nature medicine

Article

https://doi.org/10.1038/s41591-024-03249-3

Neoadjuvant nivolumab or nivolumab plus ipilimumab in early-stage triple-negative breast cancer: a phase 2 adaptive trial

Received: 2 February 2024

Accepted: 14 August 2024

Published online: 16 September 2024

Check for updates

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

Immune checkpoint inhibition (ICI) with chemotherapy is now the standard of care for stage II–III triple-negative breast cancer; however, it is largely unknown for which patients ICI without chemotherapy could be an option and what the benefit of combination ICI could be. The adaptive BELLINI trial explored whether short combination ICI induces immune activation (primary end point, twofold increase in $CD8^+$ T cells or *IFNG*), providing a rationale for neoadjuvant ICI without chemotherapy. Here, in window-of-opportunity cohorts A (4 weeks of anti-PD-1) and B (4 weeks of anti-PD-1 + anti-CTLA4), we observed immune activation in 53% (8 of 15) and 60% (9 of 15) of patients, respectively. High levels of tumor-infiltrating lymphocytes correlated with response. Single-cell RNA sequencing revealed that higher pretreatment tumor-reactive $CD8^+$ T cells, follicular helper T cells and shorter distances between tumor and $CD8^+$ T cells correlated with response. Higher levels of regulatory T cells after treatment were associated with nonresponse. Based on these data, we opened cohort C for patients with high levels of tumor-infiltrating lymphocytes (≥50%) who received 6 weeks of neoadjuvant anti-PD-1 + anti-CTLA4 followed by surgery (primary end point, pathological complete response). Overall, 53% (8 of 15) of patients had a major pathological response (<10% viable tumor) at resection, with 33% (5 of 15) having a pathological complete response. All cohorts met Simon's two-stage threshold for expansion to stage II. We observed grade ≥3 adverse events for 17% of patients and a high rate (57%) of immune-mediated endocrinopathies. In conclusion, neoadjuvant immunotherapy without chemotherapy demonstrates potential efficacy and warrants further investigation in patients with early triple-negative breast cancer. ClinicalTrials.gov registration: NCT03815890.

The addition of anti-PD-(L)1 to neoadjuvant chemotherapy has changed the treatment landscape for patients with early (stage II–III) triple-negative breast cancer (TNBC)[1]; however, all trials evaluating the efficacy of anti-PD-(L)1 in TNBC combined it with chemotherapy[1–4]. This chemotherapy backbone inevitably results in a high rate of adverse events (AEs), affects quality of life and could diminish T cell activity[5,6].

So far, no biomarkers have been established to predict which patients with early-stage TNBC will benefit from anti-PD-1. Therapy is currently given for a total duration of 1 year, although data in other tumor types have shown that a pathological complete response (pCR) can be reached after only a few weeks of treatment with immune checkpoint inhibition (ICI)[7–11]. Overtreatment prevention is an increasingly important consideration due to the high number of patients needed to treat to prevent one recurrence and increasing toxicity with more intense and longer treatments. Therefore, there is an urgent clinical need to optimize treatment schedules and improve patient selection for specific treatments[12].

e-mail: m.kok@nki.nl

While numerous studies have integrated anti-PD-(L)1 therapy with chemotherapy in early-stage TNBC[1–3,13], data on combination ICIs are limited. ICIs targeting CTLA4 have revolutionized treatment for non-small cell lung cancer[8] and melanoma[14–16]. Additionally, neoadjuvant trials across various tumor types have shown impressive major pathological response (MPR) rates when combining anti-PD-(L)1 with low-dose anti-CTLA4 (refs. 7,8,10,17). A trial in metastatic breast cancer revealed long-lasting responses after combining low-dose anti-CTLA4 with anti-PD-1 (ref. 18), which are infrequently observed with anti-PD-(L)1 alone. These findings provide a rationale to test low-dose anti-CTLA4 in combination with anti-PD-(L)1 in early TNBC.

Simultaneously with the advent of ICI, tumor-infiltrating lymphocytes (TILs) have emerged as a putative prognostic and predictive biomarker[19–22]. Patients with TNBC with high TIL levels have an excellent prognosis even without chemotherapy[19,23], suggesting that TILs reflect an endogenous antitumor T cell response. Moreover, in metastatic TNBC, high TIL levels are associated with response to ICI[24,25]. Collectively, these findings imply that TILs may serve as a tool for identifying patients with TNBC who are more likely to benefit from ICI and have a favorable prognosis, paving the way for exploring chemotherapy de-escalation. The BELLINI trial is an adaptive platform trial exploring the effect of ICI without chemotherapy starting with window-of-opportunity (WOO) cohorts with a biological end point followed by neoadjuvant cohorts with a pCR end point. This adaptive platform trial consists of sequential, single-cohort, phase 2 studies, where new cohorts can be opened based on signals obtained in previous cohorts. The first two cohorts evaluated whether 4 weeks of nivolumab (anti-PD-1, cohort A) or nivolumab and low-dose ipilimumab (anti-PD-1 and anti-CTLA4, cohort B) can lead to immune activation (primary end point). This 4-week therapy regimen was scheduled before the start of regular therapy and therefore the effect of ICI could be assessed independently of chemotherapy. Promising results in cohorts A and B among patients with high TIL levels (≥50%) led to the initiation of cohort C. In cohort C, we used a neoadjuvant design with 6 weeks of nivolumab plus low-dose ipilimumab followed by surgery to assess the pCR rate[14,26].

This trial combines anti-PD-1 with anti-CTLA4 in early breast cancer and explores what pCR rate could be achieved with ICI-only approaches and using TIL levels as an entry criterion to enrich for inflamed tumors.

Results

Design and patient characteristics

The BELLINI trial (ClinicalTrials.gov registration: NCT03815890; Fig. 1a,g and Extended Data Fig. 1a) is a preoperative, WOO, phase 2, multiple-cohort nonrandomized study in early (stage I–III) breast cancer utilizing an adaptive Simon's two-stage design[27]. Here, we report the initial results from the first two WOO cohorts exploring the immune-activating capacity of short-term neoadjuvant nivolumab ± ipilimumab (cohorts A and B, $n = 31$) in patients with ≥5% TILs as well as the initial results of cohort C that was opened based on the results of cohorts A and B. The first patient was included on 19 September 2019 and the last patient on 24 January 2023.

Cohort A ($n = 15$) received two cycles of nivolumab (240 mg) on days 1 and 15. Cohort B ($n = 15$) received two cycles of nivolumab (240 mg) on days 1 and 15, plus one cycle of ipilimumab (1 mg kg^{-1}) on day 1. To exclude patients with a poor prognosis, less likely to respond to ICI and not suitable for chemotherapy de-escalation, we enrolled patients with ≥5% TILs in cohorts A and B. Baseline characteristics were similar between cohorts A and B, except for a higher proportion of patients with positive lymph nodes in cohort B (Table 1).

The primary end point for cohorts A and B was immune activation, defined as at least a twofold increase in $CD8^+$ T cells (measured by immunohistochemistry (IHC); Extended Data Fig. 1b–f) and/or increased interferon-γ (*IFNG*) gene expression. This end point was based on the observation that increases in intratumoral $CD8^+$ T cells[25,28] and higher *IFNG* signature scores[17,29] in serially biopsied tumors are correlated with responses to anti-PD-(L)1.

Clinical response (secondary end point) in cohorts A and B was defined as partial response (PR)/complete response on magnetic resonance imaging (MRI) (RECIST v.1.1) or no viable tumor in post-treatment biopsy for patients proceeding to neoadjuvant chemotherapy. For patients directly proceeding to surgery, this was defined as PR or pCR (European Society of Mastology; EUSOMA). Other secondary end points included safety and translational analyses. MRI scans and biopsies were collected at baseline and after two ICI cycles.

Efficacy of short-term nivolumab and nivolumab + ipilimumab in early TNBC (WOO)

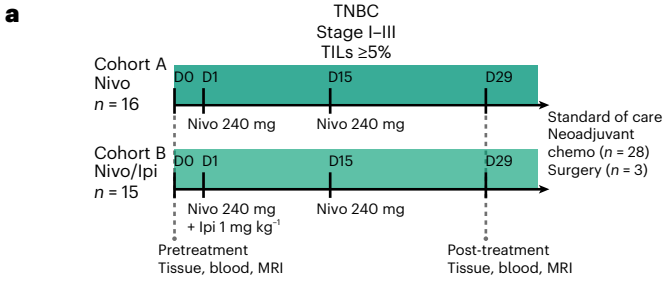
Immune activation was achieved in eight tumors (53.3%) in the nivolumab cohort (A) and nine (60%) in the nivolumab + ipilimumab cohort (B) (Fig. 1b). Therefore, both cohorts met the Simon's two-stage[27] threshold for expansion to stage II. After 4 weeks, patients proceeded to standard neoadjuvant chemotherapy followed by surgery ($n = 28$) or surgery without neoadjuvant chemotherapy ($n = 3$). Clinical response was observed in 12 of 31 patients (38.7%, 95% CI 23.7–56.2%) with 7 of 31 patients (22.6%, 95% CI 11.4–39.8%) having a PR according to RECIST v.1.1 criteria[30] (Fig. 1c,d). Ten of 31 patients had no viable tumor in the biopsy and in the three patients who underwent surgery directly after ICI, two PRs and one pCR was seen. Despite these clear pathological responses, MRI showed modest downsizing, indicating MRI underestimates early ICI response (Extended Data Fig. 1h), consistent with findings in early-stage melanoma[31], colorectal and gastroesophageal cancers[17,32]. Notably, clinical response was only observed for patients with TILs ≥ 30% (Fig. 1e) and a combined positive score (CPS) PD-L1 ≥ 20% (Fig. 1f). Patients with lower pretreatment $CD8^+$ T cell levels were more likely to achieve immune activation (Extended Data Fig. 1g), likely due to either less possibility for value doubling or to a very early immune response in highly inflamed tumors.

Fig. 1 | BELLINI trial design, efficacy data and baseline biomarkers. a, Trial design for cohorts A and B. Cohort A received two cycles of nivolumab (anti-PD-1). Cohort B received two cycles of nivolumab (anti-PD-1) and one cycle of ipilimumab (anti-CTLA4). Biopsies and blood were taken pretreatment and after 4 weeks of treatment after which patients proceeded to standard-of-care neoadjuvant chemotherapy ($n = 28$) or primary surgery ($n = 3$). CR, complete response; WES, whole-exome sequencing. **b**, Numbers of patients reaching immune activation in cohorts A ($n = 15$) and B ($n = 15$). **c**,**d**, Changes in tumor size according to the MRI for cohort A (**c**) and cohort B (**d**). The dashed line at −30% indicates radiological PR. The green bars indicate clinical responses (radiological PR and/or pathological response). Asterisks (*) represent patients with resection after ICI only ($n = 3$). pPR, pathological PR according to EUSOMA; SLD, sum of length diameters. **e**, TILs in pretreatment biopsies of patients with and without clinical response in cohorts A and B. $n = 31$ patients. **f**, Combined positive PD-L1 score (CPS) in pretreatment biopsies of patients with and without clinical response in cohorts A and B. $n = 31$ patients. **g**, BELLINI trial design for cohort C. Cohort C ($n = 15$) received two cycles of nivolumab and ipilimumab on days 1 and 21. Biopsies and blood were taken pretreatment and after 6 weeks. Patients proceeded to primary surgery ($n = 15$). **h**, pCR and MPR (<10% viable tumor left) rates in cohort C. NR, nonresponse. **i**, Changes in tumor size according to the MRI in cohort C. The dashed line at −30% indicates radiological PR. Dark blue bars show pCR. **j**, TILs in pretreatment biopsies of patients according to pCR status in cohort C. $n = 15$ patients. **k**, CPS in pretreatment biopsies for patients according to pCR status in cohort C. $n = 15$ patients. Panels **a**,**g** were created with BioRender.com. Levels of TILs calculated as average from TIL levels at diagnostic and pretreatment study (**e**,**j**). Boxplots display minimum (Q0), maximum (Q4), median (Q2) and IQR (**e**,**f**,**j**,**k**). *P* values were derived using a two-sided Mann–Whitney test.

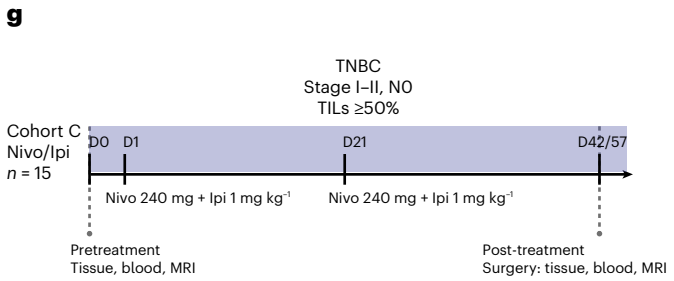
https://doi.org/10.1038/s41591-024-03249-3

Short-term neoadjuvant nivolumab + ipilimumab can induce pathological responses in patients with high TIL levels

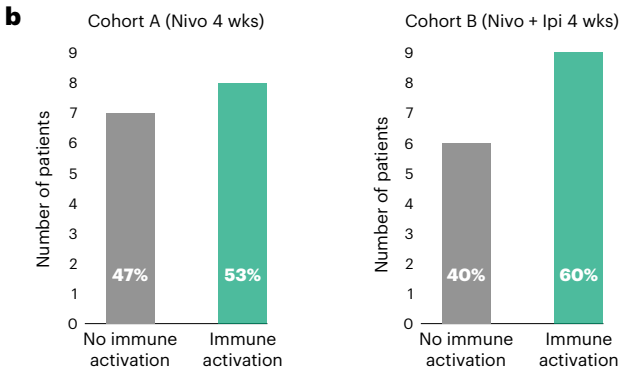
Both cohorts A and B met the predefined thresholds of the Simon's two-stage design[27], allowing expansion to stage II; however, given the promising clinical responses observed in cohorts A and B and the approval of neoadjuvant pembrolizumab plus chemotherapy[4], the study team decided not to proceed to stage II with the WOO design but to open cohort C with a true neoadjuvant design ($n = 15$; Fig. 1g). As all patients with a clinical response in cohorts A and B had high TIL levels, cohort C was opened for patients with ≥50% TILs and allowed

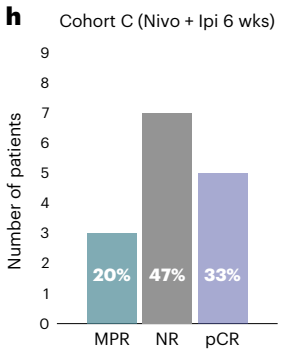


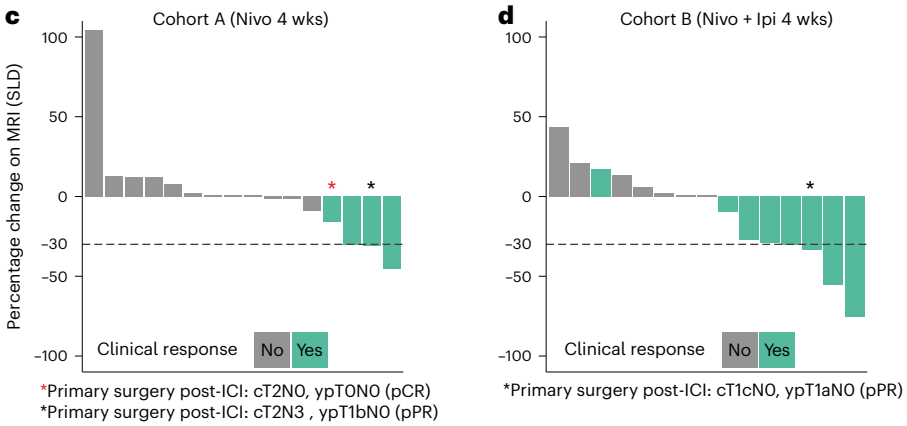
Primary end point	Secondary end points	Translational techniques
Immune activation = twofold increase in - $CD8^+$ T cells (IHC) and/or - *IFNG* expression (RNA)	Safety Clinical response = Biopsy post-ICI ($n = 28$) no viable tumor or PR/CR on MRI (RECIST 1.1) Surgery post-ICI ($n = 3$) partial or complete response (EUSOMA)	TILs, spatial analysis Bulk RNA-seq, WES Single-cell RNA-seq ctDNA Flow cytometry



Primary end point	Secondary end points	Translational techniques
pCR	Safety MPR Response on MRI (RECIST 1.1)	TILs, spatial analysis Bulk RNA-seq ctDNA

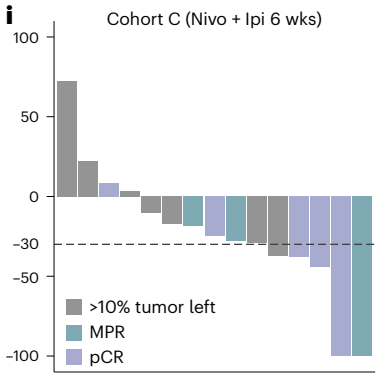


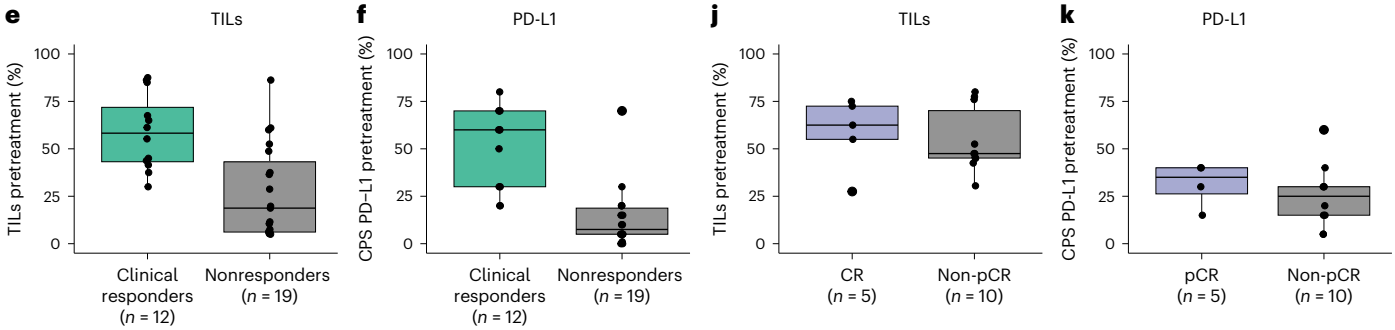




*Primary surgery post-ICI: cT2N0, ypT0N0 (pCR)
*Primary surgery post-ICI: cT2N3 , ypT1bN0 (pPR)

*Primary surgery post-ICI: cT1cN0, ypT1aN0 (pPR)





https://doi.org/10.1038/s41591-024-03249-3

Table 1 | Baseline patient characteristics

Characteristic	A: Nivo (*n* = 16)	B: Nivo-Ipi 4 weeks (*n* = 15)	C: Nivo-Ipi 6 weeks (*n*=15)
Median age, years (IQR)	48 (39.8–53.2)	50 (42.5–57.5)	51 (36.0–56.5)
WHO PS, *n* (%)			
0	16 (100)	14 (93.3)	15 (100)
1	0 (0.0)	1 (6.7)	(0.0)
Histological subtype, *n* (%)			
NST	16 (100)	13 (86.7)	14 (93.3)
Metaplastic	0 (0.0)	1 (6.7)	0 (0.0)
Lobular pleiomorphic	0 (0.0)	1 (6.7)	1 (6.7)
Tumor stage, *n* (%)			
T1	5 (31.3)	5 (33.3)	2 (13.3)
T2	10 (62.5)	9 (60.0)	13 (86.7)
T3	1 (6.2)	1 (6.7)	0 (0.0)
Nodal status, *n* (%)			
N0	13 (81.3)	5 (33.3)	15(100)[c]
N1	2 (12.5)	9 (60.0)	0 (0.0)
N3	1 (6.3)	1 (6.7)	0 (0.0)
Tumor grade[a], *n* (%)			
2	1 (6.3)	4 (26.7)	0 (0.0)
3	15 (93.8)	11 (73.3)	15 (100)
Germline BRCA1/2 mutation, *n* (%)			
Yes	3 (18.8)	3 (20.0)	4 (26.7)
No	12 (75.0)	10 (66.7)	11 (0.0)
Unknown	1 (6.3)	2 (13.3)	0 (0.0)
TILs[b], (%)			
Median (IQR)	40.8 (6.2–60.3)	37.5 (23.8–61.4)	52.5 (45.3–73.8)

[a]Tumor grade according to Bloom Richardson. [b]TILs were averaged between the diagnostic TILs score and the study pretreatment TILs score. sTILs were scored according to international guidelines[22] as a numerical variable. All samples were evaluated by at least two breast cancer pathologists and their score for each sample was averaged. [c]Cohort C allowed the inclusion of only N0 patients. Nivo, nivolumab; Ipi, ipilimumab; WHO PS, World Health Organization performance status; NST, no special type; sTILs, stromal tumor-infiltrating lymphocytes.

