On preoperative systemic treatment of muscle-invasive bladder cancer

Dorp, J. van

CHAPTER 5

High- or low-dose preoperative ipilimumab plus nivolumab in stage III urothelial cancer: the phase 1B NABUCCO trial

Jeroen van Dorp, Christodoulos Pipinikas, Britt B.M. Suelmann, Niven Mehra, Nick van Dijk, Giovanni Marsico, Maurits L. van Montfoort, Sophie Hackinger, Linde M. Braaf, Tauanne Amarante, Charlaine van Steenis, Kirsten McLay, Antonios Daletzakis, Daan van den Broek, Maaike W. van de Kamp, Kees Hendricksen, Jeantine M. de Feijter, Thierry N. Boellaard, Richard P. Meijer, Antoine G. van der Heijden, Nitzan Rosenfeld, Bas W.G. van Rhijn, Greg Jones, Michiel S. van der Heijden

Nature Medicine, 2023

ABSTRACT

Cohort 1 of the phase 1B NABUCCO trial showed high pathological complete response (pCR) rates with preoperative ipilimumab plus nivolumab in stage III urothelial cancer (UC). In cohort 2, the aim was dose adjustment to optimize responses. Additionally, we report secondary endpoints including efficacy and tolerability in cohort 2, and the association of presurgical absence of circulating tumor DNA (ctDNA) in urine and plasma with clinical outcome in both cohorts. Thirty patients received two cycles of either ipilimumab 3 mg/kg plus nivolumab 1 mg/kg (cohort 2A) or ipilimumab 1 mg/kg plus nivolumab 3 mg/kg (cohort 2B), both followed by nivolumab 3 mg/kg. We observed a pCR in six (43%) patients in cohort 2A, and a pCR in one (7%) patient in cohort 2B. Absence of urinary ctDNA correlated with pCR in the bladder (ypT0Nx), but not with progression-free survival (PFS). Absence of plasma ctDNA correlated with pCR (odds ratio (OR): 45.0; 95% confidence interval (CI): 4.9-416.5) and PFS (hazard ratio (HR): 10.4; 95% CI: 2.9-37.5). Our data suggest that high-dose ipilimumab plus nivolumab is required in stage III UC and that absence of ctDNA in plasma can predict PFS. ClinicalTrials.gov registration: NCT03387761.

INTRODUCTION

Cisplatin-based chemotherapy followed by radical cystectomy is recommended for patients with muscle-invasive UC[1,2]. pCR (defined as ypT0N0) has been observed in 24-42% of patients and is associated with modestly improved long-term survival[2,3]. However, some patients are ineligible to receive cisplatin-based chemotherapy and the prognosis of UC patients remains poor. To improve this outlook, we have treated patients with locoregionally advanced UC (stage III; cT3-4aN0 or cT1-4aN1-3) with ipilimumab 3 mg/kg (anti-CTLA-4) plus nivolumab (anti-PD-1) in cohort 1 of the NABUCCO trial (Extended Data Fig. 1)[4]. The trial met its primary endpoint of surgical resection within 12 weeks after initiation of preoperative treatment. Moreover, we observed a pCR in eleven (46%) patients, and no residual muscle-invasive disease (ypT0/Tis/Ta/T1N0) in fourteen (58%) patients[4]. Due to immunotherapy-related adverse events, six (25%) patients missed the third treatment cycle (nivolumab 3 mg/kg), resulting in a relatively low exposure to nivolumab (one cycle of nivolumab 1 mg/kg)[4]. In other cancer types, encouraging efficacy with acceptable tolerability was observed using a lower dose of preoperative ipilimumab (1 mg/kg instead of 3 mg/kg)[5-7].

In NABUCCO cohort 2, we aimed to further optimize preoperative treatment with ipilimumab plus nivolumab in terms of efficacy and tolerability. Patients were randomized to receive either two cycles of ipilimumab 3 mg/kg plus nivolumab 1 mg/kg (cohort 2A) or two cycles of ipilimumab 1 mg/kg plus nivolumab 3 mg/kg (cohort 2B), in both cohorts followed by a third cycle of nivolumab 3 mg/kg (Extended Data Fig. 1). Similar to NABUCCO cohort 1, we included patients with cT3-4aN0 or cT1-4aN1-3 UC, who were ineligible to receive cisplatin-based chemotherapy or refused (Study Protocol in Supplementary Information)[4]. Feasibility in cohort 1 (surgery <12 weeks) was the primary endpoint of the trial[4]. Efficacy - defined as pCR rate - was a secondary endpoint of the trial and the main objective in cohort 2. Other secondary endpoints reported here include immunotherapy-related adverse events (irAEs) and the association between absence of ctDNA pre-surgery and pathological response and PFS. For biomarker purposes, patients with ypT0/Ta/Tis/T1N0 were considered responders[8].

RESULTS

Thirty patients were randomized (1:1) to cohort 2A and 2B between December 5, 2019 and April 13, 2021. Baseline characteristics can be found in Supplementary Table 1; cohort 1 patients are included for reference. Twenty-six (87%) patients received all treatment cycles; the other four (13%) patients missed one or two cycles due to irAEs. Twenty-six (87%) patients underwent radical surgery. One patient progressed on-treatment and was no longer eligible for surgery (non-responder). Three patients refused radical surgery, one of whom had a transurethral resection of the bladder and a lymph node dissection showing a nodal micrometastasis (ypTxN3). This

patient was also considered a non-responder. The two remaining patients who refused surgery were considered non-evaluable in terms of pathological response (Fig. 1a), both patients are recurrence-free after 11 and 23 months.

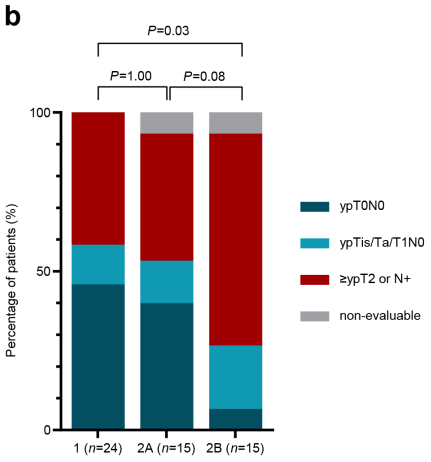
We observed a pCR in six (43%) evaluable patients treated with ipilimumab 3 mg/kg plus nivolumab 1 mg/kg (2A; 95% CI: 21-67%) and in one (7%) patient treated with ipilimumab 1 mg/kg plus nivolumab 3 mg/kg (2B; 95% CI: 0-34%; *P*=0.08; Fig. 1a-c). The pCR rate in cohort 2A was similar to cohort 1 (*P*=1.00; Fig. 1b,c). In contrast, the pCR rate in cohort 2B was significantly lower compared to cohort 1 (*P*=0.03; Fig. 1b,c). Eight (57%) patients treated with ipilimumab 3 mg/kg plus nivolumab 1 mg/kg (2A; 95% CI: 33-79%) and four (29%) patients treated with ipilimumab 1 mg/kg plus nivolumab 3 mg/kg (2B; 95% CI: 11-55%) had no residual muscle-invasive disease after surgery (ypT0/Tis/Ta/T1N0; Fig. 1a-c).

Treatment details and clinical follow-up for patients in cohort 2A and 2B (data cut-off: March 31, 2022) can be found in Extended Data Fig. 2. IrAEs were comparable to previous preoperative studies with ipilimumab plus nivolumab (Fig. 1c and Supplementary Table 2)[5-7]. Surgery was postponed in one patient due to irAEs (>12 weeks since start of therapy). Of note, the proportion of patients with a grade ≥3 irAE was significantly lower in cohort 2B (ipilimumab 1 mg/kg plus nivolumab 3 mg/kg) compared to cohort 1 (cohort 1: 58%, cohort 2B: 20%; *P*=0.026, two-sided Fisher's Exact test).

Radical cystectomy is a surgical procedure with significant morbidity and mortality[9]. Given the high rate of pCR after treatment with ipilimumab plus nivolumab, resection may not always be necessary and a bladder-sparing approach could be feasible in a subset of patients. However, current diagnostic tools cannot accurately predict a pathological response upon neoadjuvant therapy[10]. Recently, ctDNA in plasma was correlated with clinical outcome in MIBC patients treated with preoperative and adjuvant atezolizumab[11,12]. To evaluate the effect of preoperative treatment with ipilimumab and nivolumab in patients with locoregionally advanced MIBC, we assessed ctDNA in urine and plasma before and during therapy, and prior to surgery, using the commercially available RaDaR assay (Supplementary Fig. 1)[13].

Based on whole-exome sequencing of tumor and germline material, we identified tumor-specific genetic variants for all patients in NABUCCO cohort 1 (*n*=24) and isolated cell-free DNA from plasma and urine samples (Extended Data Fig. 3, Supplementary Table 3-6 and Methods)[13]. The mean estimated variant allele frequency (eVAF%) in urine was not statistically different between responders (ypT0/Tis/Ta/T1N0) and non-responders (≥ypT2 or N+) at any time point (Extended Data Fig. 4a)[8]. We observed no correlation between absence of ctDNA in urine prior to surgery and pathological response (*P*=0.39; Extended Data Fig. 4a-c). When comparing patients with and without a pCR in the bladder (ypT0Nx) specifically, we observed a correlation with absence of urinary ctDNA prior to surgery (*P*<0.01; Extended Data Fig. 4a-c). As some patients with absence of urinary ctDNA prior to surgery had ypT0N+ disease, there was no correlation with pCR (ypT0N0;

P=0.08) or clinical outcome (Extended Data Fig. 4b-d and 5), making urinary ctDNA a suboptimal biomarker for clinical decision-making.



c

	Cohort 1 (*n*=24)	Cohort 2A (*n*=15)	Cohort 2B (*n*=15)
Pathological response, *n* (%)			
ypT0N0	11/24 (46)	6/14 (43)	1/14 (7)
ypT0/Tis/Ta/T1N0	14/24 (58)	8/14 (57)	4/14 (29)
Immunotherapy-related adverse events, *n* (%)			
Any	24/24 (100)	15/15 (100)	11/15 (73)
Grade ≥3	13/24 (55)	5/15 (33)	3/15 (20)

