Multi-omics discovery of hallmark protein and lipid features of circulating small extracellular vesicles in humans

Alin Rai[1,2,3*], Kevin Huynh[1,2,3], Qi Hui Poh[1], Haoyun Fang[1,2], Bethany Claridge[1], Jonathon Cross[1], Thy Duong[1], Carla Duarte[1], Thomas H Marwick[1,2,3], Peter Meikle[1,2,3], David W Greening[1,2,3*]

[1]*Baker Heart and Diabetes Institute, Melbourne, Victoria, Australia.*

[2]*Baker Department of Cardiometabolic Health, University of Melbourne, Melbourne, Victoria, Australia.*

[3]*Baker Department of Cardiovascular Research Translation and Implementation, La Trobe University, Melbourne, Victoria, Australia*

*To whom correspondence should be addressed:

Alin Rai, PhD

Baker Heart and Diabetes Institute

75 Commercial Road, Melbourne, 3004, Australia

Tel: +61-3-8532-1217

Email: alin.rai@baker.edu.au

David W. Greening, PhD

Baker Heart and Diabetes Institute

75 Commercial Road, Melbourne, 3004, Australia

Tel: +61-3-8532-1585

Email: david.greening@baker.edu.au

ORCID

Alin Rai - 0000-0001-7994-5151; David W Greening - 0000-0001-7516-485X; Thomas H Marwick - 0000-0001-9065-0899

Keywords

extracellular vesicles, plasma, proteomics, lipidomics, hallmark features/markers, repository

Figures: 7

Abstract

Extracellular vesicles (EVs) are now being increasingly recognized as an essential signaling entity in human plasma, linking them to health and various diseases. Still, their core protein and lipid componentry, which lie at the center of EV form and function, remains poorly defined. Achieving this unmet milestone remains greatly hindered by abundant non-vesicular extracellular plasma components (non-EVs) in mass spectrometry-based analyses. Here, we performed high-resolution density gradient fractionation of over 110 human plasma samples to isolate circulating EVs, and systematically construct their quantitative proteome (4500 proteins) and lipidome (829 lipids) landscapes. This led to the discovery of a highly conserved panel of 182 proteins (ADAM10, STEAP23, STX7) and 52 lipids (PS, PIPs, Hex2Cer, PAs), providing a deep survey of hallmark molecular features and biological pathways intrinsic to circulating EVs. Our efforts also mapped the surfaceome diversity, identifying 151 proteins on EV surface. We further establish a set of 42 proteins and 114 lipids features that served as hallmark features of non-EV particles in plasma. We submit ADAM10 and PS(36:1) as conserved EV biological markers that precisely differentiates between EV and non-EV particles. Our findings, which can be explored via open-source Shiny web tool (evmap.shinyapps.io/evmap/) will serve as a valuable repository to the research community for a clearer understanding of circulating EV biology.

Introduction

Extracellular vesicles (EVs) are membrane-enclosed nanoscale particles (30–1000 nm in diameter) released by cells into their extracellular space[1,2]. By transferring bioactive cargo such as proteins, lipids, nucleic acids, and metabolites between cells[3-6], EVs execute diverse biological functions in various physiological and pathological processes. EVs are also found ubiquitously in human circulatory system[7-15]; with an estimated 50 x 10^6 particles per ml of plasma[9], EVs are now recognized as essential signaling entity in human plasma. While their precise functions in humans remain mostly elusive, circulating EVs have been implicated in essential life processes, such as immune regulation[16], inter-organ crosstalk[17,18], tissue homeostasis and regeneration[19] and coordinating physiological responses[20-22]. Dysregulating in their cargo composition is also implicated in various human diseases such as cancer[23], cardiovascular disease[24], artery calcification[20], and COVID-19 pathogenesis[25]. As minimally invasive liquid biopsies, they have also garnered interest for their potential diagnostic value[9] and real-time monitoring of therapeutic response[26].

Despite their wide biological and biomedical implications, the study of circulating EVs remains exceptionally challenging[27], with their conserved protein and lipid componentry remaining poorly characterized and an unmet milestone. Given that conserved molecular blueprint epicentres form and function of a biological system - sharing unique proximity to phenotype and pathophysiology[28] - closing this knowledge gap is critical for advancing our fundamental understanding of circulating EVs and harnessing their clinical potential. Recently, while several seminal studies[29,30] have begun to define the precise core of EVs using *in vitro* culture systems; making such discoveries for EVs circulating in human plasma remains a formidable challenge. This is mainly due to the presence of large abundance of non-EV components (such as lipoprotein particles, soluble proteins, complement proteins, immunoglobulins) in plasma that outnumber EVs by six to seven orders of magnitude[31]. Such non-EV particles co-purify with EVs which invariably limits mass spectrometry-based quantifications primarily to high-abundant plasma proteins or lipoproteins particle-associated lipids, resulting in incomplete and low-coverage data[14,23,25]. Thus, there is an ongoing international effort within the EV community[9,11,30,32-34] to develop and refine plasma EV isolation methods, with the aim to precisely define their molecular landscape[10,12,14,35,36].

A systematic construction of the conserved protein and lipid componentry of circulating EVs in humans has several implications. It will inform us on fundamental building blocks of circulating EVs, providing high-confidence molecular maps and associated biological pathways such as biogenesis (including membrane curvature, stability, cargo recruitment), release, environmental interactions, and uptake mechanisms in humans. Given that existing EV markers obtained from cell culture systems have limited conservation in humans[23] - potentially due to unique architecture and context of human tissues[37,38] – these conserved features will also serve as robust EV markers applicable to human plasma that can be rigorously implemented in large-scale population studies, advance EV purification and characterization techniques, and extend international EV guidelines[27] for standardized circulation EV research. Additionally, these conserved features will bridge the knowledge gap between humans and cultures/animal models, enhancing knowledge transferability and translatability. Other implications include informing on strategies for engineering and functionalizing EV membranes for improved drug delivery vehicles (e.g., extended half-life in circulation), and developing safer EV-based therapies.

In this study, we employed high-resolution density gradient separation to isolate a major EV sub-type, called small EVs from human plasma. We verify the enrichment strategy and EV identity using various biochemical and biophysical characterization, ensuring a high degree

of separation of EVs and non-EV particles in plasma. We then construct their detailed proteome and lipidome maps, identifying 182 proteins and 52 lipids as core protein and lipid componentry of EVs, which we refer to as EV hallmark features. We also identify 29 proteins and 114 lipids that are defining features of non-EV particles. These markers, in particular ADAM10 and PS(36:1), enable precise differentiation between EV and non-EV particles using machine-learning. To enhance data accessibility, we developed an open-source R/Shiny web tool (evmap.shinyapps.io/evmap/).

Results

Isolation of circulating EVs from human plasma

While there is no method that can isolate EVs to absolute purity, high-resolution iodixanol-based density gradient separation (DGS)[39-41] remains a powerful strategy for enriching EVs from complex biofluids such as plasma. To this end, we subjected plasma to top-down DGS **(Fig. 1A-B)** which resolved the majority of signals for abundant plasma components such as albumin, apolipoproteins, and argonaut 2 (AGO2, known to associate with non-EV RNA extracellularly[42]) to DGS fractions 1-5. In contrast, CD63 (tetraspanins that are found in small EVs) resolved in DGS fractions 6-8 (corresponding to flotation buoyancy of ~1.09 g/ml, typical of EVs) **(Fig. 1A-B)**. This resolution remained inaccessible to ultracentrifugation (100,000 *g*, referred to as p100K) alone that co-isolated large quantities of abundant plasma components **(Supplementary Fig. 1A)**.

EV particles in DGS (6-8) fractions (pEVs; plasma EVs) were morphologically intact, membrane-limited spherical vesicles, consistent with previous reports[43-46] **(Fig. 1C)**, with a mean diameter of 220.4 nm (**Supplementary Fig. 1B)**. In contrast, in p100K contained abundance of proteaceous material that formed aggregates, whereas DGS (1-3) fractions (referred to as pDGS.LD; plasma DGS light density particles) contained spherical structures without an apparent lipid membrane (resembling size and morphology of lipoprotein particles[8]).

Moreover, nanoparticle tracking analysis **(Fig. 1D)** revealed that pEVs ranged from 30-300 nm in size and displayed a net negative charge **(Fig. 1E)**, consistent with a previous report[47]. Thus, our data show that we can successfully enrich for small EVs from human plasma with minimal contamination from non-EV plasma components. We further demonstrate EV enrichment from multiple plasma samples **(Fig. 1D-E)**. From 1 ml of plasma, we obtained ~8.7 µg of pEVs (a striking >24,000 fold less compared to total plasma protein) **(Supplementary Fig. 1C)**, which enumerated to ~4.2 x 10^9 particles **(Supplementary Fig. 1D-E)**.

Constructing the proteome draft of circulating EVs in humans

We next performed MS-based proteomics analysis of pEVs from 38 human plasma samples from multiple sources **(Supplementary Fig. 2A)**. To identify EV specific proteins, we compared the pEV proteomes with those of pDGS.LD, p100K and unprocessed plasma, collectively termed as "NonEVs" (n=42) **(Fig. 2A, Supplementary Fig. 2B, Supplementary Tables 1-6)**. Using stringent peptide and protein identification criterion (1% false discovery rate), we quantified 4631 proteins in pEVs and 1678 in NonEVs **(Supplementary Fig. 2C, Supplementary Table 6)**. The size of pEV proteome dataset was comparable to those

obtained for In vitro EV proteome data from four different cell lines (4492 proteins), ensuring sufficient proteome coverage.

The pEV proteome displayed a remarkable dynamic range **(Supplementary Fig. 2D)**, including extensive coverage of low-abundant circulating proteins[48] **(Supplementary Fig. 2E)**, and was distinct to NonEV proteome **(Fig. 2B, Supplementary Fig. 2F).** Moreover, differentially abundant proteins in pEVs were enriched for terms/pathways associated with small EVs (regulation of actin cytoskeleton and endocytosis) **(Supplementary Fig. 3A-C, Supplementary Table 7-8)** and EV biogenesis proteins[29,44,46,49,50] **(Supplementary Fig. 3D),** while NonEVs contained components related to complement and coagulation, and abundant plasma proteins[51] **(Supplementary Fig. 3A-D)**, further supporting our EV enrichment pipeline at omics level.

Defining the core protein features of circulating EVs

Echoing a recent report[23], although several EV proteins (namely, 22 core proteins reported in *in vitro* EVs[29] and 94 EV marker proteins recommended by the MISEV guidelines[27,35]) were detected in pEVs **(Supplementary Fig. 4A, Supplementary Tables 6)**, many were not universally quantified **(Supplementary Fig. 4B-D)**.

On the other hand, occurrence analysis identified 259 proteins ubiquitously quantified in all 38 pEV samples (termed Category 1 proteins**, Fig. 2C, Supplementary Table 9)**; these proteins were enriched for EV related terms/processes (**Extended data 1A, Supplementary Table 10)** and encompass proteins associated with endosomal trafficking network **(Extended data 1B)** such as flotillins, integrins, and CD proteins, whose close interconnectedness is highlighted by their protein-protein interaction network **(Extended data 1C)**.

Amongst the 259 conserved proteins, we further distinguish 182 proteins present uniquely or in significantly higher abundance in pEVs compared to NonEVs: we refer to them as "pEV protein features" **(Fig. 2D, Extended data 2, Supplementary Table 9**).

To ascertain protein features of NonEVs, we set two criteria- 1) Category 1 proteins present in significantly higher abundance in NonEVs compared to pEVs (28 proteins, **Fig. 2D, Supplementary Table 9)** and 2) proteins quantified in over 70% of NonEV samples but less than 30% of pEVs samples (14 proteins, **(Extended data 3A, Supplementary Table 9)**. Collectively, these 42 proteins (28+14) will be referred to as "NonEV protein features" **(Extended data 3B)**.

Emphasizing their significance as essential EV components, the pEV protein features were enriched for classical EV pathways **(Fig. 2E, Supplementary Table 11)**. These pEV protein features are known to functionally interact to execute several EV-associated biological processes including vesicular transport, actin cytoskeleton regulation, and membrane raft assembly (**Fig. 2F)**. In contrast, NonEV protein features were enriched for "endopeptidase inhibitor activity, "Complement and coagulation cascades", and "Cholesterol metabolism" (**Fig. 2E, Supplementary Table 11)**.

We next queried whether the pEV protein features also comprised the EV surfaceome, essential interactive surface platforms on EVs[52]. To explore this, surface-accessible proteins on pEV were labelled with membrane-impermeant Sulfo-NHS-SS-Biotin, captured with avidin-coated beads, and analysed through MS-based proteome profiling (n=9) (**Supplementary Fig. 5A, Supplementary Table 12)**. Our data revealed that 151 pEV protein features were surface-accessible, including CD proteins (CD44, CD47), integrins

(ITGA2, ITGA6, ITGB1) and annexins (ANXA2, ANXA5) **(Fig. 2G, Supplementary Fig. 5B-C)**. We further confirmed their surface-accessibility in *in vitro* EVs **(Fig. 2C, Supplementary Table 12)**. In contrast, the conserved SDCBP protein remained inaccessible to biotin-capture, which is in direct agreement to its luminal localization[29].

By leveraging data from published proteomes[15,53,54], we demonstrated a remarkable conservation of pEV protein features in multiple plasma or serum EVs **(Fig. 3, A-C, Supplementary Tables 9)**. These features were also conserved in different EV sub-populations (CD81+, CD63+, and CD9+ EVs) prevalent in the circulatory system **(Fig. 3, D-F)**, with 100% quantification of at least 74 proteins. Moreover, interrogating previous published sEV proteomes from fourteen cell lines[29], 81 EV features were also conserved, displaying greater abundance in EVs compared to cells (**Supplementary Fig. 6A-B)**. Additionally, we experimentally verified their conservation in In vitro EVs (n=12) release by four cell lines **(Fig. 3, I-L)**. Finally, we validated their conservation in pEVs (n=4) and In vitro EVs (n=6) using orthogonal TMT-based isobaric multiplexed proteomics **(Fig. 3, G-H, Supplementary Table 13)**. Overall, 78 pEV proteins displayed 100% conservation, of which 63 were surface-accessible (**Supplementary Table 9)**. Moreover, 68 protein features displayed similar abundance in pEVs from individuals with and without coronary artery calcium (CAC) deposit in our EDCAD cohort **(Supplementary Table 14**), highlighting their potential conservation in disease states.

