nature medicine

Article

https://doi.org/10.1038/s41591-026-04710-1

Neoadjuvant botensilimab with or without balstilimab in pMMR and dMMR colon cancer: the phase 2 UNICORN trial

i d: 6 r h 2026

pt d: 14 pt r 2026

Published online: 02 November 2026

h or pd t s

A list of authors and their affiliations appears at the end of the paper

Preoperative immune checkpoint inhibitors (ICIs) are active in patients with deficient mismatch repair (dMMR) colon cancer (CC), whereas proficient mismatch repair (pMMR) CC has been consistently considered resistant to first-generation ICIs. In the phase 2 UNICORN trial, we evaluated short-course preoperative botensilimab (BOT; multifunctional Fc-enhanced anti-CTLA-4 monoclonal antibody) alone or in combination with the anti-PD-1 monoclonal antibody balstilimab (BOT + BAL) in 56 patients with resectable, locally advanced CC. Twenty-eight patients with pMMR CC were sequentially enrolled in BOT ($n = 14$) and BOT + BAL ($n = 14$) cohorts. Twenty-eight patients with dMMR CC followed the same sequential allocation strategy. The primary endpoint of centrally assessed pathological major response (pMR) rate (≤10% of residual viable tumor) was met in all cohorts, except pMMR BOT. Specifically, the pMR rate was 0% (95% confidence interval (CI): 0–23.2%) in pMMR BOT, 36% (95% CI: 12.8–64.9%) in pMMR BOT + BAL, 36% (95% CI: 12.8–64.9%) in dMMR BOT and 100% (95% CI: 76.8–100%) in dMMR BOT + BAL. Secondary endpoints included safety, quality of life, describing treatment-induced transcriptional dynamics, peripheral blood mononuclear cell dynamics, radiomics and radiogenomics features, fecal microbiota composition and circulating tumor DNA (ctDNA) dynamics. Eight (29%) and 15 (54%) patients treated with BOT and BOT + BAL, respectively, experienced immune-related adverse events, most commonly flu-like syndrome and diarrhea. Single-cell spatial transcriptomic analysis of matched baseline and post-treatment samples highlighted high baseline PD-L1 stromal expression and low post-treatment regulatory T cell signatures in pMMR responders, whereas dMMR responders displayed high baseline *FCGR2A* and *FCGR3A* expression on intratumoral macrophages. Clearance of ctDNA before surgery correlated with pathological response and lack of recurrence at data cutoff. Preoperative BOT + BAL elicits pathological responses in patients with pMMR or dMMR CC, supporting the combination as a promising strategy to expand ICI benefits to the early-stage setting. ClinicalTrials.gov identifier: NCT05845450.

e-mail: filippo.pietrantonio@istitutotumori.mi.it

Colorectal cancer (CRC) is a leading cause of cancer-related deaths[1,2]. Despite maximal curative-intent treatment, up to 35% of patients with localized CRC will experience disease relapse, leading to relevant mortality[3].

Neoadjuvant/perioperative strategies are established in the therapeutic algorithm of patients with non-metastatic gastroesophageal and rectal cancers, and growing evidence is being collected in CC[4–6]. The FOxTROT trial highlighted the benefit of 6-week neoadjuvant FOLFOX (5-fluorouracil/leucovorin and oxaliplatin) followed by surgery and adjuvant chemotherapy in patients with radiologically staged rT3–4 CC[7]. Interestingly, the study reported an inverse correlation between treatment activity and recurrence risk, which reached 0% in patients obtaining a pathological complete response (pCR). Expectedly, chemotherapy activity was minimal in deficient mismatch repair (dMMR) CC, in line with the relative chemoresistance of the disease[8,9]; in these patients, an ICI-based preoperative approach has consistently demonstrated excellent activity and disease-free survival (DFS) across several single-arm trials[10–12]. Accordingly, 6 months of the anti-PD-1 antibody dostarlimab is globally recommended as non-operative management (NOM) for locally advanced dMMR rectal cancer[13], and neoadjuvant/perioperative ICI-based strategies are being investigated in registrational phase 3 trials in localized dMMR CC[14,15]. By contrast, proficient mismatch repair (pMMR) CC is typically regarded as resistant to first-generation ICIs[16,17].

Botensilimab (BOT) is a multifunctional, Fc-enhanced anti-CTLA-4 monoclonal antibody designed to overcome ICI resistance in solid tumors by leveraging both the adaptive and innate immunity through co-engagement of activating Fc gamma receptors (FcγRs), particularly FcγRIIa and FcγRIIIa[18]. In addition to complete CTLA-4–ligand interaction blockade, BOT promotes FcγR-dependent activation of intratumoral myeloid and dendritic cells and improves antigen-presenting cell (APC)–T cell immune synapse formation, enhancing T cell priming and memory while depleting intratumoral regulatory T (T_{reg}) cells[18,19]. Furthermore, BOT's modified Fc-enhanced IgG1 region is engineered to limit complement-mediated toxicity associated with conventional anti-CTLA-4 antibodies[20].

Consistently, BOT and balstilimab (BAL) (anti-PD-1 antibody) have demonstrated preliminary activity in heavily pre-treated pMMR metastatic colorectal cancer (mCRC), with an overall response rate (ORR) of 20% in patients without liver metastases and, consequently, lower levels of immune suppression[21].

However, how these biological mechanisms may contribute to pathological regression in non-metastatic pMMR CC remains unresolved[19,22,23]. Notably, the early-stage setting offers a distinct immunobiological context for ICI activity, where intratumoral T cell enrichment, low tumor burden, absence of visceral disease and consequent cancer-mediated immune suppression may enhance immunotherapy benefit even in immune-cold tumors[10,24]. Nonetheless, the activity of first-generation ICIs in localized pMMR CC is limited, with only modest signals reported with dual checkpoint blockade[10,25,26], underscoring the need for mechanistically distinct strategies.

We designed UNICORN (NCT05845450), an open-label, non-randomized, window-of-opportunity, umbrella/platform phase 2 trial testing short-course preoperative treatments in patients with molecularly selected, upfront resectable CRC. UNICORN is an ongoing and expanding clinical trial enrolling patients selected for the presence of prespecified molecular profiles to detect signals of activity of targeted agents or immunotherapeutic drugs administered as short-course preoperative strategy (Extended Data Fig. 1). Here we report activity, safety and preliminary efficacy results of cohorts 4–7 with preoperative BOT or BOT + BAL in pMMR and dMMR CC.

Results

Patient population

The cohort-specific design and an overview of screening and enrollment flows are reported in Fig. 1a and Fig. 1b, respectively. From June 2023 to July 2024, 56 patients with radiologically staged rT3–4 N0/N+ M0, resectable pMMR or dMMR colon adenocarcinoma were enrolled in cohorts 4 (pMMR BOT), 5 (pMMR BOT + BAL), 6 (dMMR BOT) and 7 (dMMR BOT + BAL) of UNICORN. Twenty-eight patients with pMMR CC were sequentially enrolled in BOT ($n = 14$) or BOT + BAL ($n = 14$) cohorts; 28 patients with dMMR CC were enrolled in BOT ($n = 14$) and, subsequently, BOT + BAL ($n = 14$) cohorts.

The pMMR BOT cohort enrolled patients from 28 June 2023 to 1 January 2024; the pMMR BOT + BAL cohort enrolled patients from 10 January to 2 May 2024; the dMMR BOT cohort enrolled patients from 24 July 2023 to 14 February 2024; and the dMMR BOT + BAL cohort enrolled patients from 29 February to 24 July 2024.

The data cutoff date for the present analyses was 9 February 2026.

Baseline patient and disease characteristics are reported in Table 1. In brief, patient demographics and tumor features were in line with literature data, with most patients in the dMMR cohorts of female sex and with right-sided primary tumor location[27]. No patient with rectal cancer was enrolled in these cohorts.

Extended Data Table 1 reports baseline radiological staging as centrally assessed by a senior radiologist with expertise in the field of colorectal imaging (G. Brown).

Feasibility and treatment safety

All 56 patients started treatment and 54 (96%) completed it. Two patients did not receive the second BAL administration due to immune-related adverse events (irAEs): grade 3 hyperthyroidism and grade 2 flu-like syndrome. Surgery was performed in all patients and was timely (delay ≤14 days from day 35 due to treatment-related adverse events) in 98% of cases. All patients were, therefore, evaluable for the primary and secondary endpoints.

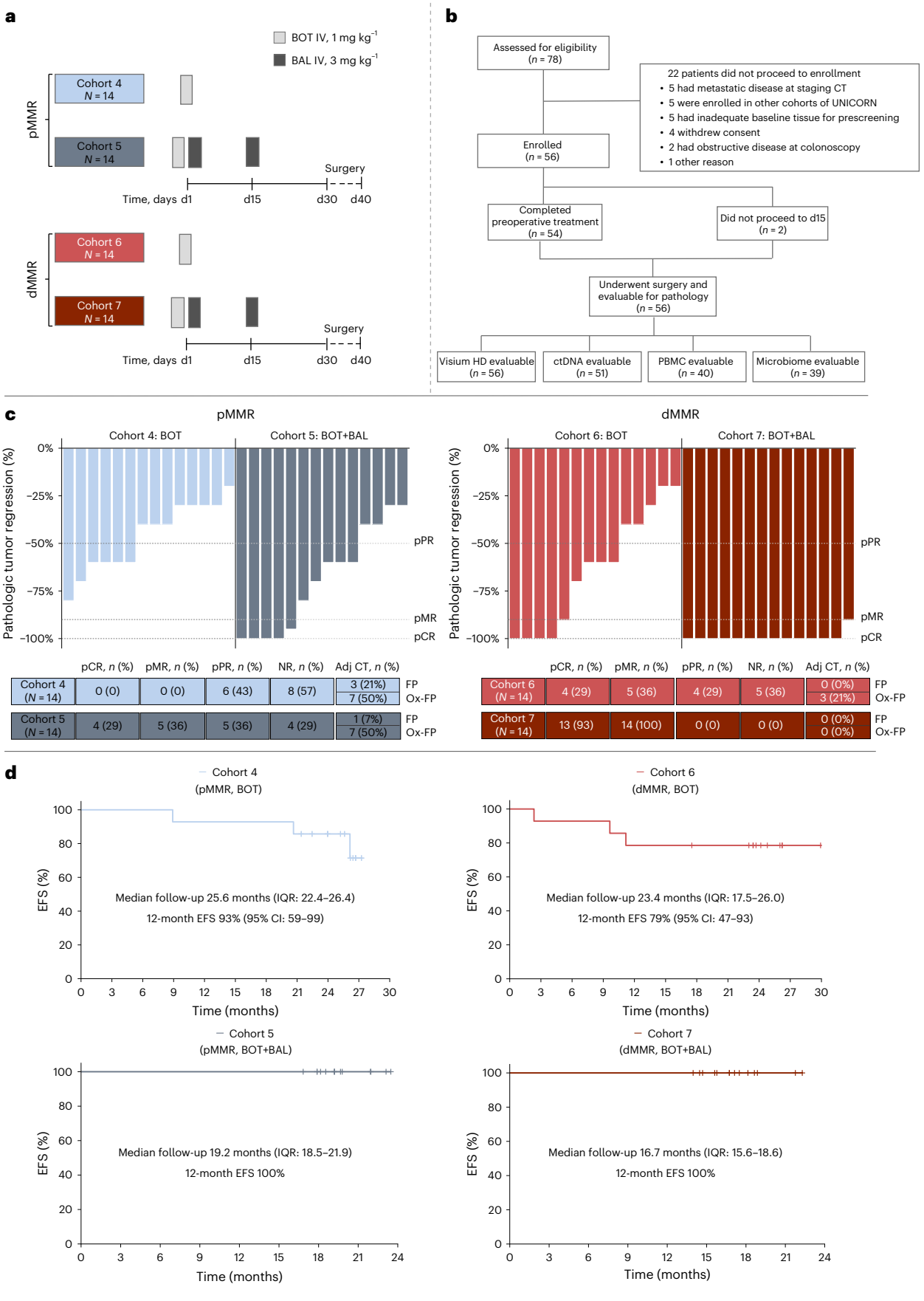
As outlined in Table 2, adverse events occurred in 30 (54%) patients and were irAEs in 23 (41%) patients. One dMMR BOT + BAL patient had G3 hyperthyroidism, One dMMR BOT + BAL patient had G3 hyperthyroidism, leading to surgery on day 50, 10 days beyond the protocol-specified surgical window and 15 days after day 35, thus not meeting the definition of timely surgery. Three patients had surgical complications unrelated to study treatment. No adverse event led to death or unresolved sequelae.

Treatment activity

The rate of pathological major response (pMR) (defined as ≤10% of residual viable tumor) was the primary endpoint of the study. As shown in Fig. 1c, in pMMR BOT, the pMR rate was 0% (Clopper–Pearson 95% confidence interval (CI): 0–23.2%). Six patients (43%)

Fig. 1 | Study design, main clinical outcomes and time to surgery. a, Study design. The lower horizontal line represents study timeline; treatment start is depicted on day 1 by a light gray-colored square (BOT) alone or associated with a dark gray-colored square (BAL). The dotted timeline between days 30 and 40 represents the protocol-specified window for surgery. **b**, CONSORT patient disposition diagram providing an overview of patients undergoing eligibility assessment, enrolled in the trial, receiving study treatment and undergoing surgery and study-related translational analyses. **c**, Waterfall plots reporting centrally assessed pathological tumor regression in each patient across treatment cohorts. Bars represent individual patients and are grouped by treatment cohort. Three horizontal dotted gray lines are distributed across the plots corresponding to the threshold for definition of pathological response, as protocol specified: the first line at the 50% threshold marks pPR; the second line represents the threshold for the definition of pMR (that is, 90% regression); and pCR is indicated by the third dotted line corresponding to 100% tumor regression. **d**, Kaplan–Meier estimates of EFS across treatment cohorts. CT, computed tomography; FP, fluoropyrimidine; IV, intravenous; Ox, oxaliplatin.

https://doi.org/10.1038/s41591-026-04710-1



https://doi.org/10.1038/s41591-026-04710-1

Table 1 | Baseline patient and disease characteristics in each cohort

	pMMR BOT *N*=14 (%)	pMMR BOT+BAL *N*=14 (%)	dMMR BOT *N*=14 (%)	dMMR BOT+BAL *N*=14 (%)
Age at enrollment, years				
Median (IQR)	69 (56–75)	65 (57–70)	63 (54–72)	66 (52–71)
Sex, *n* (%)				
Male	6 (43)	6 (43)	7 (50)	4 (29)
Female	8 (57)	8 (57)	7 (50)	10 (71)
Ethnicity, *n* (%)				
Caucasian	14 (100)	14 (100)	14 (100)	13 (93)
Latin American	0 (0)	0 (0)	0 (0)	1 (7)
ECOG performance status, *n* (%)				
0	12 (86)	13 (93)	13 (93)	14 (100)
1	2 (14)	1 (7)	1 (7)	0 (0)
Primary tumor sidedness, *n* (%)				
Right colon	6 (43)	8 (57)	11 (79)	12 (86)
Left colon	8 (57)	6 (43)	3 (21)	2 (14)
Lynch syndrome, *n* (%)				
Yes	0 (0)	0 (0)	2 (14)	1 (7)
No	14 (100)	14 (100)	12 (86)	13 (93)

had a pathological partial response (pPR, defined as >10% and ≤50% of residual tumor).

In pMMR BOT + BAL, the pMR rate was 36% (95% CI: 12.8–64.9%), with pCR (no residual tumor) observed in four (29%) patients. Five (36%) patients achieved a pPR.

In dMMR BOT, the pMR rate was 36% (95% CI: 12.8–64.9%), with pCR in four (29%) patients. Four (29%) patients achieved a pPR.

Finally, in dMMR BOT + BAL, the pMR rate was 100% (95% CI: 76.8–100%), with pCR in 13 (93%) patients.

The primary endpoint was met in all cohorts except for cohort 4 (pMMR BOT).

Longitudinal assessment of quality of life during treatment

Compliance with health-related quality of life assessments is reported in Supplementary Table 1. Two pMMR BOT + BAL and two dMMR BOT + BAL patients did not file any questionnaire. Two additional dMMR BOT + BAL patients did not provide the end-of-treatment (EOT) assessment.

In pMMR BOT + BAL, European Organization for Research and Treatment of Cancer Core Quality of Life Questionnaire (QLQ-C30) scores declined compared to baseline (77.8 points, s.d.: 13.9) on day 15 (64.4 points, s.d.: 23.6) ($P = 0.049$) and at EOT (61.1 points, s.d.: 26.7) ($P = 0.01$) (Supplementary Table 2); consistently, QLQ-C30 declines were mostly observed in patients with pMMR CC (38.1%) and in those receiving BOT + BAL (45.8%) (Extended Data Fig. 2).

Mean changes across selected QLQ-C30 and Quality of Life Questionnaire-Colorectal 29 (QLQ-CR29) domains are provided in Supplementary Figs. 3–6.

To address the missing EOT timepoints in two patients in the dMMR BOT + BAL cohort, we exploratorily performed a sensitivity analysis (last observation carried forward) in BOT + BAL-treated patients using day 15 as reference: at EOT, the score was 62.8 ± 23.2 versus 71.5 ± 19.2 at baseline (mean change of −8.7 points; Wilcoxon signed-rank $P = 0.070$).

Post-treatment radiological and pathological stagings and postoperative management

Post-treatment radiological and pathological annotations are reported in Table 3.

In pMMR BOT, 10 (71%) patients had post-treatment rT3/T4 tumors (yrT3/T4); post-treatment radiological nodal positivity (yrN+) was reported in four (29%) patients. Ten (71%) patients had post-treatment pathological T3/T4 tumors (ypT3/T4); six (43%) patients had post-treatment pathologically proven nodal disease (ypN+).

In pMMR BOT + BAL, five (36%) patients had yrT3/T4 tumors; two (14%) patients were yrN+. Five (36%) patients had ypT3/T4 tumors; four (29%) patients had ypN+.

In dMMR BOT, six (43%) patients had yrT3/T4 tumors; five (36%) patients were yrN+. Seven (50%) patients had ypT3/T4 tumors; four (29%) patients had ypN+.

Finally, in dMMR BOT + BAL, four (29%) patients had yrT3/T4 tumors; four (29%) patients were yrN+. One (7%) patient had a ypT3/T4 tumor; no patient had ypN+.

As shown in Fig. 1c, in pMMR BOT, 10 (71%) patients received adjuvant chemotherapy: three (21%) patients received fluoropyrimidine monotherapy, and seven (50%) patients received oxaliplatin-based doublets. Median adjuvant chemotherapy duration was 154 days (interquartile range (IQR): 78–165); three of 10 (30%) patients received a 3-month course of adjuvant chemotherapy.

In pMMR BOT + BAL, eight (57%) patients received adjuvant chemotherapy: one (7%) patient received fluoropyrimidine monotherapy, and seven (50%) patients received oxaliplatin-based doublets. Median adjuvant chemotherapy duration was 129 days (IQR: 87–156); five of eight (63%) patients received a 3-month course of adjuvant chemotherapy.

In dMMR BOT, three (21%) patients received oxaliplatin-based adjuvant chemotherapy. Median duration of adjuvant chemotherapy was 84 days (IQR: 75–138).

In dMMR BOT + BAL, no patient received adjuvant chemotherapy.

Treatment efficacy

At data cutoff (9 February 2026), median follow-up was 25.6 months (IQR: 22.4–26.4) for pMMR BOT; 19.2 months (IQR: 18.5–21.9) for pMMR BOT + BAL; 23.4 months (IQR: 17.5–26.0) for dMMR BOT; and 16.7 months (IQR: 15.6–18.6) for dMMR BOT + BAL.

Kaplan–Meier curves for event-free survival (EFS) in each cohort are shown in Fig. 1d. The 12-month EFS rate was 93% (95% CI: 59–99%)

https://doi.org/10.1038/s41591-026-04710-1

Table 2 | Adverse events in the overall population and according to treatment regimen

	Overall *N*=56 (%)	BOT *N*=28 (%)	BOT+BAL *N*=28 (%)
Any adverse event			
Any grade	30 (54)	12 (43)	18 (64)
G3–4	6 (11)	4 (14)	2 (7)
irAEs			
Any grade	23 (41)	8 (29)	15 (54)
G3–4	1 (2)	0 (0)	1 (4)
Flu-like syndrome			
Any grade	7 (13)	1 (4)	6 (21)
G3–4	0 (0)	0 (0)	0 (0)
Diarrhea			
Any grade	4 (7)	2 (7)	2 (7)
G3–4	0 (0)	0 (0)	0 (0)
Fatigue			
Any grade	3 (5)	3 (11)	0 (0)
G3–4	0 (0)	0 (0)	0 (0)
Pruritus			
Any grade	4 (7)	1 (4)	3 (11)
G3–4	0 (0)	0 (0)	0 (0)
Rash			
Any grade	3 (5)	0 (0)	3 (11)
G3–4	0 (0)	0 (0)	0 (0)
Hyperthyroidism			
Any grade	2 (4)	0 (0)	2 (7)
G3–4	1 (2)	0 (0)	1 (4)
Hypothyroidism			
Any grade	1 (2)	0 (0)	1 (4)
G3–4	0 (0)	0 (0)	0 (0)
Oral mucositis			
Any grade	1 (2)	1 (4)	0 (0)
G3–4	0 (0)	0 (0)	0 (0)
Myalgia			
Any grade	1 (2)	0 (0)	1 (4)
G3–4	0 (0)	0 (0)	0 (0)
Nausea			
Any grade	1 (2)	0 (0)	1 (4)
G3–4	0 (0)	0 (0)	0 (0)
GOT/GPT increase			
Any grade	1 (2)	0 (0)	1 (4)
G3–4	0 (0)	0 (0)	0 (0)
Serum creatinine increase			
Any grade	1 (2)	0 (0)	1 (4)
G3–4	0 (0)	0 (0)	0 (0)
Surgical complications			
Any grade	3 (5)	2 (7)	1 (4)
G3–4	3 (5)	2 (7)	1 (4)

G, grade; GOT, glutamic-oxoloacetic transaminase; GPT, glutamic-pyruvic transaminase.

https://doi.org/10.1038/s41591-026-04710-1

Table 3 | Post-treatment radiological and pathological correlates

	pMMR BOT *N*=14 (%)	pMMR BOT+BAL *N*=14 (%)	dMMR BOT *N*=14 (%)	dMMR BOT+BAL *N*=14 (%)
yrT				
yrT0	0 (0%)	2 (14%)	2 (14%)	5 (36%)
yrT1	0 (0%)	0 (0%)	0 (0%)	1 (7%)
yrT2	4 (29%)	7 (50%)	6 (43%)	4 (29%)
yrT3	9 (64%)	5 (36%)	5 (36%)	3 (21%)
yrT4	1 (7%)	0 (0%)	1 (7%)	1 (7%)
yrN				
yrN0	10 (71%)	12 (86%)	9 (64%)	10 (71%)
yrN+	4 (29%)	2 (14%)	5 (36%)	4 (29%)
R0 resection	14 (100%)	14 (100%)	14 (100%)	14 (100%)
ypT				
ypT0	0 (0%)	4 (29%)	4 (29%)	13 (93%)
ypT1	1 (7%)	1 (7%)	0 (0%)	0 (0%)
ypT2	3 (21%)	4 (29%)	3 (21%)	0 (0%)
ypT3	9 (64%)	5 (36%)	6 (43%)	1 (7%)
ypT4	1 (7%)	0 (0%)	1 (7%)	0 (0%)
ypN				
ypN0	8 (57%)	10 (71%)	10 (71%)	14 (100%)
ypN1	5 (36%)	3 (21%)	2 (14%)	0 (0%)
ypN2	1 (7%)	1 (7%)	2 (14%)	0 (0%)
ypTN				
ypT0N0	0 (0%)	4 (29%)	4 (29%)	13 (93%)
ypT1–2N0	0 (0%)	4 (29%)	2 (14%)	0 (0%)
ypT3–4N0	8 (57%)	2 (14%)	4 (29%)	1 (7%)
ypT(any)N1–2	6 (43%)	4 (29%)	4 (29%)	0 (0%)

in pMMR BOT, 79% (95% CI: 47–93%) in dMMR BOT and 100% in pMMR and dMMR BOT + BAL.

