Supplementary information

Reagents used for the studies

HPLC grade LC buffers, Dithiothreitol (DTT), acetonitrile (ACN), ammonium bicarbonate, trifluoretic acids (TFA), and iodoacetamide (IAA) were purchased from Thermo Fisher Scientific (Waltham, MA). Trypsin Gold, mass spectrometry grade, was purchased from Promega (Madison, WI).

Serum and Urine exosomal isolation

Urine and 250μl of urine or 100μl of serum were used to purify total exosome using total Exosome RNA and protein isolation kit (Life Technologies, Grand Island, NY) following the manufacturer's protocol with the following modifications. We used 40μl of the reagent in the kit for 100μl of serum and we washed the exosome pellet twice with PBS. Total exosome lysate was then generated in 50μl of the lysis buffer (50mM Ammonium Bicarbonate, 4M Urea, and protease cocktail) using 1.4mm ceramic beads and the Omni Bead Ruptor Homogenizer (Omni International, GA). Protein concentration in total exosome lysate was determined by the EZQ Protein Quantification Assay (Life Technology Corp. CT).

Sample Preparation for LC-MS/MS Analysis

For characterization of total urine proteome, proteins from 10ul of urine were separated using precast 4–20% Tris-Glycine SDS-PAGE gels (1.0 mm thick) (Life Technologies, Carlsbad, CA). The protein gels were stained with Coomassie Blue and the entire lane of the gel were separated into 6 fractions. Biological replicates at each time point (n=5) were run on the same gel and processed simultaneously. The gel pieces were destained,

reduced and alkylated, and dried for in-gel digestion. For in-gel digestion, the dried gel pieces were rehydrated and digested in 80 μL of 12.5 ng/μL Trypsin Gold/50 mM ammonium bicarbonate at 37 °C overnight. After the digestion was complete, condensed evaporated water was collected from tube walls by 5 s centrifugation using benchtop microcentrifuge (Eppendorf, Hauppauge, NY). The gel pieces and digestion reaction were mixed with 50 μL 2.5% TFA and rigorously mixed for 15 min. The solution with extracted peptides was transferred into a fresh tube. The remaining peptides were extracted with 80 μL 70% ACN/5% TFA mixture using rigorous mixing for 15 min. The extracts were pooled and dried to completion (1.5–2 h) in SpeedVac. The dried peptides were reconstituted in 30 μL 0.1% TFA by mixing for 5 min and stored in ice or at −20 °C prior to analysis. 125ng of trypsin was added to 5μg of total exosome lysate (1:40) along with 2mM $CaCl_2$ and incubated at 37^{o}C for 16 hours. Samples were centrifuged subsequently for 30 minutes at 14,000 rpm, and the cleared supernatants were transferred to fresh tubes to be acidified with 90% formic acid (2% final) to stop proteolysis. The soluble peptide mixtures were collected for LC-MS/MS analysis.

LC-MS/MS Analysis

The concentrated peptide mix was reconstituted in a solution of 2 % acetonitrile (ACN), 2 % formic acid (FA) for MS analysis. Peptides were loaded with the autosampler directly onto a 2cm C18 PepMap pre-column and were eluted from the 50cm x 75 μm ID PepMap RSLC C18, 2μm column using a Thermo Dionex 3000 with a 98 min gradient from 2% buffer B to 30 % buffer B (100 % acetonitrile, 0.1 % formic acid). The gradient was switched from 30 % to 85 % buffer B over 5 min and held constant for 1 minute. Finally, the gradient was changed from 85 % buffer B to 98 % buffer A (100% water, 0.1% formic acid) over 2 minutes, and then held constant at 98 % buffer A for 25 more

minutes. The application of a 2.0 kV distal voltage electrosprayed the eluting peptides directly into the mass spectrometer equipped with an Easy-spray source (Thermo Finnigan, San Jose, CA). Full mass spectra (MS) was recorded on the peptides over a 400 to 1500 m/z range at 120,000 resolution, followed by tandem mass (MS/MS) CID (collision induced dissociation) events for a total of a 3sec cycle. Charge state dependent screening was turned off, and peptides with a charge state of 2-6 were analyzed. Mass spectrometer-scanning functions and HPLC gradients were controlled by the Xcalibur data system (Thermo Finnigan, San Jose, CA). Three technical replicates were run for each sample, and MS/MS data from technical replicates were merged for subsequent database search.

