

Washington University School of Medicine

Digital Commons@Becker

2020-Current year OA Pubs

Open Access Publications

1-1-2023

Proteins in tumor-derived plasma extracellular vesicles indicate tumor origin

Meltem Barlin

Washington University School of Medicine in St. Louis

Petra Erdmann-Gilmore

Washington University School of Medicine in St. Louis

Jacqueline L Mudd

Washington University School of Medicine in St. Louis

Qiang Zhang

Washington University School of Medicine in St. Louis

Robert W Seymour

Washington University School of Medicine in St. Louis

See next page for additional authors

Follow this and additional works at: https://digitalcommons.wustl.edu/oa_4



Part of the [Medicine and Health Sciences Commons](#)

Please let us know how this document benefits you.

Recommended Citation

Barlin, Meltem; Erdmann-Gilmore, Petra; Mudd, Jacqueline L; Zhang, Qiang; Seymour, Robert W; Guo, Zhanfang; Miessner, Julia R; Goedegebuure, S Peter; Bi, Ye; Osorio, Omar A; Alexander-Brett, Jennifer; Li, Shunqiang; Ma, Cynthia X; Fields, Ryan C; Townsend, R Reid; and Held, Jason M, "Proteins in tumor-derived plasma extracellular vesicles indicate tumor origin." *Molecular and Cellular Proteomics*. 22, 1. 100476 (2023).

https://digitalcommons.wustl.edu/oa_4/2811

This Open Access Publication is brought to you for free and open access by the Open Access Publications at Digital Commons@Becker. It has been accepted for inclusion in 2020-Current year OA Pubs by an authorized administrator of Digital Commons@Becker. For more information, please contact vanam@wustl.edu.

Authors

Meltem Barlin, Petra Erdmann-Gilmore, Jacqueline L Mudd, Qiang Zhang, Robert W Seymour, Zhanfang Guo, Julia R Miessner, S Peter Goedegebuure, Ye Bi, Omar A Osorio, Jennifer Alexander-Brett, Shunqiang Li, Cynthia X Ma, Ryan C Fields, R Reid Townsend, and Jason M Held

Proteins in Tumor-Derived Plasma Extracellular Vesicles Indicate Tumor Origin

Authors

Meltem Barlin, Petra Erdmann-Gilmore, Jacqueline L. Mudd, Qiang Zhang, Robert W. Seymour, Zhanfang Guo, Julia R. Miessner, S. Peter Goedegebuure, Ye Bi, Omar A. Osorio, Jennifer Alexander-Brett, Shunqiang Li, Cynthia X. Ma, Ryan C. Fields, R. Reid Townsend, and Jason M. Held

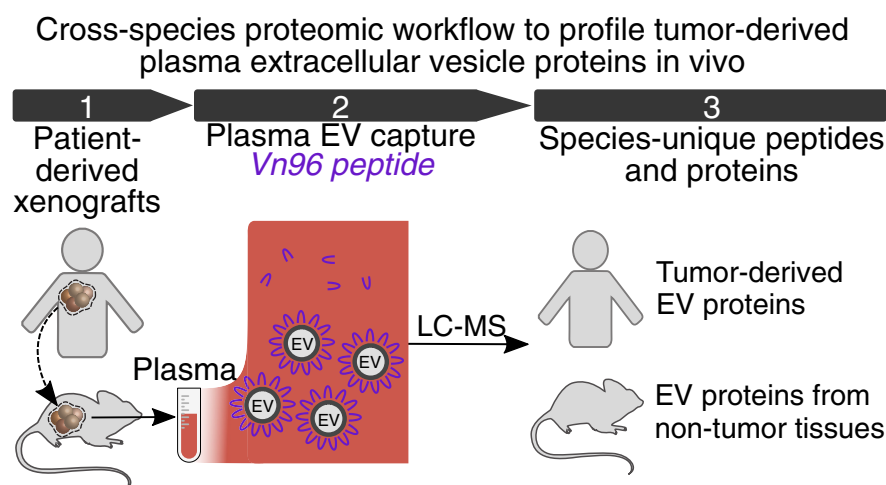
Correspondence

jheld@wustl.edu

In Brief

We report of a new cross-species proteomic approach to profile which proteins patient-derived xenograft tumors package into plasma extracellular vesicles *in vivo*. We find that 0.02 to 0.05% of proteins in plasma extracellular vesicles are derived from the tumor, with some proteins common to all four cancer models evaluated and some proteins expressed uniquely in certain cancer types. Machine learning classification can predict the underlying tumor type with greater than 92% accuracy.

Graphical Abstract



Highlights

- A cross-species profile of the proteins PDXs package into extracellular vesicles.
- About 0.02 to 0.05% of proteins in plasma extracellular vesicles are derived from the tumor.
- Protein packaging into EVs is an intrinsic property of each tumor.
- Machine learning predicted the underlying tumor type with more than 92% accuracy.



Proteins in Tumor-Derived Plasma Extracellular Vesicles Indicate Tumor Origin

Meltem Barlin¹, Petra Erdmann-Gilmore^{1,2}, Jacqueline L. Mudd^{2,3}, Qiang Zhang^{1,2}, Robert W. Seymour¹, Zhanfang Guo¹, Julia R. Miessner^{2,3}, S. Peter Goedegebuure^{2,3}, Ye Bi^{2,3}, Omar A. Osorio¹, Jennifer Alexander-Brett^{1,4}, Shunqiang Li^{1,2}, Cynthia X. Ma^{1,2}, Ryan C. Fields^{2,3}, R. Reid Townsend^{1,2}, and Jason M. Held^{1,2,5,*}

Cancer-derived extracellular vesicles (EVs) promote tumorigenesis, premetastatic niche formation, and metastasis *via* their protein cargo. However, the proteins packaged by patient tumors into EVs cannot be determined *in vivo* because of the presence of EVs derived from other tissues. We therefore developed a cross-species proteomic method to quantify the human tumor-derived proteome of plasma EVs produced by patient-derived xenografts of four cancer types. Proteomic profiling revealed individualized packaging of novel protein cargo, and machine learning accurately classified the type of the underlying tumor.

Plasma extracellular vesicles (EVs) play a functional role in cell–cell signaling, and their bioactive cargoes are sentinels of organismal homeostasis. Since EVs can be readily assayed by noninvasive liquid biopsies, they also have significant potential for disease diagnosis, prognosis, and surveillance, as well as investigating the molecular mechanisms of extracellular communication. In cancer, tumor-derived EVs facilitate premetastatic niche formation and metastasis *via* their protein cargo, which can remodel the extracellular matrix, regulate stromal cell differentiation, alter vascular permeability, suppress immune cells, and guide organotropic metastasis (1–3).

New insights into the EV biology of tumors *in vivo* are limited by the release of EVs by the other tissues in the body. These tumor-independent EVs create a high background, which confounds unambiguous determination of which EV proteins are tumor derived. For this reason, studies typically utilize EVs generated *in vitro* by cell lines for *in vivo* analysis (2). Alternatively, studies analyzing the plasma EV cargo of cancer patients may identify putative cancer biomarkers, but this approach cannot verify that the altered cargo is tumor derived. This caveat is especially relevant when the goal is identifying biomarkers of tumor treatment response since anticancer

therapy also effects nontumor tissues and likely impacts stromal EV biology.