Table 2 | Summary of adverse events

	A, Nivo (*n*=16)		B, Nivo+Ipi 4 weeks (*n*=15)		C, Nivo+Ipi 6 weeks (*n*=15)	
	Number of patients (%)					
Immune-mediated AEs	Any grade	Grade ≥3	Any grade	Grade ≥3	Any grade	Grade ≥3
Hypothyroidism[a]	6 (38)	0 (0)	7 (47)	0 (0)	6 (40)	0 (0)
Adrenal insufficiency[b]	1 (6)	0 (0)	2 (13)	1 (7)	3 (20)	1 (7)
Diabetes mellitus	0 (0)	0 (0)	1 (7)	1 (7)	0 (0)	0 (0)
Colitis	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (7)
Hepatitis[c]	0 (0)	0 (0)	2 (13)	0 (0)	3 (20)	3 (20)
Polymyalgia rheumatica	0 (0)	0 (0)	0 (0)	0 (0)	1 (7)	0 (0)
Pneumonitis	0 (0)	0 (0)	0 (0)	0 (0)	2 (13)	1 (7)

This table sums all immune-mediated AEs that required treatment with steroids or did not resolve (endocrinopathies). A detailed list of all AEs according to CTCAE criteria can be found in Extended Data Table 1. [a]All patients are still dependent on hormone replacement therapy. [b]All patients were classified as having secondary adrenal insufficiencies and all patients remain dependent on corticosteroid replacement. [c]We included all patients requiring steroids and one patient with grade 3 IR hepatitis who did not receive steroid treatment.

only patients with node-negative disease, as for this patient population chemotherapy de-escalation could be an option in the future. The treatment schedule with combination ICI for cohort C was based on our data obtained in cohorts A and B as well as on the well-established, effective and tolerable combination ICI schedule in melanoma[14,26].

Patients in cohort C underwent a 6-week treatment regimen of nivolumab and ipilimumab (administered on days 1 and 21), followed by surgery (Fig. 1g). Five patients had a pCR (33.3%, 95% CI 15.2–58.3%; Fig. 1h) with confirmed tumor-negative lymph nodes (ypT0N0). Less than 10% viable tumor remaining was seen in 3 of 15 patients (20%, 95% CI 4–48%; Fig. 1h), resulting in a total MPR rate of 8/15 patients (53%, 95% CI 27–79%). Notably, of the five patients with a pCR only one had a complete radiological response (Fig. 1i). Because of high TILs, N0 status and pCR, which are all very favorable prognostic features, all five patients with a pCR were offered the option of omitting adjuvant chemotherapy and all chose not to undergo adjuvant chemotherapy (shared decision). Patients without pCR were advised to have adjuvant chemotherapy.

Safety data and follow-up

Toxicity data are summarized in Table 2 (all events required steroids or persisted) and detailed in Extended Data Table 1. Neither neoadjuvant nivolumab nor nivolumab + ipilimumab resulted in previously unreported toxicities. All patients were monitored for (immune-related; IR) toxicities until 1 year after ICI therapy. Treatment-related AEs of any grade occurred in 41 of 46 patients (89%). A total of eight (17%) patients developed grade ≥ 3 treatment-related AEs, of which six were treated in cohort C. Except for the endocrinopathies, all AEs resolved. Notably, 19 of 46 patients (41%) developed treatment-related hypothyroidism. All patients with hypothyroidism remain dependent on replacement therapy. Six patients (13%) developed adrenal insufficiency and require ongoing corticoid replacement therapy. One patient developed a diabetic ketoacidosis and remains dependent on insulin.

All patients proceeded with tumor resection or neoadjuvant chemotherapy as scheduled. A total of 44 patients received both ICI doses and 2 patients only received one dose due to suspected immunotoxicity.

With a median follow-up duration of 32.5 months in cohorts A and B (interquartile range (IQR) 28.1–40.3 months), one patient in cohort A (cT2N0; intermediate TILs) developed a second primary tumor and one patient in cohort B (cT2N1; intermediate TILs) died from metastatic TNBC despite receiving standard-of-care (neo)adjuvant chemotherapy. The median follow-up for cohort C was 17.6 months (IQR 18.8–22.1 months). One patient (no response to ICI) refused adjuvant chemotherapy and radiotherapy and developed recurrent TNBC (pT1cNx, 80% TILs).

Pretreatment composition of the tumor microenvironment is associated with ICI response

Due to limited sample size, we compared clinical responders versus nonresponders from both cohorts (A and B) combined and not for the cohorts separately. Clinical responders in cohorts A and B had higher pretreatment TILs ($P = 0.0014$; Fig. 1e) and PD-L1 scores ($P = 8.6 \times 10^{-5}$; Fig. 1f) compared to nonresponders. $CD8^+$ T cell density was not associated with clinical response (Fig. 2a and Extended Data Fig. 1b–f). Spatial analysis showed that responders had shorter distances from tumor cells to the nearest $CD8^+$ T cells ($P = 0.00001$; Fig. 2b). Responders also exhibited a larger density of double-positive $CD8^+PD\text{-}1^+$ cells ($P = 0.02$; Extended Data Fig. 2a) and $PD\text{-}1^+$ cells ($P = 0.001$, IHC; Extended Data Fig. 2b) before treatment.

In cohort C, TILs were not different between responders and nonresponders, probably due to the more homogeneous patient population with only patients with ≥50% TILs (Fig. 1j). In line with this, patients

https://doi.org/10.1038/s41591-024-03249-3

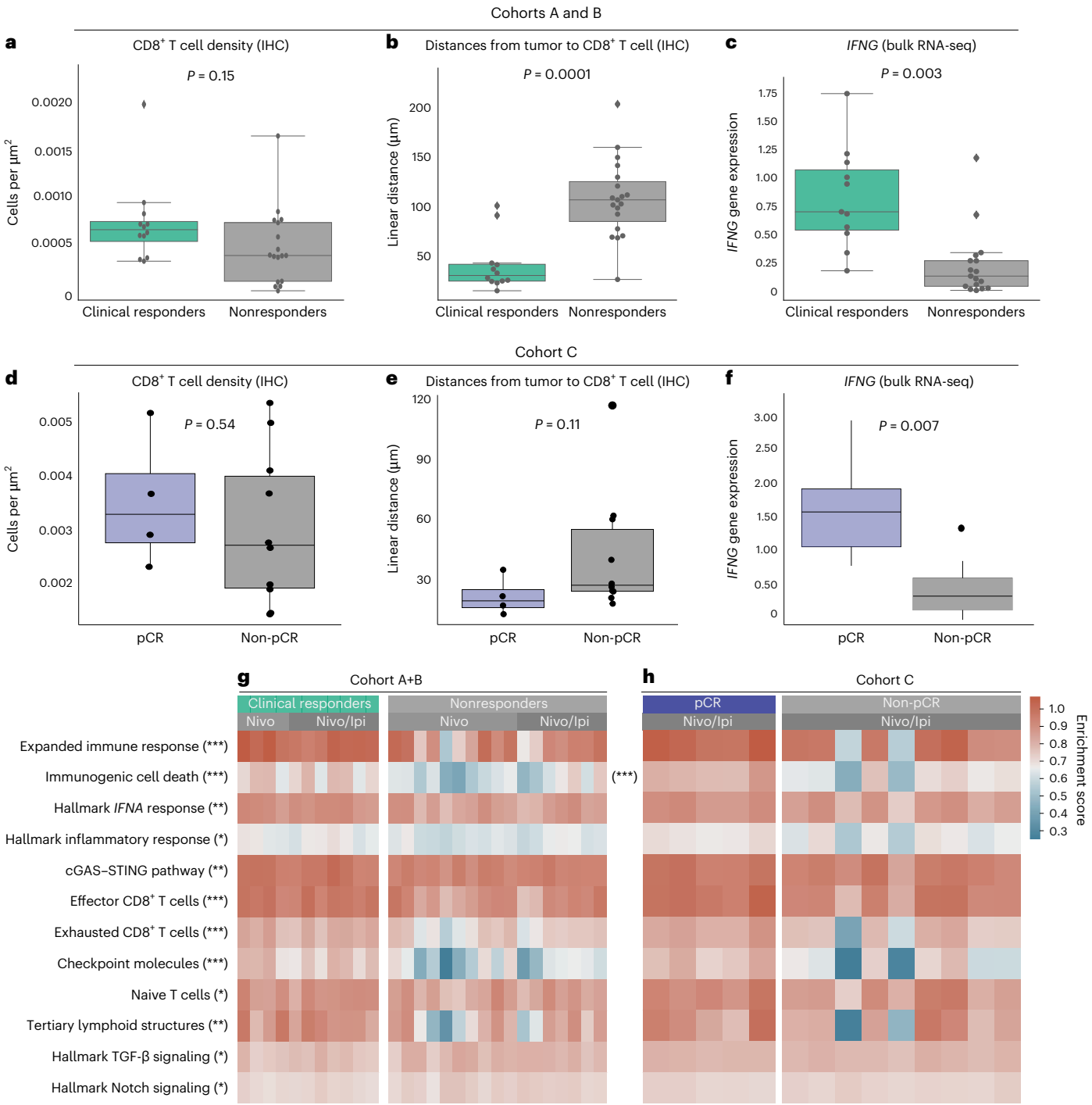


Fig. 2 | Pretreatment immune activation associated with clinical response. **a**, CD8$^+$ density (IHC) in pretreatment biopsies of patients with and without clinical response in cohorts A and B. $n = 31$ patients. **b**, Median distances (µm) from tumor cells to the nearest CD8$^+$ T cells in pretreatment biopsies of patients with and without clinical response in cohorts A and B. $n = 31$ patients. **c**, *IFNG* gene expression scores in pretreatment biopsies of patients with and without clinical response in cohorts A and B. $n = 28$ patients. **d**, CD8$^+$ density (IHC) in pretreatment biopsies of patients with and without pCR in cohort C. $n = 14$ patients. **e**, Median distances from tumor cells to the nearest CD8$^+$ T cells in pretreatment biopsies of patients with and without pCR in cohort C. $n = 14$ patients. **f**, *IFNG* gene expression scores in pretreatment biopsies of patients with and without pCR in cohort C. $n = 14$ patients. **g**,**h**, Gene set enrichment expression scores in pretreatment biopsies of patients with and without clinical response in cohorts A and B ($n = 28$ patients (**g**)) or pCR ($n = 14$ patients (**h**)) in cohort C. Heatmaps include Expanded immune signature[56], Immunogenic cell death signature[57], Hallmark *IFNA* response gene set, Hallmark inflammatory response gene set, cGAS–STING pathway gene set[58], Effector CD8$^+$ T cell gene set[59], Exhausted T cell gene set[59], Checkpoint molecules gene set[59], Naive T cell gene set[60], Tertiary lymphoid structures gene set[61], Hallmark TGF-β signaling gene set, Hallmark Notch signaling. Asterisks represent the *P* values. $^*P \le 0.05$, $^{**}P \le 0.01$, $^{***}P \le 0.001$. Reported *P* values were significant after Benjamini–Hochberg (FDR) correction at 10% significance level. Boxplots display minimum (Q0), maximum (Q4), median (Q2) and IQR (**a**–**f**). *P* values were derived using a two-sided Mann–Whitney test.

with pCR had similar PD-L1 scores, $CD8^+$ T cell density (cells per μm^2) and distances from tumor to nearest $CD8^+$ T cells as patients without pCR (Figs. 1k and 2d,e).

We found no association between tumor mutational burden and clinical response (Extended Data Fig. 2c,d). There were no statistically significant differences between clinical responders and nonresponders in TNBC subtypes[33] (Extended Data Fig. 2e).

Tumors of clinical responders harbor pre-existing inflammatory profiles and tumor-specific $CD8^+$ T cells

We conducted in-depth analyses between clinical responders and nonresponders using bulk RNA-seq (all cohorts) and single-cell RNA sequencing (scRNA-seq) and TCR sequencing (cohorts A and B) pre- and post-treatment. Bulk RNA-seq revealed higher pretreatment levels of *IFNG* gene expression ($P = 0.0003$; Fig. 2c) and inflammatory gene signatures in clinical responders ($P < 0.05$ for all, false discovery rate (FDR) 10%; Fig. 2g and Extended Data Fig. 3a–e). Clinical responders also exhibited higher gene signatures associated with immune infiltration ($P < 0.05$ for all, FDR 10%; Fig. 2g and Extended Data Fig. 3f–j). Conversely, clinical nonresponders displayed upregulation of TGF-β and Notch signaling ($P < 0.05$ for both, FDR 10%; Fig. 2d and Extended Data Fig. 3l–m). Though TIL levels and distances from tumor cells to $CD8^+$ T cells were not different in responders versus nonresponders in cohort C that included TIL high patients only, patients with pCR had significantly higher pretreatment *IFNG* gene expression (Fig. 2f) and higher scores of gene signatures related to immune response and T cell infiltration (Fig. 2h), consistent with our previous observations of a more inflammatory profile of the tumor microenvironment in clinical responders in cohorts A and B.

After scRNA-seq data preprocessing, we obtained 80,000 high-quality T cells from 52 samples (29 patients). Following unsupervised clustering of the T cells, we identified various subpopulations (Fig. 3a–d and Extended Data Fig. 4a–t), including $CD8^+$ effector T cells, $CD8^+$ tissue resident memory ($CD8^+$ T_{RM}) T cells, proliferating $CD8^+$ T cells, naive $CD4^+$ T cells, follicular helper T (T_{FH}) cells, memory $CD4^+$ T cells, regulatory T (T_{reg}) cells, $CD56^{bright}$ and $CD56^{dim}$ natural killer cells. Notably, we identified a cluster of $CD8^+$ T cells with features of tumor-specific T cells. This cluster was characterized by the highest clonality and highest expression of tumor recognition signatures derived using functional tumor recognition experiments[34,35] (Fig. 3c,d). This $CD8^+$ tumor-specific cluster was marked by high expression of tumor-reactive markers (*CD39*, *CD103* and *PDCD1*), *IFNG*, effector molecules (*GZMB*, *NKG7*, *PRF1* and *GNLY*), chemokines (*CCL5*, *CCL4*, *CXCL13* and *CCL3*) and exhaustion markers (*LAG3*, *HAVCR2*, *TIGIT*, *TOX* and *CTLA4*; Fig. 3c,d). Clinical responders exhibited higher fractions of pretreatment $CD8^+$ tumor-specific T cells (Fig. 3e). Clinical responders also had higher fractions of $CD4^+$ T_{FH} cells (Fig. 3f). The presence of tumor-specific $CD8^+$ T cells and T_{FH} in pretreatment biopsies was correlated with tumor decrease on MRI, indicating a continuous association between the abundance of these cells before treatment and the depth of the tumor response (Extended Data Fig. 4u,v). Patients with different TIL levels had similar T cell subtypes before treatment (Extended Data Fig. 4w).

Flow cytometry of blood samples (19 markers; Extended Data Table 2 and Extended Data Fig. 5a) revealed increased Ki-67^+ cells within the PD-1^+ conventional $CD4^+$ T cell population in clinical responders ($P = 0.005$; Fig. 3g). A similar trend was observed for $CD8^+$ T cells (Fig. 3h). The increased proliferation of PD-1^+CD4^+ T cells observed in the blood could also be traced back to the tumor, with responders having higher levels of Ki-67^+ T_{FH}, which was the $CD4^+$ T cell cluster with the highest PDCD1 gene expression in the tumor scRNA-seq data (Fig. 3i,l). In line with the blood data, the levels of PD-1^+ proliferating $CD8^+$ T cells were not significantly different between clinical responders and nonresponders (Fig. 3j,k), suggesting a specific role for proliferating $CD4^+$ T cells systemically as well as in the tumor microenvironment (TME).

Dynamics and post-treatment composition of the tumor microenvironment are distinct in clinical responders and nonresponders

Single-cell RNA-seq analysis revealed that although the clinical responders had higher proportions of tumor-specific $CD8^+$ T cells before treatment, after treatment their tumors included higher levels of effector $CD8^+$ T cells compared to nonresponders ($P = 0.008$; Fig. 4a,b). This suggests that effector $CD8^+$ T cells contribute to ICI-induced tumor regression and underscore the ongoing antitumor $CD8^+$ T cell response, even 4 weeks after treatment initiation.

Conversely, nonresponders had elevated memory $CD4^+$ T cells ($P = 0.05$; Fig. 4a,c) and T_{reg} cells ($P = 0.02$; Fig. 4a,d) post-treatment, potentially suggesting the involvement of T_{reg} cells in mediating resistance to ICI, consistent with previous studies[36]. Notably, we observed an association between the fraction of T_{reg} cells after treatment and the lack of response or in some patients even increase in tumor volume on MRI (Fig. 4e). This correlation was specifically mediated by activated (CD137$^+$) T_{reg} cells, rather than nonactivated T_{reg} cells (Extended Data Fig. 5b,c).

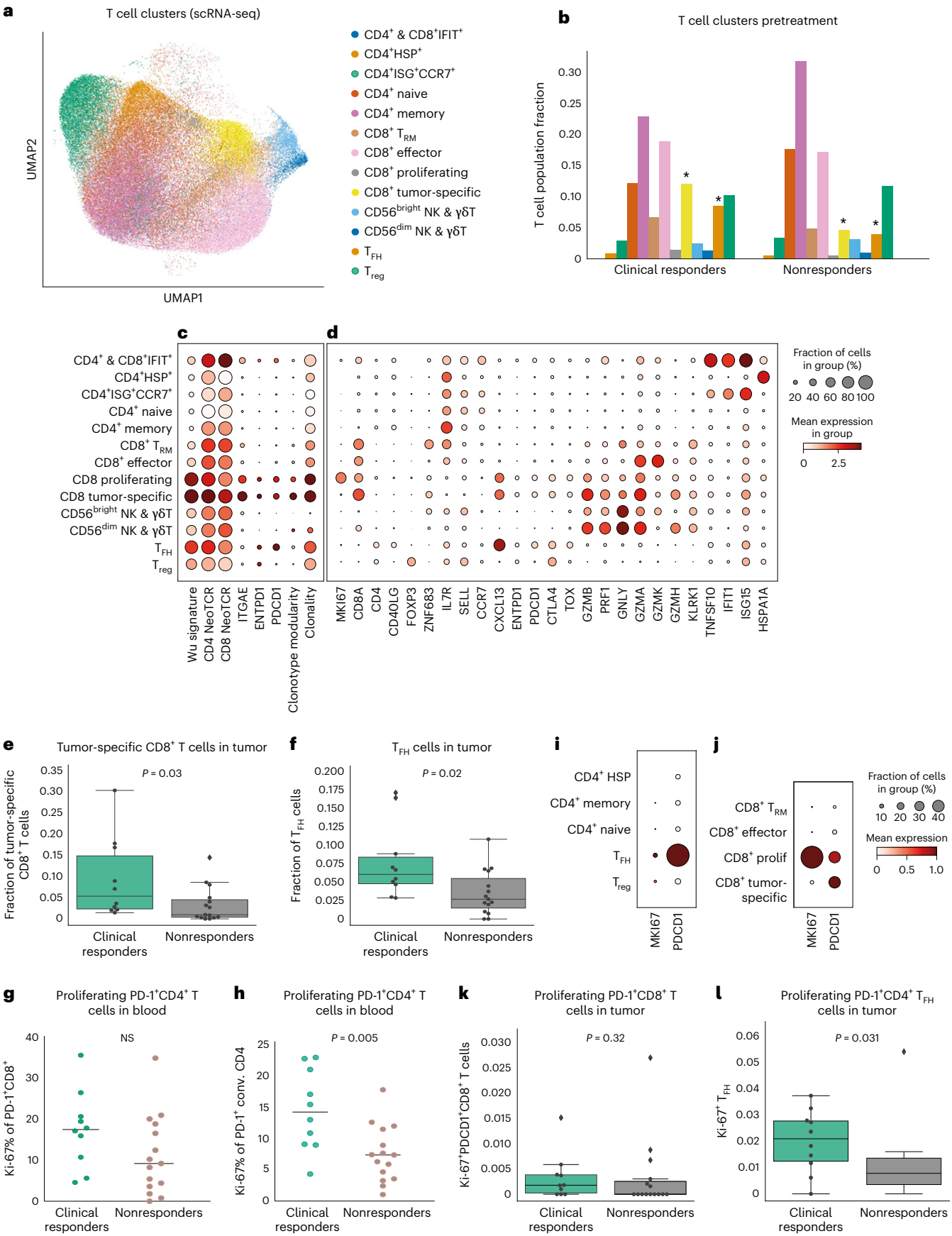
We also investigated whether the addition of anti-CTLA4 led to differential alterations in the TME compared to nivolumab monotherapy, although the study was not powered for cohort comparisons. Patients receiving nivolumab plus ipilimumab showed a reduced fold change in T_{FH} cells ($P = 0.02$; Fig. 4f), but an increased fold change in naive $CD4^+$ T cells ($P = 0.03$; Fig. 4g). Additionally, the combination ICI resulted in a decreased fold change in T_{reg} cells ($P = 0.01$; Fig. 4h) compared to monotherapy, including both activated and non-activated T_{reg} cells (Extended Data Fig. 5d,e).

ctDNA dynamics during early response to ICI

To assess the impact of short-term ICI on circulating tumor DNA (ctDNA), we conducted ctDNA analysis pretreatment and after 4 weeks (cohorts A and B) or 6 weeks (cohort C) of ICI using a tumor-informed ctDNA assay (Signatera). Despite the early tumor stages included

Fig. 3 | Pretreatment T cell profiles of the tumor microenvironment and peripheral blood associated with clinical response in cohorts A and B. **a**, UMAP representation of the T cell clusters in the scRNA-seq dataset (cohorts A and B). $n = 52$ samples from 29 patients, 80, 000 cells. NK, natural killer. **b**, Fractions of different T cell populations relative to all T cells in the pretreatment biopsies from clinical responders (left) and nonresponders (right) in cohorts A and B. **c**, Dotplot illustrating markers of different T cell clusters based on scRNA-seq data (cohorts A and B). **d**, Dotplot illustrating differences in tumor reactivity markers in different T cell clusters based on scRNA-seq data (cohorts A and B). Wu_signature, $CD8^+$ T cell tumor specificity signature[34]; CD4_NeoTCR, $CD4^+$ T cell tumor specificity signature[35]; CD8_NeoTCR, $CD8^+$ T cell tumor specificity signature[35]. **e**, Tumor-specific $CD8^+$ T cell fractions relative to all T cells in pretreatment biopsies of patients with and without clinical response (cohorts A and B). $n = 25$ patients. **f**, T_{FH} fractions relative to all T cells in pretreatment biopsies of patients with and without clinical response (cohorts A and B). $n = 25$ patients. **g**,**h**, Ki-67 expression on PD-1^+CD8^+ T cells (**g**) and conventional $CD4^+$ T cells (**h**) pretreatment in peripheral blood of patients with and without clinical response in cohorts A and B. $n = 25$ patients. **i**, Dotplot for PDCD1 and MKI67 expression in $CD4^+$ T cell clusters (tumoral, scRNA-seq, cohorts A and B). **j**, Dotplot for PDCD1 and MKI67 expression in $CD8^+$ T cell clusters (tumoral, scRNA-seq, cohorts A and B). **k**, Fraction of proliferating PD-1^+CD8^+ T cells relative to all T cells in pretreatment biopsies of patients with and without clinical response based on scRNA-seq data (cohorts A and B). $n = 25$ patients. **l**, Fraction of Ki-67^+ T_{FH} cells relative to all T cells in pretreatment biopsies of patients with and without clinical response (cohorts A and B). $n = 25$ patients. Boxplots display minimum (Q0), maximum (Q4), median (Q2) and IQR (**e**,**f**,**k**,**l**). *P* values were derived using a two-sided Mann–Whitney test. NS, not significant.

https://doi.org/10.1038/s41591-024-03249-3



(mostly I–II), pretreatment ctDNA was detected in 32 of 43 (74%) patients. After treatment, nine (21%) patients had complete ctDNA clearance, while an additional seven patients had a reduction of ≥50% in ctDNA load (mean tumor molecules (MTM) per ml; Fig. 4i,j). All clinical responders in cohorts A and B and patients with pCR/MPR ($n = 8$) in cohort C demonstrated at least a 50% drop in ctDNA or were negative for ctDNA at baseline (Fig. 4i–k).

