**Proteomic Analysis of Cerebrospinal Fluid: Towards Identification of Unique Markers for Brain-Derived Extracellular Vesicles**

Maia Norman\*1,2,3, Adnan Shami-shah\*1,2, Sydney C. D’Amaddio1,2, Benjamin G. Travis1,2, Dmitry Ter-Ovanesyan1, Tyler J. Dougan1,2,4, David R. Walt#1,2

1Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, MA, USA

2Department of Pathology, Brigham and Women’s Hospital, Harvard Medical School, Boston, MA, USA

3Department of Psychiatry, Massachusetts General Hospital, Boston, MA, USA

4Harvard-MIT Program in Health Sciences and Technology, Cambridge, MA, USA

\* Denotes equal contribution

# Denotes corresponding author

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 tumor, 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 select candidate markers for Brain-Derived EV immunocapture and internal validation of cell origin, we conducted an unbiased Olink screen, measuring 5416 unique proteins. By identifyingproteins that demonstrate a clear EV peak after fractionation with size exclusion chromatography, we were able to create a searchable database of candidate EV associated markers, both those that are cell-type specific within the brain, and those that are found across cell types and can be used as general EV markers. Using the DeepTMHMM deep learning model to differentiate predicted cytosolic, transmembrane, and secreted proteins, we found that intriguingly, only 10% of predicted transmembrane proteins have an EV peak, highlighting the critical importance of verifying EV associated markers by their elution pattern using methods such as size exclusion chromatography or density gradient centrifugation.

Keywords: Membrane association, Brain-Derived Extracellular Vesicles, Size exclusion chromatography, Proteomics, Olink

Word count: \*\*\*

**Introduction**

Extracellular vesicles (EVs) are nanometer-scale, membrane-bound compartments that contain proteins, RNAs, and metabolites endogenous to their cell of origin1. 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 specific cell types or tumors utilizing proteins that can indicate the cell origin2, annotated as transmembrane to the parent cell. While this has led to some success, such as in the case of monitoring prostate cancer3, 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 secreted4. 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 by size exclusion chromatography (SEC) or density gradient chromatography (DGC)4. Nevertheless, analyzing 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, copious amounts of lipoproteins can mask rare EV-associated proteins, while in the soluble fractions, abundant proteins like albumin create a similar problem5. Furthermore, the high levels of soluble proteins in biofluids, such as albumin, 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 pattern6. 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 4-6. Here, we sought to apply a large-scale unbiased method to identify proteins that can be used for EV immunocapture, and to corroborate EV-brain-cell origin based on internal cargo.

**Methods**

*Human sample preparation:*

One milliliter each of four healthy CSF samples (PrecisionMed) were thawed at room temperature and centrifuged at 2,000 g for 10 minutes. 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 2,000 g for 10 minutes at room temperature. The flow through from this filtration was used for downstream experiments.

*Size exclusion chromatography 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 on 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 fraction collection commenced in 0.5mL increments. When the 1mL 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 fractions 1-6 consist of PBS remaining from the previous washes.

Fractions 6-15 were transferred to separate 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 2,000 g at 4°C until all fractions were concentrated 15-fold. The concentrated fractions were brought to a volume of 97 µL with PBS..76 µL 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.

*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 in *Norman & Ter-Ovanesyan et al. (2021)4.* Abcam (ab79559, clone M38) and Biolegend (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.

