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

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 transmembrane proteins that can allude to cell origin2. 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.

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 challenging 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, abundant lipoproteins can mask rare EV-associated proteins, while in the soluble fractions, abundant proteins like albumin create a similar problem5. Furthermore, protein-rich fractions can be challenging to analyze by Western blots because the high abundance of free proteins can clog the gel. As a result, ELISAs have thus far been the best method of assessing EV fractionation pattern6. To optimize total EV isolation, our lab has previously developed highly sensitive ELISA assays to quantify both canonical EV proteins (CD9, CD63, CD81, Alix) and potential contaminants to EV preparations (Apolipoprotein B, albumin) using the Simoa platform invented in our lab4-6. Here, we sought to apply a large-scale unbiased method to identify proteins that can be used for EV immunocapture and corroboration of EV-brain-cell origin based on internal cargo.

To that end, we used a highly multiplexed proximity extension ELISA platform, Olink, to analyze thousands of proteins from microliters of biofluid with high specificity7. To assess EVs coming from the brain, we fractionated cerebrospinal fluid (CSF) using SEC to separate proteins that localize to the EV fractions (7-10) and soluble protein fractions (>13) from four healthy individuals without neurologic or psychiatric disease (see online methods). 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 that the majority of signal comes from 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 whose median value across four CSF samples was greater in fractions 9 and 10 compared to fractions 7, 11, 12, and 13 (SI tables 2 and 3). As many proteins can be found as both EV-bound and soluble isoforms, we did not choose proteins based on being greater in the early (7-10) compared to late fractions (>13), but rather selected proteins where a definable EV peak could be seen. Next, we utilized the DeepTMHMM deep learning model to predict the localization of each protein analyzed by the Olink platform and identified 953 predicted transmembrane, 3522 predicted cytosolic, and 941 predicted extracellular proteins 8. We demonstrate that 80% of predicted internal markers, 10% of transmembrane targets, and 9% of external targets have an EV peak (Figure 1c). The fact that only 10% of predicted transmembrane proteins had a definable EV peak is likely due to overwhelming signal from cleaved or secreted isoforms of these proteins. This data adds credence to the necessity of running SEC or DGC on putative immunocapture targets before proceeding to immunocapture.

Our primary interest from this dataset was in analyzing proteins predicted to be transmembrane or internal, as they can be used to isolate or define an EV’s cell of origin. To that end, 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 >0.759,10, calculated using the mean astrocyte, oligodendrocyte, microglia, neuron, and endothelial cell expression. Thus, we identified transmembrane proteins that can potentially be used in CSF to isolate cell-type specific EVs using antibodies 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 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 useful for the normalization of total EV quantity.

Utilizing the highly sensitive and specific multiplexed Olink platform on fractionated healthy CSF, we were able to identify dozens of cell-type specific EV-associated proteins both for potential immunocapture and internal analysis of brain-derived EVs. Several important caveats are of note: First, while we analyzed 5416 proteins, this remains only a quarter of the ~20,000 known to be in the human proteome11. 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 analysis unless they have a unique extracellular epitope not present on secreted forms2,4. 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 were not included in the Olink dataset. Substantial additional work is required to assess cell origin and EV association for those proteins that did meet our criteria displayed in Figure 2. Nevertheless, this dataset is an important step toward validating novel immunocapture targets for brain-derived extracellular vesicles to enable a liquid brain biopsy.

**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 *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).

8 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).

9 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).

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

11 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 deviation.

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 deviation.

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

1a.



1b.



1c.

A close-up 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