Journal of Extracellular Vesicles

SHORT COMMUNICATION OPEN ACCESS

Toward Identification of Markers for Brain-Derived Extracellular Vesicles in Cerebrospinal Fluid: A Large-Scale, Unbiased Analysis Using Proximity Extension Assays

Maia Norman[1,2,3] | Adnan Shami-shah[1,2] | Sydney C. D'Amaddio[1,2] | Benjamin G. Travis[1,2] | Dmitry Ter-Ovanesyan[1] | Tyler J. Dougan[1,2,4] | David R. Walt[1,2]

[1]Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, Massachusetts, USA | [2]Department of Pathology, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA | [3]Department of Psychiatry, Massachusetts General Hospital, Boston, Massachusetts, USA | [4]Harvard-MIT Program in Health Sciences and Technology, Cambridge, Massachusetts, USA

Correspondence: David Walt (dwalt@bwh.harvard.edu)

Received: 13 April 2024 | **Revised:** 27 December 2024 | **Accepted:** 5 February 2025

Funding: This work was supported by funding from Good Ventures/Open Philanthropy (to David R. Walt) and from work on "Brain-Derived Extracellular Vesicles for Analysis of Treatment Resistant Major Depressive Disorder" supported by Wellcome Leap as part of the Multi-Channel Psych Program (to David R. Walt). These funding agencies had no role in conceptualization, design, analysis, decision to publish or preparation of the manuscript.

Keywords: brain-derived extracellular vesicles | membrane association | Olink | proteomics | size exclusion chromatography

ABSTRACT

Extracellular vesicles (EVs) captured in biofluids have opened a new frontier for liquid biopsies. To enrich for vesicles coming from a particular cell type or tumour, scientists utilize antibodies to transmembrane proteins that are relatively unique to the cell type of interest. However, recent evidence has called into question the basic assumption that all transmembrane proteins measured in biofluids are, in fact, EV-associated. To identify both candidate markers for brain-derived EV immunocapture and cargo proteins to validate the EVs' cell of origin, we conducted an unbiased Olink screen, measuring 5416 unique proteins in cerebrospinal fluid after size exclusion chromatography. We identified proteins that demonstrated a clear EV fractionation pattern and created a searchable dataset of candidate EV-associated markers—both proteins that are cell type-specific within the brain, and proteins found across multiple cell types for use as general EV markers. We further implemented the DeepTMHMM deep learning model to differentiate predicted cytosolic, transmembrane, and external proteins and found that intriguingly, only 10% of the predicted transmembrane proteins have a clear EV fractionation pattern based on our stringent criteria. This dataset further bolsters the critical importance of verifying EV association of candidate proteins using methods such as size exclusion chromatography before downstream use of the targets for EV analysis.

1 | Introduction

Extracellular vesicles (EVs) are nanometre-scale, membrane-bound compartments that contain proteins, RNAs and metabolites endogenous to their cell of origin (Raposo and Stoorvogel 2013). As such, the content of EVs, isolated from biofluids, can serve as a molecular snapshot of the parent cell. A preponderance of EV research has focused on isolating EVs from

Maia Norman and Adnan Shami-shah contributed equally to this article.

The code used for analysis is available on GitHub. (The word GitHub should be clicable with the following link: https://github.com/Walt-Lab/ev_association_olink_analysis.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2025 The Author(s). *Journal of Extracellular Vesicles* published by Wiley Periodicals, LLC on behalf of the International Society for Extracellular Vesicles.

specific cell types or tumours utilizing proteins annotated as transmembrane and enriched in the parent cell (Shami-shah et al. 2023). While this has led to some success, such as in the case of monitoring prostate cancer (Ramirez-Garrastacho et al. 2022), studies that have sought to capture brain-derived EVs have been hampered by methodological challenges. Specifically, proteins cited as transmembrane or internal to EVs have been shown to be predominantly cleaved and secreted (Norman et al. 2021). It is, therefore, critical to validate methods to differentiate EV-associated proteins from those that are secreted and cleaved in biofluids.

Plasma and cerebrospinal fluid (CSF) EVs can be easily separated from soluble proteins using size exclusion chromatography (SEC) or density gradient chromatography (DGC) (Norman et al. 2021). Nevertheless, analysing the proteomic content of the EV and soluble protein fractions with a single biochemical technique can be difficult because the soluble protein fractions contain several orders of magnitude more protein than the EV fractions. Unbiased techniques like mass spectrometry are challenging because, in the EV fractions, lipoproteins can co-isolate and mask rare EV-associated proteins, while in the secreted protein fractions, abundant proteins like albumin create a similar problem (Ter-Ovanesyan et al. 2023). Furthermore, the high levels of abundant proteins, such as albumin in plasma, preclude the ability to use gel-based techniques such as Western blots. As a result, ELISAs have thus far been the best method of assessing EV fractionation patterns (Ter-Ovanesyan et al. 2021). In previous work, we have utilized the ultrasensitive digital ELISA platform Simoa, invented by our lab, to quantify canonical EV proteins (CD9, CD63, CD81, Alix), assess potential contaminants to EV preparations (apolipoprotein B, albumin), and evaluate individual proteins as targets for cell-type-specific enrichment (Norman et al. 2021; Ter-Ovanesyan et al. 2021, 2023). Here, we sought to apply a large-scale unbiased method to generate a much-needed dataset and establish a bioinformatic approach to identify proteins that can be used for potential immunocapture of EVs secreted by a cell-type of interest, as well as cytosolic proteins to corroborate EV-brain-cell origin.

CSF directly surrounds the brain and the spinal cord, making CSF-derived EVs more likely to contain predominantly brain-specific markers compared to plasma and other biofluids (Hladky and Barrand 2014; Shetgaonkar et al. 2022). Furthermore, CSF has approximately 200-fold lower soluble protein content compared to plasma (Fogh et al. 2020). This lowers the chance of nonspecific interactions compared to that for EVs isolated from plasma and other more complex matrices. Although estimates of the proportion of brain-derived EVs in CSF are highly limited by the lack of reliable markers, one study reported that approximately 16% of brain-specific proteins in CSF EVs were of neuronal origin while about 84% of them were of glial origin (Muraoka, Jedrychowski, et al. 2020). This makes CSF an ideal biofluid for brain-derived EV biomarker discovery analysis. By using human CSF, we seek to make strides towards a liquid biopsy of the nervous system, eventually enabling the development of minimally invasive diagnostics for neurological and psychiatric diseases.

2 | Methods

2.1 | Human Sample Preparation

For the main experimental figures utilizing Olink and Simoa technology, one millilitre each of four healthy CSF samples (PrecisionMed) were thawed at room temperature and centrifuged at 2000 × *g* for 10 min. Subsequently, the supernatant from this first centrifugation was transferred to a 0.45-μm Corning Costar Spin-X filter (Sigma-Aldrich) and centrifuged again at 2000 × *g* for 10 min at room temperature. The flow-through from this filtration was used for downstream experiments. For Figures S1–S3 (nanoparticle tracking analysis and Western blotting), one pooled lot of CSF (Innovative Research) was used to ensure enough material was available for all Western blots and nanoparticle tracking analysis without adding inter-individual variability. The CSF was processed in the same way for this pooled lot as for the individual samples used in the main figures.

