Supplementary Online Content

Kapogiannis D, Mustapic M, Shardell MD, et al. Association of extracellular vesicle biomarkers with Alzheimer disease in the Baltimore Longitudinal Study of Aging. *JAMA Neurol*. Published online July 15, 2019 doi:10.1001/jamaneurol.2019.2462

This supplementary material has been provided by the authors to give readers additional information about their work.

eMethods.

EV isolation

Blood samples were collected by venipuncture into vacutainer EDTA tubes, incubated for 10 min at room temperature, centrifuged for 15 min at 2500 g at 25°C, re-aliquoted in .5-ml aliquots and stored at -80ºC. Samples were thawed to room temperature once before processing. We defibrinated .5 ml plasma samples with 200μl of Thromboplastin-D (Pacific Hemostasis™ Thromboplastin-D Thermo Fisher Scientific cat. #100356, Hanover Park, IL), followed by 30 minutes incubation at room temperature (RT). We added 300μl of Dulbecco's calcium- and magnesium-free salt solution premixed with 3 times more than the suggested concentrations of protease inhibitor cocktail (Complete Tablets Easy pack, Roche Applied Sciences, Inc., Indianapolis, IN) and phosphatase inhibitor cocktail (Pierce Halt, Thermo-Fisher Scientific, Inc., Rockford, IL), incubated 5 minutes at RT and then centrifuged at 3,000 x g for 20 minutes at 4°C. After transferring the supernatant into clean Eppendorf tubes, we precipitated total EVs using a high-throughput and high efficiency [1] particle precipitation method adding 252μl of Exoquick® exosome solution (System Biosciences, Inc., Mountainview, CA) and gently mixing by inversion. After 60 min incubation at 4°C, suspensions were centrifuged at 1,500 x g for 20 minutes at 4°C to acquire pellets containing total EVs. Supernatants were discarded and pellets were recentrifuged at 1,500 x g for 5 min and residual supernatants were removed. The final pellets were re-suspended in .5ml of Ultra-pure distilled water (Invitrogen-Thermo-Fisher Scientific, Inc., Rockford, IL) containing 3 times the suggested concentrations of protease and phosphatase inhibitors. Next, we performed positive selection for neuronal origin by immunoprecipitation with antibodies against L1 Cell Adhesion Molecule (L1CAM), given its high and relatively specific expression in neural tissues and early research demonstrating that it is highly expressed on exosomes derived from cultured neurons [2]. Suspensions were incubated for 1 hour on a rotational mixer at 4°C with 4 μg of mouse anti-human CD171 [L1CAM biotinylated antibody (clone 5G3, eBioscience, San Diego, CA)] in 3% BSA [1:3.33 dilution of Blocker BSA 10% solution in PBS (Thermo Scientific, Inc.)] to a final added volume of 50μl; all samples were processed with the same batch of the CD171 clone 5G3 to avoid variability in antibody affinity. Next, suspensions received 15 μl of streptavidin-agarose Ultralink resin (Thermo Scientific, Inc.) in 25 μl of 3% BSA and were further incubated for 30 min at 4°C on a rotational mixer with continuous gentle mixing. Pellets were re-suspended in 200 μl of .1 M glycine-HCl, solutions were vigorously mixed for 10 sec and centrifuged at 4,500 x g for 5 min at 4°C to detach L1CAM+ EVs from the bead-antibody complex. After adjusting the pH to 7.4 with 1 M TRIS-HCl, samples were centrifuged at 500 x g

at 4°C for 5 min, and supernatants containing L1CAM+ EVs were transferred to clean tubes containing 25 μl of 3% BSA and 15 μl of 1 M TRIS-HCl (to neutralize the pH) and mixed. To lyse L1CAM+ EVs, each tube received 260 μl of M-PER mammalian protein extraction reagent (Thermo Scientific, Inc.), containing 3 times the suggested concentrations of protease and phosphatase inhibitors and underwent 2 freeze thaw cycles. Final suspensions (500 μl) were separated into 100 μl aliquots and stored at -80°C until immediately prior to assays.

EV characterization by Nanoparticle Tracking Analysis

We performed Nanoparticle tracking analysis (NTA) in all neuronal-enriched EV preparations in this study. Ten microliters were taken from the extracellular vesicle suspensions prior to lysis with M-PER and were diluted 1:1600 (for total EVs) or 1:200 (for the more dilute L1CAM+ EVs) to reach the optimal quantification range of 3–15 × 10^8/ml and 20-100 of particles per frame. Particle concentration and size distribution (diameter in nm) of the isolates was measured with Nanosight NS500 instrument (Malvern Instruments Ltd, UK). The particles were visualized by their scattering of a focused laser beam and the collection of the scattered light by a standard optical microscope fitted with a CCD video camera. Five exposures of 20 seconds each were recorded from fields chosen randomly by a computer operating with NanoSight software (NanoSight NTA 3.2), which also calculated the EV concentration and average diameter.

Supplemental Figure 1 displays NTA graphs of particles isolated from plasma by Exoquick® alone (total EVs), and neuronal-enriched EVs isolated through Exoquick® followed by immunoprecipitation with antibodies against L1CAM (L1CAM+ EVs).

EV characterization by Electron Microscopy

After biotin elution, L1CAM+ EVs were suspended intact in a coated tube containing ~200 µL of a PBS-based solution with neutral pH. For negative stain, we used 400 mesh carbon coated copper grids (EMS) floated on EVs in suspension. After 2 minutes the grids were rinsed in dH2O, stained in filtered 2% uranyl acetate (EMS) for 1 minute, blotted dry with Whatman #1 filter paper and viewed. All TEM viewing was performed on a Zeiss Libra 120 at 120KV, with an EMSIS Veleta camera. EM images of negative stained EVs are shown in Supplemental Figure 2.

EV characterization for EV and neuronal markers

To confirm that neuronal-enriched EVs carry typical EV markers we performed western blots (WB) of L1CAM+ EVs, total EVs (positive control), and EV-depleted plasma (negative control) from plasma of three participants (Figure 1B). EV-depleted plasma was supernatant after removing EV pellet after Exoquick®; total EVs were re-suspended total EVs isolated with Exoquick®. We used a Bradford assay (Cat# 23200; ThermoFisher Scientific, Houston, TX, USA) to determine total protein concentration per mL of sample and adjusted the dilution of sample loaded per well (20µl) to load the same amount of total protein. We used 4-12% NuPAGE Bis-Tris Mini gels (Cat# NPO322BOX; ThermoFisher Scientific, Houston, TX, USA) and probed for GM130 (as negative control) (Cat#ab52649; Abcam, Cambridge, MA, USA), ApoA1 (to show relative depletion of lipoproteins) (Cat#ab7613; Abcam, Cambridge, MA, USA), Alix (Cat# NBP1-90201; Novus Biologicals, Littleton, CO, USA) (as positive control, component of the ESCRT system, theoretically intravesicular), and tetraspanins (as positive controls, theoretically transmembrane) CD9 (Cat# EXOAB-CD9A-1; SBI System Biosciences, Mountain View, CA, USA), and CD81 (Cat# EXOAB-CD81A-1; SBI System Biosciences, Mountain View, CA, USA). The WBs were developed using LiCOR QuickWestern Kit (Cat# P/N 926-68100; LiCor Biosciences, Lincoln, NE, USA) and read with the Odyssey CLx and LiCOR ImageStudio software (LiCor Biosciences, Lincoln, NE, USA). The results confirm the presence of three EV markers in the L1CAM+ EV preparations, and the relative depletion of GM130 (suggesting the absence of cellular fragments/apoptotic bodies) and Apolipoprotein A1 (suggesting the relative depletion of lipoproteins from the final preparation).

In a previously published Methods paper, we showed that L1CAM+ EVs are enriched for several neuronal markers including p181-tau, neuron-specific enolase, microtubule associated protein 2 (MAP2), neuron-specific class III β-tubulin (TuJ1), p35, light neurofilaments, brain derived neurotrophic factor (BDNF), proBDNF, neuronal cell adhesion molecule NCAM and L1CAM; the degree of enrichment compared to total plasma EVs and/or a general EV sub-population ranges from 1.6-fold to 5.6-fold depending on the marker [3]. Since then, we provided further evidence for neuronal enrichment in two additional papers [4,5]. To further confirm that EVs immunoprecipitated with antibodies against L1CAM are enriched for neuronal origin compared to control EV subpopulations, we performed an ELISA for ELISA for Neurofilament light (NF-L) [Uman Diagnostics AB, Umea, Sweden (distributed by IBL International)]. We compared EVs immunoprecipitated with anti-L1CAM antibody, with EVs immunoprecipitated with anti-GLAST antibody (of presumed astrocytic origin),[6,7] EVs immunoprecipitated with anti-CD81 antibody (of

variable cellular origin), and EV-depleted plasma (supernatant after Exoquick®) for N = 4 participants (Supplemental Figure 3). EV-depleted plasma and CD81 EVs for 3 out of 4 participants were below the Lowest Limit of Quantification (LLoQ) for the assay. The results confirmed that L1CAM EVs have higher NF-L levels compared to GLAST EVs (Supplemental Figure 4).

In addition, to assess the degree of enrichment in terms of fold-diffidence in the levels of additional neuronal markers we compared L1CAM+ EVs with total plasma EVs for levels of Synaptophysin (by WB and ELISA), NCAM (by immunoblot), and L1CAM (Supplemental Figure 5). For the WB comparison (Supplemental Figure 5A, B), total EVs and L1CAM+ EVs were loaded on the NuPAGE 4-12% Bis-Tris Gel (Cat# NPO322BOX; ThermoFisher Scientific, Houston, TX, USA) and the membrane was probed with mouse monoclonal anti-Synaptophysin antibody (S5768, Sigma) and rabbit polyclonal anti-Alix antibody (NBP1-90201, Novus) that were fluorescently labeled using LI-COR quick western labeling kit (IRDye® 680RD). Total EVs were diluted six times compared to L1CAM+ EVs based on total protein amount. As a positive control, we included 4ug of rat synaptic vesicles and 8ug of rat brain lysate (SYSY; 501-TOT). Synaptophysin protein concentration was also determined using human synaptophysin (SYP) ELISA (CSB-E17406h, Cusabio) (Supplemental Figure 5C). Finally, neuronal markers L1CAM and NCAM were identified using Exo-Check™ Exosome Antibody Array (Neuro) (System Biosciences, Inc., Mountainview, CA). Total EVs and L1CAM+ EVs were isolated from plasma of 3 healthy Controls and were pooled to acquire sufficient total protein amount to load on a neuro array. Signal was detected using Chemiluminescence (ECL) solution and were quantified with the LiCOR ImageStudio software (LiCor Biosciences, Lincoln, NE, USA) (Supplemental Figure 5D). The degree of neuronal marker enrichment ranged from 4.2 – 8.2 depending on the marker and the technique.

Quantification of EV biomarkers by immunoassays (additional information)

We used Mesoscale Discovery (MSD®, Meso Scale Diagnostics, Rockville, MD) electrochemiluminescence assays to quantify total-tau (K151LAE), p181-tau (N45CB-1), p231-tau (K15121), pSer312-IRS-1 (K150HLD) and pY-IRS-1 (panTyr) (N45CA-1), and TSG101. MSD® assays require only 50 μl of sample for duplicate quantifications, which made it possible to quantify more analytes with a given sample volume, and offer greater sensitivity compared to ELISAs used in previous studies [8,9]. The MSD® total tau assay, similar to all commercially available “total tau” assays, likely detects mid-region tau [10]; however, the precise antibodies were

not disclosed by the manufacturer. The pSer312-IRS-1 and pY-IRS-1 assays share the same capture antibody but employ different detection antibodies. TSG101 is an EV marker characterizing primarily endosomal-origin EVs (i.e. exosomes) [11]. The TSG101 assay was developed in-house using MSD GOLD Streptavidin (Meso Scale Diagnostics, Rockville, MD) coated plates, anti-TSG101 (ab133586, Abcam, Cambridge, MA) as capture antibody and anti-TSG101 (H00007251-B01P, Novus Biologicals LLC, Littleton, CO) as detection antibody. The capture antibody was biotinylated prior to plate coating using EZ-Link™ Sulfo-NHS-LC-Biotin, No-Weigh™ Format (Thermo Fischer Scientific, Waltham, MA). Given the relatively low levels of Aβ42 previously observed [8], we opted to quantify Aβ42 using the SIMOA® assay (Simoa™ Aβ42 2.0 Kit, # 101664), which provides an order of magnitude higher sensitivity than MSD® or ELISA assays. The MSD® plates were read with MESO QuickPlex SQ120 imager and values were calculated using the MSD discovery workbench Software 4.0 (Meso Scale Diagnostics, Rockville, MD). The Aβ42 plates were read with SIMOA® HD-1 Analyzer and values were calculated using Simoa HD-1 Analyzer Software version 1.5.

The optimum dilution for each assay was determined using serial dilutions of test samples. For total-tau, lysed EV suspensions were diluted 1:10 with MSD diluent 35; for Aβ42, lysed EV suspensions were diluted 1:3 with kit sample diluent; for TSG101, lysed EV suspensions were diluted 1:2.5 with EV lysis buffer (M-PER); for p181-tau and p231-tau, pSer312-IRS-1 and pY-IRS-1 undiluted EV suspensions were used. EV proteins were quantified by calculating the standard curve separately for each plate using standards provided by the manufacturer for total tau, p231-tau (phosphorylated full length 441 tau), and Aβ42. For the home-made TSG101 assay, we used human TSG101 recombinant protein (H00007251-P01, Novus Biologicals LLC, Littleton, CO). For p181-tau, in collaboration with Meso Scale Diagnostics scientists, we applied the same standard as for p231-tau to calculate a standard curve, since for its production, global phosphorylation of the full length 441 tau protein at multiple residues was induced. The fitted values for p181-tau have arbitrary units expressing % phosphorylation in relation to the electrochemiluminescence signal of phosphorylated full length 441 tau (Supplemental Figure 6A and Supplemental Table 2). The lowest limit of quantification (LLOQ) (defined as the concentration of the standard with i) signal above the mean of the blank plus 9 SD of the blank, ii) Coefficient of Variability (CV) among duplicates < 20%, and iii) recovery >80% and <120%) was calculated for each plate and the highest LLOQ was used as the global LLOQ for the analyte for all subsequent analyses. Limit of detection (LOD) was defined as mean of the blank plus 2.5 SD of the blank. The LLOQ was 118.52 ng/ml for total tau; 3.12 A.U. (113.69 A.U. correspond to 1µg of phosphorylated full length 441 tau) for p231-tau; .39% signal of undiluted phosphorylated

full length 441 tau standard for p181-tau (Supplemental Figure 6A); .412 pg/ml for Aβ42; and 5ng/ml for TSG101 (Supplemental Figure 6B). For the pSer312-IRS-1 and pY-IRS-1 phospho-assays, no standards were provided; therefore, standard curve and LLOQ could not be calculated and the electrochemiluminescence signal was used for the analysis. The average Intra-Assay CVs were 11.03 % (total-tau), 4.95% (p181-tau), 3.59% (p231-tau), 6.66% (pY-IRS-1), 8.47% (pSer312-IRS-1), 1.91% (TSG101) and .12% (Aβ42).