Discussion

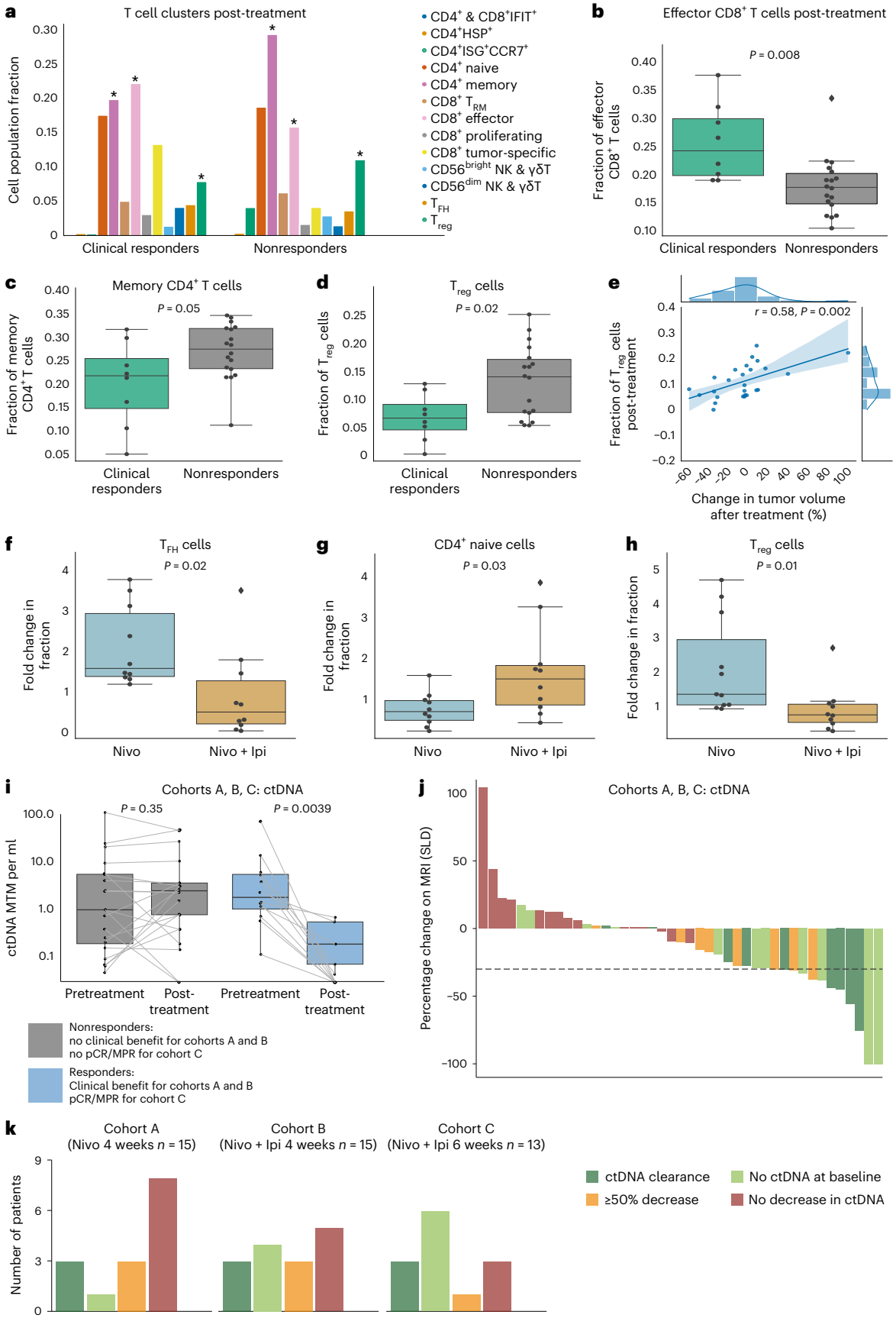
In this study, we demonstrate that neoadjuvant nivolumab, with or without ipilimumab, is a feasible chemotherapy-free regimen for patients with early-stage TNBC. We show that nivolumab ± ipilimumab induces immune activation in the majority of patients and can result in pCR and ctDNA clearance. Pre-existing inflammatory features such as higher TILs, shorter distances from $CD8^+$ T cells to the tumor and higher baseline fractions of tumor-specific $CD8^+$ T cells were associated with response. In contrast, higher fractions of T_{reg} cells post-treatment were associated with lack of response. While standard chemo-immunotherapy for TNBC with four chemotherapy agents plus anti-PD-1 is a 5-month treatment regimen leading to a 63% pCR rate, our work suggests that with only 6 weeks of anti-PD-1 plus low-dose anti-CTLA4, a 33% pCR rate may be obtained in TNBC with high TILs. This suggests that for some patients a short-term immunotherapy-first approach may be an option if confirmed by future research in larger cohorts with a more robust follow-up; however, a substantial group of patients still needs chemotherapy and/or longer treatment to obtain a pCR. Although we did not observe any unexpected toxicity, the rate of persisting endocrinopathies, in particular hypothyroidism, was high compared to reports in other tumor types or in breast cancer when anti-PD-(L)1 is added to neoadjuvant chemotherapy. Although the 33% pCR rate would allow expansion of cohortC to stage II, with 40% grade 3–4 toxicity, 40% hypothyroidism and 20% adrenal gland insufficiencies, substantial toxicity is a serious concern, especially considering the relatively good prognosis of patients with TNBC with high TILs.

The BELLINI trial has investigated the feasibility and potential efficacy of ICI without concurrent chemotherapy in early-stage TNBC. The scoring of TILs is used as an inclusion criterion to select patients with a good prognosis for whom development of de-escalated treatment regimens is most promising. Larger clinical trials also using TILs according to this workflow when including patients have recently started (NCT05929768). In addition, the ETNA trial (NCT06078384) will explore whether patients with stage I TNBC with high TILs can forgo (neo)adjuvant chemotherapy or be treated with immunotherapy alone. The larger international OPTImaL patient preference study (NCT06476119) will also allow the option of no chemotherapy for this patient population. In addition, other studies use TILs as inclusion criteria for immunotherapy-first approaches: Pop-Durva (NCT05215106) and pan-cancer NEOASIS trial (NCT06279130). Further studies that are sufficiently powered to assess long-term outcomes are needed on the use of TILs or other immune-based biomarkers as entry criteria for immunotherapy or de-escalation studies, especially as patients with lower stage TNBC and high TILs can have an excellent outcome with local treatment alone[19,37].

Immune-related endocrine disorders were the most common AEs observed. Specifically, 41% of the patients developed hypothyroidism, which, though usually easy to manage, is a permanent condition and 13% developed adrenal insufficiency, a serious long-term toxicity. Comparable neoadjuvant ICI-only studies with nivolumab + low-dose ipilimumab in head and neck squamous carcinoma, colorectal cancer, urothelial carcinoma and melanoma reported hypothyroidism in 4–8% of patients[9–11,14] and adrenal insufficiency in 0–8% of patients[9–11,14]; however, the recent largest phase 3 trial (stage III melanoma, $n = 423$) reports substantial higher rates of endocrinopathies with 23.6% hypothyroidism and 9.9% adrenal gland insufficiency[26]. Notably, for cancer types with poor prognosis such as stage III melanoma, high toxicity rates might be acceptable, whereas this is different for patient populations with more favorable outcomes. The higher rates of hypothyroidism and adrenal insufficiencies in BELLINI compared to these studies could stem from different patient demographics. Patients with TNBC are typically female and relatively young, potentially contributing to different systemic immunity and AE incidence[38]. In BELLINI, we reported all immune-mediated AEs during the first year of follow-up, with 4 of 6 patients developing adrenal insufficiency >100 days after inclusion. Trials with shorter reporting periods may miss these late events, leading to underreported delayed toxicity, especially in centers not specialized in evaluating ICI regimens. When focusing on patients with similar demographics and disease, we still observe a higher rate of endocrine AEs in BELLINI compared to neoadjuvant trials for TNBC evaluating ICI plus chemotherapy. The KEYNOTE-522 trial reported thyroid dysfunction in 22% of patients treated with anti-PD-1 plus chemotherapy[1]. Adrenal insufficiency/hypophysitis was reported for 4.5% of patients in the KEYNOTE-522 study. A recent study with an oncolytic virus without chemotherapy found that 3 of 6 patients with breast cancer developed hypothyroidism[39], which is more in line with our observations. The lower hypothyroidism rate in the KEYNOTE-522 compared to the oncolytic virus study[39] and BELLINI could suggest that chemotherapy results in partial blunting of the immune response. Last, the preselection of patients with higher TILs in BELLINI may have resulted in patients who are more likely to develop IR AEs due to different systemic immunity. We also cannot rule out the influence of chemotherapy given after ICI, where steroids are used as antiemetics. Our cohort sizes are too small to compare toxicities induced by 4-week nivolumab versus 4-week nivolumab + ipilimumab versus 6-week nivolumab + ipilimumab; however, in the latter group, we observed more non-endocrinopathies such as colitis, hepatitis and pneumonitis, whereas endocrinopathies were already remarkably high with nivolumab monotherapy. This potentially signifies that neoadjuvant ICI without chemotherapy could result in a higher rate of hypothyroidism in patients with breast

Fig. 4 | Effects of anti-PD-1 ± anti-CTLA4 on the T cell profiles in the tumor microenvironment after treatment in cohorts A and B: ctDNA data for all cohorts. a, Fractions of different T cell clusters relative to all T cells in post-treatment biopsies of patients who did (left) and did not (right) experience clinical response based on scRNA-seq data. **b**, Effector $CD8^+$ T cell fractions relative to all T cells in post-treatment biopsies versus clinical response (cohorts A and B). $n = 26$ patients. **c**, Memory $CD4^+$ T cell fractions relative to all T cells in post-treatment biopsies versus clinical response (cohorts A and B). $n = 26$ patients. **d**, T_{reg} cell fractions relative to all T cells in post-treatment biopsies versus clinical response (cohorts A and B). $n = 26$ patients. **e**, Fractions of T_{reg} cells relative to all T cells in post-treatment biopsies of patients (cohorts A and B) in relation to the change in tumor volume after treatment assessed using MRI (RECIST v.1.1). $n = 26$ patients. **f**–**h**, Fold changes in fractions of T cell populations relative to all T cells in cohort A and cohort B. $n = 22$ patients. T_{FH} cells (**f**). Naive $CD4^+$ T cells (**g**). T_{reg} cells (**h**). **i**, Changes in ctDNA levels of responding and nonresponding patients upon treatment. Patients from all cohorts (A, B and C) for whom ctDNA analysis was performed and ctDNA was detected at baseline ($n = 32$) were included. **j**, Waterfall plot of all patients for whom ctDNA analysis was performed ($n = 43$, all cohorts) colored according to the fold change in ctDNA levels in blood upon treatment. The groups represent ctDNA clearance; post-therapy decrease in ctDNA levels of 50% or more; no ctDNA at baseline; and no decrease in ctDNA. The dashed line at −30% indicates radiological PR. **k**, Barplots summarizing the number of patients for each ctDNA response category in each cohort (A, B and C). ctDNA at baseline was available for 43 of 46 patients. Boxplots display minimum (Q0), maximum (Q4), median (Q2) and IQR (**b**–**d**,**f**–**i**). *P* values (**b**–**d**,**f**–**h**) were derived using a two-sided Mann–Whitney test. *P* values in **i** were derived using a paired Wilcoxon test.

https://doi.org/10.1038/s41591-024-03249-3



cancers. Of note, it was demonstrated that immunotherapy-related thyroid dysfunction and other IR AEs are associated with improved survival in multiple cancer types[40–43]. Nevertheless, upfront prediction of risk of immunotherapy-related toxicity for individual patients is a large unmet clinical need and the burden of AEs should be evaluated in light of the prognosis of each patient[44].

The advantage of WOO studies such as BELLINI is the opportunity to evaluate promising drugs and drug combinations in an efficient manner and to analyze pre- and post-treatment tumor material that can provide insights into the therapy effects. Our primary end point of immune activation, defined as a doubling of $CD8^+$ T cells and/or *IFNG* expression, was reached in 17 of 30 patients (57%). Although both cohorts reached the >30% immune activation rate, allowing cohort expansion, we observed more doubling of $CD8^+$ T cells in patients with low pretreatment levels of these features. This could be due to the biopsy timing with deep responses at 4 weeks in tumors with high endogenous $CD8^+$ T cells and/or a 'saturation' of $CD8^+$ T cells in patients with high pretreatment values. In contrast to $CD8^+$ T cells, *IFNG* counts may double even with high pretreatment values; however, they could also be impacted by decreased antigen availability in case of tumor regression. This suggests that different biomarker approaches could apply to inflamed and noninflamed tumors. Recent insights from the developments of personalized neoadjuvant immunotherapy in melanoma indicate that patients with high pre-existing *IFNG* levels or an increase in *IFNG* signature upon treatment were most likely to benefit[45]. The disadvantage of WOO designs with short scheduled treatments is the nonguaranteed benefit for participating patients. Also, information on established end points such as pCR rate is needed before a new treatment approach will be tested in larger trials. For this reason, the adaptive BELLINI trial allowed the opening of new cohorts with established end points to bring therapies to the next step. Although allowed by the protocol and statistical analysis plan, reporting only stage I data of a Simon's two-stage design comes with the risk of false-positive findings. Similarly to cohorts A and B, cohort C also reached the threshold of sufficient responders to expand into stage II; however, given the relatively high rate of endocrinopathies, which are chronic, cohort C was not expanded to stage II. In this view, testing new anti-CTLA4-targeting antibodies, such as botensilimab[46], intentionally designed to overcome the limitations of conventional ICI such as persisting endocrinopathies could be interesting for patients with breast cancer.

When analyzing pretreatment tumor characteristics in high-TIL tumors only (cohort C), we found that the inflammatory phenotype and markers were still discriminative between responders and nonresponders and remarkably similar to the clinical responders and nonresponders in cohorts A and B. In cohort C, pCRs had higher inflammatory gene expression profiles pretreatment, including signatures for *IFNG* response, checkpoint molecules, exhausted $CD8^+$ T cells and immunogenic cell death. This suggests that, even in patients with high TILs, the profiling of baseline inflammatory status may facilitate early identification of (non)responders and should be considered in addition to TILs.

The recent publication of the tumor-specific T cell signatures[34,35] enabled us to identify and follow tumor-specific $CD8^+$ T cells in a clinical trial setting. Notably, using these signatures as a proxy for the tumor reactivity, we demonstrate that the presence of tumor-specific $CD8^+$ T cells pretreatment is linked to ICI response.

Additionally, we observed decreased fractions of T_{reg} cells in clinical responders compared to nonresponders after treatment, in line with previous reports on the role of T_{reg} cells in resistance to ICI[47]. In a resistant mouse tumor model, anti-PD-L1 therapy led to T_{reg} cell activation and T_{reg} cells were shown to be activated in the single-cell data of patients with non-small cell lung cancer and basal cell carcinoma not responding to anti-PD-(L)1 ICI[36]. In this recent study, ICI treatment induced higher expression of genes involved in T_{reg} cell-mediated immune suppression (PDCD1, CTLA4 and CD38) and cell cycle (MKI67) in T_{reg} cells from the tumors of nonresponders[36]. Together, these findings demonstrate that T_{reg} cells might play a critical role in resistance to ICI.

To date, data on combining anti-PD-(L)1 with low-dose anti-CTLA4 were lacking in early-stage breast cancer. Due to the noncomparative design and the small sample size, our data on the potential additive effect of ipilimumab should be considered exploratory. At the single-cell level, the addition of ipilimumab resulted in a lower fold change in T_{reg} cells in the TME upon treatment. We also observed a correlation between higher levels of activated T_{reg} cells post-treatment and the lack of response or in some cases even slight increase in tumor volume on MRI. This suggests that activated T_{reg} cells play a role in resistance to immune checkpoint blockade and that depleting activated T_{reg} cells could be a promising strategy for patients with TNBC who are unresponsive to anti-PD-1-based treatments. Of note, we cannot exclude that the lack of response or the increase of tumor volume observed by imaging was in part due to pseudoprogression. A growing body of literature analyzing anti-CTLA4 using in vivo models indicates that anti-CTLA4 can deplete T_{reg} cells[48]; however, whether anti-CTLA4 can deplete T_{reg} cells in human tumors remains a matter of debate[49]. A recent study by van der Leun et al. in head and neck squamous cell carcinoma also demonstrated an increase in transitional $CD8^+$ T cells and a decrease in $CD137^+$ T_{reg} cells in responders after treatment with anti-PD-1 and anti-CTLA4 therapy[50], indicating that this might be a consistent pattern across multiple tumor types.

After the results of the landmark trials in early-stage TNBC that added PD-1 blockade to standard neoadjuvant chemotherapy[1,3,51,52], our current data provide a rationale to further explore the following observations. First, we observed complete and near-complete pathological responses after only 6 weeks of treatment with ICI in patients with high TILs. This suggests that a subgroup of TNBC could be treated with chemo-free regimens if further research powered for long-term outcome analysis will confirm our results. More research is needed on the optimal selection strategy and treatment regimen, especially in view of the observed high endocrinopathy rate. It is tempting to speculate whether extending the 6-week treatment period could result in higher pCR rates and thereby reach responses similar to outcomes obtained with chemo + ICI. This can only be achieved if the accompanying toxicity does not increase; however, it remains unknown whether pCR after immunotherapy has the same prognostic value as pCR after chemotherapy. Therefore, larger trials are needed to validate the pCR rate after short-term ICI alone and to determine whether this results in excellent survival rates, as seen in other cancers[11,53]. Moreover, pCR might not be the optimal end point as KEYNOTE-522 and GeparNUEVO have indicated that the benefit of PD-1 blockade is not exclusively seen in patients with pCR[51,54]. Second, our exploratory clinical and translational data suggest that combination ICI is feasible and could potentially enhance the effects of PD-1 blockade; however, the benefit–risk ratio of such combinations should always be carefully monitored. Third, establishing the feasibility of patient inclusion based on TIL opens the door for more immune biomarker-driven trials, which is particularly important in diseases such as TNBC that include both inflamed and non-inflamed tumors. The potential integration of additional inflammation analyses, for example, using *IFNG* gene expression as well as TILs as suggested by our data, may optimize patient selection, increase pCR rates for ICI-only approaches and could help treatment personalization in the future. Last, a substantial fraction of patients achieved ctDNA clearance after short-term ICI. Given the strong prognostic value of early ctDNA decrease, as shown by the I-SPY trial[55], future studies are needed to investigate the feasibility and reliability of TIL-informed patient inclusion and the potential of ctDNA-informed therapy adjustments.

