Proteomic profiling of NCI-60 extracellular vesicles uncovers common protein cargo and cancer type-specific biomarkers

Stephanie N. Hurwitz[1],*, Mark A. Rider[1],*, Joseph L. Bundy[1],*, Xia Liu[1], Rakesh K. Singh[1], David G. Meckes, Jr.[1]

[1]Department of Biomedical Sciences, Florida State University College of Medicine, Tallahassee, FL, 32306, USA

*These authors contributed equally to this work

Correspondence to: David G. Meckes, Jr., *email:* david.meckes@med.fsu.edu

ABSTRACT

Packed with biological information, extracellular vesicles (EVs) offer exciting promise for biomarker discovery and applications in therapeutics and non-invasive diagnostics. Currently, our understanding of EV contents is confined by the limited cells from which vesicles have been characterized utilizing the same enrichment method. Using sixty cell lines from the National Cancer Institute (NCI-60), here we provide the largest proteomic profile of EVs in a single study, identifying 6,071 proteins with 213 common to all isolates. Proteins included established EV markers, and vesicular trafficking proteins such as Rab GTPases and tetraspanins. Differentially-expressed proteins offer potential for cancer diagnosis and prognosis. Network analysis of vesicle quantity and proteomes identified EV components associated with vesicle secretion, including CD81, CD63, syntenin-1, VAMP3, Rab GTPases, and integrins. Integration of vesicle proteomes with whole-cell molecular profiles revealed similarities, suggesting EVs provide a reliable reflection of their progenitor cell content, and are therefore excellent indicators of disease.

INTRODUCTION

Extracellular vesicles (EVs) represent a diverse population of communication pods released from cells. Ranging in size from 40 to 1000 nm, EVs include exosomes, microvesicles, and apoptotic bodies. Generally, exosomes are described as 40-150 nm endocytically-derived vesicles formed by intraluminal budding of multivesicular bodies (MVBs) which are released following MVB fusion with the plasma membrane. Microvesicles are generally larger than exosomes, and are shed by budding and fission events directly at the cell membrane. Varying in size, apoptotic bodies are formed by plasma membrane blebbing, and can contain packaged organelles following initiation of cell death. Further sub-populations of vesicles likely exist, reflecting the heterogeneity of cellular biology encapsulated by EVs.

Extracellular vesicles have been implicated in a number of different physiological processes, including immune system modulation, cell-to-cell signaling, and cell proliferation [1–5]. Accumulating evidence has implicated EVs as major players in the growth, invasion, and metastatic capacity of cancer cells [6, 7]. For example, exosomes and microvesicles have been demonstrated to transfer oncoproteins and nucleic acids from virally-infected cells to uninfected neighboring cells, and likely promote viral-associated tumor progression [8–11]. Systemic circulation of EVs can play a role in establishing a tumor microenvironment, providing the "soil" for cancer cell "seeding" to metastatic sites [12] and cancer patients have been shown to have increased levels of circulating EVs [13–15]. Recent research has begun to characterize specific transmembrane proteins responsible for targeted vesicle uptake by cells in common cancer type-specific metastatic sites [16]. These circulating vesicles therefore reflect a diverse form of intercellular communication that can facilitate the progression of neoplastic growth and tumor metastasis.

Though EVs likely contribute to the progression of some types of cancer, our knowledge of vesicular communication is still incomplete. One reason is due to the heterogeneity of EV sub-populations. Though

the complexity of EV populations has burdened our understanding of their roles in cell biology, the existence of a variety of EVs may be beneficial in the context of cancer diagnostics and prognostics. One challenge in diagnosing cancer is that tumors often represent a diversity of cell types and genetic mutations; a tissue biopsy is limited in its ability to reflect this diversity. However, circulating EVs are derived from an all-inclusive population of cells, and therefore have the potential to more accurately reflect the entirety of a heterogeneous tumor [17]. While some limitations of vesicle-based biopsies exist, such as disparities in EV quantities released from different tumor cells and the ability to detect small changes in the populations of circulating vesicles, EV-based detection offers alternatives to current diagnostic approaches. Current screening or monitoring tests of cancer progression, such as prostate-specific antigen (PSA; prostate), CA-125 (ovarian), alpha fetoprotein (liver), or CA19-9 (pancreatic) often lack the sensitivity or specificity to provide highly accurate clinical diagnoses [18–20]. As EVs provide membrane-bound protected cargo that can reflect cell-specific pathological processes, we and others propose that these vesicles bear great potential as circulating biomarkers that could improve the current strategies of cancer diagnosis [21–24].

A better understanding of the contents of these vesicles is crucial to the development of EV clinical applications. A current limitation in proteomic analyses of EVs from cancer cells is the narrow number of cell lines studied using comparable and reproducible methods. Here, we characterize 60 diverse human cancer cells derived from 9 distinct tissue types from the National Cancer Institute (NCI-60). The NCI-60 panel was originally compiled by the Developmental Therapeutics Program for high-throughput drug screening, and has led to a number of successful chemotherapeutic drugs used to treat cancer patients [25]. The NCI-60 has also contributed vastly to a better understanding of cancer cell biology and the identification of many novel oncogenic DNA mutations [26, 27]. Since then, the panel has become publicized for cancer research purposes, and a full whole-cell proteomic analysis of each individual cell line has been published [28]. Proteomic and RNA analyses of EVs from subsets of the NCI-60 panel have recently been investigated, providing initial characterizations of cancer vesicle contents [29–32]. Research using cell lines from the NCI-60 panel has also contributed to evidence demonstrating the roles of EVs in the growth and survival of tumor cells, multidrug resistance [33–35], immune evasion [36], cancer cell migration [37] and impact on cells in the tumor microenvironment [38]. Subsets of the panel have also been used to study general mechanisms of EV biogenesis and release from cells [39, 40]. Recently, we compared the vesicle secretion of NCI-60 cell lines using nanoparticle tracking analysis (NTA). Results highlighted differences in secretion rates and sizes of vesicles from cancer cells [41].

With this in mind, we conducted a comparative analysis of proteins from EVs secreted by the NCI-60 cells. To our knowledge, this is the largest single-study proteomic investigation of vesicles to date. In this study, 6,071 unique proteins were identified, including 213 common to all 60 cell types, which likely reflect the common machinery involved in EV biogenesis. Differentially expressed proteins were also identified. Many of these proteins are associated with tissue type, and could therefore serve as markers of EV origin or aid as future diagnostic biomarkers of cancers. To investigate proteins involved in mechanisms of EV biogenesis and secretion, the EV proteome was further analyzed to look for associations between protein accumulation (spectral counts) and vesicle secretion quantity. Finally, the proteomic analysis of NCI-60 EVs was compared to existing cellular proteome and transcriptome datasets. These analyses revealed that the EV proteome closely reflects the transcriptome and proteome of the cell of origin, supporting the hypothesis that EVs are a rich source of diagnostic and prognostic markers. Overall, this extensive proteomic dataset provides a foundation to further investigate general mechanisms of vesicle biogenesis, and demonstrates the incredible biomedical and clinical utilities of extracellular vesicles.

RESULTS AND DISCUSSION

Cancer-cell derived EVs contain core vesicle machinery

To characterize and compare extracellular vesicle proteomes, EVs were harvested from 60 cell lines (NCI-60). As pure EV sub-populations are empirically difficult to isolate, a method of enriching a broad spectrum of vesicles was used in this study to contribute to a greater understanding of global EV content. We have previously demonstrated that the ExtraPEG method enriches for EVs with a comparable purity [42] to sucrose-purified samples following growth of cells for a period of time in serum-free medium, as performed in this study [43] (Figure 1A). Vesicles harvested using this method were found to contain extracellular vesicle markers by western blot [41, 43]. Nanoparticle tracking and electron microscopy revealed sizes and morphology consistent with those previously reported for EVs [41, 43]. Here, vesicles were enriched using identical methods from individual cell lines across nine represented histological origins: breast, brain (CNS), colon, kidney, leukemia, lung, melanoma, ovary, and prostate. Collectively, 6,071 unique proteins were identified in EVs (Supplementary Table S1). To examine the overlap between known vesicular proteins, NCI-60 EV proteins were compared to those in the Vesiclepedia compendium of extracellular vesicle molecular data [44]. Nearly 4,500 proteins were previously identified in EVs (Figure 1B and Supplementary Table S2). Over 1,500

proteins not previously characterized as EV components were further discovered. Because so many unreported vesicular proteins were identified, we aimed to ensure that proteins found in this study were congruent with those previously found in EVs. To increase the stringency of our dataset and characterize common EV proteins, only those identified in at least two-thirds all cell line isolates ([NCI-60]$_{stringent}$) were compared to the Vesiclepedia compendium. These proteins showed over 97% overlap with proteins currently characterized as extracellular vesicle proteins.

