FNIH Biomarkers Consortium

Data Primer

Biomarkers Consortium CSF Proteomics Project

1. Background

The data and analysis plan described within this document represents the work of the FNIH Biomarkers Consortium Project "Use of Targeted Mass Spectrometry Proteomic Strategies to Identify CSF-Based Biomarkers in Alzheimer's Disease". This project was submitted to the Biomarkers Consortium Neuroscience Steering Committee by a subgroup of the Industry Private Partner Scientific Board (PPSB) of the Alzheimer's Disease Neuroimaging Initiative (ADNI) for execution and was managed by a Biomarkers Consortium Project Team that includes members from academia, government and the pharmaceutical industry. Funding for this project was provided by Genentech (a member of the Roche Group), Janssen, Lundbeck, Merck, and Takeda. This project is the second part of a multi-phased effort seeking to utilize samples collected by ADNI to qualify multiple peptides in cerebrospinal fluid (CSF) to diagnose patients with Alzheimer's disease (AD) and monitor disease progression. These efforts were reviewed by the ADNI Resource Allocation Review Committee (RARC) and approved by the National Institute on Aging (NIA). An earlier phase of the program focused on using a multiplexed immuno-based assay (performed by Rules Based Medicine) to characterize potential AD biomarkers in CSF.

The aim of this project was to determine the ability of a panel of peptides measured by mass spectrometry to discriminate among disease states and show changes in a longitudinal manner. Additional biomarkers with diagnostic and prognostic value are needed for AD drug development, especially in the context of clinical trials aiming to treat patients before the onset of dementia. In addition, biomarkers that can be used to monitor treatment effects in both early and established AD clinical trials could enable more efficient trial designs and facilitate understanding of therapeutic mechanism of action. A full description of the LC/MS-MRM approach to biomarker quantitation is beyond the scope of this Primer but an overview with key references is provided in Section 2.

The CSF multiplex MRM panel was developed by Caprion Biosciences in collaboration with the Biomarkers Consortium Project Team. Proteins and peptides were selected based upon their previous detection in CSF, relevance to AD, and previous results from the Rules Based Medicine (RBM) multiplex immunoassay analysis of ADNI CSF. From previous work, 5 primary targets were selected for absolute quantitation (FABPH, SCG2, VGF, NPXT2 and CHGA) and 121 targets were monitored for relative quantitation. As described in more detail below, CSF samples were digested, and analyzed by LC-MS/MS operating in MRM mode at Caprion Biosciences Inc. The final MRM panel consisted of 278 peptides representing 126 proteins and, for each peptide, two mass transitions were monitored. For the 5 absolute quantitation targets a concentration was reported using a single quantifier transition back calculated on a curve prepared from recombinant protein. These results are reported in arbitrary signal intensity units on a natural log scale. Because alternate splicing or post-translational processing could result in

biologically significant differences in the levels of two peptides from the same protein, the use of the "log peptide intensity" is recommended for further analysis.

Stable isotope-labelled (SIL) peptides were synthesized for all peptides monitored in this study. These peptides were used for method development and as normalizers for the study. Sample analysis batches include standard curve and quality control samples.

Seven hundred and fifty (750) unique frozen CSF samples were received at Caprion. These were composed of 730 longitudinal samples from the ADNI-1, ADNI-2 and ADNI-GO studies, and 20 blinded replicate aliquots, one each for 20 subjects in the study. The 20 blinded replicates were distributed throughout the MS analysis runs and used to assess assay reproducibility.

Caprion initially defined and performed the QC and processing steps described below. All work described in this primer was performed before the sample ID's were un-blinded to reveal the corresponding ADNI subject IDs. The Project Team also pre-defined a Statistical Analysis Plan for initial analysis of these biomarkers for diagnostic and prognostic utility (**Appendix I**). The Project Team will conduct this analysis once they are un-blinded, which occurs simultaneously with the posting of the un-blinded data to the LONI website. The Project Team intends to report results of the Statistical Analysis Plan through conferences and/or publications.

2. Description of Technology

MRM allows the specific and sensitive quantification of peptides and proteins in biological samples. It is the most sensitive mass spectrometry-based platform (Lange et al., 2008; Mol. Syst. Biol. 4, 222) and was demonstrated to be highly reproducible within and across laboratories and instrument platforms (Addona et al., Nat Biotechnol. (2009) (27) 633; Kennedy et al., Nat Methods (2014) (11) 149; Geyer et. Al, (2017), Mol Sys Biol. (13) 942). MRM experiments are performed on triple quadrupole (QQQ) mass spectrometers. Peptide ions are isolated in the first quadrupole (Q1), ions are fragmented in the second Q by collision induced dissociation (CID), characteristic fragments ions isolated in the third Q, and then monitored and quantified. Up to 750 peptides, covering a dynamic range of 5 orders of magnitude can be quantified in a single 30 min run. The samples in this study were analyzed on two QTRAP 6500 mass spectrometers (AB Sciex).

3. Sample Randomization and Blinding

The original study randomization was performed by Dr. Shaw's group at UPenn and was based on a sample processing workflow including CSF depletion. The 750 CSF samples were split into groups of 36 samples with all longitudinal samples from an individual contained within the same group (called run01, run02 etc. by UPenn). Eleven (11) runs were planned to be depleted using one IgY14/Supermix depletion column and ten (10) runs were to be depleted using a second depletion column.

Through subsequent assay development work, the depletion steps were deemed unnecessary and eliminated and the new sample processing workflow was designed with four runs on two Q-TRAP 6500

instruments. The groups and order of injection from the initial randomization (which assumed the use of depletion columns) were maintained and divided between the four runs as described in **Table 1**.

Table 1. Linking of sample run with randomization blocks. The (depletion column x) designation is included only to reflect that they were used in the design of the randomization although they were not used in the final processing workflow.

Caprion Run	Batch size	QTRAP Instrument	UPenn randomization designation
Run 01	186	U	run01-run05 (depletion column 1)
Run 02	186	Z	run06-run10 (depletion column 1)
Run 03	180	U	run11 (depletion column 1) and run01-run04 (depletion column 2)
Run 04	222	Z	run05-run10 (depletion column 2)

All project team members remain blinded to the participant IDs in ADNI. Thus, experimental data cannot be linked to demographic, clinical or other biomarker data for the participants until the experimental data is uploaded to the ADNI website.

4. Sample Processing and MRM Analysis Overview

The flow chart presented in **Figure 1** describes the main steps of the entire process for sample analysis. Each of these steps is described herein.

