PAPER

Check for updates

Assessment of urine sample collection and processing variables for extracellular vesicle-based proteomics†

Guiyuan Zhang, [a,b,c] Yajie Ding, [a] Hao Zhang, [a,c] Dong Wei, [a,b] Yufeng Liu, [a,b] Jie Sun, [a] Zhuoying Xie, [a] W. Andy Tao [*d] and Yefei Zhu [*a,e]

Extracellular vesicles (EVs) in urine are a promising source for developing non-invasive biomarkers. However, urine concentration and content are highly variable and dynamic, and actual urine collection and handling often is nonideal. Furthermore, patients such as those with prostate diseases have challenges in sample collection due to difficulties in holding urine at designated time points. Here, we simulated the actual situation of clinical sample collection to examine the stability of EVs in urine under different circumstances, including urine collection time and temporary storage temperature, as well as daily urine sampling under different diet conditions. EVs were isolated using functionalized EVtrap magnetic beads and characterized by nanoparticle tracking analysis (NTA), western blotting, electron microscopy, and mass spectrometry (MS). EVs in urine remained relatively stable during temporary storage for 6 hours at room temperature and for 12 hours at 4 °C, while significant fluctuations were observed in EV amounts from urine samples collected at different time points from the same individuals, especially under certain diets. Sample normalization with creatinine reduced the coefficient of variation (CV) values among EV samples from 17% to approximately 6% and facilitated downstream MS analyses. Finally, based on the results, we applied them to evaluate potential biomarker panels in prostate cancer by data-independent acquisition (DIA) MS, presenting the recommendation that can facilitate biomarker discovery with nonideal handling conditions.

1. Introduction

Extracellular vesicles (EVs) are membranous vesicles wrapped by lipid bilayers, which contain proteins, lipids, nucleic acids, metabolites, and others.[1–4] Initially believed to primarily serve the purpose of discarding unwanted substances from cells, EVs have been revealed to have vital roles in intercellular communication and are essential for normal physiological processes and pathological progression.[1,5,6] EVs are a heterogeneous population of particles, typically ranging in size from 30 to 1000 nm. During the process of EV biogenesis, EVs can encapsulate molecules that may reflect the physiological or pathophysiological state of their parental cells, and therefore EVs are a promising source of biomarkers for a variety of diseases. Compared with other fluid biopsy sources, EVs have multiple advantages: (i) EVs are present in nearly all body fluids and their cargo molecules exhibit high stability when enclosed within lipid bilayers; (ii) relatively a large number of EVs are present in most biofluids compared to circulating tumor cells (CTCs) or cell-free DNA (cfDNA); and (iii) EVs enrich functional molecules, effectively circumventing the complexity and wide dynamic range of biofluids. Various commercial products and new technologies have been developed for the separation of EVs from different biofluids. Common separation techniques include ultracentrifugation, density gradient separation, polymer-based precipitation, size exclusion, and affinity-based isolation.[7]

Urine sample storage and handling can impact the detection of analytes including EVs present in urine, while the use of different normalization methods may affect the measurement and potentially introduce biases. So far, several studies have demonstrated the stability of EVs in urine samples under various storage conditions. For example, Zhou *et al.* found that urine samples stored at −20 °C and −80 °C for 1 week or

a School of Biological Science & Medical Engineering, Southeast University, Nanjing 210096, China. E-mail: zhuyf@njmu.edu.cn, taow@purdue.edu
b Bell Mountain Molecular MedTech Institute, Nanjing 210032, China
c EVLiXiR Biotech, Nanjing 210032, China
d Departments of Chemistry and Biochemistry, Purdue University, West Lafayette, Indiana 47907, USA
e Laboratory Medicine Center, The Second Affiliated Hospital of Nanjing Medical University, Nanjing 210011, China
† Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d4an00296b

7 months showed different outcomes.[8] Freezing at −20 °C resulted in significant loss of exosomes, whereas minimal loss was observed after freezing at −80 °C. Riccardo Vago *et al.* showed that adding preservatives to urine can maintain its stability at room temperature for up to 6 months.[9] In addition, it has been reported that urine samples and isolated EVs can remain stable at −80 °C for at least 2 years.[10] However, most of these studies have focused on the impact of long-term storage on the stability of urine EVs. Currently, the short-term, temporary storage of urine samples in labs and clinics has not been systematically investigated in clinical applications. Furthermore, when collecting clinical samples, often it can be challenging to collect urine samples from patients such as those with prostate problems at specific time points. In order to solve the nonideal conditions encountered in the actual urine collection process, it is necessary to study the variations in EVs in daily urine collection and handling. Additionally, due to the highly dynamic nature of urine samples, it is important to normalize them during the sample preparation stage. Although several standardized methods have been proposed, such as exosome markers, nanoparticle tracking analysis (NTA) and Tamm–Horsfall protein,[11–15] they have drawbacks in practical applications. Therefore, a sensible and effective normalization strategy is needed to reduce the variability between samples.

In this study (Fig. 1), we first examined the stability of EVs in urine, by assessing factors such as nonideal urine collection, temporary storage temperature and short storage time, and diet. We examined several normalization methods and identified that creatine normalization is relatively effective and easier to implement to minimize variations. Finally, we profiled urine EV proteomes to screen potential biomarkers for prostate cancer (PCa), so as to provide valuable guidance for clinical applications. Currently, the detection methods for PCa mainly include serum prostate-specific antigen (PSA) measurement and digital rectal examination (DRE), with serum PSA measurement being the most widely utilized in clinical practice.[16–18] Due to the low specificity of both PSA serum measurements and DRE, a large number of unnecessary prostate biopsies are performed. Therefore, there remains a clear need for new biomarkers for the rapid and reliable diagnosis of PCa. Recent proteomic analyses based on mass spectrometry (MS) make it possible to detect, quantify and characterize cancer-specific EV biomarkers by high-throughput and in-depth proteomic analysis of urine EVs.[19] Currently, the common MS data acquisition methods include the data dependent acquisition (DDA) method and the data independent acquisition (DIA) method. Compared to DDA, DIA has several advantages, including higher coverage, sensitivity, and repeatability.[20–22] Building a library is typically required in advance for DIA data analysis, but recent advances in library-free DIA (directDIA) have allowed researchers to achieve direct peptide sequencing without library building.[23–25] In addition, compared to DDA technology, directDIA can achieve higher quantitative sensitivity, accuracy, and reproducibility.[26,27] In this study, we applied directDIA to profile urine EV proteomes from prostate cancer patients and control groups.

