**Supplemental Methods**

**Optimization of workflow for urine cfDNA profiling.** To evaluate the optimal approach to urine cfDNA extraction, we compared a protocol adapted with modifications for large fluid volumes from Shekhtman et al. to commercially produced kits from Bioo (Bioo Scientific, Austin, TX) and Zymo (Zymo Research, Irvine, CA) (1). Urine was gathered from three individuals, split into three equal aliquots, and each aliquot was extracted by one of the three methods. The extracts were then analyzed by Qubit dsDNA high sensitivity assay (ThermoFisher, Waltham, MA; Supplemental Fig. S1A) and fragment analyzer (data not shown).

To evaluate optimal preservation of urine cfDNA, we gathered 200 mL of urine from three volunteers and centrifuged it at 3500rpm for 10 minutes to pellet cells. The supernatant was then split into 25mL aliquots, which were combined with EDTA to a final concentration of 0.5mM. Two aliquots were immediately extracted for each patient as described in the Methods, and the remaining four aliquots were stored at either 4 degrees Celsius or at room temperature. Aliquots for each individual were extracted at Day 1 and Day 7. All extracts were analyzed by the Qubit dsDNA high sensitivity assay (ThermoFisher, Waltham, MA; Supplemental Figs. S1B and C).

We evaluated the variant allele fraction as a function of fragment size by two means. First, urine samples from two patients with known bladder cancer were split into four aliquots and sequenced after four processing conditions (32ng input into library preparation for each condition): unfragmented DNA, enzymatically fragmented of DNA (with the Kapa Hyperplus Kit; Roche, Basel, Switzerland), size selection for long (>500bp) fragments with 0.7x SPRI beads (keeping DNA on beads; Beckman Coulter, Brea, CA) followed by enzymatic fragmentation, size selection for short (<500bp) fragments with 0.7x SPRI beads (discarding DNA on beads) (Supplemental Figs. S3A and B). Second, urine supernatants were gathered from three patients with known bladder cancer. Extracts for each patient were then sequenced per the previously described CAPP-Seq workflow with or without prior acoustic fragmentation (Covaris, Woburn, MA), with 32ng of sheared cfDNA input into library prep for each condition (2,3) (Supplemental Fig. S3C). To analyze data from both approaches, putative driver mutations were identified by next-generation sequencing and their variant allele fractions were compared across each condition.

We evaluated the optimal method for converting the broad spectrum of fragments present in urine cfDNA extracts into a range optimal for next-generation sequencing by taking six urine cfDNA extracts from different individuals, splitting them into three equal aliquots, and comparing the fractional yield by the Qubit dsDNA high sensitivity assay (ThermoFisher, Waltham, MA) before and after: enzymatic fragmentation with the KAPA frag kit (Roche, Basel, Switzerland), 0.6 – 2.0x dual-sided size selection with SPRI beads, and acoustic fragmentation (Supplemental Fig. S3D).

We evaluated the optimal strategy for maximizing genome equivalent recovery by two means. First, we used urine cfDNA extracts from three patients with known bladder cancer and processed them under four conditions: 0.93uL of 15uM adapter input with 0.8x SPRI cleanup after ligation and PCR, 1.92uL of 15uM adapter input with 0.8x SPRI cleanup, 1.92uL of 15uM adapter input with 1.0x SPRI cleanup, and enzymatic fragmentation (with the Kapa Hyperplus Kit; Roche, Basel, Switzerland) followed by 0.8x SPRI cleanup and 1.92uL of 15uM adapter input. For each processing condition, 32ng of DNA was input into library preparation and samples were pooled for hybrid capture and sequencing in the same lane. No differences were observed among the median deduplicated depth with first three conditions (all containing unfragmented DNA), but the fourth condition (containing fragmented DNA) had a significantly increased deduplicated depth relative to other samples in the lane (p<0.05; Supplemental Fig. S3E). Second, we evaluated the median deduplicated sequencing depth achieved with a full lane of samples all processed without fragmentation, inputting 60ng into library preparation, to that achieved with a full-lane of samples processed with enzymatic fragmentation, inputting 32ng into library preparation following enzymatic fragmentation (Supplemental Fig. S3F).

