

Detection of Prostate Cancer-Specific Transcripts in Extracellular Vesicles Isolated from Post-DRE Urine

Kathryn L. Pellegrini^{1,2*}, Dattatraya Patil^{1,2}, Kristen J.S. Douglas^{1,2}, Grace Lee^{1,2}, Kathryn Wehrmeyer^{1,2}, Mersiha Torlak^{1,2}, Jeremy Clark³, Colin S. Cooper³, Carlos S. Moreno^{2,4-5}, Martin G. Sanda^{1,2*}

¹Department of Urology, Emory University School of Medicine, Atlanta, GA, USA

²Winship Cancer Institute, Emory University, Atlanta, GA, USA

³Norwich Medical School, University of East Anglia, Norwich, UK

⁴Department of Pathology and Laboratory Medicine, Emory University School of Medicine, Atlanta, GA, USA

⁵Department of Biomedical Informatics, Emory University School of Medicine, Atlanta, GA, USA

*Correspondence to: Kathryn L. Pellegrini, Ph.D. or Martin G. Sanda, M.D., Department of Urology, Emory University School of Medicine, 1365 Clifton Road NE, Building B, Suite 1400, Atlanta, GA 30322, USA. Email: kathryn.pellegrini@emory.edu; martinsanda@emory.edu

Running title: Prostate Cancer Biomarkers in Urine Vesicles

ABSTRACT

BACKGROUND: The measurement of gene expression in post-digital rectal examination (DRE) urine specimens provides a non-invasive method to determine a patient's risk of prostate cancer. Many currently available assays use whole urine or cell pellets for the analysis of prostate cancer-associated genes, although the use of extracellular vesicles (EVs) has also recently been of interest. We investigated the expression of prostate-, kidney-, and bladder-specific transcripts and known prostate cancer biomarkers in urine EVs.

METHODS: Cell pellets and EVs were recovered from post-DRE urine specimens, with the total RNA yield and quality determined by Bioanalyzer. The levels of prostate, kidney, and bladder-associated transcripts in EVs were assessed by TaqMan qPCR and targeted sequencing.

RESULTS: RNA was more consistently recovered from the urine EV specimens, with over 80% of the patients demonstrating higher RNA yields in the EV fraction as compared to urine cell pellets. The median EV RNA yield of 36.4 ng was significantly higher than the median urine cell pellet RNA yield of 4.8 ng. Analysis of the post-DRE urine EVs indicated that prostate-specific transcripts were more abundant than kidney- or bladder-specific transcripts. Additionally, patients with prostate cancer had significantly higher levels of the prostate cancer-associated genes *PCA3* and *ERG*.

CONCLUSIONS: Post-DRE urine EVs are a viable source of prostate-derived RNAs for biomarker discovery and prostate cancer status can be distinguished from analysis of these specimens. Continued analysis of urine EVs offers the potential discovery of novel biomarkers for pre-biopsy prostate cancer detection.

KEY WORDS: Prostate Cancer, Biomarkers, *PCA3*, *ERG*

INTRODUCTION

Urinary biomarkers for prostate cancer are highly sought after for their potential to discover and track disease progression non-invasively. Currently, the most well-characterized urine biomarker for prostate cancer is *PCA3*, which is a prostate-specific non-coding RNA with increased expression in prostate cancer (1,2). When combined with additional clinical parameters such as prostate volume, digital rectal examination (DRE) findings, and serum PSA, the FDA-approved *PCA3* urine assay can improve the detection of prostate cancer (3). However, the low sensitivity of the *PCA3* assay has encouraged continued investigation to identify additional urinary biomarkers (4). This is demonstrated by the pairing of urine assays for *PCA3* and *TMPRSS2:ERG*, which improves the sensitivity and specificity over that achieved by either assay alone (5-7). However, it is likely that multiplexing of additional genes predictive of prostate cancer risk will be necessary before a urine biomarker assay has the level of performance to be used in decisions prior to biopsy.

Biomarker discovery in urine has been complicated by the difficulty in obtaining sufficient quantities of intact RNA for whole transcriptome analyses. The initial identification of *PCA3* and *TMPRSS2:ERG* as prostate cancer biomarkers was based on analysis of prostatectomy tissue, while the urine assays were developed from analysis of cell pellets and whole urine (2,8,9). Another component of urinary RNAs that have recently been of interest are extracellular vesicles (EVs), which encompass microvesicles, apoptotic bodies, and exosomes (10,11). Previous investigations of EVs in the context of prostate cancer have demonstrated the presence of known prostate cancer-associated transcripts such as *KLK3*, *PCA3*, and *TMPRSS2:ERG* within urinary EVs (12-14). A recently developed prostate cancer detection test based on the expression of *PCA3* and *TMPRSS2:ERG* in urine EVs has established the potential to distinguish patients with and without prostate cancer based on their urine EV

transcript expression levels (15). In our current study, we have investigated the RNA content of cell pellets and EVs recovered from urine specimens to assess their suitability for transcriptome-wide analysis and biomarker discovery.

MATERIALS AND METHODS

Patient Cohort

Cohort participants were patients at primary care and urology clinics of Emory Healthcare, which includes The Emory Clinic, Emory University Hospital, Emory University Hospital Midtown, and Emory Saint Joseph's Hospital. Eligible patients provided informed written consent to provide specimens for our research study as part of our EDRN Clinical Validation Center cohort and the Winship Cancer Institute Prostate Cancer Tissue Bank. Demographic and clinical data were obtained through interviewer-administered questionnaires and Emory Healthcare's medical records database, and are summarized in Table 1. This study was conducted with the approval of the institutional review board of Emory University.