2. Experimental section

Detailed experimental materials, western blotting methods and EV proteome sample preparation are included in the ESI.†

2.1. Collection of samples

The urine samples were collected from the People's Hospital of Jiangsu Province. Urine samples were collected from healthy individuals, patients with prostate cancer and other patients
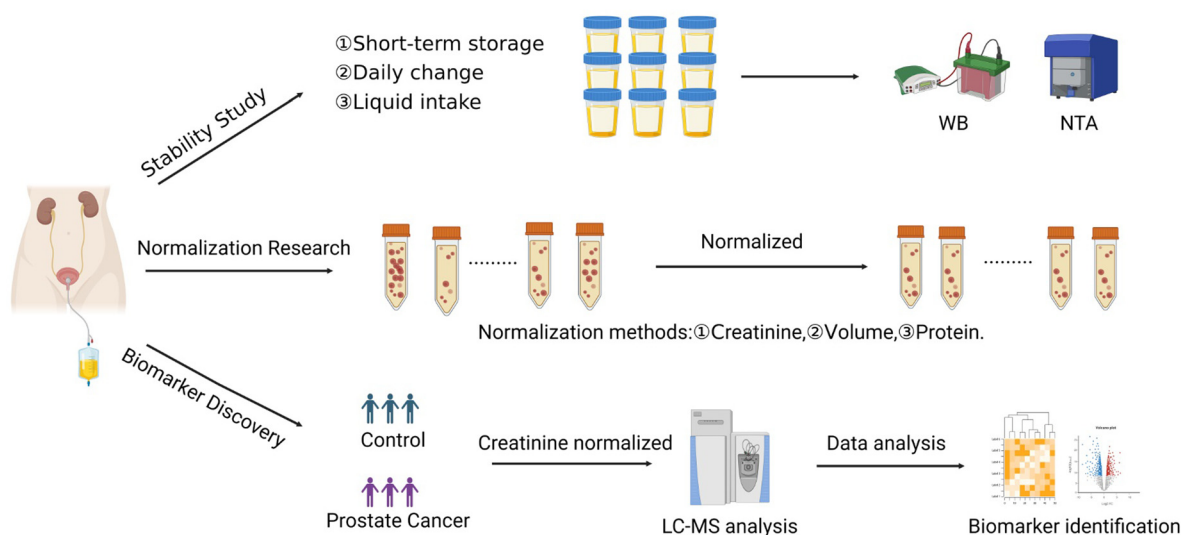

Fig. 1 Experimental design and workflow of the study. First, the stability of urine EVs during temporary storage, with daily changes and after liquid intake was studied. Then, the effects of creatinine, protein and volume normalization were evaluated. Finally, clinical samples were analyzed using mass spectrometry after normalization with creatinine, enabling the screening of potential biomarkers for prostate cancer.

with prostatitis or prostate disease. All experiments were performed according to the ethical guidelines of the local hospital. The study's protocol was approved by the Ethics Committee of the First Affiliated Hospital of Nanjing Medical University (2021-SR-167, 7 April 2021), Nanjing, China. Informed consent was obtained from human participants of this study.

We collected urine samples from 7 individuals, and for non-MS-based experiments such as western blotting, results with 3 individuals were illustrated. The individuals followed a light diet the day before sample collection. The urine collection process for individuals was conducted as follows: (1) Individuals' morning urine samples were collected. (2) Urine samples were collected in the morning, after lunch, and after dinner. Additionally, their morning urine samples were collected for three consecutive days. (3) Urine samples were collected in the morning. Subsequently, samples were collected after the individuals consumed 500 mL of water or 500 mL of Coca-Cola. (4) The individuals consumed a regular diet, and urine samples were collected at various time points throughout the day.

After collecting urine samples, they were centrifuged twice at 2500g for 10 minutes each time (to remove cell fragments and large apoptotic bodies). After centrifugation, the supernatant was stored at −80 °C for subsequent operation. The urine creatinine level was detected using a creatinine colorimetric assay kit (Cayman, 500701).

2.2. EV isolation by EVtrap

The method of EV extraction using EVtrap magnetic beads (Tymora Analytical, West Lafayette, IN, USA) was based on a previous protocol.[7] EVtrap magnetic beads were added to each urine sample in the proportion of 20 µL magnetic beads per mL of urine. The mixture was then incubated at room temperature for 1 hour, after which the supernatant was removed with the aid of a magnetic separator. Subsequently, EVs were washed once with a 1× loading buffer and three times with 1× PBS solution, respectively. Finally, EVs were obtained by eluting the EVtrap magnetic beads with 200 µL of elution buffer twice.

2.3. Nanoparticle tracking analysis (NTA)

EVs were eluted from 1 mL of the urine sample with 200 µL of 100 mM triethylamine and were diluted to 1 mL with PBS. The detection of the particle size and concentration of separated EVs was performed using ZetaView (Particle Metrix, Meerbusch, Germany) following a standard protocol. The instrument was initially calibrated using 100 nm polystyrene particles, which were diluted 250 000 times with purified water. The brightness of the instrument was set to 20, and the sensitivity and shutter were set to 70 and 100, respectively.

2.4. Characterization of EV by transmission electron microscopy (TEM)

EVs isolated from 1 mL of the urine sample were dispersed in 200 µL of PBS. Then, 10 µL of the solution was applied to a 200-mesh formvar carbon-coated copper grid, left to air dry naturally, and subsequently incubated with a 2% phosphotungstic acid solution (pH = 7.0) for 2–3 minutes at room temperature for negative staining. Finally, the EVs were imaged using transmission electron microscopy (TEM) using a Hitachi H-8100 electron microscope (Hitachi, Tokyo, Japan).

