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,2 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,8,9

1Department of Cancer Biology, University of Texas MD Anderson Cancer Center, Houston, TX, USA; 2Department of Human Genetics, University of Chicago, Chicago, IL, USA; 3Department of Genomic Medicine, University of Texas MD Anderson Cancer Center, Houston, TX, USA; 4Feinberg School of Medicine, Northwestern University, Chicago, IL, USA; 5Johns Hopkins Greenberg Bladder Cancer Institute, Baltimore, MD, USA; 6Department of Urology, University of Texas MD Anderson Cancer Center, Houston, TX, USA; 7Department of Pathology, University of Texas MD Anderson Cancer Center, Houston, TX, USA; 8School of Bioengineering, Rice University, Houston, TX, USA; 9Department 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.11–14 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.15

Received 15 January 2021; accepted 21 May 2021; https://doi.org/10.1016/j.omtm.2021.05.010.
10These 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.1–34 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.32 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.33 Size exclusion chromatography (SEC), which is performed by passive gravity, 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, 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 × 1011 mL−1 versus 3.6 × 1010 mL−1, respectively; Figures 1A–1C). Western blot analysis identified exosome-associated proteins, including tetraspan 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–Horsfall 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–Sidák 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.
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 × 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 BC-related hotspot regions in six genes (TERT, FGFR3, PIK3CA, TP53, HRAS, and KDM6A), designed based on a previous study (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 (≥100×) 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.44–47 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
shown in Table 4. Somatic mutations in genes commonly associated with BC are shown in Table 5. Analysis of exoDNA samples revealed 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.

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.

**Table 1. Distribution of identified somatic variations**

| Tissue source      | Patient ID | All somatic variants | Somatic driver variants |
|--------------------|------------|----------------------|-------------------------|
|                    |            | 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.

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

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 30 and 50 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 30 and 50 UTRs

---

**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                            | 1,109      | 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                         | 16.3999       |                 |
| 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.

---

**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

(P1 and P2 and P6, respectively).
have recently been associated with high-penetrance hereditary disorders. Significant polymorphisms in the 5’ regions and in the 3’ UTRs are linked to glioma, colon, breast, and ovarian cancer. 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. 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 LAT51 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 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| T  | –  |
| KLK10         | rs11345999 | no UTR3| U, T   | T  | T  | U, S, T| T  | –  |
| IGF2          | rs58312807 | yes UTR3| –     | T  | T  | U, S, T| T  | –  |
| AKT1          | rs1130214  | no UTR5| U, T   | S, T| T  | –  | T  | –  |
| FSCA          | rs2976396  | yes UTR3| –     | U, T| T  | –  | U, T| T  |
| PTK2          | rs13258775 | yes UTR5| –     | T  | –  | T  | U, T| T  |
| PLEKHG2       | rs251860   | no UTR5| U, T   | T  | T  | T  | U  | –  |
| ETV6          | rs1051782  | no UTR3| –     | T  | –  | T  | T  | T  |
| RASSF2        | rs2422978  | no UTR3| –     | –  | –  | T  | T  | T  |
| TBX3          | rs1061651  | yes UTR3| –     | –  | T  | 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.

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

| Ranking | Gene     | Total variants | CGI/BC | P1  | P2  | P4  | P5  | P6  |
|---------|----------|----------------|--------|-----|-----|-----|-----|-----|
| 1       | KLK10    | 28             | no     | 4   | 6   | 3   | 7   | 0   |
| 2       | IGF1R    | 13             | yes    | 6   | 2   | 3   | 1   | 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              | yes    | 0   | 1   | 1   | 3   | 2   |
| 8       | SYNF2    | 7              | yes    | 1   | 1   | 1   | 1   | 3   |
| 9       | CREBL2   | 6              | no     | 2   | 1   | 1   | 1   | 1   |
| 10      | PLEKHG2  | 6              | no     | 1   | 1   | 2   | 1   | 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.
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 exosomal contents deviate from control levels. Finally, exosome testing based on urine exoDNA and intended target populations.  