CSF Sample Processing Method

Seven hundred and fifty (750) unique samples were included in this study. Sample aliquots were shipped to Caprion on dry ice and stored at -80˚C until use. Due to the presence of endogenous proteins in CSF and the difficulty of sourcing CSF with low levels of the 5 absolute quantitation targets, standard curve samples were prepared using recombinant proteins (CMGA, SCG2, VGF, NPTX2, FABPH) in a BSA fortified buffer (0.2 mg/mL BSA). Quality control (QC) samples were prepared in a CSF pool comprised of 300 individual donors, representative of the study samples, supplied by ADNI and prepared in Caprion's facilities. Aliquots of the standard and QC samples were frozen at -80°C until use.

The 750 study samples were processed in 4 batches, with longitudinal samples from each individual subject processed in the same batch. After thawing, 50µL of each sample was denatured with trifluoroethanol (Sigma) followed by proteolytic digestion with trypsin (Promega) at an approximate 1:25 protease to protein ratio. The samples were then acidified with trifluoroacetic acid and SIL peptides spiked in. The peptides were desalted using Oasis MCX desalting plates (Waters) and aliquoted into two replicate mass spectrometry (MS) plates which were dried by vacuum evaporation and stored at -20°C prior to MS analysis.

Two mass spectrometers were used in the analysis of the samples. Runs 01 and 03 were analyzed on one instrument (QTRAP U) while runs 02 and 04 were analyzed on a separate instrument (QTRAP Z). Prior to analysis, the two mass spectrometers were cross-validated by testing the backup plates from 3 precision and accuracy runs.

Each processing batch consisted of:

1. Standard curves (8 non-zero levels, plus blank) in duplicate;
2. Quality control (QC) samples in CSF pool (3 non-zero levels) in triplicate;
3. CSF pool spiked with SIL peptides in replicates of 6 (to determine endogenous levels);
4. Reference LLOQ samples in triplicate;
5. Double blanks and carryover blanks,
6. Samples for use during retention time correction and system suitability test (SST) of the LC-MS
7. 180-222 study samples

   Note: Standard and QC samples were prepared using recombinant proteins for the 5 absolute quantitation targets only. SIL peptides used as internal standards and for relative quantitation were added during processing as described in **Figure 1**.

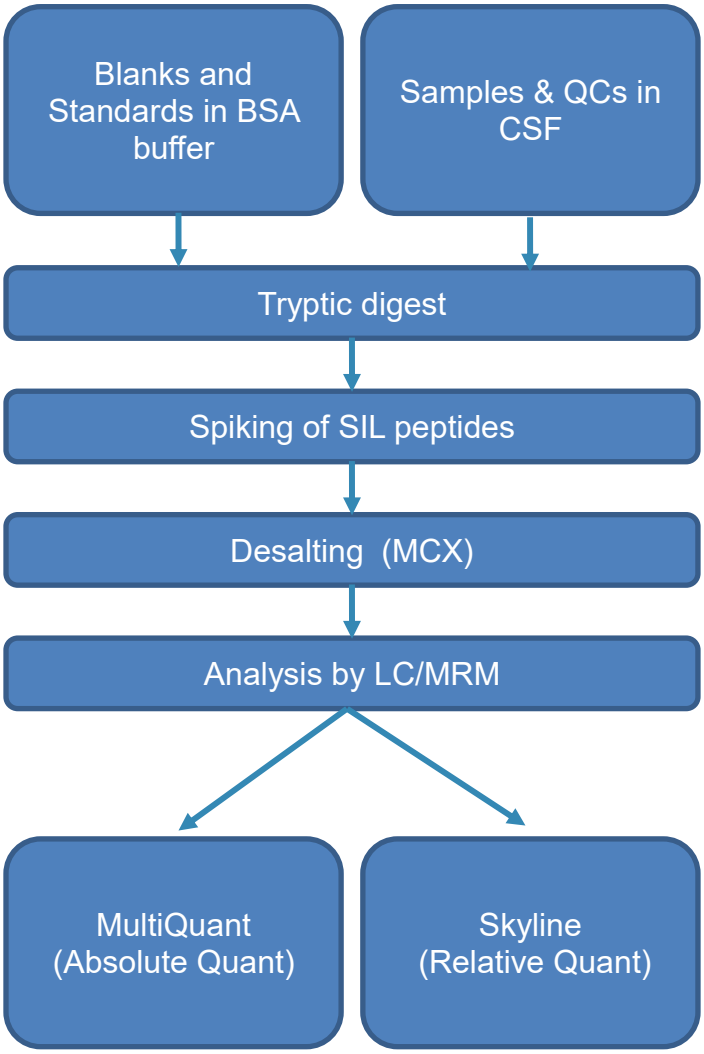

Figure 1: Processing workflow

MRM Analysis of CSF Samples

The multiplexed MRM panel used in this project consisted of 268 peptides targeting 121 relative quantitation protein targets as well as 10 peptides targeting 5 proteins (FABPH, SCG2, VGF, NPTX2 and CHGA) which were selected for absolute quantitation. These targets were selected based on previous work with the Biomarkers Consortium. For each of the 278 peptides, a synthetic SIL peptide was ordered and used as internal standard during the analysis.

Absolute quantitation was performed using the surrogate peptide approach. A single peptide/transition was used for quantitation and was selected based on the following criteria: sensitivity, chromatographic performance, lack of interference, linearity, precision, and accuracy.

Commercial CSF samples were sourced and processed in-parallel with standard curve and QC samples to assess the range of endogenous concentrations of the target proteins. Peptide selection criteria are outlined in **Table 2**. Standard curve and QC levels were adjusted to reflect the measured endogenous levels. Peptides with sufficient sensitivity at the lower limit of quantitation (LLOQ) level were assessed for linearity and accuracy through the curve range. Interference, defined as a non-specific signal at the retention time of the peptide, was assessed in BSA buffer. To assess interference in CSF, the transition area ratio and peak shape of the endogenous peptide was compared to that of the SIL peptide. Precision was assessed for CSF QCs, as the use of a surrogate matrix for the curve precludes an assessment of true accuracy in CSF. Finally, carryover was assessed in BSA by injecting a blank sample immediately following the upper limit of quantitation (ULOQ) sample. Two peptides per protein meeting the criteria were kept, and the peptide showing the best overall performance used as quantifier.

Table 2: Peptide selection criteria.