Six EFS events occurred: three in pMMR BOT and three in dMMR BOT. We exploratorily assessed EFS according to response in BOT cohorts (Extended Data Fig. 3): two and four events occurred in patients achieving a pPR and a non-response (NR), respectively. One dMMR BOT patient died of progressive disease.

Aims of the translational analyses in UNICORN

The window-of-opportunity design of UNICORN provided a unique framework to interrogate the biological correlates of short-course preoperative ICIs in treatment-naive patients with localized CC, enabling paired assessment of baseline and post-treatment biological samples. Within this framework, we performed signal-seeking analyses aimed at describing baseline disease and patient features as well as treatment-associated tumor and host remodeling to help generate hypotheses on potential correlates of treatment activity. With due limitations coming from the small sample size, findings were examined pooled together and separately in pMMR and dMMR tumors, given the distinct immunobiology and differential sensitivity to ICIs. These analyses were intended to support a biological interpretation of the observed clinical and pathological activities rather than establish any definitive mechanistic conclusions or validate predictive biomarkers and should, therefore, be regarded as purely hypothesis generating.

Single-cell spatial transcriptomics on matched pre-treatment and post-treatment tissue samples

We generated single-cell-resolved spatial transcriptomic maps using 10x Genomics Visium HD from matched pre-treatment biopsies and surgical specimens from all 56 patients. Representative hematoxylin and eosin (H&E) images and corresponding cell type maps from a responder and a non-responder are shown in Fig. 2a and Fig. 2b, respectively. Across samples, we identified 18 epithelial, stromal and immune cell states, supported by the expression of canonical lineage markers (Fig. 2c and Extended Data Fig. 4a). The cohort overview integrated cellular composition with treatment, MMR status, consensus molecular subtype (CMS) and colorectal intrinsic subtype (CRIS) classifications and tumor regression grade (TRG) (Fig. 2c); the distributions of CMS and CRIS assignments at baseline and after treatment are summarized in Extended Data Fig. 4c,e.

Surgical specimens displayed different cellular compositions according to pathological response, MMR status and treatment: both pMMR and dMMR responders displayed lower proportions of tumor and mitochondria-high tumor cells than non-responders, accompanied by differences in immune and stromal populations (Fig. 2d and Extended Data Fig. 4d).

In dMMR tumors, T cell and B cell fractions increased significantly from baseline to surgery, with more pronounced increases among responders than non-responders (Fig. 2e). Consistent with the proposed FcγR-dependent mechanism of action of BOT, dMMR responders also showed higher baseline expression of the activating FcγR genes *FCGR2A* and *FCGR3A* within the myeloid population, potentially indicating a pre-existing myeloid context permissive to BOT activity; this pattern was not observed in pMMR tumors (Extended Data Fig. 4l). By contrast, response-associated differences in the T_{reg} gene set score were not significant at either baseline or surgery in dMMR tumors (Fig. 2f).

In pMMR tumors, neither T cell nor B cell fractions changed significantly across treatment (Fig. 2e). Instead, response was associated with

a qualitative change in the immune state: compared to non-responders, pMMR responders had numerically higher T_{reg} scores at baseline but significantly lower scores at surgery (Fig. 2f). This inversion pattern was also evident in the pooled analysis (Extended Data Fig. 4f).

We next examined whether these distinct immune dynamics were accompanied by differences in spatial organization. Each segmented cell was assigned to the tumor, adjacent stroma (defined as stroma within approximately 100 µm of the tumor boundary) or distant stroma (Fig. 2g). Computationally defined tumor masks closely corresponded to tumor regions identified on H&E sections (Extended Data Fig. 4n).

In pooled baseline analyses, responders showed higher inflammatory myeloid cell abundance and numerically higher macrophage abundance in the adjacent stroma ($P = 0.0538$), whereas plasma B cells were enriched in non-responders (Extended Data Fig. 4g). These baseline differences were not significant when evaluated separately within either MMR subgroup. Both pMMR and dMMR responders showed lower abundance of macrophages in the adjacent stroma at surgery (Fig. 2g).

Differential expression analysis of sample-by-compartment pseudobulk profiles, including MMR status as a covariate, identified response-associated immune regulatory genes predominantly within the stromal compartments (Fig. 2h).

In pMMR tumors, baseline *CD274* (encoding PD-L1) expression was significantly higher in responders than in non-responders in both the adjacent and distant stroma but not within the tumor compartment, indicating that the response-associated PD-L1 signal was spatially restricted to the stromal compartments (Fig. 2h).

In dMMR tumors, baseline *CD274* expression was not associated with response in any spatial compartment (Extended Data Fig. 4h).

Spatial relationships between lymphoid and myeloid populations also differed according to response (Fig. 2i).

In pMMR tumors, T cell–B cell distance decreased across treatment in responders but increased in non-responders, whereas changes in T cell–macrophage distance did not differ significantly.

In dMMR tumors, neither T cell–B cell nor T cell–macrophage distances showed significant response-associated dynamics. Pooled analyses similarly showed significant reductions in both T cell–B cell and T cell–macrophage distances among responders (Extended Data Fig. 4i).

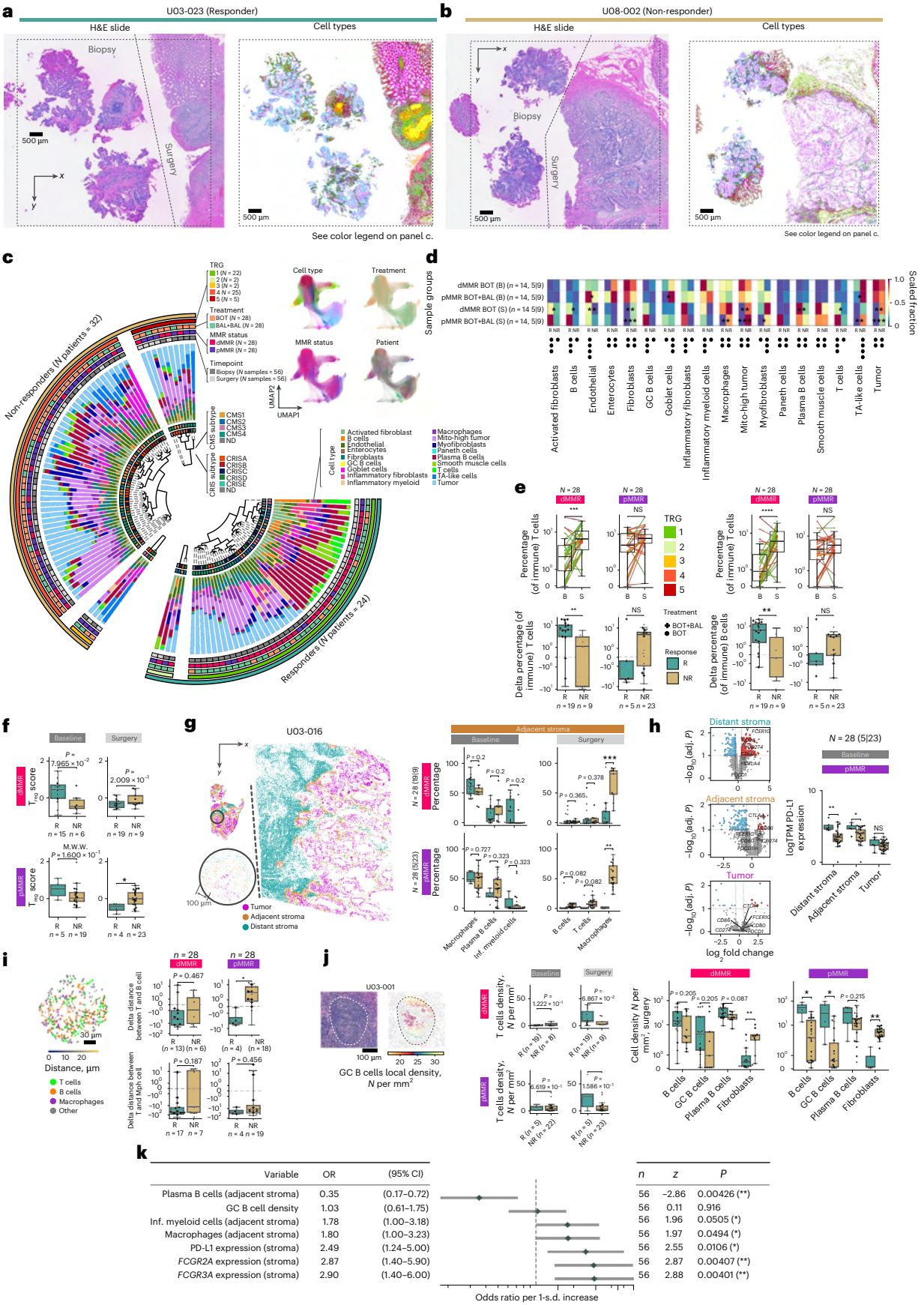
Local density analysis provided complementary information on the spatial accumulation of immune populations (Fig. 2j).

At surgery, pMMR responders had higher B cell and germinal center B cell densities and lower fibroblast density than non-responders, whereas plasma B cell density was numerically, but not significantly, higher. In dMMR tumors, B cell, germinal center B cell and plasma B cell densities were directionally higher in responders but did not differ significantly, whereas fibroblast density was significantly lower. In pooled longitudinal analyses, responders showed an inversion in germinal center B cell and T cell densities, with lower densities at baseline but higher densities at surgery; B cell and plasma B cell densities were also higher and fibroblast density was lower in responders at surgery (Extended Data Fig. 4j). Germinal center B-cell-rich foci co-localized with morphologically defined tertiary lymphoid structures on H&E sections, although the baseline prevalence of germinal center B-cell-positive foci did not differ significantly according to response (Extended Data Fig. 4k). Cell contact analysis further showed stronger tumor–T cell co-occurrence in responders at surgery, consistent with greater T cell access to residual tumor regions (Extended Data Fig. 4m).

Finally, pooled exploratory univariable logistic regression identified baseline adjacent stromal macrophage and inflammatory myeloid cell abundance as positive candidate biomarkers of response, whereas adjacent stromal plasma B cell abundance was negatively associated

Fig. 2 | Single-cell spatial transcriptomics on matched pre-treatment and post-treatment tissue samples. a,**b**, Representative H&E-stained sections and corresponding segmented cell maps colored by assigned cell type. Each Visium HD capture area contains a pre-treatment biopsy and a post-treatment resection section. Scale bars, 500 µm. The same spatial transcriptomic and image analysis workflow was applied independently to all 56 patient pairs; **a** and **b** show one representative responder pair and one representative non-responder pair, respectively. The displayed sections were analyzed once and were not used as technical replicates. **c**, Circular cohort overview of $n = 112$ samples from 56 patients, separated into responders ($n = 24$; TRG 1–2) and non-responders ($n = 32$; TRG 3–5). Concentric annotations show TRG, treatment, MMR status, timepoint, CMS, CRIS and cell type composition; dendrograms show unsupervised clustering within response groups. UMAP embeddings of the latent space are colored by cell type, treatment, MMR status and patient. **d**, Heatmap of scaled cell type fractions in the dMMR BOT and pMMR BOT + BAL groups at biopsy and surgery, comparing responders and non-responders. Asterisks mark significant responder-versus-non-responder differences within a sample group; the dot matrix summarizes the number and direction of groups in which each cell type is enriched. **e**, Top, paired biopsy-to-surgery changes in T cell (left) and B cell (right) fractions among immune cells, stratified by MMR status; lines are colored by TRG. Bottom, change in cell fraction (surgery minus biopsy) by response and MMR status; symbols indicate treatment. **f**, T_{reg} gene set scores in responders and non-responders at baseline and surgery, stratified by MMR status. **g**, Representative compartment assignment. Cells were assigned to tumor, adjacent stroma (approximately 100 µm from the tumor boundary) or distant stroma. Right, selected adjacent stromal cell type fractions by response, timepoint and MMR status. **h**, Baseline differential expression analysis between responders and non-responders using sample-by-compartment pseudobulk profiles with MMR status as a covariate. Volcano plots show $\log_2$ fold change and Benjamini–Hochberg-adjusted *P* value for distant stroma, adjacent stroma and tumor; selected immune regulatory genes are labeled. Right, baseline *CD274* (PD-L1) expression in pMMR samples by compartment and response. **i**, Representative cell pair distance map and change in T cell–B cell and T cell–macrophage distances from biopsy to surgery, stratified by MMR status and response. Negative values indicate reduced separation after treatment. **j**, Representative H&E section and local GC B cell density map. Box plots compare local cell densities by response, timepoint and MMR status; only selected comparisons are shown. **k**, Univariable logistic regression analyses of selected baseline spatial candidate biomarkers. Each model included one continuous predictor and pathological response as the binary outcome; $n = 56$ biologically independent patients for each model. All continuous predictors were standardized, and odds ratios, therefore, represent the change in the odds of response per 1-s.d. increase in the predictor. Diamonds denote maximum-likelihood odds ratio estimates, and horizontal lines denote 95% Wald CIs; the dashed vertical line denotes an odds ratio of 1. Exact two-sided Wald *P* values and *z*-statistics are shown. Models were exploratory and were not adjusted for multiple comparisons. For all quantitative analyses, *N* denotes biologically independent patients. Fifty-six patients contributed one pre-treatment biopsy section and one post-treatment resection section (112 sections in total). Each point represents one patient-level value derived from the indicated section; paired analyses include only patients with measurements at both timepoints. Cell-level, bin-level, neighborhood-level and compartment-level measurements were summarized within each patient section before statistical testing, and no technical replicate was treated as an independent observation. Box plots show the median as the center line, the 25th and 75th percentiles as the box limits and the most extreme observations within 1.5× the IQR as the whiskers. All individual patient values, including values beyond the whiskers, are shown; minima and maxima, therefore, correspond to the lowest and highest displayed points. Independent responder–non-responder comparisons used two-sided Mann–Whitney *U*-tests. Paired biopsy–surgery comparisons used two-sided Wilcoxon signed-rank tests. Benjamini–Hochberg adjustment was applied only where explicitly stated. Exact *n*, test statistics, effect estimates, 95% CIs and exact *P* values are reported for each comparison partially in the panel and fully in the Source Data. adj. *P*, adjusted *P* value; B, biopsy; GC, germinal center; Inf., infiltrating; M.W.W., Mann–Whitney–Wilcoxon; Mph, macrophage; NR, non-responder (TRG 3–5); NS, not significant; OR, odds ratio; R, responder (TRG 1–2); S, surgery; TA, transit-amplifying; TPM, transcripts per million.

https://doi.org/10.1038/s41591-026-04710-1



https://doi.org/10.1038/s41591-026-04710-1

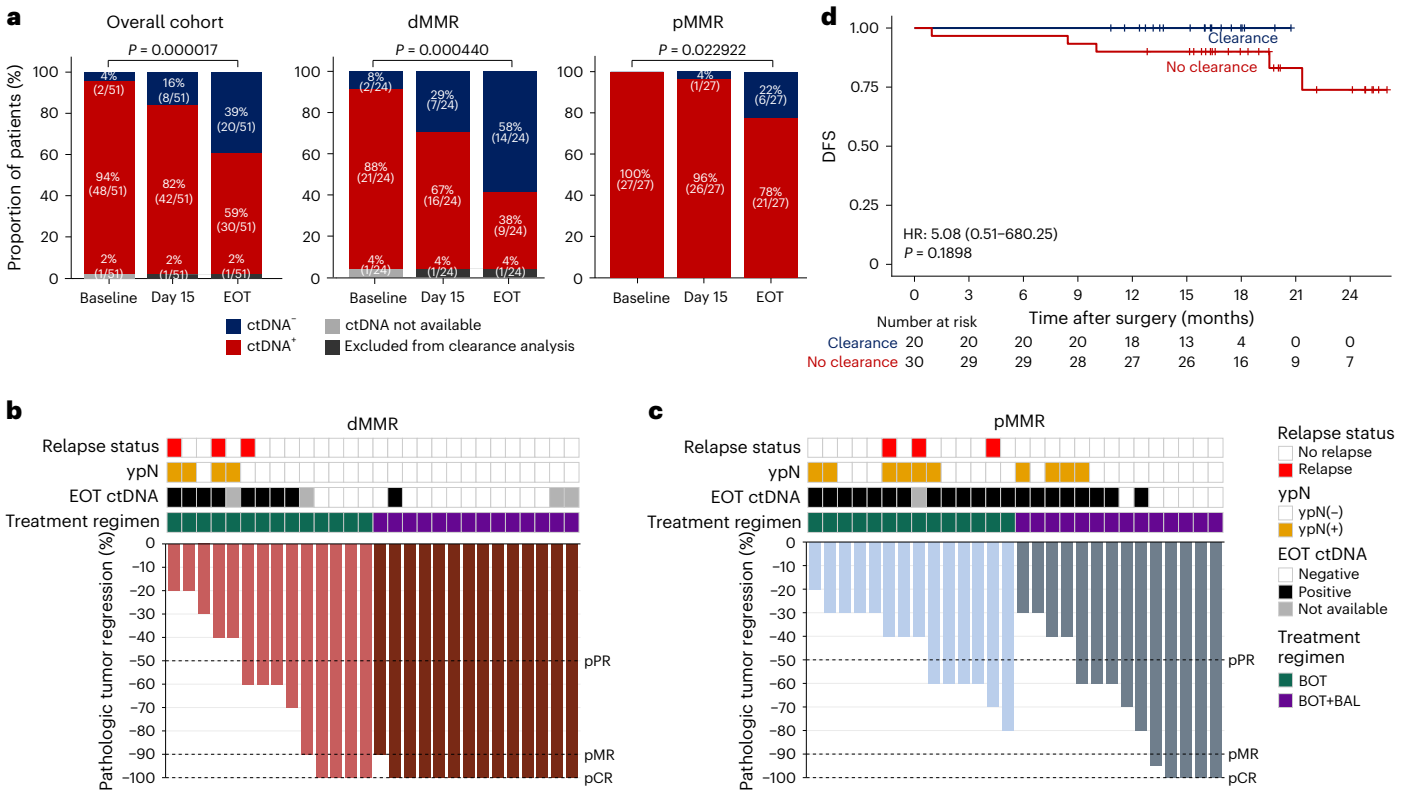


Fig. 3 | Longitudinal ctDNA dynamics across cohorts and association with clinical/pathological outcomes. a, Proportion of patients with ctDNA data during preoperative treatment. Stacked bar plots show ctDNA$^+$ (red) and ctDNA$^-$ (blue) at baseline, day 15 and EOT in the evaluable cohort (left) and according to molecular status (dMMR and pMMR). Percentages and absolute patient numbers are shown within bars. Statistical comparisons between baseline and EOT were performed using a two-sided Fisher's exact test (overall cohort: $P = 0.000017$; dMMR: $P = 0.000440$; pMMR: $P = 0.022922$). No adjustment for multiple comparisons was applied, given the exploratory nature of the analysis. *N* values at each timepoint are shown within the figure. **b**,**c**, Waterfall plots depicting pathological tumor regression (%) in individual patients with dMMR tumors (**b**) and pMMR tumors (**c**). Bars represent the percentage of pathologic tumor regression. Annotation tracks above the plots indicate preoperative regimen (BOT in green, BOT + BAL in purple), pathological lymph node status at surgery (positive in yellow), EOT ctDNA status (ctDNA$^+$ in black) and relapse status (relapses shown in red). These panels use the same patient-level framework as Fig. 1, re-annotated with EOT ctDNA status, ypN status, treatment regimen and relapse status to contextualize molecular response at the individual patient level. **d**, Kaplan–Meier analysis of DFS stratified by ctDNA clearance status. Patients were grouped according to whether they achieved ctDNA clearance (blue line) or not (red line). Tick marks indicate censored observations. Time is shown in months from surgery. Hazard ratios and two-sided *P* values comparing post-neoadjuvant ctDNA$^+$ versus ctDNA$^-$ patients were estimated using Cox proportional hazards regression with Firth's correction, landmarked from the date of surgery, with DFS as the endpoint. Twelve-month and 24-month DFS rates were estimated from the surgery landmark timepoint. HR, hazard ratio.

with response (Fig. 2k). Effect estimates were imprecise, with an odds ratio of 26.19 (95% CI: 1.01–679.67) for macrophages and 21.89 (95% CI: 0.99–482.57) for inflammatory myeloid cells. Stromal *CD274* expression showed a positive but non-significant association with response (odds ratio = 9.54; 95% CI: 0.69–132.35; $P = 0.09$).

Fecal ecosystem analysis

We analyzed baseline and post-treatment fecal samples to address two exploratory questions: whether baseline gut microbiota features differed across MMR status, treatment cohorts and pathological response and whether short-course preoperative ICIs were associated with gut microbiota remodeling during treatment.

Among 39 (70%) patients with evaluable stool samples, we assessed microbial species-interacting group (SIG) scores as a summary ecological measure of microbial communities associated with ICI outcomes, beta diversity to evaluate global differences in microbial community structure, linear discriminant analysis effect size (LEfSe) to identify taxa differentially represented across groups and paired Bray–Curtis similarity to quantify microbiota dynamics across treatment.

At baseline, dMMR tumors showed numerically lower SIG scores compared to pMMR tumors, although this difference was not statistically significant, and beta diversity analyses did not highlight any clear separation of microbial communities according to MMR status (Extended Data Fig. 5a)[28].