2.2 | SEC and Fraction Processing

Sepharose CL-6B resin (GE Healthcare) was washed with an equal volume of PBS 3 times. For each wash, the resin was allowed to settle at 4°C overnight before the PBS was poured off and replaced. Following the washes, the resin was stored in an equal volume of PBS.

Econo-Pac Chromatography columns (Bio-Rad) were prepared immediately prior to fractionation. For each sample, washed resin was poured into a column to achieve a resin bed volume of 10.2 mL. A polyethylene bed support (Bio-Rad) was inserted into the top of the resin to compress to a bed volume of 10 mL. The packed resin was then washed with 20 mL of PBS. Immediately following the elution of the wash, 1 mL of each CSF sample was added to the respective column, and fractions were collected in 0.5 mL increments. When the 1 mL of CSF had flowed through, 0.5 mL of PBS was added to the column sequentially until fractions 1–15 were collected. Fractions 1–5 were discarded to avoid redundancy as EVs generally begin to elute in fractions 7 or 8 when using a 10 mL Sepharose 6B column.

Each fraction (6–15) was transferred to 10 kDa MWCO Amicon Ultra Centrifugal Filters (Sigma-Aldrich) and diluted to a total volume of 1.5 mL with PBS. These fractions were then centrifuged at 2000 × *g* at 4°C until all fractions were concentrated 15-fold. The concentrated fractions were brought to a volume of 97 μL with PBS. A 76 μL aliquot was transferred to a 96-well plate supplied by Olink. Triton X-100 was added to a final concentration of 1% by volume, and the plate was stored at −80°C. The remaining 21 μL of fraction volume was used to measure CD81 by Simoa.

2.3 | Simoa CD81 Sample Analysis

The Simoa analysis was performed according to the manufacturer's instruction. Reagent preparation and assay parameters were followed as described previously by Norman et al. (2021).

Abcam (anti-CD81 ab79559, clone M38) and Biolegend (anti-CD81 349502, clone 5A6) were used as capture and detector antibodies, respectively. Human recombinant CD81 from Origene (TP317508) was used in the calibration curve. Data analysis was performed using GraphPad Prism version 10.1.1.

2.4 | Nanoparticle Tracking Analysis

Separate CSF fractions 7–10 and 11–15 were collected using SEC, as described above. This 2 mL volume was condensed using a 10 kDa MWCO Amicon Ultra Centrifugal Filter (Sigma-Aldrich) to a volume of 500 µL in PBS. Extracellular vesicle particle size and number were characterized using the NanoSight LM10 (Malvern Panalytical). A 500 µL of sample was injected, and five 1-min videos were captured at 24.98 fps with a detection threshold of 2, at a fixed temperature of 25°C. Parameters were determined based on the manufacturer's software manual and performed by the NTA 3.4 Build software v3.4.4.

2.5 | Western Sample Analysis

SEC was performed as above with 8 mL of pooled CSF. Each mL was loaded on its own column. Respective fractions were pooled and concentrated using 10 kDa MWCO Amicon Ultra Centrifugal Filters (Sigma-Aldrich). For fractions 6–12, one-sixteenth of the concentrated fractions were loaded per gel. For fractions 13–15, protein input was normalized to fraction 12 to avoid overloading the gel. The fractions and human brain cerebellum whole tissue lysate (HBL) (Novus Biologicals) were denatured with 4× LDS and, for certain targets, reduced with DTT (see table below). Subsequently, CSF and HBL samples were heated at 70°C for 10 min, run at 150 V for 70 min on 4%–12% Bolt Bis-Tris Plus gels (Thermo Fisher Scientific), and transferred to nitrocellulose membranes using the iBlot 3 Dry Blotting System (Thermo Fisher Scientific). The membranes were blocked for 30 min at 4°C and incubated with primary antibodies overnight. The next day, membranes were washed, incubated with secondary antibody (Bethyl Laboratories) for 1 h at 4°C, and washed again. Nonspecific signals were assessed by probing CSF SEC fractions and HBL with the corresponding secondary antibodies (anti-mouse IgG, anti-rabbit IgG, or anti-rat IgG) without the application of primary antibody. For primary and secondary antibody dilutions, as well as membrane blocking, a PBS-T solution of 5% milk (weight by volume) with 1% Tween was used. All washes were performed with PBS-T (1% Tween) in cycles of three 7-min washes (except SLC16A1, which was incubated in PBS-T six times per wash). Specifics on primary antibodies used and dilutions can be found in the table below. After the final wash, blots were developed using the ProSignal Femto substrate kit (Genesee Scientific) and imaged with a Sapphire Biomolecular Imager (Azure Biosystems) Table 1.

2.6 | Olink Sample Analysis

Samples were shipped on dry ice to the Broad Institute in Cambridge, MA, for analysis by the Olink HT platform, which measures 5416 unique proteins using highly multiplexed proximity extension assays. Pairs of antibodies with unique,

TABLE 1 | Western blot antibodies used for target verification.

Target	Primary antibody clone	Primary antibody vendor	Primary antibody species	Reducing conditions?	1:25 Diluted HBL volume loaded (µL)	Primary antibody dilution	Secondary antibody dilution
CD9	MM2/57	Millipore sigma	Mouse	No	1	1:1000	1:2000
CD63	H5C6	BD Biosciences	Mouse	No	10	1:1000	1:2000
CD81	M38	Thermo Fisher Scientific	Mouse	No	3	1:666	1:1000
AQP1	EPR11588(B)	Abcam	Rabbit	Yes	4	1:10000	1:1000
SLC16A1	E7A2K	Cell Signaling Technology	Rabbit	Yes	1.75	1:1000	1:1000
FCAR	EPR4622(2)	Abcam	Rabbit	Yes	20	1:1000	1:1000
CHRM3	580011	R&D Systems	Mouse	Yes	20	1:1000	1:1000
TSPAN	458811	R&D Systems	Rat	Yes	20	1:1000	1:1000

complementary oligonucleotides, called proximity probes, each specific to a unique protein of interest, bind to their target antigens. After binding the target, the oligonucleotide probes encounter each other due to physical proximity and hybridize, resulting in the formation of an immuno-complex. The resulting hybridized proximity probes can be amplified by DNA polymerase, creating a DNA amplicon that can be detected by quantitative PCR (qPCR) or next-generation sequencing (NGS) techniques (Shami-shah et al. 2023; Wik et al. 2021). Samples were run with a single replicate for each protein except for GBP1 and MAP2K1, which were run in Blocks 3, 4 and 5 to check correlation between blocks.