Figure 1 | Treatment and response details for patients in cohort 2. a, Patient details for cohort 2A and 2B. Patients are organized based on response. cTNM-stage before treatment and ypTNM-stage after surgery are shown above and below the plot, respectively. **b,** Bar chart showing pathological responses for all cohorts. Cohort 1 vs 2A: *P*=1.00; cohort 1 vs 2B: *P*=0.03; cohort 2A vs 2B: *P*=0.08. *P*-values: two-sided Fisher's Exact test comparing pCR rate (ypT0N0) between cohorts. **c,** Table showing pathological responses for evaluable patients and immunotherapy-related adverse events. *Ipi: ipilimumab, Nivo: nivolumab, TURBT: transurethral resection of the bladder tumor, LND: lymph node dissection.*

In plasma, the mean eVAF% was not statistically different between responders and non-responders at baseline. We observed a significantly lower mean plasma eVAF% prior to surgery in responders compared to non-responders (P<0.01; Fig. 2a). ctDNA was not detectable in the latest available plasma sample prior to surgery in thirteen (93%) responders and in four (40%) non-responders (P<0.01; Fig. 2b-d). To validate these observations, we also assessed plasma ctDNA collected from patients treated in the Netherlands Cancer Institute in cohort 2A and 2B (n=17; Supplementary Table 3-6 and Extended Data Fig. 6). CtDNA was not detectable in plasma prior to surgery in five (100%) responders and in two (18%) non-responders (P<0.01; Fig. 2b-d) showing a similar correlation between absence of ctDNA in plasma prior to surgery and pathological response as observed in cohort 1. Taken together, the OR for cohort 1+2 combined was 45.0 (95% CI: 4.9-416.5; Fig. 2b-d). Finally, we analyzed PFS for patients (n=24+17) with and without detectable ctDNA in the latest available plasma sample. We observed a strong association between absence of ctDNA and PFS (HR: 10.4; 95% CI: 2.9-37.5; Fig. 2e and Extended Data Fig. 5).

DISCUSSION

There are several limitations to our findings. Most importantly, the sample size was relatively small. Although the response rate was significantly different between cohort 1 and cohort 2B, suggesting a difference in pCR between high and low ipilimumab schedules, we could not show a statistical difference in response rates between cohort 2A and 2B. Still, given the magnitude of the difference in pCR rate between the cohorts and the results of the CheckMate-032 trial, favoring ipilimumab 3 mg/kg plus nivolumab over ipilimumab 1 mg/kg plus nivolumab in the metastatic setting[14], we decided against a further expansion of our study. Ultimately, the efficacy of preoperative ipilimumab 3 mg/kg in combination with nivolumab in urothelial cancer will need to be confirmed in a randomized trial. The underlying biological mechanism of the context-dependent dose-response association for anti-CTLA-4 therapy is currently unknown.

The precise value of urinary ctDNA as a biomarker for response was not fully addressed in our study, as it was assessed in a limited number of samples (only cohort 1). Whereas our results suggest absence of urinary ctDNA after preoperative therapy is unable to accurately predict PFS or OS, it may provide a useful biomarker to assess bladder-specific pCR (ypT0). Interestingly, some patients having a pCR in the bladder still had detectable urinary ctDNA, which could potentially be derived from residual non-viable tumor tissue. Assessment of ctDNA in plasma and urine in larger studies, including later time points, may provide additional granularity on ctDNA dynamics and help to build personalized strategies to de-escalate

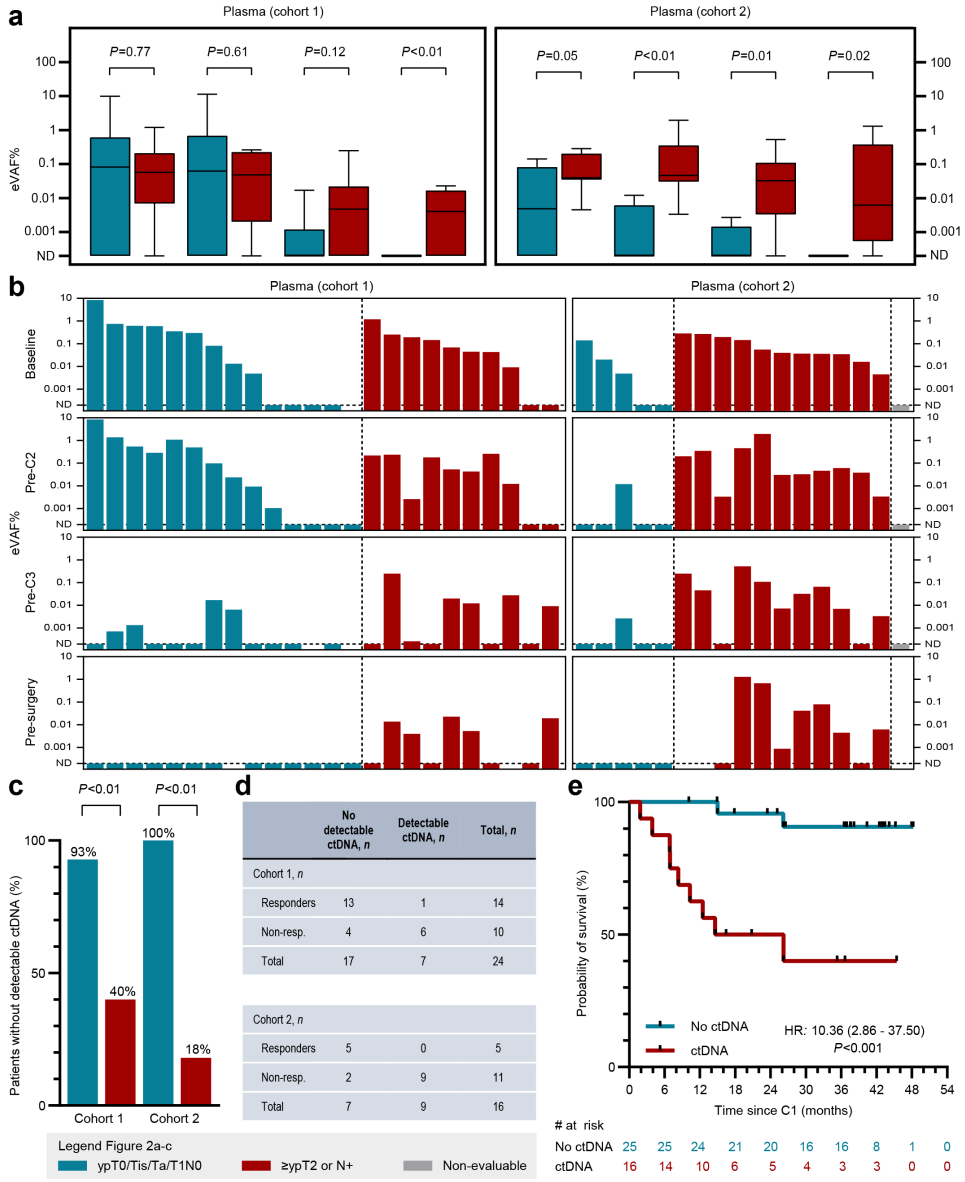


Figure 2 | ctDNA assessment in plasma and relation to outcome. a-c, Data for patients in cohort 1 (left, *n*=24) and cohort 2 (right, *n*=17). Blue bars: responders (ypT0/Tis/Ta/T1N0), red bars: non-responders (≥ypT2 and/or N+), grey bars: non-evaluable patients, missing bars: missing samples. Non-evaluable patients were excluded in responders versus non-responders comparisons. Undetectable values are plotted as 0.0002. **a,** Mean eVAF% in plasma for responders and non-responders before (cohort 1: *P*=0.77; cohort 2: *P*=0.05) and during treatment, and pre-surgery (cohort 1: *P*<0.01; cohort 2: *P*=0.02). Boxplots show median value and 25th and 75th percentiles. Whiskers show minimum/maximum values. *P*-values: two-sided Mann-Whitney test. **b,** Individual eVAF% in plasma before and during treatment, and pre-surgery. **c,** Percentage of patients with no detectable ctDNA in latest available plasma sample pre-surgery for responders and non-responders per cohort. Cohort 1: *P*<0.01; cohort 2: *P*<0.01. *P*-values: two-sided Fisher's Exact test comparing patients with detectable ctDNA pre-surgery in responders to non-responders for cohort 1 and 2 separately. **d,** Number of

patients with and without detectable ctDNA in latest available plasma sample pre-surgery for responders and non-responders per cohort. **e,** Progression-free survival for patients with and without detectable ctDNA in latest available plasma sample pre-surgery. Hazard ratio: 10.36; 95% confidence interval: 2.86-37.50; *P*<0.001 using a two-sided Log-rank (Mantel-Cox) test. *ND: not detectable, Pre-C2/3: prior to second/third treatment cycle.*

UC treatment. Clinical strategies could include a wait-and-see approach with longitudinal ctDNA monitoring in patients having absence of ctDNA in both plasma and urine after preoperative therapy.

In conclusion, we confirmed in an independent cohort that preoperative treatment with ipilimumab 3 mg/kg - in combination with nivolumab - leads to high pCR rates in stage III UC, whereas ipilimumab 1 mg/kg appeared less effective. The absence of pre-surgery ctDNA in plasma predicts pathological response and PFS and could potentially guide clinical decision-making to select patients for bladder-sparing strategies after treatment with ipilimumab plus nivolumab. The ongoing INDIBLADE trial (ClinicalTrials.gov registration: NCT05200988) will assess the efficacy of preoperative ipilimumab plus nivolumab followed by chemoradiation therapy as a bladder-sparing strategy.

ACKNOWLEDGEMENTS

We would like to thank all patients and their families who participated in the NABUCCO trial, all staff involved in the care for patients participating in this trial in the Netherlands Cancer Institute, the University Medical Center Utrecht and the Radboud University Medical Center. We thank the clinical trial teams at participating centers and all staff involved in the collection, processing and storage of trial samples and tumor material. We thank A. Gil-Jimenez for processing the DNA sequencing data and C. Blank for input on the trial design. We thank Bristol-Myers Squibb for funding of the trial (CA209-9Y4) and for providing the study drugs. The KWF Dutch Cancer Society provided institutional funding.