	Criteria	
Parameter	**Endogenous transitions**	**SIL transitions***
Linearity	R > 0.98	N/Ap
Accuracy of standard levels	<25% bias, <30% at LLOQ	N/Ap
Interference in blank BSA sample	<20% of LLOQ peak area	<5% of SIL peak area
Interference in CSF pool sample	Match SIL peak area ratio, peak shape	N/Ap
Signal/Noise ratio	>5:1 at LLOQ	N/Ap
Reproducibility	CV of peak area ratio in ULOQ samples <20%	
Carryover	<20% of LLOQ peak area	<5% of SIL Peak area

* SIL criteria differ from endogenous as the signal intensity is uniformly high

Selecting a single, well characterized, peptide/transition as quantifier avoids discrepancies in reported concentrations between peptides and ensures reproducibility between analytical runs. A second

transition was monitored for QC purposes. A second peptide was monitored for trouble shooting purposes only.

Relative quantitation was performed using the peak area ratio of the analyte/SIL peptides. Data treatment and quality control for relative quantitation are described in detail below.

Prior to analysis, retention time correction was performed on a naive CSF pool sample spiked with SIL peptides to ensure analyte peaks fell within the retention time window used during analysis. Subsequently a system suitability test (SST) was performed to assess sensitivity, repeatability, interference and carryover of the LC-MS system. The SST was performed as follows:

1. One blank injection, to assess interference
2. One lower limit of quantitation (LLOQ) sample, to assess sensitivity
3. Three upper limit of quantitation (ULOQ) samples, to assess repeatability
4. Two carryover blanks, to assess carryover

These samples were prepared in BSA buffer as the signal of the endogenous proteins in CSF precluded the use of CSF. System suitability criteria are described in **Table 2a** and are assessed for the 5 absolute quantitation transitions used as quantifiers, **Table 2b**.

Table 2a: SST criteria.

	Criteria	
Parameter	**Endogenous transitions**	**SIL transitions***
Interference in Blank sample	<20% of LLOQ peak area	<5% of SIL peak area
Signal/Noise ratio	>5:1 at LLOQ	N/Ap
Reproducibility	CV of peak area ratio in ULOQ samples <20%	
Carryover	<20% of LLOQ peak area	<5% of SIL Peak area

* SIL criteria differ from endogenous as the signal intensity is uniformly high

Table 2b: Quantifier transitions.

Protein	Peptide	Transition 1 (T1)	Transition 2 (T2)
CMGA	EDSLEAGLPLQVR	713.88	612.4
FABP3	SLGVGFATR	454.26	707.4
VGF	VLEYLNQEK	568.30	923.5
NPTX2	TNYLYGK	429.72	643.4
SCG2	THLGEALAPLSK	618.85	444.3

Sample analysis was initiated after a successful system suitability test run. The samples were injected by processing batch. The processed samples were re-solubilized with 10 µL of a reconstitution solution containing 5 internal standard peptides (ISP) each at 100 ng/mL. These 5 ISP elute at different retention times during the chromatographic gradient and are used to monitor instrument performance during sample analysis. Seven (7) µL of material was injected, per sample, onto a NanoAcquity UPLC (Waters) coupled to a 6500 QTRAP mass spectrometer (AB Sciex). Peptide separation was achieved using a 500 µm x 10 mm, 2.7 µm particle size Halo Peptides ES C18 column (Canada Life Science). The LC gradient used is shown in **Table 3**. The flow rate was 18 µL/min.

Table 3: Liquid Chromatography (LC) gradient.

Time	%A *	%B **
Initial	92.5	7.5
0.2	92.5	7.5
25	72	28
25.6	40	60
26.6	40	60
26.61	92.5	7.5

* Eluent A: Water:dimethylsulfoxide (DMSO) 97:3v/v, +0.2% Formic Acid

** Eluent B: Acetonitrile:DMSO 97:3 v/v, + 0.2% Formic Acid

Assay performance was monitored per injection batch using pre-defined acceptance criteria which were outlined in the study plan.

MRM Measurement Quality for Relative Quantitation Targets

During peak integration of the MRM raw data, each transition was evaluated to assess measurement quality (ex. signal-to-noise ratio, presence of interference, etc.). Based on these quality filters, transition light-to-heavy peak area ratio values were flagged in the corresponding filtered datasets as follows:

- Values replaced by 0: reported if both light transitions were not detected above the limit of detection (LOD). The LOD is defined as 3 S/N (3 times signal-to-noise ratio) and was based on peak height.
- Values replaced by NA: reported for a peptide if a signal above the LOD was detected but at least one of the following conditions was met:
    - The signal of at least one of the 2 transitions of the heavy peptide was below 3 S/N
    - The transitions light-to-heavy ratios did not co-elute
    - The light-to-heavy ratio of transition #1 and the light-to-heavy ratio of transition #2 differed by more than 60%

5. Statistical Analysis

Definitions

PQC samples: Process Quality Control samples are replicates of a pooled sample, similar to the study samples, and used to monitor reproducibility of sample processing and analysis.
LH: ratio of the light peptide (endogenous) transition area over its heavy (SIL) counterpart, for each transition. Samples for which the heavy peptide transitions are not detected will not be considered.
Missing Peak: LH replaced by a missing value (NA). Such values will be treated as "missing at random" during the statistical analysis.
Reliable Peak: Measurements that passed all quality filters during peak integration (i.e. LH not replaced by 0 or NA) are considered as reliable.
Detection Rate: Calculated for each transition, it is the proportion of reliable samples. The detection rate calculation is calculated among all study samples and stratified by status.
Boxplot: A graphical display of sample distribution characterized by a box and whiskers. Boxes are delimited at the top by the third quartile (Q3) and at the bottom by the first quartile (Q1). The thick black line within a box represents the median (Q2). Whiskers extend to the most extreme data point which is no more than 1.5 times the interquartile range (Q3-Q1) from the box.
Barplot: A graphical display characterized by a bar for each level of interest. An example of level of interest is sample ID. The height of the bar represents the characteristic of interest to display.
Scatterplot: A graphical display to visualize the relationship between two numerical variables. A scatterplot consists of a series of points on a two-dimensional plot where the two axis positions are determined by the values of the variables of interest.

Data Transformation for Outliers and Pattern Detection

For LH values that were replaced by 0, this replacement was considered as a flag and not as a value to use for analysis. The original peak integration value was used for analysis.

Unless mentioned otherwise, base-2 log of LH, referred to as log-LH, was used for the statistical analyses. To avoid issues with taking the logarithm of 0, 1 was added to light areas before division by heavy areas. Values replaced by NA as described above were propagated to these re-calculated ratios.