Proteomic profiling of EVs has greatly improved our understanding of EV cargo by providing a more complete and nuanced understanding of EV biology and markers (4). Proteomic analysis facilitated a recent reassessment of exosome composition and discovered that annexin A1 is a specific marker for EVs shed from the plasma membrane (5). Proteomic profiling also contributed to the recent discovery of exomeres, an exosomal subpopulation of nonmembranous nanoparticles approximately 35 nm in size (6). Since there are no agreed upon protein markers present or absent in all EV subpopulations (7), global protein profiling provides a more complete picture of EVs than candidate approaches such as Western blotting.

The chemical information captured by mass spectrometry (MS)–based proteomic technologies can be coupled with bioinformatics to uniquely answer a range of important biological questions. For example, stable isotope–labeled amino acids can be used to encode quantitative information about protein expression (8, 9), chemical labeling (10), or investigating aspects of protein biochemistry such as protein turnover (11). Mass shifts can indicate the presence and position of post-translational modifications (12). In addition, the amino acid sequence information provided by MS can be used to deconvolute the proteomes of multiple species present in a single location or sample. This cross-species proteomic approach has broad applications for metaproteomics (13), but it has also been used to investigate cancer xenograft models where human-specific peptide and protein sequences are derived from the tumor and murine-specific sequences are from the mouse stroma. This approach has been used to investigate how tumors educate the stroma (14) and the integrated response of the tumor and stroma to anticancer drugs (15). One of the biggest challenges of cross-species

From the ¹Department of Medicine, ²Siteman Cancer Center, ³Department of Surgery, ⁴Department of Pathology and Immunology, and ⁵Department of Anesthesiology, Washington University School of Medicine in St Louis, St Louis, Missouri, USA

*For correspondence: Jason M. Held, jheld@wustl.edu.

analysis is bioinformatically assigning taxon-specific peptides, but several software tools have been recently developed to facilitate species-specific deconvolution (16, 17).

In the present study, we utilized cancer patient-derived xenograft (PDX) models to determine which proteins are packaged by patient tumors *in vivo* into EVs using an approach that integrates plasma EV enrichment, cross-species proteomics, and machine learning-based classification. We proteomically profiled 14 PDX across four cancer types (supplemental Table S1), and tumors were derived from a mix of primary tumors and those of metastatic origin in the patients. The method uses a commercial EV isolation kit and open source software tools to streamline and maximize its accessibility to investigate many aspects of tumor EV biology *in vivo*.

EXPERIMENTAL PROCEDURES

Experimental Design and Statistical Rationale

PDX Generation and Growth—Animal studies were performed under an approved animal studies protocol at the Washington University School of Medicine. Sample N and biological replicates are described in supplemental Table S1.

For breast PDX generation, female homozygous nude mice (Charles River Laboratories; catalog no.: 088) were injected with 1×10^6 patient-derived breast tumor cells mixed with an equal volume of Matrigel media (BD Biosciences; catalog no.: 354234) and 10% fetal bovine serum (Fisher Scientific; catalog no.: Mt35010CV) in RPMI (Fisher Scientific; catalog no.: SH30027LS) into the fourth mammary fat pad.

Pancreatic ductal carcinoma (PDAC), colorectal cancer (CRC), and melanoma PDXs were generated in female NSG NOD.Cg-Prkdc^{scid} IL2rg^{tm1Wjl}/SzJ mice (Jackson Laboratory; catalog no.: 005557). To establish tumor growth, mice were anesthetized with isoflurane, a subcutaneous nick was made in each flank, and a small tumor fragment coated with Matrigel was transferred into each subcutaneous pocket (implanted tumor size: 1 mm \times 1 mm \times 1 mm). Nicks were closed with a small amount of GLUture topical adhesive.

Plasma Collection—Avertin (Sigma-Aldrich; catalog no.: T48402-25G) at a dose of 500 to 1000 mg/kg or ketamine (100 mg/kg) plus xylazine (10 mg/kg) was given intraperitoneally for cardiac puncture blood collection with either 25G needles (Fisher Scientific; catalog no.: 14-829-2C) or 26G needles (Fisher Scientific; catalog no.: 14-823-2E). At least 200 μ l of peripheral blood was collected in K3 EDTA tubes (Sarstedt; catalog no.: 41.1504.105), centrifuged at 1.5 to 2g, 4 °C for 10 min, and immediately transferred into Nalgene Cryogenic Tubes at room temperature (RT; Thermo Fisher Scientific; catalog no.: 5000-1020) and stored at -80 °C. Plasma was collected from PDX models no sooner than the second serial passage to minimize the influence of residual nondividing and nontumorigenic stroma on analysis.

EV Preparation—EV capture was performed using the ME kit from New England Peptide. Manufacturer instructions were followed with the following modifications. Aliquots of frozen mouse plasma were thawed on ice, vortexed, and spun at 15,000g for 10 min. About 200 μ l of plasma was added to a new 1.7 ml microcentrifuge tube (Axygen; catalog no.: MCT-175-C) containing 200 μ l of PBS (Gibco; catalog no.: 370011044) (10 \times , pH 7.4; diluted to 1 \times with LC-MS grade water) and mixed by gentle inversion. EV precipitation was initiated by addition of 8 μ l Vn96 peptide stock (prepared per manufacturer instructions) followed by mixing *via* inversion and incubation at RT with end-over-end rotation for 1 h. Samples were centrifuged at 17,000g for 15 min at

4 °C to collect the EVs at the bottom of the tube. The supernatant was carefully removed, and the pellet was resuspended in 400 μ l of PBS. The samples were spun again at 17,000g for 15 min at 4 °C. The EV pellet was washed two more times with PBS as described previously. After the last spin, the pellets were stored at -80 °C until peptide preparation. All steps were performed at RT unless otherwise noted. The default brake settings of the centrifuge were used.

EV Size Analysis—Dynamic light scattering (DLS) was used to analyze size distribution of EVs isolated from mouse serum. Following precipitation with Vn96 peptide, the pellet was solubilized in PBS/8% trehalose (from 30% w/v trehalose dihydrate stock; Hampton Research) with incubation in a Branson B200 ultrasonic bath for 10 min. Sample was then centrifuged at 14,000g and analyzed on a Malvern NanoS instrument at 25 °C. A representative intensity profile is shown in supplemental Fig. S1. Cumulant size (intensity-weighted Z-average) was reproducible between runs, and polydispersity was low, which indicated a relatively homogeneous distribution of particles. These results are consistent with measurements from cell line-derived EVs purified by size-exclusion chromatography and validated by transmission electron microscopy (18).