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/s41591-024-03249-3.

References

1. Schmid, P. et al. Event-free survival with pembrolizumab in early triple-negative breast cancer. *N. Engl. J. Med.* **386**, 556–567 (2022).
2. Loibl, S. et al. A randomised phase II study investigating durvalumab in addition to an anthracycline taxane-based neoadjuvant therapy in early triple-negative breast cancer: clinical results and biomarker analysis of GeparNuevo study. *Ann. Oncol.* **30**, 1279–1288 (2019).
3. Mittendorf, E. A. et al. Neoadjuvant atezolizumab in combination with sequential nab-paclitaxel and anthracycline-based chemotherapy versus placebo and chemotherapy in patients with early-stage triple-negative breast cancer (IMpassion031): a randomised, double-blind, phase 3 trial. *Lancet* **396**, 1090–1100 (2020).
4. Schmid, P. et al. Pembrolizumab for early triple-negative breast cancer. *N. Engl. J. Med.* **382**, 810–821 (2020).
5. Gustafson, C. E. et al. Immune cell repertoires in breast cancer patients after adjuvant chemotherapy. *JCI Insight* **5**, e134569 (2020).
6. Mariniello, A. et al. Platinum-based chemotherapy attenuates the effector response of CD8 T cells to concomitant PD-1 blockade. *Clin. Cancer Res.* https://doi.org/10.1158/1078-0432.CCR-23-1316 (2023).
7. Blank, C. U. et al. Neoadjuvant versus adjuvant ipilimumab plus nivolumab in macroscopic stage III melanoma. *Nat. Med.* **24**, 1655–1661 (2018).
8. Cascone, T. et al. Neoadjuvant nivolumab or nivolumab plus ipilimumab in operable non-small cell lung cancer: the phase 2 randomized NEOSTAR trial. *Nat. Med.* **27**, 504–514 (2021).
9. Vos, J. L. et al. Neoadjuvant immunotherapy with nivolumab and ipilimumab induces major pathological responses in patients with head and neck squamous cell carcinoma. *Nat. Commun.* **12**, 7348 (2021).
10. van Dijk, N. et al. Preoperative ipilimumab plus nivolumab in locoregionally advanced urothelial cancer: the NABUCCO trial. *Nat. Med.* **26**, 1839–1844 (2020).
11. Chalabi, M. et al. Neoadjuvant immunotherapy in locally advanced mismatch repair-deficient colon cancer. *N. Engl. J. Med.* **390**, 1949–1958 (2024).
12. Bianchini, G., De Angelis, C., Licata, L. & Gianni, L. Treatment landscape of triple-negative breast cancer - expanded options, evolving needs. *Nat. Rev. Clin. Oncol.* **19**, 91–113 (2022).
13. Gianni, L. et al. LBA19 Event-free survival (EFS) analysis of neoadjuvant taxane/carboplatin with or without atezolizumab followed by an adjuvant anthracycline regimen in high-risk triple negative breast cancer (TNBC): NeoTRIP Michelangelo randomized study. *Ann. Oncol.* https://doi.org/10.1016/j.annonc.2023.10.009 (2023).
14. Rozeman, E. A. et al. Identification of the optimal combination dosing schedule of neoadjuvant ipilimumab plus nivolumab in macroscopic stage III melanoma (OpACIN-neo): a multicentre, phase 2, randomised, controlled trial. *Lancet Oncol.* **20**, 948–960 (2019).
15. Robert, C. et al. Ipilimumab plus dacarbazine for previously untreated metastatic melanoma. *N. Engl. J. Med.* **364**, 2517–2526 (2011).
16. Hodi, F. S. et al. Improved survival with ipilimumab in patients with metastatic melanoma. *N. Engl. J. Med.* **363**, 711–723 (2010).
17. 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 (2020).
18. Adams, S. et al. A multicenter phase II Trial of ipilimumab and nivolumab in unresectable or metastatic metaplastic breast cancer: cohort 36 of dual anti-CTLA-4 and anti-PD-1 blockade in rare tumors (DART, SWOG S1609). *Clin. Cancer Res.* **28**, 271–278 (2022).
19. de Jong, V. M. T. et al. Prognostic value of stromal tumor-infiltrating lymphocytes in young, node-negative, triple-negative breast cancer patients who did not receive (neo) adjuvant systemic therapy. *J. Clin. Oncol.* **40**, 2361–2374 (2022).
20. Loi, S. et al. Tumor-Infiltrating lymphocytes and prognosis: a pooled individual patient analysis of early-stage triple-negative breast cancers. *J. Clin. Oncol.* **37**, 559–569 (2019).
21. Denkert, C. et al. Tumour-infiltrating lymphocytes and prognosis in different subtypes of breast cancer: a pooled analysis of 3771 patients treated with neoadjuvant therapy. *Lancet Oncol.* **19**, 40–50 (2018).
22. Salgado, R. et al. The evaluation of tumor-infiltrating lymphocytes (TILs) in breast cancer: recommendations by an International TILs Working Group 2014. *Ann. Oncol.* **26**, 259–271 (2015).
23. Park, J. H. et al. Prognostic value of tumor-infiltrating lymphocytes in patients with early-stage triple-negative breast cancers (TNBC) who did not receive adjuvant chemotherapy. *Ann. Oncol.* **30**, 1941–1949 (2019).
24. Loi, S. et al. Association between biomarkers and clinical outcomes of pembrolizumab monotherapy in patients with metastatic triple-negative breast cancer: KEYNOTE-086 exploratory analysis. *JCO Precis. Oncol.* **7**, e2200317 (2023).
25. Voorwerk, L. et al. Publisher correction: immune induction strategies in metastatic triple-negative breast cancer to enhance the sensitivity to PD-1 blockade: the TONIC trial. *Nat. Med.* **25**, 1175 (2019).
26. Blank, C. U. et al. Neoadjuvant nivolumab and ipilimumab in resectable stage III melanoma. *N. Engl. J. Med.* https://doi.org/10.1056/NEJMoa2402604 (2024).
27. Simon, R. Optimal two-stage designs for phase II clinical trials. *Control. Clin. Trials* **10**, 1–10 (1989).
28. Tumeh, P. C. et al. PD-1 blockade induces responses by inhibiting adaptive immune resistance. *Nature* **515**, 568–571 (2014).
29. Higgs, B. W. et al. Interferon-γ messenger RNA signature in tumor biopsies predicts outcomes in patients with non-small cell lung carcinoma or urothelial cancer treated with durvalumab. *Clin. Cancer Res.* **24**, 3857–3866 (2018).
30. Eisenhauer, E. A. et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur. J. Cancer* **45**, 228–247 (2009).
31. Menzies, A. M. et al. Pathological response and survival with neoadjuvant therapy in melanoma: a pooled analysis from the International Neoadjuvant Melanoma Consortium (INMC). *Nat. Med.* **27**, 301–309 (2021).
32. Verschoor, Y. L. et al. Neoadjuvant atezolizumab plus chemotherapy in gastric and gastroesophageal junction adenocarcinoma: the phase 2 PANDA trial. *Nat. Med.* https://doi.org/10.1038/s41591-023-02758-x (2024).
33. Lehmann, B. D. et al. Refinement of triple-negative breast cancer molecular subtypes: implications for neoadjuvant chemotherapy selection. *PLoS ONE* **11**, e0157368 (2016).
34. Oliveira, G. et al. Phenotype, specificity and avidity of antitumour CD8 T cells in melanoma. *Nature* **596**, 119–125 (2021).
35. Lowery, F. J. et al. Molecular signatures of antitumor neoantigen-reactive T cells from metastatic human cancers. *Science* **375**, 877–884 (2022).

36. van Gulijk, M. et al. PD-L1 checkpoint blockade promotes regulatory T cell activity that underlies therapy resistance. *Sci. Immunol.* **8**, eabn6173 (2023).
37. Geurts, V. C. M. et al. Tumor-infiltrating lymphocytes in patients with stage I triple-negative breast cancer untreated with chemotherapy. *JAMA Oncol.* https://doi.org/10.1001/jamaoncol.2024.1917 (2024).
38. Unger, J. M. et al. Sex differences in risk of severe adverse events in patients receiving immunotherapy, targeted therapy, or chemotherapy in cancer clinical trials. *J. Clin. Oncol.* **40**, 1474–1486 (2022).
39. Nguyen, V. P. et al. A pilot study of neoadjuvant nivolumab, ipilimumab and intralesional oncolytic virotherapy for HER2-negative breast cancer. *Cancer Res. Commun.* https://doi.org/10.1158/2767-9764.crc-23-0145 (2023).
40. Maher, V. E. et al. Analysis of the association between adverse events and outcome in patients receiving a programmed death protein 1 or programmed death ligand 1 antibody. *J. Clin. Oncol.* **37**, 2730–2737 (2019).
41. Eggermont, A. M. M. et al. Association between immune-related adverse events and recurrence-free survival among patients with stage III melanoma randomized to receive pembrolizumab or placebo: a secondary analysis of a randomized clinical trial. *JAMA Oncol.* **6**, 519–527 (2020).
42. Beaufils, M. et al. Dysthyroidism during immune checkpoint inhibitors is associated with improved overall survival in adult cancers: data mining of 1385 electronic patient records. *J. Immunother. Cancer* **11**, e006786 (2023).
43. Street, S. et al. The positive effect of immune checkpoint inhibitor-induced thyroiditis on overall survival accounting for immortal time bias: a retrospective cohort study of 6596 patients. *Ann. Oncol.* **32**, 1050–1051 (2021).
44. Groha, S. et al. Germline variants associated with toxicity to immune checkpoint blockade. *Nat. Med.* **28**, 2584–2591 (2022).
45. Reijers, I. L. M. et al. IFN-γ signature enables selection of neoadjuvant treatment in patients with stage III melanoma. *J. Exp. Med.* **220**, e20221952 (2023).
46. Bullock, A. J. et al. Botensilimab plus balstilimab in relapsed/refractory microsatellite stable metastatic colorectal cancer: a phase 1 trial. *Nat. Med.* https://doi.org/10.1038/s41591-024-03083-7 (2024).
47. Blomberg, O. S. et al. Neoadjuvant immune checkpoint blockade triggers persistent and systemic T activation which blunts therapeutic efficacy against metastatic spread of breast tumors. *Oncoimmunology* **12**, 2201147 (2023).
48. Simpson, T. R. et al. Fc-dependent depletion of tumor-infiltrating regulatory T cells co-defines the efficacy of anti-CTLA-4 therapy against melanoma. *J. Exp. Med.* **210**, 1695–1710 (2013).
49. Sharma, A. et al. Anti-CTLA-4 immunotherapy does not deplete FOXP3$^+$ regulatory T cells (Tregs) in human cancers. *Clin. Cancer Res.* **25**, 1233–1238 (2019).
50. van der Leun, A. M. et al. Dual immune checkpoint blockade induces analogous alterations in the dysfunctional CD8$^+$ T cell and activated T_{reg} compartment. *Cancer Discov.* https://doi.org/10.1158/2159-8290.CD-22-0851 (2023).
51. Loibl, S. et al. Neoadjuvant durvalumab improves survival in early triple-negative breast cancer independent of pathological complete response. *Ann. Oncol.* **33**, 1149–1158 (2022).
52. Gianni, L. et. al. Pathologic complete response (pCR) to neoadjuvant treatment with or without atezolizumab in triple-negative, early high-risk and locally advanced breast cancer: NeoTRIP Michelangelo randomized study. *Ann. Oncol.* **33**, 534–543 (2022).
53. Rozeman, E. A. et al. Survival and biomarker analyses from the OpACIN-neo and OpACIN neoadjuvant immunotherapy trials in stage III melanoma. *Nat. Med.* **27**, 256–263 (2021).
54. Pusztai, L. et al. Event-free survival by residual cancer burden with pembrolizumab in early-stage TNBC: exploratory analysis from KEYNOTE-522. *Ann. Oncol.* **35**, 429–436 (2024).
55. Magbanua, M. J. M. et al. Circulating tumor DNA in neoadjuvant-treated breast cancer reflects response and survival. *Ann. Oncol.* **32**, 229–239 (2021).
56. Ayers, M. et al. IFN-γ-related mRNA profile predicts clinical response to PD-1 blockade. *J. Clin. Invest.* **127**, 2930–2940 (2017).
57. Garg, A. D., De Ruysscher, D. & Agostinis, P. Immunological metagene signatures derived from immunogenic cancer cell death associate with improved survival of patients with lung, breast or ovarian malignancies: a large-scale meta-analysis. *Oncoimmunology* **5**, e1069938 (2016).
58. Hu, X. E. et al. Clinical and biological heterogeneities in triple-negative breast cancer reveals a non-negligible role of HER2-low. *Breast Cancer Res.* **25**, 34 (2023).
59. Bagaev, A. et al. Conserved pan-cancer microenvironment subtypes predict response to immunotherapy. *Cancer Cell* **39**, 845–865.e7 (2021).
60. Gangaev, A. et al. Identification and characterization of a SARS-CoV-2 specific CD8$^+$ T cell response with immunodominant features. *Nat. Commun.* **12**, 1–14 (2021).
61. Cabrita, R. et al. Tertiary lymphoid structures improve immunotherapy and survival in melanoma. *Nature* **577**, 561–565 (2020).

Iris Nederlof[1,22], Olga I. Isaeva[1,22], Manon de Graaf[1], Robbert C. A. M. Gielen[1], Noor A. M. Bakker[1,2], Adrianne L. Rolfes[1], Hannah Garner[1,2], Bram Boeckx[3,4], Joleen J. H. Traets[1], Ingrid A. M. Mandjes[5], Michiel de Maaker[6], Thomas van Brussel[3,4], Maksim Chelushkin[1], Elisa Champanhet[1], Marta Lopez-Yurda[5], Koen van de Vijver[7,8], José G. van den Berg[8], Ingrid Hofland[6], Natasja Klioueva[9], Ritse M. Mann[10], Claudette E. Loo[10], Frederieke H. van Duijnhoven[11], Victoria Skinner[11], Sylvia Luykx[12], Emile Kerver[13], Ekaterina Kalashnikova[14], Marloes G. J. van Dongen[15], Gabe S. Sonke[15], Sabine C. Linn[15],

https://doi.org/10.1038/s41591-024-03249-3

Christian U. Blank[15,16], Karin E. de Visser[1,2,17], Roberto Salgado[18,19], Lodewyk F. A. Wessels[2,20], Caroline A. Drukker[11], Ton N. Schumacher[2,16,21], Hugo M. Horlings[8], Diether Lambrechts[3,4] & Marleen Kok[1,15]

[1]Division of Tumor Biology and Immunology, The Netherlands Cancer Institute, Amsterdam, the Netherlands. [2]Oncode Institute, Utrecht, the Netherlands. [3]Laboratory for Translational Genetics, Department of Human Genetics, KU Leuven, Leuven, Belgium. [4]VIB Center for Cancer Biology, Leuven, Belgium. [5]Biometrics Department, The Netherlands Cancer Institute, Amsterdam, the Netherlands. [6]Core Facility Molecular Pathology & Biobanking, The Netherlands Cancer Institute, Amsterdam, the Netherlands. [7]Department of Pathology, UZ Gent – Universitair Ziekenhuis Gent, Gent, Belgium. [8]Department of Pathology, The Netherlands Cancer Institute, Amsterdam, the Netherlands. [9]Department of Pathology, OLVG Hospital, Amsterdam, the Netherlands. [10]Department of Radiology, The Netherlands Cancer Institute, Amsterdam, the Netherlands. [11]Department of Surgical Oncology, The Netherlands Cancer Institute, Amsterdam, the Netherlands. [12]Medical Oncology Department, Tergooi Hospital – locatie Hilversum, Hilversum, the Netherlands. [13]Department of Oncology, OLVG Hospital, Amsterdam, the Netherlands. [14]Natera, San Carlos, CA, USA. [15]Department of Medical Oncology, The Netherlands Cancer Institute, Amsterdam, the Netherlands. [16]Division of Molecular Oncology and Immunology, The Netherlands Cancer Institute, Amsterdam, the Netherlands. [17]Department of Immunology, Leiden University Medical Center, Leiden, the Netherlands. [18]Department of Pathology, ZAS hospitals, Antwerp, Belgium. [19]Division of Research, Peter MacCallum Cancer Centre, Melbourne, Victoria, Australia. [20]Division of Molecular Carcinogenesis, The Netherlands Cancer Institute, Amsterdam, the Netherlands. [21]Department of Hematology, Leiden University Medical Center, Leiden, the Netherlands. [22]These authors contributed equally: Iris Nederlof, Olga I. Isaeva. e-mail: m.kok@nki.nl

Methods

Patients

Patients in cohorts A and B were eligible for enrollment if they were at least 18 years of age and had stage I–III (clinical tumor stage T1c-3 and nodal stage N0–3, according to the primary tumor regional lymph node staging criteria of the American Joint Committee on Cancer, 7th edition) TNBC with confirmation of estrogen receptor (ER) and HER2 negativity (ER < 10% and HER2 0, 1 or 2 in the absence of amplification as determined by in situ hybridization) on a biopsy from the primary tumor in the breast; newly diagnosed, previously untreated disease; a WHO PS score[62] of 0 or 1 and adequate organ functions. The TIL percentage is needed to be 5% or more. To ensure balanced enrollment based on TIL levels, each cohort included five patients with low (5–10%), five patients with intermediate (11–49%) and five patients with high (≥50%) TIL levels. Patients with concurrent ipsilateral, bilateral or multifocal primary tumors were also eligible for enrollment. For cohort C, patients had to meet the same criteria, but the nodal stage had to be N0, tumor stage T1c–T2 and TILs had to be 50% or more. The intention for cohort C was to explore the potential feasibility of chemotherapy de-escalation in patients with high TILs. As withholding adjuvant capecitabine for high-risk patients and/or escalating locoregional treatment for patients with more extensive disease was undesired, cohort C included only patients who were lymph node-negative.

Exclusion criteria included history of immunodeficiency, autoimmune disease or conditions requiring immunosuppression (>10 mg d^{-1} prednisone or equivalent); other immunosuppressive medications intake within 28 days of study drug administration; chronic or recurring infections; occult breast cancer; fertility preservation due to breast cancer diagnosis; active hepatitis B virus or hepatitis C virus infection; clinically overt cardiovascular disease; or previous systemic anticancer treatment.

Trial design and treatments

The BELLINI trial (Preoperative Trial for Breast Cancer With Nivolumab in Combination With Novel IO; ClinicalTrials.gov registration: NCT03815890) is a single center, nonblinded, nonrandomized, noncomparative phase 2 study designed to evaluate the feasibility and efficacy of checkpoint inhibition before regular neoadjuvant therapy or surgery in patients with primary breast cancer. Cohorts for prespecified breast cancer subgroups are opened in a sequential manner. Here we report the first three TNBC cohorts for patients who were treated with nivolumab (cohort A) or nivolumab + ipilimumab for 4 (cohort B) or 6 (cohort C) weeks. Cohort A had nivolumab monotherapy, 240 mg on day 1 (D1) and D15. Cohort B had nivolumab + ipilimumab 1 mg kg^{-1} on D1 and nivolumab 240 mg on D15. Cohort C had nivolumab + ipilimumab 1 mg kg^{-1} on D1 and D21. Regular therapy, consisting of neoadjuvant chemotherapy or primary surgery, started on D29 and onwards. Given the poor prognosis of patients with low TIL levels and the hypothesis that these women will probably not be the super-responders to ICI, patients were only eligible with TILs ≥ 5%. A threshold of 5% TILs was selected to exclude true immune-deserted tumors. Equal distribution of patients with different levels of tumor of infiltrating lymphocytes over the cohorts was ensured by inclusion of five patients with low TIL (5–10%), five patients with intermediate TIL (11–49%) and five patients with high TIL (≥50%) scores per cohort.

After cohorts A (in the protocol defined as cohort 1B) and B (in the protocol defined as cohort 2B) the protocol was amended to open cohort C (in the protocol defined as cohort 3B). Cohort C had the same inclusion criteria as cohort A and B, except that only inclusion of patients with clinically node-negative disease and with TIL levels of 50% or higher was allowed. With the amendment to open cohort C, the WOO design was changed into a true neoadjuvant design with all patients proceeding to surgery after the immunotherapy. After completing the interim analysis of cohorts A and B, an amendment was approved to use pCR as a primary end point instead of immune activation for cohort C and subsequent cohorts (see details on end points below).

Ethics statement

All patients provided written informed consent before enrollment. This investigator-initiated trial was designed by the Netherlands Cancer Institute (NKI).

The trial was conducted in accordance with the protocol, Good Clinical Practice standards and the Declaration of Helsinki. The full protocol, amendments and the informed consent form were approved by the medical ethical committee of the NKI.

End points

Cohorts A and B. The primary end point for cohorts A and B is immune activation following two cycles of neoadjuvant ICI, defined as a twofold increase in $CD8^+$ T cells assessed via immunohistochemistry and/or an increase in *IFNG* gene expression. High-quality paired biopsies are necessary for the evaluability of this primary end point.