Thus, our study defines highly conserved protein features of circulating EVs in human plasma (**Table 1** highlights top 25 EV or NonEV protein features).

The lipidome draft of circulating EVs in humans

Next, we defined the lipidome landscape of circulating EVs. The majority of membrane lipids fall under glycosphingolipids, sphingolipids and sterols (predominantly cholesterol in mammals). Therefore, to capture this lipid diversity, we employed a high-throughput targeted-lipidomics platform interrogating 829 lipids representing 40 lipid classes within these three major groups **(Fig. 4A, Supplementary Fig. 7, Supplementary Table 15)**. Because the lipidome of small EVs from cells are not as well defined compared to their proteome counterpart, we also performed lipidome analysis of In vitro EVs and parental cells using the same platform **(Supplementary Fig. 7A-B, Supplementary Table 15)**. We reasoned that circulating EV lipid features should ideally be enriched in In vitro EVs compared to cells.

Principal Component Analysis revealed that pEVs and NonEVs particles (p100K/pDGS.LD) lipidomes were distinct **(Fig. 4B)**. Highly-abundant lipids in both pEVs and In vitro EVs included cholesterol (COH), phosphatidylcholines (PC), phosphatidylethanolamine (PE), and sphingomyelin (SM) **(Fig. 4C)**; lipids that are well-recognized major structural components in eukaryotic cell membranes[55] with housekeeping structural functions. For example, PC forms a planar bilayer with COH, ensuring membrane stability and integrity, while the incorporation of conical PE and SM imposes a curvature stress crucial for membrane budding, fission and fusion. These findings align with previous reports highlighting COH as one of the most abundant EV lipids (40–60% of total lipids)[56].

Over 30% of measured lipid species displayed differential abundance (adjusted p values <0.05, fold change >1.5) between pEVs and NonEVs (p100K/pDGS.LD) **(Fig. 4D, Supplementary Fig. 8A-B, Supplementary Table 16)**. Lipids enriched in pEVs are linked to EV biogenesis[56] and include dihydroceramides (dhCer)[57], trihexosylceramides (Hex3Cer),

dihexosylceramides (Hex2Cer) (**Supplementary Fig. 8C)**. These lipids are similarly enriched in In vitro EVs *versus* their parental cells **(Fig. 4D, Supplementary Fig. 8D-E, Supplementary Table 16)**. In contrast, lipids enriched in NonEVs include the lipid classes cholesterol ester (CE), triacylglycerol (TG) and coenzyme Q10 (CoQ10)[58], major components of lipoprotein particles. This further supports our EV enrichment pipeline at lipidomics level.

The core lipidome of circulating EVs

To identify pEV lipid features, we performed K-means clustering of differentially abundant lipids between pEVs *versus* NonEVs and In vitro EVs *versus* cells **(Fig. 5A, Supplementary Tables 17-18)** which resulted in four major clusters. Representing pEV lipid features were Cluster c4 lipids (52 lipids) enriched in pEVs (compared to NonEVs) and In vitro EVs (compared to cells). In contrast, representing NonEV lipid features were Cluster c1 lipids (114 lipids) enriched in NonEVs and cells.

Notably, we observed a class-level co-regulation of pEV lipid features **(Fig. 5B),** supporting previous reports that lipid species from the same pathway correlated/co-regulated[59]. For example, lipids from PS and PE class were co-enriched in EVs, which is indicative of their co-localization to external leaflets of EV membranes and co-regulating similar cellular processes[60,61]. Another example includes SM and ceramides co-enriched in EVs which is indicative of SM-ceremide axis driven EV biogenesis[57,62].

Emphasizing their significance as essential EV lipid components, Cluster c4 lipid features **(Extended data 4)** were linked to LION Ontology terms such as bilayer membrane and plasma membrane **(Fig. 5C)**, but more specifically with EV features such as glycophosphoserines (known to decorate the outer membrane/leaflet of EVs) and low lateral diffusion (a distinct physical property of lipid rafts[63] that are hubs for EV formation). Conversely, Cluster c1 NonEV lipid features **(Extended data 4)** were linked to lipid storage/droplets, typical of lipoprotein particles.

To validate the association of Cluster c4 lipids with circulating EVs, and Cluster c1 with NonEV particles in plasma, we subjected twelve fractions of plasma DGS (n=3 plasma samples) to the same targeted lipidomics platform **(Fig. 5D-E, Supplementary Fig. 9A-B, Supplementary Table 18-19)**. The normalized abundance of lipids across the fractions were plotted (z-scored within each fraction) **(Fig. 5F)**. Indeed, Cluster c4 lipids were enriched in DGS fractions 6 and 7, where pEVs resolve. In contrast, Cluster c1 lipids were enriched in DGS fractions 1-4, where the majority of NonEV particles resolve, reinforcing the specificity of Cluster c4 lipids for circulating EVs in plasma.

Thus, our study defines highly conserved lipid features of circulating EVs in human plasma (**Table 2** highlights selected EV or NonEV lipid features).

Biological protein and lipid markers for circulating EVs

We next investigated the ability of the EV and NonEV protein feature panel to distinguish between pEVs vs NonEV particles using machine learning (Naïve Bayes algorithm). For this, pEV and NonEV proteomes were evenly partitioned based on sample type into training set (70% of the samples) and a validation set (remaining 30% of samples). By bootstrapping resampling method employed with 25 resampling iterations, the model achieved absolute

accuracy (**Supplementary Fig. 10A)**. Our model also achieved excellent accuracy (97%) in EV particle identification in an independent test set comprising of pEVs and NonEVs proteomes from additional 16 plasma samples (**Supplementary Fig. 10A)**. Moreover, a panel comprising 151 surface accessible EV protein features were also able to distinguish between EV and NonEV particles with 100% accuracy (**Supplementary Fig. 10B)**. Of the 182 EV protein features, Recursive Feature Elimination based random forest algorithm identified ADAM10 as a prominent feature contributing to model performance **(Supplementary Table 20)**. Moreover, ADAM10 protein showed exclusive and absolute quantification in pEVs as well as In vitro EVs, compared to NonEVs (**Fig. 6A**). Moreover, within each EV proteome the abundance of ADAM10 was up to 7.5-fold higher compared to the median intensity (**Fig. 6B, Supplementary Table 21**); this relative abundance of ADAM10 within each proteome alone can distinguish between pEVs vs NonEV particles (**Supplementary Fig. 10C)**.

Similarly, using a Neural Network algorithm ('nnet') for machine learning, a panel comprising of EV and NonEV lipid features were also able to distinguish between EVs (pEVs or In vitro EV lipidomes) vs NonEVs (p100K, pDGS.LD or cell lipidomes) (**Supplementary Fig. 11A)**. Moreover, this model achieved 97% accuracy in classifying EV particles across DGS plasma fractions (**Supplementary Fig. 11B)**. Using Recursive Feature Elimination, we identified PS(36:1) (EV lipid feature) and CE(18:0) as prominent lipid features **(Supplementary Table 20)**, highlighting their marker potential.

Indeed, PS(36:1) lipid is one of the most abundant lipids present in both pEVs and In vitro EVs **(Fig. 4C**), making them amenable to robust and reliable measurements in diverse analytic tools. Moreover, PS(36:1) and CE(18:0) lipid abundances, and importantly their relative abundances, concur with EV protein CD63 and lipoprotein APOB100 resolution on plasma DGS fractions **(Fig 6. D-E)**. Our machine learning ('nnet') model, based on PS(36:1) and CE(18:0) relative abundance within each lipidome serving as a feature **(Supplementary Table 22)**, can distinguish EVs (pEVs or In vitro EV, DGS fractions 6-7 lipidomes) vs NonEVs (p100K, pDGS.LD or DGS fractions 1-5 lipidomes) with absolute accuracy (**Fig. 6F)**. Because EV and NonEV enrichments along 12 fractions represent a continuum, which can be inferred from PS(36:1) and CE(18:0) relative abundances, our data suggests that these lipids can thus serve as a marker for assessment of EV purity. Importantly, quantitative measurement would be possible in subsequent studies due to the availability of isotope labelled internal standards for these two species, which assists in translation across different laboratories and instrumental setups.

We next investigated whether PS and ADAM10 constitute protein and lipid biological marker componentry on circulating EVs. For this, we employed previously reported[64] strategy of using extracellular region of T-cell immunoglobulin domain and mucin domain-containing protein 4 (Tim4) immobilized on magnetic beads to directly capture (in the presence of calcium ions) PS+ EVs from plasma (n=23), which was subsequently released (by adding EDTA) and subjected to MS proteomics **(Fig. 6G, Supplementary Table 23)**. Compared to mock capture (beads alone) or unprocessed plasma, PS+ EV proteomes displayed 100% quantification of ADAM10 across all twenty-three samples **(Fig. 6G)**. Additional EV protein features (e.g., FLOT2, RAB7A) were also conserved in these proteomes, which support their EV identity.

Thus, our data show that ADAM10 and the ratio combination of PS(36:1) and CE(18:0) serves as highly conserved and reliable biological markers for EVs in human plasma.

R/Shiny web tool for EV proteome and lipidome data

Lastly, to facilitate easy access to our data and enhance re-use, we have developed an open-source R/Shiny web tool (evmap.shinyapps.io/evmap/). This tool allows users to quickly interrogate our proteome **(Supplementary Fig. 12)** and lipidome datasets **(Supplementary Fig. 13)** for their molecule(s) of interest. This tool also allows assessment of feature conservation in published studies, surface-accessibility of EV proteins, and construct network analysis, retrieve Gene Ontology and KEGG pathways for selected protein features. Lipid features can also be quickly interrogated for their abundance in EVs vs NonEV particles and their distribution in plasma DGS lipidome fractions. We anticipate that this tool will enable broad utilization of our data and will serve as a valuable repository for the broader EV community.

Discussion

In our study, we integrated multi-omics investigation to systematically resolve the core protein and lipid componentry of circulating EVs in humans **(Figure 7)**. Our discovery includes a conserved set of 182 proteins and 52 lipids intrinsic to circulating EVs, and a panel of 29 proteins and 114 lipids that are NonEV features in plasma, which together serves as biological markers for EV research applicable to human samples. This extensive protein and lipid landscapes, which can be easily accessed with the Shiny web tool, will be instrumental to the EV community for clearer understanding of circulating EV biology. This includes the survey of circulating EV surfaceome, comprise of 151 EV protein features, that could be potentially exploited for antibody-based capture of circulating EVs based on their surface protein expression(s). The highly conserved protein ADAM10 on circulating EVs is also included in this EV surfaceome. The role of ADAM10 as conserved and robust EV marker is also supported by various studies that have reported ADAM10 expression in circulating EVs in human plasma, different biofluids, and EVs released by diverse tissue cell types using various biochemical analysis, different EV isolation strategies. Moreover, ADAM10 was also highly conserved in different circulating EV subpopulations (based on surface expression of tetraspanins, CD9, CD63, CD81[54]). The surface localization of ADAM10 is also corroborated in previous reports[65]. On the other hand, EV lipid features PS[15] has already been developed for capturing circulating EVs using TIM4-based magnetic beads.

Our data also provides “high-confidence” molecular leads for studying EV biology in multimodal organisms, particularly targeting EV biogenesis to curb EV-driven pathogenesis in multimodal organisms such as the mouse[62] which remains largely unexplored. For instance, EV protein feature STEAP3 (also known as TSAP6) has been shown to regulate EV biogenesis in mice[66]. A holistic analysis of the core protein and lipid componentry can also provide insights into biological pathways co-involved in proteins and lipids; one such biology includes the potential involvement of lipid raft biology in human EVs. For instance, EV features such as flotillins[67], SM and PS, along with cholesterol that are the most abundant lipids in EVs, are core components of lipid raft domains[63]. Assembly of flotillins in SM/ceremide[63]-rich microdomains induces curvature stress[68] to regulate EV biogenesis and cargo selection[57,69] in ESCRT[70]- and syntenin/syndecan-independent mechanisms[71]. Moreover, proteins that make the cortical actin (KEGG term enriched in EV core proteins) helps maintain and remodel these domains[63]. Accumulating evidence suggests that lipid rafts are nidus of EV biogenesis and are present in EVs as functional units[69]. Ubiquitous identification of these features in pEVs highlights similar EV biogenesis mechanisms conserved for circulating EVs in humans.

A major caveat in our study is that it primarily focuses on bulk preparation of small EVs present in human plasma. Cells release heterogeneous EV populations[35,44,72,73] that display functional diversity[74]. They can be categorized based on their origin[44,75,76], density[44,49,73], biochemical composition[30,49] and size[46,49]. Since current EV isolation strategies cannot purify a specific subset of EVs to homogeneity, operational terms such as small EVs (30-150 nm) and large EVs (100-1000 nm) are encouraged where evidence for subcellular origin cannot be ascertained. In our study, small EVs potentially include both exosomes (endosomal-derived) and microvesicles (shed from plasma membrane) evident from our EV protein markers that include markers of both EV subtype (TSG101, ANXA2). Thus, our EV markers are applicable in a broad sense to small EVs; their specificity in exosomes *versus* microvesicles can potentially provide greater resolution in human EV biology and thus warrants future investigation. Moreover, EV heterogeneity is reported in circulating EVs too based on surface expression of tetraspanins, CD9, CD63, CD81[54]. Nonetheless, our EV protein features show high conservation in all three EV subpopulations[54]. While bulk EV study provides a broader understanding of the complex interactions and processes involving EVs within the body that specific EV subtypes might not fully capture, it is essential to study EV subpopulation-specific core features in humans, as this avenue holds great potential for advancing precise diagnostics and therapeutics.

In summary, we identify core protein and lipid componentry of circulating EVs in humans. We submit these features as “hallmark molecular features” of circulating EVs that serve as valuable resource to the EV community and a tool to enhance the quality of human EV-research.

