Molecular Therapy
Methods & Clinical Development
Original Article

Unique somatic variants in DNA from urine exosomes of individuals with bladder cancer

Xunian Zhou,[1,10] Paul Kurywchak,[1,10] Kerri Wolf-Dennen,[1] Sara P.Y. Che,[1] Dinanath Sulakhe,[2] Mark D'Souza,[2] Bingqing Xie,[2] Natalia Maltsev,[2] T. Conrad Gilliam,[2] Chia-Chin Wu,[3] Kathleen M. McAndrews,[1] Valerie S. LeBleu,[1,4] David J. McConkey,[5] Olga V. Volpert,[1] Shanna M. Pretzsch,[6] Bogdan A. Czerniak,[7] Colin P. Dinney,[6] and Raghu Kalluri[1,8,9]

[1]Department of Cancer Biology, University of Texas MD Anderson Cancer Center, Houston, TX, USA; [2]Department of Human Genetics, University of Chicago, Chicago, IL, USA; [3]Department of Genomic Medicine, University of Texas MD Anderson Cancer Center, Houston, TX, USA; [4]Feinberg School of Medicine, Northwestern University, Chicago, IL, USA; [5]Johns Hopkins Greenberg Bladder Cancer Institute, Baltimore, MD, USA; [6]Department of Urology, University of Texas MD Anderson Cancer Center, Houston, TX, USA; [7]Department of Pathology, University of Texas MD Anderson Cancer Center, Houston, TX, USA; [8]School of Bioengineering, Rice University, Houston, TX, USA; [9]Department of Molecular and Cellular Biology, Baylor College of Medicine, Houston, TX, USA

Bladder cancer (BC), a heterogeneous disease characterized by high recurrence rates, is diagnosed and monitored by cystoscopy. Accurate clinical staging based on biopsy remains a challenge, and additional, objective diagnostic tools are needed urgently. We used exosomal DNA (exoDNA) as an analyte to examine cancer-associated mutations and compared the diagnostic utility of exoDNA from urine and serum of individuals with BC. In contrast to urine exosomes from healthy individuals, urine exosomes from individuals with BC contained significant amounts of DNA. Whole-exome sequencing of DNA from matched urine and serum exosomes, bladder tumors, and normal tissue (peripheral blood mononuclear cells) identified exonic and 3′ UTR variants in frequently mutated genes in BC, detectable in urine exoDNA and matched tumor samples. Further analyses identified somatic variants in driver genes, unique to urine exoDNA, possibly because of the inherent intra-tumoral heterogeneity of BC, which is not fully represented in random small biopsies. Multiple variants were also found in untranslated portions of the genome, such as microRNA (miRNA)-binding regions of the *KRAS* gene. Gene network analyses revealed that exoDNA is associated with cancer, inflammation, and immunity in BC exosomes. Our findings show utility of exoDNA as an objective, non-invasive strategy to identify novel biomarkers and targets for BC.

INTRODUCTION

Bladder cancer (BC) is a widespread and costly disease with little progress in early detection because of the paucity of active screening methods. One quarter of individuals with BC are diagnosed when the disease has already progressed to muscle-invasive BC (MIBC) or metastatic stages, for which there are no effective treatments.[1] BC arises from field cancerization of the entire urothelium, resulting in molecular and cellular heterogeneity, which is represented incompletely by small specimens used for staging and diagnosis.[2] White-light cystoscopy, a fairly invasive standard procedure, is used to evaluate and monitor BC, but its reliability for detection and diagnosis of early-stage BC is modest.[3–6] Cystoscopy and cytology can be very accurate in the hands of an experienced urologist at a high-volume academic center; however, in smaller general practices, most individuals are not diagnosed until they have progressed to MIBC. Incorrect staging leads to higher recurrence rates and can be fatal because of the critical differences in clinical management of MIBC and non-muscle-invasive BC (NMIBC).[7] The high recurrence rates, typical of field cancerization, necessitate regular surveillance with cystoscopy and cytology, making BC the most expensive cancer on a lifetime-per-individual basis.[7–9] Thus, more objective and sensitive monitoring strategies for BC, with improved prognostic capacity, remain a high priority.

Urine is an attractive liquid biopsy candidate for BC because of its direct contact with the tumor. In addition, urine may overcome the limitations posed by the paucity of tissue specimens in NMIBC and better reflect the molecular heterogeneity of BC than small biopsies. There are currently six US Food and Drug Administration (FDA)-approved commercial tests for BC detection and surveillance; however, their sensitivity and specificity for recurrent disease (35%–75% and 76%–94%) do not dramatically exceed those of cystoscopy (49%–93% and 47%–96%).[9,10] Recent studies assessing exfoliated cells and cell-free DNA (cfDNA) in urine samples of individuals with BC lend further support to use of urine cfDNA as a biomarker for BC.[11–16] A retrospective study employing next-generation sequencing and digital droplet PCR of serially collected NMIBC samples revealed an association between higher levels of tumor DNA in the urine and disease progression.[13]

Received 15 January 2021; accepted 21 May 2021;
https://doi.org/10.1016/j.omtm.2021.05.010.

[10]These authors contributed equally

Correspondence: Raghu Kalluri, MD, PhD, Department of Cancer Biology, University of Texas MD Anderson Cancer Center, Houston, TX, USA.
E-mail: rkalluri@mdanderson.org

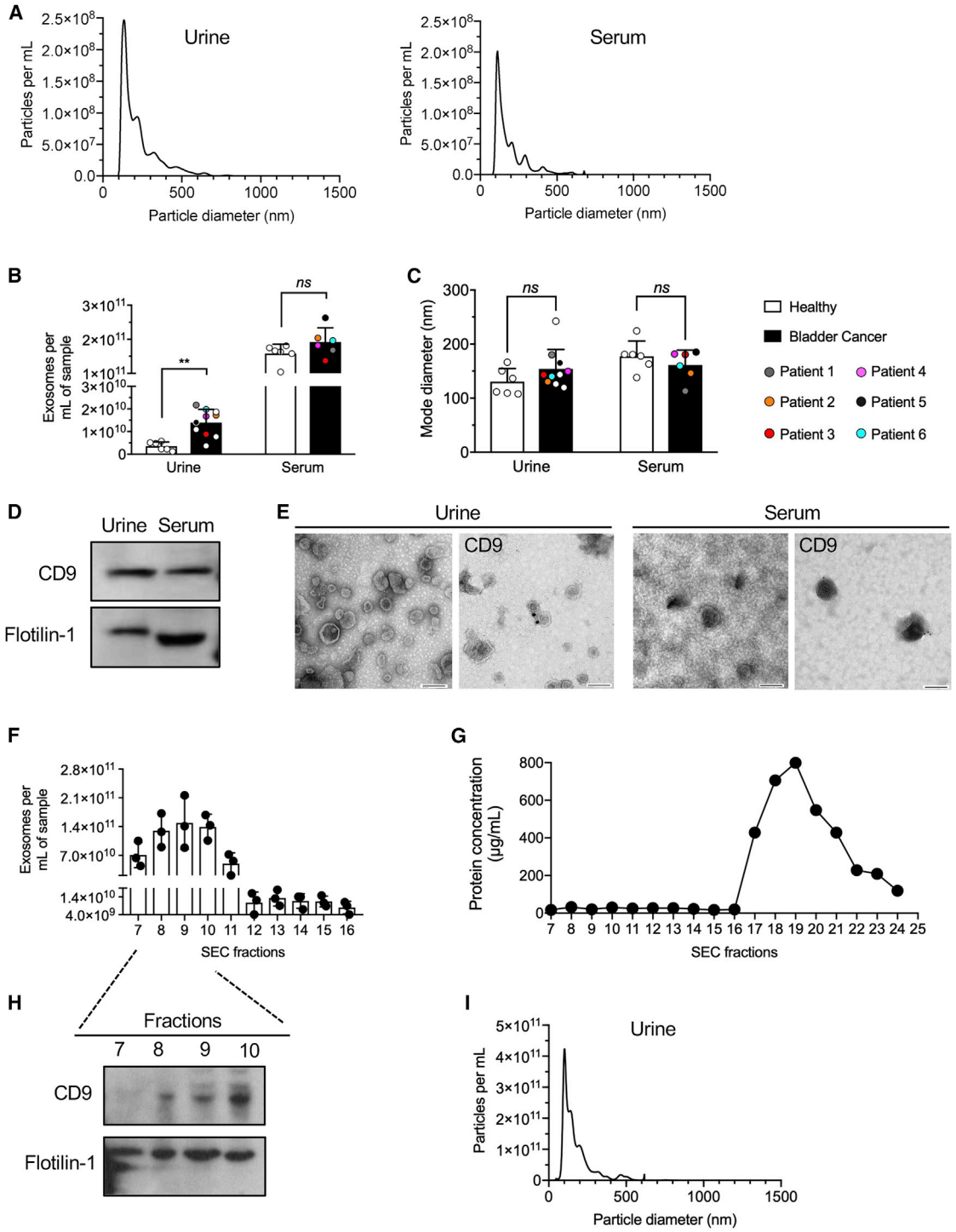


Figure 1. Characterization of exosome isolates from urine and sera of healthy samples and individuals with BC

(A) Representative graphs for nanoparticle tracking analyses of exosomes from urine and serum of a healthy sample. (B) Urine exosomes were isolated from 4 mL of healthy sample urine and 4 mL urine from individuals with BC. Exosomes number was determined by nanoparticle tracking analysis and normalized to input volume (particles per mL). (C) Exosome mode diameter as determined by nanoparticle tracking analysis. (D) Western blot analysis of exosome lysates from healthy human urine and serum probed for the exosome markers CD9 and flotillin-1. (E) TEM and CD9 immunogold staining of urine and serum exosomes from healthy samples. Scale bars, 100 nm. (F) Urine exosomes were isolated from 35 mL urine from individuals with BC. Ultracentrifuged exosome pellets were subjected to SEC, and different fractions were collected. The concentrations

(legend continued on next page)

The source, mechanism, and kinetics of cfDNA release in benign and malignant disease remain subjects of investigation and debate; however, deep sequencing studies suggested low frequency of tumor-associated mutations in cfDNA.[17–20] Deep sequencing of cfDNA remains a challenge because of low abundance and a high degree of fragmentation, with an average length of 150–200 base pairs (bp), which is even lower (10–200 bp) in urine because of a more than 100-fold increase in DNase I activity.[17,21–25] Achieving high global sequencing coverage of these low-quality and low-quantity samples necessitates development of new technologies and limits clinical use of cfDNA.[18,20]

Exosomes are 40- 150-nm vesicles formed by inward reverse budding of the endosomal membrane.[26,27] The resultant multivesicular bodies fuse with the plasma membrane and release exosomes (vesicles) into extracellular space.[28–30] Growing evidence from our laboratory and others points to circulating (serum and plasma) exosomes as a rich source of genomic DNA in individuals with cancer.[31–34] It is becoming increasingly clear that exosomal DNA (exoDNA) is valuable biomarker platform for cancer. Enrichment and analysis of exoDNA may benefit from more enhanced signals with possibly greater nucleotide fragment length than cfDNA as well as a greater capacity to detect cancer-specific DNA compared with non-cancer cell DNA.[35,36] Previous studies have shown that serum and plasma exosomes from individuals with pancreatic cancer contain fragments of genomic DNA ranging from 100 bp to 17 kb in length that collectively (in a population of exosomes) span the entire genome.[31] It has also been reported that exoDNA from individuals with metastatic pancreatic cancer showed a higher frequency of mutant *KRAS* than exosomes from individuals with local disease.[37] Moreover, exoDNA could capture DNA to detect somatic mutations and copy number variations (CNV) in BC but exhibited poor purity because commercial kits were used for exosome isolation.[38] Size exclusion chromatography (SEC), which is performed by passive gravity flow, is often used to enrich the exosome with better yield and purity, and the integrity and biological activity of the exosomes are highly preserved.[39,40] Miranda et al.[41] suggest that DNA associated with the urine exosome fractions is an exogenous contaminant that could be removed by incubation with DNase I. Bryzgunova et al.[42] also found small (less than 50 pg/mL) amounts of DNA in urine exosomes form healthy samples and individuals with prostate cancer; however, this study examined neither DNase resistance nor the diagnostic value of exosome-associated DNA. Here we evaluated the DNA content of urine exosomes isolated from individuals with BC via ultracentrifugation (UC) with or without SEC, all of which were treated with DNase I to exclude contamination with extraluminal DNA. We found that exosomes from the urine of individuals with BC contained significant amounts of DNase-resistant DNA, likely sequestered in the exosomes lumen. Interestingly, urine from individuals with BC was characterized by significantly higher exosome concentrations and higher exoDNA content (per volume and per particle) compared with healthy samples. We performed whole-exome sequencing of urine exoDNA, matched serum exoDNA, and DNA from tumors and peripheral blood mononuclear cells (PBMCs; normal control) from six individuals with BC. Comparative variant analysis of urine and serum exoDNA using customized bioinformatics pipelines revealed superior potential of urine exoDNA for capture of somatic mutations, with multiple distinctive driver variants in genes typically associated with BC, which could be validated by targeted Sanger sequencing of the tumor DNA, and a subset of mutations in the microRNA (miRNA)-binding regions.

Our results, although only based on 6 individuals, demonstrate the potential diagnostic and therapeutic value of urine exoDNA for BC. Our studies provide ways to further develop and refine such techniques and analyses to generate robust, rigorous, and reproducible assays for BC management.

RESULTS

Characterization of urine and serum exosomes

Matched urine, serum, tumor, and PBMC samples were obtained from six individuals with BC (patient 1 [P1]–[P6]) (Table S1). Histological findings, tumor grade, and tumor stage, along with representative H&E-stained tumor sections, are shown in Table S1 and Figure S1. Exosomes enriched from urine and serum by UC (Figures S2A and S2B) were characterized by NanoSight. Exosomes from urine and serum specimens from healthy samples revealed similar sizes (mode diameter, nanometers), but total exosome concentrations were higher in the sera compared with urine ($1.6 \times 10^{11} \pm 2.8 \times 10^{10}$ mL^{-1} versus $3.6 \times 10^{9} \pm 1.9 \times 10^{9}$ mL^{-1}, respectively; Figures 1A–1C). Western blot analysis identified exosome-associated proteins, including tetraspanin CD9 and flotillin-1, in urine and serum exosomes (Figure 1D; Figure S3A). Transmission electron microscopy (TEM) of exosomal preparations showed similar vesicular structures in the serum and urine (Figure 1E), and immunogold labeling confirmed CD9 on urine and serum exosomes (Figure 1E). Exosomes from urine of individuals with BC (P8–P10) were also evaluated following enrichment using SEC (Figure S2A; Table S1). Void volume (fractions 1–6) was discarded because low numbers of particles were present, and all of the remaining fractions were retained for analysis. Fractions 7–10 showed significant enrichment in exosomes, as measured by NanoSight (Figure 1F). Protein concentration measurements in each fraction showed that later fractions (17–24) contained protein contaminants (Figure 1G), presumed to be Tamm-Horshafall protein (THP) and albumin (Figure S3B), whereas fractions 7–10 were associated with the exosome markers CD9 and flotillin-1 (Figure 1H; Figure S3C). Moreover, NanoSight analysis of the pooled fractions 7–10 revealed a size distribution characteristic

of 7–16 SEC fractions were determined by nanoparticle tracking analysis. (G) Quantification of relative protein concentration in fractions 7–24 via microBCA assay. (H) Western blot of CD9 and flotillin-1 of the pooled 7–11 SEC fractions. (I) Representative graphs for nanoparticle tracking analyses of exosomes from fractions 7–10 of SEC urine exosomes. Data are expressed as mean ± SD, with the exception of (C), which is expressed as mode ± SD. Multiple t tests with Holm-Sidak post hoc analysis was performed independently for all urine and serum datasets and represented as single graphs in (B) and (C). $^{**}p < 0.01$, *ns*, not significant.

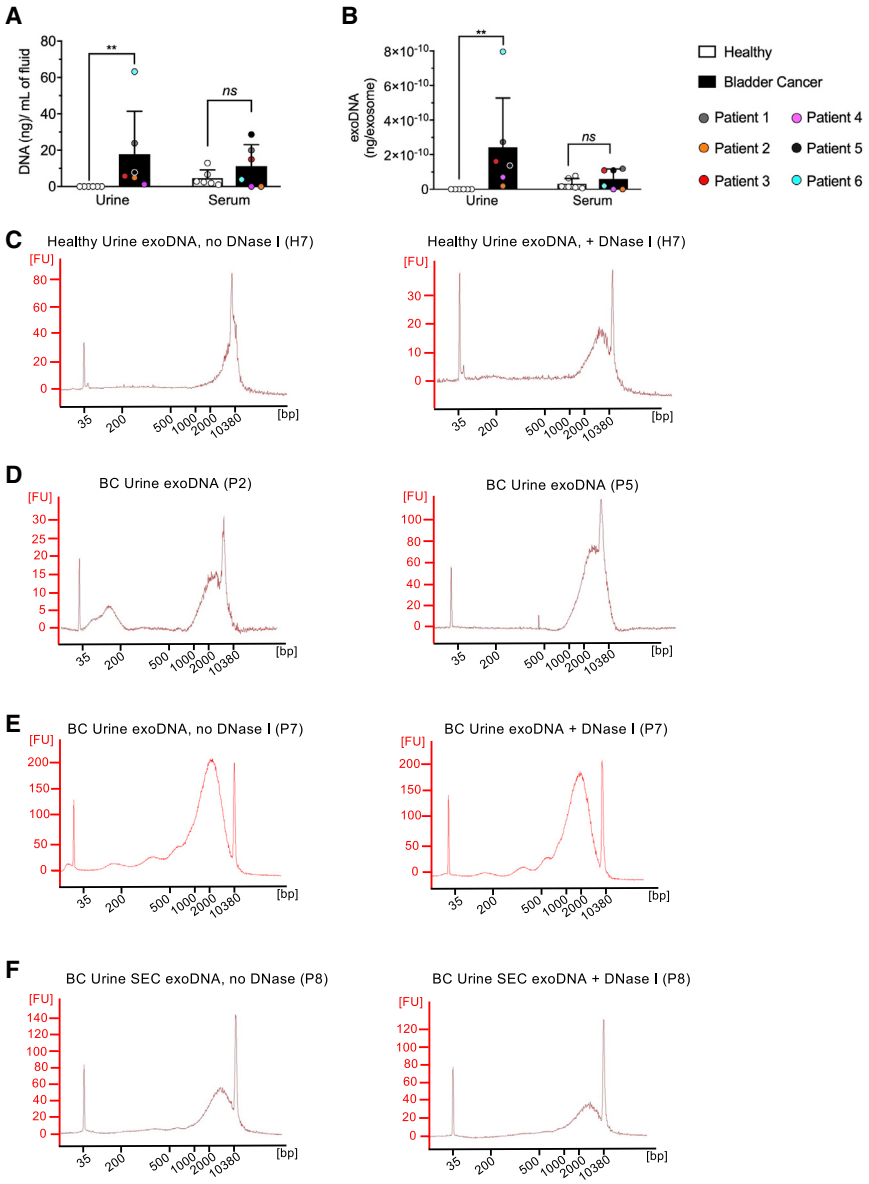


Figure 2. Comparative analysis of DNA preparations from exosomes and matched tumor tissues

DNA was isolated after DNase I treatment as described, and preparation quality was assessed using capillary electrophoresis (Bioanalyzer). (A and B) exoDNA measurements in urine and serum normalized (A) per mL biological fluid or (B) per particle. (C) DNA was isolated from urine exosomes from healthy sample 7 (H7), which were left intact (left) or pre-treated with DNase I (right) to eliminate exogenous DNA, and the resultant DNA fragments were analyzed by capillary electrophoresis. (D) DNA was isolated from tumor biopsies from P2 and P5 and analyzed by capillary electrophoresis. (E) DNA was isolated from urine exosomes from P7 without DNase I (left) or with DNase I (right) to eliminate exogenous DNA, and the resultant DNA fragments were analyzed by capillary electrophoresis. (F) DNA was isolated from urine SEC exosomes from P8, which were left intact (left) or pre-treated with DNase I (right) to eliminate exogenous DNA, and the resultant DNA fragments were analyzed by capillary electrophoresis. Mann-Whitney *U* tests was used to determine statistical significance. **$p < 0.01$, *ns*, not significant.

of exosomes (Figure 1I). Although extracellular vesicles are heterogeneous, and a mixed population of vesicles is inherently captured with all known methodologies, our results demonstrate specific enrichment of exosomes from urine and serum of healthy samples and individuals with BC.