Outliers and Pattern Detection

The impact of various sample processing steps was evaluated. The sample processing variables (SPVs) included:

- Digestion (plate ID and well position)
- Mass spectrometer (instrument, plate ID and well position)
- Analytical run
- Order of injection, to monitor sensitivity loss

Relative Quantitation Data Set

Signal from SIL peptides was investigated for each transition to ensure sufficient detection. Transitions with a large proportion of samples with low signal were excluded from the analysis. The following 34 transitions (ordered alphabetically) were excluded from further analyses because they were detected in no samples, study or PQC:

A4_HUMAN.WYFDVTEGK.y7+
A4_HUMAN.WYFDVTEGK.y8+
AFAM_HUMAN.FLVNLVK.b2+
ALDOA_HUMAN.ALQASALK.b3+
ALDOA_HUMAN.ALQASALK.y6+
CD44_HUMAN.EQWFGNR.y3+
CD44_HUMAN.EQWFGNR.y4+
CNTP2_HUMAN.HELQHPIIAR.y3+
CNTP2_HUMAN.HELQHPIIAR.y5+
CO2_HUMAN.DFHINLFR.b3+
CO2_HUMAN.DFHINLFR.y5+
FAM3C_HUMAN.SPFEQHIK.y3+
FAM3C_HUMAN.SPFEQHIK.y6+
FMOD_HUMAN.IPPVNTNLENLYLQGNR.y5+
FMOD_HUMAN.IPPVNTNLENLYLQGNR.y6+
KLK10_HUMAN.ALQLPYR.b3+
KLK10_HUMAN.ALQLPYR.y4+
LIGO1_HUMAN.ATVPFPFDIK.y5+
LIGO1_HUMAN.ATVPFPFDIK.y7+
PON1_HUMAN.SFNPNSPGK.y6+
PON1_HUMAN.SFNPNSPGK.y7+
PRDX1_HUMAN.ADEGISFR.y5+
PRDX3_HUMAN.HLSVNDLPVGR.b2+
PRDX3_HUMAN.HLSVNDLPVGR.y4+
PRDX6_HUMAN.LIALSIDSVEDHLAWSK.y4+
PRDX6_HUMAN.LIALSIDSVEDHLAWSK.y8+
SHSA5_HUMAN.KFVWSEER.y5+
SHSA5_HUMAN.KFVWSEER.y6+
SORC3_HUMAN.AVASQWPEELASAR.y4+
SORC3_HUMAN.AVASQWPEELASAR.y8+
TTHY_HUMAN.TSESGELHGLTTEEEFVEGIYK.y5+
TTHY_HUMAN.TSESGELHGLTTEEEFVEGIYK.y7+
VTDB_HUMAN.EFSHLGK.y3+
VTDB_HUMAN.EFSHLGK.y5+

An investigation to identify poorly detected samples was conducted. **Figure 2** shows the per-sample transition detection rate of each sample separately for PQC samples, study samples, and study sample replicates in the relative quantitation data set. In study samples, almost 95% of the transitions targeted were detected.

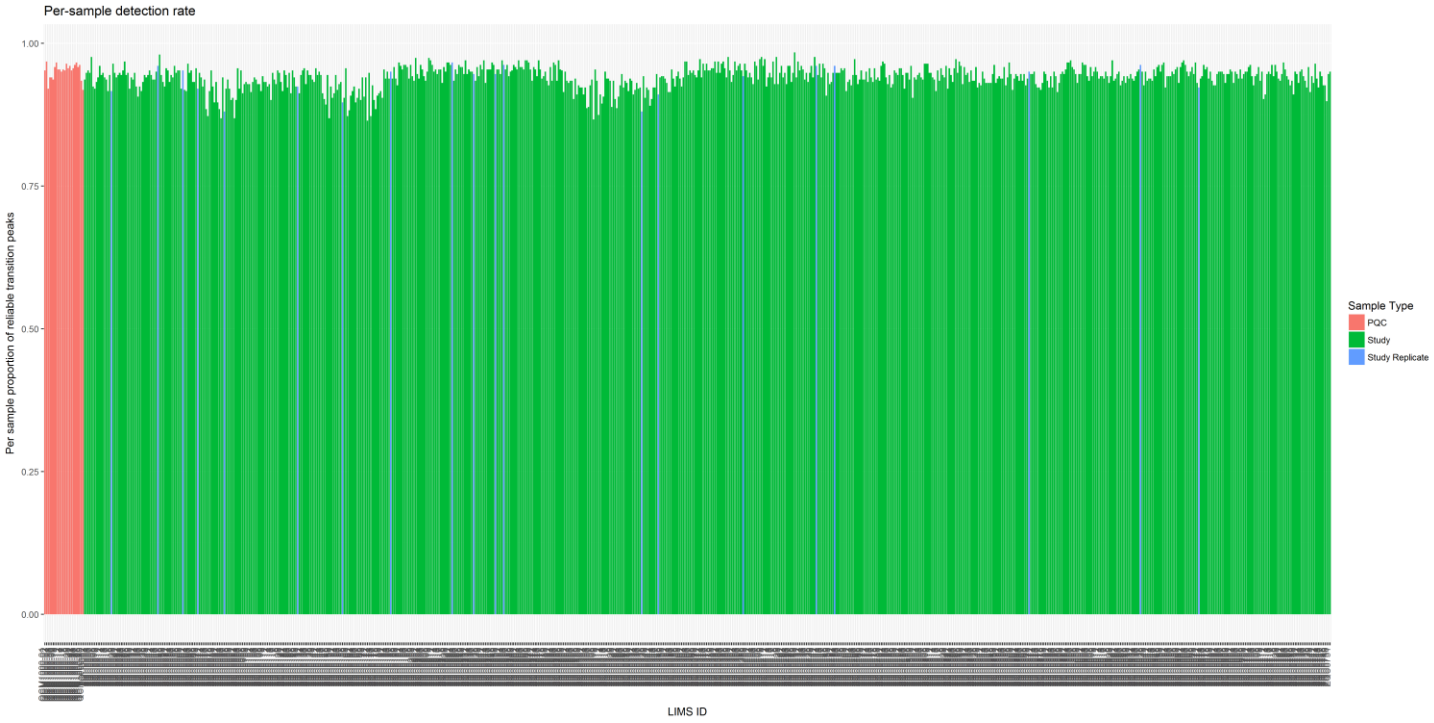

Figure 2: Per-sample relative-quantitated transition detection rate.