2.5. Data independent acquisition-mass spectrometry (DIA-MS) analysis and data processing

Peptides were analyzed with a QE HF-X (Thermo Fisher Scientific, USA) using nanospray, equipped with a Thermo Scientific EASY-nLC 1200 system. Phase A was 0.1% formic acid in water and phase B was 80% acetonitrile (ACN)/0.1% formic acid, and the analytical column containing C18 packing (2.2 µm, 100 Å; Michrom Bioresources) had a flow rate of 300 nL min^{-1}, with a linear 23 min gradient from 5% to 40% B, followed by a 7 min wash gradient. Mass spectrometry was performed in the DIA mode. The detailed parameters are as follows: the full MS resolution was set to 120 000, the scan range was set to 400–1200 m/z, the AGC target was 3e6, and the maximum IT was 30 ms. The DIA resolution was set to 15 000, the AGC target was 1e6, the maximum IT was 20 ms, and the isolation window was 8 m/z.

Thermo raw files were processed using the Direct-DIA analysis function of Spectronaut™ software 17 (Biognosys, Schlieren, Switzerland). The analysis was performed using the version of the human UniProt database downloaded on March 15, 2023. The default parameters of the software were used to search the directDIA database: trypsin/P was set, up to 3 missed cleavages were allowed, carbamoylmethylation (+57.02 Da) was chosen as the fixed modification and methionine oxidation (+15.99 Da) and acetylation (+42.01 Da) were chosen as variable modifications, protein and peptide false discovery rates (FDRs) were set to 1%, MS/MS was used for quantitative analysis. Perseus software was used to analyze the differential expression of the search data, the volcano map and heat map were generated ($p < 0.05$, t test $S_0 = 0$, |log$_2$ (fold change) |>1 is a differential protein), and the missing values were filled according to the normal distribution.

2.6. Statistical analysis and data availability

Statistical analysis was performed using Perseus software. All protein abundances were transformed using the binary logarithm, and the samples were categorized into a control group (negative samples) and a prostate cancer group (Gleason score > 6). Proteins with abundance frequencies higher than 50% across all samples were considered valid values. Missing values were imputed with smaller values, drawn from a normal distribution with a downshift of 1.8 standard deviation (SD) and a width of 0.3 SD. Subsequently, the abundances were median-normalized for each sample. Analysis of variance (ANOVA) was used for the multiple comparison test with $S_0 = 0$ and $p = 0.05$. Finally, |log$_2$ (fold change)|>1 is a differential protein. Statistical analysis was conducted using one-way analysis of variance (ANOVA) to assess the significance of differences. SPSS (v24.0) was employed for statistical analysis. A

p-value <0.05 was considered statistically significant. All the pictures were drawn using GraphPad Prism (v8.0), Origin2022 (v9.9) and R (4.2.3).

3.   Results and discussion

3.1   EVtrap facilitates the isolation and evaluation of EV stability

To assess the effect of urine collection and processing variables, we need to have a reliable and reproducible method for EV isolation and evaluation. Our group has recently introduced functionalized magnetic beads, EVtrap, to isolate EVs based on chemical affinity, demonstrating effective EV capture by utilizing the properties of hydrophilic and lipophilic groups, which exhibit specific affinity for EVs.[7] To further confirm the suitability of EVtrap in isolating EVs from urine to facilitate evaluating preanalytical variables, EVs isolated by EVtrap were characterized using WB, NTA, and TEM. In general, it is recommended to use both positive EV and negative markers to verify the purity of EV capture,[28] such as common EV markers CD9, CD81, CD63, and TSG101, and calnexin, an endoplasmic reticulum protein, as a negative marker.[29] In the WB experimental results, the EV markers CD9, CD81, and TSG101 were clearly detected in samples from multiple individuals (illustrated with n = 3), while the calnexin signal (negative control) was not observed (Fig. 2A). The results demonstrate the successful isolation and specificity of EV isolation by EVtrap from urine. The particle number and size distribution of EVs were further determined using nanoparticle tracking analysis (NTA). As depicted in Fig. 2B, urinary EVs isolated by EVtrap were pre-

dominantly distributed in the range of 30–400 nm. To visualize the size and structure of the captured EVs, TEM was also performed (Fig. 2C and D) to reveal clear observation of the original EV membrane and three-dimensional structure.

3.2   Effect of temporary urine storage temperature and time on EV stability

In previous studies, the effect of long-term storage temperatures (*e.g.*, −20 °C and −80 °C), time and storage format on the quantity and quality of EVs has been investigated. EVs and their proteome appear stable in longitudinal samples of subjects over a prolonged time-period at −80 °C, but freezing and thawing can cause some loss of EVs or artifacts.[8,10] However, there is no systemic investigation of EV stability in the context of clinical settings. During actual urine collection clinically, immediate analysis of urine samples may not always be feasible, such as in a busy hospital or local clinic. Once urine samples are collected, hospitals and clinics may not have freezers available for storage immediately either. Instead, samples are temporarily kept at room temperature or in a general refrigerator at least for multiple hours before being transported to a freezer. Different temporary storage conditions can greatly affect the EV concentration in urine. Therefore, it is necessary to evaluate the stability of urine storage at different temporary storage temperatures and time periods. We simulated the actual clinical situation for urine collection and its effect on EV stability, by carrying out the investigation at 4 °C and room temperature (25 °C) for 6 and 12 hours, respectively, while urine samples immediately stored at −80 °C served as the control, with at least 3 biological repeats for each condition. As illustrated in Fig. 3A and B, there was basically no
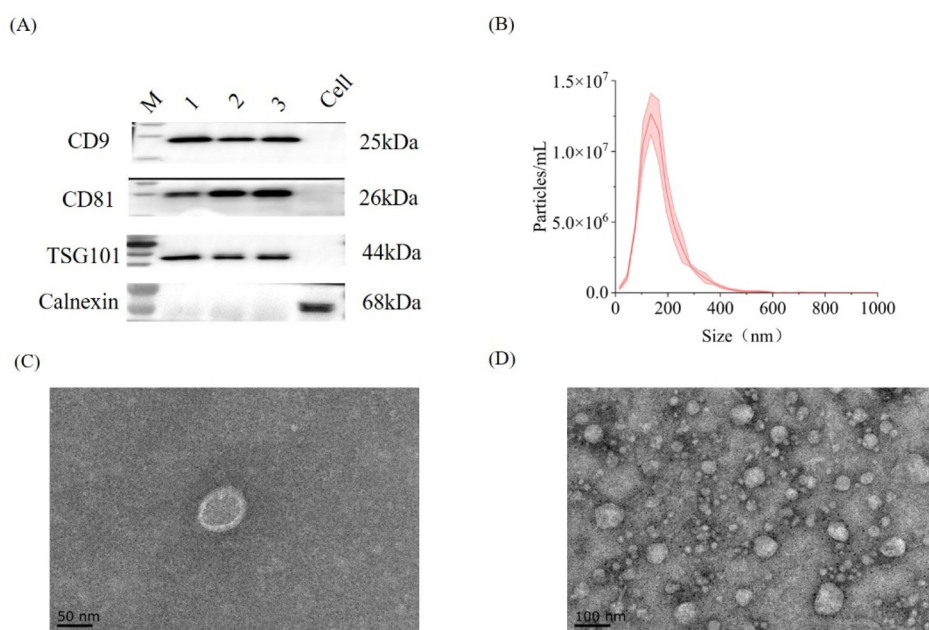

