

Supplementary Methods

Urine cell-free DNA extraction: Urine samples were collected in cups pre-filled with 1-2 mL of 0.5 M EDTA. Shortly following collection, cell-free DNA (cfDNA) was extracted from 22 to 90 ml of urine with Q-sepharose resin slurry (GE Healthcare, Chicago, Illinois) as described previously¹. Urine cfDNA was quantified using the Qubit dsDNA High Sensitivity Assay kit (Thermo Fisher Scientific, Waltham, Massachusetts). cfDNA quality was assessed by running on an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, California).

Germline DNA extraction: A peripheral blood sample was collected from each subject using EDTA tubes. Plasma-depleted whole blood (PDWB) was collected by centrifugation and then frozen at -80°C for isolation of germline DNA. Germline DNA was extracted from 50 to 100 ul of PDWB using the QIAmp DNA Micro Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. DNA was then quantified by the Qubit dsDNA High Sensitivity Assay to determine yield (Thermo Fischer, Waltham, Massachusetts).

Single nucleotide variant analysis from urine cfDNA: Only non-silent mutations with duplex support and with no germline support were considered when querying minimal residual disease (MRD) from urine cfDNA, as previously described¹. The maximum variant allele frequency (VAF) was selected as the metric representing urine tumor DNA by uCAPP-Seq, and correlated with MRD status in the surgical specimen. Non-silent SNVs with > 2.3% VAF¹ are represented in the **Supplementary Fig. 3** heatmap. We additionally calculated the inferred tumor mutational burden (iTMB) using our uCAPP-Seq results by applying the method described previously¹. Briefly, we utilized our TMB gene

panel, which is 387 kb in size and covers 536 genes, and applied the equation determined previously by linear regression while accounting for potential dropout in order to infer exome-wide TMB¹.

Machine learning model to predict pathologic complete response (pCR) and

survival: We implemented a random forest model for the prediction of pCR, which we validated using leave-one-out cross-validation (LOOCV). pCR was assessed by board-certified genitourinary surgical pathologists at Washington University blinded to the study results, and defined as stages pT0N0, pTaN0/N1, or pTisN0. We used the maximum VAF, iTMB, and ULP-WGS-inferred tumor fraction in urine cfDNA, which were combined together into one urine tumor DNA feature for the random forest model via multiplication followed by square-root of the product. Other features included age, gender, ethnicity, smoking status, receipt of neoadjuvant chemotherapy, and tumor invasion status (**Supplementary Fig. 4**). We used the Python scikit-learn package (v0.24.2)² to implement the random forest algorithm, with the following parameters: `n_estimators = 2000`; `criterion = gini`; `bootstrap = True`. Performance of the model after LOOCV for predicting pCR was assessed by receiver operating characteristic (ROC) area under the curve (AUC) analysis. LOOCV model predictions were also stratified by Kaplan-Meier analysis from the time of surgical resection for progression-free survival (PFS) and overall survival (OS). The model was additionally generated using an independent test set and validated in a held-out cohort (**Supplementary Fig. 7**). Furthermore, we calculated feature importance levels by assessing mean decrease in impurity³, to determine how classifications of pCR versus no pCR were affected if a particular feature was left out of the model.

Power and statistical analyses: We powered the current study assuming a substantial difference in urine tumor DNA levels between patients who achieved pCR or healthy donors, compared to patients with no pCR. Assuming a large effect size estimated by Cohen's $f = 0.5$, we accrued subjects to this study until there were at least 14 subjects per group (groups = healthy donors, bladder cancer with pCR, bladder cancer with no pCR) in order to detect a difference between healthy or pCR, and no pCR with an estimated power of 80% and significance level of 0.05 as determined by 1-way ANOVA. Patient characteristics such as age, gender, ethnicity, smoking, tumor stage, neoadjuvant chemotherapy and histology were statistically compared between groups of pCR and no pCR patients using Fisher's exact test for categorical variables and Student t test for normally distributed continuous variables (**Supplementary Table 6**). SNV-derived maximum VAFs, inferred tumor mutational burden, and CNA-inferred tumor fraction levels in urine cell-free DNA from patients with localized bladder cancer were statistically compared between groups of pCR and no pCR using the Mann-Whitney U test (**Fig. 1B-D; Supplementary Figs. 5A-C and 6A-C**). The Python scikit-learn package (v0.24.2) was used for random forest modeling with LOOCV (**Supplementary Fig. 4A**) or with separate training and validation datasets (**Supplementary Fig. 7**). ROC analysis was carried out to assess the performance of the LOOCV random forest model and the corresponding AUC was calculated for the full cohort of 74 localized bladder cancer patients (**Fig. 1E**) and for MIBC patients (**Supplementary Fig. 5D**). MRD predictions based on the LOOCV random forest model were compared to surgical ground-truth by Fisher's exact test (**Fig. 1F; Supplementary Fig. 5E**). Survival curves for PFS and OS were analyzed by the Kaplan-Meier method and statistical differences were measured by

log-rank test (**Fig. 2; Supplementary Figs. 5F-G, 6D-E and 7**). The Mantel-Haenszel method was used to estimate hazard ratios. Cox proportional hazards model (PHM) univariate and multivariate analyses were developed to assess both PFS and OS (**Supplementary Tables 7 and 8**). In addition to random forest model prediction, hematocrit, body mass index, and urine cfDNA concentration were included in the multivariate models. For OS, there were no deaths during the follow-up period among patients predicted by the random forest model to achieve pCR. Given this, the assumption of proportional hazards was not met. We performed all Kaplan-Meier and Cox regression analyses starting from the time of surgery. The reverse Kaplan-Meier method was used to calculate median follow-up time (**Supplementary Table 1**). All statistical analyses were performed using Prism 9 (GraphPad Software, San Diego, California) or SAS version 9.4 (SAS, Cary, North Carolina).

References for Supplementary Methods

1. Chauhan, P.S. *et al.* Urine tumor DNA detection of minimal residual disease in muscle-invasive bladder cancer treated with curative-intent radical cystectomy: A cohort study. *PLoS Med* **18**, e1003732 (2021).
2. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
3. Breiman L. , F.J.H., Olshen R.A, Stone C.J. . *Classification And Regression Trees*, (Routledge, New York 1984).