Because technical replicates of study samples were included in this dataset, it was possible to assess the degree of non-biological technical variability by comparing intensities of replicates to their study sample pair mate. For each transition, the mean %CV between replicates and their study sample pair mate was calculated. The mean across all transitions of this average %CV was 16.3%. 95% of transitions had a mean %CV between 10.4% and 28.1%.

Per sample averages of transitions were examined with regard to the SPVs, using boxplots and scatterplots. In addition, per sample average of reliable transition peaks was also examined. Both were performed for light areas, heavy areas and LH ratios, and independently for PQC samples, study samples and study sample replicates. This investigation allowed the identification of samples that did not behave like the others, on average (low detection included), and may pinpoint problematic sample processing steps.
While heavy areas were uncharacteristically large for samples run in the first half of analytical run 2, no samples were identified as log-LH outliers by these analyses. Nonetheless, some sample processing variables, such as digestion plate ID and analytical run, appeared to have an impact on average log-LH values by visual inspection. **Figure 3** illustrates average transition intensities as a function of digestion plate ID in the relative quantitation data set. This figure demonstrates that while digestion plates had an effect on average light and heavy peptide peak areas, the average light-to-heavy area ratio remained

relatively stable throughout the experiment. Moreover, study sample replicates were not affected any differently than subject samples run on those digestion plates.

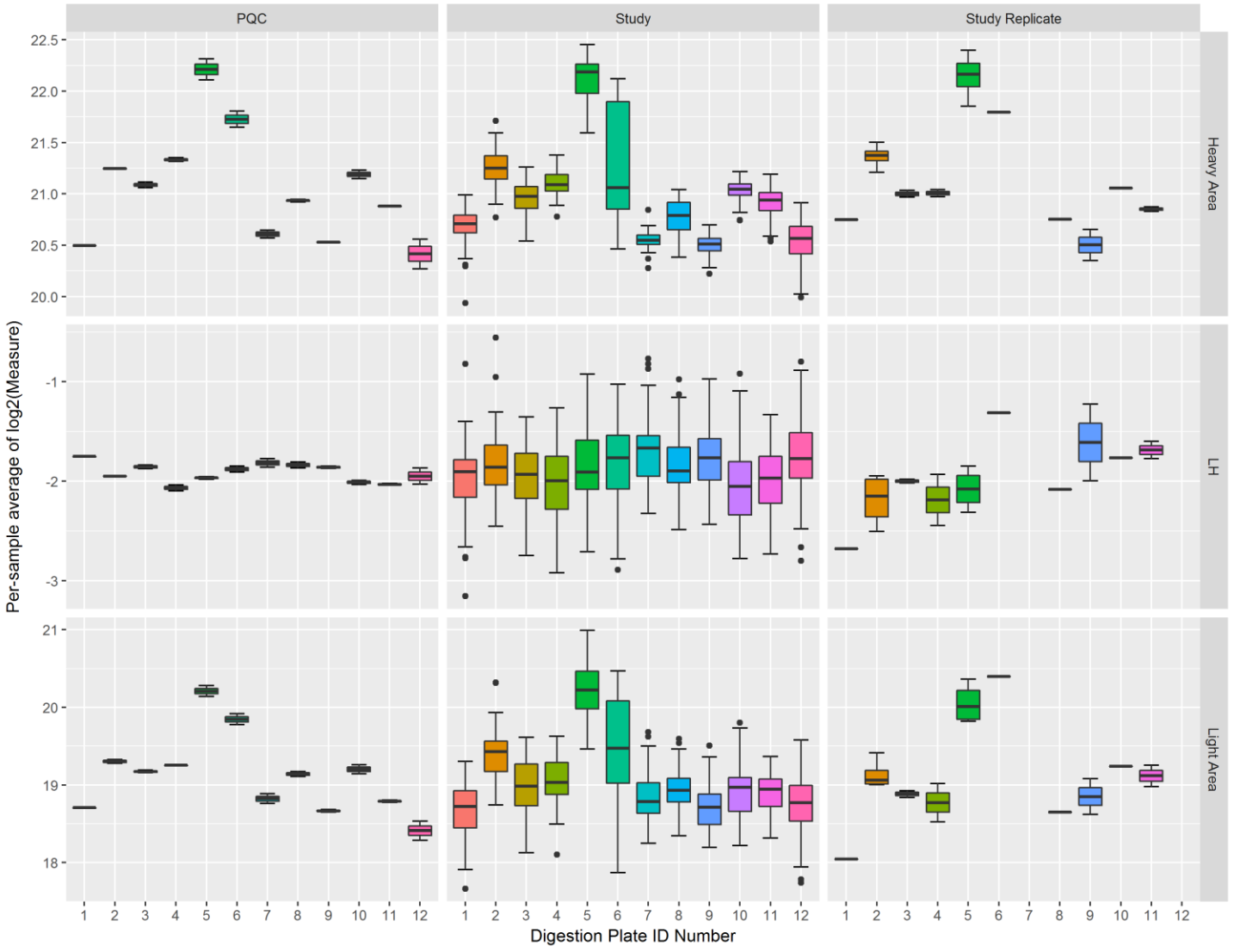

Figure 3: Per-sample average relative quantitation as a function of digestion plate ID.

Per-sample standard deviations of transition log-LH were also examined, where standard deviation was calculated as though the transitions were independent. The investigation was done independently for PQC samples, study samples and study sample replicates. This investigation allowed the identification of samples that behaved differently due to a subset of transitions only. It may also be a sign of suspicious transitions. If, for example, intensity values in a subset of transitions were suppressed in a digestion batch, the samples in that batch would have higher per-sample standard deviations of transition log-LH.

As above, there appeared to be within-run order effects, but no samples were identified as log-LH outliers. Nonetheless, sample processing variables related to instrument, analytical run and digestion plate ID appeared to have an impact on per-sample log-LH variability, by visual inspection. **Figure 4** illustrates per-sample standard deviations as a function of digestion plate ID in the relative quantitation data set. This figure demonstrates that digestion plates had an effect on the peptide-to-peptide variability of light and heavy peak areas within samples, but the peptide-to-peptide variability in light-to-heavy area ratios remained relatively stable throughout the experiment. The first 3 digestion plates had slightly lower per-sample variability, indicating that the light-to-heavy ratios of peptides measured in each sample processed in those batches were more similar to each other than in samples processed in subsequent batches, but this represented an approximately 7% change in peptide-to-peptide intensity variation. Moreover, study sample replicates were not affected any differently than subject samples run on those digestion plates: the variability between peptides in a study sample replicate was approximately equivalent to the subject samples run on the same digestion plate.