Fig. 2 Characterization of EVs isolated from urine by EVtrap. (A) Western blot results of urine EVs captured by EVtrap in individuals (n = 3). The cell extract was used as the control. CD9, CD81 and TSG101 were selected as biomarkers of EVs, and calnexin was used as a negative control; (B) NTA results of urine EVs captured by EVtrap (n = 3); and (C and D) TEM images of urine EVs from a selected individual.
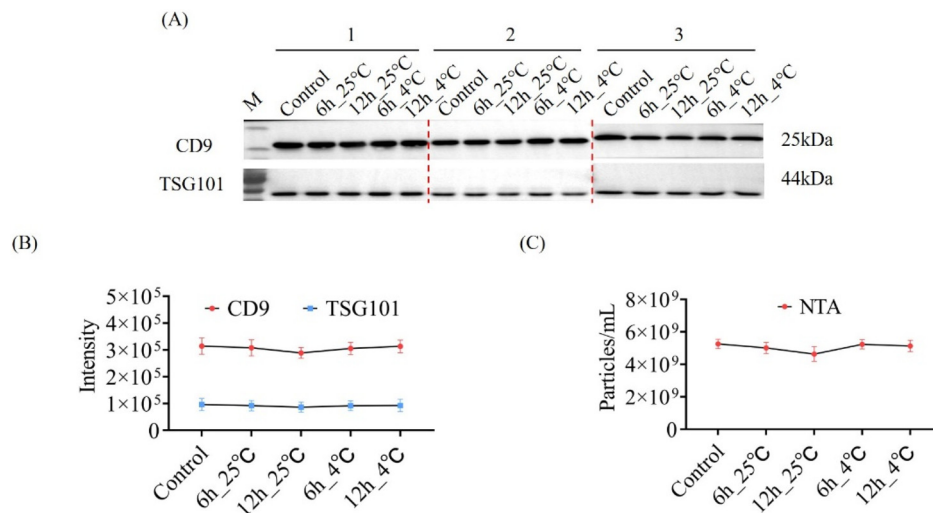
Fig. 3 Characterization of urine EVs under different temperature storage conditions. (A) Western blotting results of urine EVs captured by EVtrap ($n = 3$); (B) Grayscale values of CD9 and TSG101 bands in WB results ($n = 3$); and (C) NTA measurement ($n = 3$).

change in the WB signals of CD9 and TSG101 after the urine had been stored at 25 °C for 6 hours and 4 °C for 6 hours and 12 hours, while the levels of CD9 and TSG101 decreased after being stored at 25 °C for 12 hours. In addition, NTA was also performed to detect EVs in urine under these conditions. As can be seen from Fig. 3C, the EV concentration measured by NTA is overall consistent with the WB results. This decrease in EVs could be attributed to the possible degradation or fragmentation of EVs in urine under prolonged room temperature conditions.[30] We used a general linear model (one-way mixed ANOVA) to investigate the effects of storage time and temperature on the EV sample. Compared with the control group, we did not observe any significant difference between the two groups in terms of the analyzed parameters ($p > 0.05$). The results indicate that EVs remained relatively stable for 6 hours at 25 °C or for 12 hours at 4 °C.

3.3 Effect of the urine sampling time and diet

The important physiological roles of urine are attributed to its highly dynamic and volatile nature and vulnerability to external conditions (exercise, diet, *etc.*).[31] While the general guideline for urine collection, for example, the first morning specimen, facilitates minimizing the variation, the actual sample collection and handling can be quite inconsistent and difficult to implement. For example, it is difficult to collect urine samples from patients with prostate issues at designated time points so a non-ideal compromise is typically made by collecting urine samples anytime when the patient is ready. To mimic the nonideal situation during the actual urine collection, we explored the variation of urine EVs on the same day, for three consecutive days, and after consuming a large volume of liquid. As illustrated in Fig. S1A–D,† there are noticeable changes in the signals of CD9 and TSG101, whether urine samples were collected over one day or morning urine for three consecutive days from the same individuals. The CV

values for CD9 and TSG101 were above 10%, with the highest being 36.02%.

To further evaluate the dynamics of urine and its EV content that is affected by dietary conditions, the urine samples collected from three individuals after consuming 500 mL of water or soft drink were compared with their normal first morning urine. Not surprisingly, the results revealed a significant decrease in the EV amount in urine after the individuals consumed a large amount of liquids (Fig. S2†). These findings clearly illustrated the dynamic nature of urine samples and their content including urinary EVs and their susceptibility to external influences.[31] Therefore, it is advisable for patients to refrain from consuming excessive amounts of water or beverages before the collection of urine samples.