Urine Collection and Processing

For patients undergoing prostate biopsy or radical prostatectomy, all specimens (including serum for PSA testing) were collected prior to these procedures. Following an attentive DRE performed by a clinical care provider, each patient provided a first-catch urine sample directly into 30 mL specimen containers. All urine specimens were transported on crushed ice to the laboratory for processing within 4 hours of collection. Urine specimens were centrifuged for 5 minutes at 700g at 6°C to pellet the cells and the urine supernatant was transferred to a 50mL conical tube for further processing. The urine cell pellet was transferred to a non-stick microfuge tube, snap frozen, and stored at -80°C. At this point, urine supernatants processed with the Post-Freeze Filtration protocol had 0.5M EDTA added to a final concentration of 40mM and

were stored at -80°C with the filtration processing to be performed at a later date (generally within two weeks of collection), while urine supernatants processed using the Same Day Filtration protocol continued through the following procedure on the day of collection (Supplementary Figure 1A). Urine supernatants were centrifuged at 2000g for 5 minutes at room temperature to remove cell debris before being filtered through a 0.8 µm syringe filter. The filtered supernatant specimens were then centrifuged through an Amicon Ultra-15 Centrifugal Filter Unit with Ultracel-100 membrane at 3400g at 21°C for 10 minutes. The flow-through was discarded and the centrifugation step repeated until the entire sample had been passed through the filter unit. The filter unit was then washed with 15 mL of PBS and centrifugation at 3400g at 21°C for 10 minutes. The concentrated sample collected in the filter unit, which contained the EVs, was transferred to a non-stick microfuge tube and the volume was made up to 200 µL with PBS if required. The filter unit was rinsed with 700 µL of RLT Buffer with DTT (RLT/DTT prepared according to instructions in the Qiagen RNeasy Micro Kit) and this was also added to the microfuge tube containing the EVs. The total recovered volume of EV sample plus RLT/DTT solution was measured and 100% ethanol was added to a final concentration of 35%. The samples were then vortexed to disrupt the vesicles and either stored at -80°C if RNA extraction was to be performed at a later date (samples processed with the Same Day Filtration protocol; Supplementary Figure 1A), or they were immediately carried through to the RNA isolation protocol below (samples processed with the Post-Freeze Filtration protocol; Supplementary Figure 1A).

RNA Isolation

Total EV RNA was extracted using the Qiagen RNeasy Micro Kit and total urine cell pellet RNA was extracted using the Qiagen AllPrep DNA/RNA Micro Kit. Each of these methods were carried out according to manufacturer's instructions, and included an on-column DNase I treatment. To elute the RNA, the spin column was transferred to a non-stick microfuge tube

containing 1 μL of GlycoBlue solution (1 $\mu\text{g}/\mu\text{L}$), and then 20 μL of RNase-free water (pre-warmed to 45°C) was added directly to the spin column membrane. Following incubation at room temperature for 5 minutes, the spin column was centrifuged for 1 minute at 20000g. The Agilent RNA Pico Kit and Bioanalyzer were used to determine RNA concentration and quality.

Transmission Electron Microscopy (TEM)

A 5 μL aliquot of each urine EV sample was deposited onto a 400 mesh carbon coated copper grid that had been glow-discharged for 20 seconds. Grids containing EV samples were placed in a covered glass dish to allow vesicles to settle. After 5 minutes, each grid was washed by briefly touching the sample side to three drops of de-ionized water. Excess water was then removed from the grid with filter paper, and the sample was stained with 1% phosphotungstic acid (pH 6.5) for 20 seconds. Phosphotungstic acid was removed with filter paper and the samples were left to air dry for 10 minutes. Imaging was done on a JEOL JEM-1400 transmission electron microscope equipped with a 2k x 2k Gatan US1000 CCD camera.

Gene Expression Analysis by TaqMan qPCR

The transcripts of interest were analyzed using TaqMan Gene Expression Assays (details included in Supplementary Table 1). Up to 20 ng of RNA (average: 13.1 ng, range: 4.6-20 ng) was converted into cDNA with the NuGEN Ovation PicoSL WTA System V2, and the reaction product was purified with the Qiagen MinElute Reaction Cleanup Kit. The cDNA yield for each sample was measured on a NanoDrop using the ssDNA setting, which indicated an average cDNA yield of 4.8 μg (range: 3.2-6.5 μg). TaqMan Gene Expression Assays were set up using TaqMan Universal Master Mix II (without Uracil N-Glycosylase) with 20 ng of cDNA/reaction. Data was normalized to *RAB7A*. Cycle threshold values for non-detected assays were set to 40.

Measurement of Gene Expression by Precise Assay

Gene expression was also assessed using the multiplex Precise Assay technology developed by Cellular Research. In this method, each RNA transcript is labeled with a unique molecular index during reverse transcription allowing measurement of the number of transcripts in the original starting material (16,17). Each reaction was performed using 1 ng of RNA with each specimen run in triplicate. Amplification was conducted for a targeted set of genes according to manufacturer instructions, which includes 15 cycles of first-round amplification followed by 12 cycles of nested, second-round, amplification (approximately 300 different genes were targeted in this analysis). The libraries were sequenced at the HudsonAlpha Genomic Services Lab to produce 150 bases of paired-end reads with an average of over 300000 reads mapped per sample. The 150 base sequence length is sufficient to determine gene identity from the initial read, with sample and molecular index information derived from the paired read. Sequence data was processed using the Cellular Research Precise Analysis Pipeline v2.4 on the Seven Bridges Genomics platform and molecular index counts were used to perform expression analysis. Molecular index counts were retained if the average over the three technical replicates was at least 5, and the data was normalized to the reference gene, *RAB7A*. When fewer than 5 molecular index counts were detected, the relative expression was set to 0.