LEfSe identified differentially represented taxa between patients with pMMR CC and patients with dMMR CC, supporting biological heterogeneity of the baseline fecal ecosystem across molecular subgroups, in line with the literature (Extended Data Fig. 5b)[29].

Across cohorts, SIG scores varied numerically with higher values in pMMR BOT compared to dMMR BOT + BAL, whereas beta diversity showed substantial overlap without a clearly distinct global microbial configuration (Extended Data Fig. 5c,d).

Within the pMMR cohorts, no clear cohort-specific microbial pattern was identified.

Among patients with dMMR CC, baseline microbial profiles showed partial separation between BOT and BOT + BAL cohorts, although global beta diversity comparisons did not reach statistical significance (Extended Data Fig. 5e). Furthermore, LEfSe analysis identified differentially represented taxa between dMMR BOT and dMMR BOT + BAL patients, with *Bacteroides uniformis* enriched in the dMMR BOT + BAL cohort (Extended Data Fig. 5f,g).

Finally, paired with paired longitudinal samples showed reduced Bray–Curtis similarity across all cohorts, consistent with

treatment-associated microbiome remodeling during the preoperative window (Extended Data Fig. 5h).

Longitudinal ctDNA tracking

A summary of individual patient timelines, clinical/pathological annotations and longitudinal ctDNA data is presented in Extended Data Fig. 6.

Among patients with pMMR CC, ctDNA data were available in 27 of 28 (96%). Rates of ctDNA positivity (ctDNA$^+$) were 100% at baseline, 96% on day 15 and 78% at EOT ($P = 0.023$) (Fig. 3a). Clearance of ctDNA (the transition from a detectable ctDNA status to an undetectable one) at EOT was observed in six of 27 (22%) patients with pMMR CC. In detail, ctDNA clearance at EOT was observed in all five pMMR responders ($P < 0.001$) and in one patient who experienced a pPR (30% residual viable tumor). All patients with pMMR CC who experienced ctDNA clearance were treated with BOT + BAL.

Among patients with dMMR CC, ctDNA data were available in 24 of 28 (86%). Rates of ctDNA$^+$ were 88% at baseline, 67% on day 15 and 38% at EOT ($P < 0.001$) (Fig. 3a). Clearance of ctDNA at EOT was observed in 74% of dMMR responders ($P < 0.001$). Of 19 dMMR pathological responders, three (16%) patients did not have any evaluable plasma for ctDNA analysis, and one (5%) patient consistently tested negative from the baseline timepoint and was, thus, excluded from the clearance analysis.

One patient with dMMR CC treated with BOT + BAL who achieved a pCR was consistently ctDNA$^+$ from baseline to EOT. In this patient, a decrease in mean tumor molecules per milliliter of plasma (MTM/ml) was observed across treatment, being 170 MTM/ml at baseline, 120 MTM/ml on day 15 and 1.37 MTM/ml at EOT. After surgery, this patient's ctDNA tested negative and was relapse free at data cutoff.

Across MMR status, nearly all ctDNA-evaluable non-responders had ctDNA$^+$ at EOT (Fig. 3b,c).

At data cutoff, no DFS event occurred in ctDNA-cleared patients (Fig. 3d), and both ctDNA clearance and pMR/pCR were inversely associated with recurrence ($\varphi \approx -0.27$ for each) (Extended Data Fig. 7).

At EOT, higher ctDNA levels were observed in patients who later recurred ($P = 0.046$; Extended Data Fig. 8e) and in patients without a pCR ($P < 0.001$; Extended Data Fig. 8f), whereas lower levels were observed in BOT + BAL-treated patients ($P = 0.0012$; Extended Data Fig. 8g) and in patients with dMMR CC ($P = 0.0059$; Extended Data Fig. 8h).

Peripheral blood mononuclear cell dynamics

Overall, 40 patients (17 with pMMR CC and 23 with dMMR CC) were evaluable for peripheral blood mononuclear cell (PBMC) analysis (Supplementary Table 3): responders had higher baseline central memory T cells (31.0% versus 25.9%; $q = 0.009$) and lower terminally differentiated effector memory T cells (5.9% versus 9.2%; $q = 0.031$) (Supplementary Fig. 8). Across timepoints, responders displayed minimal variations in PD-1$^+$ lymphocytes while showing enrichment in Ki67$^+$GZMB$^+$CD8$^+$ lymphocytes (Supplementary Figs. 9 and 10).

At baseline, pMMR responders showed higher CD8$^+$ T cell fractions compared to pMMR non-responders, whereas, in patients with dMMR CC, baseline PBMC differences according to response did not delineate a clear peripheral immune profile (Supplementary Fig. 11).

In the pMMR cohorts, responders also showed a relative increase in CCR7$^+$CD45RA$^-$ central memory CD8$^+$ T cells across treatment compared to non-responders (day 15: −5.2% versus −17.7%, $q = 0.098$; EOT: +9.8% versus −6.0%; $q = 0.100$) (Supplementary Fig. 12).

Genomic correlates of ICI activity in pMMR CC

As an exploratory analysis, we sought to investigate the association between the status of commonly altered genes in CC and treatment activity in the pMMR cohorts, by leveraging the protocol-specified molecular prescreening. Extended Data Fig. 9 represents an OncoPrint summarizing the genomic landscape of patients with pMMR CC ($n = 28$) matched with selected clinicopathological variables. Three (11%) patients had high tumor mutational burden (TMB), as defined by TMB ≥ 10 mutations per megabase (mut/Mb): one patient enrolled in pMMR BOT (TMB 12.4 mut/Mb) and two patients enrolled in pMMR BOT + BAL (TMB 12.9 mut/Mb and 10.5 mut/Mb). Of these, only one patient treated with BOT + BAL (TMB 10.5 mut/Mb) met the primary endpoint with a pCR.

Overall, alterations in *APC*, *TP53* and *KRAS* were the most frequent events, occurring in 24 of 28 (86%), 22 of 28 (79%) and 19 of 28 (68%) patients. Among 19 tumors with *KRAS* mutations, *KRAS* mutations at G12 were most commonly observed ($n = 14$; 74%): these included four $KRAS^{G12V}$, three $KRAS^{G12A}$, two $KRAS^{G12D}$, two $KRAS^{G12R}$, two $KRAS^{G12S}$ and one $KRAS^{G12C}$. Five patients had tumors with *KRAS* mutations not affecting codon 12: two $KRAS^{A146T}$, one $KRAS^{A146P}$, one $KRAS^{G13D}$ and one $KRAS^{Q61P}$. Across five pMMR responders, five had *TP53*-mutated tumors, four had *APC* mutations and three displayed *KRAS* mutations. The following *KRAS* mutations were observed in pMMR responders: one $KRAS^{G12A}$, one $KRAS^{G12S}$ and one $KRAS^{A146T}$. A granular overview of all mutational findings across the pMMR cohorts is provided in Supplementary Table 5.

Discussion

In UNICORN, preoperative BOT + BAL was safe and active in patients with localized pMMR and dMMR CC candidates for upfront surgery.

The implementation of immunotherapy in the treatment algorithm of patients with pMMR CC has historically been limited due to low immune cell infiltration and intrinsic immune resistance, with trials investigating conventional ICIs yielding underwhelming results[16,30–32].

As a notable exception, the C-800-01 trial highlighted the anti-tumor activity of BOT + BAL in patients with pre-treated pMMR mCRC without active liver metastases[21], warranting further investigation. In this population, the phase 2 ACTIVATE trial[33] randomized patients to standard of care (SOC; regorafenib or trifluridine/tipiracil), BOT or BOT + BAL. With an ORR of 0%, 9% and 19% in SOC, BOT and BOT + BAL arms, respectively, the study underscored both an independent clinical activity of BOT and the relative contribution of PD-1 blockade to the activity of the combination.

Accordingly, we sought to determine whether Fc-enhanced CTLA-4 blockade alone could induce pathological regression in localized pMMR CC, a setting in which conventional ICIs have historically demonstrated limited activity[10,25,26]. Given the lack of activity of anti-PD-1 monotherapy in pMMR CC[10,25,26], as well as the known sensitivity of dMMR tumors to PD-1 blockade[11,13], a BAL monotherapy cohort was expected to provide limited additional insight. Indeed, the partial responses observed with the monotherapy in pMMR CC supports the hypothesis that BOT may be able to initiate biologically meaningful anti-tumor remodeling, but concurrent PD-1 blockade is required to consolidate and deepen this effect.

As a contextualization of our findings, among 31 patients with pMMR CC in NICHE, ipilimumab and nivolumab led to a 20% pMR rate[26] and a 93% 3-year DFS, and patients who progressed were ctDNA$^+$ before surgery. In NEST-1, the promising pathological activity of 4-week preoperative BOT + BAL in 10 patients with localized CC prompted further evaluation[34] and was subsequently confirmed within NEST-2, where 8-week preoperative BOT + BAL yielded a pMR rate of 47% in pMMR CC, with no recurrence at data cutoff[34].

Despite different sample size and therapeutic backbones, limiting cross-trial comparisons, the activity of preoperative BOT + BAL is especially promising when compared to oxaliplatin-based doublets. In FOxTROT, FOLFOX exerted limited pathological regressions across molecular subtypes, with a pMR rate of 8% in pMMR CC[7], underlining the need to develop alternative strategies to enhance anti-tumor activity.

In UNICORN, the numerically lower use of adjuvant chemotherapy in BOT + BAL-treated patients and the shorter duration of oxaliplatin-based doublets may represent valuable indirect benefits of treatment-mediated tumor downstaging. This observation stresses

a further potential advantage of preoperative ICIs in reducing adjuvant chemotherapy use, duration and consequent toxicity, despite requiring confirmation in the context of a randomized trial. Sparing chemotherapy-associated toxicity is especially crucial in the curative setting, given the potentially irreversible neurotoxicity of oxaliplatin, leading to often lifelong interference with quality of life.

Although baseline ctDNA$^+$ may potentially reflect a relatively more advanced disease[26], ctDNA dynamics during treatment may provide an early and minimally invasive readout of therapeutic activity in patients with localized CC undergoing neoadjuvant ICIs. Although this hypothesis requires further validation, its biological plausibility supports the exploration of ctDNA as a tool to inform clinical trials investigating adaptive strategies, including NOM and adjuvant chemotherapy de-escalation.

Notably, the observation of ctDNA clearance in a pMMR partial responder leverages the hypothesis that longer treatment duration may enhance preoperative ICI activity and challenges the concept that pathological response should be considered as a continuous rather than a dichotomous variable in this setting, with ctDNA-cleared deep partial responders potentially sharing favorable outcomes with major responders.

A rare opportunity of UNICORN was the systematical acquisition of paired pre-treatment and post-treatment tissue samples, allowing to capture early spatial immune remodeling induced by short-course ICIs. The positive correlation between baseline *CD274* stromal expression and pathological response in pMMR CC is notable, given the lack of predictive biomarkers of ICI activity in this setting, but this observation will require prospective validation in larger cohorts using clinically available PD-L1 immunohistochemistry-based assays.

Furthermore, among pMMR responders, the observed T_{reg} inversion pattern may suggest that pathological activity may not necessarily require the absence of baseline immunosuppressive niches but, rather, hints to the ability of BOT + BAL to actively remodel the tumor microenvironment during treatment. This observation is consistent with the proposed FcγR-mediated T_{reg} depletion as a defining feature of BOT's Fc-enhanced design[18,19]. Nonetheless, this finding should be interpreted cautiously, as transcriptional spatial signatures do not directly demonstrate FcγR-dependent T_{reg} depletion and may also reflect changes in T_{reg} state, localization or surrounding immune contexture.

Higher *FCGR2A* and *FCGR3A* expression has previously been linked with BOT activity and improved survival but without clear cellular resolution[18]: by localizing elevated *FCGR2A* and *FCGR3A* expression to intratumoral macrophages of dMMR CC, we suggest a potential association of BOT activity with engagement of tumor-resident myeloid cells.

Interestingly, BOT-related diarrhea/colitis was uncommon in UNICORN, which may at least partly be due to the short treatment duration as well as low BOT dose level, in line with evidence in the metastatic setting where the incidence of diarrhea/colitis was consistently higher with incremental BOT dosage[33].

Finally, gut microbiome profiling provided limited insights into the fecal ecosystem of patients receiving preoperative ICIs. Although *B. uniformis* has previously been linked with ICI benefit in solid tumors[35], the taxonomic associations in UNICORN should be regarded as hypothesis generating given the sample size and multiple potential confounders inherent to fecal microbiome studies.

This study has numerous additional limitations. First, the comparability of efficacy outcomes is burdened by limited sample size, differential follow-up times across cohorts and the non-randomized study design. Second, the enrolled patients were almost exclusively Caucasians, potentially limiting global applicability of the results. Additionally, given the sample size and data-driven selection of candidate biomarkers, we further stress that all of our translational findings should be interpreted as purely hypothesis generating. From a statistical point of view, findings should be interpreted with particular caution in cohorts meeting the primary endpoint at the decision boundary, and we acknowledge that the family-wise error rate across parallel hypothesis tests may be increased in UNICORN.

That withstanding, these constraints are bound to the window-of-opportunity design of the study, which was conducted to purely seek signals of activity of short-course preoperative treatment in locally advanced CC. The activity of BOT + BAL in pMMR CC positively compares to similar trials in this setting. As opposed to NEST, UNICORN was conducted as a multicenter trial, supporting the broader feasibility of the proposed strategy. In addition, NICHE and NEST did not employ radiological eligibility criteria and included patients with stage I CC, potentially inflating pathological activity and exposing patients with favorable prognosis after upfront surgery to an overtreatment. Interestingly, UNICORN and NICHE highlighted similar genomic correlates of treatment benefit in pMMR CC, with *TP53* mutations observed in most responders, but this should be interpreted cautiously due to the overall high prevalence of *TP53* mutations in CRC and the limited sample size of both studies.

Finally, our findings should be interpreted within the context of a rapidly evolving neoadjuvant landscape, in which ongoing trials of chemoimmunotherapeutic strategies may define potentially highly active comparator approaches[36].

Nonethless, the promising activity of BOT + BAL supports this combination as a leading candidate to advance immunotherapy in the neoadjuvant setting and, potentially, to enable organ-sparing strategies in patients achieving multimodally assessed deep clinical responses integrating radiology, endoscopy and longitudinal ctDNA.

In line with our findings, UNICORN was amended to investigate the role of 8-week neoadjuvant BOT + BAL in a larger cohort of 29 patients with locally advanced pMMR CC (Supplementary Fig. 13). This will provide further insights into the role of the combination as a purely neoadjuvant approach and inform the conduct of a global, randomized phase 3 trial of neoadjuvant BOT + BAL versus upfront surgery in locally advanced pMMR CC.

In conclusion, UNICORN provides preliminary compelling evidence that short-course preoperative BOT + BAL can elicit deep pathological responses in resectable pMMR and dMMR CC.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41591-026-04710-1.

References

1. Bray, F. et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **74**, 229–263 https://doi.org/10.3322/caac.21834 (2024).
2. Siegel, R. L., Wagle, N. S. & Jemal, A. Leading cancer deaths in people younger than 50 years. *JAMA* **335**, 632–634 https://doi.org/10.1001/jama.2025.25467 (2026).
3. Grothey, A. et al. Duration of adjuvant chemotherapy for stage III colon cancer. *N. Engl. J. Med.* **378**, 1177–1188 https://doi.org/10.1056/NEJMoa1713709 (2018).
4. Lordick, F. et al. Gastric cancer: ESMO Clinical Practice Guideline for diagnosis, treatment and follow-up. *Ann. Oncol.* **33**, 1005–1020 https://doi.org/10.1016/j.annonc.2022.07.004 (2022).
5. Scott, A. J. et al. Management of locally advanced rectal cancer: ASCO Guideline. *J. Clin. Oncol.* **42**, 3355–3375 https://doi.org/10.1200/JCO.24.01160 (2024).
6. Hofheinz, R.-D. et al. Localised rectal cancer: ESMO Clinical Practice Guideline for diagnosis, treatment and follow-up. *Ann. Oncol.* **36**, 1007–1024 https://doi.org/10.1016/j.annonc.2025.05.528 (2025).

7. Morton, D. et al. Preoperative chemotherapy for operable colon cancer: mature results of an international randomized controlled trial. *J. Clin. Oncol.* **41**, 1541–1552 https://doi.org/10.1200/JCO.22.00046 (2023).
8. Sargent, D. J. et al. Defective mismatch repair as a predictive marker for lack of efficacy of fluorouracil-based adjuvant therapy in colon cancer. *J. Clin. Oncol.* **28**, 3219–3226 https://doi.org/10.1200/JCO.2009.27.1825 (2010).
9. André, T. et al. Pembrolizumab in microsatellite-instability–high advanced colorectal cancer. *N. Engl. J. Med.* **383**, 2207–2218 https://doi.org/10.1056/NEJMoa2017699 (2020).
10. Chalabi, M. et al. Neoadjuvant immunotherapy leads to pathological responses in MMR-proficient and MMR-deficient early-stage colon cancers. *Nat. Med.* **26**, 566–576 https://doi.org/10.1038/s41591-020-0805-8 (2020).
11. Chalabi, M. et al. Neoadjuvant immunotherapy in locally advanced mismatch repair–deficient colon cancer. *N. Engl. J. Med.* **390**, 1949–1958 https://doi.org/10.1056/NEJMoa2400634 (2024).
12. De Gooyer, P. G. M. et al. Neoadjuvant nivolumab and relatlimab in locally advanced MMR-deficient colon cancer: a phase 2 trial. *Nat. Med.* **30**, 3284–3290 https://doi.org/10.1038/s41591-024-03250-w (2024).
13. Cercek, A. et al. PD-1 blockade in mismatch repair–deficient, locally advanced rectal cancer. *N. Engl. J. Med.* **386**, 2363–2376 https://doi.org/10.1056/NEJMoa2201445 (2022).
14. Xu, R. H. et al. 119TiP A randomized, controlled, multicenter phase III study of IBI310 (anti-CTLA-4 antibody) plus sintilimab (anti-PD-1 antibody) as neoadjuvant treatment for resectable microsatellite instability-high/mismatch repair-deficient (MSI-H/dMMR) colon cancer. *Ann. Oncol.* **35**, S1447 https://doi.org/10.1016/j.annonc.2024.10.141 (2024).
15. Starling, N. et al. AZUR-2, a phase III, open-label, randomized study of perioperative dostarlimab monotherapy vs standard of care in previously untreated patients with T4N0 or stage III dMMR/MSI-H resectable colon cancer. *J. Clin. Oncol.* **42**, TPS240 https://doi.org/10.1200/JCO.2024.42.3_suppl.TPS240 (2024).
16. Le, D. T. et al. PD-1 blockade in tumors with mismatch-repair deficiency. *N. Engl. J. Med.* **372**, 2509–2520 https://doi.org/10.1056/NEJMoa1500596 (2015).
17. Overman, M. J. et al. Durable clinical benefit with nivolumab plus ipilimumab in DNA mismatch repair–deficient/microsatellite instability–high metastatic colorectal cancer. *J. Clin. Oncol.* **36**, 773–779 https://doi.org/10.1200/JCO.2017.76.9901 (2018).
18. Chand, D. et al. Botensilimab, an Fc-enhanced anti–CTLA-4 antibody, is effective against tumors poorly responsive to conventional immunotherapy. *Cancer Discov.* **14**, 2407–2429 https://doi.org/10.1158/2159-8290.CD-24-0190 (2024).
19. Waight, J. D. et al. Selective FcγR co-engagement on APCs modulates the activity of therapeutic antibodies targeting T cell antigens. *Cancer Cell* **33**, 1033–1047 https://doi.org/10.1016/j.ccell.2018.05.005 (2018).
20. Iwama, S. et al. Pituitary expression of CTLA-4 mediates hypophysitis secondary to administration of CTLA-4 blocking antibody. *Sci. Transl. Med.* **6**, 230ra45 https://doi.org/10.1126/scitranslmed.3008002 (2014).
21. Schlechter, B. L. et al. Extended follow-up of botensilimab plus balstilimab in an expanded cohort of microsatellite-stable metastatic colorectal cancer without active liver metastases. *Clin. Cancer Res.* https://doi.org/10.1158/1078-0432.CCR-26-1610 (2026)
22. O'Day, S. et al. 398 AGEN1181, an Fc engineered anti-CTLA-4 antibody, demonstrates clinical activity, alone or in combination with balstilimab (anti-PD-1), and broadens the therapeutic potential of CTLA-4 therapy. *J. Immunother. Cancer* **8** https://doi.org/10.1136/jitc-2020-SITC2020.0398 (2020).
23. Tanne, A. J. et al. Abstract 1878: Fc-enhanced anti-CTLA-4 antibody, AGEN1181: new mechanistic insights for potent antitumor immunity and combination potential in treatment-resistant solid tumors. *Cancer Res.* **81**, 1878 https://doi.org/10.1158/1538-7445.AM2021-1878 (2021).
24. Mlecnik, B. et al. Integrative analyses of colorectal cancer show immunoscore is a stronger predictor of patient survival than microsatellite instability. *Immunity* **44**, 698–711 https://doi.org/10.1016/j.immuni.2016.02.025 (2016).
25. Verschoor, Y. L. et al. Neoadjuvant nivolumab, ipilimumab, and celecoxib in MMR-proficient and MMR-deficient colon cancers: final clinical analysis of the NICHE study. *J. Clin. Oncol.* **40**, 3511 https://doi.org/10.1200/JCO.2022.40.16_suppl.3511 (2022).
26. Tanm, P. B. et al. Neoadjuvant immunotherapy in mismatch-repair-proficient colon cancers. *Nature* **648**, 726–735 https://doi.org/10.1038/s41586-025-09679-4 (2025).
27. Halvarsson, B., Anderson, H., Domanska, K., Lindmark, G. & Nilbert, M. Clinicopathologic factors identify sporadic mismatch repair–defective colon cancers. *Am. J. Clin. Pathol.* **129**, 238–244 https://doi.org/10.1309/0PP5GDRTXUDVKAWJ (2008).
28. Derosa, L. et al. Custom scoring based on ecological topology of gut microbiota associated with cancer immunotherapy outcome. *Cell* **187**, 3373–3389 https://doi.org/10.1016/j.cell.2024.05.029 (2024).
29. Segata, N. et al. Metagenomic biomarker discovery and explanation. *Genome Biol.* **12**, R60 https://doi.org/10.1186/gb-2011-12-6-r60 (2011).
30. Overman, M. J. et al. Nivolumab in patients with metastatic DNA mismatch repair-deficient or microsatellite instability-high colorectal cancer (CheckMate 142): an open-label, multicentre, phase 2 study. *Lancet Oncol.* **18**, 1182–1191 https://doi.org/10.1016/S1470-2045(17)30422-9 (2017).
31. Eng, C. et al. Atezolizumab with or without cobimetinib versus regorafenib in previously treated metastatic colorectal cancer (IMblaze370): a multicentre, open-label, phase 3, randomised, controlled trial. *Lancet Oncol.* **20**, 849–861 https://doi.org/10.1016/S1470-2045(19)30027-0 (2019).
32. Chen, E. X. et al. Effect of combined immune checkpoint inhibition vs best supportive care alone in patients with advanced colorectal cancer: the Canadian Cancer Trials Group CO.26 study. *JAMA Oncol.* **6**, 831 https://doi.org/10.1001/jamaoncol.2020.0910 (2020).
33. Fakih, M. et al. Preliminary results from a randomized, open-label, phase 2 study of botensilimab (BOT) with or without balstilimab (BAL) in refractory microsatellite stable metastatic colorectal cancer with no liver metastases (MSS mCRC NLM). *J. Clin. Oncol.* **43**, 23 https://doi.org/10.1200/JCO.2025.43.4_suppl.23 (2025).
34. Hissong, E. et al. Neoadjuvant botensilimab (BOT) plus balstilimab (BAL) in resectable mismatch repair proficient (pMMR) and deficient (dMMR) colorectal cancer (CRC): NEST clinical trial update. *J. Clin. Oncol.* **43**, 207 https://doi.org/10.1200/JCO.2025.43.4_suppl.207 (2025).
35. Weis, A. M. et al. A capsular polysaccharide from a healthy human microbiota member activates a Lag-3-NK cell axis to restrain colon cancer and augment immunotherapy. *Cell Rep.* **44**, 116172 https://doi.org/10.1016/j.celrep.2025.116172 (2025).
36. Zhang, J. et al. Neoadjuvant mFOLFOXIRI chemotherapy with or without cadonilimab versus mFOLFOX6 alone in locally advanced colorectal cancer: a randomized phase II study (OPTICAL2). *Med* **7**, 101141 (2026).

