Optimizing immunotherapy in locoregional and metastatic urothelial cancer

Dijk, N. van

Part II

Preoperative checkpoint immunotherapy in urothelial cancer

4

Preoperative ipilimumab plus nivolumab in locoregionally advanced urothelial cancer: the NABUCCO trial

Nick van Dijk, Alberto Gil-Jimenez, Karina Silina, Kees Hendricksen, Laura A Smit, Jeantine M de Feijter, Maurits L van Montfoort, Charlotte van Rooijen, Dennis Peters, Annegien Broeks, Henk G van der Poel, Annemarie Bruining, Yoni Lubeck, Karolina Sikorska, Thierry N Boellaard, Pia Kvistborg, Daniel J Vis, Erik Hooijberg, Ton N Schumacher, Maries van den Broek, Lodewyk F A Wessels, Christian U Blank, Bas W van Rhijn, Michiel S van der Heijden

Nature Medicine; October 2020

ABSTRACT

Preoperative immunotherapy with anti-PD1 plus anti-CTLA4 antibodies has shown remarkable pathological responses in melanoma1 and colorectal cancer2. In NABUCCO (ClinicalTrials.gov: NCT03387761), a single-arm feasibility trial, 24 patients with stage III urothelial cancer (UC) received two doses of ipilimumab and two doses of nivolumab, followed by resection. The primary endpoint was feasibility to resect within 12 weeks from treatment start. All patients were evaluable for the study endpoints and underwent resection, 23 (96%) within 12 weeks. Grade 3–4 immune-related adverse events occurred in 55% of patients and in 41% of patients when excluding clinically insignificant laboratory abnormalities. Eleven patients (46%) had a pathological complete response (pCR), meeting the secondary efficacy endpoint. Fourteen patients (58%) had no remaining invasive disease (pCR or pTisN0/pTaN0). In contrast to studies with anti-PD1/PD-L1 monotherapy, complete response to ipilimumab plus nivolumab was independent of baseline $CD8^+$ presence or T-effector signatures. Induction of tertiary lymphoid structures upon treatment was observed in responding patients. Our data indicate that combined CTLA-4 plus PD-1 blockade might provide an effective preoperative treatment strategy in locoregionally advanced UC, irrespective of pre-existing $CD8^+$ T cell activity.

INTRODUCTION

Recurrence rates after surgical resection of muscle-invasive UC are high, and many patients die from their disease (3). Although neoadjuvant cisplatin-based chemotherapy shows impressive responses, including 22–40% pCR (4) , the absolute survival benefit is only 5% at 5 years (5) , accompanied by substantial toxicity. Given the high rate of distant recurrences after surgical resection of UC, there is a great need for more effective systemic treatment approaches. Immune checkpoint inhibitors have shown durable responses in the metastatic setting (6–12). Response to checkpoint inhibitors appears higher when treating in first-line rather than second-line metastatic disease and when treating patients with only lymph node metastases (13), providing the rationale for immunotherapeutic treatment in earlier disease settings. A study testing neoadjuvant ipilimumab showed that this drug can be safely administered before cystectomy and does not preclude resection (14,15). Recently, promising pCR rates were observed in trials testing preoperative anti-PD-1/PD-L1 in UC of the bladder (16–18). However, complete responses were primarily observed in less advanced tumors and in tumors with pre-existing $CD8^+$ T cell immunity (16,18). In the NABUCCO study, we investigated whether the addition of anti-CTLA-4 to PD-1 blockade is feasible as preoperative treatment in locoregionally advanced (stage III) UC. Furthermore, we studied whether this combination treatment could broaden efficacy to more advanced tumors and tumors with limited baseline $CD8^+$ T cell immunity. Given the recently published work correlating B cells and tertiary lymphoid structures (TLSs) to immunotherapy response (19,20), we also assessed baseline B cell presence and TLS dynamics.

METHODS

Study design and patient population

NABUCCO (ClinicalTrials.gov: NCT03387761) is an investigator-initiated, prospective, single-arm trial testing the feasibility of attenuated preoperative ipilimumab 3 mg kg^{-1} (day 1), ipilimumab 3 mg kg^{-1} plus nivolumab 1 mg kg^{-1} (day 22) and nivolumab 3 mg kg^{-1} (day 43), followed by cystectomy or nephro/urethrectomy with appropriate lymph node dissection (21). A total of 24 patients with stage III resectable UC (cT3-4aN0M0 and cT1-4aN1-3M0) according to American Joint Committee on Cancer guidelines were enrolled at the Netherlands Cancer Institute (**Supplementary Fig. 1**). Patients were 18 years of age or older, were ineligible for cisplatin or refused cisplatin-based chemotherapy, were anti-CTLA-4/PD-1/PD-L1 naive, had a World Health Organization performance score of 0–1, and had diagnostic transurethral resection (TUR) blocks available. Key exclusion criteria included documented severe autoimmune disease, chronic

infectious disease and use of systemic immunosuppressive medications. All patients provided written informed consent. Surgery was scheduled in weeks 9–11 from the first drug infusion. Baseline staging was based on diagnostic pathology and thorax/abdomen/pelvis computed tomography (CT). CT-based response assessment occurred in weeks 8–10. Additional baseline and on-treatment magnetic resonance imaging (MRI) assessment was optional. Non-related adverse events and irAEs were graded and reported throughout the study. In case of serious irAEs after the second treatment cycle, the remaining dose of nivolumab was not administered so as to not interfere with the timing of surgery. The sponsor of the trial was the Netherlands Cancer Institute. Bristol-Myers Squibb provided the study drugs and funding for the trial. The trial is part of the Dutch Uro-Oncology Study Group. A data safety monitoring board was established, consisting of Thomas Powles (medical oncologist, Barts Cancer Centre) and Joost Boormans (urologist, Erasmus Medical Centre). This study was approved by the institutional review board (IRB) of the Netherlands Cancer Institute and was executed in accordance with the protocols and Good Clinical Practice Guidelines defined by the International Conference on Harmonization and the principles of the Declaration of Helsinki. After meeting the pre-specified primary and secondary endpoints in the current analysis ($n = 24$), the study was expanded with new dose exploratory cohorts (IRB approved in September 2019).

Study endpoints and statistics

The primary endpoint of this trial was feasibility, testing whether preoperative ipilimumab and nivolumab is feasible within 12 weeks from first infusion and does not delay surgical resection, as this is an endpoint that is clinically meaningful for this population. We set the desired resection rate by 12 weeks at 90% of patients. The lower statistical boundary for futility was set at 60%. For 24 patients and one-sided $\alpha = 0.025$, the power of rejecting 60% resection rate, under the alternative of 90%, is 91%. Treatment efficacy (pCR) was selected as a secondary endpoint and defined by the percentage of pCR (complete absence of neoplastic cells, pT0N0) at surgical resection. For the efficacy endpoint, we considered the disease burden of this cohort, which was higher than a typical neoadjuvant cohort, and cisplatin ineligibility/refusal. We determined that a 40% pCR rate would be desirable, whereas 14% or lower would not warrant further investigation of treatment. Under those assumptions, 24 evaluable patients provided 90% power to detect treatment efficacy with a one-sided α of 0.05. At least seven responders are required. Histopathological assessment was performed on the resected primary tumor and lymph node specimens, and pathological staging was done according to international standards and protocols. In addition, percentage of residual vital tumor tissue was analyzed and calculated. Other secondary endpoints included translational endpoints and safety and are extensively described in the study protocol. Unless specified other-

wise, the Wilcoxon signed-rank test was used for pairwise comparisons. R 3.4.4 was used for statistical analysis, together with the packages tidyverse 1.2.1 and ggpubr 0.2.1.

PD-L1 assessment, multiplex immunofluorescence analyses and immunohistochemistry

Baseline PD-L1 immunohistochemistry was performed on formalin-fixed paraffin-embedded (FFPE) sections by a PD-L1 IHC 22C3 1/40 dilution pharmDx qualitative immunohistochemical assay on a DAKO Autostainer 48 system at the Tergooi hospital laboratory. After staining, PD-L1 and hematoxylin and eosin slides were uploaded to Slide Score (www.slidescore. com) for manual scoring. An experienced uro-pathologist determined the CPS, and PD-L1 positivity was qualified as CPS > 10 (16). Analysis of tumor immune cell infiltrates (anti-CD3 (1:400 dilution Clone P7, Thermo Fisher Scientific), anti-CD8 (1:100 dilution Clone C8/144B, Dako), anti-CD68 (1:500 dilution Clone KP1, Dako), anti-FoxP3 (1:50 dilution Clone 236A/47, Abcam), anti-CD20 (1:500 dilution Clone L26, Dako) and anti-PanCK (1:100 dilution Clone AE1AE3, Thermo Fisher Scientific) was performed by multiplex immunofluorescence technology on a Ventana Discovery Ultra automated stainer using PerkinElmer opal seven-color dyes. In short, 3-µm FFPE sections were cut and heated at 75 °C for 28 min and subsequently deparaffinized in EZ Prep solution (Ventana Medical Systems). Using Cell Conditioning 1 (CC1, Ventana Medical Systems), heat-induced antigen retrieval was conducted at 95 °C for 32 min. Further analysis was done by VECTRA image acquisition (Akoya Biosciences, v3.0) and HALO (Indica Labs, v2.3) image analysis. Tumor and stroma regions were classified by HALO automated tissue segmentation. TLS composition was analyzed in tissue sections by seven-plex multiplex immunofluorescence[27] using multispectral microscopy and the following antibodies and dilutions: CD21 (1:5,000 2G9, Leica), DC-LAMP (1:1,000 1010E1.01, Dendritics), CD23 (1:1,000 SP3, Abcam), PNAd (1:5,000 MECA-79, BioLegend), CD20 (1:5,000 L26, Dako) and CD3 (1:1,000 SP7, Thermo Fisher Scientific). Area segregation was done by the Inform tissue segmentation algorithm. Additional information on antibodies used is provided in the Life Sciences Reporting Summary.

For immunohistochemistry (including co-stainings), slides were cut at 3 µm and dried overnight. Slides were transferred to the Ventana Discovery Ultra Autostainer. Slides were then deparaffinized within the instrument, and antigen retrieval was carried out at 95 °C for 64 min. For the double staining of CD20 (yellow) followed by CD27 (purple), the CD20 was detected in the first sequence using clone L26 (1:800, 32 min at 37 °C; Agilent). CD20-bound antibody was visualized using Ready-to-Use anti-mouse NP (Ventana Medical Systems) for 12 min at 37 °C, followed by anti-NP AP (Ventana Medical Systems) 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, CD27 was detected

using clone EPR8569 (1:3,000 dilution, 32 min at 37 °C; Abcam). CD27 was visualized using Ready-to-Use anti-mouse HQ (Ventana Medical Systems) for 12 min at 37 °C, followed by anti-HQ horseradish peroxidase (Ventana Medical Systems) 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). For the double staining of CD4 (yellow) followed by Ready-to-Use BCL6 (purple), the CD4 was detected in the first sequence using clone SP35 (1:25, 32 min at 37 °C; Cell Marque). CD4-bound antibody was visualized using the Roche Diagnostics Ready-to-Use anti-mouse NP (Ventana Medical Systems) for 12 min at 37 °C, followed by anti-NP AP (Ventana Medical Systems) 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, BCL6 was detected using clone GI191E/A8 (Ready-to-Use, Ventana Medical Systems; 32 min at 37 °C). BCL6 was visualized using the Roche Diagnostics Ready-to-Use anti-mouse HQ (Ventana Medical Systems) for 12 min at 37 °C, followed by the Roche Diagnostics Ready-to-Use anti-HQ horseradish peroxidase (Ventana Medical Systems) 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). CXCL13 was stained using the polyclonal goat R&D Systems antibody Clone AF801, in a 1:400 dilution. CXCL13-bound antibody was visualized using the Roche Diagnostics Ready-to-Use OmniMap anti-goat horseradish peroxidase (Ventana Medical Systems) for 12 min at 37 °C, followed by the Chromomap DAB Detection Kit. Slides were counterstained with hematoxylin and bluing reagent (Ventana Medical Systems). All immunohistochemistry slides were uploaded to SlideScore for manual scoring. Additional information on antibodies used is provided in the Life Sciences Reporting Summary.

For biomarker analysis, tumors with CR (defined as pCR, pTis or pTaN0) were compared to non-CR tumors. We included noninvasive disease in the CR definition, as this is generally thought to be cured by surgery (30). Figure panel colors are optimized for colorblind readers.