Statistical Analysis

Descriptive statistics of variables focused on medians, interquartile ranges (IQR), frequencies, and proportions as appropriate. Significance was determined using Wilcoxon rank sum test, Welch's two-sample t-test, or ANOVA, with all analysis and plotting conducted in R v3.2. The boxplots show the median and IQR, with 1.5x IQR represented by whiskers and outliers marked with dot points.

RESULTS

Post-DRE Urine Extracellular Vesicles Contain More RNA than Urine Cell Pellets

We began our assessment of urinary RNAs by characterizing the yield and quality of RNA isolated from the cell pellet and EV fractions of post-DRE urine collected from 105 patients. For the majority of patients (86 patients, 81.9%), more RNA was recovered from the EVs than from the cell pellet (Figure 1A). RNA yields from the cell pellets were typically very low, with the majority of the cell pellets producing less than 10 ng of RNA (74 specimens, 70.5%), which included 23 specimens (21.9%) where less than 1 ng of RNA was recovered. In contrast, very few of the EV specimens had less than 1 ng of RNA (2 specimens, 1.9%), and only 24 (22.9%) produced less than 10 ng of RNA.

Overall, the median EV RNA yield of 36.4 ng (IQR: 11.6-80.9 ng) was significantly higher than the median cell pellet RNA yield of 4.8 ng (IQR: 1.3-13.5 ng; Wilcoxon rank sum test $p < 0.001$; Figure 1B). RNA quality as measured by RNA Integrity Number (RIN) was also generally higher for the EV specimens, as the majority of EV specimens (89 specimens, 84.8%), but fewer than half of the cell pellet specimens (46 specimens, 43.8%) had a RIN ≥ 5 (Figure 1C). The samples with low RINs included 17 cell pellets and 3 EV specimens that were too low in concentration or quality for a RIN to be determined (Figure 1C). The Bioanalyzer profiles for EV and cell pellet RNAs both demonstrated the presence of 18S and 28S rRNAs, confirming that the detected RNA was not from bacterial contaminants in the urine (Supplementary Figure 2). After focusing our attention on the EV specimens, we investigated the relationship between RNA yield and cancer status, finding no significant differences in EV RNA yields for patients with no evidence of disease, or low risk, intermediate risk, or high risk prostate cancer (ANOVA $p = 0.275$; Figure 1D). Characterization of the EV fraction of post-DRE urine by TEM demonstrated the presence of larger microvesicles (~200 nm) and smaller particles likely to be exosomes (30-100 nm, Figure 2).

RNA from Extracellular Vesicles Can Still be Recovered After Freezing

To improve the feasibility of EV assays for clinical sites without on-site laboratory facilities we investigated whether EVs could be recovered from frozen urine supernatants, allowing the more time-consuming filtration step to be carried out later. For 82 of the specimens, filtration was performed on the same day as collection (Same Day Filtration), while 23 of the specimens were frozen after removing the cell pellet and EV RNA isolation was carried out separately (Post-Freeze Filtration; Supplementary Figure 1A). The median RNA yield was 36.1 ng (IQR: 11.6-79.3 ng) for specimens processed with the Same Day Filtration protocol, which was not significantly different from the 47.0 ng (IQR: 12.6-95.2 ng) median EV RNA yield from the Post-Freeze Filtration protocol (Wilcoxon rank sum test $p = 0.667$; Supplementary Figure 1B). For the majority of samples, the time between freezing and EV/RNA isolation was less than one week.

Extracellular Vesicle RNA Yields are Significantly Lower in Urine Collected Without DRE

We also determined that the application of a DRE immediately prior to urine collection was necessary for the recovery of sufficient EV RNA. Urine collections from 36 consecutive patients were randomly performed either with or without a prior DRE. The median EV RNA yield from post-DRE urine was 36.9 ng (IQR: 11.6-79.3 ng), which was similar to that observed in the larger cohort described above (Supplementary Figure 3). However, urine collected without a DRE had a significantly lower median EV RNA yield of 3.1 ng (IQR: 1.7-4.0 ng; Wilcoxon rank sum test $p < 0.001$; Supplementary Figure 3).

Post-DRE Urinary Extracellular Vesicles are Enriched for Prostate-Specific Transcripts

As urine may contain molecules originating from tissues throughout the genitourinary system, we used TaqMan qPCR to assess genes considered to have prostate-, kidney-, or bladder-specific expression to better understand the relative contribution of each of these tissues. These genes were selected based on a previous study that performed whole transcriptome RNA sequencing on EVs from urine collected without a prior DRE to investigate a range of transcripts

representing different regions of the kidney and bladder (18). Overall, the prostate-specific transcripts had much higher expression levels than any of the kidney- or bladder-specific transcripts we assessed. The highest expressed transcript was *HOXB13* at a median normalized expression level of 0.058 (IQR: 0.042-0.076), followed by *KLK2* (median normalized expression of 0.025, IQR: 0.019-0.035), and *KLK3* (median normalized expression of 0.016, IQR: 0.009-0.025; Figure 3). The lowest expressed prostate-specific transcript we assessed was *TRPM8*, which had a median normalized expression level of 0.005 (IQR: 0.004-0.007). In contrast, the highest expressed of the kidney- and bladder-specific transcripts, *UPK2*, had a median normalized expression of 0.002 (IQR: 0.001-0.004; Figure 3). This is over 2-fold lower than the level observed for *TRPM8*, and almost 30 times lower than *HOXB13*. The values for all other kidney- and bladder-specific transcripts we assessed were less than 0.001, with a number of them not detectable in all 60 samples (*AQP2*, *CUBN*, *NPHS2*, and *SLC34A1* were detected in 52, 59, 40, and 58 specimens, respectively).