Urine and serum exosomes contain large DNA fragments

When exosomes in healthy urine and sera samples (H1–H6) and those of individuals with BC were quantified by NanoSight and the numbers adjusted for input volumes (particles $\times$ mL^{-1}), significantly higher exosome concentrations were noted in urine from individuals with BC compared with healthy samples (Figure 1B). No significant size differences were observed between healthy samples and individuals with BC for urine or serum exosomes (Figure 1C). Total intraluminal exoDNA in the samples (following DNase I treatment to remove non-luminal DNA) showed higher exoDNA in the urine of individuals with BC compared with healthy samples (Figure 2A). When exoDNA was adjusted for the concentration of exosomes per sample (normalized per input volume or per vesicle number), higher exoDNA content was observed in urine exosomes of individuals with BC compared with healthy samples, whereas serum-derived exosomes showed no significant differences in exoDNA (Figure 2B). Capillary electrophoresis was used to assess the quality (length range and average size) of urine exoDNA fragments from healthy samples (H7 and H8) before and after DNase I treatment to eliminate exogenous (non-luminal) DNA potentially associated with exosomes (Figure 2C; Figure S3D). Comparison of urine exoDNA isolated from an untreated healthy sample (H7) (377.64 pg/μL) with the same sample treated with DNase I prior to exoDNA isolation (352.48 pg/μL) suggested that the majority of the DNA fragments are localized within the exosomal lumen and are shielded from enzymatic degradation (Figure 2C). Comparison of urine exoDNA isolated from another healthy sample (H8) with DNase I pretreatment (280.38 pg/μL) or without DNase I pre-treatment (294.85 pg/μL) showed a similar result (Figure S3D). The quality of BC urine exoDNA was also compared with DNA isolated from matched tumor tissue and serum samples (Figure 2D; Figures S3E and S3F). As expected, tumor tissue yielded high-quality DNA, with fragments ranging between 1,521–12,216 bp, with 80% of the

fragments larger than 7,000 bp (Figure S3E). Urine and serum exoDNA displayed similar DNA profiles, with urine exoDNA fragments ranging from 1,593–16,295 bp (80% in the 2,000- 6,000-bp range) and serum exoDNA ranging from 1,508–29,640 bp (80% of fragments in the 2,000- 6,000-bp range) (Figure 2D; Figures S3E and S3F). The yield of DNase I-treated BC urine exoDNA (404.38 pg/μL) showed 52.6% loss compared with non-treated exosomes (853.91 pg/μL), suggesting that around half of the DNA fragments are localized in the exosomal lumen (Figure 2E). ExoDNA from urine samples processed using SEC showed a similar fragment sizes, from 1,027–15,172 bp (80% fragments in the 2,000- 6,000-bp range) compared with exoDNA obtained from UC (Figure 2F). DNase I-treated exosomes, following SEC enrichment, yielded exoDNA (86.78 pg/μL) and 49.9% loss compared with non-treated exosomes (173.21 pg/μL), suggesting that around half of the DNA fragments were protected from degradation (Figure 2F).

Urine exoDNA is suitable for PCR amplification and Sanger sequencing

To determine whether exoDNA from urine of individuals with BC could be used to identify hotspot mutations, urine exoDNA and genomic DNA from matched tumors and PBMCs (normal tissue control) were PCR amplified and subjected to Sanger sequencing. 18 primer sets were utilized for targeting known BC-related hotspot regions in six genes (*TERT*, *FGFR3*, *PIK3CA*, *TP53*, *HRAS*, and *KDM6A*), designed based on a previous study[14] (Table S2). Successful amplification of each target region was confirmed by gel electrophoresis followed by Sanger sequencing. For each of the six targets, a positive result was observed with urine exoDNA from at least one affected individual (Figure S4A). Urine exoDNA from P1 showed the highest representation of queried genes, with positive PCR amplification for all targets. Sanger sequencing performed for a subset of targets in PBMC, tumor, and urine exoDNA resulted in base calls with clear peaks for all samples, indicative of high-quality sequencing suitable for detecting mutations. A representative comparison of base calls in one sample set in a mutational hotspot region of *TP53* is shown in Figure S4B.

ExoDNA sequencing can be affected negatively by non-uniform whole-genome amplification

The work flow for sequencing analyses is summarized in Figure S5. Whole-exome sequencing (WES) data generated by Illumina HiSeq 3000 using 76-bp paired-end reads yielded a high mean target coverage ($\geq 100\times$) for most samples (Table S3). However, for some exoDNA samples, the median coverage was poor, suggesting non-uniform coverage for a subset of targets, which could potentially skew variant calling, causing under- and over-estimation of variant frequency. This could be attributed to the flaws in the whole-genome amplification (WGA) procedure prior to library preparation, which was used for exoDNA samples to generate suitable DNA concentrations for WES, based on current technology requirements. Similar low-template WGA procedures have been known to cause amplification bias and poor coverage, especially in samples with varying quality.[43–45] This diminished coverage in some samples was also reflected in the total number of variants identified per sample (see below) as well as in concordance and contamination data. Quality control of sequencing results using the Conpair tool yielded contamination values of 0.33%–0.75% in tumor samples for the 5 patients (Tables S4A–S4F). Even low contamination levels (0.5% and above) can severely affect calling for somatic mutations (https://github.com/nygenome/Conpair) and diminish the specificity of the calling. Given excessive numbers of somatic variants generated by primary filtering against normal tissue (PBMC) sequences, we modified further analysis by evaluating minor allelic frequency (MAF) after filtering of germline variants using PBMC sequencing data (Table 1), an approach approved by the Standards and Guidelines for the Interpretation and Reporting of Sequence Variants in Cancer.[46,47]

Variant analysis revealed superior mutation capture in urine exoDNA samples

DNA contamination is a frequent problem in sequencing studies, which may lead to genotyping errors and reduced power for association testing. This includes within-species contamination in multiple-subject studies and cross-species contamination. The latter can be detected and eliminated during alignment of sequence reads. Within-species contamination is harder to detect and can compromise the quality of the analysis, especially in low-pass sequencing studies, but can also affect deep sequencing.[48] Indeed, we observed varying read depths for specific variants (Table S5). It was therefore critical for our discovery-based analysis to maximize the sensitivity and efficiency of the data screening for potential artifacts. We therefore applied multiple levels of stringency, comparison with normal tissue counterparts followed by MAF analysis, to eliminate potential artifacts because of sample contamination. It is, however, possible that observed fluctuations in coverage may reflect non-uniform representation of the host genome by a non-homogeneous population of exosomes. Tables 1 and 2 demonstrate efficient elimination of germline mutations using this approach despite poor coverage. Figure S4A and Table S5 point to substantial variability between the frequency and distribution of somatic variations between the samples obtained from the same affected individual. Some variations were identified in tumor tissue and urine samples, whereas others were specific only for urine or tumor.

We performed a systems-level bioinformatics analysis of the NGS data obtained from normal tissue, tumors, and urine exosomes of the individuals with BC. Identification and comparative analyses of somatic mutations in tumors and exoDNA were carried out on all levels of system organization, including genomic (SNPs), gene level, functional categories, molecular pathways, and organismal (between affected individuals). Serum samples were not taken into consideration when pathway analysis was performed because of the low quality of the sequencing data. Such systems-level analyses support identification of the molecular mechanisms and relevant molecular networks potentially contributing to carcinogenesis rather than individual SNPs and genes identified by routine reductionist analyses. Moreover, it allows detection of the individual-specific variations associated with the particular pathways and functional modules, providing a foundation for individualized treatment strategies. The goal of the analysis was

Table 1. Distribution of identified somatic variations

		All somatic variants			Somatic driver variants		
Tissue source	Patient ID	UTR3	UTR5	Exonic	UTR3	UTR5	Exonic
Tumor tissue	P1	2,217	365	13	103	18	2
	P2	2,488	426	13	127	23	0
	P4	2,238	439	7	95	20	0
	P5	2,589	453	13	130	22	0
	P6	2,141	419	6	88	17	0
Urine exosomes	P1	3,442	470	25	188	28	1
	P2	2,189	345	25	105	22	2
	P4	287	75	93	7	4	3
	P5	1,215	192	23	85	4	3
	P6	1,229	172	30	55	11	2
Serum exosomes	P1	118	44	62	5	5	1
	P2	1,047	175	62	72	16	1
	P4	460	141	68	19	5	4
	P5	766	142	8	57	11	1
	P6	150	48	52	6	3	2

In the discovery-based approach, somatic variants in all samples were identified by filtering against all variants found in the normal tissue (PBMC) sequence, followed by using the minor allelic frequency (MAF) test. P1–P6, patients 1–6.

(1) to identify and characterize the known BC markers and (2) to predict additional markers driving carcinogenesis in single individuals. The latter was done by predicting the somatic mutations, functional categories, and molecular networks potentially contributing to cancer progression in each individual under consideration. Comparative analyses of the markers identified in different individuals as well as of the markers identified in tumor and liquid biopsies of the same individual were also performed, as described under Materials and methods and illustrated in Figure S5A.

Comparison of tumor samples revealed all variants shared among three of five individuals with one variant common among four individuals (Table 3). *STK11*/rs10415095 was common for P1, P2, P4, and P5. *KLK10* and *IGF1R* were the most commonly mutated somatic driver genes across all tumor samples (Table 4). The mutation frequency of other potential driver genes, including *IGF2*, *AKT1*, *AKAP13*, *ELAC2*, *RASSF2*, *SYNPO2*, *CREB3L2*, and *PLEKHG2*, are shown in Table 4. Somatic mutations in genes commonly associated with BC are shown in Table 5. Analysis of exoDNA samples revealed superior capture of BC tumor mutations in DNA isolated from urine versus serum exosomes as well as identification of variants that were not detectable in the matched tumor tissues. *KRAS* variants were unique to urine samples in three of five individuals (Table 6). When comparing somatic variants of cancer-associated genes (drivers) identified in tumor samples with those found in matched exoDNA, 22%–74% of total variants were found in tumor samples, 7%–54% were found in the urine exoDNA (U), and 3%–17% in serum samples (excluding those with poor median target coverage). Additionally, samples from single individuals showed up to 21% overlap between tumor and urine exoDNA (T/U), 0%–6% were shared between tumor and serum exoDNA (T/S), and 0%–8% were common between urine and serum samples (U/S). 0%–6% were common between tumor, urine exoDNA, and serum exoDNA (T/U/S) (Figure 3A). Last, in the two individuals with the highest quality of WES data (P2 and P5), 6% of somatic driver variants were shared between tumor, urine, and serum (Figure 3A; Tables 2). Comparative analysis of all somatic variants showed a similar distribution, with 26%–70% of all variants unique to tumor samples, 11%–50% unique to urine exoDNA, and 4%–15% unique to serum exoDNA samples. Up to 20% of somatic variants were shared between tumor and urine samples, 0%–5% between tumor and serum samples, 0%–6% between urine and serum samples, and 0%–4% between tumor, urine, and serum in individual patients (Figure 3B). Our results suggest that analysis of urine and serum exosomes is potentially needed to accurately represent the full mutational spectrum of the tumor tissue.

Urine exoDNA analysis showed somatic driver variants in BC-associated genes

Using the hypothesis-based approach, we sought mutations in genes altered frequently in BC according to The Cancer Genome Atlas (TCGA) database; specifically, *RXRA*, *TP53*, and *FGFR3* (Table 5). We identified somatic variants in the 3′ untranslated regions (3′ UTR) of *FGFR3*/rs3135904 for P1 (urine only) and P2 (tumor and urine). For *TP53*, P4 showed somatic variant rs1800372 (exonic sequence) in urine exoDNA only, and somatic variant rs193920817 (exonic) was found in tumor DNA of P5. Another somatic variant, rs28934578 (exonic), was identified in the tumor of P2. We also found 3′ UTR somatic variants in tumor DNA and urine exoDNA of P2 and P6; 3′ UTR somatic variants of *RXRA*, rs1045570, and rs55645907 were identified in tumor DNA and urine exoDNA

Table 2. Frequency of identified somatic variations by DNA source

Source	Patient ID	Total somatic variations	All somatic variations (%)	Total drivers	All drivers (%)
Tumor tissue	P1	1,447	25.8763	62	21.5278
	P2	1,944	36.1944	89	32.4818
	P4	2,605	69.9517	111	72.549
	P5	2,325	52.13	108	44.6281
	P6	2,267	57.2619	93	54.0698
Urine exosomes	P1	2,778	49.6781	155	53.8194
	P2	1,453	27.0527	62	22.6277
	P4	390	10.4726	10	6.5359
	P5	788	17.6682	48	19.8347
	P6	1,100	27.7847	52	30.2325
Tumor and urine exosomes	P1	1,143	20.4399	60	20.8333
	P2	662	12.3255	33	12.0438
	P4	48	1.2889	3	1.9608
	P5	424	9.5067	17	7.0248
	P6	341	8.6133	16	9.3023
Serum exosomes	P1	207	3.7011	9	3.125
	P2	629	11.711	38	13.8686
	P4	568	15.2524	25	16.3399
	P5	433	9.7085	26	10.7438
	P6	231	5.8348	11	6.3953
Tumor and serum exosomes	P1	1	0.0179	0	0
	P2	188	3.4996	15	5.4745
	P4	84	2.2556	3	1.9608
	P5	249	5.583	15	6.1983
	P6	4	0.101	0	0
Urine and serum exosomes	P1	12	0.2146	1	0.3472
	P2	300	5.5856	22	8.0292
	P4	29	0.7787	1	0.6536
	P5	117	2.6233	13	5.3719
	P6	12	0.3031	0	0
Tumor, urine, and serum exosomes	P1	4	0.07	1	0.3472
	P2	195	3.6306	15	5.4745
	P4	0	0	0	0
	P5	124	2.7803	15	6.1983
	P6	4	0.101	0	0

Somatic variants in all samples were identified by filtering against all variants found in the normal tissue (PBMC) sequences, followed by the minor allelic frequency (MAF) test. The frequencies are calculated based on the total number of mutations found in all samples. P1–P6, patients 1–6.

(P1 and P2 and P6, respectively). We further validated select variants (rs28934578, rs193920817, rs3135904, and rs55645907) by PCR amplification-based Sanger sequencing of tumor DNA (Figure S6).

Variant analysis of BC individual panels reveals a high proportion of mutations in UTRs

Comparison of the mutational profiles in tumor samples revealed a range of 107–152 total somatic driver variants in individual samples, with 5–12 variants shared between at least two samples (Table S6). Variant analysis using sequencing data obtained using DNA isolated from tumor tissues and exoDNA, with matched PBMC DNA as the reference sequence, followed by Genome Analysis Toolkit (GATK)-based variant calling, revealed multiple somatic variants in 3′ and 5′ UTRs in all samples, including intronic, intergenic, and UTR3, which were more prevalent than variations in the exonic regions (Table S7; Figure S5). The non-coding sequence variants in 3′ and 5′ UTRs

Table 3. Somatic variants in potential driver genes common for multiple tumor samples

Gene	dbSNP ID	CGI/BC	RefSeq	P1	P2	P4	P5	P6
STK11/LKB1	rs10415095	no	UTR3	U, T	T	T	S, T	–
KLK10	rs11343599	no	UTR3	U, T	T	–	U, S, T	–
IGF2	rs58312807	yes	UTR3	–	T	T	U, S, T	–
AKT1	rs1130214	no	UTR5	U, T	S, T	–	T	–
PSCA	rs2976396	yes	UTR3	–	U, T	T	–	U, T
PTK2	rs13258775	yes	UTR5	–	T	–	T	U, T
PLEKHG2	rs251860	no	UTR3	U	T	T	T	U
ETV6	rs1051782	no	UTR3	–	T	–	T	T
RASSF2	rs2422978	no	UTR3	–	–	T	T	T
TBX3	rs1061651	yes	UTR3	–	–	T	T	T

Somatic variants in all samples were identified by filtering against all variants found in the normal tissue (PBMC) sequence, followed by using the minor allelic frequency (MAF) test. The resultant variants were ranked by their prevalence across all tumor samples (P1–P6) and by their frequency in multiple samples from a single patient (tumor, urine, and serum). All somatic variants were then analyzed for association with BC based on the Cancer Gene Index (CGI/BC). Only somatic variants present in at least three of five individuals are included in the table. The localization of the variants to exonic or untranslated (UTR) sequences is indicated (RefSeq). P1–P6, patients 1–6.

have recently been associated with high-penetrance hereditary disorders. Significant polymorphisms in the 5′ regions[49–52] and in the 3′ UTRs are linked to glioma, colon, breast, and ovarian cancer.[53–56] A recent study that provides means of a unified analytic framework to prioritize such non-coding variants revealed over 130 potentially deleterious polymorphisms in breast and ovarian carcinoma.[57] We found six of the UTR variants to localize in the miRNA binding domains of potential driver genes, suggesting that these mutations may interfere with miRNA binding to gene transcripts and, therefore, prevent post-transcriptional regulation and promote cancer progression (Table 7). Of note, we have also found a number of 3′ UTR and 5′ UTR variants in genes associated with BC (Table 5).