REFERENCES

1. Witjes, J.A., *et al.* European Association of Urology Guidelines on Muscle-invasive and Metastatic Bladder Cancer: Summary of the 2020 Guidelines. *Eur Urol* 79, 82-104 (2021).
2. Advanced Bladder Cancer Meta-analysis, C. Neoadjuvant chemotherapy in invasive bladder cancer: update of a systematic review and meta-analysis of individual patient data advanced bladder cancer (ABC) meta-analysis collaboration. *Eur Urol* 48, 202-205; discussion 205-206 (2005).
3. Pfister, C., *et al.* Randomized Phase III Trial of Dose-dense Methotrexate, Vinblastine, Doxorubicin, and Cisplatin, or Gemcitabine and Cisplatin as Perioperative Chemotherapy for Patients with Muscle-invasive Bladder Cancer. Analysis of the GETUG/AFU V05 VESPER Trial Secondary Endpoints: Chemotherapy Toxicity and Pathological Responses. *Eur Urol* 79, 214-221 (2021).
4. van Dijk, N., *et al.* Preoperative ipilimumab plus nivolumab in locoregionally advanced urothelial cancer: the NABUCCO trial. *Nat Med* 26, 1839-1844 (2020).
5. 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).
6. 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).
7. 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).
8. Rosenblatt, R., *et al.* Pathologic downstaging is a surrogate marker for efficacy and increased survival following neoadjuvant chemotherapy and radical cystectomy for muscle-invasive urothelial bladder cancer. *Eur Urol* 61, 1229-1238 (2012).
9. Stimson, C.J., *et al.* Early and late perioperative outcomes following radical cystectomy: 90-day readmissions, morbidity and mortality in a contemporary series. *J Urol* 184, 1296-1300 (2010).
10. Becker, R.E.N., *et al.* Clinical Restaging and Tumor Sequencing are Inaccurate Indicators of Response to Neoadjuvant Chemotherapy for Muscle-invasive Bladder Cancer. *Eur Urol* 79, 364-371 (2021).
11. Powles, T., *et al.* ctDNA guiding adjuvant immunotherapy in urothelial carcinoma. *Nature* 595, 432-437 (2021).
12. Szabados, B., *et al.* Final Results of Neoadjuvant Atezolizumab in Cisplatin-ineligible Patients with Muscle-invasive Urothelial Cancer of the Bladder. *Eur Urol* 82, 212-222 (2022).
13. Flach, S., *et al.* Liquid BIOpsy for MiNimal RESidual DiSease Detection in Head and Neck Squamous Cell Carcinoma (LIONESS)-a personalised circulating tumour DNA analysis in head and neck squamous cell carcinoma. *Br J Cancer* 126, 1186-1195 (2022).
14. 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).

METHODS

Study design and population

NABUCCO (ClinicalTrials.gov: NCT03387761) is a multicenter investigator-initiated, prospective phase 1B trial which was conducted in two subsequent phases. In the first phase (NABUCCO cohort 1) 24 patients with resectable locoregionally advanced urothelial cancer (UC; stage III; cT3-4aN0 or cT1-4aN1-3) were enrolled at the Netherlands Cancer Institute and treated with preoperative ipilimumab 3 mg/kg (day 1), ipilimumab 3 mg/kg plus nivolumab 1 mg/kg (day 22) and nivolumab 3 mg/kg (day 43), followed by cystectomy or nephroureterectomy with lymph node dissection[4]. In the second phase (NABUCCO cohort 2A and 2B) 30 patients with resectable locoregionally advanced UC were enrolled at the Netherlands Cancer Institute, the University Medical Center Utrecht and the Radboud University Medical Center between December 5, 2019 and April 13, 2021. Patients were randomized (1:1, open-label) and treated with either (2A) ipilimumab 3 mg/kg plus nivolumab 1 mg/kg (day 1 and 22) and nivolumab 3 mg/kg (day 43), or (2B) ipilimumab 1 mg/kg plus nivolumab 3 mg/kg (day 1 and 22) and nivolumab 3 mg/kg (day 43), both followed by cystectomy or nephroureterectomy with appropriate lymph node dissection (Extended Data Fig. 1). Disease staging was performed according to American Joint Committee on Cancer guidelines. Patients were ≥18 years of age and were either ineligible for cisplatin or refused cisplatin-based chemotherapy, have not received prior treatment with anti-CTLA-4, anti-PD-1 or anti-PD-L1 agents, had a World Health Organization performance score of 0–1, and had diagnostic transurethral resection bladder tumor (TURBT) blocks available (or other baseline tumor material in case of upper tract tumors). Key exclusion criteria included documented severe autoimmune disease, chronic infectious disease and use of systemic immunosuppressive medications. All patients provided written informed consent. No financial compensation was offered to patients except to cover travel expenses to and from the hospital. Surgery was scheduled in weeks 9–12 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 pelvic magnetic resonance imaging (MRI) assessment was optional. All adverse events and immunotherapy-related adverse events were graded and reported throughout the study according to CTCAE v5.0. The sponsor of the trial was the Netherlands Cancer Institute. Bristol-Myers Squibb provided the study drugs and funding for the trial (CA209-9Y4). 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 (METC-AVL, currently named NedMEC) 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. For more information, refer to the study protocol (Study Protocol in Supplementary Information).

CLINICAL DATA ACQUISITION AND INTERPRETATION

Clinical data was collected in electronic Case Report Forms (eCRF) using Tenalea (version 18.1) at the Netherlands Cancer Institute, University Medical Center Utrecht and Raboud University Medical Center and further processed and analyzed in the Netherlands Cancer Institute. Clinical follow-up visits were planned for all patients at 1, 4 and 8 weeks after surgery. CT-scans and follow-up visits were planned according to local protocol but at a minimum included CT scans 12 and 24 months after surgery. The date of the first treatment cycle (C1) was used as date of reference to determine progression-free and overall survival. Data cut-off was done on March 31, 2022. All eligible patients were contacted prior to this date and last date of contact was used as date of reference to determine progression-free and overall survival. Progression was defined as the first evident sign of recurrence of disease through clinical, radiological and/or pathological assessment. Patients who were alive and had no evident signs of progression were censored on the last day of contact before the data cut-off date for progression-free and overall survival analysis. All deceased patients - regardless of cause of death - were considered an event for overall survival analysis and were censored on the date of death for progression-free survival analysis if there were no evident signs of progression.

DNA ISOLATION AND SEQUENCING

TURBT, cystectomy and/or nephroureterectomy material was stored as formalin-fixed, paraffin embedded (FFPE) tissue blocks in The Netherlands Cancer Institute after surgery. Additional tumor material from participating hospitals was requested as FFPE tissue blocks or as cut slides from University Medical Center Utrecht, Radboud University Medical Center or other referring hospitals.

Tumor area and tumor cell percentage for every specimen was assessed by a trained pathologist. DNA was isolated from baseline tumor material (5-10x 10 µm slides) using the AllPrep FFPE DNA/RNA Kit (Qiagen), using the QIAcube according to manufacturer's protocol. Germline DNA was extracted from peripheral blood mononuclear cells from every patient prior to start of treatment (baseline) using the QIAsymphony DSP DNA Midi kit (Qiagen). 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 Hyper DNA Library Kit (KAPA Biosystems) was used for library preparation. IDT Human Exome capture v2.0 (xGen Exome Research Panel v2.0) was used for exome enrichment. Libraries were sequenced with 100 base pairs paired-end reads using the Novaseq 6000 (Illumina) with a high output mode PE 100.

RAW TUMOR DNA SEQUENCING DATA ANALYSIS

Raw whole exome sequencing (WES) data were transferred to Inivata as BAM files and converted to FASTQ file format. A custom analytical pipeline based on the GATK best practices was used to process the FASTQ files by first aligning the sequencing reads to the human genome (hg38) using bwa mem (version 0.7.17) and marking duplicates by Picard (version 2.23.8), followed by a

proprietary algorithm performing somatic variant calling. Germline variants were removed using information available from public single nucleotide polymorphism (SNP) databases and variants were filtered and prioritized based on multiple parameters including allele frequency and depth. Patient WES data were subjected to full quality control by assessing key sequencing metrics (Supplementary Table 4). A matched germline genomic DNA sample available from all patients was also subjected to WES to ensure that germline variants were filtered out. The median read depth per sample based on usable (i.e. not duplicated, on-target) reads was 72 (range: 8-107) for tumor WES compared to a median of 71 (range: 46-113) for germline WES. The estimated purity of the tumor material was a median fraction of 0.41 (range: 0.18-0.70).

RADAR PATIENT-SPECIFIC ASSAY DESIGN AND PANEL QUALIFICATION

RaDaR is a multiplex PCR, next generation sequencing (NGS) liquid biopsy assay built on the InVision® platform[15] that uses a tumor-informed approach for the sensitive detection of circulating tumor DNA (ctDNA) in a patient's blood circulation. Refer to the paragraph below (*RaDaR assay analytical validation*) for more details. Tumor WES-derived somatic variants (single nucleotide variants (SNPs) and indels) were prioritized using a proprietary algorithm to build patient-specific primer panels of up to 48 primer pairs each capturing at least one somatic variant. Variant prioritization and selection criteria take into consideration the variant set most suitable for the sensitive detection of plasma ctDNA in the patient for whom the assay was designed for as well as ensuring that the selected primer pairs are well suited for multiplex PCR allowing the efficient amplification of the target ctDNA. Each designed RaDaR panel was synthesized (Integrated DNA Technologies, Coralville, IA) and combined with a fixed primer panel of common population-specific SNPs for quality control purposes during the NGS testing. Panel qualification involved NGS testing of a reference DNA standard, a negative amplification control and FFPE tumor tissue DNA on iSeq100 (Illumina). Panels with 8 or more tumor-confirmed variants detected at sufficient read depth were considered as qualified. RaDaR is an US, EU and UK registered trademark of Inivata, Ltd.

PLASMA AND URINE COLLECTION AND PROCESSING

Serial blood samples (*n*=159) from 41 patients were collected by standard phlebotomy techniques in 2x10mL K2-EDTA tubes from up to 4 different time points, namely prior to start of treatment (day 1, baseline), prior to each subsequent treatment cycle (day 22 and 43) and prior to surgery. Blood samples were processed to plasma within two hours after collection with an initial centrifugation at 380x g for 20 min followed by a second centrifugation of isolated plasma aliquots at 20,000x g for 10 min to remove any remaining cellular debris and stored at -80 °C. Urine samples (*n*=93) were collected together with plasma samples for patients in cohort 1 (*n*=24). Urine samples were spun down to remove any cells and cellular debris processed at 380x g for 20 min and stored at -80 °C.