The relative abundances of the amplicon, as measured by NGS, are then converted to normalized protein expression (NPX) values. The Olink panel includes plate, sample, and extension (ExtCtrl) controls. To ensure robustness, the NPX calculation accounts for variability in the different controls measured in the panel and includes a log2 transformation of the data. The number of matched sequence reads (counts) generated by NGS is first normalized by the number of counts for the extension control of the sample, and then log2 transformed as follows:

$$ExtNPX_{i,\,j} = \log 2\left(\frac{Counts\left(Sample_j Assay_i\right)}{Counts\left(ExtCtrl_j\right)}\right)$$

where $\mathrm{ExtNPX}_{i,j}$ is the NPX, normalized by the counts of the extension control specific to assay i measured in sample j. The median value of ExtNPX of the plate controls is then used to adjust for variability between plates, allowing comparison of relative protein abundances across different plates (Wik et al. 2021):

$NPX_{i,\,j} = ExtNPX_{i,\,j} - median(ExtNPX_{plate_h\ control})$ control is the quality control measure collected from plate h, and $\mathrm{NPX}_{i,j}$ is the reported NPX value for sample j, analysed using assay i on plate h.

Further details on NPX value generation can also be found on the Olink website.

2.7 | Data Analysis Methods

The reported NPX values have an arbitrary unit and reflect the relative concentrations of the analysed proteins in the sample of interest. All analysis was conducted in Python (version 3.11.5) using Visual Studio Code (Microsoft Corporation, Redmond, Washington). The HT panel includes 5420 proteins, including 5416 unique proteins and two assays processed in triplicate, measuring relative concentrations of MAP2K1 and GBP1, to ensure accuracy of data collection. Given that no calibration curve is included, all NPX values presented in fraction data are relative values only. The data points were all linearized, and two assays that were processed in triplicate were removed from downstream high-throughput analysis.

The HT panel from Olink measures two negative controls for each assay. Olink recommends against calculating a limit of detection (LOD) with fewer than 10 negative controls in a dataset, so we instead considered the fixed LODs made available by Olink. The fixed LOD calculation is based on 24–36 negative controls, ensuring a more robust calculation to minimize the higher variation among negative controls. This approach is consistent with the recommendations from Olink, which reports that values below LOD are unlikely to increase the risk of false positive discoveries and may be beneficial for biomarker discovery. They also highlight that filtering data based solely on LOD may remove meaningful signals, especially when a protein is well expressed in one group but undetectable in another. Therefore, excluding data points below LOD would prevent us from including potentially useful proteins in our analyses. The LOD data are available in Table S2 but were not considered in downstream analyses.

2.8 | Fractionation Analysis

Four individual fractionated CSF samples (fractions 6–15) were submitted to Olink for analysis. Only fractions 7, 9, 10, 11, 12 and 13 were used to verify whether a protein exhibited the fractionation pattern typical for EV-associated proteins (Norman et al. 2021). For each fraction of interest, the median NPX was calculated for each protein. A protein was considered to have a fractionation pattern typical of EV-associated proteins if the medians of fractions 9 and 10 were greater than the medians for fractions 7, 11, 12 and 13.

2.9 | Protein Localization

The proteins in the Olink panel were computationally determined to be transmembrane, internal, or external using DeepTMHMM. DeepTMHMM is a deep learning model-based algorithm that uses a hidden Markov model to predict subcellular localization of a protein in a cell. The model calculated a probability for each amino acid in each protein and returned the highest probability domain for each amino acid, allowing the most likely localization of the overall protein to be determined. Using this model, each amino acid was characterized as:

1. Cytosolic
2. Alpha transmembrane helix
3. Beta transmembrane barrel
4. Signalling peptide
5. External to the cell and any secreted vesicles or exosomes

Information regarding signalling peptides was not considered, as they are largely cleaved from the protein when it enters the endoplasmic reticulum, and, therefore, are unlikely to be present in the epitope of the protein found in EVs (Liaci and Forster 2021). We classified a protein as internal to the cell if all its amino acids were characterized as cytosolic, and we classified a protein as external if all its amino acids were characterized as being outside the cell or on secreted vesicles Figure 1. Because proteins containing a transmembrane domain also contain domains found internal and external to the cell, a protein was classified as transmembrane if it contained one or more amino acids characterized as an alpha transmembrane helix or a beta transmembrane barrel. Because