DNA sequencing

DNA and RNA were isolated in parallel from baseline and on-treatment FFPE tumor material using the Qiagen AllPrep FFPE DNA/RNA Kit. Germline DNA was extracted from peripheral blood mononuclear cells using the QIAsymphony DSP DNA Midi kit. Tumor depositions in the lymph nodes were isolated by using laser microdissection. DNA sequencing was carried out following the Human IDT Exome Target Enrichment Protocol. Covaris shearing was used to fragment the DNA to 200–300 base pairs. KAPA HTP DNA Library Kit was used for library preparation. IDT Human Exome capture V1.0 (xGen Exome Research Panel v1.0) was used for exome enrichment. Libraries were sequenced

with 100 base pairs paired-end reads using the Hiseq 2500 (Illumina) with a high output mode PE 100. Burrows–Wheeler aligner 37 was used to align the raw reads to the human reference genome GRCh38. Duplicated reads were marked using MarkDuplicates (http://broadinstitute.github.io/picard), and, lastly, GATK BaseRecalibrator was used to recalibrate the quality scores. Somatic single-nucleotide variants (SNVs) and short insertions and deletions (indels) were called for tumor samples matched with germline samples using Strelka 2.9.2 (31). Only variants with a high reliability score (HighEVS) and allele frequency greater than 5% were maintained and annotated using SnpEff v4.3t (build 201711-24 10:18). TMB was assessed as the number of somatic variants annotated as non-synonymous, non-intronic or non-intergenic. For biomarker discovery, we retrieved variants classified as indels, missense, frameshift, stop codon gain or loss, splice acceptors or donors, transcription ablation, exon loss and structural interaction variant. Copy number calling was done using CNVkit v0.9.6 (32), and copy number ratios were estimated against a pooled normal reference. Gene copy number ratios (cnr) were calculated with a weighted average copy number ratio per gene. Deep deletions were defined as $\log_2(\text{cnr}) < -0.7$, shallow deletions as $-0.7 \leq \log_2(\text{cnr}) < -0.5$ and amplifications as $\log_2(\text{cnr}) > 1$. All genomic data were analyzed using the R packages VariantAnnotation v1.24.5 and ComplexHeatmap v1.17.1. For associations between somatic alterations in DDR genes (22) and response, we excluded variants with a clinical significance labeled as Benign or Likely Benign, variants previously reported as population/germline single-nucleotide polymorphisms (SNPs) (TOPMED and CAF annotations with alternate allele frequency >5%) and variants annotated as Tolerated with the SIFT predicted score. We explored alterations in DNA mismatch repair machinery (*MLH1*, *MSH2*, *MSH6*, *PMS1* and *PMS2*), nucleotide excision repair (*ERCC2*, *ERCC3*, *ERCC4* and *ERCC5*), homologous recombination (*BRCA1*, *MRE11A*, *NBN*, *RAD50*, *RAD51*, *RAD51B*, *RAD51D*, *RAD52* and *RAD54L*), Fanconi anemia (*BRCA2*, *BRIP*, *FANCA*, *FANCC*, *PALB2*, *RAD51C* and *BLM*), checkpoint (*ATM*, *ATR*, *CHEK1*, *CHEK2* and *MDC1*) and others (*POLE*, *MUTYH*, *PARP1* and *RECQL4*). Lymph node and primary tumors with discrepant response were sequenced using three different pools to reach >150× exomic coverage. Somatic SNVs and short indels for each tumor–germline and lymph node–germline sample pairs were called using Strelka. For each patient, we retrieved all variants that had a high reliability score (HighEVS) in variant calling for at least one of the tumor or lymph node samples, recomputed the allele frequencies from the bam files and filtered by allele frequency greater than 5%.

RNA sequencing

The TruSeq Stranded mRNA Sample Prep Kit from Illumina was used to generate strand-specific libraries according to the manufacturer's instructions. Twelve cycles of polymerase chain reaction (PCR) were done to amplify the 3' end-adenylated and adapter-ligated RNA. Quality and quantity of the total RNA from FFPE was assessed

by the 2100 Bioanalyzer using a 7500 Nano chip (Agilent). The percentage of RNA fragments with >200-nt fragment distribution values (DV200) was determined using the region analysis method according to the manufacturer's instructions (Illumina, technical-note-470-2014-001). The libraries were subsequently diluted and pooled equimolar in a multiplex sequencing pool. Storage was done at −20 °C. The pooled libraries were enriched for target regions using the probe Coding Exome Oligos set (45 MB) according to the manufacturer's instructions (Illumina, no. 1000000039582v01). Briefly, complementary DNA (cDNA) libraries and magnetic bead-bound capture probes were combined, followed by hybridization using a denaturation step of 95 °C for 10 min and an incubation step from 94 °C to 58 °C having a ramp of 18 cycles with 1-min incubation and 2 °C per cycle. The enriched fraction was subjected to two stringency washes, an elution step and a second round of enrichment, followed by a cleanup using AMPure XP beads (Beckman, A63881) and PCR amplification of ten cycles. The target enriched pools were analyzed on a 2100 Bioanalyzer using a 7500 chip (Agilent), diluted and subsequently pooled equimolar into a multiplex sequencing pool. A HiSeq 2500 System was used to sequence libraries with 65 base pair single-end reads in Illumina high output mode using V4 chemistry (Illumina). HISAT2 aligner was used to align the raw reads to the human reference genome GRCh38 and to quantify gene expression using default parameters. For quality control, samples with a low cDNA yield, low sequencing coverage (reads <9 million) and fraction of aligned reads to human reference genome less than 85% were filtered out (six baseline samples). Genes with low gene expression variability (null counts in >80% of the samples and/or read counts mean <1) were filtered out. EdgeR v3.20.9 (33) was used to calculate the normalization factors to account for variable library sizes. Only genes with maximum counts per million (CPM) greater than 1 and with a CPM greater than 1 in at least three samples were considered for biomarker analysis (~7,000 genes). Differential expression (DE) was tested using limma v3.34.9 (34). First, we used the voom function to account for different weights from the mean–variance relationships. Then, we built linear models to test DE between different baseline/ on-treatment response groups. We corrected for multiple hypothesis testing by modeling the density-based false discovery rate (FDR) from the *P* value distribution using fdrtool (35). Genes that scored an FDR of less than 8% were considered significant.

Bioinformatic algorithm to detect TLS in multiplex immunofluorescence

TLS clusters in multiplex immunofluorescence (CD3, CD8, FoxP3, CD20 and CD68) data were detected using DBSCAN v1.1-4. The cell phenotypes $CD20^{+}$, $CD3^{+}CD8^{-}FoxP3^{-}$ and $CD3^{+}FoxP3^{+}CD8^{-}$ were used to identify putative TLS clusters in stromal regions. These same three phenotypes were quantified residing either inside TLS-detected regions or in stromal/tumoral regions. Mann–Whitney U tests were used to test for difference in TLS composition between pre- and post-treatment samples of phenotypes

frequencies $CD20^+$, $CD3^+CD8^-FoxP3^-$ and $CD3^+CD8^-FoxP3^+$. The difference in TLS Treg ($CD3^+CD8^-FoxP3^+$)/naive T cell ($CD3^+CD8^-FoxP3^-$) ratio was explored with a Mann–Whitney U test between pre/post samples of CR and non-CR tumors. The $CD3^+CD8^-FoxP3^-$ and CD20 location ratios (in TLS/in tumor+stroma) was tested for difference in ratio between pre/post samples in CR and non-CR tumors with a Mann–Whitney U test.

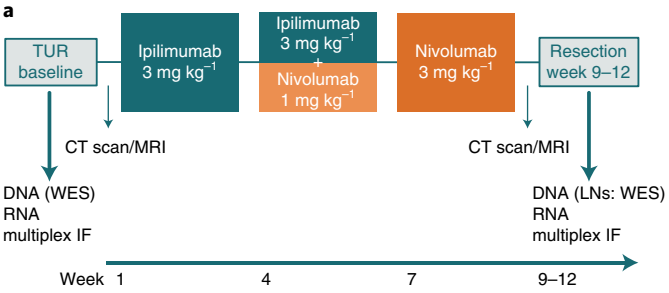
RESULTS

In the NABUCCO study **(Fig. 1a)**, 24 patients with stage III UC were enrolled **(Fig. 1b)** between February 2018 and February 2019 and treated with ipilimumab 3 mg kg^{-1} (day 1), ipilimumab 3 mg kg^{-1} plus nivolumab 1 mg kg^{-1} (day 22) and nivolumab 3 mg kg^{-1} (day 43), followed by resection[21]. The median postoperative follow-up was 8.3 months (interquartile range (IQR), 4.7–13.1 months). Fourteen (58%) patients had node-negative disease at baseline (cT3-4aN0), and ten (42%) patients had baseline lymph node metastases (cT2-4aN1-3). Patients were ineligible for cisplatin or refused cisplatin-based chemotherapy.

All 24 patients were evaluable for the primary endpoint, which was surgical resection within 12 weeks after initiation of preoperative therapy. All patients had surgical resection, despite inclusion of patients with bulky/extensive disease. Twenty-three (96%; 95% confidence interval (CI), 79–100%) patients underwent resection within 12 weeks, thus meeting the study's primary endpoint. One patient had a delay of 4 weeks due to immune-related hemolysis. Grade 3–4 immune-related adverse events (irAEs) occurred in 55% of patients (**Supplementary Table 1**) and in 41% of patients when excluding clinically insignificant laboratory abnormalities (mainly lipase elevation). Eighteen (75%) patients received all three treatment cycles. Six (25%) patients received only two cycles owing to irAEs. There was no treatment-related mortality.

After ipilimumab plus nivolumab, most resected primary tumors displayed extensive tumor regression at histopathological examination **(Fig. 2a)**. Eleven (46%; 95% CI, 26–67%) patients had pCR (pT0N0), meeting the pre-specified secondary efficacy endpoint. Fourteen (58%; 95% CI, 37–77%) patients had no residual invasive cancer (pCR or small CIS/pTa focus) after treatment. Two additional patients (8%) achieved a major pathological response (MPR), defined as less than 10% residual vital tumor and pN018, exhibiting extensive immune infiltration **(Fig. 2b)**. A 50% (95% CI, 29–71%) pCR rate was observed in patients without baseline lymph node metastases (cT3-4aN0), compared to 40% (95% CI, 12–73%) pCR in patients with clinically suspected node-positive disease (cT24aN1-3). Interestingly, four patients achieved pCR/MPR in the primary tumor, whereas a resistant

a



b

Baseline characteristics	Total (n = 24)
Median age, years (range)	65 (50–81)
Male sex, n (%)	18 (75%)
Clinical T stage, (%)	
▪ cT3-4aN0M0	14 (58%)
▪ cT2-4aN1–3M0[1]	10 (42%)
Cisplatin eligibility status n (%)	
▪ Cisplatin ineligible	13 (54%)
▪ Refusal of cisplatin-based chemotherapy[2]	11 (46%)
WHO, n (%)	
▪ WHO 0	21 (88%)
▪ WHO 1	3 (12%)
Median tumor volume, cm^3 (range)	
▪ Primary tumor[3]	19.9 (0–153)
PD-L1 combined positive score (CPS)	
▪ Median PD-L1 CPS, %	32
▪ PD-L1 CPS > 10, n (%)	15 (63%)
▪ PD-L1 CPS < 10, n (%)	9 (37%)

[1]One patient had bladder + upper tract disease; all others had bladder-only disease.
[2]Cisplatin-eligible according to Galsky criteria. Patients had relative contraindications or refused cisplatin-based chemotherapy
[3]Assessed through multiparametric magnetic resonance

Fig. 1 | NABUCCO study design and baseline study population characteristics. **a,** NABUCCO study treatment and time points of radiological assessment and sample collection. **b,** Baseline characteristics of patients enrolled in the study.

CT, computed tomography; IF, immunofluorescence; LN, lymph node; MRI, magnetic resonance imaging; TUR, transurethal resection; WHO, World Health Organization.

micrometastasis was observed in a concurrent lymph node. No specific genetic resistance mechanisms could be identified in discrepant mutations between the primary tumor and unresponsive lymph nodes upon whole-exome sequencing (WES; **Extended Data Fig. 1**). At the time of analysis, two patients (both non-pCR) had relapsed; one of these patients died of metastatic disease **(Fig. 2c)**.

We examined several published biomarkers for immunotherapy response in our cohort. For these biomarker analyses, we compared complete response (CR, defined as pCR or CIS/pTa) to non-CR. The CR rate in PD-L1-positive tumors (combined positive score (CPS) >10) was 73% (95% CI, 45–92%), compared to 33% (95% CI, 7-70%) in PD-L1-negative tumors (P = 0.15). Analysis of somatic mutations by WES **(Extended Data Fig. 2)** of pre-treatment tumor tissue and germline DNA showed a trend toward a higher tumor mutational burden (TMB) in tumors achieving CR to ipilimumab plus nivolumab compared to non-CR tumors (P = 0.056; **Fig. 3a**). We also assessed mutations in a set of DNA damage response (DDR) genes (22). Alterations in DDR genes were more frequently observed in responding tumors than in non-CR tumors (P = 0.03; **Fig. 3b**). Recent data from studies in advanced and localized UC indicate that transforming growth factor-beta (TGF-β) signaling is associated with atezolizumab unresponsiveness (18,23). In line with these data, we found a TGF-β expression signature to be associated with non-response **(Fig. 3c)**. This result suggests that TGF-β-mediated inhibition of the anti-cancer immune response cannot be overcome by the addition of anti-CTLA-4 to PD-1 blockade.

Previous studies in UC found that response to preoperative anti-PD1/PD-L1 monotherapy is primarily observed in tumors exhibiting pre-existing CD8+ T cell immunity (16,18). This restriction could limit the general applicability of anti-PD1/PD-L1 monotherapy in the preoperative setting. We assessed whether the addition of anti-CTLA-4 enables response in tumors with limited pre-existing T cell immunity in our cohort. Using quantitative multiplex immunofluorescence, we observed no correlation between baseline CD8+ T cell density and response to ipilimumab plus nivolumab **(Fig. 3d)**. The CR rate in both the lowest and highest CD8+ cell density quartile was 67%. We further explored the correlation between pre-existing immunity and response by transcriptome analysis of baseline tumor tissue. No difference was observed between CR and non-CR tumors in baseline interferon gamma (IFN-γ), tumor inflammation signature (TIS) and CD8+ T cell effector signatures **(Fig. 3e)**. Thus, our data suggest that the addition of anti-CTLA-4 to PD-1 blockade can induce pCR in tumors irrespective of baseline CD8+ T cell immunity.