Quality control measures

The following quality control measures were undertaken as part of this study.

i) **Censoring criteria based on LLOQ and %CV.** Six (6) samples for p181-tau, 165 samples for total tau, 7 samples for p-231tau, 34 samples for Aβ42 and 14 samples for TSG101 had %CVs ≥ 20% and were excluded from the analysis independent of their levels. For total-tau and p181-tau all samples were above the LLOQ and within the linear range of the standard curve. For p-231tau, 201 samples were below the LLOQ; for Aβ42, 64 samples were below the LLOQ; and for TSG101, 32 samples were below the LLOQ. If their values were above the LOD (.137 pg/ml for Aβ42, .78 A.U. for p231-tau and 5 ng/ml for TSG101) and they had CV < 20%, these samples were considered truly low and were assigned the LLOQ value (42/64 samples for Aβ42, 198/201 samples for p231-tau and 32/32 samples for TSG101), whereas if CV ≥ 20%, they were excluded from the analysis (22/64 samples for Aβ42 and 3/201 samples for p231-tau).

   For pSer312-IRS-1 and pY-IRS-1, given that no standard curve was available, quality control had to rely exclusively on CVs. Therefore, for these markers, we adopted a more stringent threshold and all samples with electrochemiluminescence signal CV ≥ 15% were excluded from the analysis (130 and 70 samples respectively).

ii) **Between plates variability.** To assess plate to plate variability for all assays, we included two internal standards in all plates (EVs from a Control participant); between plate CVs were under 20% for all assays. In addition, for the Aβ42 assay, we used two internal quality controls of low (2pg/ml) and high (60 pg/ml) Aβ42 concentration provided by the manufacturer; CVs for these quality controls were under 15%.

iii) **Within-Operator variability for assays.** To assess day to day variability for assay performance for total tau and Aβ42 a subset of samples was measured twice on two different days by the same operator. Coefficient of determination (R^2) for total tau (supplemental Figure 7A) was .852 ($p < .0001$), for Aβ42 (supplemental Figure 7B) it was

.981 ($p < .0001$).

iv) **Within-Operator variability for the entire methodology.** To assess day to day variability for EV isolation followed by biomarker quantification, L1CAM+ EVs were isolated twice from a subset of samples and p181-tau and pSer312-IRS-1 were measured in both isolated replicates by the same operator. Coefficient of determination (R^2) was .751 ($p < .0001$) for p181-tau (Supplemental Figure 7C) and .898 ($p < .0001$) for pSer312-IRS-1 (Supplemental Figure 7D).

Between-Operator variability for the entire methodology (Supplemental Figure 8). To assess combined variability between and within operators, three different lab members performed EV isolation followed by biomarker quantification in two biological replicates for four test subjects. For each one of the four test subjects, six plasma aliquots were thawed, mixed and divided into six new aliquots. Each of the three lab members received two of these aliquots for four test subjects (2 * 4 = 8), two negative controls (PBS) and a negative control spiked with a test EV sample (PBS-sp). Each lab member performed EV isolation separately and measured p181-tau and TSG101 in those samples. Coefficients of determination (R^2) were .98 (1^{st} vs. 2^{nd}), .96 (1^{st} vs. 3^{rd}) and .98 (2^{nd} vs. 3^{rd}) between operators for p181-tau and .87 (1^{st} vs. 2^{nd}), .92 (1^{st} vs. 3^{rd}) and .94 (2^{nd} vs. 3^{rd}) for TSG101. Coefficients of variation (%CV) were 11.59 for p181-tau and 11.62 for TSG101.

Statistical Analyses (additional information)

We performed a random split of the data to create a training set (2/3 of BLSA data) and a test set (1/3 of BLSA data). Wilcoxon rank-sum tests and side-by-side boxplots were used to compare individual EV biomarker levels between future AD and Control participants. For each biomarker, we compared values from the last visit prior to symptom onset, within-person averages and slopes over time. Within-person slopes were computed using linear mixed-effects models with log-transformed biomarkers and random intercepts and slopes. To visualize age-specific retrospective longitudinal biomarker changes for future AD and Control participants, we performed locally weighted scatterplot smoothing (LOESS) regression and plotted locally weighted sum of squares smoothing splines. LOESS was selected since it is a non-parametric exploratory approach that fits a smooth curve without requiring pre-specification of the functional model form and provides some protection from model mis-specification. The biomarkers were skewed, therefore they were log-transformed for analyses, and the predicted values were back-transformed in the figures.

Using data from 887 samples from 350 BLSA participants, we used mixed-models to compute person-specific slopes and means for each log-transformed nEV biomarker. 241 BLSA participants with complete information on all predictors (complete demographic data and mean, slope, and last visit data for all nEV biomarkers) at ≥1 visit contributed to model building. nEV biomarkers at the last visit, within-person mean, and within-person slope were selected as candidates to determine which aspects of a participant's biomarker history are most predictive of AD (the latest available value, the long-run average, or the rate of change) while keeping the number of candidates manageable and accommodating participants with different history lengths, similar to a clinical context. We compared participants who were included in prediction model building to those who were excluded with respect to AD status, sex, age, and year of birth. Continuous variables were compared using t-tests, dichotomous variables were compared using Fisher's exact tests (Supplemental Table 8).

We performed a random split of the data to create a training set (2/3 of BLSA data; 161 participants) and a test set (1/3 of BLSA data; 80 participants). In the training set, we performed stepwise logistic regression with internal cross-validation and receiver operating characteristic analysis to identify a model discriminating future AD cases from Controls; model fit was assessed in the test set. To assess how EV biomarkers perform in AD prediction individually and collectively, we built 10 models to predict AD diagnosis as functions of the following predictors: Model 1) age, sex, and sample type (to account for the fact that 83 samples were serum rather than plasma); Model 2) Model 1 predictors plus measures of EV concentration and average diameter (to assess whether these NTA parameters inform AD classification in their own right and as a normalizers for differential EV yield for subsequent models); Model 3) Model 2 predictors plus measures of TSG101; Model 4) Model 2 predictors plus measures of total tau; Model 5) Model 2 predictors plus measures of p231-tau; Model 6) Model 2 predictors plus measures of p181-tau; Model 7) Model 2 predictors plus measures of pY-IRS1; Model 8) Model 2 predictors plus measures of pSer312-IRS1; Model 9) Model 2 predictors plus measures of Aβ42; Model 10) Model 2 predictors plus the most predictive measures of individual EV biomarkers. We considered 12 measures of each biomarker as candidate predictors: last preclinical visit measurement, within-individual average from first until the last preclinical measurement; within-individual slope (rate of change) from first until the last preclinical measurement; interactions of the first three measures (last visit, average, slope) with sex; interactions of the first three measures with age; and interactions of the first three measures with age and sex. Thus, for each biomarker, we considered 4,095 (2^{12} -1) combinations of measures. All biomarkers were log-transformed; age was centered at 81 years.

Model 1 was fit using logistic regression to appropriately handle the case-control design; we computed area under the receiver operating characteristic (ROC) curve; and we performed internal leave-10%-out cross-validation to compute a cross-validated area under the curve (AUC). We used a similar approach for Model 2, except that we first selected the measures of EV concentration that maximized cross-validated AUC (cvAUC) given Model 1 predictors; then given EV concentration functions, we selected the measures of average EV diameter that maximized cvAUC. Lastly, we performed a reduction step to determine whether a submodel produced a higher cvAUC. Single-protein models (Models 3 to 9) were built by selecting the set of protein measures that maximized cvAUC given the Model 2 predictors; a reduction step determined whether a submodel produced a higher cvAUC. Model 10 was built by first identifying the set of single-protein model measures that maximized cvAUC, then recursively adding proteins until cvAUC no longer increased; once again, a reduction step determined whether a submodel produced a higher cvAUC. By this process, Model 10 ended up including functions of age, sex, sample type, EV concentration, EV average diameter, TSG101, total tau, pY-IRS1, pSer312-IRS1, p181-tau, and p231-tau (but not Aβ42). To visualize and compare the models' ability to discriminate between participants with and without AD, we plotted ROC curves for each model; we also included side-by-side boxplots of participants' risk scores. Model 10 performance was further evaluated by selecting the threshold risk score that minimizes the distance from the ROC curve from the top left corner and computing sensitivity (proportion above the threshold risk score among AD cases), specificity (proportion below the threshold risk score among Controls), and odds ratio to address the case-control design. Performance statistics (ROC, sensitivity, specificity, odds ratio) were calculated separately for the BLSA training set, the BLSA test set and for another cohort, JH ADRC training and test sets.

We chose to optimize cvAUC in the BLSA to avoid over-fitting and enhance validity. Nonparameteric tests compared AUC and cvAUC between models [12]. Models were built and compared using R Statistical Software version 3.4.0 (https://www.R-project.org/). Cross-validated estimates were derived by implementing logistic regression using the SuperLearner package with stratified sampling [13]. ROC and AUC analysis was performed using the pROC package [14]. P-values are shown without adjustment for multiple comparisons. A Bonferroni correction can be implemented by multiplying p-values by the number of tests for adjustment (23 tests using estimates derived from all data and 23 tests using cross-validated estimates), or by dividing a chosen significance level by the number of tests (e.g., $.05/23 = 2.17 \times 10^{-3}$).

The biomarkers selected from analysis of BLSA training data were externally validated in the JH ADRC cohort. Owing to differences in the BLSA and JH ADRC study designs and

participant characteristics, we split the JH ADRC cohort into training (2/3 of participants) and test sets (1/3 of participants). We fit the model using the training data and assessed its performance using test data.

EV biomarker associations with cognition

In the BLSA, we sought to determine cross-sectional and longitudinal associations between EV biomarkers and verbal memory, executive function, attention, language, and visual spatial function. Verbal memory was defined as the composite z scores of California Verbal Learning Test (CVLT) learning/immediate free recall and CVLT long delayed recall. Executive Function was defined as the composite z score of Trail Making Part B and Digit Span backwards tests. Attention was defined as the composite z score of Trail Making Part A and Digit Span forwards tests. Language was defined as the composite z score of letter fluency and category fluency tests. Visuospatial function was defined as the composite z score of the Card Rotations Test and Clock Drawing to Command. The z-scores were computed based on this sample's mean and standard deviation at the first preclinical visit. We applied linear-mixed models with cognitive measure as the dependent variable and log-transformed values of EV Aβ42, p181-tau, p231-tau, total tau, pSer312-IRS1 and pY-IRS1 at the first preclinical visit were used as predictors; Baseline Age, $(\text{Baseline Age})^2$, sex, race (white/non-white), and education years were used as covariates. (The models were fit separately for each cognitive measure and EV predictor.) Fixed effects included Baseline age, $(\text{Baseline Age})^2$, sex, race, years of education, time from baseline in years, and interactions of time with Baseline Age, race, sex, and years of education. Random effects included intercept and time interval with unstructured covariance.

eTable 1. BLSA and JH ADRC Study sample characteristics

BLSA	**Full cohort**	**Future AD participants (preclinical)**	**Control participants**	**p-value**
Number of participants (%)	350	128 (36.57)	222 (63.43)	
Number of person-visits (%)	887	304 (34.27)	583 (65.73)	
Number of women participants (%)	178 (50.86)	68 (53.13)	110 (50.45)	0.519
Number of non-White participants, n (%)	58 (16.57)	11 (8.59)	47 (21.17)	.0023
Education (yrs), mean (SD)	16.84 (2.54)	16.82 (2.48)	16.85 (2.58)	.913
Baseline age (yrs), mean (SD), range	77.26 (7.36) 56.51- 101.83	79.09 (7.02) 58.30-101.83	76.20 (7.36) 56.51-92.06	.0004
Follow-up time from first biomarker assessment to last biomarker assessment (yrs), median, (IQR), [range]	4.00 (2.13, 4.91) [0.00, 9.73]	3.95 (2.16, 4.89) [0.00, 9.73]	4.03 (2.10, 4.91) [0.00, 9.20]	.64
Follow-up time from first biomarker assessment to AD symptom onset (yrs), median, (IQR), [range]*		4.07 (3.06, 5.37) [0.03, 9.94]		
Mini Mental State Exam at last preclinical visit, mean (SD)	28.0 (1.82)	27.52 (1.83)	28.39 (1.75)	<.0001
JH ADRC	**Full cohort**	**Clinical AD participants**	**Control participants**	**p-value**
Number of participants (%)	64	35 (54.69)	29 (45.31)	
Number of person-visits (%)	127	69 (54.33)	58 (45.67)	
Number of women participants (%)	41 (64.06)	18 (51.43)	23 (79.31)	.0353
Number of non-White participants, n (%)	16 (25.00)	7 (2 .00)	9 (31.03)	.389
Education (yrs), mean (SD)	15.88 (2.65)	16.40 (2.60)	15.24 (2.60)	.081
Baseline age (yrs), mean (SD), range	73.17 (8.34) 54 - 87	74.03 (8.73) 54 - 87	72.14 (7.86) 56-86	.371

