with cold tumors treated with standard NAC achieved a numerically higher pCR rate (55%) than patients with hot tumors receiving the same chemotherapy regimen (30%), although this difference did not reach statistical significance (Fisher's exact test, $p = 0.31$, estimated odds ratio of 2.23, 95% CI: 0.48–11.13). This inverse trend underscores that bulk inflammation scores did not enrich for ICI responders and failed to identify the specific tumor biology associated with chemosensitivity.

There were no significant differences in progression-free survival across treatment groups ($p = 0.7$; reference hot NAC: hot ICI hazard ratio [HR] = 1.50, 95% CI: 0.56–4.04; cold NAC HR = 1.43, 95% CI: 0.48–4.26; Figure 1C) or in OS ($p > 0.6$; reference hot NAC: hot ICI HR = 2.16, 95% CI: 0.44–10.76; cold NAC HR = 1.45, 95% CI: 0.20–10.59; Figure S1B). We did not find significant differences in the TIS score of responders vs. non-responders in any of the treatment arms, although there was a tendency for higher scores in responders to ICI ($p = 0.097$) (Figure 1D). Finally, the pCR rate was similar to that observed in other studies assessing ICI in an unselected population of MIBC, suggesting that the TIS, as used in this trial, did not sufficiently enrich the treated population with responders. A retrospective analysis of the outcome data of DUTRENEO shows that higher TIS thresholds would have increased substantially the fraction of responding patients at the expense of excluding most patients who responded to the neoadjuvant immunotherapy (Figure 1E).

A comprehensive multi-omics characterization of the DUTRENEO cohort

We employed a multi-omics approach to characterize the pre-treatment tumor landscape and investigate the molecular determinants of response (Figure 1A). Formalin-fixed paraffin-embedded

tissues obtained during TURBT were analyzed via whole exome sequencing (WES) ($n = 44$), bulk RNA sequencing (RNA-seq) ($n = 66$), and high-resolution spatial transcriptomics with Xenium ($n = 20$, Figure 1A). The Xenium dataset profiled 377 genes across 5,399,355 single cells (median, ~224,000 cells per sample), providing a deep architectural map of the tumor microenvironment.

To determine whether the patient cohort is representative of primary untreated MIBC, we assessed cancer driver mutations and molecular subtypes. The frequency of driver mutations in key bladder cancer genes (e.g., *TP53*, *FGFR3*, *PIK3CA*, and *KMT2D*) aligns with larger public cohorts¹⁹ (Figure S2A). Consensus molecular subtype distribution (derived from bulk RNA-seq) is also consistent with prior reports,²⁰ apart from a known bias toward luminal subtypes inherent to TURBT sampling: LumP ($n = 20$), LumNS ($n = 9$), LumU ($n = 7$), Stromal-rich ($n = 9$), and Ba/Sq ($n = 12$) (Figure S2B). As expected, *FGFR3* mutations were enriched in LumP tumors. Subtype distribution was not associated with treatment allocation or response (Figures S2C–S2E).

Comparison of hot and cold tumors revealed that *TP53*, *RB1*, and *ARID1A* mutations were more frequent in hot tumors, whereas *FGFR3* and *STAG2* mutations were enriched in cold tumors¹⁹ (Figure S2A). Differential expression analysis confirmed that the TIS stratification successfully captured the intended biological phenotype: hot tumors displayed significant overexpression of immune-related genes (e.g., *IDO1* and *IFNG*) and enrichment of immune pathways (e.g., T cell receptor signaling and natural killer [NK]-mediated cytotoxicity) (Figure S3). Conversely, cold tumors were enriched for *FGFR3* expression and luminal subtypes (Figures S2A–S2C).

To resolve the spatial architecture underlying these bulk profiles, we used Xenium to map cell types and their neighborhoods. The cellular composition was highly heterogeneous (Figure 1F). Among hot tumors, composition did not differ significantly by treatment arm (Figure S4A). Importantly, cell composition correlated strongly with TIS score, confirming that “hot” tumors indeed contained higher fractions of immune cells, including B cells (3.7% vs. 0.9%, $p = 0.015$), CD4 T cells (4.6% vs. 1.4%, $p = 0.004$), Tregs (2.1% vs. 0.7%, $p = 0.001$), and CD8 T cells (4% vs. 0.7%, $p = 0.0004$) (Figures 1F–1H; Figure S4B). Furthermore, the median distance between epithelial cells and their nearest immune neighbor was significantly shorter in hot tumors (Figure 1I). Collectively, these data demonstrate that the TIS signature successfully discriminated inflamed from non-inflamed phenotypes and captured immune cell abundance. Therefore, its failure to predict clinical benefit suggests that immune pathway activity alone is insufficient, pointing to spatial organization as a critical determinant of response.

Immune checkpoint gene expression and response to ICI

Prior studies suggest that checkpoint gene expression may serve as a biomarker for checkpoint blockade efficacy.²¹ We investigated whether these markers were associated with clinical outcome in the DUTRENEO cohort. Protein levels of PDL1, assessed using standard immunohistochemistry (IHC), were not associated with clinical features or response to therapy (Figure S5). In contrast,

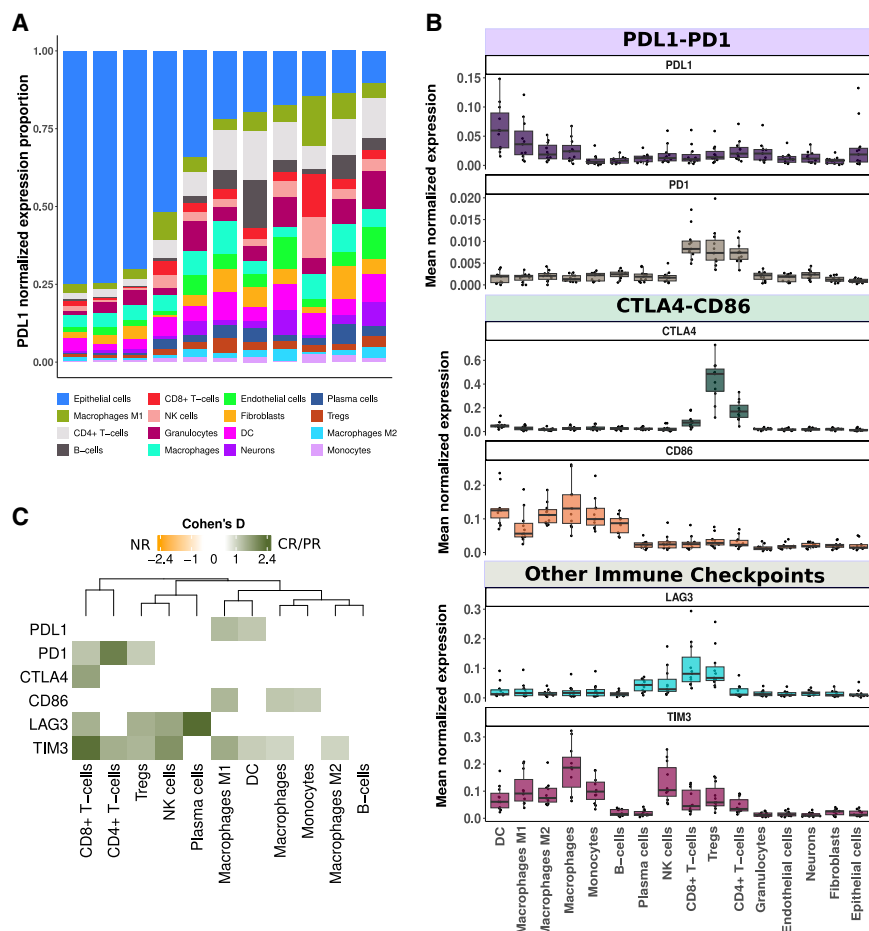


Figure 2. Gene expression landscape of immune checkpoint genes and their partners in ICI-treated patients

(A) Cell-type distribution of PDL1-expressing cells based on normalized expression. Each bar represents one ICI-treated patient, and colors indicate the relative contribution of each annotated cell type to total normalized PDL1 expression.

(B) Distribution of the mean normalized expression of PDL1, PD1, CTLA4, CD86, LAG3, and TIM3 across samples for each cell type. Boxplots show the median, interquartile range, and 1.5 $\times$ interquartile range whiskers; points represent individual patient-level mean expression values.

(C) Heatmap showing standardized mean differences (Cohen's d) in checkpoint gene expression or its binding partner between responders and non-responders, calculated separately within each annotated cell type based on the patient-level values shown in (B). Higher Cohen's d values in dark green indicate higher mean expression in the relevant cell type in responders (CR/PR), whereas higher values toward orange indicate higher mean expression in non-responders (NR). White cells indicate either Cohen's d values below the visualization threshold ($-1 < \text{Cohen's } d < 1$, set to 0 for display) or cell type-gene combinations that were not considered biologically relevant based on the expression patterns shown in (B). Group differences were assessed using two-sided Wilcoxon rank-sum tests; raw p values are reported.

bulk RNA-seq data revealed that tumors of responders harbored significantly higher transcript levels of *CD274* (PDL1) ($\log_2$ fold-change >1.8 , $p = 0.001$) and *CTLA4* ($\log_2$ fold-change >2 , $p = 0.002$) compared with non-responders (Figure S3).

To dissect the cellular origins of this bulk signal, we analyzed cell-type contributions using Xenium. *PDL1* expression displayed a bimodal distribution with respect to its cellular source: in 4/11 patients, $>50\%$ of *PDL1* transcripts originated from cancer cells ("epithelial dominant"), whereas in 6/11 patients, $<25\%$ came from cancer cells (Figure 2A). Importantly, this source-defined bimodality was not solely explained by total epithelial cell abundance and did not independently stratify response (Figure S6A). Rather, tumor-level differences in the relative contribution of PDL1 transcripts reflect a combination of cellular composition and cell-type-specific transcriptional regulation across epithelial and immune compartments, rather than differences in per-cell expression alone. While *CTLA4* expression was predominantly Treg derived (13%–34% of transcripts; Figure S6B), it was the expression specifically within CD8 $^+$ T cells (contributing 2%–30%) that strongly associated with ICI response ($p = 0.0043$, Figure S6C).

We further examined the mean expression of checkpoints (*PDL1*, *CTLA4*, *LAG3*, and *TIM3*) and their binding partners (*PD1* and *CD86*) across all cell types (Figure 2B). To quantify

the specific differences between responders and non-responders, we calculated the standardized mean difference (Cohen's d) for each gene within its relevant cell types. This analysis revealed that the "inflamed" signal observed in bulk RNA-seq arises from a coordinated upregulation across multiple compartments rather than a single cell type. For instance, responders displayed significantly higher mean expression of *PD1* on CD4 $^+$ T cells ($p = 0.008$) and *CTLA4* on CD8 $^+$ T cells ($p = 0.03$, Figure 2C). Similarly, *TIM3* expression on CD8 $^+$ T cells and *LAG3* expression on plasma/NK cells were significantly elevated in responders ($p = 0.03$ and $p = 0.008$, respectively; Figure 2C). Collectively, these single-cell data resolve the mechanism behind the bulk RNA-seq association: responders are characterized by a coordinated increase in checkpoint intensity across specific adaptive and innate populations.