Peptide Preparation—Samples were digested as previously described (19, 20) using a modification of the filter-aided sample preparation method (21). The EV pellets were solubilized with 30 μ l SDS buffer (4% [w/v] SDS, 100 mM Tris-HCl [pH 8.0]). The samples were reduced by addition of 50 mM DTT with heating to 95 °C for 10 min. The reduced samples were mixed with 200 μ l of 100 mM Tris-HCl buffer, pH 8.5 containing 8 M urea (UA buffer), transferred on top chamber of a 30,000 molecular weight cutoff filtration unit (Millipore; catalog no.: MRCF0R030), and spun in a microcentrifuge at 14,000g for 10 min. An additional 200 μ l of 100 mM Tris-HCl buffer, pH 8.5, containing 8 M urea (UA buffer) was added to the top chamber of the filter unit, and the filter was spun at 14,000g for 15 to 20 min in a microcentrifuge (Eppendorf 5424; Eppendorf, catalog no.: 2231000767). The flow through was discarded, and the proteins were alkylated by addition of 100 μ l of 50 mM iodoacetamide (Pierce; catalog no.: A39271) in UA buffer to the top chamber of the filtration unit and gyrating at 550 rpm in the dark at RT for 30 min using a thermomixer (Thermomixer R; Eppendorf). The filter was spun at 14,000g for 15 min, and the flow through was discarded. Unreacted iodoacetamide was washed through the filter with two sequential additions of 200 μ l of 100 mM Tris-HCl buffer, pH 8.5 containing 8 M urea, and centrifuged at 14,000g for 15 to 20 min after each addition. The urea buffer was exchanged into digestion buffer (DB), 50 mM ammonium bicarbonate buffer, pH 8. Two sequential additions of DB (200 μ l) with centrifugation after each addition to the top chamber were performed. The top filter units were transferred to a new collection tube, 100 μ l DB containing 1 μ AU of LysC (Wako Chemicals; catalog no.: 129-02541) was added, and samples were digested at 37 °C. After 2 h of LysC digestion, 1 μ g of sequencing-grade trypsin (Promega; catalog no.: V5113) was added, and samples were digested overnight at 37 °C. The filters were spun at 14,000g for 15 min to collect the peptides in the flow through. The filter was washed with 50 μ l 100 mM ammonium bicarbonate buffer, and the wash was collected with the peptides. In preparation for desalting, peptides were acidified to pH 2 with 1% (v/v) TFA. The peptides were desalted using two microtips (porous graphite carbon; Glygen BIOMEKNT3CAR) on a Beckman robot (Biomek NX), as previously described (22). The peptides were eluted with 60 μ l of 60% (v/v) acetonitrile in 0.1% TFA (v/v) and dried in a Speed-Vac (Thermo Fisher Scientific, model no.: Savant DNA 120 concentrator) after adding TFA to 5% (v/v). The peptides were dissolved in 20 μ l of 1% (v/v) acetonitrile in water. An aliquot (10%) was removed for quantification using the Pierce Quantitative Fluorometric Peptide Assay kit (Thermo Fisher Scientific; catalog no.: 23290). The remaining peptides were transferred to autosampler vials (Sun-Sri; catalog no.:

200046), dried, and stored at -80°C for LC–MS analysis. The average peptide yield across samples was $31.5\ \mu\text{g}$ per ml of plasma.

Lipid Quantitation—Lipid quantitation was performed using the SPV assay per the manufacturer's instructions (Cell Biolabs; catalog no: STA-613). The average lipid yield across samples was $59.6\ \mu\text{g}$ per ml of plasma, and the average peptide:lipid ratio was 0.53.

NanoLC–MS/MS—The samples in formic acid (FA; 1%) were loaded ($2.5\ \mu\text{l}$) onto a $75\ \mu\text{m}$ i.d. \times $50\ \text{cm}$ Acclaim PepMap 100 C18 RSLC column (Thermo Fisher Scientific) on an EASY nanoLC (Thermo Fisher Scientific) at a constant pressure of 700 bar at 100% mobile phase A (1% FA). Prior to sample loading, the column was equilibrated to 100% mobile phase A for a total of $11\ \mu\text{l}$ at 700 bar pressure. Peptide chromatography was initiated with mobile phase A (1% FA) containing 2% mobile phase B (100% acetonitrile [MeCN] and 1% FA) for 5 min, then increased to 20% B over 100 min, to 32% B over 20 min, to 95% B over 1 min and held at 95% B for 19 min, with a flow rate of 250 nL/min. The data were acquired in data-dependent acquisition mode. The full scan mass spectra were acquired with a Q-Exactive mass analyzer with a scan range of $m/z = 325$ to 1500 and a mass resolving power set to 70,000. Ten data-dependent high-energy collisional dissociations were performed with a mass resolving power set to 17,500, a fixed lower value of m/z 100, an isolation width of 2 Da, and a normalized collision energy setting of 27. The maximum injection time was 60 ms for parent-ion analysis and product-ion analysis. The target ions that were selected for MS/MS were dynamically excluded for 15 s. The automatic gain control was set at a target value of $1\text{e}6$ ions for full MS scans and $1\text{e}5$ ions for MS2. Peptide ions with charge states of 1 or >8 were excluded for high-energy collisional dissociation acquisition.

Protein Identification—LC–MS data were searched against MaxQuant (<https://www.maxquant.org/>) search engine (23) (version 1.6.17.0). MaxQuant was set to search against a concatenated UniProt (version March 2020) database of human (20,365 entries), mouse (17,033 entries), and common contaminant proteins (cRAP; version 1.0; January 1, 2012; 116 entries). Enzyme cleavage specificity was trypsin/P with a maximum of four missed cleavages allowed. The MS2 database searches were performed with a fragment ion mass tolerance of 20 ppm and a parent ion tolerance of 20 ppm. Carbamidomethylation of cysteine was specified in MaxQuant as a fixed modification. Deamidation of asparagine, formation of pyroglutamic acid from N-terminal glutamine, acetylation of protein N terminus, oxidation of methionine, and pyrocarbamidomethylation of N-terminal cysteine were specified as variable modifications. Peptides and proteins were filtered at 1% false discovery rate by searching against a reversed database. Peptides not detected in all three replicates of at least two samples were filtered out to ensure robust detection. The ratios of peptides were calculated in relative to the MaxQuant label-free quantitation (LFQ) intensities of peptides across all samples (24), and the medians were taken to represent protein ratios. The ratios of peptides and proteins were then transformed such that the median under each sample was zero on a \log_2 scale. The offsets used in the median centering of peptide and protein ratios were applied to scale intensity values accordingly.

LC–MS Data Analysis—ProteoClade (16) was used to assign peptides as human unique, mouse unique, or species shared. Species-shared peptides were removed for analysis since the tissue origin is ambiguous. Genes assigned by detection of only a single gene-unique peptide sequence were confirmed and required to have either been detected by (1) either top four most observed peptides in the gene in the PeptideAtlas database (25) or (2) at least 5000 total observations of the peptide in PeptideAtlas. The MS1 signal of human peptides was quantified using Skyline (26) for the most accurate LFQ results. Peptides with low signal intensity or interferences were removed, as previously described (10). Skyline MS1 quantitation was

used in Figures 1, E, F, and 2. The heatmap was generated with the Seaborn package using default parameters with Python 3.9.6. Genes were matched to the ExoCarta database (27) for assignment in supplemental Table S2.