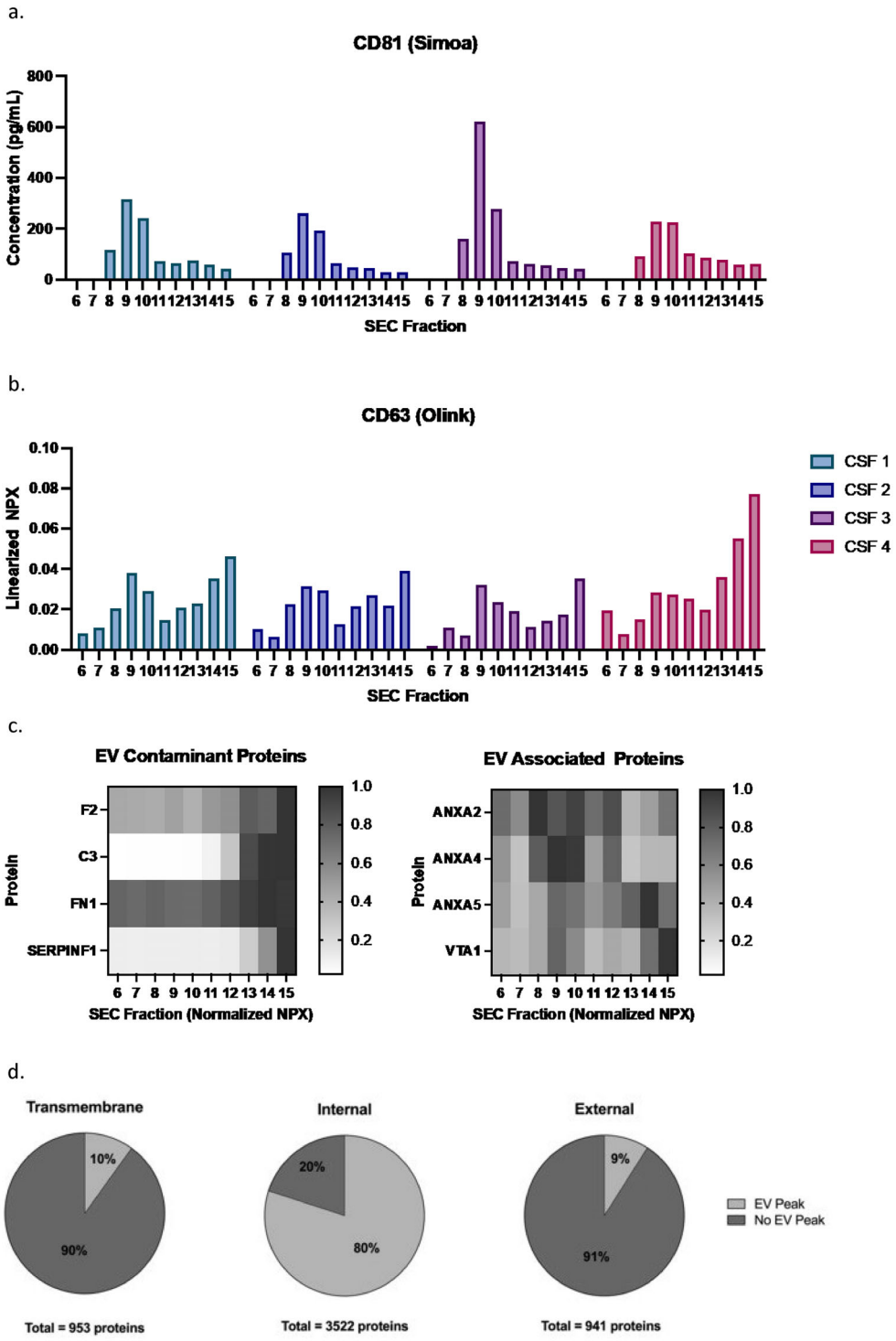

FIGURE 1 | CSF SEC fractionation as a measure of EV association. (a) Quantification of CSF fractions using a Simoa assay for CD81. Four individual healthy CSF samples were fractionated using SEC, and each fraction was analysed by Simoa. A Mann–Whitney *U* test performed comparing fractions 9 and 10 with fractions 7, 11, 12 and 13 in all samples combined showed fractions 9 and 10 are significantly greater than fractions 7, 11, 12 and 13 ($p < 0.0005$). (b) Quantification of CSF fractions using the Olink assay for CD63. Four individual healthy samples were fractionated using SEC, and each fraction was analysed by the Olink HT panel. The Mann–Whitney *U* test performed comparing fractions 9 and 10 with fractions 7, 11, 12 and 13 in all samples combined showed fractions 9 and 10 are significantly greater than fractions 7, 11, 12 and 13 ($p < 0.0005$). (c) Heat maps showing normalized NPX values for each SEC fraction for four representative previously published EV contaminants and four EV-associated proteins in the Olink panel. Of note, the EV contaminant proteins F2, C3, FN1 and SERPINF1 (PEDF) all have increasingly high NPX values predominantly in the late free protein fractions. EV-associated proteins ANXA2, ANXA4, ANXA5 and VTA1 show EV-associated fractionation patterns with high NPX values in fractions 9 and 10. Anxa5 and VTA1 also show NPX signals in later fractions 14 and 15, suggesting possibly soluble protein isoforms for these proteins. (d) Percentage of Deep TMHMM predicted transmembrane, internal, and external targets quantified by Olink as having an EV fractionation pattern in CSF. CSF, cerebrospinal fluid; NPX, normalized protein expression; PEDF, pigment epithelium-derived factor; SEC, size exclusion chromatography.

of the budding mechanisms by which EVs are secreted (Teng and Fussenegger 2020), it is largely assumed that proteins would have the same cytosolic, transmembrane, or extracellular domains in both EVs and the cell. However, additional validation techniques would be necessary to confirm the localization of proteins relative to EVs, a problem we address through SEC fractionation analysis as described previously.

2.10 | EV-Associated Protein Identification

This pipeline was used to identify proteins that may be associated with EVs. Proteins were labelled as internal to EVs if they met the fractionation criteria and were identified as internal using DeepTMHMM as described previously. The same criteria were followed to identify transmembrane and external proteins associated with EVs. This yielded a list that was further narrowed by selecting proteins considered to be cell type-specific based on the Tau score and BrainRNA-Seq dataset as described below. Each protein was assigned an "EV Association Score," which was calculated as the ratio of the median NPX for the EV fractions (fractions 9 and 10), and the median NPX for fractions 7, 11, 12 and 13. This value is shown on the *y*-axis of Figure 2.

2.11 | Cell-Type-Specificity

The BrainRNA-Seq atlas reports fragments per kilobase per million mapped fragments (FPKM), collected via RNA sequencing (Zhang et al. 2016). The mean FPKM of each gene was used for mature astrocytes, neurones, oligodendrocytes, endothelial cells, and microglia. Foetal astrocytes were excluded from analysis. Tau specificity score is used to determine cell-type-specificity of genes, as it gives a numerical indication of the relative specificity of a gene across different cell types or tissues. Scores range between 0 and 1, where 0 indicates that a gene is ubiquitously expressed in all cell types, and 1 indicates that a gene is entirely expressed in a single cell type (Kryuchkova-Mostacci and Robinson-Rechavi 2017). A gene was considered specific to a given cell type if it had a Tau specificity score of greater than 0.75 and if the mean FKPM was highest in the cell type of interest relative to the other cell types. Tau specificity scores were calculated using the following formula (Kryuchkova-Mostacci and Robinson-Rechavi 2017):

$$\tau = \frac{\sum_{i=1}^{n} (1 - \hat{x}_i)}{n - 1}$$

$$\hat{x}_i = \frac{x_i}{\max\limits_{1 \le i \le n} (x_i)}$$

$$x_i = expression\ of\ the\ gene\ of\ interest\ in\ tissue\ i$$

$$n = number\ of\ tissues$$

The opposite is also true—by identifying proteins with a low Tau specificity score, < 0.25, we selected genes that are ubiquitously expressed in all cell types. The genes were then mapped to proteins using data obtained from the UniProt website. This data is included in Table S5.

2.12 | Brain Organ Specificity

The GTEx database provides median gene-level expression transcripts per million (TPM) by tissue (Kowal et al. 2016). The tissues were grouped as described in Table 2 below:

The median TPM of each organ group was used to calculate the tissue specificity of each gene. The tau specificity score was used to determine the organ specificity of each gene, as it gives a numerical indication of the relative specificity of a gene across different organ groups. Scores range between 0 and 1, where 0 indicates that a gene is ubiquitously expressed in all tissue, and 1 indicates that a gene is entirely specific to a single tissue type (Kryuchkova-Mostacci and Robinson-Rechavi 2017). This data is included in Table S5 for reference but is not considered in quantifying cell-type-specificity as shown in Figure 2.

3 | Results

We used a highly multiplexed proximity extension assay platform from Olink to analyse thousands of proteins from microlitres of biofluid with high specificity (Olink 2020). To assess EVs coming from the brain, we fractionated CSF from healthy individuals using SEC to separate proteins that peak in the early EV fractions from those that peak in the late secreted protein fractions (Thery et al. 2018; Welsh et al. 2024).