CFDNA EXTRACTION AND QUANTIFICATION

Plasma and urine samples were sent to Inivata (Cambridge, UK) for further processing and

analysis. Cell-free DNA (cfDNA) was isolated from 3-4 mL plasma using a solid phase reverse immobilization magnetic bead protocol on a Hamilton Microlab STAR automated platform. A *no-template* control sample was included in each extraction batch to monitor for contamination. cfDNA was isolated from 8 mL urine supernatant using the QIAsymphony SP protocol following the manufacturer's instructions (Qiagen). The isolated cfDNA was quantified on a QX200 ddPCR system (BioRad, Hercules, CA) using a custom 108 bp digital droplet PCR assay (ddPCR) measuring the amplification of the RPP30 gene (ribonuclease P/MRP subunit P30). The primers and probes used were as follows:

- Forward = 5'-GGAGGTGGAGGAGGAGGATA-3';
- Reverse = 5'-ACGGAATACAGAACCCATGACT-3';
- Probe = 5'-FAM/AGCCTTGAG/ZEN/ AGACGAGAACCTGT/IABkF Q-3'[16].

CFDNA RADAR PROFILING

Following panel qualification, personalized RaDaR assays were used in multiplex PCR amplification of plasma and urine cfDNA. For plasma, median input as measured by ddPCR was 6,900 amplifiable copies per sample (range: 2,000 to 20,224 copies). For urine, median input was 20,020 amplifiable copies per sample (range: 200-25,224). In plasma, sequencing depth was a median value of 241,930 reads per variant per sample (range: 81,661 to 1,506,623) and a total median of 11,743,870 reads per sample across all variants (range: 4,001,393 to 72,317,915). In urine, sequencing depth was a median value of 238,805 reads per variant per sample (range: 874 to 953,341) and a total median of 11,512,230 reads per sample across all variants (range: 37,594 to 46,713,722). Control DNA was also tested to ensure that germline variants and confounding signals derived from age-related CHIP somatic variants (clonal hematopoiesis of indeterminate potential) were filtered out. The amplicon products were then barcoded and the resulting libraries were pooled, combined with PhiX and subjected to ultra-deep sequencing on the Illumina NovaSeq 6000 platform (Illumina Inc., San Diego, USA) to allow sensitive detection of low ctDNA levels.

RADAR DATA ANALYSIS, VARIANT FILTERING AND CTDNA STATUS CALLING

RaDaR data were analyzed in a multistep process. Raw sequencing base calls were converted to FASTQ format and then demultiplexed using bcl2fastq2 (version 2.20) before being aligned to the human reference genome using bwa mem. Proprietary software was used to identify primer pairs and count mutant and reference bases for each variant. Individual variants were subjected to strict quality control and only those detected in the tumor DNA sequencing but absent in the control genomic DNA sequencing using the RaDaR panel were considered in the process of calling plasma samples ctDNA positive or negative. Variants absent in tumor DNA when tested with the personalized panel were filtered as these are typically WES false positives or variants that have failed to amplify, sequence or align to the target region. The median number of tumor confirmed variants was 47 (range: 18 to 51), corresponding to a median of 98% confirmed variants (range: 35%-100%), with 83% of samples having at least 40 confirmed variants (Supplementary Table 4 and 6).

To obtain evidence for the presence or absence of ctDNA at the sample level, the statistical significance of the observed mutant counts for each of the remaining variants was assessed using a statistical framework incorporating the entire set of personalized variants passing the above filter. This framework compared the sequencing counts of each variant to a model of noise for each individual variant and was locked prior to the analyses described in this study. A sample was classified as "ctDNA positive" when its cumulative statistical score was above a pre-set threshold defined during the assay's analytical development and validation of the locked assay[13]. Refer to the paragraph below (*RaDaR assay analytical validation*) for more details. Further in-depth details of the statistical model to determine the ctDNA status of each sample, the pre-set statistical thresholds and the process of calculating the estimated variant allele frequency are all core part of the proprietary code and fully IP-protected and as such cannot be disclosed.

The tumor fraction for each sample was estimated and reported as the "estimated variant allele frequency" (eVAF; Supplementary Table 5). The estimated 95% limit of detection (LoD95, column "Est. lod" in Supplementary Table 5) for each sample has been assessed using linear interpolation of pre-tabulated assay validation data for 2,000 and 20,000 input copies for different number of variants: the estimated LoD per sample had a median value of 57 ppm (0.0057%) and a range of 11.4 to 120 ppm (0.0011 to 0.0120 %). In this study, plasma detection and quantification of these variants by RaDaR was used as a proxy to assess ctDNA dynamics during preoperative treatment and evaluate response to treatment and correlation with progression-free survival.

RADAR ASSAY ANALYTICAL VALIDATION

The RaDaR assay was analytically validated in Inivata's CLIA laboratory (RTP, North Carolina) by trained operators. Personalized RaDaR assays were designed and synthesized for 5 lung cancer patient samples and 2 breast cancer (HCC1395 & HCC1954) and 1 colon cancer (SW480) cell lines. The lung cancer patients were part of the INI-001 study (ClinicalTrials.gov: NCT02906852). All patients had had ctDNA previously detected in their blood using InVisionFirst-Lung®. To determine analytical specificity, the 8 personalized RaDaR panels were tested against 69 blood samples from 18 healthy donors drawn into Streck cfDNA blood collection tubes. No calls were made in any of the 69 samples. A further 20 blood samples from 5 of the same healthy donors were drawn into EDTA tubes and run against the breast and colon cancer cell line RaDaR assays and again no calls were made, thereby demonstrating 100% specificity (Supplementary Fig. 1a).

LoD of the RaDaR assay was tested using the two breast cancer and one colon cancer cell lines. Cell line-derived DNA was diluted into normal DNA to produce contrived samples mimicking ctDNA at a range of levels near the target LoD of the RaDaR assay. To achieve this, each of the cell lines along with normal DNA from the Genome in a Bottle sample NA12878 was sheared to approximately 160 bp in length to mimic cfDNA fragmentation. NA12878 was selected as it has been extensively characterized with other sequencing methods. The fragmented DNA was quantified in copies per microliter by digital PCR (copies/µl) to enable accurate dilution. DNA was then serially diluted into

normal DNA samples from 0.008%, 0.004%, 0.002% to 0.001% variant allele fraction respectively. Each cell line dilution was analyzed in duplicate on 5 separate runs using 20,000 amplifiable copies of the genome as measured by digital PCR. The runs were performed by two operators with two reagent lots (Supplementary Fig. 1b). In total, each dilution was assessed 30 times (10 replicates of each cell line). Sensitivity of the RaDaR assay was measured using Probit-based analysis. Using this approach, the LoD95 (the limit where 95% of samples are detected) of the RaDaR assay is 0.0011% variant allele fraction. Good reproducibility was also observed between runs, operators, and reagent lots.

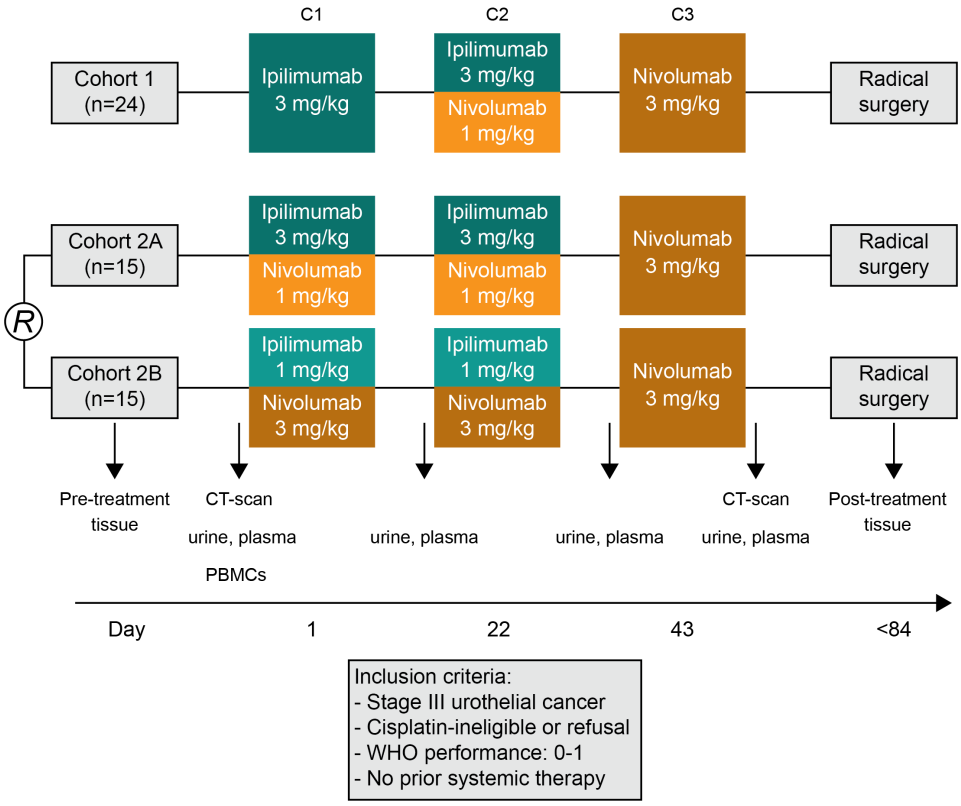
To assess the impact of the number of available variants on the LoD, results from the cell line dilution study were randomly sub-sampled (bootstrapped) removing different numbers of variants before calling. As the number of variants decreases, the sensitivity decreased in a predictable fashion (Supplementary Fig. 1c). For example, the LoD with 48 variants (0.0011%) is close to half the LoD when using half the number of variants (24 variant LoD95 = 0.0023%). This showed that targeting 48 variants improves sensitivity of the RaDaR assay. Furthermore, this demonstrated that in instances where less than 48 variants are available, the assay is still very sensitive, and this sensitivity is predictable based on the number of variants.