Linear discriminant analysis (LDA) and machine learning were performed using both Python 3.9 and R 4.1. Human protein expression was first standardized by mean centering and scaled to unit variance using the StandardScaler function of scikit-learn 1.0. The LDA plot was generated using the MASS package (28), and classification was performed using Auto-SKlearn 2.0 (29) with only the LDA classifier as detailed in the code supplement. About 25% of the dataset was held out for validation and not used in model generation, and accuracy was determined using the `sklearn.metrics.accuracy_scoring` function.

RESULTS

VN96 Captures EV Subpopulation Ranging from Exomeres to Microvesicles

Our approach to identify *bona fide* human proteins packaged into plasma EVs by patient tumors *in vivo* utilizes cross-species proteomic profiling of plasma EVs captured with the VN96 peptide (30) in patient-derived xenograft (PDX) mouse models (Fig. 1A). After parsing the amino acid sequence of the identified peptides by species (16), human-unique peptide sequences detected in EVs verify the protein's origin to be the tumor and not the stroma or secondary mouse tissues (14). At least three biological replicates of 14 PDX lines were profiled. PDXs were generated from a mix of primary tumors and those of metastatic origin in the patient across four cancer types: breast, PDAC, CRC, and melanoma (supplemental Table S1). Tumor attributes are listed in supplemental Table S1. Six nontumor-bearing “no PDX” mice were assayed as controls.

The size of isolated plasma EVs enriched by the VN96 EV capture technology was determined by DLS. A major advantage to DLS is that it samples a broad range of vesicle sizes (1–10,000 nm diameter), which may be present in complex specimens. Most EVs were between 35 and 400 nm with a peak near 165 nm (supplemental Fig. S1). This size distribution indicated robust enrichment of all EV subpopulations including exomeres, exosomes, and microvesicles but not larger apoptotic bodies or intact cells (5). It is possible that some EVs are not precipitated *via* the VN96 method, but prior studies have found that the proteome of VN96-enriched EVs is similar to those isolated by ultracentrifugation (31).

Proteomic profiling of the enriched EVs assigned 5556 peptides with at least six peptide spectral matches below a 1% false discovery rate across the dataset to 1001 genes (supplemental Tables S1, S2, and supplemental Fig. S2). Common contaminant proteins were filtered out using the CRAPome (Contaminant Repository for Affinity Purification) dataset, such as keratins, which could be introduced by sample handling (32). However, human albumin was not filtered out, as discussed later.

About 65.1% of the identified peptide sequences were mouse unique, and 1.0% were human unique based on

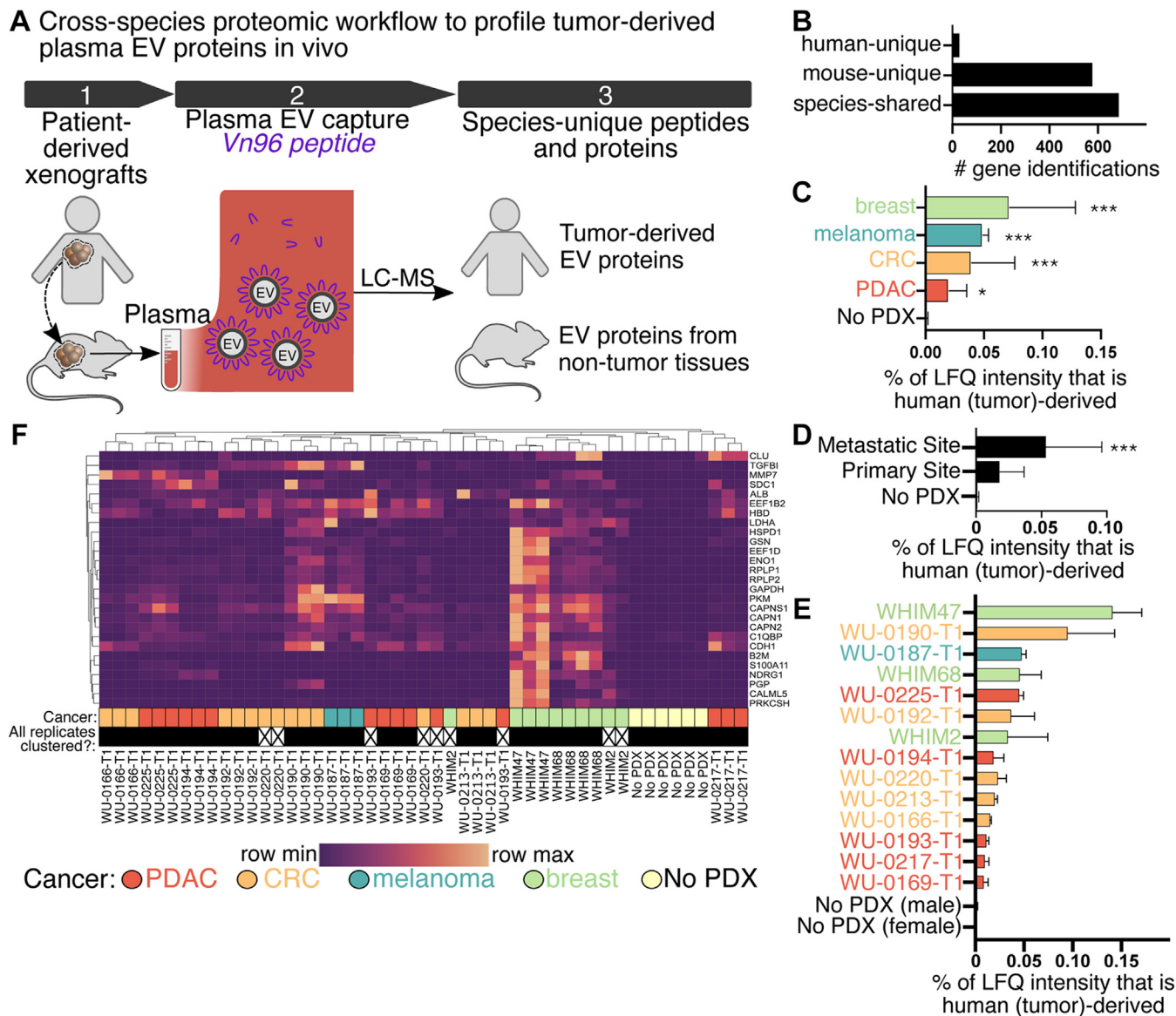


FIG. 1. Patient tumor-derived EV proteins are an intrinsic property of the individual patient tumor and indicative of the tumor type and metastatic origin. *A*, overview schematic of cross-species proteomic profiling of tumor-derived plasma EV proteins. *B*, number of species-specific assignments. *C*, total human-unique (tumor-derived) protein intensity by tumor type, *(D)* tumor site, and *(E)* individual PDX lines. *F*, heatmap of tumor-derived proteins across PDX lines and biological replicates. *Filled black squares* indicate each biological replicates clustered together. * $p < 0.05$, *** $p < 0.005$, based on Kruskal–Wallis one-way ANOVA, and *N* is detailed in [supplemental Table S1](#). EV, extracellular vesicle; PDX, patient-derived xenograft.

cross-species parsing with ProteoClade (16). With regard to gene assignment, 44.6% of peptides were uniquely assigned to mouse, whereas 2.2% were uniquely assigned to human genes (Fig. 1B). Consistent with size analysis, many common EV markers were robustly detected using mouse or species-shared peptides including HSPA5 (BIP), PDCD6IP (ALIX), and HSPA8 (supplemental Table S2). In addition, we detected markers of specific subpopulations such as ANXA1, a specific indicator of EVs shed from the plasma membrane (5), as well as exomere markers including MTHFD1 and IHD1 (6).