Cellular neighborhoods and spatial architecture of response

To determine whether spatial organization influences ICI efficacy beyond simple immune abundance, we analyzed the topology of the tumor microenvironment. First, we calculated the median distance between epithelial cells and their nearest neighboring cell types within a 200- μm radius. The spatial arrangement of these cells proved highly discriminatory: epithelial cells were

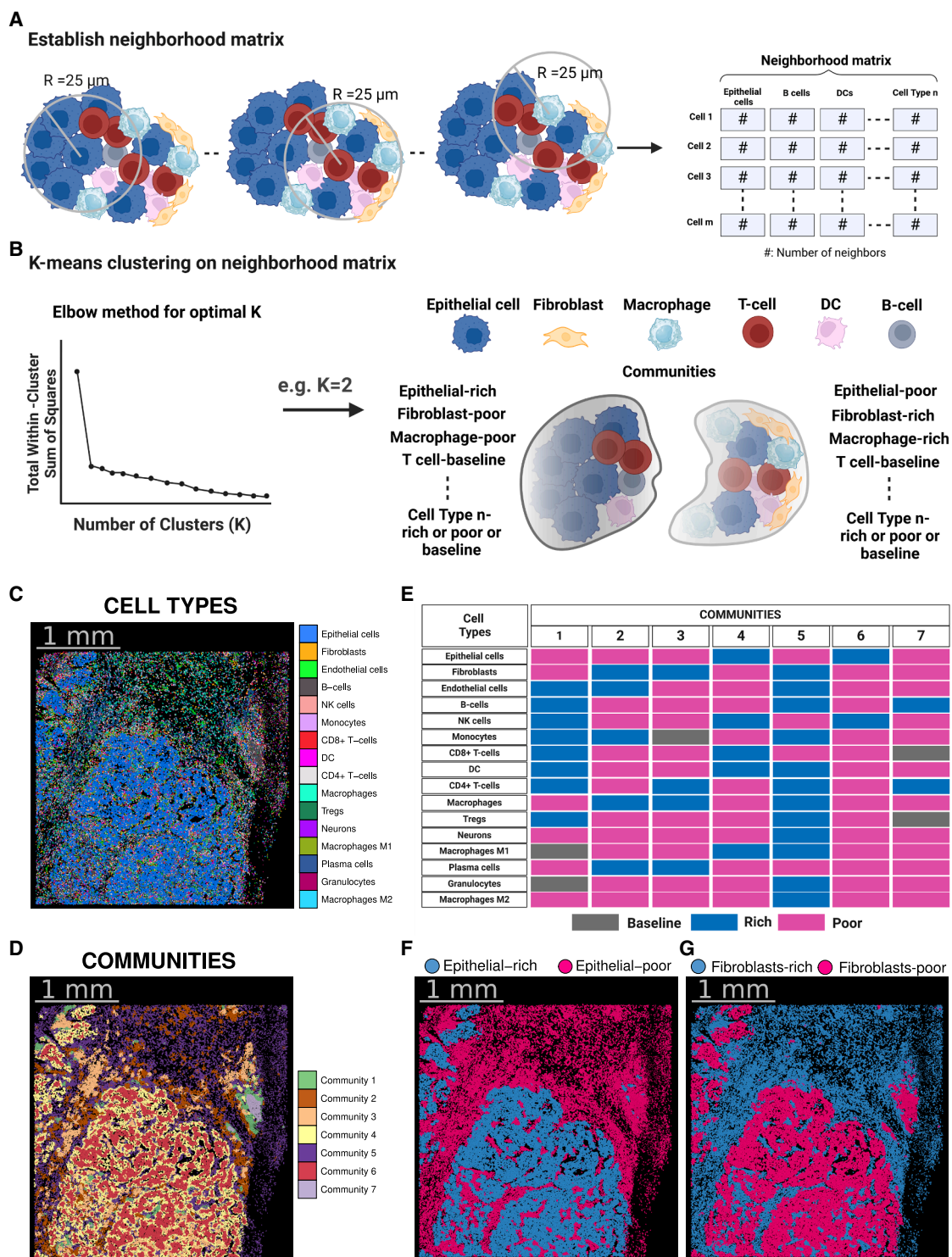


Figure 3. Neighborhood analysis strategy

(A and B) Scheme of the neighborhood analysis for the identification of communities and their cellular composition. First, a neighborhood radius of 25 μ m is established around each cell, within which the number of cells of a given type are counted to form a neighborhood matrix (A). Next, the elbow method is applied on the neighborhood matrix for each sample to determine an optimum number of clusters (i.e., communities) that will be used for *k*-means clustering. The relative abundance of each cell type within each community is finally determined, revealing its level of abundance as “baseline” (neutral), “rich” (high abundance), or “poor” (low abundance) (B).

(legend continued on next page)

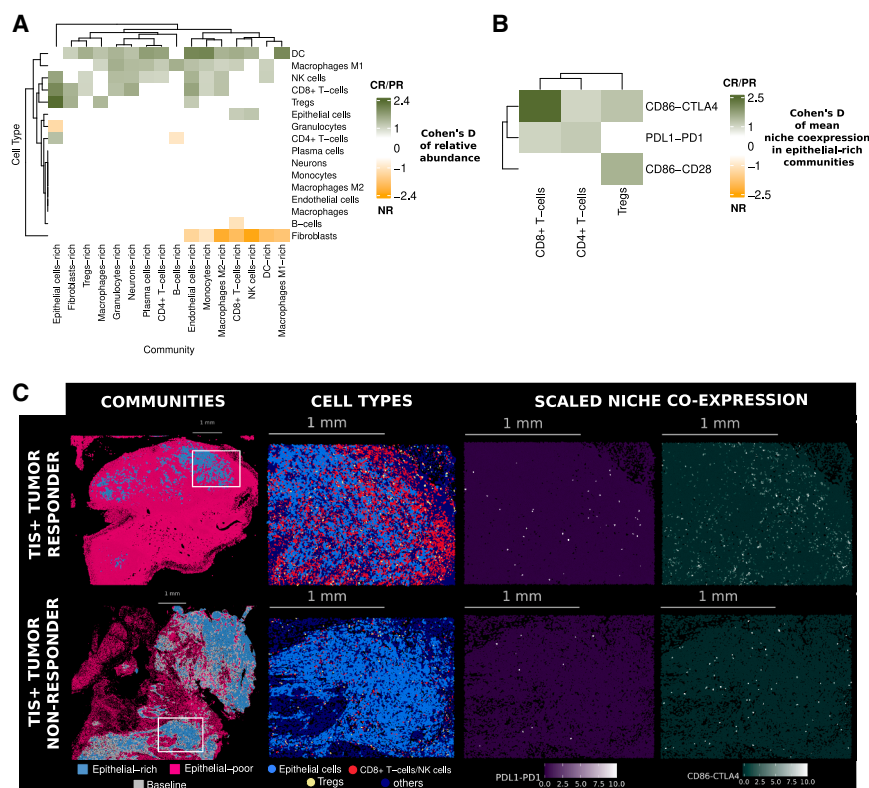


Figure 4. Cellular neighborhoods predict ICI response

(A) Heatmap showing Cohen's d of the relative abundance of each cell type in each cell-rich community, comparing the two response groups among cases from the ICI Xenium cohort. Higher Cohen's d values in dark green indicate higher relative abundance of a given cell type in each community in responders (CR/PR), whereas higher values toward orange indicate higher relative abundance in non-responders (NR). Group differences were assessed using two-sided Wilcoxon rank-sum tests; raw *p* values are reported. (B) Heatmap showing Cohen's d of the mean niche co-expression of the ligand-receptor (LR) pairs of interest. Higher Cohen's d values in dark green indicate higher niche co-expression of an LR pair around a given T cell type in responders compared with non-responders, whereas higher values toward orange indicate higher niche co-expression in non-responders. Group differences were assessed using two-sided Wilcoxon rank-sum tests; raw *p* values are reported. For both heatmaps in (A) and (B), white cells indicate Cohen's d values below the visualization threshold ($-1 < \text{Cohen's } d < 1$, set to 0 for display).

(C) Selected findings in the tumors of one ICI responder (top) and one non-responder (bottom). From left to right: the segmented cells from epithelial-rich vs. epithelial-poor communities are shown, followed by the cell-type composition of the insets from the previous community images,

with a focus on epithelial cells, CD8 T cells/NK cells, and Tregs. The last two images refer to the scaled niche co-expression of LR pairs CTLA4-CD86 and PDL1-PD1 in the regions from the insets. Scale bars, 1 mm.

significantly closer to adaptive immune cells (mean, 54 vs. 75 μm , $p < 0.001$, Figure S7A) and innate immune cells (mean, 53 vs. 66 μm , $p < 0.05$, Figure S7A) in responders than in non-responders. By contrast, global immune cell abundance showed limited association with outcome, with fibroblasts being the only cell type whose overall abundance was significantly higher in non-responders (Figure S7B), consistent with a role for stromal exclusion. Together, these findings indicate that response to ICI depends less on the total abundance of immune cells than on their spatial proximity to epithelial cells.

To investigate local interaction networks, we next defined the spatial scale most relevant for direct cell-cell communication. We found that the majority of nearest-neighbor events occurred within a 25- μm radius (Figure S7C), establishing this as the operative window for subsequent neighborhood analyses. We therefore quantified the cellular composition within a 25- μm radius around each cell (Figure 3A) and used *k*-means clustering to identify cellular communities with similar neighborhood composition (Figure 3B; Figure S8). We then quantified the enrichment or depletion of each cell type within each community, allowing us to classify communities as "rich," "neutral," or "poor" for a given

cell type (Figure 3B). Finally, we grouped all cell type-rich communities within each sample for downstream comparisons (Figures 3C–3G).

This analysis revealed marked differences in local tissue architecture between responders and non-responders. Responders were characterized by an enrichment of dendritic cells, M1 macrophages, CD8⁺ T cells, Tregs, and NK cells, particularly within epithelial-rich communities (Figure 4A). These associations were especially pronounced for CD8⁺ T cells and Tregs ($p = 0.017$ and $p = 0.004$, respectively) in epithelial-rich communities, and for dendritic cells in monocyte-rich and endothelial-rich communities ($p = 0.004$ and $p = 0.008$, respectively). By contrast, non-responders showed a higher relative abundance of fibroblasts within macrophage-rich ($p = 0.03$ for M1 and $p = 0.017$ for M2 types), NK-rich ($p = 0.008$), and CD8-rich ($p = 0.03$) communities (Figure 4A), reinforcing the concept of fibroblast-mediated immune isolation.

Given these distinct architectural constraints, we hypothesized that therapeutic efficacy depends on the interplay between neighborhood composition and localized checkpoint engagement. To test this, we quantified the niche co-expression of ligand-receptor (LR) pairs in epithelial-rich communities,

(C–G) Example of the community identification and annotation workflow shown in (B) for a section of a hot tumor, including the overlaid cell types (C), the overlaid sample-specific communities defined by *k*-means clustering (D), the abundance of each cell type across these communities in the whole sample (E), and the same pre-defined communities highlighted according to their annotation as epithelial rich or epithelial poor (F) and fibroblast rich or fibroblast poor (G). Scale bars, 1 mm (C, D, F, and G).

focusing on the immediate vicinity of T cells. We define “niche co-expression” as the mean co-expression of a ligand and its receptor across the three neighboring cells of a target cell in epithelial-rich communities, thereby capturing the local signaling environment (Figure S9A). While *CD86-CD28* mean niche co-expression did not differ by response (Figure 4B), the mean niche co-expression of *CD86-CTLA4* around $CD8^+$ T cells was higher in responders ($p = 0.004$, Figures 4B and 4C). Similarly, *PDL1-PD1* niche co-expression around non-regulatory $CD4^+$ T cells was higher in responders ($p = 0.015$, Figures 4B and 4C), and this association became even stronger when response was stringently defined as pCR ($p = 0.009$, Figures S9B and S9C). Collectively, these findings indicate that response to ICI is governed not merely by the presence of immune cells, but by their capacity to infiltrate epithelial neighborhoods and engage in localized checkpoint signaling.

Spatially confined checkpoint signaling in responding patients

We hypothesized that the functional impact of immune checkpoint expression depends on the specific cellular community in which it occurs. To test this, we systematically compared the expression of checkpoints and their ligands within heterotypic “cell-type rich” vs. “cell-type poor” communities. As a proof of concept, we examined *PDL1* expression in M1 macrophages located within endothelial-rich vs. endothelial-poor niches (Figures 5A and 5B). In non-responders, M1 macrophages in endothelial-rich communities contributed significantly more to total *PDL1* expression than those in endothelial-poor zones (adjusted $p = 0.01$, Figure 5C), demonstrating that local neighborhood composition modulates immune gene output.

Extending this analysis to all ligand-receptor interactions, we observed a striking difference in network topology between groups. Responders displayed a “spatially constrained” signaling architecture, where high-intensity checkpoint expression was restricted to a few key biological niches (Figures 5D–5G, top). The strongest determinant of response was the confinement of *PD1* and *CTLA4* expression to $CD8^+$ T cells specifically within fibroblast-poor communities (Figures 5D and 5E). This reinforces our earlier finding that fibroblast exclusion is a prerequisite for effective T cell engagement. Similarly, *LAG3* and *TIM3* expression in responders was selectively enriched in NK and $CD8^+$ T cells within these same fibroblast-poor niches (Figures 5F and 5G), suggesting a spatially coordinated adaptive immune response.