Time interval between 2 visits (yrs), mean (SD), range	1.09 (.16) .88-1.77	1.11 (.18) .95 – 1.77	1.07 (.15) .88-1.61	.250
Mini Mental State Exam at first visit, mean (SD)	26.55 (3.72)	23.89 (3.03)	29.76 (.64)	<.001

Continuous variables were compared using t test and categorical variables were compared using Fisher exact test. In the BLSA cohort, time interval before symptom onset refers to AD participants; a similar interval between first-last visit was used to select Controls.

eTable 2. Electrochemiluminescence units for p231-tau and p181- tau and corresponding concentration of phosphorylated full length 441 tau standard used for calibration

	Phosphorylated full length 441 tau (ng/ml)	p231-tau (electrochemiluminescence units/well)	p181-tau (% electrochemiluminescence units/well of undiluted standard)
standard 1	1000	50	100
standard 2	250	13	25
standard 3	63	3.1	6.25
standard 4	16	.78	1.56
standard 5	3.9	.2	.39
standard 6	.98	.049	.10
standard 7	.24	.012	.02
standard 8	0	0	0

For The undiluted standard was assigned a 100% phosphorylation value and 5 additional calibrators were generated by 5 additional 4-fold serial dilutions

eTable 3. Stability of L1CAM+ EV biomarkers over time

	Control participants			AD participants		
Pearson's correlation	**r**	**R^2**	**p-value**	**r**	**R^2**	**p-value**
EV concentration	.376	.141	3.05E-07	.349	.122	9.20E-05
EV mean diameter	.649	.421	2.01E-22	.483	.233	2.07E-08
Size mode diameter	.611	.373	2.06E-19	.385	.148	1.30E-05
Aβ42	.529	.280	8.65E-11	.265	.070	1.20E-02
p181-tau	.833	.694	5.39E-46	.689	.475	6.34E-18
Total tau	.760	.578	1.83E-26	.846	.716	5.75E-23
p231-tau	.788	.621	1.00E-37	.668	.446	3.39E-16
TSG101	.786	.618	6.66E-37	.679	.461	1.37E-16
pSer312-IRS1	.733	.537	1.47E-24	.765	.585	1.23E-19
pY-IRS1	.783	.613	6.62E-32	.697	.486	2.75E-16

Within-subject correlations between values of EV biomarkers at the earliest and the latest preclinical visits for Control and AD participants. Correlations were estimated using the Pearson’s correlation test.

eTable 4. Inter-correlations between means of EV biomarkers across all visits

Correlation of mean biomarkers									
	EV Concentration	**EV average Diameter**	**TSG101**	**Total tau**	**p231-tau**	**p181-tau**	**pY-IRS1**	**pSer312-IRS1**	**A(**
EV Concentration	1	.033	**-.207**	.063	.081	.072	- .058	- .147	.0
EV average Diameter	.033	1	**.183**	.099	**.191**	**.378**	**.149**	**.127**	.0
TSG101	**- .207**	**.183**	1	.038	**.454**	**.260**	**.384**	**.608**	**- .**
Total tau	.063	.099	.038	1	**.282**	.090	**- .136**	**- .134**	**.2**
p231-tau	.081	**.191**	**.454**	**.282**	1	**.424**	**.373**	**.289**	**.1**
p181-tau	.072	**.378**	**.2**	.090	**.424**	1	**.578**	**.545**	.0

			60						
pY-IRS1	- .058	**.149**	**.384**	**- .136**	**.373**	**.578**	1	**.605**	**- .**
pSer312-IRS1	**- .147**	**.127**	**.608**	**- .134**	**.289**	**.545**	**.605**	1	- .
Aβ42	.085	.019	**-.145**	**.219**	**.147**	.022	**- .299**	**- .101**	1

p-value:

	EV Concentration	**EV average Diameter**	**TSG101**	**Total tau**	**p231-tau**	**p181-tau**	**pY-IRS1**	**pSer312-IRS1**	**Aβ42**
EV Concentration	0	.616	.001	.331	.212	.269	.371	.022	.187
EV average Diameter	.616	0	.004	.125	.003	.000	.021	.049	.767
TSG101	.001	.004	0	.555	< .001	< .001	< .001	< .001	.024
Total tau	.331	.125	.555	0	.000	.162	.035	.038	.001
p231-tau	.212	.003	< .001	< .001	0	< .001	< .001	< .001	.022
p181-tau	.269	< .001	< .001	.162	< .001	0	< .001	< .001	.729

pY-IRS1	.371	.021	< .001	.035	< .001	< .001	0	< .001	.000
pSer312-IRS1	.022	.049	< .001	.038	< .001	< .001	< .001	0	.117
Aβ42	.187	.767	.024	.001	.022	.729	< .001	.117	0

Correlations were estimated using the Pearson's correlation test. Bolded r-correlation coefficients correspond to p-value < .05.

eTable 5. Inter-correlations between EV biomarkers at the last preclinical visit

Inter-correlations of EV biomarkers at last preclinical visit									
	EV Concentration	**EV average diameter**	**TSG101**	**Total tau**	**p231-tau**	**p181-tau**	**pY-IRS1**	**pSer312-IRS1**	**Aβ42**
EV Concentration	1.000	.008	- .102	.066	**.158**	.027	- .017	- .044	.034
EV average Diameter	.008	1.000	.158	.065	**.158**	**.254**	.067	.028	- .028
TSG101	- .102	**.158**	1.000	- .001	**.441**	**.266**	**.403**	**.583**	- .125
Total tau	.066	.065	- .001	1.000	**.298**	.111	- .110	**- .128**	**.190**
p231-tau	**.158**	**.158**	**.441**	**.298**	1.000	**.460**	**.373**	**.305**	.120
p181-tau	.027	**.254**	**.266**	.111	**.460**	1.000	**.562**	**.539**	- .023
pY-IRS1	- .017	.067	**.403**	- .110	**.373**	**.562**	1.000	**.619**	**- .251**
pSer312-IRS1	- .044	.028	**.583**	**- .128**	**.305**	**.539**	**.619**	1.000	- .102
Aβ42	.034	- .028	- .125	**.190**	.120	- .023	**- .251**	- .102	1.000
p-value:									
	EV Concentration	**EV average diameter**	**TSG101**	**Total tau**	**p231-tau**	**p181-tau**	**pY-IRS1**	**pSer312-IRS1**	**Aβ42**

EV Concentration	0	.906	.115	.305	.014	.672	.790	.500	.599
EV average Diameter	.906	0	.014	.315	.014	< .001	.299	.664	.660
TSG101	.115	.014	0	.985	< .001	< .001	< .001	< .001	.052
Total tau	.305	.315	.985	0	< .001	.085	.087	.047	.003
p231-tau	.014	.014	< .001	< .001	0	< .001	< .001	< .001	.063
p181-tau	.672	.000	< .001	.085	< .001	0	< .001	< .001	.724
pY-IRS1	.790	.299	< .001	.087	< .001	< .001	0	< .001	< .001
pSer312-IRS1	.500	.664	< .001	.047	< .001	< .001	< .001	0	.113
Aβ42	.599	.660	.052	.003	.063	.724	< .001	.113	0

Correlations were estimated using the Pearson's correlation test. Bolded r-correlation coefficients correspond to p-value < .05.

eTable 6. Inter-correlations between the slopes of EV biomarkers

Inter-correlations of EV biomarker slopes									
	EV Concentration	**EV average diameter**	**TSG101**	**Total tau**	**p231-tau**	**p181-tau**	**pY-IRS1**	**pSer312-IRS1**	**Aβ42**
EV Concentration	1.000	.034	**- .219**	**.127**	.068	.078	- .077	- .164	.096
EV average Diameter	.034	1.000	.085	.051	.087	**.189**	.052	.017	- .016
TSG101	**- .219**	.085	1.000	.106	**.472**	**.237**	**.337**	**.538**	- .105
Total tau	**.127**	.051	.106	1.000	**.283**	.097	**- .151**	- .093	**.191**
p231-tau	.068	.087	**.472**	**.283**	1.000	**.424**	**.360**	**.239**	.093
p181-tau	.078	**.189**	**.237**	.097	**.424**	1.000	**.568**	**.533**	- .018
pY-IRS1	- .077	.052	**.337**	**- .151**	**.360**	**.568**	1.000	**.573**	**- .312**
pSer312-IRS1	- .164	.017	**.538**	- .093	**.239**	**.533**	**.573**	1.000	- .089
Aβ42	.096	- .016	- .105	**.191**	.093	- .018	**- .312**	- .089	1.000
P-value:									
	EV Concentration	**EV average diameter**	**TSG101**	**Total tau**	**p231-tau**	**p181-tau**	**pY-IRS1**	**pSer312-IRS1**	**Aβ42**

EV Concentration	0	.596	.001	.050	.292	.229	.231	.011	.138
EV average Diameter	.596	0	.189	.434	.178	.003	.419	.792	.810
TSG101	.001	.189	0	.100	< .001	< .001	< .001	< .001	.105
Total tau	.050	.434	.100	0	< .001	.134	.019	.150	.003
p231-tau	.292	.178	< .001	.000	0	.000	< .001	< .001	.151
p181-tau	.229	.003	< .001	.134	< .001	0	< .001	< .001	.780
pY-IRS1	.231	.419	< .001	.019	< .001	< .001	0	< .001	< .001
pSer312-IRS1	.011	.792	< .001	.150	< .001	< .001	< .001	0	.166
Aβ42	.138	.810	.105	.003	.151	.780	< .001	.166	0

Correlations were estimated using the Pearson's correlation test. Bolded r-correlation coefficients correspond to p-value < .05.

eTable 7. Cross-sectional and longitudinal associations of nEV biomarkers and composite cognitive scores across all BLSA participants

		Cross-Sectional Effect			Longitudinal Effect		
Cognitive Domains	**Predictors**	**β**	**SE**	**p-value**	**β**	**SE**	**p-value**
Verbal Memory	Log Aβ42	- .029	.143	.840	**.062**	**.026**	**.020**
	Log p181-tau	**- .435**	**.214**	**.043**	.006	.037	.878
	Log p231-tau	- .351	.230	.128	- .069	.041	.101
	Log total tau	.197	.235	.403	.001	.051	.987
	log pSer312-IRS1	**- .405**	**.112**	**<.001**	- .020	.019	.314
	log pY-IRS1	- .186	.095	.053	- .024	.021	.238
Attention	Log Aβ42	- .069	.112	.540	.004	.023	.850
	Log p181-tau	**- .424**	**.172**	**.014**	- .004	.028	.900
	Log p231-tau	- .186	.190	.328	.017	.032	.585
	Log total tau	.164	.189	.386	- .039	.036	.288
	log pSer312-IRS1	- .143	.093	.126	.000	.016	.980
	log pY-IRS1	.032	.076	.675	**- .042**	**.017**	**.013**
Executive Function	Log Aβ42	.005	.121	.966	.025	.023	.288
	Log p181-tau	**- .448**	**.186**	**.017**	- .039	.030	.189
	Log p231-tau	- .098	.205	.634	- .044	.034	.198
	Log total tau	.334	.205	.105	- .021	.040	.592
	log pSer312-IRS1	**- .222**	**.099**	**.025**	- .014	.016	.399
	log pY-IRS1	- .010	.079	.897	- .009	.017	.607
Language	Log Aβ42	- .102	.119	.392	**.050**	**.020**	**.011**
	Log p181-tau	- .198	.184	.283	- .023	.026	.381
	Log p231-tau	.158	.200	.429	- .025	.030	.403
	Log total tau	.320	.193	.099	.019	.035	.598

	log pSer312-IRS1	- .106	.098	.279	- .024	.014	.081
	log pY-IRS1	- .023	.080	.774	- .012	.016	.466
Visuospatial	Log Aβ42	.069	.112	.538	.002	.023	.930
	Log p181-tau	**- .409**	**.169**	**.016**	.011	.029	.714
	Log p231-tau	**- .376**	**.184**	**.042**	.007	.034	.827
	Log total tau	- .224	.191	.242	.071	.039	.067
	log pSer312-IRS1	- .176	.090	.051	.005	.016	.736
	log pY-IRS1	.011	.075	.880	- .016	.018	.374

Significant associations are in bold font when p-value < .05.

eTable 8. Missing values assessment

Feature	Included in Prediction Model N=241 Mean (SD) or Number (%)	Excluded from Prediction Model N=109 Mean (SD) or Number (%)	P-value
Alzheimer's Disease	82 (34.0%)	46 (42.2%)	.15
Female Sex	127 (52.7%)	51 (46.8%)	.36
Age (years)	8 .5 (7.2)	8 .0 (7.8)	.59
Year of Birth (year)	1923 (12)	1922 (14)	.47

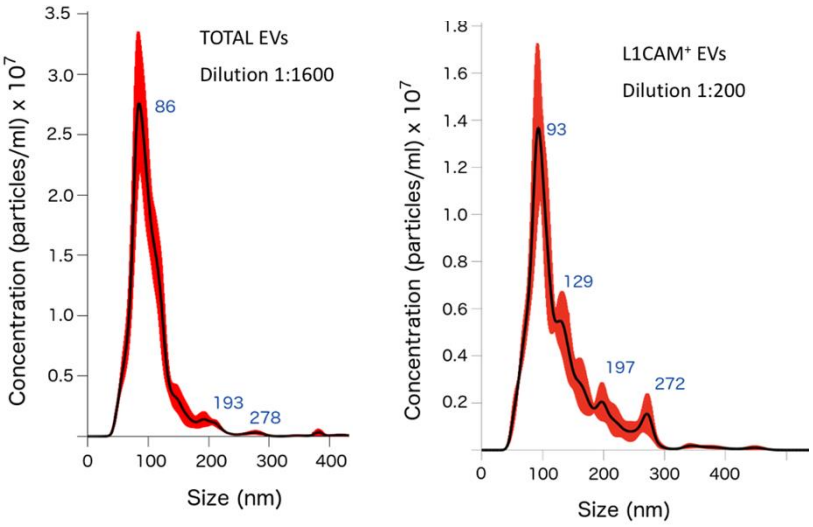
Participants were included in prediction model building, if they had at least one visit prior to AD symptom onset with valid values for all nEV biomarkers and complete demographic data. P-values are from t-tests or Fisher's exact tests.

eTable 9. Performance statistics of Final Model 10 for Alzheimer Disease Prediction

Risk Score Estimates	Model 10 Performance for Risk Score >.407		
	Sensitivity (95% CI)	Specificity (95% CI)	Odds Ratio (95% CI)
Training data[a]	0.818 (0.686-0.905)	0.858 (0.774-0.916)	27.3 (11.8-68.9)
Internally cross-validated training data[b]	0.673 (0.532-0.790)	0.792 (0.701-0.863)	7.8 (3.8-16.7)
Test data[c]	0.556 (0.356-0.740)	0.887 (0.763-0.953)	9.8 (3.3-32.8)

[a]Training data refer to 2/3 of the Baltimore Longitudinal Study of Aging sample.
[b]Internal cross-validation refers to leave-10%-out of training data.
[c]Test data refer to 1/3 of the Baltimore Longitudinal Study of Aging sample used for external validation.