Baseline B cell presence has been associated with immunotherapy response and prognosis in several malignancies (19,20,24), although other studies suggest immune-suppressive capacities of the intratumoral B cell compartment (25,26). Upon exploratory

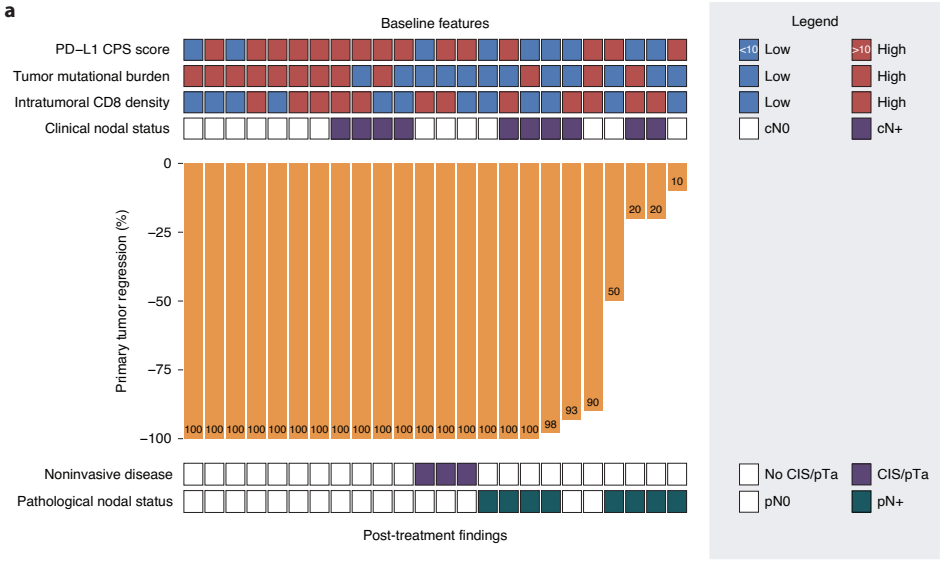
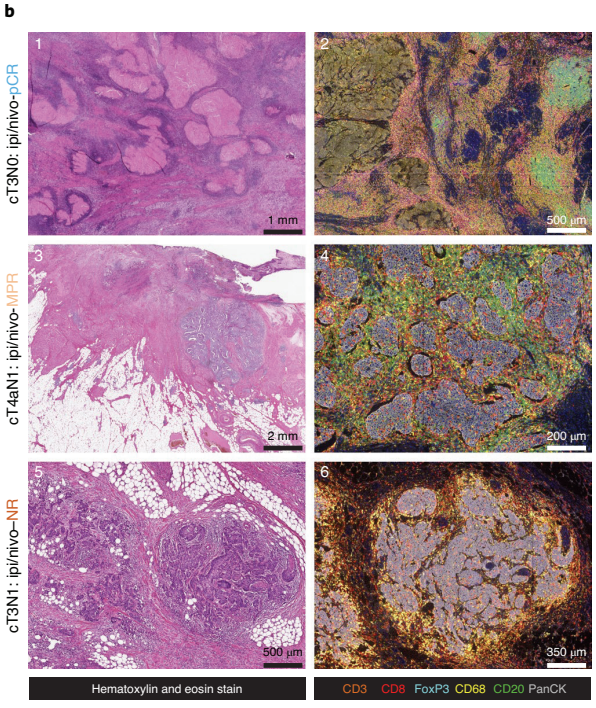
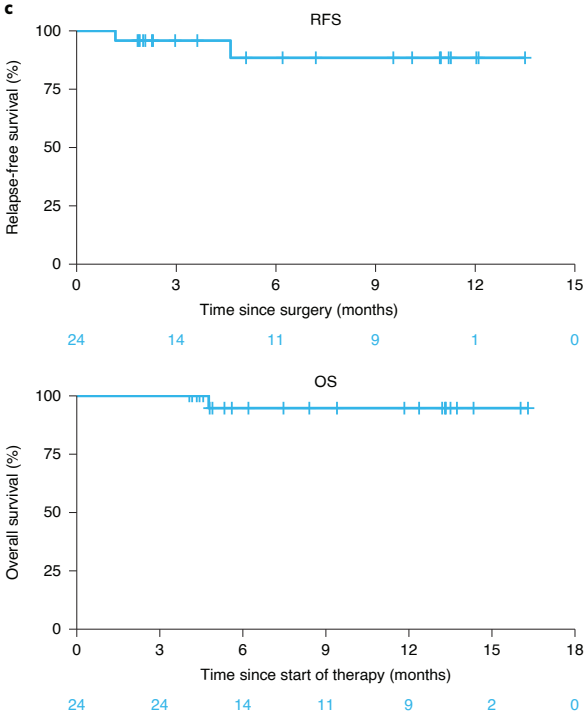


Fig. 2 | Pathological tumor regression and outcome to preoperative ipilimumab plus nivolumab. **a**, Percentage of pathological tumor regression per primary bladder tumor after preoperative ipilimumab plus nivolumab, based on the percentage of residual viable tumor area, for the complete cohort. Baseline clinical and biomarker features are annotated for each patient in the top section. Median mutational burden cutoffs; below median (low <9.6) versus above median (high ≥9.6) mutations per megabase. Intratumoral median CD8 cell density assigned as below median (low <145.6) versus above median (High ≥145.6) cells per mm^2. Histopathological findings in the surgical resection specimen upon ipilimumab plus nivolumab are listed in the lower boxes. **b**, pCR example, displaying large fields of necrosis surrounded by abundant CD8 T cells and TLS in multiplex immunofluorescence images (boxes 1 and 2). MPR (<10% residual viable tumor), characterized by reactive changes and small remaining tumor nests with substantial immune cell infiltration after treatment (boxes 3 and 4). Non-response, displaying intact tumor fields and immune infiltration at the tumor margin (boxes 5 and 6). **c**, Kaplan–Meier analysis of recurrence-free survival (RFS) and overall survival (OS). At a median postoperative follow-up of 8.3 months, two (8%) patients relapsed (both non-responders). One patient died 6 weeks after surgery due to rapid progressive disease. Experiments and scorings related to the presented micrographs were conducted once.


b
cT3N0: ipi/nivo-pCR
cT4aN1: ipi/nivo-MPR
cT3N1: ipi/nivo–NR
1
2
3
4
5
6
1 mm
500 µm
2 mm
200 µm
500 µm
350 µm
Hematoxylin and eosin stain
CD3 CD8 FoxP3 CD68 CD20 PanCK


c
RFS
Relapse-free survival (%)
100
75
50
25
0
0 3 6 9 12 15
Time since surgery (months)
24 14 11 9 1 0
OS
Overall survival (%)
100
75
50
25
0
0 3 6 9 12 15 18
Time since start of therapy (months)
24 24 14 11 9 2 0

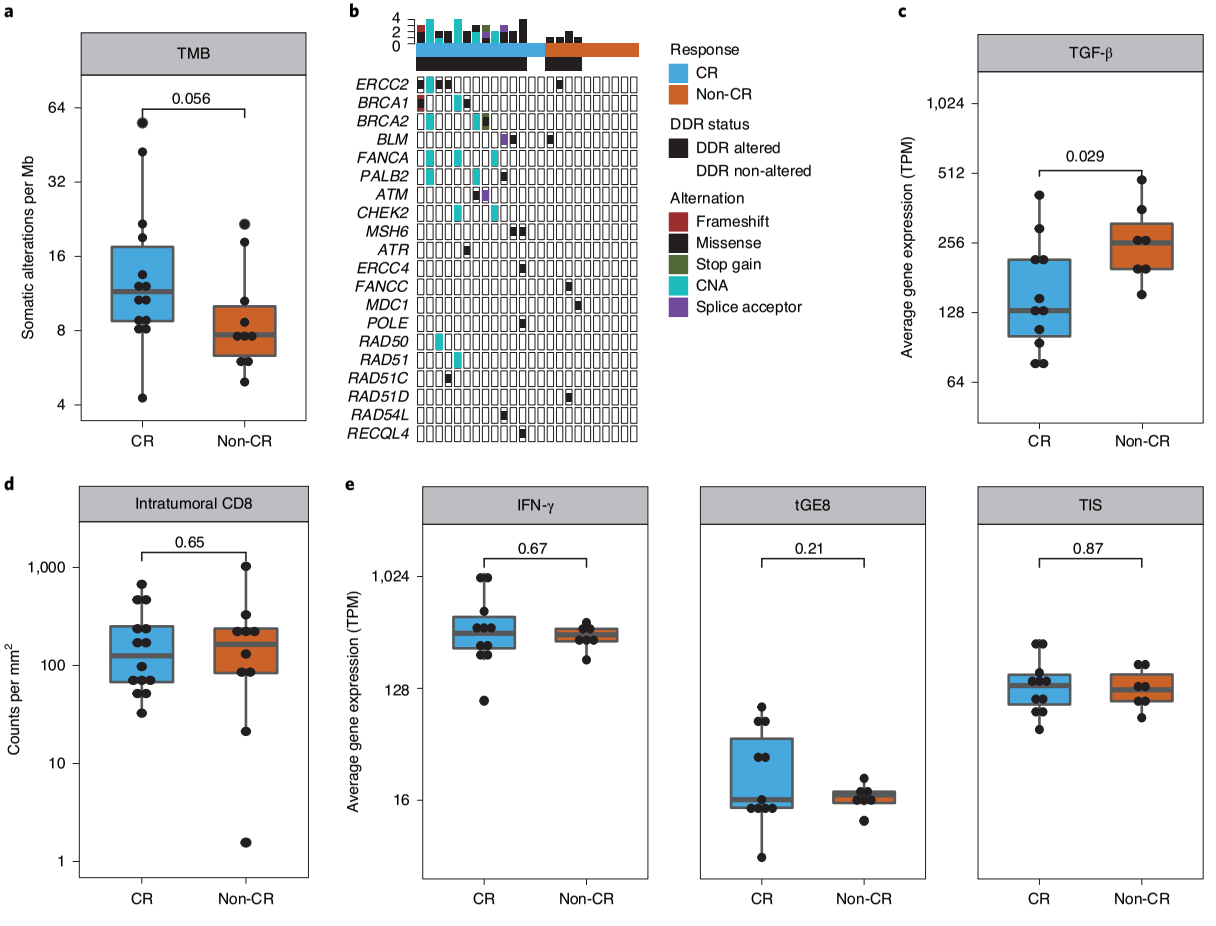
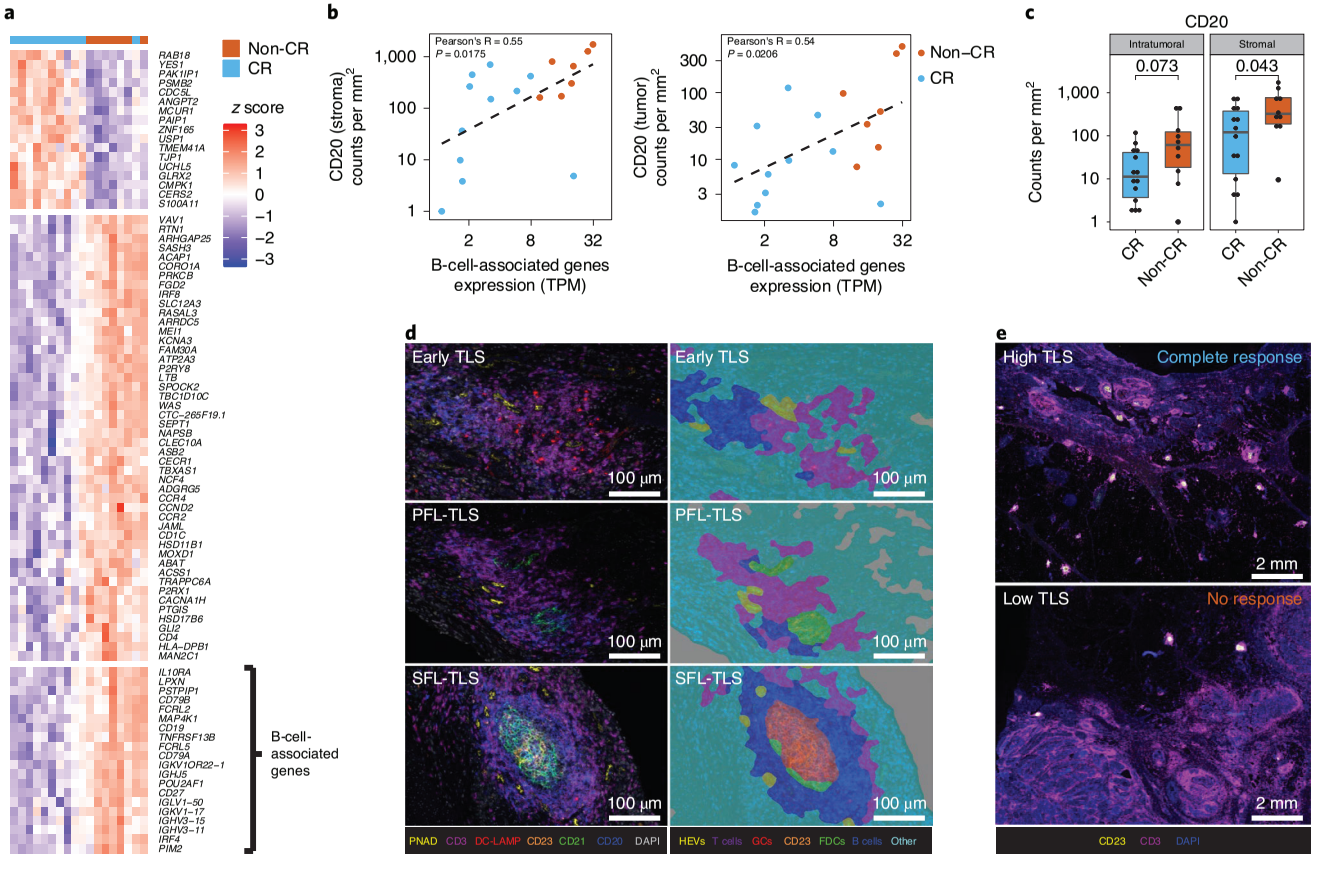

a
TMB
0.056
Somatic alterations per Mb
64
32
16
8
4
CR
Non-CR
b
4
2
0
ERCC2
BRCA1
BRCA2
BLM
FANCA
PALB2
ATM
CHEK2
MSH6
ATR
ERCC4
FANCC
MDC1
POLE
RAD50
RAD51
RAD51C
RAD51D
RAD54L
RECQL4
Response
CR
Non-CR
DDR status
DDR altered
DDR non-altered
Alternation
Frameshift
Missense
Stop gain
CNA
Splice acceptor
c
TGF-β
0.029
Average gene expression (TPM)
1,024
512
256
128
64
CR
Non-CR
d
Intratumoral CD8
0.65
Counts per mm2
1,000
100
10
1
CR
Non-CR
e
IFN-γ
0.67
Average gene expression (TPM)
1,024
128
16
CR
Non-CR
tGE8
0.21
CR
Non-CR
TIS
0.87
CR
Non-CR