In contrast, non-responders exhibited a “diffuse” signaling topology, characterized by widespread checkpoint gene expression across multiple heterogeneous niches (Figures 5D–5G, bottom; higher number of network edges). In these tumors, *PDL1* and *PD1* expression was driven by communities rich in M1 macrophages, monocytes, or endothelial cells, rather than adaptive immune niches (Figure 5D). Crucially, *CTLA4* expression in T cells was significantly higher in communities rich in monocytes and macrophages but depleted in epithelial communities (Figure 5E). This spatial sequestration of checkpoint-expressing T cells away from cancer cells (epithelial-poor niches) suggests that in non-responders, immune activation occurs in the wrong spatial compartment, generating “inflammatory noise” without

direct anti-tumor engagement. These findings indicate that the spatial confinement of immune signaling to relevant epithelial and fibroblast-poor niches, rather than the total magnitude of expression, is the hallmark of effective ICI response.

A generalizable framework for spatial study design

The next critical step for the field is to understand how to optimize spatial assays for scalable deployment in clinical trials in a cost-effective manner. Currently, this translation often defaults to minimizing costs via tissue microarrays (TMAs) with restricted sampling areas (~1-mm cores). However, it remains unclear whether such limited sampling strikes the correct balance between throughput and the reproducibility of complex spatial findings. To address this, we sought to establish a quantitative framework for converting high-dimensional spatial discovery into a feasible validation design, effectively asking what parameters would be required to power a prospective clinical trial.

We adopted a two-stage optimization strategy: first, identifying the spatial feature that best correlates with outcome in the discovery cohort and second, systematically defining the minimal tissue area and gene content required to measure that feature reproducibly. To identify the optimal target feature, we screened the full Xenium dataset for the metric with the highest discriminatory potential. While we identified complex neighborhood-based associations, the median shortest $CD8^+$ T cell-to-cancer cell distance emerged as the most robust correlate of response (area under the curve [AUC] = 0.90, 95% CI: 0.69–1.00, DeLong), outperforming other immune cell distances and abundance metrics (Figure 6A). Leave-one-out LASSO stability analysis confirmed $CD8^+$ T cell proximity as the most consistently selected feature (90.9% selection frequency; Figure 6B), validating its selection as the primary target for optimization.

Having identified this architectural feature, we next determined the minimal assay specifications needed to measure it, explicitly testing the validity of the field’s reliance on small tissue areas. We performed a systematic “downsampling” simulation to stress-test the stability of the $CD8$ -cancer distance metric. First, regarding gene content, we progressively downsized the annotation reference from the full 377-gene panel to 17 genes (Figure 6C). As the gene set was progressively reduced, annotation accuracy declined gradually but remained robust. Even at 47 genes, an ~8-fold reduction from the original 377-gene panel, the classifier retained substantial discriminative power, achieving an AUC of 0.75 (95% CI: 0.75–0.75, DeLong) for $CD8^+$ T cell identification and 0.87 (95% CI: 0.87–0.87, DeLong) for epithelial cells, representing strong agreement with the ground-truth labels despite considerable panel compression. Using these reduced annotations, we re-evaluated our spatial biomarker (median shortest $CD8^+$ T cell-to-cancer distance) and find that predictive performance remains relatively stable from the ground truth (377 genes, AUC = 0.90) down to 17 genes (AUC = 0.80; 95% CI: 0.41–1.00, DeLong), with an inflection at 77 genes (AUC = 0.87; 95% CI: 0.60–1.00, DeLong; Figure 6D). A control analysis using random gene sets yielded a mean AUC of 0.51 (Figure S10), confirming that the signal depends on biologically specific lineage markers.

Second, we addressed the critical question of sampling area, exploring the minimal region required to recapitulate our spatially

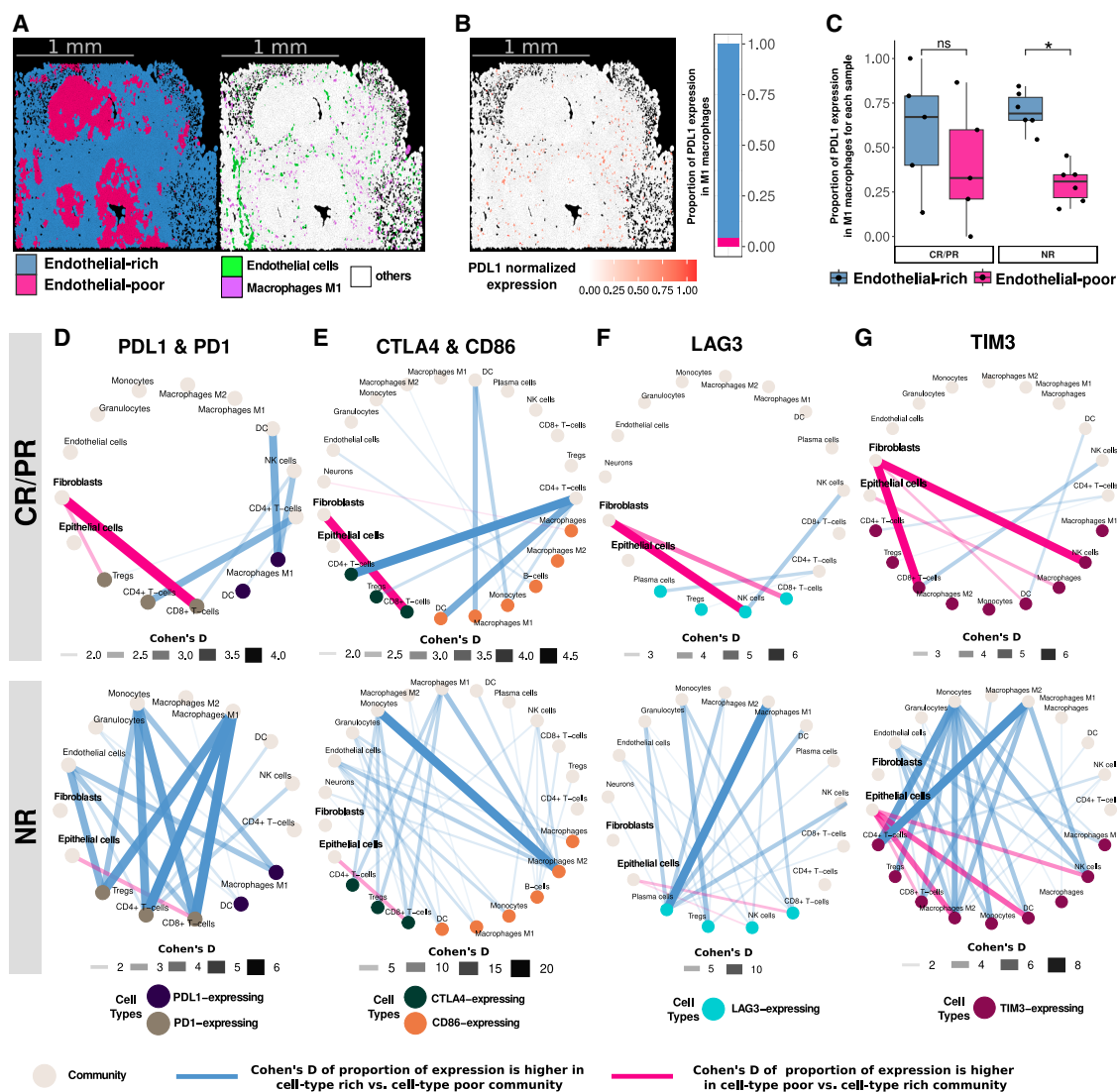


Figure 5. Spatially constrained checkpoint signaling associates with response to ICI

(A–C) Example illustrating how checkpoint output from a defined immune cell type varies according to community context in a section of a hot tumor. Specifically, these panels show the distribution of PDL1 transcripts in M1 macrophages across endothelial-rich and endothelial-poor communities, quantified as the relative share of PDL1 expression contributed by these two community types. (A) Overlay of endothelial-rich and endothelial-poor communities (left), followed by an overlay of endothelial cells (green), M1 macrophages (purple), and all other cell types (white) (right). Scale bars, 1 mm. (B) Overlay of normalized PDL1 expression, followed by a barplot showing the relative contribution of PDL1+ M1 macrophages in endothelial-rich and endothelial-poor communities within that tissue section. Scale bar, 1 mm. (C) The same relative contributions shown in (B), quantified across whole tissue sections from all ICI-treated samples and stratified by response ($n = 11$). Boxplots show the median, interquartile range, and $1.5 \times$ interquartile range whiskers; points represent individual patients. Statistical significance was assessed using two-sided Wilcoxon rank-sum tests comparing endothelial-rich and endothelial-poor community contexts within each response group, followed by Benjamini-Hochberg correction. ns, adjusted $p > 0.05$; *adjusted $p < 0.05$.

(D–G) Circular networks of the significant differences in magnitude, as calculated by Cohen's d , of the relative contribution of immune marker expression in relevant cell types by cell-rich and cell-poor communities in responders (top) or non-responders (bottom) for PDL1 and PD1 (D), CTLA4 and CD86 (E), LAG3 (F), and TIM3 (G). Nodes labeled as “community” are colored light brown, while the other nodes are colored based on the cell types relevant for each immune marker. Edge colors denote the direction of spatial enrichment: steel blue indicates that the proportion of expression is higher in cell-type-rich than in cell-type-poor communities, whereas magenta indicates that the proportion of expression is higher in cell-type-poor than in cell-type-rich communities. Edge widths indicate the magnitude of the association based on absolute Cohen's d . Statistical significance was assessed using two-sided Wilcoxon rank-sum tests comparing cell-type-rich and cell-type-poor community contexts within each response group, followed by Benjamini-Hochberg correction; only significant contrasts with adjusted $p < 0.05$ are displayed.

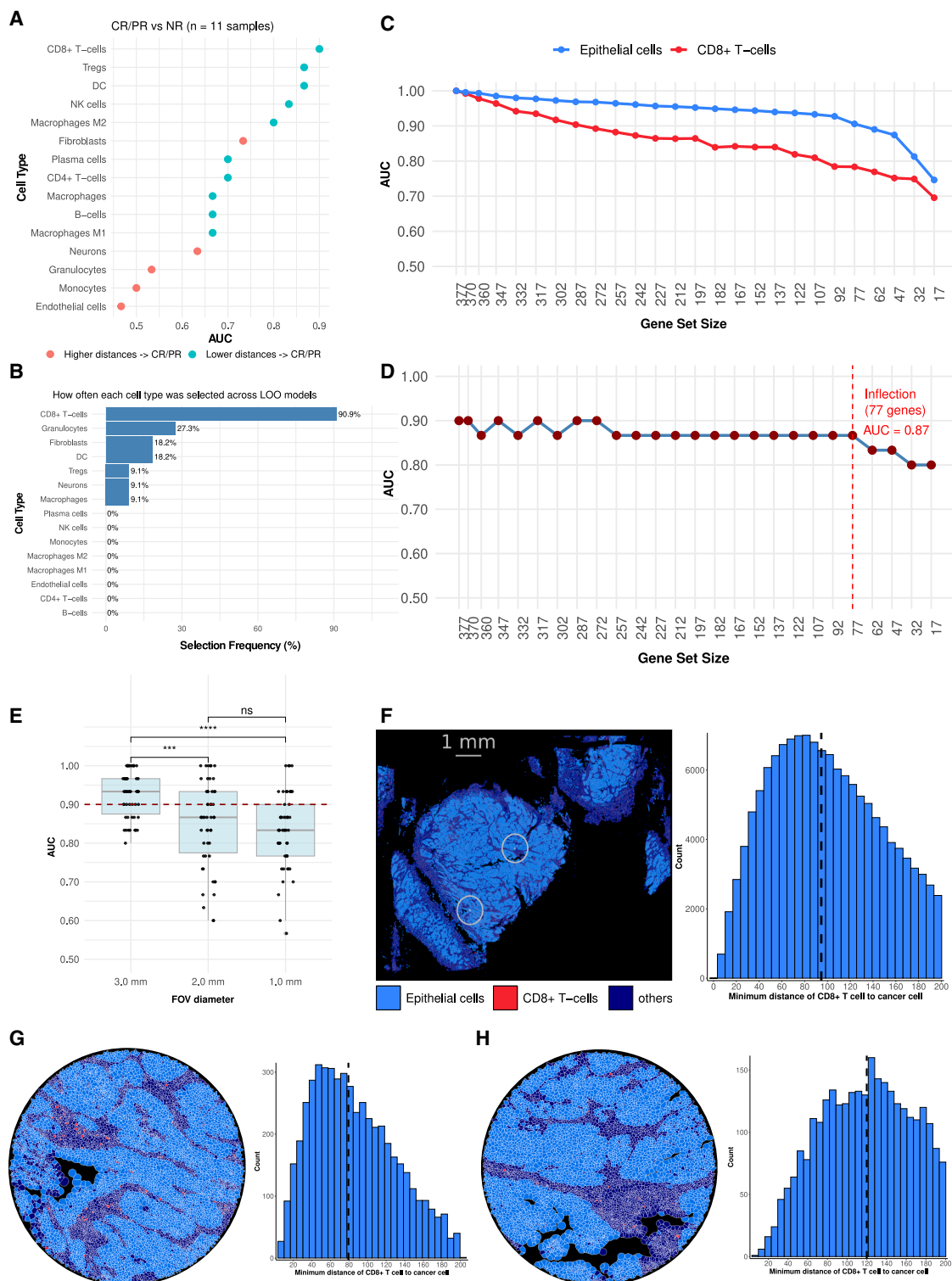


Figure 6. A practical roadmap for clinical research deployment of spatial Xenium biomarkers

