RSC Advances

PAPER

View Article Online
View Journal | View Issue

Cite this: *RSC Adv.*, 2026, **16**, 22252

LOESS-based normalization workflow for targeted HDL glycoproteomics in an Alzheimer's disease cohort

Brian V. Hong,[a] Yiyun Liu,[b] Armin Oloumi,[b] Joanne K Agus,[a] Yasmine Bouchibti,[b] Lee-Way Jin,[c] Izumi Maezawa,[c] Danielle J. Harvey,[d] Carlito B. Lebrilla[b] and Angela M. Zivkovic *[a]

High-density lipoprotein (HDL) carries proteins and glycoproteins involved in lipid metabolism and inflammatory regulation, yet quantitative characterization of HDL-associated peptides in Alzheimer's disease (AD) cohorts remains challenging due to small biological effect sizes superimposed on substantial technical variability. We applied a Locally Estimated Scatterplot Smoothing (LOESS)-based drift correction and internal-standard-guided normalization workflow to targeted multiple reaction monitoring (MRM) glycoproteomic data generated from HDL isolates collected from 194 participants spanning the cognitive spectrum. Of the 164 transitions originally targeted, 59 features passed quality control (QC), and 21 HDL-associated peptide and glycopeptide features showed consistent signal across all 194 samples; these 21 analytes were used for analysis. Normalization improved analytical reproducibility, reducing median HDL pooled QC coefficients of variation from 69.1% to 55.2%. *APOE* genotype analyses identified six peptides with statistically significant differences between *APOE3/E3* and *APOE3/E4* carriers, five of which remained statistically significant after false-discovery rate correction, and all six of which remained significant in covariate-adjusted models, whereas disease-related differences within *APOE3/E3* carriers were modest and did not remain statistically significant after covariate adjustment. These findings demonstrate that LOESS-based drift correction combined with feature-specific internal-standard selection stabilizes quantitative HDL glycoproteomic measurements and support downstream comparisons. This workflow provides a practical framework for QC-informed normalization in targeted glycoproteomics and highlights *APOE*-associated variation in HDL peptides within an aging clinical cohort.

Received 17th February 2026
Accepted 20th April 2026

DOI: 10.1039/d6ra01426g

rsc.li/rsc-advances

1. Introduction

Glycosylation changes in cancer are often large, allowing group differences to be resolved using conventional analytical pipelines. For example, advanced cancers exhibit marked shifts in serum glycan profiles that can reliably separate patient groups.[1] In contrast, glycomic and glycoproteomic alterations associated with Alzheimer's disease (AD) and related dementias tend to be more modest in magnitude and heterogeneous across cohorts, particularly in peripheral blood. Recent serum and whole-blood glycomic studies report modest average changes between AD patients and age-matched controls, with alterations sometimes restricted to specific subgroups rather than the entire disease population.[2,3] These small effect sizes highlight the need for analytical strategies that improve signal-to-noise and reduce the risk that technical variability obscures true biological variation.

Older individuals with neurodegenerative disease also display substantial inter-individual variability due to aging, comorbidities, and metabolic factors. A recent population of 1516 adults reported that aging is accompanied by a shift toward simpler N-glycan profiles, including higher levels of agalactosylated and other low-complexity structures, with individuals exhibiting greater metabolic comorbidity disproportionately represented in this "simple-glycan" cluster.[4] Such background variation complicates efforts to detect disease-specific glycosylation signatures in AD and increases the likelihood that apparent group differences reflect unrelated physiological processes rather than disease biology. Even well-studied circulating AD markers display inconsistent patterns across disease stage and biological compartments. For example, plasma clusterin has been reported to be elevated in AD dementia, but findings in mild cognitive impairment (MCI) are inconsistent, and serum and cerebrospinal fluid concentrations show only modest or absent correspondence.[5] These factors

[a]*Department of Nutrition, University of California-Davis, Davis, CA 95616, USA. E-mail: amzivkovic@ucdavis.edu*
[b]*Department of Chemistry, University of California-Davis, Davis, CA 95616, USA*
[c]*Department of Pathology and Laboratory Medicine, School of Medicine, University of California-Davis, Davis, CA 95817, USA*
[d]*Department of Public Health Sciences, University of California-Davis, Davis, CA 95616, USA*

underscore the analytical challenges inherent to biomarker discovery in heterogeneous aging populations.

High-density lipoproteins (HDL) carry a defined set of apolipoproteins and immune-related proteins involved in lipid transport and inflammation, including apolipoprotein A-1 (ApoA1), apolipoprotein E (ApoE), apolipoprotein D (ApoD), apolipoprotein M (ApoM), complement components, and acute-phase proteins.[6] HDL has also been linked to vascular and amyloid-related processes relevant to AD physiology.[7] In our previous work, HDL isolated from individuals with MCI and AD exhibited reduced lecithin-cholesterol acyltransferase (LCAT) activity and smaller particle size, and both measures were related to cognitive performance.[8] These findings support the idea that HDL provides a controlled biochemical matrix in which subtle AD-related alterations may be easier to resolve than in unfractionated plasma or serum.

The analytical challenge addressed in this study is that glycoproteomic measurements of HDL often contain small biological effect sizes superimposed on substantial technical variability, including run-order drift, plate-to-plate differences, and intensity-dependent noise. Standard normalization approaches are frequently inadequate for data of this type. Demonstrating an effective normalization strategy for HDL-targeted glycoproteomic measurements is therefore important because this matrix exhibits distinct drift and variability patterns and no standardized workflow has yet been established. Here, we evaluated a combined strategy incorporating internal standards and Locally Estimated Scatterplot Smoothing (LOESS)-based correction to improve peptide-level reproducibility in a large HDL glycoproteomic dataset from 194 participants. We assessed performance using objective metrics, including coefficients of variation, stabilization rates, and preservation of biologically interpretable signal distributions. This workflow provides a generalizable approach for improving data quality in glycoproteomic studies where disease-associated differences are subtle and technically challenging to detect.