Clinical response. As a secondary end point for cohorts A and B, we evaluated the clinical response. Clinical response is defined as having a radiological and/or pathological response.

Radiological signs of response. At least a 30% decrease on MRI (PR according to RECIST v.1.1, not confirmed). The target (or index) lesion is defined as the largest enhancing lesion. In case of multifocality or multicentricity the largest mass and/or nonmass enhancement was measured in the axial–sagittal or coronal plane and defined as target/index lesion. In these cases, the total area occupied by the tumor (including all masses and nonmass enhancement) was also measured. The total tumor area was used for the RECIST measurements.

Pathological signs of response. Pathological response could be studied in biopsies from 28 patients due to the WOO design. The absence of viable tumor after 4 weeks of therapy in the post-treatment biopsy was classified as a clinical response. For patients proceeding to surgery this was defined as partial or pCR, according to the EUSOMA criteria.

Cohort C. The primary end point for cohort C is pCR, defined as no viable tumor remaining in the breast and lymph nodes (ypT0N0)[63]. MPR (the secondary end point) is a frequently used surrogate end point for efficacy in neoadjuvant trials evaluating immune checkpoint blockade across cancer types[8,11,26]. MPR was defined as ≤10% of residual viable tumor in the surgical specimen[17,64,65] or no viable tumor in the breast but residual tumor cells in the lymph nodes.

All cohorts (A, B and C). Secondary end points included feasibility, safety and radiological response. Feasibility was determined based on any treatment-related complications that led to a delay in chemotherapy or primary surgery beyond 6 weeks from the start of therapy. All patients were closely monitored for AEs for 100 days after the administration of the last study treatment, following the Common Terminology Criteria for Adverse Events (CTCAE) v.5 (ref. 66). In addition, we reported all immune-related AEs in the first year of follow-up. Radiological response was assessed according to the RECIST v.1.1 guidelines, but not confirmed.

Statistical analysis

For this exploratory, hypothesis-generating study, no formal sample size calculation was performed for efficacy because there were no data on the efficacy of neoadjuvant immunotherapy in breast cancer at the time of the design of this study. For cohorts A and B, the null hypothesis of a true immune activation in ≤30% of patients was tested against a one-sided alternative. For cohort C, design was identical with the exception of null hypothesis being pCR in ≤30% of patients tested against a one-sided alternative. For 80% power, at a one-sided significance level of 0.05, 15 patients were accrued per cohort to be evaluated in the first stage. If there were 5 or fewer responses among these 15 patients, the cohort was closed for futility. Otherwise, the cohort could be expanded

with 31 additional patients, reaching a total of 46. We decided to publish after stage I, which was allowed by protocol, due to the observation that very early responses to ICI without chemotherapy are possible in TNBC, which warrants efforts to de-escalate therapy for a subset of patients, in contrast to the current therapy escalation for all patients with TNBC. The median follow-up time was obtained using a reverse Kaplan–Meier method. Analyses were performed using R[67] v.4.2.1.

Pathology assessments and IHC analyses

All patients underwent baseline tumor staging, consisting of ultrasound of the breast, axilla and periclavicular region and MRI imaging of the breast. Positron emission tomography and computed tomography imaging was performed in all participants to confirm the clinical stage. Pretreatment tumor histological biopsies (four core biopsies, 14G needle) were taken for all patients and post-treatment tissue was either obtained through a biopsy (three core biopsies, 14G needle) for patients continuing neoadjuvant chemotherapy ($n = 28$) and the surgical specimen was used for those undergoing surgery right after the ICI study treatment ($n = 3$). Histopathological examination of biopsies and resection specimens was carried out by five experienced breast cancer pathologists (H.M.H., R.S., K.v.d.V., J.v.d.B. and N.K.). 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 identified on routine hematoxylin and eosin (H&E) staining. Formalin-fixed paraffin-embedded (FFPE) tissue sections were used for H&E staining and for immunohistochemical analysis of CD8 (C8/144B, DAKO), PD-L1 (22C3, DAKO) and PD-1 (NAT105, Roche Diagnostics). The percentage of tumor cells and TILs was assessed by pathologists trained for TIL assessment on H&E-stained slides according to the international standard from the International Immuno-Oncology Biomarker Working Group[22] (see www.tilsinbreastcancer.org for all guidelines on TIL assessment in solid tumors). After a pathologist provided an initial TIL score, an 'expert TIL score' was generated as a consensus score from at least two out of four trained pathologists using slidescore.com for online scoring (www.slidescore.com). TIL scores for inclusion were scored on the diagnostic biopsy of the patient to allow for stratification of patients (low ≥ 5–10%, intermediate = 11–49% and high ≥ 50%).

Immunohistochemistry

IHC of the FFPE tumor samples was performed on a BenchMark Ultra autostainer (Ventana Medical Systems). The double stain was performed on a Discovery Ultra autostainer. In brief, paraffin sections were cut at 3 µm, heated at 75 °C for 28 min and deparaffinized in the instrument with EZ prep solution (Ventana Medical Systems). Heat-induced antigen retrieval was carried out using Cell Conditioning 1 (CC1, Ventana Medical Systems) for 48 min at 95 °C (PD-L1) or 64 min at 95 °C (PD-1/CD8 double). PD-L1 was detected using clone 22C3 (1:40 dilution, 1 h at room temperature, Agilent/DAKO, lot 11654144). Bound antibody was detected using the OptiView DAB Detection Kit (Ventana Medical Systems). Slides were counterstained with Hematoxylin and Bluing Reagent (Ventana Medical Systems).

For the double-staining PD-1 (Yellow) followed by CD8 (Purple), PD-1 was detected in the first sequence using clone NAT5 (Ready-to-Use, 32 min at 37 °C, Roche Diagnostics, lot 11654144). The PD-1-bound antibody was visualized using anti-mouse NP (Ventana Medical Systems, Ready-to-Use dispenser, lot K09956) for 12 min at 37 °C followed by anti-NP AP (Ventana Medical Systems, Ready-to-Use dispenser, lot J23971) for 12 min at 37 °C, followed by the Discovery Yellow detection kit (Ventana Medical Systems). In the second sequence of the double-staining procedure, CD8 was detected using clone C8/144B (1:200 dilution, 32 min at 37 °C, Agilent, lot 41527763). CD8 was visualized using anti-mouse HQ (Ventana Medical systems, Ready-to-Use dispenser, lot K20711) for 12 min at 370 °C followed by anti-HQ HRP (Ventana Medical Systems, Ready-to-Use dispenser, lot K22062) for 12 min at 37 °C, followed by the Discovery Purple Detection kit (Ventana Medical Systems). Slides were counterstained with Hematoxylin and Bluing Reagent (Ventana Medical Systems). A PANNORAMIC 1000 scanner from 3DHISTECH was used to scan the slides at a ×40 magnification.

Distance analysis between tumor and CD8$^+$ T cells

Spatial analysis was performed on the pretreatment biopsies of all included patients. The stained slides were scanned and image analysis was performed with the HALO image analysis software from Indica Labs, v.3.4.2986.185 (cohorts A and B) and v.3.6.4134 (cohort C). Within HALO, the multiplex IHC module was used to phenotype and quantify CD8$^+$ cells. Cell segmentation was performed by the detection of hematoxylin (detection weight of 1) and PD-1 (detection weights 0.045 for cohorts A and B; and 0.5 for cohort C) and CD8 for cohort C (detection weight of 0.5) staining, utilizing a nuclear segmentation aggressiveness of 0.045. Minimal intensity thresholds to consider a cell positive for a marker were set for hematoxylin (0), PD-1 (0.25 for cohorts A and B and 0.1 for cohort C) and CD8 (0.1) separately. Biopsies were analyzed in total, while for resection specimens the analysis was restricted to representative tumor beds as annotated by a breast cancer pathologist. The quantified levels of CD8$^+$ and PD-1$^+$CD8$^+$ cells were corrected for the analyzed tissue area (cells per µm^2).

Artificial intelligence tumor classifiers (Object Phenotyper, HALO AI) were developed to discriminate between tumor and nontumor cells in cohorts A and B and in cohort C. Individual cells were segmented (nuclei seg BF v.1.0.0), and the classifiers were trained by annotating single cells as tumor or nontumor. The annotations were guided by marked tumor regions on H&E-stained slides by a trained breast cancer pathologist. The classifiers were finalized with 20,000 iterations and a cross-entropy of 0.009 (cohort A and B) and >10,000 iterations and cross-entropy of 0.021 (cohort C).

Merging the results of the multiplex IHC and tumor classifier enabled the visualization of the spatial distribution of tumor and CD8$^+$ cells (Extended Data Fig. 1b–f). Using the nearest neighborhood analysis, the average distance between the tumor and immune cells was quantified by taking the mean of the distances between every tumor cell and its nearest cell of the above-mentioned immune phenotypes in the pretreatment biopsies (Extended Data Fig. 1f). Distances from tumor cells to the nearest CD8$^+$ T cells were taken as a measure of proximity of CD8$^+$ T cells to the tumor.

DNA and RNA isolation

DNA and RNA were extracted from fresh-frozen, pre- and post-treatment tumor material using the AllPrep DNA/RNA kit (QIAGEN) for frozen material, following the manufacturer's protocol, in a QIAcube (QIAGEN). Germline DNA was isolated from patient peripheral blood mononuclear cells using the DNeasy Blood & Tissue kit (QIAGEN).

Bulk RNA sequencing

Total RNA quality control. Quality and quantity of the total RNA was assessed by the 2100 BioAnalyzer using a Nano chip (Agilent). Total RNA samples having a RIN > 8 were subjected to library generation.

TruSeq stranded mRNA library generation

Strand-specific libraries were generated using the TruSeq stranded mRNA sample preparation kit (Illumina, RS-122-2101/2) according to the manufacturer's instructions (Illumina, document no. 1000000040498 v00). In brief, polyadenylated RNA from intact total RNA was purified using oligo-dT beads. Following purification, the RNA was fragmented, random primed and reverse transcribed using SuperScript II Reverse Transcriptase (Invitrogen, part no. 18064-014) with the addition of Actinomycin D. Second-strand synthesis was performed using Polymerase I and RNaseH with replacement of dTTP for dUTP. The generated cDNA fragments were 3' end adenylated and ligated to Integrated DNA Technologies (IDT) xGen UDI(10 bp)-UMI(9 bp) paired-end sequencing adaptors (Integrated DNA Technologies) and subsequently amplified

by 12 cycles of PCR. The libraries were analyzed on a 2100 BioAnalyzer using a 7500 chip (Agilent), diluted and pooled equimolar into a multiplex sequencing pool.

Sequencing

The libraries were sequenced with 54 paired-end reads on a NovaSeq 6000 using S1 Reagent kit v.1.5 (100 cycles) (Illumina).

Data analysis

RNA-seq data were aligned to GRCh38 with STAR[68] v.2.7.1a, with the twopassMode = 'Basic'. FPKM were obtained with RSeQC[69] v.4.0.0 FPKM_count.py and subsequently normalized to transcripts per million. Data quality was assessed with FastQC[70] v.0.11.5, FastQ Screen[71] v.0.14.0, the Picard CollectRnaSeqMetrics[72,73] and RSeQC[69] v.4.0.0 read_distribution.py and read_duplication.py and were found to be suitable for the downstream analysis. TNBCtype[74] was used for the Lehmann subtype classification[75]. The Gseapy[76] v.1.0.3 ssgsea tool with the sample_norm_method = 'rank' was used for gene set signature scoring. For the signature analysis, *P* values were significant after FDR correction (Benjamini–Hochberg) at a 10% significance level. Data were analyzed with Python[77] v.3.10.5. Pandas[78,79] v.2.0.0 and numpy[80] v.1.22.4 were used for data handling. Matplotlib[72] v.3.5.2, seaborn[81] v.0.12.2 and statannotations[82] v.0.5.0 were used for plotting.

Whole-exome sequencing

For each sample the amount of double-stranded DNA was quantified by using the Qubit dsDNA HS Assay kit (Invitrogen, cat. no. Q32851). A maximum amount of 2 µg double-stranded genomic DNA was fragmented by covaris AFA technology to obtain fragment sizes of 200–300 bp. Samples were purified using Agencourt AMPure XP Reagent (Beckman Coulter, cat. no. A63881) in a 2× reaction volume settings according to the manufacturer's instructions. The fragmented DNA was quantified and qualified on a BioAnalyzer system using the DNA7500 assay kit (Agilent Technologies cat no. 5067- 1506). With a maximum input amount of 1 µg fragmented DNA, next-generation sequencing library preparation for Illumina sequencing was performed using the KAPA HTP Prep kit (KAPA Biosystems, KK8234) in combination with xGen UDI-UMI adaptors (IDT). During the library amplification step, four cycles of PCR were performed to obtain enough yield for the exome enrichment assay. All DNA libraries were quantified on a BioAnalyzer system using the DNA7500 assay kit. Exome enrichment was performed on library pools of six unique dual indexed libraries, 500 ng each, using the xGen Exome Hyb Panel v.2 (IDT, cat. no. 10005152) and xGen Hybridization Capture Core Reagents according to manufacturer's protocol, with hybridization time adjusted to 16 h and ten cycles of PCR performed during post-capture PCR. All exome enriched library pools were quantified on a BioAnalyzer system using the DNA7500 assay kit, pooled equimolar to a final concentration of 10 nM and subjected to paired-end 100-bp sequencing on an Illumina Novaseq 6000 instrument using a NovaSeq 6000 S4 Reagent Kit v.1.5 (Illumina, 20028313), according to the manufacturer's instructions.

Data analysis

Sequencing reads were aligned to the human reference GRCh38 (Ensemble, v.105) using BWA[83] v.0.7.17. Duplicated reads were marked using Picard[73] MarkDuplicates v.2.25.0, after which quality scores were recalibrated using GATK4 (ref. 84) BaseRecalibrator v.4.2.2.0. Single-nucleotide variants and short insertions and deletions (indels), were called using GATK4 (ref. 84) Mutect2 v.4.2.2.0 on the tumor samples matched with germline samples. Subsequently, variants were filtered by the PASS filter, and annotated using Ensembl Variant Effect Predictor 105. The maftools[85] v.2.10.5 package was used for the analysis. Tumor mutational burden was calculated by summarizing the total number of nonsynonymous somatic mutations with a minimal variant allele frequency of 20%. Data were analyzed with Python[77] v.3.10.5 and R[67] v.4.1.3. Pandas[78,79] v.2.0.0 was used for data handling. maftools[85] v.2.10.5, Matplotlib[72] v.3.5.2, seaborn[81] v.0.12.2 and statannotations[82] v.0.5.0 were used for plotting.

scRNA-seq and TCR sequencing

Preparation of the single-cell suspension. Following biopsy or obtaining resection specimens, samples were rapidly processed for scRNA-seq. Samples from cohort A were minced on ice and frozen in 10% dimethylsulfoxide FCS at −80 °C. Within 4 weeks after freezing, samples were defrosted in 37 °C medium. Samples from cohort B were minced on ice and immediately processed for single-cell sequencing (not frozen), which did not result in a batch effect.

Samples were transferred to a tube containing 1 ml digestion medium containing collagenase P (2 mg ml^{-1}, Thermo Fisher Scientific) and DNase 1 (10 U $µl^{-1}$, Sigma) in RPMI (Thermo Fisher Scientific). Samples were incubated for 20 min at 37 °C and were pipetted up and down every 5 min for 30 s. Next, samples were filtered on a 40-µm nylon mesh (Thermo Fisher Scientific) and directly after the same volume of ice cold PBS containing 0.04% BSA was added. Following centrifugation at 300*g* and 4 °C for 5 min, the supernatant was removed and discarded, and the cell pellet was resuspended in red cell blood lysis buffer for 5 min at room temperature and then centrifuged again at 300*g* at 4 °C for 5 min. The supernatant was removed and discarded and the pellet was resuspended in PBS containing 0.04% BSA. Next, 10 µl of this cell suspension was counted using an automated cell counter (ChemoMetec NucleoCounter NC-200) to determine the concentration of live cells. The entire procedure was usually completed within 1 h and 15 min.

scRNA-seq data acquisition and preprocessing. Libraries for scRNA-seq were generated using the Chromium Single Cell 5′ library and Gel Bead & Multiplex kit from 10x Genomics. We aimed to profile 10,000 cells per library if a sufficient number of cells was retained during dissociation. All libraries were sequenced on a HiSeq4000 or NovaSeq 6000 until sufficient saturation was reached.

Data analysis. After quality control, raw sequencing reads were aligned to the human reference genome GRCh38 and processed to a matrix representing the unique molecular identifiers' per-cell barcode per gene using Cell Ranger (10x Genomics, v.2.0). The data were analyzed with scanpy[86] v.1.9.3 and Seurat[87] v.3. Cellbender[88] v.0.3.0 was used for eliminating technical artifacts and cells above the quality cutoff of 0.5 were filtered out. Cells with mitochondrial RNA content >0.25, the number of genes <200 or >6,000 and <400 counts were filtered out. After normalization, regression for the number of unique molecular identifiers, percentage mtRNA, sample ID, cell cycle, hypoxia, interferon content and cell stress was performed on the 2,000 most variable genes followed by principal-component analysis. Next, a Uniform Manifold Approximation and Projection (UMAP) was generated and clustering was performed at resolution of 0.2 using the 30 most informative components. Major cell types were identified based on canonical marker genes.

For T cell subclustering, the T cells were selected from the full Seurat object and the analysis described above was repeated with ten principal components based on the elbow plot and clusters were identified at a resolution of 0.6 and were annotated based on breast cancer tissue-specific marker genes[89]. Cells expressing markers of other cell types (immunoglobulins and hemoglobin) were filtered out. Principal-component analysis was calculated on highly variable genes with $k = 30$. Clustering was performed with Phenograph[90] with $k = 30$. Cluster identification was performed based on canonical marker genes. Signature scores were calculated with sc.tl.score_genes. Groups were compared to sc.tl.rank_genes_groups, with method = 'wilcoxon' and use_raw = True. EnrichR[91,92] was used for the pathway enrichment analysis. Activated T_{reg} cells were defined based on the level of *CD137* gene expression >0.5 in the T_{reg} cell population. PD-1^{+}Ki-$67^{+}$$CD4^{+}$ cells

were defined based on the level of *MKI67* gene expression >0 in the T_{FH} cell population. Scirpy[93] v.0.11.2 was used for the TCR analysis. Clonotypes were defined based on the amino acid structure. Clonality was calculated as (1 − normalized Shannon entropy). Data were analyzed with Python[77] v.3.10.5. Pandas[78,79] v.2.0.0 and numpy[80] v.1.22.4 were used for data handling. Matplotlib[72] v.3.5.2, seaborn[81] v.0.12.2, sc-toolbox[94] v.0.12.3 and statannotations[82] v.0.5.0 were used for plotting.

ctDNA analysis. A proprietary bioinformatics tissue variant calling pipeline was used to select a set of 16 high-ranked, patient-specific, somatic, clonal single-nucleotide variants from whole-exome sequencing. The Signatera amplicon design pipeline was used to generate multiplex PCR (mPCR) primer pairs for the given set of 16 variants. For cfDNA library preparation, up to 20,000 genome equivalents of cfDNA from each plasma sample were used. The cfDNA was end-repaired, A-tailed and ligated with custom adaptors, followed by amplification (20 cycles) and purified using Ampure XP beads (Agencourt/Beckman Coulter). A proprietary mPCR methodology was used to run patient-specific assays. Sequencing was performed on these mPCR products on an Illumina HiSeq 2500 Rapid Run (50 cycles) using the Illumina Paired End v.2 kit with an average read depth of >100,000× per amplicon. All paired-end reads were merged using Pear v.0.9.8 software and mapped to the hg19 reference genome with Novoalign v.2.3.4 (http://www.novocraft.com/). Plasma samples with at least two variants with a confidence score above a predefined algorithm threshold were defined as ctDNA-positive.

Flow cytometry of fresh blood. Flow cytometry was performed as previously described[95]. In brief, fresh blood samples were processed and analyzed within 24 h after blood draw. Peripheral blood was collected in EDTA vacutainers (BD) and subjected to red blood cell lysis (lysis buffer, dH_2O, NH_4Cl, $NaHCCO_3$ and EDTA). Cells were suspended in PBS containing 0.5% BSA and 2 mM EDTA and counted using the NucleoCounter NC-200 (Chemometec) automated cell counter. To obtain absolute white blood cell counts per ml human blood, the total amount of post-lysis cells was divided by the volume (ml) of blood obtained from the patient. For surface antigen staining, cells were first incubated with human FcR Blocking Reagent (1:100 dilution, Miltenyi) for 15 min at 4 °C and then incubated with fluorochrome-conjugated antibodies for 30 min at 4 °C. For intracellular antigen staining, cells were fixed with Fixation/Permeabilization solution 1× (Foxp3/Transcription Factor Staining Buffer Set, eBioscience) for 30 min at 4 °C and stained with fluorochrome-conjugated antibodies in Permeabilization buffer 1× (eBioscience) for 30 min at room temperature. Viability was assessed by staining with either 7AAD staining solution (1:10 dilution; eBioscience) or Zombie Red Fixable Viability kit (1:800 dilution, BioLegend). Data acquisition was performed on an LSRII SORP flow cytometer (BD Biosciences) using Diva software and data analysis was performed using FlowJo v.10.6.2. The gating strategy is displayed in Extended Data Fig. 5a.