(A) Predictive performance (AUC) for the median minimum distance of each immune and stromal cell type to epithelial (cancer) cells within a 200- μ m neighborhood, across the full Xenium dataset. Orange dots indicate cell types with larger distances in responders; cyan dots indicate cell types with shorter distances in responders. AUC values and 95% confidence intervals were calculated using the DeLong method.

(B) Leave-one-out LASSO feature selection stability showing the frequency of each cell type's selection as a predictive feature.

(legend continued on next page)

informed findings from the original (approximate) 5×5 -mm region per patient. We simulated clinical sampling constraints by randomly selecting 50 fields of view (FOVs) per sample at diameters of 3, 2, and 1 mm, ensuring adequate epithelial content and tissue coverage. Using ground-truth cell-type annotations, the median shortest CD8⁺ T cell-to-cancer distance within 3-mm FOVs closely recapitulated full-sample predictive performance for therapy response (mean AUC = 0.93; Figure 6E), with typical per-FOV 95% confidence bounds spanning approximately 0.79–1.00 (DeLong). Smaller sampling areas showed a measurable loss of predictive signal: 2-mm and 1-mm FOVs yielded lower mean AUCs of 0.85 and 0.83, respectively ($p < 0.001$), with correspondingly wider per-FOV confidence ranges (mean CI bounds ≈ 0.61 –1.00 and 0.57–1.00). These results indicate that spatial biomarker performance degrades with restricted tissue area, although substantial discriminative information is retained even at 1- to 2-mm scales. Visual inspection confirmed that 1-mm regions frequently failed to capture the relevant CD8⁺ infiltration heterogeneity observed at the whole-sample level (Figure 6F), leading to high intra-sample variance (Figures 6G and 6H).

In conclusion, while the specific parameters identified here (CD8-cancer distance, ≥ 77 genes, FOV ≥ 3 mm in diameter) are intrinsic to this therapeutic setting, the optimization pathway itself is generalizable. These data explicitly demonstrate that standard 1-mm TMAs may be underpowered to validate spatial neighborhood metrics. By rigorously defining the minimum information content required to preserve the signal of interest before designing a validation cohort, this framework provides a roadmap for moving from exploratory spatial profiling to scalable, statistically powered clinical trials.

DISCUSSION

The DUTRENEO trial represents one of the few prospective, biomarker-driven randomized studies in MIBC. While the primary endpoint was not met, demonstrating that the 18-gene TIS signature did not provide sufficient resolution to improve patient selection for neoadjuvant immunotherapy compared with chemotherapy,¹⁵ the study generated the largest single-cell spatial transcriptomics dataset in a clinical bladder cancer cohort to date. This allowed us to perform a “molecular autopsy” of the trial, revealing that the failure of bulk inflammation biomarkers stems from their inability to resolve the spatial architecture of immune response. We show that while bulk signatures can detect the presence of immune cells, they are blind to the

barriers, such as fibroblast-mediated exclusion and excessive immune-to-tumor distances, that ultimately dictate efficacy. These findings extend the immune phenotype framework established in bladder cancer by Mariathasan et al.,¹³ who showed that stromal exclusion and fibroblast TGF- β activity are associated with ICI resistance and concluded that much of the spatial biology of bladder cancer remains to be characterized at higher resolution.

The pCR rates observed in this study align with recent unselected neoadjuvant ICI trials,^{11,12} confirming that TIS-based stratification did not enrich responders. Our spatial analysis explains why: the “inflamed” phenotype captured by bulk RNA-seq is architecturally distinct in responders vs. non-responders. In responders, immune signaling is spatially constrained, characterized by the infiltration of CD8⁺ T cells into epithelial neighborhoods and the localized co-expression of checkpoints (e.g., CTLA4 and PD1) within specific fibroblast-poor niches. In contrast, non-responders display a diffuse signaling topology, where inflammatory markers are widely expressed but physically sequestered away from cancer cells by stromal barriers. Crucially, standard PDL1 IHC failed to capture these nuances.

Importantly, this limitation of IHC is not simply a matter of marker choice but reflects a fundamental constraint of low-plex spatial readouts. In our data, accurate preservation of CD8⁺ T cell and epithelial annotations, and of the resulting CD8-cancer distance metric, required multi-gene annotation panels on the order of ~ 77 genes. This finding indicates that single-plex or low-plex IHC, while indispensable for clinical practice, is unlikely to fully recapitulate the cell identity fidelity and neighborhood structure required to quantify proximity-based spatial biomarkers. Thus, multi-gene spatial profiling is likely not merely additive but probably necessary to capture the architectural information underlying immune efficacy. These findings suggest that the next generation of biomarkers must move beyond quantifying how much inflammation is present to characterizing where it is located,^{22–27} supporting the shift toward “functional spatial biomarkers.”^{28,29}

The landscape of neoadjuvant therapy is rapidly evolving, with the recent NIAGARA study establishing perioperative durvalumab plus chemotherapy as a standard⁷ and emerging data supporting enfortumab vedotin plus pembrolizumab.³⁰ However, the shift toward these potent combinations does not obviate the need for biomarkers; rather, it demands more precise tools to spare patients unnecessary toxicity and identify those unlikely to respond to specific modalities. We propose that the

(C) Annotation accuracy (AUC) for matching reduced-gene-panel SingleR cell-type labels to ground-truth labels from the full 377-gene panel, for CD8⁺ T cells and epithelial cells. AUC values and 95% confidence intervals were calculated using the DeLong method.

(D) Predictive performance (AUC) of the median CD8⁺ T cell-to-cancer minimum distance for treatment response when annotations are derived from downsized gene panels. Dashed red line marks the inflection point. AUC values and 95% confidence intervals were calculated using the DeLong method.

(E) Predictive performance (AUC) of the median CD8⁺ T cell-to-cancer minimum distance for treatment response within randomly selected fields of view (FOVs) of 3 mm, 2 mm, and 1 mm in diameter (50 per sample) using ground-truth annotations. Each point represents the AUC calculated from one randomly sampled FOV set. Boxplots show the median, interquartile range, and 1.5 $\times$ interquartile range whiskers. Dashed red line marks the ground-truth AUC from the full Xenium regions. Statistical significance was assessed using pairwise two-sided Wilcoxon rank-sum tests. ns, $p > 0.05$; *** $p \leq 0.001$; **** $p \leq 0.0001$.

(F–H) Examples of spatial heterogeneity in CD8⁺ T cell-to-cancer distances at the 1-mm FOV scale. (F) Whole-sample view for one example tumor showing epithelial cells (blue), CD8⁺ T cells (red), and other cell types (dark blue); distribution of minimum distances is shown to the right of the sample. Scale bar, 1 mm. (G) Representative 1-mm FOV from the same tumor with denser CD8⁺ T cell infiltration. (H) Another 1-mm FOV with sparser CD8⁺ T cell infiltration from the same sample. Circular FOVs in (G and H) are 1 mm in diameter. Dashed black lines mark the median minimum distances.

architectural constraints identified here, specifically fibroblast-mediated exclusion and physical immune isolation, are fundamental biological barriers that are likely to impact the efficacy of any PD-1/PDL1-based backbone, regardless of the combination partner. For example, the “cold” luminal tumors that responded preferentially to chemotherapy in our control arm represent a distinct biological entity that may require different therapeutic strategies than the “immune-excluded” phenotype. Spatial profiling offers a framework to disentangle these mechanisms, potentially guiding the integration of antibody-drug conjugates, chemotherapy, and ICI in future decision-making.

A major challenge in the field is bridging the gap between high-plex discovery and clinical implementation. Recent perspectives emphasize that the critical bottleneck is identifying the “minimum information” needed to stratify patients and optimizing “low-plex” panels that are economically viable.²⁹ Furthermore, current validation efforts often default to TMAs with restricted sampling areas (~1-mm cores), potentially underpowering spatial analysis. Our study directly addresses these challenges by providing a quantitative framework for “target panel curation.” We demonstrate that 1-mm cores fail to reproduce the spatial signals observed in whole-tissue analysis, leading to high intra-sample variance. To overcome this, we estimated practical design parameters for evaluating the CD8⁺ T cell proximity signature, identifying that annotation panels of approximately 77 genes and tissue sampling diameters of ~3 mm were sufficient to retain most of the predictive signal in this dataset. These thresholds provide the field with an empirically derived roadmap for designing scalable, statistically powered validation cohorts.

In summary, the DUTRENEO trial provides a definitive prospective test of the TIS biomarker, demonstrating that bulk gene signatures are insufficient to guide patient selection in this setting. By combining this clinical result with a high-resolution molecular autopsy, we reveal the architectural constraints of immunotherapy response and provide a rigorous, data-driven roadmap for designing the next generation of spatially informed clinical trials.

Limitations of the study

Several limitations should be considered when interpreting our results. The spatial analyses presented here are associative and exploratory and are not intended to define an immediately deployable clinical biomarker. Rather, they aim to provide a framework for understanding why bulk inflammation-based biomarkers fail and how spatial features may inform future biomarker design. Xenium profiling was performed on pretreatment TURBT specimens, which may not fully capture residual disease biology at the time of systemic therapy, and each assay samples only a portion of the tumor area. In addition, the targeted gene panel, while enriched for immune and epithelial programs, cannot resolve all phenotypic states. Finally, spatial associations were derived from patient-level summaries and do not establish causality. Nevertheless, the persistence of the CD8-tumor distance association among TIS-high tumors supports the added value of spatial organization beyond immune abundance alone.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Francisco X. Real (freal@cnio.es).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- Xenium raw data have been deposited at Zenodo as <https://doi.org/10.5281/zenodo.19693316>. Bulk RNA-seq and WES raw data have been deposited at SRA under GSE328930 and PRJNA1458521 accession numbers, respectively. All data are publicly available as of the date of publication.
- The scripts used to generate the Xenium results have been deposited at Zenodo as <https://doi.org/10.5281/zenodo.19608038> and are publicly available as of the date of publication.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

The DUTRENEO study was supported by an unrestricted educational grant by AstraZeneca; Nanostring provided reagents for TIS score. The work was supported, in part, by grants from Fundación Científica de la Asociación Española Contra el Cáncer (GCB14142293REAL (to F.X.R., D.C., A.F., and N.M.) and PRYGN223005REAL (to F.X.R.) E.P.-P., M.S., and M.E. are funded by the AECC project LABAE20038PORT. M.S. is supported by “la Caixa” Foundation (LCF/BQ/DR22/11950021). E.P.-P. is funded by the Spanish Ministry of Science (RYC2019-026415-I MICIU/AEI/10.13039/501100011033, “El FSE invierte en tu futuro,” and PID2024-159258OB-I00) and Fundación FERO-ASEICA (BFERO2022.6). CNIO is supported by MCIU as a Centro de Excelencia Severo Ochoa (grant SEV-2015-0510). IJC is supported by MCIU as a Centro de Excelencia Severo Ochoa (CEX2023-001258-S, MCIN/AEI/10.13039/501100011033). We thank Leticia Rodríguez, Natalia del Pozo, and the CNIO Histopathology and Genomics Unit for excellent technical support; Jedd Wolchock and Antoni Ribas for discussions; Javier Carmona and Joaquín Mateo for valuable comments on earlier versions of the manuscript; Joakim Lundberg for insightful discussions on spatial biomarker design; and the CRIS Cancer Foundation and APICES for their contributions to the implementation of the study.