2. Results and discussion

2.1 Participant characteristics

Participant characteristics for the 194 individuals are summarized in Table S1. The cohort included 83 controls, 40 individuals with MCI, and 71 with AD dementia. All experiments were performed in accordance with the Guidelines of the Declaration of Helsinki, and experiments were approved by the ethics committee at the University of California at Davis. Informed consents were obtained from human participants of this study. Age differed modestly across groups, whereas sex, ethnicity, BMI, and hypertension were broadly similar. Several metabolic variables (for example BMI and history of diabetes, hypercholesterolemia, and hypertension) showed group differences or trends and had incomplete data. Because these clinical variables represent downstream metabolic or vascular conditions that may themselves be influenced by cognitive status or *APOE* genotype, they were not used as covariates in peptide-level models. All participants had complete age and ethnicity data and were included in covariate-adjusted models. The final analytic dataset included *APOE3/E3* and *APOE3/E4* carriers only, with both genotype groups represented across all diagnostic groups.

2.2 LOESS drift correction of internal standards

Our approach builds on and extends prior work on QC-based LOESS normalization in mass-spectrometry omics. Comparative evaluations across multi-omics datasets have identified LOESS and related QC-anchored LOESS procedures, together with Probabilistic Quotient Normalization (PQN) and median scaling, as among the most robust strategies for improving QC feature consistency while preserving treatment- and time-related variance in metabolomics, lipidomics, and proteomics.[9] In parallel, previous HDL glycoproteomic workflows from our group and others have typically relied on either a single synthetic internal standard or on normalization of glycopeptides to nonglycosylated peptides from the same protein, without explicit drift modeling, and have generally been applied in relatively small pilot cohorts (~5–80 participants).[10–13] In contrast, the present study applies a QC-informed, multi-internal-standard framework in 194 older adults. Rather than applying LOESS directly to all features, we fitted LOESS models to each internal-standard peptide as a function of injection order within pooled HDL QC injections and used the LOESS-corrected internal standards for feature-level scaling. This design accounts for run-order drift across the full acquisition sequence and allows each analyte to be normalized to the internal standard that minimizes QC variance, which is particularly valuable in a heterogeneous matrix such as HDL where no single internal standard performs optimally for all targets.

Fig. 1 outlines the QC-informed normalization workflow used in this study, providing the basis for the LOESS-based drift correction and downstream analyses described below. The LOESS-based drift correction applied to the internal-standard peptides in the pooled HDL QC samples effectively removed non-linear intensity fluctuations associated with injection order, producing a stabilized baseline suitable for downstream normalization (Fig. 2). Across the QC panel, the median coefficient of variation decreased from 69.1% before normalization to 55.2% after LOESS correction and internal-standard scaling, with 55 of 59 QC-passing features (93.2%) exhibiting improved reproducibility. Pre- and post-normalization CVs for all QC-passing features are provided in Table S2. Of the 59 QC-robust features, 21 demonstrated measurable signal across the 194 participant samples and were carried forward into differential-expression analyses (Table S3). The feature-specific Sigma peptide yielding the lowest post-normalization CV for each analyte is listed in Table S4. These QC metrics indicate that the LOESS- and internal-standard-based normalization framework effectively stabilized technical variation in the HDL glycoproteomic panel.

2.3 Differential expression analysis of HDL-associated glycopeptides

Across the 21 HDL-associated peptide and glycopeptide features that passed normalization and QC filtering, six showed nominally significant differences between *APOE3/E3* and *APOE3/E4* carriers,

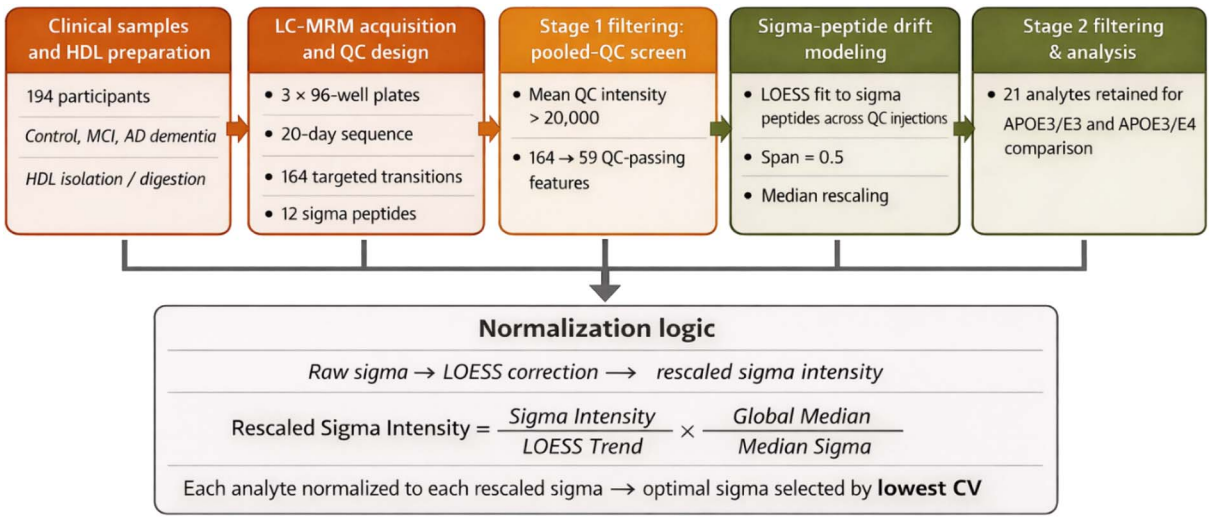

Fig. 1 Schematic of LOESS normalization workflow for targeted HDL glycoproteomics.

regardless of diagnosis status (Fig. 3). These included A1AT_271 MC_5402 ($p = 0.021$), APOD_VLNQELR ($p = 0.011$), APOE_AATVGSLAGQPLQER_z2 ($p < 0.001$), APOE_LGPLVEQGR ($p < 0.001$), APOM_AFLLTPR ($p = 0.011$), and APOM_SLTSCLDSK ($p = 0.002$). Across these six peptides, median $\log_2$ intensities were consistently lower in *APOE3/E4* compared with *APOE3/E3* carriers, with genotype-associated differences ranging from −0.27 to −0.65 $\log_2$ units, corresponding to approximately 17 to 36% lower abundance in *APOE3/E4* carriers. The largest difference was observed for A1AT_271 MC_5402 (−0.65 $\log_2$ units, ~36% lower).