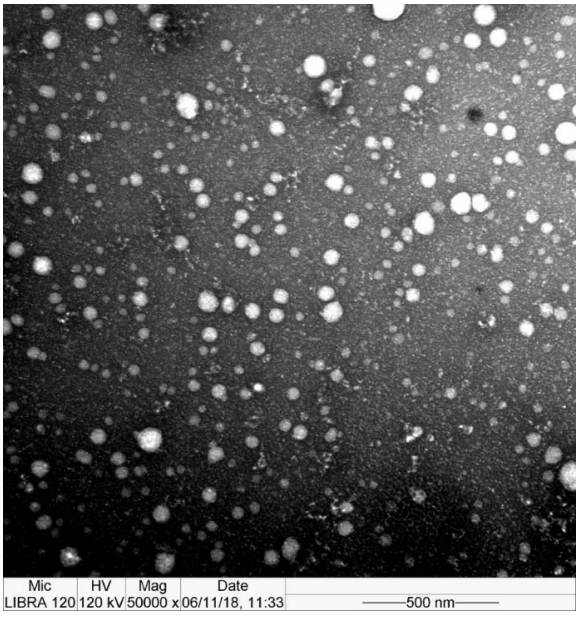
eFigure 1. Nanoparticle tracking analysis (NTA) of total EVs isolated by Exoquick® alone and neuronal-enriched EVs isolated through Exoquick® followed by immunoprecipitation with antibodies against L1CAM (L1CAM+ EVs) from a typical plasma sample.



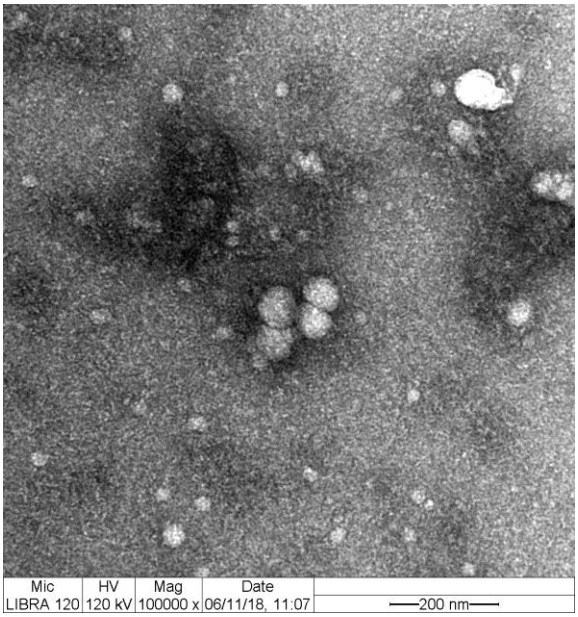
The graphs display particle concentration/ml against size (diameter in nm)

eFigure 2. Electron Microscopy of neuronal-enriched plasma EVs

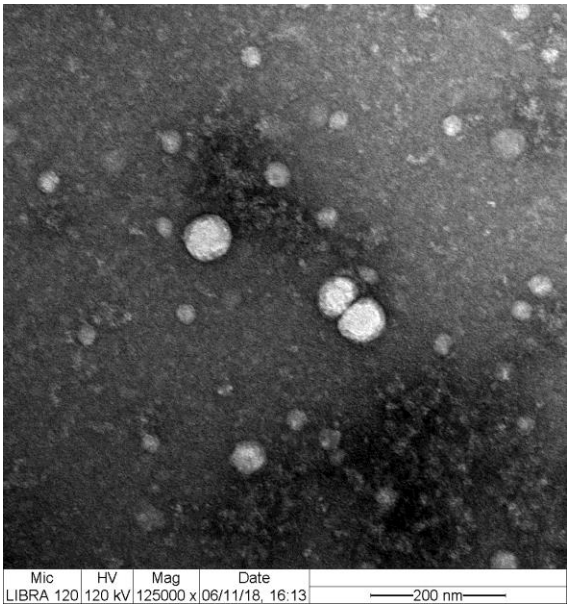
A



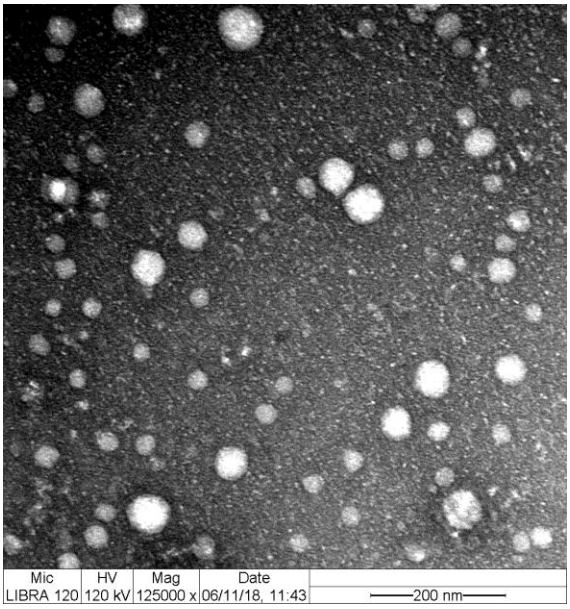
B



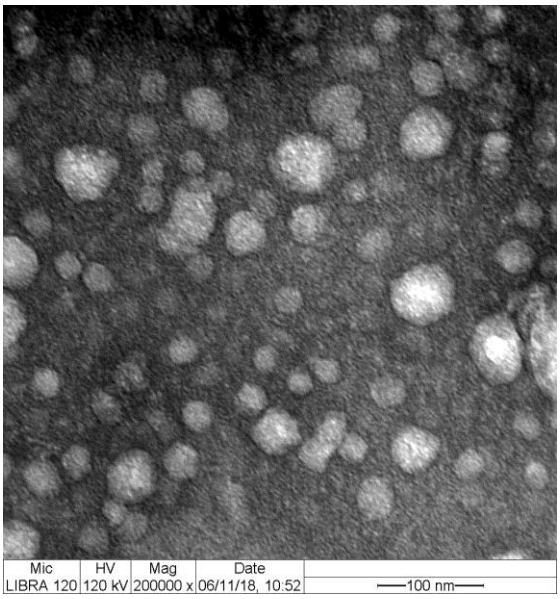
C



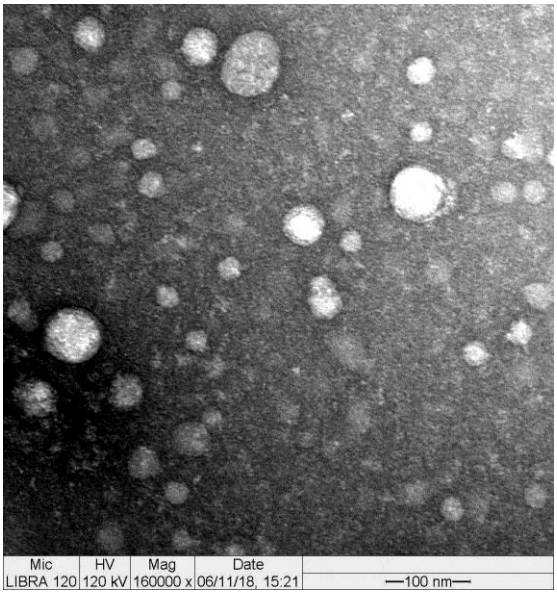
D



E

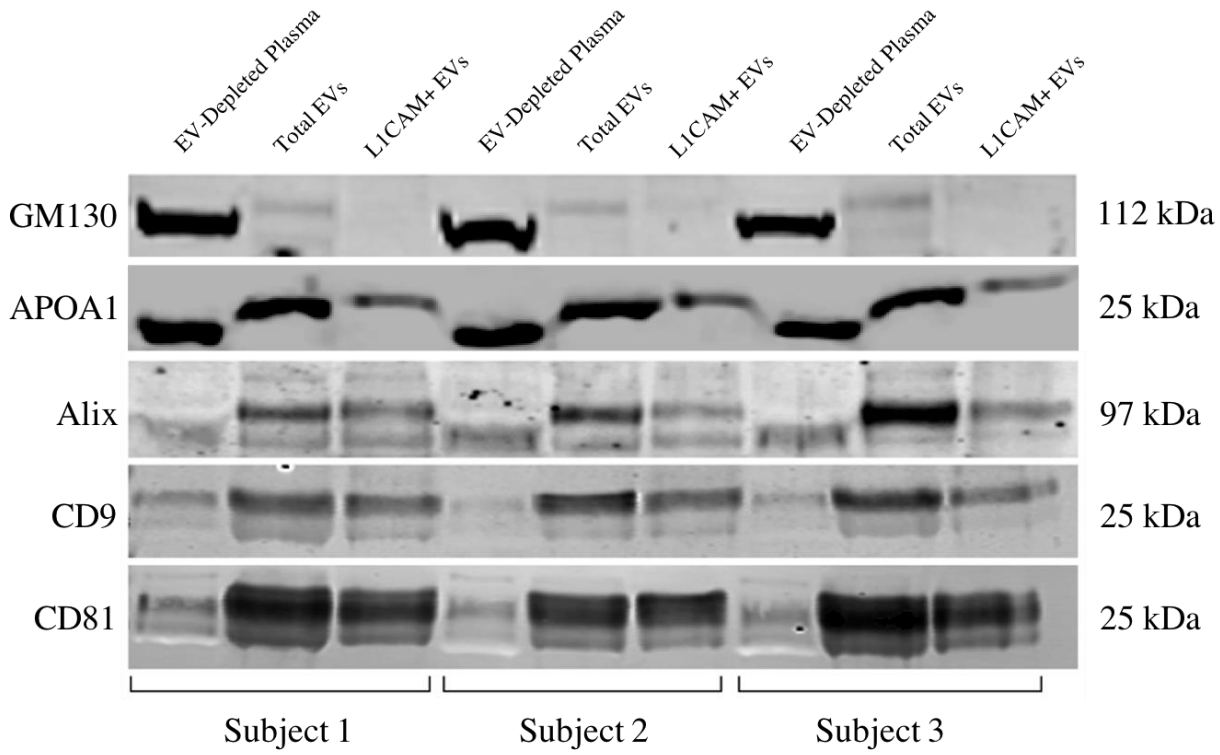


F



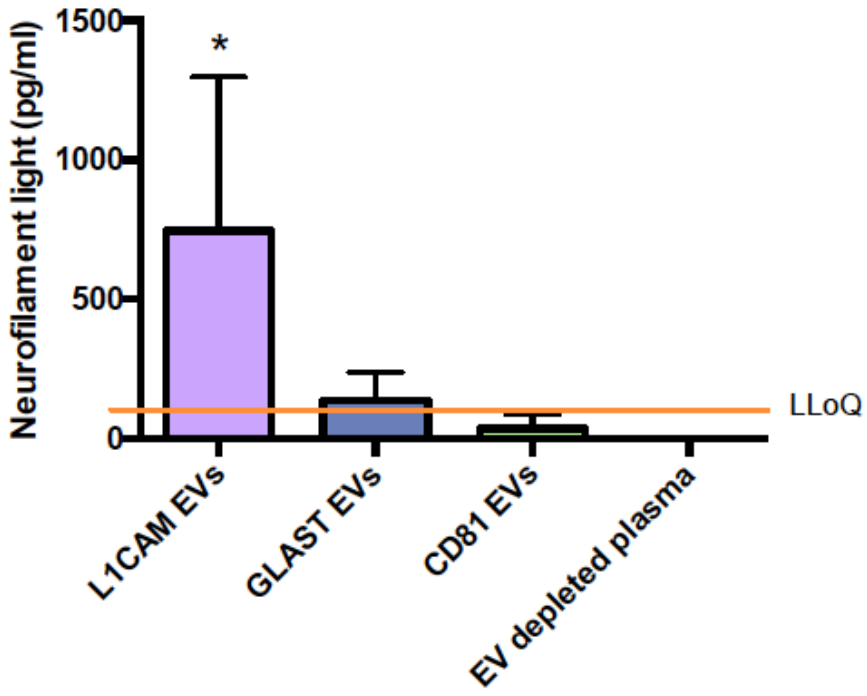
Characteristic EM images (negative staining) of neuronal-enriched plasma EVs from three participants in lower to higher magnifications (A: 500 nm; B, C, D: 200 NM; E, F: 100 nm)

