## **Proteomic analysis of urinary microvesicles and exosomes in medullary sponge**

# **kidney disease and autosomal dominant polycystic kidney disease**

Maurizio Bruschi<sup>1\*</sup>, Simona Granata<sup>2\*</sup>, Laura Santucci<sup>1\*</sup>, Giovanni Candiano<sup>1</sup>, Antonia Fabris<sup>2</sup>, Nadia Antonucci<sup>2</sup>, Andrea Petretto<sup>3</sup>, Martina Bartolucci<sup>3</sup>, Genny Del Zotto<sup>3</sup>, Francesca Antonini<sup>3</sup>, Gian Marco Ghiggeri<sup>4</sup>, Antonio Lupo<sup>2</sup>, Giovanni Gambaro<sup>2</sup>, Gianluigi

# Zaza<sup>2</sup>

<sup>1</sup> Laboratory on Molecular Nephrology, IRCCS Istituto Giannina Gaslini, Genoa, Italy.

 $2$  Renal Unit, Department of Medicine, University-Hospital of Verona, Italy

<sup>3</sup> Core Facilities, IRCCS Istituto Giannina Gaslini, Genoa, Italy.

<sup>4</sup> Division of Nephrology, Dialysis and Transplantation, IRCCS Istituto Giannina Gaslini, Genoa, Italy.

#### **Table of contents for supplemental materials**

### **Supplemental methods**

**Supplemental Figure 1.** Age and eGFR of all study partecipants included in the study.

**Supplemental Figure 2.** Characterization of isolated exosomes and microvesicles

**Supplemental Figure 3.** Gene Ontology annotation of urinary extracellular vesicle proteins.

**Supplemental Figure 4.** Multidimensional scaling analysis of extracellular vesicles from the urine of medullary sponge kidney (MSK) and autosomal dominant polycystic kidney disease (ADPKD) patients.

**Supplemental Figure 5.** Sample clustering and trait indicators.

**Supplemental Figure 6.** Venn diagram of statistically significant differences in protein abundance in the different types of extracellular vesicles from medullary sponge kidney (MSK) or autosomal dominant polycystic kidney disease (ADPKD) patients.

**Supplemental Figure 7.** Proteins network interaction.

**Supplemental Figure 8.** Gene Ontology enrichment analysis for core discriminatory proteins in the extracellular vesicles of medullary sponge kidney (MSK) and autosomal dominant polycystic kidney disease (ADPKD) patients.

**Supplemental Table 1.** List of all significant proteins identified using mass spectrometry

**Supplemental Table 2. ELISA cutoff, sensitivity, specificity and likelihood ratio.** 

## **References for supplemental materials**

## **Supplemental methods**

### *Mass spectrometry: Instrumentation*

The desalted peptides were dried by speed vacuum and resuspended in 2% acetonitrile containing 0.2% formic acid (FA). They were separated on a 50-cm reversed-phase Easy Spray column (75-μm internal diameter × 50 cm; 2 μm/100 Å C18) on an Ultimate 3000 RSLCnano system (Thermo Fisher Scientific, Waltham, MA, USA) with a binary buffer system comprising buffer A (0.1% FA) and buffer B (80% acetonitrile, 5% dimethylsulfoxide, 0.1% FA). The program comprised a 70-min gradient (2–45% buffer B) at a flow rate of 250 nl per min, with the column temperature maintained at 60°C. The chromatography system was coupled to an Orbitrap Fusion Tribrid mass spectrometer (Thermo Fisher Scientific), acquiring data in Charge Ordered Parallel Ion aNalysis (CHOPIN) mode [1]. The precursors were ionized using an EASY-spray source held at +2.2 kV and the inlet capillary temperature was held at 300°C. Single MS survey scans were performed over the mass window 375–1500 *m/z* with an AGC target of 250,000, a maximum injection time of 50 ms, and a resolution of 120,000 at 200 *m/z*. Monoisotopic precursor selection was enabled for peptide isotopic distributions, precursors of  $z = 2-5$ were selected for 2 s of cycle time, and dynamic exclusion was set to 25 s with a  $\pm$ 10 ppm window set around the precursor. The following CHOPIN conditions were applied: a) if the precursor charge state is 2, then follow with collision-induced dissociation (CID) and scan in the ion trap with an isolation window of 1.8, CID energy of 35% and a rapid ion trap scan rate; b) if the precursor charge state is 3–5 and precursor intensity >500,000, then follow with higher-energy C-trap dissociation (HCD) and scan in the Orbitrap with an isolation window of 1.8, HCD energy of 28% and a resolution of 15000; c) if the precursor charge state is 3–5 and precursor intensity <500,000, then follow with CID as described for option

(a). For all  $MS<sup>2</sup>$  events, the following options were set: "Injection lons for All Available Parallelizable Time" with an AGC target value of 4000 and a maximum injection time of 250 ms for CID, or an AGC target value of 10,000 and a maximum injection time of 40 ms for HCD.