AUTHOR CONTRIBUTIONS

Conception and design, E.G., M.S., E.A., D.C., A.F., E.P.-P., N.M., I.D., and F.X.R.; financial support, E.G., D.C., A.F., E.P.-P., N.M., E.P., and F.X.R.; administrative support, F.X.R.; patient recruitment and provision of study materials, E.G., O.R., D.C., J.P., A.F., T.A.-G., M.A.-M., J.B., M.A.C., M.D., I.G., F.G.-R., P.M., A. Prat, and I.D.; collection and assembly of data, E.G., M.S., D.G., E.P., E.A., A.A., J.M.d.V., P.G., J.F.G., D.G., M.M., A. Prat, E.P.-P., N.M., and F.X.R.; data analysis and interpretation, E.G., M.S., D.G., E.A., A.A., J.M.d.V., R.B., A.M.-O., J.F.G., D.G., M.M., J.M.P., N.M., I.D., E.P.-P., and F.X.R.; manuscript writing, E.G., M.S., E.A., E.P.-P., and F.X.R.; final approval of manuscript, E.G., M.S., D.G., E.A., O.R., A.A., D.C., J.P., J.M.d.V., A.F., T.A.-G., R.B., A.M.-O., M.A.-M., J.B., M.A.C., M.D., P.G., I.G., J.F.G., X.G.d.M., D.G., F.G.-R., M.M., P.M., J.M.P., A. Pinto, A. Prat, N.M., I.D., E.P.-P., and F.X.R.; accountable for all aspects of the work, E.G., E.P., and F.X.R.

DECLARATION OF INTERESTS

E.G. has received honoraria for speaker engagements and advisory roles or funding of continuous medical education from Adacap, AMGEN, Angelini, Astellas, AstraZeneca, Bayer, Blueprint, Bristol Myers Squibb, Caris Life Sciences, Celgene, Clovis-Oncology, Eisai, Eusa Pharma, Genetracer, Guardant

Health, HRA-Pharma, IPSEN, ITM-Radiopharma, Janssen, Lexicon, Lilly, Merck KGaA, MSD, Nanostring Technologies, Natera, Novartis, ONCODNA (Biosequence), Palex, Pharmamar, Pierre Fabre, Pfizer, Roche, Sanofi-Genzyme, Servier, Taiho, and Thermo Fisher Scientific.

O.R. has received honoraria for speaker engagements and advisory roles or funding of continuous medical education from Astellas, Bristol Myers Squibb, IPSEN, Pfizer, and MSD. D.C. has received honoraria for consulting or advisory role of CME from Pfizer, Novartis, AstraZeneca, Astellas, Bayer, Bristol Myers Squibb, Eisai, Eusa Pharma, MSD, Merck KG&A, Roche, Gilead, Ipsen, Janssen, Excelsis, and Genentech.

J.P. has received honoraria for speaker engagements and advisory roles or funding of continuous medical education from Astellas, AstraZeneca, Bayer, Bristol Myers Squibb, Eisai, Eusa Pharma, IPSEN, Janssen, Merck KGaA, MSD, Novartis, Pfizer, Roche, and Gilead. J.P. has received research grants from Pfizer, Astellas, and Merck.

A.F. has received honoraria for an advisory council or committee from Janssen, Astellas, Bayer, Merck, and EUSA; grants or funds from AstraZeneca, Roche, and Astellas; travel accommodation fees from Janssen, Bristol Myers Squibb, Merck, and Roche; and TAG funding, honoraria, and nonfinancial or other support from IPSEN, Adacup, Pfizer, Sanofi, Eisai, Lilly, Bayer, Johnson & Johnson, Astellas, Novartis, and Roche.

X.G.d.M. has received honoraria for speaker engagements or advisory roles from Astellas, Bristol Myers Squibb, Deciphera, Eisai, GSK, IPSEN, Merck KGaA, MSD, Pharmamar, Pfizer, Recordati, and Roche. X.G.d.M. has received research grants from AstraZeneca and Incyte. F.G.R. has received grants/research funding from Combat Medical and Roche; has been involved in clinical trials with UroGen, Pharmedica, Astellas, Johnson & Johnson, Roche, Combat Medical, Taris Bio, Janssen, Pfizer, AstraZeneca, Bristol Myers Squibb, Seagen, Fidia, and Cepheid; has received consulting/advisory fees from Pfizer, AstraZeneca, Johnson & Johnson, Nucleix, Taris Bio, Bristol Myers Squibb, Roche, Combat Medical, and Janssen; has received honoraria/speaker fees from Taris Bio, Combat Medical, Nucleix, Palex, Astellas, Bristol Myers Squibb, Merck, Janssen, AstraZeneca, and Pfizer; and has received travel support from Johnson & Johnson, AstraZeneca, Janssen, Ipsen, and Pfizer.

P.M. has received honoraria from advisory boards with Pfizer, Astellas, Bayer, Novartis, Janssen, Merck, and MSD and travel support from Bayer, Merck, and Pfizer.

A. Pinto has received research funding from Pfizer and BMS; compensation for speaker fees or advisory boards from Janssen, Astellas, Bayer, MSD, Roche, and Ipsen; and compensation for travel expenses from Janssen, MSD, Bayer, and Ipsen.

A. Prat reports advisory and consulting fees from Reveal Genomics, AstraZeneca, Daiichi-Sankyo, Roche, Pfizer, Novartis, and Peptomyc and institutional financial interests from AstraZeneca, Novartis, Roche, and Daiichi Sankyo. A. Prat is a stockholder and employee of Reveal Genomics; patents filed include PCT/EP2016/080056, PCT/EP2022/086493, PCT/EP2023/060810, EP23382703, and EP23383369.

I.D. has participated in compensated advisory boards sponsored by Astellas, Bristol Myers Squibb, Immunomedics, MSD, Roche, Bicycle Therapeutics, and Gilead. He has also taken part as an invited speaker in compensated educational activities sponsored by Astellas, Bayer, Bristol Myers Squibb, Jansen, Merck, Pfizer, and Roche. Bayer, AstraZeneca, and Gilead have covered expenses related to travel, accommodation, and registration of medical conferences for I.D. He serves as a member of the steering committee of two advanced bladder cancer trials sponsored by GILEAD and is the leading investigator of an investigator-initiated study funded by the same company. His department has received research grants from AstraZeneca and Roche.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this work, the authors used ChatGPT to reduce the character count. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS**
 - Study participants
- **METHOD DETAILS**
 - Study population and trial design
 - Immunohistochemical assays
 - Whole exome sequencing
 - Bulk RNA-seq and bioinformatics analysis
 - Spatial transcriptomics - Xenium *in situ* gene expression
 - Post-xenium H&E staining
 - Xenium data preprocessing
 - Cell type assignment
 - Cell-to-cell distance and neighborhood analysis
 - Whole-sample proximity analysis
 - Gene panel downsizing for annotation
 - Field-of-view (FOV) downsizing
 - Graphical figures
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
 - Clinical trial analyses
 - Xenium analyses
 - Effect size estimation using Cohen's *d*
 - ROC analyses

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.xcr.2026.102878>.

Received: February 22, 2025

Revised: August 25, 2025

Accepted: May 26, 2026

Published: June 23, 2026

REFERENCES

1. Bray, F., Laversanne, M., Sung, H., Ferlay, J., Siegel, R.L., Soerjomataram, I., and Jemal, A. (2024). Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* 74, 229–263.
2. Stein, J.P., Lieskovsky, G., Cote, R., Groshen, S., Feng, A.C., Boyd, S., Skinner, E., Bochner, B., Thangathurai, D., Mikhail, M., et al. (2001). Radical cystectomy in the treatment of invasive bladder cancer: long-term results in 1,054 patients. *J. Clin. Oncol.* 19, 666–675.
3. Advanced Bladder Cancer ABC Meta-analysis Collaboration (2005). Neoadjuvant chemotherapy in invasive bladder cancer: update of a systematic review and meta-analysis of individual patient data advanced bladder cancer (ABC) meta-analysis collaboration. *Eur. Urol.* 48, 202–206.
4. Moschini, M., Karnes, R.J., Sharma, V., Gandaglia, G., Fossati, N., Dell'Oglio, P., Cucchiar, V., Capogrosso, P., Shariat, S.F., Damiano, R., et al. (2016). Patterns and prognostic significance of clinical recurrences after radical cystectomy for bladder cancer: A 20-year single center experience. *Eur. J. Surg. Oncol.* 42, 735–743.
5. Grossman, H.B., Natale, R.B., Tangen, C.M., Speights, V.O., Vogelzang, N.J., Trump, D.L., deVere White, R.W., Sarosdy, M.F., Wood, D.P., Jr., Rathgavan, D., and Crawford, E.D. (2003). Neoadjuvant chemotherapy plus cystectomy compared with cystectomy alone for locally advanced bladder cancer. *N. Engl. J. Med.* 349, 859–866.
6. Yin, M., Joshi, M., Meijer, R.P., Glantz, M., Holder, S., Harvey, H.A., Kaag, M., Fransen van de Putte, E.E., Horenblas, S., and Drabick, J.J. (2016). Neoadjuvant Chemotherapy for Muscle-Invasive Bladder Cancer: A Systematic Review and Two-Step Meta-Analysis. *Oncologist* 21, 708–715.