Gene network reconstruction through identification of driver gene mutations

Network reconstruction of driver genes with mutations shared across at least four affected individuals revealed significant interactions between cancer-associated driver genes, centered primarily around the *AKT1* pathway (Figure 4). In addition, network analysis using matched urine exoDNA and tumor samples from P1 showed significant overlap with the network generated using tumor sequencing data and a tight network centered around pathways with strong cancer relevance, which include oncogenes *AKT1/2*, *BCL2*, *KRAS*, *MDM2*, *PDGFRB*, *AXIN1*, and *IGF1R*; tumor-suppressive genes *LATS1* and *BRCA1*, and immunomodulatory genes *IL4R*, *CXCL12*, and *IL6R* (Figure 5).

DISCUSSION

In this study, we compared the urine exoDNA content from individuals with BC and compared it to relevant tumor and normal tissue (PBMC) DNA. Exosomes were treated with DNase I prior to DNA isolation and subsequent analysis; thus, our results are representative of intraluminal exoDNA rather than extraluminal cfDNA on the exosome surface, which could be co-precipitated with exosome pellets.

Table 4. Mutation frequency of potential driver genes in tumor samples

Ranking	Gene	Total variants	CGI/BC	P1	P2	P4	P5	P6
1	*KLK10*	20	no	4	6	3	7	0
2	*IGF1R*	15	yes	6	2	3	3	1
3	*IGF2*	8	yes	0	1	2	4	1
4	*AKT1*	8	no	2	1	3	2	0
5	*AKAP13*	8	no	3	0	1	3	1
6	*ELAC2*	7	no	2	2	2	1	0
7	*RASSF2*	7	no	0	1	1	3	2
8	*SYNPO2*	7	yes	1	1	1	1	3
9	*CREB3L2*	6	no	2	1	1	1	1
10	*PLEKHG2*	6	no	1	1	1	2	1

Somatic variants were determined by filtering variants found in the tumor sequences against all variants in the matched normal tissue (PBMC) sequences and additionally filtered using minor allelic frequency (MAF) analysis and somatic variants in a specific gene identified across multiple individuals counted. Genes are ranked based on the total number of variants found in a specific gene across tumor samples. The top nine variants are shown. P1–P6, patients 1–6.

Table 5. Somatic variants in genes commonly associated with BC

Gene	dbSNP ID	CGI or COSMIC	RefSeq Func	P1	P2	P4	P5	P6
RXRA	*rs55645907*	no	UTR3	–	T, U, S	T	–	T, U
	rs1045570	no	UTR3	T, U	S	–	–	–
	rs4842194	no	UTR3	–	T, U	–	–	–
	rs34109509	no	UTR3	–	T, S	–	–	–
	rs35280127	no	UTR3	–	T	–	–	–
TP53	*rs193920817*	no	exonic	–	–	–	T	–
	rs1800372	no	exonic	–	–	U	–	–
	rs28934578	no	exonic		T			
FGFR3	*rs3135904*	yes	UTR3	U	T, U	–	–	–

In the hypothesis-based approach, somatic variants in the genes commonly associated with BC (as determined from the TCGA database) were sought by filtering variants found in the sequences from tumor DNA, urine, and serum exoDNA against all variants in the normal tissue (PBMC) sequence. No MAF analysis was required. The identified variants were then ranked by prevalence across tumor samples and by prevalence in the DNA from multiple sources (tumor, serum, and urine) from a single individual. T, tumor; U, urine; S, serum. The variants validated by Sanger sequencing are shown in italics (see also Figure S5). P1–P6, patients 1–6.

Although the DNA amount is lower after DNase I treatment, previous studies have shown that DNase-treated exosome samples have higher coverage compared with non-DNase I-treated samples.[58] Based on previous work and the paucity of biomarkers for detection and monitoring, we chose BC, in which cancer exosomes could be reasonably expected to be enriched in the urine. Our work demonstrated elevated exosome content in urine of individuals with BC compared with healthy control individuals, in agreement with a study using an integrated double-ultrafiltration device conjugated to a nanochip.[59] Moreover, the intraluminal exoDNA content was significantly higher in exosomes isolated from urine of individuals with BC compared with that of healthy samples, but this finding needs to be confirmed in multiple cohorts of affected individuals. In contrast to urine exosomes of healthy samples, exosomes derived from sera of healthy samples or individuals with BC contained significant amounts of intraluminal exoDNA (normalized per vesicle). Although the intraluminal exoDNA concentration normalized to exosomes particle numbers from urine of individuals with BC appears to be higher compared with sera of individuals with BC, a direct comparison of distinct biological fluids (sera and urine) remains challenging, considering the respective limitations of biological fluid-specific protocols for enrichment, which may distinctly affect possible contaminant composition, DNA degradation, and vesicle integrity. Although the exosome concentration in sera did not vary significantly between individuals with cancer and healthy controls, based on the limited samples studied here, a trend for more exosomes in the sera of individuals with BC was observed. Notably, a recent report indicated that exosome concentrations were increased proportionally in serum and urine of individuals with BC with higher disease stages.[60]

More work is needed to carefully determine how urine exoDNA recapitulates the mutational landscape of tumors compared with DNA isolated from cells or cfDNA, which are also present in urine.[25] Our study indicates that the length of exoDNA fragments is in the same size range as genomic DNA isolated from the tumor tissue and significantly larger than cfDNA, especially when isolated from urine (1–200 bp).[17,21–25] Previous studies have demonstrated that mutational analysis of cfDNA predicts BC recurrence with 53% accuracy.[61] In our study, urine exoDNA was representative of up to 50% of all somatic variants identified. In both cases, the sensitivity and accuracy need improvement to be used for reliable clinical testing. Another important question is the potential utility of the biomarkers identified using urine exoDNA. These points are beyond the scope of this limited pilot study. However, our results encourage expanded analyses, including longitudinally collected samples from affected individuals, to determine the time point when exosome numbers and exoDNA contents deviate from control levels. Finally, controlled and defined clinical trials, with balanced groups of affected individuals and observing relevant biological variables and progression stages, are needed to establish the accuracy of biomarker testing based on urine exoDNA and intended target populations.

We employed WES over Whole Genome Sequencing (WGS) to assess the genome-wide content of the exosomes because preliminary WGS experiments indicated that a fraction of total reads represented DNA of non-human origin (data not shown), likely reflecting the presence of commensal gut bacteria[62] or contaminating vulvovaginal bacteria acquired in the process of voiding.[62,63] Because WES uses capture baits to enrich for exonic sequences, this strategy was relied upon to also enrich for human DNA. This proved to be effective, yielding a sequence mapping rate of at least 96% in five of six urine exoDNA samples. Furthermore, the sequence reads in exoDNA samples were spread across all chromosomes, extending our previous observation that exosome populations contain DNA that collectively spans the entire genome from serum[31] to urine exosomes.

Because the amount of DNA in urine exosomes of healthy samples is low, it is likely that exoDNA from urine of individuals with BC is derived primarily from tumor cells. This could therefore result in a higher signal-to-noise ratio compared with that in cfDNA or serum exoDNA. The poor representation of BC DNA in serum exosomes suggests that bladder tumor cells may shed exosomes into the urine

Table 6. Individual-specific variants identified in urine exoDNA

P1		P2		P4		P5		P6	
Gene	Variants	Gene	Variants	Gene	Variants	Gene	Variants	Gene	Variants
CTSB	3	*AHRR*	4	*CCL16*	1	*VHL*	5	*RNF213*	3
ETV6	3	*GAS7*	3	*DPH1*	1	*AKAP13*	3	*KRAS*	2
KRAS	3	*RNF213*	3	*ESR2*	1	*GPI*	2	*PCM1*	2
LDLR	3	*KRAS*	2	*GDF15*	1	*TEP1*	2	*VHL*	2
PLG	3	*CCL16*	2	*MT1B*	1	*FGF1*	1	*CDH1*	1

Somatic variants were filtered as above, and variants in individual variant sets found in urine exoDNA were ranked based on the total number of variants found in that gene in each individual. P1–P6, patient 1–6.

at a higher rate than into the circulation because, in bladder tumors, a larger surface is directly exposed to the bladder lumen compared with blood vessels, as supported by previous findings.[13,16] Considering the metastatic profile of some advanced stages of BC, cancer-associated exosomes are likely a proportionally minor population compared with other cell-derived exosomes in the serum. In contrast, urine-derived exosomes may proportionally capture more cancer-associated exosomes compared with other exosomes in the urine. Therefore, urine exosomes may be a superior source for BC-related biomarkers. Our study suggests that combined sequencing of tumor biopsy DNA and urine exoDNA can be a better representation of the genetic heterogeneity of tumors than small biopsy specimens alone. The unique subsets of mutations found specifically in urine exoDNA argue in support of this concept, but expanded analysis of a larger dataset of affected individuals is needed to validate the utility of urine exoDNA as a biomarker.

UTR variants are often excluded from WES bioinformatics pipelines, which are generally focused on coding regions, but the significance of UTR mutations for mRNA and non-coding RNA regulation in cancer is increasingly appreciated.[64–69] Indeed, despite somatic driver variants being primarily found in UTRs and intronic regions, our computational analysis of mutational profiles in the exoDNA of individuals with BC using LynxKB software tools[70] implicated multiple cancer-associated pathways, including those specific for BC. Among them were pathways supporting cancer cells themselves as well as those more attributed to the tumor microenvironment (TME; angiogenic factors, inflammatory chemokines, and their cognate receptors). We found multiple mutations in the coding regions of *TP53*. However, even genes commonly mutated in BC showed a prevalence of mutations in 3′ UTRs rather than in exonic regions. A stringent filtering procedure was used, employing normal tissue controls (PBMCs), to ensure identification of true somatic variants rather than germline mutations. Thus, the prevalence of variants in the 3′ UTRs in this particular cohort likely reflects a true pattern and requires further investigation in multiple extended cohorts of affected individuals.

The most affected cancer-driving nodes found in urine exoDNA and tumor tissue DNA across the majority of the individuals included *AKT1-3*, *BCR*, *FOXO3*, *IGF2*, *KRAS*, and *MTOR/RPTOR*, all of which affect cancer cell proliferation, survival, and metabolism.[71–79] Frequent mutations in the *SMO/WNT/FZD* module found in 3 of 6 patients suggest that a fraction of cells underwent epithelial-to-mesenchymal transition[80] or potentially activated the tumor stroma.[81,82] The significant overlap between the main driver nodes involved in cancer progression and the TME lend further support to the validity of using urine exoDNA as a non-invasive biomarker for BC and potentially other cancer types. Angiogenesis-related genes with mutations found in urine exoDNA and in matched tumor samples included *VEGFB*, *PDGFRA*, and *PDGFRB*. In addition, a significant mutational burden in the *IL4R* and *IL6R* genes could potentially augment an inflammatory microenvironment. Last, analysis of mutational profiles using the SomaMIR database identified mutations in the miRNA binding domains of multiple cancer-associated genes, including two in the 3′ UTR of the *KRAS* gene.

We demonstrated that genomic DNA can be found in exosomes from urine of individuals with BC but not in those of healthy samples. This urine exoDNA can be used to identify cancer-specific mutational profiles that partially match the profiles of parental tumor and pathway signatures characteristic for BC. The genomic information would also be harnessed to inform future implementation and design personalized medicine. For instance, a better understanding of DNA sequencing may directly affect our efforts to achieve successful medical treatment, including cancer diagnostics, individual cancer prevention, risk assessment, personalized pharmacogenomics-based therapy, and post-therapy surveillance.[83] Furthermore, our research suggests that urine exoDNA is superior to serum exoDNA for mutational analysis of BC. Finally, urine exoDNA contains subsets of mutations that have not been found in matched tumor specimens, likely because of the limited representation of the highly heterogeneous bladder tumor tissue in a small biopsy. If this is the case, then urine exoDNA is a complementary tool to examine the biology of BC. In addition to allowing serial assessment of bladder tumor genetics, urine exoDNA may also reveal additional driver mutations that could be significant for accurate prognosis and a more appropriate clinical treatment strategy. Nevertheless, substantial variability of the frequency and distribution of somatic mutations is observed even in the same individual, which indicates that a larger cohort of samples is necessary to fully realize the potential of urine exoDNA as a biomarker for BC.

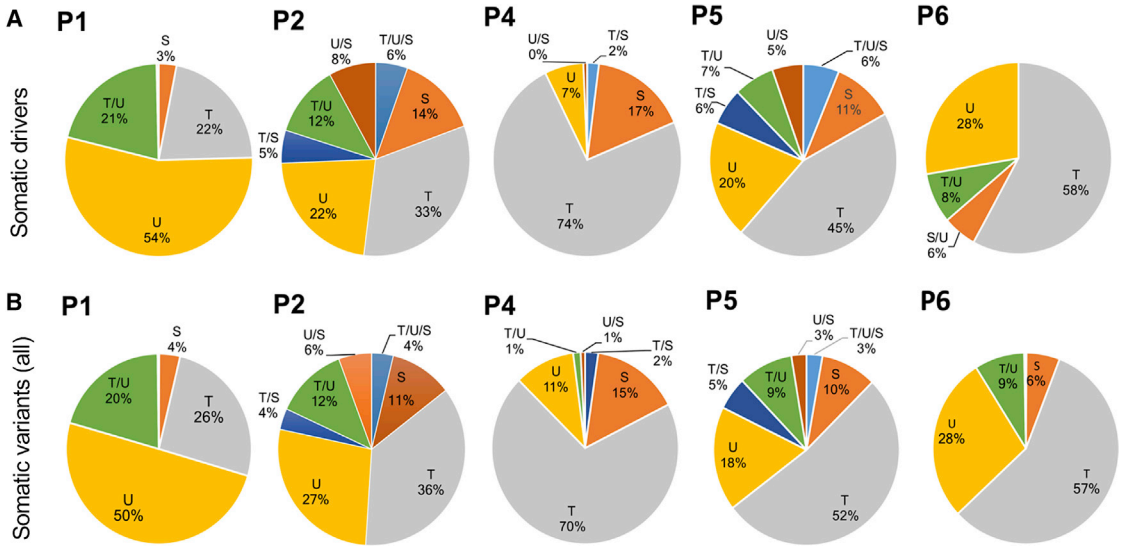


Figure 3. Overlap in somatic variants identified in the urine exoDNA, serum exoDNA, and tumors of individuals with BC

(A) Comparative analysis of somatic driver gene variants in P1, P2, and P4–P6 (tumor DNA sample (T), urine exoDNA samples (U), serum exoDNA samples (S), tumor and urine exoDNA (T/U), tumor and serum exoDNA (T/S), urine and serum samples (U/S), and tumor, urine exoDNA, and serum exoDNA (T/U/S). (B) Comparative analysis of all somatic variants in P1, P2, and P4–P6.

MATERIALS AND METHODS

Information about affected individuals and specimen collection

All samples were from the Bladder SPORE Tissue Bank (University of Texas MD Anderson Cancer Center). Collection and analysis were approved by the institutional review board (protocol PA15-0970), informed consent was provided, and the samples were properly de-identified. Exosomes used for sequencing were isolated from six affected individuals: three with NMIBC (P1–P3) and three with MIBC (P4–P6). At the time of the study, two of individuals with NMIBC (P1 and P2) had progressed to MIBC and were given neo-adjuvant therapy. Blood and urine were collected prior to transurethral biopsy, except for two blood samples (those from P2 and P5 were collected 32 and 15 days after biopsy). P3 and P4 were treated with neo-adjuvant intravesical Bacillus Calmette-Guerin (BCG), and the others received 3–5 cycles of chemotherapy. Additionally, another four BC urine-derived exosomes were included, one individual with MIBC (P7), and three with NMIBC (P8–P10). P7 was given neo-adjuvant therapy, and the other three were treated with BCG. For each biopsy, H&E staining and staging were performed by a blinded Bladder Core pathologist. Urines from healthy samples were purchased from Bioreclamation IVT (Baltimore, MD), and sera were considered institutional review board (IRB) exempt and obtained from the University of Texas MD Anderson Cancer Center Blood Bank.

Exosome isolation using UC and SEC

A schematic of the methodology employed for urine exosome enrichment is shown in Figure S1. Urine samples (4 mL for healthy samples, 4–35 mL for individuals with BC) were centrifuged for 10 min at 17,000 × *g* (first centrifugation), and supernatants were placed on ice. Tamms-Horsfall glycoprotein pellets were dissolved in DTT (200 mg mL^{-1}) 10 min at 37°C to liberate exosomes and centrifuged again for 10 min at 17,000 × *g* (second centrifugation). The supernatants from the first and second centrifugation were combined and the pellets discarded. Supernatants were passed through 0.20-μm filters (431219, Corning, NY, USA) to remove larger vesicles and debris and ultracentrifuged for 3 h at 200,000 × *g* at 4°C. The ultracentrifuged pellets were washed once with 11 mL PBS (200,000 × *g*, 3 h at 4°C) before use in the various assays or stored at −80°C before evaluation. When DNase I treated, the exosomes were treated as detailed under DNA extraction.

SEC exosomes were obtained from the ultracentrifuged pellets (obtained from 35 mL of urine from individuals with BC). Then 250 μL PBS was loaded onto temperature-equilibrated qEV size exclusion columns (552382, Izon Science, Christchurch, New Zealand) after washing with pre-filtered PBS. The qEV columns were processed according to the manufacturer's instructions. Each fraction (500 μL) was concentrated with a 30-kDa Amicon Ultra 0.5-mL centrifugal filter (UFC803024, EMD Millipore, Burlington, MA, USA) and centrifuged at 14,000 ×*g* for 10 min at room temperature (RT) to a final volume of 23 μL. The fractions were combined before DNase I treatment as detailed under DNA extraction.

For serum exosomes, frozen sera (1 mL) were thawed on ice, adjusted to 11 mL with PBS, passed through a 0.22-μm syringe filters (6782-1302, GE Healthcare, Chicago, IL, USA), and enriched by UC at 200,000 × *g* for 3 h at 4°C, followed by a wash step with 11 mL PBS once (200,000 × *g* for 3 h at 4°C). When DNase I treated, the exosomes were treated as detailed under DNA extraction.