The entire proteome of NCI-60-derived EVs was further characterized systematically using qualitative and quantitative analyses. An average of nearly 1,900 proteins were identified per cell line across all tissue types (Figure 1C). Calculation of the median logarithmic abundance of proteins revealed a normal distribution of spectral counts across the panel (Supplementary Figure 1D). The number of total EV proteins found within each tissue was similar across the panel (Figure 1D), with the exception of the prostate cancer group, likely explained by the underrepresentation of this tissue type (n=2). Strikingly, 213 proteins were identified that were common to every vesicle sample, representing the core NCI-60 EV proteome (Supplementary Table S3). These proteins likely reflect essential proteins packaged into EVs from many different cellular origins, and provide insight into general mechanisms of EV biogenesis, entry, protein trafficking, and secretion. Proteins found in at least one cell line from each exclusive tissue type were also compiled to represent tissue-specific markers (Supplementary Table S4). Notably, 165 proteins were exclusively found in leukemia-derived EVs, while fewer unique proteins were in other tissue types. As EVs from many cancer cells have been shown to be enriched in functional integrins [16, 45], these data may reflect detectable differences between cancer EVs originating from circulating hematopoietic cells versus those typically attached to basement membrane matrices through integrin linkers.
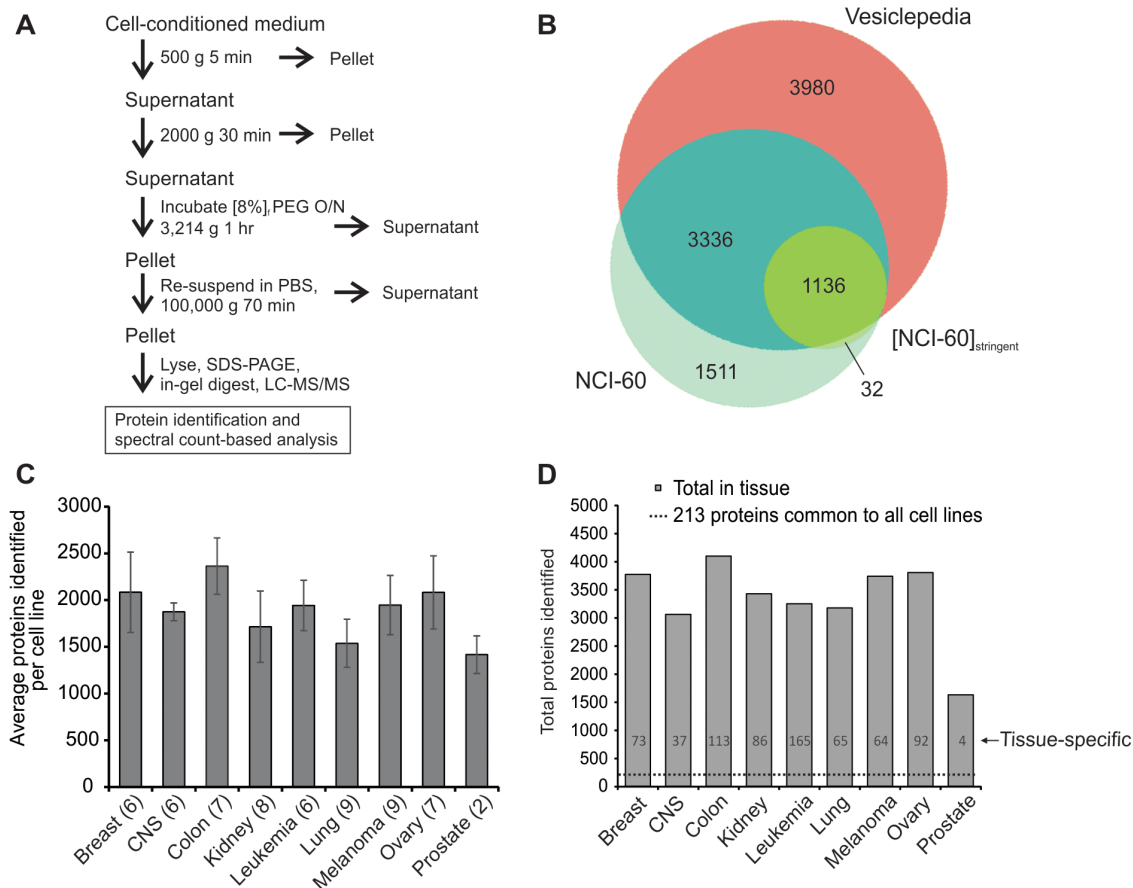

Figure 1: Proteomic analysis of extracellular vesicles secreted by the NCI-60 cells. A. Centrifugation protocol and general workflow of EV enrichment for LC-MS/MS analysis. **B.** Venn diagram of proteins identified in EV samples in the NCI-60 and NCI-60]$_{stringent}$ datasets compared to the Vesiclepedia database of proteins. See also Supplementary Table S2. **C.** Average spectral counts per cell line across tissue types. Parentheses indicate the number of cell lines represented in each tissue type. Data are represented as mean ± STD. **D.** Total proteins including common proteome (dotted line) and tissue-specific proteins identified across the nine histological origins represented. See also Supplementary Table S3 and S4.

Enrichment analysis highlights the subcellular localization and function of cancer cell extracellular vesicle proteins

To characterize the cargo abundant in the majority of cancer EVs, proteins found in the $[\text{NCI-60}]_{\text{stringent}}$ dataset were further analyzed. Functional and pathway analyses were conducted using the Database for Annotation, Visualization, and Integrated Discovery (DAVID) v6.7. Not surprisingly, proteins enriched in protein localization, transport, and vesicular functions were identified in our data set (Figure 2A). Many ribosomal proteins involved in translation processes were also enriched, similar to results seen in previous studies [32, 46]. Ribosomal components may facilitate cell-to-cell communication by directly translating mRNAs present in EVs following fusion with target cells. Pathway analysis revealed proteins to be enriched in RNA processing and proteolytic processes, as well as cytoskeletal and endocytic, pathways (Figure 2B). Comparison to the Vesiclepedia database revealed the majority of proteins to have a subcellular localization in endolysosomal or cytoplasmic compartments (Figure 2C).

Altogether, these analyses demonstrate an abundance of proteins with recognized functions in protein and vesicle trafficking from the endosomal pathway, and cytoskeletal involvement that likely plays a role in both exosome and microvesicle secretion. The presence of proteins involved in RNA processing suggests an active process of RNA sorting and packaging, and offers insight into how biologically active RNA messages are communicated between cells.

A common method of characterizing EVs relies on the presence of accepted protein markers enriched in vesicle populations. Recent research has described different subpopulations of EVs from cells by identifying vesicle specific protein markers [47]. We compared commonly used exosome and microvesicle markers described across EVs derived from the NCI-60 cells (Figure 2D). Historically, these markers have been considered universal, however, only tetraspanin CD81, Alix, and HSC70 were found across all samples. Tetraspanins CD63 and CD9, as well as TSG101, Syntenin-1, and Flotillin-1 were identified in at least two-thirds of the samples. MMP-2, a previously
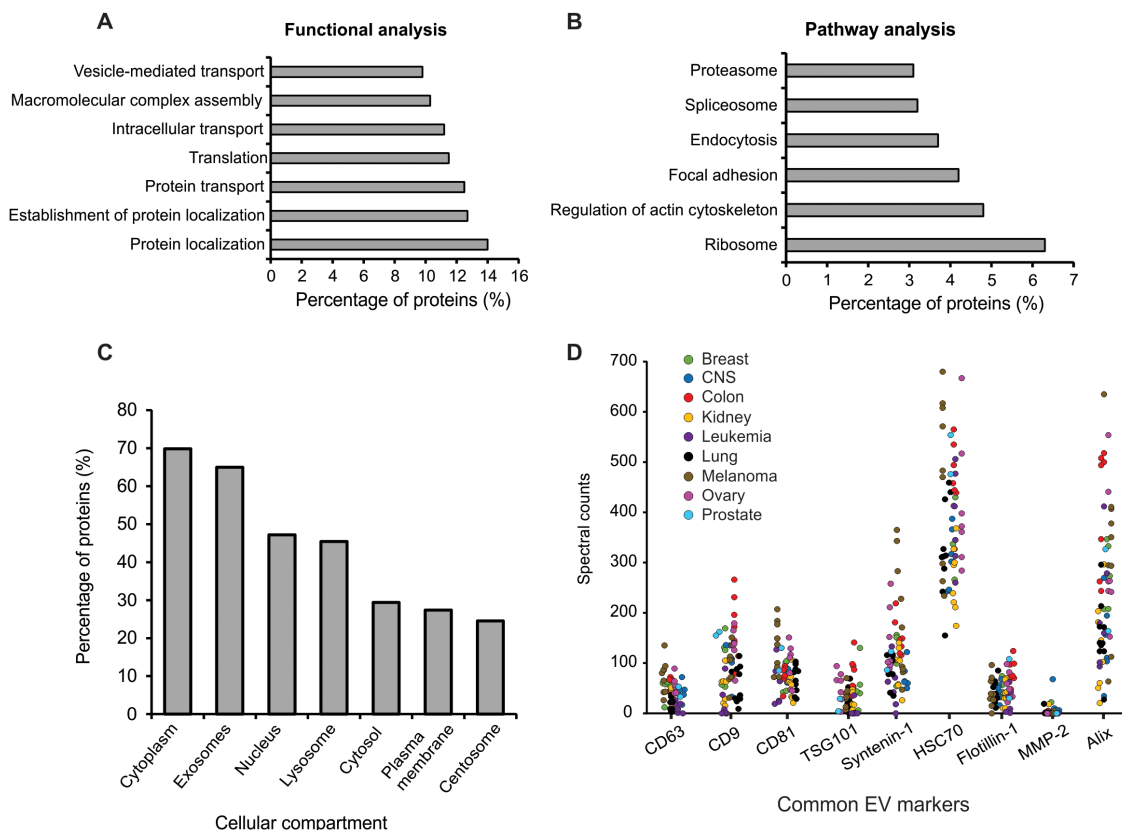

Figure 2: Enrichment analysis of EV proteins identified. Proteins identified in the $[\text{NCI-60}]_{\text{stringent}}$ dataset were used for enrichment analyses. **A.** Functional and **B.** pathway enrichment analysis of EV proteins using the DAVID database (GOTERM_BP_FAT and KEGG_PATHWAY). All terms were significant ($p < 0.05$) following Benjamini correction. **C.** Subcellular localization enrichment using FunRich. All terms were significant ($p < 0.001$) following Benjamini correction. **D.** Spectral counts of common vesicle protein markers measured across all NCI-60 EV samples.

reported microvesicle marker [32] was found in only 16 EV samples across the panel. Many of these protein markers have historically been used to describe and quantify extracellular vesicles regardless of their cellular localization or method of harvest. However, data presented here show wide variation in the levels of traditional EV markers, with some being completely undetectable in certain EV preparations. We subsequently analyzed EV proteins identified in this study to characterize those common to all cells in the NCI-60 panel. Proteins within this dataset involved in vesicle-mediated transport and protein localization were identified, and were largely enriched in GTPase function, including Rab proteins 1A, 2A, 5C, 6A, 7A, 8A, 10, 11B, and 14 (Supplementary Table S5). These proteins likely represent more universal markers of EVs that future EV researchers should consider for characterization. Notably, as various cells may package these proteins to different degrees within the same number of EVs, the correlation of vesicle quantity across cell lines to any one of these markers is complex. This poses challenges for researchers, as quantitative protein assays such as ELISA or immunoblot analyses are often used to determine vesicle quantity following EV enrichment.