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

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2026

Filippo Ghelardi[1], Margherita Ambrosini[1], Alessandra Raimondi[1], Andrei Tyshevich[2], Federica Palermo[1], Francesca Bergamo[3], Vincenzo Nasca[1], Paolo Manca[1], Ilaria Mulas[2], Lisa Derosa[4], Gina Brown[5], Claudia Chiodoni[6], Massimo Di Maio[7], Samantha Di Donato[8], Anna Chiaramonte[1], Lorenzo Castagnoli[9], Stephanie Ni[10], Darryl Nousome[10], Erik Spickard[10], Antony Tin[10], Robert Lentz[10], Adham Jurdi[10], Giovanna Sabella[11], Daniele Spada[12], Stefano Tamberi[13,14], Alessandro Passardi[15], Federica Morano[1], Emiliano Tamburini[16], Chiara Rossi[11], Luca Azzolin[2], Francesco Cambuli[2], Antonia Martinetti[1], Elisa Sottotetti[1], Andrew Allan Almonte[4], Valerio Iebba[4], Chiara Ratti[6], Riccardo Cerantola[3], Özlem Kutlu[17], Eleonora Cristarella[1], Serenella M. Pupa[9], Cecilia Villa[1], Eleonora Perissinotto[3], Margherita Galassi[18], Andrea Barina[19], Fabrizio Vittadello[20], Marcello Guaglio[21], Gianluca Baiocchi[22], Andrea Sartore-Bianchi[23,24], Riccardo Rosati[25], Chiara Cremolini[26,27], Isacco Montroni[21], Sara Lonardi[3], Andrea Sottoriva[2] & Filippo Pietrantonio[1] ✉

[1]Department of Medical Oncology, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy. [2]Computational Biology Research Centre, Human Technopole, Milan, Italy. [3]Department of Oncology, Veneto Institute of Oncology IOV-IRCCS, Padova, Italy. [4]Tumor Immunology and Cancer Immunotherapy, ClinicObiome, Université Paris-Saclay, Gustave Roussy, INSERM UMR1367, Villejuif, France. [5]Department of Surgery and Cancer, Imperial College London, London, UK. [6]Molecular Immunology Unit, Department of Experimental Oncology, Amadeolab, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy. [7]Department of Oncology, University of Turin, Azienda Ospedaliero-Universitaria Città della Salute e della Scienza di Torino, Turin, Italy. [8]Department of Medical Oncology, General Hospital, Prato, Italy. [9]Microenvironment and Biomarkers of Solid Tumors Unit, Department of Experimental Oncology, Amadeolab, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy. [10]Natera, Austin, TX, USA. [11]First Pathology Unit, Department of Pathology and Laboratory Medicine, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy. [12]Medical Oncology Unit, Azienda Socio-Sanitaria Territoriale of Cremona, Cremona Hospital, Cremona, Italy. [13]Oncology Unit, Santa Maria delle Croci Hospital Ravenna, AUSL Romagna, Ravenna, Italy. [14]Department of Medical Sciences DIMEC, University of Bologna, Bologna, Italy. [15]Department of Medical Oncology, IRCCS Istituto Romagnolo per lo Studio dei Tumori 'Dino Amadori', Meldola, Italy. [16]Department of Oncology and Palliative Care, Cardinale Giovanni Panico General Hospital, Tricase, Italy. [17]Department of Medical Oncology, Izmir City Hospital, Izmir, Turkey. [18]Centrale Produzione Farmaci, Hospital Pharmacy, National Cancer Institute of Milan, Milan, Italy. [19]Surgical Oncology of Digestive Tract, Veneto Institute of Oncology, IOV-IRCCS, Padua, Italy. [20]Department of Surgery, Hospital Sant'Antonio, Azienda Ospedale Università di Padova, Padova, Italy. [21]Colorectal Surgery Unit, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy. [22]Surgical Clinic, Department of Experimental and Clinical Sciences, University of Brescia, Brescia, Italy. [23]Department of Oncology and Hemato-Oncology, University of Milan, Milan, Italy. [24]Department of Hematology, Oncology and Molecular Medicine, ASST Grande Ospedale Metropolitano Niguarda, Milan, Italy. [25]Gastrointestinal Surgery Unit, IRCCS San Raffaele Scientific Institute, Milan, Italy. [26]Department of Oncology, University Hospital of Pisa, Pisa, Italy. [27]Department of Translational Research and New Technologies in Medicine, University of Pisa, Pisa, Italy.
✉e-mail: filippo.pietrantonio@istitutotumori.mi.it

Methods

Ethical approval and consenting

The trial was conducted in accordance with Good Clinical Practice guidelines, International Council for Harmonization Good Clinical Practice (ICH-GCP) standards, the ethical principles of the Declaration of Helsinki and all applicable regulatory requirements. The study was authorized by the Italian Medicines Agency (AIFA) through the Clinical Trials Information System (CTIS) under EU Regulation Number 536/2014 (EU CT number 2022-501308-90-00; CTIS ID: 1351), after favorable assessment of part I and part II of the application, including approval by the competent Ethics Committee 'Comitato Etico per la sperimentazione clinica delle province di Verona e Rovigo' on 19 April 2023. All participating centers were activated only after completion of the required site-specific authorization procedures. All patients provided written informed consent before any study-specific procedures.

Study registration

The study was prospectively registered at ClinicalTrials.gov (NCT05845450) on 26 April 2023, before the enrollment of the first patient. The original study protocol and statistical analysis plan are provided in the Supplementary Information. Analyses not prespecified in the protocol are explicitly identified as exploratory.

Study design and treatment plan

UNICORN (NCT05845450) is an ongoing multicenter, non-randomized, open-label, parallel assignment, window-of-opportunity, umbrella/platform phase 2 clinical trial enrolling patients with non-metastatic and radiologically staged rT3–T4, N0 or N+ CRC initially prescreened for the presence of specific molecular actionable alterations or profiles and then enrolled and treated with a matched short-course preoperative treatment. Study design is depicted in Extended Data Fig. 1. Sex and gender were not considered in the study design. No compensation was provided to participants.

The molecular prescreening consists of the following:

1. Evaluation of MMR status by immunohistochemistry (IHC) staining of four MMR proteins (MLH1, PMS2, MSH2 and MSH6). Tumoral nuclear staining is evaluated to assess tumor MMR status, and retained protein expression in the surrounding stroma and lymphoid cells serves as the internal positive control. Deficient MMR (dMMR) is defined as a complete loss of nuclear staining in tumor cells of at least one of the MMR proteins. Conversely, tumors showing preserved nuclear expression of all four proteins are classified as proficient MMR (pMMR).
2. Next-generation sequencing through Oncomine Comprehensive Assay (OCA) Plus panel for the X5 Ion Torrent Personal Genome Platform (a full list of analyzed genes is provided below in Appendix A). The assay simultaneously evaluates 510 genes and can detect single-nucleotide variations, insertions and deletions (indels), copy number variations (CNVs), gene fusions (both known and novel) and splice variants.
3. HER2 IHC with or without *HER2* silver in situ hybridization (SISH), with testing algorithm and positivity defined according to American Society of Clinical Oncology/College of American Pathologists/American Society for Clinical Pathology (ASCO/CAP/ASCP) guidelines for gastric and gastroesophageal cancers.

All molecular analyses are centrally performed at the Coordinating Center (Fondazione IRCCS Istituto Nazionale dei Tumori). Whenever a prespecified molecular actionable genomic alteration or profile is identified, the patient will be eligible for the matching cohort and, after enrollment, will receive a specific short-course preoperative treatment. Each cohort has a targeted enrollment of 14 patients, as per the statistical design. In brief, 10 cohorts were currently terminated or ongoing in the frame of the UNICORN study. Patients with HER2^{+} tumors are enrolled in cohort 1 for a short-course preoperative treatment with the HER2-directed antibody–drug conjugate trastuzumab deruxtecan[37,38]. Cohort 2 enrolls patients with *POLE*/*POLD1*-mutated tumors and ultra-mutated status (TMB ≥ 100 mut/Mb) to receive preoperative treatment with the anti-PD-L1 monoclonal antibody durvalumab[39]. Patients with EGFR-dependent tumors, defined as tumors arising from the left side of the colon or the rectum; *RAS*/*BRAF* wild-type status; and 'hyperselected' because of the lack of molecular alteration included in the previously published PRESSING panel (*HER2* amplification/activating mutations, *MET* amplification, *NTRK*/*ROS1*/*ALK*/*RET* fusions, *PIK3CA* exon 20 mutations, *PTEN* inactivating mutations and *AKT1* mutations) were included in cohort 3 (completed) and received the anti-EGFR monoclonal antibody panitumumab[40]. Patients selected for the presence of pMMR/microsatellite stable (MSS) status and absence of HER2 overexpression/amplification, absence of pathogenic mutations in the *POLE* and *POLD1* proofreading domains associated with ultra-mutated status were enrolled to receive preoperative treatment with the anti-CTLA-4 antibody BOT as monotherapy (cohort 4) or in combination with the anti-PD-1 monoclonal antibody BAL (cohort 5). Similarly, patients selected for the presence of dMMR/microsatellite instability-high (MSI-H) status and absence of pathogenic mutations in the *POLE* and *POLD1* proofreading domains associated with ultra-mutated status were enrolled to receive BOT with or without BAL across cohorts 6 and 7, respectively. Patients with $KRAS^{G12C}$ mutation are enrolled in cohort 8 to receive the selective KRAS(G12C) inhibitor sotorasib plus the anti-EGFR agent panitumumab[41]. After completion of cohorts 4 and 5, two additional cohorts are now enrolling patients with pMMR/MSS tumors to receive short-course preoperative treatment with the EP4 receptor antagonist vorbipiprant, either as monotherapy (cohort 9) or in association with the anti-PD-1 monoclonal antibody nivolumab (cohort 10)[42]. Finally, after completion of cohort 3, two additional cohorts are now enrolling patients with pMMR/MSS, *RAS*/*BRAF* wild-type status, hyperselected (regardless of *MET*), left-sided (cohort 11) or right-sided (cohort 12) CRC to receive preoperative subcutaneous amivantamab.

At the time of this paper redaction, six cohorts have completed enrollment (cohorts 3, 4, 5, 6 and 9).

This paper reports the results of cohorts 4, 5, 6 and 7, which enrolled patients with pMMR/MSS disease treated with BOT alone (cohort 4) or in combination with BAL (cohort 5) or patients with dMMR/MSI-H tumors treated with BOT alone (cohort 6) or with the addition of BAL (cohort 7). Specific molecular inclusion criteria include:

Cohorts 4 and 5: patients selected for the presence of pMMR/MSS status, absence of HER2 overexpression/amplification and absence of pathogenic mutations in the *POLE* and *POLD1* proofreading domains associated with ultra-mutated status. Patients were enrolled in cohort 5 upon completion of cohort 4.

Cohorts 6 and 7: patients selected for the presence of dMMR/MSI-H status and absence of pathogenic mutations in the *POLE* and *POLD1* proofreading domains associated with ultra-mutated status. Patients were enrolled in cohort 7 upon completion of cohort 6.

Patient population

Patients aged 18 years or older with previously untreated histologically confirmed colorectal adenocarcinoma that is judged radically resectable and is radiologically staged as rT3–4, N0/N+, M0. Patients with rectal cancer not requiring preoperative radiotherapy based on MDT assessment were allowed; for patients with rectal cancer, contrast-enhanced magnetic resonance imaging (MRI) of the pelvis was mandatory. The main exclusion criteria were metastatic disease and risk of bowel obstruction, unless adequately managed through surgical creation of a protective ostomy or placement of an endoluminal stent. A complete list of eligibility criteria is available below.

Eligibility criteria

General inclusion criteria. All of the following inclusion criteria must be met for a patient to be included in the study:

- Provide a signed and dated informed consent document prior to any study procedure
- Age ≥18 years at time of informed consent
- Eastern Cooperative Oncology Group (ECOG) performance status of 0 or 1
- Histologically confirmed colorectal adenocarcinoma that is judged as initially resectable with elective surgery aimed at radical intent with R0 margins as per multidisciplinary team (MDT) assessment
- Radiological stage cT3–4, N0–2, M0 using computed tomography as in the pivotal FOxTROT study
- Patients with rectal cancer candidate for R0 resection, not requiring preoperative radiotherapy based on MDT assessment, with the following characteristics on high-resolution thin-slice (3-mm) contrast-enhanced MRI:
  - ≤T3a defined at the MRI (perivisceral fat infiltration <2 mm) and clinical N0
  - Upper-medium, defined as tumors with distal margin ≥5 cm from the anal verge
  - Absence of mesorectal fascia invasion, as defined as a distance ≥1 mm between tumor and the mesorectal fascia
- Absence of distant metastases at baseline as defined by negativity of chest/abdomen/pelvis contrast-enhanced computed tomography
- Able to provide enough archival formalin-fixed paraffin-embedded (FFPE) tumor specimen that is already available from initial diagnostic procedures for the purpose of molecular prescreening
- Presence of one of the selected molecular profiles/alterations after central prescreening and necessary for the assignment to a matching treatment cohort
- No prior systemic treatment for CRC or neoadjuvant radiation therapy for rectal cancer
- Adequate bone marrow function characterized by the following at screening:
  - Absolute neutrophil count $\geq 1.5 \times 10^9$ per liter
  - Platelet count $\geq 100 \times 10^9$ per liter
  - Hemoglobin ≥9.0 g dl^{-1}. Note that transfusions will be allowed to achieve this level provided the patient has not received more than two units of red blood cells in the prior week to achieve this criterion.
- Adequate renal function characterized by serum creatinine ≤1.5× upper limit of normal (ULN) or calculated by Cockroft–Gault formula or directly measured creatinine clearance ≥50 ml min^{-1} at screening.
- Adequate hepatic function characterized by the following at screening:
  - Serum total bilirubin ≤1.5× ULN and <2 mg dl^{-1}. Note that patients who have a total bilirubin level 1.5× ULN will be allowed if their indirect bilirubin level is ≤1.5× ULN.
  - Alanine aminotransferase and/or aspartate aminotransferase ≤2.5× ULN
- Women of childbearing potential must have a negative blood pregnancy test at the baseline visit. For this trial, women of childbearing potential are defined as all women after puberty, unless they are postmenopausal for at least 12 months, surgically sterile or sexually inactive. A postmenopausal state is defined as no menses for 12 months without an alternative medical cause. A high follicle-stimulating hormone (FSH) level in the postmenopausal range may be used to confirm a postmenopausal state in women not using hormonal contraception or hormonal replacement therapy. However, in the absence of 12 months of amenorrhea, a single FSH measurement is insufficient.
- Patients and their partners must be willing to avoid pregnancy during the trial and until a specific time interval after the last trial treatment: 4 months after the last dose of trastuzumab deruxtecan; 3 months for durvalumab, BOT and BAL; and 2 months for panitumumab. Male patients with female partners of childbearing potential and female patients of childbearing potential must, therefore, be willing to use adequate contraception as approved by the investigator (barrier contraceptive measure or oral contraception).
- Willing and able to comply with scheduled visits, treatment plan, laboratory tests and other study procedures

Cohort-specific inclusion criteria

Cohort 4 (pMMR BOT): pMMR/MSS status and absence of molecular characteristics allowing enrollment in cohorts 1 and 2 (that is, HER2 overexpression/amplification and absence of pathogenic mutations in the *POLE* and *POLD1* proofreading domains associated with ultra-mutated status, respectively). Patients with tumor characteristics of EGFR dependency will be eligible for cohort 4 only after completion of cohort 3 and upon discussion with the sponsor.

Cohort 5 (pMMR BOT + BAL; opened sequentially after completion of cohort 4): pMMR/MSS status and absence of molecular characteristics allowing enrollment in cohorts 1 and 2 (that is, HER2 overexpression/amplification and absence of pathogenic mutations in the *POLE* and *POLD1* proofreading domains associated with ultra-mutated status, respectively). Patients with tumor characteristics of EGFR dependency will be eligible for cohort 5 only after completion of cohort 3 and upon discussion with the sponsor.

Cohort 6 (dMMR BOT): dMMR/MSI-H status and absence of POLE/D1 proof-read domain pathogenic mutation associated with ultra-mutated status

Cohort 7 (dMMR BOT + BAL; opened sequentially after completion of cohort 6): dMMR/MSI-H status and absence of POLE/D1 proof-read domain pathogenic mutation associated with ultra-mutated status

General exclusion criteria

Patients meeting any of the following criteria will be excluded from the study:

- Distant metastases at any site
- Risk criteria for obstructing disease at radiology or endoscopy as defined in the pivotal FOxTROT study
- Need to receive neoadjuvant radiation or chemoradiation in patients with rectal cancer
- Patients with known hypersensitivity to the study drug of the assigned cohort or to its excipients or to drugs belonging to the same drug class
- Previous or concurrent malignancy within 2 years of study entry
- Impaired cardiovascular function or clinically significant cardiovascular diseases, including any of the following: history of acute myocardial infarction, acute coronary syndromes (including unstable angina, coronary artery bypass graft, coronary angioplasty or stenting) ≤6 months prior to start of study treatment, symptomatic congestive heart failure (that is, grade 2 or higher) and history or current evidence of clinically significant cardiac arrhythmia and/or conduction abnormality ≤6 months prior to start of study treatment, except atrial fibrillation and paroxysmal supraventricular tachycardia
- Known history of HIV infection
- Active infection, including tuberculosis (clinical evaluation that includes clinical history, physical examination and radiographic

findings and tuberculosis testing in line with local practice), hepatitis B (known positive HBV surface antigen (HBsAg) result) and hepatitis C (HCV). Patients with a past or resolved HBV infection (defined as the presence of hepatitis B core antibody (anti-HBc) and absence of HBsAg) are eligible only in case of negativity of HBV DNA. Patients positive for HCV antibody are eligible only if polymerase chain reaction is negative for HCV RNA.
- Other severe acute or chronic diseases that may increase the risk associated with study participation or study drug administration or that may interfere with the interpretation of study results and, in the judgment of the investigator, would make the patient an inappropriate candidate for the study
- Any psychiatric condition that would prohibit the understanding or rendering of informed consent and that would limit compliance with trial requirements
- Women in pregnancy or lactation condition, women with childbearing potential or sexually active men not willing to use adequate contraception during whole study period
- Use of any disallowed drug (section 9.2 of protocol v.1.0, dated 3 December 2021) that cannot be suspended ≤1 week before study treatment start

Cohort-specific exclusion criteria

Cohorts 4 (pMMR BOT), 5 (pMMR BOT + BAL), 6 (dMMR BOT) and 7 (dMMR BOT + BAL):

- History of autoimmune diseases or history of bone marrow or organ transplantation that requires immunosuppressive therapy. The following are exceptions to this criterion:
  - Patients with vitiligo or alopecia
  - Patients with hypothyroidism (for example, after Hashimoto syndrome) stable on hormone replacement
  - Any chronic skin condition that does not require systemic therapy
  - Patients with celiac disease controlled by diet alone
- History of active primary immunodeficiency
- Any condition requiring systemic treatment with corticosteroids at doses equal to or higher than 10 mg daily of prednisone or equivalents or other immunosuppressive drugs within 14 days from inclusion in the study. The following medications are exceptions to this criterion:
  - Intranasal, inhaled, topical steroids or local steroid injections (for example, intra-articular injection)
  - Systemic corticosteroids at physiologic doses not to exceed 10 mg daily of prednisone or its equivalent
  - Steroids as premedication for hypersensitivity reactions (for example, computed tomography scan premedication)
- Administration of live vaccines within 4 weeks from inclusion in the study. Note that patients, if enrolled, should not receive a live vaccine while receiving study drug(s) and up to 30 days after the last dose of study drug(s).
- Prior treatment with anti-CTLA4, anti-PD-1 or anti-PD-L1 therapeutic antibody or pathway-targeting agents

Study treatments

Cohorts 4 and 6 were opened concomitantly. Enrollment in cohort 5 was initiated after completion of cohort 4, and cohort 7 was opened after completion of cohort 6. Patients enrolled in cohorts 4 and 6 received the anti-CTLA4 monoclonal antibody BOT, intravenously, at the dose of 1 mg kg^{-1} on day 1.

Patients enrolled in cohorts 5 and 7 received the anti-CTLA4 monoclonal antibody BOT, intravenously, at the dose of 1 mg kg^{-1} on day 1 and the anti-PD-1 monoclonal antibody BAL, intravenously, at the dose of 3 mg kg^{-1} on days 1 and 15.

In case of onset of diarrhea considered to be related to treatment with BOT, the anti-tumor necrosis factor (TNF) agent infliximab was administered as rescue medication within 72 hours from the onset at the dose of 5 mg kg^{-1}, according to the predefined toxicity management algorithm (Appendix B).