The Tumor-Derived EV Proteome Varies by Cancer Type and Individual PDX

Quantitation of species-specific peptides was performed using the MS1 precursor intensities determined by LFQ. Across all 14 PDX lines, an average of 0.041% of the total MS1 signal intensity was from human-unique peptides in EVs isolated from PDX-bearing mice, compared with only 0.001% in plasma EVs from control nontumor-bearing mice. Breast tumors had the highest amount of tumor-derived protein, greater than 0.05% of all EV proteins detected (Fig. 1C). PDAC

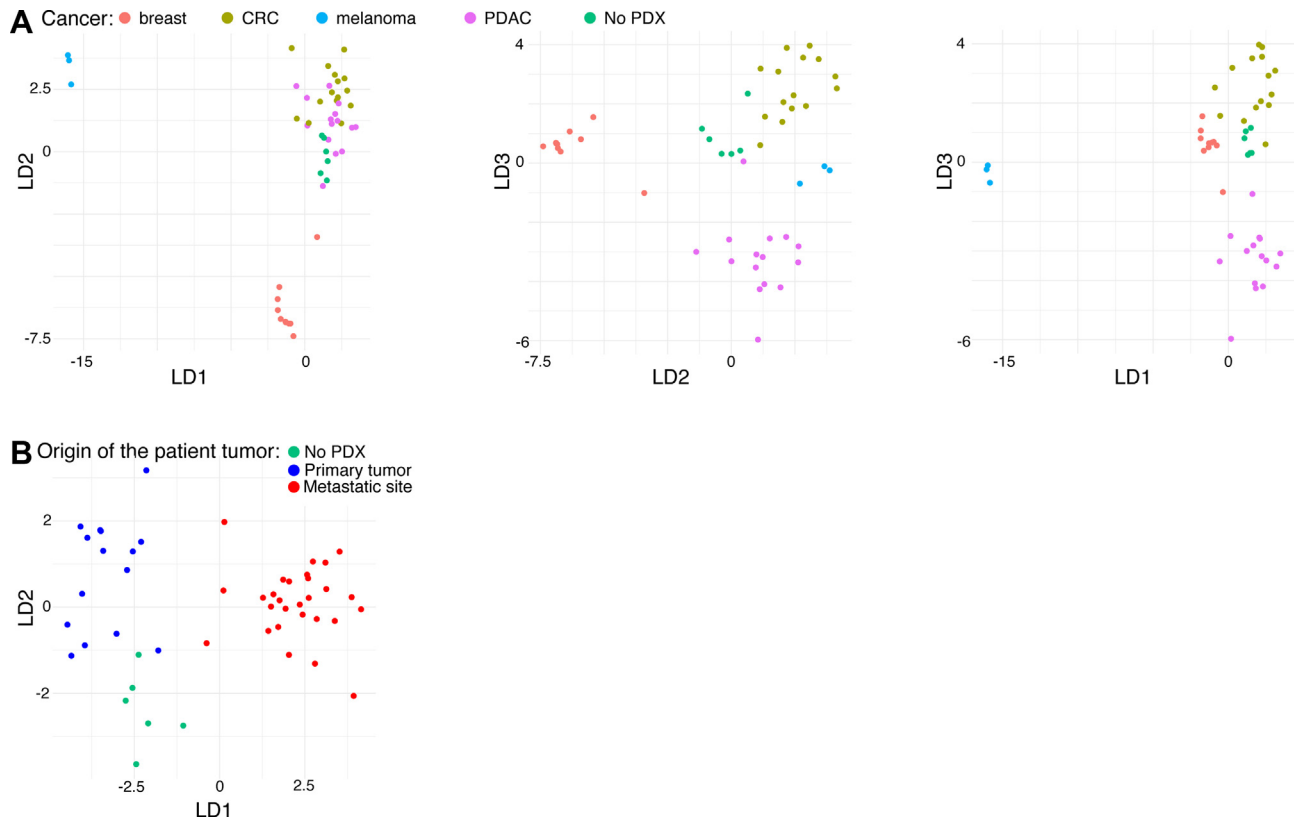


FIG. 2. **Tumor classification via the tumor-derived EV proteome using machine learning.** A, LDA classification of PDXs based on tumor type or (B) origin of the patient tumor. EV, extracellular vesicle; LDA, linear discriminant analysis; PDX, patient-derived xenograft.

tumors produced the fewest tumor-derived EV proteins, on average 0.02% of all EV proteins. PDXs derived from metastatic sites released more EV protein than PDXs derived from primary sites (Fig. 1D), consistent with the known role of EVs in metastasis and formation of the metastatic niche (1–3).

The packaging of proteins into EVs by individual PDXs was consistent across biological replicates. For example, while the overall amount of tumor-derived proteins in EVs varied widely across the entire PDX set, they were similar within biological replicates of each PDX (Fig. 1E). A heatmap of human protein expression demonstrated that all biological replicates for 11 of the 14 PDX clusters next to each other (Fig. 1F). Statistically, the Spearman correlation (r^2) of tumor-derived proteins within replicates of individual PDXs was 0.76, significantly higher than across the PDX cohort (0.42, $p < 0.001$). The consistent packaging of proteins into EVs by individual PDXs likely reflects the intrinsic biology of each tumor, but it is also possible that these patterns are associated with certain molecular subtypes.

We next determined if the tumor-derived EV proteome could be used to classify the underlying cancer present. The open source automated machine learning Python library Auto-Sklearn 2.0 (29) was used to generate models, and standard training and “holdout” testing methodology assessed accuracy (33). To avoid overfitting, we used only the LDA classifier, which predicted the underlying tumor type with greater than 92% accuracy. As

shown in Figure 2, the single LDA classification algorithm is capable of a high degree of separation of tumor types and metastatic origin based on the tumor-derived EV proteome.

One caveat of this classification analysis is that while PDAC and breast cancers were a mix of primary tumors and those of metastatic origin, all CRC and melanoma PDXs were of metastatic origin (supplemental Table S1). It is also possible that the site of PDX implantation plays a role, since the breast PDXs were orthotopic, whereas other PDX models were implanted subcutaneously. However, application of this cross-species proteomics method to additional cancer types and larger xenograft cohorts will likely reveal further insight into the EV biology of cancer subtypes and metastasis.