Reporting summary

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

Data availability

DNA and RNA-seq data are stored in the European Genome–Phenome Archive (EGAS50000000567 (RNA-Seq) and EGAS50000000568 (WES)). Sequencing data and source data supporting the findings of this study will be made available from the corresponding author (m.kok@nki.nl) for academic use, within the limitations of the provided informed consent. Data will not be made available for commercial use. A first response to the request will be sent in <4 weeks. Data requests will be reviewed by the corresponding author and Institutional Review Board of the NKI and after approval, applying researchers will have to sign a data transfer agreement with the NKI.

Code availability

No custom developed code was used for the analysis of the study data. All relevant references are included in Methods.

References

62. Oken, M. M. et al. Toxicity and response criteria of the Eastern Cooperative Oncology Group. *Am. J. Clin. Oncol.* **5**, 649–655 (1982).
63. Litton, J. K. et al. Standardized definitions for efficacy end points in neoadjuvant breast cancer clinical trials: NeoSTEEP. *J. Clin. Oncol.* **41**, 4433–4442 (2023).
64. Blakely, C. M. & McCoach, C. E. Role of MPR as an early signal for efficacy in neoadjuvant studies. *Clin. Cancer Res.* **26**, 3499–3500 (2020).
65. Cascone, T. et al. A phase I/II study of neoadjuvant cisplatin, docetaxel, and nintedanib for resectable non-small cell lung cancer. *Clin. Cancer Res.* **26**, 3525–3536 (2020).
66. *Common Terminology Criteria for Adverse Events (CTCAE) Version 5.0* (US Department of Health and Human Services, 2017); https://ctep.cancer.gov/protocoldevelopment/electronic_applications/docs/ctcae_v5_quick_reference_5x7.pdf
67. *R: A Language and Environment for Statistical Computing* https://www.R-project.org/ (R Foundation for Statistical Computing, 2022).
68. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
69. Wang, L., Wang, S. & Li, W. RSeQC: quality control of RNA-seq experiments. *Bioinformatics* **28**, 2184–2185 (2012).
70. *FastQC A Quality Control tool for High Throughput Sequence Data* (Babraham Bioinformatics, 2023); https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
71. Wingett, S. W. & Andrews, S. FastQ Screen: a tool for multi-genome mapping and quality control. *F1000Res.* **7**, 1338 (2018).
72. Hunter, J. D. Matplotlib: a 2D graphics environment. *Comput. Sci. Eng.* **9**, 90–95 (2007).
73. *Picard Toolkit. GitHub Repository* (Broad Institute, 2019); https://broadinstitute.github.io/picard
74. Chen, X. et al. TNBCtype: a subtyping tool for triple-negative breast cancer. *Cancer Inform.* **11**, 147–156 (2012).
75. Lehmann, B. D. et al. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *J. Clin. Invest.* **121**, 2750–2767 (2011).
76. Fang, Z., Liu, X. & Peltz, G. GSEApy: a comprehensive package for performing gene set enrichment analysis in Python. *Bioinformatics* **39**, btac757 (2022).
77. van Rossum, G. *Python Reference Manual* (CWI, 1995).
78. The Pandas Development Team. pandas-dev/pandas: Pandas (v.2.0.0). *Zenodo* https://doi.org/10.5281/zenodo.10957263 (2024).
79. McKinney, W. Data structures for statistical computing in Python. In *Proc. 9th Python in Science Conference* (SciPy, 2010).
80. Harris, C. R. et al. Array programming with NumPy. *Nature* **585**, 357–362 (2020).
81. Waskom, M. et al. Mwaskom/seaborn: v.0.8.1. *Zenodo* https://doi.org/10.5281/zenodo.883859 (2017).
82. Charlier, F. et al. trevismd/statannotations: v.0.5. *Zenodo* https://doi.org/10.5281/zenodo.7213391 (2022).
83. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
84. van der Auwera, G. & O'Connor, B. D. *Genomics in the Cloud: Using Docker, GATK, and WDL in Terra* (O'Reilly Media, 2020).
85. Mayakonda, A., Lin, D.-C., Assenov, Y., Plass, C. & Koeffler, H. P. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res.* **28**, 1747–1756 (2018).
86. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 1–5 (2018).
87. Hao, Y. et al. Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat. Biotechnol.* https://doi.org/10.1038/s41587-023-01767-y (2023).

88. Fleming, S. J. et al. Unsupervised removal of systematic background noise from droplet-based single-cell experiments using CellBender. *Nat. Methods* https://doi.org/10.1038/s41592-023-01943-7 (2023).
89. Bassez, A. et al. A single-cell map of intratumoral changes during anti-PD1 treatment of patients with breast cancer. *Nat. Med.* **27**, 820–832 (2021).
90. Levine, J. H. et al. Data-driven phenotypic dissection of AML reveals progenitor-like cells that correlate with prognosis. *Cell* **162**, 184–197 (2015).
91. Chen, E. Y. et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinform.* **14**, 128 (2013).
92. Kuleshov, M. V. et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90–W97 (2016).
93. Sturm, G. et al. Scirpy: a Scanpy extension for analyzing single-cell T-cell receptor-sequencing data. *Bioinformatics* **36**, 4817–4818 (2020).
94. Heumos, L. et al. schillerlab/sc-toolbox: a collection of project templates and useful functions for single-cell data analysis with Scanpy. *GitHub* https://github.com/schillerlab/sc-toolbox (2021).
95. Blomberg, O. S. et al. IL-5-producing CD4 T cells and eosinophils cooperate to enhance response to immune checkpoint blockade in breast cancer. *Cancer Cell* **41**, 106–123.e10 (2023).

Acknowledgements

We are grateful to the participants and their families for participating in the trial. We thank all supporting clinical trial staff, in particular nurse specialists and the Departments of Medical Oncology, Surgery, Radiology and Pathology of the participating centers. We thank the NKI trial laboratory for handling incoming blood samples. We are grateful to A. Broeks and the Core Facility of Molecular Pathology and Biobanking for the storage and handling of human tumor material and to the Genomics Core Facility for DNA and RNA sequencing. We acknowledge the supporting staff of the Department of Medical Oncology and the diagnostic laboratory. We thank C. Klaver and M. Duijst for blood sample experiments and the Flow Cytometry Facility at the NKI for support in these experiments. We thank B. Stegenga from Bristol Myers Squibb (BMS) for trial support and arranging the supply of nivolumab and ipilimumab. R.S. is supported by the Breast Cancer Research Foundation (grant no. 17-194). Research at the NKI is supported by institutional grants of the Dutch Cancer Society and of the Dutch Ministry of Health, Welfare and Sport. Research in the laboratory of M.K. is funded by the Netherlands Organization for Scientific Research (VIDI 09150172010043) and Victoria's Secret Global Fund for Women's Cancers Rising Innovator Research Grant, in Partnership with Pelotonia & American Association for Cancer Research (23-30-73-KOK). Trial costs were supported by BMS. The funders had no role in study design, data collection or analysis, decision to publish or preparation of the manuscript.

Author contributions

I.N. and O.I.I. contributed equally to this work as shared co-first authors. M.d.G., R.C.A.M.G. and N.A.M.B. contributed equally to this work as shared co-second authors. I.N. wrote the study protocol, coordinated trial procedures, performed wet laboratory experiments for scRNA-seq analyses, analyzed and interpreted clinical and translational data of the trial. O.I.I. designed, performed and interpreted computational analyses of the DNA, bulk and scRNA-seq data, analyzed and interpreted translational data. I.N., O.I.I. and M.K. wrote the paper. N.A.M.B., E.C. and H.G. were responsible for blood sample processing and analysis, supervised by K.E.d.V. and M.K., and H.G. designed the flow cytometry panel. M.d.G. performed the spatial analyses and helped with collection and analysis of clinical data. R.C.A.M.G. and A.L.R. coordinated trial procedures and collected clinical data for cohort C. B.B., J.G.H.T. and M.C. performed bioinformatics analyses and contributed to their design. I.A.M.M. was the clinical projects manager. M.d.M. processed FFPE for IHC and isolated DNA and RNA from tissue biopsies. T.v.B. performed wet laboratory experiments for scRNA-seq analyses. M.L.-Y. performed statistical analysis of the trial data. J.G.v.d.B., N.K., H.M.H., K.v.d.V. and R.S. performed the TIL and histological scoring of the pathology slides. I.H. developed and performed double CD8-PD-1 staining. R.M.M. and C.E.L. revised MRI scans and, together with colleagues, were involved in taking biopsies. E. Kalashnikova organized the ctDNA experiments. E Kerver, F.H.v.D., V.S., S.L., C.A.D., M.G.J.v.D., G.S.S., S.C.L. and M.K. were the main treating physicians. I.N., I.A.M.M., G.S.S., T.N.S., C.U.B., S.C.L. and M.K. wrote the trial protocol. H.M.H. supervised the computational pathology analyses. L.F.A.W. and D.L. supervised computational analyses. M.K. was the principal investigator of the trial, supervised all the analyses presented in the paper and acquired funding. All authors edited and approved the paper.

Competing interests

R.M.M. reports research grants from Siemens Healhtineers, Bayer Healthcare, Screenpoint Medical, Beckton & Dickinson, PA Imaging, Lunit and Koning, and is an advisory board member for Screenpoint, Bayer, Siemens and Guerbet, all outside the scope of this work. E. Kalashnikova is an employee of Natera. G.S.S. reports research funding to the institute from Merck, Agendia, AstraZeneca, Roche and Novartis and a consulting role for Novartis, Seattle Genetics and Biovica, outside the submitted work. S.C.L. reports research funding to the institute from Roche/Genentech, AstraZeneca, BMS, Tesaro, Merck, Immunomedics, Eurocept Pharmaceuticals, Agendia and Novartis and a consulting role and travel grant from Daiichi Sankyo, outside this work. C.U.B. has received research grants from Novartis, BMS and NanoString, is a paid advisory board member for BMS, MSD, Roche, Novartis, GlaxoSmithKline, AstraZeneca, Pfizer, Lilly, GenMab and Pierre Fabre and holds ownership interest in Uniti Card, Neon Therapeutics and Forty Seven, all outside this submitted work. K.E.d.V. reports research funding from Roche and is a consultant for Macomics, outside the scope of this work. R.S. reports nonfinancial support from Merck and BMS, research support from Merck, Puma Biotechnology and Roche and personal fees from Roche, BMS and Exact Sciences for advisory boards, all outside the scope of this paper. L.F.A.W. reports funding to the institute from Genmab BV. T.N.S. is an advisor for Allogene Therapeutics, Asher Bio, Merus, Neogene Therapeutics and Scenic Biotech; is a stockholder in Allogene Therapeutics, Asher Bio, Cell Control, Celsius, Merus and Scenic Biotech; and is venture partner at Third Rock Ventures, all outside of the current work. M.K. reports research funding to the institute from BMS, Roche and AstraZeneca/MedImmune and an advisory role/speakers' fee (all compensated to the institute) for Alderaan, BMS, Domain Therapeutics, Medscape, Roche, MSD and Daiichi Sankyo, outside the submitted work. Natera provided nonfinancial support to this study. The other authors declare no competing interests.

Additional information

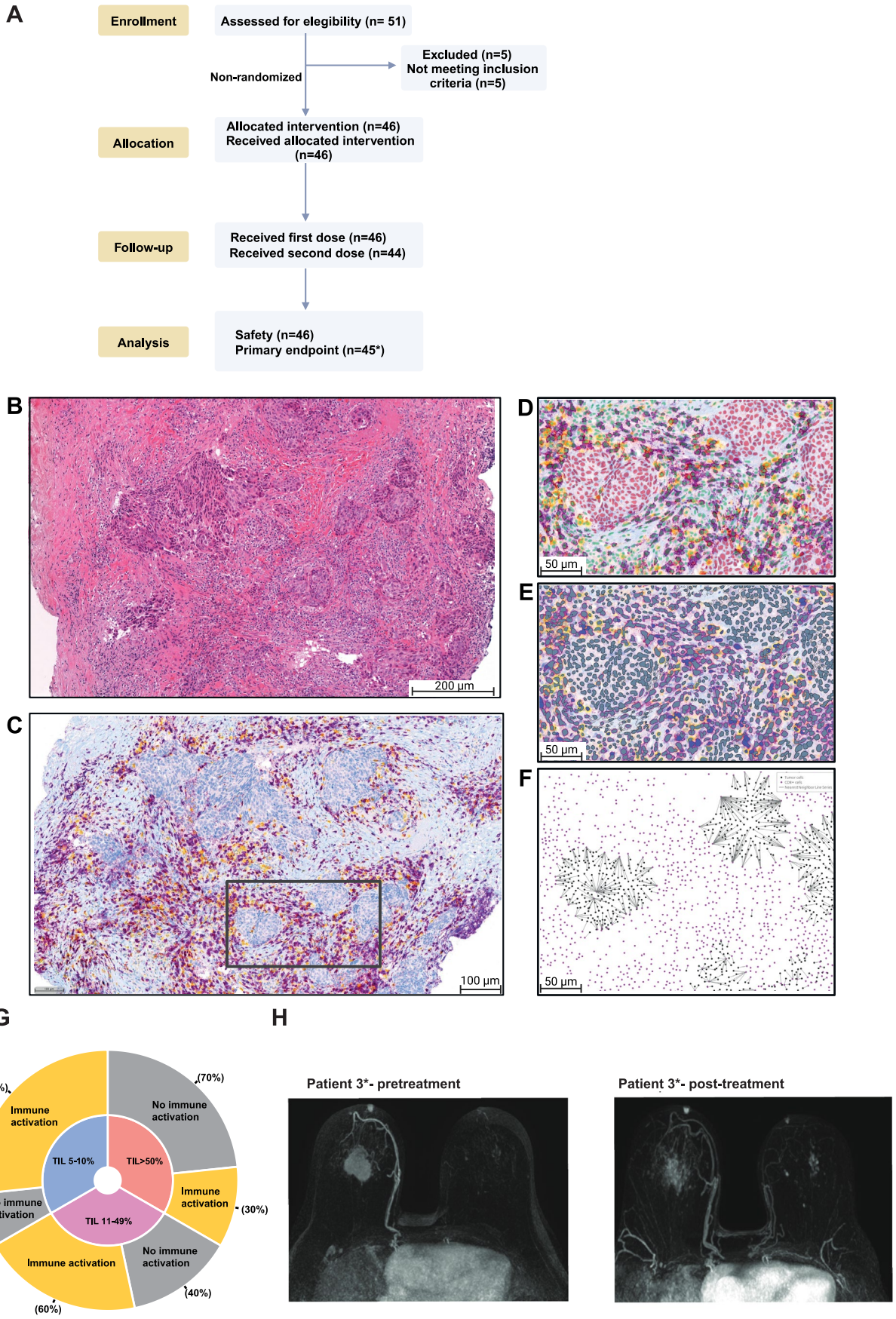
Extended data is available for this paper at https://doi.org/10.1038/s41591-024-03249-3.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41591-024-03249-3.

Correspondence and requests for materials should be addressed to Marleen Kok.

Peer review information *Nature Medicine* thanks Melinda Telli and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editors: Ulrike Harjes and Saheli Sadanand, in collaboration with the *Nature Medicine* team.

https://doi.org/10.1038/s41591-024-03249-3

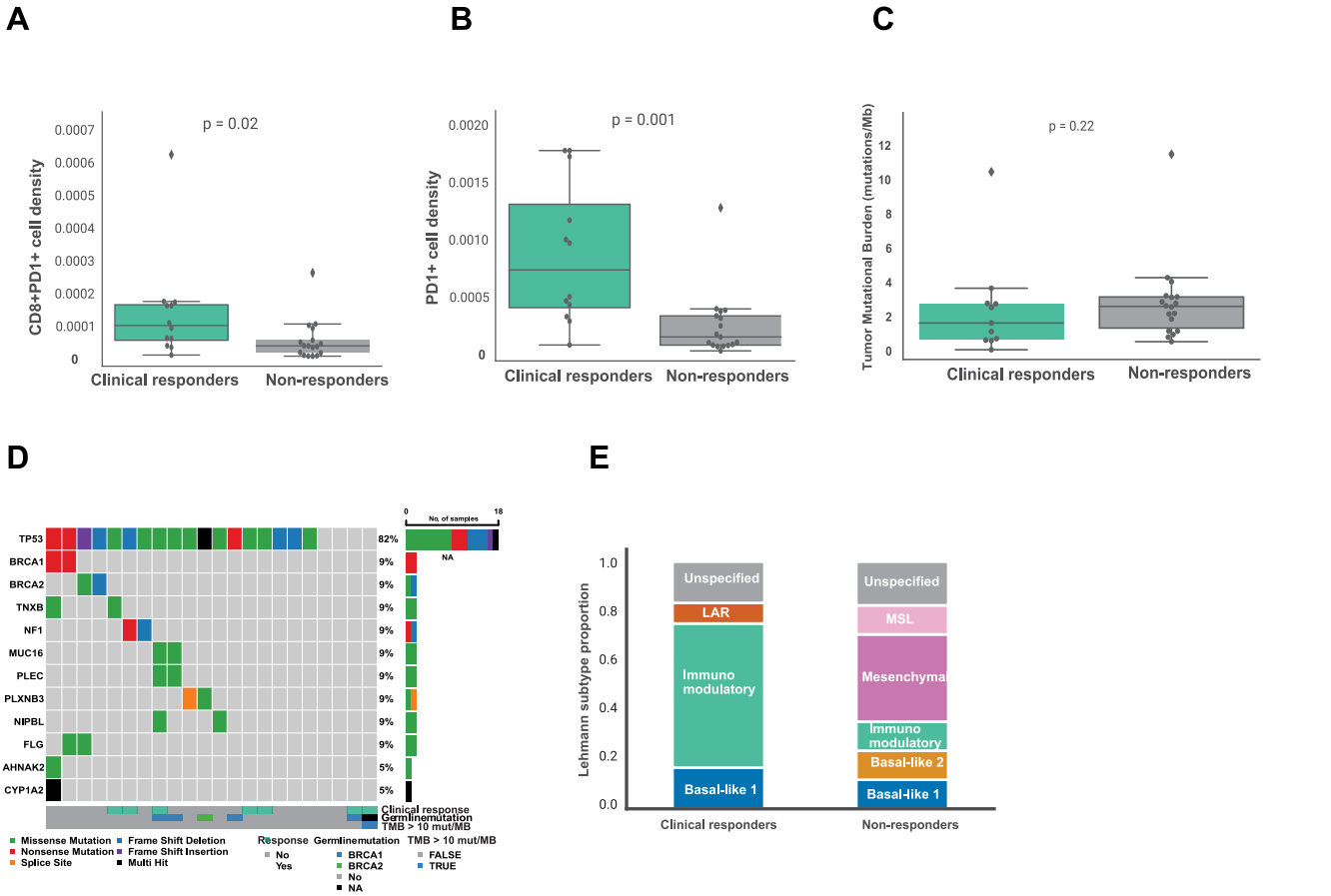


Extended Data Fig. 1 | See next page for caption.