7. Powles, T., Catto, J.W.F., Galsky, M.D., Al-Ahmadie, H., Meeks, J.J., Nishiyama, H., Vu, T.Q., Antonuzzo, L., Wiechno, P., Atduiev, V., et al. (2024). Perioperative Durvalumab with Neoadjuvant Chemotherapy in Operable Bladder Cancer. *N. Engl. J. Med.* 391, 1773–1786.
8. Huo, J., Ray-Zack, M.D., Shan, Y., Chamie, K., Boorjian, S.A., Kerr, P., Jana, B., Freedland, S.J., Kamat, A.M., Mehta, H.B., and Williams, S.B. (2019). Discerning Patterns and Quality of Neoadjuvant Chemotherapy Use Among Patients with Muscle-invasive Bladder Cancer. *Eur. Urol. Oncol.* 2, 497–504.
9. Powles, T., Kockx, M., Rodriguez-Vida, A., Duran, I., Crabb, S.J., Van Der Heijden, M.S., Szabados, B., Pous, A.F., Gravis, G., Herranz, U.A., et al. (2019). Clinical efficacy and biomarker analysis of neoadjuvant atezolizumab in operable urothelial carcinoma in the ABACUS trial. *Nat. Med.* 25, 1706–1714.
10. Necchi, A., Raggi, D., Gallina, A., Madison, R., Colecchia, M., Lucianò, R., Montironi, R., Giannatempo, P., Farè, E., Pederzoli, F., et al. (2020). Updated Results of PURE-01 with Preliminary Activity of Neoadjuvant Pembrolizumab in Patients with Muscle-invasive Bladder Carcinoma with Variant Histologies. *Eur. Urol.* 77, 439–446.
11. Gao, J., Navai, N., Alhalabi, O., Siefker-Radtke, A., Campbell, M.T., Tidwell, R.S., Guo, C.C., Kamat, A.M., Matin, S.F., Araujo, J.C., et al. (2020). Neoadjuvant PD-L1 plus CTLA-4 blockade in patients with cisplatin-ineligible operable high-risk urothelial carcinoma. *Nat. Med.* 26, 1845–1851.
12. van Dijk, N., Gil-Jimenez, A., Silina, K., Hendricksen, K., Smit, L.A., de Feijter, J.M., van Montfoort, M.L., van Rooijen, C., Peters, D., Broeks, A., et al. (2020). Preoperative ipilimumab plus nivolumab in locoregionally advanced urothelial cancer: the NABUCCO trial. *Nat. Med.* 26, 1839–1844.
13. Mariathasan, S., Turley, S.J., Nickles, D., Castiglioni, A., Yuen, K., Wang, Y., Kadel, E.E., III, Koepfen, H., Astarita, J.L., Cubas, R., et al. (2018). TGF β attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. *Nature* 554, 544–548.
14. Grasso, C.S., Tsoi, J., Onyshchenko, M., Abril-Rodriguez, G., Ross-Macdonald, P., Wind-Rotolo, M., Champhekar, A., Medina, E., Torrejon, D.Y., Shin, D.S., et al. (2020). Conserved Interferon- γ Signaling Drives Clinical Response to Immune Checkpoint Blockade Therapy in Melanoma. *Cancer Cell* 38, 500–515.e3.
15. Ayers, M., Lunceford, J., Nebozhyn, M., Murphy, E., Loboda, A., Kaufman, D.R., Albright, A., Cheng, J.D., Kang, S.P., Shankaran, V., et al. (2017). IFN- γ -related mRNA profile predicts clinical response to PD-1 blockade. *J. Clin. Invest.* 127, 2930–2940.
16. Danaher, P., Warren, S., Lu, R., Samayoa, J., Sullivan, A., Pekker, I., Wallen, B., Marincola, F.M., and Cesano, A. (2018). Pan-cancer adaptive immune resistance as defined by the Tumor Inflammation Signature (TIS): results from The Cancer Genome Atlas (TCGA). *J. Immunother. Cancer* 6, 63.
17. Damotte, D., Warren, S., Arrondeau, J., Boudou-Rouquette, P., Mansuet-Lupo, A., Biton, J., Ouakrim, H., Alifano, M., Gervais, C., Bellesoeur, A., et al. (2019). The tumor inflammation signature (TIS) is associated with anti-PD-1 treatment benefit in the CERTIM pan-cancer cohort. *J. Transl. Med.* 17, 357.
18. Reijers, I.L.M., Rao, D., Versluis, J.M., Menzies, A.M., Dimitriadis, P., Wouters, M.W., Spillane, A.J., Klop, W.M.C., Broeks, A., Bosch, L.J.W., et al. (2023). IFN- γ signature enables selection of neoadjuvant treatment in patients with stage III melanoma. *J. Exp. Med.* 220, e20221952. <https://doi.org/10.1084/jem.20221952>.
19. Robertson, A.G., Kim, J., Al-Ahmadie, H., Bellmunt, J., Guo, G., Cherniack, A.D., Hinoue, T., Laird, P.W., Hoadley, K.A., Akbani, R., et al. (2017). Comprehensive Molecular Characterization of Muscle-Invasive Bladder Cancer. *Cell* 171, 540–556.e25.
20. Kamoun, A., de Reyniès, A., Allory, Y., Sjöndahl, G., Robertson, A.G., Seiler, R., Hoadley, K.A., Groeneveld, C.S., Al-Ahmadie, H., Choi, W., et al. (2020). A Consensus Molecular Classification of Muscle-invasive Bladder Cancer. *Eur. Urol.* 77, 420–433.
21. Doroshow, D.B., Bhalla, S., Beasley, M.B., Sholl, L.M., Kerr, K.M., Gnjatic, S., Wistuba, I.I., Rimm, D.L., Tsao, M.S., and Hirsch, F.R. (2021). PD-L1 as a biomarker of response to immune-checkpoint inhibitors. *Nat. Rev. Clin. Oncol.* 18, 345–362.
22. Hammerl, D., Martens, J.W.M., Timmermans, M., Smid, M., Trapman-Jansen, A.M., Foekens, R., Isaeva, O.I., Voorwerk, L., Balcioglu, H.E., Wijers, R., et al. (2021). Spatial immunophenotypes predict response to anti-PD1 treatment and capture distinct paths of T cell evasion in triple negative breast cancer. *Nat. Commun.* 12, 5668.
23. Wang, X.Q., Danenberg, E., Huang, C.-S., Egle, D., Callari, M., Bermejo, B., Dugo, M., Zamagni, C., Thill, M., Anton, A., et al. (2023). Spatial predictors of immunotherapy response in triple-negative breast cancer. *Nature* 621, 868–876.
24. Färkkilä, A., Gulhan, D.C., Casado, J., Jacobson, C.A., Nguyen, H., Kochupurakkal, B., Maliga, Z., Yapp, C., Chen, Y.-A., Schapiro, D., et al. (2020). Immunogenomic profiling determines responses to combined PARP and PD-1 inhibition in ovarian cancer. *Nat. Commun.* 11, 1459.
25. Ma, X., Guo, Z., Wei, X., Zhao, G., Han, D., Zhang, T., Chen, X., Cao, F., Dong, J., Zhao, L., et al. (2021). Spatial Distribution and Predictive Significance of Dendritic Cells and Macrophages in Esophageal Cancer Treated With Combined Chemoradiotherapy and PD-1 Blockade. *Front. Immunol.* 12, 786429.
26. Williams, H.L., Frei, A.L., Koessler, T., Berger, M.D., Dawson, H., Michielin, O., and Zlobec, I. (2024). The current landscape of spatial biomarkers for prediction of response to immune checkpoint inhibition. *npj Precis. Oncol.* 8, 178.
27. Ghiringhelli, F., Bibeau, F., Greillier, L., Fumet, J.-D., Ilie, A., Monville, F., Laugé, C., Catteau, A., Boquet, I., Majdi, A., et al. (2023). Immunoscore immune checkpoint using spatial quantitative analysis of CD8 and PD-L1 markers is predictive of the efficacy of anti-PD1/PD-L1 immunotherapy in non-small cell lung cancer. *EBioMedicine* 92, 104633.
28. Gong, D., Arbesfeld-Qiu, J.M., Perrault, E., Bae, J.W., and Hwang, W.L. (2024). Spatial oncology: Translating contextual biology to the clinic. *Cancer Cell* 42, 1653–1675.
29. Engblom, C., and Lundberg, J. (2025). Putting cancer immunotherapy into spatial context in the clinic. *Nat. Biotechnol.* 43, 471–476.
30. Vulsteke, C., Adra, N., Danchaivijit, P., Sabadash, M., Rodriguez-Vida, A., Zhang, Z., Atduiev, V., Göger, Y.E., Rausch, S., Kang, S.-H., et al. (2026). Perioperative Enfortumab Vedotin and Pembrolizumab in Bladder Cancer. *N. Engl. J. Med.* 394, 1257–1269. <https://doi.org/10.1056/NEJMoa2511674>.
31. Galsky, M.D., Hahn, N.M., Rosenberg, J., Sonpavde, G., Hutson, T., Oh, W.K., Dreicer, R., Vogelzang, N., Sternberg, C.N., Bajorin, D.F., and Bellmunt, J. (2011). Treatment of patients with metastatic urothelial cancer “unfit” for Cisplatin-based chemotherapy. *J. Clin. Oncol.* 29, 2432–2438.
32. Eckstein, M., Cimadamore, A., Hartmann, A., Lopez-Beltran, A., Cheng, L., Scarpelli, M., Montironi, R., and Gevaert, T. (2019). PD-L1 assessment in urothelial carcinoma: a practical approach. *Ann. Transl. Med.* 7, 690.
33. DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* 43, 491–498.
34. Benjamin, D., Sato, T., Cibulskis, K., Getz, G., Stewart, C., and Lichtenstein, L. (2019). Calling Somatic SNVs and Indels with Mutect2. Preprint at bioRxiv. <https://doi.org/10.1101/861054>.
35. Kandath, C., Gao, J., Qwangmsk, Mattioni, M., Struck, A., Boursin, Y., Penson, A., and Chavan, S. (2018). mskcc/vcf2maf: vcf2maf v1.6.16. Zenodo. <https://doi.org/10.5281/ZENODO.593251>.
36. Mayakonda, A., Lin, D.-C., Assenov, Y., Plass, C., and Koeffler, H.P. (2018). Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res.* 28, 1747–1756.

37. Cerami, E., Gao, J., Dogrusoz, U., Gross, B.E., Sumer, S.O., Aksoy, B.A., Jacobsen, A., Byrne, C.J., Heuer, M.L., Larsson, E., et al. (2012). The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* 2, 401–404.
38. Smith, T., Heger, A., and Sudbery, I. (2017). UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res.* 27, 491–499.
39. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550.
40. Aran, D., Looney, A.P., Liu, L., Wu, E., Fong, V., Hsu, A., Chak, S., Naikawadi, R.P., Wolters, P.J., Abate, A.R., et al. (2019). Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat. Immunol.* 20, 163–172.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
PD-L1 rabbit monoclonal antibody (clone E1L3N)	Cell Signaling Technology	Cat#13684
Biological samples		
Human FFPE pre-treatment TURBT tumor specimens from DUTRENEO trial participants	This paper; participating hospitals	DUTRENEO clinical trial; EudraCT: 2017-002246-68
Critical commercial assays		
Covaris truXTRAC FFPE total NA kit	Covaris	N/A
QuantSeq 3' mRNA-Seq Library Prep Kit (FWD) for Illumina	Lexogen	Cat#015
Xenium Decoding Consumables (1 run, 2 slides)	10x Genomics	Cat#1000487
Xenium Decoding Reagents (1 run, 2 slides)	10x Genomics	Cat#1000461
Xenium Slides & Sample Prep Reagents (2 slides, 2 rxns)	10x Genomics	Cat#1000460
Xenium Human Multi-tissue and Cancer Panel (377 genes)	10x Genomics	Cat#1000629
Deposited data		
DUTRENEO: bulk biomarker-guided therapy selection and spatially resolved profiling of immunotherapy response in muscle-invasive bladder cancer. Xenium Data.	This paper	https://doi.org/10.5281/zenodo.19606472
DUTRENEO: bulk biomarker-guided therapy selection and spatially resolved profiling of immunotherapy response in muscle-invasive bladder cancer. Bulk RNA-Seq Data.	This paper	GSE328930
DUTRENEO: bulk biomarker-guided therapy selection and spatially resolved profiling of immunotherapy response in muscle-invasive bladder cancer. WES Data.	This paper	PRJNA1458521
Software and algorithms		
Varca pipeline	CNIO Bioinformatics Unit	https://github.com/cnio-bu/varca
NIS-Elements	Nikon	RRID:SCR_014329
Genome Analysis Toolkit (GATK) Best Practices workflow	Broad Institute	https://gatk.broadinstitute.org/
BWA aligner	SourceForge project	https://bio-bwa.sourceforge.net/
Mutect2	Broad Institute	https://gatk.broadinstitute.org/hc/en-us/articles/360037593851-Mutect2
vcf2maf	Memorial Sloan Kettering Cancer Center	https://github.com/mskcc/vcf2maf
maftools (R package)	Bioconductor	https://bioconductor.org/packages/maftools
STAR aligner	GitHub	https://github.com/alexdobin/STAR
UMI-tools	UMI-tools project	https://umi-tools.readthedocs.io/
featureCounts	Subread package	https://subread.sourceforge.net/
DESeq2 (R package)	Bioconductor	https://bioconductor.org/packages/DESeq2
Seurat	Satija Lab	v5; https://satijalab.org/seurat/
Xenium Analyzer software	10x Genomics	v1.5.0.3
SingleR (R package)	Bioconductor	https://bioconductor.org/packages/SingleR
dbscan (R package)	CRAN	https://cran.r-project.org/package=dbscan
ClusterR (R package)	CRAN	https://cran.r-project.org/package=ClusterR

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
glmnet (R package)	CRAN	https://cran.r-project.org/package=glmnet
pROC (R package)	CRAN	https://cran.r-project.org/package=pROC
ggplot2 (R package)	tidyverse/CRAN	https://ggplot2.tidyverse.org/
ComplexHeatmap (R package)	Bioconductor	https://bioconductor.org/packages/ComplexHeatmap
ggraph (R package)	CRAN	https://cran.r-project.org/package=ggraph
igraph (R package)	CRAN	https://igraph.org/r/
BioRender	BioRender	https://www.biorender.com/
Inkscape	Inkscape Project	https://inkscape.org/
Deposited Code	This paper	https://doi.org/10.5281/zenodo.19608038
Other		
NanoString nCounter platform	NanoString Technologies	N/A
Xenium Analyzer instrument	10x Genomics	N/A

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Study participants

The multicenter DUTRENEO phase 2 trial included patients with cT2–T4a and/or N1+ urothelial bladder cancer, residual disease after TURBT, no distant metastasis, ECOG 0–1, and cisplatin eligibility according to Galsky's criteria. Pre-treatment FFPE TURBT specimens were used for molecular stratification and spatial analyses. Among 101 screened patients, 73 received treatment. The treated cohort had a median age of 66 years and was predominantly male. All participants provided written informed consent. The trial was approved by the Ethics Committee of Hospital Universitario Ramón y Cajal (EudraCT 2017-002246-68).