eFigure 3. Western Blot characterization of neuronal-enriched plasma EVs



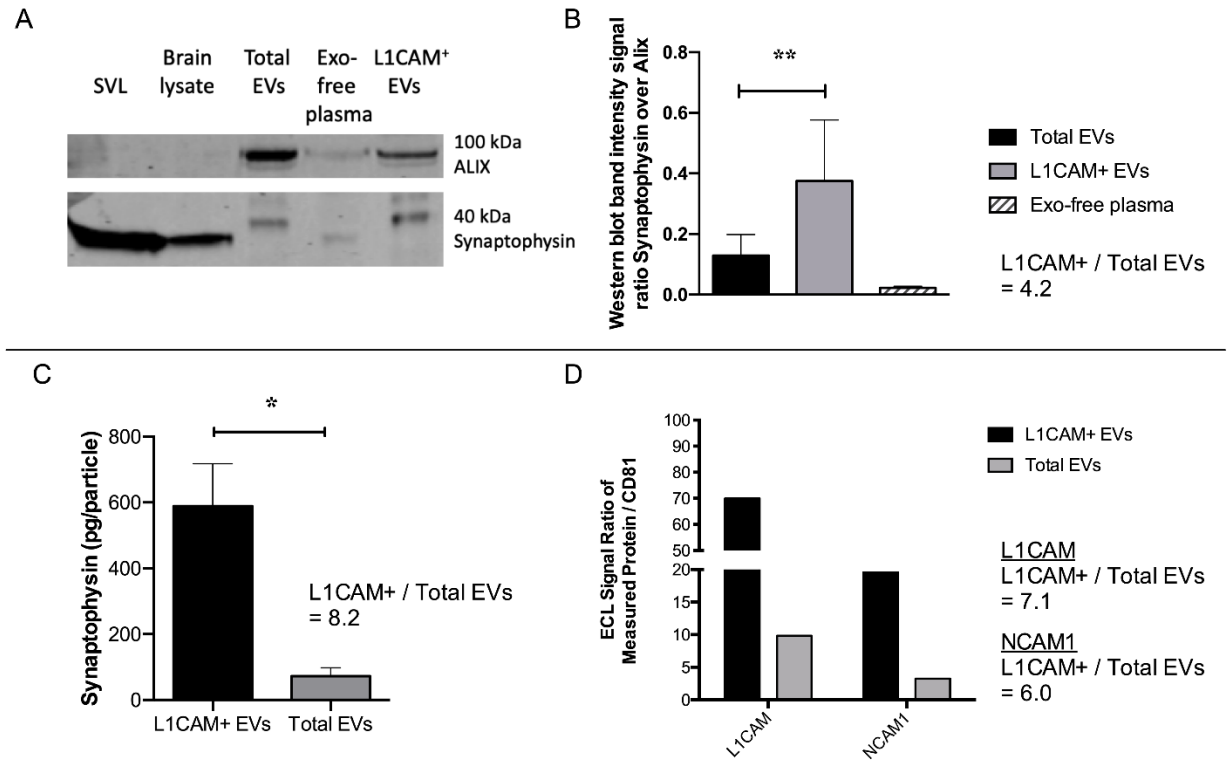
Western Blots of neuronal-enriched (L1CAM+) plasma EVs compared to Total EVs (isolated with Exoquick®) and EV-depleted plasma (supernatant after Exoquick®) for N = 3 participants. L1CAM+ EVs compared to EV-depleted plasma had lower GM130 ($p = .003$) and ApoA1 ($p = .003$) and higher Alix ($p < .001$), CD9 ($p = .003$) and CD81 ($p < .001$). L1CAM+ EVs compared to Total EVs had similar levels of GM130, Alix, and CD81, but lower levels of CD9 ($p = .02$) and ApoA1 ($p = .006$).

eFigure 4. Neuronal enrichment of L1CAM+ EVs



ELISA for Neurofilament light (NF-L) for N = 4 participants. L1CAM EVs had higher levels compared to CD81 EVs (p = .03) and a trend for higher levels compared to GLAST EVs (p = .06).

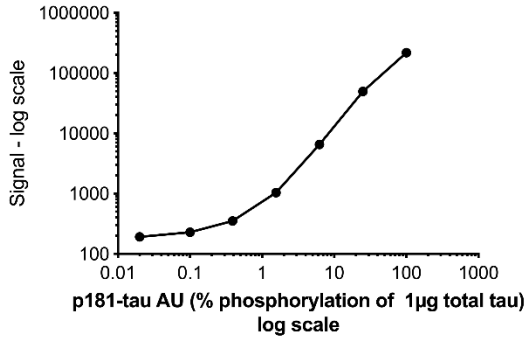
eFigure 5. Degree of Neuronal enrichment of L1CAM+ EVs compared to Total plasma EVs



A. Characteristic WB. Lanes: membrane probed using anti-Alix antibody (top, 100 kDa band) and anti-Synaptophysin (bottom, 40 kDa band). Rows: 1) Synaptic vesicle lysate (SVL); 2) Brain lysate (rat); 3) Total EVs (TEV); 4) EV-free plasma (Exo-free, i.e. supernatant after EV isolation by Exoquick); 5) L1CAM+ EVs. B. WB quantification for N = 4 healthy Controls; band intensity for Synaptophysin was normalized by the ALIX band intensity and comparison was made by unpaired t-test comparing Total and L1CAM+ EVs (**p= .004). C. Synaptophysin ELISA results for Total and L1CAM+ EVs from 3 healthy Controls (comparison made using unpaired t-test, *p= .017). D. Chemiluminescence (ECL) signal intensity for L1CAM and NCAM using the Exo-Check Exosome Antibody Array (Neuro) and quantified by Image Studio for Total EVs and L1CAM+ EVs pooled from 3 healthy Controls.

eFigure 6. Representative standard curves and determinations of the Lowest Limit of Quantification (LLoQ) for the A) p181-tau and B) TSG101 electrochemiluminescence assays

A

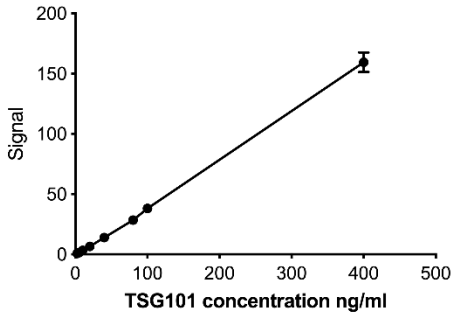


Concentration (AU)	Signal 1	Signal 2	Signal Mean	Signal SD	Signal %CV	% Recovery	Calculated concentration mean	Calculated concentration %CV
100	222124	213361	217743	6196	2.85	98.81	98.81	3.44
25	50447	48916	49682	1083	2.18	104.73	26.18	1.65
6.25	6904	6137	6521	542	8.32	96.81	6.05	5.93
1.56	1093	979	1036	81	7.78	96.71	1.51	6.60
0.39	346	357	352	8	2.21	119.85	0.47	3.55
0.1	226	232	229	4	1.85	149.16	0.15	10.73
0.02	187	197	192	7	3.68	NaN	NaN	NaN
BLANK	188	205	197	12	6.12	NaN	NaN	NaN

BLANK Mean	196.50
Mean (blank) + 9*SD (blank)	304.69

1. Signal above 304.69
2. Signal %CV < 15%
3. % Recovery: 80%<>120%

B

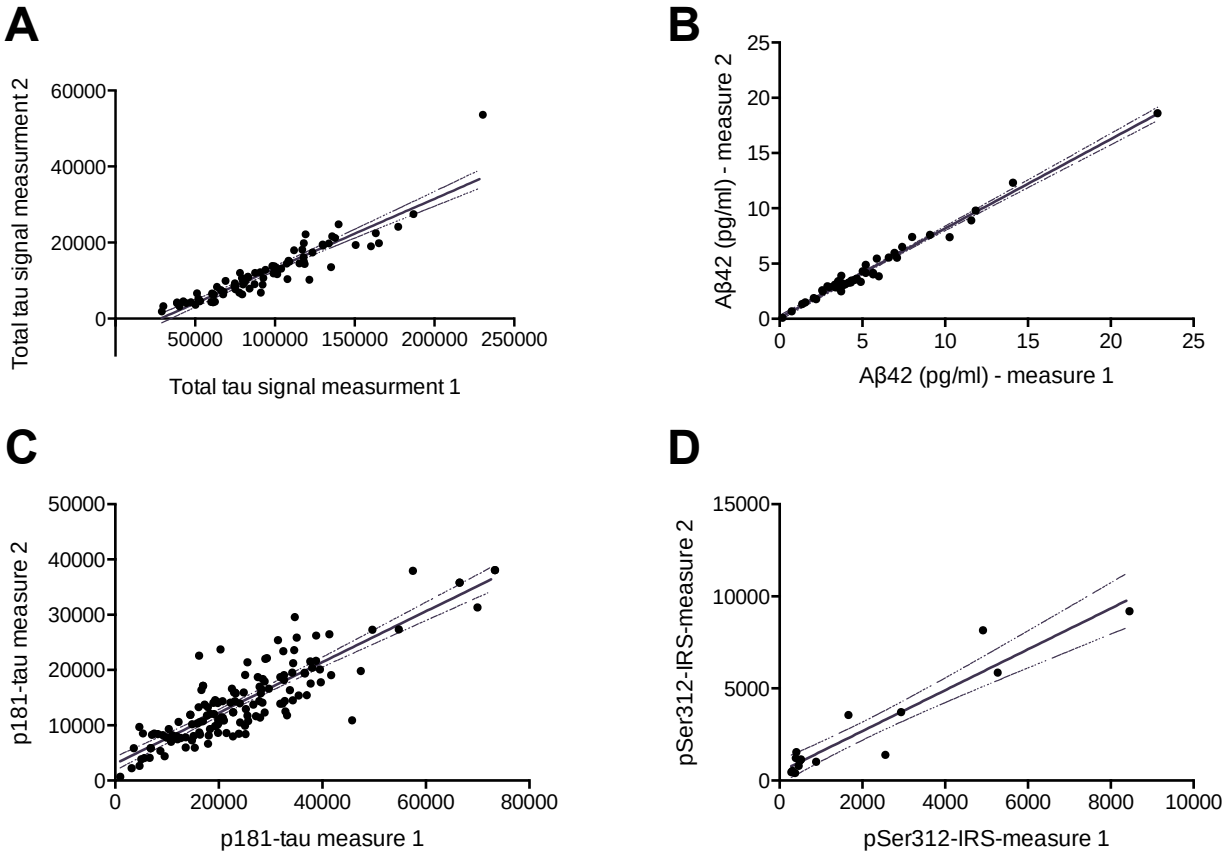


Concentration (ng/ml)	Signal 1	Signal 2	Signal Mean	Signal SD	Signal %CV	% Recovery	Calculated concentration mean	Calculated concentration %CV
400	153676	165248	159462	8183	5.13	100.29	401.17	4.88
100	37247	38962	38105	1213	3.18	103.06	103.06	3.02
80	27773	28972	28373	848	2.99	97.41	77.93	2.83
40	13326	14495	13911	827	5.94	99.27	39.71	5.61
20	6442	6664	6553	157	2.40	97.79	19.56	2.24
10	3318	3296	3307	16	0.47	103.45	10.35	0.43
5	1454	1510	1482	40	2.67	99.61	4.98	2.39
BLANK	676	659	668	12	1.80	99.80	2.50	1.50

BLANK Mean	667.50
Mean (blank) + 9*SD (blank)	775.69

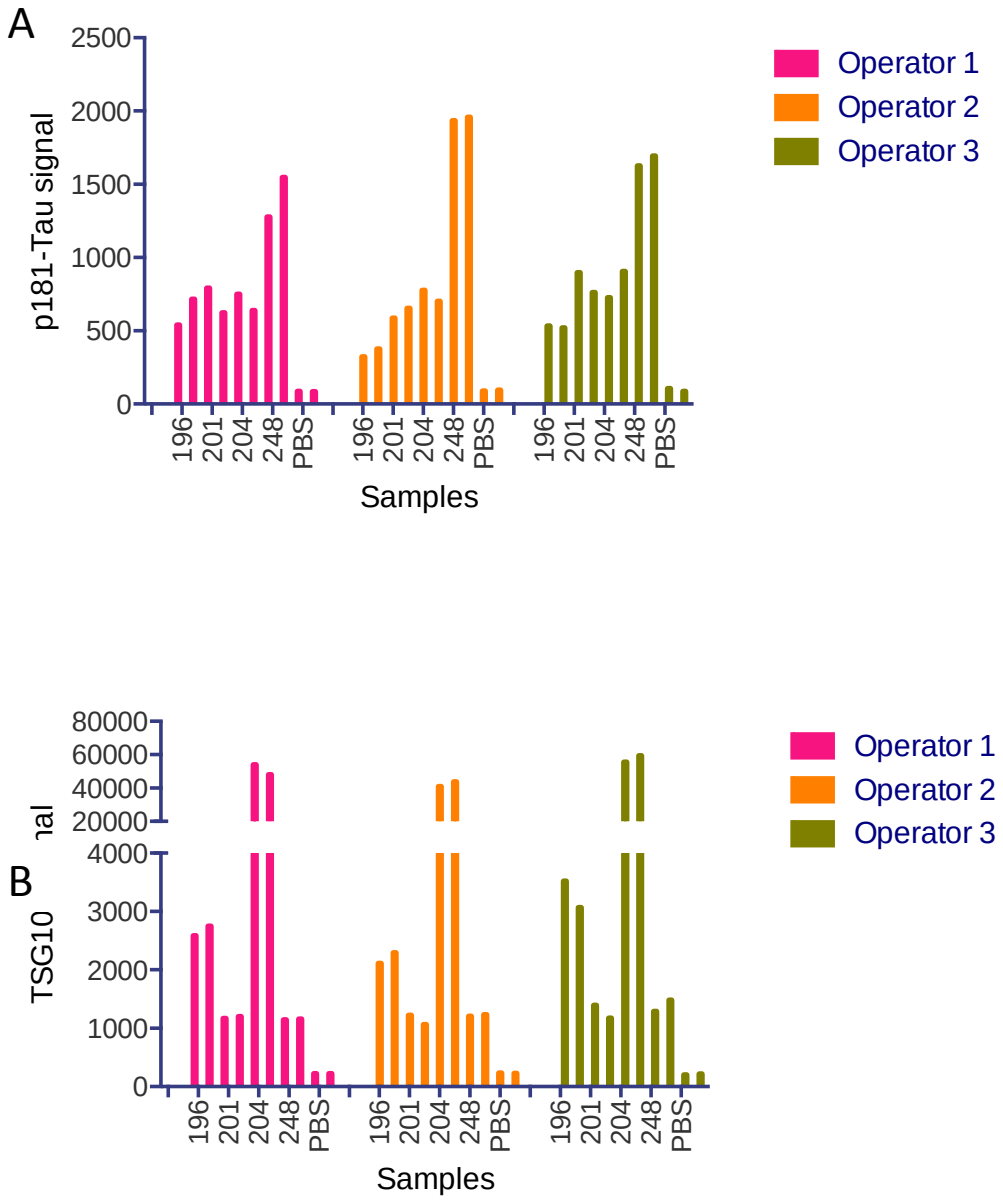
1. Signal above 775.69
2. Signal %CV < 15%
3. % Recovery: 80%<>120%