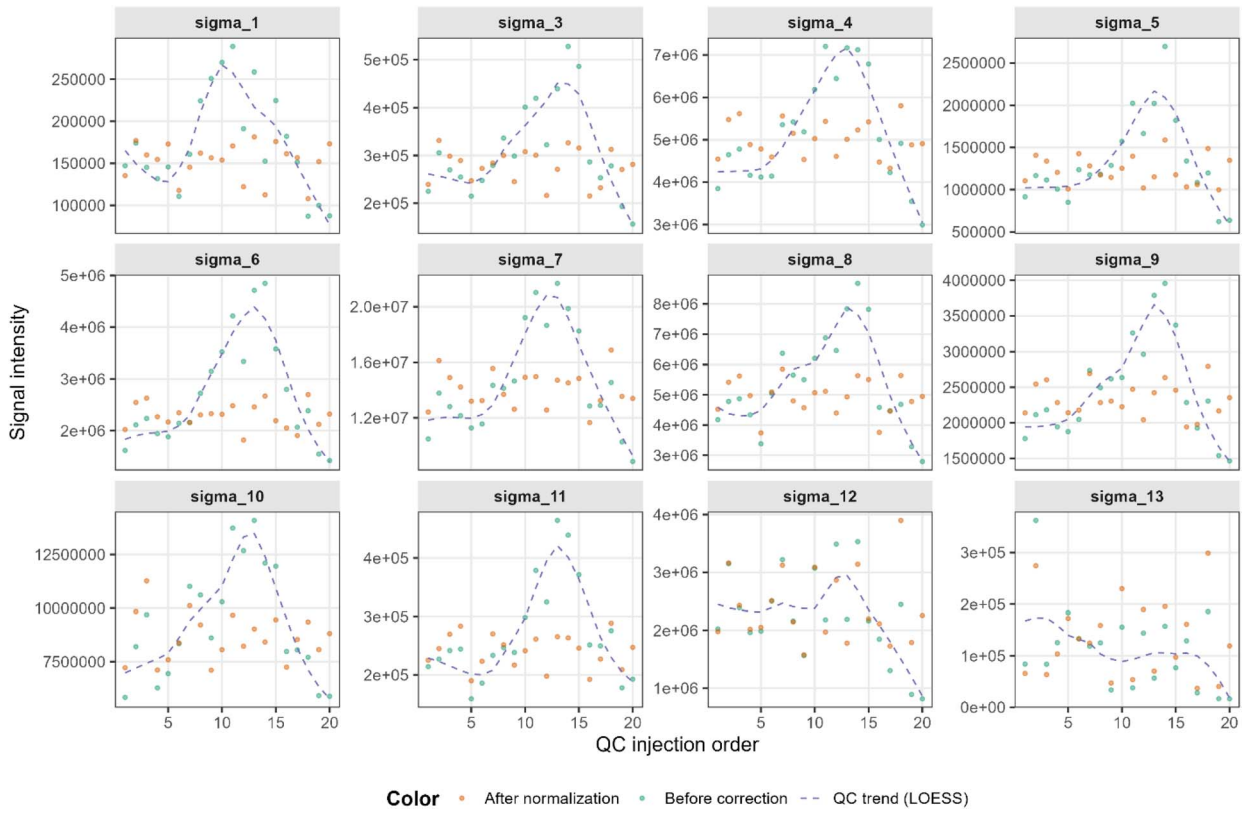

Fig. 2 LOESS drift correction of internal standard (Sigma) peptides across the analytical run. Raw intensities for 12 Sigma internal standards (green points) demonstrated pronounced non-linear injection-order drift, with elevated mid-run signal followed by progressive decline. LOESS smoothing (blue dashed line) modeled the systematic trend for each internal standard using pooled HDL QC injections distributed across all plates. Application of LOESS-based correction and internal-standard rescaling effectively removed this drift, producing stabilized post-normalization intensities (orange points) that no longer followed the injection-order trajectory. Each facet represents one Sigma peptide, displayed on its raw intensity scale.

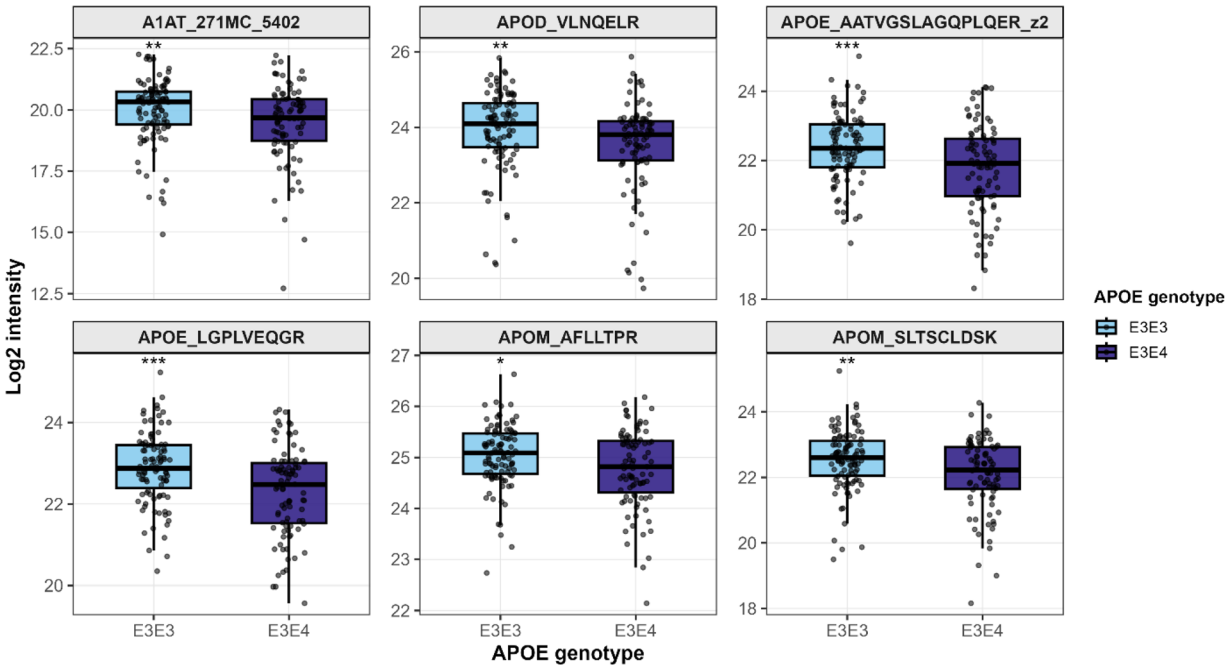

Fig. 3 Differential expression of HDL-associated peptides between *APOE3/E3* and *APOE3/E4* carriers regardless of diagnosis status. Box-and-scatter plots show $\log_2$-normalized intensities for six peptides with nominal genotype-associated differences. Asterisks indicate nominal significance from unadjusted Wilcoxon rank-sum tests ($p < 0.05$). FDR-adjusted q values and covariate-adjusted linear-model p values are reported in Table S5. All six features also remained significant after covariate adjustment for age, ethnicity, and clinical status. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$.