The *TERT* promoter region is known for its high GC content, making it difficult to include in multiplex panels (4-6). We addressed this problem by including an excess of oligonucleotide probes for this region relative to other regions, bringing its deduplicated sequencing depth to same level as the panel-wide sequencing depth (median of 1718 vs. 1627; p=ns).

**Tissue processing.**Patient slides were examined by a pathologist to identify regions of tumor and normal tissue, and DNA was extracted from corresponding Formalin Fixed, Paraffin Embedded tissue blocks using the QIAamp DNA FFPE tissue kit (Qiagen, Hilden, Germany). DNA extracts were acoustically sheared to mode of ~200bp prior to library preparation, as described previously (2). Tumor samples with at least one possible driver mutation over 5% variant allele fraction identified by sequencing were used for tumor-informed genotyping and all other analyses (49/51 samples).

**Criteria for variant calling.**SNV calling in cfDNA required passage of our adaptive variant calling threshold for read support, as previously described (2,7), with a maximum requirement of 8 supporting reads and a minimum 0.5% allele fraction requirement, as calibrated from the highest driver mutation identified in the urine cfDNA of 33 young, healthy controls (Supplemental Fig. S6). Indels causing frameshifts in tumor suppressors were called after filtering out variants immediately adjacent to homopolymers of 3 or more repeat bases, with a cutoff for single-base indels of 1.88% as determined from the highest allele fraction single-base frameshift indel identified in a tumor suppressor in 33 healthy control cfDNA samples. Complex indels (involving four or more non-repetitive bases) were called down to an allele fraction of 1%. To reduce the risk of calling germline SNVs or indels as somatic mutations in the cfDNA, all variants above 0.0001 population frequency in the gnomAD database (8) were filtered and only those in driver genes were called above 20% allele fraction.Criteria for tumor-naive detection, as described in the text, were then applied to the resulting cfDNA genotyping data, with the exclusion of variants in genes in our panel that are primarily associated with hematopoietic processes (*JAK2*, *IDH1*, *IDH2*).

 Tumor genotyping began with several steps to remove putative germline SNPs. These included filtering against cored germline tissue from FFPE blocks where available (42/51 cases) with variants above 40% allele fraction in normal tissue designated as “germline”; a requirement for enrichment of putative somatic variants by at least 20% in tumor DNA relative to germline to account for tumor contamination in cored germline samples; and filtering out variants in the gnomAD database at a population frequency > 0.0001 (8). All variants passing these filters above 5%, and all “oncogenic” SNVs above 0.5%, were used for the analyses in Figs. 1 and 2. For tumor-informed classification, we used only driver gene variants to further reduce the risk of using residual germline SNPs for diagnostic classification. Driver gene SNVs identified in tumor were searched for in the corresponding cfDNA sample and a Monte Carlo-based statistical test described previously was applied to detect significant enrichment of tumor-associated variants in urine cfDNA (2,7).

Across tumor and cfDNA samples, copy number variation was quantified using CNVKit after normalizing to 33 healthy control cfDNA samples for cfDNA CNV detection and 9 germline FFPE samples that had no evidence of tumor contamination for tumor CNV detection (9). Samples were excluded from CNV analysis if their median depth was below 500 (n=10 tumor cases). The spread of bin-level copy number ratios was determined for each sample via an interquartile range (IQR), and samples with IQR greater than 2 standard deviations from the mean of their group were excluded as technical outliers (n=2 tumor cases and 7 cfDNA cases). APOBEC mutational signature enrichment was assessed via the deconstructSigs R package (10) across all synonymous and non-synonymous variants after filtering out SNPs with the gnomAD database (population frequency > 0.0001) (8) (Supplemental Fig. S5).