3.4 Evaluation of EV normalization using urine creatinine and the total EV protein amount

As urine collection and handling often is nonideal, it is necessary to normalize urine samples before or during EV analysis. Several EV normalization strategies have been proposed, *e.g.*, NTA, CD9 biomarkers, *etc.*, but they all present certain challenges. For example, NTA requires expensive instruments and the result shows bias in detecting EVs smaller than 70 nm, while the use of CD9 and other EV markers through immunoassay is time-consuming and less practical. Creatinine has been widely used to correct the excretion rate or urine concentration of a variety of endogenous and exogenous substances.[32] On the other hand, MS-based analyses typically involve the use of the total protein/peptide amount to normalize samples. We attempted to investigate which urine normalization method is sufficient to minimize pre-analytical variables. To examine the differences among various normalization methods in downstream MS analyses, we normalized the urine samples from the same individual at different time points on 3 consecutive days with 3 biological replicates using creatinine, volume, and

total protein amount. The protein content was quantified using a BCA Protein Assay Kit (Beyotime, P0010). The total protein amount in a 1 mL urine sample for each individual over three days is as follows: Individual 1 (Day1–Day3: 7.65 µg, 5.14 µg, 7.12 µg, 5.63 µg, 4.95 µg, 4.80 µg, 7.43 µg, 6.22 µg, and 7.57 µg); Individual 2 (Day1–Day3: 7.56 µg, 6.38 µg, 4.79 µg, 5.25 µg, 5.62 µg, 6.08 µg, 7.40 µg, 6.51 µg, and 5.21 µg); and Individual 3 (Day1–Day3: 4.63 µg, 7.34 µg, 5.14 µg, 4.24 µg, 5.20 µg, 5.71 µg, 6.52 µg, 6.60 µg, and 4.21 µg). As EV isolation methods may affect the purity of EVs, low EV purity could result in the dominance of contaminating proteins in total protein measurement.[33–35] In our study, we employed EVtrap to capture EVs from urine samples. Fig. S3† shows that the correlation between the total protein content and the number of EVs ranges from 0.78 to 0.84, indicating a good correlation.

The normalization efficiency was assessed by western blotting experiments based on two known EV markers, CD9 and TSG101. As shown in Fig. S4A-4I,† without creatinine normalization (volume normalization only), there was significant fluctuation in the levels of CD9 and TSG101 at different time points. When compared with volume and total protein normalization, the intensities of CD9 and TSG101 bands were basically identical after urine normalization using creatinine. To better compare the effect of normalization, we used the coefficient of variation to observe the level of volatility. As can be seen from Fig. S4J,† when urine is normalized with creatinine, the dispersion of the nine points is low, and the coefficients of variation (CV) of CD9 and TSG101 are 8.02% and 6.89%, respectively. However, when volume normalization and total protein normalization were used, the dispersion of the nine points was

higher, with the coefficients of variation (CVs) for volume normalization being 19.43% and 29.11%, and the CV values for total protein normalization being 17.55% and 15.76%, respectively. The data indicate that normalizing urine with creatinine is more efficient in reducing the variation from the same individuals. EVs from biofluids, as nano- and micro-particles, are typically coated with protein corona during the circulation and the total protein amount can be affected by the EV size distribution and properties of the EV surface. This may explain why it is not a better strategy to normalize urine based on the total protein amount. As shown in Fig. 4A–C, when urine creatinine normalization was applied, the variation of CD9 and TSG101 at different time points was significantly reduced. The CV values for CD9 and TSG101 in the volume-normalized samples were approximately 17% on average. However, after normalization with creatinine, the CVs of CD9 and TSG101 signals decreased to around 6% (Fig. 4D). These results indicate that normalizing urine samples using creatinine could effectively reduce inter-sample variability. The observation supports the relationship between the content of EVs in urine and creatinine values, consistent with previous literature.[11,36] Therefore, in clinical applications in which urine sample collection and handling is nonideal, creatinine normalization can be used to minimize the inter-sample variability for EV-based studies.

3.5 Profiling urinary EV proteomes of prostate cancer patients through DIA-MS analyses

Finally, we used the guideline from our observation to profile urinary EV proteomes of prostate cancer patients and age-matched controls. Urine samples from 30 prostate cancer
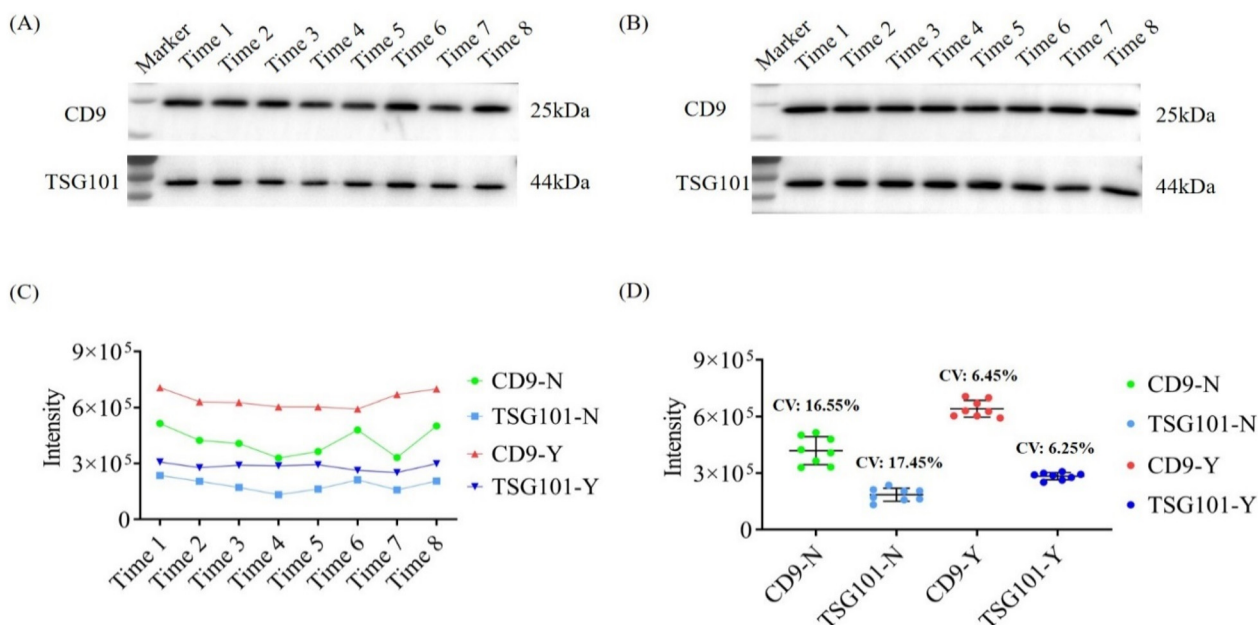