Table 7. Mutated miRNA binding sites in driver genes

			Tumor					Urine					Serum								
			Patient ID					Patient ID					Patient ID					Total			
Gene	dbSNP ID	miRNA ID	1	2	4	5	6	1	2	4	5	6	1	2	4	5	6	T	U	S	
PAX2	rs67035383	miR-185-5p					x											1			
KRAS	rs9266	miR-181-5p							x		x								2		
KRAS	rs712	miR-877-5p	x		x													2			
CSF1R	rs3828609	miR-155-5p				x												1			
PCM1	rs1057016	miR-599		x								x						1	1		
ADAM7	rs3173956	miR-382-3p			x													1			

The top six somatic variants identified in the 3′ regions of potential driver genes were identified using the discovery-based approach and additionally annotated using the SomaMIR database. The occurrence of specific variants in the indicated samples is shown for each individual.

Nanoparticle tracking analysis

Exosome suspensions (10 μL) were diluted in cell-culture-grade H_2O and loaded, via syringe pump, on a NanoSight (LM10, Malvern Instruments, UK). Tracking was performed at 25°C with the camera level set at 13–16 for urine samples and 12–13 for serum samples to ensure readings at a similar setting. Three 30-s videos per sample were used to determine the size range and concentration of the particles.

Immunogold labeling and TEM

Serum and urine exosomes were washed by two UC cycles in PBS at 200,000 × *g* for 3 h at 4°C. Exosomes were suspended in 50 μL PBS with 2.5% electron-microscopy-grade glutaraldehyde. Immunogold labeling with anti-CD9 antibody (Table S8) and TEM were performed as described previously.[84]

Western blot analysis

UC exosome pellets from urine and serum as well as the different fractions from SEC were lysed for 1 h on ice in 100 μL urea buffer (8 M urea, 2.5% SDS, 5 μg/mL leupeptin, 1 μg mL^{-1} pepstatin, and 1 mM phenylmethylsulphonyl fluoride). Lysates were cleared by centrifugation (22,000 × *g*, 15 min at 4°C), and protein concentration was determined with a microBCA kit (23235, Thermo Fisher, Waltham, MA, USA). Each sample was measured in duplicate, and concentration was determined against a standard curve (BSA dilutions). Lysates were resolved on 4%–12% Tris-Bis gel (NP0321PK2, Thermo Fisher Scientific, MA, USA) using 1× 2-(N-morpholino)-ethanesulfonic acid (MES) running buffer (NP0002, Invitrogen, Carlsbad, CA, USA) before transferred to a polyvinylidene fluoride (PVDF) membrane (IPVH00010, Millipore, Burlington, MA, USA), as described previously.[85] Membranes were blocked for 1 h at RT with 5% non-fat dry milk in TBS-T (1× Tris-buffered saline [TBS] and 0.05% Tween-20) and incubated overnight at 4°C with primary antibodies in 2% milk in TBS-T (Table S8). Membranes were washed four times for 15 min each time with TBS-T and incubated with appropriate secondary antibodies for 1 h at RT (see Table S8 for dilutions). West-Q Pico enhanced chemiluminescence (ECL) solution (GenDEPOT W3652020, TX, USA) and Amersham Hyperfilm ECL (28906835, GE Healthcare, IL, USA) were used for protein detection. Uncropped non-adjusted images of the blots are shown in Figure S3; strips of lanes were uniformly adjusted into a gray scale.

DNA extraction

Exosome DNA was extracted from the UC and serum pellets using the QIAamp MiniElute kit (57414, QIAGEN, Hilden, Germany). Prior to lysis, when indicated, exosome samples were incubated with DNase I (25 U/mL, 9PIM610, Promega, Madison, WI, USA) for 30 min at 37°C, and the reaction was terminated by incubation with DNase stop solution (Promega) for 5 min at 65°C to remove any remaining external DNA associated with the surface of exosomes (Figure S2). For isolation of exosome DNA from pooled SEC fractions 7–10, 500 μL of each fraction was concentrated to 23 μL using Amicon Ultra 0.5-mL centrifugal filters (UFC803024, EMD Millipore, Burlington, MA, USA), followed by DNA extraction using UltraPure phenol:chloroform:isoamyl alcohol (25:24:1, v/v) (15593031, Thermo Fisher Scientific, Waltham, MA, USA) based on the manufacturer's protocol. DNA from PBMCs (recovered from 1 mL blood) and ~1 mg tumor tissue was extracted with the DNeasy Blood and Tissue Kit (69506, QIAGEN, Hilden, Germany) according to the manufacturer's instructions. DNA concentration was measured with the Qubit 3.0 high-sensitivity dsDNA kit (Q32854, Thermo Fisher Scientific, Waltham, MA, USA), and the size range was assessed using the Bioanalyzer 2100 High Sensitivity DNA Kit (5067-4626, Agilent, Santa Clara, CA, USA). All DNA samples were stored at −20°C.

Targeted PCR and Sanger sequencing

To demonstrate the applicability of exoDNA for sequencing analysis, DNA from urine exosomes, tumor tissue, and PBMCs were PCR amplified (25-μL reaction volume; for primers, see Table S2) using KAPA2G Robust Hot Start DNA polymerase (KK5522, KAPA Biosystems, Basel, Switzerland) in a T100 thermocycler (Bio-Rad, Hercules, CA, USA). PCR products were separated on 1% agarose gel (1 h at 100 V) and purified using Wizard SV Gel and the PCR Clean-Up System (A9281, Promega, Madison, WI, USA). For validation of somatic variants identified by WES, tumor DNA was amplified using Phusion high-fidelity DNA polymerase (F530, Thermo Fisher, Waltham, MA, USA). Sanger sequencing was performed with amplification primers unless indicated otherwise (Table S2), at the MD Anderson Cancer Center Advanced Technology Genomics Core. Cycling details were as follows: one cycle of 95°C for 3 min; 35 cycles of 95°C for 15 s, 60°C for 15 s, and 72°C for 30 s; and one cycle of 72°C for 5 min.[14] Some of the primer sets were

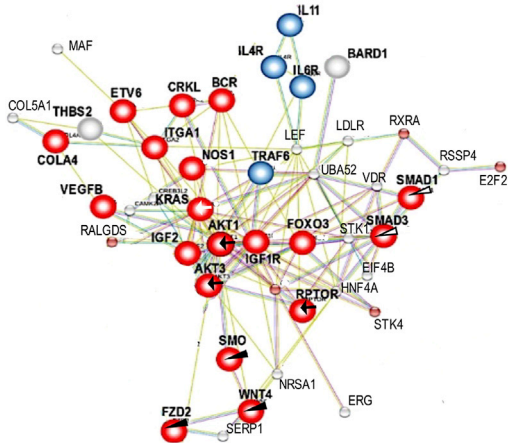


Figure 4. Common network for driver genes shared between at least four affected individuals

Network reconstruction of mutated driver genes (Pathways in Cancer, Kyoto Encyclopedia of Genes and Genomes (KEGG) shows highly significant interactions of cancer-associated driver genes, with most abundant clustering around the *AKT* pathways. Major oncogenic nodes are shown in red. Note multiple alterations in *AKT* pathways (black arrows), *KRAS* pathway (white arrow), *Wnt* pathway (black ars), and transforming growth factor β (TGF-β) pathway (white ars).

designed previously.[14] The sequences were aligned and probed for somatic variants with DNAStar SeqMan Pro v.12.3.1.

DNA library preparation and WES

Prior to library preparation, serum, and urine exoDNA were used in a WGA reaction using the REPLI-g Mini Kit (150025, QIAGEN, Hilden, Germany). Library preparation and WES were performed at the Advanced Technology Genomics Core (University of Texas MD Anderson Cancer Center). Samples (12–15 μg) were submitted for sequencing. DNA capture/library preparation were performed using the SureSelect Clinical Exome Kit V2 (5190-9501, Agilent, Santa Clara, CA, USA), followed by sequencing on an Illumina HiSeq 3000. Sequencing quality metrics (coverage, concordance, and contamination) are provided in Tables S3 and S4. The quality threshold included the following: mapping rate ($\geq$95%), duplicate mapped reads ($\leq$ 25%), mean coverage ($\geq$100×), and median coverage ($\geq$50×) of WES total reads and coverage. PBMC DNA was not available for P3, and this individual was excluded from further bioinformatics analysis. Raw sequencing metrics revealed a mean target coverage of 158–198× in PBMC samples, 138–162× in tumor samples, 31–334× in urine exosome samples, and 14–187× in serum exosome samples. Median target coverage ranged from 124–156× in PBMC samples, 103–127× in tumor samples, 1–138× in urine exosome samples, and 0–24× in serum samples. Median target coverage was likely reduced in urine and serum exosome samples because of whole-genome amplification being employed before library preparation, which is known to create bias in sequence fragment representation. In samples with poor mean and median coverage, the total number of identified variants was reduced (bold values, Table S4A), which had additional negative impact for concordance and contamination (bold values, Table S4B-E). Concordance data in samples with adequate target coverage indicate that tumor, urine, and serum samples were indeed from single affected individuals and not swapped.

Quality control, sequence alignment, and variant calling

Estimates of concordance and contamination for matched sample-normal (PBMC) pairs were performed using Conpair[86] for detection of sample swaps and cross-individual contamination in WES experiments (Tables S4A–S4E). Identification of somatic mutations using WES data from PBMC DNA, tumor tissue DNA, and urine and serum exoDNA was performed in accordance with the Standards and Guidelines for the Interpretation and Reporting of Sequence Variants in Cancer: Joint Consensus Recommendation of the Association for Molecular Pathology, American Society of Clinical Oncology, and College of American Pathologists, as described by Li et al.[46] and Chang et al.[47] Globus Genomics,[87] a Galaxy-based platform that uses Amazon Web Services for scalable computation and storage resources, was used for reference genome alignment and GATK-based best practices pipeline for variant calling.[88] The raw fastQ files for all affected individuals were aligned to a reference human genome (hg19) using Burrows-Wheeler Aligner-Maximal Exact Match (BWA-MEM). The aligned Binary Alignment Map (BAM) files were re-ordered, and read groups were added using the Picard tool.[89] Finally, the variants were called using GATK's HaplotypeCaller.[90] The resulting variants in the form of VCF files were annotated using ANNOtate VARiation (ANNOVAR).[91]

Enrichment and analysis

Two types of analysis were performed: hypothesis-based and discovery-based. In hypothesis-based analysis, a limited number of genes commonly associated with BC was analyzed for somatic variants. In this case, our analysis was based on filtering of the sequencing data against germline variants using the sequences generated using normal tissue counterparts (PBMCs). In a discovery-based approach we sought known and unknown somatic variants by unbiased sequence analysis. In this instance, because of contamination issues, which can significantly affect variant calling (in our case, it was an excessive number of variants), we added MAF analysis following filtering of the data against normal tissue samples (PBMCs) to further categorize variations as germline or somatic variations. This approach is approved by the Guidelines for the Interpretation and Reporting of Sequence Variants in Cancer (see above). MAF values less than or equal to 0.01 (1%) were classified as germline and MAF values above 0.01 as somatic. The sources used were the 1000 Genomes Project (1000 g2015aug_ALL) and Exome Sequencing Project (ESP) 6500 (ESP6500si_ALL).

ANNOVAR[91] was used to annotate the Variant Call Format (VCF) files for each sample, adding these values (when available) for each variation. Each variation may have values from both sources, either one, or neither. We used only the 1000 Genomes value if it was present and ESP6500 if not, if both values were missing the variation was not reported as germline or somatic. This additional level of stringency (MAF) was introduced to minimize errors because of poor quality of the data as determined by Conpair analysis.

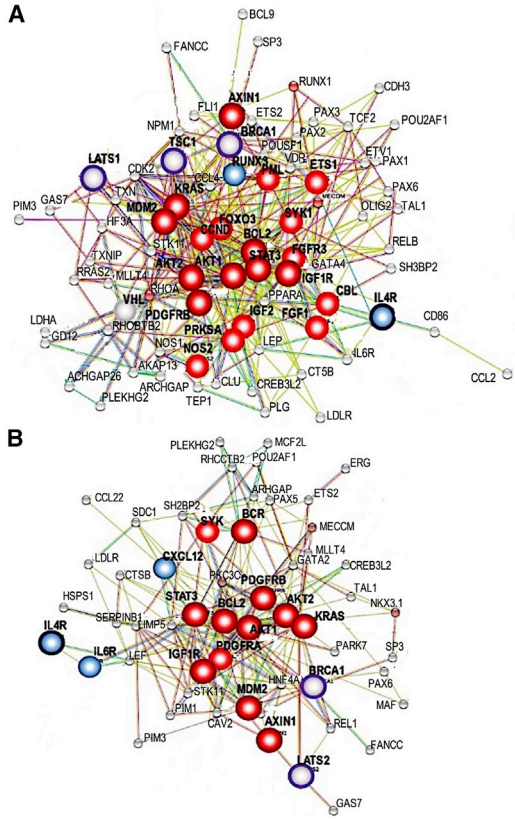


Figure 5. Networks and driver genes in urine exoDNA and matched tumor DNA of P1

(A) Networks and driver genes in urine exoDNA and matched tumor sample of P1. (B) Networks and driver genes in urine exoDNA and matched tumor sample of P1. Note tight network clustering around *KRAS* and *AKT* pathways with mutated tumor suppressor pathways (*TSC1* and *BRCA1*) and altered *TP53* pathway (*MDM2*).

Identified variations were annotated with additional information from the UniProt database[92,93] and Cancer Gene Index. Somatic driver mutations with strong or potential clinical significance were designated when the somatic mutation was identified in a gene that was (1) annotated as an oncogene or tumor suppressor gene by UniProt keywords, (2) associated with bladder carcinogenesis (Cancer Gene Index), or (3) had an entry in the Catalogue of Somatic Mutations In Cancer (COSMIC) database. An overview of this bioinformatics approach, with the blocks of customized pipelines used for analysis, is presented as a flowchart in Figure S5.

Analysis of somatic mutations with strong or potential clinical significance was performed as follows. All genes containing somatic mutations were annotated using information from the Lynx Knowledge Base (LynxKB).[70] Enrichment analysis to discover over-represented functional categories and molecular pathways in the identified gene sets was done using Lynx enrichment analysis tools and ToppGene.[70,94] Reconstruction of molecular networks and pathways harboring somatic variations of strong clinical and diagnostic significance was also performed using the Lynx suite of tools.[70] STRING 10 was used as an underlying global network for network-based gene prioritization.[95] Identification of miRNAs potentially interacting with mutated UTRs was performed using information from the SomaMIR 2.0 database and ToppGene.[94,96]

Comparative analysis of somatic mutations in single affected individuals was performed using customized analytical pipelines developed in-house specifically for this purpose (Figure S4). Somatic mutations identified in samples from the same affected individual were compared with establish variations unique to a particular sample and those shared among two or more samples belonging to one individual. Comparative analysis of somatic mutations between individuals was performed using additional customized analytical pipelines developed in-house. The results of analyses were visualized using InteractiVenn.[97]

Statistical analysis

All data are expressed as mean values ± SD. Statistical analysis was performed using GraphPad Prism v.7. Multiple t tests with Holm-Sidak correction or non-parametric unpaired, two-tailed Mann-Whitney tests were performed, and the p values are listed in the figures (**$p \leq 0.05$).

Data availability

All primary data generated in this study are available in the supplemental materials or from the corresponding author upon reasonable request.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.omtm.2021.05.010.

ACKNOWLEDGMENTS

We are grateful to Ken Dunner Jr. for performing the TEM analysis in the High Resolution Electron Microscopy Facility (institutional core grant NCI CA016672). DNA sequence analyses and library construction were performed at the Advanced Technology Genomics Core at MD Anderson Cancer Center and supported by core grant NCI CA016672. The graphical abstract was created with BioRender. This work was supported by a developmental grant from the GU SPORE (NCI CA091846) and NCI (R01CA231465 to R.K.). P.K. was supported by T32 CA186892-2. T.C.G. was funded by Mr. and Mrs. Lawrence Hilibrand, the Boler Family Foundation, and in part by NIH/NINDS NS050375. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

AUTHOR CONTRIBUTIONS

X.Z. and P.K. designed, performed, and interpreted experiments. P.K. helped prepare figures and the manuscript. K.W.-D. helped with experiments associated with exosome isolation, characterization, and DNA isolation from the clinical samples. S.P.Y.C. helped with the Sanger sequencing data analysis. D.S., M.D., B.X., N.M., T.C.G., and C.-C.W. designed the custom bioinformatics pipeline and performed the bioinformatics analysis. K.M.M. edited the manuscript. V.S.L.

designed experiments, prepared figures, and edited the manuscript. D.J.M. and O.V.V. helped interpret experiments and draft and edit the manuscript and established collaborator partnerships. S.M.P. and C.P.D. provided all samples used. B.A.C. served as the blinded Bladder Core pathologist, provided H&E photos, and evaluated staging of individuals with BC. R.K. conceptualized the study, designed the experiments, and provided intellectual input for data interpretation.

DECLARATION OF INTERESTS

MD Anderson Cancer Center and R.K. hold patents in the area of exosome biology and are licensed to Codiak Biosciences Inc. MD Anderson Cancer Center and R.K. are stock equity holders in Codiak Biosciences Inc. R.K. receives research support from Codiak Biosciences Inc. and serves as a member of the board of directors. V.S.L. served as a paid consultant for Codiak Biosciences Inc.