https://doi.org/10.1038/s41591-024-03249-3

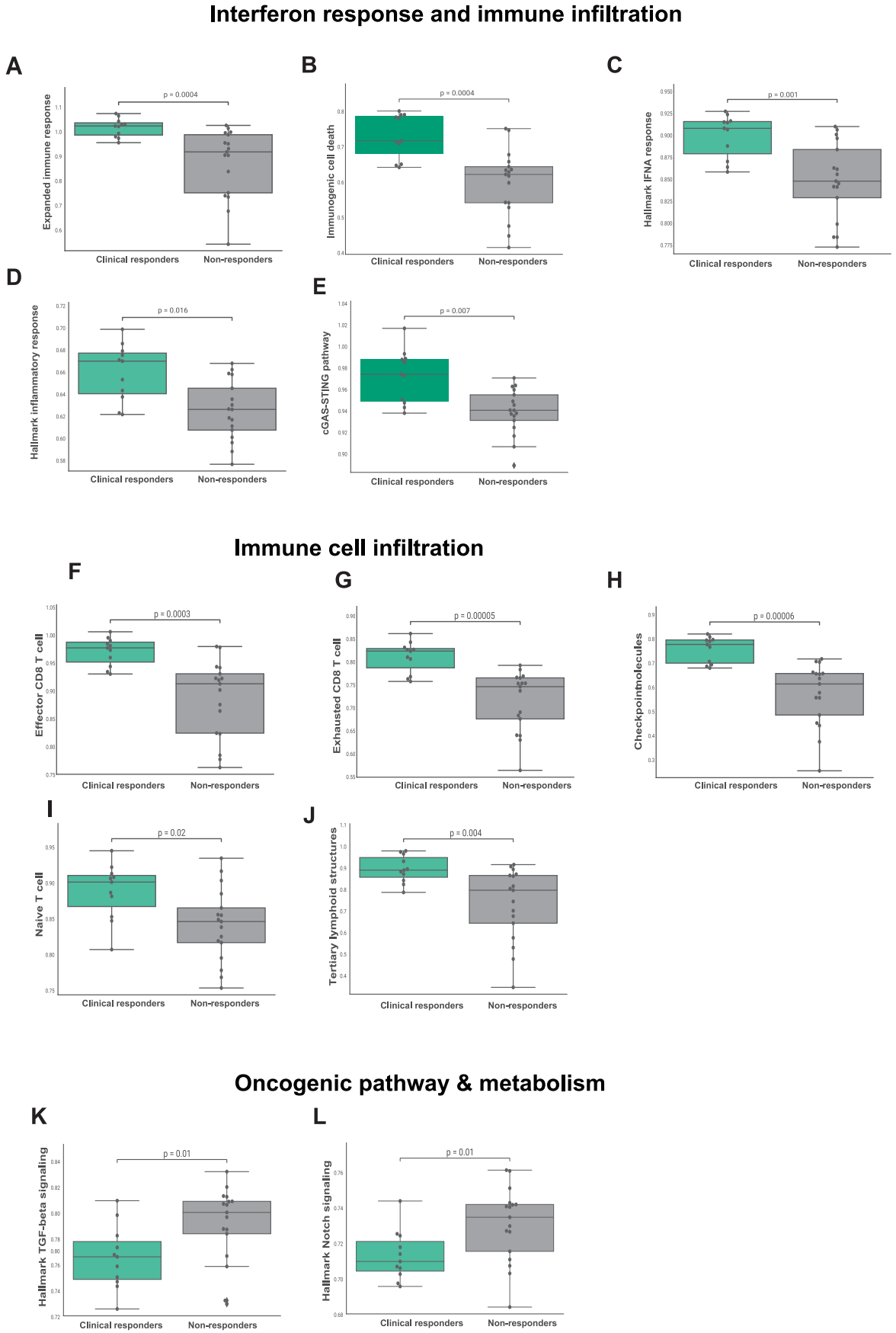
Extended Data Fig. 1 | IHC CD8 + T cell analysis. a. CONSORT Flow Diagram. Consort diagram of patients eligible, recruited, numbers followed up and included in analysis. *max 15 patients per cohort analyzed for primary end point **b**. H&E-stained image, corresponding to CD8/PD-1 stained tissue under **C**. **c**. Representative example of a CD8/PD-1 double-stained tissue (haematoxylin = blue, PD-1 = yellow, CD8 = purple). **d**. Representative example of the performance of the AI-based tumor cell classifier Tumor classification (red) and nontumor cells (green). **e**. Example of cell segmentation and tumor phenotype assignment. Cell with purple border = CD8+ cell, yellow border = PD-1+ cell, orange border = PD-1 + CD8+ cell. **f**. Corresponding distance analysis in the same tissue area as under **D** and **E**. The grey lines represent the shortest distance from a tumor cell to its nearest CD8 + T cell **g**. Proportions of patients reaching immune activation stratified according to TIL levels at inclusion in cohorts A and B. 10 patients had 5–10% TILs, 10 patients 11–49% TILs and 10 patients had 50% or more TILs. **h**. Pretreatment and post-treatment MRI images of patient #3 with a pathological complete response (pCR) at surgery after ICI only (cT2N0, ypT0N0). Figure **A** was created with BioRender.com. In **A-B**, one biopsy was analyzed per patient.

https://doi.org/10.1038/s41591-024-03249-3



Extended Data Fig. 2 | Baseline tumor microenvironment features and genomic profile of cohorts A and B. a. PD-1 + CD8 + T cell density in pretreatment biopsies of patients with and without who did and did not experience clinical response in cohorts A and B. n = 31 patients. **b.** PD-1+ cell density in pretreatment biopsies of patients with and without who did and did not experience clinical response in cohorts A and B. n = 31 patients. **c.** Tumor mutational burden (TMB) in pretreatment biopsies of patients with and without clinical response in cohorts A and B. n = 30 patients. Boxplots display a minimum (Q0), a maximum (Q4), a median (Q2) and the interquartile range. Data were analyzed by a two-sided Mann–Whitney test. **d.** Oncoplot of TMB (mutations per megabase (Mb)) and top mutated genes in cohorts A and B. **e.** Proportions of Lehmann et al. subtypes[33] in patients with and without clinical response in cohorts A and B. MSL, mesenchymal stem-like; LAR, luminal androgen receptor.

https://doi.org/10.1038/s41591-024-03249-3

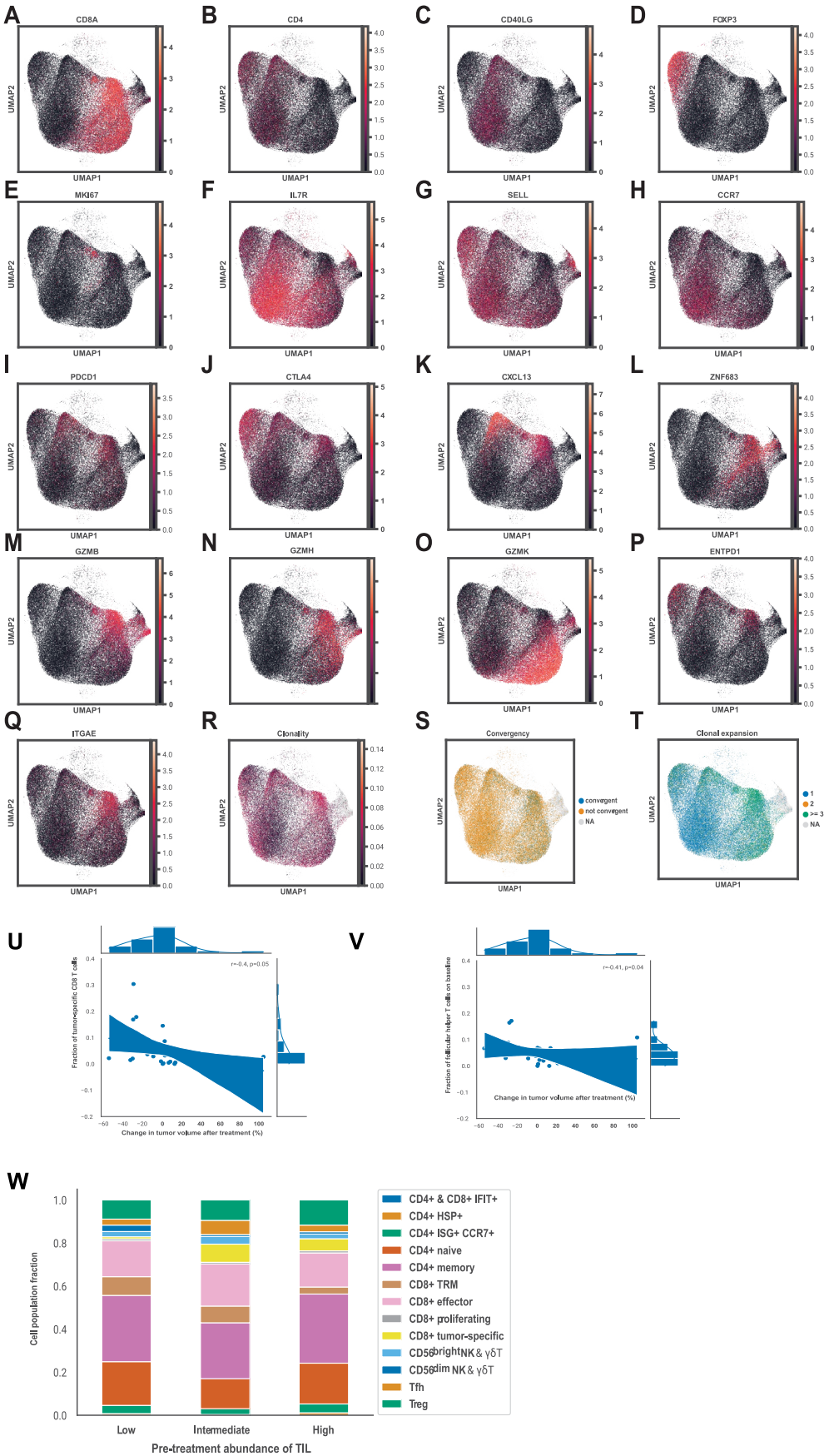


Extended Data Fig. 3 | See next page for caption.

https://doi.org/10.1038/s41591-024-03249-3

Extended Data Fig. 3 | Gene signatures in pretreatment biopsies associated with clinical response in cohorts A and B. a–**l**. Gene set expression scores in pretreatment biopsies of patients with and without clinical response in cohorts A and B. n = 28 patients. **a**. Expanded immune signature from Ayers et al.[56] **b**. Immunogenic cell death signature[57]. **c**. Hallmark IFNA response gene set. **d**. Hallmark inflammatory response gene set. **e**. cGAS-STING pathway gene set[58]. **f**. Effector CD8 + T cell gene set[59]. **g**. Exhausted T cell gene set[59]. **h**. Checkpoint molecules gene set[59]. **i**. Naive T cell gene set[60]. **j**. Tertiary lymphoid structures gene set[61]. **k**. Hallmark TGF-beta signaling gene set. **l**. Hallmark Notch signaling. In **A**–**L**, boxplots display a minimum (Q0), a maximum (Q4), a median (Q2) and the interquartile range. P values were derived using a two-sided Mann–Whitney test. Reported p values were significant after Benjamini–Hochberg (FDR) correction at 10% significance level.

https://doi.org/10.1038/s41591-024-03249-3

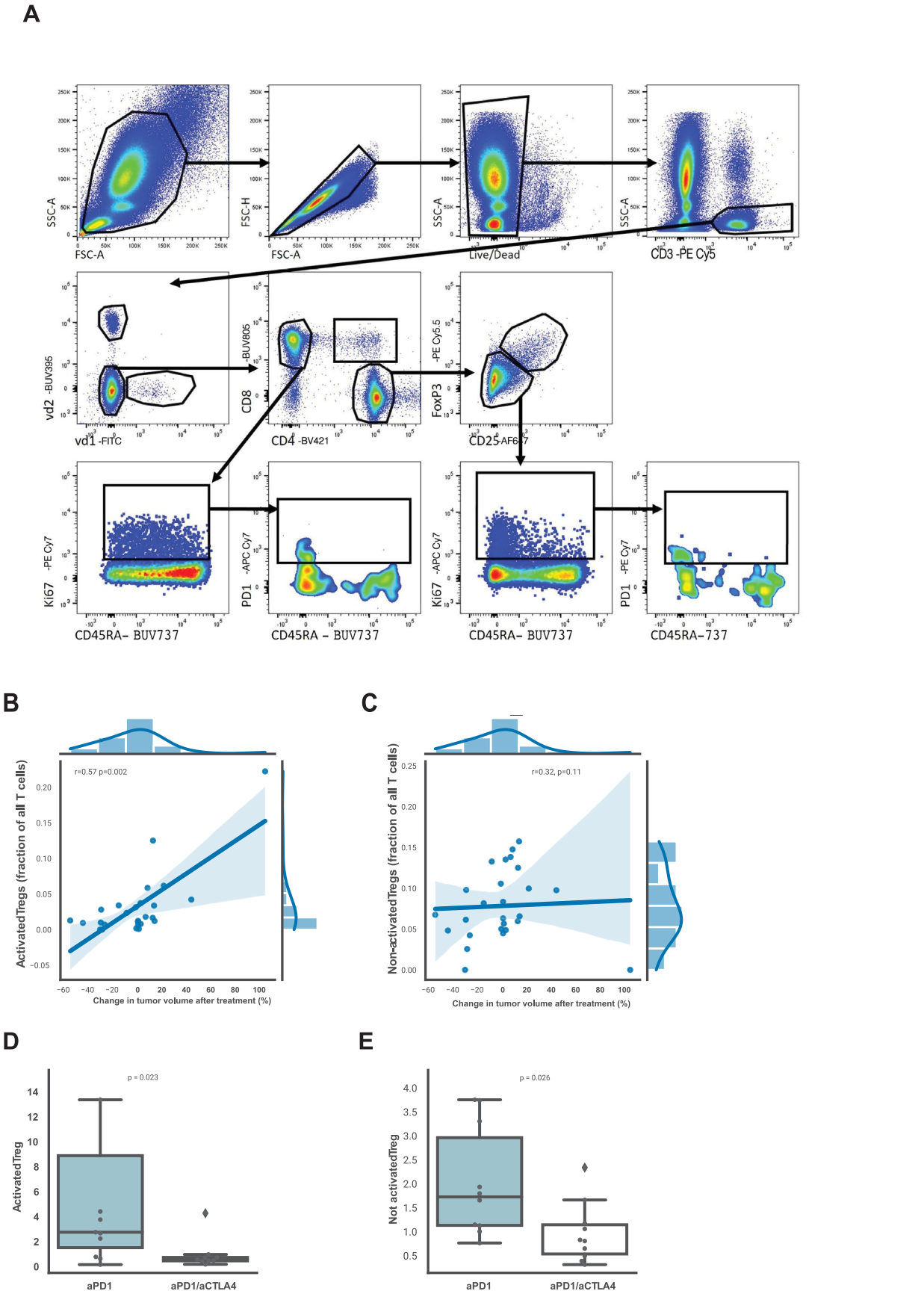


Extended Data Fig. 4 | See next page for caption.

https://doi.org/10.1038/s41591-024-03249-3

Extended Data Fig. 4 | Single-cell RNA-Seq pretreatment tumor microenvironment profile of the cohorts A and B. a-q. UMAP representations of the marker gene expression in the dataset. **a.** CD8A. **b.** CD4. **c.** CD40LG **d.** FOXP3 **e.** MKI67 **f.** IL7R. **g.** SELL. **h.** CCR7. **i.** PDCD1. **j.** CTLA4. **k.** CXCL13. **l.** ZNF683. **m.** GZMB. **n.** GZMH. **o.** GZMK. **p.** ENTPD1. **q.** ITGAE. **r.** UMAP representation of the T cell clonality in the dataset. **s.** UMAP representation of the T cell clone convergence in the dataset. **t.** UMAP representation of the T cell clonal expansion in the dataset. **u.** Fractions of tumor-reactive CD8 + T cells relative to all T cells in pretreatment biopsies of patients based on single-cell RNA-Seq data in relation to the change in tumor volume after treatment based on RECIST 1.1 in cohorts A and B. **v.** Fractions of Tfh cells relative to all T cells in pretreatment biopsies of patients based on single-cell RNA-Seq data in relation to the change in tumor volume after treatment based on RECIST 1.1 in cohorts A and B. **w.** Fractions of different T cell clusters relative to all T cells based on single-cell RNA-Seq data in pretreatment biopsies of patients who had low (5–10%), intermediate (11–49%) and high (>=50%) presence of tumor-infiltrating lymphocytes before treatment in cohorts A and B. In **U-V**, correlation was estimated with Spearman's rank correlation coefficient, two-sided, with 95% confidence interval for the regression estimate.

https://doi.org/10.1038/s41591-024-03249-3



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

https://doi.org/10.1038/s41591-024-03249-3

Extended Data Fig. 5 | Gating strategy used for the flow cytometry data analysis and activated and non-activated Tregs in cohorts A and B. a. Gating strategy used for the flow cytometry data analysis. **b**. Spearman correlation between fraction of activated Tregs and the change in tumor size on MRI (%). **c**. Spearman correlation between fraction of non-activated Tregs and the change in tumor size on MRI (%). Activated Tregs were defined as activated by the expression of CD137. **d**–**e**. Fold change in activated (**d**) and non-activated (**e**) Tregs after anti-PD-1 or anti-PD-1/anti-CTLA4 therapy. n = 22 patients. In B-C, correlation was estimated with Spearman's rank correlation coefficient, two-sided, with 95% confidence interval for the regression estimate. In D–E, boxplots display a minimum (Q0), a maximum (Q4), a median (Q2) and the interquartile range. P values were derived using a two-sided Mann–Whitney test.

https://doi.org/10.1038/s41591-024-03249-3

Extended Data Table 1 | Full list of adverse events

	A: Nivolumab (N=16)			B: Nivo+Ipi 4 wks (N=15)			Nivo+Ipi 6 wks (N=15)		
Adverse event	Grade 1-2	Grade 3	Grade 4	Grade 1-2	Grade 3	Grade 4	Grade 1-2	Grade 3	Grade 4
Worst grade adverse event per patient	12 (75%)	1 (6%)	0 (0%)	13 (86.6%)	1 (7%)	1 (7%)	9 (60%)	5 (33%)	1 (7%)
Endocrine disorders									
Thyroid dysfunction	7 (44%)	1 (6%)[a]	0 (0%)	9 (60%)	0 (0%)	0 (0%)	9 (60%)	0 (0%)	0 (0%)
Hypothyroidism	6 (38%)	0 (0%)	0 (0%)	7 (47%)	0 (0%)	0 (0%)	6 (40%)	0 (0%)	0 (0%)
Hyperthyroidism	4 (25%)	1 (6%)	0 (0%)	8 (53%)	0 (0%)	0 (0%)	9 (60%)	0 (0%)	0 (0%)
Adrenal insufficiency*	1 (6%)	0 (0%)	0 (0%)	1 (7%)	1 (7%)[b]	0 (0%)	2 (13%)	1 (7%)[f]	0 (0%)
Diabetes Mellitus (Immune mediated)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (7%)[b]	0 (0%)	0 (0%)	0 (0%)
Gastro-intestinal									
Abdominal pain	0 (0%)	0 (0%)	0 (0%)	1 (7%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Diarrhea	0 (0%)	0 (0%)	0 (0%)	1 (7%)	0 (0%)	0 (0%)	1 (7%)	0 (0%)	0 (0%)
Nausea	1 (6%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	2 (13%)	0 (0%)	0 (0%)
Colitis	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (7%)[g]	0 (0%)
Rectal Ulcer	0 (0%)	0 (0%)	0 (0%)	1 (7%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Laboratory test									
Elevated liver function tests	0 (0%)	0 (0%)	0 (0%)	1 (7%)	0 (0%)	0 (0%)	7 (47%)	0 (0%)	0 (0%)
Hyperphosphatemia	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (7%)	0 (0%)	0 (0%)
Hypophosphatemia	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (7%)	0 (0%)	0 (0%)
Lymphocyte count decreased	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (7%)	0 (0%)	0 (0%)
Troponine T increased	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (7%)	0 (0%)	0 (0%)
Immune related hepatitis	0 (0%)	0 (0%)	0 (0%)	2 (13%)	0 (0%)	0 (0%)	0 (0%)	2(13%)[c,d]	1 (7%)[h]
Musculo-skeletal									
Arthralgia	3 (19%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Myalgia	2 (12%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Back pain	1 (6%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Immune mediated polymyalgia rheumatica	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (7%)	0 (0%)	0 (0%)
Cardiopulmonal									
Chest-pain	0 (0%)	0 (0%)	1 (6%)[a]	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Ejection fraction decreased #	1 (6%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Pneumonitis	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (7%)	1 (7%)[e]	0 (0%)
Upper respiratory infection	0 (0%)	0 (0%)	0 (0%)	1 (7%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Cough	0 (0%)	0 (0%)	0 (0%)	1 (7%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Other									
Allergic reaction	1 (6%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Anemia	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	3 (20%)	0 (0%)	0 (0%)
Dry eye	1 (6%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (7%)	0 (0%)	0 (0%)
Dry mouth	2 (12%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	4 (27%)	0 (0%)	0 (0%)
Dry skin	1 (6%)	0 (0%)	0 (0%)	2 (13%)	0 (0%)	0 (0%)	1 (7%)	0 (0%)	0 (0%)
Fatigue	1 (6%)	0 (0%)	0 (0%)	1 (7%)	0 (0%)	0 (0%)	1 (7%)	0 (0%)	0 (0%)
Flu like symptoms	2 (12%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (7%)	0 (0%)	0 (0%)
Headache	2 (12%)	0 (0%)	0 (0%)	1 (7%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Infusion related reaction	2 (12%)	0 (0%)	0 (0%)	1 (7%)	0 (0%)	0 (0%)	1 (7%)	0 (0%)	0 (0%)
Skin rash	2 (12%)	0 (0%)	0 (0%)	1 (7%)	0 (0%)	0 (0%)	4 (27%)	0 (0%)	0 (0%)
Photosensitivity	1 (6%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Itching	0 (0%)	0 (0%)	0 (0%)	1 (7%)	0 (0%)	0 (0%)	1 (7%)	0 (0%)	0 (0%)
Cervical lymphadenopathy	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (7%)	0 (0%)	0 (0%)