Expression Patterns of Tumor-Derived Proteins and Their Roles in EVs

Investigating the packaging of individual proteins revealed new insights into tumor-derived EVs *in vivo*. For example, several proteins were pan-cancer markers present in EVs from nearly all PDX models, including ALB, C1QBP, CDH1, and PKM (Fig. 3A). CDH1 has been shown to localize to the surface of tumor EVs and to heterodimerize with vascular endothelial-cadherin on the surface of endothelial cells to promote signaling (34). While previous studies utilized ovarian and prostate cell lines *in vivo*, these results demonstrate that

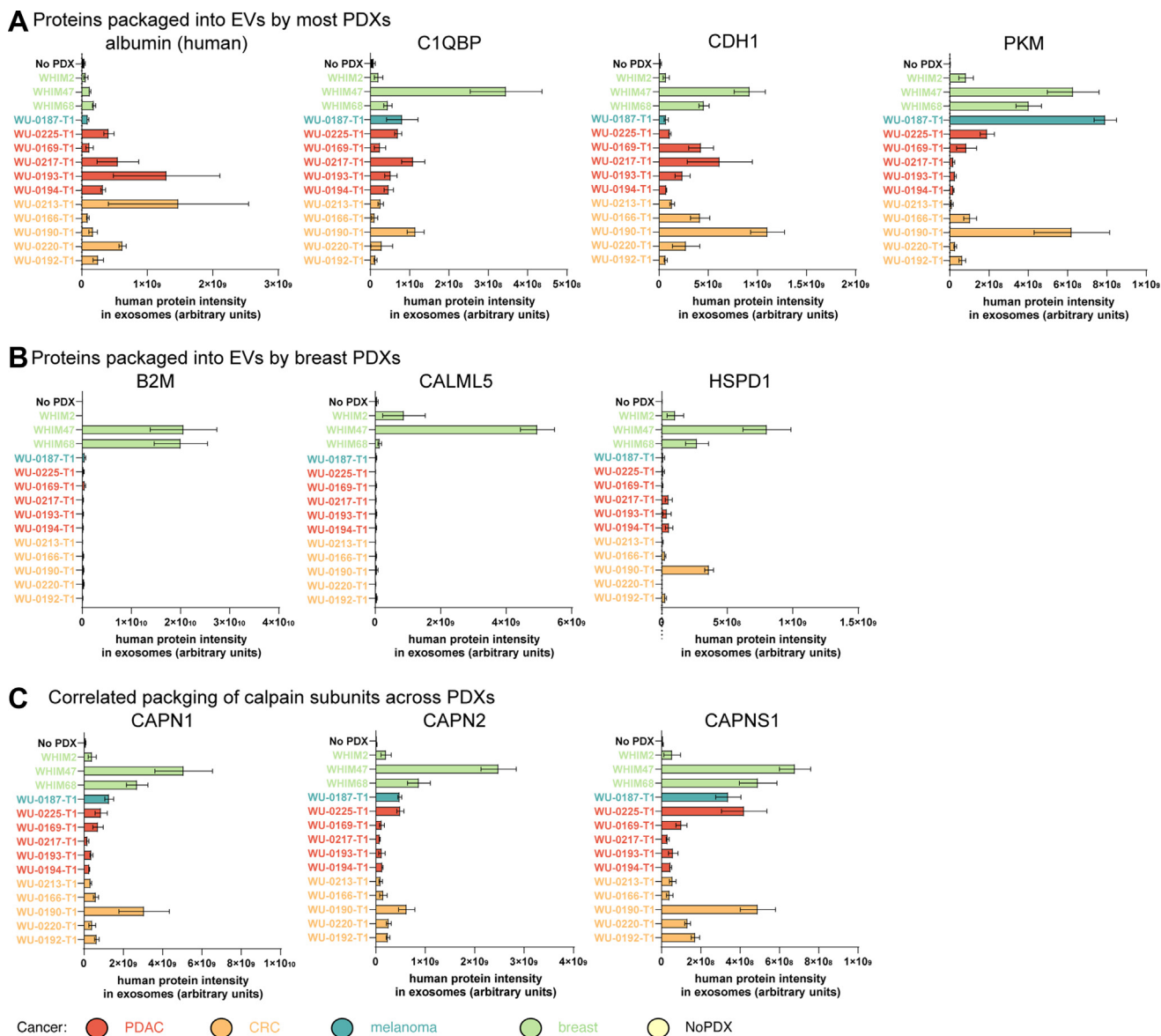


FIG. 3. **Abundance of tumor-derived proteins in EVs reveals pan-cancer markers and cancer-unique indicators.** A, relative human protein abundance of newly identified and known components of cancer EVs that pan-cancer markers present in most tumor-derived EVs. B, human proteins detected from EVs specific for breast cancer. C, relative human protein abundance of members of the calpain protein complex in EVs. N is detailed in supplemental Table S1. EV, extracellular vesicle.

patient-derived tumors of many cancer types generate CDH1-positive EVs.

Albumin's expression in EVs is not well characterized, primarily because albumin is typically filtered out as a nonspecifically bound contaminant. Since our method can distinguish the origin of the albumin, it is able to provide new insight into tumor-dependent packaging of albumin into EVs. We previously confirmed that albumin is expressed in several of these PDX models using cross-species proteomics (16) and thus is available for packaging by the tumor. It has also been shown that albumin is found in distinct EV subpopulations (35). Furthermore, lymphoma and leukemia cell

lines package albumin into small and lipid-like entities that are likely EVs, which inhibit T-cell activation, proliferation, and function (36). Since albumin is an abundant component of nearly all PDX-derived EVs, its immunosuppressive role in EVs (36) may have unanticipated molecular and clinical importance.

The packaging of some proteins was limited to certain cancer subtypes, especially breast cancer (Fig. 3B). For example, B2M, CALML5, and HSPD1 are largely breast cancer-specific markers, but even these differ in their expression across the different three breast cancer PDXs. Many other proteins thought to be highly abundant in EVs also

did not have a pan-cancer expression, including GAPDH, HSPD1, and ENO1 (supplemental Fig. S3).

Tumor-derived proteases in EVs facilitate remodeling of the extracellular matrix to promote the metastatic niche (1). Therefore, we noted the presence of MMP7 and three calpain components, CAPN1, CAPN2, and CAPNS1, as tumor-derived EV protein cargo (Figs. 3C and S3). While calpains are not well characterized in EVs, the highly correlated expression profiles of the three calpain component across the 14 PDXs suggest that these known interaction partners (37) are likely packaged into EVs together preconfigured as a complex (Fig. 3C). In contrast, MMP7 had very different expression pattern, which was almost nonoverlapping with calpains (supplemental Fig. S3). This indicates that cancers package proteases into EVs in a highly individualized manner, which may explain their varied effect on the ECM (1) and also provide new molecular targets to inhibit formation of the metastatic niche and metastasis.

DISCUSSION

We report a new cross-species proteomic approach to detect *bona fide* tumor-derived proteins in plasma EVs. This method uses commercial EV capture reagents and open source software, thus it can be broadly applied to assay cancer EVs in xenograft models. We demonstrate that this approach can identify *bona fide* tumor-derived proteins *in vivo*, including new protein complexes, and allows classification of the cancer type of the underlying PDX tumor with greater than 92% accuracy.