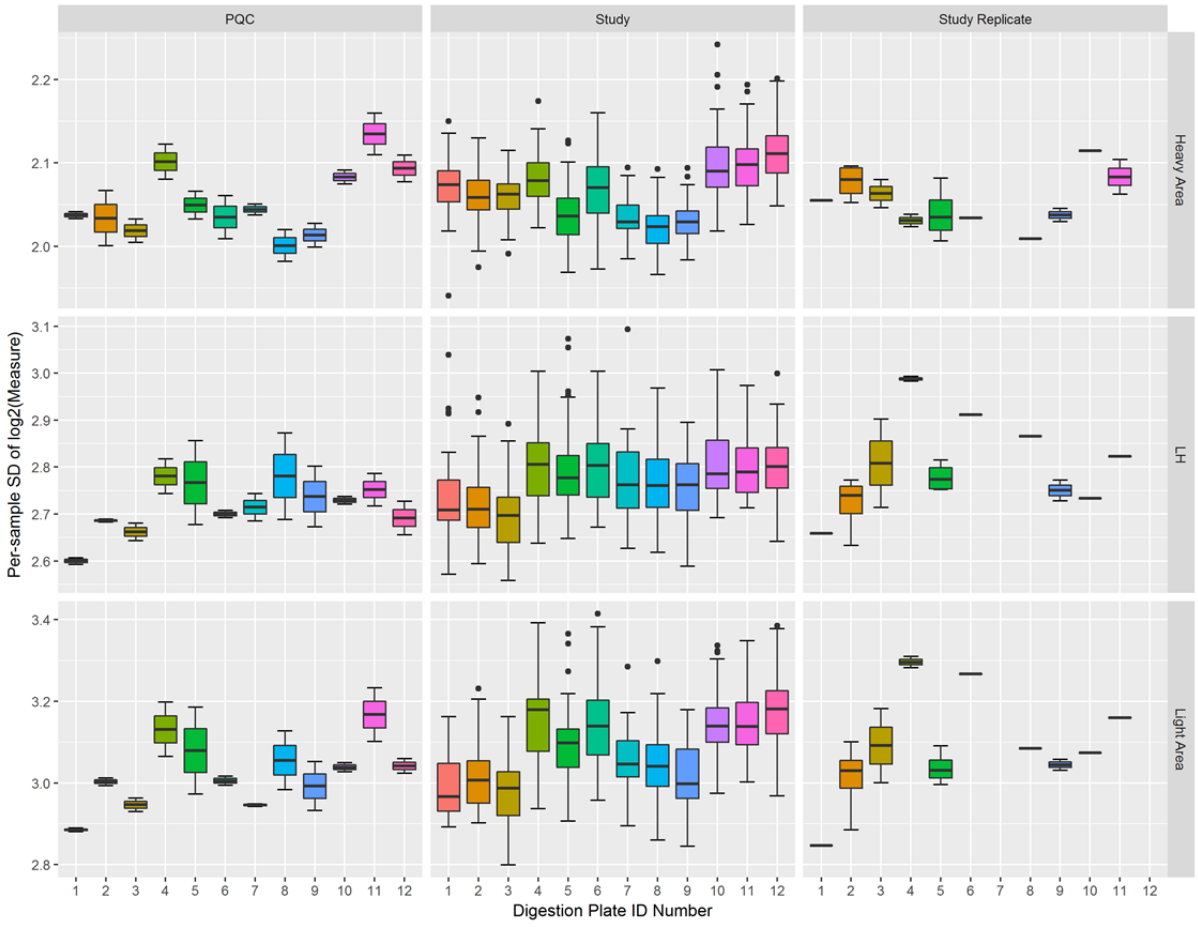

Figure 4: Per-sample standard deviation in relative quantitation as a function of digestion plate ID.

To determine if there was a sample bias, such as samples of systematically higher or lower expression, sample boxplots were created, where each box represented the per-sample distribution of transition log LH. No samples stood out as distributional outliers.

A principal component analysis (PCA) was applied to the doubly-centered log-LH. Double centering is the operation after which both sample and transition log-LH averages are 0.

PCA's cannot accept missing values. To overcome this shortcoming, a Probabilistic PCA (PPCA) method was used. A PPCA combines an Expectation-Maximization (EM) approach to PCA with a probabilistic model which is based on the assumption that latent variables and noise are normally distributed (Stacklies, et al., 2007). Such models can incorporate missing values. The modeling of PPCA assumes that missing values are missing at random.

The number of principal components to retain was decided as 2 via a scree plot – which is a scatter plot of the explained variance (i.e. eigenvalue) by principal component. The number of components corresponded to the point of inflection in eigenvalues for the relative-quantitation (log-LH) PPCA.

The components were visually inspected to assess the presence of any patterns in the PPCA. Because no outlier samples were flagged when evaluating per-sample detection rates, log-LH averages, log-LH standard deviations, or sample distributions, a new iteration excluding those samples was not performed to re-assess previously identified patterns in the PCA. A small number of outlier analytes were identified, which could all be traced to blood proteins HBA (hemoglobin subunit alpha), HBB (hemoglobin subunit beta), PRDX2 (peroxiredoxin), and SAMP (serum amyloid p-component). Removing these analytes and samples flagged as pink did not change the results shown below.

A variance-component analysis (VCA) was performed to formally assess which SPVs had an impact on the data variance and principal components. This was done by fitting a linear model predicting retained principal component values with SPVs as predictors and performing model selection while optimizing the Aikaike Information Criterion. Digestion plate ID and sample injection order had significant effects on variance components according to this analysis.

A variance partition analysis (VPA) (Hoffman & Schadt, 2016) was performed to formally assess which of the SPVs had an impact on data variance. For this analysis, both principal components and transition log-LH were used as inputs. Digestion plate ID and run-order effects were nontrivial contributors to variance in both principal components and transition log-LH values. **Figure 5** illustrates the percentage of variance in each component explained by each SPV in the relative quantitation data set.