To define our EV fractions, adhering to MISEV 2018[18] and 2023[19] guidelines, we analysed 20% of each fraction using our previously validated Simoa assay for CD81 to demonstrate that EVs predominantly eluted in fractions 9 and 10 (Figure 1a) (Norman et al. 2021; Ter-Ovanesyan et al. 2021, 2023). The remaining 80% of each fraction was analysed using the Olink HT platform, which quantifies 5416 unique proteins (Table S1). We analysed the fractionation pattern of CD63 using data from the Olink assay and demonstrated a peak in signal in fractions 9 and 10 (Figure 1b). We performed nanoparticle tracking analysis to show that EV-sized particle counts are increased in fractions 7–10 (Figure S1). Next, we performed Western blots of CD9, CD63 and CD81 on fractionated CSF and demonstrated peak signals in fractions 9 and 10 for all three tetraspanins. Of note, in the Olink data, CD63 had a second later peak, which was not observed in Western blotting, indicating this peak may be caused by nonspecific binding in the setting of high protein abundance in the later soluble protein fractions. Finally, in agreement with the literature (Kowal et al. 2016; You et al. 2022; Jeppesen et al. 2019; Hallal et al. 2022) and MISEV 2018[18], we also report several previously identified generic EV markers, Annexins A2 (ANXA2), A4 (ANXA4), and A5 (ANXA5), and Vacuolar protein sorting-associated protein VTA1 homolog (VTA1), and non-EV contaminant markers fibronectin (FN1), prothrombin (F2), pigment epithelium-derived factor (PEDF, also known as SERPINF1), and complement C3 (C3) (Figure 1c) included in our Olink pipeline.

To identify targets that could be effective for EV immunocapture or for the analysis of EV cargo, we selected all proteins where the median NPX value across CSF samples was greater in both fractions 9 and 10 compared to fractions 7, 11, 12 and 13 (Table S3). Because many proteins can be found as both

TABLE 2 | Classification of tissues used to characterize brain specificity.

Group	GTEx portal category
Brain	Brain_Amygdala Brain_Anterior_cingulate_cortex_BA24 Brain_Caudate_basal_ganglia Brain_Cerebellar_Hemisphere Brain_Cerebellum Brain_Cortex Brain_Frontal_Cortex_BA9 Brain_Hippocampus Brain_Hypothalamus Brain_Nucleus_accumbens_basal_ganglia Brain_Putamen_basal_ganglia Brain_Spinal_cord_cervical_c-1 Brain_Substantia_nigra Nerve_Tibial Pituitary
Heart	Heart_Atrial_Appendage Heart_Left_Ventricle
Small intestine	Small_Intestine_Terminal_Ileum Small_Intestine_Terminal_Ileum_Lymphode_Aggregate Small_Intestine_Terminal_Ileum_Mixed_Cell
Colon	Colon_Sigmoid Colon_Transverse Colon_Transverse_Mixed_Cell Colon_Transverse_Mucosa Colon_Transverse_Muscularis
Liver	Liver Liver_Hepatocyte Liver_Mixed_Cell Liver_Portal_Tract
Pancreas	Pancreas Pancreas_Acini Pancreas_Islets Pancreas_Mixed_Cell
Esophagus	Esophagus_Gastroesophageal_Junction Esophagus_Mucosa Esophagus_Muscularis
Stomach	Stomach Stomach_Mixed_Cell Stomach_Mucosa Stomach_Muscularis
Kidney	Kidney_Cortex Kidney_Medulla
Adipose	Adipose_Subcutaneous Adipose_Visceral_Omentum
Artery	Artery_Aorta Artery_Coronary Artery_Tibial
Skin	Skin_Not_Sun_Exposed_Suprapubic Skin_Sun_Exposed_Lower_leg
Muscle	Muscle_Skeletal
Cervix	Cervix_Ectocervix Cervix_Endocervix

(Continues)

TABLE 2 | (Continued)

Group	GTEx portal category
Lung	Lung
Spleen	Spleen
Testis	Testis
Breast	Breast_Mammary_Tissue
Ovary	Ovary
Prostate	Prostate
Thyroid	Thyroid
Bladder	Bladder
Uterus	Uterus
Vagina	Vagina
Cell culture	Cells_Cultured_fibroblasts Cells_EBV-transformed_lymphocytes
Fallopian tube	Fallopian_Tube
Minor salivary gland	Minor_Salivary_Gland
Adrenal gland	Adrenal_Gland
Whole blood	Whole_Blood

EV-bound and soluble isoforms, we did not consider relative protein abundance in fractions 14 and 15 in our criteria, but rather selected proteins where a definable EV fractionation pattern could be seen. The signal from EV-associated proteins begins to peak from fraction 8 and reaches its highest point in fractions 9 and 10. With minimal to no signal observed in fractions 6 and 7, we treat these two fractions as internal controls. However, due to proportional signals from fractions 6 and 7 based on both CD81 Simoa (Figure 1a) and CD63 Olink (Figure 1b) assays, we used fraction 7, rather than the combination of fractions 6 and 7, in our EV fractionation pattern selection criteria. Next, we utilized the DeepTMHMM deep learning model to differentiate cytosolic, transmembrane, and external proteins. Running this model on each protein analysed by the Olink platform, we categorized them into 953 predicted transmembrane, 3522 predicted cytosolic, and 941 predicted external proteins (Hallgren et al. 2022). We demonstrate that 80% of predicted cytosolic proteins, 10% of transmembrane proteins, and 9% of external proteins have a definable EV fractionation pattern (Figure 1d).

The HT panel from Olink measures two negative controls for each assay. Olink recommends against calculating a lower LOD with fewer than 10 negative controls in a dataset, so we instead considered the fixed LODs made available by Olink, available in Table S2. The fixed LOD calculation is based on 24–36 negative controls, ensuring a more robust calculation to minimize the higher variation among negative controls. When thresholding our data using the LODs, we observed a significant loss of many targets. In total, without considering LOD, our analysis pipeline identifies 57 unique transmembrane and internal proteins associated with EVs. However, when we threshold using the fixed LOD, this number drops to three proteins across the five cell types. This is consistent with the recommendations from Olink, which reports that values below LOD are unlikely to increase the risk of false positive discoveries and may risk eliminating informative biomarkers. They also highlight that filtering data based solely on LOD may remove meaningful signals, especially when a protein is well expressed in one group but undetectable in another. For example, Aquaporin 1 (AQP1), a transmembrane protein specific to astrocytes, is eliminated from consideration when thresholding due to all fractions except fraction 9 being below LOD. However, we independently validated its fractionation pattern via Western blot (Figure S3) and received results similar to those reported through Olink. Therefore, excluding data points below LOD could prevent us from including potentially useful proteins in our analyses.