EV proteomes cluster by tissue type and contain proteins unique to cancer type

Next we aimed to compare vesicle proteomes across individual cancer cells in the NCI-60 panel. As extracellular vesicles have been suggested to carry proteins that reflect their progenitor cell, it was hypothesized that EVs released from cells of the same tissue type would share similarities in protein content. Principal component analysis (PCA) demonstrated that EV proteins clustered based on tissue type (Figure 3A). Of note, one leukemia cell line (K562) did not cluster with the remaining cancer EVs (Supplementary Figure 1C). The K562 cell line is an erythroleukemia derived from the pleural effusion of a patient with chronic myelogenous leukemia (CML), and is positive for the Philadelphia translocation on chromosome 22, creating the chimeric BCR/ABL fusion gene. The BCR/ABL gene has been demonstrated to downregulate many cell adhesion molecules [48], which may, in part, explain the divergence from other cell-derived EVs. For instance, in this study, ICAM3, an adhesion molecule abundantly expressed in leukocytes, was found in high levels among all leukemia-derived EVs except the K562 EVs. In light of these findings, the K562 cell line was excluded from all subsequent analyses.

Unsupervised hierarchal clustering was used to further examine the congruity of EV proteomes from cells of the same tissue origin. Samples from colon, kidney, leukemia, lung, and melanoma cancers clustered closely within tissue type (Figure 3B). Breast, CNS, and ovarian cancer EVs demonstrated sub-clustering within each

tissue type, suggesting similarities between cancer types that may not be universal across the tissue of origin. Part of these unique clustering patterns may reflect metastatic potential of particular cell lines. For example, highly metastatic breast cancer cells (BT-549 and MDA-MB-231) cluster together away from other non-metastatic breast cell lines.

Strikingly, a number of proteins were observed to be differentially expressed in vesicles secreted from adherent cell lines compared to cells in suspension culture. For example, Agrin (*AGRN*), a basement membrane glycoprotein that contains heparan and chondroitin sulfate residues was absent only in the leukemia-derived EVs (Figure 3C). On the other hand, adhesion molecule ICAM3 was present predominately in EVs secreted from leukemia cancer lines grown in suspension (Figure 3D). As integrin and heparan sulfate proteoglycan receptors have been demonstrated to play a significant role in the uptake of vesicles into cells [16, 49], these observations provide new targets for future studies of tissue-specific mechanisms of vesicular protein trafficking and targeting of EVs to recipient cells.

Recently, a urine-based exosome diagnostic assay has been demonstrated to predict high-grade prostate cancer among men with elevated PSA levels [50]. Likewise, glypican-1, a heparan sulfate proteoglycan, was found to be detected in only cancer-derived exosomes, and was further shown to be correlated with pancreatic cancer progression in patients, providing a non-invasive early diagnostic tool for pancreatic cancer [23]. In the NCI-60 panel, glypican-1 was identified in only 35 of 60 cancer EV isolates. This suggests an even stronger likelihood of novel protein markers identified across *all* samples in this study (or all samples within a tissue-type) to be useful as early diagnostic markers of cancers. Furthermore, the evidence showing differences in EV contents from circulating cells (leukocytes) compared to basement membrane-adhered cells is particularly valuable when considering liquid biopsy techniques to isolate vesicles, as proteins such as ICAM3 could serve as a tool to distinguish blood cell-borne EVs from those secreted into circulation by organ-derived or metastatic cells.

Extracellular vesicle-based liquid biopsies also carry the potential for early detection of cancer cells that normally have limited access to blood circulation. The proteomic analysis of NCI-60 EVs confirmed the presence of premelanosome protein (*PMEL*) in vesicles secreted from all melanoma cancer cell types (Figure 3E). This melanocyte-specific transmembrane glycoprotein has previously been shown to be sorted into endosomes for exosomal secretion [51]. As melanocytes are ordinarily confined to the epidermal layers of the skin, access to deeper blood vessels is not usually achieved unless vertical growth of cancerous cells occurs. Therefore, a circulating (plasma) melanocyte-specific exosomal protein

marker could serve as an early indication of various types of invasive melanoma growth.

Moreover, several vesicular proteins were identified in a very small population of cancer lines. For instance, Tenascin XB (*TNXB*), an extracellular matrix glycoprotein was found in abundant levels only in PC-3 cells, a high grade prostate adenocarcinoma line. Although Tenascin XB was not identified in whole cell proteomic profiling from PC-3 cells [28], relatively high transcript levels of this protein in PC-3 cells have been described previously [52]. Another EV protein, periostin was recently identified as a metastatic breast cancer vesicular marker [22]. In our
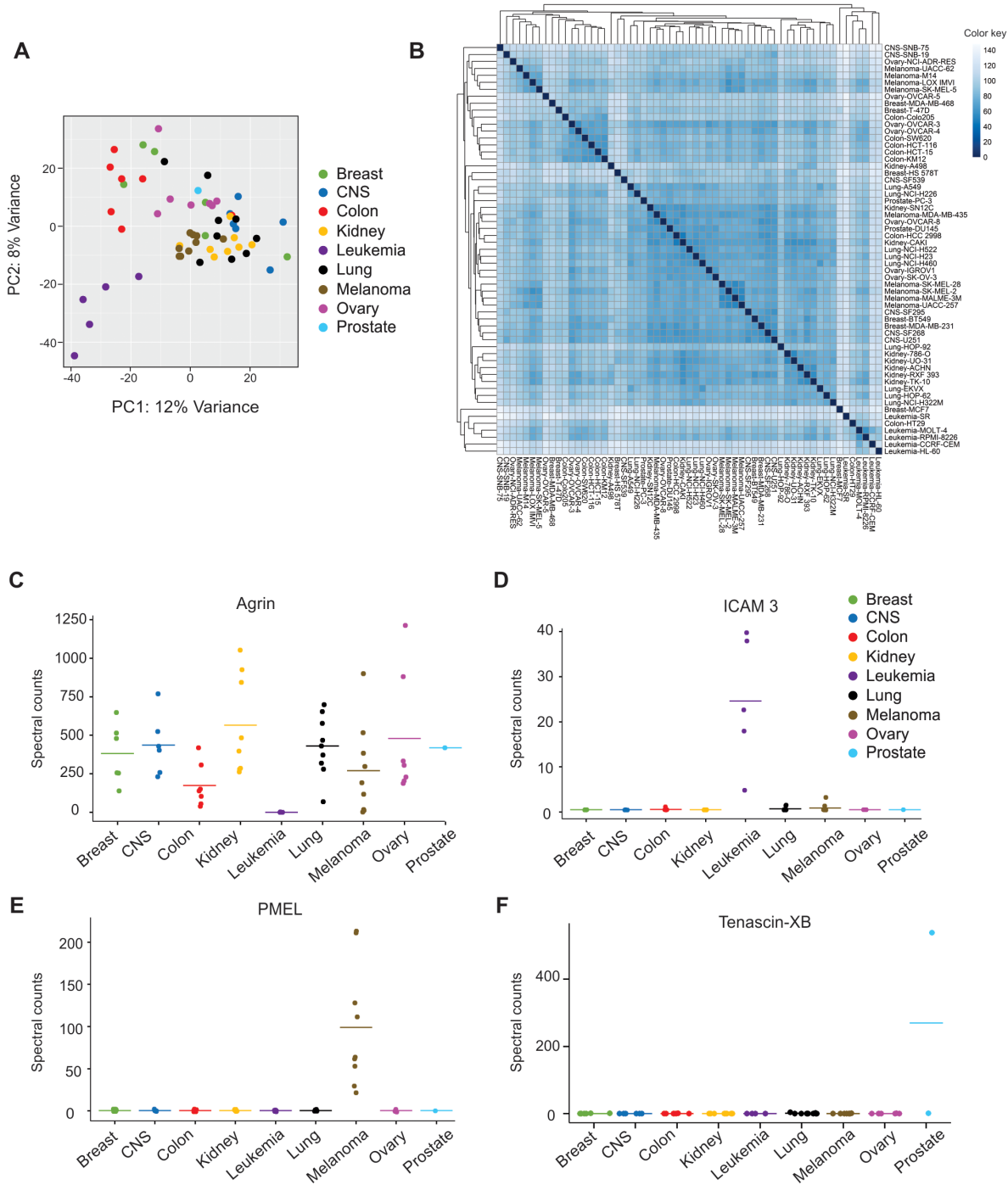

Figure 3: Differential expression of proteins found in NCI-60 cell-derived extracellular vesicles. A. PCA plot based on variant vesicle proteins across tissue types. **B.** Unsupervised hierarchal clustering of cell lines based on EV proteomic profiles. Spectral count-based differential expression of **C.** Agrin, **D.** Intercellular adhesion molecule (ICAM) 3, **E.** Premelanosome protein (*PMEL*), and **F.** Tenascin-XB. See also Supplementary Figure 1 and Supplementary Table S6.

study, the presence of periostin was confirmed in both metastatic breast cancer lines (MDA-MB-231 and HS 578T), but was not found in other non-metastatic breast cancer-derived EVs. Additional proteins including raftilin (a lipid-raft regulating protein), fibulin-7 (an adhesion molecule), and plasminogen activator inhibitor 1 (a serine protease inhibitor that has previously been implicated in aggressive tumor growth [53, 54]) were exclusively found in metastatic breast cancer EVs. Likewise, latent-TGFβ-binding protein-1 was identified preferentially in lung cancer cells with high invasive capacity (A549, HOP-62, and HOP-92) [55]. In all, over 1,500 proteins were found to be differentially expressed across the 60 EV samples (Supplementary Table S6). Interestingly, comparison of whole cell protein expression reported by Gholami et al. to EV expression in Figure 3C-3E [28] revealed these differentially expressed proteins to be largely conserved between cell and vesicle isolates. For instance, ICAM3 was chiefly absent in non-leukemic tumor cells, while Agrin was underrepresented in leukemia-derived cells (Supplementary Figure S2). Likewise, PMEL was found in substantial levels in melanoma cell lines compared to other cancer cells. These findings suggest that differentially expressed proteins found in cancer EVs may reflect cellular phenotypes. Furthermore, the specificity of protein content in vesicles from individual cancer cell types promises great potential in further investigation of these novel markers for early cancer detection and prognostic monitoring.