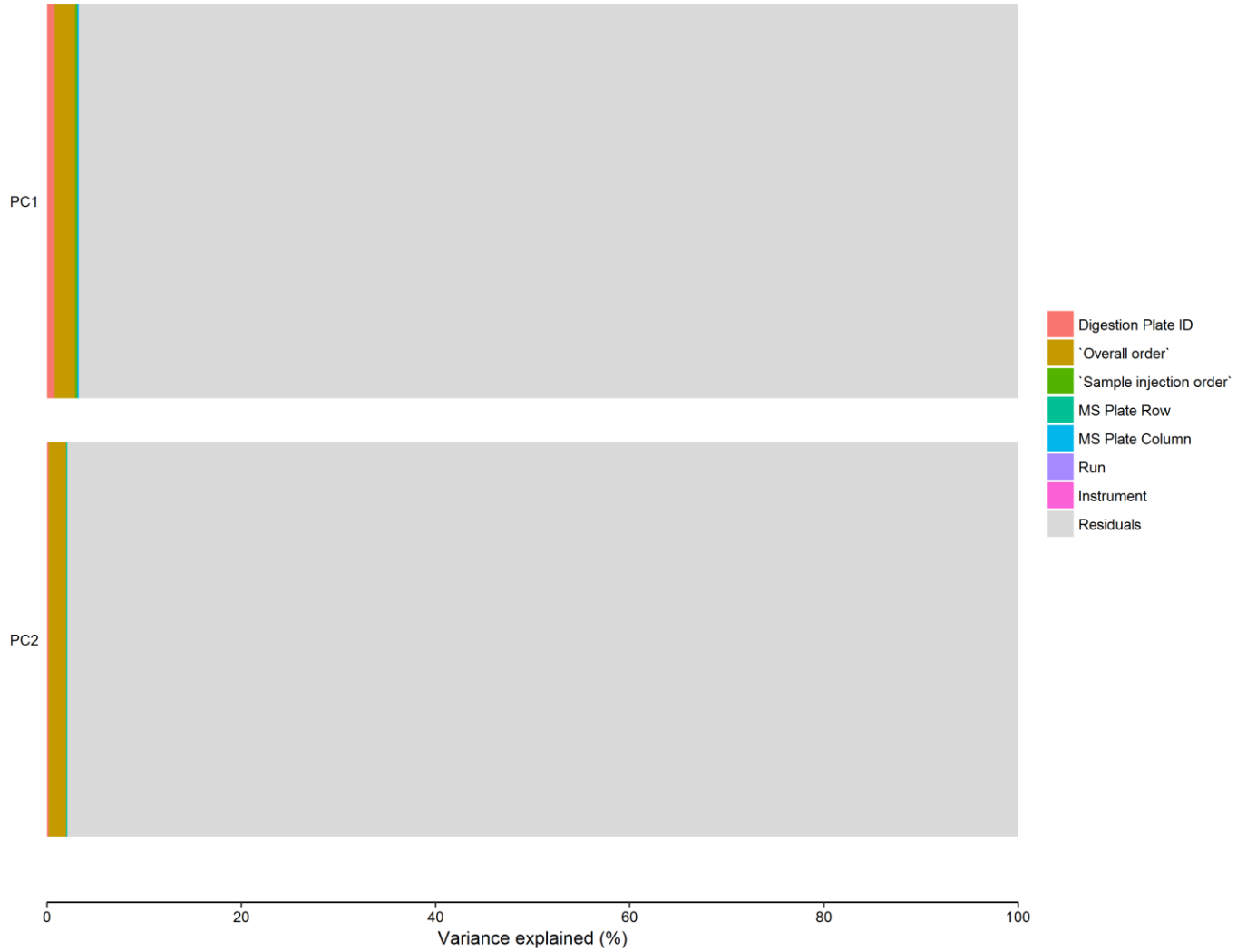

Figure 5: Variance in relative-quantitated components explained by each SPV.

Absolute Quantitation Data Set

As with the relative quantitation data, an investigation was conducted to determine whether any samples were poorly detected. **Figure 6** shows the per-sample detection rate for absolute-quantitation peptides separately for study samples and study sample replicates. In study samples, almost 100% of the peptides targeted were detected.

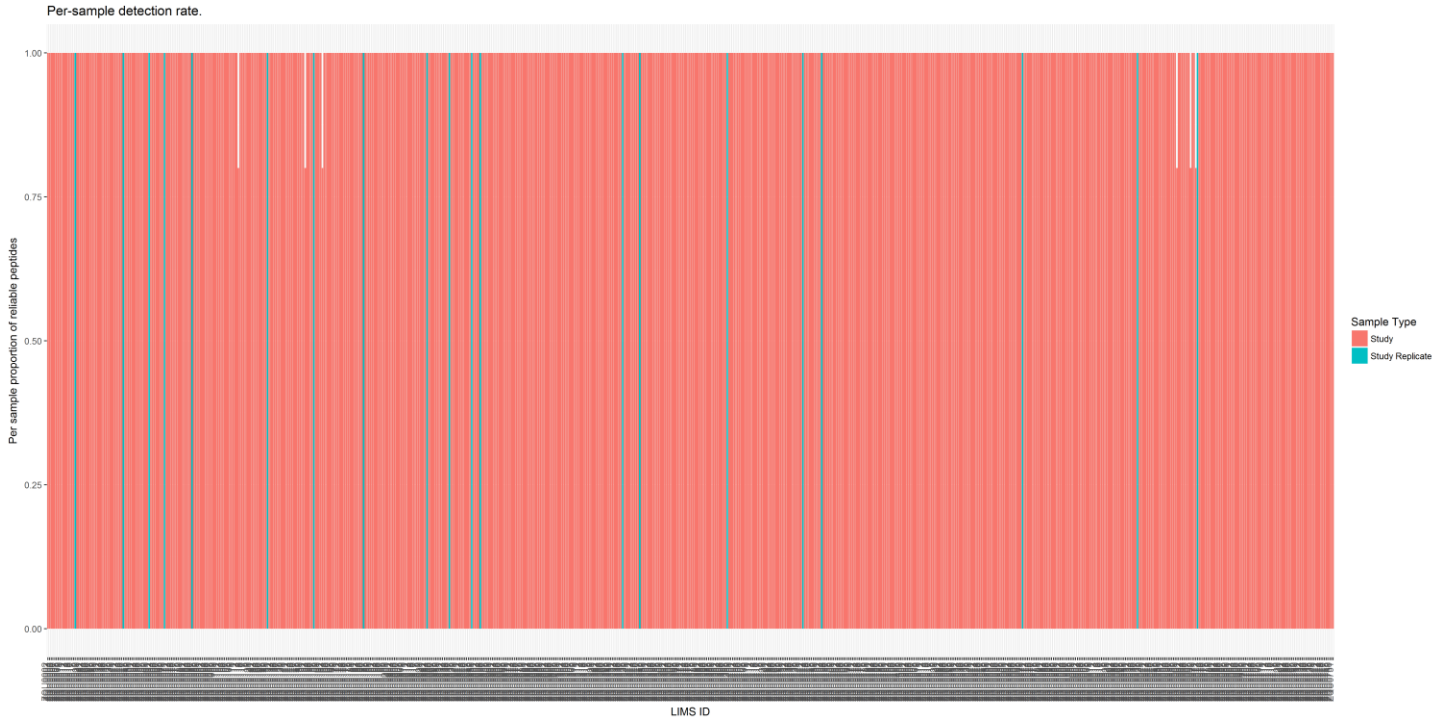

Figure 6: Per-sample absolute-quantitation detection rate.