Five of the six peptides met the significance threshold after FDR correction ($q < 0.05$), and all six remained statistically significant in covariate-adjusted linear models that accounted for age, ethnicity, and clinical status, with adjusted p values ranging from 5.33×10^{-5} to 0.019. Complete statistical results, including unadjusted p values from Wilcoxon rank-sum tests, FDR-adjusted q values, covariate-adjusted p values, and direction of effect, are provided in Table S5.

Although prior HDL glycoproteomic studies have not stratified by *APOE* genotype, our previous work has shown that site-specific glycoforms of HDL-associated ApoE and other glycoproteins are strongly associated with HDL cholesterol efflux capacity and immunomodulatory function, suggesting that relatively small changes in HDL glycoprotein composition can have functional consequences.[14] Related intervention work in children has further demonstrated that nutritionally induced changes in HDL glycopeptide profiles, including A1AT and other acute-phase proteins, track with changes in cholesterol efflux capacity, reinforcing the functional relevance of HDL glycoprotein composition.[13] In the present cohort, we did not directly quantify ApoE glycopeptides but instead observed genotype effects on ApoE peptides and an A1AT glycopeptide, consistent with the idea that *APOE* genotype may influence both ApoE content and the broader HDL glycoprotein milieu. A1AT is an anti-inflammatory acute-phase component of HDL, whereas ApoM is a lipid-binding HDL apolipoprotein linked to endothelial and lipoprotein signaling.[6] Lower abundance of these peptides in *APOE3/E4* carriers may therefore reflect broader genotype-associated differences in inflammatory regulation and HDL functional composition, although the present data do not establish mechanisms. These peptide-level differences are qualitatively compatible with epidemiologic data showing that plasma and HDL ApoE concentrations vary by *APOE* genotype and that higher ApoE content in specific HDL subspecies has been associated with better cognitive performance and lower dementia risk.[15] Our results are not intended to establish clinical biomarkers; rather, they demonstrate that once technical variation is controlled, subtle genotype-associated differences in HDL-associated peptides can be detected in a moderately sized cohort.

2.4 Disease-group differences within *APOE3/E3*

Disease-related differences were examined within *APOE3/E3* carriers. Two analytes showed nominally statistically significant variation across clinical groups. APOA1_LAEYHAK differed between control and MCI participants ($p = 0.036$), and APOM_AFLLTPR also showed a difference between control and MCI ($p = 0.044$), with both analytes exhibiting lower abundance in MCI than in controls (Fig. 4). No significant differences were observed for control *versus* AD or for MCI *versus* AD. In covariate-adjusted linear models that included age and ethnicity, none of the disease-related associations remained significant after FDR correction. *APOE3/E4* carrier status did not modify disease-related patterns; no peptide or glycopeptide showed significant clinical-group differences within *APOE3/E4* carriers, consistent with the exploratory Wilcoxon results. This is consistent with the broader difficulty of identifying strong peripheral protein or lipid signatures of AD in bulk blood

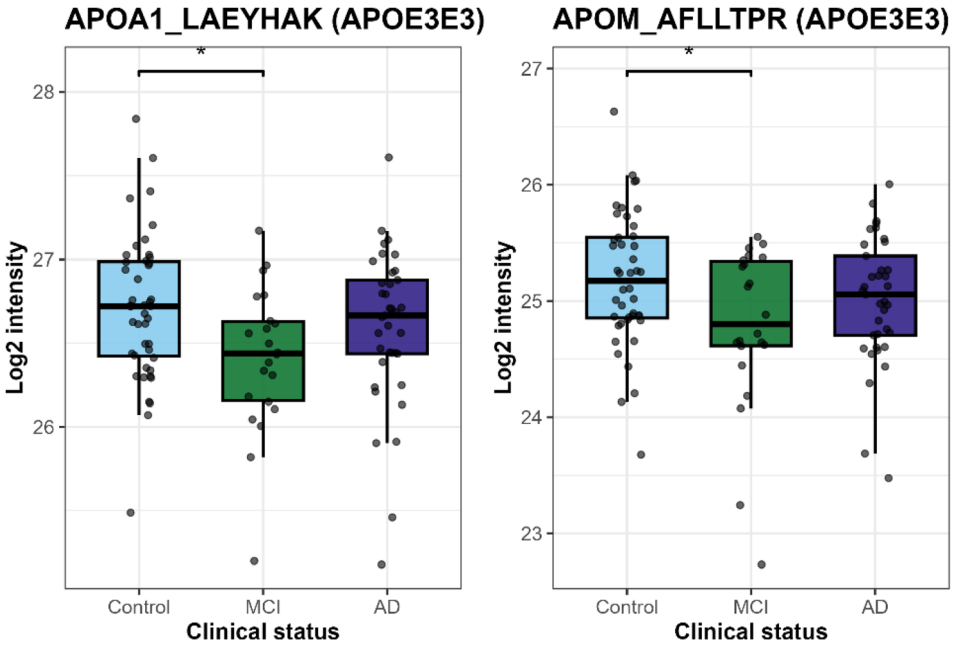

Fig. 4 Comparison of HDL-associated peptide intensities across clinical groups (Control, MCI, AD) within *APOE3/E3* carriers. Two peptides (APOA1_LAEYHAK and APOM_AFLLTPR) showed nominal differences between Control and MCI by Wilcoxon rank-sum testing. No other pairwise comparisons were significant, and no disease-related associations remained significant after covariate adjustment. *$P < 0.05$.

compartments, where effect sizes tend to be modest and often require very large cohorts or multiplexed biomarker panels to reach clinical significance. Because the analytical panel was applied to a single bulk HDL fraction, any disease-related signal would be distributed across heterogeneous HDL subspecies, attenuating detectable group differences. In addition, because the cohort was of moderate size and covariate adjustment was limited to age, ethnicity, and clinical status, residual confounding by unmeasured metabolic, vascular, medication-related, or other clinical factors cannot be excluded. The present results should therefore be interpreted as demonstrating the utility of the normalization framework within this cohort rather than as definitive estimates of genotype- or disease-specific effect sizes.