Conserved vesicular proteins are correlated with EV secretion

Recently, we described relative extracellular vesicle secretion quantities across the NCI-60 panel [41]. As many EV proteins in samples across the study were identified in different quantities, we hypothesized that some of these proteins are likely involved in vesicle biogenesis and therefore correlate to the total number of vesicles secreted by cells. To investigate proteins involved in common pathways of EV formation, levels of proteins in the [NCI-60]$_{stringent}$ were compared to previously collected vesicle secretion quantities [41] (particles per cell; Supplementary Table S7) by weighted gene coexpression network analysis (WGCNA).

In this analysis, hierarchal clustering of proteins (Figure 4A) demonstrated inter-related expression patterns and produced 15 clusters of highly related proteins (modules) that were detected by dynamic tree cut, an optimal method used to detect clusters of data within a dendrogram [56]. Modules were then correlated to vesicle secretion patterns across the panel (Supplementary Table S8). The yellow module containing 88 proteins was most significantly correlated with particle secretion, and therefore served as our target protein cluster (Figure 4B).

Here, protein significance is defined as the correlation of the protein expression profile across the NCI-60 panel with particle secretion levels. Module membership further measures the correlation of protein expression patterns across the members of the yellow module. We found protein significance and module membership to be positively correlated in the yellow module ($p = 0.006$) (Figure 4C). These findings suggest that proteins clustered into the module show interconnected profiles of expression in vesicles that positively correlated with the number of vesicles secreted by cells.

Enrichment analysis of the yellow module demonstrated proteins were significantly enriched in cell adhesion and growth, GTPase activity, and cell surface receptor signaling (Figure 4D), and included CD63, CD81, VAMP3, syntenin-1, and SEC22B, among other vesicular proteins. Strikingly, 25 of the yellow module proteins were identified in EVs from every cancer cell in the panel (Table 1), supporting the hypothesis that commonly identified EV components likely play a role in EV biogenesis. In light of the variation in current EV markers seen (Figure 2D), these represent important proteins that could more accurately compare vesicle quantities across a diversity of cell lines and certainly warrant future investigation.

Cancer vesicle proteomes reflect the molecular composition of progenitor cells

Given the clinical utility of using extracellular vesicles for cancer diagnostics, we investigated the relationships between EV protein composition and whole cell content. Previously, cellular protein and transcript expression profiles were compared using co-inertia analysis (CIA) to examine the concordance between these molecular datasets across the NCI-60 panel [28]. Here vesicle protein levels were similarly compared to cellular protein and RNA expression. In Figure 5A, each of the three datasets (vesicle proteome, cellular proteome, and cellular transcriptome) is plotted for individual cell lines. Markers represent the relative position of a cell line in respective proteome or transcriptome space, where the divergence of datasets is delineated by connecting vectors. Molecular factors driving tissue-dependent clustering were plotted as overlapping protein (cell and vesicle) and RNA transcript data in the same orientation (Figure 5B). In variable space, RNA or protein coordinates farther from the origin are more highly expressed in cell lines projected in the same direction. Whole cell RNA profiling showed the strongest ability to differentiate samples, demonstrated by farther distances of RNAs from the origin in variable space. However, vesicle protein was observed to show similar magnitudes of projection from the origin as whole cell protein, demonstrating the ability of EV proteins to

reflect cell protein profiling. A histogram of eigenvalues (Figure 5C) demonstrated that the first and second co-inertia axes represented 49% of the total variance (sum of the eigenvalues) seen in the datasets plotted, accounting for 31% and 18% of the variance respectively. The three datasets were also examined to consider how much variance of the eigenvalues was contributed by each dataset (Figure 5D). No single dataset contributed to both co-inertia axes alone, indicating that the analysis examining the relationships between vesicle proteomes and cellular transcriptomic and proteomic profiles was dependent on all datasets.

In general, melanoma, leukemia and colon cells demonstrated tissue-type clustering based on their unique proteomic and transcriptomic profiles, similar to clusters previously observed (Figure 3A-3B). The vesicle protein dataset showed considerable association with the whole cell molecular profiles, indicated by the short and randomly oriented arrows (Figure 5A). Overall, the similarity of vesicle protein to cellular RNA and protein
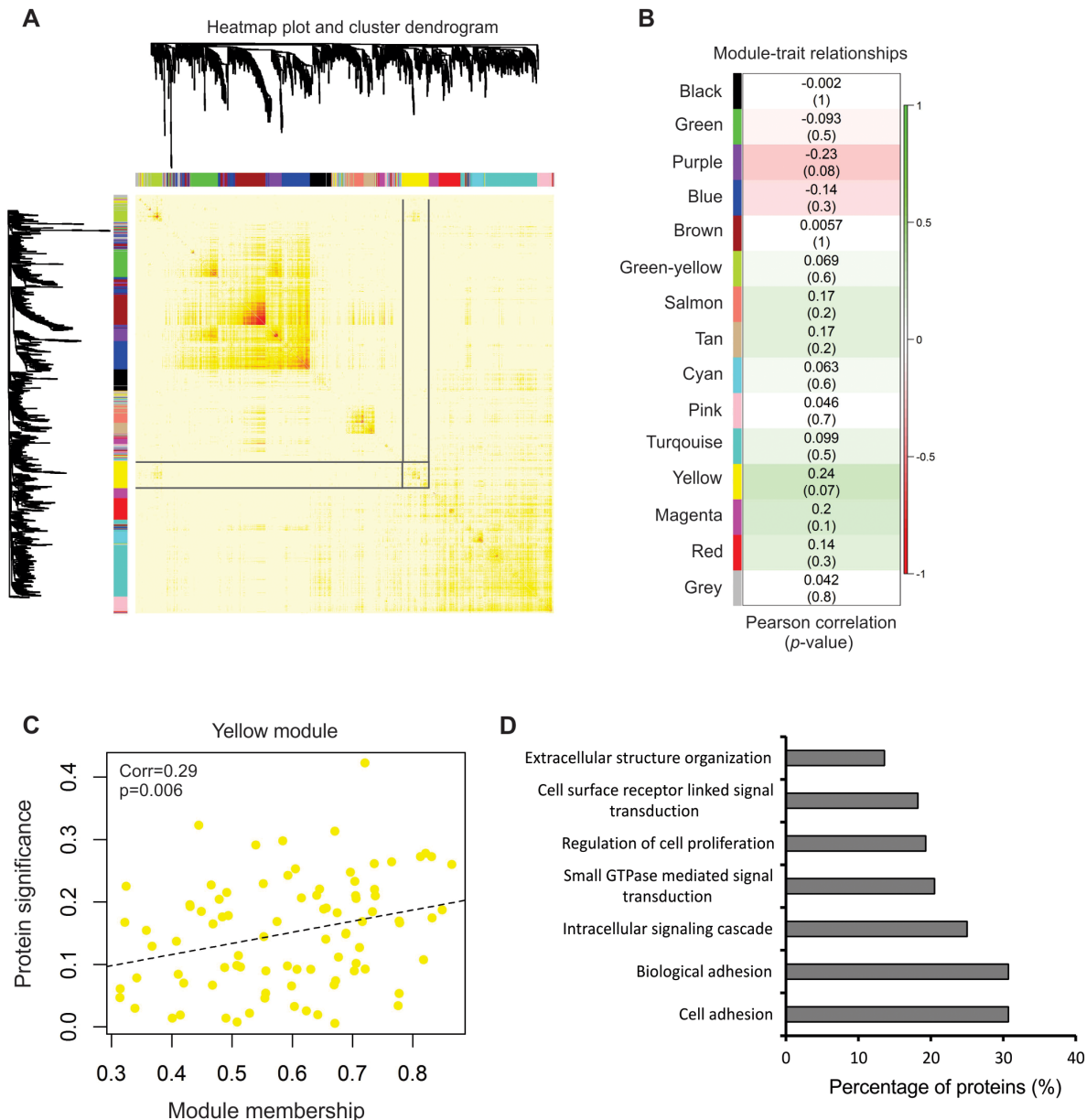

Figure 4: Network analysis of protein content with vesicle secretion. A. Network heatmap plot of topological overlap depicting protein dendrogram and module assignment. Targeted yellow module is highlighted. See also Supplementary Figure 3. **B.** Heatmap of module-trait correlation containing correlation and relevant *p*-values of modules detected. See also Supplementary Table S7 and S8. **C.** Scatter plot of membership in the yellow module and protein significance corresponding to vesicle secretion quantities across the NCI-60. A significant positive correlation between protein significance and module membership was determined (*p* = 0.006). **D.** GOTERM_BP_FAT analysis of biological processes enriched in the yellow module protein dataset.

Table 1: Common cancer proteins associated with vesicle secretion

Protein ID	Gene symbol	Protein name	Average spectral count per sample
P07355	*ANXA2*	Annexin A2	288.53
P18085	*ARF4*	ADP-ribosylation factor 4	18.05
P35613	*BSG*	Basigin	79.12
P13987	*CD59*	CD59 glycoprotein	34.40
P60033	*CD81*	CD81 antigen	83.18
P05534	*HLA-A*	HLA class I histocompatibility antigen, A-24 alpha chain	92.82
P01112	*HRAS*	GTPase HRas	40.58
P0DMV9	*HSPA1B*	Heat shock 70 kDa protein 1B	142.08
P04792	*HSPB1*	Heat shock protein beta-1	74.40
P05556	*ITGB1*	Integrin beta-1	168.10
Q08380	*LGALS3BP*	Galectin-3-binding protein	550.05
P08590	*MYL3*	Myosin light chain 3	10.32
P60660	*MYL6*	Myosin light polypeptide 6	72.50
P22392	*NME2*	Nucleoside diphosphate kinase B	75.72
P01111	*NRAS*	GTPase NRas	7.08
P51148	*RAB5C*	Ras-related protein Rab-5C	51.18
P51149	*RAB7A*	Ras-related protein Rab-7a	64.45
P63000	*RAC1*	Ras-related C3 botulinum toxin substrate 1	43.45
P11234	*RALB*	Ras-related protein Ral-B	38.67
P61224	*RAP1B*	Ras-related protein Rap-1b	111.35
P61586	*RHOA*	Transforming protein RhoA	81.78
O75396	*SEC22B*	Vesicle-trafficking protein SEC22b	21.85
P37802	*TAGLN2*	Transgelin-2	87.45
P61077	*UBE2D3*	Ubiquitin-conjugating enzyme E2 D3	7.58
Q15836	*VAMP3*	Vesicle-associated membrane protein 3	28.08