REFERENCES

1. Charlton, M.E., Adamo, M.P., Sun, L., and Deorah, S. (2014). Bladder cancer collaborative stage variables and their data quality, usage, and clinical implications: a review of SEER data, 2004-2010. Cancer *120* (*Suppl 23*), 3815–3825.
2. Thomsen, M.B.H., Nordentoft, I., Lamy, P., Vang, S., Reinert, L., Mapendano, C.K., Høyer, S., Ørntoft, T.F., Jensen, J.B., and Dyrskjøt, L. (2017). Comprehensive multiregional analysis of molecular heterogeneity in bladder cancer. Sci. Rep. *7*, 11702.
3. Gosnell, M.E., Polikarpov, D.M., Goldys, E.M., Zvyagin, A.V., and Gillatt, D.A. (2018). Computer-assisted cystoscopy diagnosis of bladder cancer. Urol. Oncol. *6*, 8.e9–8.e15.
4. Grossman, H.B., Gomella, L., Fradet, Y., Morales, A., Presti, J., Ritenour, C., Nseyo, U., and Droller, M.J.; PC B302/01 Study Group (2007). A phase III, multicenter comparison of hexaminolevulinate fluorescence cystoscopy and white light cystoscopy for the detection of superficial papillary lesions in patients with bladder cancer. J. Urol. *178*, 62–67.
5. Fradet, Y., Grossman, H.B., Gomella, L., Lerner, S., Cookson, M., Albala, D., and Droller, M.J.; PC B302/01 Study Group (2007). A comparison of hexaminolevulinate fluorescence cystoscopy and white light cystoscopy for the detection of carcinoma in situ in patients with bladder cancer: a phase III, multicenter study. J. Urol. *178*, 68–73, discussion 73.
6. Mowatt, G., Zhu, S., Kilonzo, M., Boachie, C., Fraser, C., Griffiths, T.R., N'Dow, J., Nabi, G., Cook, J., and Vale, L. (2010). Systematic review of the clinical effectiveness and cost-effectiveness of photodynamic diagnosis and urine biomarkers (FISH, ImmunoCyt, NMP22) and cytology for the detection and follow-up of bladder cancer. Health technology assessment (Winchester, England) *14*, 1–331, iii–iv.
7. Sanli, O., Dobruch, J., Knowles, M.A., Burger, M., Alemozaffar, M., Nielsen, M.E., and Lotan, Y. (2017). Bladder cancer. Nat. Rev. Dis. Primers *3*, 17022.
8. Herr, H.W. (2015). Role of Repeat Resection in Non-Muscle-Invasive Bladder Cancer. J. Natl. Compr. Canc. Netw. *13*, 1041–1046.
9. Mbeutcha, A., Lucca, I., Mathieu, R., Lotan, Y., and Shariat, S.F. (2016). Current Status of Urinary Biomarkers for Detection and Surveillance of Bladder Cancer. Urol. Clin. North Am. *43*, 47–62.
10. Schmitz-Dräger, C., Bonberg, N., Pesch, B., Todenhöfer, T., Sahin, S., Behrens, T., Brüning, T., and Schmitz-Dräger, B.J. (2016). Replacing cystoscopy by urine markers in the follow-up of patients with low-risk non-muscle-invasive bladder cancer?-An International Bladder Cancer Network project. Urol. Oncol. *34*, 452–459.
11. Kompier, L.C., Lurkin, I., van der Aa, M.N.M., van Rhijn, B.W.G., van der Kwast, T.H., and Zwarthoff, E.C. (2010). FGFR3, HRAS, KRAS, NRAS and PIK3CA mutations in bladder cancer and their potential as biomarkers for surveillance and therapy. PLoS ONE *5*, e13821.
12. Togneri, F.S., Ward, D.G., Foster, J.M., Devall, A.J., Wojtowicz, P., Alyas, S., Vasques, F.R., Oumie, A., James, N.D., Cheng, K.K., et al. (2016). Genomic complexity of urothelial bladder cancer revealed in urinary cfDNA. Eur. J. Hum. Genet. *24*, 1167–1174.
13. Birkenkamp-Demtröder, K., Nordentoft, I., Christensen, E., Høyer, S., Reinert, T., Vang, S., Borre, M., Agerbæk, M., Jensen, J.B., Ørntoft, T.F., and Dyrskjøt, L. (2016). Genomic Alterations in Liquid Biopsies from Patients with Bladder Cancer. Eur. Urol. *70*, 75–82.
14. Ward, D.G., Baxter, L., Gordon, N.S., Ott, S., Savage, R.S., Beggs, A.D., James, J.D., Lickiss, J., Green, S., Wallis, Y., et al. (2016). Multiplex PCR and Next Generation Sequencing for the Non-Invasive Detection of Bladder Cancer. PLoS ONE *11*, e0149756.
15. Dahmcke, C.M., Steven, K.E., Larsen, L.K., Poulsen, A.L., Abdul-Al, A., Dahl, C., and Guldberg, P. (2016). A Prospective Blinded Evaluation of Urine-DNA Testing for Detection of Urothelial Bladder Carcinoma in Patients with Gross Hematuria. Eur. Urol. *70*, 916–919.
16. Patel, K.M., van der Vos, K.E., Smith, C.G., Mouliere, F., Tsui, D., Morris, J., Chandrananda, D., Marass, F., van den Broek, D., Neal, D.E., et al. (2017). Association Of Plasma And Urinary Mutant DNA With Clinical Outcomes In Muscle Invasive Bladder Cancer. Sci. Rep. *7*, 5554.
17. Heitzer, E., Ulz, P., and Geigl, J.B. (2015). Circulating tumor DNA as a liquid biopsy for cancer. Clin. Chem. *61*, 112–123.
18. Wang, J.F., Pu, X., Zhang, X., Chen, K., Xi, Y., Wang, J., Mao, X., Zhang, J., Heymach, J.V., Antonoff, M.B., et al. (2018). Variants with a low allele frequency detected in genomic DNA affect the accuracy of mutation detection in cell-free DNA by next-generation sequencing. Cancer *124*, 1061–1069.
19. Koeppel, F., Blanchard, S., Jovelet, C., Genin, B., Marcaillou, C., Martin, E., Rouleau, E., Solary, E., Soria, J.C., André, F., and Lacroix, L. (2017). Whole exome sequencing for determination of tumor mutation load in liquid biopsy from advanced cancer patients. PLoS ONE *12*, e0188174.
20. Chang, Y., Tolani, B., Nie, X., Zhi, X., Hu, M., and He, B. (2017). Review of the clinical applications and technological advances of circulating tumor DNA in cancer monitoring. Ther. Clin. Risk Manag. *13*, 1363–1374.
21. Macanovic, M., and Lachmann, P.J. (1997). Measurement of deoxyribonuclease I (DNase) in the serum and urine of systemic lupus erythematosus (SLE)-prone NZB/NZW mice by a new radial enzyme diffusion assay. Clin. Exp. Immunol. *108*, 220–226.
22. Nadano, D., Yasuda, T., and Kishi, K. (1993). Measurement of deoxyribonuclease I activity in human tissues and body fluids by a single radial enzyme-diffusion method. Clin. Chem. *39*, 448–452.
23. Bryzgunova, O.E., and Laktionov, P.P. (2015). Extracellular Nucleic Acids in Urine: Sources, Structure, Diagnostic Potential. Acta Naturae *7*, 48–54.
24. Melkonyan, H.S., Feaver, W.J., Meyer, E., Scheinker, V., Shekhtman, E.M., Xin, Z., and Umansky, S.R. (2008). Transrenal nucleic acids: from proof of principle to clinical tests. Ann. N Y Acad. Sci. *1137*, 73–81.
25. Su, Y.H., Wang, M., Brenner, D.E., Ng, A., Melkonyan, H., Umansky, S., Syngal, S., and Block, T.M. (2004). Human urine contains small, 150 to 250 nucleotide-sized, soluble DNA derived from the circulation and may be useful in the detection of colorectal cancer. J. Mol. Diagn. *6*, 101–107.
26. Kalluri, R., and LeBleu, V.S. (2020). The biology, function, and biomedical applications of exosomes. Science *367*, eaau6977.
27. Xunian, Z., and Kalluri, R. (2020). Biology and therapeutic potential of mesenchymal stem cell-derived exosomes. Cancer Sci. *111*, 3100–3110.
28. Kowal, J., Tkach, M., and Théry, C. (2014). Biogenesis and secretion of exosomes. Curr. Opin. Cell Biol. *29*, 116–125.
29. Brinton, L.T., Sloane, H.S., Kester, M., and Kelly, K.A. (2015). Formation and role of exosomes in cancer. Cell. Mol. Life Sci. *72*, 659–671.
30. Kalluri, R. (2016). The biology and function of exosomes in cancer. J. Clin. Invest. *126*, 1208–1215.
31. Kahlert, C., Melo, S.A., Protopopov, A., Tang, J., Seth, S., Koch, M., Zhang, J., Weitz, J., Chin, L., Futreal, A., and Kalluri, R. (2014). Identification of double-stranded genomic DNA spanning all chromosomes with mutated KRAS and p53 DNA in the serum exosomes of patients with pancreatic cancer. J. Biol. Chem. *289*, 3869–3875.
32. Yang, S., Che, S.P., Kurywchak, P., Tavormina, J.L., Gansmo, L.B., Correa de Sampaio, P., Tachezy, M., Bockhorn, M., Gebauer, F., Haltom, A.R., et al. (2017). Detection of

mutant KRAS and TP53 DNA in circulating exosomes from healthy individuals and patients with pancreatic cancer. Cancer Biol. Ther. *18*, 158–165.

33. Thakur, B.K., Zhang, H., Becker, A., Matei, I., Huang, Y., Costa-Silva, B., Zheng, Y., Hoshino, A., Brazier, H., Xiang, J., et al. (2014). Double-stranded DNA in exosomes: a novel biomarker in cancer detection. Cell Res. *24*, 766–769.

34. LeBleu, V.S., and Kalluri, R. (2020). Exosomes as a Multicomponent Biomarker Platform in Cancer. Trends Cancer *6*, 767–774.

35. Kahlert, C. (2019). Liquid Biopsy: Is There an Advantage to Analyzing Circulating Exosomal DNA Compared to cfDNA or Are They the Same? Cancer Res. *79*, 2462–2465.

36. Kalluri, R., and LeBleu, V.S. (2016). Discovery of Double-Stranded Genomic DNA in Circulating Exosomes. Cold Spring Harb. Symp. Quant. Biol. *81*, 275–280.

37. Allenson, K., Castillo, J., San Lucas, F.A., Scelo, G., Kim, D.U., Bernard, V., Davis, G., Kumar, T., Katz, M., Overman, M.J., et al. (2017). High prevalence of mutant KRAS in circulating exosome-derived DNA from early-stage pancreatic cancer patients. Ann. Oncol. *28*, 741–747.

38. Lee, D.H., Yoon, H., Park, S., Kim, J.S., Ahn, Y.H., Kwon, K., Lee, D., and Kim, K.H. (2018). Urinary Exosomal and cell-free DNA Detects Somatic Mutation and Copy Number Alteration in Urothelial Carcinoma of Bladder. Sci. Rep. *8*, 14707.

39. Oeyen, E., Van Mol, K., Baggerman, G., Willems, H., Boonen, K., Rolfo, C., Pauwels, P., Jacobs, A., Schildermans, K., Cho, W.C., and Mertens, I. (2018). Ultrafiltration and size exclusion chromatography combined with asymmetrical-flow field-flow fractionation for the isolation and characterisation of extracellular vesicles from urine. J. Extracell. Vesicles *7*, 1490143.

40. Lozano-Ramos, I., Bancu, I., Oliveira-Tercero, A., Armengol, M.P., Menezes-Neto, A., Del Portillo, H.A., Lauzurica-Valdemoros, R., and Borràs, F.E. (2015). Size-exclusion chromatography-based enrichment of extracellular vesicles from urine samples. J. Extracell. Vesicles *4*, 27369.

41. Miranda, K.C., Bond, D.T., McKee, M., Skog, J., Păunescu, T.G., Da Silva, N., Brown, D., and Russo, L.M. (2010). Nucleic acids within urinary exosomes/microvesicles are potential biomarkers for renal disease. Kidney Int. *78*, 191–199.

42. Bryzgunova, O.E., Zaripov, M.M., Skvortsova, T.E., Lekchnov, E.A., Grigor'eva, A.E., Zaporozhchenko, I.A., Morozkin, E.S., Ryabchikova, E.I., Yurchenko, Y.B., Voitsitskiy, V.E., and Laktionov, P.P. (2016). Comparative Study of Extracellular Vesicles from the Urine of Healthy Individuals and Prostate Cancer Patients. PLoS ONE *11*, e0157566.

43. Stranska, J., Jancik, S., Slavkovsky, R., Holinkova, V., Rabcanova, M., Vojta, P., Hajduch, M., and Drabek, J. (2015). Whole genome amplification induced bias in the detection of KRAS-mutated cell populations during colorectal carcinoma tissue testing. Electrophoresis *36*, 937–940.

44. Sho, S., Court, C.M., Winograd, P., Lee, S., Hou, S., Graeber, T.G., Tseng, H.R., and Tomlinson, J.S. (2017). Precision oncology using a limited number of cells: optimization of whole genome amplification products for sequencing applications. BMC Cancer *17*, 457.

45. Borgström, E., Paterlini, M., Mold, J.E., Frisen, J., and Lundeberg, J. (2017). Comparison of whole genome amplification techniques for human single cell exome sequencing. PLoS ONE *12*, e0171566.

46. Li, M.M., Datto, M., Duncavage, E.J., Kulkarni, S., Lindeman, N.I., Roy, S., Tsimberidou, A.M., Vnencak-Jones, C.L., Wolff, D.J., Younes, A., and Nikiforova, M.N. (2017). Standards and Guidelines for the Interpretation and Reporting of Sequence Variants in Cancer: A Joint Consensus Recommendation of the Association for Molecular Pathology, American Society of Clinical Oncology, and College of American Pathologists. J. Mol. Diagn. *19*, 4–23.

47. Chang, V.Y., Basso, G., Sakamoto, K.M., and Nelson, S.F. (2013). Identification of somatic and germline mutations using whole exome sequencing of congenital acute lymphoblastic leukemia. BMC Cancer *13*, 55.

48. Jun, G., Flickinger, M., Hetrick, K.N., Romm, J.M., Doheny, K.F., Abecasis, G.R., Boehnke, M., and Kang, H.M. (2012). Detecting and estimating contamination of human DNA samples in sequencing and array-based genotype data. Am. J. Hum. Genet. *91*, 839–848.

49. Chen, H., Sun, B., Zhao, Y., Song, X., Fan, W., Zhou, K., Zhou, L., Mao, Y., and Lu, D. (2012). Fine mapping of a region of chromosome 11q23.3 reveals independent locus associated with risk of glioma. PLoS ONE *7*, e52864.

50. Gimm, O., Dziema, H., Brown, J., Hoang-Vu, C., Hinze, R., Dralle, H., Mulligan, L.M., and Eng, C. (2001). Over-representation of a germline variant in the gene encoding RET co-receptor GFRalpha-1 but not GFRalpha-2 or GFRalpha-3 in cases with sporadic medullary thyroid carcinoma. Oncogene *20*, 2161–2170.

51. Peng, M., Bakker, J.L., Dicioccio, R.A., Gille, J.J., Zhao, H., Odunsi, K., Sucheston, L., Jaafar, L., Mivechi, N.F., Waisfisz, Q., and Ko, L. (2013). Inactivating Mutations in GT198 in Familial and Early-Onset Breast and Ovarian Cancers. Genes Cancer *4*, 15–25.

52. Diskin, S.J., Capasso, M., Diamond, M., Oldridge, D.A., Conkrite, K., Bosse, K.R., Russell, M.R., Iolascon, A., Hakonarson, H., Devoto, M., and Maris, J.M. (2014). Rare variants in TP53 and susceptibility to neuroblastoma. J. Natl. Cancer Inst. *106*, dju047.

53. Bagheri, F., Mesrian Tanha, H., Mojtabavi Naeini, M., Ghaedi, K., and Azadeh, M. (2016). Tumor-promoting function of single nucleotide polymorphism rs1836724 (C3388T) alters multiple potential legitimate microRNA binding sites at the 3′-untranslated region of ErbB4 in breast cancer. Mol. Med. Rep. *13*, 4494–4498.

54. Garcia, A.I., Buisson, M., Damiola, F., Tessereau, C., Barjhoux, L., Verny-Pierre, C., Sornin, V., Dondon, M.G., Eon-Marchais, S., Caron, O., et al.; GENESIS investigators (2016). Mutation screening of MIR146A/B and BRCA1/2 3′-UTRs in the GENESIS study. Eur. J. Hum. Genet. *24*, 1324–1329.

55. You, H.S., Fadriquela, A., Sajo, M.E.J., Bajgai, J., Ara, J., Kim, C.S., Kim, S.K., Oh, J.R., Shim, K.Y., Lim, H.K., and Lee, K.J. (2017). Wound Healing Effect of Slightly Acidic Electrolyzed Water on Cutaneous Wounds in Hairless Mice via Immune-Redox Modulation. Biol. Pharm. Bull. *40*, 1423–1431.

56. Jiang, Y., Qin, Z., Hu, Z., Guan, X., Wang, Y., He, Y., Xue, J., Liu, X., Chen, J., Dai, J., et al. (2013). Genetic variation in a hsa-let-7 binding site in RAD52 is associated with breast cancer susceptibility. Carcinogenesis *34*, 689–693.

57. Mucaki, E.J., Caminsky, N.G., Perri, A.M., Lu, R., Laederach, A., Halvorsen, M., Knoll, J.H., and Rogan, P.K. (2016). A unified analytic framework for prioritization of non-coding variants of uncertain significance in heritable breast and ovarian cancer. BMC Med. Genomics *9*, 19.

58. Lázaro-Ibáñez, E., Lässer, C., Shelke, G.V., Crescitelli, R., Jang, S.C., Cvjetkovic, A., García-Rodríguez, A., and Lötvall, J. (2019). DNA analysis of low- and high-density fractions defines heterogeneous subpopulations of small extracellular vesicles based on their DNA cargo and topology. J. Extracell. Vesicles *8*, 1656993.

59. Liang, L.G., Sheng, Y.F., Zhou, S., Inci, F., Li, L., Demirci, U., and Wang, S. (2017). An Integrated Double-Filtration Microfluidic Device for Detection of Extracellular Vesicles from Urine for Bladder Cancer Diagnosis. Methods Mol. Biol. *1660*, 355–364.

60. Elsharkawi, F., Elsabah, M., Shabayek, M., and Khaled, H. (2019). Urine and Serum Exosomes as Novel Biomarkers in Detection of Bladder Cancer. Asian Pac. J. Cancer Prev. *20*, 2219–2224.

61. Di Meo, A., Bartlett, J., Cheng, Y., Pasic, M.D., and Yousef, G.M. (2017). Liquid biopsy: a step forward towards precision medicine in urologic malignancies. Mol. Cancer *16*, 80.

62. Lewis, D.A., Brown, R., Williams, J., White, P., Jacobson, S.K., Marchesi, J.R., and Drake, M.J. (2013). The human urinary microbiome; bacterial DNA in voided urine of asymptomatic adults. Front. Cell. Infect. Microbiol. *3*, 41.

63. Whiteside, S.A., Razvi, H., Dave, S., Reid, G., and Burton, J.P. (2015). The microbiome of the urinary tract–a role beyond infection. Nat. Rev. Urol. *12*, 81–90.