eFigure 7. Within-Operator variability for assays and the entire methodology



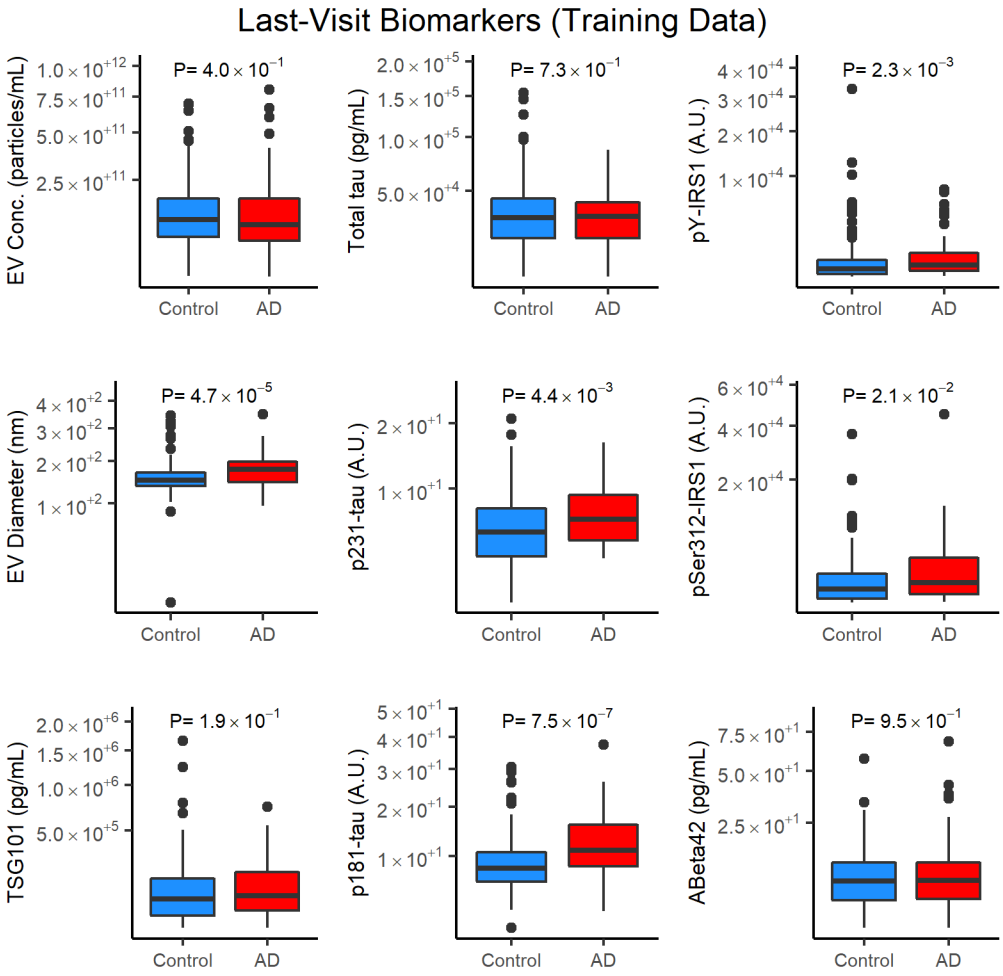
A, B. To assess day to day variability for assay performance for total tau and Aβ42 a subset of lysed L1CAM+ EVs preparations was measured twice on two different days by the same operator. Coefficient of determination (R^2) was .852 ($p < .0001$) for total tau (A) and .981 ($p < .0001$) for Aβ42 (B). C, D. To assess day to day variability for EV isolation followed by biomarker quantification, L1CAM+ EVs were isolated twice from a subset of samples and p181-tau and pSer312-IRS-1 were measured in both isolated replicates by the same operator. Coefficient of determination (R^2) was .751 ($p < .0001$) for p181-tau (C) and .898 ($p < .0001$) for pSer312-IRS-1 (D).

eFigure 8. Between-Operator variability for the entire methodology



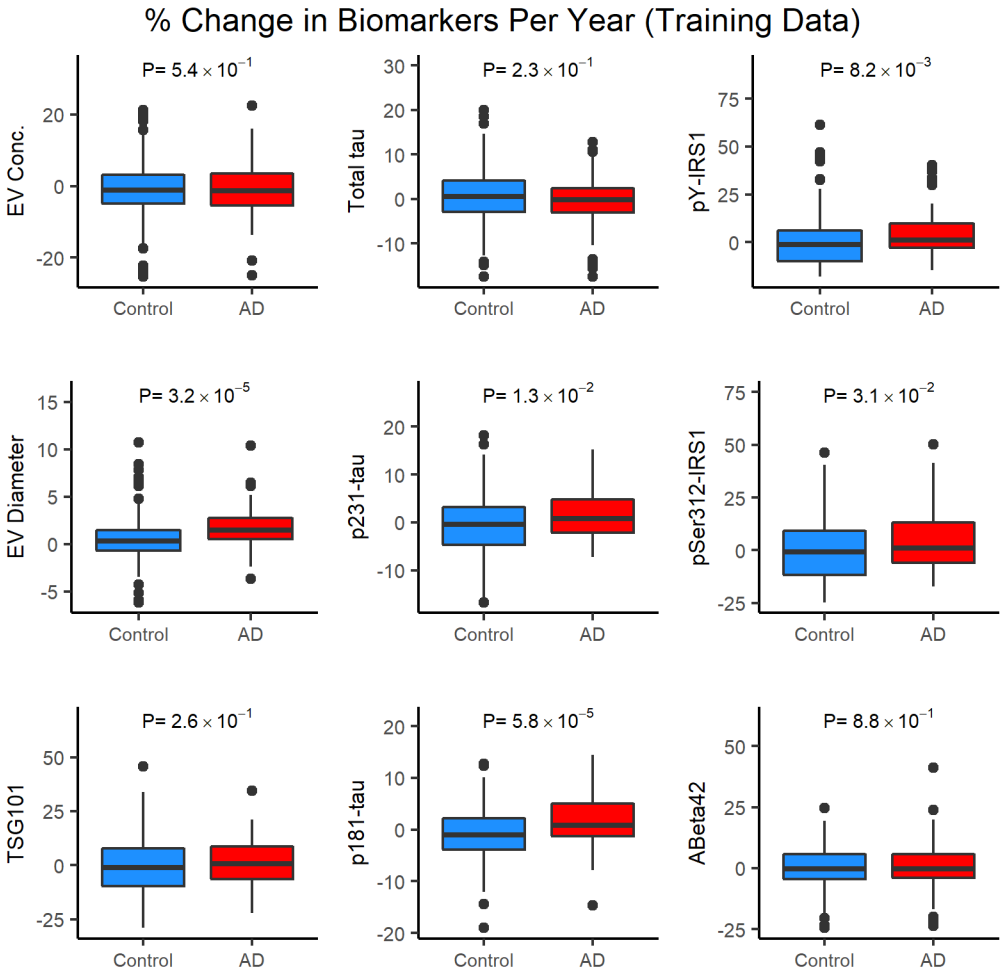
To assess variability between and within operators, three different lab members performed EV isolation followed by biomarker quantifications in two biological replicates for four test subjects (196, 201, 204, 248). For each one of the four test subjects, six plasma aliquots were thawed, mixed and divided into six new aliquots. Each of the three lab members received two of these aliquots for four test subjects, as well as two negative control samples (PBS). Each lab member performed EV isolation separately and measured p181-tau (A) and TSG101 (B). Coefficients of determination (R^2) were .98 (1^{st} vs. 2^{nd}), .96 (1^{st} vs. 3^{rd}) and .98 (2^{nd} vs. 3^{rd}) between operators for p181-tau and .87 (1^{st} vs. 2^{nd}), .92 (1^{st} vs. 3^{rd}) and .94 (2^{nd} vs. 3^{rd}) for TSG101. Coefficients of variation (%CV) were 11.59 for pTau-181 and 11.62 for TSG101.

eFigure 9. EV biomarker levels at the last visit before AD onset for AD and Control BLSA participants



The y-axis depicts biomarker values on the original scale and unit of measurement: E10 particles/mL for EV Concentration; nm for EV average diameter; A.U. for p231-tau, p181-tau, pY-IRS1, and pSer312-IRS1; and pg/mL for total tau and Aβ42. Boxes depict the median and upper and lower quartiles; error bars depict 1.5 x interquartile range and single dots depict outliers. The statistical significance for the Wilcoxon rank-sum tests comparison between AD and Control participants is shown separately in each panel.

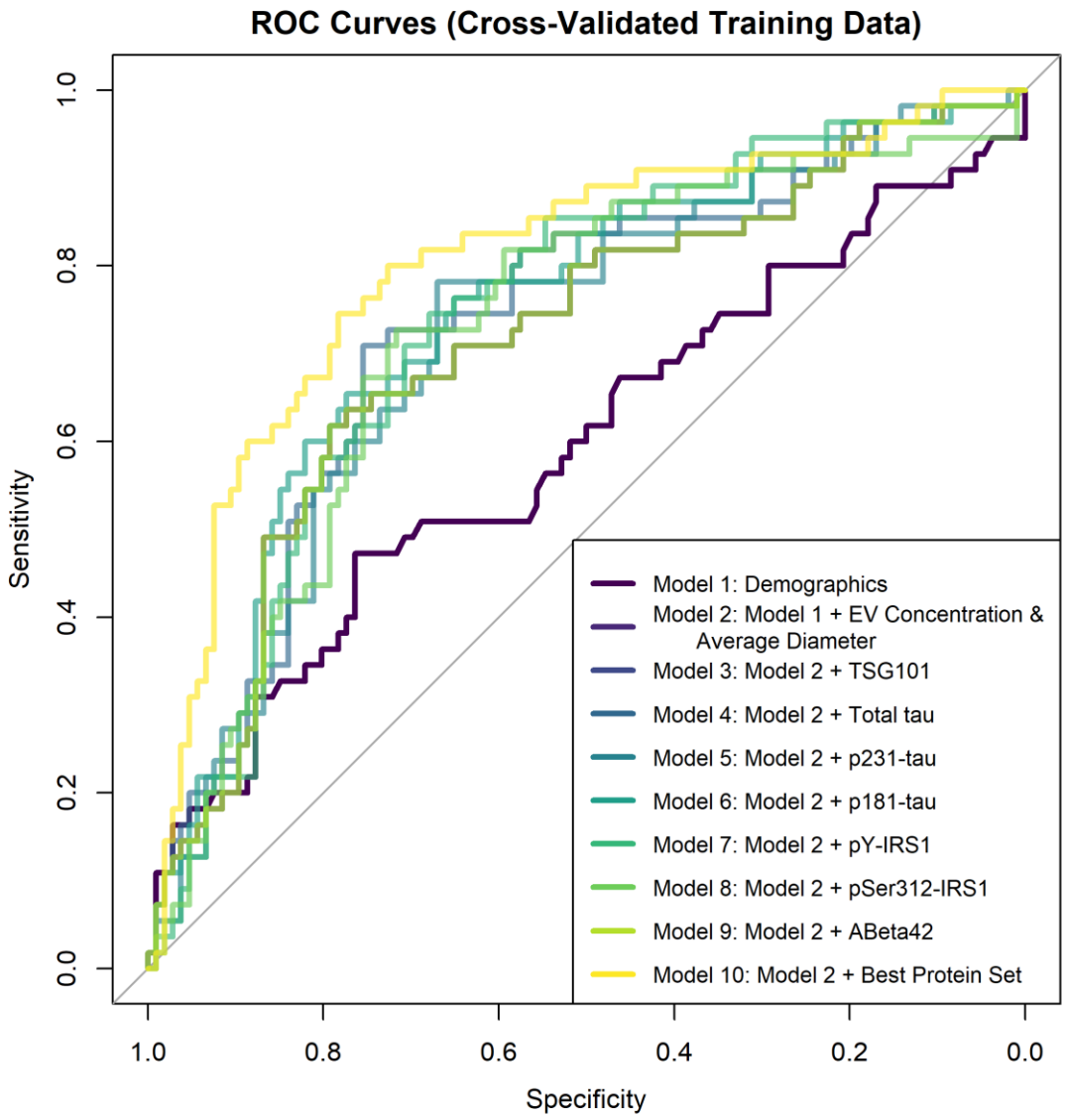
eFigure 10. EV biomarker slopes (percent change in levels of log-transformed biomarker per year) for future AD and Control BLSA participants



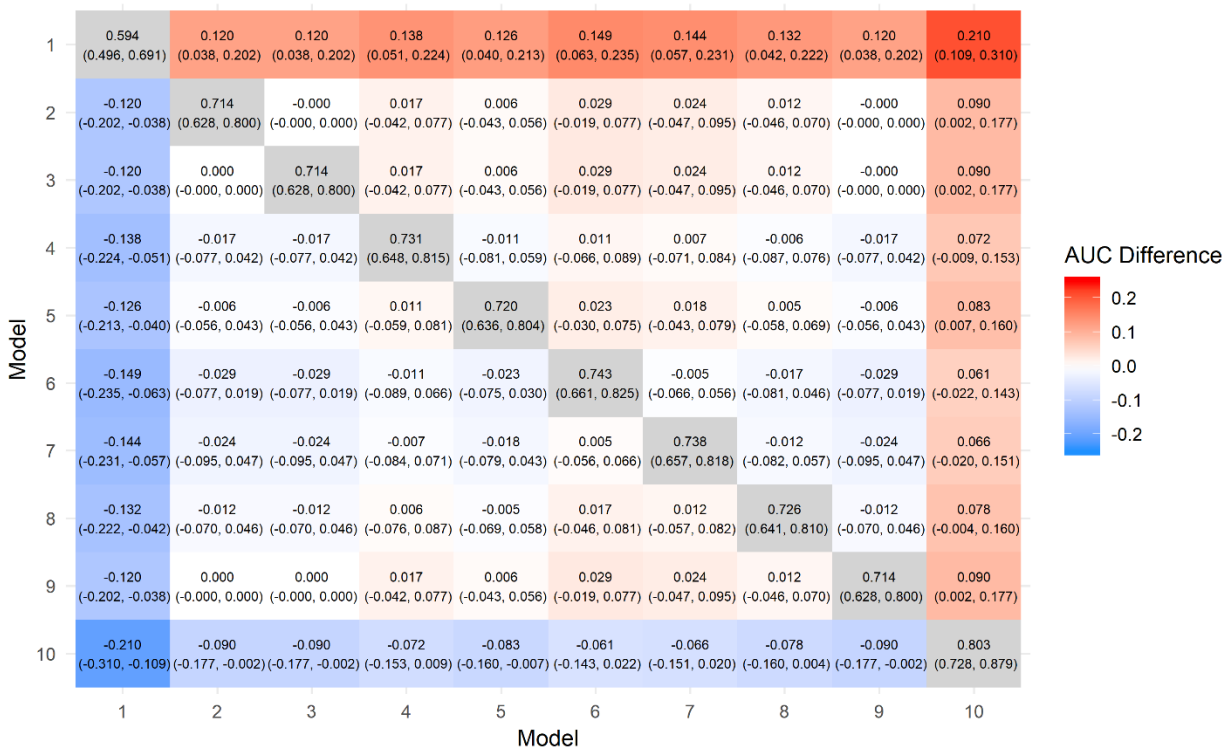
The y-axis depicts biomarker values on the original scale and unit of measurement: E10 particles/mL for EV Concentration; nm for EV average diameter; A.U. for p231-tau, p181-tau, pY-IRS1, and pSer312-IRS1; and pg/mL for total tau and Aβ42. Boxes depict the median and upper and lower quartiles; error bars depict 1.5 x interquartile range and single dots depict outliers. The statistical significance for the Wilcoxon rank-sum tests comparison between AD and Control participants is shown separately in each panel.

eFigure 11. Comparison of Alzheimer's Disease risk prediction models using internal leave-10%-out cross-validation

A.