resulted in RV coefficients of 0.63 and 0.57, respectively. Notably, several cell lines demonstrated greater variation between vesicle protein and whole cell profiles, including two lung lines (NCI-H322M and HOP-62), one colon cell (HCC-2998), and an ovarian cell (IGROV1). To more closely examine proteins found in discordance between whole cell and EV proteomes, proteins found in the core vesicle proteome were compared to reported cellular expression data [28]. In total, 144 proteins identified in the core vesicle proteome in this study were similarly identified in cell lines, as reported by Gholami et al. Comparison of expression levels between whole cell and vesicular isolates revealed several outlying proteins enriched in vesicles or cells alone (Figure 5E–5H and Supplementary Table S9). While proteins such as Actin (ACTG1) or Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) were found abundantly in both cells and EVs, tubulin proteins (TUBB4B and TUBA1C) were enriched in lung cancer vesicles compared to cells, and actin-binding protein Moesin (MSN) was enriched in vesicles from IGROV1 and HCC-2998 cells. Strikingly, Galectin-3-binding protein (LGALS3BP) was enriched in multiple EV isolates compared to respective progenitor cells. These proteins probably represent targets of selective vesicular packaging, and future focus will likely advance our understanding of vesicular trafficking mechanisms.
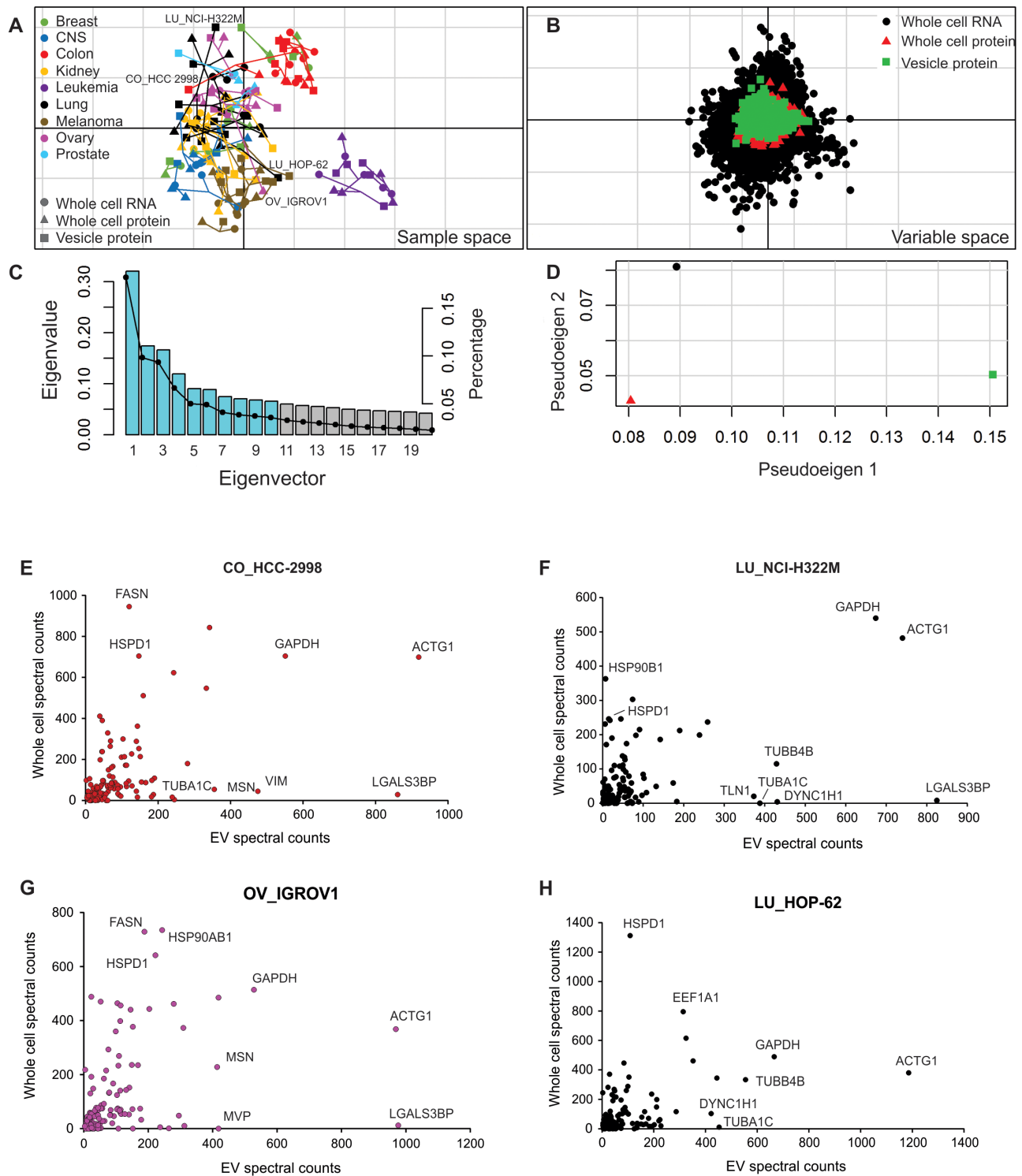
Figure 5: Comparison of vesicular proteome with cellular proteome and transcriptome. A. Co-inertia analysis of cellular proteome and transcriptome with vesicular proteome across the NCI-60 panel. Vectors represent the proportional divergence between data sets within each cell line. **B.** Coordinates of the proteins or RNA from each data set plotted in the same orientation as the co-inertia analysis. **C.** Histogram of eigenvalues. Blue bars represent the absolute values of eigenvalues retained in the analysis. Black dots represent the proportion of variance found within each eigenvector. **D.** Plot of pseudo-eigenvalues space, indicating the variance of eigenvalues contributed by each of the three datasets. Comparison of expression levels of core vesicular proteins with whole cell protein expression from representative cells lines: **E.** HCC-2993 (colon); **F.** NCI-H322M (lung); **G.** IGROV1 (ovary); and **H.** HOP-62 (lung). See also Supplementary Table S9.

Altogether, the congruency of tissue dependent clusters across datasets and resemblance between the vesicle and cellular molecular profiles substantiates the ability of EV profiles to discriminate between tissue types and supports a significant clinical value of EVs as circulating biomarkers. As EVs can closely reflect the molecular profile of their cellular origin, they therefore represent circulating biological archetypes of healthy or diseased cells.

CONCLUSIONS

In the past decade, it has become clear that extracellular vesicles are an important component of cellular communication informing fields including cancer biology, immunology, and virology. The data here provide a means to better understand the protein cargo involved in this ubiquitous messaging system with implications for further use in clinical diagnostic practice. In the characterization of vesicular protein contents from the NCI-60 cells, we have expanded the list of known vesicle proteins by nearly twenty percent (based on the Vesiclepedia database of characterized vesicle proteins), significantly growing our foundation of knowledge of the nature of this communication system. Of these proteins identified, 213 were common to all cell lines. This core EV proteome likely represents conserved structural and signaling components of vesicles, and molecular factors involved in vesicle biogenesis and secretion. In support of this, 25 of the 213 proteins were found to be positively correlated with EV secretion levels. One of the largest obstacles currently faced by EV researchers is accurately defining these vesicles subtypes. The data presented in this study offer a substantial step forward in the ability to define EV populations, and to enrich vesicles for studies in all areas of EV research. As our ability to isolate distinct vesicle sub-populations continues to improve, this substantial database will likely provide an important backbone for understanding more unique vesicle contents.

In this study, critical differences in extracellular vesicle contents were also identified which may directly lead to the discovery of biomarkers of cancer, ultimately affecting diagnosis and prognosis of the increasingly prevalent disease. As emerging large-scale "omic" datasets progressively demonstrate significant potential in the future of individualized medicine, it is expected that extracellular vesicles will play an important role in modern health and disease-state surveillance. In this study, the close resemblance of vesicle and cellular molecular profiles suggests that extracellular vesicles reliably represent their progenitor cells. They are therefore excellent candidates for bearing biomarkers, and will increasingly find a place amongst clinical strategies for combating diseases like cancer.

EXPERIMENTAL METHODS

Cell culture

Sixty cell lines from the National Cancer Institute (NCI-60) were acquired from the NCI Developmental Therapeutics Program. Cells were seeded based on growth rate (doubling time) to achieve 80-90% confluence after two to three days. Cells were cultured in RPMI-1640 medium (Lonza, 12-702Q) supplemented with 10% fetal bovine serum (FBS; Seradigm, 1400-500), 2 mM L-glutamine (Corning, 25-005-CI), 100 units penicillin/ streptomycin (Corning, 30-002-CI), and 100 units:100 µg/mL:0.25 µg/mL antibiotic/antimycotic (Corning, 30-002-CI). Following the two to three day culture, complete medium was aspirated and cells were washed with warmed sterile phosphate buffered saline (PBS). To minimize contaminating serum proteins in downstream mass spectrometry analysis and aid in the identification of lower abundance EV proteins, cells were cultured in serum-free medium for a further 48 hours before EV enrichment. Previous results have demonstrated that EVs harvested using this method are as pure as sucrose-cushion purified vesicles [43].

EV enrichment and protein quantification

Following the 48 hour serum-free culture, medium was harvested. Cell viability at the time of harvest was measured by counting cells stained with 0.2% trypan blue in PBS (Sigma, T8154) or AO/PI (Nexcelom Bioscience, CS2-0106) with an automated cell counter (Cellometer Vision, v2.1.4.2, Nexcelom Bioscience). We have previously demonstrated that cell viability does not significantly contribute to the number or size of particles secreted by the NCI-60 cells when viabilities are maintained greater than 85% [41]. In this study, cell viabilities following the forty-eight hour serum-free culture approximated those observed previously in complete medium. EVs were enriched using the ExtraPEG method previously described [43]. Briefly, cell-conditioned medium was centrifuged at 500 g for five minutes to pellet and discard cells, followed by 2,000 g for 30 minutes to remove cellular debris. Supernatant was pooled from several flasks to amass sufficient material (200-500 mL) for downstream proteomic analysis. A 1:1 volume of 2X PEG solution [16% (w/v) polyethylene glycol, 1 M NaCl] was added. Samples were inverted to mix, then incubated overnight. The next day, the medium/PEG mixture was centrifuged at 3,214 g for one hour. Crude vesicle pellets were resuspended in 1 mL of particle-free PBS and re-pelleted by ultracentrifugation at 100,000 g for 70 minutes. Final pellets were lysed in strong lysis buffer [5% SDS, 10 mM EDTA, 120 mM Tris-HCl pH 6.8, 2.5% β-mercaptoethanol, 8 M urea] with the addition of HALT protease inhibitor (Thermo, 78438).