Our primary interest in using this dataset was to identify proteins that can be used to isolate or define an EV's cell of origin. Therefore, we overlaid the Olink data with the BrainRNA-Seq atlas and selected proteins that were enriched in a specific brain cell-type—as defined by having a Tau specificity score > 0.75[14], calculated using the mean astrocyte, oligodendrocyte, microglia, neuron, and endothelial cell expression levels. Thus, we identified candidate transmembrane and external proteins that can potentially be used in CSF to isolate cell-type-specific brain-derived EVs as well as candidate cytosolic proteins that can be analysed as internal EV cargo to confirm cell-type-specificity following immunocapture (Figure 2).

Finally, we identified a set of proteins that demonstrate a clear EV fractionation pattern but are not specific to a given cell type as defined by a Tau score < 0.25 (Table S4). These latter proteins can be used to normalize total EV quantity.

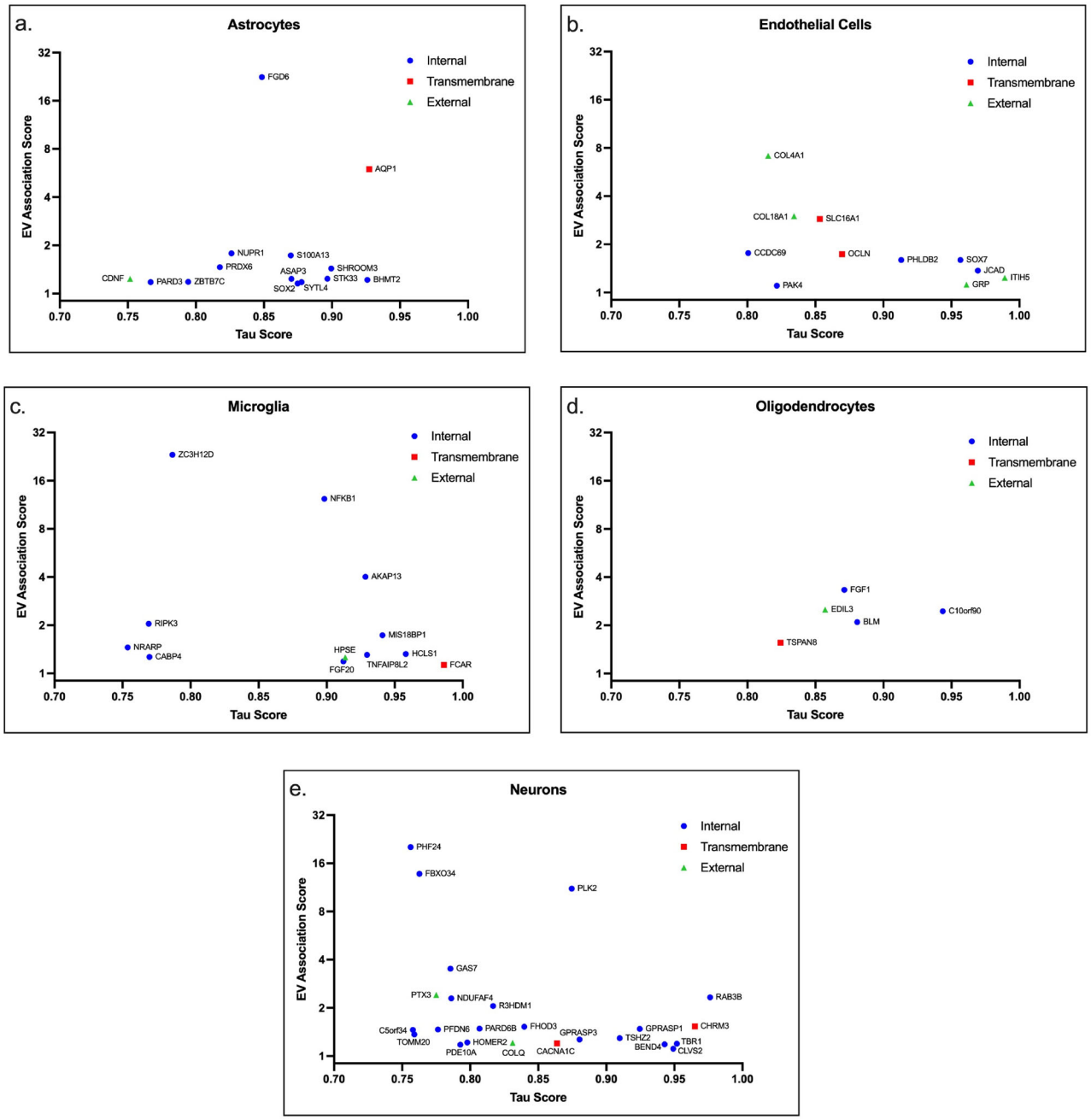

FIGURE 2 | Cell-type-specificity of proteins that show an EV-associated fractionation pattern. The EV Association Score (EV-associated NPX signals in fractions 9 and 10 were greater than NPX signals in fractions 7, 11, 12 and 13) and calculated Tau Score of > 0.75 for each identified transmembrane (red), internal (blue) and external (green) proteins that demonstrated an EV-associated fractionation pattern for (a) astrocytes, (b) endothelial cells, (c) microglia, (d) oligodendrocytes, and (e) neurons. NPX, normalized protein expression.

4 | Conclusions

By utilizing the highly sensitive and specific multiplexed Olink platform on SEC-fractionated healthy CSF, we identified cell-type-specific proteins that may be associated with EVs and can be used both for potential EV immunocapture and for the analysis of the luminal protein cargo of brain-derived EVs. Furthermore, we demonstrate that 90% of predicted transmembrane proteins did not have a definable EV fractionation pattern, which we speculate is due to overwhelming signals from cleaved or secreted isoforms of these proteins. Such targets are likely not viable for use in EV immunocapture. Conversely, some targets identified as external were highly EV-associated (e.g., EDIL3) and are likely bound tightly to the extravesicular surface, making them potential immunocapture targets.

There are several important caveats to this work: First, although we analysed 5416 targets, this remains only a quarter of the ~20,000 proteins known to be in the human protein coding genome (Aebersold et al. 2018). Second, many proteins are known to be found in both secreted and transmembrane forms. In some cases, the abundance of the secreted form can mask an EV

peak in fractions 9 and 10 after SEC. Without the ability to separate the peak in fractions 9 and 10, which includes EVs and associated proteins, from the soluble protein peak in fractions 14 and 15, these proteins are likely not useful for EV enrichment unless they have a unique extracellular epitope absent on the cleaved and secreted forms (Shami-shah et al. 2023; Norman et al. 2021). Third, while proximity extension assays lower the chance of nonspecific binding as can occur with ELISAs, the soluble protein fractions have substantially more protein, increasing the chance for nonspecific binding interactions to produce a signal, as was seen in our CD63 comparison between Olink and Western blot. Thus, our analysis is useful for identifying potential EV-associated proteins but cannot rule out EV association for proteins that do not meet our stringent criteria or that are not included in the Olink HT panel. This dataset supports the necessity of running SEC or DGC on putative immunocapture targets before proceeding to EV immunocapture. Fourth, our cell-type-specificity analysis accounts for within-brain specificity but does not include specificity for cell types outside of the brain. As can be seen using the GTex database (Consortium 2020) (overlayed in Table S5), some of the cell-type-specific targets in Figure 2 are highly specific to the brain (e.g., TBR1, FGF1 and SOX2), while others are also expressed on several cell types outside the brain (e.g., NDUFAF4, TOMM20 and PRDX6). Therefore, for EV analysis of CSF, our specificity criteria are likely sufficient, as the majority of CSF proteins come from within the central nervous system. However, if future work is to utilize these targets for analysis in blood, more stringent criteria overlaying the GTEx database would need to be included to ensure cell-type-specificity.