Methods

Generation of cell conditioned media

SW480 (CCL-288, ATCC), SW620 (CCL-227, ATCC) and LIM1863 cells[77] (Ludwig Institute for Cancer Research, Melbourne) cells were cultured in RPMI-1640 (Life Technologies). MDA MB 231 (HTB-26, ATCC) was cultured in DMEM (Life Technologies). Complete culture media included media supplemented with 5% (v/v) Fetal Bovine Serum (FBS, Life Technologies) and 1% (v/v) Penicillin Streptomycin (Pen/Strep, Life Technologies) at 37 °C with 10% CO_2 as described[44,52]. Cells were passaged with trypsin-EDTA (Gibco). Cells were cultured in CELLine AD-1000 Bioreactor classic flasks (Integra Biosciences) for conditioned media generation as described[52].

Plasma preparation

Human blood plasma samples were obtained from Australian Red Cross Lifeblood or EDCAD study. For Red Cross samples, ethical permits were obtained from the ethical committee of the Australian Red Cross Blood Service Human Research Ethics Committee, and La Trobe University Human Ethics Committee (HEC19485). Blood samples were collected via aseptic venipuncture into commercial EDTA-containing sampling containers at room temperature and centrifuged at 4,200 *g* for 10 min. The top plasma supernatant was carefully collected, and further centrifuged at 5,000 *g* (15 min) at room temperature. The supernatant was immediately isolated and stored as 1 ml aliquots at -80°C until further use. For EDCAD samples, ethics permit was approved through Human Research Ethics Committee (HREC) at Baker Heart and Diabetes Institute and by the Alfred Hospital Ethics Committee (EDCAD-PMS, #492/20). Blood was collected via venipuncture into a sterile EDTA tube and centrifuged at 2,700 g at room temperature for 13 min. A small volume (1.6 ml) of the supernatant (plasma) was transferred into a clean sterile tube and Butylated Hydroxytoluene solution (BHT) added to a final concentration of 100 µM. The Plasma/BHT sample was mixed and equally transferred into 3 x 0.75 ml Fluid X cryotubes. Plasma/BHT samples were snap frozen on dry ice prior to being transferred and stored at -80°C until use. For all plasma collections, only ~80% of top plasma supernatant was collected to avoid disturbing the buffy coat. Samples with potential hemolysis, cellular contamination and high fat/lipid content were not used for this study.

Ultracentrifugation of plasma

To generate crude 100K plasma pellet, plasma samples were thawed on ice and subjected to ultracentrifugation at 100,000 *g* for 1 hr or 18 hr at 4 °C (TLA-55 rotor; Optima MAX-MP Tabletop Ultracentrifuge, Beckman Coulter). Pellets were washed once in 1.0 ml PBS, resuspended in 100 µl PBS and stored at -80°C until further use.

Direct density-gradient separation

Plasma (500 µl) was directly overlaid on top of discontinuous gradient of OptiPrep™ (40% (3 ml), 20% (3 ml), 10% (3 ml) and 5% (2.5 ml) (diluent: 0.25 M sucrose / PBS solution)) and ultracentrifuged at 100,000 *g* for 18 hr (4 °C, 41 Ti rotor; Optima XPN Ultracentrifuge). Twelve 1 ml fractions were carefully collected top-bottom, diluted in PBS (2 ml) and ultracentrifuged at 100,000 *g* (1 h, 4°C, TLA-55 rotor; Optima MAX-MP Tabletop Ultracentrifuge). Fraction densities were determined as previously described[44]. Pellets were suspended in 100 µl PBS and stored at -80 °C until further use.

For isolation of cell culture-derived EVs, conditioned media was centrifuged at 500 *g* (5 min, 4 °C), 2,000 *g* (10 min, 4 °C), and 10,000 *g* (30 min, 4 °C, SW28 rotor; Optima XPN Ultracentrifuge, Beckman Coulter) to pellet large EVs[52]. The supernatant (1 ml) was

subjected to direct DGS for EV isolation as described for plasma above. EV pellets were reconstituted in 100 µl PBS and stored at -80 °C until further use.

Immunoblotting

Samples were subjected to protein quantification (microBCA™ Protein Assay Kit (23235, Thermo Fisher Scientific)), Western blotting (iBlot 2 Dry Blotting System, Thermo Fisher Scientific) was performed as described[44,52]. Rabbit antibodies raised against, Albumin (ab207327, Abcam), AGO2 (ab186733, Abcam) and Fibronectin (ab2413, Abcam) were used. Mouse antibodies CD63 (556019, BD Pharmingen™), CD81 (555675, BD Pharmingen™), APOB100 (3715-3-250, Mabtech), APOA1 (3710-3-1000, Mabtech) and ALIX (2171S, CST) were used. Secondary antibodies used were IRDye 800 goat anti-mouse IgG or IRDye 700 goat anti-rabbit IgG (1:15000, LI-COR Biosciences).

Biophysical characterization of EVs

Cryo-electron microscopy (Tecnai G2 F30) on samples (1 µg) was performed as described[46]. Briefly, Plasma 100K, D-DGS fractions 1-3 (pooled) and Plasma EVs (DGS fractions 6-8, pooled) (~1 µg) were loaded onto to glow-discharged C-flat holey carbon grids (ProSciTech Pty Ltd) that, after blotting away excess liquid, were plunge-frozen in liquid ethane and subsequently mounted in a Gatan cryoholder (Gatan, Inc.) in liquid nitrogen. Images were then acquired at 300 kV using a Tecnai G2 F30 (FEI) in low-dose mode. EV samples (~1 µg; 1:1000 dilution) were prepared in PBS (#14190-144, Thermo Fisher Scientific)[78] and particle size distribution and zeta potential (surface charge) determined by nanoparticle tracking analysis (ZetaView, Particle Metrix, PMX-120; 405 nm laser diode) according to manufacturers' instructions.

Surface biotin-labelling of EVs and proteomic sample preparation

EV surface proteins were captured using biotinylated using Pierce™ Cell Surface Biotinylation and Isolation Kit (A44390, Thermo Fisher Scientific), digested with trypsin (Promega, V5111), and peptides analyzed using Q-Exactive HF-X as described[52].

Label-free proteomics sample preparation

Label-free proteomics sample preparation (~2-10 µg in 20-50 µl) was performed as described[52], using single-pot solid-phase-enhanced sample preparation (SP3) workflow[79]. Briefly, samples were solubilized (1% (v/v) sodium dodecyl sulphate (SDS), 50 mM HEPES pH 8.0), incubated at 95 °C for 5 min and cooled. Samples were reduced with 10 mM dithiothreitol (DTT) for 45 min at 25 °C followed by alkylation with 20 mM iodoacetamide for 30 min at 25 °C in the dark. The reaction was quenched to a final concentration of 20 mM DTT. Magnetic beads were prepared by mixing SpeedBeads™ magnetic carboxylate modified particles (65152105050250, 45152105050250, Cytiva) at 1:1 (v:v) ratio and washing twice with 200 µl MS-water. Magnetic beads were reconstituted to a final concentration of 100 µg/µl. Magnetic beads were added to the samples at 10:1 beads-to-protein ratio, and 100% ethanol (EA043, ChemSupply) added for a final concentration of 50% ethanol (v/v). The sample bead mix were incubated at room temperature with constant shaking (1000 rpm). Protein-bound magnetic beads were washed three times with 200 µl of 80% ethanol and reconstituted in 50 mM TEAB and digested with trypsin (Promega, V5111) at a 1:50 enzyme-to-substrate ratio for 16 h at 37 °C with constant shaking (1000 rpm). The peptide mixture was acidified to a final concentration of 2% formic acid (pH <2) and centrifuged at 20,000 *g* for 1 min. The peptide digests were frozen at -80°C and dried by vacuum centrifugation (Savant SPD121P, Thermo Fisher Scientific), reconstituted in 0.07%

trifluoroacetic acid, and quantified by Fluorometric Peptide Assay (23290, Thermo Fisher Scientific) as per manufacturer's instruction.

TMT-based proteomics sample preparation

For TMT-based labelling was performed as described[80,81], using 10-plex TMT according to manufacturer's instructions with minor modifications (Thermo Fisher Scientific, 90406, lot UG287488/278919). The 10-plex experiment comprised 9 different chemical tags for EV peptide labelling, with the tenth tag used for generating reference channel made by a pooled peptide digest from all samples. A list of the sample labelling strategy is available in MassIVE proteomeXchange (MSV000094307). Peptide samples (15.5 µg) were labelled with TMT10plex™ Isobaric Label (90406, Thermo Fisher Scientific) at 4:1 label-to-peptide ratio for 2 hr at room temperature and quenched with 0.5% (v/v) hydroxylamine for 30 min at room temperature. Labelled peptide samples were acidified with 4% (v/v) formic acid (FA) and pooled into a new microtube. Pooled samples were desalted using Sep-Pak tC18 96-well µElution (186002318, Waters) and elution lyophilized by vacuum-based speedVac. Peptides were then fractionated into 20 fractions (with increasing concentration of ACN from 2-50%) using high pH reversed-phase chromatography using in-house SPE StageTips (SDB-RPS, Empore) as described[82]. Peptide eluates were lyophilized by vacuum-based SpeedVac, reconstituted in 0.07% TFA and quantified using Colormetric peptide assay. Peptide fractions with low yields were pooled.

Nano liquid chromatography–tandem mass spectrometry

Peptides (label-free and TMT-labelled) were analyzed on a Dionex UltiMate NCS-3500RS nanoUHPLC coupled to a Q-Exactive HF-X hybrid quadrupole-Orbitrap mass spectrometer equipped with nanospray ion source in positive mode as described.[52,80,83,84]

For label-free proteomics, peptides were loaded (Acclaim PepMap100 C18 3 µm beads with 100 Å pore-size, Thermo Fisher Scientific) and separated (1.9-µm particle size C18, 0.075 × 250 mm, Nikkyo Technos Co. Ltd) with a gradient of 2–28% acetonitrile containing 0.1% formic acid over 49 or 95 mins at 300 nl min^{-1} at 55°C (butterfly portfolio heater, Phoenix S&T). A list of the run details is available in MassIVE proteomeXchange (MSV000094307).

For data-dependent acquisition (DDA), MS1 scan was acquired from 350–1,650 *m/z* (60,000 resolution, 3×10^6 automatic gain control (AGC), 128 msec injection time) followed by MS/MS data-dependent acquisition (top 25) with collision-induced dissociation and detection in the ion trap (30,000 resolution, 1×10^5 AGC target, 27 msec injection time, 28% normalized collision energy, 1.3 *m/z* quadrupole isolation width). Unassigned, 1, 6-8 precursor ions charge states were rejected and peptide match disabled. Selected sequenced ions were dynamically excluded for 30 s. Data was acquired using Xcalibur software v4.0 (Thermo Fisher Scientific).

For DIA[85], full scan MS were performed in the m/z range of 350 to 1100 m/z with a 60,000 resolution, using an automatic gain control (AGC) of 3×10^6, maximum injection time of 50 ms and 1 microscan. MS2 was set to 15,000 resolution, 1e6 AGC target and the first fixed mass set to 120 m/z. Default charge state set to 2 and recorded in centroid mode. Total scan windows (38 windows; 49 min gradient, 63 windows, 95 min gradient) with staggered isolation window from 350 to 1100 m/z were applied with 28% normalized collision energy as described[86].

TMT labelled peptides were separated (1.9-µm particle size C18, 0.075 × 250 mm, Nikkyo Technos Co. Ltd) with a gradient of 4-28% ACN containing 0.1% FA over 110 min at 300 nL/min at 55 °C (butterfly portfolio heater, Phoenix S&T). The MS1 scan was acquired from 300 to 1650 m/z (60,000 resolution, 3e6 AGC, 128 ms injection time) followed by MS/MS data-dependent acquisition of the top 15 ions with HCD (30,000 resolution, 1e5 AGC, 60 ms injection time, 33 NCE, 0.8 m/z isolation width). Unassigned, 1, 6-8 precursor ions charge states were rejected and peptide match disabled. Selected sequenced ions were dynamically excluded for 30 s.

MS-based proteomics data (including sample/label annotation) is deposited to the ProteomeXchange Consortium via the MassIVE partner repository and available via MassIVE with identifier (MSV000094307).

Proteomic data processing

For DDA and TMT proteomics, peptide identification and quantification were performed using MaxQuant (v1.6.6.0-v1.6.14) with its built-in search engine Andromeda as described[52] against *homo sapiens* (UP000005640) including contaminants database. For LFQ-based analyses (DDA), cysteine carbamidomethylation was selected as a fixed modification and N-terminal acetylation and methionine oxidations as variable modifications. Data was processed using trypsin/P as the proteolytic enzymes with up to 2 missed cleavage sites allowed. Further processing using match between runs (MBR) and label free quantification (LFQ) algorithm were employed. Peptides were identified with an initial precursor mass deviation of up to 7 ppm and a fragment mass deviation of 20 ppm with a minimum length of 7 amino acids. False discovery rate (FDR) was 0.01 for both the protein and peptide by searching against a reverse database. For TMT-based analyses, reporter ion MS2 (TMT10plex) settings were employed. For biotin surface proteome analysis additional Thioacyl (DSP) (C(3)H(4)OS)) was employed[46].

For DIA proteomics, data processing was performed using DIA-NN[87] (v1.8). Spectral libraries were predicted using the neural network deep learning algorithm employed in DIA-NN with Trypsin/P, allowing up to 1 missed cleavage. The precursor change range was set to 1-4, and the m/z precursor range was set to 300-1800 for peptides consisting of 7-30 amino acids with N-term methionine excision and cysteine carbamidomethylation enabled as a fixed modification with 0 maximum number of variable modifications. The mass spectra were analyzed using default settings with a false discovery rate (FDR) of 1% for precursor identifications and match between runs (MBR) enabled for replicates. The resulting output files contaminants and reverse identifications were removed.