Fig. 3 | pCR to ipilimumab plus nivolumab is independent of baseline CD8+ T cells and inflammatory signatures. **a**, Association between response groups and TMB. TMB is measured as the number of somatic, non-synonymous and exonic mutations, based on WES. The panel shows the TMB distribution in CR ($n = 14$) and non-CR ($n = 10$) tumors. **b**, Association between somatic alterations in DDR genes[17] and response (CR, $n = 14$ and non-CR, $n = 10$). We excluded variants with a clinical significance labeled as Benign or Likely Benign (CLINVAR annotation), variants previously reported as population/ germline SNPs (TOPMED and CAF annotations with alternate allele frequency >5%) and variants annotated as Tolerated with the SIFT predicted score. Patients were stratified by response (CR and non-CR) and DDR status (DDR altered and DDR non-altered) in a contingency table, and statistical significance was tested using a two-sided Fisher's exact test. **c**, Comparison of average TGF-β-induced genes[18] at baseline between CR ($n = 11$) and non-CR ($n = 7$) tumors. **d**, Baseline comparison of CD8 density per mm^2 between CR ($n = 14$) and non-CR ($n = 10$) tumors. Baseline intratumoral CD8 density per mm^2 was analyzed by quantitative multiplex immunofluorescence. **e**, Comparison of average gene expression for the IFN-γ signature (Ayers et al., 2017), the eight-gene cytotoxic T cell transcriptional signature (tGE8, Mariathasan et al.[23]) and TISs (Ayers et al.[29]) between CR ($n = 11$) and non-CR ($n = 7$) tumors. Box plots in all panels represent the median and 25th and 75th percentiles. The whiskers expand from the hinge to the largest value not exceeding 1.5× IQR from the hinge. Complete responders are displayed in blue, whereas non-responders are displayed in orange. A two-sided *t*-test was used for comparisons of the gene expression features between CR and non-CR tumors, and a two-sided Mann–Whitney test was used for comparison of the TMB between CR and non-CR tumors. The *P* value is presented in between box plots. All statistical tests were two sided. No adjustments were made for multiple comparisons. CNA, copy number aberration. SNP, single-nucleotide polymorphism.


a
Non-CR
CR
z score
3
2
1
0
−1
−2
−3
RAB18
YES1
PAK1IP1
PSMB2
CDC5L
ANGPT2
MCUR1
PAIP1
ZNF165
USP1
TMEM41A
TJP1
UCHL5
GLRX2
CMPK1
CERS2
S100A11
VAV1
RTN1
ARHGAP25
SASH3
ACAP1
CORO1A
PRKCB
FGD2
IRF8
SLC12A3
RASAL3
ARHGDC5
MEI1
KCNA3
FAM30A
ATP2A3
P2RY8
LTB
SPOCK2
TBC1D10C
WAS
CTC−265F19.1
SEPT1
NAPSB
CLEC10A
ASB2
CECR1
TBXAS1
NCF4
ADGRG5
CCR4
CCND2
CCR2
JAML
CD1C
HSD11B1
MOXD1
ABAT
ACSS1
TRAPPC6A
P2RX1
CACNA1H
PTGIS
HSD17B6
GLI2
CD4
HLA−DPB1
MAN2C1
IL10RA
LPXN
FSTPIP1
CD79B
FCRL2
MAP4K1
CD19
TNFRSF13B
FCRL5
CD79A
IGKV1OR22−1
IGHJ5
POU2AF1
CD27
IGLV1−50
IGKV1−17
IGHV3−15
IGHV3−11
IRF4
PIM2
B-cell-associated genes
b
Pearson's R = 0.55
P = 0.0175
CD20 (stroma) counts per mm2
1,000
100
10
1
2
8
32
B-cell-associated genes expression (TPM)
Pearson's R = 0.54
P = 0.0206
CD20 (tumor) counts per mm2
300
100
30
10
3
Non−CR
CR
c
CD20
Intratumoral
Stromal
0.073
0.043
Counts per mm2
1,000
100
10
1
CR
Non-CR
d
Early TLS
100 µm
PFL-TLS
SFL-TLS
PNAD CD3 DC-LAMP CD23 CD21 CD20 DAPI
HEVs T cells GCs CD23 FDCs B cells Other
e
High TLS
Complete response
2 mm
Low TLS
No response
CD23 CD3 DAPI

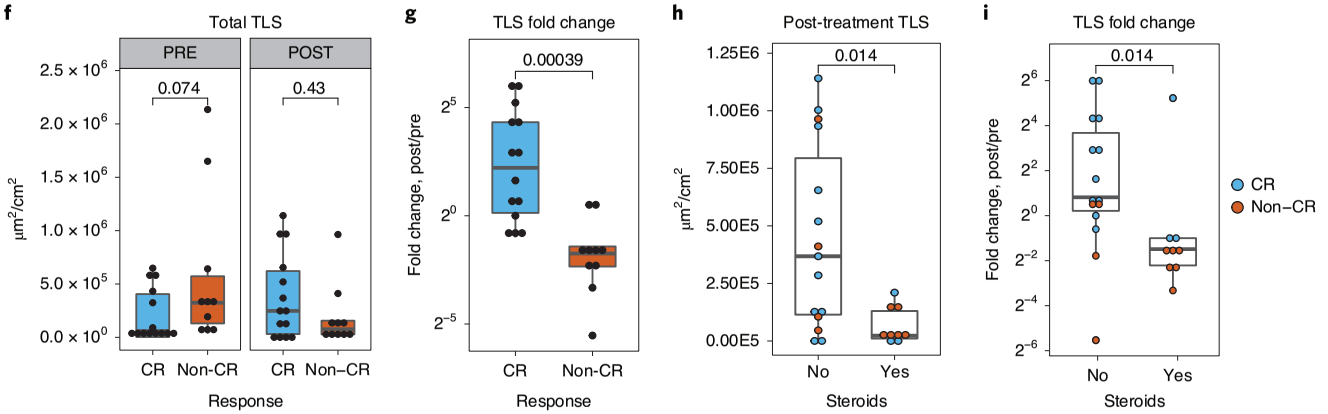


Fig. 4 | B cell analysis and assessment of TLS dynamics upon preoperative ipilimumab plus nivolumab. **a**, Hierarchical clustering of differentially expressed genes (FDR<8%) at baseline between CR (blue, $n=11$) and non-CR (orange, $n=7$) tumors, showing upregulation of B-cell-related genes in patients with non-CR tumors. Gene expression levels are represented as CPM, mean centered and scaled (z scores). Higher expression levels (red) and lower expression levels (blue) are indicated per gene. For each gene, baseline differential expression between response groups was tested fitting a linear model to response using weighted least squares (limma R package). Contrasts that compared CR and non-CR tumors were fitted. For each gene, the density-based FDR from the P-value distribution was computed to correct for multiple hypothesis testing. Only genes that scored an FDR<8% are shown on the heat map. **b**, The differentially expressed B-cell-associated genes (derived from **a**) correlate with tumor and stroma CD20+ immune cell counts per mm^2 by multiplex immunofluorescence analysis. A two-sided Pearson's moment correlation test was used to analyze this correlation in CR ($n=11$) and non-CR ($n=7$) tumors. Pearson's coefficient and correlation P value are shown in the plot. **c**, Baseline CD20+ immune cell densities per mm^2 in tumor and stroma by multiplex immunofluorescence analysis between CR (blue, $n=14$) and non-CR (orange, $n=10$) tumors stratified by response categories for tumor and stroma. **d**, Tissue segmentation (PerkinElmer) was used to segregate TLS areas. Early TLS displays dense B cell clusters without follicular dendritic cells (FDCs) or germinal centers (GCs). Primary follicle-like TLSs (PFL-TLSs) harbor FDC networks while lacking GCs. Secondary follicle-like TLSs (SFL-TLSs) exhibit active GC reactions. **e**, Example images of the TLS landscape in a CR versus non-CR tumor after therapy **f**, TLS development was quantified as normalized TLS area (square microns per tissue square centimeters). Pre-treatment and post-therapy samples were compared between CR ($n=14$) and non-CR ($n=10$) tumors. **g**, Immunotherapy-induced changes in normalized TLS area were assessed as fold change (post/pre) and compared between the CR (blue, $n=14$) and non-CR (orange, $n=10$) tumors. **h**, Normalized TLS area in post-therapy samples between patients who received steroids (>20 mg per day before surgery, $n=9$) and patients who received no steroids ($n=15$). A comparison was made using a two-sided Mann–Whitney test. **i**, Immunotherapy-induced fold changes in normalized TLS areas were assessed as fold changes (post/pre) in patients who received steroids (>20 mg per day before surgery, $n=9$) and in patients who received no steroids ($n=15$). Box plots in all panels represent the median and 25th and 75th percentiles. The whiskers expand from the hinge to the largest not exceeding 1.5× IQR from the hinge.

analysis of differentially expressed genes, we found that upregulation of B-cell-related genes at baseline correlated with non-response **(Fig. 4a)**. Furthermore, these differentially expressed B cell genes were found to positively correlate with CD20+ B cell counts in multiplex immunofluorescence analysis **(Fig. 4b)**. In non-responding tumors, stromal B cells were more abundant than in responding tumors (P = 0.043; **Fig. 4c**), whereas the density of B cells in the tumor compartment was numerically higher (P = 0.07). The increased B cell presence in non-CR tumors was observed irrespective of pre-existing CD8+ T cell immunity **(Extended Data Fig. 3)**. B cells have been studied in conjunction with TLS composition (20). We used multiplex immunofluorescence staining and image analysis to visualize TLS composition and quantify TLS maturation stages **(Fig. 4d,e)**. No correlation was observed between baseline TLS numbers and response **(Fig. 4f)**, although immature TLSs were higher in non-CR tumors **(Extended Data Fig. 3)**. On-treatment analysis revealed enrichment in TLS (P < 0.001; **Fig. 4g**) in CR tumors. Enrichment was observed across all TLS maturation stages **(Extended Data Fig. 4)**. Expression of TLS-related genes and a recently established signature (19) related to B cell/TLS presence in melanoma increased upon treatment **(Extended Data Fig. 5)**, particularly in CR tumors. TLS development was dampened in patients receiving corticosteroids for immunotherapy toxicity management (P = 0.014; **Fig. 4h,i)**, as observed in lung cancer (27). TLS were further characterized by the presence of CD27+ B cells **(Extended Data Fig. 6)**, T follicular helper cells, BCL6+CD4− cells and expression of CXCL13 **(Extended Data Fig. 7)**. We observed no increase of CD27+ B cells and T follicular helper cells in TLS after therapy. Early induction of these cells might have been missed owing to the fact that samples were taken relatively late in the anti-cancer immunotherapy response, after tumor clearance. To further study TLS dynamics, a bioinformatic algorithm was developed to annotate the TLS compartment in multiplex immunofluorescence data. Using this algorithm, we found that Tregs were reduced in TLS upon treatment with ipilimumab plus nivolumab (**Extended Data Fig. 8**). These exploratory findings suggest that the presence of B cells and TLSs at baseline does not predict for ipilimumab plus nivolumab response in UC, although TLS induction is observed in responding patients.