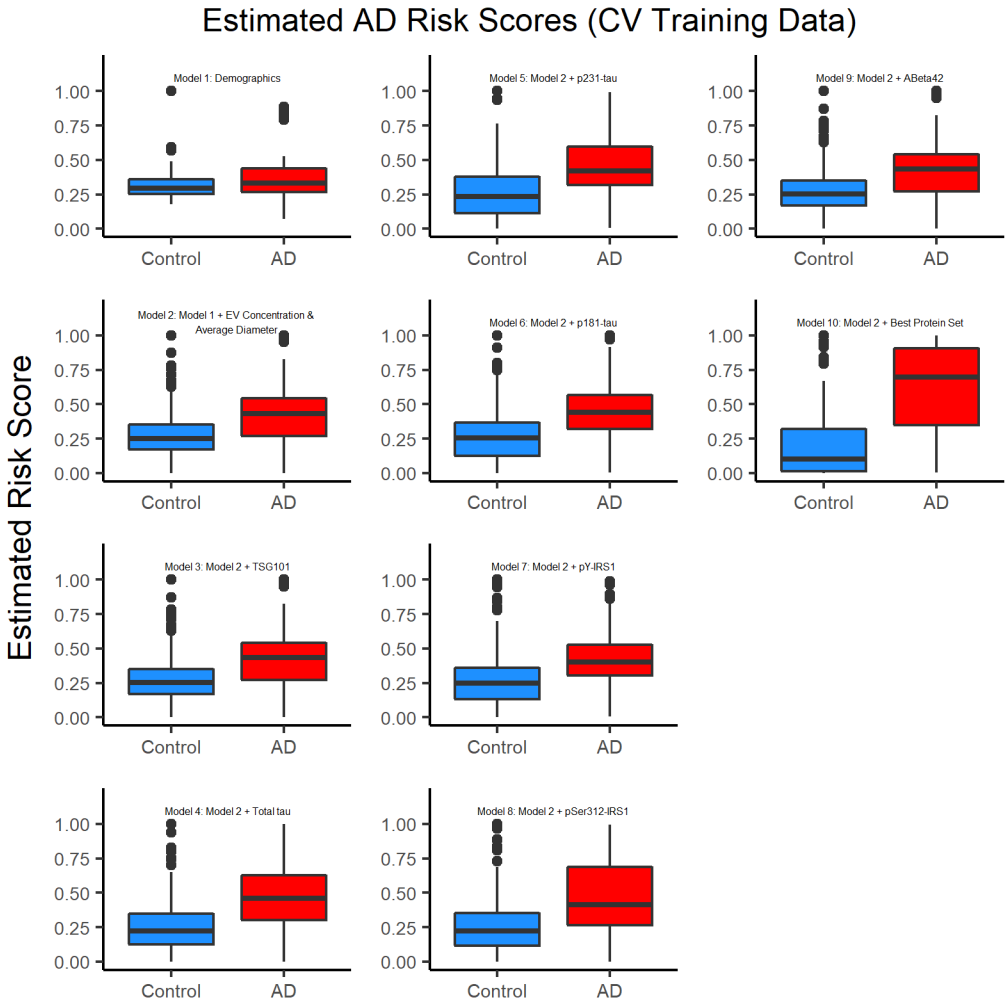


B.



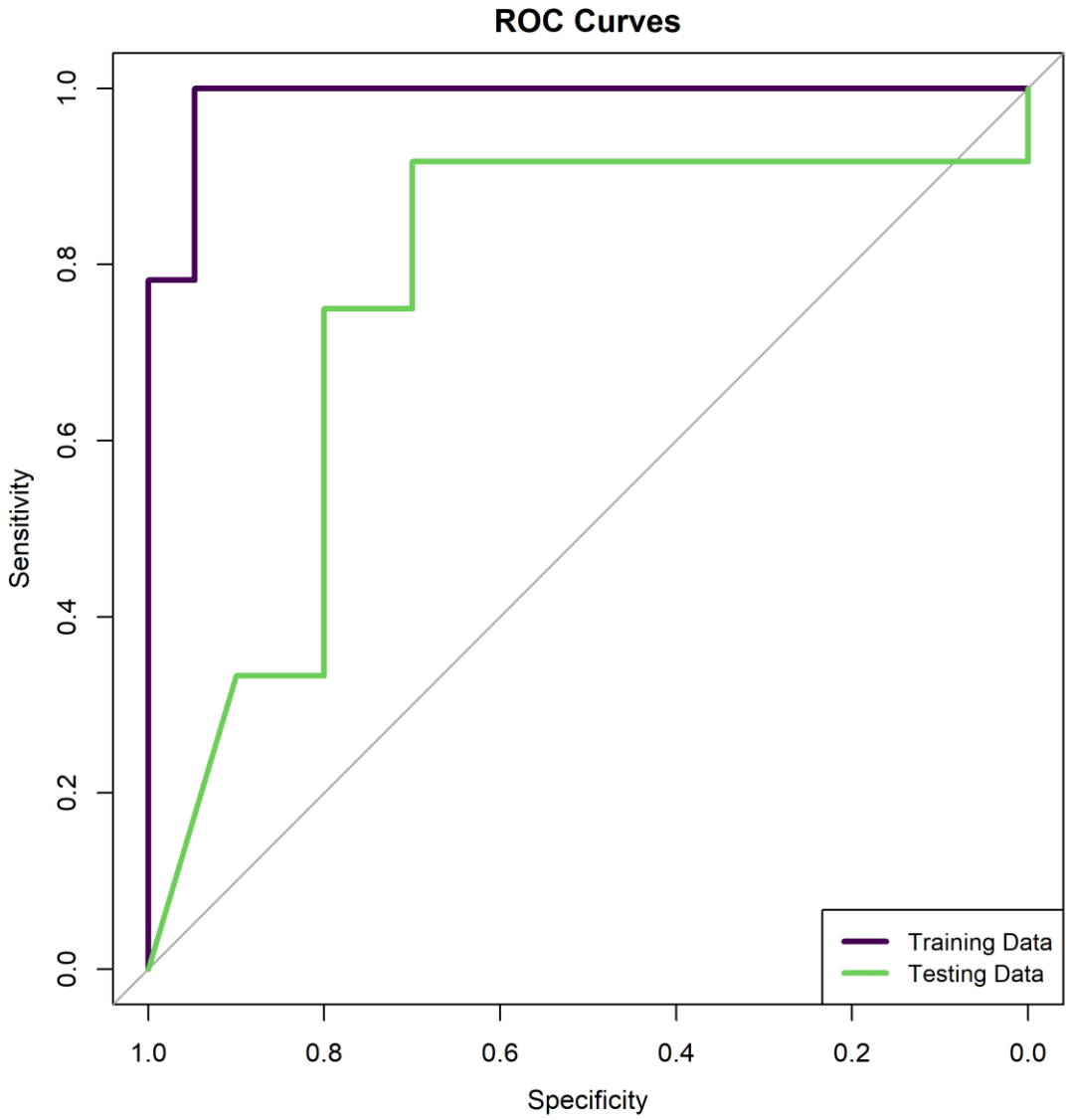
A) ROC curves for 10 models. Model 1: age, sex, and plasma/serum sample type; Model 2: Model 1 + EV Concentration & Average Diameter; Model 3: Model 2 + TSG101; Model 4: Model 2 + total tau; Model 5: Model 2 + p231-tau; Model 6: Model2 + p181-tau; Model 7: Model 2 + pY-IRS1; Model 8: Model 2 + pSer312-IRS1; Model 9: Model 2 + Aβ42; Model 10: Model 2 + Best Protein Set (functions of TSG101, total tau, pY-IRS1, pSer312-IRS1, p181-tau, Aβ42). B) Heat map of differences in AUC between models with 95% confidence intervals (off diagonal elements: column model *minus* row model). Diagonal elements show model AUC with 95% confidence intervals

eFigure 12. Boxplots of risk scores from the internally leave-10%-out cross-validated prediction models (BLSA training set)



The y-axis depicts estimated risk scores estimates using cross-validated data for AD and Control participants based on each model (range 0 -1). Risk scores were computed by converting logistic regression linear predictors to a 0-to-1 scale via the expit transformation: expit(x)=exp(x)/[1 + exp(x)]. Boxes depict the median and upper and lower quartiles; error bars depict 1.5 x interquartile range and single dots depict outliers.

eFigure 13. Receiver operating characteristic analysis for classification of JH ADRC participants into AD cases or Controls



ROC curves for the training (2/3) and test (1/3) sets of data. The model chosen based on training data included functions of nEV Concentration, nEV Average Diameter, TSG101, total tau, p181-tau, P-231-tau, pY-IRS1, pSer312-IRS1, and Aβ42. AUC for Training data is 98.9% and for Test data is 76.7%.

eResults.

EV biomarker differences between future AD and Control participants (whole BLSA cohort)

Wilcoxon rank-sum tests showed that future AD compared to Control BLSA participants at the last visit before AD onset had higher EV average diameter ($p < .001$); higher p231-tau ($p = .004$); higher p181-tau ($p < .001$); higher pY-IRS1 ($p = .002$); and higher pSer312-IRS1 ($p = .021$) (Supplemental Figure 9).

EV biomarker associations with cognition

Linear mixed model analysis revealed multiple cross-sectional and longitudinal associations between preclinical levels of several EV biomarkers and composite cognitive performance scores (Supplemental Table 7).

Assessment of missing values

Supplemental Table 8 describes an assessment of missing values in the BLSA by comparing BLSA participants who were included in prediction model building to those who were excluded from prediction model building with respect to AD status (vs. control status), sex, age, and year of birth.

AD compared to Control participants also had higher slopes (percent change in levels per year) of EV average diameter ($p < .001$); higher p231-tau ($p = .013$); higher p181-tau ($p < .001$); higher pY-IRS1 ($p = .008$); and higher pSer312-IRS1 ($p = .031$) (Supplemental Figure 10).

Stability of EV biomarkers over time

All biomarkers showed high within-subject stability over the time interval of the analyzed samples (Supplemental Table 3). This stability over time was similar for future AD and Control subjects except for Aβ42. These results agree with a recent finding for total EVs [15] and suggest that individuals have unique signatures in terms of concentration and composition of circulating EVs.

Inter-correlations between EV biomarkers

Supplemental Table 4 describes inter-correlations between means of EV biomarkers across all subjects and visits. Supplemental Tables 5 and 6 describe inter-correlations between EV biomarkers for last preclinical visits and slopes.

eDiscussion.

L1CAM is a transmembrane protein highly expressed in the brain, but also expressed in other tissues and cells including lymphoid and myeloid monocytic cells, renal tubule epithelial cells, and intestinal crypt cells (see the protein atlas entry for L1CAM: https://www.proteinatlas.org/ENSG00000198910-L1CAM/tissue). The basis for considering L1CAM+ EVs to be enriched for neuronal origin is the fact that they contain many-fold higher levels of multiple neuronal markers compared to total plasma EVs and subpopulations of plasma EVs immunoprecipitated by alternative antibodies (e.g. anti-IgG2a, CD81, GLAST). In a methods paper undertaken in conjunction with the present study, we showed that L1CAM+ EVs are enriched for several neuronal markers [3]. In recent publications, we showed that L1CAM+ EVs are enriched for multiple neuronal proteins based on targeted and untargeted proteomics [4], and demonstrated high-yield immunocapture of known neuronal EVs (rat neuronal culture supernatant EVs) resuspended in human EV-depleted plasma (see supplemental material in [5]). Here, we provide additional evidence for neuronal enrichment in terms of levels of Neurofilament light, Synaptophysin, L1CAM and NCAM (Supplemental Figures 4 and 5). The degree of neuronal marker enrichment achieved by immunoprecipitation against L1CAM ranges from 4.2 – 8.2-fold depending on the marker and the technique used to assess it. We are hopeful that future technological developments and refinements in the nEV isolation methodology may provide even greater level of enrichment and, perhaps, an even greater signal to noise ratio for biomarker studies.