STATISTICAL ANALYSIS

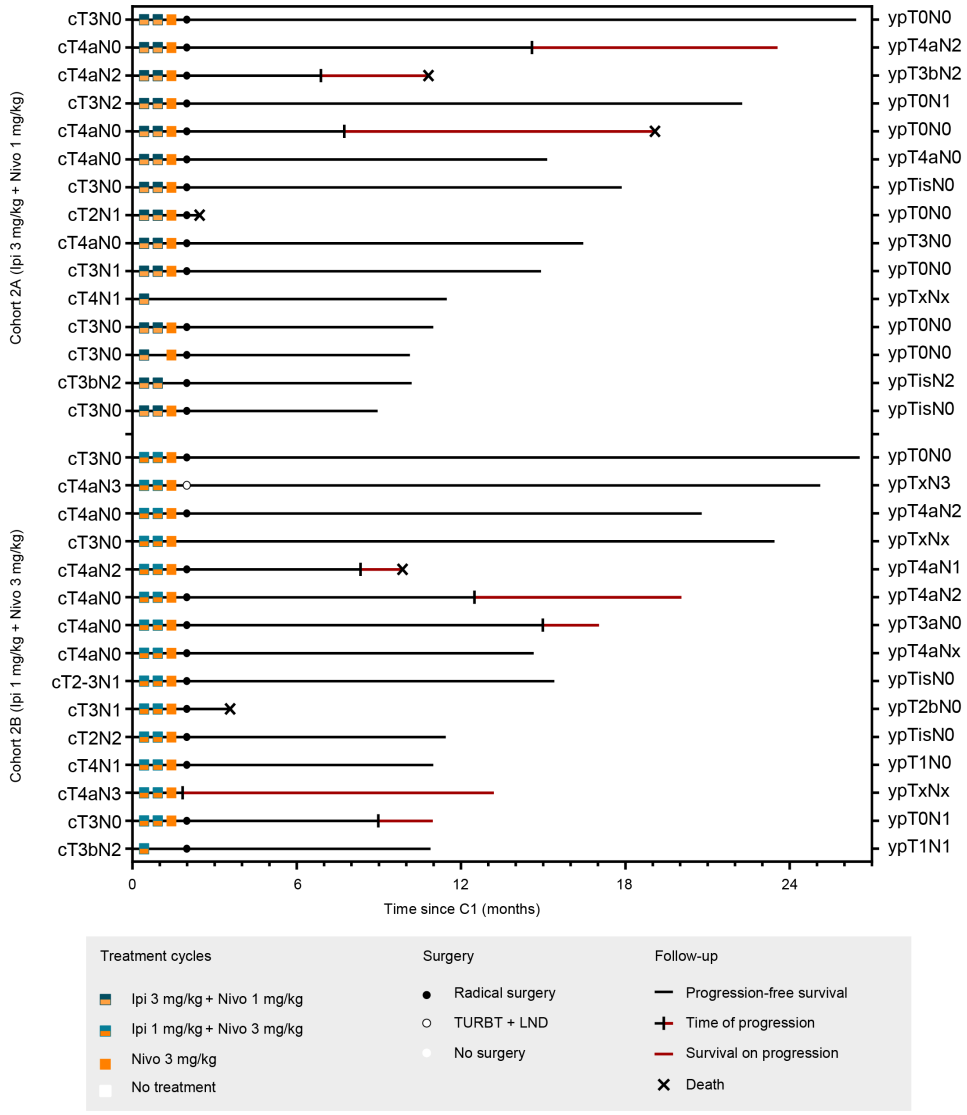
Information on statistical methodology related to individual figure panels can be found in the corresponding figure legends. General statistical methodology is discussed in this subsection. The interpretation of the secondary endpoint of pCR in cohort 2 was mainly exploratory to support further expansion or a randomized trial. Nevertheless, a power calculation was done to support patient numbers. We assumed a comparator pCR rate for chemotherapy of 21% in this population. Based on the results in cohort 1, we presumed a target of 55% pCR for cohort 2A and 2B, which is higher than the observed pCR rate in cohort 1, as we hypothesized that introducing both ipilimumab and nivolumab from the start of treatment may increase response rates. With 15 patients per treatment arm, cohort 2A and 2B would have a power of 82% to exclude a pCR rate of 21%, with a two-sided alpha of 0.05. All 95% confidence intervals for pCR rate for the separate cohorts were calculated by modified Wald method. ctDNA measurements shown in Figure 2 and Extended Data Fig. 4 are based on a single measurement per time point per patient, as described in the RaDaR sequencing analysis subheading. All statistical tests described are two-sided. No corrections for multiple hypothesis testing have been conducted. No tests for normality of data were performed. Statistical tests were performed using R (version 4.2.0) and GraphPad Prism (version 9.0.2).

METHODS REFERENCES

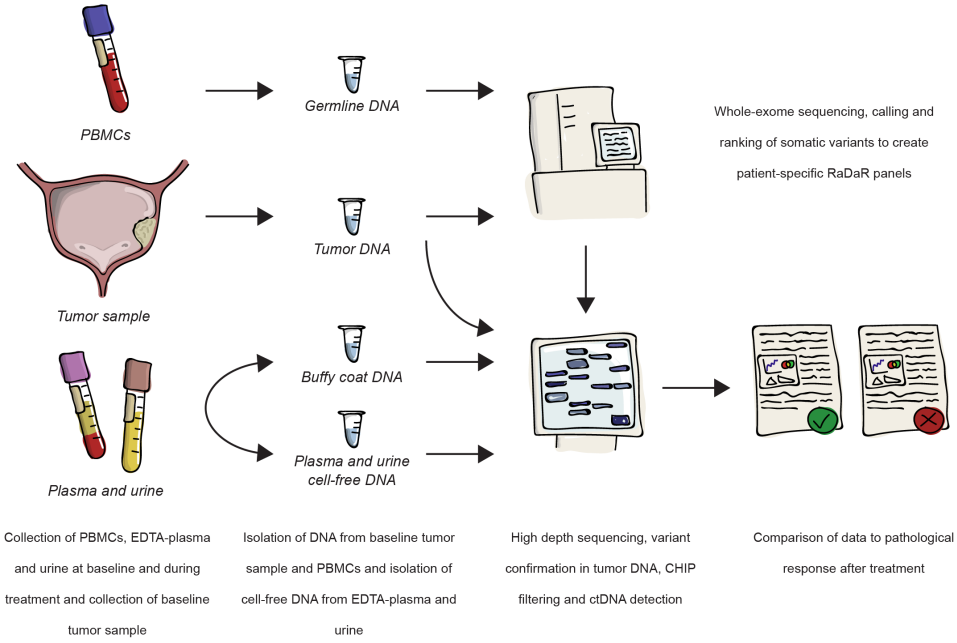
15. Plagnol, V., *et al.* Analytical validation of a next generation sequencing liquid biopsy assay for high sensitivity broad molecular profiling. *PLoS One* **13**, e0193802 (2018).
16. Gale, D., *et al.* Development of a highly sensitive liquid biopsy platform to detect clinically-relevant cancer mutations at low allele fractions in cell-free DNA. *PLoS One* **13**, e0194630 (2018).



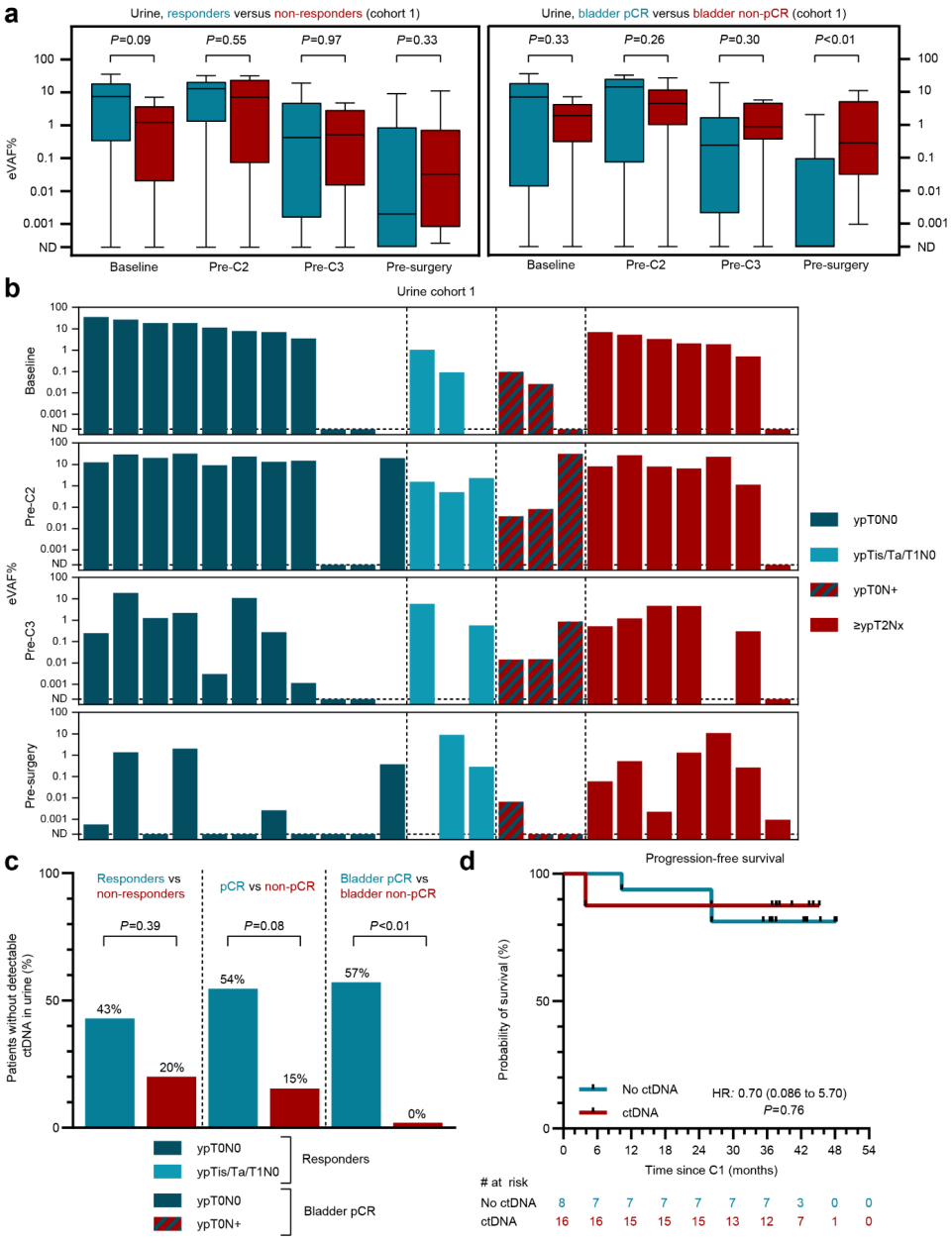
Extended Data Figure 1 | NABUCCO study design. Cohort 1 is shown at the top, cohort 2A and cohort 2B are shown at the bottom. Arrows correspond with timing of study procedures and apply to all cohorts. Peripheral blood mononuclear cells (PBMCs), urine and plasma samples were collected <24 hours prior to corresponding treatment cycle or radical surgery. *C1/2/3: first, second and third treatment cycle.*