Database Search And Interpretation Of MS/MS Data

Tandem mass spectra from raw files were searched against a human protein database using the Proteome Discoverer 1.4 (Thermo Finnigan, San Jose, CA). The Proteome Discoverer application extracts relevant MS/MS spectra from the raw file and determines the precursor charge state and the quality of the fragmentation spectrum. The Proteome Discoverer probability-based scoring system rates the relevance of the best matches found by the SEQUEST algorithm. The mouse database was downloaded as FASTA-formatted sequences from Uniprot protein database (database released on December, 2014). The peptide mass search tolerance was set to 10ppm. A minimum sequence length of 7 amino acids residues was required. Only fully tryptic peptides were considered. To calculate confidence levels and FDR, Proteome Discoverer generates a decoy database containing reverse sequences of the non-decoy protein database and performs the search against this concatenated database (non-decoy + decoy). The discriminant score was set at 1% FDR determined based on the number of accepted

decoy database peptides to generate protein lists for this study. Spectral counts used to identify each protein were used for expression profiling analysis.

Spectral Count Profile Analysis

Spectral counts of identified proteins (computed as above) were loaded into the Qlucore Omics Explorer (Qlucore AB, Sweden) software. Data was filtered for uninformative features. Statistical analysis of quantifiable proteins was conducted using the linear model for microarray (LIMMA) package for group comparisons and FDR-based correction for multiple comparisons.

Additional MS Bioinformatics Analysis (DDA and DIA)

DDA data were analyzed using a multi-pass workflow that combines spectral library search, database search, and blind PTM search methods. The detail of this workflow will be described elsewhere. Briefly, all spectra were first clustered using MS-Cluster so redundant MS/MS spectra were grouped together. Next all clustered spectra were searched against a mouse spectral library downloaded from National Institute of Standards and Technology (NIST, ver. 01/14/2010) using M-SPLIT with 2.0 Da parent mass tolerance and 0.5 Da fragment mass tolerance. Peptide-spectrum-matches (PSMs) were filtered at 1% FDR using the Target/Decoy approach (TDA). Spectra not identified by M-SPLIT were then searched against a mouse sequence database (from UniProt, released in December 2014) using MSGFDB with 50ppm precursor mass tolerance, allowing Carbxymethlation as fixed modification. PSMs were again filtered at 1% FDR using TDA. Finally the leftover spectra were searched against the mouse sequence database using a blind post-translational (PTM) search tool MOD by allowing up to 200 Da in the mass of the PTM. The results were filtered at 1% FDR.

DIA data was analyzed with precursor isolation window of 25.0 Da and 0.05 Da fragment mass tolerance using MSPLIT-DIA, a spectral library search method that

identifies multiplexed MS/MS spectra from DIA data. The details of the algorithm will be described elsewhere. Briefly, a spectral library was built from the results from the DDA runs. To ensure the quality of the library we filtered the results at 1% peptide-level FDR. For each peptide the PSM with the lowest spectral probability was taken as the representative spectrum. The spectral library was then used to identify peptides in DIA data by matching each library spectrum against the multiplexed MS/MS spectra and returning all Spectrum-Spectrum-Matches (SSMs) with significant similarity. The statistical significance of the SSMs was estimated at 1% FDR using TDA. Protein quantification was performed via spectral counting. Spectral count for a peptide was computed as the number of PSMs that was identified to be that peptide. Spectral count for a protein was then computed as the sum of spectral count of all its proteotypic peptides (i.e. peptides that map to only one protein sequence in the database).