Lipid extraction and liquid chromatography mass spectrometry

Lipids were extracted from samples using a modified single-phase butanol/methanol extraction method as described[88]. In brief, the samples (2-5 µg protein from lyophilized sample reconstituted in 10ul of water) were mixed with 100 µl of 1:1 butanol:methanol containing the internal standards, the samples were vortexed, sonicated on a sonicator bath (1 h) and centrifuged (13,000 *g*, 10 min). The supernatants were transferred into glass vials and stored at −80 °C. Prior to mass spectrometry analysis, samples were thawed for 1 hr at room temperature, vortexed and sonicated on the sonicator bath for 15 min and left to sit at 20 °C for 2 h prior to analysis. Analysis of lipid extracts was performed on an Agilent 6495C QQQ mass spectrometer with an Agilent 1290 series HPLC system. Mass spectrometry settings and transitions for each lipid class are provided (Supplementary Table 24) adapted from previous methodology[89].

Mass spectrometer running conditions were gas temperature 150 °C, gas flow rate 17 l/min, nebulizer gas pressure 20 psi, sheath gas temperature 200 °C, capillary voltage 3500 V and sheath gas flow 10 l/min. Isolation widths for Q1 and Q3 were set to "unit" resolution (0.7 amu). The Agilent 1290 Infinity II HPLC conditions were as follows: The composition of running solvents A and B comprised 50:30:20 and 1:9:90 water, acetonitrile and isopropanol, respectively. Solvent A was buffered with 10mM ammonium formate with medronic acid, while Solvent B was buffered with 10mM ammonium formate. A ZORBAX eclipse plus C18 (2.1 × 100 mm 1.8 µm, Agilent) column was used. Solvents were run at a flow rate of 0.4 ml/min with the column compartment temperature set to 45 °C.

The solvent gradient was as follows: starting at 15% B and increasing to 50% B over 2.5 min (while diverting to waste for the first min), then to 57% B over 0.1 min, to 70% B over 9.4 min, to 80% B over 0.5 min, to 90% B over 3.5 min and finally to 100% over 0.1 min. The solvent was then held at 100% B for 0.9 min and returned to 15% B over 0.1 min and held for 2.9 min (a combined total of 20 min run time).

Lipid data integration and statistical analysis

Raw MS data was analyzed using MassHunter Quant 10.0 (Agilent Technologies). Peak area of the lipid species was normalized to their respective internal standards to generate relative concentration data per sample. For lipids that appear to have isomeric separation on the chromatography, they are designated with the (a) and (b) annotations to highlight different elution orders. Lipid class totals were generated by summing the individual species within each class. MS-based lipidomic sample/label annotation is deposited to the ProteomeXchange Consortium via the MassIVE partner repository and available via MassIVE with identifier (MSV000094307).

Isolation of extracellular vesicles using affinity to TIM4

To capture PS-positive EVs from plasma, human TIM4-Fc protein (FUJIFILM Wako Pure Chemical Corporation) was biotinylated with EZ-Link™ Maleimide-PEG11-Biotin (Thermo Fisher Scientific) and conjugated to Dynabeads™ MyOne™ Streptavidin C1 magnetic beads (65001, Thermo Fisher Scientific) as described[90]. Briefly, plasma (200 µL), diluted with 300 µl of PBS, was supplemented with heparin sodium (4U, Thermo Fisher Scientific) and $CaCl_2$ (final concentration of 2 mM) and incubated with TIM4 affinity beads for 16 hr at 4°C with gentle rotation. Beads were collected and washed 5x with 1 mL Tween-TBS (0.05% Tween-20, 2 ml $CaCl_2$). EVs were eluted in 50 µl PBS containing 1 mM EDTA and subjected to proteomic analyses using Q-Exactive HF-X as described above.

Data processing and statistical analysis

Software tools used for this study are freely available as open-source R packages (https://www.r-project.org). No newly generated software or custom code were used in this current study and hence the codes have not been deposited in a public repository but is available from the corresponding author(s) upon request.

For key analyses, proteome and lipidome data sets were analyzed using R package Differential Enrichment analysis of Proteomics data (DEP).[91] Proteins identified in at least 70% of one biological group were selected for downstream analysis. Using DEP, the data was background corrected and normalized by variance stabilizing transformation (vsn, which also log2-transforms the data), followed by imputation of missing values, whereby missing at random data were 'knn' imputed and missing not at random data were 'MinProb' imputed.

Protein-wise linear models combined with empirical Bayes statistics were used for the differential enrichment analysis, whereby the raw p-values were adjusted to correct for multiple testing using Benjamini-Hochberg method. Differentially abundant proteins or lipids were clustered by k-means clustering using DEP package. The PCA plot, Pearson Correlation matrix, volcano plots, log2 centred bar plots and overlap bar plots were also generated using DEP. Heatmaps were generated using ComplexHeatmap package[92]. Box plots and scatter plots were generated using RStudio package ggplot2 (https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-985X.2010.00676_9.x). Cytoscape[93] was used to generate Ontology map[94] (plugin v3.7.1). Bioconductor package clusterProfiler 4.0[95] was used to perform Ontology or KEGG pathway enrichment analysis, or gene set enrichment (KEGG) analysis, with default parameters used to identify significantly enriched gene sets.The pathway-based data integration and visualization was constructed using R package pathview.[96] For identification of lipid-associated terms enriched in lipidomes, LION web-based ontology enrichment tool[97] was used.

For conserved protein identification analysis, proteins were classified as detected or not detected across samples (protein abundance was not considered). In surface biotin-labelling of EVs, proteins detected in at least 6 out of 9 experiments were considered as pEV surface proteins (protein abundance was not considered). For annotating surface proteins into different categories, SURFY [98]-based categorical annotation of cell surface proteins were employed. To identify pEV lipid features, we performed K-means clustering of differentially abundant lipids between pEVs *versus* NonEVs and In vitro EVs *versus* cells using DEP package.

We employed caretEnsemble R package to assess the ability of these protein features (182 pEV protein features and 42 NonEV protein features, using Naïve Bayes algorithm), surface protein features (comprising of 151 surface proteins associating with pEV surface and 32 surface proteins associating with NonEVs, using Naïve Bayes algorithm) or lipid features (114 pEV lipid features and 52 NonEV lipid features, using Neural Network algorithm ('nnet')) to distinguish between distinguish between pEV and NonEV particles. The pEV and NonEV proteome/lipidome datasets were evenly partitioned based on sample type into training set (70% of the samples) and a validation set (remaining 30% of samples). For protein features based machine learning, proteomes of pEVs (n=16) and pDGS.LD (NonEV particles, n=16) from as 16 plasma samples of the EDCAD cohort served as independent test set. For lipid features based machine learning, lipidomes of DGS plasma fractions served as independent test set. A bootstrapping resampling method was employed with 25 resampling iterations to train the models. The performance of the models were evaluated using class probabilities and the two-class summary function. The preprocessing steps of centering and scaling were applied to the predictors before training the models. A confusion matrix that compare predicted classes to actual classes based on machine learning algorithm used were generated to visualize the predictive performance of the model. For identification of ADAM10 protein or PS(36:1) and CE(18:0) lipids as biomarker candidates, Recursive Feature Elimination (RFE) provided by the caret R package for feature selection using default options was used.

The Shiny web application (https://evmap.shinyapps.io/evmap/) powered by R and hosted on shinyapps.io, was created using the R packages shiny, gplots, and ComplexHeatmap, and offers feature selections and visualizations for EV protein and lipid feature conservations in circulating EVs.

Statistics and reproducibility

At least three independent biological replicates were used to generate the data. The raw p-values were adjusted to correct for multiple testing using Benjamini-Hochberg method and reported in the corresponding figure legends and supplementary tables.

Data Availability

The data supporting the findings of this study are available within the manuscript, as supplementary information files or as source data files. The raw MS files and the search/identification files obtained using MaxQuant/DIA-NN have been deposited to the ProteomeXchange Consortium via the MassIVE partner repository and available via MassIVE with identifier (MSV000094307).

Reviewer access to data repository can be made through MassIVE; MSV000094307 and reviewer portal: ftp://MSV000094307@massive.ucsd.edu

Acknowledgements

D.W.G laboratory is supported by research funds from National Health and Medical Research Council (NHMRC, MRF2015523, APP1141946), National Heart Foundation (NHF, 105072), Helen Amelia Hains Fellowship (D.W.G) and Department of Defense (PR230065). The Baker Heart & Diabetes Institute acknowledges support by the Victorian State Government Operational Infrastructure funding. We thank Eric Hanssen and the Bio21 Molecular Science and Biotechnology Institute for assisting cryo-electron microscopy (University of Melbourne). K.H is supported by NHMRC Emerging Leadership Fellowship (EL1). Q.H.P, H.F and B.C are supported through an Australian RFP Scholarship, in addition to Bright Sparks Foundation. All figures were made in BioRender.com.

Author Contribution

A.R. and D.W.G conceptualised the idea. A.R. designed the experiments and wrote the manuscript. A.R., Q.H.P., H.F., B.C., J.C., T.D., C.D., and D.W.G. performed experiments. A.R. performed bioinformatics analysis. A.R., K.H., and D.W.G., analysed the results. C.D., and T.H.M., provided resources. A.R., K.H, B.C., T.H.M., P.M., and D.W.G. authors reviewed manuscript for submission.

Competing Interests

The authors have no conflicts to declare.

References

1. Xu, R *et al.* Extracellular vesicles in cancer - implications for future improvements in cancer care. *Nat Rev Clin Oncol* **15**, 617-638 (2018).
2. van Niel, G, D'Angelo, G & Raposo, G. Shedding light on the cell biology of extracellular vesicles. *Nat Rev Mol Cell Biol* **19**, 213-228 (2018).
3. Dixson, AC, Dawson, TR, Di Vizio, D & Weaver, AM. Context-specific regulation of extracellular vesicle biogenesis and cargo selection. *Nat Rev Mol Cell Biol* **24**, 454-476 (2023).
4. Valadi, H *et al.* Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells. *Nat Cell Biol* **9**, 654-9 (2007).
5. Skog, J *et al.* Glioblastoma microvesicles transport RNA and proteins that promote tumour growth and provide diagnostic biomarkers. *Nat Cell Biol* **10**, 1470-6 (2008).
6. Peinado, H *et al.* Melanoma exosomes educate bone marrow progenitor cells toward a pro-metastatic phenotype through MET. *Nat Med* **18**, 883-91 (2012).
7. Iannotta, D, A, A, Kijas, AW, Rowan, AE & Wolfram, J. Entry and exit of extracellular vesicles to and from the blood circulation. *Nat Nanotechnol* (2023).
8. Yuana, Y *et al.* Cryo-electron microscopy of extracellular vesicles in fresh plasma. *J Extracell Vesicles* **2**(2013).
9. Arraud, N *et al.* Extracellular vesicles from blood plasma: determination of their morphology, size, phenotype and concentration. *J Thromb Haemost* **12**, 614-27 (2014).
10. Karimi, N *et al.* Detailed analysis of the plasma extracellular vesicle proteome after separation from lipoproteins. *Cell Mol Life Sci* **75**, 2873-2886 (2018).
11. Andaluz Aguilar, H, Iliuk, AB, Chen, IH & Tao, WA. Sequential phosphoproteomics and N-glycoproteomics of plasma-derived extracellular vesicles. *Nat Protoc* **15**, 161-180 (2020).
12. Dong, L *et al.* Comprehensive evaluation of methods for small extracellular vesicles separation from human plasma, urine and cell culture medium. *J Extracell Vesicles* **10**, e12044 (2020).
13. Toth, EA *et al.* Formation of a protein corona on the surface of extracellular vesicles in blood plasma. *J Extracell Vesicles* **10**, e12140 (2021).
14. Veerman, RE *et al.* Molecular evaluation of five different isolation methods for extracellular vesicles reveals different clinical applicability and subcellular origin. *J Extracell Vesicles* **10**, e12128 (2021).
15. Muraoka, S *et al.* Comprehensive proteomic profiling of plasma and serum phosphatidylserine-positive extracellular vesicles reveals tissue-specific proteins. *iScience* **25**, 104012 (2022).
16. Adamczyk, AM *et al.* Extracellular vesicles from human plasma dampen inflammation and promote tissue repair functions in macrophages. *J Extracell Vesicles* **12**, e12331 (2023).
17. Wang, J *et al.* Extracellular vesicles mediate the communication of adipose tissue with brain and promote cognitive impairment associated with insulin resistance. *Cell Metab* **34**, 1264-1279 e8 (2022).
18. Ying, W *et al.* Adipose Tissue Macrophage-Derived Exosomal miRNAs Can Modulate In Vivo and In Vitro Insulin Sensitivity. *Cell* **171**, 372-384 e12 (2017).
19. Sahu, A *et al.* Regulation of aged skeletal muscle regeneration by circulating extracellular vesicles. *Nat Aging* **1**, 1148-1161 (2021).
20. Viegas, CSB *et al.* Chronic Kidney Disease Circulating Calciprotein Particles and Extracellular Vesicles Promote Vascular Calcification: A Role for GRP (Gla-Rich Protein). *Arterioscler Thromb Vasc Biol* **38**, 575-587 (2018).
21. Safdar, A, Saleem, A & Tarnopolsky, MA. The potential of endurance exercise-derived exosomes to treat metabolic diseases. *Nat Rev Endocrinol* **12**, 504-17 (2016).
22. Deatherage, BL & Cookson, BT. Membrane vesicle release in bacteria, eukaryotes, and archaea: a conserved yet underappreciated aspect of microbial life. *Infect Immun* **80**, 1948-57 (2012).