Plasma EVs are frequently used for biomarker studies because of their easy accessibility and the important mechanistic role they play in cancer. However, the EVs found in plasma can be derived from nearly any tissue, not just those associated with disease. This leads to challenges of interpreting plasma EV profiling results, since the tissue of origin is unknown. It also leads to sensitivity challenges, since the tissue or tumor may not produce many exosomes. Our approach provides clear assessment into the tumor origin, but it has several challenges to consider. First, the sensitivity is directly related to EV production by the tumor, a challenge shared by all plasma EV approaches. Second, there is significant species conservation between mouse and human, which limits the number of human-unique peptides. For example, 35% of identified peptides have sequences shared between human and mouse. Together, this resulted in only 2.2% of the identified peptides being assigned to human. On one hand, this limits the sensitivity of the method, but we are able to clearly demonstrate that most plasma EVs are not derived from the tumor, at least in PDX models. This result suggests that many cancer biomarkers in patients may be derived from the stroma as opposed to the tumor itself, though further investigation is needed.

PDXs are gaining popularity to study cancer and can be used to model cancer growth, metastasis, and drug response (38, 39). Further application of this EV proteomics approach could provide new insights into the basic biology and diagnostic utility of

cancer-derived EVs by defining new cancer biomarkers, determining how patient tumors are responding to anticancer therapies, and delineating drug resistance mechanisms.

DATA AVAILABILITY

The proteomic results have been deposited to ProteomeXchange consortium with the dataset ID of PXD028662. To view these data, follow the link at the ProteomeXchange page for the dataset to UCSD's Massive and login as the site instructs using username: MSV000088128_reviewer and password: Held_005. Public access will become available once the article is accepted.

The experiment has been uploaded to EV-Track with entry ID EV220125. This may be accessed via the following EV-TRACK URL: <http://evtrack.org/review.php>. Please use the EV-TRACK ID (EV220125) and the last name of the first author (Barlin) to access our submission.

CODE AVAILABILITY

A viewable Jupyter lab notebook used for data analysis is included as a supplemental file [DataAnalysisCode.pdf](#).

Supplemental data—This article contains [supplemental data](#).

Acknowledgments—The proteomic experiments were performed at the Washington University Proteomics Shared Resource, R Reid Townsend MD, PhD., Director and Robert Sprung, PhD., Co-Director). The expert technical assistance of Dr. Yiling Mi, Alan Davis, and Rose Connors is gratefully acknowledged. The Washington University Proteomics Shared Resource is supported in part by the WU Institute of Clinical and Translational Sciences (NCATS UL1 TR000448), the Mass Spectrometry Research Resource (National Institute of General Medical Sciences grant nos.: P41 GM103422 and R24 GM136766), and the Siteman Comprehensive Cancer Center support grant (National Cancer Institute grant no.: P30 CA091842).

Funding and additional information—This publication was supported by the National Cancer Institute of the National Institutes of Health Award Number: U54 CA224083 (to S. L., C. X. M., R. C. F., and J. M. H.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Author contributions—S. L., C. X. M., R. C. F., R. R. T., and J. M. H. conceptualization; S. L., C. X. M., R. C. F., R. R. T., and J. M. H. methodology; P. E.-G., J. L. M., Z. G., J. R. M., S. P. G., Y. B., O. A. O., J. A.-B., S. L., C. X. M., R. C. F., R. R. T., and J. M. H. investigation; J. A.-B., S. L., C. X. M., R. C. F., R. R. T., and J. M. H. supervision; M. B., Q. Z., R. W. S., and J. M. H. formal analysis; M. B. and J. M. H. writing—original draft.

Conflict of interest—Dr. Li has received license fee from Envigo and research funding from Pfizer, Takeda Oncology, and Zenopharm not associated with this article.

Abbreviations—The abbreviations used are: CRC, colorectal cancer; DB, digestion buffer; DLS, dynamic light scattering; EV, extracellular vesicle; FA, formic acid; LDA, linear discriminant analysis; LFQ, label-free quantitation; MS, mass spectrometry; PDAC, pancreatic ductal carcinoma; PDX, patient-derived xenograft; RT, room temperature.

Received June 13, 2022, and in revised form, October 12, 2022
Published, MCPRO Papers in Press, December 5, 2022, <https://doi.org/10.1016/j.mcpro.2022.100476>