METHOD DETAILS

Study population and trial design

Figure 1A shows the study design. DUTRENEO is a multicenter, open-label, study aimed to assess the value of molecular patient stratification to improve treatment selection in patients with muscle-invasive bladder cancer. In addition, the study aims to compare the antitumor activity of tremelimumab and durvalumab in comparison with standard platinum-based NAC in patients with high inflammatory scores. Patients were eligible if they had urothelial muscle-invasive bladder cancer defined by cT2–T4a and/or lymph node involvement (N1+) staging with residual disease after TURBT, no evidence of distant metastasis and fit to receive a cisplatin-based treatment by Galsky's criteria.³¹ Additional inclusion criteria included Eastern Cooperative Oncology Group performance status of 0 or 1, absence of previous chemotherapy for muscle-invasive disease, adequate blood counts and liver function - including a calculated creatinine clearance of >40 mL/min - and HbA1c ≤ 8%. Main exclusion criteria were histology of pure adenocarcinoma, pure squamous cell carcinoma, or predominant small cell carcinoma or sarcomatoid features and evidence of distant metastasis. Current or prior use of immunosuppressive medication for an immune-related disease was also considered an exclusion criterion. The DUTRENEO trial employed a Simon two-stage design with a prespecified pCR threshold for expansion. As this threshold was not reached, the trial did not proceed to the second stage.

Based on an estimated positivity rate of the TIS in half of the patients and to detect an improvement in pCR rate from 30% with standard cisplatin-based chemotherapy up to 50% with durvalumab + tremelimumab (α : 0.1; β : 0.1). It was considered that 22 patients in each arm would be needed in a first phase. If at least 8 pCR were observed among the first 22 patients recruited in the investigational ICI arm, 24 additional patients would be subsequently recruited in the ICI arm. A dropout rate of 10% was assumed, therefore the total sample size needed to fulfill the objectives of the trial were 99 patients. A prospective and centralized TIS based on 18-gene IFN- γ signaling-related expression (see results) was obtained on TURBT samples to classify patients. One-hundred-one patients were screened for study eligibility. Of them, 28 resulted in screening failures [concurrent metastases (5); abnormal blood chemistries (2); abnormal QT interval (3); insufficient tumor sample to assess TIS (2); intercurrent infection (1); later refusal to participate (3); others (12, including 8 patients for which a slot was not available in the corresponding treatment arm)].

To establish a threshold to categorize tumors as TIS+ vs. TIS-, we used TURBT samples ($n = 65$) from a retrospective series of patients with MIBC who received cisplatin-based NAC in two hospitals participating in the DUTRENEO trial. The NanoString nCounter platform was used and TIS scoring was carried out following an established algorithm. A threshold that discriminated the samples with the lowest 30% scores was established to reach a balance between case enrichment and recruitment rate feasibility.

Subjects whose tumors were classified as cold were offered standard platinum-based NAC with gemcitabine 1250 mg/m² on days 1 and 8, and cisplatin 70 mg/m² on day 1, every 3 weeks for a total of three cycles or dose dense-MVAC with methotrexate 30 mg/m² on day 1, vinblastine 3 mg/m² on day 2, doxorubicin 30 mg/m² on day 2, and cisplatin 70mg/m² on day 2 with associated granulocyte colony-stimulating factor (G-CSF) from day 3 to day 9, every 2 weeks for a total of four cycles, according to investigator's choice. Patients whose tumors were classified as hot were randomized to receive either standard cisplatin-based NAC as mentioned above, according to investigator's choice, or the combination of durvalumab (1500 mg) and tremelimumab (75 mg) intravenously every 4 weeks for a total of three cycles before surgery. A cystectomy was requested to be performed 4 to 8 weeks after the end of neo-adjuvant treatment. The primary endpoint was pCR, defined as no evidence of residual disease based on independent pathological review of the surgical specimen of durvalumab plus tremelimumab arm in comparison with standard NAC after cystectomy in the intention-to-treat population (pCR = ypT0N0). The presence of carcinoma *in situ* (CIS) in the cystectomy specimen was considered residual disease and not compatible with pCR (ypT0N0). Secondary endpoints included disease-free survival, overall survival, and safety profile. Follow-up after surgery was performed following ESMO guidelines rules and site protocols.

The sponsor of the study was the CRIS Cancer Foundation (<https://criscancer.org/en/>), thanks to an unrestricted independent grant from Astra Zeneca. The study was conducted in accordance with Good Clinical Practice Guidelines and the Declaration of Helsinki and in accordance with the applicable regulatory requirements, in particular the ICH tripartite harmonized standards for good clinical practice 1996 and the Royal Decree of Clinical Trials 1090/2015, by which clinical trials with medicinal products are regulated in Spain, which fully incorporates the Regulation (EU) No 536/2014 of the European Parliament and of the Council of 16 April 2014.

All patients signed the informed consent form to be enrolled in the trial, approved by the Ethics Committee of the Hospital Universitario Ramón y Cajal on September 03, 2018 (EudraCT number: 2017-002246-68).

Immunohistochemical assays

Blocks from pre-treatment FFPE tumor samples were obtained from the participating hospitals. Sections (4 micron) were deparaffinized and stained in a centralized laboratory with primary antibodies detecting PD-L1 (rabbit monoclonal antibody E1L3N, Cell Signaling Technologies, #13684). Immunostaining scoring was performed blindly by an expert pathologist (AA), following standard protocols used for patient selection in the clinical setting.³²

Whole exome sequencing

H-E sections from pre-treatment FFPE TURBT samples were used to select areas with a high tumor cellularity (>60%). The same tumor regions were selected for macrodissection for WES and RNA-Seq. Data from tumor samples were analyzed from FASTQ files using Varca pipeline in tumor-only mode (<https://github.com/cnio-bu/varca>). This pipeline implements GATK Best-practices workflow.³³ Briefly, reads were aligned to hg38 reference genome using BWA aligner and somatic mutations were called using Mutect2.³⁴ VCF files were filtered using an hg38 panel of normal samples for both exomes and whole genomes generated from the 1000 Genomes Project and annotated and transformed to Mutation Annotation Format (MAF) with vcf2maf.³⁵ Downstream analyses were carried out using R package maftools.³⁶ Due to the high degree of noise derived from FFPE tumour-only samples and the relatively small sample size in each group, we filtered the dataset using somatic mutations included in cBioportal³⁷ (bladder cancer datasets) to focus on the top mutated genes.

Bulk RNA-seq and bioinformatics analysis

RNA was extracted using the Covaris truXTRAC FFPE total NA kit; RNA quality was assessed using bioanalyzer. Total RNA was converted into cDNA sequencing libraries with the "QuantSeq 3' mRNA-Seq Library Prep Kit (FWD) for Illumina" (Lexogen, Cat.No. 015). Briefly, library generation was initiated by reverse transcription with oligodT priming, followed by a random-primed second strand synthesis. A UMI Second Strand Synthesis module was used, whereby random primers featured a 6 nucleotide long Unique Molecular Identifier 5' tag to be addressed at data analysis. Sequencing was performed using an Illumina NextSeq 550 instrument. Reads were aligned to GRCh38 genome using STAR aligner. UMIs were extracted and demultiplexed using UMI-tools software³⁸ to improve transcript quantification and correct for PCR artifacts related to poor-quality samples. Transcript quantification was carried out with featureCounts and differential expression analysis was done using DESeq2 package.³⁹ Molecular subtype calls were based on the consensus molecular classification of MIBC.²⁰

Spatial transcriptomics - Xenium *in situ* gene expression

The Xenium workflow began by sectioning 5µm FFPE tissue sections onto the Xenium slides, according to the "Xenium FFPE Tissue Preparation Guide" (CG000578 Rev C, 10X Genomics). Briefly, the sections were cut, floated in an RNase-free water bath at 42°, and carefully placed onto the capture area of a Xenium slide (PN- 3000941). After sectioning, the slides were incubated at 42° per 3 h and kept in a sealed bag with desiccators at room temperature overnight.

The next day, the slides were incubated at 60° for 2 h, and were processed following the "Xenium *In Situ* for FFPE- Deparaffinization and Decrosslinking" protocol (CG000580 Rev D, 10X Genomics). After deparaffinization, the Xenium slides were assembled into Xenium cassettes, which allow for the incubation of slides on the Xenium Thermocycler Adapter inside a Thermocycler machine with a closed lid for optimal temperature control, and a decrosslinking step was performed. The slides were then processed using the "Xenium *In Situ* Gene Expression" user guide (CG000582 Rev D, 10X Genomics). After decrosslinking PBS-T washes, the samples

were incubated at 50°C overnight for probe hybridization approximately 19 h with the Xenium Human Multi-tissue and Cancer Panel, which targets 377 human markers for bladder, breast, colon, heart, kidney, liver, lung, pancreas, etc. This was followed by a series of washing steps, a post-hybridization wash, a ligation step, an amplification step and finally the slides were treated with an autofluorescence quencher and a nuclei staining. At the end of the third day of sample preparation, the slides were loaded into the Xenium Analyzer. The first step in the instrument consists in a sample scan, where images of the fluorescent nuclei in each section are given, and these images allow the user to select and determine the regions to be included in the analysis. Each run in the Xenium Analyzer lasted 4 days approximately. After that, the instrument was emptied of consumables and the slides carefully removed. PBS-T was added to the slides and a post-run H&E staining was performed. We designed experiments to place between 3 and 4 samples per Xenium slide, with a total of 6 slides that correspond to 3 Xenium Runs in the Instrument, we were able to process and analyze a total of 20 samples.

Post-xenium H&E staining

Slides were processed following post-xenium analyzer H&E protocol (CG000613 Rev A, 10X Genomics) A quencher removal solution was prepared immediately before use by dissolving sodium hydrosulfite in Milli-Q water to a final concentration of 10 mM. The slides were gently immersed in the quencher removal solution using a slide mailer and incubated for 10 min at room temperature. This was followed by three washes in Milli-Q water.

The sections were then stained with Gill's hematoxylin for 20 min, followed by three washes in Milli-Q water. Bluing solution was applied for 1 min, followed by a Milli-Q water wash. The slides were subsequently dehydrated through graded ethanol solutions (70% and 95%), counterstained with alcoholic eosin for 2 min, further dehydrated to xylene, and finally mounted with coverslips.

Brightfield histology images were taken using 10X and 20× objectives (Plan APO) on a Nikon Eclipse Ti2, images were stitched together using NIS-Elements software (Nikon) and exported as tiff files.

Xenium data preprocessing

Gene expression matrices of all 20 Xenium samples generated by Xenium Analyzer (v1.5.0.3) were imported into Seurat (v5). Cells were filtered out if they had less than 4 genes in each sample, thus minimizing false positive downstream cell type assignment as a result of potentially imperfect cellular segmentation. Finally, the gene expression in each sample was normalized with the SCTransform function from Seurat.

Cell type assignment

Cells were annotated using the R package SingleR,⁴⁰ which performs unbiased identification of cell types in single cell transcriptomics data based on built-in reference transcriptomes of pure cell types. In our case, we used the “BlueprintEncodeData” reference dataset which included a number of different cell types. To minimize false positive assignments, we restricted the “main labels” of the reference to cell types that may populate bladder tumors: namely, epithelial cells, fibroblasts, smooth muscle, pericytes, neurons, monocytes, CD4 T cells, CD8 T cells, NK cells, B cells, macrophages, endothelial cells, dendritic cells, neutrophils, and eosinophils.