64. Devanna, P., Chen, X.S., Ho, J., Gajewski, D., Smith, S.D., Gialluisi, A., Francks, C., Fisher, S.E., Newbury, D.F., and Vernes, S.C. (2018). Next-gen sequencing identifies non-coding variation disrupting miRNA-binding sites in neurological disorders. Mol. Psychiatry *23*, 1375–1384.

65. Iuliano, R., Vismara, M.F.M., Dattilo, V., Trapasso, F., Baudi, F., and Perrotti, N. (2013). The role of microRNAs in cancer susceptibility. BioMed Res. Int. *2013*, 591931.

66. Ryan, B.M., Robles, A.I., and Harris, C.C. (2010). Genetic variation in microRNA networks: the implications for cancer research. Nat. Rev. Cancer *10*, 389–402.

67. Melton, C., Reuter, J.A., Spacek, D.V., and Snyder, M. (2015). Recurrent somatic mutations in regulatory regions of human cancer genomes. Nat. Genet. *47*, 710–716.

68. Poulos, R.C., Sloane, M.A., Hesson, L.B., and Wong, J.W.H. (2015). The search for cis-regulatory driver mutations in cancer genomes. Oncotarget *6*, 32509–32525.

69. Diederichs, S., Bartsch, L., Berkmann, J.C., Fröse, K., Heitmann, J., Hoppe, C., Iggena, D., Jazmati, D., Karschnia, P., Linsenmeier, M., et al. (2016). The dark matter of the cancer genome: aberrations in regulatory elements, untranslated regions, splice sites, non-coding RNA and synonymous mutations. EMBO Mol. Med. *8*, 442–457.

70. Sulakhe, D., Xie, B., Taylor, A., D'Souza, M., Balasubramanian, S., Hashemifar, S., White, S., Dave, U.J., Agam, G., Xu, J., et al. (2016). Lynx: a knowledge base and an analytical workbench for integrative medicine. Nucleic Acids Res. *44* (D1), D882–D887.

71. Sathe, A., and Nawroth, R. (2018). Targeting the PI3K/AKT/mTOR Pathway in Bladder Cancer. Methods Mol. Biol. *1655*, 335–350.

72. Roy, S., Pradhan, D., Ernst, W.L., Mercurio, S., Najjar, Y., Parikh, R., Parwani, A.V., Pai, R.K., Dhir, R., and Nikiforova, M.N. (2017). Next-generation sequencing-based molecular characterization of primary urinary bladder adenocarcinoma. Mod. Pathol. *30*, 1133–1143.

73. Hedegaard, J., Lamy, P., Nordentoft, I., Algaba, F., Høyer, S., Ulhøi, B.P., Vang, S., Reinert, T., Hermann, G.G., Mogensen, K., et al. (2016). Comprehensive Transcriptional Analysis of Early-Stage Urothelial Carcinoma. Cancer Cell *30*, 27–42.

74. Sellar, R., and Losman, J.A. (2017). Targeting Aberrant Signaling in Myeloid Malignancies: Promise Versus Reality. Hematol. Oncol. Clin. North Am. *31*, 565–576.

75. Wang, Y., Kang, X.L., Zeng, F.C., Xu, C.J., Zhou, J.Q., and Luo, D.N. (2017). Correlations of Foxo3 and Foxo4 expressions with clinicopathological features and prognosis of bladder cancer. Pathol. Res. Pract. *213*, 766–772.

76. Byun, H.M., Wong, H.L., Birnstein, E.A., Wolff, E.M., Liang, G., and Yang, A.S. (2007). Examination of IGF2 and H19 loss of imprinting in bladder cancer. Cancer Res. *67*, 10753–10758.

77. Al Hussain, T.O., and Akhtar, M. (2013). Molecular basis of urinary bladder cancer. Adv. Anat. Pathol. *20*, 53–60.

78. Knowles, M.A., Platt, F.M., Ross, R.L., and Hurst, C.D. (2009). Phosphatidylinositol 3-kinase (PI3K) pathway activation in bladder cancer. Cancer Metastasis Rev. *28*, 305–316.

79. Garcia, J.A., and Danielpour, D. (2008). Mammalian target of rapamycin inhibition as a therapeutic strategy in the management of urologic malignancies. Mol. Cancer Ther. *7*, 1347–1354.

80. Mao, X.W., Xiao, J.Q., Xu, G., Li, Z.Y., Wu, H.F., Li, Y., Zheng, Y.C., and Zhang, N. (2017). CUL4B promotes bladder cancer metastasis and induces epithelial-to-mesenchymal transition by activating the Wnt/β-catenin signaling pathway. Oncotarget *8*, 77241–77253.

81. Castellone, M.D., and Laukkanen, M.O. (2017). TGF-beta1, WNT, and SHH signaling in tumor progression and in fibrotic diseases. Front. Biosci. (Schol. Ed.) *9*, 31–45.

82. Macheda, M.L., and Stacker, S.A. (2008). Importance of Wnt signaling in the tumor stroma microenvironment. Curr. Cancer Drug Targets *8*, 454–465.

83. Hou, Y.C., Yu, H.C., Martin, R., Cirulli, E.T., Schenker-Ahmed, N.M., Hicks, M., Cohen, I.V., Jönsson, T.J., Heister, R., Napier, L., et al. (2020). Precision medicine integrating whole-genome sequencing, comprehensive metabolomics, and advanced imaging. Proc. Natl. Acad. Sci. USA *117*, 3053–3062.

84. Kamerkar, S., LeBleu, V.S., Sugimoto, H., Yang, S., Ruivo, C.F., Melo, S.A., Lee, J.J., and Kalluri, R. (2017). Exosomes facilitate therapeutic targeting of oncogenic KRAS in pancreatic cancer. Nature *546*, 498–503.

85. Zhou, X., Xiao, F., Sugimoto, H., Li, B., McAndrews, K.M., and Kalluri, R. (2021). Acute kidney injury instigates malignant renal cell carcinoma via CXCR2 in mice with inactivated Trp53 and Pten in proximal tubular kidney epithelial cells. Cancer Res. *81*, 2690–2702.

86. Bergmann, E.A., Chen, B.J., Arora, K., Vacic, V., and Zody, M.C. (2016). Conpair: concordance and contamination estimator for matched tumor-normal pairs. Bioinformatics *32*, 3196–3198.

87. Madduri, R.K., Sulakhe, D., Lacinski, L., Liu, B., Rodriguez, A., Chard, K., Dave, U.J., and Foster, I.T. (2014). Experiences Building Globus Genomics: A Next-Generation Sequencing Analysis Service using Galaxy, Globus, and Amazon Web Services. Concurr. Comput. *26*, 2266–2279.

88. Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., et al. (2013). From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr. Protoc. Bioinformatics *43*, 11.10.1–11.10.33.

89. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

90. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 1297–1303.

91. Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. *38*, e164.

92. Pundir, S., Martin, M.J., and O'Donovan, C. (2017). UniProt Protein Knowledgebase. Methods Mol. Biol. *1558*, 41–55.

93. The UniProt Consortium (2017). UniProt: the universal protein knowledgebase. Nucleic Acids Res. *45* (D1), D158–D169.

94. Chen, J., Bardes, E.E., Aronow, B.J., and Jegga, A.G. (2009). ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. Nucleic Acids Res. *37*, W305-11.

95. Szklarczyk, D., Morris, J.H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., Santos, A., Doncheva, N.T., Roth, A., Bork, P., et al. (2017). The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. Nucleic Acids Res. *45* (D1), D362–D368.

96. Bhattacharya, A., and Cui, Y. (2016). SomamiR 2.0: a database of cancer somatic mutations altering microRNA-ceRNA interactions. Nucleic Acids Res. *44* (D1), D1005–D1010.

97. Heberle, H., Meirelles, G.V., da Silva, F.R., Telles, G.P., and Minghim, R. (2015). InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. BMC Bioinformatics *16*, 169.

OMTM, Volume *22*

Supplemental information

Unique somatic variants in DNA from urine exosomes of individuals with bladder cancer

Xunian Zhou, Paul Kurywchak, Kerri Wolf-Dennen, Sara P.Y. Che, Dinanath Sulakhe, Mark D'Souza, Bingqing Xie, Natalia Maltsev, T. Conrad Gilliam, Chia-Chin Wu, Kathleen M. McAndrews, Valerie S. LeBleu, David J. McConkey, Olga V. Volpert, Shanna M. Pretzsch, Bogdan A. Czerniak, Colin P. Dinney, and Raghu Kalluri

Supplementary Material

Unique somatic variants in the DNA from urine exosomes of bladder cancer patients

Xunian Zhou[1,‡], Paul Kurywchak[1,‡], Kerri Wolf-Dennen[1], Sara P.Y. Che[1], Dinanath Sulakhe[2], Mark D'Souza[2], Bingqing Xie[2], Natalia Maltsev[2], T. Conrad Gilliam[2], Chia-Chin Wu[3], Kathleen M. McAndrews[1], Valerie S. LeBleu[1,4], David J McConkey[5], Olga Volpert[1], Shanna Pretzsch[6], Bogdan A Czerniak[7], Colin Dinney[6], Raghu Kalluri[1,8,9,*]

[1]Department of Cancer Biology, University of Texas MD Anderson Cancer Center, Houston, TX

[2]Department of Human Genetics, University of Chicago, Chicago, IL

[3]Department of Genomic Medicine, University of Texas MD Anderson Cancer Center, Houston, TX

[4]Feinberg School of Medicine, Northwestern University, Chicago, IL

[5]Johns Hopkins Greenberg Bladder Cancer Institute, Baltimore, MD

[6]Department of Urology, University of Texas MD Anderson Cancer Center, Houston, TX

[7]Department of Pathology, University of Texas MD Anderson Cancer Center, Houston, TX

[8]School of Bioengineering, Rice University, Houston, TX

[9]Department of Molecular and Cellular Biology, Baylor College of Medicine, Houston, TX

[‡]These authors contributed equally

*Corresponding author:

Raghu Kalluri, MD, PhD

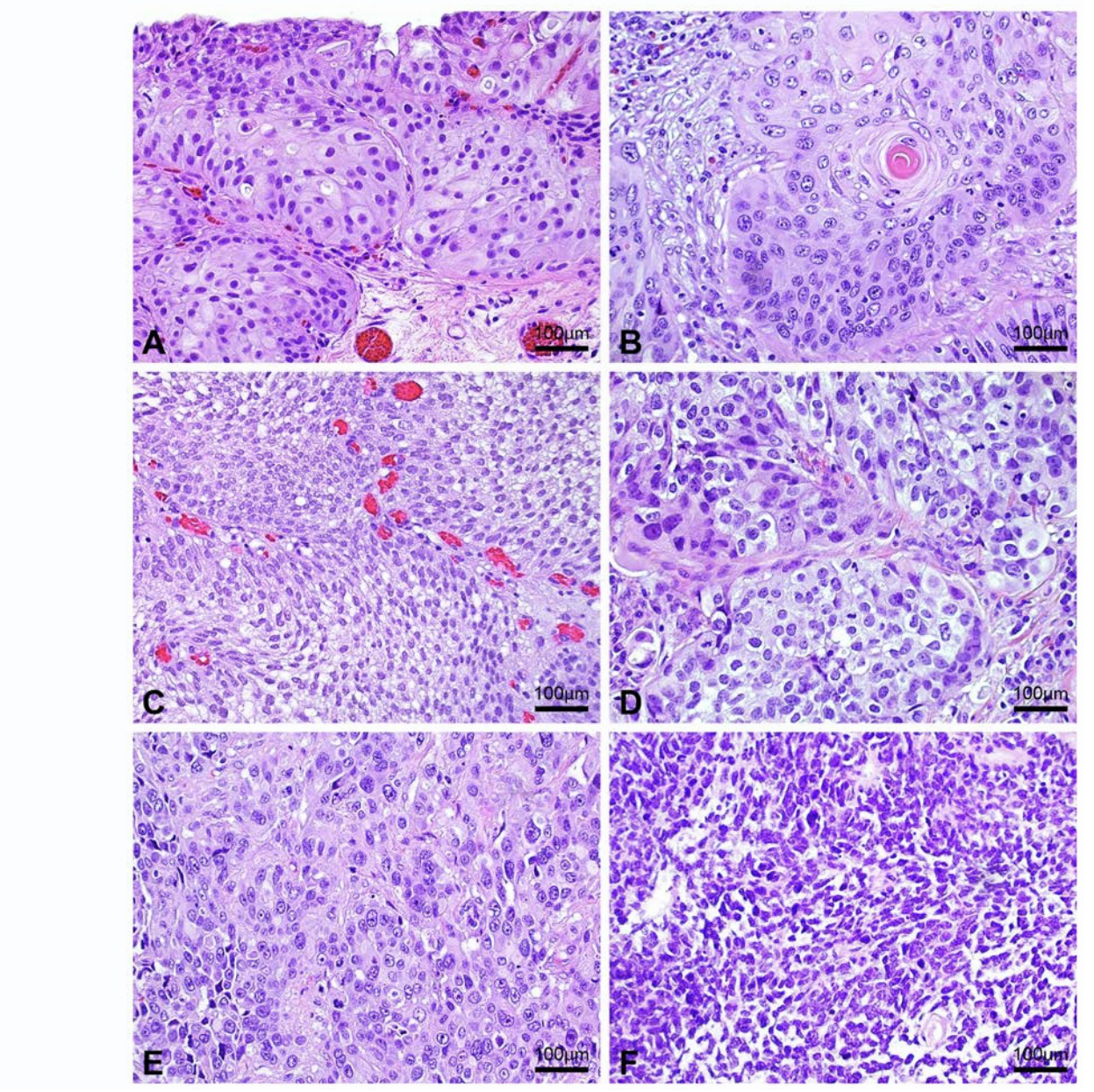
Email: rkalluri@mdanderson.org

Supplementary data with analyses for individual patients

https://doi:10.17632/kg2ccb425s.1

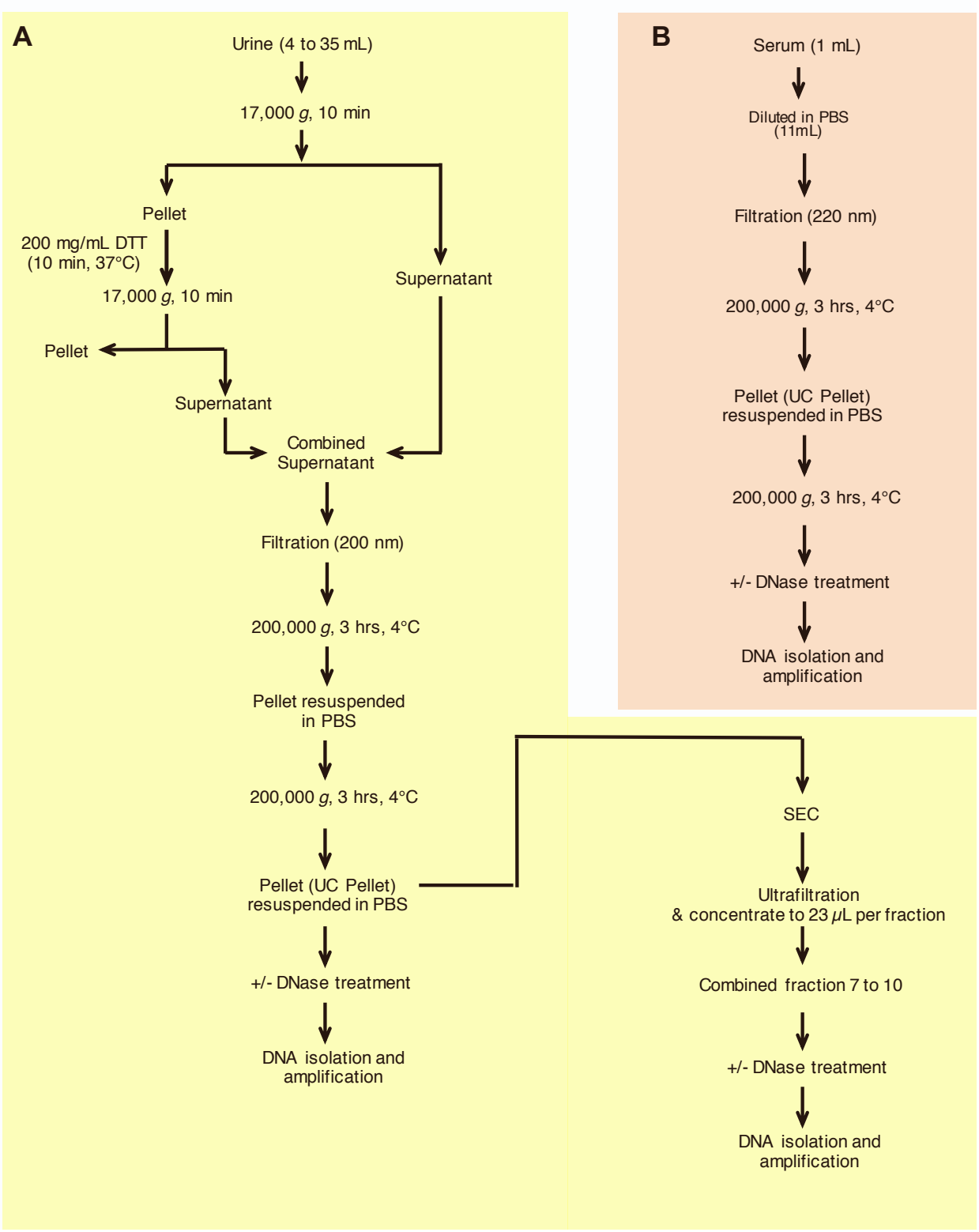
The files contain (a) annotations of the genes containing somatic variations; (b) the results of the enrichment analyses for the identification of the functional categories and pathways over-represented in the set of genes containing somatic variations; (c) the reconstructions of the molecular networks of genes identified in every sample (e.g. tumor or urine), and (d) the predictions of the miRNAs potentially targeting the UTRs containing somatic variations.