23. Hoshino, A *et al.* Extracellular Vesicle and Particle Biomarkers Define Multiple Human Cancers. *Cell* **182**, 1044-1061 e18 (2020).
24. Zhou, R, Bozbas, E, Allen-Redpath, K & Yaqoob, P. Circulating Extracellular Vesicles Are Strongly Associated With Cardiovascular Risk Markers. *Front Cardiovasc Med* **9**, 907457 (2022).
25. Lam, SM *et al.* A multi-omics investigation of the composition and function of extracellular vesicles along the temporal trajectory of COVID-19. *Nat Metab* **3**, 909-922 (2021).
26. Shao, H *et al.* Protein typing of circulating microvesicles allows real-time monitoring of glioblastoma therapy. *Nat Med* **18**, 1835-40 (2012).
27. Welsh, JA *et al.* Minimal information for studies of extracellular vesicles (MISEV2023): From basic to advanced approaches. *J Extracell Vesicles* **13**, e12404 (2024).
28. Kim, MS *et al.* A draft map of the human proteome. *Nature* **509**, 575-81 (2014).
29. Kugeratski, FG *et al.* Quantitative proteomics identifies the core proteome of exosomes with syntenin-1 as the highest abundant protein and a putative universal biomarker. *Nat Cell Biol* **23**, 631-641 (2021).
30. Jeppesen, DK *et al.* Reassessment of Exosome Composition. *Cell* **177**, 428-445 e18 (2019).
31. Takov, K, Yellon, DM & Davidson, SM. Comparison of small extracellular vesicles isolated from plasma by ultracentrifugation or size-exclusion chromatography: yield, purity and functional potential. *J Extracell Vesicles* **8**, 1560809 (2019).
32. Baranyai, T *et al.* Isolation of Exosomes from Blood Plasma: Qualitative and Quantitative Comparison of Ultracentrifugation and Size Exclusion Chromatography Methods. *PLoS One* **10**, e0145686 (2015).
33. Sodar, BW *et al.* Low-density lipoprotein mimics blood plasma-derived exosomes and microvesicles during isolation and detection. *Sci Rep* **6**, 24316 (2016).
34. Yuana, Y, Levels, J, Grootemaat, A, Sturk, A & Nieuwland, R. Co-isolation of extracellular vesicles and high-density lipoproteins using density gradient ultracentrifugation. *J Extracell Vesicles* **3**(2014).
35. Thery, C *et al.* Minimal information for studies of extracellular vesicles 2018 (MISEV2018): a position statement of the International Society for Extracellular Vesicles and update of the MISEV2014 guidelines. *J Extracell Vesicles* **7**, 1535750 (2018).
36. Vergauwen, G *et al.* Robust sequential biophysical fractionation of blood plasma to study variations in the biomolecular landscape of systemically circulating extracellular vesicles across clinical conditions. *J Extracell Vesicles* **10**, e12122 (2021).
37. Thippabhotla, S, Zhong, C & He, M. 3D cell culture stimulates the secretion of in vivo like extracellular vesicles. *Sci Rep* **9**, 13012 (2019).
38. Rocha, S *et al.* 3D Cellular Architecture Affects MicroRNA and Protein Cargo of Extracellular Vesicles. *Adv Sci (Weinh)* **6**, 1800948 (2019).
39. Liangsupree, T, Multia, E & Riekkola, M-L. Modern isolation and separation techniques for extracellular vesicles. *Journal of Chromatography A* **1636**, 461773 (2021).
40. Lötvall, J *et al.* Minimal experimental requirements for definition of extracellular vesicles and their functions: a position statement from the International Society for Extracellular Vesicles. *J Extracell Vesicles* **3**, 26913 (2014).
41. Kashkanova, AD *et al.* Label-free discrimination of extracellular vesicles from large lipoproteins. *J Extracell Vesicles* **12**, e12348 (2023).
42. Geekiyanage, H, Rayatpisheh, S, Wohlschlegel, JA, Brown, R, Jr. & Ambros, V. Extracellular microRNAs in human circulation are associated with miRISC complexes that are accessible to anti-AGO2 antibody and can bind target mimic oligonucleotides. *Proc Natl Acad Sci U S A* **117**, 24213-24223 (2020).
43. Rai, A *et al.* Exosomes Derived from Human Primary and Metastatic Colorectal Cancer Cells Contribute to Functional Heterogeneity of Activated Fibroblasts by Reprogramming Their Proteome. *Proteomics* **19**, e1800148 (2019).
44. Rai, A *et al.* Secreted midbody remnants are a class of extracellular vesicles molecularly distinct from exosomes and microparticles. *Commun Biol* **4**, 400 (2021).

45. Rai, A, Greening, DW, Xu, R, Suwakulsiri, W & Simpson, RJ. Exosomes Derived from the Human Primary Colorectal Cancer Cell Line SW480 Orchestrate Fibroblast-Led Cancer Invasion. *Proteomics* **20**, e2000016 (2020).
46. Xu, R, Greening, DW, Rai, A, Ji, H & Simpson, RJ. Highly-purified exosomes and shed microvesicles isolated from the human colon cancer cell line LIM1863 by sequential centrifugal ultrafiltration are biochemically and functionally distinct. *Methods* **87**, 11-25 (2015).
47. Midekessa, G *et al.* Zeta Potential of Extracellular Vesicles: Toward Understanding the Attributes that Determine Colloidal Stability. *ACS Omega* **5**, 16701-16710 (2020).
48. Farrah, T *et al.* A high-confidence human plasma proteome reference set with estimated concentrations in PeptideAtlas. *Mol Cell Proteomics* **10**, M110 006353 (2011).
49. Kowal, J *et al.* Proteomic comparison defines novel markers to characterize heterogeneous populations of extracellular vesicle subtypes. *Proc Natl Acad Sci U S A* **113**, E968-77 (2016).
50. Minciacchi, VR *et al.* Large oncosomes contain distinct protein cargo and represent a separate functional class of tumor-derived extracellular vesicles. *Oncotarget* **6**, 11327-41 (2015).
51. Anderson, NL & Anderson, NG. The human plasma proteome: history, character, and diagnostic prospects. *Mol Cell Proteomics* **1**, 845-67 (2002).
52. Rai, A, Fang, H, Claridge, B, Simpson, RJ & Greening, DW. Proteomic dissection of large extracellular vesicle surfaceome unravels interactive surface platform. *Journal of Extracellular Vesicles* **10**, e12164 (2021).
53. Nunez Lopez, YO *et al.* Proteomics and Phosphoproteomics of Circulating Extracellular Vesicles Provide New Insights into Diabetes Pathobiology. *Int J Mol Sci* **23**(2022).
54. Karimi, N, Dalirfardouei, R, Dias, T, Lotvall, J & Lasser, C. Tetraspanins distinguish separate extracellular vesicle subpopulations in human serum and plasma - Contributions of platelet extracellular vesicles in plasma samples. *J Extracell Vesicles* **11**, e12213 (2022).
55. Harayama, T & Riezman, H. Understanding the diversity of membrane lipid composition. *Nat Rev Mol Cell Biol* **19**, 281-296 (2018).
56. Skotland, T, Sagini, K, Sandvig, K & Llorente, A. An emerging focus on lipids in extracellular vesicles. *Adv Drug Deliv Rev* **159**, 308-321 (2020).
57. Trajkovic, K *et al.* Ceramide triggers budding of exosome vesicles into multivesicular endosomes. *Science* **319**, 1244-7 (2008).
58. Stocker, R, Bowry, VW & Frei, B. Ubiquinol-10 protects human low density lipoprotein more efficiently against lipid peroxidation than does alpha-tocopherol. *Proc Natl Acad Sci U S A* **88**, 1646-50 (1991).
59. Jha, P *et al.* Genetic Regulation of Plasma Lipid Species and Their Association with Metabolic Phenotypes. *Cell Syst* **6**, 709-721 e6 (2018).
60. Stafford, JH & Thorpe, PE. Increased exposure of phosphatidylethanolamine on the surface of tumor vascular endothelium. *Neoplasia* **13**, 299-308 (2011).
61. Leite, NB *et al.* PE and PS Lipids Synergistically Enhance Membrane Poration by a Peptide with Anticancer Properties. *Biophys J* **109**, 936-47 (2015).
62. Dinkins, MB, Dasgupta, S, Wang, G, Zhu, G & Bieberich, E. Exosome reduction in vivo is associated with lower amyloid plaque load in the 5XFAD mouse model of Alzheimer's disease. *Neurobiol Aging* **35**, 1792-800 (2014).
63. Sezgin, E, Levental, I, Mayor, S & Eggeling, C. The mystery of membrane organization: composition, regulation and roles of lipid rafts. *Nat Rev Mol Cell Biol* **18**, 361-374 (2017).
64. Nakai, W *et al.* A novel affinity-based method for the isolation of highly purified extracellular vesicles. *Sci Rep* **6**, 33935 (2016).
65. Mu, W, Rana, S & Zoller, M. Host matrix modulation by tumor exosomes promotes motility and invasiveness. *Neoplasia* **15**, 875-87 (2013).
66. Lespagnol, A *et al.* Exosome secretion, including the DNA damage-induced p53-dependent secretory pathway, is severely compromised in TSAP6/Steap3-null mice. *Cell Death Differ* **15**, 1723-33 (2008).
67. Cvjetkovic, A *et al.* Detailed Analysis of Protein Topology of Extracellular Vesicles-Evidence of Unconventional Membrane Protein Orientation. *Sci Rep* **6**, 36338 (2016).

68. Babuke, T & Tikkanen, R. Dissecting the molecular function of reggie/flotillin proteins. *Eur J Cell Biol* **86**, 525-32 (2007).
69. Skryabin, GO, Komelkov, AV, Savelyeva, EE & Tchevkina, EM. Lipid Rafts in Exosome Biogenesis. *Biochemistry (Mosc)* **85**, 177-191 (2020).
70. Kajimoto, T, Okada, T, Miya, S, Zhang, L & Nakamura, S. Ongoing activation of sphingosine 1-phosphate receptors mediates maturation of exosomal multivesicular endosomes. *Nat Commun* **4**, 2712 (2013).
71. Baietti, MF *et al.* Syndecan-syntenin-ALIX regulates the biogenesis of exosomes. *Nat Cell Biol* **14**, 677-85 (2012).
72. Ji, H *et al.* Deep sequencing of RNA from three different extracellular vesicle (EV) subtypes released from the human LIM1863 colon cancer cell line uncovers distinct miRNA-enrichment signatures. *PLoS One* **9**, e110314 (2014).
73. Tauro, BJ *et al.* Two distinct populations of exosomes are released from LIM1863 colon carcinoma cell-derived organoids. *Mol Cell Proteomics* **12**, 587-98 (2013).
74. Tkach, M, Kowal, J & Thery, C. Why the need and how to approach the functional diversity of extracellular vesicles. *Philos Trans R Soc Lond B Biol Sci* **373**(2018).
75. Al-Nedawi, K *et al.* Intercellular transfer of the oncogenic receptor EGFRvIII by microvesicles derived from tumour cells. *Nat Cell Biol* **10**, 619-24 (2008).
76. Wiklander, OP *et al.* Extracellular vesicle in vivo biodistribution is determined by cell source, route of administration and targeting. *J Extracell Vesicles* **4**, 26316 (2015).
77. Whitehead, RH, Jones, JK, Gabriel, A & Lukies, RE. A new colon carcinoma cell line (LIM1863) that grows as organoids with spontaneous differentiation into crypt-like structures in vitro. *Cancer Res* **47**, 2683-9 (1987).
78. Lozano, J *et al.* Scalable Generation of Nanovesicles from Human-Induced Pluripotent Stem Cells for Cardiac Repair. *Int J Mol Sci* **23**(2022).
79. Hughes, CS *et al.* Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. *Nat Protoc* **14**, 68-85 (2019).
80. Fatmous, M, Rai, A, Poh, QH, Salamonsen, LA & Greening, DW. Endometrial small extracellular vesicles regulate human trophectodermal cell invasion by reprogramming the phosphoproteome landscape. *Front Cell Dev Biol* **10**, 1078096 (2022).
81. Notaras, M *et al.* Neurodevelopmental signatures of narcotic and neuropsychiatric risk factors in 3D human-derived forebrain organoids. *Mol Psychiatry* **26**, 7760-7783 (2021).
82. Ruprecht, B, Zecha, J, Zolg, DP & Kuster, B. High pH Reversed-Phase Micro-Columns for Simple, Sensitive, and Efficient Fractionation of Proteome and (TMT labeled) Phosphoproteome Digests. *Methods Mol Biol* **1550**, 83-98 (2017).
83. Kompa, AR *et al.* Sustained subcutaneous delivery of secretome of human cardiac stem cells promotes cardiac repair following myocardial infarction. *Cardiovasc Res* **117**, 918-929 (2021).
84. Greening, DW *et al.* Chronic methamphetamine interacts with BDNF Val66Met to remodel psychosis pathways in the mesocorticolimbic proteome. *Mol Psychiatry*, In Press (2019).
85. Cross, J, Rai, A, Fang, H, Claridge, B & Greening, DW. Rapid and in-depth proteomic profiling of small extracellular vesicles for ultralow samples. *Proteomics*, e2300211 (2023).
86. Fang, H & Greening, DW. An Optimized Data-Independent Acquisition Strategy for Comprehensive Analysis of Human Plasma Proteome. *Methods Mol Biol* **2628**, 93-107 (2023).
87. Demichev, V, Messner, CB, Vernardis, SI, Lilley, KS & Ralser, M. DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. *Nat Methods* **17**, 41-44 (2020).
88. Huynh, K *et al.* Concordant peripheral lipidome signatures in two large clinical studies of Alzheimer's disease. *Nat Commun* **11**, 5698 (2020).
89. Huynh, K *et al.* High-Throughput Plasma Lipidomics: Detailed Mapping of the Associations with Cardiometabolic Risk Factors. *Cell Chem Biol* **26**, 71-84 e4 (2019).
90. Yoshida, T, Ishidome, T & Hanayama, R. High Purity Isolation and Sensitive Quantification of Extracellular Vesicles Using Affinity to TIM4. *Curr Protoc Cell Biol* **77**, 3 45 1-3 45 18 (2017).

91. Zhang, X *et al.* Proteome-wide identification of ubiquitin interactions using UbIA-MS. *Nat Protoc* **13**, 530-550 (2018).
92. Gu, Z, Eils, R & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847-9 (2016).
93. Shannon, P *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**, 2498-504 (2003).
94. Merico, D, Isserlin, R, Stueker, O, Emili, A & Bader, GD. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. *PLoS One* **5**, e13984 (2010).
95. Wu, T *et al.* clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation (N Y)* **2**, 100141 (2021).
96. Luo, W & Brouwer, C. Pathview: an R/Bioconductor package for pathway-based data integration and visualization. *Bioinformatics* **29**, 1830-1 (2013).
97. Molenaar, MR *et al.* LION/web: a web-based ontology enrichment tool for lipidomic data analysis. *Gigascience* **8**(2019).
98. Bausch-Fluck, D *et al.* The in silico human surfaceome. *Proc Natl Acad Sci U S A* **115**, E10988-E10997 (2018).