#### *Mass spectrometry: Data analysis*

Raw MS files were processed within the MaxQuant v1.6.0.16environment [2] using the MaxLFQ algorithm for label-free quantification and the integrated Andromeda search engine with a false discovery rate (FDR) <0.01 at the protein and peptide levels. The search included variable modifications for oxidized methionine (M), acetylation (protein N-terminus), and fixed carbamidomethyl modifications (C). Up to two missed cleavages were allowed for protease digestion. Peptides with at least six amino acids were considered for identification, and 'match between runs' was enabled with a matching time window of 1 min to allow the quantification of  $MS<sup>1</sup>$  features which were not identified in each individual measurement. Peptides and proteins were identified using the UniProt FASTA *Homo sapiens* database (August 2017).

### *Dynamic light scattering*

The size of the exosomes and microvesicles were determined by dynamic light scattering (DLS) using a Zetasizer nano ZS90 particle sizer at a 90° fixed angle (Malvern Instruments, Worcestershire, UK). The particle diameter was calculated using the Stokes– Einstein equation [3]. For particle sizing in solution, exosome or microvesicle aliquots were diluted in 10% PBS and analyzed at a constant 25°C. The data were acquired and analyzed using Dispersion Technology Software (Malvern Instruments).

4

### *Western blot*

The protein content of microvesicles and exosomes was measured by Bicinchoninic acid assay. Samples (5 µg total protein for all type of extracellular vesicles) were separated by sodium dodecylsulfate polyacrylamide gel electrophoresis (SDS-PAGE) on an 8–16% acrylamide gradient and then transferred to a nitrocellulose membrane. The membrane was blocked, rinsed and labeled with one of the following primary human antibodies diluted in 1% (w/v) bovine serum albumin (BSA) in PBS containing 0.15% (v/v) Tween-20 (PBS-T): monoclonal anti-CD63 (Novus Biological, Littleton, CA, USA, 1:1000 clone H5C6), monoclonal anti-CD81 (Novus Biological, 1:1000 clone 1D6), or monoclonal anti-CD45 (LifeSpan BioSciences, Seattle, WA, USA, 1:1000 clone 3G4). After rinsing in PBS-T, the membrane was probed with secondary antibodies conjugated to horseradish peroxidase (diluted 1:10,000 in 1% (w/v) BSA in PBS-T). Chemiluminescence was monitored using the ChemiDoc Touch Imaging System (Bio-Rad, Hercules, CA, USA) and the signal was quantified by densitometry using Image Lab software (Bio-Rad).

#### *Flow cytometry: Instrumentation*

Before exosomes acquisition, the LSRFortessa X-20 instrument (Becton Dickinson Franklin Lakes, NJ, USA) was calibrated using CS&T beads (Becton Dickinson) and carefully washed with double-distilled water for at least 1 h. Front scatter (FSC) and side scatter (SSC) were set to log scale as recommended [4], and voltages were adjusted to the highest values to exclude background noise from PBS [4], which was purified by passing through 0.22-µm UltrafreeVR-MC/DuraporeVR-PVDF centrifugal filter units (Merk, Darmstadt, Germany). This allowed the detection of all different dimensions (0.5, 0.24, 0.2, 0.16 µm) of SSC MegaMix beads (BioCytex, Stago Group, France). For every sample

5

acquisition, Rainbow Fluorescent Particles (Spherotech, Lake Forest, IL, USA) were used to adjust all channel voltages thus maintaining voltage consistency. Monoclonal mouse anti-human CD63 (IgG1, clone H5C6), CD81 (IgG1, clone JS-81) and IgG1 isotype controls (clone MOPC-21) were sourced from Becton Dickinson whereas CD133PE-Cy7 (IgG1, clone7) was sourced from BioLegend (San Diego, CA, USA). FITC-conjugated phalloidin was provided by Sigma-Aldrich (St. Louis, MO, USA).

## *Flow cytometry: Analysis*

 Exosomes, identified as particles smaller than 0.16 µm in diameter when compared to SSC MegaMix Beads, were gated according to their SSC, and 100,000 events for each sample were acquired at a low sample pressure and a low flow rate of 8–12 µl/min [4]. Fluorescence Minus One (FMO) analysis was used to achieve the correct size cutoff [5]. FCS files were exported from Fortessa X20 and data were evaluated using Kaluza software (Beckman Coulter, Brea, CA, USA).