Supplementary Figures



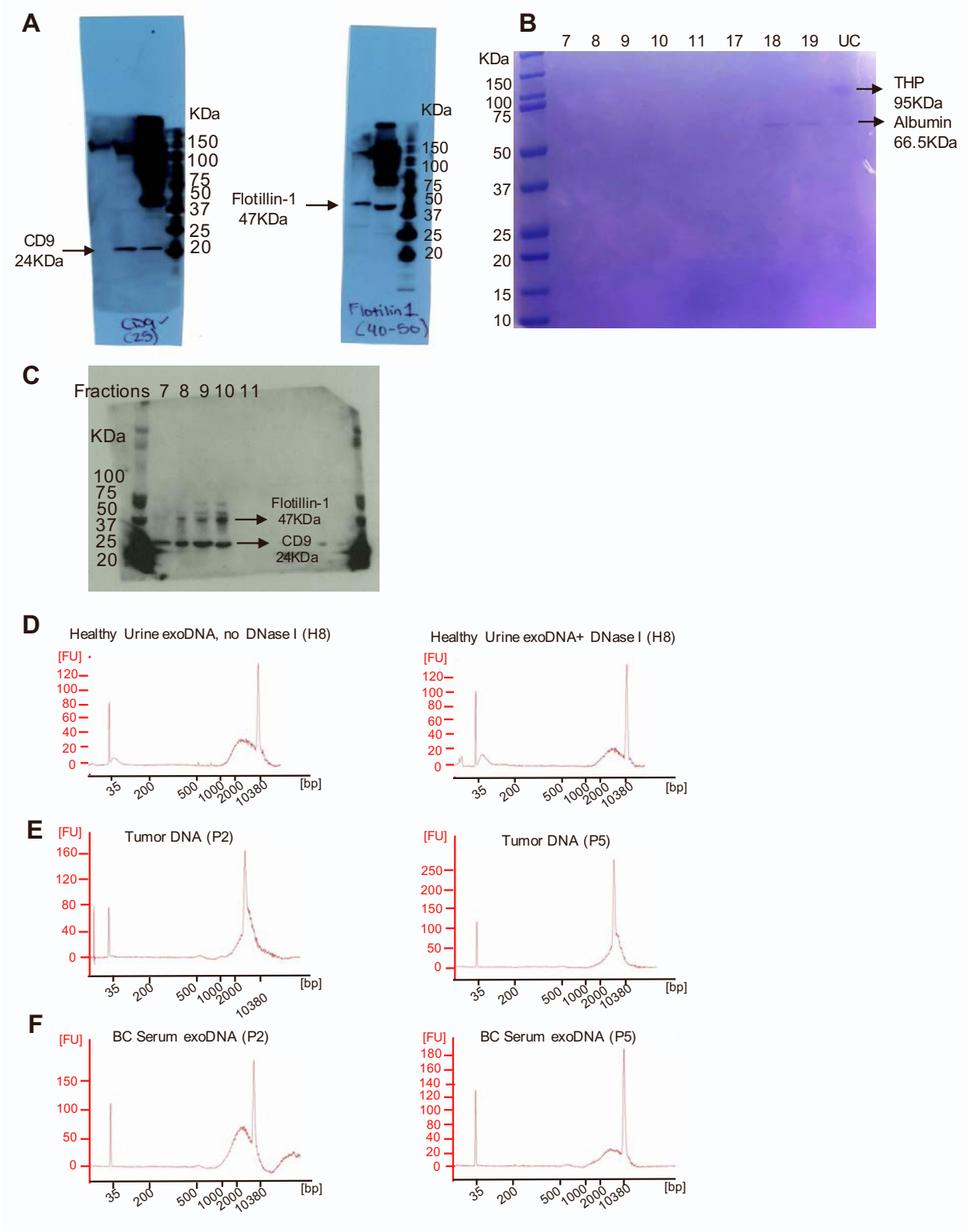
Supplementary Figure 1. Representative H&E images of bladder cancer patients

Representative H&E images of the indicated bladder cancer patient; please see **Supplementary Table 1** for details. (**A**) Papillary urothelial carcinoma, low grade (P1). (**B**) Urothelial carcinoma with squamous differentiation, high grade (P2). (**C**) Papillary urothelial carcinoma, low grade (P3). (**D**) Urothelial carcinoma, high grade (P4). (**E**) Urothelial carcinoma, high grade (P5). (**F**) Small cell variant of urothelial carcinoma (P6). Scale bar, 100 μm.



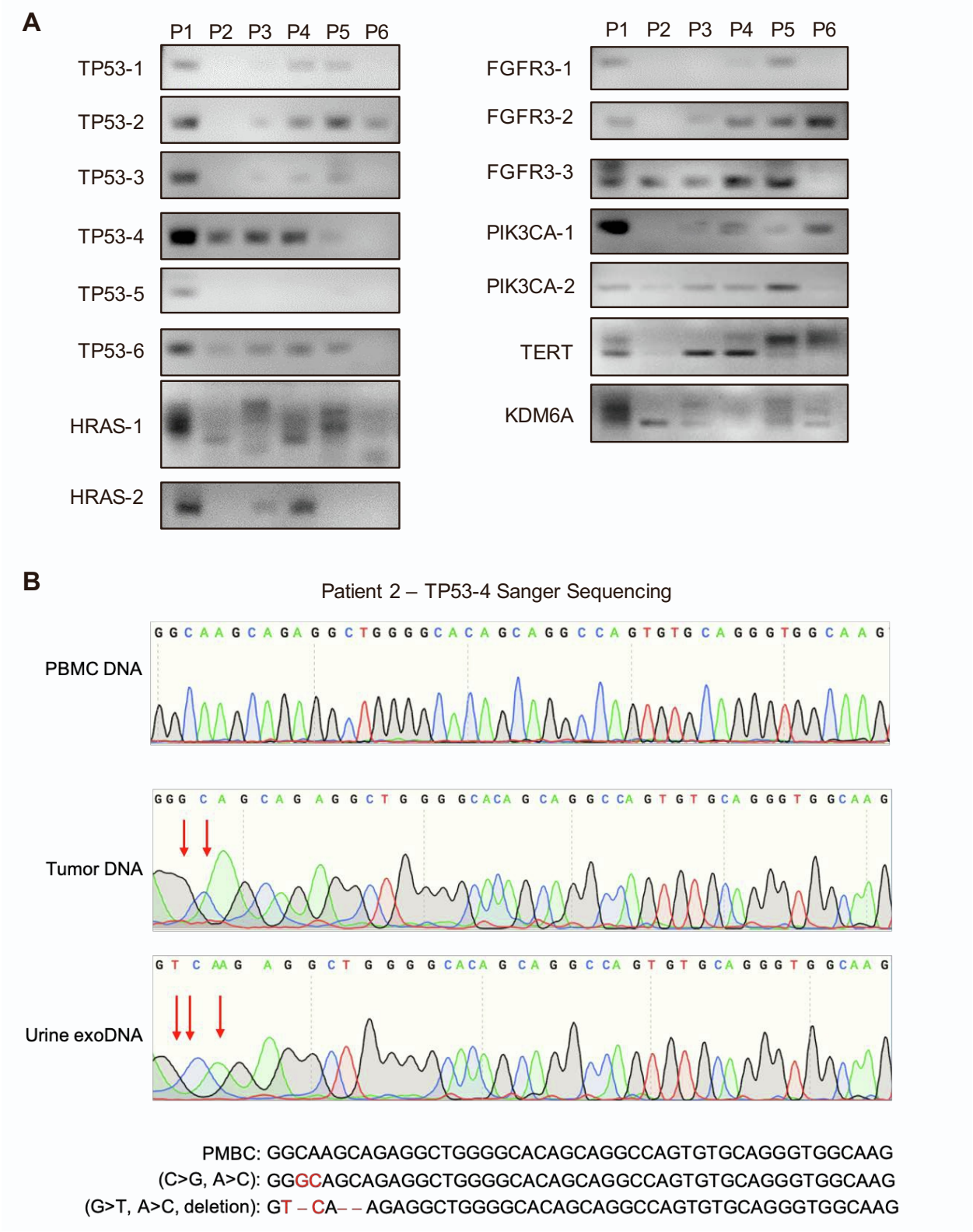
Supplementary Figure 2. Diagram of urine exosome isolation and purification procedure. Healthy sample and BC patient urine are thawed from -80°C storage and centrifuged to remove protein contaminants

and other debris. The resulting pellet contains Tamm-Horsfall (THS) glycoprotein, which is known to trap exosomes. These exosomes can be liberated by digesting Tamm-Horsfall in DTT. Liberated exosomes, now in the supernatant following an additional spin, are combined with the supernatant of the first spin. The combined supernatants are then processed using filtration and ultracentrifugation (UC) to collect exosomes for downstream analysis. The exosomes from UC (UC pellets) were also subjected to size exclusion chromatography and fractions 7-10 were pooled for further experiments.



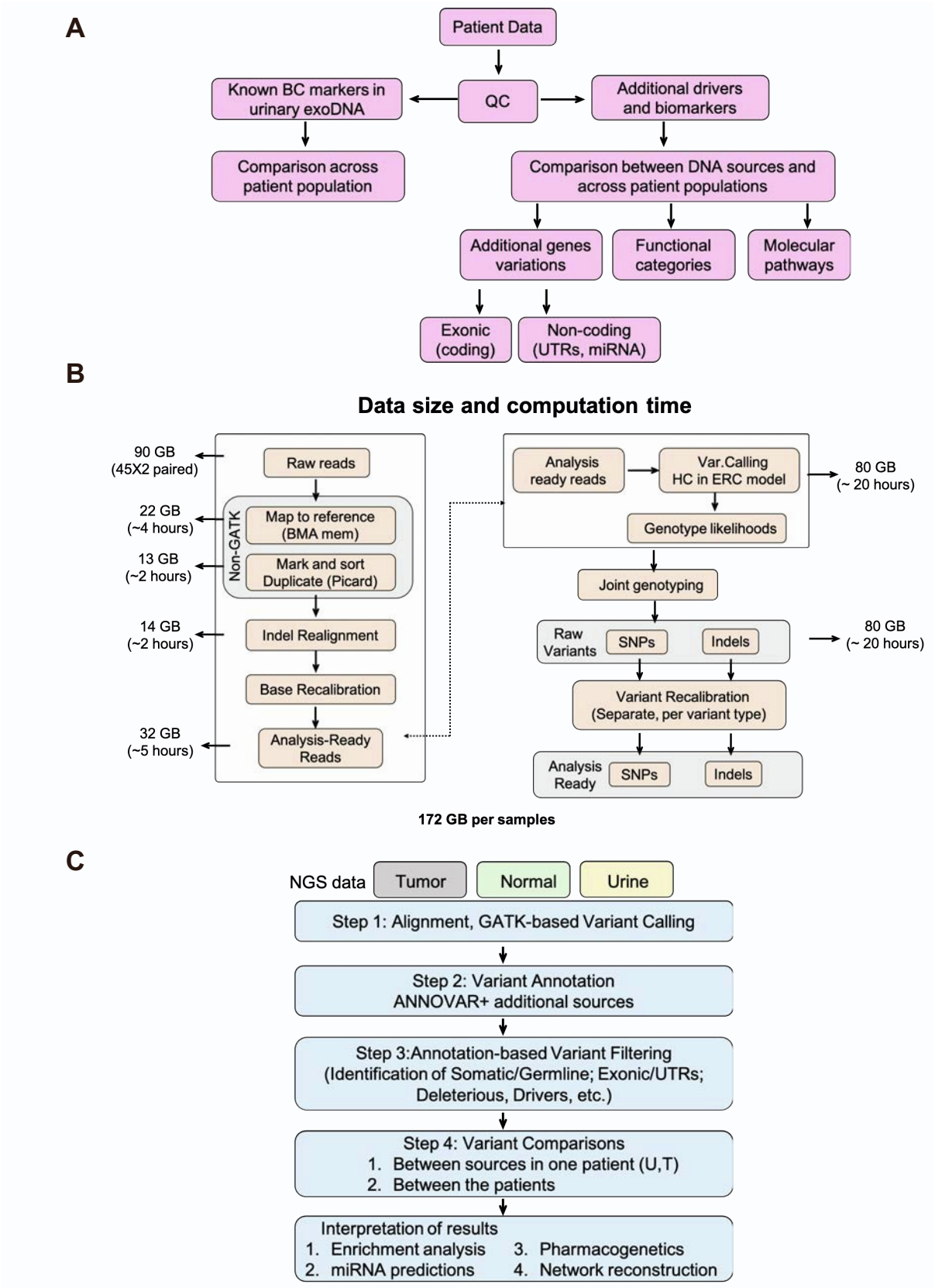
Supplementary Figure 3. Full-size autographs of the Western blots and Coomassie staining. (**A**) Lane 1: Healthy human urine exosome lysate. Lane 2: Healthy human serum exosome lysate. (**B**) Coomassie

blue staining of the fractions 7-11, 17-19 and UC exosome pellets. The presence of albumin was found in fractions 18 and 19. THP and albumin were present in the UC pellet. (**C**) Fraction 7-11 of urine exosome lysate. Molecular weights are indicated in writing under each autograph. (**D**) DNA was isolated from healthy sample 8 (H8) without DNase I (left) or with DNase I (right) to eliminate exogenous DNA and resultant DNA fragments were analyzed by capillary electrophoresis. (**E-F**) DNA was isolated from BC tumor biopsies (**E**), matched serum (**F**) from P2 and P5 and analyzed by capillary electrophoresis.

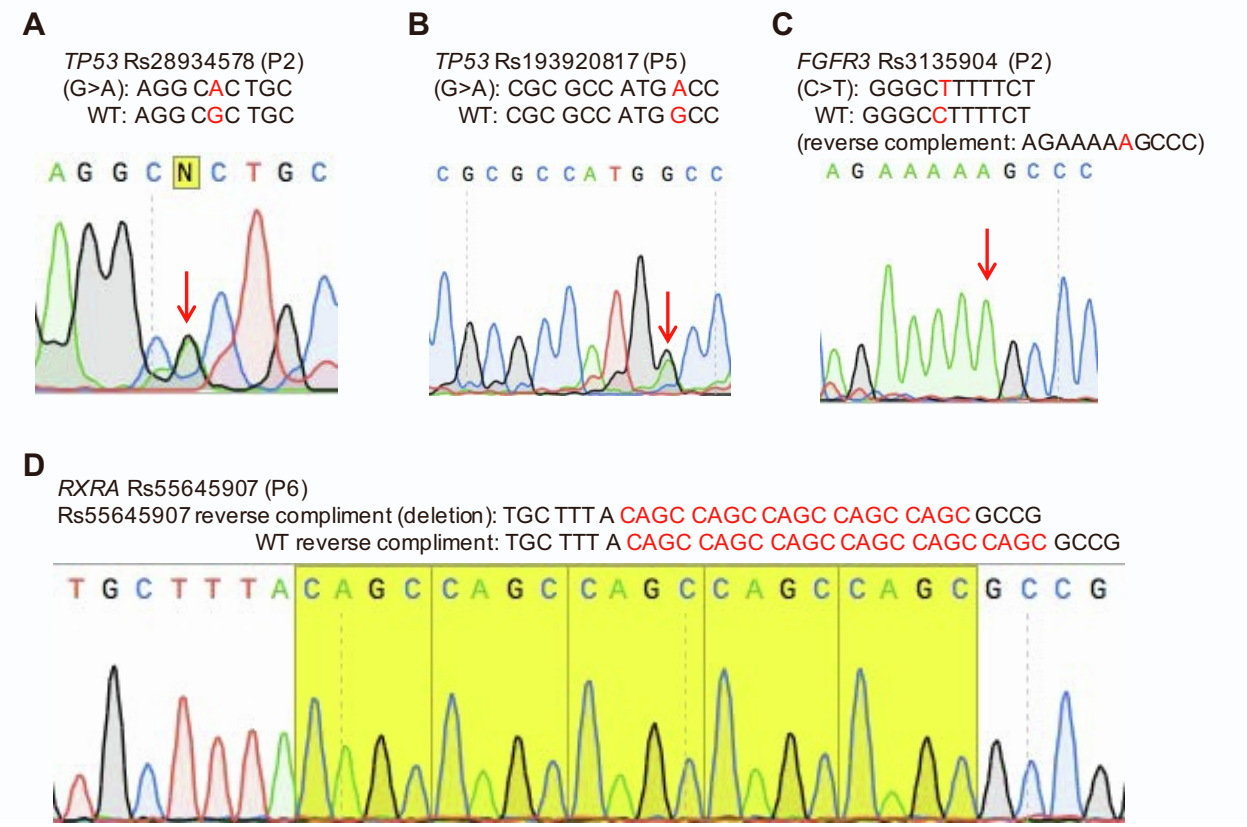


Supplementary Figure 4. PCR & Sanger sequencing of BC hotspots in patient samples. (**A**) Gel electrophoresis of PCR hotspot targets in urine exoDNA samples in six BC patients. (**B**) Representative

Sanger sequencing alignment of *TP53* hotspot PCR product in PBMC, tumor, and urine exoDNA from patient 2.


A
Patient Data
Known BC markers in urinary exoDNA
QC
Additional drivers and biomarkers
Comparison across patient population
Comparison between DNA sources and across patient populations
Additional genes variations
Functional categories
Molecular pathways
Exonic (coding)
Non-coding (UTRs, miRNA)
B
Data size and computation time
90 GB (45X2 paired)
Raw reads
22 GB (~4 hours)
Non-GATK
Map to reference (BMA mem)
13 GB (~2 hours)
Mark and sort Duplicate (Picard)
14 GB (~2 hours)
Indel Realignment
Base Recalibration
32 GB (~5 hours)
Analysis-Ready Reads
Analysis ready reads
Var.Calling HC in ERC model
80 GB (~ 20 hours)
Genotype likelihoods
Joint genotyping
Raw Variants
SNPs
Indels
80 GB (~ 20 hours)
Variant Recalibration (Separate, per variant type)
Analysis Ready
SNPs
Indels
172 GB per samples
C
NGS data
Tumor
Normal
Urine
Step 1: Alignment, GATK-based Variant Calling
Step 2: Variant Annotation
ANNOVAR+ additional sources
Step 3:Annotation-based Variant Filtering
(Identification of Somatic/Germline; Exonic/UTRs;
Deleterious, Drivers, etc.)
Step 4: Variant Comparisons
1. Between sources in one patient (U,T)
2. Between the patients
Interpretation of results
1. Enrichment analysis
2. miRNA predictions
3. Pharmacogenetics
4. Network reconstruction

Supplementary Figure 5. Bioinformatics analysis using NGS data for normal and tumor tissues, serum and urine exoDNA. (**A**) Analytic workflow; (**B**) High-throughput data analysis; (**C**) Downstream analysis.



Supplementary Figure 6. Sanger sequencing confirmed select variants in genes frequently mutated in bladder cancer. Tumor DNA isolated for the P2, P5 and P6 sample sets, which showed identical variants in the urine exoDNA and tumor tissue. DNA was amplified with primers flanking the identified variants and Sanger sequencing performed (for primers see Supplementary Table 2). (**A, B**) Confirmed variant for exonic variants in *TP53* sequence (P2 and P5, respectively). (**C**) Confirmed exonic variant in *FGFR3* sequence (P2). (**D**) Confirmed deletion in 3'UTR region in *RXRA* (P6).