Known Prostate Cancer Biomarkers Are Detectable in Post-DRE Urine Vesicles

To determine whether gene expression in post-DRE urine EVs differed between patients with and without prostate cancer, we investigated the expression of the known prostate cancer-associated genes *KLK3*, *PCA3*, and *ERG* with TaqMan qPCR. No difference in *KLK3* expression was observed when comparing patients with no evidence of disease (N.E.D.) to patients with low risk (GS6) or aggressive (GS7+) prostate cancer (ANOVA $p = 0.672$; Figure 4A). *PCA3* was also robustly detected in all patients, with significantly higher levels observed in patients with prostate cancer. Mean expression of *PCA3* was 4.5-fold and 4.8-fold higher for patients with GS6 and GS7+ prostate cancer, respectively, as compared to patients with no evidence of disease (Welch's two-sample t-test $p = 0.002$ and $p < 0.001$; Figure 4B). However, mean *PCA3* expression was not significantly different when comparing the GS6 and GS7+ groups (1.1-fold, Welch's two-sample t-test $p = 0.831$; Figure 4B). Patients with GS7+ prostate

cancer also had significantly higher levels of *ERG* as compared to patients with no evidence of disease (10.5-fold, Welch's two-sample t-test $p = 0.027$; Figure 4C). Patients with GS6 prostate cancer showed a 6.4-fold increase over patients with no evidence of disease, although this was not significant (Welch's two-sample t-test $p = 0.116$). There was also no significant difference between the GS6 and GS7+ patient groups (1.6-fold, Welch's two-sample t-test $p = 0.433$; Figure 4C).

Prostate-Specific Transcripts Are Also Detectable Via Multiplex Precise Assay

After performing the individual assays, we were interested in investigating a novel method for multiplex transcript analysis that could expand our ability to discover and detect urinary RNAs. The Precise Assay technology encodes each RNA transcript with a unique molecular index during reverse transcription before any amplification has occurred. Analysis of these molecular index sequences is intended to distinguish the number of molecules in the starting material from PCR duplicates arising during amplification (16,17). A comparison of the raw data generated by the individual TaqMan qPCR assays and the multiplex Precise Assay is provided in Supplementary Figure 4.

In agreement with the TaqMan qPCR analysis, the multiplex Precise Assay demonstrated that the prostate-specific transcripts, *KLK3*, *KLK2*, *HOXB13*, and *TRPM8*, were more abundantly expressed than the kidney- or bladder-specific transcripts (Figure 5A). However, the highest expressed transcript detected in the Precise Assay was *KLK3*, at approximately 10-fold higher abundance than the next highest expressed transcript, *KLK2*. In contrast, *KLK3* and *KLK2* had very similar expression levels when assessed by TaqMan qPCR, where the highest expressed transcript was *HOXB13* (Figure 3). Analysis of the read-to-molecule ratio of the prostate-associated genes and *RAB7A* indicated that most of the transcripts were amplified at similar levels, with approximately 10 sequence reads for each molecule. However, *HOXB13* showed a

relatively low ratio (3.8 reads/molecule) suggesting that amplification of this transcript in the Precise Assay was less efficient than the other transcripts (Figure 5D). The Precise Assay also demonstrated very low expression of the kidney- and bladder-specific transcripts, with none being detected in all 60 samples. In particular, *NPHS2*, *PTH1R*, and *SLC34A1* were not detected in any sample and are shown with a normalized expression of zero (Figure 5A).

When we focused our analysis on the prostate cancer-associated genes, the largest difference between the TaqMan and Precise approaches was that *ERG* could not be detected by Precise Assay. However, the Precise Assay demonstrated similar patterns of expression to TaqMan qPCRs when assessing differences in *KLK3* and *PCA3* amongst each of the patient groups. While *KLK3* was very highly expressed, there was no significant difference between patients with GS6 or GS7+ prostate cancer when compared to patients with no evidence of disease (ANOVA $p = 0.864$; Figure 5B). The rate of detection for *PCA3* was lower by Precise Assay, where it was only detected in 50% of the patients with no evidence of disease (6 out of 12), 71% of the patients with GS6 prostate cancer (10 out of 14), and 77% of patients with GS7+ prostate cancer (20 out of 26). Mean expression of *PCA3* was 2.5-fold and 3.3-fold higher for patients with GS6 and GS7+ prostate cancer, respectively, as compared to patients with no evidence of disease, although only the increase in the GS7+ group was statistically significant (Welch's two-sample t-test $p = 0.052$ and $p = 0.003$; Figure 5C). As observed with the TaqMan qPCR assay, there was no significant difference in *PCA3* expression between patients with GS6 and GS7+ prostate cancer (1.4-fold, Welch's two-sample t-test $p = 0.327$; Figure 5C).

DISCUSSION

While urine assays for *PCA3* and *TMPRSS2:ERG* have provided more options and improved detection for men at risk of prostate cancer, the discovery of additional urinary biomarkers is needed to improve the sensitivity and specificity of this approach before it could have a

meaningful role in pre-biopsy decision-making. This study was focused on understanding how to recover RNAs from post-DRE urine for comprehensive discovery studies. Our initial findings established that urine EVs had significantly higher quantities of intact RNA than urine cell pellets, regardless of the cancer-status of the patients. In particular, usable amounts of RNA were most reliably recovered from the EV fraction of post-DRE urine, indicating that urine EVs are more suitable for further expression analyses. Additionally, during our optimization of the urine EV isolation protocol, we found the yield of EV RNA to be unaffected following frozen storage of urine supernatants, suggesting that urine EVs are stable at low storage temperatures and during freeze-thaw. This is further supported by previous observations that exosomes and microvesicles remain intact throughout repeated freeze-thaw cycles (19-21). This demonstrates the potential for implementation of urine EV-based assays at clinics lacking on-site laboratory facilities, and the possibility of isolating urine EVs from archived urine supernatants.