Due to the NGS readout, Olink has a wide dynamic range of 10 logs (fg-mg/mL) while requiring as little as 2 μL of sample input (Shami-shah et al. 2023; Olink 2020). In contrast, depending on the instrument type, mass spectrometry has a much narrower dynamic range of 4–5 logs (Tang et al. 2004; Marshall et al. 2013). This narrower range necessitates greater sample input, depletion of higher abundant contaminant proteins (e.g., albumin, lipoproteins, immunoglobulins), and more complex sample processing and cleanup (leading to additional sample loss) to detect lower abundance proteins. Therefore, while discovery-based mass spectrometry is a powerful technology, its lower dynamic range results in the preferential detection of highly abundant proteins, limiting the ability to access the low-abundance EV proteome.

While previous research has explored EVs derived from brain tissues, cell-type-specific media collected from induced pluripotent stem cells of various cell types, and CSF-derived EVs without any consideration of cell-type-specificity (Muraoka, Jedrychowski, et al. 2020; You et al. 2022; Muraoka, DeLeo, et al. 2020), to our knowledge, this dataset provides the first unbiased proteomic profiling of EV association on a large scale, making it a valuable resource for future EV biomarker discovery. Additionally, with the growing importance of cell type-specific EVs in liquid biopsy for hard-to-biopsy organs (e.g., the brain), we have created a computational approach based on stringent criteria to discover potential cell type-specific brain-derived EV biomarkers. While many of the proteins we identified as EV-associated have been described previously in the literature (Oshikawa-Hori et al. 2021; Hoshino et al. 2020; Gupta et al. 2022), substantial additional work is required to assess cell origin for those proteins that meet the criteria displayed in Figure 2. In future work, we plan to validate these candidate proteins for each cell type, prioritizing those with the highest Tau and EV association scores. Validation will require immunocapture with antibodies to a transmembrane or external protein and analysis of proposed internal targets with Simoa following a proteinase protection assay. We nonetheless feel that this process of target validation should be done collaboratively within the EV community. Therefore, this dataset is an important and powerful new resource for identifying novel targets for brain-derived EVs.

AUTHOR CONTRIBUTIONS

Maia Norman: Conceptualization (lead), data curation (lead), formal analysis (lead), funding acquisition (lead), investigation (lead), methodology (lead), project administration (lead), writing–original draft (lead), writing–review and editing (lead). **Adnan Shami-shah**: Conceptualization (lead), data curation (lead), formal analysis (lead), investigation (lead), methodology (lead), project administration (lead), writing–original draft (lead), writing–review and editing (lead). **Sydney C. D'Amaddio**: Formal analysis (lead), methodology (lead), resources (lead), software (lead), writing–review and editing (supporting). **Benjamin G. Travis**: Data curation (lead), methodology (lead), writing–review and editing (supporting). **Dmitry Ter-Ovanesyan**: Formal analysis (supporting), investigation (supporting), methodology (supporting). **Tyler J. Dougan**: Methodology (supporting), Software (supporting). **David R. Walt**: Conceptualization (supporting), funding acquisition (lead), project administration (lead), Writing–review and editing (lead).

Acknowledgements

Consent

All human samples utilized in this work were purchased from commercial sources. All patients were appropriately consented. The use of these samples was approved by the Mass General Brigham IRB.

Conflicts of Interest

David R. Walt is a founder and equity holder in Quanterix. His interests were reviewed and are managed by Mass General Brigham in accordance with their conflicts of interest policies.

Data Availability Statement

Full dataset with raw data is included in Table S1.

References

Aebersold, R., J. N. Agar, I. J. Amster, et al. 2018. "How Many Human Proteoforms Are There?" *Nature Chemical Biology* 14: 206–214. https://doi.org/10.1038/nchembio.2576.

Consortium, G. T. 2020. "The GTEx Consortium Atlas of Genetic Regulatory Effects Across Human Tissues." *Science* 369, no. 6509: 1318–1330. https://doi.org/10.1126/science.aaz1776.

Fogh, J. R., A. M. Jacobsen, T. Nguyen, K. D. Rand, and L. R. Olsen. 2020. "Investigating Surrogate Cerebrospinal Fluid Matrix Compositions for Use in Quantitative LC-MS Analysis of Therapeutic Antibodies in the Cerebrospinal Fluid." *Analytical and Bioanalytical Chemistry* 412, no. 7: 1653–1661. https://doi.org/10.1007/s00216-020-02403-3.

Marshall, A. G., G. T. Blakney, T. Chen, et al. 2013. "Mass Resolution and Mass Accuracy: How Much Is Enough?" *Mass Spectrometry (Tokyo)* 2: S0009. https://doi.org/10.5702/massspectrometry.S0009.

Gupta, K., K. A. Brown, M. L. Hsieh, et al. 2022. "Necroptosis Is Associated With Rab27-Independent Expulsion of Extracellular Vesicles Containing

RIPK3 and MLKL." *Journal of Extracellular Vesicles* 11, no. 9: e12261. https://doi.org/10.1002/jev2.12261.

Hallal, S., A. Tuzesi, G. E. Grau, M. E. Buckland, and K. L. Alexander. 2022. "Understanding the Extracellular Vesicle Surface for Clinical Molecular Biology." *Journal of Extracellular Vesicles* 11, no. 10: e12260, https://doi.org/10.1002/jev2.12260.

Hallgren, J., K. Tsirigos, M. D. Pedersen, et al. 2022. "DeepTMHMM Predicts Alpha and Beta Transmembrane Proteins Using Deep Neural Networks." Preprint, bioRxiv, April 10. https://doi.org/10.1101/2022.04.08.487609.

Hladky, S. B., and M. A. Barrand. 2014. "Mechanisms of Fluid Movement Into, Through and Out of the Brain: Evaluation of the Evidence." *Fluids Barriers CNS* 11, no. 1: 26. https://doi.org/10.1186/2045-8118-11-26.

Hoshino, A., H. S. Kim, L. Bojmar, et al. 2020. "Extracellular Vesicle and Particle Biomarkers Define Multiple Human Cancers." *Cell* 182, no. 4: 1044–1061.e1018. https://doi.org/10.1016/j.cell.2020.07.009.