Extended Data Figure 2 | Treatment, response and follow-up details for every patient treated in cohort 2A and 2B. Swimmer plot with treatment and follow-up details for every patient in cohort 2A and 2B. Patients are organized per cohort and based on inclusion date. Time is relative to the date of first treatment of ipilimumab plus nivolumab (C1). Clinical TNM-stage prior to start of treatment is shown on the left side of the plot. Pathological TNM-stage after surgery is shown on the right side of the plot. Orange and green boxes represent treatment cycles administered. Black dots represent radical surgery. Open circles represent alternative surgery (transurethral resection bladder tumor and pelvic lymph node dissection). Black horizontal lines represent progression-free survival. Vertical lines indicate disease progression. Red horizontal lines represent survival after disease progression. Black cross indicates death. Two fatal surgery-related adverse events occurred within 90 days after surgery, both hemorrhage after an intra-operative complication. After extensive review, no relation with ipilimumab and/or nivolumab was established. *Ipi: ipilimumab, Nivo: nivolumab, TURBT: transurethral resection of the bladder tumor, LND: lymph node dissection, C1: first treatment cycle.*

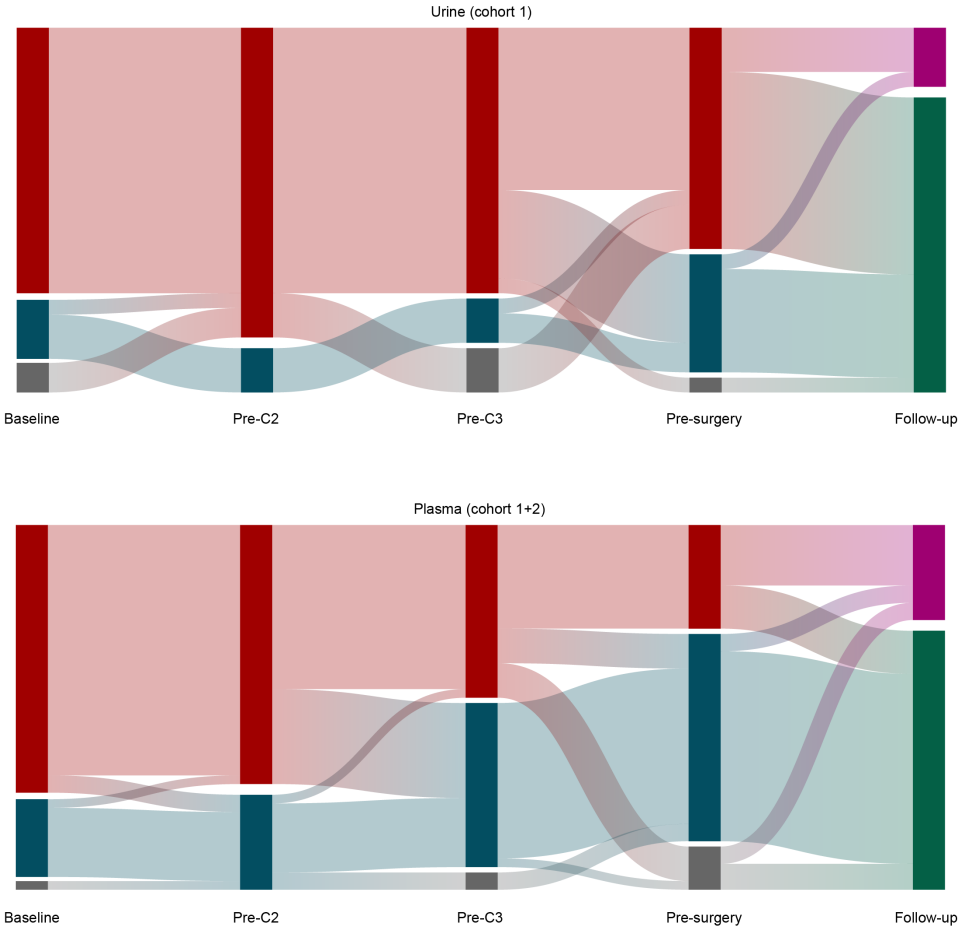


Extended Data Figure 3 | Workflow of RaDaR panel generation and ctDNA assessment in plasma and urine. DNA was isolated from baseline tumor sample and from PBMCs and used for whole-exome sequencing to identify tumor-specific somatic mutations. A personalized ctDNA assay was developed for each patient. Cell-free DNA and buffy coat DNA from EDTA-plasma and urine samples were analyzed together with tumor DNA using the personalized assays to confirm somatic mutations and perform CHIP filtering. *PBMCs: peripheral blood mononuclear cells, ctDNA: circulating tumor DNA, CHIP: clonal hematopoiesis of indeterminate potential.*

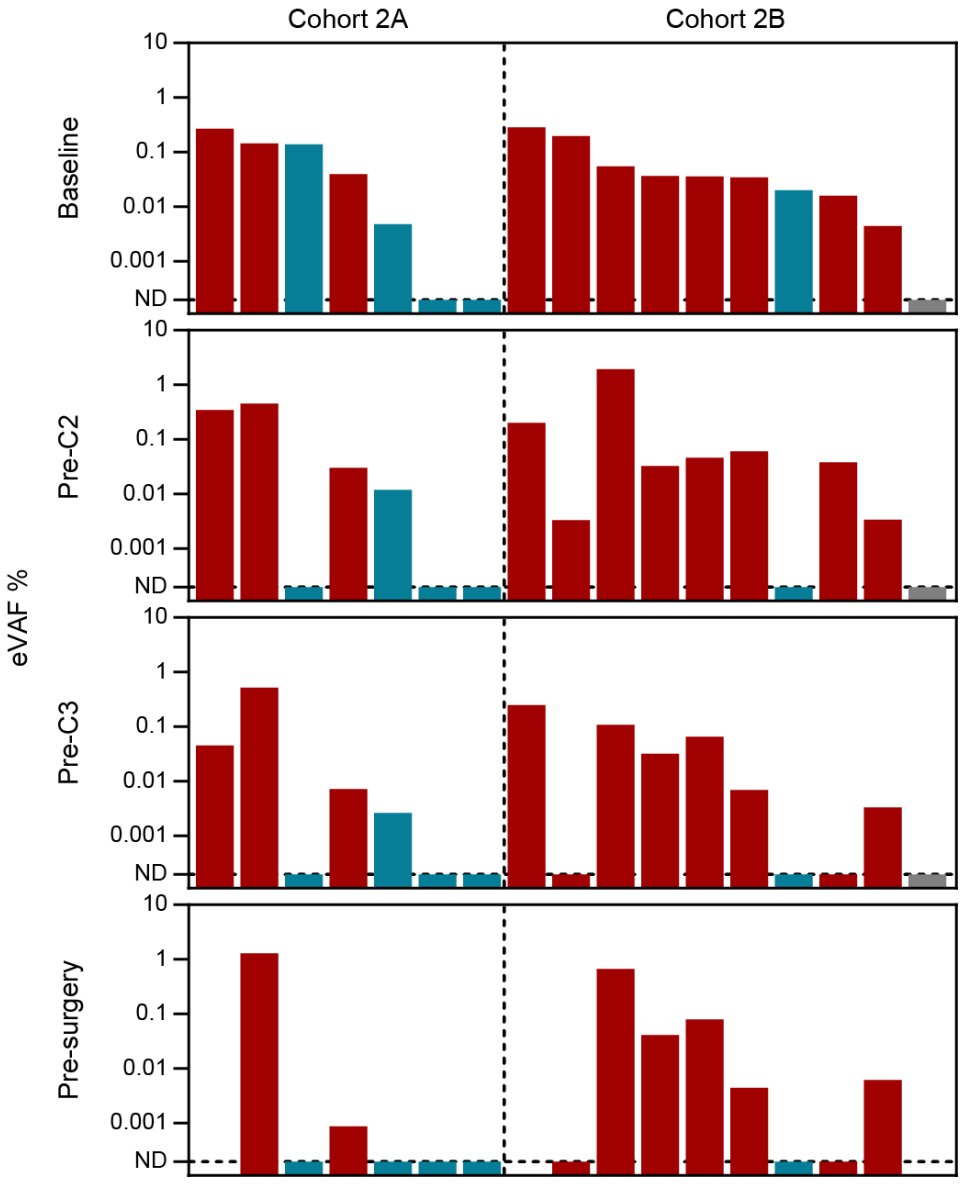


Extended Data Figure 4 | ctDNA assessment in urine and relation to outcome. Urine was assessed in patients in cohort 1 (*n*=24). **a,** Mean eVAF% in urine before and during treatment, and pre-surgery for responders and non-responders (left; baseline: P=0.09; pre-surgery: P=0.33) and for patients with and without a pathological complete response in the bladder (ypT0Nx versus ypTis/Ta/T1-4aNx) (right; baseline: P=0.33; pre-surgery: P<0.01). Boxplots show median value and 25th and 75th percentiles. Whiskers show minimum/maximum values. P-values: two-sided Mann-Whitney test. **b,** Individual data points for eVAF% in urine before and during treatment, and pre-surgery. Dark blue bars: patients with a pathological complete response (pCR, ypT0N0), light blue bars: patients with no residual muscle-invasive disease (ypTis/Ta/T1N0), red bars: patients

with muscle-invasive disease in the bladder (≥ypT2Nx), Mixed dark blue and red bars: patients with a pCR in the bladder and lymph node metastases (ypT0N+), missing bars: missing samples. Undetectable values are plotted as 0.0002. **c,** Percentage of patients with no detectable ctDNA in urine at latest available sample pre-surgery for patients categorized in different ways. Responders vs non-responders: P=0.39; pCR vs non-pCR: P=0.08; bladder pCR vs bladder non-pCR: P<0.01. P-values: two-sided Fisher's Exact test comparing the proportion of patients with detectable ctDNA pre-surgery in responders (ypTis/Ta/T1N0) compared to non-responders (left), patients with or without a pCR (ypT0N0; middle) and patients with or without a pathological complete response in the bladder (ypT0Nx; right). **d,** Progression-free survival for patients with and without detectable ctDNA in urine measured at latest available sample pre-surgery. Hazard ratio (HR) and P-value: Hazard ratio: 0.70; 95% confidence interval: 0.086-5.70; P=0.76 using a two-sided Log-rank (Mantel-Cox) test. *ND: not detectable, Pre-C2/3: prior to second or third treatment cycle.*