Study procedures and assessments

Patients underwent study-defined molecular prescreening on archival tumor tissue samples from the initial diagnostic colonoscopy (B1 sample). An optional colonoscopy with biopsy (B2 sample) was performed if archival tumor tissue was insufficient to complete prescreening analyses and to perform all other protocol-defined mandatory analyses included in the secondary and exploratory endpoints. Baseline radiological assessment was performed within 3 weeks (21 days) before treatment start through a chest/abdomen/pelvis computed tomography scan with intravenous contrast, and pelvic MRI was mandatory for patients with rectal cancer to confirm eligibility. For patients with contraindications to iodinated contrast agents due to a history of allergic reaction, a contrast-enhanced MRI scan of the abdomen/pelvis and a chest computed tomography scan without contrast could have been performed as baseline radiological assessments. Baseline evaluation of serum tumor markers (carcinoembryonic antigen (CEA) and carbohydrate antigen 19.9 (CA19.9)) was performed.

After study treatment, during the EOT window (between days 29 and 35 from treatment start), tumor radiological reassessment was performed using the same imaging technique (computed tomography or MRI) employed at baseline. Serum tumor markers (CEA and CA19.9) were also assessed during the EOT window. Patients underwent radical surgery on day 35 ± 5.

After surgery, patients received standard postoperative management as per national and international guidelines and according to the evaluation of a central MDT. The MDT consisted of a team of experts in the field of colorectal and gastrointestinal oncology, including medical and radiation oncologists, surgeons, radiologists and pathologists. The MDT gathered weekly at the Coordinating Center (Fondazione IRCCS Istituto Nazionale dei Tumori) and provided a thorough evaluation of deidentified patient and tumor characteristics (including baseline radiological images, patient demographics and clinical history as well as locally assessed pathological reports). A personalized suggestion regarding the most adequate postoperative management, based on major international guidelines, was, therefore, submitted to the Participating Center referring the specific patient.

Safety assessments included monitoring of adverse events, clinical laboratory tests (chemistry, hematology, coagulation and urinalysis), vital signs, physical examinations and electrocardiogram as applicable. Health-related quality of life was evaluated by the administration of patient-reported outcome (PRO) questionnaires (EORTC QLQ-C30, EORTC QLQ-CR29 and EuroQol 5-Dimension 5-Level (EQ-5D-5L)) at baseline prior to treatment start, during preoperative treatment at day 15 and at EOT.

Collection of biological samples

Biological samples were centrally collected for prespecified translational analyses. Tumor tissue samples included archival tumor tissue collected from diagnostic colonoscopy (B1 sample), tissue collected on the optional pre-treatment colonoscopy (B2 sample) and tissue collected at surgery (S1 sample). Tissue samples were collected in FFPE tumor tissue blocks.

Fresh-frozen tissue samples were collected with B2 and S1 samples, whenever feasible.

Longitudinal plasma samples and PBMCs were collected at the following timepoints: baseline, day 1, day 15, EOT, before surgery, postoperative (week 4–12 after surgery, before any postoperative treatments, if indicated) and during follow-up visits until disease relapse,

death, consent withdrawal or study closure. Whole blood samples were collected at the baseline timepoint. Stool samples were collected at baseline and at EOT.

Study endpoints

The primary endpoint was centrally assessed pMR rate defined as the percentage of patients, relative to the total of enrolled patients in the intention-to-treat population for each cohort, achieving a pCR or a pMR as per central pathological review.

Surgical samples were analyzed at the Coordinating Center by a blinded senior pathologist with expertise in the field of gastrointestinal pathology (G.S.). The pathologist was blinded to cohort allocation and, therefore, blinded to both treatment regimen and molecular status. Primary tumors and lymph nodes were staged according to the American Joint Committee on Cancer (8th edition). Resected tumors were examined in their entirety, and regression of resected tumors was assessed by estimating the percentage of residual viable tumor of the macroscopically identifiable tumor bed as well as lymph nodes with routine H&E staining. Pathological complete response (pCR) was defined as no residual viable tumor in the tumor bed and lymph nodes; pathological major response (pMR) was defined as ≤10% of residual viable tumor; pathological partial response (pPR) was defined as >10% but ≤50% residual viable tumor; and no response (NR) was defined as >50% residual viable tumor. These cutoffs were chosen drawing from previous pivotal trials with preoperative ICIs in patients with dMMR CRC, such as the NICHE trials[10–12].

Secondary endpoints included:

1. Safety and tolerability
2. The impact of treatment on health-related quality of life, defined by PRO questionnaires EORTC QLQ-C30, EORTC QLQ-CR29 and EQ-5D-5L
3. The characterization of treatment-induced transcriptional dynamics
4. The evaluation of the changes in systemic immunity with longitudinal analysis of PBMCs
5. The correlation of pCR/pMR with ctDNA mutational profiles and its clearance after the short-course preoperative treatment
6. The evaluation of the prognostic and/or predictive impact of radiogenomic or radiomic signatures in specific molecular subgroups from the analysis of the baseline and post-treatment tumor imaging
7. The evaluation of the prognostic and/or predictive value of microbiota

Explorative endpoints included overall survival, defined as the time from treatment start to the date of death due to any cause, and EFS, defined as the time from treatment start to the first occurrence of any of the following events: disease progression that precluded curative surgery, R2 surgery, local or distant recurrence, occurrence of a second primary cancer or death from any cause.

Statistical analysis

For each cohort according to a Fleming single-stage design and selecting the design parameters p_0 (pCR + pMR rate in the null hypothesis—that is, with preoperative chemotherapy such as FOLFOX) = 0.15 and p_1 (pCR + pMR rate in the alternative hypothesis—that is, with the short-course preoperative targeted treatment) = 0.45, considering a β error of 20% and an α error of 5%, a total of 14 patients were planned to be enrolled in each cohort. The null hypothesis (p_0 = 0.15) was defined based on historical data from neoadjuvant chemotherapy studies across colon and rectal cancers. In CC, the FOxTROT trial reported low rates of deep pathological response (approximately 8% pCR + pMR), whereas, in rectal cancer, higher pCR rates of approximately 20–22% were observed with neoadjuvant FOLFOX in the PROSPECT trial. Therefore, a value of 15% was selected as a clinically realistic and conservative estimate of the expected activity of chemotherapy alone across colon and rectal cancers. The null hypothesis would be rejected and the experimental treatment judged promising if at least five pCRs or pMRs were observed in each cohort. All the analyses for the study endpoints were performed for each cohort in the intention-to-treat population, which includes every patient who has been assigned to each cohort. Safety data were summarized on the safety analysis set, which includes all of the patients who have received at least one administration of the study drug. Descriptive statistics were used for all variables, as appropriate. Continuous variables were summarized by the number of observations, median, minimum, maximum and IQR, and categorical variables were summarized by frequency counts and percentages for each category. Median follow-up was estimated with reverse Kaplan–Meier. Statistical analyses for each cohort were performed with R software (v.4.2.0, 22 April 2022).

Health-related quality-of-life analyses

PRO analyses. PROs were assessed by means of the EORTC QLQ-C30, evaluating global health status, functional domains and symptoms burden; the EORTC QLQ-CR29, a site-specific module designed specifically for patients with CRC; and the EQ-5D-5L, which includes a utility index and a visual analog scale (VAS) score. The assessments were conducted on day 1, before treatment start, on day 15 and at EOT.

EORTC scoring guidelines were followed to calculate the mean scores for each item and to manage missing data. The QLQ-C30 is a 30-item questionnaire that has been validated in CRC[43]. It includes five functional domains (physical, role, emotional, cognitive and social functioning), nine symptom domains (fatigue, nausea/vomiting, pain, dyspnea, insomnia, appetite loss, constipation, diarrhea and financial impact) and a global health status/quality of life (GHS/QoL) scale[44]. Scores for the functional and symptom domains are standardized on a 0–100 scale. Higher scores on the functional and GHS/QoL scales reflect better functioning and health-related quality of life, whereas higher symptom scores indicate greater symptom burden[45,46]. The QLQ-CR29 is a validated, 29-item, CRC-specific questionnaire used with the QLQ-C30 that assesses disease-related and treatment-related symptoms and functioning. It measures gastrointestinal and urinary symptoms/functioning, abdominal and pelvic pain, treatment-related side effects, stoma-related concerns, psychosocial domains (body image and anxiety) and sexual functioning. Scores range from 0 to 100, with higher values representing greater symptom burden or worse functioning.

For global quality of life from the QLQ-C30 questionnaire, mean scores with standard deviation were described at days 1 and 15 and EOT and compared using the Wilcoxon signed-rank test. The score changes at day 15 and EOT were categorized as 'improved' for increases of ≥10 points compared to baseline in at least one of the two timepoints, 'worsened' in case of no improvement and a decrease of ≥10 points in at least one of the two assessments and 'stable' in those cases not meeting the overspecified criteria. Finally, mean changes from baseline at each scheduled assessment were used to describe changes for all the other items of the questionnaires.

QLQ-C30 and QLQ-CR29 analyses were conducted in the general population, in each individual cohort, according to molecular profile (pMMR versus dMMR) and according to treatment (BOT versus BOT + BAL).

Single-cell spatial transcriptomics on matched pre-treatment and post-treatment tissue samples

Visium HD data processing. Spatial transcriptomics data were generated using the 10x Genomics Visium HD platform. Libraries were processed with Space Ranger v.3.1.3 using the recommended SLURM configuration. Each web summary report was reviewed, and samples showing suboptimal alignment were manually realigned using Loupe Browser v.8.1.2.

To correct residual misalignment between transcript-derived spatial coordinates and histological morphology, we implemented an in-house registration step within the following workflow. For each sample, the H&E image was read from the multiscale OME-Zarr representation at a downsampled pyramid level, and the hematoxylin component was extracted by RGB-to-HED color deconvolution. In parallel, Visium HD bin positions were converted into spatial geometries using the Space Ranger full-resolution pixel coordinates and scale factors, and each bin was weighted by its total unique molecular identifier (UMI) count. These weighted bin geometries were rasterized into a transcript count/density image in the same coordinate frame as the histology image. The transcript density raster and hematoxylin image were then cropped to the tissue-containing region, intensity normalized and contrast enhanced using adaptive histogram equalization. The transcript density image was additionally Gaussian smoothed and log transformed to reduce sparsity and emphasize broad tissue-level structure. Registration was performed with SimpleITK using the hematoxylin channel as the fixed image and the transcript density raster as the moving image. A centered two-dimensional similarity transform was initialized and optimized with a correlation metric and regular-step gradient descent, yielding a transformation describing residual translation, rotation and isotropic scaling between the molecular and morphological signals. The resulting transform was converted to an affine parameterization and applied directly to the original Visium HD full-resolution bin coordinates. Corrected coordinates were written to a new bin position parquet file and propagated to downstream bin-to-cell reconstruction steps. For quality control, the workflow also exported the raw hematoxylin and transcript density arrays, the fitted transform and a composite OME-TIFF containing the fixed hematoxylin image, the unregistered transcript density image and the registered transcript density image. Cell segmentation was performed using CellViT++ with the SAM-H architecture and the Lizard classifier[47]. Gene counts were aggregated at the single-cell level using a modified bin2cell algorithm[48], where each bin was assigned to the cell mask overlapping more than 80% of its area. To account for cytoplasmic RNA, labels were expanded by two bins. These steps yielded a single-cell resolution count matrix for each sample. Cell segmentation quality was assessed by manual visual quality control.

Visium HD data analysis. Quality control was performed using Scanpy[49] on an initial dataset of 9,939,846 cells segmented within capture areas. Cells with fewer than 50 counts, fewer than 30 detected genes or more than 20% mitochondrial transcripts were removed. Genes with fewer than 10 counts or detected in fewer than 100 cells were excluded, and cells with fewer than 10 expressed genes were filtered out. After filtering, a total of 8,065,561 cells across 112 samples were retained for downstream analyses.

Batch correction was performed using the ResolVI framework[50] in unsupervised mode with patient identity as a batch key. After training, latent representations were extracted, neighborhood graphs ($k = 15$) and Leiden clusters (resolution = 0.5) were computed and uniform manifold approximation and projection (UMAP) embeddings were generated using rapids-singlecell[51]. Cell type annotations were assigned based on cluster marker genes (from Scanpy rankings) and validated by histomorphological inspection (for example, tumor cells in malignant regions, normal epithelial cells in glandular areas and lymphocytes within tertiary lymphoid structures). Representative marker genes used for annotation included *FAP*, *PDPN*, *COL1A1* and *THY1* for activated fibroblasts; *DCN* and *LUM* for fibroblasts; *IL6*, *CXCL12* and *CXCL14* for inflammatory fibroblasts; *ACTA2*, *TAGLN* and *MYL9* for myofibroblasts and smooth muscle cells; *PECAM1*, *VWF* and *CDH5* for endothelial cells; *MS4A1*, *CD79A* and *CD79B* for B cells; *BCL6* and *AICDA* for germinal center B cells; *MZB1*, *SDC1*, *XBP1* and *JCHAIN* for plasma B cells; *CD3D*, *CD3E*, *CD2* and *TRAC* for T cells; *CD68*, *CD163*, *CSF1R* and *MSR1* for macrophages; *S100A8*, *S100A9*, *IL1B* and *CXCL8* for inflammatory myeloid cells; *EPCAM*, *KRT8*, *ALPI* and *FABP1* for enterocytes; *MUC2*, *SPDEF* and *TFF3* for goblet cells; *LYZ*, *DEFA5* and *DEFA6* for Paneth cells; *MKI67*, *TOP2A* and *PCNA* for transit-amplifying cells; *EPCAM*, *KRT8* and *KRT19* for tumor cells; and mitochondrial genes including *MT-** for mito-high tumor cells. Clusters with similar transcriptional profiles were merged. Gene selection for cell type annotation was based on canonical marker genes consistently reported in prior single-cell and spatial transcriptomic studies of CRC and normal intestinal tissue[52,53].

Spatial distances between cells were calculated using the KDTree algorithm from scipy.spatial[54]. Cells located more than 440 µm (4,000 pixels) from any target cell type were excluded. The rationale for this cutoff was to exclude spatially isolated cells unlikely to participate in biologically meaningful local interactions and to minimize edge artifacts arising from sparse tissue regions or segmentation boundaries.

For T_{reg} score calculation, subset data of T cells expression were pseudobulked using decoupleR[55]. Pseudobulk profiles were generated by aggregating raw counts across T cells grouped by sample. Only those with at least 10 cells and 500 counts were kept. Gene expression of the profiles was calculated as the sum of initial gene counts. Score was calculated on normalized, log-scaled values of the following set of genes: *IKZF2*, *FOXP3*, *CCR4*, *IL2RA*, *CTLA4*, *ENTPD1* and *LRRC32*, using sc.tl.score genes (average). Gene selection was based on canonical T_{reg} marker genes consistently reported in prior transcriptomic studies[56].

Tissue compartments were defined using 'shapely' morphological operations. Tumor compartments were defined using spatial coordinates and segmentation masks of malignant epithelial cells identified from spatial transcriptomic data. For each sample, segmentation polygons corresponding to tumor cells were rasterized onto a binary image grid in tissue coordinate space using a pixel size of 20 units (2.2 µm). The resulting binary masks were smoothed using morphological image-processing operations, including binary closing (disk radius = 10 pixels) to connect adjacent tumor regions and fill small gaps, followed by hole filling and removal of small, disconnected regions (<500 pixels). A subsequent binary opening step (disk radius = 2 pixels) was applied to remove small protrusions and boundary irregularities while preserving the overall tumor architecture. Smoothed tumor boundaries were extracted from the processed raster masks and converted back into vector polygons for downstream spatial analyses and visualization. Adjacent stromal compartments were subsequently defined by radial expansion of the tumor compartment boundary (as a 100-µm/909-pixel annulus around the tumor mask). Resulting compartment annotations were visually inspected on the corresponding H&E images to ensure consistency with tissue morphology and invasive tumor margins, with all remaining cells classified as distant stroma. Distance thresholds were selected based on empirical optimization and histomorphological inspection to approximate biologically relevant tumor–stroma interfaces while preserving contiguous tissue architecture.

Pseudobulking and differential expression analysis were performed with decoupler[55] and pydeseq2[57] packages. For differential expression analysis, pseudobulk samples were constructed by aggregating counts across cells belonging to the same sample/cell type/compartment/response depending on the analysis. Only those with at least 10 cells and 500 counts were kept. Local cell density was computed as the number of a given cell type within a sliding circular window of radius 200 µm (1,818 pixels), normalized by window area. Cell–cell contact scores were derived using interlabel connections from spateo[58]. Contact scores were interpreted as normalized probability of neighboring interactions between cell types.

Statistical testing and visualizations. All statistical comparisons were performed using two-tailed Mann–Whitney–Wilcoxon (*U*) tests, with Benjamini–Hochberg correction for multiple comparisons, implemented via the statannotations package.

Data visualization utilized matplotlib[59], Scanpy[49], seaborn, pyCirclize, shapely and geopandas. Final figure layouts were composed in Figma using vector or PNG exports of the plots.

Longitudinal ctDNA tracking

Longitudinal ctDNA analysis was performed on 51 patients using Signatera Genome assay. In brief, germline DNA (gDNA) isolated from FFPE tumor tissue and gDNA isolated from matched whole blood collected in K2 EDTA blood collection tubes were processed for tumor tissue characterization using whole-genome sequencing (WGS). Through the WGS process, tumor-specific somatic variants were identified after the subtraction of germline and clonal hematopoiesis of indeterminate potential (CHIP) variants identified from the matched normal genomic sequence. From these candidate variants, 64+ single-nucleotide variants (SNVs) were selected to detect ctDNA in cell-free DNA (cfDNA) isolated from whole blood (plasma) samples collected using Streck blood collection tubes. A library was generated from plasma cfDNA, wherein the initial starting cfDNA (and the subpopulation of ctDNA, if present) molecules were amplified. An aliquot of this cfDNA library was used as input for the amplification of the 64+ patient-specific targets using multiplex polymerase chain reaction (mPCR). The individual patient mPCR products were barcoded and then pooled with other patient mPCR products, and the pool was sequenced on the Illumina NovaSeq X instrument. Sequenced data that passed the quality control went through the ctDNA calling algorithm. The WGS component of the test was performed once for variant selection, and the plasma assessment was performed routinely to analyze a patient's cfDNA.

PBMC dynamics

Cryopreserved PBMCs were thawed, counted and assessed for viability before being divided among three staining panels—T cell, T_{reg} cell and myeloid-derived suppressor cell (MDSC)—at $1.5–2 \times 10^6$ cells per tube. Cells were washed with staining buffer consisting of 1× PBS supplemented with 2% FBS and 2 mM EDTA. For surface staining, cells were incubated with the appropriate fluorochrome-conjugated antibodies for 30 minutes at 4 °C. For the T_{reg} panel, cells were first incubated with an anti-CCR7 antibody for 15 minutes at 37 °C and subsequently stained with the remaining surface antibodies. After surface staining, intracellular markers were assessed after fixation and permeabilization using the eBioscience FOXP3/Transcription Factor Staining Buffer Set (catalogue number 00-5523-00), according to the manufacturer's instructions, followed by incubation with the corresponding intracellular antibodies. As the MDSC panel did not include intracellular staining, cells were fixed with paraformaldehyde after surface staining and before acquisition. Samples were acquired on a FACSCelesta flow cytometer (BD Biosciences), and data were analyzed using FlowJo v.10 (BD Biosciences). For antibodies used, see Supplementary Table 3. The abundance of PBMC subpopulations at baseline was expressed as the relative fraction within five major leukocyte subtypes ($CD8^+$ lymphocytes, natural killer cells, conventional CD4-2 lymphocytes, T_{reg} lymphocytes and myeloid cells). The variation between the baseline and the subsequent timepoints was expressed as a relative percent change from the baseline abundance. Differences in the enrichment of baseline PBMC subpopulations and their variations across timepoints were modeled with the Wilcoxon signed-rank test. The respective *P* values were adjusted using the Benjamini–Hochberg procedure. Given the exploratory nature of these analyses and the limited sample size, in line with common practice in discovery-oriented cytometry studies (ref. 59), we considered a false discovery rate (FDR) threshold of 0.10 to retain sensitivity for signal detection[60].

Fecal ecosystem analysis

Gut microbiome profiling. *Stool sample collection and handling.* Stool samples were self-collected by patients into DNA/RNA Shield Fecal Collection Tubes (Zymo Research), which stabilize nucleic acids at ambient temperature and prevent microbial community shifts during transport. Samples were stored at −80 °C within 1 hour after collection and shipped on dry ice to a single sequencing facility (PreBiomics), thereby eliminating inter-laboratory technical variability. Baseline samples were defined as those collected ≤10 days after treatment initiation. No patients were on antibiotics during this study.

DNA extraction, library preparation and sequencing. DNA was extracted from stool samples using the DNeasy 96 PowerSoil Pro QIAcube HT Kit (Qiagen), which combines chemical lysis with mechanical disruption by bead-beating on a TissueLyser II (Qiagen). Mechanical bead-beating is critical for efficient lysis of Gram-positive and spore-forming bacteria, which are underrepresented when using chemical-only extraction protocols. DNA quantity was assessed using the Quant-iT 1X dsDNA Assay Kit, High Sensitivity (Life Technologies), and fragment size distribution was evaluated using the D5000 ScreenTape Assay (Agilent Technologies) on a TapeStation 4150 system (Agilent Technologies).

Shotgun metagenomic sequencing libraries were prepared using the Illumina DNA Prep, (M) Tagmentation Kit (Illumina), in combination with DNA/RNA UD Indexes (Illumina), following the manufacturer's instructions. Amplified libraries were purified using the double-sided bead purification procedure described in the Illumina protocol. The final library pool was quantified using the Qubit 1X dsDNA HS Assay Kit (Life Technologies) on a Qubit 3.0 Fluorometer (Life Technologies), and fragment size distribution was reassessed on the TapeStation 4150 system. Library pools were sequenced on an Illumina NovaSeq X Plus platform using paired-end 2 × 150-base pair reads, targeting an average sequencing depth of 15 gigabases per sample. Negative and internal positive controls underwent the same extraction, library preparation and sequencing protocols in parallel with patient samples. Sequencing depth averaged approximately 61.7 million paired-end reads (range, 14.3–97.4 million paired-end reads) across all patient samples included in the analysis. PhiX spike-in was utilized during sequencing as a standard quality check to calculate alignment efficiency and sequencing error rates.

Sequencing quality control and host read removal. Preprocessing and quality control of raw sequencing reads were performed using the SegataLab standalone preprocessing pipeline (https://github.com/SegataLab/preprocessing). Adapter trimming and read-level quality filtering were performed with TrimGalore v.0.6.6 (a wrapper around Cutadapt and FastQC). Reads were discarded if they had a mean Phred quality score below 20, a post-trimming length below 75 base pairs or contained more than five ambiguous nucleotides (N bases). These thresholds ensure that only high-confidence reads contribute to downstream taxonomic profiling and prevent spurious species assignments driven by low-quality sequence data. Human-derived reads were identified and removed by mapping against the GRCh38 (hg38) human reference genome using Kraken v.2.3.4.3., and the proportion of human-derived reads per sample was recorded as a quality metric. This left an average of 58.3 million paired-end reads per sample (range, 5.88–94.8 million paired-end reads) to be analyzed downstream.