Figure Legends

Figure 1. Isolation and characterization of EVs from human plasma. A. Workflow for density gradient separation (DGS) of human plasma (0.5 ml) for EV isolation and characterization. **B.** Western blot analysis of twelve DGS fractions (vol:vol matched) of human plasma with antibodies against indicated proteins. **C.** Cryogenic electron microscopic images of p100K, DGS fractions (1-3) and (6-8). Scale bar, 500 nm (n=3). **D.** Size distribution (particle diameter (nm)) of plasma DGS fractions (1-3) and (6-8) (n=7) based on nanoparticle tracking analysis. Small EVs released by SW620 cells (In vitro EVs (n=7)) were also analyzed. **E.** The zeta potential of DGS fractions (1-3) and (6-8) at 11 positions throughout the sample cell (top panel), or stationary layers in the sample cell (bottom panel) (n=5). **F.** Western blot analysis of unprocessed plasma, p100K and DGS fractions 6-8 (pooled). Protein load per lane as indicated. **G-H.** Western blot analysis of indicated DGS fractions from 6 different pooled plasma samples (n=6 independent plasma samples).

Figure 2. Mapping the core proteome of circulating EVs in humans. A. Proteome landscape construction of plasma EVs (pEVs, n=38) and NonEVs (n=42). **B.** Principal component analysis of quantified proteins. **C**. Occurrence analysis of proteins in 38 pEV proteomes where Category 1 proteins represent ubiquitously quantified proteins. Gene Ontology pathways enriched in each category proteins are indicated. **D.** Scatter plot representing differential abundance ($p<0.05$) of Category 1-3 proteins in pEVs compared to NonEVs. **E.** KEGG pathways enriched ($p<0.05$) in pEV or NonEV protein features. **F.** Protein–protein interaction networks of selected EV protein features (quantified in <30% of NonEV datasets) analysed using STRING-DB and Cytoscape; each node represents a protein. Functional enrichment (GO cellular component) was performed using Cytoscape. Nodes with red border are enriched for "Extracellular exosome" term. **G.** Surface-accessible Category 1 proteins of pEVs categorised based on their molecular/functional annotation. C/R/T represents cluster of differentiation (CDs), receptors and transporters. Heatmap for abundance of indicated proteins in pEVs (n=9) and In vitro EVs (n=3).

Figure 3. Conservation of circulating EV protein features. A. Heatmap depicts quantification of EV protein features in our pEV and NonEV proteome datasets (only protein features quantified in <30% of NonEV datasets were used). **B.** Heatmap depicts quantification of EV protein features in previous reported proteomes for circulating EVs: (A-C) EV preparations from plasma or serum sourced from healthy individuals or those with known pathology[15,53], and (D-F) EV subtypes (CD81+, CD63+, and CD9+ EVs) prevalent in the circulatory system[54]. As an orthogonal validation, heatmap also depicts conservation of EV protein features (G-J) using TMT-based isobaric multiplexing of pEVs (n=4) and In vitro EV proteomes (n=6), and (I-L) using label-free quantitative proteomics for In vitro EVs (n=3 per cell line).

Figure 4. Lipidome landscape of circulating EVs. A. Workflow for lipidomic analysis, interrogating 829 lipids representing 40 lipid classes within three major groups. **B.** Principal-component analysis of lipidome data showing group clustering and separation of lipidomes (pEVs; n=12, p100K; n= 6, pDGS.LD; n=5, In vitro EVs; n=12, cells; n=12). **C.** Scatter plot depicts abundance of 829 lipids in pEVs and In vitro EVs with lipids belonging to cluster 3 displaying high abundance in both pEVs and In vitro EVs. **D.** Scatter plot showing relative abundance of lipids in pEVs vs p100K/pDGS.LD and In vitro EVs vs cell lipidome data sets. Blue circles (lipid markers) represent lipids with significantly greater abundance (fold change >1.5 & $p<0.05$). Red circles (exclusion lipids) represent lipids with significantly lower

abundance (fold change <-1.5 & p<0.05). Grey circles represent lipids that do not meet the above criteria. Pie chart depicts number of lipids (%) with differential abundance between the data sets.

Figure 5. Conserved lipid features of circulating EVs in humans. A. Heatmap depicting K-means clustering of differentially abundant lipids from Figure 4D. **B.** Network map of lipids grouped based on lipid classes. Blue circles represent EV-associated lipids clusters c2, c3 and c4, whereas red circles represent NonEV-associated lipid cluster c1. Grey circles are lipids with similar abundance between EVs and NonEVs. **C.** LION Ontologies enriched in EV lipid features (cluster c4 lipids) or NonEV lipid features (cluster c1 lipids). **D.** Density gradient separation of plasma. The twelve fractions were subjected to lipidomics analysis. **E.** Principal component analysis of lipidomes of 12 fractions (n=3). **F.** Box plots depict abundance of cluster 1-4 lipids for DGS fractions. Y axis represents Z-scored abundance (MS-based abundance for each lipid - mean abundance) / standard deviation (Z score normalization). Grey lines mark either fractions 6-7 (corresponding to pEV fractions) or fractions 1-3 (corresponding to pDGS.LD fractions). Lipids from cluster c4 and cluster c1 are depicted.

Figure 6. ADAM10 protein and phosphatidylserine lipid are biological markers for circulating EVs in humans. A. Heatmap depicting pEV protein features that are uniquely identified in pEVs compared to NonEVs and their conservation, in particular for ADAM10, in previously published circulating EV proteomes[15,53] (A-C) and (D-F) EV subtypes (CD81+, CD63+, and CD9+ EVs) in plasma[54], and in our TMT-based isobaric multiplexing of pEVs (n=4) and In vitro EVs proteomes (n=6) as well as label-free proteomes of (I-L) small EVs from 4 cell lines (n=3 per cell line). Ratio of ADAM10 intensity and corresponding sample median intensity for each proteome is depicted in **B** as a boxplot. **C** Box plot depicting normalised intensities (abundance log2-tranformed) and raw intensities of PS(36:1) and CE(18:0) in indicated lipidome datasets. **D.** Abundance of PS(36:1) and CE(18:0) in lipidome datasets across 12 DGS fractions (n=3). Grey points represent rest of the quantified lipids. **E.** Top panel: Relative abundance of PS(36:1) and CE(18:0) in lipidome datasets across 12 DGS fractions (n=3). Middle panel: Median of lipidome datasets across 12 DGS fractions. Bottom panel: Western blot analysis of 12 fractions for indicated proteins. **F.** Confusion matrix using Neural Network algorithm ('nnet') classifier of training set (70%) and validation set (30%) using relative abundance of PS(36:1) versus CE(18:0). **G.** Workflow for capturing PS+ EVs from human plasma using TIM4-magnetic bead conjugate. Captured EVs (n=23), along with mock capture using beads alone (n=3) or unprocessed plasma (n=3), were subjected to proteomics analysis. Heatmap depicts absolute and exclusive quantification of ADAM10 protein in captured EVs.

Figure 7. Hallmark molecular features of circulating EVs. A. Resolving the protein and lipid componentry intrinsic to circulating EVs. These features are also conserved in EVs released by cultured cells. We submit EV features, namely ADAM10 and phosphatidyl serine PS(36:1), as biological marker for confident EV identification and potentially purity assessment. The freely accessibly Shiny webtool will facilitate easy interrogation (conservation, abundance report, surface localization, and enrichments amongst others) of the multiomics data to encourage reuse.

Extended Data Legends

Extended data 1. Characterization of conserved proteins in pEVs. **A**. Gene Ontology pathways and KEGG pathways enriched ($p<0.05$) in each cluster. **B.** KEGG endocytosis pathway graph rendered by Pathview where category 1 proteins are indicated as high and category2-3 proteins are indicated as low. **C.** Protein–protein interaction networks of cluster 1 proteins analysed using STRING-DB and Cytoscape. Each node represents a protein. Nodes with red border are enriched for "vesicular pathway".

Extended data 2. Heatmap of pEV protein features (182 proteins)

Extended data 3. Mapping the core proteome of NonEVs in humans. A. Occurrence analysis of proteins in 38 pEV proteomes in our data set where Category 1 proteins are ubiquitously quantified. **B.** Heatmap of conserved protein features in NonEVs.

Extended data 4. Heatmap of EV cluster 1 and cluster 4 lipids in EVs and NonEVs.

Supplementary Figure Legends

Supplementary Figure 1. Characterization of EVs from human plasma. **A**. Ultracentrifugation of human plasma. The protein yields of plasma and isolated EV pellets are indicated. Western blot analysis of the crude EV pellets for indicated proteins. **B.** Cryo EM image-based size distribution of plasma EVs and previously reported sizes for small (s-EV) and large EVs (L-EVs)[52]. **C**. Protein yield (µg) of pEVs per ml of plasma. **D.** Nanoparticle tracking analysis based total particle counts in DGS fractions (1-3) and DGS fractions (6-8) fractions/pEVs per ml of plasma (n=7). **E**. Comparative summary of biophysical characterization of pEVs and NonEVs.

Supplementary Figure 2. Construction and characterization of circulating EV proteome landscape. **A**. Workflow depicting sample details and MS-acquisition modes for constructing proteome landscape of plasma EVs (pEVs). Different experimental groups are indicated as Set, totalling six sets. **B**. Classification of different proteomes (from sets 1-5) into pEVs or nonEV groups. Bar plot depicts number of proteins identified within each proteome. Box plot represents normalised protein abundance (log10 intensity). **C.** Venn diagram of proteins identified indicated proteomes. **D.** Protein ranked by their abundance within indicated proteomes in Set 1. Proteins typical reported as EVs proteins are highlighted. **E.** Correlation of proteins in plasma, pDGS.LD and pEVs proteomes (median protein intensities of replicates) to the published concentration of the same proteins[48]. The black lines are linear regression models (the grey shaded regions represent 95% confidence interval). **F.** Heatmap of Pearson correlation matrix of protein abundance between proteomes.

Supplementary Figure 3. Classical hallmark features of EVs preserved in circulating EV proteome. **A.** Heatmap of differentially abundant proteins ($p<0.05$, log2 FC >1.5) with k-means clustering. **B.** KEGG pathways and Gene Ontology pathways enriched in each cluster ($p<0.05$) **C.** Leading edge analysis indicated for selected GSEA-KEGG pathways enriched in pEVs vs NonEVs. **D.** Bar plots showing relative abundance (log2 centred MS-based intensities) of differentially abundant EV and abundant plasma proteins.

Supplementary Figure 4. **Mapping EV markers in circulating EV proteome**. **A.** Pair-wise comparison of differentially abundant proteins ($p<0.05$, FC > 1.5) depicted as a volcano plot where 22 core proteins established for In vitro EVs[29] and ninety-four EV marker proteins recommended by MISEV[35] guidelines are depicted. **B**. Heatmap showing quantification of current EV marker proteins in pEVs. **C**. Occurrence analysis for current EV marker proteins in 38 pEV proteomes. **D**. Heatmap of selected EV proteins depicting their identification and abundance in 38 pEV proteomes.

Supplementary Figure 5. **Surfaceome of circulating EVs in human plasma. A.** Workflow for capture and identification of EV surface proteins. **B.** Categorization of pEV surface proteins and their relative abundance in pEVs vs pDGS.LD proteomes. **C.** Surface-accessible Category 1-3 proteins of pEVs categorised based on their molecular/functional annotation. C/R/T represents cluster of differentiation (CDs), receptors and transporters.

Supplementary Figure 6. **Conservation of circulating EV protein features in cell culture-derived EVs published proteomes. A.** Relative abundance of EV protein features in sEVs vs their parental cell proteome as reported[29], whereby EVs were isolated from culture media using ultracentrifugation (UC). **B.** Relative abundance of EV protein features in sEVs vs their

parental cell proteome as reported[29], where EVs were isolated using density gradient separation (DG), size-exclusion chromatography (SEC), and UC.

Supplementary Figure 7. Lipidome profiling of circulating EVs. A. Lipid abundance (MS-based intensity) for In vitro EVs, pEVs, p100K, pDGS.LD, and cells (WCL). **B.** Heatmap of Pearson correlation matrix of lipid abundance between datasets.

Supplementary Figure 8. Lipid species co-enriched in circulating EVs and EVs from cell culture. A. Scatter plot showing relative abundance of lipids in pEVs vs p100K and pEVs vs pDGS.LD. Lipids in blue are enriched in pEVs. **B.** Heatmap depicting differentially abundant lipids in pEVs vs NonEVs. **C.** Scatter plot showing relative abundance of lipids in pEVs vs p100K/pDGS.LD and In vitro EVs vs WCL lipidome data sets. Blue circles (lipid markers) represent lipids with significantly greater abundance (Log2 FC > 0.5 & p value < 0.05) in pEVs vs pDGS.LD and In vitro EVs vs cells. Red circles (exclusion lipids) represent lipids with significantly lower abundance (Log2 FC < -0.5 & p value < 0.05) in pEVs vs pDGS.LD and In vitro EVs vs cells. Grey circles represent lipids that do not meet the above criteria. **D-E.** Bar plots showing relative abundance (log2 centred intensities) of differentially abundant ($p<0.05$) lipids in indicated lipidomes.

Supplementary Figure 9. PS lipid class is enriched in EVs in plasma. A Lipid abundance (normalized) across 12 fractions of density gradient separated plasma. **B.** Box plots depict abundance of cluster 1-4 lipids for DGS fractions for each biological replicate (R1-3). Y axis represents Z-scored abundance (MS-based abundance for each lipid - mean abundance) / standard deviation (Z score normalization). Blue lines highlights fractions 6-7 (corresponding to pEV fractions), whereas red line depicts fractions 1-3 (corresponding to pDGS.LD fractions).