Our analysis of EV RNA also demonstrated the high quality of the RNA, which included the 18S and 28S rRNA peaks. Previous studies investigating RNA recovered from urine exosomes and vesicles have not established a consensus on the expected EV RNA profile. While many have shown that urine EVs can contain the full complement of RNAs, with small RNAs, mRNAs, and rRNAs all present (18,22,23), a number of other studies have only observed small RNAs in urine EVs, with rRNAs completely absent (24,25). In one of these studies, rRNAs that were originally assumed to be eukaryotic were actually found to be 16S and 23S prokaryotic rRNA peaks from contaminating bacteria in the urine (25). We have confirmed that the rRNAs in our post-DRE urine EV specimens are eukaryotic, indicating that mixed urine EV populations can contain the full complement of RNAs and that contaminating prokaryotic species are unlikely to have contributed significantly to our observed RNA yields. However, it would not be appropriate to use our results to demonstrate the relative presence of small RNAs to larger RNA species as

the RNA extraction method we have used is not intended for the recovery of small RNAs (< 200 nucleotides).

One of the complications of urine-based biomarkers is the possibility for genes expressed by other tissues in the genitourinary tract to contribute to the detected signal. However, we found prostate-specific genes to be present at much higher levels than any of the kidney- or bladder-specific genes indicating that urinary EVs are enriched for prostate-specific RNAs. The urine in this study was collected following a DRE, which may have increased the presence of prostate-derived cells/vesicles/molecules in the urine. This is supported by a number of previous studies that have demonstrated increased expression of prostate cancer biomarkers, such as *KLK3*, *PCA3*, and *TMPRSS2:ERG*, in urine following a DRE (12,13). However, a similar abundance of prostate-associated RNAs has been observed in exosomes isolated from healthy donor urine collected without a prior DRE (18), suggesting that other factors other than the DRE may be responsible for this effect. Another recent study found significantly increased levels of the exosomal markers CD9 and CD63 in urine collected following a DRE as compared to urine collected without a DRE, while there were significantly decreased levels of these exosomal markers in urine collected from patients that had undergone a prostatectomy or from women (26). This indicates that the prostate may contribute a large proportion of the vesicles found in urine of men, and that an enrichment of prostate-derived transcripts would be expected. Further characterization with tissue-specific surface markers could provide further insight into the origins of urinary EVs. This would be of particular importance for biomarker discovery in kidney or bladder conditions, where an abundance of prostate-derived EVs may cause urine EV biomarker assays to perform very differently for men and women.

Having established post-DRE urine EVs as a viable source of prostate-derived RNAs, we next investigated the detection of prostate cancer-associated transcripts. *PCA3* and *ERG* expression was significantly higher in patients with prostate cancer as compared to those without prostate cancer. This is in agreement with previous studies, which have demonstrated that these genes are expressed at higher levels in the urine vesicles of patients with prostate cancer and substantiates the rationale for using post-DRE urine EV RNA as a biomarker source (15,27). *KLK3* was very highly expressed in urine EVs from all patients, but showed no significant difference between the three patient groups. Given that *KLK3* is not specific for cancer and is currently used as a normalizer gene for the *PCA3* and *TMPRSS2:ERG* urine assays (3), this consistent expression across all of the patient groups fits with its expected performance and supports the validity of studying urine EVs.

Our experience with the Precise Assay indicated that there was some loss of detection sensitivity when using the multiplex platform, as opposed to the robust detection in all or most of the specimens when using individual TaqMan qPCR assays. In particular, none of the kidney and bladder transcripts could be detected in all 60 specimens when using the Precise Assay, and a number of the transcripts (*NPHS2*, *PTH1R*, and *SLC34A1*) were not detected in any sample with this method. The number of specimens with detectable *PCA3* and *ERG* was also reduced when using the Precise Assay. There are a number of possible reasons for the differences in sensitivity of the two methods, although a large contributing factor may be the low RNA input (1 ng) used in the Precise Assay. While being able to use less RNA is appealing due to ensuring that more specimens are able to meet the required input, it does introduce complications when trying to detect transcripts of lower abundance, with lack of detection and high variability between replicates becoming more likely. Another factor that may have influenced the comparative performance of the two assays are the relative efficiencies of the reverse transcription and amplification for each method. This may be best demonstrated by the

difference in the apparent abundance of *HOXB13*, which was less efficiently amplified by the Precise Assay than the other prostate-specific transcripts. Future iterations of this assay will need to be carefully designed to ensure that all transcripts amplified concurrently can be adequately and accurately detected.

Although this study was limited to known urine biomarkers while we focused on characterizing the RNA and investigating the implementation of this technique, the Precise Assay technology that we have used is capable of detecting up to 150 genes simultaneously. Prognostic biomarker panels assessed in biopsy or radical prostatectomy tissue generally assess 20-30 genes to create an overall risk score (28-31) and the combination of the *PCA3* and *TMPRSS2:ERG* urine assays demonstrates that the performance of urine biomarker panels can also be improved with the addition of more genes (5-7). This proof of principle study has better characterized the RNA recovered from post-DRE urine EVs, and demonstrated the feasibility of performing multiplex gene expression analyses in this specimen type. We are currently expanding our analysis to include additional prostate cancer-associated genes that have not previously been investigated in urine. The ability to recover sufficient urinary RNA to conduct comprehensive gene expression analysis is likely to accelerate the development of urine biomarker panels for the accurate determination of prostate cancer risk prior to biopsy.