*Olink sample analysis:*

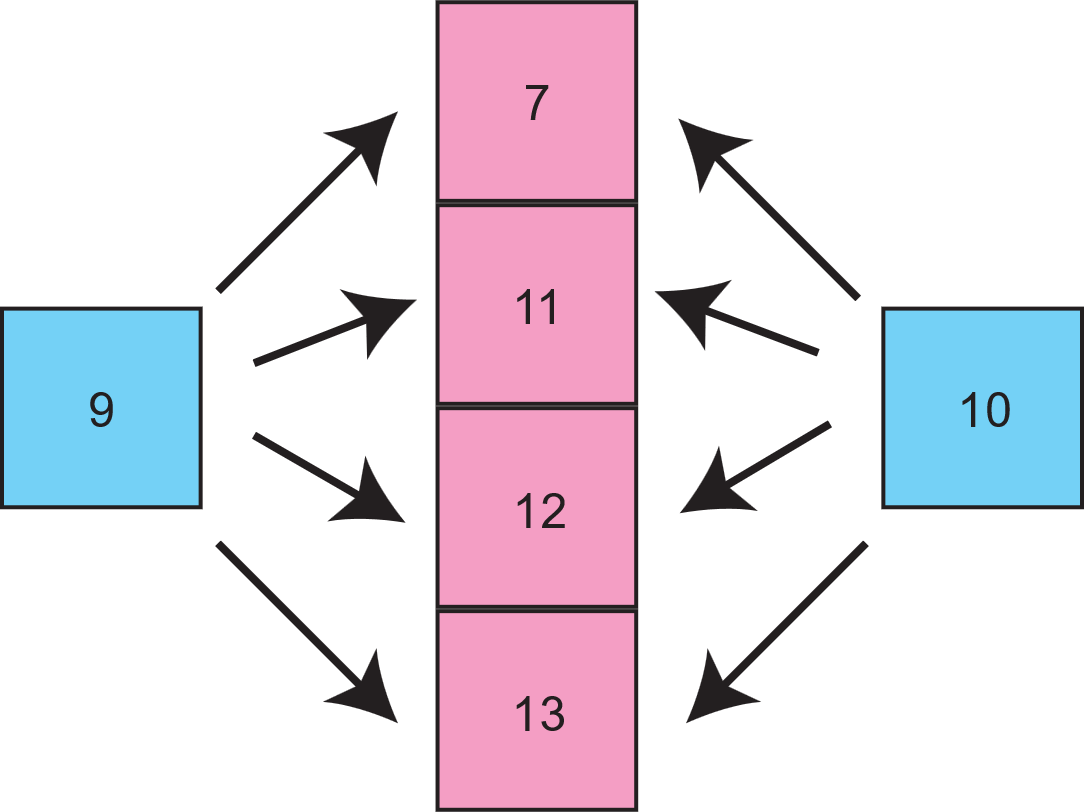
Samples were shipped on dry ice to The Broad Institute in Cambridge, MA for analysis by the Olink HT platform, which measures 5416 unique proteins.