Taxonomic profiling. Taxonomic profiling was performed on the host-depleted, quality-controlled reads using MetaPhlAn v.4.1.1 (ref. 61) using default parameters against the CHOCOPhlAnSGB_202103 (Jan21) marker gene database. MetaPhlAn 4 leverages a curated set of clade-specific marker genes derived from more than one million prokaryotic reference and metagenome-assembled genomes, enabling taxonomic classification at the species genome bin (SGB) level. SGBs represent species-level clusters that encompass both characterized (that is, at least one cultivated reference) and previously uncharacterized (that is, lacking cultured representatives) microbial lineages.

All downstream analyses were conducted at the SGB level unless otherwise specified.

Bioinformatics analysis. All bioinformatics analyses were conducted in R v.4.5.2. For analyses requiring feature filtering, and to reduce the influence of spurious low-abundance detections, SGBs were retained only if present at non-zero relative abundance in at least two samples within the relevant analysis subset.

Normalization strategy. MetaPhlAn 4 generates relative abundance profiles that are compositional in nature, with abundances constrained to sum to 100% within each sample. These relative abundance profiles were used directly for beta diversity and differential abundance analyses. No rarefaction (that is, subsampling of reads to a common sequencing depth) was applied because it discards a substantial proportion of sequencing data and has been shown to introduce unnecessary variability and reduce statistical power in comparative microbiome analyses[28].

Beta diversity analysis. Microbial community dissimilarity between samples (beta diversity) was quantified using Bray–Curtis dissimilarity calculated from SGB-level relative abundance profiles, using the vegdist() function from the vegan R package. Bray–Curtis dissimilarity was used because it incorporates both taxon presence/absence and the relative abundance information and is widely applied in compositional microbiome analyses. Principal coordinate analysis (PCoA) was performed on the resulting dissimilarity matrix to visualize inter-sample relationships in reduced-dimensional space.

Between-group differences in overall community composition were evaluated using two complementary non-parametric multivariate tests: permutational multivariate analysis of variance (PERMANOVA), implemented via the adonis2() function in vegan, and analysis of similarities (ANOSIM), each performed with 999 permutations. PERMANOVA partitions the total variance in the dissimilarity matrix among groups and tests significance by permutation; the R^2 statistic is reported as a measure of the proportion of variance explained by group membership. ANOSIM tests whether within-group dissimilarities are smaller than between-group dissimilarities and reports an R statistic (ranging from −1 to +1, where $R = 0$ indicates no compositional difference between groups and $R = \pm 1$ indicates complete separation). To evaluate potential confounding by primary tumor sidedness, sensitivity analyses were performed by restricting comparisons within left-sided and right-sided tumor strata separately.

Longitudinal beta diversity analysis was restricted to patients with at least two longitudinal stool samples available. For each such patient, Bray–Curtis dissimilarity was calculated between the on-treatment sample (sample 2, collected during treatment) and the paired baseline sample (sample 1). Dissimilarity values were converted to normalized Bray–Curtis similarity (1 − dissimilarity) to facilitate interpretation, where a value of 1.0 indicates identical community composition and 0 indicates complete dissimilarity. Within-patient changes in community composition over time were assessed using paired Wilcoxon signed-rank tests within each cohort.

Differential abundance analysis: LEfSe. LEfSe was applied to SGB-level relative abundance profiles to identify microbial taxa discriminating between predefined clinical groups (for example, MSI-H versus MSS tumors; cohort 6 versus cohort 7). Relative abundance profiles were rescaled by a constant factor of 10^6 to express values in parts-per-million units, following the standard LEfSe input convention[29]. This rescaling is a linear transformation of relative abundances and does not alter the compositional structure of the data nor does it constitute rarefaction or subsampling to one million reads.

LEfSe proceeds in three sequential steps: (1) a non-parametric Kruskal–Wallis test identifies features that differ significantly across the primary class of interest; (2) pairwise Wilcoxon rank-sum tests are applied within subclasses to assess consistency of the direction of effect across subgroups; and (3) linear discriminant analysis (LDA) is used to estimate the effect size of each feature that passed the non-parametric filters. LDA effect sizes were estimated using 30 bootstrap iterations, each performed on a random 67% subsample (subsampling fraction $f = 0.67$) of the data. This bootstrapping procedure is an integral part of the standard LEfSe algorithm and serves to improve the stability and reproducibility of LDA effect size estimates by reducing sensitivity to individual samples, but it does not constitute a separate statistical test. The maximum log-transformed LDA score was set to 3.0. Features were considered discriminative if they passed both non-parametric filters and had an absolute LDA score ≥2.0 (corresponding to a 100-fold difference in relative abundance between groups on the $\log_{10}$ scale).

SIG score calculation. SIG scores were calculated for each sample from their respective MetaPhlAn 4 SGB-level relative abundance profiles, as previously described[28]. SIG scores are composite indices derived from the relative abundances of a defined set of co-occurring microbial species identified as ecologically and clinically relevant in the context of ICI response. Comparisons of SIG scores between MMR status groups (MSI-H versus MSS) and across cohorts were performed using the Wilcoxon rank-sum test. For multigroup comparisons (for example, across cohorts 4, 5, 6 and 7), pairwise Wilcoxon rank-sum tests were applied as post hoc contrasts. The baseline relative abundance of *B. uniformis* SGB1836 was analyzed using Wilcoxon rank-sum tests.

List of protocol amendments

The initial study protocol (v.1.0, 3 December 2021) included cohorts 1–7, including all four cohorts reported in this paper (cohorts 4–7). Subsequent substantial amendments expanded the broader UNICORN platform without modifying the design, treatment regimens, sample size, endpoints or prespecified analyses of cohorts 4–7. Protocol v.2.0 (20 May 2023) introduced cohort 8, evaluating sotorasib plus panitumumab in *KRAS* G12C-mutated CRC, and incorporated updated safety information for BOT + BAL, together with changes to investigational medicinal product supply and regulatory documentation. Version 3.0 (21 October 2024) introduced cohorts 9 and 10, evaluating vorbipiprant with or without nivolumab in pMMR/MSS CRC, added participating centers and incorporated updated clinical and safety information. Version 4.0 (10 April 2025) introduced cohorts 11 and 12, evaluating subcutaneous amivantamab in molecularly hyperselected left-sided and right-sided pMMR/MSS CRC as well as UNICORN part 2 cohort 1, evaluating an extended-duration regimen of BOT + BAL. Version 4.1 (17 September 2025) included administrative and other non-substantive protocol refinements. None of these amendments affected the conduct or interpretation of the cohorts presented here.

Reporting summary

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

Data availability

Deidentified individual participant data underlying the results reported in this article, along with a data dictionary and relevant analysis code, will be made available. Requests should be submitted to the corresponding author (at filippo.pietrantonio@istitutotumori.mi.it) and must include a detailed research proposal, analysis plan and evidence of institutional affiliation. Proposals will be reviewed to ensure scientific validity and compliance with ethical and legal requirements. A response will be provided within approximately 8 weeks.

Data will be shared with qualified researchers for non-commercial purposes after approval and execution of a data use agreement and in accordance with applicable data protection regulations.

Study protocol and registration details are publicly available at ClinicalTrials.gov (NCT05845450). The raw paired-end FASTQ

sequencing data generated in this study from 56 CRC tissue samples using the 10x Genomics Visium HD platform have been deposited in the European Genome-phenome Archive (EGA) under accession number EGAD50000002908 (study EGAS50000002020). The data are available under controlled access and can be requested through the Data Access Committee associated with the dataset. Source data are provided with this paper.

Code availability

The code for additional Visium HD data processing is available at https://github.com/sottorivalab/rescope. The code used for the analysis and visualizations is available at https://github.com/sottorivalab/UNICORN_ST_analysis.

The code and data used for ctDNA analyses are available at https://github.com/Natera-TMED/Unicorn.

The code and data used for the fecal ecosystem analyses are available at https://github.com/andrewalmonte/UNICORN.

References

37. Siena, S. et al. Trastuzumab deruxtecan (DS-8201) in patients with HER2-expressing metastatic colorectal cancer (DESTINY-CRC01): a multicentre, open-label, phase 2 trial. *Lancet Oncol.* **22**, 779–789 https://doi.org/10.1016/S1470-2045(21)00086-3 (2021).
38. Raghav, K. et al. Trastuzumab deruxtecan in patients with HER2-positive advanced colorectal cancer (DESTINY-CRC02): primary results from a multicentre, randomised, phase 2 trial. *Lancet Oncol.* **25**, 1147–1162 https://doi.org/10.1016/S1470-2045(24)00380-2 (2024).
39. Ambrosini, M. et al. Immune checkpoint inhibitors for *POLE* or *POLD1* proofreading-deficient metastatic colorectal cancer. *Ann. Oncol.* **35**, 643–655 https://doi.org/10.1016/j.annonc.2024.03.009 (2024).
40. Cremolini, C. et al. Negative hyper-selection of metastatic colorectal cancer patients for anti-EGFR monoclonal antibodies: the PRESSING case–control study. *Ann. Oncol.* **28**, 3009–3014 https://doi.org/10.1093/annonc/mdx546 (2017).
41. Fakih, M. G. et al. Sotorasib plus panitumumab in refractory colorectal cancer with mutated *KRAS* G12C. *N. Engl. J. Med.* **389**, 2125–2139 https://doi.org/10.1056/NEJMoa2308795 (2023).
42. Pietrantonio, F. et al. The prostaglandin EP4 antagonist vorbipiprant combined with PD-1 blockade for refractory microsatellite-stable metastatic colorectal cancer: a phase Ib/IIa trial. *Clin. Cancer Res.* **31**, 649–658 https://doi.org/10.1158/1078-0432.CCR-24-2611 (2025).
43. Cocks, K. et al. Content validity of the EORTC quality of life questionnaire QLQ-C30 for use in cancer. *Eur. J. Cancer* **178**, 128–138 https://doi.org/10.1016/j.ejca.2022.10.026 (2023).
44. Aaronson, N. K. et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. *J. Natl Cancer Inst.* **85**, 365–376 https://doi.org/10.1093/jnci/85.5.365 (1993).
45. Whistance, R. N. et al. Clinical and psychometric validation of the EORTC QLQ-CR29 questionnaire module to assess health-related quality of life in patients with colorectal cancer. *Eur. J. Cancer* **45**, 3017–3026 https://doi.org/10.1016/j.ejca.2009.08.014 (2009).
46. Gujral, S. et al. Assessing quality of life in patients with colorectal cancer: an update of the EORTC quality of life questionnaire. *Eur. J. Cancer* **43**, 1564–1573 https://doi.org/10.1016/j.ejca.2007.04.005 (2007).
47. Hörst, F. et al. CellViT++: energy-efficient and adaptive cell segmentation and classification using foundation models. *Comput. Methods Programs Biomed.* **277**, 109206 https://doi.org/10.1016/j.cmpb.2025.109206 (2026).
48. Polański, K. et al. Bin2cell reconstructs cells from high resolution Visium HD data. *Bioinformatics* **40**, btae546 https://doi.org/10.1093/bioinformatics/btae546 (2024).
49. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 https://doi.org/10.1186/s13059-017-1382-0 (2018).
50. Ergen, C. & Yosef, N. ResolVI: addressing noise and bias in spatial transcriptomics. *Nat. Methods* https://doi.org/10.1038/s41592-026-03212-9 (2026).
51. Nolet, C. et al. Accelerating single-cell genomic analysis with GPUs. Preprint at *bioRxiv* https://doi.org/10.1101/2022.05.26.493607 (2022).
52. Su, A. et al. The single-cell spatial landscape of stage III colorectal cancers. *npj Precis. Oncol.* **9**, 101 https://doi.org/10.1038/s41698-025-00853-5 (2025).
53. Virtanen, P. et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 https://doi.org/10.1038/s41592-019-0686-2 (2020).
54. Badia-i-Mompel, P. et al. decoupleR: ensemble of computational methods to infer biological activities from omics data. *Bioinform. Adv.* **2**, vbac016 https://doi.org/10.1093/bioadv/vbac016 (2022).
55. Bhairavabhotla, R. et al. Transcriptome profiling of human FoxP3$^+$ regulatory T cells. *Hum. Immunol.* **77**, 201–213 https://doi.org/10.1016/j.humimm.2015.12.004 (2016).
56. Muzellec, B., Teleńczuk, M., Cabeli, V. & Andreux, M. PyDESeq2: a python package for bulk RNA-seq differential expression analysis. *Bioinformatics* **39**, btad547 https://doi.org/10.1093/bioinformatics/btad547 (2023).
57. Qiu, X. et al. Spatiotemporal modeling of molecular holograms. *Cell* **187**, 7351–7373 https://doi.org/10.1016/j.cell.2024.10.011 (2024).
58. Hunter, J. D. Matplotlib: a 2D graphics environment. *Comput. Sci. Eng.* **9**, 90–95 https://doi.org/10.1109/MCSE.2007.55 (2007).
59. Weber, L. M., Nowicka, M., Soneson, C. & Robinson, M. D. diffcyt: differential discovery in high-dimensional cytometry via high-resolution clustering. *Commun. Biol.* **2**, 183 https://doi.org/10.1038/s42003-019-0415-5 (2019).
60. Blanco-Míguez, A. et al. Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn 4. *Nat. Biotechnol.* **41**, 1633–1644 https://doi.org/10.1038/s41587-023-01688-w (2023).
61. McMurdie, P. J. & Holmes S. Waste not, want not: why rarefying microbiome data is inadmissible. *PLoS Comput. Biol.* **10**, e1003531 https://doi.org/10.1371/journal.pcbi.1003531 (2014).

Acknowledgements

We thank the patients and their families and G. Abolafio from the Flow Cytometry Service for the acquisition of the samples and K. Winter and A. Romanova for invaluable support to the Computational Biology Research Center at Human Technopole.

Author contributions

F.G. contributed to study conceptualization, patient management, data collection and curation, statistical and translational analyses, data interpretation, figure preparation, literature review and paper writing. M.A. contributed to protocol writing, patient management, clinical data collection, clinical data and quality-of-life analyses, data interpretation and paper revision. A.R. contributed to protocol writing, patient management, clinical data collection and paper revision. A.T. led and performed spatial transcriptomic and computational analyses and contributed to data interpretation and paper revision. F. Palermo served as main study coordinator and contributed to study management, data curation and paper revision. F.B., S.D.D., D.S., S.T., A.P., R.C., A.B., F.V., M. Guaglio, G. Baiocchi, A.S.-B., R.R., C. Cremolini, I. Montroni and S.L. contributed to patient enrollment,

clinical management and paper revision. V.N. and P.M. contributed to data analyses and paper revision. I. Mulas contributed to Visium sample preparation and tissue processing. L.D., G. Brown, C. Chiodoni, M.D.M., S.N., D.N., E. Spickard, A. Tin, R.L., A.J., A.A.A., V.I., C. Ratti and C.V. contributed to translational and/or statistical analyses and paper revision. G.S. contributed to pathological analyses and central pathology review. A.C. contributed to study coordination and data management. L.C., F.M., C. Rossi, L.A., F.C., E.C., S.M.P. and M. Galassi contributed to data interpretation and paper revision. A.M. and E. Sottotetti coordinated biological sample collection, processing and storage. Ö.K. contributed to case report form management and data curation. E.P. contributed to statistical analyses, data interpretation and paper revision. A.S. conceived and supervised the spatial transcriptomic analyses and contributed to data interpretation and paper revision. F. Pietrantonio conceived and designed the study, wrote the protocol, supervised the project, interpreted data and critically revised the paper. All authors reviewed and approved the final version of the paper.

Funding

This study was sponsored and promoted by Fondazione GONO GI. Agenus Inc. provided drug supply and partial financial support. This research was partially funded by Italian Ministry of Health 'Ricerca Corrente' funds. This study was also supported by the Associazione Italiana per la Ricerca contro il Cancro (AIRC) to A.S. (28961). We acknowledge ERC Consolidator Award funding to A.S. (101125077). L.D. is supported by the European Union's Horizon 2020 research and innovation program (grant agreement no. 825410), Horizon Europe (grant agreement no. 101095604, PREVALUNG-EU), RHU program ANR-21-RHUS-0017 IMMUNOLIFE, the ARC (MICROBIONT-PREDICT 2021; Made-IT 2023), the Seerave Foundation, the Association Robert Debré pour la Recherche Médicale, the AACR 2025 Career Development Award in Lung Cancer Research and the Prix Sœurs Lucie et Olga Fradiss award 2025. No funding entity other than Fondazione GONO GI, the study sponsor, had any role in study design or conduct, data analysis or interpretation, or the decision to submit the manuscript for publication.

Competing interests

F.G. reports speaker honoraria and travel accommodations/expenses from Agenus. M.A. reports advisory role activities for Bayer. A.R. reports consulting or advisory roles for MSD and Nordic Bioscience and travel accommodations/expenses from Amgen and Merck Serono. P.M. reports travel accommodations/expenses from Merck Serono. F.B. reports participation in advisory boards for AAA Novartis, Teysuno, Takeda and Merck and personal honoraria as an invited speaker from Eli Lilly, MSD, Bayer, Bristol Myers Squibb, Pierre Fabre, AstraZeneca and Merck. F.M. reports personal honoraria from Pierre Fabre, Incyte and Accademia della Medicina; institutional research grants from Incyte; and travel accommodations/expenses from Daiichi Sankyo/AstraZeneca, Amgen and Pierre Fabre. M.D.M. reports honoraria from AstraZeneca, Boehringer Ingelheim, Janssen, MSD, Novartis, Pfizer, Roche, GlaxoSmithKline, Takeda, Viatris, Eisai and Daiichi Sankyo for consultancy or participation on advisory boards and institutional funding for clinical trials or contracted research from BeiGene, Exelixis, MSD, Pfizer and Roche. A.S.-B. reports participation on advisory boards for Amgen, Bayer, Pierre Fabre, Servier and Takeda. E.T. reports advisory board participation for MSD, Merck Serono, Amgen, AstraZeneca, Daiichi Sankyo, Bristol Myers Squibb, Servier and Takeda. C. Cremolini reports advisory board or consultant roles with AbbVie, Astellas, AstraZeneca, Bayer, Bicara Therapeutics, Bristol Myers Squibb, GlaxoSmtihKline, Johnson & Johnson, Eli Lilly, Merck, Mirati, MSD, Nordic Pharma, Roche, Pfizer, Pierre Fabre, Revolution Medicines, Rottapharm, Shionogi, Takeda and Tempus; invited speaker honoraria from AbbVie, Amgen, Bayer, Merck Serono, MSD, Pierre Fabre, Servier and Takeda; and research grants from Amgen, Merck, Pierre Fabre, Roche, Seagen/Pfizer, Servier and Tempus. I. Montroni reports faculty roles for Olympus; faculty and advisory board participation for Fresenius Kabi; and advisory board participation for Incyte. G. Baiocchi reports relationships with Stryker and B. Braun in the field of fluorescence-guided surgery. S.L. reports personal honoraria as an invited speaker from Amgen, AstraZeneca, Bristol Myers Squibb, Incyte, GlaxoSmithKline, Eli Lilly, Merck Serono, MSD, Pierre Fabre, Roche and Servier and participation on advisory boards for Amgen, Astellas, AstraZeneca, Bayer, BeOne, Bristol Myers Squibb, Daiichi Sankyo, Fosun Pharma, Gilead, GlaxoSmithKline, Incyte, Eli Lilly, Merck Serono, MSD, Nimbus Therapeutics, Rottapharm, Servier and Takeda. F. Pietrantonio reports research funding to the institution from Eli Lilly, Bristol Myers Squibb, Incyte, AstraZeneca, Amgen, Agenus, Rottapharm, Johnson & Johnson, GlaxoSmithKline, Tempus and BeOne; personal honoraria from BeOne, Daiichi Sankyo, Astellas, Ipsen, AstraZeneca, Servier, Bayer, Takeda, Johnson & Johnson, Bristol Myers Squibb, MSD, Amgen, Merck Serono, Pierre Fabre, Incyte, Jazz Pharmaceuticals, Medscape and VJ Oncology; advisory/consultancy roles for Bristol Myers Squibb, MSD, Amgen, Pierre Fabre, Johnson & Johnson, Servier, Bayer, Takeda, Astellas, GlaxoSmithKline, Daiichi Sankyo, Pfizer, BeOne, Jazz Pharmaceuticals, Incyte, Rottapharm, Merck Serono, Italfarmaco, Gilead, AstraZeneca, Agenus, Revolution Medicines, 3T Biosciences, Dolon, Clarivate and Simon Kucher; and travel expenses from Amgen, Merck Serono, Pierre Fabre, Servier, Astellas, Incyte, Johnson & Johnson and Agenus. The other authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41591-026-04710-1.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41591-026-04710-1.

Correspondence and requests for materials should be addressed to Filippo Pietrantonio.

Peer review information *Nature Medicine* thanks Cathy Eng, Niels Halama and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editor: Anna Ranzoni, in collaboration with the *Nature Medicine* team.

Reprints and permissions information is available at www.nature.com/reprints.