DISCUSSION

We found that ipilimumab plus nivolumab is feasible and highly active as preoperative treatment in UC. Moreover, high pCR rates were observed in patients with locoregionally advanced disease. Previous studies testing preoperative anti-PD1/PD-L1 monotherapy in UC were conducted mainly in patients with less advanced tumors (cT2-3N0), showing encouraging pCR rates. Substantially lower pCR rates were observed with anti-PD-L1 in patients with more advanced disease (18). Our results are in line with findings in

locoregionally advanced melanoma, where the addition of anti-CTLA-4 to anti-PD-1 monotherapy before surgery broadened response and resulted in superior efficacy and outcome (1,28). Furthermore, we found that response to preoperative ipilimumab plus nivolumab was independent of baseline $CD8^+$ T cell presence and inflammatory signatures. This observation contrasts with findings from trials using anti-PD1/PD-L1 monotherapy, where pCR was primarily observed in tumors exhibiting pre-existing $CD8^+$ T cell immunity (16,18). A limitation of our study is the small sample size, which was powered to show feasibility of preoperative ipilimumab plus nivolumab and to determine an efficacy signal that would justify further clinical studies. Additionally, the treatment schedule can be further refined to find the optimal balance between toxicity and efficacy.

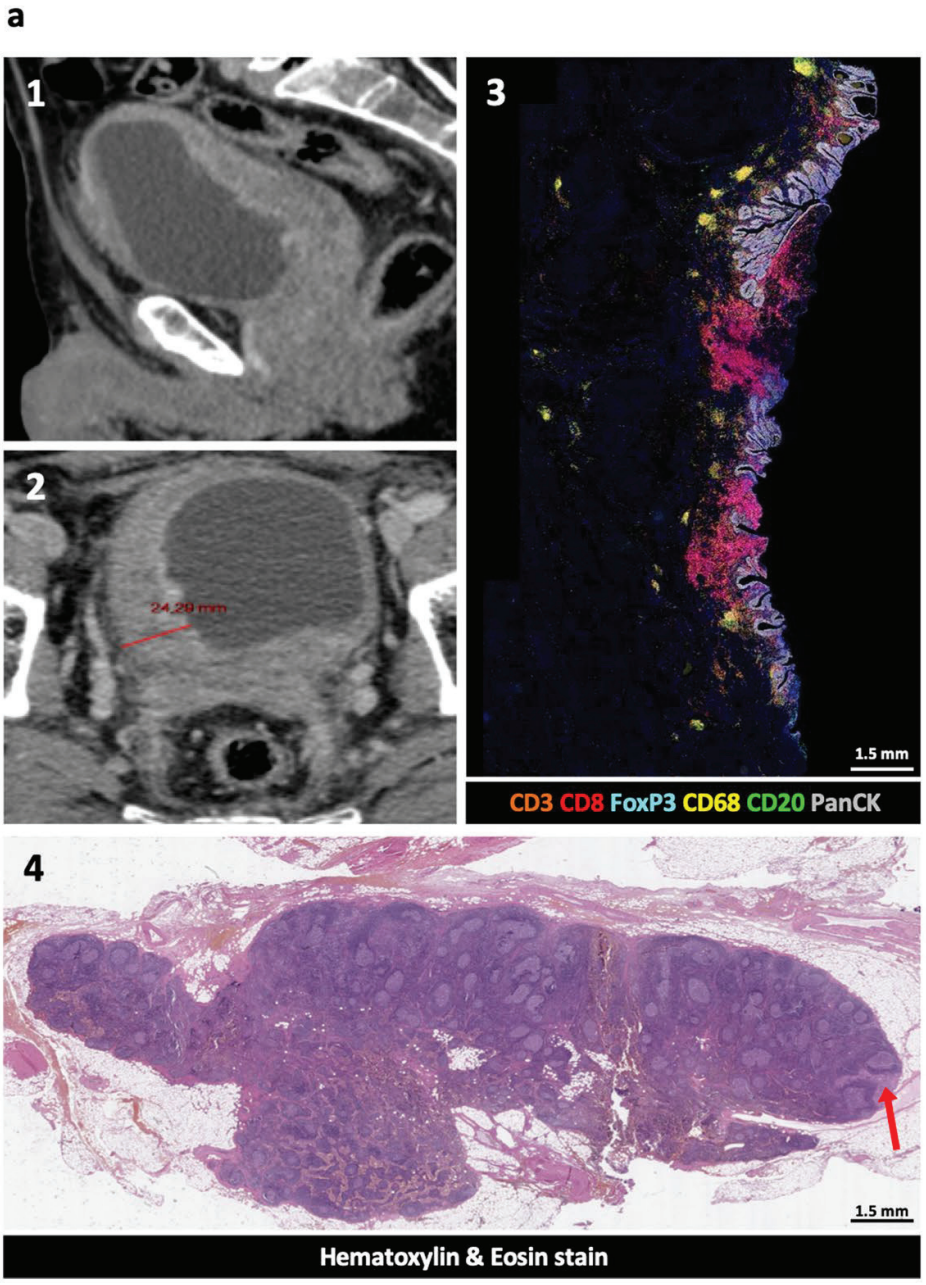
In conclusion, combined CTLA-4 plus PD-1 blockade might provide an effective pre-operative treatment strategy in locoregionally advanced UC, regardless of pre-existing $CD8^+$ T cell activity at baseline. The therapeutic efficacy in this cohort, with poor prognosis and high unmet clinical need, warrants further development through randomized trials.

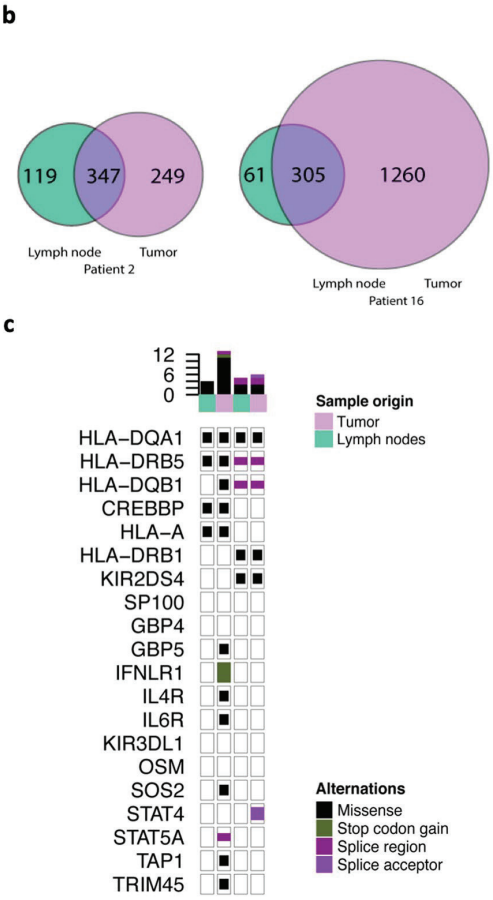
REFERENCES

1. 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).
2. 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).
3. Stein, J. P. et al. Radical cystectomy in the treatment of invasive bladder cancer: long-term results in 1,054 patients. *J. Clin. Oncol.* **19**, 666–675 (2001).
4. Zargar, H. et al. Multicenter assessment of neoadjuvant chemotherapy for muscle-invasive bladder cancer. *Eur. Urol.* **67**, 241–249 (2014).
5. Advanced Bladder Cancer (ABC) Meta-analysis Collaboration. Neoadjuvant chemotherapy in invasive bladder cancer: update of a systematic review and meta-analysis of individual patient data. *Eur. Urol.* **48**, 202–205 (2005).
6. Bellmunt, J. et al. Pembrolizumab as second-line therapy for advanced urothelial carcinoma. *N. Engl. J. Med.* **376**, 1015–1026 (2017).
7. Balar, A. V. et al. Atezolizumab as first-line treatment in cisplatin-ineligible patients with locally advanced and metastatic urothelial carcinoma: a single-arm, multicentre, phase 2 trial. *Lancet* **389**, 67–76 (2017).
8. Sharma, P. et al. Nivolumab in metastatic urothelial carcinoma after platinum therapy (CheckMate 275): a multicentre, single-arm, phase 2 trial. *Lancet Oncol.* **18**, 312–322 (2017).
9. Rosenberg, J. E. et al. Atezolizumab in patients with locally advanced and metastatic urothelial carcinoma who have progressed following treatment with platinum-based chemotherapy: a single-arm, multicentre, phase 2 trial. *Lancet* **387**, 1909–1920 (2016).
10. Massard, C. et al. Safety and efficacy of durvalumab (MEDI4736), an anti-programmed cell death ligand-1 immune checkpoint inhibitor, in patients with advanced urothelial bladder cancer. *J. Clin. Oncol.* **34**, 3119–3125 (2016).
11. Powles, T. et al. Atezolizumab versus chemotherapy in patients with platinum-treated locally advanced or metastatic urothelial carcinoma (IMvigor211): a multicentre, open-label, phase 3 randomised controlled trial. *Lancet* **391**, 748–757 (2018).
12. Sharma, P. et al. Nivolumab alone and with ipilimumab in previously treated metastatic urothelial carcinoma: CheckMate 032 nivolumab 1 mg/kg plus ipilimumab 3 mg/kg expansion cohort results. *J. Clin. Oncol.* **37**, 1608–1616 (2019).
13. Balar, A. V. et al. First-line pembrolizumab in cisplatin-ineligible patients with locally advanced and unresectable or metastatic urothelial cancer (KEYNOTE-052): a multicentre, single-arm, phase 2 study. *Lancet Oncol.* **18**, 1483–1492 (2017).
14. Carthon, B. C. et al. Preoperative CTLA-4 blockade: tolerability and immune monitoring in the setting of a presurgical clinical trial. *Clin. Cancer Res.* **16**, 2861–2871 (2010).
15. Liakou, C. I. et al. CTLA-4 blockade increases IFNγ-producing $CD4^{+}ICOShi$ cells to shift the ratio of effector to regulatory T cells in cancer patients. *Proc. Natl Acad. Sci. USA* **105**, 14987–14992 (2008).
16. Necchi, A. et al. Pembrolizumab as neoadjuvant therapy before radical cystectomy in patients with muscle-invasive urothelial bladder carcinoma (PURE-01): an open-label, single-arm, phase II study. *J. Clin. Oncol.* **36**, 3353–3360 (2018).
17. Necchi, A. et al. Updated results of PURE-01 with preliminary activity of neoadjuvant pembrolizumab in patients with muscle-invasive bladder carcinoma with variant histologies. *Eur. Urol.* **77**, 439–446 (2019).

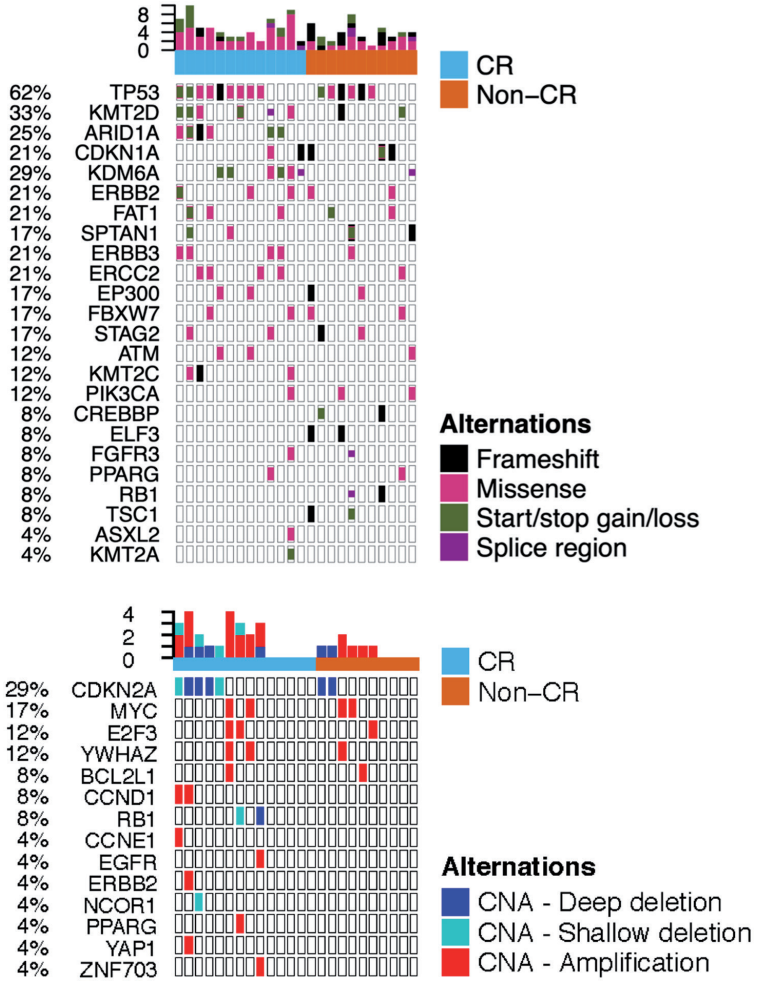
18. Powles, T. et al. Clinical efficacy and biomarker analysis of neoadjuvant atezolizumab in operable urothelial carcinoma in the ABACUS trial. *Nat. Med.* **25**, 1706–1714 (2019).
19. Cabrita, R. et al. Tertiary lymphoid structures improve immunotherapy and survival in melanoma. *Nature* **577**, 561–565 (2020).
20. Helmink, B. A. et al. B cells and tertiary lymphoid structures promote immunotherapy response. *Nature* **577**, 549–555 (2020).
21. Meerveld-Eggink et al. Short-term CTLA-4 blockade directly followed by PD-1 blockade in advanced melanoma patients: a single-center experience. *Ann. Oncol.* **28**, 862–867 (2017).
22. Teo, M. Y. et al. Alterations in DNA damage response and repair genes as potential marker of clinical benefit from PD-1/PD-L1 blockade in advanced urothelial cancers. *J. Clin. Oncol.* **36**, 1685–1694 (2018).
23. Mariathasan, S. et al. TGFβ attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. *Nature* **554**, 544–548 (2018).
24. Hollern, D. P. et al. B cells and T follicular helper cells mediate response to checkpoint inhibitors in high mutation burden mouse models of breast cancer. *Cell* **179**, 1191–1206 (2019).
25. Largeot, A., Pagano, G., Gonder, S., Moussay, E. & Paggetti, J. The B-side of cancer immunity: the underrated tune. *Cells* **8**, 449 (2019).
26. Griss, J. et al. B cells sustain inflammation and predict response to immune checkpoint blockade in human melanoma. *Nat. Commun.* **10**, 4186 (2019).
27. Silina, K. et al. Germinal centers determine the prognostic relevance of tertiary lymphoid structures and are impaired by corticosteroids in lung squamous cell carcinoma. *Cancer Res.* **78**, 1308–1320 (2018).
28. Amaria, R. N. et al. Neoadjuvant immune checkpoint blockade in high-risk resectable melanoma. *Nat. Med.* **24**, 1649–1654 (2018).
29. Ayers, M. et al. IFN-γ–related mRNA profile predicts clinical response to PD-1 blockade. *J. Clin. Invest.* **127**, 2930–2940 (2017).
30. Zargar, H. et al. Multicenter assessment of neoadjuvant chemotherapy for muscle-invasive bladder cancer. *Eur. Urol.* **67**, 241–249 (2015).
31. Kim, S. et al. Strelka2: fast and accurate calling of germline and somatic variants. *Nat. Methods* **15**, 591–594 (2018).
32. Talevich, E., Shain, A. H., Botton, T. & Bastian, B. C. CNVkit: genome-wide copy number detection and visualization from targeted DNA sequencing. *PLoS Comput. Biol.* **12**, e1004873 (2016).
33. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2009).
34. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
35. Strimmer, K. fdrtool: a versatile R package for estimating local and tail area-based false discovery rates. *Bioinformatics* **24**, 1461–1462 (2008).