This table sums all treatment-related adverse events for all patients receiving at least 1 dose of ICI. Percentages for all grade adverse events total more than 100% due to more AEs per patient. Some AEs may be related to one another and were reported as one entity, always reporting the highest grade; for example, thyroid dysfunction includes hyperthyroidism and hypothyroidism, adrenal insufficiency with hypophysitis were graded as adrenal insufficiency. Elevated liver function tests were defined as an increased ALAT, ASAT, GGT, alkaline phosphatase, bilirubin, aPPT and/or INR. If elevated liver function tests were occurring simultaneously, only one AE reporting the highest grade was reported. IR hepatitis was based on an increase in liver function tests requiring treatment with immunosuppressive medication. * All patients were classified as having secondary adrenal insufficiencies and all patients are still dependent on corticosteroid replacement. # Ejection fraction decreased during ddAC treatment, however IR-myocarditis could not be ruled out. A total of 8 patients from cohorts A, B and C developed grade 3/4 treatment-related adverse events. a) ID 14 (Cohort A) - Hyperthyroidism (grade 3) and chest pain (grade 3). Patient was diagnosed with hyperthyroidism leading to an orthostatic tremor requiring hospitalization of the patient. Symptoms were treated with beta-blockers. b) ID 36 (Cohort B) - Hyperglycemia with diabetic ketoacidosis (grade 4) and adrenal insufficiency (grade 3). The patient remains insulin-dependent and dependent on corticosteroid replacement therapy. c) ID 44 (Cohort C) - IR hepatitis (grade 3) with primary biliary cholangitis, treated with corticosteroids. d) ID 47 (Cohort C) - IR hepatitis (grade 3), patient did not receive corticosteroid treatment. e) ID 55 (Cohort C) - Pneumonitis (grade 3) with suspicion of pulmonary sarcoidosis, treated with corticosteroids. f) ID 59 (Cohort C)- Adrenal insufficiency (grade 3), still dependent on corticosteroid replacement therapy. g) ID 65 (Cohort C) - Colitis (grade 3), treated with corticosteroids. h) ID 66 (Cohort C)- IR hepatitis (grade 4), treated with corticosteroids.

https://doi.org/10.1038/s41591-024-03249-3

Extended Data Table 2 | Antibody overview

Human flow cytometry antibodies					
Antigen	**Fluorochrome**	**Clone**	**Dilution**	**Company**	**Catalog number**
CD3	PE Cy5	UCHT1	1:200	BD Bioscience	555334
CD4	BV421	RPA-T4	1:100	BD Bioscience	562424
CD8	BUV805	SK1	1:200	BD Bioscience	612754
Pan γδ TCR	PE	11F2	1:100	BD Bioscience	555717
vδ1	FITC	TS8.2	1:100	Thermofisher	TCR2730
vδ2	BUV395	B6	1:100	BD Bioscience	748582
FoxP3	PE Cy5.5	FJK-16s	1:50	Thermofisher	35-5773-82
CCR7	APC R700	150503	1:50	BD Bioscience	565868
CD45RA	BUV737	HI100	1:400	BD Bioscience	612846
CD25	AF647	BC96	1:100	BioLegend	302618
PD-1	APC Cy7	EH12.2H7	1:100	BioLegend	329922
CTLA-4	PE CF594	BNI3	1:200	BD Bioscience	562742
IL-17	PerCP Cy5.5	N49-653	1:50	BD Bioscience	560799
IFNγ	BV785	4S.B3	1:200	BioLegend	502542
TNFα	PE Cy7	Mab11	1:400	BioLegend	502930
CD27	BV786	L128	1:100	BD Bioscience	563327
TIGIT	PerCP Cy5.5	A151536	1:100	BioLegend	372718
Ki-67	PE Cy7	B56	1:50	BD Bioscience	561283
CTLA-4	PE CF594	PE/Dazzle594	1:200	BioLegend	369616

nature portfolio

Corresponding author(s): Marleen Kok

Last updated by author(s): *YYYY-MM-DD*

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	No software was used for data collection
Data analysis	For DNA sequencing, sequencing reads were aligned to the human reference GRCh38 (Ensemble, v. 105) using BWA90 0.7.17. Duplicated reads were marked using Picard MarkDuplicates 2.25.0, after which quality scores were recalibrated using GATK4 BaseRecalibrator 4.2.2.0. Single-nucleotide variants (SNVs) and short insertions and deletions (indels), were called using GATK4 Mutect2 4.2.2.0 on the tumor samples matched with germ line samples. Subsequently, variants were filtered by the PASS filter, and annotated using Ensembl Variant Effect Predictor. maftools 2.10.5 package was used for the analysis. Tumor mutational burden (TMB) was calculated by summarizing the total number of non-synonymous somatic mutations with a minimal variant allele frequency of 20%. Data were analyzed with Python 3.10.5 and R 4.1.3. Pandas 2.0.0 was used for data handling. maftools 2.10.5, Matplotlib 3.5.2, seaborn 0.12.2, and statannotations 0.5.0 were used for plotting. For ctDNA analysis, a proprietary bioinformatics tissue variant calling pipeline (Natera Inc) was used to select a set of 16 high-ranked, patient-specific, somatic, clonal single nucleotide variants (SNVs) from WES. The Signatera amplicon design pipeline was used to generate mPCR primer pairs for the given set of 16 variants. All paired-end reads were merged using Pear 0.9.8 software and mapped to the hg19 reference genome with Novoalign version 2.3.4 (http://www.novocraft.com/). Plasma samples with at least 2 variants with a confidence score above a predefined algorithm threshold were defined as ctDNA-positive. For single-cell RNA-Seq, after quality control, raw sequencing reads were aligned to the human reference genome GRCh38 and processed to a matrix representing the UMI's per cell barcode per gene using Cell Ranger (I0x Genomics, v2.0). The data were analyzed with scanpy 1.9.3 and Seurat v3. Cellbender 0.3.0 was used for eliminating technical artifacts, and cells above the quality cutoff of 0.5 were filtered out. Cells with mitochondrial RNA content >0.25, the number of genes <200 or >6000 and <400 counts were filtered out. After normalization, regression for the number of UMIs, percentage mt-RNA, sample ID, cell cycle, hypoxia, interferon content and cell stress was performed on the 2000 most variable genes followed by a principal component analysis. Next a UMAP was generated and clustering was performed at resolution 0.2 using

the 30 most informative components. Major cell types were identified based on canonical marker genes.
For the T cell subclustering, the T cells were selected from the full Seurat object and the analysis described above was repeated with 10 principal components based on the elbow plot and clusters were identified at a resolution of 0.6 and were annotated based on breast cancer tissue-specific marker genes. Cells expressing markers of other cell types (immunoglobulins, hemoglobin) were filtered out. PCA was calculated on highly variable genes with k=30. Clustering was performed with Phenograph with k=30. Cluster identification was performed based on canonical marker genes. Signature scores were calculated with sc.tl.score_genes. Groups were compared with sc.tl.rank_genes_groups, with method='wilcoxon' and use_raw=True. EnrichR was used for the pathway enrichment analysis. Activated Tregs were defined based on the level of CD137 gene expression >0.5 in the Treg cell population. PDl +Ki67+CD4+ cells were defined based on the level of MKl67 gene expression >0 in the Tfh cell population. Scirpyl 0.11.2 was used for the TCR analysis. Clonotypes were defined based on the amino acid structure. Clonality was calculated as (1 - normalized Shannon entropy). Data were analyzed with Python 3.10.5. Pandas 2.0.0 and numpy 1.22.4 were used for data handling. Matplotlib 3.5.2, seaborn 0.12.2, sc-toolboxlOl 0.12.3 and statannotations 0.5.0 were used for plotting.
For bulk RNA-Seq, data were aligned to GRCh38 with STAR 2.7.1a, with the twopassMode='Basic'. FPKM were obtained with RSeQC 4.0.0 FPKM_count.py and subsequently normalized to transcripts per million. Data quality was assessed with FastQC 0.11.5, FastQ Screen 0.14.0, the Picard CollectRnaSeqMetrics and RSeQC 4.0.0 read_distribution.py and read_duplication.py and were found to be suitable for the downstream analysis. TNBCtype was used for the Lehmann subtype classification. The Gseapy 1.0.3 ssgsea tool with the sample_norm_method='rank' was used for gene set signature scoring. For the signature analysis, p-values were significant after FDR correction (Benjamini–Hochberg) at 10% significance level. Data were analyzed with Python 3.10.5. Pandas 2.0.0 and numpy 1.22.4 were used for data handling. Matplotlib 3.5.2, seaborn 0.12.2 and statannotations 0.5.0 were used for plotting.

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

DNA and RNA sequencing data will be stored in the European Genome–Phenome Archive upon publication. Sequencing data and source data supporting the findings of this study will be made available from the corresponding author (m.kok@nki.nl) for academic use, within limitations of the provided informed consent. Data will not be made available for commercial use. A first response to the request will be sent in <4 weeks. Data requests will be reviewed by the corresponding author and Institutional Review Board of the NKI, and, after approval, applying researchers will have to sign a data transfer agreement with the NKI.

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	All participants are of female sex, by DNA and by identification
Reporting on race, ethnicity, or other socially relevant groupings	NA, we do not report on race, ethnicity or socially relevant groupings.
Population characteristics	Patients in cohorts A and B were eligible for enrollment if they were at least 18 years of age and had stage 1-111 (clinical tumor stage Tlc-3, nodal stage N0-3, according to the primary tumor regional lymph node staging criteria of the American Joint Committee on Cancer, 7th edition) triple negative breast cancer with confirmation of estrogen receptor and HER2 negativity (ER<l0% and HER2 0, 1 or 2 in the absence of amplification as determined by in situ hybridization) on a biopsy from the primary tumor in the breast; newly diagnosed, previously untreated disease; a WHO performance status score68 of 0 or 1 and adequate organ functions. The Tl Ls percentage needed to be 5% or more. Patients with concurrent ipsilateral, bilateral, or multifocal primary tumors were also eligible for enrollment. For cohort C, patients had to meet the same criteria, but the nodal stage had to be NO, tumor stage Tlc-T2, and Tlls had to be 50% or more. Exclusion criteria included history of immunodeficiency, autoimmune disease or conditions requiring immunosuppression (>10 mg daily prednisone or equivalent); other immunosuppressive medications intake within 28 days of study drug administration; chronic or recurring infections; occult breast cancer; fertility preservation due to breast cancer diagnosis; active hepatitis B virus or hepatitis C virus infection; clinically significant cardiovascular disease; previous systemic anti-cancer treatment.
Recruitment	The BELLINI trial (full title: Pre-operative Trial for Breast Cancer With Nivolumab in Combination With Novel 10; NCT03815890) is a single center, non-blinded, non-randomized, non-comparative phase II study designed to evaluate the feasibility and efficacy of checkpoint inhibition before regular neoadjuvant therapy or surgery in patients with primary breast cancer. Cohorts for prespecified breast cancer subgroups are opened in a sequential manner. Patients were asked to be screened for BELLINI at the routine diagnostic visits at the treatment center. Patients willing to participate may have had motivation to de-escalate chemotherapy, even though this was not the primary endpoint of the trial. We do not know if this could affect the outcomes.
Ethics oversight	All patients provided written informed consent before enrollment. This investigator-initiated trial was designed by the

Ethics oversight	Netherlands Cancer Institute (NKI). The trial was conducted in accordance with the protocol, Good Clinical Practice standards and the Declaration of Helsinki. The full protocol, amendments, and the informed consent form were approved by the medical ethical committee of the Netherlands Cancer Institute (NKI, Amsterdam).

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	For this exploratory, hypothesis-generating study, no formal sample size calculation was performed for efficacy because there was no data on the efficacy of neoadjuvant immunotherapy in breast cancer at the time of the design of this study. For cohorts A and B, the null hypothesis of a true immune activation in ≤30% of patients was tested against a one-sided alternative. For cohort C, design was identical with the exception of null hypothesis being pCR in ≤30% of patients tested against a one-sided alternative. For 80% power, at a one-sided significance level of 0.05, 15 patients were accrued per cohort to be evaluated in the first stage. If there were 5 or less responses among these 15 patients, the cohort was closed for futility. Otherwise, the cohort could be expanded with 31 additional patients, reaching a total of 46. We decided to publish after stage I, which was allowed by protocol, due to the observation that very early responses to ICI without chemotherapy are possible in TNBC, which warrants efforts to de-escalate therapy for a subset of patients, in contrast to the current therapy escalation for all TNBC patients. Median follow-up time was obtained using the reverse Kaplan-Meier method. Analyses were performed using R73 v.4.2.1.
Data exclusions	Patients in cohorts A and B were eligible for enrollment if they were at least 18 years of age and had stage I-III (clinical tumor stage T1c-3, nodal stage N0-3, according to the primary tumor regional lymph node staging criteria of the American Joint Committee on Cancer, 7th edition) triple negative breast cancer with confirmation of estrogen receptor and HER2 negativity (ER<10% and HER2 0, 1 or 2 in the absence of amplification as determined by in situ hybridization) on a biopsy from the primary tumor in the breast; newly diagnosed, previously untreated disease; a WHO performance status score68 of 0 or 1 and adequate organ functions. The TILs percentage needed to be 5% or more. Patients with concurrent ipsilateral, bilateral, or multifocal primary tumors were also eligible for enrollment. For cohort C, patients had to meet the same criteria, but the nodal stage had to be N0, tumor stage T1c-T2, and TILs had to be 50% or more. Exclusion criteria included history of immunodeficiency, autoimmune disease or conditions requiring immunosuppression (>10 mg daily prednisone or equivalent); other immunosuppressive medications intake within 28 days of study drug administration; chronic or recurring infections; occult breast cancer; fertility preservation due to breast cancer diagnosis; active hepatitis B virus or hepatitis C virus infection; clinically significant cardiovascular disease; previous systemic anti-cancer treatment. No patients were excluded from analysis after registration for the trial.
Replication	The current study was not attempted to be replicated as this is a human phase 2 clinical trial, stage I. However, the most important findings will be tested in the second stage of the trial.
Randomization	Patients were not randomized in this phase II non-randomized trial.
Blinding	Since the BELLINI is the first trial testing apD1/aCTLA-4 in early stage TNBC, physicians needed to know what patients received in order to ensure safety. Pathologists were blinded for clinical outcome and treatment group upon scoring of the biopsies.

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

Human flow cytometry antibodies
Antigen Fluorochrome Clone Dilution Company Catalog number
CD3 PE Cy5 UCHT1 1:200 BD Bioscience 555334
CD4 BV421 RPA-T4 1:100 BD Bioscience 562424
CD8 BUV805 SK1 1:200 BD Bioscience 612754
Pan γδ TCR PE 11F2 1:100 BD Bioscience 555717
vδ1 FITC TS8.2 1:100 Thermofisher TCR2730
vδ2 BUV395 B6 1:100 BD Bioscience 748582
FoxP3 PE Cy5.5 FJK-16s 1:50 Thermofisher 35-5773-82
CCR7 APC R700 150503 1:50 BD Bioscience 565868
CD45RA BUV737 HI100 1:400 BD Bioscience 612846
CD25 AF647 BC96 1:100 BioLegend 302618
PD-1 APC Cy7 EH12.2H7 1:100 BioLegend 329922
CTLA-4 PE CF594 BNI3 1:200 BD Bioscience 562742
IL-17 PerCP Cy5.5 N49-653 1:50 BD Bioscience 560799
IFNγ BV785 4S.B3 1:200 BioLegend 502542
TNFα PE Cy7 Mab11 1:400 BioLegend 502930
CD27 BV786
L128 1:100 BD Bioscience 563327
TIGIT
PerCP Cy5.5 A151536 1:100 BioLegend 372718
Ki-67
PE Cy7 B56 1:50 BD Bioscience 561283
CTLA-4
PE CF594 PE/Dazzle594 1:200 BioLegend 369616

Immunohistochemistry
PDL1 clone 22C3 (1/40 dilution, 1 hour at RT, Agilent/DAKO, Lot11654144)
PD1 clone NAT5 (Ready-to-Use, 32 minutes at 370C, Roche Diagnostics, Lot11654144).

The PD1-bound antibody was visualized using Anti-Mouse NP (Ventana Medical systems, Ready to Use dispenser, LotK09956) followed by Anti-NP AP (Ventana Medical systems, Ready to Use dispenser, LotJ23971)

CD8 clone C8/144B (1/200 dilution, 32 minutes at 37C, Agilent, Lot41527763).
PD-L1 (22C3, DAKO) and PD-1 (NAT105, Roche Diagnostics). CD8 was visualized using Anti-Mouse HQ (Ventana Medical systems, Ready to Use dispenser, LotK20711), followed by Anti-HQ HRP (Ventana Medical systems, Ready to Use dispenser, LotK22062)

Validation

All flow cytometry antibodies are commercially available and described in extended data table 2, including company, dilution and catalogue number for further information and background on each antibody.
For immunohistochemistry: CD8 and PD-L1 are diagnostic markers and were validated on human diagnostic tissue by the local pathology department. PD-1 has been validated for research purposes and validated on human tonsil and appendix tissue by the local pathology department. All antibodies used are commercially available and validated by manufacturer. Antibodies for flow cytometry were further validated for target species (human) using FMO or isotype controls where necessary. All antibodies were titrated to identify optimal staining concentration

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

NCT03815890

Study protocol

The full trial protocol can be requested for academic purposes at IRB of the Netherlands Cancer Institute and the corresponding author (Marleen Kok)

Data collection

All patients were enrolled in the Netherlands Cancer Institute, Amsterdam. The first patient was enrolled on the 19th of September 2019 and the last on the 24th of January 2023. The study is still recruiting new patients for new cohorts.

Outcomes

Cohorts A and B:
The primary endpoint for cohorts A and B is immune activation following two cycles of neoadjuvant ICI, defined as a 2-fold increase in CD8+ T cells assessed via immunohistochemistry and/or an increase in IFNG gene expression. High-quality paired biopsies are necessary for the evaluability of this primary endpoint.
As a secondary endpoint for cohorts A and B, we evaluated the clinical response.
Clinical response was defined as:
Radiological signs of response: At least a 30% decrease on MRI (partial response (PR) according to RECIST 1.1, not confirmed)
AND/OR

Pathological signs of response: Pathological response could be studied in biopsies from 28 patients due to the window of opportunity design. Absence of viable tumor after four weeks of therapy in the post-treatment biopsy was classified as a clinical response. For patients proceeding to surgery this was defined as partial or complete pathological response, according to the European Society of Mastology (EUSOMA criteria).
Cohort C:
The primary endpoint for cohort C is pathological complete response (pCR), defined as no viable tumor remaining in the breast and lymph nodes (ypT0N0)69. Major pathologic response (MPR, secondary endpoint) is a frequently used surrogate endpoint for efficacy in neoadjuvant trials evaluating immune checkpoint blockade across cancer types9,12,27. MPR was defined as ≤10% of residual viable tumor in the surgical specimen 18,70,71 or no viable tumor in the breast but residual tumor cells in the lymph nodes.
All Cohorts (A, B, C):
Secondary endpoints included feasibility, safety, and radiological response. Feasibility was determined based on any treatment-related complications that led to a delay in chemotherapy or primary surgery beyond six weeks from the start of therapy. All patients were closely monitored for adverse events (AEs) for 100 days after the administration of the last study treatment, following the Common Terminology Criteria for Adverse Events (CTCAE) v.572. In addition, we reported all immune-related adverse events in the first year of follow-up. Radiological response was assessed according to the RECIST 1.1 guidelines, but not confirmed.

Plants

Seed stocks

Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.

Novel plant genotypes

Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.

Authentication

Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.

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

Fresh blood samples were processed and analyzed within 24 hours after blood draw. Peripheral blood was collected in EDTA vacutainers (BD) and subjected to red blood cell lysis (lysis buffer: dH2O, NH4Cl, NaHCCO3, EDTA). Cells were suspended in PBS containing 0.5% BSA and 2mM EDTA and counted using the NucleoCounter NC-200 (Chemometec) automated cell counter. To obtain absolute white blood cell (WBC) counts per mL of human blood, the total amount of post-lysis cells was divided by the volume (mL) of blood obtained from the patient. For surface antigen staining, cells were first incubated with human FcR Blocking Reagent (1:100 Miltenyi) for 15 min at 4°C and then incubated with fluorochrome-conjugated antibodies for 30 min at 4°C. For intracellular antigen staining, cells were fixed with Fixation/Permeabilization solution 1X (Foxp3/ Transcription Factor Staining Buffer Set, eBioscience) for 30 min at 4°C and stained with fluorochrome-conjugated antibodies in Permeabilization buffer 1X (eBioscience) for 30 min at room temperature. Viability was assessed by staining with either 7AAD staining solution (1:10; eBioscience) or Zombie Red Fixable Viability Kit (1:800, BioLegend).

Instrument

Data acquisition was performed on an LSRII SORP flow cytometer (BD Biosciences)

Software

Diva software and data analysis was performed using FlowJo 10.6.2. Gating strategy is displayed in Extended Data Fig. 5A.

Cell population abundance

We report the percentage of a proliferating T cells within a pre-specified T-cell population in figure 3G and H; it;s therefore a relative measure and not an absolute count.

Gating strategy

Gating strategy is displayed in extended Data fig. 5A
) T cell panel gating strategy to identify proliferating T cells
T cells (CD3+,) vd1 negative and vd2 negative T cells
(CD3+, vd1-vd2+)
CD8 T cells (CD3+, vd1-vdd2-, CD8+, CD4-) --> next, Ki67High for proliferating fraction of CD8 T cells (CD3+, vd1-vdd2-, CD8+, CD4-, Ki67High) --> Last, PD1 positive fraction (CD3+, vd1-vdd2-, CD8+, CD4-, Ki67High, PD1High)
conventional CD4 T cells (CD3+, vd1-vd2-, CD8-, CD4+, FoxP3-) --> Next, Ki67High for proliferating fraction of CD4 T cells (CD3 +, vd1-vd2-, CD8-, CD4+, FoxP3-, Ki67high) --> last, PD1 positive fraction (CD3+, vd1-vd2-, CD8-, CD4+, FoxP3-, Ki67high,

PD1high)

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