CONCLUSIONS

We have demonstrated that post-DRE urine EVs are a more reliable source of RNA than urine cell pellets, that prostate-specific transcripts are enriched in these EVs, and that the current urinary prostate cancer biomarkers, *PCA3* and *ERG*, are present at significantly higher levels in the urine EVs of prostate cancer patients. These findings indicate that prostate cancer status is reflected in urinary EVs, and that continued analysis of urine EVs offers the potential to discover additional biomarkers for prostate cancer detection.

ACKNOWLEDGEMENTS

The authors would like to thank Dr. Rebecca Arnold and Dr. John Petros for their leadership of the Winship Cancer Institute Prostate Cancer Tissue Bank. Funding from the National Cancer Institute Early Detection Research Network (U01 CA113913) and the Movember Foundation (part of the GAP1 Global Prostate Cancer Biomarker Initiative) supported this study. Research reported in this publication was also supported in part by the Emory Integrated Genomics Core Shared Resource of the Winship Cancer Institute of Emory University (NIH/NCI award P30 CA138292), and the Robert P. Apkarian Integrated Electron Microscopy Core (subsidized by the Emory College of Arts and Sciences and Emory University School of Medicine with additional support provided by NIH UL1 TR000454) with data described here gathered on the JEOL JEM-1400 120kV TEM (supported by NIH S10 RR025679). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The protocols for urine processing, EV purification, and RNA extraction from cell pellet and EVs were provided by the Cooper lab at the University of East Anglia, UK as part of the Movember Foundation GAP1 Global Prostate Cancer Biomarker Initiative.

CONFLICT OF INTEREST

The authors have no conflicts to disclose.

REFERENCES

1. Bussemakers MJ, van Bokhoven A, Verhaegh GW, Smit FP, Karthaus HF, Schalken JA, Debruyne FM, Ru N, Isaacs WB. DD3: a new prostate-specific gene, highly overexpressed in prostate cancer. *Cancer Res* 1999;59(23):5975-5979.
2. Hessels D, Klein Gunnewiek JM, van Oort I, Karthaus HF, van Leenders GJ, van Balken B, Kiemeny LA, Witjes JA, Schalken JA. DD3(PCA3)-based molecular urine analysis for the diagnosis of prostate cancer. *Eur Urol* 2003;44(1):8-15.
3. Groskopf J, Aubin SM, Deras IL, Blase A, Bodrug S, Clark C, Brentano S, Mathis J, Pham J, Meyer T, Cass M, Hodge P, Macairan ML, Marks LS, Rittenhouse H. APTIMA PCA3 molecular urine test: development of a method to aid in the diagnosis of prostate cancer. *Clin Chem* 2006;52(6):1089-1095.
4. Wei JT, Feng Z, Partin AW, Brown E, Thompson I, Sokoll L, Chan DW, Lotan Y, Kibel AS, Busby JE, Bidair M, Lin DW, Taneja SS, Viterbo R, Joon AY, Dahlgren J, Kagan J, Srivastava S, Sanda MG. Can Urinary PCA3 Supplement PSA in the Early Detection of Prostate Cancer? *J Clin Oncol* 2014;32(36):4066-4072.
5. Tomlins SA, Aubin SM, Siddiqui J, Lonigro RJ, Sefton-Miller L, Miick S, Williamsen S, Hodge P, Meinke J, Blase A, Penabella Y, Day JR, Varambally R, Han B, Wood D, Wang L, Sanda MG, Rubin MA, Rhodes DR, Hollenbeck B, Sakamoto K, Silberstein JL, Fradet Y, Amberson JB, Meyers S, Palanisamy N, Rittenhouse H, Wei JT, Groskopf J, Chinnaiyan AM. Urine TMPRSS2:ERG fusion transcript stratifies prostate cancer risk in men with elevated serum PSA. *Sci Transl Med* 2011;3(94):94ra72.
6. Leyten GH, Hessels D, Jannink SA, Smit FP, de Jong H, Cornel EB, de Reijke TM, Vergunst H, Kil P, Knipscheer BC, van Oort IM, Mulders PF, Hulsbergen-van de Kaa CA, Schalken JA. Prospective multicentre evaluation of PCA3 and TMPRSS2-ERG gene fusions as diagnostic and prognostic urinary biomarkers for prostate cancer. *Eur Urol* 2014;65(3):534-542.
7. Tomlins SA, Day JR, Lonigro RJ, Hovelson DH, Siddiqui J, Kunju LP, Dunn RL, Meyer S, Hodge P, Groskopf J, Wei JT, Chinnaiyan AM. Urine TMPRSS2:ERG Plus PCA3 for Individualized Prostate Cancer Risk Assessment. *Eur Urol* 2016;70(1):45-53.
8. Laxman B, Tomlins SA, Mehra R, Morris DS, Wang L, Helgeson BE, Shah RB, Rubin MA, Wei JT, Chinnaiyan AM. Noninvasive detection of TMPRSS2:ERG fusion transcripts in the urine of men with prostate cancer. *Neoplasia* 2006;8(10):885-888.
9. Hessels D, Smit FP, Verhaegh GW, Witjes JA, Cornel EB, Schalken JA. Detection of TMPRSS2-ERG fusion transcripts and prostate cancer antigen 3 in urinary sediments may improve diagnosis of prostate cancer. *Clin Cancer Res* 2007;13(17):5103-5108.
10. Junker K, Heinzelmann J, Beckham C, Ochiya T, Jenster G. Extracellular Vesicles and Their Role in Urologic Malignancies. *Eur Urol* 2016;70(2):323-331.
11. Nawaz M, Camussi G, Valadi H, Nazarenko I, Ekstrom K, Wang X, Principe S, Shah N, Ashraf NM, Fatima F, Neder L, Kislinger T. The emerging role of extracellular vesicles as biomarkers for urogenital cancers. *Nat Rev Urol* 2014;11(12):688-701.
12. Dijkstra S, Birker IL, Smit FP, Leyten GH, de Reijke TM, van Oort IM, Mulders PF, Jannink SA, Schalken JA. Prostate cancer biomarker profiles in urinary sediments and exosomes. *J Urol* 2014;191(4):1132-1138.
13. Hendriks RJ, Dijkstra S, Jannink SA, Steffens MG, van Oort IM, Mulders PF, Schalken JA. Comparative analysis of prostate cancer specific biomarkers PCA3 and ERG in whole urine, urinary sediments and exosomes. *Clin Chem Lab Med* 2016;54(3):483-492.