REFERENCES

1. Das, A., Mohan, V., Krishnaswamy, V. R., Solomonov, I., and Sagi, I. (2019) Exosomes as a storehouse of tissue remodeling proteases and mediators of cancer progression. *Cancer Metastasis Rev.* **38**, 455–468
2. Hoshino, A., Costa-Silva, B., Shen, T.-L. L., Rodrigues, G., Hashimoto, A., Mark, M. T., et al. (2015) Tumour exosome integrins determine organotropic metastasis. *Nature* **527**, 329–335
3. Xu, R., Rai, A., Chen, M., Suwakulsiri, W., Greening, D. W., and Simpson, R. J. (2018) Extracellular vesicles in cancer — implications for future improvements in cancer care. *Nat. Rev. Clin. Oncol.* **15**, 617–638
4. Kugeratski, F. G., Hodge, K., Lilla, S., McAndrews, K. M., Zhou, X., Hwang, R. F., et al. (2021) *Quantitative Proteomics Identifies the Core Proteome of Exosomes with Syntenin-1 as the Highest Abundant Protein and a Putative Universal Biomarker*. Springer, US
5. Jeppesen, D. K., Fenix, A. M., Franklin, J. L., Higginbotham, J. N., Zhang, Q., Zimmerman, L. J., et al. (2019) Reassessment of exosome composition. *Cell* **177**, 428–445.e18
6. Zhang, H., Freitas, D., Kim, H. S., Fabijanic, K., Li, Z., Chen, H., et al. (2018) Identification of distinct nanoparticles and subsets of extracellular vesicles by asymmetric flow field-flow fractionation. *Nat. Cell Biol.* **20**, 332–343
7. Witwer, K. W., and Théry, C. (2019) Extracellular vesicles or exosomes? On primacy, precision, and popularity influencing a choice of nomenclature. *J. Extracell. Vesicles* **8**, 213–228
8. Ong, S. E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A., et al. (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell Proteomics* **1**, 376–386
9. Zecha, J., Satpathy, S., Kanashova, T., Avanesian, S. C., Kane, M. H., Clauser, K. R., et al. (2019) TMT labeling for the masses: a robust and cost-efficient, in-solution labeling approach. *Mol. Cell Proteomics* **18**, 1468–1478
10. Behring, J. B., van der Post, S., Mooradian, A. D., Egan, M. J., Zimmerman, M. I., Clements, J. L., et al. (2020) Spatial and temporal alterations in protein structure by EGF regulate cryptic cysteine oxidation. *Sci. Signal.* **13**, eaay7315
11. Price, J. C., Guan, S., Burlingame, A., Prusiner, S. B., and Ghaemmghami, S. (2010) Analysis of proteome dynamics in the mouse brain. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 14508–14513
12. Huang, K., Wu, Y., Primeau, T., Wang, Y.-T., Gao, Y., McMichael, J. F., et al. (2019) Regulated phosphosignaling associated with breast cancer subtypes and druggability. *Mol. Cell Proteomics* **18**, 1630–1650
13. Grassl, N., Kulak, N. A., Pichler, G., Geyer, P. E., Jung, J., Schubert, S., et al. (2016) Ultra-deep and quantitative saliva proteome reveals dynamics of the oral microbiome. *Genome Med.* **8**, 1–13
14. Wang, X., Mooradian, A. D., Erdmann-Gilmore, P., Zhang, Q., Viner, R., Davies, S. R., et al. (2017) Breast tumors educate the proteome of stromal tissue in an individualized but coordinated manner. *Sci. Signal.* **10**, eaam8065
15. Rajeeve, V., Vendrell, I., Wilkes, E., Torbett, N., and Cutillas, P. R. (2014) Cross-species proteomics reveals specific modulation of signaling in cancer and stromal cells by Phosphoinositide 3-kinase (PI3K) inhibitors. *Mol. Cell Proteomics* **13**, 1457–1470
16. Mooradian, A. D., Van Der Post, S., Naegle, K. M., and Held, J. M. (2020) ProteoClade: a taxonomic toolkit for multispecies and metaproteomic analysis. *PLoS Comput. Biol.* **16**, 1–12
17. Saltzman, A. B., Leng, M., Bhatt, B., Singh, P., Chan, D. W., Dobrolecki, L., et al. (2018) GpGrouper: a peptide grouping algorithm for gene-centric inference and quantitation of bottom-up proteomics data. *Mol. Cell Proteomics* **17**, 2270–2283
18. Katz-Kiriakos, E., Steinberg, D. F., Kluender, C. E., Osorio, O. A., Newsom-Stewart, C., Baronia, A., et al. (2021) Epithelial IL-33 appropriates exosome trafficking for secretion in chronic airway disease. *JCI Insight* **6**, e136166
19. Erde, J., Loo, R. R. O., and Loo, J. A. (2017) Improving proteome coverage and sample recovery with enhanced FASP (eFASP) for quantitative proteomic experiments. *Met. Mol. Biol.* **1550**, 11–18
20. Kim, D. I., Jensen, S. C., Noble, K. A., Kc, B., Roux, K. H., Motamedchaboki, K., et al. (2016) An improved smaller biotin ligase for BioID proximity labeling. *Mol. Biol. Cell* **27**, 1188–1196
21. Wiśniewski, J. R., Zougman, A., Nagaraj, N., and Mann, M. (2009) Universal sample preparation method for proteome analysis. *Nat. Met.* **6**, 359–362
22. Chen, Z. W., Fuchs, K., Sieghart, W., Townsend, R. R., and Evers, A. S. (2012) Deep amino acid sequencing of native brain GABAA receptors using high-resolution mass spectrometry. *Mol. Cell Proteomics* **11**, 1–18
23. Cox, J., and Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372
24. Cox, J., Hein, M. Y., Luber, C. A., Paron, I., Nagaraj, N., and Mann, M. (2014) Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol. Cell Proteomics* **13**, 2513–2526
25. Schwenk, J. M., Omenn, G. S., Sun, Z., Campbell, D. S., Baker, M. S., Overall, C. M., et al. (2017) The human plasma proteome draft of 2017: building on the human plasma PeptideAtlas from mass spectrometry and complementary assays. *J. Proteome Res.* **16**, 4299–4310
26. MacLean, B., Tomazela, D. M., Shulman, N., Chambers, M., Finney, G. L., Frewen, B., et al. (2010) Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **26**, 966–968
27. Keerthikumar, S., Chisanga, D., Ariyaratne, D., al Saffar, H., Anand, S., Zhao, K., et al. (2016) ExoCarta: a web-based compendium of exosomal cargo. *J. Mol. Biol.* **428**, 688–692
28. Venables, W., and Ripley, B. (2002) *Modern Applied Statistics with S*, Fourth, Springer, NY
29. [preprint] Feurer, M., Eggensperger, K., Falkner, S., Lindauer, M., and Hutter, F. (2020) Auto-sklearn 2.0: hands-free AutoML via meta-learning. *arxiv*. <https://doi.org/10.48550/arXiv.2007.04074>
30. Ghosh, A., Davey, M., Chute, I. C., Griffiths, S. G., Lewis, S., Chacko, S., et al. (2014) Rapid isolation of extracellular vesicles from cell culture and biological fluids using a synthetic peptide with specific affinity for heat shock proteins. *PLoS One* **9**, e110443
31. Bijnsdorp, I.v., Maxouri, O., Kardar, A., Schelfhorst, T., Piersma, S. R., Pham, T.v., et al. (2017) Feasibility of urinary extracellular vesicle proteome profiling using a robust and simple, clinically applicable isolation method. *J. Extracell. Vesicles* **6**, 1313091
32. Mellacheruvu, D., Wright, Z., Couzens, A. L., Lambert, J. P., St-Denis, N. A., Li, T., et al. (2013) The CRAPome: a contaminant repository for affinity purification-mass spectrometry data. *Nat. Met.* **10**, 730–736
33. Vabalas, A., Gowen, E., Poliakov, E., and Casson, A. J. (2019) Machine learning algorithm validation with a limited sample size. *PLoS One* **14**, 1–20
34. Tang, M. K. S., Yue, P. Y. K., Ip, P. P., Huang, R., Lai, H., Cheung, A. N. Y., et al. (2018) Soluble E-cadherin promotes tumor angiogenesis and localizes to exosome surface. *Nat. Commun.* **9**, 2270
35. Kowal, J., Arras, G., Colombo, M., Jouve, M., Morath, J. P., Prindl-Bengtson, B., et al. (2016) Proteomic comparison defines novel markers to characterize heterogeneous populations of extracellular vesicle subtypes. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E968–E977
36. Graner, M. W., Likhacheva, A., Davis, J., Raymond, A., Brandenberger, J., Romanoski, A., et al. (2004) Cargo from tumor-expressed albumin inhibits T-cell activation and responses. *Cancer Res.* **64**, 8085–8092

37. Huttlin, E. L., Bruckner, R. J., Paulo, J. A., Cannon, J. R., Ting, L., Baltier, K., et al. (2017) Architecture of the human interactome defines protein communities and disease networks. *Nature* **545**, 505–509
38. Mundt, F., Rajput, S., Li, S., Ruggles, K.v., Mooradian, A. D., Mertins, P., et al. (2018) Mass spectrometry-based proteomics reveals potential roles of NEK9 and MAP2K4 in resistance to PI3K inhibition in triple-negative breast cancers. *Cancer Res.* **78**, 2732–2746
39. Huang, K., Li, S., Mertins, P., Cao, S., Gunawardena, H. P., Ruggles, K.v., et al. (2017) Proteogenomic integration reveals therapeutic targets in breast cancer xenografts. *Nat. Commun.* **8**, 14864