*Data Analysis Methods:*

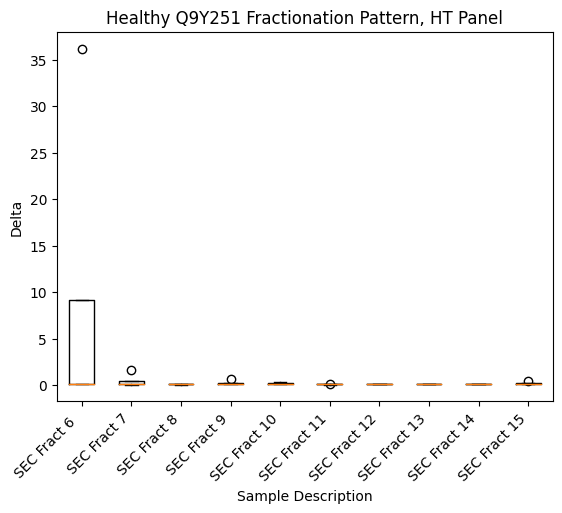
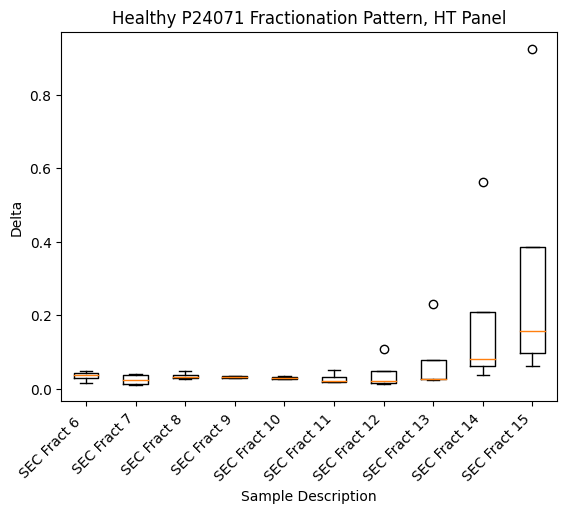
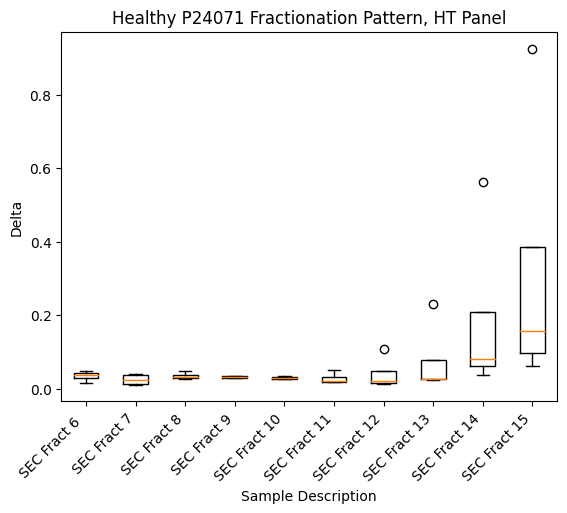
Protein levels were reported in normalized protein expression values (NPX). NPX is an arbitrary unit on a Log2 scale and reflects the relative concentrations of selected proteins in the sample of interest7.All analysis was conducted in Python (version 3.11.5) using Visual Studio Code (Microsoft Corporation, Redmond, Washington). Methods for Olink data pre-processing is available on their website (<https://www.olink.com/content/uploads/2022/04/data-analysis-npx-manager.pdf>), and includes normalization. The HT panel includes 5420 proteins, including 5416 unique proteins and two assays processed in triplicate. No lower limit of detection was reported by Olink, so it is assumed that all data points are above the lower limit of detection and thus given a value greater than zero. The data points were further normalized using Olink’s negative controls and linearized. Two assays that were processed in triplicate were removed from downstream high throughput analysis.

*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 proteins4. For each fraction of interest, the median NPX was calculated for each protein. A protein was considered to have the 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.



If we don’t include fraction 6 in our analysis, these all meet our fractionation criteria:



*Protein Localization:*

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

1. Cytosolic

2. Alpha transmembrane helix

3. Beta transmembrane barrel

4. Signaling peptide

5. External to the cell and any secreted vesicles or exosomes

We classified a protein as internal if all of its amino acids were characterized as cytosolic and we classified a protein as transmembrane if any of the amino acids of the protein were characterized as an alpha transmembrane helix or a beta transmembrane barrel. Proteins were classified as secreted if all of its amino acids were characterized as outside the cell or secreted vesicles.

*EV-Associated Protein Identification:*

This pipeline was used to identify proteins that may be associated with EVs, either as internal cargo or as a transmembrane protein. Proteins were labeled as internal to EVs if they met the fractionation criteria and were identified as internal using DeepTMHMM. The same criteria were followed to identify transmembrane proteins associated with EVs. This yielded a list which was further narrowed by selecting only the 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.

*Cell-Type Specificity:*

The BrainRNA-Seq atlas reports fragments per kilobase per million mapped fragments (FPKM), collected via single-cell RNA sequencing8. The mean FPKM of each gene was used for mature astrocytes, neurons, oligodendrocytes, endothelial cells, and microglia. Fetal astrocytes were excluded from analysis. 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 formula9:

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 supplemental tables 4 and 5.

**Results**

We used a highly multiplexed proximity extension assay platform, Olink, to analyze thousands of proteins from two microliters of biofluid with high specificity10. To assess EVs coming from the brain, we fractionated CSF using SEC to separate proteins that localize to the early EV fractions and late soluble protein fractions from healthy individuals. We analyzed 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)4-6. The remaining 80% of each fraction was analyzed using the Olink HT platform, which quantifies 5416 unique proteins (SI table 1). To further define the EV peak, we analyzed the fractionation pattern of CD63 using data from the Olink assay and demonstrated a clear peak in signal in fractions 9 and 10 (Figure 1b).