**Genotyping analyses.** Tumor and cfDNA concordance was assessed in patients in the early-detection group with paired tumor tissue available and at least one cfDNA variant identified by genotyping (n=18; Fig. 1C). The cumulative distribution of mutations per patient within our panel space was analyzed in the 2017 TCGA dataset for urothelial carcinoma, using the published MuTect SNV calls and compared to the observed distribution in our study (11). Samples in our study used for this and other genotyping analyses included all tumor samples sequenced, including some for which paired cfDNA was not analyzed, and all cfDNA samples with at least one variant identified by genotyping. Genotypes from these samples also formed the basis for the data in Figs. 1B, 2A, and 2B. utDNA percentages were calculated based on tumor-informed genotyping results, when available, using the average allele fraction across all reporters. In cfDNA cases without paired tumor available, utDNA percentages were calculated based on the average allele fraction of driver mutations.

 Putative driver mutations were defined as single-nucleotide variants (SNVs) classified as “oncogenic” in the OncoKB database (12), truncating mutations in tumor suppressors, copy number loss (<1.5) of a tumor suppressor or copy number gain (>3) of an oncogene, and mutations at one of the four previously characterized hotspots in the *TERT* and *PLEKHS1* promoters (13).

**Study Design.** Our hypothesis was that utDNA would have higher sensitivity than cytology for detecting the presence of BLCA, both in a diagnostic and surveillance context. The diagnostic performance in both study groups was assessed with a case-control study design. Based on a null hypothesis that cytology and utDNA would have the same sensitivity, we targeted 34 patients to achieve a power of 90% and two-sided alpha of 0.05 to detect a difference between 34% sensitivity for cytology and 72% sensitivity of utDNA, based on previously published sensitivities for cytology and tumor-naive cell-free tumor DNA detection in the plasma (3,14). We included more BLCA cases (n=54) in the early-stage BLCA group given the anticipated difficulty of detecting the higher fraction of low grade, early-stage tumors and roughly the targeted number (n=37) of BLCA cases in the surveillance group. We additionally performed an exploratory analysis to establish the mutational concordance of utDNA with paired tumor biopsy.

**Statistical Analyses.**To generate the receiver operating characteristic curves in Fig. 3B and 4B for tumor-informed genotyping, we first performed monitoring (as previously described) of patient-specific variants identified from the tumors of patients in the diagnostic group on the urine of 33 young, healthy volunteers to calibrate a threshold for the ctDNA detection index, a previously described parameter analogous to a false positive rate (2,3). Based on this analysis, the threshold was set at a Monte Carlo p-value of < 0.00225 as detected and >0.00225 as not detected. This threshold was then applied to all other analyses across both groups. Monitoring was performed on the corresponding urine sample taken prior to cystoscopy of patients with biopsy-proven tumor (to evaluate sensitivity in Fig. 3B) or supernatant from the concurrent cytology specimen in patients who developed recurrent disease in our study interval (to evaluate sensitivity for Fig. 4B). We also performed monitoring with these same variants on the urine of 33 approximately risk-matched controls (to evaluate specificity for Fig. 3B) and on supernatant from the concurrent cytology specimen of patients with at least 9 months of negative clinical follow-up (to evaluate specificity for Fig. 4B). To generate the ROC curves in Figs. 3B and 4B for tumor-naive genotyping, we used the highest allele fraction “oncogenic” variant identified in each sample’s cfDNA by our adaptive variant caller but without the 0.5% cutoff, for both cases and controls. In calculating the sensitivity and specificity of cytology, “negative” and “atypical” results were considered clinically negative while “suspicious” and “positive” results were considered clinically positive. In calculating the sensitivity of cystoscopy for Fig 4C, we evaluated reports from cystoscopies performed at the same clinical visit in which the urine sample we analyzed was gathered and placed them into “negative,” “equivocal,” and “positive” categories based on the diagnosis of “recurrent tumor” in the bottom line summary.