14. Nilsson J, Skog J, Nordstrand A, Baranov V, Mincheva-Nilsson L, Breakefield XO, Widmark A. Prostate cancer-derived urine exosomes: a novel approach to biomarkers for prostate cancer. *Br J Cancer* 2009;100(10):1603-1607.
15. McKiernan J, Donovan MJ, O'Neill V, Bentink S, Noerholm M, Belzer S, Skog J, Kattan MW, Partin A, Andriole G, Brown G, Wei JT, Thompson IM, Jr., Carroll P. A Novel Urine Exosome Gene Expression Assay to Predict High-grade Prostate Cancer at Initial Biopsy. *JAMA Oncol* 2016;2(7):882-889.
16. Fu GK, Hu J, Wang PH, Fodor SP. Counting individual DNA molecules by the stochastic attachment of diverse labels. *Proc Natl Acad Sci U S A* 2011;108(22):9026-9031.
17. Fu GK, Xu W, Wilhelmy J, Mindrinos MN, Davis RW, Xiao W, Fodor SP. Molecular indexing enables quantitative targeted RNA sequencing and reveals poor efficiencies in standard library preparations. *Proc Natl Acad Sci U S A* 2014;111(5):1891-1896.
18. Miranda KC, Bond DT, Levin JZ, Adiconis X, Sivachenko A, Russ C, Brown D, Nusbaum C, Russo LM. Massively parallel sequencing of human urinary exosome/microvesicle RNA reveals a predominance of non-coding RNA. *PLoS One* 2014;9(5):e96094.
19. Yuana Y, Boing AN, Grootemaat AE, van der Pol E, Hau CM, Cizmar P, Buhr E, Sturk A, Nieuwland R. Handling and storage of human body fluids for analysis of extracellular vesicles. *J Extracell Vesicles* 2015;4:29260.
20. Jayachandran M, Miller VM, Heit JA, Owen WG. Methodology for isolation, identification and characterization of microvesicles in peripheral blood. *J Immunol Methods* 2012;375(1-2):207-214.
21. Sokolova V, Ludwig AK, Hornung S, Rotan O, Horn PA, Epple M, Giebel B. Characterisation of exosomes derived from human cells by nanoparticle tracking analysis and scanning electron microscopy. *Colloids Surf B Biointerfaces* 2011;87(1):146-150.
22. Crescitelli R, Lasser C, Szabo TG, Kittel A, Eldh M, Dianzani I, Buzas EI, Lotvall J. Distinct RNA profiles in subpopulations of extracellular vesicles: apoptotic bodies, microvesicles and exosomes. *J Extracell Vesicles* 2013;2:20677.
23. Miranda KC, Bond DT, McKee M, Skog J, Paunescu TG, Da Silva N, Brown D, Russo LM. Nucleic acids within urinary exosomes/microvesicles are potential biomarkers for renal disease. *Kidney Int* 2010;78(2):191-199.
24. Bryzgunova OE, Zaripov MM, Skvortsova TE, Lekchnov EA, Grigor'eva AE, Zaporozhchenko IA, Morozkin ES, Ryabchikova EI, Yurchenko YB, Voitsitskiy VE, Laktionov PP. Comparative Study of Extracellular Vesicles from the Urine of Healthy Individuals and Prostate Cancer Patients. *PLoS One* 2016;11(6):e0157566.
25. Tataruch-Weinert D, Musante L, Kretz O, Holthofer H. Urinary extracellular vesicles for RNA extraction: optimization of a protocol devoid of prokaryote contamination. *J Extracell Vesicles* 2016;5:30281.
26. Duijvesz D, Versluis CY, van der Fels CA, Vredenburg-van den Berg MS, Leivo J, Peltola MT, Bangma CH, Pettersson KS, Jenster G. Immuno-based detection of extracellular vesicles in urine as diagnostic marker for prostate cancer. *Int J Cancer* 2015;137(12):2869-2878.
27. Motamedinia P, Scott AN, Bate KL, Sadeghi N, Salazar G, Shapiro E, Ahn J, Lipsky M, Lin J, Hruby GW, Badani KK, Petrylak DP, Benson MC, Donovan MJ, Comper WD, McKiernan JM, Russo LM. Urine Exosomes for Non-Invasive Assessment of Gene Expression and Mutations of Prostate Cancer. *PLoS One* 2016;11(5):e0154507.
28. Cuzick J, Swanson GP, Fisher G, Brothman AR, Berney DM, Reid JE, Meshner D, Speights VO, Stankiewicz E, Foster CS, Moller H, Scardino P, Warren JD, Park J, Younus A, Flake DD, 2nd, Wagner S, Gutin A, Lanchbury JS, Stone S. Prognostic value of an RNA expression signature derived from cell cycle proliferation genes in patients with prostate cancer: a retrospective study. *Lancet Oncol* 2011;12(3):245-255.