## *CD133 ELISA assay*

CD133 expression in urinary exosomes was determined by homemade ELISA. Clean exosome fraction recovered from 16 ml of urine was solubilized in 10 μl of mild detergent solution i.e. 1%(v/v) Nonident P-40 (NP-40), 0.5%(v/v) Tween-20 in PBS and stored at - 80°C until use. 96-well maxi-sorp-nunc-immuno plates (ThermoFisher Scientific, MA, USA) were coated overnight at 4°C with Anti-CD133 antibody (United States Biological, USA) diluted 1:10 in PBS. After blocking with 3%(w/v) Bovine serum Albumin (BSA) in PBS, 100 μl of diluted sample (1:10) were added per well and incubated overnight at 4°C. Samples were, then, washed five times with PBS and 0.15%(v/v) Tween-20 (PBS-T) and incubated

4 hours with anti-CD133 mouse monoclonal antibody (United States Biological, USA, clone 2F8C5) diluted 1:1000 with 1%BSA(w/v) in PBS-T. After three washes with PBS-T, conjugated HRP anti-mouse IgG diluted 1:5000 were added and incubated 1 hour. Samples were washed again three times with PBS-T and one time with PBS before adding the peroxidase substrate (TMB, Bio-Rad). The reaction was stopped with a  $H_2SO_4$ solution. Absorbance at 450 nm was measured using Mark microplate Absorbance Spectrophotometer (Bio-Rad). To standardize the response of the antibodies, a pool of highly positive control was used. The optical density results were expressed as Relative Unit per ml (RU/ml).

#### *Statistical analysis*

After normalization, mass spectrometry data were analyzed by unsupervised hierarchical clustering using multidimensional scaling (MDS) with k-means and Spearman's correlation to identify outliers and the dissimilarity between samples. The normalized expression profiles of the proteins were then used to construct the co-expression network using the weighted gene co-expression network analysis (WGCNA) package in R [6]. A weighted adjacency matrix was constructed using the power function. After choosing the appropriate β parameter of power (with the value of independence scale set to 0.8) the adjacency matrix was transformed into a topological overlap matrix (TOM), which measures the network connectivity of all the proteins. To classify proteins with co-expression profiles into protein modules, hierarchical clustering analysis was conducted according to the TOM dissimilarity with a minimum size of 30 proteins per module. To identify the relationship between each module and clinical trait, we used module eigengenes (MEs) and calculated the correlation between MEs and each clinical trait and their statistical significance

corrected for multiple interactions. A heat map was then used to visualize the degree and significance of each relationship.

To identify the hub proteins of modules that maximize the discrimination between the selected clinical traits, we applied a non-parametric Mann–Whitney U test, machine learning methods such as non-linear support vector machine (SVM) learning, and partial least squares discriminant analysis (PLS-DA). For the Mann–Whitney U test, proteins were considered to be significantly differentially expressed between the two conditions with power of 80% and an adjusted P-value ≤0.05 after correction for multiple interactions (Benjamini-Hochberg) and a fold change of ≥2. In addition, the proteins needed to show at least 70% identity in the samples in one of two conditions. Volcano plots were used to visualize this analysis. In SVM learning, a fourfold cross-validation approach was applied to estimate the prediction and classification accuracy. Besides, the whole matrix was randomly divided into two part. One for learning (65%) and the other one (35%) to verify the accuracy of the prediction. Finally, the resulting core panel of hub proteins was uploaded to Cytoscape to construct a protein–protein interaction network and identify the principal biological processes and pathways involved in the modules corresponding to each clinical trait. Gene Ontology (GO) annotations were extracted from the UniProt, Reactome, KEGG and ClueGO databases and presented on a heat map and twodimensional scatter plot.

For ELISA data analysis, the Kruskal-Wallis test was used to assess differences in CD133 protein levels among the three study groups, and the results were expressed as medians and IQ ranges. A value of  $P < 0.05$  was considered to be statistically significant. Received operating characteristic (ROC) curves were generated to assess the diagnostic efficiency of each assay. Youden's index was used to identify the cutoff [7]. All statistical tests were performed using the latest version of software package *R* available at the time of the experiments.

8



## **Supplemental Figure 1. Age and eGFR of the study participants.**

Dot plots represent **(A)** age and **(B)** eGFR measured by CKD-EPI Equations of all study participants. Full triangles, squares and circles indicate male subjects while triangles, squares and circles with diagonal stripes indicate female patients. The line represents the mean value.





Plot of exosomes **(A)** and microvesicles **(B)** size distribution, as evaluated by dynamic light scattering. The plot shows a Gaussian distribution profile with a mean peak at 90  $\pm$  5 nm or 1000  $\pm$  70 nm respectively for exosomes or microvesicles. No statistical differences were observed between the exosomes or microvesicles isolated from the urine of medullary sponge kidney (MSK) and autosomal dominant polycystic kidney disease (ADPKD) patients. **(C)** Representative western blot analysis of exosomes and microvesicles isolated from the urine of medullary sponge kidney (MSK) and autosomal dominant polycystic kidney disease (ADPKD) patients. Whole exosomes (lanes 1–2, 5–6 and 9–10) and microvesicles (lanes 3–4, 7–8 and 11–12) from MSK (lanes 1, 3, 5, 7, 9 and 11) and ADPKD (lanes 2, 4, 6, 8, 10 and 12) patients were analyzed by detecting CD63 (lanes 1–4), CD81 (lanes 5–8) and CD45 (lanes 9–12). Stain-free technology was used as loading control.