A key implication of these findings is that bulk HDL measurements, even when informed by targeted glycoproteomics, may obscure subclass-specific changes that are biologically important. HDL is a structurally and compositionally heterogeneous family of particles, with established subclasses that differ in density, size, and protein cargo.[6] Glycomic and glycoproteomic analyses have shown that HDL carries a highly sialylated and compositionally distinct glycome compared with plasma and that proteins such as A1AT, APOC3, and clusterin exhibit HDL-enriched glycoforms, underscoring glycosylation as an additional axis of HDL specialization.[10,11] Proteomic analyses of HDL subclasses have shown that large and small HDL fractions carry distinct protein complements and exhibit divergent functional capacities, including differences in cholesterol efflux and antioxidant activity.[16,17] Consistent with this, size-based isolation of HDL demonstrates that ApoE is substantially more abundant in the largest HDL fractions than in smaller particles, indicating marked enrichment of ApoE on large HDL in plasma.[18] Emerging work on ApoE-containing HDL and other protein-defined HDL subspecies links these subclass-specific proteotypes to cardiometabolic and vascular outcomes and to dementia risk.[18–20] In parallel, large cohort and imaging studies show that peripheral ApoE concentration and isoform balance in plasma are associated with dementia-associated mortality and with AD-relevant measures of gray matter volume and cerebral glucose metabolism.[21,22] Isoform-resolved analyses comparing central and peripheral ApoE further demonstrate that these pools are only partially coupled, with *APOE3/E4* heterozygotes exhibiting higher ApoE4/ApoE3 ratios in cerebrospinal fluid and brain but lower ratios in plasma.[23] These observations suggest that subclass-specific HDL proteotypes, and particularly ApoE-enriched HDL in plasma, may also be relevant to cognitive decline and dementia. In that context, the present study can be viewed as a bulk-HDL proof of principle: it establishes and tests a normalization framework that can be directly extended to HDL subclass-resolved datasets generated using newer isolation methods and higher-coverage proteotyping workflows. Future subclass-resolved HDL studies that combine improved isolation of large and small HDL fractions with QC-informed normalization frameworks like the one described here may be well positioned to localize ApoE-, A1AT, and ApoC3-containing glycoproteoforms to discrete HDL subclasses and to relate those subclass-specific profiles to cognitive trajectories and AD biomarkers.

3. Conclusions

This study implemented and evaluated a QC-informed normalization framework for targeted HDL glycoproteomics

in an aging cohort spanning cognitively normal controls, MCI, and AD dementia. The analytical sequence comprised about three weeks of continuous acquisition, monitoring 164 targeted transitions in 194 participants, with pooled HDL QC injections interleaved throughout to capture injection-order-dependent drift. Combining LOESS-based correction of internal-standard peptides in pooled HDL QC samples with feature-specific internal-standard selection at the transition level improved analytical stability for HDL-associated peptides and glycopeptides carried forward for downstream analyses. For a dataset of this scale, implementing transition-level drift modeling and internal-standard-based normalization corresponds to months of systematic data processing and QC review. Baseline behavior was estimated separately for each monitored transition and then propagated across the cohort. This work shows that rigorous, QC-based normalization can stabilize HDL-targeted glycoproteomic measurements across extended acquisition periods and enables detection of *APOE* genotype-associated differences in HDL-associated peptide and glycopeptide abundances. The LOESS- and internal-standard-based workflow described here is generalizable to other targeted lipoprotein assays and can be integrated with emerging HDL subclass-specific isolation and data-independent acquisition-based proteotyping strategies to support larger, more deeply phenotyped cohorts. As HDL isolation, subclass separation, and quantitative proteomics continue to advance, incorporating such normalization strategies will improve the reliability of HDL-focused glycoproteomic studies and enable detailed dissection of how HDL composition, *APOE* genotype, and subclass distribution intersect in the pathobiology of neurodegenerative disease.

4. Experimental

4.1 Participants and clinical characterization

This cross-sectional analysis utilized plasma samples from 194 older adults enrolled in a longitudinal cohort and maintained in the associated biorepository, including cognitively normal controls, individuals with MCI, and patients with AD dementia, as previously described.[8] Participants underwent standardized neuropsychological assessment and consensus clinical diagnosis. Cognitive status (control, MCI, AD dementia) was assigned using established criteria based on clinical history, neurological examination, functional status, and cognitive testing. Venous blood was collected in EDTA tubes, processed, and plasma was aliquoted and stored at −80 °C until analysis. All 194 individuals provided HDL samples with non-missing signal for each of the 21 analytes included in downstream analyses and had complete age and ethnicity data for covariate-adjusted models. Other metabolic and vascular clinical variables (for example BMI and histories of diabetes and hypercholesterolemia), several of which had incomplete data and all of which may represent downstream phenotypes influenced by cognitive status or *APOE* genotype, were not used as covariates.

All procedures were approved by the Institutional Review Board of the University of California, Davis, and all participants or their legally authorized representatives provided written informed consent.

4.2 HDL isolation

HDL particles were isolated from frozen EDTA plasma using a two-step procedure consisting of sequential flotation ultracentrifugation followed by size exclusion chromatography, as described previously.[18] In brief, 500 μL of thawed plasma were underlaid beneath potassium bromide (KBr) density solution (d = 1.006 g mL^{-1}) in OptiSeal tubes and centrifuged at 110 000 rpm at 15 °C to remove chylomicrons and very-low-density lipoproteins. The infranatant was then adjusted to d = 1.21 g mL^{-1} with KBr and centrifuged again to separate lipoproteins from denser plasma proteins. The fraction with density 1.006–1.21 g mL^{-1}, containing IDL, LDL, and HDL, was recovered, dialyzed into phosphate-buffered saline, and loaded onto a Superdex 200 Increase 10/300 GL size exclusion column on an ÄKTA FPLC system.[18] Fractions corresponding to the HDL peak were pooled based on the UV absorbance chromatogram and used for subsequent proteomic and glycoproteomic analyses. All samples in the present study were processed using the same isolation protocol and batch structure.