EXTENDED DATA FIGURES

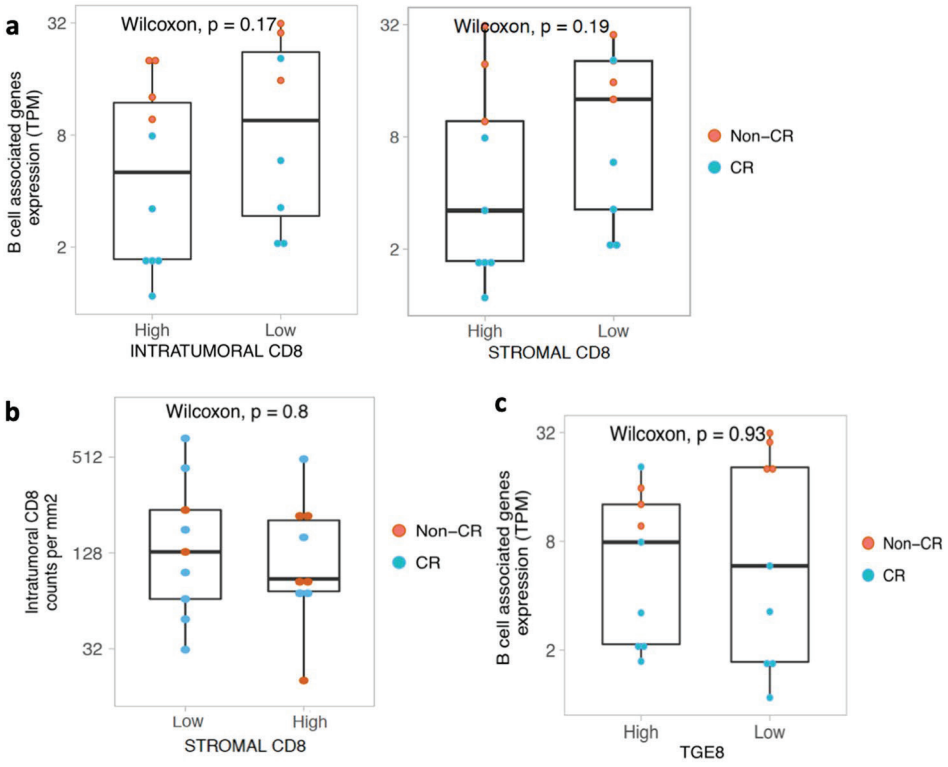




Extended Data Fig. 1 | Assessment of response discrepancies between the responding primary bladder tumor and unresponsive local lymph node micrometastases upon ipilimumab plus nivolumab. **a1-2**. Imaging shows a cT4aN1 bladder tumor (enlarged lymph node not shown). Pathological assessment revealed a complete response in the bladder, as shown by multiplex immunofluorescent analysis **(a-3)** and a non-responding lymph node micrometastasis, annotated by a red arrow **(a-4)**. Experiments and scorings related to the presented micrographs were conducted once. **b,** Whole-exome sequencing was employed to assess whether lymph node micrometastases are genetically distinct from the primary tumor and could explain unresponsiveness (n=2 patients). The figure shows overlap between somatic variants identified in the baseline tumor (purple) and the post-treatment lymph node metastasis (green). **c,** Genomic alterations in genes related to interferon gamma signaling, JAK/STAT signaling or antigen presentation machinery for baseline tumor samples (purple, n=2) and matching lymph nodes metastases (green, n=2). Sample labels and genetic alteration type are displayed in figure legends. No specific genetic cause for resistance could be identified in the discrepant mutations.

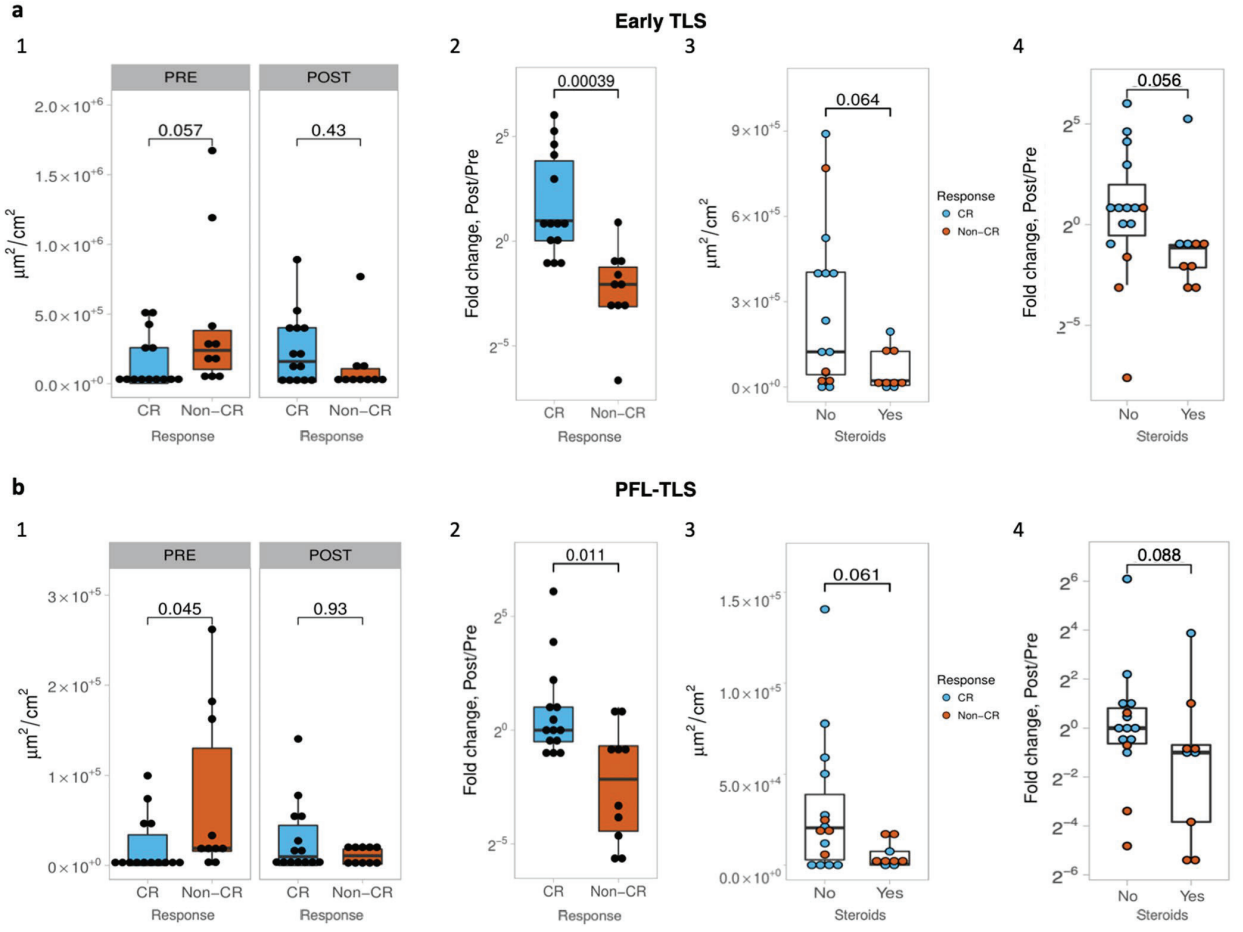


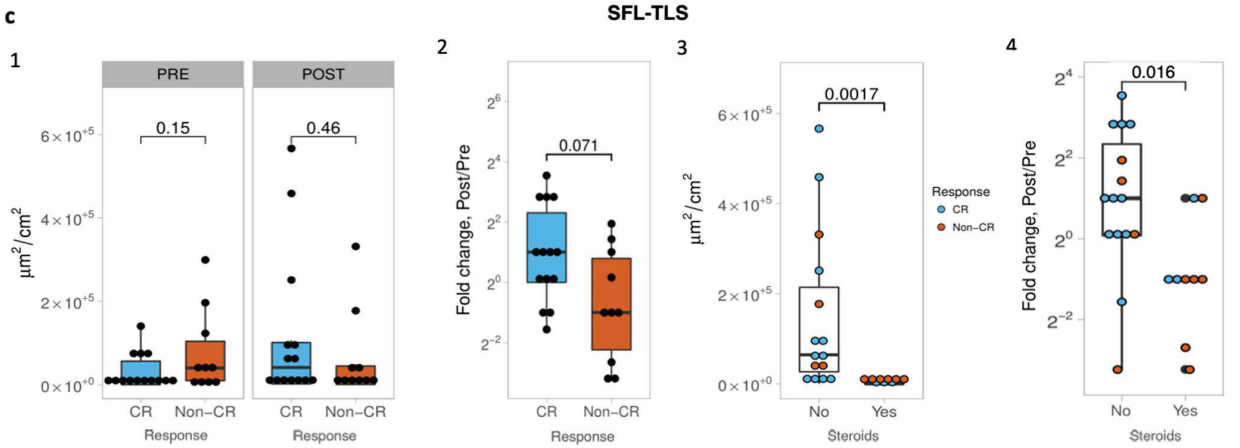
Extended Data Fig. 2 | Analysis of baseline genomic alterations by whole-exome sequencing. Whole-exome sequencing of pretreatment tumor tissue and germline DNA was employed to identify somatic mutations in baseline tumors (n=24). An oncoprint figure of genetic alterations in significantly mutated bladder cancer genes (by TCGA, *Robertson et al, Cell 2018*) is shown. Alterations are clustered by response categories, including 14 CR and 10 non-CR tumors. Sample labels and genetic alteration type are displayed in the figure. Abbreviations: CR: complete response, non-CR: non complete response, CNA: copy number alterations.



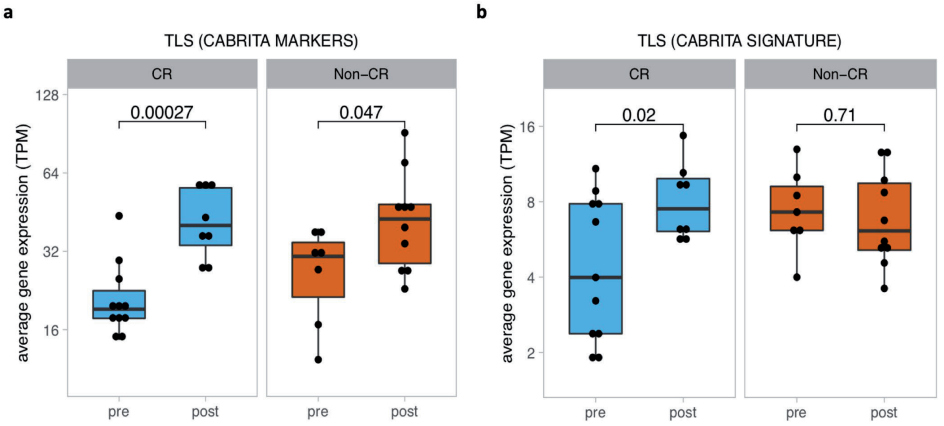
Extended Data Fig. 3 | Assessment of interdependence between baseline B cell presence and CD8 T cell infiltration. **a**, Average expression of B-cell-related differentially expressed genes at baseline stratified by intratumoral and stromal $CD8^+$ higher and lower than median groups on multiplex immunofluorescence (mIF). Gene expression levels are represented as transcripts per million (TPM), mean-centered and scaled (Z-scores). **b**, Intratumoral CD8 density per mm^2 by multiplex immunofluorescence, stratified by stromal CD20 higher and lower than median groups on multiplex immunofluorescence (mIF). **c**, Average expression of B-cell-associated genes at baseline stratified by average TGE8 signature groups (higher or lower than median) showing similar expression of B-cell-related genes in the TGE8 signature groups. Gene expression levels are represented as transcripts per million (TPM), mean-centered and scaled (Z-scores). A Wilcoxon signed tank test was used for all comparisons. The P-value is presented in-between boxplots.