Supplementary Tables

Supplementary Table 1. Bladder cancer patient information.

Information		Biopsy			Cystectomy		Treatment	
Patient ID	**Gender**	**Histology**	**Grade**	**Invasive**	**Positive Nodes (Y/N)**	**LVI (Y/N)**	**Neo-adjuvant Chemo**	**BCG/Intravesical Tx**
P1	Male	Papillary UCC	Low	Non-invasive	No	No	Ifos/Adria(2) DD-MVAC(4)	No
P2	Female	Squamous	High	Non-invasive	No	No	CGI-5 cycles	No
P3	Female	Papillary UCC	Low	Non-invasive	-	-	No	BCG/Mitomycin
P4	Male	UCC	High	Invasive	No	No	GEM/Cis 4 cycles	Yes (BCG)
P5	Male	UCC	High	Invasive	Yes	No	DD-MVAC 4 cycles	No
P6	Male	Small Cell/UCC	-	Invasive	No	No	Ifos/Adria(1) Etop/Cis (3)	No
P7	Male	UCC	High	invasive	Yes	-	Tax/Carbo, Gem/Cis, Pembro	No
P8	Male	Papillary	Low	Non-invasive	-	-	N/A	Yes (BCG)
P9	Male	Papillary	High	Non-invasive	-	-	N/A	Yes (BCG)
P10	Male	UCC, CIS	High	Non-invasive	No	No	No	Yes (BCG)

Supplementary Table 2. Primer sets for PCR amplification and Sanger sequencing.

Gene	Use	Forward – 5’	Reverse – 3’
FGFR3-1	PCR	CATGTCTTTGCAGCCGAGGA	GGCAGCTCAGAACCTGGTAT
FGFR3-2	PCR	GTGACCGAGGACAACGTGAT	TCGGTCAAACAAGGCCTCAG
FGFR3-3	PCR	CCCTGAGCGTCATCTGCC	ACCTTGCTGCCATTCACCTC
HRAS-1	PCR	GCGCCAGGCTCACCTCTAT	CTGGGCCTGGCTGAGCA
HRAS-2	PCR	ACTGGTGGATGTCCTCAAAAGA	AGAGGCTGGCTGTGTGAACT
KDM6A	PCR	ACACAACCAGCATTTACTTTTCCT	ATTGGCCAAAGGCTGCCC
TP53-1	PCR	GGCAACTGACCGTGCAAGT	TGCTGTCCCCGGACGATATT
TP53-2	PCR	AAGAAGCCCAGACGGAAACC	TCACCCATCTACAGTCCCCC
TP53-3	PCR	AACCCCTCCTCCCAGAGAC	CCAGGCCTCTGATTCCTCAC
TP53-4	PCR	TATGGAAGAAATCGGTAAGAGGTGG	ATCTTGGGCCTGTGTTATCTCC
TP53-5	PCR	CTGAGGCATAACTGCACCCT	TCCTTACTGCCTCTTGCTTCTC
TP53-6	PCR	GCTGCTCACCATCGCTATCT	TACTCCCCTGCCCTCAACAA
PIK3CA-1	PCR	CATCTGTGAATCCAGAGGGGAA	AGCACTTACCTGTGACTCCAT
PIK3CA-2	PCR	ACATTCGAAAGACCCTAGCCTT	AATCGGTCTTTGCCTGCTGA
TERT	PCR	AGTGGATTCGCGGGCACAGA	CAGCGCTGCCTGAAACTC
FGFR3-4	PCR	CCTGAAGATGGGAGCCTTTAC	CCTGGGACACACAGCAATTA
FGFR3-5	Sequencing	AGGCTGGACGTACATTCTTG	
RXRA	PCR	TGAGCCTCATACCTGTACCA	CTCTGTGGCATCTTCACTCC
RXRA	Sequencing	GGTGGCTAATGAGCTGATGTTA	
TP53-7	PCR	CATCACACCCTCAGCATCTC	GCCAGACCTAAGAGCAATCA
TP53-8	Sequencing	CATCACACCCTCAGCATCTC	

Supplementary Table 3. Total reads and coverage achieved by whole exome sequencing. Mapping rate (≥95%), Duplicate mapped reads (≤25%), Mean coverage (≥100X), Median coverage (≥50X) WES total reads & coverage. PBMC DNA was not available for patient 3 and this patient was excluded from further bioinformatics analysis. Raw sequencing metrics revealed a mean target coverage of 158-198X in PBMC samples, 138-162X in tumor samples, 31-334X in urine exosome samples, and 14-187X in serum exosome samples. Median target coverage ranged from 124-156X in PBMC samples, 103-127X in tumor samples, 1-138X in urine exosome samples, and 0-24X in serum samples. Median target coverage was likely reduced in urine and serum exosome samples due to whole-genome amplification being employed before library preparation, which is known to create bias in sequence fragment representation.

Type	Patient ID	Total reads x 10^6	Mapping rate (%)	Duplicate Mapped Reads (%)	Mean Coverage	Median Coverage	100X (%Targets)	75X (%Targets)	50X (%Targets)
PBMCs	1	166.57	99.78	12.75	196.35	155.77	68.03	73.68	77.9
PBMCs	2	162.04	99.8	13.23	198.36	156.26	68.36	73.86	77.78
PBMCs	4	150.11	99.8	11.66	186.92	150.57	67.18	73.29	77.63
PBMCs	5	142.86	99.76	13.26	168.41	132.65	63.18	71.1	76.72
PBMCs	6	138.96	99.7	13.2	158.76	124.98	61.02	70	76.31
Tumor	1	135.69	99.73	10.89	161.8	127.42	61.48	70.12	76.21
Tumor	2	131.81	99.73	11.3	157.93	118.28	57.96	67.85	75.18
Tumor	3	119.83	99.69	12.57	138.7	103.83	52.03	64.7	74.27
Tumor	4	126.69	99.73	11	153.08	114.17	56	65.81	73.64
Tumor	5	133.66	99.72	10.78	162.23	124.24	60.33	69.43	75.91
Tumor	6	128.38	99.73	13.91	149.83	109.06	54.27	65.52	74.28
Urine ExoDNA	1	313.38	99.28	18.11	334.59	138.64	39.59	47.82	58.56
Urine ExoDNA	2	247.14	98.38	24.03	233.48	63.12	23.67	29.86	39.25
Urine ExoDNA	3	238.27	96.4	42.73	162.47	6.86	8.88	10.95	14.43
Urine ExoDNA	4	239.51	78.11	76.43	31.29	1.14	1.58	2.08	2.98
Urine ExoDNA	5	249.37	98.75	32.85	199.52	20.47	11.62	14.94	20.64
Urine ExoDNA	6	260.87	98.9	36.09	185.7	26.2	9.64	13.4	20.28
Serum ExoDNA	1	163.31	87.06	75.18	13.99	0	0.6	0.77	1.08
Serum ExoDNA	2	255.72	97.39	39.18	187.52	23.95	24.08	29.04	36.44
Serum ExoDNA	3	259.5	94.78	47.39	141.96	3.93	12.94	15.42	19.33
Serum ExoDNA	4	264.42	95.95	65.09	148.16	0	5.97	6.93	8.46

Serum ExoDNA	5	255.88	93.14	47.93	122.54	21.31	20.08	25.14	32.88
Serum ExoDNA	6	256.7	89.06	78.58	33.8	0	1.29	1.57	2.12

Below quality threshold

Supplementary Table 4. Total variants, concordance and contamination analysis. (**A**) Total variants by sample, (**B**) Concordance, PBMCs (Normal, N) vs Tumor (T), (**C**) Concordance, PBMCs (Normal, N) vs Urine (U), (**D**) Concordance, PBMCs (Normal, N) vs Serum (S), (**E**) Contamination, PBMCs (Normal, N) vs Tumor (T), Urine (U) and Serum (S), (**F**) Summary of QU analysis.

A. Total variants by sample

	P1	P2	P4	P5	P6	Total Variants
PBMCs	1,055,918	954,048	879,234	908,288	948,607	
Serum	**28,197**	515,271	**140,414**	438,430	**34,241**	1,156,553
Tumor	944,394	934,908	1,053,203	874,486	872,261	4,679,252
Urine	1,214,430	851,080	**92,435**	524,199	637,053	3,319,197
Total variants	3,242,939	3,255,307	2,165,286	2,745,403	2,492,162	

B. Concordance, PBMCs (Normal, N) vs Tumor (T)

	P1N	P2N	P4N	P5N	P6N
P1T	**99.90%**	39.11%	40.33%	40.62%	39.06%
P2T	39.21%	**99.84%**	37.81%	39.78%	37.43%
P4T	40.31%	37.67%	**99.86%**	40.08%	38.07%
P5T	40.65%	39.78%	39.98%	**99.88%**	39.09%
P6T	38.88%	37.13%	38.12%	38.59%	**99.92%**

C. Concordance, PBMCs (Normal, N) vs Urine (U)

	P1N	P2N	P4N	P5N	P6N
P1U	**98.86%**	32.74%	34.62%	35.01%	31.30%
P2U	34.90%	**98.56%**	34.19%	36.00%	31.89%
P4U	44.85%	45.05%	***67.56%***	48.70%	43.95%
P5U	41.67%	39.62%	40.75%	**95.03%**	39.17%
P6U	35.83%	32.78%	34.25%	36.22%	**92.19%**

D. Concordance, PBMCs (Normal, N) vs Serum (S)

	P1N	P2N	P4N	P5N	P6N
P1S	***49.05%***	71.52%	42.95%	42.13%	38.98%
P2S	41.59%	**97.94%**	42.21%	42.08%	39.36%
P4S	55.02%	52.07%	***66.84%***	55.54%	52.81%
P5S	38.79%	38.42%	38.84%	**98.20%**	35.84%
P6S	51.79%	51.97%	50.82%	48.18%	***49.14%***

E. Contamination, PBMCs (Normal, N) vs Tumor (T), Urine (U) and Serum (S)

	N	T	U	S
P1	0.37%	0.51%	**1.38%**	***98.87%***
P2	0.36%	**0.70%**	**1.31%**	**2.35%**
P4	0.32%	0.33%	***94.87%***	***99.51%***
P5	0.55%	**0.75%**	**2.78%**	**1.55%**
P6	0.43%	**0.65%**	**4.48%**	***98.99%***

F. Summary of QU analysis. Conc: concordance; Cont: contamination.

	N		T		U		S	
	Conc	**Cont**	**Conc**	**Cont**	**Conc**	**Cont**	**Conc**	**Cont**
P1	OK	Possible	OK	Possible	OK	Possible	Possible swap	Massive
P2	OK	Possible	OK	Possible	Possible swap	Massive	Negative	Massive
P4	OK	OK	OK	OK	Possible swap	Massive	Possible swap	Massive
P5	OK	Possible	OK	Possible	OK	Possible	Negative	Massive
P6	OK	Possible	OK	Possible	OK	Possible	Possible swap	Massive

Supplementary Table 5. Read depth per individual somatic variants identified in the study. Normal: matched PBMCs.

Variation ID	Normal (PBMCs)					Tumor					Urine					Serum				
	P1	P2	P4	P5	P6	P1	P2	P4	P5	P6	P1	P2	P4	P5	P6	P1	P2	P4	P5	P6
rs10415095						4	3	11	7		27								10	
rs11343599						4	4		2		8			10					6	
rs58312807	4				4	3	5	3	5	4	4			10	11				17	
rs1130214						6	5		3		27						18			
rs2976396	6			8			5	6		5	193	2212		226	95					
rs13258775			4				5		6	3					3					
rs251860							3	6	2		3				5					
rs1051782	4						5		3	3										
rs2422978	5							4	6	6	3									
rs55645907	3			10		2	7	3		2	6	8			4		5			
rs1045570				5		3			4		9						10			
rs4842194	7		3	6		4	5		12		8	6								
rs34109509							3										25			
rs35280127	7						7				14									
rs3135904			3	8	3		2	9	3	3	37	20		18						
rs28934578							151													
rs193920817									128											
rs1800372		115					58					723	5				442			
rs9266												12			32					
rs712		8		4	6	7		3		6					3					
rs3828609									2											
rs1057016							2								2					
rs3173956	4				5			2			3									
rs704010					7	2	2		16		30			22	4					
rs10875943	10					7	3	3			17			6			158			
rs10248903	3						3		3		15	2			2					
rs7931342			5			3	7	3	4											
rs2981582	2						4	3			5	68								
rs5768709		2				4	3				8	3								
rs7832232	2						2		2		5	18								
rs1883924		4				3					7	5							5	
rs4939827		2					2					9			2				4	
rs4986938	125					111					557		5			8				
rs7504990					5						429	206		49	225					
rs4132601	3					5			4	3										
rs2367202					3	3	3				2									
rs3176336	2					2				11	6									
rs5030625	3					2			4			5								
rs9340799					5		2			3		4								
rs2234693					3					2		4			3					

rs1138272		117					60					844					187			
rs11611238		36					33					88					2			
rs17632542				266					264					198					751	
rs17634425				3					5					4					7	
rs1799939				302					291					589					11	
rs1801270		251					232					52					33			
rs2107425		5								3		17					61			
rs2479106		4					5					3					10			

Supplementary Table 6. Total shared & unique somatic variants in driver genes in tumor samples.

Shaded cells represent the number of variant unique for a specific patient. Clear cells show the overlap between the two patients.

	P1	P2	P4	P5	P6	Total
P1	90	6	11	9	5	124
P2	6	109	12	10	7	144
P4	11	12	71	11	6	117
P5	9	10	11	110	5	152
P6	5	7	6	5	81	107
	Overlap of variants between patients					
	Variants unique for a specific patient					

Supplementary Table 7. Number of variations predicted by GATK. GATK-based variant analysis using DNA isolated from tumor tissues, exoDNA from urine and serum, and matched normal (PBMC) DNA as the reference sequence.

Variation type	Normal (PBMCs)				
	P1	P2	P4	P5	P6
Exonic	23,292	23,647	22,878	22,966	24,052
Exonic; splicing	11	9	10	14	13
ncRNA_splicing	16	18	16	17	22
ncRNA_UTR3	239	214	209	222	250
ncRNA_UTR5	90	75	82	79	84
ncRNA_UTR5; ncRNA_UTR3	1	1	1	1	1
Splicing	146	134	137	148	151
UTR3	12,858	12,177	11,258	11,830	12,801
UTR5	4,870	4,801	4,492	4,675	4,685
UTR5;UTR3	11	12	8	11	12
Downstream	8,610	7,678	6,795	7,201	7,924
Intergenic	551,475	488,321	450,182	462,446	474,249
Intronic	401,210	367,773	337,914	351,450	375,700
ncRNA_exonic	5,476	5,321	4,928	5,228	5,236
ncRNA_intronic	34,464	31,578	28,947	30,253	31,093
Upstream	12,666	11,777	10,960	11,305	11,916
Upstream; downstream	483	512	417	442	418

Variation type	Tumor				
	P1	P2	P4	P5	P6
Exonic	23,252	23,792	22,765	22,953	23,097
Exonic; splicing	11	10	10	14	13
ncRNA_splicing	20	14	17	20	21
ncRNA_UTR3	217	212	220	226	233
ncRNA_UTR5	87	76	73	78	85
ncRNA_UTR5; ncRNA_UTR3	1	1	1	1	0
Splicing	140	141	131	143	152
UTR3	11,895	12,001	11,498	11,969	11,480
UTR5	4,646	4,666	4,432	4,539	4,438
UTR5; UTR3	10	11	7	10	9
Downstream	7,476	7,326	8,141	7,105	7,133
Intergenic	487,245	478,321	556,140	436,611	437,281
Intronic	361,076	360,185	399,713	345,189	342,511
ncRNA_exonic	5,296	5,199	4,997	5,089	4,984
ncRNA_intronic	31,316	31,262	33,656	29,313	29,470
Upstream	11,290	11,180	10,977	10,787	10,934
Upstream; downstream	416	511	425	439	420

Variation type	Urine				
	P1	P2	P4	P5	P6
Exonic	23,689	24,205	9,773	22,064	23,359
Exonic; splicing	10	9	5	9	13
ncRNA_splicing	25	25	6	18	19
ncRNA_UTR3	250	229	41	147	143
ncRNA_UTR5	85	69	13	48	54
ncRNA_UTR5; ncRNA_UTR3	1	8	0	0	1
Splicing	168	183	87	193	240
UTR3	13,970	11,370	1,587	7,705	7,973
UTR5	4,684	4,008	642	2,726	2,610
UTR5;UTR3	10	11	2	7	11
Downstream	11,068	7,868	672	4,876	4,728
Intergenic	635,479	417,525	39,441	252,902	318,299
Intronic	465,532	343,203	35,790	206,794	250,256
ncRNA_exonic	6,013	4,905	755	3,656	3,308
ncRNA_intronic	39,214	26,926	2,612	16,571	20,251
Upstream	13,688	10,054	959	6,144	5,572
Upstream; downstream	544	482	50	339	216

Variation type	Serum				
	P1	P2	P4	P5	P6
Exonic	4,609	23,779	7,080	25,240	3,758
Exonic; splicing	4	9	6	13	3
ncRNA_splicing	0	15	2	12	3
ncRNA_UTR3	15	171	48	140	7
ncRNA_UTR5	3	48	5	35	1
ncRNA_UTR5; ncRNA_UTR3	0	0	0	0	0
Splicing	30	235	81	281	52
UTR3	656	8,523	1,961	7,980	710
UTR5	306	3,164	832	3,098	319
UTR5; UTR3	0	7	1	6	0
Downstream	122	4,431	1,170	3,627	243
Intergenic	10,336	239,596	69,886	189,446	14,448
Intronic	10,740	208,731	51,924	185,453	12,924
ncRNA_exonic	303	3,278	1,127	3,097	294
ncRNA_intronic	771	16,402	4,660	13,690	955
Upstream	288	6,570	1,568	5,985	495
Upstream; downstream	14	312	63	327	29

Supplementary Table 8. Antibodies used in this study.

Primary antibodies							
Antigen	Specificity	Host	Vendor	Catalog No	Lot No	Used in	Dilution
CD9	Human, Mouse, Rat	Rabbit	Abcam	Ab92726	GR237847-20	TEM, WB	1:300, 1:1000
Flotillin-1	Human	Rabbit	Santa Cruz	sc25506	H1914	WB	1:300
Secondary antibodies							
Ab Type	Specificity	Host	Conjugate	Vendor	Catalog No	Used in	Dilution
IgG	Rabbit	Goat	HRP	Sigma	A0545	WB	1:2000