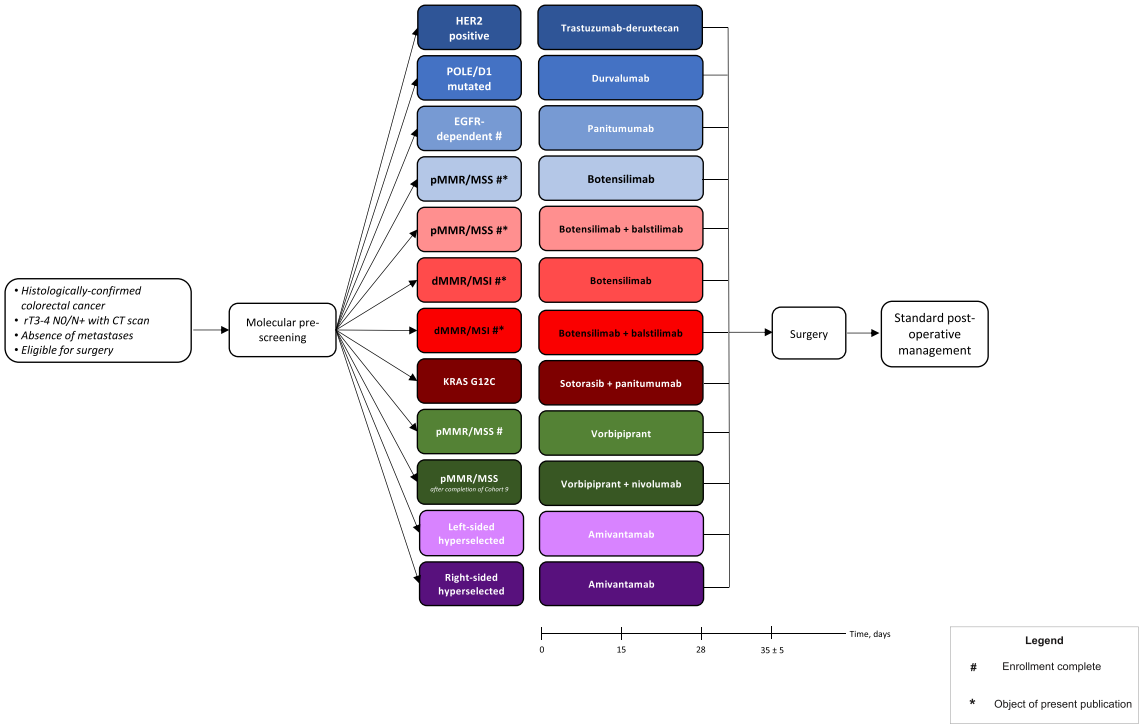
https://doi.org/10.1038/s41591-026-04710-1

Extended Data Table 1 | Centrally assessed baseline radiological staging in each treatment Cohort

	pMMR BOT N=14 (%)	pMMR BOT+BAL N=14 (%)	dMMR BOT N=14 (%)	dMMR BOT+BAL N=14 (%)
rT status (%)				
rT2	3 (21%)	4 (29%)	1 (7%)	5 (36%)
rT3	10 (71%)	10 (71%)	12 (86%)	7 (50%)
rT4	1 (7%)	0 (0%)	1 (7%)	2 (14%)
rN status (%)				
rN0	8 (57%)	9 (64%)	7 (50%)	9 (64%)
rN+	6 (43%)	5 (36%)	7 (50%)	5 (36%)

List of abbreviations: BAL, Balstilimab; BOT, Botensilimab; dMMR, deficient mismatch repair; pMMR, proficient mismatch repair; rT, radiological primary tumor staging; rN, radiological nodal staging.

https://doi.org/10.1038/s41591-026-04710-1



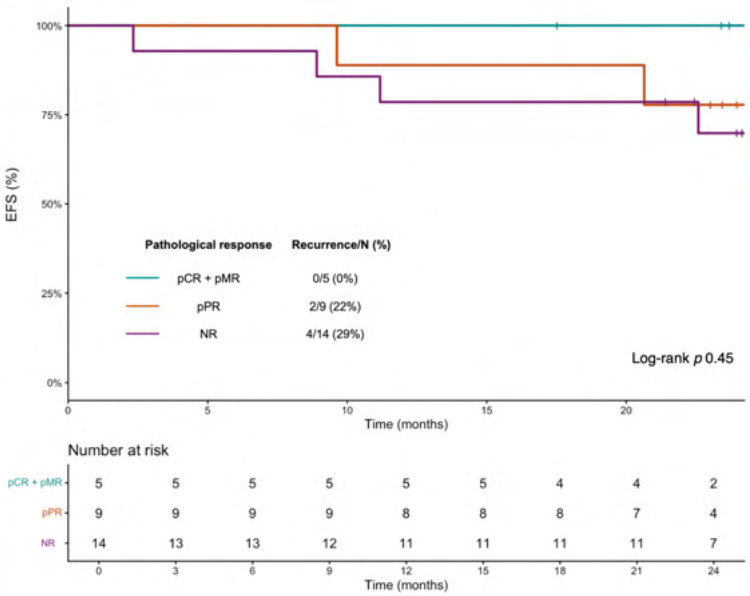
Extended Data Fig. 1 | UNICORN Study Design and biomarker-driven treatment allocation. Study design for the whole UNICORN trial by Fondazione GONO GI (ClinicalTrials.gov: NCT05845450). Molecular pre-screening is centrally conducted at Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy. Upon the identification of a protocol-specified molecular profile and conduction of screening evaluations, patients are enrolled and treated in the specific treatment Cohort. After the 28-day short-course, tailored treatment and radiological reassessment, patients are submitted to curative-intent surgery on day 35 + /- 5 followed by standard, guideline-recommended, multidisciplinarily assessed post-operative management.

https://doi.org/10.1038/s41591-026-04710-1



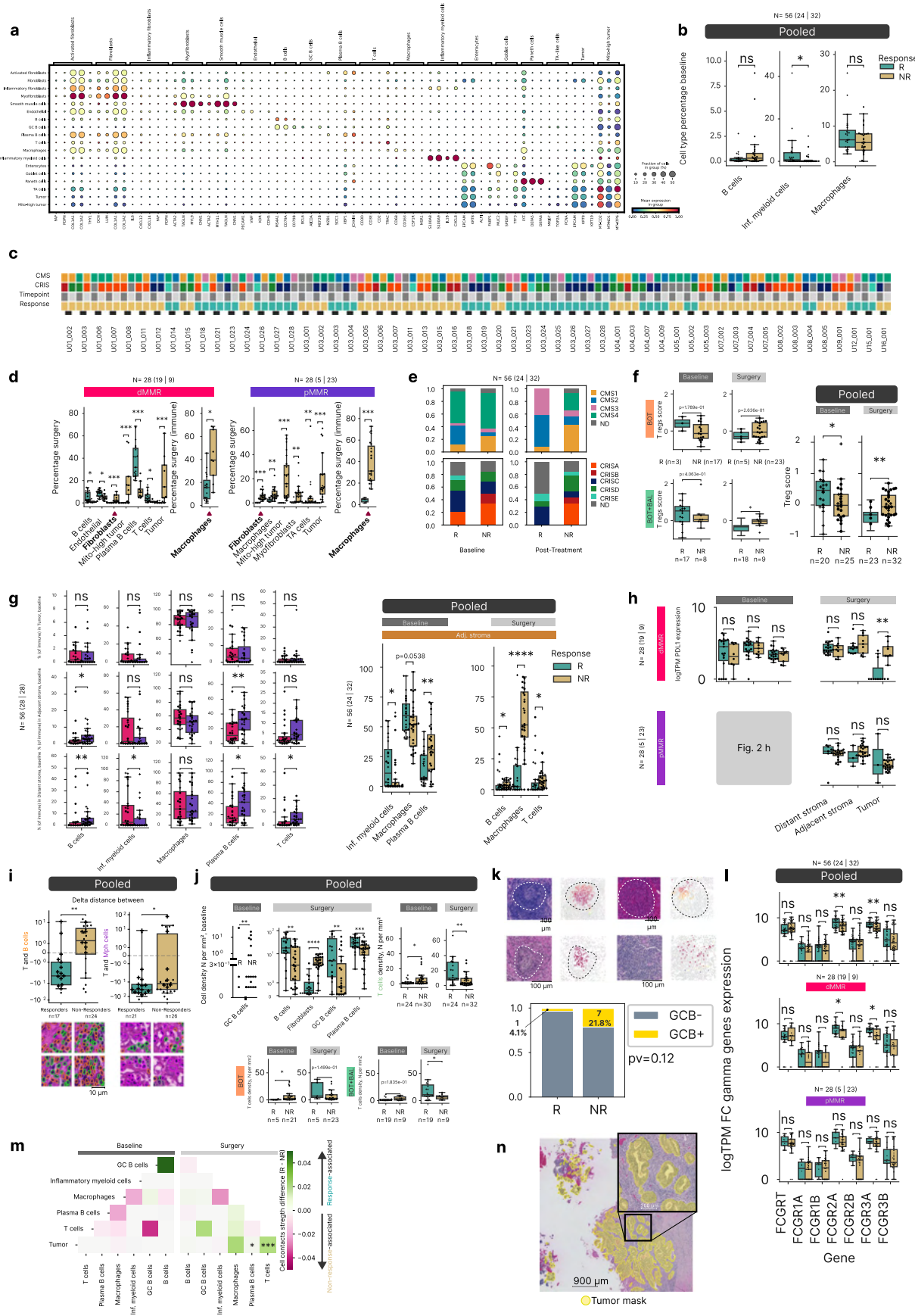
Extended Data Fig. 2 | QLQ-C30 dynamics in the overall population, according to treatment Cohort, molecular status and treatment regimen. **a-d**, Distribution of the frequency of patients, relative to the evaluable populations completing at least two timepoints of the QLQ-C30 assessment, with either improvement, stability or worsening global QOL score in the overall population (**a**), according to treatment Cohort (**b**), molecular profile (**c**) and treatment regimen (**d**).

https://doi.org/10.1038/s41591-026-04710-1



Extended Data Fig. 3 | EFS according to treatment response in patients treated with BOT monotherapy. Kaplan-Meier curves for EFS in patients treated with BOT monotherapy (n = 28) according to pathological response to treatment. Patients meeting the primary objective of pathological complete or major response (pCR + pMR) are represented by the blue line, patients achieving a pathological partial response (pPR) and non-responders (NR) are represented by the orange and purple lines, respectively.

https://doi.org/10.1038/s41591-026-04710-1

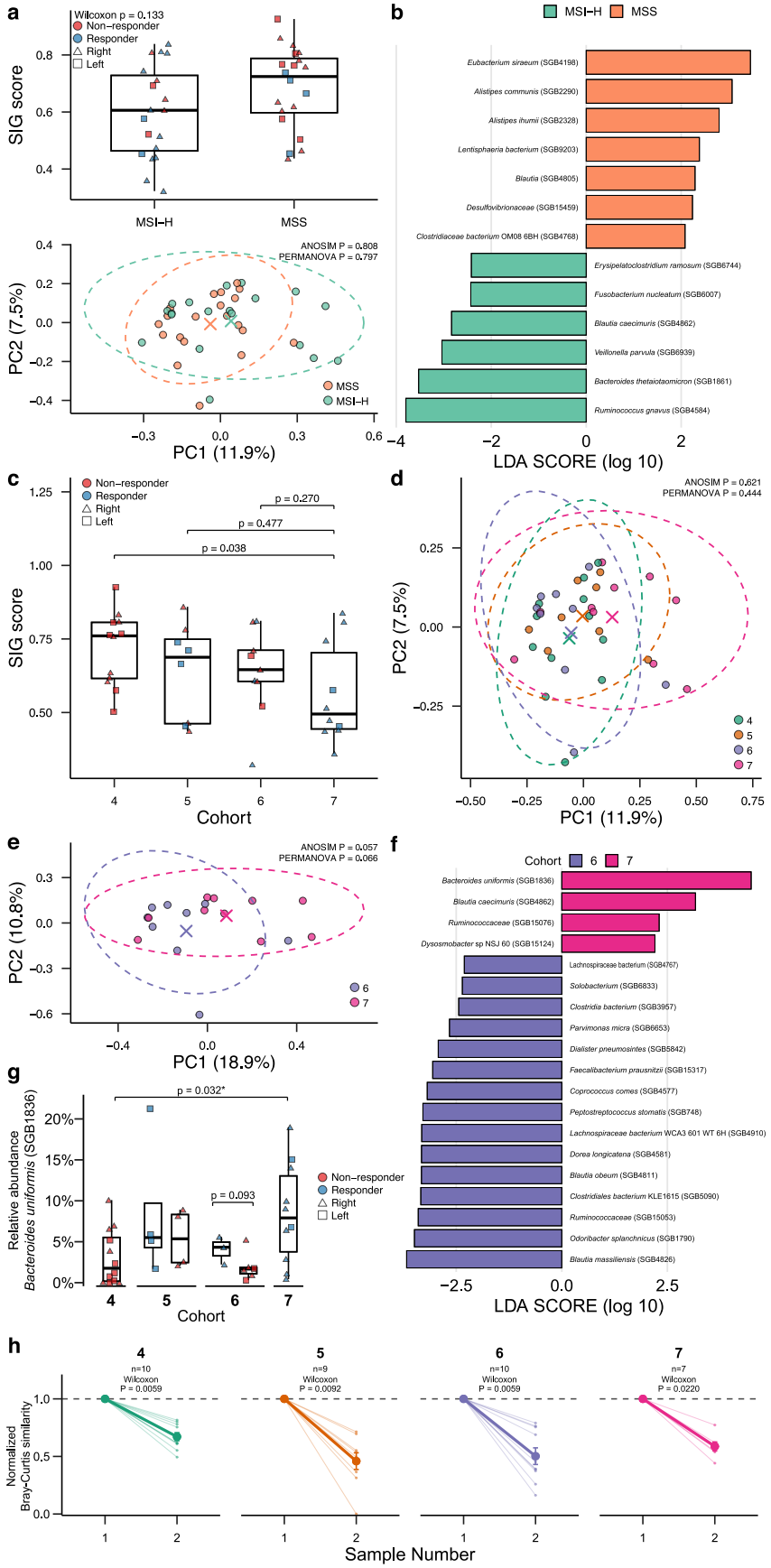


Extended Data Fig. 4 | See next page for caption.

https://doi.org/10.1038/s41591-026-04710-1

Extended Data Fig. 4 | Baseline immune contexture, molecular subtypes and response-associated spatial remodeling. a, Dot plot of canonical marker-gene expression across annotated cell types. Dot size indicates the fraction of cells expressing each gene and color indicates mean normalized expression. **b**, Pooled baseline fractions of B cells, inflammatory myeloid cells and macrophages in responders and non-responders. **c**, Heat map of CMS subtype, CRIS subtype, timepoint and pathological response across paired patient samples. **d**, Post-treatment cell-type fractions at surgery in responders and non-responders, stratified by MMR status. Left panels show fractions among all cells; right panels show macrophage fractions among immune cells. Red triangles mark cell types concordant between molecular subtypes. P values were obtained using two-sided Mann-Whitney-Wilcoxon tests with Benjamini-Hochberg correction. **e**, Proportional distribution of CMS1-4 and CRIS A-E assignments at baseline and after treatment in responders and non-responders; ND, not determined. **f**, Treg gene-set scores at baseline and surgery. Left, analyses stratified by treatment arm; right, pooled analysis. **g**, Left, baseline immune-cell composition in tumor, adjacent stroma and distant stroma, comparing dMMR and pMMR samples. Percentages are relative to all immune cells in the indicated compartment. Right, pooled response-group comparisons of selected adjacent-stromal cell types at baseline and surgery. **h**, PD-L1 expression (logTPM) by response, MMR status, timepoint and tissue compartment. The baseline pMMR comparison is shown in Fig. 2h and is not duplicated here. **i**, Pooled change in T-cell-B-cell and T-cell-macrophage distances from biopsy to surgery by response. Negative values indicate reduced separation. Representative contact maps are shown below; scale bar, 10 µm. **j**, Local cell-density analyses. Top, pooled baseline and surgery comparisons for GC B cells, B cells, fibroblasts, plasma B cells and T cells. Bottom, T-cell-density comparisons stratified by BOT and BOT + BAL treatment. **k**, Representative H&E sections and corresponding spatial maps showing GC B-cell-rich foci within morphologically defined tertiary lymphoid structures. Dashed circles delineate the structures; scale bars, 100 µm. Stacked bars show the baseline frequency of sections with or without GC B-cell-positive foci in responders and non-responders (1 of 24 versus 7 of 32; p = 0.12, two-sided Fisher's exact test). **l**, Baseline Fcγ-receptor-gene expression in a patient-level myeloid pseudobulk comprising inflammatory myeloid cells and macrophages, shown for the pooled cohort and stratified by MMR status. Raw counts from the two myeloid populations were combined within each patient before TPM normalization and log transformation. Each patient therefore contributed one value per gene; **m**, Heat maps showing differences in cell–cell contact strength between responders and non-responders (R − NR) at baseline and surgery. Positive values (green) indicate stronger contacts in responders; negative values (magenta) indicate stronger contacts in non-responders. Asterisks indicate statistically significant differences between response groups. **n**, Representative H&E section with the computationally assigned tumor mask overlaid in yellow; scale bars are shown in the panel. For all quantitative analyses, n denotes biologically independent patients. Each point represents one patient-level value derived from the indicated section; paired analyses include only patients with measurements at both timepoints. Cell-, bin-, neighbourhood- and compartment-level measurements were summarized within each patient section before statistical testing, and no technical replicate was treated as an independent observation. Box plots show the median as the centre line, the 25th and 75th percentiles as the box limits and the most extreme observations within 1.5× the interquartile range as the whiskers. All individual patient values, including values beyond the whiskers, are shown; minima and maxima therefore correspond to the lowest and highest displayed points. Independent responder–non-responder comparisons used two-sided Mann–Whitney U tests. Paired biopsy–surgery comparisons used two-sided Wilcoxon signed-rank tests. Benjamini–Hochberg adjustment was applied only where explicitly stated. Exact n, test statistics, effect estimates, 95% confidence intervals and exact P values are reported for each comparison in the panel and Source Data. R, responder (TRG 1–2); NR, non-responder (TRG 3–5); B, biopsy; S, surgery; Mph, macrophage.

https://doi.org/10.1038/s41591-026-04710-1

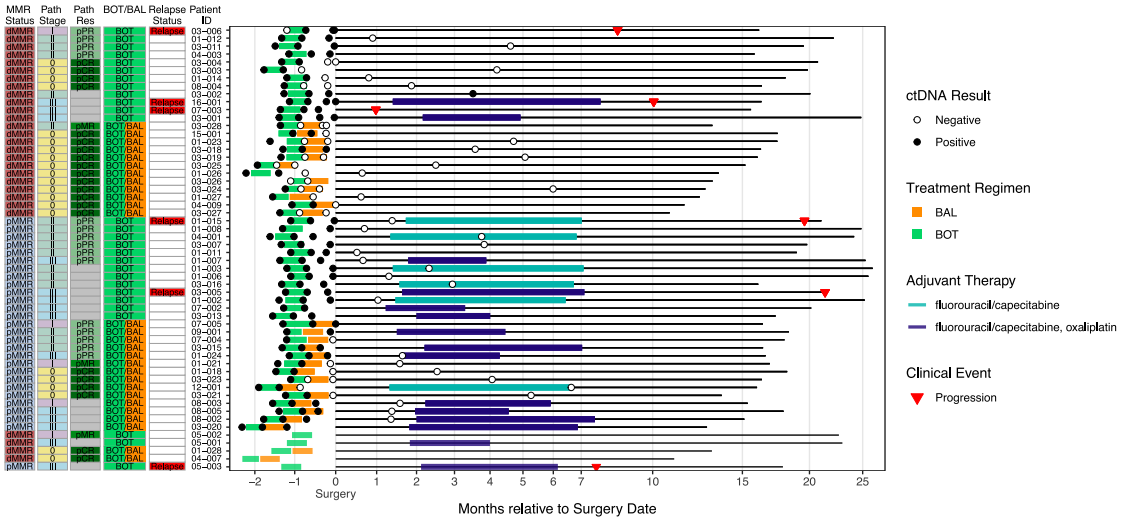


Extended Data Fig. 5 | See next page for caption.

https://doi.org/10.1038/s41591-026-04710-1

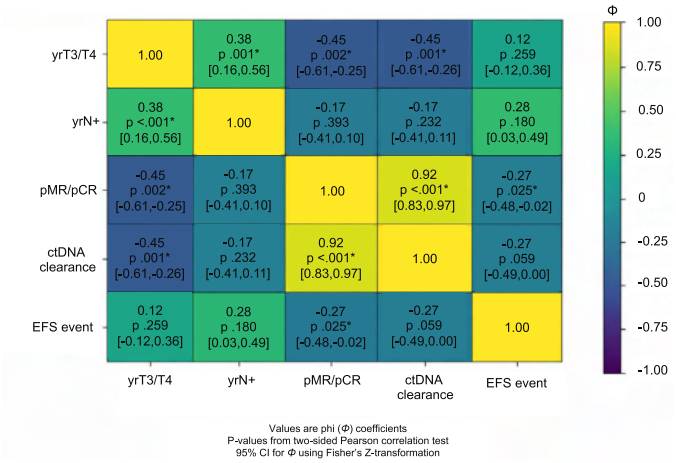
Extended Data Fig. 5 | Fecal ecosystem analysis. a, Baseline species-interacting group (SIG) scores in 19 patients with dMMR/MSI-H tumours and 20 with pMMR/MSS tumours (n = 39 independent biological replicates; one metagenomic profile per patient) (top). Wilcoxon rank-sum test: W = 244, P = 0.133; estimated difference = 0.09 (95% CI, -0.018–0.197; r_{rb} = 0.284). Principal coordinate analysis (PCoA) of Bray-Curtis distances by MMR status is shown below. ANOSIM: R = -0.022, P = 0.808; PERMANOVA: pseudo-F(1,37) = 0.841, R^2 = 0.022, P = 0.797. **b**, Linear discriminant analysis effect size (LEfSe) identifying bacterial species differentially enriched between dMMR and pMMR colorectal cancers at baseline; bars show LDA scores. **c**, Baseline SIG scores across Cohorts 4 to 7. Cohort 7 (n = 10) was compared with Cohort 4 (n = 12), Cohort 5 (n = 8), and Cohort 6 (n = 9) using Wilcoxon rank-sum tests. Cohort 4 versus 7: W = 92, P = 0.038, estimated difference = 0.167 (95% CI, 0.014–0.321; r_{rb} = 0.533); Cohort 5 versus 7: W = 48.5, P = 0.477, estimated difference = 0.049 (95% CI, -0.08–0.283; r_{rb} = 0.213); Cohort 6 versus 7: W = 59, P = 0.27, estimated difference = 0.098 (95% CI, -0.119–0.248; r_{rb} = 0.311). **d**, PCoA according to cohort. ANOSIM: R = -0.015, P = 0.621; PERMANOVA: pseudo-F(3,35) = 1.006, R^2 = 0.079, P = 0.444. **e**, PCoA of Cohorts 6 and 7. ANOSIM: R = 0.114, P = 0.057; PERMANOVA: pseudo-F(1,17) = 1.411, R^2 = 0.077, P = 0.066. **f**, LEfSe analysis comparing baseline samples between Cohorts 6 and 7. **g**, Box plot showing relative abundance of *B. uniformis* SBG1836. Cohort 4 comprised 12 (100%) non-responders; Cohort 5, 4/8 (50%) responders; Cohort 6, 3/9 (33.3%) responders; and Cohort 7, 10 (100%) responders. Cohort 7 responders versus Cohort 4 non-responders: W = 93, P = 0.032, estimated difference = 5.444 (95% CI, 0.354–9.915; r_{rb} = 0.55). Cohort 6 responders versus non-responders: W = 16, P = 0.093, estimated difference = 2.53 (95% CI, -3.012–5.288; r_{rb} = 0.778). For panels **a**, **c**, and **g**, the unit of analysis was the individual patient. Each patient contributed one faecal specimen. Box plots show the median and interquartile range (IQR), with whiskers extending to 1.5 x IQR. Colours indicate response status, and triangles and squares indicate right- and left-sided primary tumours, respectively. **h**, Normalised Bray–Curtis similarity between each patient's first two available chronological faecal samples with usable species profiles. Unlike the baseline SIG analyses, which selected the sample closest to treatment initiation collected no later than day +10, this analysis did not apply a baseline timing restriction. Differences in sample timing and availability of longitudinal pairs explain the different cohort sample sizes between panels c and h (Sample 1, chronologically first available sample; Sample 2, chronologically second available sample) within each cohort. Thin lines connect paired samples from individual patients; large, coloured points and thick lines show cohort means, with error bars indicating ± SEM. Sample 1 is the reference and therefore has a normalised similarity of 1 and SEM of 0. Sample 1 and Sample 2 were compared within each cohort using paired Wilcoxon signed-rank tests. Cohort 4: V = 55, P = 0.006, estimated decrease = 0.324 (95% CI, 0.223–0.416); Cohort 5: V = 45, P = 0.009, estimated decrease = 0.506 (95% CI, 0.376–0.732); Cohort 6: V = 55, P = 0.006, estimated decrease = 0.489 (95% CI, 0.274–0.684); Cohort 7: V = 28, P = 0.022, estimated decrease = 0.418 (95% CI, 0.304–0.51). Paired r_{rb} = 1 in all cohorts. All Wilcoxon tests were two-sided and performed without adjustment for multiple comparisons; degrees of freedom are not defined for these tests. Effect estimates for Wilcoxon comparisons are Hodges–Lehmann location shifts. All PCoA group differences were assessed by ANOSIM and PERMANOVA using 999 permutations without multiple-comparison adjustment. For these upper-tail permutation tests, the conventional one- versus two-sided distinction does not apply, and confidence intervals were not estimated.