We ran the singleR function on the “counts” slot with the “fine labels” as the reference cell types. These included the aforementioned main labels as well as additional subtypes: namely, Tregs, CD4 Tcm, CD4 Tem, CD8 Tcm, CD8 Tem, plasma cells, naive B cells, memory B cells, class-switched memory B cells, macrophages M1, macrophages M2, and mv endothelial cells. We assessed the quality of assignment with the “plotScoreHeatmap” function from the SingleR package to inspect the confidence of cell clusters when assigned to a given (sub)type. Despite the high confidence, the limited number of genes in the Xenium panel led us to adopt a more robust approach for downstream analysis by grouping all CD4 subtypes (except for Tregs) under CD4 T cells and those of CD8 subtypes under CD8 T cells. Moreover, we grouped all B cell subtypes (except plasma cells) under B cells, and “mv endothelial cells” under “endothelial cells”. Similarly, pericytes and smooth muscles were joined together, whereas neutrophils and eosinophils were grouped under “granulocytes”. Accordingly, the final list of cell types which we could confidently represent in our Xenium samples are: epithelial cells, endothelial cells, fibroblasts, smooth muscle, neurons, granulocytes, monocytes, CD4 T cells, Tregs, CD8 T cells, NK cells, B cells, plasma cells, macrophages, macrophages M1, macrophages M2, and DCs. Cells that had a low confidence score by SingleR were labeled “unassigned”. Finally, we ran differential expression analysis for the final labels on the SCT-normalized, merged Xenium data of all 20 samples, using the FindAllMarkers with the default “wilcox” test. This provided complementary evidence for the confidence of the assigned cell types, where some of the classical or known markers could be clearly associated with the corresponding cell type (e.g., EPCAM/KRTs for epithelial cells, CD3D/CD8A for CD8⁺ T cells, MS4A1 for B cells, PECAM1 for endothelial cells, etc.).

Cell-to-cell distance and neighborhood analysis

To calculate the median nearest-neighbor distance of each cell type to epithelial cells in each of the Xenium samples, we used the frNN function from the “dbscan” R package. We first established a 200-μm radius around each epithelial cell and calculated the distance of the nearest-neighboring cell type to each epithelial cell within that radius. Then, we took the median nearest-neighbor distance of each cell type to all epithelial cells within a sample. As for the neighborhood analysis, we counted the number of cells of any given type within a 25 μm radius of each cell (as justified in Figure S7C), yielding a neighborhood count matrix for each sample (Figure 3A). Next, we used MiniBatchKmeans function from the R package “ClusterR” (batch_size = 500, max_iters = 500) to get

the total within-cluster sum of squares (wcscs) to determine an optimum number of clusters for each sample's neighborhood matrix using the elbow method (Figure S8). We performed k-means clustering on the neighborhood matrix with the kmeans function in base R, based on the selected number of clusters "k" (i.e., communities) for each sample.

To establish if a community is "rich" (enriched), "neutral", or "poor" (depleted) in any given cell type, we first created Seurat objects (specifying min.feature = 1) for each of the neighborhood count matrices and normalized them with SCTransform. Then, we set the clusters (communities) as the identity in these objects, and ran FindAllMarkers function from Seurat (test.use = "wilcox", logfc.threshold = 0, min.pct = 0, and min.diff.pct = 0) on the normalized neighborhood cell type assay. To systematically represent the enrichment status of each cell type within at least one community across all samples for downstream analysis, we adopted a lenient classification approach. Specifically, avg_log2FC values greater than, equal to, or less than zero were categorized as "rich", "baseline", or "poor", respectively.

Whole-sample proximity analysis

We quantified the predictive value of spatial proximity between immune cell populations and epithelial (cancer) cells across the entire Xenium dataset. For each sample, we extracted cell centroids from the Xenium segmentation output and linked them to SingleR-based fine-label annotations (BlueprintEncode reference; see cell type assignment). Using the dbscanfrNN function (eps = 200 μ m) in R, we identified all neighboring cells within a 200 μ m radius of each index cell. Each neighbor pair was annotated with the reference cell type and the neighbor cell type. We then computed, for every unique triplet (reference, neighbor, and cell ids), the minimum Euclidean distance between the reference cell and any neighbor of the given type. We restricted this table to Epithelial cells as reference ids and each immune population as neighbor ids, then calculated the median minimum distance per sample for each reference-neighbor combination. Predictive performance for clinical response (CR/PR vs. NR) was quantified using the area under the ROC curve (AUC; pROC package). For stability analysis, we performed a leave-one-out (LOO) outer loop. In each LOO fold, we fit a binomial LASSO (glmnet, alpha = 1) on the training samples with 5-fold cross-validation to select λ (lambda.min), refit at that λ on the training set, and recorded features with non-zero coefficients (excluding the intercept). Selection frequency was defined as the percentage of LOO folds in which each feature was selected.

Gene panel downsizing for annotation

We evaluated whether the predictive biomarker could be retained when the number of genes used for cell type annotation was systematically reduced. Reduced panels were generated from a precomputed differential expression analysis table containing, for each broad category (CD8⁺ T-cells, Epithelial cells, Others), the adjusted *p*-value (p_val_adj), log₂ fold-change (avg_log2FC), and detection rate differences (pct.1, pct.2). Genes were ranked within each category by a composite score:

$$\text{score} = -\log_{10}(\text{padj} + 10^{-10}) \times |\text{avg_log2FC}| \times (\text{pct.1} - \text{pct.2}).$$

The DEA contained a maximum of 87, 67, and 312 unique genes for CD8, Epithelial, and Others, respectively (total = 466 containing shared DE genes between the 3 groups; total unique = 377). We targeted a descending sequence of total panel sizes {377, 362, 347, ...17} in 15-gene steps. For each target size *T*, we allocated category quotas in proportion to their full-panel representation:

$$n_{\text{CD8}} = T \times 87/377; n_{\text{Epithelial}} = T \times 67/377; n_{\text{Others}} = T - n_{\text{CD8}} - n_{\text{Epithelial}}$$

Top-ranked genes were selected per category, excluding any already selected from previous categories (CD8 excluded from Epi; CD8+Epi excluded from Others). The union of category selections defined the panel for that *T*. Because of cross-category overlap and exclusion rules, the realized panel size could differ slightly from *T*; panels are denoted in figures by their realized size. For each downsized panel, SingleR (fine labels) was rerun on the Xenium counts matrix filtered to the selected genes (the reference remained unchanged). Fine labels were collapsed post hoc to CD8⁺ T-cells, Epithelial cells, and Others. Annotation accuracy was computed as the AUC for matching downsized labels to the 377-gene ground truth. Using these downsized annotations, we recomputed the spatial biomarker (median CD8→Epithelial minimum distance) and its AUC versus clinical response.

Regarding the control analysis, we generated 50 random 17-gene panels by sampling without replacement from the set of 45 genes discarded between N377 and N332. Each random panel was processed identically to the downsized panels to yield a null distribution of AUCs.

Field-of-view (FOV) downsizing

To simulate reduced tissue area, we identified candidate circular FOVs of radii 1500 μ m (3 mm diameter), 1000 μ m (2 mm), and 500 μ m (1 mm) within each sample using centroid coordinates and segmentation polygons from the Xenium outputs. For each radius, 2000 random circle centers were tested, and a candidate was retained if: 1) total cells $\geq$ 200; 2) epithelial cells $\geq$ 30; 3) cell area coverage $\geq$ 40% of circle area. From the candidate set, 50 FOVs were randomly selected per radius per sample. All cells within each FOV were identified by centroid inclusion. Using ground truth annotations, we recalculated the CD8⁺ T cell-to-epithelial median minimum distance for each FOV. The AUC for treatment response prediction was computed separately for each of the 50 FOVs. For qualitative illustration, we visualized the segmentation boundaries overlaid with cell type annotations (Seurat's ImageDimPlot) for one full sample and two 1 mm FOVs, and plotted the distributions of CD8→Epithelial minimum distances for each region.

Graphical figures

Individual panels with statistical data were generated with the R packages “ggplot2”, “pheatmap”, “ComplexHeatmap”. Circular network plots were generated with the R packages “ggraph” and “igraph”. Figures 1A, 3A, and 3B were generated using BioRender, and all individual panels were put together using Inkscape.

QUANTIFICATION AND STATISTICAL ANALYSIS

Clinical trial analyses

The primary clinical endpoint was pathological complete response (pCR; ypT0N0 in the randomized TIS-high arm). Secondary endpoints included pathological downstaging, progression-free survival (PFS), and overall survival (OS). For all translational analyses, patients were classified as responders or non-responders based on the presence or absence of pathological downstaging, regardless of whether they achieved pCR. Categorical variables, including pCR, downstaging, and baseline clinical characteristics, were compared using Chi-Square test or Fisher’s exact test (when expected cell counts were <5), accordingly, with effect sizes estimated as odds ratios and 95% confidence intervals (CI). Continuous baseline variables (e.g., age) were compared using two-sided, two-sample t-tests when normally distributed or Wilcoxon rank-sum tests otherwise. Survival curves for PFS and OS were generated separately, and the log rank test was used to compare the three treatment arms. Associations between clinical covariates and survival outcomes were estimated using univariable and multivariable Cox proportional hazards models, performed separately for PFS and OS. Multivariable models included treatment arm, age, gender, T stage, nodal status, and smoking history as prespecified covariates; hazard ratios with 95% CI are reported for each. All tests were two-sided with $\alpha = 0.05$.

Xenium analyses

All Xenium-derived spatial measurements were aggregated to patient-level summary values to ensure independence of observations. Group differences in continuous spatial quantities, including median nearest-neighbor distances, ligand-receptor niche co-expression scores, and community-level marker contributions, were evaluated using two-sided Wilcoxon rank-sum tests unless otherwise indicated. For analyses shown in Figures 2C, 4A, and 4B, the corresponding raw *p*-values are reported. For the checkpoint compartmentalization analysis shown in Figure 5, comparisons were performed within each response group by testing marker-expression contributions in cell-type-rich versus the corresponding cell-type-poor community contexts using two-sided Wilcoxon rank-sum tests. *P*-values were adjusted for multiple testing using the Benjamini–Hochberg procedure; only contrasts with FDR <0.05 were considered significant and displayed in the circular networks.

Effect size estimation using Cohen’s *d*

Cohen’s *d* was used as a standardized effect size to quantify the magnitude and direction of contrasts in two distinct analytical contexts.

1. Responder vs. non-responder contrasts. These effect sizes allowed direct comparison across heterogeneous spatial metrics and complemented non-parametric hypothesis tests by highlighting biologically meaningful differences that may not be fully captured by *p*-values alone.
2. Cell-rich vs. cell-poor community contrasts within each response group. In the checkpoint compartmentalization analyses (Figure 5), Cohen’s *d* quantified the extent to which marker expression from a specific cell type was biased toward cell-type-rich versus cell-type-poor spatial communities, separately in responders and non-responders.

For each marker–cell-type–community pair, we compared (1) the proportion of marker expression contributed by cell-rich communities, and (2) the proportion contributed by cell-poor communities across patients *within* a response group. Cohen’s *d* captured both the direction of spatial enrichment (rich > poor or poor > rich) and its magnitude. Statistical significance for these within-group community contrasts was assessed using paired comparisons with Benjamini–Hochberg FDR correction, and only significant contrasts (FDR <0.05) were displayed in circular network visualizations. In these figures, edge width encodes |Cohen’s *d*| and edge color distinguishes enrichment toward cell-rich (steel blue) or cell-poor (magenta) communities. These visualizations are descriptive summaries of spatial bias, not outputs of correlation tests or network inference models.

ROC analyses

The discriminative ability of spatial biomarkers, particularly the median CD8⁺ T cell-to-cancer minimum distance, was evaluated using receiver operating characteristic (ROC) curves and the AUC. AUC is a non-parametric measure equivalent to the Mann-Whitney U statistic and quantifies the probability that a randomly chosen responder has a more favourable biomarker value than a non-responder. AUC values and 95% CI were calculated using the DeLong method. These ROC-based analyses were applied consistently across the gene-panel downsizing and area-downsizing experiments, in which AUCs and their DeLong confidence intervals were used to quantify how annotation depth and sampling area affect biomarker performance.