4.3 Glycoproteomic data acquisition, quality control, and normalization

HDL isolates were prepared for targeted multiple reaction monitoring (MRM) LC-MS analysis by reduction, alkylation, and overnight tryptic digestion in ammonium bicarbonate buffer using an established in-plate workflow.[12] Briefly, samples were processed in a 96-well format with freshly prepared 50 mM ammonium bicarbonate, reduced with dithiothreitol (DTT) at elevated temperature, alkylated with iodoacetamide (IAA) in the dark, and digested with sequencing-grade trypsin for 18 h at 37 °C. Proteolysis was quenched by acidification (formic acid (FA)) prior to MRM LC-MS analysis.[12] This reduction, alkylation, and digestion approach is also consistent with established in-plate digestion procedures used in glycoproteomic MRM workflows in serum and related matrices.[24] MRM LC-MS analyses were performed on an Agilent HPLC triple quadrupole (QQQ) mass spectrometry (MS) platform. Chromatographic separation was carried out using a reversed-phase C18 column under conditions consistent with previously published HDL and plasma glycoproteomic MRM workflows from our group.[12,24] A concise summary of the LC-MRM acquisition and run-design parameters used in the current study is provided in Table S6.

A panel of 12 commercially available synthetic peptide internal standards ("Sigma peptides") was spiked into all samples and pooled QC injections at a constant amount immediately prior to LC-MS analysis to monitor instrument response and support normalization. HDL digests were subsequently analyzed through LC-MS using a targeted MRM glycoproteomic workflow designed to quantify a panel of HDL-associated peptide and glycopeptide features. Glycopeptide targets were initially identified and validated from Orbitrap data searched with Byonic, and glycopeptide-spectrum matches were filtered using the Byonic score, retaining assignments with a score ≥ 50, at which tandem MS spectra provided sufficient fragment ion coverage to support peptide backbone and glycan

structure assignments.[24] For each monitored transition, chromatographic peaks were integrated only when signal exceeded local background noise, with baselines inspected and manually adjusted when necessary. In addition, a pooled HDL sample-preparation quality control (QC1) was generated by combining equal-volume aliquots of HDL from all 194 study participants and processed using the same HDL isolation and digestion procedures as the study samples. The 194 participant samples were analyzed in three 96-well plates over a 20 day period (20 September to 9 October 2023) in a series of overnight runs. A commercial human HDL preparation (Sigma) was injected at the start of each acquisition day to monitor long-term instrument stability. QC1 was injected at the beginning of each plate-row block (well position 1 of rows C through H on plates 1–3; for example P1C1, P1D1 … P3E1), approximately every ten study samples, yielding 20 QC1 injections distributed across the full acquisition sequence. The Sigma peptides and interleaved QC1 injections were used to characterize run-order drift and to monitor potential plate and row-position effects across the acquisition, supporting the LOESS-based normalization and feature-level internal-standard assignment described below.

Of the 164 transitions originally targeted, 59 features passed the QC mean-intensity threshold of 20 000 in pooled HDL and were detected in the participant dataset. This threshold was applied to exclude low-abundance features with unstable signal in the pooled HDL quality-control injections. Among these, 21 features had above baseline signal intensities across all 194 participant samples and were thus included in downstream analyses. Filtering therefore proceeded in two stages: first, a pooled-QC screen to remove low-abundance or unstable transitions, and second, a participant-level screen to retain analytes with quantifiable signal across the full cohort for comparative modeling. These features comprised 11 non-glycosylated peptides and 10 glycopeptides or glycoforms (see Table S3). Glycopeptide compositions are encoded by a numeric string appended to the peptide identifier in the format Protein_site_code. The four-digit composition codes used here (for example 1101 or 5401) denote the numbers of hexose (Hex), *N*-acetylhexosamine (HexNAc), fucose (Fuc), and *N*-acetylneuraminic acid or sialic acid (Neu5Ac), respectively. Thus, APOC3_94_1101 describes a glycopeptide of apolipoprotein C-III with a glycan consisting of 1 Hex, 1 HexNAc, 0 Fuc, and 1 Neu5Ac attached at asparagine position 94 in the peptide. For extended five-digit codes (for example 11 000), the first four positions follow the same convention, and the fifth digit denotes an additional monosaccharide as defined by the Byonic glycan library used in the discovery experiments (see Table S2). Zero intensities were retained in the raw and QC summaries and were interpreted as signals below the limit of detection for that analyte in that sample. For inferential models using $\log_2$ intensities, we added a constant offset of 1 to each corrected intensity ($\log_2$ (corrected_intensity + 1)) before transformation. All 21 analytes had measurable values across all participants after normalization.

Normalization was based on a LOESS-corrected internal-standard framework developed using the pooled QC data. For each Sigma peptide, a LOESS model was fitted to its intensity as a function of QC injection order across the full acquisition sequence, and LOESS-corrected Sigma peptide intensities were then normalized by the median LOESS-corrected intensity for that Sigma peptide and rescaled to the global raw median intensity of that Sigma peptide. LOESS smoothing was implemented using the base *R* stats::loess function with a smoothing span of 0.5, and the resulting fitted values were used as predicted intensity trends. A smoothing span of 0.5 was selected empirically after inspection of the pooled QC drift profiles because it captured the broad non-linear injection-order trend while avoiding overfitting to short-range fluctuations in the internal-standard signals. Observed Sigma intensities were divided by the corresponding LOESS trends to remove injection-order drift and then scaled by the ratio of the global median to the median LOESS-corrected intensity for that Sigma peptide:

$$\text{Rescaled Sigma Intensity} = \frac{\text{Sigma Intensity}}{\text{Loess Trend}} \times \frac{\text{Global Median}}{\text{Median Sigma}}$$

Median-based scaling was used because the median LOESS-corrected Sigma intensity provides a robust estimate of the typical internal-standard signal across QC injections, reducing the influence of occasional outlier injections while preserving the overall raw intensity scale *via* rescaling to the global median. These corrected Sigma peptides served as internal standards for feature-level normalization. Each analyte was normalized to each Sigma peptide in turn, and the sigma peptide that minimized QC-based coefficient of variation (CV) was selected as the optimal internal standard. This feature-specific internal-standard assignment was applied to all participant samples. The resulting LOESS-normalized, internal-standard-scaled analyte intensities were $\log_2$-transformed for statistical analysis. QC-derived feature summaries used to assess analytical reproducibility, evaluate normalization performance, and select feature-specific internal standards are provided in Dataset S1. Participant-level peptide and glycopeptide intensity data for the 21 analytes quantified across all 194 participants, provided prior to LOESS correction and transformation, are provided in Dataset S2.