Fig. 4 Characterization of EVs in urine at 8 different time points during the day. (A) Western blotting results of urine with volume normalization only; (B) Western blotting results of urine with creatinine normalization; (C) Grayscale values of CD9 and TSG101 bands when urine creatinine was not normalized (volume normalization) and normalized; and (D) Coefficients of variation of CD9 and TSG101 with or without urinary creatinine normalization ($n = 8$, N: not normalized (volume normalization only), and Y: normalized).

patients and 30 non-prostate cancer patients as the control group were collected. In particular, prostate cancer patients have challenges in holding urine at designated time points, and therefore we did not restrict specific sample collection time. Creatinine normalization was performed on all urine samples. After normalization, EVs in urine were enriched using EVtrap and lysated and proteins were extracted and denatured using the phase transfer surfactant (PTS) buffer, followed by reduction, alkylation and trypsin digestion to obtain peptides for LC-MS analysis. A total of 47 334 unique peptides representing 4939 proteins were identified. Label-free quantification (LFQ) was performed to screen significant proteins specific to prostate cancer by profiling proteomes in the patient and control groups. The volcano plot (Fig. 5A, p-value < 0.05, $\log_2$|Fold change|>1) illustrates that a total of 57 significantly differential proteins were identified, including 51 upregulated proteins and 6 downregulated proteins. Among the upregulated proteins, at least 19 have been previously reported to be associated with prostate cancer (INAFM1, GLS, PDHA1, IGFBP1, HLA-C, EPB41, SDC1, NDUFV2, C4BPB, FGFR3, MMP11, GSTT1, TRIM28, POSTN, ADIRF, AKAP4, ANO1, GOLM1, and SUCLA2). Notably, GOLM1, a Golgi membrane protein, is upregulated in clinical prostate cancer and has been found to be superior to serum prostate-specific antigen (PSA) in diagnosing prostate cancer.[37] POSTN is an exploratory periosteal protein that is highly expressed in prostate cancer. It promotes the proliferation and invasion of PCa cells and plays an important role in osteoblast metastasis of PCa.[38]

To further evaluate the 57 EV proteins differentially identified in prostate cancer, we performed a cluster analysis for the 30 prostate cancer patients and 30 control groups. As seen in the heat map (Fig. 5B, p-value < 0.05, t test $S_0 = 0$), 51 upregulated proteins and 6 downregulated proteins were visualized. We conducted the Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis on the 51 upregulated proteins (Fig. S5,† p-value < 0.05). The KEGG pathway analysis showed that these proteins were related to central carbon metabolism in cancer, metabolic pathways, carbon metabolism and others. In particular, it is intriguing to the identification of four proteins (GLS, PDHA1, EPB41, and FGFR3) in EVs involved in the central carbon metabolism in the cancer pathway.

Boruta is a feature selection method based on random forest, which randomly destroys each real feature sequentially, evaluates the importance of each feature, and iteratively deletes the features with low correlation to find the best variables.[39,40] We further performed a signature protein screen for 51 upregulated proteins by using the Boruta package. As shown in Fig. S6A,† we screened a total of 16 feature proteins (highlighted in green). In order to understand the ability of each feature protein to distinguish between prostate cancer and non-prostate cancer, we assessed the importance of the 16 feature proteins and visualized them in a variable importance plots map (Fig. S6B†). The top 5 proteins are ranked in terms of importance as AKAP4, TRIM28, FGFR3, MMP11 and ATP5PB. Among them, it has been reported that AKAP4 is abnormally expressed in PCa and can be used as a potential biomarker for PCa.[41] We utilized receiver operating characteristic (ROC) curves to evaluate the discriminatory capability of these 5 proteins in distinguishing between prostate cancer
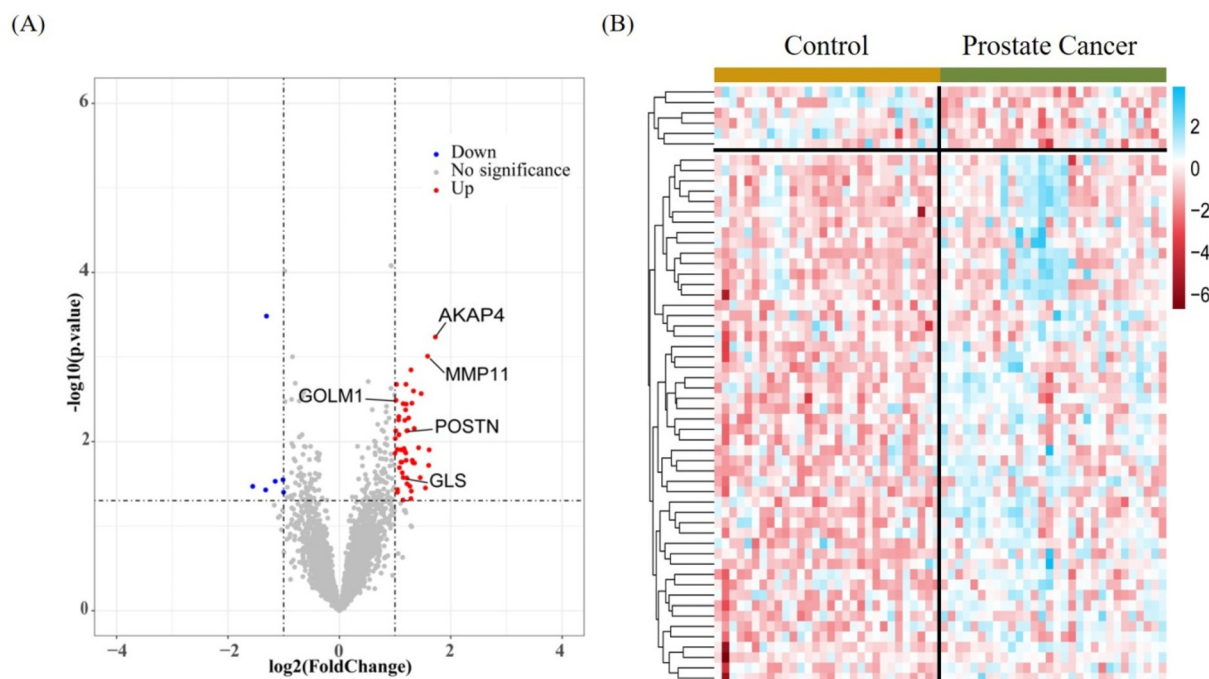