https://doi.org/10.1038/s41591-026-04710-1



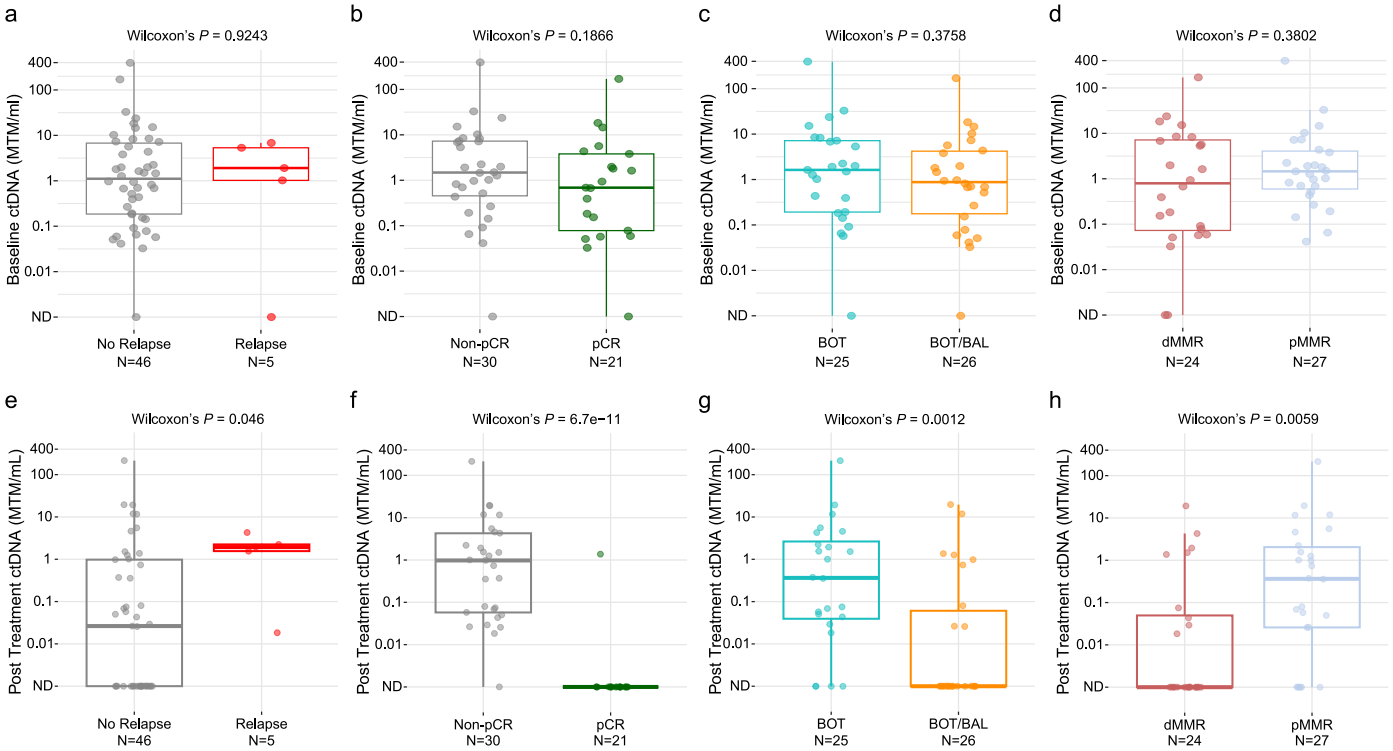
Extended Data Fig. 6 | Swimmer plot of patients with available ctDNA along clinical-pathological annotations. Swimmer plot summarizing individual patient timelines from preoperative treatment initiation through surgery and follow-up. Each horizontal line represents one patient. Columns to the right indicate MMR status, pathological stage, pathological response, treatment regimen, and relapse status. Circles denote ctDNA assessments, with filled symbols indicating ctDNA-positive and open symbols indicating ctDNA-negative results. Red triangles indicate time of disease recurrence. Time is shown in months relative to surgery.

https://doi.org/10.1038/s41591-026-04710-1



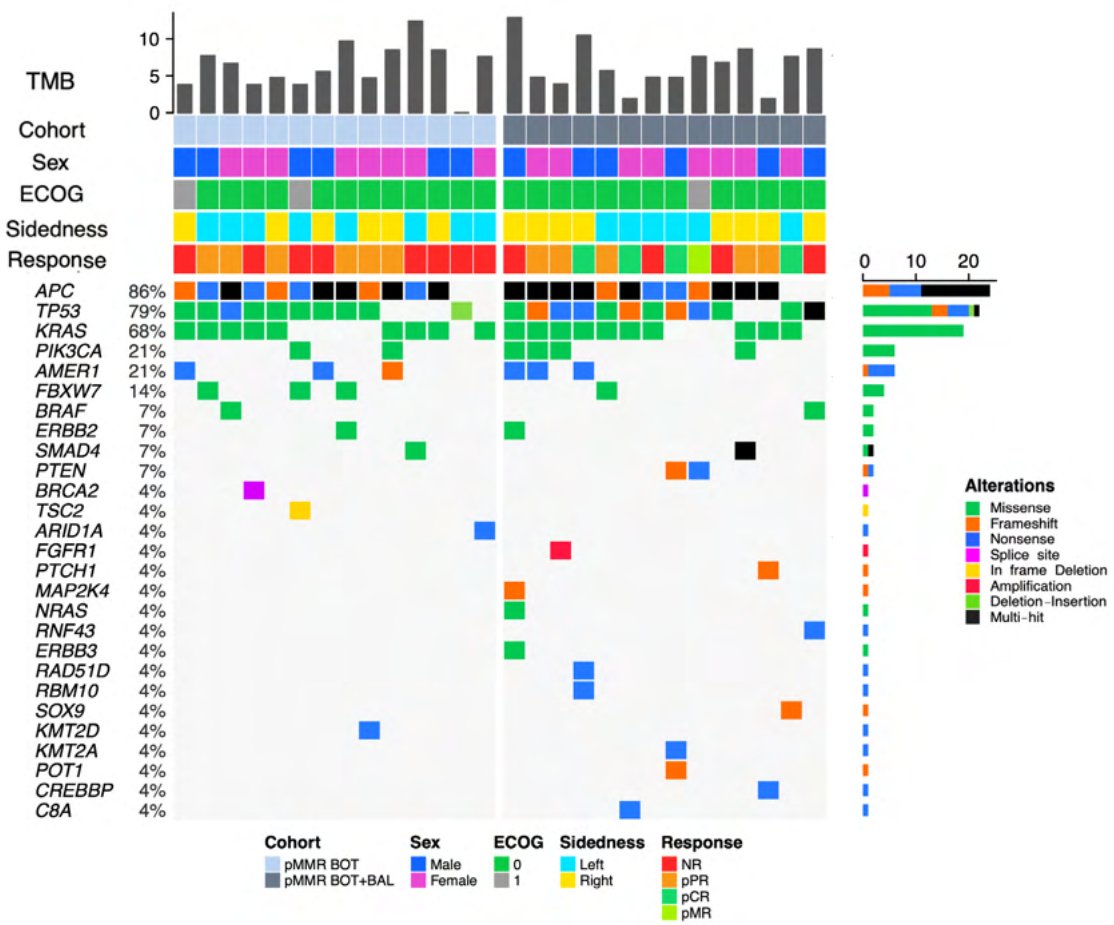
Extended Data Fig. 7 | Correlation analysis of radiological post-treatment staging, pathological response, ctDNA clearance and survival. Heatmap showing pairwise associations between post-treatment radiological staging variables (yrT3/T4 and yrN+), pathological major or complete response, ctDNA clearance and event-free survival (EFS). Correlations were quantified using the phi (φ) coefficient for binary variables. P values were calculated using two-sided Pearson correlation tests for binary variables, and 95% confidence intervals for φ were computed using Fisher's Z-transformation. No adjustment was made for multiple comparisons. Positive φ values indicate a positive association between two categories, whereas negative values indicate an inverse association. Larger absolute φ values indicate stronger associations.

https://doi.org/10.1038/s41591-026-04710-1



Extended Data Fig. 8 | Baseline and post-treatment ctDNA levels according to clinical outcome, pathological response, treatment regimen, and molecular status. Post-treatment ctDNA levels measured as MTM/mL in individual patients. ctDNA levels are shown stratified by disease recurrence status (**a, e**), pathological response (**b, f**), treatment regimen (**c, g**), and molecular status (**d, h**). Panels **a**–**d** and **e**-**h** display ctDNA levels at baseline and EOT, respectively. Each point represents an individual patient. Box plots indicate the median and interquartile range, with whiskers extending to 1.5× the interquartile range. ctDNA values are displayed on a logarithmic scale; samples with no detectable ctDNA are plotted at the lower limit of detection (ND). Statistical comparisons were performed using two-sided Wilcoxon rank-sum tests, with exact *P* values reported in the figure.

https://doi.org/10.1038/s41591-026-04710-1



Extended Data Fig. 9 | Genomic landscape of pMMR CC. Oncoprint showing recurrent genomic alterations in patients with pMMR colon cancer treated across pMMR BOT and pMMR BOT + BAL Cohorts. Each column represents an individual patient and each row represents a recurrently altered gene, ordered by decreasing alteration frequency. The upper bar plot shows tumor mutational burden (TMB) for each patient. Clinical and pathological annotation tracks indicate treatment cohort, sex, ECOG performance status, primary tumor sidedness and pathological response category. Colored boxes within the oncoprint indicate the presence and type of genomic alteration, as specified in the legend 'alterations' on the right side of the figure. The stacked horizontal bar plots on the right report the absolute number of alterations observed for each gene across the cohort, with colors corresponding to the alteration classes shown in the legend. List of abbreviations: BOT, botensilimab; BAL, balstilimab; pMMR, proficient mismatch repair; TMB, tumor mutational burden; pCR, pathological complete response; pMR, pathological major response; pPR, pathological partial response; NR, non-response.

nature portfolio

Corresponding author(s): Filippo Pietrantonio

Last updated by author(s): Aug 21, 2026

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	MyHealth clinical trial data management system, UNICORN Platform version 2.12.0; BD FACSDiva for flow-cytometry acquisition; 10x Genomics Loupe Browser version 8.1.2 for review and realignment of Visium HD data.
Data analysis	R version 4.2.0 for clinical statistical analyses; R version 4.5.2 for microbiome analyses; 10x Genomics Space Ranger version 3.1.3; Loupe Browser version 8.1.2; TrimGalore version 0.6.6; Kraken version 2.3.4.3; MetaPhlAn version 4.1.1; FlowJo version 10.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Deidentified individual participant data underlying the results reported in this Article, along with a data dictionary and relevant analysis code will be made available. Requests should be submitted to the corresponding author (at Filippo.pietrantonio@istitutotumori.mi.it) and must include a detailed research proposal, analysis

plan, and evidence of institutional affiliation. Proposals will be reviewed to ensure scientific validity and compliance with ethical and legal requirements. A response will be provided within approximately 8 weeks.
Data will be shared with qualified researchers for non-commercial purposes following approval and execution of a data use agreement, and in accordance with applicable data protection regulations.
Study protocol and registration details are publicly available at ClinicalTrials.gov (NCT05845450). The raw paired-end FASTQ sequencing data generated in this study from 56 colorectal cancer tissue samples using the 10x Genomics Visium HD platform have been deposited in the European Genome-phenome Archive (EGA) under accession number EGAD50000002908 (Study EGAS50000002020). The data are available under controlled access and can be requested through the Data Access Committee associated with the dataset.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	Sex was recorded as a baseline demographic variable from the medical record. The study included 23 male and 33 female participants. Sex and gender were not used as eligibility or stratification variables and were not considered in the sample-size calculation. Outcomes were not analysed by sex because the trial was not designed or powered for sex-disaggregated efficacy analyses and each cohort included only 14 patients. Sex-disaggregated baseline data are reported in Table 1. Gender identity was not collected. The findings are not intended to apply to a single sex or gender.
Reporting on race, ethnicity, or other socially relevant groupings	Race/ethnicity was collected as a baseline demographic characteristic and is reported descriptively in Table 1. Of the 56 participants, 55 were recorded as Caucasian and 1 as Latin American. The study was not designed or powered for race- or ethnicity-specific analyses. The limited racial and ethnic diversity of the study population restricts the generalizability of the findings.
Population characteristics	The study included 56 adults with resectable, non-metastatic, radiologically staged rT3–T4, N0 or N+, pMMR/MSS or dMMR/MSI-H colon adenocarcinoma enrolled in four molecularly defined treatment cohorts. Each cohort included 14 patients. Median age (IQR) was 69 (56–75), 65 (57–70), 63 (54–72) and 66 (52–71) years in the pMMR BOT, pMMR BOT+BAL, dMMR BOT and dMMR BOT+BAL cohorts, respectively. The study included 23 male and 33 female participants; 55 participants were recorded as Caucasian and 1 as Latin American. Baseline disease characteristics, ECOG performance status, previous diagnoses and study treatments are reported in Table 1 and the Methods.
Recruitment	Eligible patients were prospectively recruited at participating Italian oncology centres after molecular prescreening and assessment of the protocol-defined eligibility criteria. Enrolment occurred from 28 June 2023 to 24 July 2024: cohort 4, 28 June 2023–1 January 2024; cohort 5, 10 January–2 May 2024; cohort 6, 24 July 2023–14 February 2024; and cohort 7, 29 February–24 July 2024. Cohorts were enrolled sequentially as specified in the protocol. All participants provided written informed consent before any study-specific procedure, and no compensation was provided. Potential sources of selection bias include recruitment at specialist centres, molecular prescreening, restriction to patients considered suitable for curative surgery and voluntary consent to participate. The almost exclusively Caucasian study population may further limit generalizability.
Ethics oversight	The study was authorized by the Italian Medicines Agency (AIFA) through the Clinical Trials Information System under EU Regulation No. 536/2014 (EU CT number 2022-501308-90-00; CTIS ID 1351), following favourable assessment of Parts I and II, including approval by the competent Ethics Committee, "Comitato Etico per la sperimentazione clinica delle province di Verona e Rovigo", on 19 April 2023. All participating centres were activated only after completion of the required site-specific authorization procedures. The study was conducted in accordance with the Declaration of Helsinki, ICH-GCP and applicable regulatory requirements. All participants provided written informed consent before any study-specific procedure.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	A Fleming single-stage design was applied independently to each cohort. Assuming a pathological complete or major response rate of 0.15 under the null hypothesis and 0.45 under the alternative hypothesis, with a one-sided type I error rate of 5% and 80% power, 14 patients were required per cohort. The experimental treatment was considered promising if at least 5 of 14 patients achieved a pathological complete or major response. Four cohorts were included in this report, for a total planned sample size of 56 patients.
Data exclusions	No enrolled patient was excluded from the primary intention-to-treat analysis. The intention-to-treat population included all patients assigned to each cohort. The safety analysis set included all patients who received at least one dose of study treatment. For individual translational analyses, only patients with an available and technically evaluable sample were included; the corresponding sample numbers and any assay-specific quality-control exclusions are reported in the Methods and figure legends. No post hoc exclusion was applied on the basis of study outcome.

Replication	Clinical observations were obtained in a prospective multicentre trial using protocol-defined procedures and centrally assessed endpoints. Laboratory and computational analyses incorporated assay-specific technical and biological quality-control procedures, as described in the Methods. Spatial transcriptomic annotations and segmentation outputs were subjected to histomorphological and visual quality control. Flow-cytometry analyses were performed using predefined antibody panels and gating procedures. All planned experimental analyses that passed the prespecified technical quality-control criteria were successfully completed; missing or non-evaluable samples and the resulting sample sizes are reported for each analysis.
Randomization	This was a non-randomized, open-label study. Cohorts 4 and 6 were opened concomitantly, whereas cohorts 5 and 7 were opened sequentially after completion of enrolment in cohorts 4 and 6, respectively. Patients were assigned according to tumour mismatch-repair status and the cohort available for enrolment; no random allocation procedure was used.
Blinding	Participants and treating investigators were not blinded because this was an open-label, non-randomized trial. The central pathological assessment of the primary endpoint was performed by a senior gastrointestinal pathologist who was blinded to cohort allocation and therefore to both treatment regimen and molecular status. Blinding was not applicable to treatment administration and most exploratory laboratory analyses. Where analyses were not blinded, standardized assay procedures, predefined quality-control criteria and computational pipelines were used to reduce subjective assessment.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☐	☒ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	Antibodies used for multicolour flow-cytometry analysis of PBMCs are listed in Supplementary Table 3. For each antibody, the table reports the target antigen, fluorochrome, clone, supplier, catalogue number and staining amount or dilution. Lot numbers and RRIDs are reported where available and applicable.
Validation	All antibodies are validated by the manufacturer for flow cytometry in human samples. In this study, appropriate positive and negative controls confirmed specific staining and FMO staining has been done for selected antibodies.

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	ClinicalTrials.gov: NCT05845450, prospectively registered on 26 April 2023, before enrolment of the first participant. EU Clinical Trial number: 2022-501308-90-00; CTIS ID: 1351.
Study protocol	The original study protocol and statistical analysis plan are provided in the Supplementary Information. The trial registration record is available at https://clinicaltrials.gov/study/NCT05845450. The initial protocol was version 1.0 dated 3 December 2021. Subsequent protocol amendments and their dates are listed in the Methods. None of the amendments modified the design, treatment regimens, sample size, endpoints or prespecified analyses of cohorts 4–7 reported in this Article
Data collection	Clinical, pathological, radiological, patient,-reported outcome and translational data were prospectively collected at participating Italian centres at protocol-defined timepoints between 28 June 2023 and 24 July 2024. Baseline assessment included molecular prescreening, clinical evaluation, serum tumour markers and contrast-enhanced chest–abdomen–pelvis CT performed within 21 days before treatment; pelvic MRI was required for patients with rectal cancer. Tumour tissue, blood, plasma and stool samples were collected at the timepoints specified in the protocol. Surgery and postoperative follow-up assessments were conducted according to the protocol. Central pathological assessment was performed at the Coordinating Centre by a senior gastrointestinal pathologist blinded to cohort allocation. The data cut-off for the analyses reported here was 9 February 2026.
Outcomes	The primary outcome was the centrally assessed pathological complete or major response rate in the intention-to-treat population of each cohort. Pathological complete response was defined as no residual viable tumour in the tumour bed or lymph nodes, and pathological major response as ≤10% residual viable tumour. Resected primary tumours and lymph nodes were reviewed centrally by

a senior gastrointestinal pathologist blinded to cohort allocation. Secondary outcomes included safety and tolerability; health-related quality of life assessed using the EORTC QLQ-C30, EORTC QLQ-CR29 and EQ-5D-5L questionnaires; treatment-induced transcriptional changes; longitudinal PBMC immune changes; associations between pathological response and ctDNA profiles or clearance; radiomic/radiogenomic analyses; and microbiome analyses. Exploratory outcomes included overall survival and event-free survival, as defined in the protocol and Methods.

Plants

Seed stocks: NA

Novel plant genotypes: NA

Authentication: NA

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation: Peripheral blood was collected into four 10-ml K2-EDTA tubes and processed within 1 h of collection. After centrifugation at 1,600g for 10 min at room temperature with the brake disabled, plasma was carefully removed without disturbing the cellular interface. The volume removed was replaced with an equal volume of room-temperature PBS, and the tubes were gently inverted three to four times. Under sterile conditions, 3 ml of Ficoll-Paque Plus was added to each of four 12-ml Leucosep tubes and centrifuged at 1,000g for 2 min to allow the Ficoll to pass below the filter. The plasma-depleted blood was then gently layered into the Leucosep tubes and centrifuged at 800g for 15 min at room temperature without braking. The PBMC-containing interface was collected, pooling the material from two Leucosep tubes into each of two 15-ml tubes. Cells were washed twice with 10 ml PBS by centrifugation at 1,500 rpm for 10 min at 4 °C. Residual erythrocytes were lysed by resuspending each PBMC pellet in 1 ml of 1× lysis buffer and incubating for 10 min, followed by a final wash with 10 ml PBS under the same centrifugation conditions. Each pellet was subsequently resuspended in 1 ml FBS and 1 ml freezing mixture containing 80% FBS and 20% DMSO, resulting in a final cryopreservation medium containing 90% FBS and 10% DMSO. The cell suspensions were divided equally into four cryovials, placed in a CoolCell controlled-rate freezing container overnight at −80 °C and subsequently transferred to cryoboxes for storage at −80 °C until shipment to the coordinating centre.

Instrument: Samples were acquired using a Celesta BD Cytometer

Software: FlowJo version 10 (BD Biosciences)

Cell population abundance: The abundance of peripheral blood mononuclear cells (PBMCs) subpopulations at baseline was expressed as the relative fraction within 5 major leukocytes subtype (CD8+ lymphocytes, NK cells, conventional CD4 2 lymphocytes, T regulatory (Treg) lymphocytes and Myeloid Cells).

Gating strategy: Gating strategy is presented in Supplementary Figure 14

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.