Vesicular protein was quantified using the fluorescence-based EZQ™ Kit (Thermo, R33200).

SDS PAGE and in-gel digestion

For protein purification and separation by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE), 20 μg of vesicular protein from each sample was loaded into a 4-20% polyacrylamide gel (Lonza, 59511). Following electrophoresis, gels were fixed and Coomassie-stained as previously detailed [57]. Samples were fractionated by cutting gel lanes into five sections. Sections were subsequently subdivided into 1 mm^3 cubes before trypsin-digesting as described [43].

Protein identification and quantification

Following protein digestion, samples were submitted to the Florida State University Translational Science Laboratory for liquid chromatography tandem mass spectrometry (LC-MS/MS) analysis. The digest was freeze-dried and resuspended in 30 μL 0.1% FA and analyzed by liquid chromatography tandem mass spectrometry (LC-MS). An externally calibrated Thermo LTQ Orbitrap Velos nLC-ESI-LTQ-Orbitrap (high-resolution electrospray tandem mass spectrometer) was used with the following parameters: A 2 cm trap column of 100 μm internal diameter (i.d.) (SC001 Easy Column from Thermo-scientific) was followed by a 10 cm analytical column of 75 μm i.d. (SC200 Easy Column from Thermo-scientific). Both trap column and analytical column had C18-AQ packaging. Separation was carried out using Easy nanoLC II (Thermo-Scientific) with a continuous, vented column configuration. A 5 μL sample was aspirated into a 20 μL sample loop and loaded onto the trap. The flow rate was set to 300 nL/min for separation on the analytical column. Mobile phase A was composed of 99.9 H$_2$O (EMD Omni Solvent), and 0.1% formic acid, and mobile phase B was composed of 99.9% ACN and 0.1% formic acid. A 1 hour linear gradient from 0% to 45% B was performed. The LC eluent was directly nano-sprayed into an LTQ Orbitrap Velos mass spectrometer (Thermo Scientific). During the chromatographic separation, the LTQ Orbitrap Velos was operated in a data-dependent mode and under direct control of the Xcalibur software (Thermo Scientific). The mass spectrometry data were acquired using the following parameters: 10 data-dependent collisional-induced-dissociation (CID) MS/MS scans per full scan. All measurements were performed at room temperature and three technical replicates were run for each sample.

Raw data collected from each of five fractions were pooled by sample and analyzed using MaxQuant (v1.5.3.30). The mass spectrometry data were analyzed using the integrated Andromeda peptide search engine and a recent (March 2016) UniProt knowledgebase reviewed (Swiss-prot) human protein database. The database was appended with a list of common contaminants in MaxQuant, and search parameters used were either the default settings for this version of the software, or as follows. Instrument type was set to Orbitrap, using label free quantitation, and first search peptide tolerance set to 10 ppm, and main search peptide tolerance at 4.5 ppm. Digestion mode was set specific for trypsin, with a maximum of two missed cleavages. Fixed modifications included only carbamidomethyl (C), and variable modifications included oxidation (M), N-terminal acetylation, and phosphorylation (STY). A maximum of five modifications were allowed per peptide. False discovery rate was set to 0.01. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium [58] via the PRIDE partner repository.

Primary analyses were conducted with the MaxQuant output data set (NCI-60). This dataset represents all proteins identified in the study. To examine proteins found in the majority of samples that reflect a more commonly shared cancer EV proteome, a more stringent protein set was used, and defined as [NCI-60]$_{stringent}$. This data set includes only proteins found in at least two thirds of all EV samples.

Differential expression as a function of tissue origin

For differential expression and network analyses, raw spectral count data were exported from the MaxQuant software and imported into R statistical framework. The spectral count based method has been demonstrated to provide a reliable representation of protein abundance with a linear dynamic range over several orders of magnitude with similar sensitivity as ion peak intensity quantitation [59–61]. Utilization of spectral counts was necessary for our downstream pipeline analyses using DeSeq2 for the identification of differentially expressed proteins and normalization for WGCNA and co-inertia analyses. Raw peak intensity data are reported in Supplementary Table 1.

DESeq2 is a package originally designed for analyzing read counts from RNA-sequencing datasets and recently used to analyze other forms of biological count data, including LC-MS/MS spectral count data [62]. Normalization of spectral counts with DESeq2 considerably reduced the difference in sample depth between samples (Supplementary Figure 1A and 1B). PCA analysis of the top 500 variant proteins revealed one sample with substantially different protein expression profiles (K562) (Supplementary Figure 1C), which was removed from further analysis. After normalization, a likelihood ratio test was performed across all samples in DESeq2 to identify proteins with significantly different patterns of expression between samples of different tissue

origin. Proteins that returned an FDR-adjusted *p*-value (q value) of less than 0.05 were considered significant.

Weighted gene coexpression network analysis (WGCNA)

The WCGNA was originally implemented for the analysis of transcriptomic microarray data, but has recently been used to analyze RNA-seq and LC-MS/MS proteomics datasets [63, 64]. To identify proteins that are associated with the regulation of vesicle secretion, a network analysis was performed with the WGCNA package in the R statistical framework [65]. Tutorials for WGCNA can be found at http://labs.genetics.ucla.edu/horvath/CoexpressionNetwork/. WGCNA is a multi-step analysis which involves: 1) generating a coexpression matrix for gene (or protein) expression, 2) transforming protein correlations into a adjacency matrix for network construction, 3) grouping together proteins (modules) that show high co-correlation, and 4) correlating an eigenprotein, the first principle component of each module, with a biological trait of interest. Importantly, WGCNA avoids the extensive FDR-correction needed to control for false positives in protein-wise differential expression analyses, as there are many fewer modules in the entire network than proteins.

Normalized [NCI-60]$_{stringent}$ count data were transformed in DESeq2 with a variance stabilizing transformation using the "VST" function, and used as input for WGCNA. The resulting count matrix was used as input for WGCNA. We used the one-step network construction and module detection function "blockwiseModules." Modules of co-regulated proteins across the vesicle protein abundance profiles of the NCI-60 cell lines were identified by first grouping proteins (nodes) together into clusters of highly co-correlated members based on their degree of co-correlation (edges). For measuring the coexpression similarity between proteins, we used a signed Pearson correlation. For adjacency matrix construction, soft thresholding was used. The scale-free fitting index approached a local maximum at $\beta = 12$ (Supplementary Figure S3). This threshold approximately achieves scale free topology, meaning that the node degree distribution approximately follows the power law. Therefore, the co-correlation matrix was raised to the power of 12 to create the adjacency matrix. Finally, the adjacency matrix was converted into a topological overlap matrix, and the corresponding dissimilarities (1 – the topological overlap matrix) was used for module detection via hierarchical clustering. The minimum number of proteins required per module was set to 30, and the mergeCutHeight was set to 0.25, such that sufficiently similar modules were merged revealing 15 modules. Module eigenproteins, a synthetic measure of module belonging, were then related to the quantity of vesicles secreted per cell previously collected [28] by calculating Pearson correlation coefficients.

The module most significantly associated with particle secretion was retained for enrichment analysis.

Comparison between vesicle and whole-cell molecular profiles

Co-inertia analysis (CIA; [66, 67]) was used to examine similarities between the vesicle proteome and cellular proteome and transcriptome across the NCI-60 cell lines. Microarray gene expression data was downloaded from the NCBI Gene Expression Omnibus [68] (accession number GSE32474; [69, 70]). Whole-cell proteome data was obtained from the NCI60 proteome resource (http://proteomics.wzw.tum.de/nci60; [28]). Spectral count data from the whole-cell dataset was processed analogously to the EV proteome dataset for comparability: raw spectral count data published by Moghaddas Gholami et al. were normalized with DESeq2 following removal of all proteins not found with at least one count in 40 or more samples, and further transformed with the VST function. NCI-60 cell lines not included in all three datasets were excluded from this analysis, leaving 56 samples in total. Differences between samples and molecular assays were analyzed and visualized with co-inertia analysis using the R omicade4 package [71] in the R statistical framework.

Protein set enrichment analysis

To identify proteins previously found in EVs, the Vesiclepedia database of vesicular proteins was downloaded (Version 3, 9 Jan 2015) from microvesicles.org/download. The Database for Annotation, Visualization, and Integrated Discovery (DAVID) v6.7 [72, 73] was used for functional (GOTERM_BP_FAT) and pathway (KEGG_PATHWAY) analyses of proteins found in the [NCI-60]$_{stringent}$ dataset (Figure 2A-2B). Enrichment of cellular compartment terms (Figure 2C) was analyzed using FunRich v3 [74].

For further interpretation of the target WGCNA module, enrichment of biological processes was performed using DAVID: GOTERM_BP_FAT. Proteins entered into the WGCNA analysis were used as the background dataset, and proteins within the target module were used as the target list. All terms with a *p*-value (Benjamini or Benjamini-Hochberg adjusted) less than 0.05 were considered significant and ranked by the number of proteins identified in the group.

ACKNOWLEDGMENTS

of Health (RO1CA204621 and R15CA188941) awarded to D.G.M.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

Author contributions

D.G.M, M.A.R., and S.N.H. designed the study, performed experiments, and interpreted results. X.L. prepared samples for mass spectrometry analyses performed by R.K.S. Computational, bioinformatics, and statistical analyses were performed by S.N.H. and J.L.B. The manuscript was written by S.N.H. and D.G.M. with contributions from all authors. D.G.M. conceived and supervised the project.