Fig. 5 Quantitative proteomics of urine EVs from control and prostate cancer groups. (A) Volcano plot comparison of the differentially expressed proteins. (B) Heatmap of the significantly regulated proteins of the control and prostate cancer groups.

patients and non-prostate cancer patients. As shown in Fig. S7,† the area under the curve (AUC) values of all 5 feature proteins are below 0.8. In cancer diagnosis, individual indicators may exhibit uncertainty, and there is a growing trend towards combining multiple indicators for joint analysis. To enhance the overall predictive ability, we integrated the 5 proteins into a combined panel. By employing the random forest algorithm, we constructed an ROC curve and calculated its corresponding area under the curve (AUC). As illustrated in Fig. S8,† the results demonstrated that the panel comprising the 5 feature proteins achieved a high AUC value of 0.988, indicating their good performance in distinguishing prostate cancer from non-prostate cancer patients. Therefore, these 5 proteins show promise as potential biomarkers for prostate cancer for future validation and investigation with larger cohorts of samples.

4. Conclusions

There is increasing interest in using urinary EVs as the source to develop disease biomarkers. Previous studies have carefully investigated the effects of preanalytical variables on EV stability during long-term storage. Overall, urinary EVs and their cargoes such as proteome are relatively stable over a prolonged time-period, suggesting the feasibility of developing reliable non-invasive diagnostic/prognostic biomarkers. However, actual collection and handling of urine samples and temporary storage in clinics and hospitals before long-term storage may have profound effects on EV stability. In addition, depending on clinical settings and patient conditions, urine collection and handling often is not ideal. In this study, we simulated the actual clinical sample collection to investigate the stability of EVs when urine samples were collected and stocked on the spot and examined different normalization methods to minimize fluctuations. These explorations provide valuable guidance for clinical urine sample collection and handling, and demonstrated the identification of a panel of potential protein biomarkers for the diagnosis of PCa that may facilitate standardization of urine EV biomarker research and development.

Data availability

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium *via* the PRIDE partner repository with the data set identifier PXD042489.

Author contributions

G. Z., W. A. T. and Y. Z. participated in the planning, data generation and data interpretation. Y. D. contributed to the data analysis. H. Z., Y. L., J. S. and D. W. were involved in the clinical sample collection and data interpretation. W. A. T., Y. Z. and G. Z. contributed to writing the manuscript. Z. X. was involved in funding acquisition. All authors have read and approved the manuscript.

Conflicts of interest

The authors declare no conflicts of interest.

Acknowledgements

References

1 A. Schubert and M. Boutros, *Mol. Oncol.*, 2021, **15**, 3–26.
2 C. Ciardiello, L. Cavallini, C. Spinelli, J. Yang, M. Reis-Sobreiro, P. de Candia, V. R. Minciacchi and D. Di Vizio, *Int. J. Mol. Sci.*, 2016, **17**, 175.
3 R. Kalluri and V. S. LeBleu, *Science*, 2020, **367**, 640.
4 D. W. Greening and R. J. Simpson, *Proteomics*, 2021, **21**, e2100126.
5 A. Alberro, L. Iparraguirre, A. Fernandes and D. Otaegui, *Int. J. Mol. Sci.*, 2021, **22**, 8163.
6 C. Marar, B. Starich and D. Wirtz, *Nat. Immunol.*, 2021, **22**, 560–570.
7 X. Wu, L. Li, A. Iliuk and W. A. Tao, *J. Proteome Res.*, 2018, **17**, 3308–3316.
8 H. Zhou, P. S. Yuen, T. Pisitkun, P. A. Gonzales, H. Yasuda, J. W. Dear, P. Gross, M. A. Knepper and R. A. Star, *Kidney Int.*, 2006, **69**, 1471–1476.
9 R. Vago, G. Radano, D. Zocco and N. Zarovni, *Sci. Rep.*, 2022, **12**, 17663.
10 K. Barreiro, O. P. Dwivedi, S. Valkonen, P. H. Groop, T. Tuomi, H. Holthofer, A. Rannikko, M. Yliperttula, P. Siljander, S. Laitinen, E. Serkkola, T. Af Hallstrom, C. Forsblom, L. Groop and M. Puhka, *J. Extracell. Vesicles*, 2021, **10**, e12158.
11 C. J. Blijdorp, T. A. Hartjes, K. Y. Wei, M. H. van Heugten, D. M. Bovee, R. P. J. Budde, J. van de Wetering, J. G. J. Hoenderop, M. E. van Royen, R. Zietse, D. Severs and E. J. Hoorn, *J. Extracell. Vesicles*, 2022, **11**, e12181.
12 P. Fernandez-Llama, S. Khositseth, P. A. Gonzales, R. A. Star, T. Pisitkun and M. A. Knepper, *Kidney Int.*, 2010, **77**, 736–742.
13 D. L. Heavner, W. T. Morgan, S. B. Sears, J. D. Richardson, G. D. Byrd and M. W. Ogden, *J. Pharm. Biomed. Anal.*, 2006, **40**, 928–942.
14 E. H. Koritzinsky, J. M. Street, R. R. Chari, D. M. Glispie, T. R. Bellomo, A. M. Aponte, R. A. Star and P. S. T. Yuen, *Am. J. Physiol.: Renal Physiol.*, 2019, **317**, F1098–F1110.
15 W. Oosthuyzen, N. E. Sime, J. R. Ivy, E. J. Turtle, J. M. Street, J. Pound, L. E. Bath, D. J. Webb, C. D. Gregory, M. A. Bailey and J. W. Dear, *J. Physiol.*, 2013, **591**, 5833–5842.

16 M. Ramirez-Garrastacho, C. Bajo-Santos, A. Line, E. S. Martens-Uzunova, J. M. de la Fuente, M. Moros, C. Soekmadji, K. A. Tasken and A. Llorente, *Br. J. Cancer*, 2022, **126**, 331–350.

17 K. Fujita, H. Kume, K. Matsuzaki, A. Kawashima, T. Ujike, A. Nagahara, M. Uemura, Y. Miyagawa, T. Tomonaga and N. Nonomura, *Sci. Rep.*, 2017, **7**, 42961.