Normalization for differential EV yield across samples (i.e. differential recovery of particle concentrations and sizes) was handled through the painstaking process of subjecting each sample to NTA to calculate its average EV diameter and concentration. Their NTA profiles, as well as the EM images, suggest that L1CAM+ EVs were a mixture of exosomes and larger microvesicles. Having both types of EVs in our isolations is by no means unwanted, since the origin of biomarkers may also be dual. Moreover, exosomes and microvesicles are progressively being seen as a continuum in terms of size distribution, content and biological functions [11]. EV concentration and EV average diameter were entered as covariate into the model for individual biomarkers, as well as into the final model. This approach differs from the approach taken in the original studies that relied for normalization on measurements of CD81 [8,9], but follows the practice in our more recent studies [16]. Covarying these variables in the model allows us to explicitly assess the variability explained by these normalizers rather than mask it into a ratio. Moreover, normalizing by EV concentration and EV average diameter is supported by the fact that they are

remarkably stable over time in both AD and Control participants (Supplemental Table 1), as previously observed for total plasma EVs [15]. Finally, the logical basis for normalizing by CD81 was undermined by the recent findings that this protein (and in fact any tetraspanin) is present in only a fraction of plasma EVs and is not particularly enriched in smaller EVs in the exosome range [11]. Among EV markers present in the exosomal fraction of plasma EVs, the most consistently expressed was the ESCRT-associated TSG101, TSG101 and, for this reason, we developed an assay for it and measured it in all samples. It is notable that TSG101 showed strong correlations with p231-tau, p181-tau, pY-IRS1, and pSer312-IRS1 (but not Aβ42) suggesting that these biomarkers may also be derived primarily from endosomal-origin EVs. The fact that TSG101 did not differ between AD and Control participants suggests that biomarker differences that contribute to AD diagnosis cannot be attributed to a difference in general EV content. Nevertheless, the finding of higher EV average diameter of neuronal-enriched EVs in AD is novel, potentially relevant to pathogenic mechanisms in AD and should be the subject of future research.

Combining biomarkers in ratios or composite models, rather than examining them separately, seems to be key for achieving high diagnostic accuracy [17-20], perhaps due to the pathogenic complexity of AD that cannot be reduced into a single process. In our study, the final model (which incorporated measures of longitudinal TSG101, total tau, pY-IRS1, pSer312-IRS1, and p181-tau) performed better than models including each one of these nEV biomarkers separately. This may reflect the complex etiopathogenesis of AD, which involves multiple pathways. The biomarker panel examined in this study reflects critical events in AD pathogenesis: brain amyloidosis (Aβ42), insulin resistance (pSer312-IRS1, pY-IRS1), tau hyperphosphorylation (p181-tau, p231-tau), and neurodegeneration (total tau). In general, studies reporting classification AUCs > 90% often define groups based on “gold standard” biomarkers (typically amyloid PET) [17,18]. Although defining cases based on established biomarkers decreases the risk of mis-classification, there are ecological validity advantages for assessing biomarker performance against clinical AD diagnosis, as we did in the present study. The potential for a degree of misclassification in the cohorts we studied suggests that our results are conservative estimates of model performance.

Aggregation-competent full-length tau is present in CSF and plasma EVs [10]. EVs released by neurons with tau deposits contain tau and can be uptaken by unaffected neurons, which subsequently develop similar deposits, [21] mediating the spread of tau pathology [22].. Regarding total tau, studies measuring it in plasma have shown it to be a weaker biomarker than p-tau [23,24]. Multiple lines of evidence suggest that total tau in biofluids reflects a different aspect of disease pathogenesis compared to p-tau; in the A/T/N classification scheme, CSF p-tau is considered

evidence for hyperphosphorylated tau pathology (T), whereas total tau signifies neurodegeneration (N) [25,26]. A recent two-cohort study suggests that total tau soluble in plasma is a modest predictive biomarker of AD, even though levels between future AD patients and Controls showed significant overlap [27].

Tau phosphorylation by a variety of kinases is the key molecular event leading to tau aggregation and involves several Thr and Ser residues in a progressive sequence. Phosphorylation of Thr231 is an early event, whereas phosphorylations of Thr181 and Ser396 are both considered late events [28,29]. To capture both early and late events in the development of tau pathology, and given reports that CSF p231-tau outperforms p181-tau as a diagnostic biomarker [30-32], we chose to assay Thr231 and Thr181, whereas the initial study probed Thr181 and Ser396 [8]. We found that the diagnostic performance of p181-tau exceeded that of p231-tau and that p181-tau was associated with cognition. All three tau biomarkers (p181-tau, p231-tau and total tau) entered the final model suggesting that they may provide complementary information. A recent study of plasma p181-tau (similarly measured by an electrochemiluminescence assay) showed that its levels were higher in AD compared to Control participants and were associated with brain uptake of both Aβ and tau ligands [24]. Whereas the diagnostic performance of soluble p181-tau was higher than previously seen, it was nevertheless lower than the performance of p181-tau in neuronal-enriched EVs in the present study. Besides its performance in AD prediction, higher p181-tau was cross-sectionally (but not longitudinally) associated with lower verbal memory, attention, executive function, and visuospatial function suggesting an association already established during the preclinical period and not subject to change during the transition to clinical disease.

Regarding Aβ42, we previously showed that most EV-associated Aβ is located on the outer membrane surface rather than being true intravesicular cargo [33]. This is further suggested by the fact that Aβ42, unlike the (likely intravesicular) cargo of p231-tau, p181-tau, pY-IRS1, and pSer312-IRS1, showed borderline/negative correlations with the constitutional exosomal marker TSG101. Interestingly, despite the lack of significant differences between future AD and Control participants and despite the fact that it was not included in the model, Aβ42 showed longitudinal relationships with verbal memory and language suggesting that nEV Aβ42 may reflect preclinical cognitive changes.

It should be noted that levels of pY-IRS1 were higher in future AD cases than Control participants (Fig 2) dissimilar to the pattern seen in the index study [9]. This surprising result may be due to a difference in the Tyr epitopes recognized by the detection antibody in the electrochemiluminescence assay used in the present study compared to the previously used

ELISA; this is further suggested by the moderate correlation between the two assays (R^2 = .18; data not shown).

The biomarker panel examined in this study potentially reflects some critical events in AD events: brain amyloidosis (Aβ42), insulin resistance (pSer312-IRS1 and pY-IRS1), tau hyperphosphorylation (p181-tau and p231-tau), and neurodegeneration (total tau). These events do not exhaust the spectrum of AD pathophysiology and, as we have shown in recent studies, additional pathogenenic events may in fact be interrogated through EV biomarkers including astrocyte-mediated toxicity [6,7], lysosomal dysfunction [34], inadequate stress-response mechanisms [35] and synaptic degeneration . Whereas sample availability and workload considerations prevented us from examining all these candidates in the present study, future studies should examine whether combinations of even more EV markers further improve prediction performance or track clinical disease. Moreover, future technological developments and refinements of the technique for neuronal enrichment (e.g. different or additional antibodies for positive selection) may provide further improvements in its ability to detect neuronal signals in the noise of circulating EVs of multiple origins.

eReferences.

1. Saenz-Cuesta M, Arbelaiz A, Oregi A, et al. Methods for extracellular vesicles isolation in a hospital setting. *Front Immunol.* 2015;6:50.
2. Faure J, Lachenal G, Court M, et al. Exosomes are released by cultured cortical neurones. *Mol Cell Neurosci.* 2006;31(4):642-648.
3. Mustapic M, Eitan E, Werner JK, Jr., et al. Plasma Extracellular Vesicles Enriched for Neuronal Origin: A Potential Window into Brain Pathologic Processes. *Front Neurosci.* 2017;11:278.
4. Pulliam L, Sun B, Mustapic M, Chawla S, Kapogiannis D. Plasma neuronal exosomes serve as biomarkers of cognitive impairment in HIV infection and Alzheimer's disease. *J Neurovirol.* 2019.
5. Athauda D, Gulyani S, Karnati H, et al. Utility of Neuronal-Derived Exosomes to Examine Molecular Mechanisms That Affect Motor Function in Patients With Parkinson Disease: A Secondary Analysis of the Exenatide-PD Trial. *JAMA Neurol.* 2019.
6. Goetzl EJ, Schwartz JB, Abner EL, Jicha GA, Kapogiannis D. High complement levels in astrocyte-derived exosomes of Alzheimer disease. *Ann Neurol.* 2018;83(3):544-552.
7. Goetzl EJ, Mustapic M, Kapogiannis D, et al. Cargo proteins of plasma astrocyte-derived exosomes in Alzheimer's disease. *FASEB J.* 2016;30(11):3853-3859.
8. Fiandaca MS, Kapogiannis D, Mapstone M, et al. Identification of preclinical Alzheimer's disease by a profile of pathogenic proteins in neurally derived blood exosomes: A case-control study. *Alzheimers Dement.* 2015;11(6):600-607 e601.
9. Kapogiannis D, Boxer A, Schwartz JB, et al. Dysfunctionally phosphorylated type 1 insulin receptor substrate in neural-derived blood exosomes of preclinical Alzheimer's disease. *FASEB J.* 2015;29(2):589-596.
10. Guix FX, Corbett GT, Cha DJ, et al. Detection of Aggregation-Competent Tau in Neuron-Derived Extracellular Vesicles. *Int J Mol Sci.* 2018;19(3).
11. Kowal J, Arras G, Colombo M, et al. Proteomic comparison defines novel markers to characterize heterogeneous populations of extracellular vesicle subtypes. *Proceedings of the National Academy of Sciences of the United States of America.* 2016;113(8):E968-977.
12. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics.* 1988;44(3):837-845.
13. van der Laan MJ, Polley EC, Hubbard AE. Super learner. *Stat Appl Genet Mol.* 2007;6.
14. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S plus to analyze and compare ROC curves. *Bmc Bioinformatics.* 2011;12.
15. Eitan E, Green J, Bodogai M, et al. Age-Related Changes in Plasma Extracellular Vesicle Characteristics and Internalization by Leukocytes. *Sci Rep.* 2017;7(1):1342.
16. Eitan E, Tosti V, Suire CN, et al. In a randomized trial in prostate cancer patients, dietary protein restriction modifies markers of leptin and insulin signaling in plasma extracellular vesicles. *Aging Cell.* 2017;16(6):1430-1433.

17. Nakamura A, Kaneko N, Villemagne VL, et al. High performance plasma amyloid-beta biomarkers for Alzheimer's disease. *Nature.* 2018;554(7691):249-254.
18. Hansson O, Seibyl J, Stomrud E, et al. CSF biomarkers of Alzheimer's disease concord with amyloid-beta PET and predict clinical progression: A study of fully automated immunoassays in BioFINDER and ADNI cohorts. *Alzheimers Dement.* 2018.
19. Burnham SC, Rowe CC, Baker D, et al. Predicting Alzheimer disease from a blood-based biomarker profile: A 54-month follow-up. *Neurology.* 2016;87(11):1093-1101.
20. Varma VR, Oommen AM, Varma S, et al. Brain and blood metabolite signatures of pathology and progression in Alzheimer disease: A targeted metabolomics study. *PLoS Med.* 2018;15(1):e1002482.
21. Wang Y, Balaji V, Kaniyappan S, et al. The release and trans-synaptic transmission of Tau via exosomes. *Mol Neurodegener.* 2017;12(1):5.
22. Saman S, Kim W, Raya M, et al. Exosome-associated tau is secreted in tauopathy models and is selectively phosphorylated in cerebrospinal fluid in early Alzheimer disease. *J Biol Chem.* 2012;287(6):3842-3849.
23. Mattsson N, Zetterberg H, Janelidze S, et al. Plasma tau in Alzheimer disease. *Neurology.* 2016;87(17):1827-1835.
24. Mielke MM, Hagen CE, Xu J, et al. Plasma phospho-tau181 increases with Alzheimer's disease clinical severity and is associated with tau- and amyloid-positron emission tomography. *Alzheimers Dement.* 2018.
25. Jack CR, Jr., Bennett DA, Blennow K, et al. A/T/N: An unbiased descriptive classification scheme for Alzheimer disease biomarkers. *Neurology.* 2016;87(5):539-547.
26. Jack CR, Jr., Bennett DA, Blennow K, et al. NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. *Alzheimers Dement.* 2018;14(4):535-562.
27. Pase MP, Beiser AS, Himali JJ, et al. Assessment of Plasma Total Tau Level as a Predictive Biomarker for Dementia and Related Endophenotypes. *JAMA Neurol.* 2019.
28. Luna-Munoz J, Chavez-Macias L, Garcia-Sierra F, Mena R. Earliest stages of tau conformational changes are related to the appearance of a sequence of specific phospho-dependent tau epitopes in Alzheimer's disease. *J Alzheimers Dis.* 2007;12(4):365-375.
29. Nakamura K, Greenwood A, Binder L, et al. Proline isomer-specific antibodies reveal the early pathogenic tau conformation in Alzheimer's disease. *Cell.* 2012;149(1):232-244.
30. Spiegel J, Pirraglia E, Osorio RS, et al. Greater specificity for cerebrospinal fluid P-tau231 over P-tau181 in the differentiation of healthy controls from Alzheimer's disease. *J Alzheimers Dis.* 2016;49(1):93-100.
31. Buerger K, Teipel SJ, Zinkowski R, et al. CSF tau protein phosphorylated at threonine 231 correlates with cognitive decline in MCI subjects. *Neurology.* 2002;59(4):627-629.
32. Ewers M, Buerger K, Teipel SJ, et al. Multicenter assessment of CSF-phosphorylated tau for the prediction of conversion of MCI. *Neurology.* 2007;69(24):2205-2212.
33. Eitan E, Hutchison ER, Marosi K, et al. Extracellular Vesicle-Associated Abeta Mediates Trans-Neuronal Bioenergetic and Ca(2+)-Handling Deficits in Alzheimer's Disease Models. *NPJ Aging Mech Dis.* 2016;2.
34. Goetzl EJ, Boxer A, Schwartz JB, et al. Altered lysosomal proteins in neural-derived plasma exosomes in preclinical Alzheimer disease. *Neurology.* 2015;85(1):40-47.

35. Goetzl EJ, Boxer A, Schwartz JB, et al. Low neural exosomal levels of cellular survival factors in Alzheimer's disease. *Ann Clin Transl Neurol.* 2015;2(7):769-773.