REFERENCES

1. Bobrie A, Colombo M, Raposo G, Théry C. Exosome secretion: molecular mechanisms and roles in immune responses. Traffic. 2011; 12: 1659-68. doi: 10.1111/j.1600-0854.2011.01225.x.

2. Théry C, Zitvogel L, Amigorena S. Exosomes: composition, biogenesis and function. Nat Rev Immunol. 2002; 2: 569-79. doi: 10.1038/nri855.

3. Mathivanan S, Ji H, Simpson RJ. Exosomes: extracellular organelles important in intercellular communication. J Proteomics. 2010; 73: 1907-20. doi: 10.1016/j.jprot.2010.06.006.

4. Théry C. Exosomes: secreted vesicles and intercellular communications. F1000 Biol Rep. 2011; 3: 15. doi: 10.3410/b3-15.

5. Raposo G, Stoorvogel W. Extracellular vesicles: exosomes, microvesicles, and friends. J Cell Biol. 2013; 200: 373-83. doi: 10.1083/jcb.201211138.

6. Zhang X, Yuan X, Shi H, Wu L, Qian H, Xu W. Exosomes in cancer: small particle, big player. J Hematol Oncol. 2015; 8: 83. doi: 10.1186/s13045-015-0181-x.

7. Park JE, Tan HS, Datta A, Lai RC, Zhang H, Meng W, Lim SK, Sze SK. Hypoxic tumor cell modulates its microenvironment to enhance angiogenic and metastatic potential by secretion of proteins and exosomes. Mol Cell Proteomics. 2010; 9: 1085-99. doi: 10.1074/mcp.M900381-MCP200.

8. Meckes DG, Shair KH, Marquitz AR, Kung CP, Edwards RH, Raab-Traub N. Human tumor virus utilizes exosomes for intercellular communication. Proc Natl Acad Sci U S A. 2010; 107: 20370-5. doi: 10.1073/pnas.1014194107.

9. Meckes DG, Raab-Traub N. Microvesicles and viral infection. J Virol. 2011; 85: 12844-54. doi: 10.1128/jvi.05853-11.

10. Meckes DG. Exosomal communication goes viral. J Virol. 2015; 89: 5200-3. doi: 10.1128/jvi.02470-14.

11. Pegtel DM, Cosmopoulos K, Thorley-Lawson DA, van Eijndhoven MA, Hopmans ES, Lindenberg JL, de Gruijl TD, Würdinger T, Middeldorp JM. Functional delivery of viral miRNAs via exosomes. Proc Natl Acad Sci U S A. 2010; 107: 6328-33. doi: 10.1073/pnas.0914843107.

12. Costa-Silva B, Aiello NM, Ocean AJ, Singh S, Zhang H, Thakur BK, Becker A, Hoshino A, Mark MT, Molina H, Xiang J, Zhang T, Theilen TM, et al. Pancreatic cancer exosomes initiate pre-metastatic niche formation in the liver. Nat Cell Biol. 2015; 17: 816-26. doi: 10.1038/ncb3169.

13. Kim HK, Song KS, Park YS, Kang YH, Lee YJ, Lee KR, Ryu KW, Bae JM, Kim S. Elevated levels of circulating platelet microparticles, VEGF, IL-6 and RANTES in patients with gastric cancer: possible role of a metastasis predictor. Eur J Cancer. 2003; 39: 184-91. doi: 10.1016/S0959-8049(02)00596-8.

14. Baran J, Baj-Krzyworzeka M, Weglarczyk K, Szatanek R, Zembala M, Barbasz J, Czupryna A, Szczepanik A. Circulating tumour-derived microvesicles in plasma of gastric cancer patients. Cancer Immunol Immunother. 2010; 59: 841-50. doi: 10.1007/s00262-009-0808-2.

15. Silva J, Garcia V, Rodriguez M, Compte M, Cisneros E, Veguillas P, Garcia JM, Dominguez G, Campos-Martin Y, Cuevas J, Peña C, Herrera M, Diaz R, et al. Analysis of exosome release and its prognostic value in human colorectal cancer. Genes Chromosomes Cancer. 2012; 51: 409-18. doi: 10.1002/gcc.21926.

16. Hoshino A, Costa-Silva B, Shen TL, Rodrigues G, Hashimoto A, Tesic Mark M, Molina H, Kohsaka S, Di Giannatale A, Ceder S, Singh S, Williams C, Soplop N, et al. Tumour exosome integrins determine organotropic metastasis. Nature. 2015; 527: 329-35. doi: 10.1038/nature15756.

17. Kalluri R. The biology and function of exosomes in cancer. J Clin Invest. 2016; 126: 1208-15. doi: 10.1172/JCI81135.

18. Wilbur J. Prostate cancer screening: the continuing controversy. Am Fam Physician. 2008; 78: 1377-84. PMID: 19119557.

19. Fritsche HA, Bast RC. CA 125 in ovarian cancer: advances and controversy. Clin Chem. 1998; 44: 1379-80. PMID: 9665412.

20. Yang GB. Clinical value of serum cancer antigen 19-9 as a tumor screening marker among healthy individuals. J BUON. 2015; 20: 1612-6. PMID: 26854460.

21. Meckes DG, Gunawardena HP, Dekroon RM, Heaton PR, Edwards RH, Ozgur S, Griffith JD, Damania B, Raab-Traub N. Modulation of B-cell exosome proteins by gamma herpesvirus infection. Proc Natl Acad Sci U S A. 2013; 110: E2925-33. doi: 10.1073/pnas.1303906110.

22. Vardaki I, Ceder S, Rutishauser D, Baltatzis G, Foukakis T, Panaretakis T. Periostin is identified as a putative metastatic

marker in breast cancer-derived exosomes. Oncotarget. 2016; 7:74966-78. doi: 10.18632/oncotarget.11663.

23. Melo SA, Luecke LB, Kahlert C, Fernandez AF, Gammon ST, Kaye J, LeBleu VS, Mittendorf EA, Weitz J, Rahbari N, Reissfelder C, Pilarsky C, Fraga MF, et al. Glypican-1 identifies cancer exosomes and detects early pancreatic cancer. Nature. 2015; 523: 177-82. doi: 10.1038/nature14581.

24. Boukouris S, Mathivanan S. Exosomes in bodily fluids are a highly stable resource of disease biomarkers. Proteomics Clin Appl. 2015; 9: 358-67. doi: 10.1002/prca.201400114.

25. Shoemaker RH. The NCI60 human tumour cell line anticancer drug screen. Nat Rev Cancer. 2006; 6: 813-23. doi: 10.1038/nrc1951.

26. Weinstein JN. Spotlight on molecular profiling: "Integromic" analysis of the NCI-60 cancer cell lines. Mol Cancer Ther. 2006; 5: 2601-5. doi: 10.1158/1535-7163. MCT-06-0640.

27. Weinstein JN. Integromic analysis of the NCI-60 cancer cell lines. Breast Dis. 2004; 19: 11-22. doi: 10.3233/BD-2004-19103.

28. Moghaddas Gholami A, Hahne H, Wu Z, Auer FJ, Meng C, Wilhelm M, Kuster B. Global proteome analysis of the NCI-60 cell line panel. Cell Rep. 2013; 4: 609-20. doi: 10.1016/j.celrep.2013.07.018.

29. Sinha A, Ignatchenko V, Ignatchenko A, Mejia-Guerrero S, Kislinger T. In-depth proteomic analyses of ovarian cancer cell line exosomes reveals differential enrichment of functional categories compared to the NCI 60 proteome. Biochem Biophys Res Commun. 2014; 445: 694-701. doi: 10.1016/j.bbrc.2013.12.070.

30. Staubach S, Razawi H, Hanisch FG. Proteomics of MUC1-containing lipid rafts from plasma membranes and exosomes of human breast carcinoma cells MCF-7. Proteomics. 2009; 9: 2820-35. doi: 10.1002/pmic.200800793.

31. Ji H, Greening DW, Barnes TW, Lim JW, Tauro BJ, Rai A, Xu R, Adda C, Mathivanan S, Zhao W, Xue Y, Xu T, Zhu HJ, et al. Proteome profiling of exosomes derived from human primary and metastatic colorectal cancer cells reveal differential expression of key metastatic factors and signal transduction components. Proteomics. 2013; 13: 1672-86. doi: 10.1002/pmic.201200562.

32. Keerthikumar S, Gangoda L, Liem M, Fonseka P, Atukorala I, Ozcitti C, Mechler A, Adda CG, Ang CS, Mathivanan S. Proteogenomic analysis reveals exosomes are more oncogenic than ectosomes. Oncotarget. 2015; 6: 15375-96. doi: 10.18632/oncotarget.3801.

33. Valenzuela MM, Ferguson Bennit HR, Gonda A, Diaz Osterman CJ, Hibma A, Khan S, Wall NR. Exosomes Secreted from Human Cancer Cell Lines Contain Inhibitors of Apoptosis (IAP). Cancer Microenviron. 2015; 8: 65-73. doi: 10.1007/s12307-015-0167-9.

34. Kong JN, He Q, Wang G, Dasgupta S, Dinkins MB, Zhu G, Kim A, Spassieva S, Bieberich E. Guggulsterone and bexarotene induce secretion of exosome-associated breast cancer resistance protein and reduce doxorubicin resistance in MDA-MB-231 cells. Int J Cancer. 2015; 137: 1610-20. doi: 10.1002/ijc.29542.

35. Shedden K, Xie XT, Chandaroy P, Chang YT, Rosania GR. Expulsion of small molecules in vesicles shed by cancer cells: association with gene expression and chemosensitivity profiles. Cancer Res. 2003; 63: 4331-7. PMID: 12907600.

36. Clayton A, Mitchell JP, Court J, Linnane S, Mason MD, Tabi Z. Human tumor-derived exosomes down-modulate NKG2D expression. J Immunol. 2008; 180: 7249-58. doi: 10.4049/jimmunol.180.11.7249.

37. Sung BH, Ketova T, Hoshino D, Zijlstra A, Weaver AM. Directional cell movement through tissues is controlled by exosome secretion. Nat Commun. 2015; 6: 7164. doi: 10.1038/ncomms8164.