Jeppesen, D. K., A. M. Fenix, J. L. Franklin, et al. 2019. "Reassessment of Exosome Composition." *Cell* 177, no. 2: 428–445.e418. https://doi.org/10.1016/j.cell.2019.02.029.

Kowal, J., G. Arras, M. Colombo, et al. 2016. "Proteomic Comparison Defines Novel Markers to Characterize Heterogeneous Populations of Extracellular Vesicle Subtypes." *Proceedings of the National Academy of Sciences* 113, no. 8: E968–E977. https://doi.org/10.1073/pnas.1521230113.

Kryuchkova-Mostacci, N., and M. Robinson-Rechavi. 2017. "A Benchmark of Gene Expression Tissue-Specificity Metrics." *Briefings in Bioinformatics* 18, no. 2: 205–214. https://doi.org/10.1093/bib/bbw008.

Liaci, A. M., and F. Forster. 2021. "Take Me Home, Protein Roads: Structural Insights Into Signal Peptide Interactions During ER Translocation." *International Journal of Molecular Sciences* 22, no. 21: 11871. https://doi.org/10.3390/ijms222111871.

Muraoka, S., A. M. DeLeo, M. K. Sethi, et al. 2020. "Proteomic and Biological Profiling of Extracellular Vesicles From Alzheimer's Disease Human Brain Tissues." *Alzheimer's & Dementia* 16, no. 6: 896–907. https://doi.org/10.1002/alz.12089.

Muraoka, S., M. P. Jedrychowski, K. Yanamandra, S. Ikezu, S. P. Gygi, and T. Ikezu. 2020. "Proteomic Profiling of Extracellular Vesicles Derived From Cerebrospinal Fluid of Alzheimer's Disease Patients: A Pilot Study." *Cells* 9, no. 9: 1959. https://doi.org/10.3390/cells9091959.

Norman, M., D. Ter-Ovanesyan, W. Trieu, et al. 2021. "L1CAM is Not Associated With Extracellular Vesicles in Human Cerebrospinal Fluid or Plasma." *Nature Methods* 18: 631–634. https://doi.org/10.1038/s41592-021-01174-8.

Oshikawa-Hori, S., N. Yokota-Ikeda, H. Sonoda, Y. Sasaki, and M. Ikeda. 2021. "Reduced Urinary Release of AQP1- and AQP2-bearing Extracellular Vesicles in Patients With Advanced Chronic Kidney Disease." *Physiological Reports* 9, no. 17: e15005, https://doi.org/10.14814/phy2.15005.

Olink. 2020. "PEA—A High-Multiplex Immunoassay Technology With qPCR or NGS Readout." White Paper. Olink. ⟨https://www.olink.com/content/uploads/2021/09/olink-white-paper-pea-a-high-multiplex-immunoassay-technology-with-qpcr-or-ngs-readout-v1.0.pdf.

Ramirez-Garrastacho, M., C. Bajo-Santos, A. Line, et al. 2022. "Extracellular Vesicles as a Source of Prostate Cancer Biomarkers in Liquid Biopsies: A Decade of Research." *British Journal of Cancer* 126, no. 3: 331–350. https://doi.org/10.1038/s41416-021-01610-8.

Raposo, G., and W. Stoorvogel. 2013. "Extracellular Vesicles: Exosomes, Microvesicles, and Friends." *Journal of Cell Biology* 200, no. 4: 373–383. https://doi.org/10.1083/jcb.201211138.

Shami-Shah, A., M. Norman, and D. R. Walt. 2023. "Ultrasensitive Protein Detection Technologies for Extracellular Vesicle Measurements." *Molecular & Cellular Proteomics* 22, no. 6: 100557. https://doi.org/10.1016/j.mcpro.2023.100557.

Shetgaonkar, G. G., S. M. Marques, C. E. M. DCruz, R. J. A. Vibhavari, L. Kumar, and R. K. Shirodkar. 2022. "Exosomes as Cell-Derivative Carriers in the Diagnosis and Treatment of Central Nervous System Diseases." *Drug Delivery and Translational Research* 12, no. 5: 1047–1079. https://doi.org/10.1007/s13346-021-01026-0.

Tang, K., J. S. Page, and R. D. Smith. 2004. "Charge Competition and the Linear Dynamic Range of Detection in Electrospray Ionization Mass Spectrometry." *Journal of the American Society for Mass Spectrometry* 15, no. 10: 1416–1423. https://doi.org/10.1016/j.jasms.2004.04.034.

Teng, F., and M. Fussenegger. 2020. "Shedding Light on Extracellular Vesicle Biogenesis and Bioengineering." *Advanced Science (Weinh)* 8, no. 1: 2003505. https://doi.org/10.1002/advs.202003505.

Ter-Ovanesyan, D., T. Gilboa, B. Budnik, et al. 2023. "Improved Isolation of Extracellular Vesicles by Removal of Both Free Proteins and Lipoproteins." *eLife* 12: e86394. https://doi.org/10.7554/eLife.86394.

Ter-Ovanesyan, D., M. Norman, R. Lazarovits, et al. 2021. "Framework for Rapid Comparison of Extracellular Vesicle Isolation Methods." *eLife* 10: e70725. https://doi.org/10.7554/eLife.70725.

Thery, C., K. W. Witwer, E. Aikawa, et al. 2018. "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." *Journal of Extracellular Vesicles* 7, no. 1: 1535750. https://doi.org/10.1080/20013078.2018.1535750.

Welsh, J. A., D. C. I. Goberdhan, L. O'Driscoll, et al. 2024. "Minimal Information for Studies of Extracellular Vesicles (MISEV2023): From Basic to Advanced Approaches." *Journal of Extracellular Vesicles* 13: e12404. https://doi.org/10.1002/jev2.12404.

Wik, L., N. Nordberg, J. Broberg, et al. 2021. "Proximity Extension Assay in Combination With Next-Generation Sequencing for High-Throughput Proteome-Wide Analysis." *Molecular & Cellular Proteomics* 20: 100168. https://doi.org/10.1016/j.mcpro.2021.100168.

You, Y., S. Muraoka, M. P. Jedrychowski, et al. 2022. "Human Neural Cell Type-Specific Extracellular Vesicle Proteome Defines Disease-Related Molecules Associated With Activated Astrocytes in Alzheimer's Disease Brain." *Journal of Extracellular Vesicles* 11, no. 1: e12183. https://doi.org/10.1002/jev2.12183.

Zhang, Y., S. A. Sloan, L. E. Clarke, et al. 2016. "Purification and Characterization of Progenitor and Mature Human Astrocytes Reveals Transcriptional and Functional Differences With Mouse." *Neuron* 89, no. 1: 37–53. https://doi.org/10.1016/j.neuron.2015.11.013.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.