Clinical differences were initially assessed between cases that did and did not recur in the surveillance group in univariate analyses using Fisher's exact test for binary variables, the Chi-square test for categorical variables, or the Mann Whitney test for continuous variables. These included the tumor-naive classification status; patient age, sex, and smoking history; tumor stage, grade, and morphology; prior surgical or intravesicular treatment; and the interval from the last prior treatment to the specimen we analyzed. Differences found to be statistically significant between groups were the tumor-naive classification status, tumor grade, and the interval from last prior treatment to the analyzed urine specimen. These three variables were then analyzed in a multivariate context using logistic regression for significance and an adjusted odds ratio was calculated for each factor (Supplemental Table S13).

1. Shekhtman EM, Anne K, Melkonyan HS, Robbins DJ, Warsof SL, Umansky SR. Optimization of transrenal DNA analysis: detection of fetal DNA in maternal urine. Clin Chem **2009**;55(4):723-9 doi 10.1373/clinchem.2008.113050.

2. Newman AM, Bratman SV, To J, Wynne JF, Eclov NC, Modlin LA*, et al.* An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage. Nat Med **2014**;20(5):548-54 doi 10.1038/nm.3519.

3. Newman AM, Lovejoy AF, Klass DM, Kurtz DM, Chabon JJ, Scherer F*, et al.* Integrated digital error suppression for improved detection of circulating tumor DNA. Nat Biotechnol **2016**;34(5):547-55 doi 10.1038/nbt.3520.

4. Springer SU, Chen CH, Rodriguez Pena MDC, Li L, Douville C, Wang Y*, et al.* Non-invasive detection of urothelial cancer through the analysis of driver gene mutations and aneuploidy. Elife **2018**;7 doi 10.7554/eLife.32143.

5. Ward DG, Baxter L, Gordon NS, Ott S, Savage RS, Beggs AD*, et al.* Multiplex PCR and Next Generation Sequencing for the Non-Invasive Detection of Bladder Cancer. PLoS One **2016**;11(2):e0149756 doi 10.1371/journal.pone.0149756.

6. Patel KM, van der Vos KE, Smith CG, Mouliere F, Tsui D, Morris J*, et al.* Association Of Plasma And Urinary Mutant DNA With Clinical Outcomes In Muscle Invasive Bladder Cancer. Sci Rep **2017**;7(1):5554 doi 10.1038/s41598-017-05623-3.

7. Newman AM, Lovejoy AF, Klass DM, Kurtz DM, Chabon JJ, Scherer F*, et al.* Integrated digital error suppression for improved detection of circulating tumor DNA. Nature Biotechnology **2016**.

8. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T*, et al.* Analysis of protein-coding genetic variation in 60,706 humans. Nature **2016**;536(7616):285-91 doi 10.1038/nature19057.

9. Talevich E, Shain AH, Botton T, Bastian BC. CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing. PLoS Comput Biol **2016**;12(4):e1004873 doi 10.1371/journal.pcbi.1004873.

10. Rosenthal R, McGranahan N, Herrero J, Taylor BS, Swanton C. DeconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. Genome Biol **2016**;17:31 doi 10.1186/s13059-016-0893-4.

11. Robertson AG, Kim J, Al-Ahmadie H, Bellmunt J, Guo G, Cherniack AD*, et al.* Comprehensive Molecular Characterization of Muscle-Invasive Bladder Cancer. Cell **2017**;171(3):540-56 e25 doi 10.1016/j.cell.2017.09.007.

12. Chakravarty D, Gao J, Phillips SM, Kundra R, Zhang H, Wang J*, et al.* OncoKB: A Precision Oncology Knowledge Base. JCO Precis Oncol **2017**;2017 doi 10.1200/PO.17.00011.

13. Weinhold N, Jacobsen A, Schultz N, Sander C, Lee W. Genome-wide analysis of noncoding regulatory mutations in cancer. Nat Genet **2014**;46(11):1160-5 doi 10.1038/ng.3101.

14. Lotan Y, Roehrborn CG. Sensitivity and specificity of commonly available bladder tumor markers versus cytology: results of a comprehensive literature review and meta-analyses. Urology **2003**;61(1):109-18; discussion 18.