4.4 Statistical analysis

All statistical analyses were performed in *R* (version 4.3.2) using $\log_2$-transformed, LOESS-normalized, and internal-standard-scaled intensities. Differential expression was evaluated at the individual feature level to preserve peptide- and glycoform-specific variation. To examine disease-related differences, analyses were conducted within *APOE3/E3* and *APOE3/E4* carriers to reduce confounding by genotype. For each of the 21 analytes quantified across the 194 participants, unadjusted pairwise Wilcoxon rank-sum tests were performed for control *versus* MCI, control *versus* AD, and MCI *versus* AD. These *p* values were used for visualization and descriptive interpretation and were not FDR-corrected because they served as exploratory follow-up comparisons rather than primary inferential tests.

For *APOE* genotype comparisons, unadjusted differences between *APOE3/E3* and *APOE3/E4* carriers were first assessed for

each analyte using Wilcoxon rank-sum tests. These unadjusted p values are shown in Fig. 2 and reported in Table S5 and they were used for Benjamini–Hochberg false-discovery rate (FDR) correction across the 21 analytes. Covariate-adjusted analyses were performed using feature-specific linear models with $\log_2$ intensity as the outcome and *APOE* genotype (*E3/E3 vs. E3/E4*) as the primary predictor, adjusting for age, ethnicity, and clinical status. The adjusted p values from these models are also reported in Table S5. Clinical-status differences were further evaluated within each genotype stratum using linear models with $\log_2$ intensity as a function of clinical status, age, and ethnicity. For each analyte, the p value for clinical status was obtained from an F-test and corrected for multiple testing within each genotype stratum.

Age and ethnicity were included as covariates because they represent upstream demographic factors known to influence peptide abundance and were available for all 194 participants. Other clinical characteristics (BMI, diabetes, history of diabetes, history of hypercholesterolemia, history of hypertension) were not included as covariates because they exhibited incomplete data and reflect metabolic states that may be influenced by cognitive diagnosis, making them unsuitable for adjustment in these peptide-level models.

Author contributions

Brian V. Hong (BVH): conceptualization; methodology; formal analysis; data curation; visualization; writing – original draft; writing – review & editing. Yiyun Liu (YL): methodology; investigation; data curation; visualization. Armin Oloumi (AO): methodology; investigation; data curation; writing – review & editing. Joanne K. Agus (JKA): methodology; investigation; data curation; writing – review & editing. Yasmine Bouchibti (YB): methodology; investigation; data curation; writing – review & editing. Lee-Way Jin (LWJ): investigation; resources; funding acquisition; writing – review & editing. Izumi Maezawa (IM): investigation; resources; writing – review & editing. Danielle J. Harvey (DJH): investigation; resources; writing – review & editing. Carlito B. Lebrilla (CBL): conceptualization; methodology; resources; supervision; writing – review & editing; funding acquisition. Angela M. Zivkovic (AMZ): conceptualization; methodology; resources; supervision; writing – review & editing; funding acquisition; project administration.

Conflicts of interest

There are no conflicts to declare.

Data availability

The data supporting this study are included in the Supplementary Information. Raw mass spectrometry data generated from human plasma samples are not publicly available due to ethical and confidentiality restrictions. Supplementary information is available. See DOI: https://doi.org/10.1039/d6ra01426g.

Acknowledgements

This work was supported by the National Institute on Aging of the National Institutes of Health (R01AG062240), the UC Davis Alzheimer's Disease Research Center (P30AG072972), and the USDA National Institute of Food and Agriculture Hatch Fund Project (CA-D-NUT-2242-H). We thank the participants and staff of the UC Davis Alzheimer's Disease Research Center for recruitment, clinical evaluation, and sample collection.

References

1 Z. Kyselova, Y. Mechref, P. Kang, J. A. Goetz, L. E. Dobrolecki, G. W. Sledge, *et al.*, Breast Cancer Diagnosis and Prognosis through Quantitative Measurements of Serum Glycan Profiles, *Clin. Chem.*, 2008, **54**(7), 1166–1175, DOI: 10.1373/clinchem.2007.087148.

2 S. Onigbinde, J. Solomon, C. D. Gutierrez-Reyes, O. Daramola, M. Fowowe, M. Adeniyi, *et al.*, Serum N-Glycan Profiling Identifies Candidate Glycan Biomarkers for Early Detection and Prediction of Alzheimer's Disease, *J. Proteome Res.*, 2025, **24**(9), 4417–4436, DOI: 10.1021/acs.jproteome.5c00018.

3 R. Z. Zhou, S. Gaunitz, B. E. Kirsebom, B. Lundin, M. Hellström, A. Jejcic, *et al.*, Blood N-glycomics reveals individuals at risk for cognitive decline and Alzheimer's disease, *EBioMedicine*, 2025, **113**, 105598, DOI: 10.1016/j.ebiom.2025.105598.

4 Ó. Lado-Baleato, J. Torre, R. O'Flaherty, M. Alonso-Sampedro, I. Carballo, C. Fernández-Merino, *et al.*, Age-Related Changes in Serum N-Glycome in Men and Women—Clusters Associated with Comorbidity, *Biomolecules*, 2023, **14**(1), 17, DOI: 10.3390/biom14010017.

5 C. Yang, H. Wang, C. Li, H. Niu, S. Luo and X. Guo, Association between clusterin concentration and dementia: a systematic review and meta-analysis, *Metab. Brain Dis.*, 2019, **34**(1), 129–140, DOI: 10.1007/s11011-018-0325-0.

6 A. Kontush, M. Lindahl, M. Lhomme, L. Calabresi, M. J. Chapman, and W. S. Davidson, Structure of HDL: Particle Subclasses and Molecular Components, In *High Density Lipoproteins [Internet]*, ed. Von Eckardstein A., Kardassis D., Springer International Publishing, Cham, 2015. pp. 3–51. (Handbook of Experimental Pharmacology). Available from: http://link.springer.com/10.1007/978-3-319-09665-0_1 doi:10.1007/978-3-319-09665-0_1.

7 E. B. Button, J. Robert, T. M. Caffrey, J. Fan, W. Zhao and C. L. Wellington, HDL from an Alzheimer's disease perspective, *Curr. Opin. Lipidol.*, 2019, **30**(3), 224–234, DOI: 10.1097/MOL.0000000000000604.