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\(^71\) 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 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  |
|       | rs1045570  | no            | UTR3        | T, | U  | S  |    |    |
|       | rs4842194  | no            | UTR3        | T, | U  |    |    |    |
|       | rs34109509 | no            | UTR3        | T, | S  |    |    |    |
|       | rs35280127 | no            | UTR3        |    | T  |    |    |    |
| TP53  | rs193920817| no            | exonic      |    |    |    |    |    |
|       | rs1800372  | no            | exonic      |    |    |    |    |    |
|       | 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.
alone. The unique subsets of mutations found specifically in urine suggest the genetic heterogeneity of tumors than small biopsy specimens. Biopsy DNA and urine exoDNA can be a better representation of biomarkers. Our study suggests that combined sequencing of tumor-associated exosomes compared with other exosomes in the urine. Urine-derived exosomes may proportionally capture more cancer-associated exosomes compared with other cell-derived exosomes in the serum. In contrast, exosomes are likely a proportionally minor population compared which are generally focused on coding regions, but the significance of UTR variants is often excluded from WES bioinformatics pipelines, of urine exoDNA as a biomarker. 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. 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 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. 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 or potentially activated the tumor stroma. 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, PDGFRα, 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. 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.

Table 6. Individual-specific variants identified in urine exoDNA

| P1 | Variants | Gene | P2 | Variants | Gene | P4 | Variants | Gene | P5 | Variants | Gene | P6 | 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 | ES2 | 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.
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.

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.
Table 7. Mutated miRNA binding sites in driver genes

| Gene | dbSNP ID | miRNA ID | Tumor | Urine | Serum | Total |
|------|----------|----------|-------|-------|-------|-------|
|      |          |          | Patient ID | Patient ID | Patient ID |       |
| PAX2 | rs76053583 | miR-185-5p | 1     | 2     | 4     | 6     | 1     |
|      | x        | x        |       |       |       |       |       |
| KRAS | rs7212   | miR-181-5p |       |       |       |       | 2     |
|      | x        | x        |       |       |       |       |       |
| CSF1R| rs8328609 | miR-155-5p |       |       |       | 1     |       |
|      | x        | x        |       |       |       |       |       |
| PCM1 | rs1057016 | miR-599   |       |       |       | 1     | 1     |
|      | x        | x        |       |       |       |       |       |
| ADAM7| rs3173956 | miR-382-3p |       |       |       | 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 H2O 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 pepstatin, and 1 mM phenylmethylsulphonyl fluoride). Lysates were cleared by centrifugation (22,000 × g, 15 min at 4°C), and protein concentration was determined against a standard curve (BSA dilutions). Lysates were resolved for 1 h at 100 V and 10 cm on 4–12% Tris-Bis gel (NP0321PK2, Thermo Fisher Scientific, MA, USA) using 1 x 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.93 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, UK). 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 a 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...
Quality control, sequence alignment, and variant calling

Estimates of concordance and contamination for matched sample-normal (PBMC) pairs were performed using Conpair\(^6\) 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 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\(^92\) 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 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 arrow), and transforming growth factor β (TGF-β) pathway (white arrow).

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 (≥95%), duplicate mapped reads (≤25%), mean coverage (≥100x), and median coverage (≥50x) 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.
Identified variations were annotated with additional information from the UniProt database\textsuperscript{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).\textsuperscript{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.\textsuperscript{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.\textsuperscript{70} STRING 10 was used as an underlying global network for network-based gene prioritization.\textsuperscript{95} Identification of miRNAs potentially interacting with mutated UTRs was performed using information from the SomaMIR 2.0 database and ToppGene.\textsuperscript{94}

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.\textsuperscript{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 ≤ 0.05).**

**Data availability**

All primary data generated in this study are available in the supplementary 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 collaborative 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 MD Anderson Cancer Center and R.K. hold patents in the area of muscle invasive bladder cancer.
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. LeBlue, V.S., and Kalluri, R. (2020). Exosomes as a Multicomponent Biomarker Platform in Cancer. Trends Cancer 6, 767–774.

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

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

37. Allemson, K., Castillo, J., San Lucas, F.A., Secolo, 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. Oyen, E., Van Mol, K., Baggerman, G., Willems, H., Boonen, K., Rolfi, C., Pauwels, P., Jacobs, A., Schildermons, K., Choc, 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. Lorenzo-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, T.D., McKee, M., Song, J., and Skog, J. (2015). Nucleic acids within urinary exosomes/microvesicles are potential biomarkers for renal disease. Kidney Int. 78, 191–199.

42. Byrygnowa, O.E., Zaripov, M.M., Skvortsova, T.E., Lekchnov, E.A., Grigor’eva, A.E., Zaporozhchenko, I.A., Meroktin, E.S., Ryabchikova, E.L., Yurchenko, Y.B., Voitsitskaya, 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. Stranks, 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. Borojevic, D., and Hainaut, P. (2018). Urinary Exosome and Cell-Free DNA: A Novel Biomarker for Early Detection of Colon Cancer. Cancer Prev. 20, 2219–2224.