Supplementary Figure 10. Machine learning classification of EV and NonEV particles in plasma using protein features. A. Confusion matrix (using the ensemble model) of training set (70%), remaining validation set (30%), and independent test set, for cluster 1 proteins, which include top 97 proteins enriched in EVs (FC>5) and 28 exclusion proteins. Probability scores for sample classification into EVs using ensemble model on independent test set. Greater scores indicate higher confidence in predicting sample belonging to EV class. **B.** Confusion matrix using nnet classifier of validation set (30%) and independent test set, for 151 pEV surface proteins. Bar plot show probability scores for sample classification into EVs using nnet model on independent test set. **C.** Confusion matrix using nnet classifier of validation set (30%) and independent test set, for ADAM10 vs sample median intensities for each proteome as depicted in Figure 6B.

Supplementary Figure 11. Machine learning classification of EV and NonEV particles in plasma using lipid features. A Confusion matrix using Neural Network algorithm ('nnet') classifier of training set (70%), validation set (30%) and independent test set, for cluster c1 and c4 lipids (upper panel), and top 20 c1 and c4 lipid features following reclusive feature elimination (lower panel). **B.** Bar plot indicated probability scores for DGS fraction lipidomes classification into EVs using 'nnet' model based on top 20 c1 and c4 lipid features on independent test set. Greater scores indicate higher confidence in predicting DGS fractions belonging to EV class.

Supplementary Figure 12. **Interrogating protein of interest expression in circulating EVs using the Shiny app.**

Supplementary Figure 13. **Interrogating lipid of interest expression in circulating EVs using the Shiny app.**

Supplementary Figure 14. Uncropped Western blot images.