All statistical tests were two-sided. Analyses include CR (blue; n=11) and non-CR (orange; n=7) tumors. All boxplots display the median and 25th and 75th percentiles. The whiskers expand from the hinge to largest value not exceeding 1.5× IQR from the hinge. No adjustments were made for multiple comparisons. Abbreviations: CR: complete response, non-CR: non complete response.

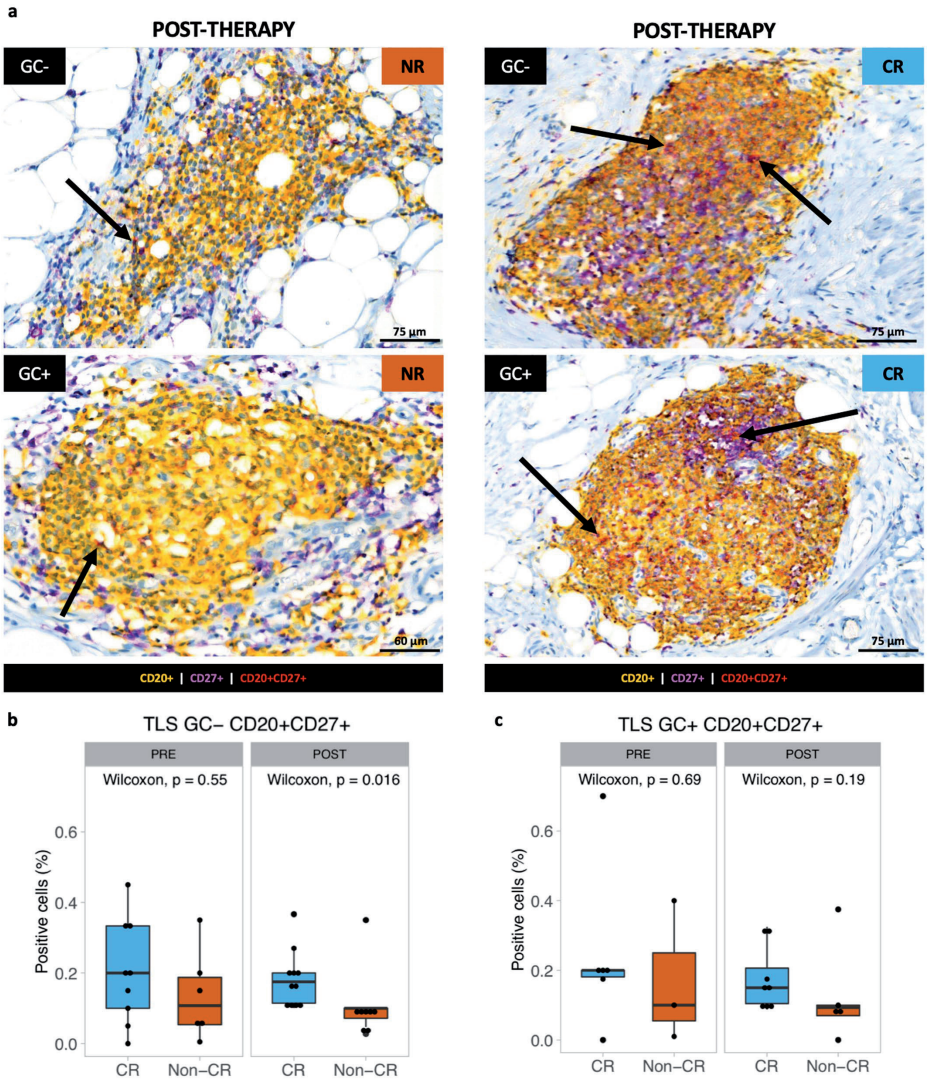

a
Early TLS
1
PRE
POST
0.057
0.43
µm²/cm²
CR
Non–CR
Response
2
0.00039
Fold change, Post/Pre
3
0.064
Response
CR
Non–CR
No
Yes
Steroids
4
0.056
b
PFL-TLS
0.045
0.93
0.011
0.061
0.088



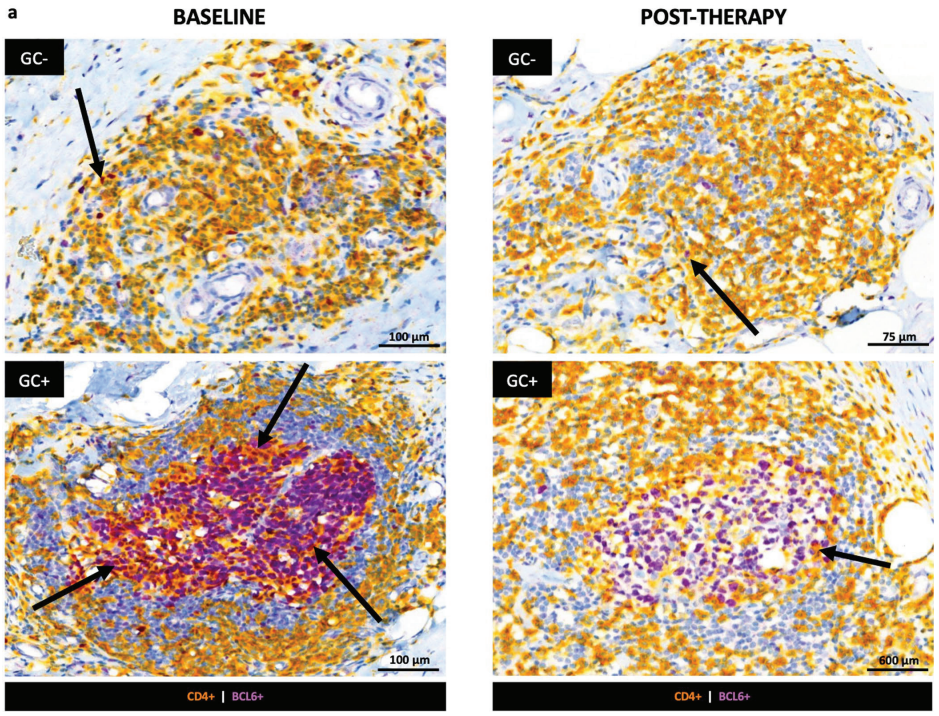
Extended Data Fig. 4 | Dynamics of tertiary lymphoid structure spectrum upon immunotherapy for response and steroid groups. Upon multiplex immunofluorescent staining and segregation of tertiary lymphoid structure (TLS) areas, **a**, Early-TLS, **b**, Primary follicle-like TLS and **c**, Secondary follicle-like TLS were quantified as normalized TLS area (square microns per tissue square centimeter). For each TLS maturation stage, four different analysis were performed; **1)** Comparison of normalized TLS areas in baseline and post-therapy samples between response groups by Mann Whitney test, **2)** Normalized TLS area assessed as fold change (post/pre) upon treatment between response groups, **3)** Normalized TLS area comparison in post-therapy samples between patients receiving steroids and no steroids, **4)** Normalized TLS area assessed as fold change (post/pre) upon treatment in patients that received steroids and no steroids. Unless otherwise noted, all boxplots display the median and 25th and 75th percentiles. The whiskers expand from the hinge to largest value not exceeding 1.5× IQR from the hinge. A Mann Whitney test was used for comparisons between response and steroid groups. The P-value is presented in-between boxplots. All statistical tests were two-sided. Analyses have 14 CR and 10 non-CR, or 9 steroids (>20mg a day prior to surgery) and 15 no steroid patients. No adjustments were made for multiple comparisons. Abbreviations: TLS: tertiary lymphoid structures, CR: complete response, non-CR: no complete response.

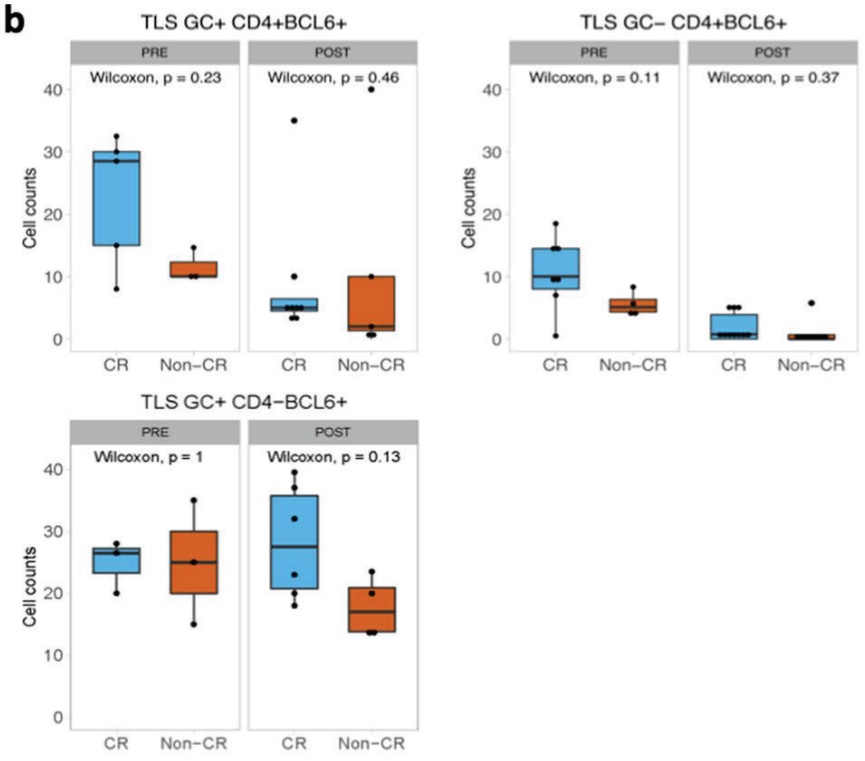


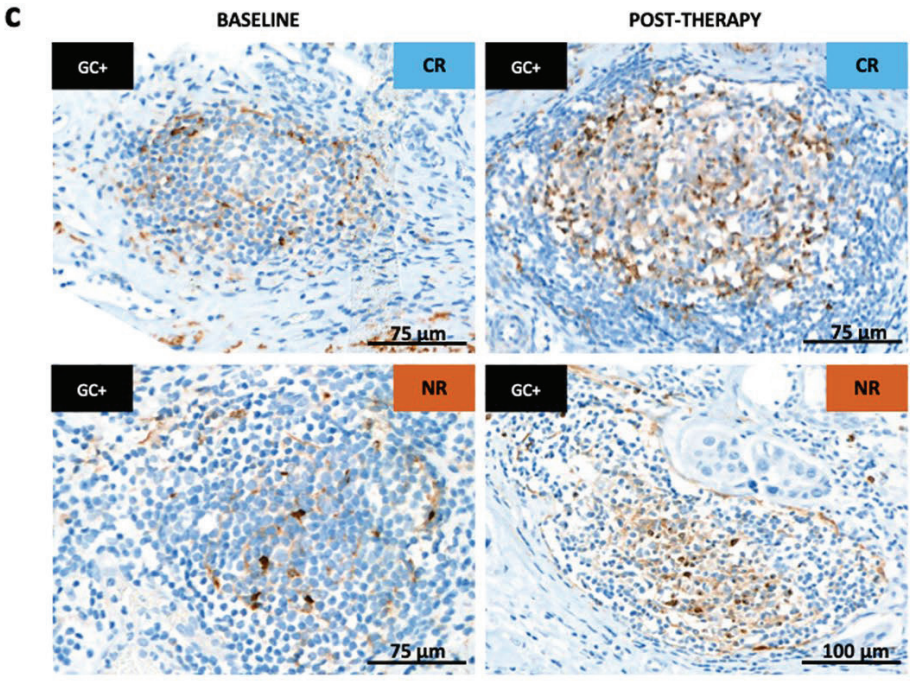
Extended Data Fig. 5 | Exploratory assessment of TLS markers and a TLS signature upon immunotherapy for response groups. **a**, Average gene expression for TLS-related genes (*CCL19, CCL21, CXCL13, CCR7, SELL, LAMP3, CXCR4, CD86, BCL6*), as published by Cabrita et al. [19]. Pre- and post-treatment samples were compared for CR (n=11 pre-treatment, n=8 post-treatment) and non-CR (n=7 pre-treatment, n=10 post-treatment) tumors using a two-sided t-test. Baseline expression was not significantly different between CR and non-CR (p=0.28). **b**, TLS gene signature (*Cabrita et al., Nature 2020*) derived from genes specifically upregulated in $CD8^{+}CD20^{+}$ metastasized melanoma tumors (*CD79B, CD1D, SKAP1, CETP, EIF1AY, RBP5, PTGDS*). *LAT and CCR6* genes were lowly expressed and thus removed from the analysis. Gene signatures were compared between baseline and post- treatment samples for CR and non-CR tumors. Baseline expression was not significantly different between CR and non-CR (p=0.05). Boxplots in all panels represent the median and 25th and 75th percentiles. The whiskers expand from the hinge to largest value not exceeding 1.5× IQR from the hinge. CR samples are marked in blue, while non-CR samples are displayed in orange. A t-test test was used for comparisons between CR and non-CR. The p-value is presented in-between boxplots. All statistical tests were two-sided. All analyses involved 11 CR and 7 non-CR tumors in pre-treatment samples, and 10 CR and 8 non-CR tumors in post-treatment samples. Abbreviations: TLS: tertiary lymphoid structures, CR: complete response, Non-CR: non complete response.