29. Klein EA, Cooperberg MR, Magi-Galluzzi C, Simko JP, Falzarano SM, Maddala T, Chan JM, Li J, Cowan JE, Tsiatis AC, Cherbavaz DB, Pelham RJ, Tenggara-Hunter I, Baehner FL, Knezevic D, Febbo PG, Shak S, Kattan MW, Lee M, Carroll PR. A 17-gene Assay to Predict Prostate Cancer Aggressiveness in the Context of Gleason Grade Heterogeneity, Tumor Multifocality, and Biopsy Undersampling. *Eur Urol* 2014;66(3):550-560.
30. Erho N, Crisan A, Vergara IA, Mitra AP, Ghadessi M, Buerki C, Bergstralh EJ, Kollmeyer T, Fink S, Haddad Z, Zimmermann B, Sierocinski T, Ballman KV, Triche TJ, Black PC, Karnes RJ, Klee G, Davicioni E, Jenkins RB. Discovery and validation of a prostate cancer genomic classifier that predicts early metastasis following radical prostatectomy. *PLoS One* 2013;8(6):e66855.
31. Long Q, Xu J, Osunkoya AO, Sannigrahi S, Johnson BA, Zhou W, Gillespie T, Park JY, Nam RK, Sugar L, Stanimirovic A, Seth AK, Petros JA, Moreno CS. Global transcriptome analysis of formalin-fixed prostate cancer specimens identifies biomarkers of disease recurrence. *Cancer Res* 2014;74(12):3228-3237.

FIGURE LEGENDS

Figure 1. RNA yield from the pellet and extracellular vesicle (EV) fraction of post-DRE urine. (A) Urine was collected following a DRE and processed to isolate the cell pellet and EV fraction. For each sample, the amount of RNA extracted is represented as a stacked bar, with the amount isolated from the EVs shown in dark gray while the amount of RNA isolated from the cell pellet is shown in light gray (n = 105). The y-axis has been split to improve visualization of the low yield samples. (B) Summary data of the data in A is represented with boxplots (outliers are not shown). The RNA yield was significantly higher in the EVs ($p < 0.001$). (C) The frequency of RNA quality is shown for EVs (upper plot) and cell pellets (lower plot). The RIN scale goes from 1 to 10, with samples of particularly low quality or concentration unable to have RINs determined (shown as N/A). Representative Bioanalyzer plots are shown in Supplementary Figure 2. (D) The yield of EV RNA was not significantly different amongst patients with no evidence of disease (N.E.D.) or various levels of prostate cancer risk ($p = 0.275$; n = 34, 28, 36, and 11; one outlier in the GS7 group at 506 ng not shown).

Figure 2. Analysis of urinary extracellular vesicles (EVs) by TEM. The EV fractions recovered from two different patients were negatively-stained and imaged by TEM. The observed particles include larger microvesicles (arrows) as well as smaller vesicles likely to be exosomes (arrowheads; scale bar on both images represents 200 nm).

Figure 3. Prostate-specific genes are enriched in post-DRE urine EVs. The expression of various prostate-, kidney-, or bladder-specific transcripts were measured by TaqMan qPCR and are shown relative to the expression of *RAB7A* (n = 60). The boxplots for *AQP2*, *CUBN*, *NPHS2*, and *SLC34A1* include data for samples where no amplification was observed and the cycle threshold was set to 40 (8, 1, 20, and 2 samples, respectively).

Figure 4. Prostate cancer-associated RNAs are detectable and informative in post-DRE urine EVs. The expression levels of (A) *KLK3*, (B) *PCA3*, and (C) *ERG* were measured by TaqMan qPCR assay and normalized to *RAB7A*. The data is shown for patients with no evidence of disease (N.E.D., n = 12), Gleason score 6 prostate cancer (GS6, n = 14), or Gleason score ≥ 7 prostate cancer (GS7+, n = 26). The *ERG* plot includes data for 9 samples where no amplification was observed and the cycle threshold was set to 40 (2, 2, and 5 samples for the N.E.D., GS6, and GS7+ groups, respectively). In comparison to the N.E.D. group, the GS6 and GS7+ groups both had significantly higher expression of *PCA3* ($p = 0.002$ and $p < 0.001$, respectively), while only the GS7+ group had significantly higher expression of *ERG* ($p = 0.027$). There was no significant difference in *KLK3* expression between the any of the groups ($p = 0.672$).

Figure 5. Precise Assay measurement of transcripts also shows enrichment of prostate-specific and prostate cancer-associated transcripts. (A) The same prostate-, kidney-, and bladder-specific transcripts shown in Figure 3 were assessed by Precise Assay and normalized to *RAB7A* (n = 60). *KLK3* is shown on a different scale because it was present at much higher levels than the other transcripts. The boxplots for *AQP2*, *CUBN*, *NPHS2*, *PTH1R*, *SLC12A1*, *SLC12A3*, *SLC34A1*, *UPK1B*, and *UPK2* include data for samples where no transcripts were detected and the normalized expression was set to 0 (17, 7, 60, 60, 15, 54, 60, 17, and 22 samples, respectively). Transcripts for (B) *KLK3* and (C) *PCA3* were measured by Precise Assay and normalized to *RAB7A* (n = 12, 14, and 26 for the N.E.D., GS6, and GS7+ groups, respectively). The *PCA3* plot includes data for 14 samples where no transcripts were detected and normalized expression was set to 0 (6, 4, and 6 samples for the N.E.D., GS6, and GS7+ groups, respectively). GS7+ patients had significantly higher expression of *PCA3* as compared to the N.E.D. group ($p = 0.003$). (D) The ratio of sequence reads to molecular index count is

shown for all of the prostate-associated genes and the normalizer *RAB7A* (means and 95% confidence intervals shown).