To identify targets that could be effective for EV immunocapture or for the analysis of EV cargo, we selected all proteins where the median normalized protein expression (NPX) value across CSF samples was greater in both fractions 9 and 10 compared to fractions 7, 11, 12, and 13 (SI tables 2 and 3). Because many proteins can be found as both 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 peak could be seen. Next, we utilized the DeepTMHMM deep learning model to differentiate cytosolic, transmembrane, and secreted proteins. Running DeepTMHMM on each protein analyzed by the Olink platform we categorized them into 953 predicted transmembrane, 3522 predicted cytosolic, and 941 predicted secreted proteins 11. We demonstrate that 80% of predicted soluble proteins, 10% of transmembrane proteins, and 9% of secreted proteins have an EV peak (Figure 1c).

Our primary interest in using this dataset was to identify proteins that are either transmembrane or internal to EVs, as they 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 >0.758,12, calculated using the mean astrocyte, oligodendrocyte, microglia, neuron, and endothelial cell expression levels. Thus, we identified transmembrane proteins that can potentially be used to isolate cell-type specific EVs from CSF, as well as cytosolic proteins that can be used to confirm cell-type specificity following immunocapture (Figure 2). Finally, we identified a set of internal and transmembrane proteins that peak in the EV fractions but are not specific to a given cell type as defined by a Tau score <0.25 (Supplemental tables 4 and 5, respectively). These latter proteins can be used to normalize total EV quantity.

**Conclusions**

Utilizing the highly sensitive and specific multiplexed Olink platform on SEC-fractionated healthy CSF, we were able to identify dozens of cell-type specific proteins that may be associated with EVs and used both for potential immunocapture and internal analysis of brain-derived EVs. Furthermore, we demonstrate that in our dataset which we hypothesize is due to

Several important caveats are of note: First, although we analyzed 5416 targets, this remains only a quarter of the ~20,000 proteins known to be in the human protein coding genome13. Therefore, our list of proteins that meet EV fractionation criteria is not exhaustive, as there are additional proteins not in the Olink panel that may meet criteria when analyzed. Second, many proteins are known to be both secreted and transmembrane. In some cases, the abundance of the secreted form can mask an EV peak. Without the ability to separate the EV peak, these proteins are likely not useful for EV enrichment unless they have a unique extracellular epitope not present on cleaved and secreted forms 2,4. Third, this dataset was collected using a single method on samples collected from four individuals. Therefore, we cannot definitively claim that proteins that meet our fractionation criteria are EV-associated—orthogonal validation of these targets using a broader collection of samples is necessary to confirm whether these proteins are associated with EVs. Similarly, if a protein does not meet our fractionation criteria, this does not necessarily mean that the protein is not EV-associated. The purpose of this dataset is to identify quality candidates for downstream targeted analysis to validate EV association, as it is not feasible to analyze such a large number of proteins using a targeted approach. We cannot rule out EV association for proteins that do not meet our criteria or were not included in the Olink HT panel. Finally, while proximity extension assays lower the chance of nonspecific binding in ELISAs, the soluble protein fractions have substantially more protein, increasing the chance for nonspecific binding interactions to produce a signal. Thus, our analysis is useful for identifying potential EV-associated proteins but cannot rule out EV association for proteins that do not meet our criteria or that are not included in the Olink HT platform. Nonetheless, this dataset does add credence to the necessity of running SEC or DGC on putative immunocapture targets before proceeding to immunocapture. Substantial additional work is required to assess cell origin and EV association for those proteins that met our criteria displayed in Figure 2. Nevertheless, this dataset is an important new resource for identifying novel targets for brain-derived extracellular vesicles.

**Funding**

This work was supported by funding from Good Ventures/Open Philanthropy (to D. R. W.) 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 D. R. W.). These funding agencies had no role in conceptualization, design, analysis, decision to publish or preparation of the manuscript.

**Author contribution**

MN, AS and DRW conceptualized the study. Experiments were performed by BGT, MN and AS. Data analysis schema was developed by MN, AS, TJD and DT. Analysis code was written by SCD. All authors contributed to manuscript writing.