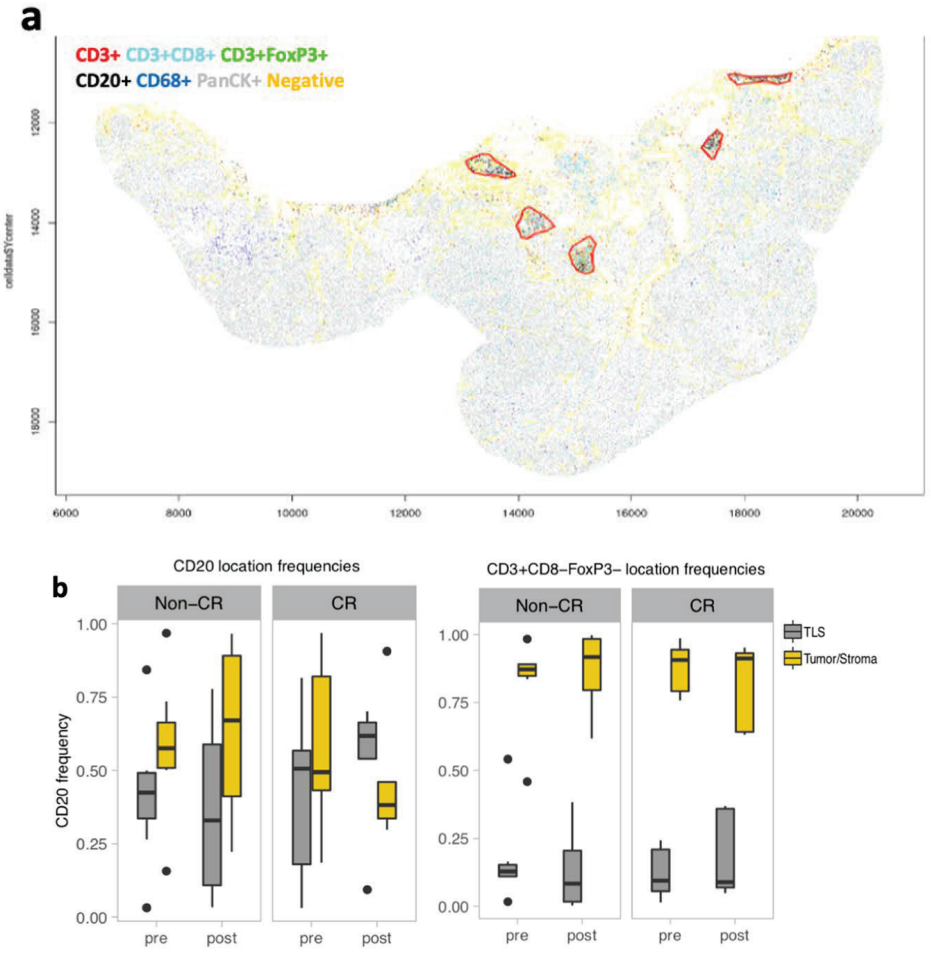
Extended Data Fig. 6 | Assessment of CD27+ B-cells in TLS upon immunotherapy for response groups. **a**, Example images of CD20 (yellow) and CD27 (purple) IHC co-staining, revealing CD27+ B-cells (CD20+CD27+) in red, as indicated by the black arrow. Experiments and scorings related to the presented micrographs were conducted once. **b/c.** Baseline and post-treatment comparison of the mean percentage of CD20+CD27+ cells in germinal center (GC) negative (**b**) and GC+ (**c**) TLS per patient between CR (n=9 GC- pre-treatment, n=6 GC+ pre-treatment, n=11 GC- post-treatment, n=8 GC+ post-treatment) and non-CR (n=6 GC- pre-treatment, n=3 GC+ pre-treatment, n=8 GC- post-treatment, n=5 GC+ post-treatment) tumors. The percentage CD20+CD27+ cells in the CD20+ population in TLS was estimated by a pathologist (L.S.). Boxplots represent the median and 25th and 75th percentiles. The whiskers expand from the hinge to largest value not exceeding 1.5× IQR from the hinge. Complete responders are marked in blue, while non-responders are displayed in orange. A Wilcoxon signed tank test was used to compare the percentage of CD20+CD27+ between CR and non-CR tumors. The P-value is presented in-between boxplots. All statistical tests were two-sided. No adjustments were made for multiple comparisons. Abbreviations: CR: complete response, NR: no response, GC: germinal center.


a
BASELINE
POST-THERAPY
GC-
GC+
100 µm
75 µm
600 µm
CD4+ | BCL6+

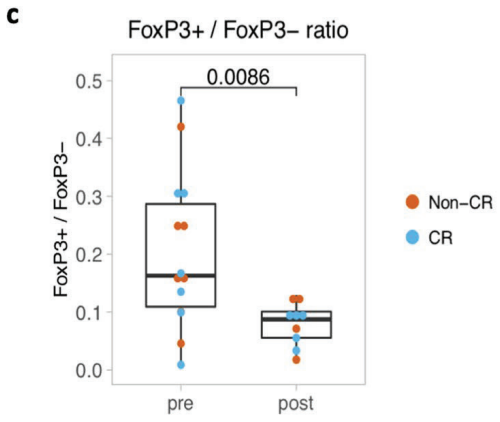

b
TLS GC+ CD4+BCL6+
TLS GC− CD4+BCL6+
TLS GC+ CD4−BCL6+
PRE
POST
Wilcoxon, p = 0.23
Wilcoxon, p = 0.46
Wilcoxon, p = 0.11
Wilcoxon, p = 0.37
Wilcoxon, p = 1
Wilcoxon, p = 0.13
Cell counts
CR
Non−CR

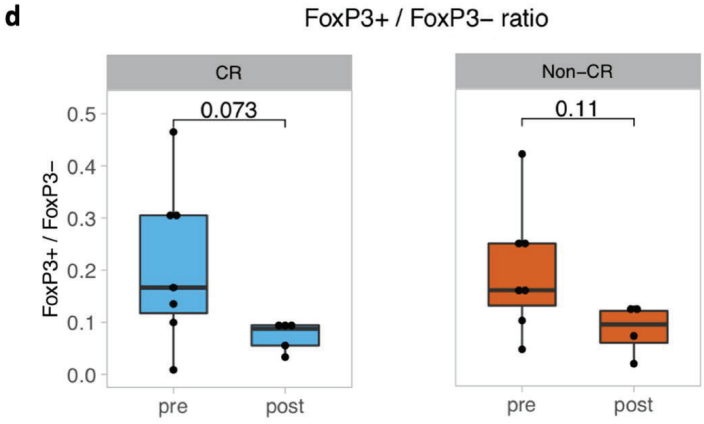


Extended Data Fig. 7 | Assessment of CD4+ T-cells and CD4+BCL6+ follicular T helper cells in tumor and TLS regions upon immunotherapy. **a,** Example images of CD4 (yellow) and BCL6 (purple) IHC co-stainings, showing CD4+BCL6+ follicular T helper cells pre- and post-treatment, characterized by a purple nucleus and deep orange cytoplasmatic staining. **b,** Mean absolute CD4+BCL6+ cell counts in co-stains for mature and immature TLS and CD4−BCL6+ cell counts for mature TLS only; for CR tumors (n=7 GC- pre-treatment, n=5 GC+ pre-treatment, n=10 GC- post-treatment, n=8 GC+ post-treatment) and non-CR tumors (n=4 GC- pre-treatment, n=3 GC+ pre-treatment, n=8 GC- post-treatment, n=5 GC+ post-treatment). Co-stainings were assessed and scored (number of cells per TLS) by a pathologist. **c,** Representative example images of CXCL13 (brown) in TLS for CR and non-CR tumors by immunohistochemistry in pre- and post-treatment specimens. CXCL13 positivity is clearly present in TLS, emphasizing that TLS are characterized by CXCL13 expression. No post-treatment differences were observed between response groups (Quantified data not shown). Unless otherwise noted, all boxplots display the median and 25th and 75th percentiles. The whiskers expand from the hinge to largest value not exceeding 1.5× IQR from the hinge. Complete responders are marked in blue, while non-responders are displayed in orange. A Wilcoxon signed tank test was used to compare the CD4+BCL6+ counts between CR and non-CR tumors. The P-value is presented in-between boxplots. All statistical tests were two-sided. No adjustments were made for multiple comparisons. Experiments and scorings related to the presented micrographs in A and C were conducted once. Abbreviations: CR: complete response, non-CR: non complete response, GC: germinal center.



Extended Data Fig. 8 | a, Data map displaying immune cell subsets by digital multiplex immunofluorescent analysis in baseline tissue in a spatial context. A bioinformatic algorithm (methods) was developed to identify and segment TLS-like structures, as annotated in red in the example map. Tumor is depicted in gray (panCK) **b,** Distribution of CD3+CD8− and CD20+ cells over the analysed area, between algorithm assigned tumor bed (orange) and TLS (grey) pre- and post-therapy, in CR (n=7 pre, n=5 post) and non-CR (n=7 pre, n=4 post) patients. **c,** The ratio of FOXP3+/FOXP3- cells was calculated within the TLS CD3+CD8- compartment upon algorithmic TLS segmentation and quantitation of immune cell subsets in multiplex immunofluorescence images (Methods). Pre-treatment and post-therapy samples were compared in the complete cohort (C; n=14 pre, 9 post) p=0.0086) or between CR (n=7 pre, 5 post; p=0.073) and non-CR (n=7 pre, 4 post; p=0.11) **d,** by Mann Whitney test. Unless otherwise noted, all boxplots display the median and 25th and 75th percentiles. The whiskers expand from the hinge to largest value not exceeding 1.5× IQR from the hinge. Complete responders are marked in blue, while non-responders are displayed in orange. A Wilcoxon signed tank test was used to compare the CD4+BCL6+ counts between CR and non-CR tumors. The P-value is presented in-between boxplots. All statistical tests were two-sided. No adjustments were made for multiple comparisons. Abbreviations: CR: complete response, non-CR: non complete response, GC: germinal center.

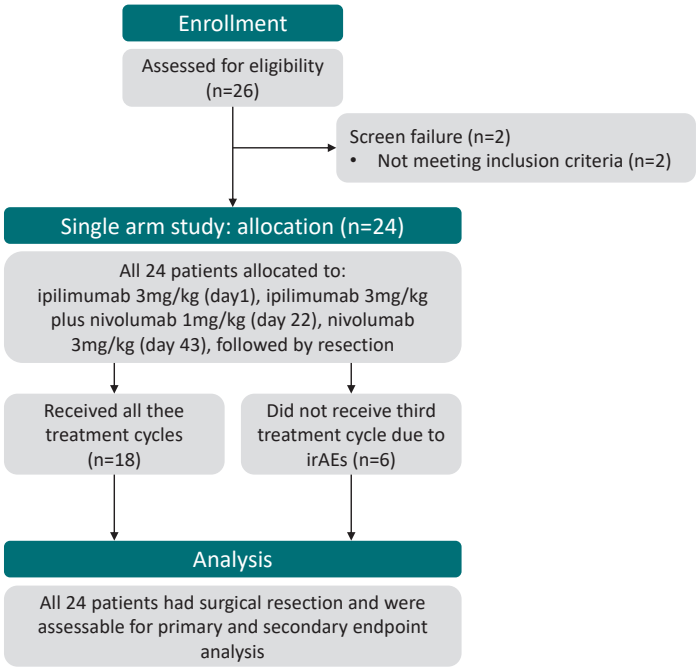

c
FoxP3+ / FoxP3– ratio
0.0086
FoxP3+ / FoxP3–
0.5
0.4
0.3
0.2
0.1
0.0
pre
post
Non–CR
CR


d
FoxP3+ / FoxP3– ratio
CR
Non–CR
0.073
0.11
FoxP3+ / FoxP3–
0.5
0.4
0.3
0.2
0.1
0.0
pre
post
pre
post

SUPPLEMENTARY MATERIAL

Supplementary Table 1. Immune-related adverse events registered up to data cut-off

	All grade, n (%)	Grade 3/4, n (%)
Any IRAE	24 (100%)	13 (55%)
Lipase increased	8 (33%)	6 (25%)
Fatigue	6 (25%)	1 (4%)
Alanine aminotransferase increased	5 (21%)	3 (12%)
Diarrhea	5 (21%)	3 (12%)
Aspartate aminotransferase increased	4 (17%)	1 (4%)
Skin rash	5 (21%)	0 (0%)
Alkaline phosphatase increased	3 (12%)	1 (4%)
Colitis	3 (12%)	2 (8%)
Gamma-glutamyltransferase increased	3 (12%)	2 (8%)
Headache	3 (12%)	0 (0%)
Hyperthyreoidism	3 (12%)	0 (0%)
Serum amylase increased	3 (12%)	0 (0%)
Adrenal insufficiency	2 (8%)	0 (0%)
Hypothyreoidism	2 (8%)	0 (0%)
Toxicoderma, DRESS syndrome	1 (4%)	1 (4%)
Blood bilirubin increased	1 (4%)	1 (4%)
Peripheral motor neuropathy	1 (4%)	1 (4%)
Hemolysis	1 (4%)	1 (4%)
Hyperglycemia	1 (4%)	1 (4%)



Supplementary Figure 1. Consort diagram