Extended Data Figure 5 | Sankey Diagrams showing the ctDNA dynamics and relation with outcome. Top: ctDNA in urine in patients in cohort 1 (*n*=24). Bottom: ctDNA in plasma in patients in cohort 1 and 2 combined (*n*=41). Red, blue and grey bars represent the number of patients with detectable, undetectable or missing ctDNA measurements, respectively, at different timepoints (Baseline, Pre-C2, Pre-C3 or Pre-surgery). In follow-up, red and blue bars represent the number of patients with and without disease progression, respectively. *Pre-C2/3: prior to second or third treatment cycle.*



Extended Data Fig. 6 | ctDNA assessment in plasma in cohort 2A and 2B. ctDNA assessment in plasma in cohort 2A and 2B.

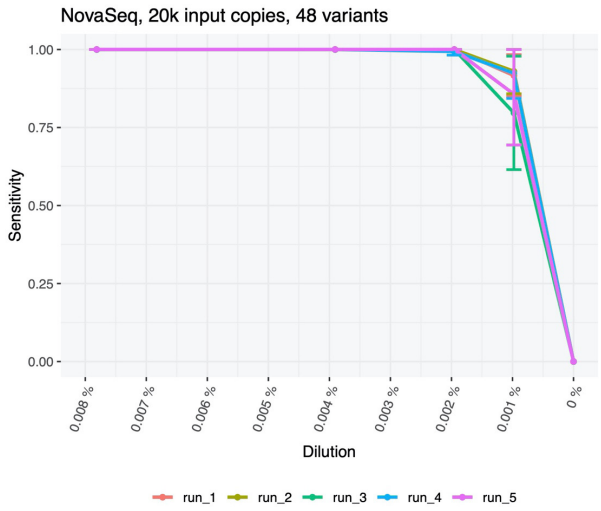
SUPPLEMENTARY DATA

Supplementary Table 4-6 and NABUCCO study protocol are available online only.

a

Tube type	Panel	Healthy donor blood tubes	Successfully tested	Detected n (%)
EDTA	HCC1395	8	8	0 (0%)
EDTA	HCC1954	8	8	0 (0%)
EDTA	SW480	4	4	0 (0%)
Streck	HCC1395	20	20	0 (0%)
Streck	HCC1954	20	20	0 (0%)
Streck	SW480	9	9	0 (0%)
Streck	5 NSCLC patients	20	20	0 (0%)

b



c

Number of variants	LoD_{95} VAF
48	0.00114%
44	0.00128%
40	0.00138%
36	0.00152%
32	0.00172%
28	0.00190%
24	0.00228%
20	0.00276%
16	0.00336%
12	0.00438%
8	0.00640%

Supplementary Figure 1 | RaDaR assay analytical validation. a, Summary of the samples and panels used to determine RaDaR assay specificity. **b**, Dilution-based analysis of RaDaR sensitivity using two breast cancer and one colon cancer cell line. Each line represents the result of an independent RaDaR run performed

in the lab with the three different cell lines, all analyzed in duplicate (*n*=2x3). Each line shows the mean and standard deviation (SD) of the sensitivity values of the three cell lines (*n*=5 separate experiments). **c**, Analysis of the impact of the number of variants on the limit of detection (LoD). Variants were randomly for all three cell lines removed before calling was performed and Probit was used to determine LoD with these sub-sampled data sets.

Supplementary Table 1 | Baseline characteristics. Baseline characteristics for patients from cohort 1, 2A and 2B. *P*-values: two-sided Chi-square tests for all parameters except *age*, for which the *P*-value is calculated using a two-sided ANOVA test. No correction for multiple hypothesis testing has been conducted.

Baseline characteristics	Cohort 1 (*n*=24)	Cohort 2A (*n*=15)	Cohort 2B (*n*=15)	Statistics
Median age, years (IQR)	66 (64-71)	70 (65-73)	68 (64-72)	*P*=0.59
Male sex, *n* (%)	18/24 (75%)	11/15 (73%)	12/15 (80%)	*P*=0.90
Clinical T-stage, *n* (%)				*P*=0.50
cT3-4aN0M0	14/24 (58%)	8/15 (53%)	7/15 (47%)	
cT1-4aN1-3M0	10/24 (42%)	7/15 (47%)	8/15 (53%)	
Cisplatin-eligibility status *n* (%)				*P*=0.45
Cisplatin-ineligible[a]	13/24 (54%)	10/15 (67%)	11/15 (73%)	
Refusal of cisplatin-based chemotherapy	11/24 (46%)	5/15 (33%)	4/15 (27%)	
WHO performance score, *n* (%)				*P*=0.13
WHO 0	21/24 (88%)	9/15 (60%)	12/15 (80%)	
WHO 1	3/24 (12%)	6/15 (40%)	3/15 (20%)	
Primary tumor localization				*P*=0.97
Lower tract	24/24 (100%)[b]	14/15 (93%)	15/15 (100%)	
Upper tract	1/24 (4%)	1/15 (7%)	0/15 (9%)	

[a] *Cisplatin-ineligible according to Galsky criteria*
[b] *1 patient in cohort 1 had combined upper and lower tract involvement*

Supplementary Table 2 | Immunotherapy-related adverse events. Table showing the frequency of all adverse events, the frequency of immunotherapy-related adverse events (i.e. with a 'possible', 'likely' or very likely' relation with ipilimumab plus nivolumab), and the frequency of specific immunotherapy-related adverse events. Data is shown separately for cohort 1, 2A and 2B and for 'any grade' events and 'grade ≥3 events'.

Adverse events	Cohort 1 (*n*=24)		Cohort 2A (Ipi 3 mg/kg + Nivo 1 mg/kg, *n*=15)		Cohort 2B (Ipi 1 mg/kg + Nivo 3 mg/kg, *n*=15)	
	All grade	Grade ≥3	All grade	Grade ≥3	All grade	Grade ≥3
Any adverse event, *n* (%)	24/24 (100%)	22/24 (92%)	15/15 (100%)	9/15 (60%)	14/15 (93%)	9/15 (60%)
Any immunotherapy-related adverse event, *n* (%)	23/24 (96%)	14/24 (58%)	15/15 (100%)	6/15 (40%)	11/15 (73%)	3/15 (20%)
Immunotherapy-related adverse events, *n* (%)						
Lipase increased	8/24 (33%)	6/24 (25%)	5/15 (33%)	2/15 (13%)	0/15 (0%)	0/15 (0%)
Diarrhea	5/24 (21%)	3/24 (13%)	4/15 (27%)	1/15 (7%)	3/15 (20%)	1/15 (7%)
Alanine aminotransferase increased	5/24 (21%)	3/24 (13%)	3/15 (20%)	0/15 (0%)	3/15 (20%)	1/15 (7%)
GGT increased	3/24 (13%)	2/24 (8%)	3/15 (20%)	1/15 (7%)	4/15 (27%)	1/15 (7%)
Colitis	3/24 (13%)	2/24 (8%)	1/15 (7%)	0/15 (0%)	1/15 (7%)	0/15 (0%)
Fatigue	6/24 (25%)	1/24 (4%)	6/15 (40%)	0/15 (0%)	3/15 (20%)	1/15 (7%)
Aspartate aminotransferase increased	4/24 (17%)	1/24 (4%)	4/15 (27%)	0/15 (0%)	2/15 (13%)	1/15 (7%)
Alkaline phosphatase increased	3/24 (13%)	1/24 (4%)	3/15 (20%)	0/15 (0%)	4/15 (27%)	1/15 (7%)
Bilirubin increased	1/24 (4%)	1/24 (4%)	0/15 (0%)	0/15 (0%)	1/15 (7%)	0/15 (0%)
Cardiomyopathy	1/24 (4%)	1/24 (4%)	0/15 (0%)	0/15 (0%)	0/15 (0%)	0/15 (0%)
Hemolysis	1/24 (4%)	1/24 (4%)	0/15 (0%)	0/15 (0%)	0/15 (0%)	0/15 (0%)
Hyperglycemia	1/24 (4%)	1/24 (4%)	0/15 (0%)	0/15 (0%)	0/15 (0%)	0/15 (0%)
Peripheral motor neuropathy	1/24 (4%)	1/24 (4%)	0/15 (0%)	0/15 (0%)	0/15 (0%)	0/15 (0%)
Toxicodermia	1/24 (4%)	1/24 (4%)	0/15 (0%)	0/15 (0%)	0/15 (0%)	0/15 (0%)
Hyperthyroidism	3/24 (13%)	0/24 (0%)	3/15 (20%)	1/15 (7%)	3/15 (20%)	0/15 (0%)
Dermatitis	0/24 (0%)	0/24 (0%)	3/15 (20%)	1/15 (7%)	1/15 (7%)	0/15 (0%)
Hyponatremia	0/24 (0%)	0/24 (0%)	1/15 (7%)	1/15 (7%)	1/15 (7%)	0/15 (0%)
Hypopituitarism	0/24 (0%)	0/24 (0%)	1/15 (7%)	1/15 (7%)	0/15 (0%)	0/15 (0%)
COVID-19 infection	0/24 (0%)	0/24 (0%)	0/15 (0%)	0/15 (0%)	1/15 (7%)	1/15 (7%)
Renal insufficiency	0/24 (0%)	0/24 (0%)	0/15 (0%)	0/15 (0%)	1/15 (7%)	1/15 (7%)
Maculo-papular rash	4/24 (17%)	0/24 (0%)	2/15 (13%)	0/15 (0%)	1/15 (7%)	0/15 (0%)
Erythema multiforme	4/24 (17%)	0/24 (0%)	0/15 (0%)	0/15 (0%)	0/15 (0%)	0/15 (0%)
Pruritis	3/24 (13%)	0/24 (0%)	4/15 (27%)	0/15 (0%)	3/15 (20%)	0/15 (0%)
Amylase increased	3/24 (13%)	0/24 (0%)	4/15 (27%)	0/15 (0%)	0/15 (0%)	0/15 (0%)
Headache	3/24 (13%)	0/24 (0%)	2/15 (13%)	0/15 (0%)	0/15 (0%)	0/15 (0%)
Endocrine disorders, unspecified	3/24 (13%)	0/24 (0%)	0/15 (0%)	0/15 (0%)	0/15 (0%)	0/15 (0%)
Hypothyroidism	2/24 (8%)	0/24 (0%)	4/15 (27%)	0/15 (0%)	1/15 (7%)	0/15 (0%)
Adrenal insufficiency	2/24 (8%)	0/24 (0%)	1/15 (7%)	0/15 (0%)	0/15 (0%)	0/15 (0%)
Dry skin	2/24 (8%)	0/24 (0%)	1/15 (7%)	0/15 (0%)	0/15 (0%)	0/15 (0%)
Nausea	2/24 (8%)	0/24 (0%)	0/15 (0%)	0/15 (0%)	1/15 (7%)	0/15 (0%)
Flu-like symptoms	2/24 (8%)	0/24 (0%)	0/15 (0%)	0/15 (0%)	0/15 (0%)	0/15 (0%)
Infusion-related reaction	1/24 (4%)	0/24 (0%)	2/15 (13%)	0/15 (0%)	1/15 (7%)	0/15 (0%)