18 F. Crocetto, G. Russo, E. Di Zazzo, P. Pisapia, B. F. Mirto, A. Palmieri, F. Pepe, C. Bellevicine, A. Russo, E. La Civita, D. Terracciano, U. Malapelle, G. Troncone and B. Barone, *Cancers*, 2022, **14**, 3272.

19 L. A. Erozenci, F. Bottger, I. V. Bijnsdorp and C. R. Jimenez, *FEBS Lett.*, 2019, **593**, 1580–1597.

20 M. Dong, T. M. Lih, M. Ao, Y. Hu, S. Y. Chen, R. V. Eguez and H. Zhang, *Anal. Chem.*, 2021, **93**, 13774–13782.

21 . R. B. Kitata, J. C. Yang and Y. J. Chen, *Mass Spectrom. Rev.*, 2022, e21781.

22 Y. Yang and L. Qiao, *Proteomics*, 2023, **23**, e2200046.

23 L. K. Pino, S. C. Just, M. J. MacCoss and B. C. Searle, *Mol. Cell. Proteomics*, 2020, **19**, 1088–1103.

24 D. Mehta, S. Scandola and R. G. Uhrig, *Anal. Chem.*, 2022, **94**, 793–802.

25 J. Zhao, Y. Yang, H. Xu, J. Zheng, C. Shen, T. Chen, T. Wang, B. Wang, J. Yi, D. Zhao, E. Wu, Q. Qin, L. Xia and L. Qiao, *npj Biofilms Microbiomes*, 2023, **9**, 4.

26 D. B. Bekker-Jensen, O. M. Bernhardt, A. Hogrebe, A. Martinez-Val, L. Verbeke, T. Gandhi, C. D. Kelstrup, L. Reiter and J. V. Olsen, *Nat. Commun.*, 2020, **11**, 787.

27 H. Zhang, G. Y. Zhang, W. C. Su, Y. T. Chen, Y. F. Liu, D. Wei, Y. X. Zhang, Q. Y. Tang, Y. X. Liu, S. Z. Wang, W. C. Li, A. Wesselius, M. P. Zeegers, Z. Y. Zhang, Y. H. Gu, W. A. Tao and E. Y. Yu, *Molecules*, 2022, **27**, 8155.

28 C. Thery, K. W. Witwer, E. Aikawa, M. J. Alcaraz, J. D. Anderson, R. Andriantsitohaina, A. Antoniou, P. Zimmermann, A. M. Zivkovic, D. Zocco and E. K. Zuba-Surma, *J. Extracell. Vesicles*, 2018, **7**, 1535750.

29 R. Kalluri and K. M. McAndrews, *Cell*, 2023, **186**, 1610–1626.

30 B. Qin, Q. Zhang, X. M. Hu, T. Y. Mi, H. Y. Yu, S. S. Liu, B. Zhang, M. Tang, J. F. Huang and K. Xiong, *J. Cell Physiol.*, 2020, **235**, 7663–7680.

31 U. Erdbrugger, C. J. Blijdorp, I. V. Bijnsdorp, F. E. Borras, D. Burger, B. Bussolati, J. B. Byrd, A. Clayton, J. W. Dear, J. M. Falcon-Perez, C. Grange, A. F. Hill, H. Holthofer, E. J. Hoorn, G. Jenster, C. R. Jimenez, K. Junker, J. Klein, M. A. Knepper, E. H. Koritzinsky, J. M. Luther, M. Lenassi, J. Leivo, I. Mertens, L. Musante, E. Oeyen, M. Puhka, M. E. van Royen, C. Sanchez, C. Soekmadji, V. Thongboonkerd, V. van Steijn, G. Verhaegh, J. P. Webber, K. Witwer, P. S. T. Yuen, L. Zheng, A. Llorente and E. S. Martens-Uzunova, *J. Extracell. Vesicles*, 2021, **10**, e12093.

32 H. Jurdáková, R. Górová, G. Addová, A. Šalingová and I. Ostrovský, *Anal. Biochem.*, 2018, **549**, 113–118.

33 J. Chen, P. Li, T. Zhang, Z. Xu, X. Huang, R. Wang and L. Du, *Front. Bioeng. Biotechnol.*, 2021, **9**, 811971.

34 L. M. Doyle and M. Z. Wang, *Cells*, 2019, **8**, 727.

35 W. M. Xu, A. Li, J. J. Chen and E. J. Sun, *J. Membr. Biol.*, 2023, **256**, 25–34.

36 C. J. Blijdorp, O. A. Z. Tutakhel, T. A. Hartjes, T. P. P. van den Bosch, M. H. van Heugten, J. P. Rigalli, R. Willemsen, U. M. Musterd-Bhaggoe, E. R. Barros, R. Carles-Fontana, C. A. Carvajal, O. J. Arntz, F. A. J. van de Loo, G. Jenster, M. C. Clahsen-van Groningen, C. A. Cuevas, D. Severs, R. A. Fenton, M. E. van Royen, J. G. J. Hoenderop, R. J. M. Bindels and E. J. Hoorn, *J. Am. Soc. Nephrol.*, 2021, **32**, 1210–1226.

37 S. Varambally, B. Laxman, R. Mehra, Q. Cao, S. M. Dhanasekaran, S. A. Tomlins, J. Granger, A. Vellaichamy, A. Sreekumar, J. Yu, W. Gu, R. Shen, D. Ghosh, L. M. Wright, R. D. Kladney, R. Kuefer, M. A. Rubin, C. J. Fimmel and A. M. Chinnaiyan, *Neoplasia*, 2008, **10**, 1285–1294.

38 C. Y. Sun, Y. Y. Mi, S. Y. Ge, Q. F. Hu, K. Xu, Y. J. Guo, Y. F. Tan, Y. Zhang, F. Zhong and G. W. Xia, *Front. Oncol.*, 2021, **11**, 795712.

39 C. Kong, Y. Zhu, X. Xie, J. Wu and M. Qian, *Front. Immunol.*, 2023, **14**, 1184700.

40 M. Kursa and W. Rudnicki, *J. Stat. Softw.*, 2010, **36**, 1–13.

41 M. Chiriva-Internati, Y. Yu, L. Mirandola, N. D'Cunha, F. Hardwicke, M. J. Cannon, E. Cobos and W. M. Kast, *Prostate*, 2012, **72**, 12–23.