8 B. V. Hong, J. Zheng, J. K. Agus, X. Tang, C. B. Lebrilla, L. W. Jin, *et al.*, High-Density Lipoprotein Changes in Alzheimer's Disease Are APOE Genotype-Specific, *Biomedicines*, 2022, **10**(7), 1495, DOI: 10.3390/biomedicines10071495.

9 C. Y. Tseng, J. A. Salguero, J. D. Breidenbach, E. Solomon, C. K. Sanders, T. Harvey, *et al.*, Evaluation of normalization strategies for mass spectrometry-based multi-omics

datasets, *Metabolomics*, 2025, **21**(4), 98, DOI: 10.1007/s11306-025-02297-1.

10 J. Huang, H. Lee, A. M. Zivkovic, J. T. Smilowitz, N. Rivera, J. B. German, *et al.*, Glycomic analysis of high density lipoprotein shows a highly sialylated particle, *J. Proteome Res.*, 2014, **13**(2), 681–691, DOI: 10.1021/pr4012393.

11 M. J. Kailemia, W. Wei, K. Nguyen, E. Beals, L. Sawrey-Kubicek, C. Rhodes, *et al.*, Targeted Measurements of O- and N-Glycopeptides Show That Proteins in High Density Lipoprotein Particles Are Enriched with Specific Glycosylation Compared to Plasma, *J. Proteome Res.*, 2018, **17**, 834–845.

12 X. Tang, M. Wong, J. Tena, C. Zhu, C. Rhodes, Q. Zhou, *et al.*, Quantitative glycoproteomics of high-density lipoproteins, *RSC Adv.*, 2022, **12**(29), 18450–18456, DOI: 10.1039/D2RA02294J.

13 B. V. Hong, C. Zhu, M. Wong, R. Sacchi, C. H. Rhodes, J. W. Kang, *et al.*, Lipid-Based Nutrient Supplementation Increases High-Density Lipoprotein (HDL) Cholesterol Efflux Capacity and Is Associated with Changes in the HDL Glycoproteome in Children, *ACS Omega*, 2021, **6**(47), 32022–32031, DOI: 10.1021/acsomega.1c04811.

14 C. Zhu, M. Wong, Q. Li, L. Sawrey-Kubicek, E. Beals, C. H. Rhodes, *et al.*, Site-Specific Glycoprofiles of HDL-Associated ApoE are Correlated with HDL Functional Capacity and Unaffected by Short-Term Diet, *J. Proteome Res.*, 2019, **18**(11), 3977–3984, DOI: 10.1021/acs.jproteome.9b00450.

15 M. Koch, S. T. DeKosky, M. Goodman, J. Sun, J. D. Furtado, A. L. Fitzpatrick, *et al.*, Association of Apolipoprotein E in Lipoprotein Subspecies With Risk of Dementia, *JAMA Netw. Open*, 2020, **3**(7), e209250, DOI: 10.1001/jamanetworkopen.2020.9250.

16 Y. Zhang, S. M. Gordon, H. Xi, S. Choi, M. A. Paz, R. Sun, *et al.*, HDL subclass proteomic analysis and functional implication of protein dynamic change during HDL maturation, *Redox Biol.*, 2019, **24**, 101222, DOI: 10.1016/j.redox.2019.101222.

17 C. Huang, J. Zhang, J. Huang, H. Li, K. Wen, J. Bao, *et al.*, Proteomic and functional analysis of HDL subclasses in humans and rats: a proof-of-concept study, *Lipids Health Dis.*, 2023, **22**(1), 86, DOI: 10.1186/s12944-023-01829-9.

18 J. J. Zheng, J. K. Agus, B. V. Hong, X. Tang, C. H. Rhodes, H. E. Houts, *et al.*, Isolation of HDL by sequential flotation ultracentrifugation followed by size exclusion chromatography reveals size-based enrichment of HDL-associated proteins, *Sci. Rep.*, 2021, **11**(1), 16086, DOI: 10.1038/s41598-021-95451-3.

19 T. Vaisar, I. Babenko, K. V. Horvath, K. Niisuke and B. F. Asztalos, Relationships between HDL subpopulation proteome and HDL function in overweight/obese people with and without coronary heart disease, *Atherosclerosis*, 2024, 397, DOI: 10.1016/j.atherosclerosis.2024.118565.

20 F. M. Sacks, J. D. Furtado and M. K. Jensen, Protein-based HDL subspecies: Rationale and association with cardiovascular disease, diabetes, stroke, and dementia, *Biochim. Biophys. Acta Mol. Cell Biol. Lipids*, 2022, **1867**(9), 159182, DOI: 10.1016/j.bbalip.2022.159182.

21 K. L. Rasmussen, A. Tybjærg-Hansen, B. G. Nordestgaard and R. Frikke-Schmidt, Plasma levels of apolipoprotein E, APOE genotype, and all-cause and cause-specific mortality in 105 949 individuals from a white general population cohort, *Eur. Heart J.*, 2019, **40**(33), 2813–2824, DOI: 10.1093/eurheartj/ehz402.

22 H. M. Nielsen, K. Chen, W. Lee, Y. Chen, R. J. Bauer, E. Reiman, *et al.*, Peripheral apoE isoform levels in cognitively normal APOE ε3/ε4 individuals are associated with regional gray matter volume and cerebral glucose metabolism, *Alzheimer's Res. Ther.*, 2017, **9**(1), 5, DOI: 10.1186/s13195-016-0231-9.

23 A. T. Baker-Nigh, K. G. Mawuenyega, J. G. Bollinger, V. Ovod, T. Kasten, E. E. Franklin, *et al.*, Human Central Nervous System (CNS) ApoE Isoforms Are Increased by Age, Differentially Altered by Amyloidosis, and Relative Amounts Reversed in the CNS Compared with Plasma, *J. Biol. Chem.*, 2016, **291**(53), 27204–27218, DOI: 10.1074/jbc.M116.721779.

24 A. Oloumi, S. T. Le, Y. Liu, S. Herbert, A. Ji-Xu, A. A. Merleev, *et al.*, Pemphigus-Associated Desmoglein-Specific IgG1 and IgG4 Have a Dominant Agalactosylated Glycan Modification, *J. Invest. Dermatol.*, 2024, **144**(11), 2584–2587.e6, DOI: 10.1016/j.jid.2024.03.044.