The pie charts show the distribution of cellular component annotations in the different conditions. The percentage distribution of cellular component categories is similar between the exosomes of medullary sponge kidney (MSK) and autosomal dominant polycystic kidney disease (ADPKD) patients, and between the microvesicles of medullary sponge kidney (MSK) and autosomal dominant polycystic kidney disease (ADPKD) patients.

**Supplemental Figure 4. Multidimensional scaling analysis of extracellular vesicles from the urine of medullary sponge kidney (MSK) and autosomal dominant polycystic kidney disease (ADPKD) patients**.



Two-dimensional scatter plot of MDS analysis of exosomes (solid symbol) and microvesicles (open symbol) of medullary sponge kidney (MSK) (red triangle) and autosomal dominant polycystic kidney disease (ADPKD) (black square) samples. Ellipsis indicates 95% confidence interval. No outliers were detected.



## **Supplemental Figure 5. Sample clustering and trait indicators.**

In the upper panel, the clustering of samples was based on the label-free quantification of proteins identified by mass spectrometry. In the lower panel, the color intensity was proportional to the trait indicator classification, i.e. the type of pathology and extracellular vesicle. No outliers were detected.

**Supplemental Figure 6. Venn diagram of statistically significant differences in protein abundance in the different types of extracellular vesicles from medullary sponge kidney (MSK) or autosomal dominant polycystic kidney disease (ADPKD) patients**.



Venn diagram shows common and exclusive peptides. The numbers represent the distinct proteins in the overlapping and non-overlapping areas.



## **Supplemental Figure 7. Network of proteins interaction.**

Network of proteins interaction. The diagram report all the interaction between the 255 statistically significant proteins. Circles and grey lines correspond respectively, to proteins and their interactions. Circles size and colours are related respectively, to the number of protein interactions and the relative expression, after Z-score, in the four conditions i.e. ADPKD exosomes (dark red) and microvesicles (light red), and MSK exosome (dark blue) and microvesicles (light blue). Besides, the proteins are grouped into four classes (dotted ellipses) in function of the type of proteins interaction and their ability to distinguish the four conditions. These classes correspond from outside to inside, respectively, to proteins: 1) linked by only co-expression and co-localization interactions, 2) with also physical interactions, 3) also linked by biochemical pathway and 4) that maximize the discrimination between the four conditions.

**Supplemental Figure 8. Gene Ontology enrichment analysis for core discriminatory proteins in the extracellular vesicles of medullary sponge kidney (MSK) and autosomal dominant polycystic kidney disease (ADPKD) patients**.



The –log10 (P value) of each term is shown on the x-axis and the enriched GO terms are shown on the y-axis. The size of each circle is proportional to the number of proteins associated with each GO term.

**Supplemental Table 1.** List of all significant proteins identified using mass spectrometry. The symbol "+" identify the significant proteins in each experimental comparison. Fold change value is reported as mean and standard deviation.

























**Supplemental Table 2.** Relative Operating Characteristic curve analysis value of CD133 ELISA assay for the comparison of autosomal dominant polycystic kidney disease (ADPKD) vs medullary sponge kidney (MSK), ADPKD vs Healthy Controls and MSK vs Healthy Controls.



## **References for Supplemental materials**

1. Davis S, Charles PD, He L, Mowlds P, Kessler BM, Fischer R. Expanding proteome coverage with charge ordered parallel ion analysis (CHOPIN) combined with broad specificity proteolysis. J Proteome Res. 2017; 16(3):1288-99

2. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol. 2008;26(12):1367-72

3. Mallol R, Rodríguez MA, Heras M, Vinaixa M, Plana N, Masana L, Morris GA, Correig X. Particle size measurement of lipoprotein fractions using diffusion-ordered NMR spectroscopy. Anal Bioanal Chem. 2012; 402(7):2407-15

4. Inglis HC, Danesh A, Shah A, Lacroix J, Spinella PC, Norris PJ. Techniques to improve detection and analysis of extracellular vesicles using flow cytometry. Cytometry A. 2015; 87(11):1052-63

5. Cossarizza A, Chang HD, Radbruch A, Akdis M, Andrä I, Annunziato F, et al., Guidelines for the use of flow cytometry and cell sorting in immunological studies. Eur J Immunol. 2017; 47(10):1584-1797

6. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9:559

7. Hughes G. Youden's Index and the Weight of Evidence Revisited. Methods Inf Med. 2015;54(6):576-7