References

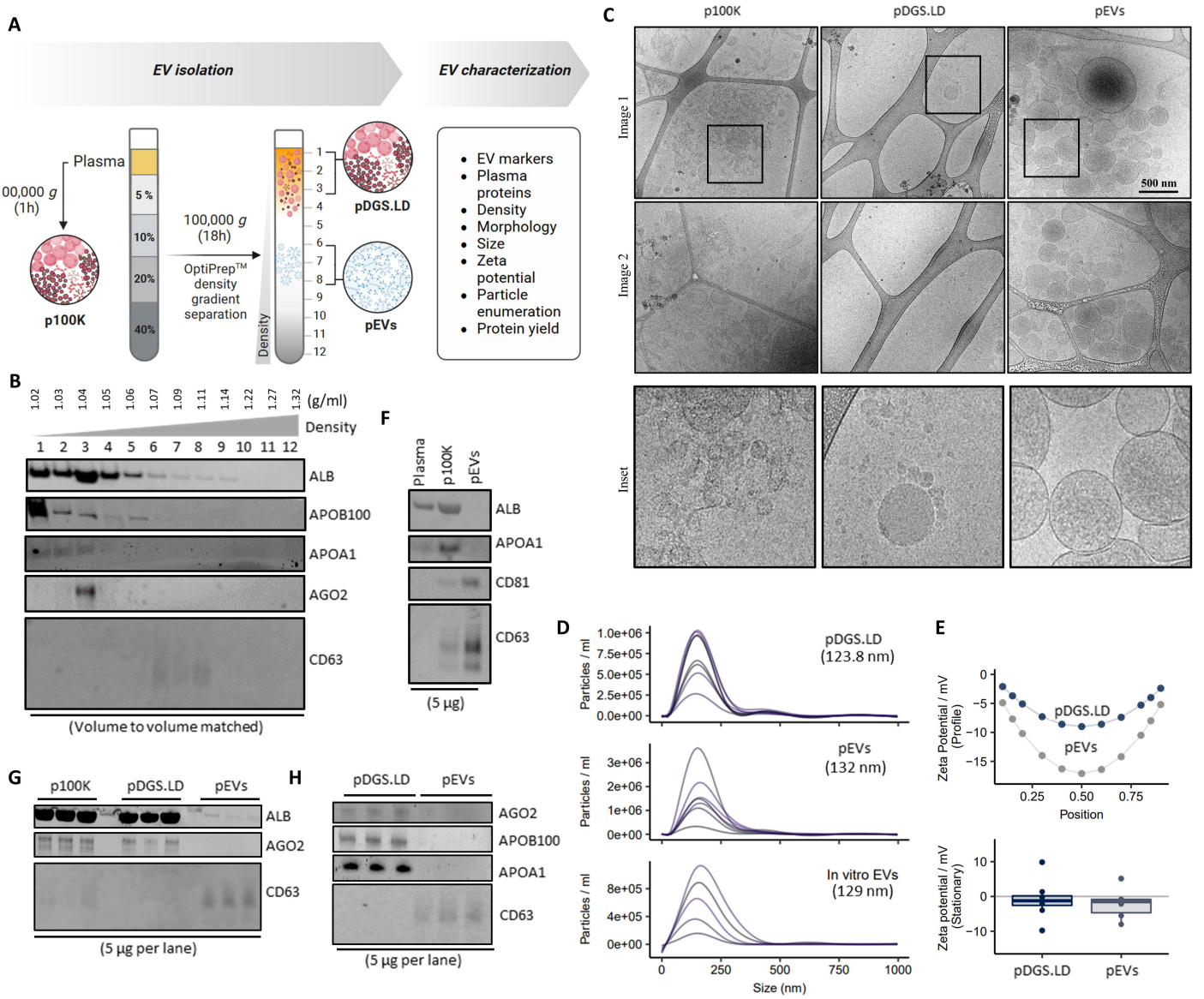

Figure 1. Isolation and characterization of EVs from human plasma

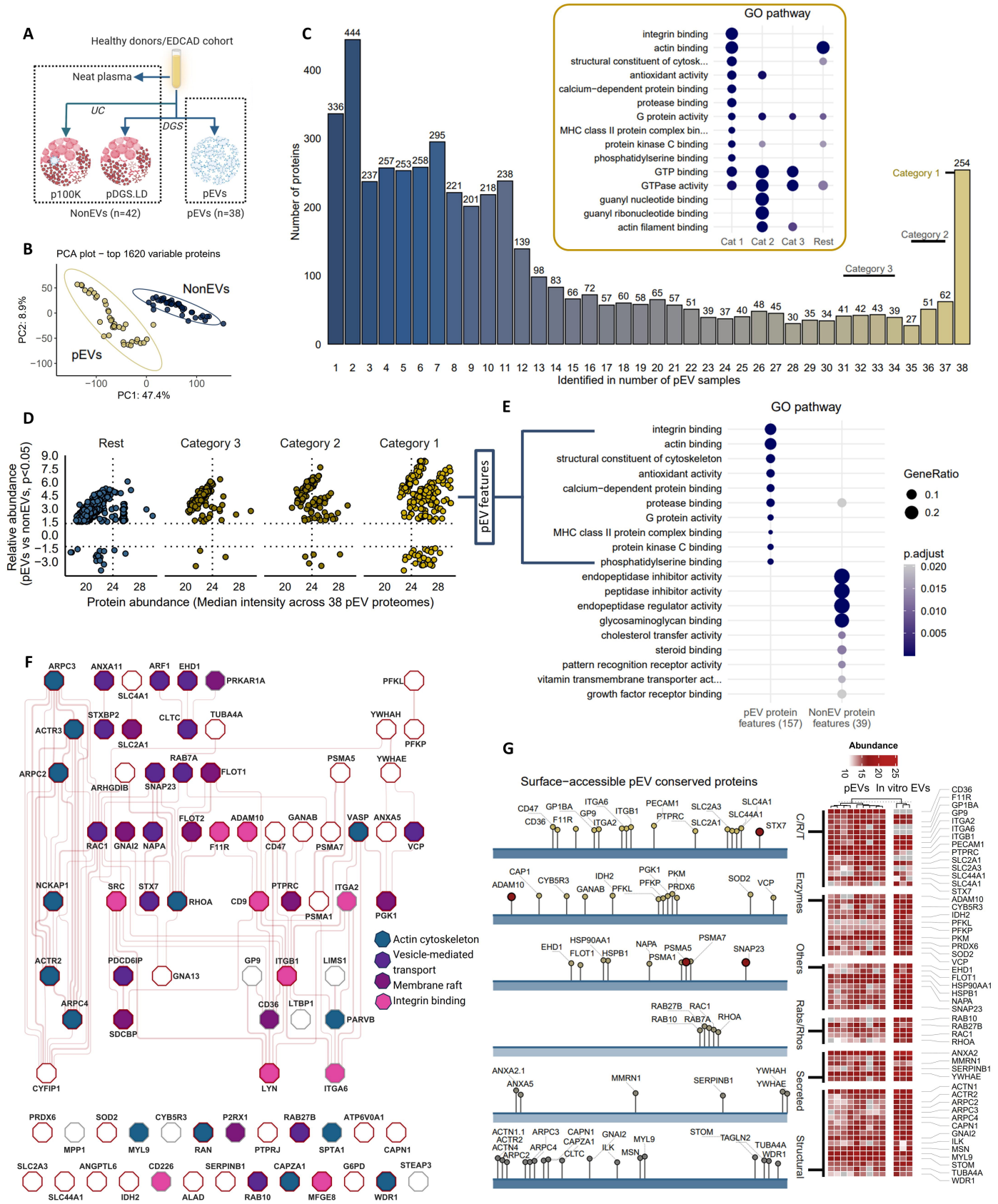

Figure 2. Mapping the core proteome of circulating EVs in humans

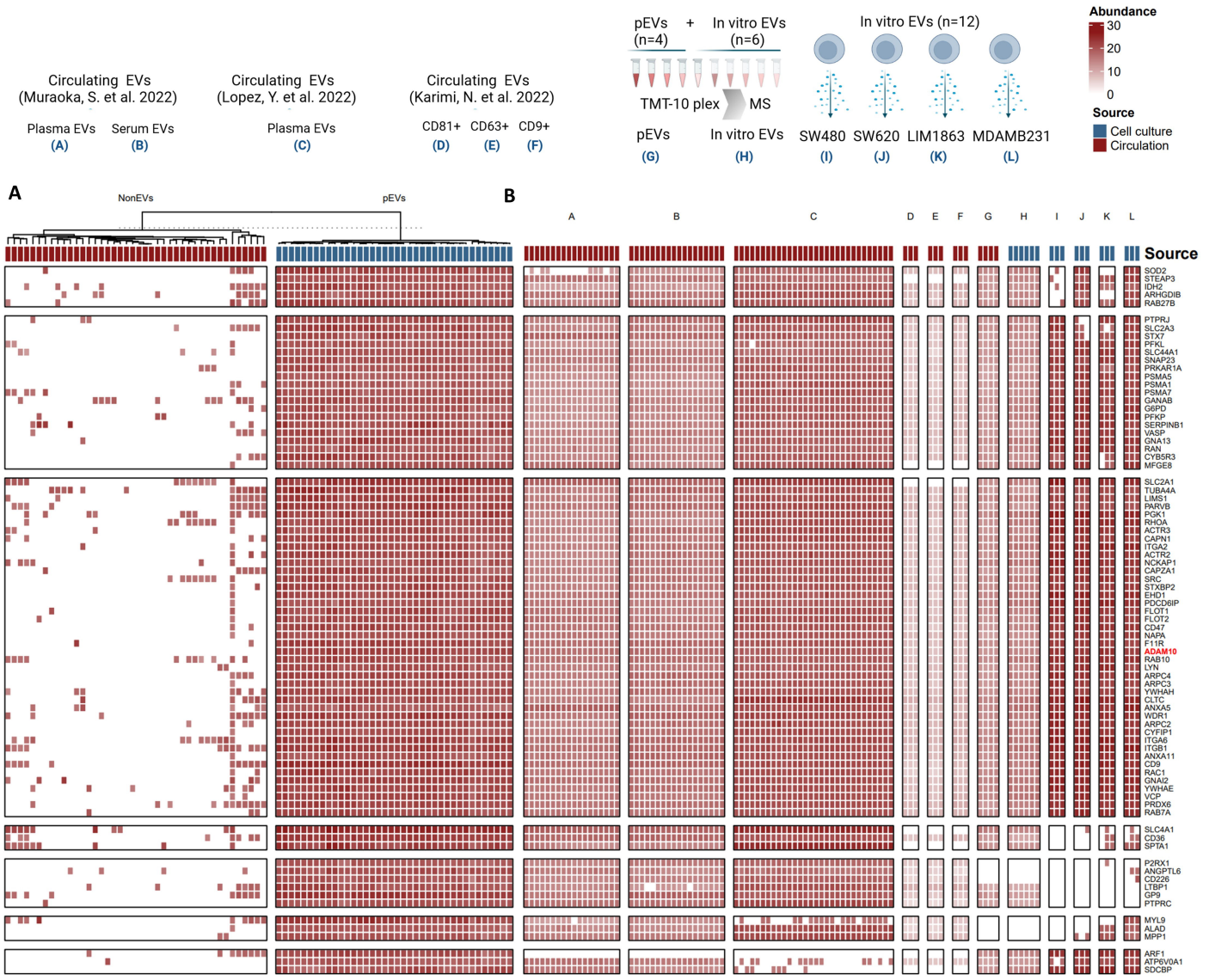

Figure 3. Conservation of circulating EV protein features

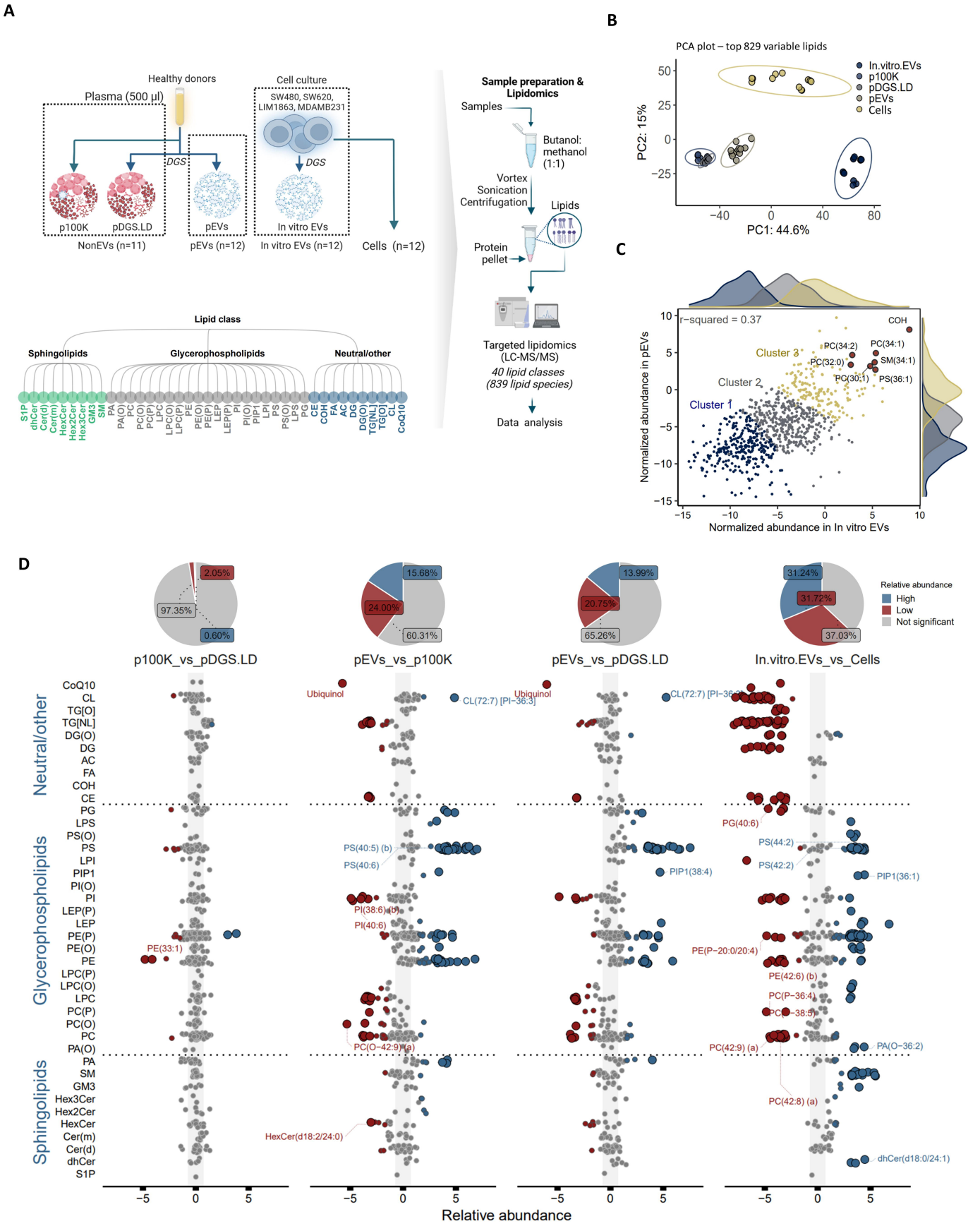

Figure 4. Lipidome landscape of circulating EVs

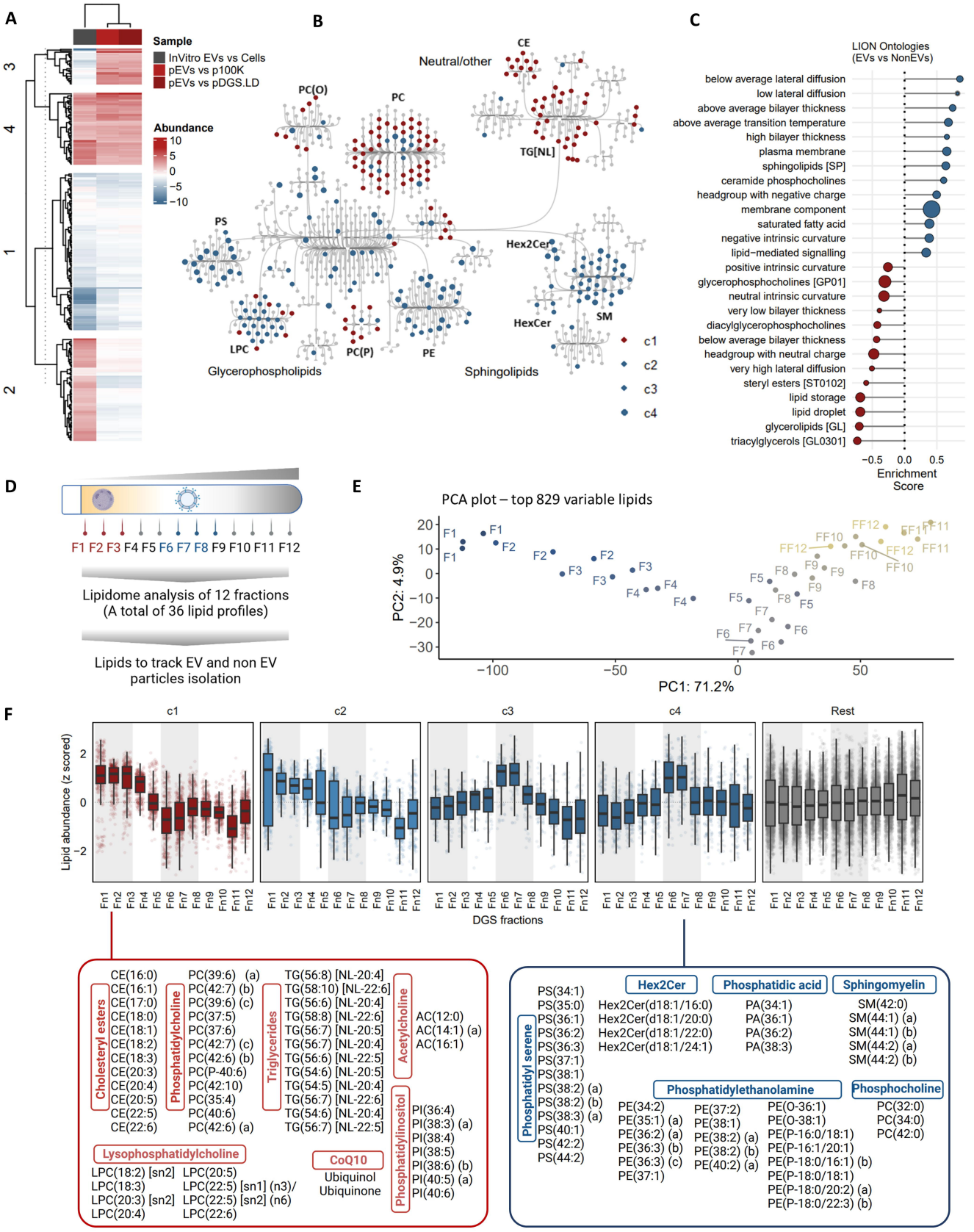

Figure 5. Conserved lipid features of circulating EVs in humans

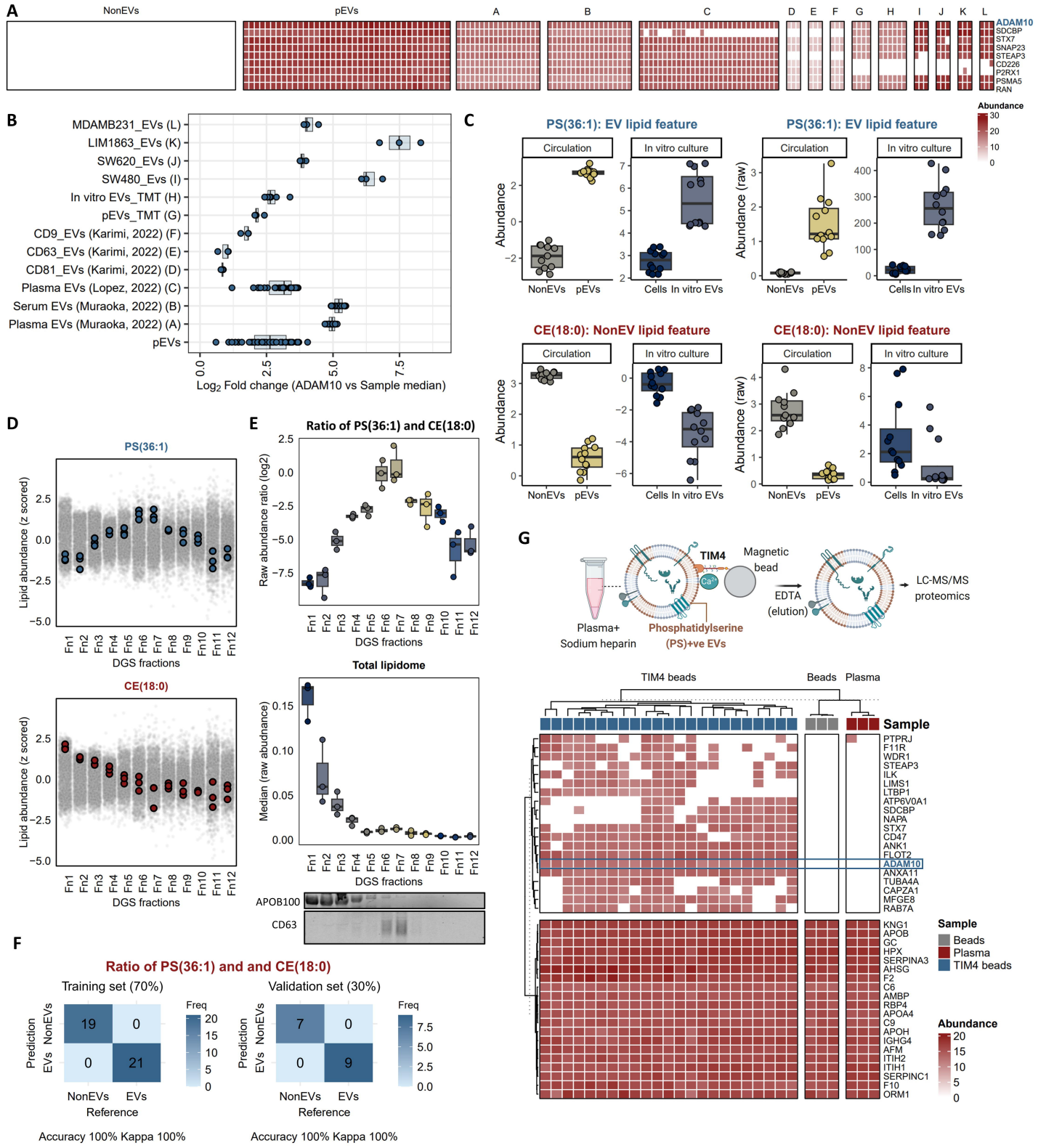

Figure 6. ADAM10 protein and phosphatidyl serine lipid serve as biological markers for human circulating EVs

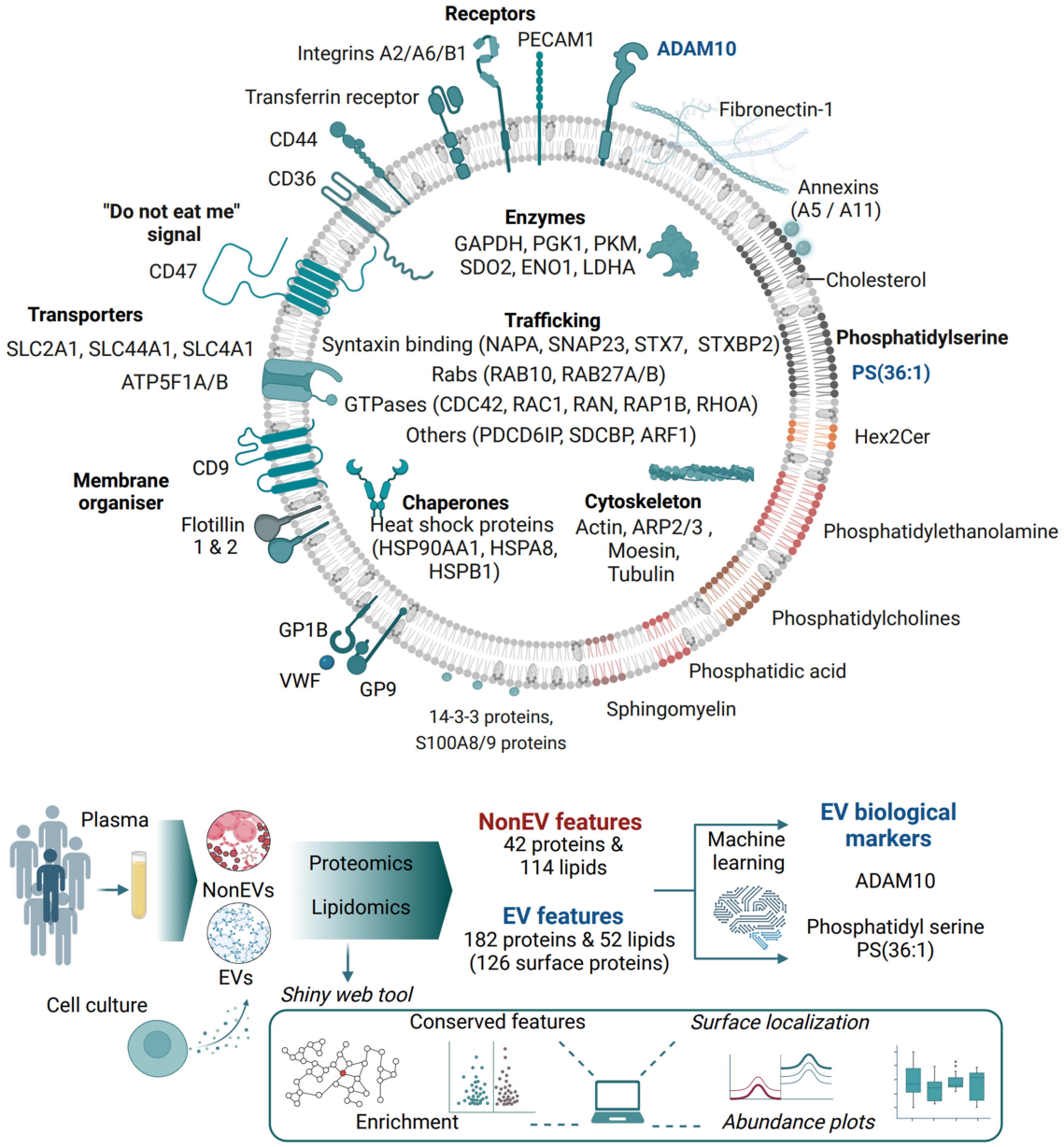

Figure 7. Hallmark molecular features of circulating EVs