Supplementary Table 2 | Continued.

Adverse events	Cohort 1 (*n*=24)		Cohort 2A (Ipi 3 mg/kg + Nivo 1 mg/kg, *n*=15)		Cohort 2B (Ipi 1 mg/kg + Nivo 3 mg/kg, *n*=15)	
	All grade	Grade ≥3	All grade	Grade ≥3	All grade	Grade ≥3
Conjunctivitis	1/24 (4%)	0/24 (0%)	1/15 (7%)	0/15 (0%)	0/15 (0%)	0/15 (0%)
Dry mouth	1/24 (4%)	0/24 (0%)	1/15 (7%)	0/15 (0%)	0/15 (0%)	0/15 (0%)
Alopecia	1/24 (4%)	0/24 (0%)	0/15 (0%)	0/15 (0%)	0/15 (0%)	0/15 (0%)
Arthralgia	1/24 (4%)	0/24 (0%)	0/15 (0%)	0/15 (0%)	0/15 (0%)	0/15 (0%)
Infections, unspecified	1/24 (4%)	0/24 (0%)	0/15 (0%)	0/15 (0%)	0/15 (0%)	0/15 (0%)
Mucosal infection	1/24 (4%)	0/24 (0%)	0/15 (0%)	0/15 (0%)	0/15 (0%)	0/15 (0%)
Muscle weakness in lower extremity	1/24 (4%)	0/24 (0%)	0/15 (0%)	0/15 (0%)	0/15 (0%)	0/15 (0%)
Rash	1/24 (4%)	0/24 (0%)	0/15 (0%)	0/15 (0%)	0/15 (0%)	0/15 (0%)
Skin exanthema	1/24 (4%)	0/24 (0%)	0/15 (0%)	0/15 (0%)	0/15 (0%)	0/15 (0%)
Skin induration	1/24 (4%)	0/24 (0%)	0/15 (0%)	0/15 (0%)	0/15 (0%)	0/15 (0%)
Urinary incontinence	1/24 (4%)	0/24 (0%)	0/15 (0%)	0/15 (0%)	0/15 (0%)	0/15 (0%)
Dysgeusia	0/24 (0%)	0/24 (0%)	2/15 (13%)	0/15 (0%)	1/15 (7%)	0/15 (0%)
Malaise	0/24 (0%)	0/24 (0%)	2/15 (13%)	0/15 (0%)	1/15 (7%)	0/15 (0%)
Fever	0/24 (0%)	0/24 (0%)	2/15 (13%)	0/15 (0%)	0/15 (0%)	0/15 (0%)
Back pain	0/24 (0%)	0/24 (0%)	1/15 (7%)	0/15 (0%)	1/15 (7%)	0/15 (0%)
Dry eye	0/24 (0%)	0/24 (0%)	1/15 (7%)	0/15 (0%)	1/15 (7%)	0/15 (0%)
Abdominal pain	0/24 (0%)	0/24 (0%)	1/15 (7%)	0/15 (0%)	0/15 (0%)	0/15 (0%)
Edema in extremity	0/24 (0%)	0/24 (0%)	1/15 (7%)	0/15 (0%)	0/15 (0%)	0/15 (0%)
Eosinophils increased	0/24 (0%)	0/24 (0%)	1/15 (7%)	0/15 (0%)	0/15 (0%)	0/15 (0%)
Eye disorders, unspecified	0/24 (0%)	0/24 (0%)	1/15 (7%)	0/15 (0%)	0/15 (0%)	0/15 (0%)
Gastrointestinal symptoms, unspecified	0/24 (0%)	0/24 (0%)	1/15 (7%)	0/15 (0%)	0/15 (0%)	0/15 (0%)
Hypophysitis	0/24 (0%)	0/24 (0%)	1/15 (7%)	0/15 (0%)	0/15 (0%)	0/15 (0%)
LDH increased	0/24 (0%)	0/24 (0%)	1/15 (7%)	0/15 (0%)	0/15 (0%)	0/15 (0%)
Leukocytosis	0/24 (0%)	0/24 (0%)	1/15 (7%)	0/15 (0%)	0/15 (0%)	0/15 (0%)
Muscle and joint stiffness	0/24 (0%)	0/24 (0%)	1/15 (7%)	0/15 (0%)	0/15 (0%)	0/15 (0%)
Pain, unspecified	0/24 (0%)	0/24 (0%)	1/15 (7%)	0/15 (0%)	0/15 (0%)	0/15 (0%)
Platelets increased	0/24 (0%)	0/24 (0%)	1/15 (7%)	0/15 (0%)	0/15 (0%)	0/15 (0%)
Skin disorders, unspecified	0/24 (0%)	0/24 (0%)	1/15 (7%)	0/15 (0%)	0/15 (0%)	0/15 (0%)
Skin hypopigmentation	0/24 (0%)	0/24 (0%)	1/15 (7%)	0/15 (0%)	0/15 (0%)	0/15 (0%)
Subcutaneous lesions in upper extremity	0/24 (0%)	0/24 (0%)	1/15 (7%)	0/15 (0%)	0/15 (0%)	0/15 (0%)
Thyroiditis	0/24 (0%)	0/24 (0%)	1/15 (7%)	0/15 (0%)	0/15 (0%)	0/15 (0%)
Anemia	0/24 (0%)	0/24 (0%)	0/15 (0%)	0/15 (0%)	1/15 (7%)	0/15 (0%)
Hepatitis	0/24 (0%)	0/24 (0%)	0/15 (0%)	0/15 (0%)	1/15 (7%)	0/15 (0%)
Hypercalcemia	0/24 (0%)	0/24 (0%)	0/15 (0%)	0/15 (0%)	1/15 (7%)	0/15 (0%)
Investigations, unspecified	0/24 (0%)	0/24 (0%)	0/15 (0%)	0/15 (0%)	1/15 (7%)	0/15 (0%)
Myositis	0/24 (0%)	0/24 (0%)	0/15 (0%)	0/15 (0%)	1/15 (7%)	0/15 (0%)
Pain in extremity	0/24 (0%)	0/24 (0%)	0/15 (0%)	0/15 (0%)	1/15 (7%)	0/15 (0%)
Peripheral sensory neuropathy	0/24 (0%)	0/24 (0%)	0/15 (0%)	0/15 (0%)	1/15 (7%)	0/15 (0%)
Urticaria	0/24 (0%)	0/24 (0%)	0/15 (0%)	0/15 (0%)	1/15 (7%)	0/15 (0%)

Supplementary Table 3 | Baseline characteristics for patients with and without ctDNA assessment. Baseline characteristics for patients in the primary ctDNA cohort (all from cohort 1), from the validation cohort (all from cohort 2A and 2B), and for patients for which ctDNA was not assessed (all from cohort 2A and 2B).

Baseline characteristics	Primary cohort (plasma and urine assessment, *n*=24)	Validation cohort (plasma assessment only, *n*=17)	Remaining patients (no ctDNA assessment, n=13)
Institute, *n* (%)			
Netherlands Cancer Institute	24/24 (100%)	17/17 (100%)	1/13 (7%)
University Medical Center Utrecht	0/24 (0%)	0/17 (0%)	10/13 (77%)
Radboud University Medical Center	0/24 (0%)	0/17 (0%)	2/13 (15%)
Median age, years (IQR)	66 (64-71)	68 (61-73)	68 (67-72)
Male sex, *n* (%)	18/24 (75%)	14/17 (82%)	9/13 (69%)
Clinical T-stage, *n* (%)			
cT3-4aN0M0	14/24 (58%)	12/17 (71%)	3/13 (23%)
cT1-4aN1-3M0	10/24 (42%)	5/17 (29%)	10/13 (77%)
Cisplatin-eligibility status, *n* (%)			
Cisplatin-ineligible[a]	13/24 (54%)	14/17 (82%)	7/13 (54%)
Refusal of cisplatin-based chemotherapy	11/24 (46%)	3/17 (18%)	6/13 (46%)
WHO performance score, *n* (%)			
WHO 0	21/24 (88%)	14/17 (82%)	7/13 (54%)
WHO 1	3/24 (12%)	3/17 (18%)	6/13 (46%)
Primary tumor localization			
Lower tract	24/24 (100%)[b]	17/17 (100%)	12/13 (92%)
Upper tract	1/24 (4%)	0/17 (0%)	1/13 (8%)

[a] Cisplatin-ineligible according to Galsky criteria

[b] 1 patient in cohort 1 had combined upper and lower tract involvement