**Data availability statement**

Full dataset with raw data is included as supplemental file 1.

**Conflict of interest disclosure**

D.R.W. is a founder and equity holder in Quanterix. His interests were reviewed and are managed by BWH and Partners HealthCare in accordance with their conflict of interest policies.

**Patient consent statement**

All human samples were collected commercially in accordance with the IRB. All patients were appropriately consented.

**References**

1 Raposo, G. & Stoorvogel, W. Extracellular vesicles: exosomes, microvesicles, and friends. *J Cell Biol* **200**, 373-383, doi:10.1083/jcb.201211138 (2013).

2 Shami-Shah, A., Norman, M. & Walt, D. R. Ultrasensitive Protein Detection Technologies for Extracellular Vesicle Measurements. *Mol Cell Proteomics* **22**, 100557, doi:10.1016/j.mcpro.2023.100557 (2023).

3 Ramirez-Garrastacho, M. *et al.* Extracellular vesicles as a source of prostate cancer biomarkers in liquid biopsies: a decade of research. *Br J Cancer* **126**, 331-350, doi:10.1038/s41416-021-01610-8 (2022).

4 Norman, M. *et al.* L1CAM is not associated with extracellular vesicles in human cerebrospinal fluid or plasma. *Nat Methods* **18**, 631-634, doi:10.1038/s41592-021-01174-8 (2021).

5 Ter-Ovanesyan, D. *et al.* Improved isolation of extracellular vesicles by removal of both free proteins and lipoproteins. *Elife* **12**, doi:10.7554/eLife.86394 (2023).

6 Ter-Ovanesyan, D. *et al.* Framework for rapid comparison of extracellular vesicle isolation methods. *Elife* **10**, doi:10.7554/eLife.70725 (2021).

7 Diorio, C. *et al.* Proteomic profiling of MIS-C patients indicates heterogeneity relating to interferon gamma dysregulation and vascular endothelial dysfunction. *Nat Commun* **12**, 7222, doi:10.1038/s41467-021-27544-6 (2021).

8 Zhang, Y. *et al.* Purification and Characterization of Progenitor and Mature Human Astrocytes Reveals Transcriptional and Functional Differences with Mouse. *Neuron* **89**, 37-53, doi:10.1016/j.neuron.2015.11.013 (2016).

9 Kryuchkova-Mostacci, N. & Robinson-Rechavi, M. A benchmark of gene expression tissue-specificity metrics. *Brief Bioinform* **18**, 205-214, doi:10.1093/bib/bbw008 (2017).

10 *PEA – a high-multiplex immunoassay technology with qPCR or NGS readout*, <<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>> (2020).

11 Hallgren, J. *et al.* DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks. *bioRxiv*, 2022.2004.2008.487609, doi:10.1101/2022.04.08.487609 (2022).

12 Camargo, A., Vasconcelos, A., Fiamenghi, M., Pereira, G. & Carazzolle, M. (Research Square, 2020).

13 Aebersold, R. *et al.* How many human proteoforms are there? *Nat Chem Biol* **14**, 206-214, doi:10.1038/nchembio.2576 (2018).

**Figure 1: CSF SEC fractionation as a measure of EV association**

1a. Quantification of CSF fractions using a Simoa assay for CD81. Four samples each run with 2 technical replicates. Bar graphs indicate median of the 4 samples while the error bars represent standard deviations.

1b. Quantification of CSF fractions using the Olink assay for CD63. Four samples each run with 2 technical replicates. Bar graphs indicate median of the 4 samples while the error bars represent standard deviations.

1c. Percentage of Deep TMHMM predicted transmembrane, internal, and secreted targets quantified by Olink as having an EV peak in CSF.

1a.



1b.



1c.

A comparison of a pie chart

Description automatically generated

**Figure 2: Cell-type specificity of proteins that peak in the EV fractions**

Calculated Tau Score and EV Association Score for each identified transmembrane (red) and internal (blue) protein that peaked in the EV fractions for a. Astrocytes b. Microglia c. Neurons and d. Oligodendrocytes.

A collage of graphs

Description automatically generated