38. Webber JP, Spary LK, Sanders AJ, Chowdhury R, Jiang WG, Steadman R, Wymant J, Jones AT, Kynaston H, Mason MD, Tabi Z, Clayton A. Differentiation of tumour-promoting stromal myofibroblasts by cancer exosomes. Oncogene. 2015; 34: 290-302. doi: 10.1038/onc.2013.560.

39. Phuyal S, Hessvik NP, Skotland T, Sandvig K, Llorente A. Regulation of exosome release by glycosphingolipids and flotillins. FEBS J. 2014; 281: 2214-27. doi: 10.1111/febs.12775.

40. Kosaka N, Iguchi H, Hagiwara K, Yoshioka Y, Takeshita F, Ochiya T. Neutral sphingomyelinase 2 (nSMase2)-dependent exosomal transfer of angiogenic microRNAs regulate cancer cell metastasis. J Biol Chem. 2013; 288: 10849-59. doi: 10.1074/jbc.M112.446831.

41. Hurwitz SN, Conlon MM, Rider MA, Brownstein NC, Meckes DG. Nanoparticle analysis sheds budding insights into genetic drivers of extracellular vesicle biogenesis. J Extracell Vesicles. 2016; 5: 31295. doi: 10.3402/jev.v5.31295.

42. Webber J, Clayton A. How pure are your vesiclesŒ J Extracell Vesicles. 2013; 2. doi: 10.3402/jev.v2i0.19861.

43. Rider MA, Hurwitz SN, Meckes DG. ExtraPEG: A Polyethylene Glycol-Based Method for Enrichment of Extracellular Vesicles. Scientific Reports. 2016; 6: 23978. doi: 10.1038/srep23978 http://www.nature.com/articles/srep23978#supplementary-information.

44. Kalra H, Simpson RJ, Ji H, Aikawa E, Altevogt P, Askenase P, Bond VC, Borràs FE, Breakefield X, Budnik V, Buzas E, Camussi G, Clayton A, et al. Vesiclepedia: a compendium for extracellular vesicles with continuous community annotation. PLoS Biol. 2012; 10: e1001450. doi: 10.1371/journal.pbio.1001450.

45. Liu Y, Cao X. Organotropic metastasis: role of tumor exosomes. Cell Res. 2016; 26: 149-50. doi: 10.1038/cr.2015.153.

46. Zhu Y, Chen X, Pan Q, Wang Y, Su S, Jiang C, Li Y, Xu N, Wu L, Lou X, Liu S. A Comprehensive Proteomics Analysis Reveals a Secretory Path- and Status-Dependent

Signature of Exosomes Released from Tumor-Associated Macrophages. J Proteome Res. 2015; 14: 4319-31. doi: 10.1021/acs.jproteome.5b00770.

47. Kowal J, Arras G, Colombo M, Jouve M, Morath JP, Primdal-Bengtson B, Dingli F, Loew D, Tkach M, Théry C. Proteomic comparison defines novel markers to characterize heterogeneous populations of extracellular vesicle subtypes. Proc Natl Acad Sci U S A. 2016; 113: E968-77. doi: 10.1073/pnas.1521230113.

48. Jongen-Lavrencic M, Salesse S, Delwel R, Verfaillie CM. BCR/ABL-mediated downregulation of genes implicated in cell adhesion and motility leads to impaired migration toward CCR7 ligands CCL19 and CCL21 in primary BCR/ABL-positive cells. Leukemia. 2005; 19: 373-80. doi: 10.1038/sj.leu.2403626.

49. Christianson HC, Svensson KJ, van Kuppevelt TH, Li JP, Belting M. Cancer cell exosomes depend on cell-surface heparan sulfate proteoglycans for their internalization and functional activity. Proc Natl Acad Sci U S A. 2013; 110: 17380-5. doi: 10.1073/pnas.1304266110.

50. McKiernan J, Donovan MJ, O'Neill V, Bentink S, Noerholm M, Belzer S, Skog J, Kattan MW, Partin A, Andriole G, Brown G, Wei JT, Thompson IM, et al. A Novel Urine Exosome Gene Expression Assay to Predict High-grade Prostate Cancer at Initial Biopsy. JAMA Oncol. 2016; 2: 882-9. doi: 10.1001/jamaoncol.2016.0097.

51. van Niel G, Bergam P, Di Cicco A, Hurbain I, Lo Cicero A, Dingli F, Palmulli R, Fort C, Potier MC, Schurgers LJ, Loew D, Levy D, Raposo G. Apolipoprotein E Regulates Amyloid Formation within Endosomes of Pigment Cells. Cell Rep. 2015; 13: 43-51. doi: 10.1016/j.celrep.2015.08.057.

52. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson Å, Kampf C, Sjöstedt E, Asplund A, Olsson I, Edlund K, Lundberg E, et al. Proteomics. Tissue-based map of the human proteome. Science. 2015; 347: 1260419. doi: 10.1126/science.1260419.

53. McMahon GA, Petitclerc E, Stefansson S, Smith E, Wong MK, Westrick RJ, Ginsburg D, Brooks PC, Lawrence DA. Plasminogen activator inhibitor-1 regulates tumor growth and angiogenesis. J Biol Chem. 2001; 276: 33964-8. doi: 10.1074/jbc.M105980200.

54. Fortenberry YM, Brandal SM, Carpentier G, Hemani M, Pathak AP. Intracellular Expression of PAI-1 Specific Aptamers Alters Breast Cancer Cell Migration, Invasion and Angiogenesis. PLoS One. 2016; 11: e0164288. doi: 10.1371/journal.pone.0164288.

55. Hsu YC, Yuan S, Chen HY, Yu SL, Liu CH, Hsu PY, Wu G, Lin CH, Chang GC, Li KC, Yang PC. A four-gene signature from NCI-60 cell line for survival prediction in non-small cell lung cancer. Clin Cancer Res. 2009; 15: 7309-15. doi: 10.1158/1078-0432.CCR-09-1572.

56. Langfelder P, Zhang B, Horvath S. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package

57. Meckes DG. Affinity purification combined with mass spectrometry to identify herpes simplex virus protein-protein interactions. Methods Mol Biol. 2014; 1144: 209-22. doi: 10.1007/978-1-4939-0428-0_14.

58. Vizcaíno JA, Deutsch EW, Wang R, Csordas A, Reisinger F, Ríos D, Dianes JA, Sun Z, Farrah T, Bandeira N, Binz PA, Xenarios I, Eisenacher M, et al. ProteomeXchange provides globally coordinated proteomics data submission and dissemination. Nat Biotechnol. 2014; 32: 223-6. doi: 10.1038/nbt.2839.

59. Liu H, Sadygov RG, Yates JR. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. Anal Chem. 2004; 76: 4193-201. doi: 10.1021/ac0498563.

60. Old WM, Meyer-Arendt K, Aveline-Wolf L, Pierce KG, Mendoza A, Sevinsky JR, Resing KA, Ahn NG. Comparison of label-free methods for quantifying human proteins by shotgun proteomics. Mol Cell Proteomics. 2005; 4: 1487-502. doi: 10.1074/mcp.M500084-MCP200.

61. McIlwain S, Mathews M, Bereman MS, Rubel EW, MacCoss MJ, Noble WS. Estimating relative abundances of proteins from shotgun proteomics data. BMC Bioinformatics. 2012; 13: 308. doi: 10.1186/1471-2105-13-308.

62. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014; 15: 550. doi: 10.1186/s13059-014-0550-8.

63. Zhang L, Liu YZ, Zeng Y, Zhu W, Zhao YC, Zhang JG, Zhu JQ, He H, Shen H, Tian Q, Deng FY, Papasian CJ, Deng HW. Network-based proteomic analysis for postmenopausal osteoporosis in Caucasian females. Proteomics. 2016; 16: 12-28. doi: 10.1002/pmic.201500005.

64. Kommadath A, Bao H, Arantes AS, Plastow GS, Tuggle CK, Bearson SM, Guan lL, Stothard P. Gene co-expression network analysis identifies porcine genes associated with variation in Salmonella shedding. BMC Genomics. 2014; 15: 452. doi: 10.1186/1471-2164-15-452.

65. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008; 9: 559. doi: 10.1186/1471-2105-9-559.

66. Dolédec S, Chessel D. (1994). Co-inertia analysis: an alternative method for studying species-environment relationships. Freshwater Biology), pp. 277-94.

67. Culhane AC, Perrière G, Higgins DG. Cross-platform comparison and visualisation of gene expression data using co-inertia analysis. BMC Bioinformatics. 2003; 4: 59. doi: 10.1186/1471-2105-4-59.

68. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, Yefanov A, Lee H, Zhang N, et al. NCBI

for R. Bioinformatics. 2008; 24: 719-20. doi: 10.1093/bioinformatics/btm563.

GEO: archive for functional genomics data sets--update. Nucleic Acids Res. 2013; 41: D991-5. doi: 10.1093/nar/gks1193.

69. Pfister TD, Reinhold WC, Agama K, Gupta S, Khin SA, Kinders RJ, Parchment RE, Tomaszewski JE, Doroshow JH, Pommier Y. Topoisomerase I levels in the NCI-60 cancer cell line panel determined by validated ELISA and microarray analysis and correlation with indenoisoquinoline sensitivity. Mol Cancer Ther. 2009; 8: 1878-84. doi: 10.1158/1535-7163.MCT-09-0016.

70. Kohn KW, Zeeberg BM, Reinhold WC, Pommier Y. Gene expression correlations in human cancer cell lines define molecular interaction networks for epithelial phenotype. PLoS One. 2014; 9: e99269. doi: 10.1371/journal.pone.0099269.

71. Meng C, Kuster B, Culhane AC, Gholami AM. A multivariate approach to the integration of multi-omics datasets. BMC Bioinformatics. 2014; 15: 162. doi: 10.1186/1471-2105-15-162.

72. Huang dW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 2009; 37: 1-13. doi: 10.1093/nar/gkn923.

73. Huang dW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009; 4: 44-57. doi: 10.1038/nprot.2008.211.

74. Pathan M, Keerthikumar S, Ang CS, Gangoda L, Quek CY, Williamson NA, Mouradov D, Sieber OM, Simpson RJ, Salim A, Bacic A, Hill AF, Stroud DA, et al. FunRich: An open access standalone functional enrichment and interaction network analysis tool. Proteomics. 2015; 15: 2597-601. doi: 10.1002/pmic.201400515.