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

46. Lewis, D.A., Brown, R., Williams, J., White, P., Jacobson, S.K., and Habib, M. (2014). Integrated Double-Filtration Microfluidic Device for Detection of Extracellular Vesicles from Urine for Bladder Cancer Diagnosis. Methods Mol. Biol. 1660, 355–364.

47. Lewis, D.A., Brown, R., Habib, M., and Habib, M. (2014). The human urinary microbiome: bacterial DNA in voided urine of asymptomatic adults. Front. Cell. Infect. Microbiol. 4, 136.

48. 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.

49. 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.

50. Gimm, O., Dziema, H., Brown, J., Hoang-Vu, C., Hinze, R., Drahle, H., Mulligan, L.M., and Eng, C. (2001). Over-representation of a germline variant in the gene encoding RET co-receptor GFRα1 but not GFRα2 or GFRα3 in cases with sporadic medullary thyroid carcinoma. Oncogene 20, 2161–2170.

51. Peng, M., Bakker, J.L., DiCicco, R.A., Gillis, J.J., Zhao, H., Oduseni, K., Susek, T., Jaafer, L., Mivechi, N.F., Waisfisz, Q., and Ko, L. (2013). Inactivating Mutations in G1T189 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 Marks, J.M. (2014). Rare variants in TP53 and susceptibility to neuroblastoma. J. Natl. Cancer Inst. 106, djt067.
68. Poulos, R.C., Sloane, M.A., Hesson, I.B., and Wong, I.W.H. (2015). The search for cis-regulatory driver mutations in cancer genomes. Oncotarget 6, 32509–32525.
69. Diederichs, S., Bartlsch, L., Berkman, J.C., Fröse, K., Heitmann, I., Hoppe, C., Iggens, D., Jazmati, D., Karschena, 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, 355–350.
72. Roy, S., Pradhan, D., Ernst, W.L., Mercurio, S., Najjar, Y., Parikh, R., Parwani, A.V., Reim, 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 IFG2 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/beta-catenin signaling pathway. Oncotarget 8, 77241–77253.
81. Castellone, M.D., and Laukkanen, M.O. (2017). TGF-β/beta, WNT, and SHH signaling in tumor progression and in fibrotic diseases. Front. Biosci. (Schol. Ed.) 9, 31–45.
82. Macheda, M.L., and Stacke, 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, J.V., Jonsson, 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., LeBlouc, 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., Vacc, V., and Zody, M.C. (2016). Compair: concordance and contamination estimator for matched tumor-normal pairs. Bioinformatics 32, 3196–3198.
87. Madduri, R.K., Sulakhe, D., Lacinska, L., Liu, B., Rodriguez, A., Chard, K., Dave, U.J., and Foster, I.T. (2014). Experiences Building Globsus 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., Harl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shrik, 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., et al. (2010). ANNOVAR: functional annotation of genome variation data. Nucleic Acids Res. 38, e164.
91. 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, 3198–3207.
92. Bhattacharya, A., and Cui, Y. (2016). SomamiR 2.0: a database of cancer somatic mutation and non-coding RNA and synonymous mutations. EMBO Mol. Med. 8, 3198–3207.
93. Szklarczyk, D., Morris, J.H., Cook, H., Czupnik, G., Kudla, Z., Morone, R., Tramitz, H., et al. (2017). The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. Nucleic Acids Res. 46, D451–D456.
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–W311.
95. McDuck, E.A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garinella, K., Altshuler, D., Gabrail, 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.
96. Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 38, e164.
97. Pundir, S., Martin, M.J., and O’Donovan, C. (2017). UniProt Protein Knowledgebase. Methods Mol. Biol. 1558, 41–55.
98. The UniProt Consortium (2017). UniProt: the universal protein knowledgebase. Nucleic Acids Res. 45 (D1), D158–D169.
99. Chen, J., Bardes, E.E., Aronow, B.J., and Heg, A.G. (2009). ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. Nucleic Acids Res. 37, W305–W311.
100. Szklarczyk, D., Morris, J.H., Cook, H., Kuhn, M., Wyder, S., Sanschagrin, P., et al. (2015). The STRING database in 2015: protein-protein association networks, made broadly accessible. Nucleic Acids Res. 43, D406–D415.
101. Bhattacharya, A., and Cui, Y. (2016). SomamiR 2.0: a database of cancer somatic mutations altering microRNA-ceRNA interactions. Nucleic Acids Res. 45 (D1), D1005–D1010.
102. 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.