Because technical replicates of study samples were included in this dataset, it was possible to assess the degree of non-biological technical variability by comparing intensities of replicates to their study sample pair mate. As a measure of this type of variability, the table below provides the mean coefficient of variation (CV) between a study replicate and its study sample pair mate, expressed as a percentage.

Table 1: Mean coefficient of variation between study replicate and study sample pair mate

Protein	analyte	Mean %CV
CMGA_HUMAN	EDSLEAGLPLQVR_713.88_612.4	15.04823
FABPH_HUMAN	SLGVGFATR_454.26_707.4	14.76556
NPTX2_HUMAN	TNYLYGK_429.72_643.4	17.95823
SCG2_HUMAN	VLEYLNQEK_568.30_923.5	21.83761
VGF_HUMAN	THLGEALAPLSK_618.85_444.3	14.89699

Per-sample average intensities were also evaluated for absolute-quantitation data. While no samples were identified as outliers, some sample processing variables like run order, instrument, analytical run, and digestion plate ID appeared to have an impact on average intensity levels by visual inspection. **Figure 7** illustrates average absolute intensities as a function of digestion plate ID.

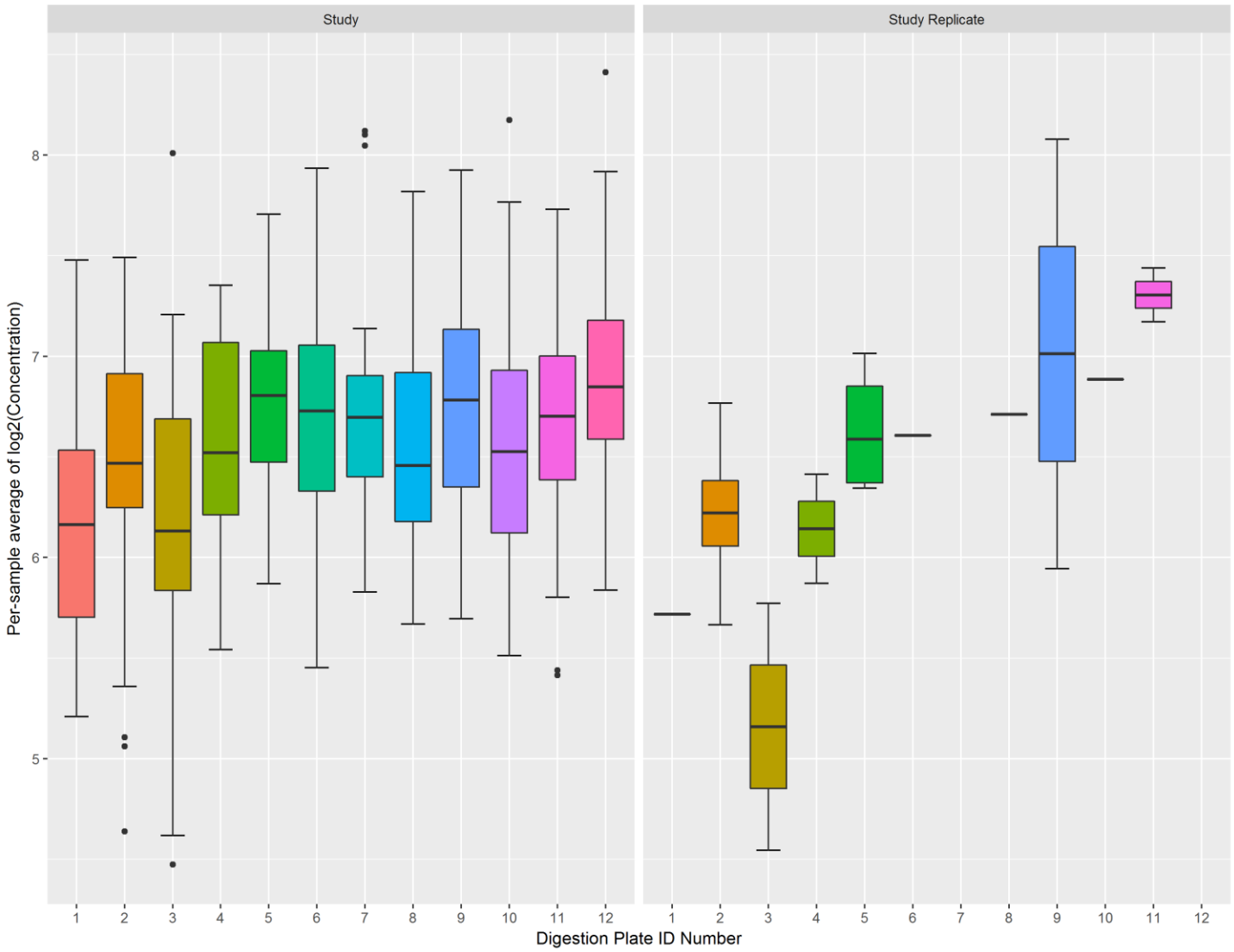

Figure 7: Per-sample average absolute intensity as a function of digestion plate ID.

Additionally, per-sample standard deviations in absolute intensities were examined as a function of SPVs. This provides an indication of whether SPVs affected a subset of peptide intensities. If, for example, the intensity of a single peptide was suppressed by digestion batch, the per-sample standard deviation would be increased for the samples processed in that batch.

As above, no sample outliers were identified with respect to absolute intensity variability, but some sample processing variables like run order, instrument, analytical run, and digestion plate ID appeared to have an impact by visual inspection. **Figure 8** is an example illustrating the effect of digestion plate ID.

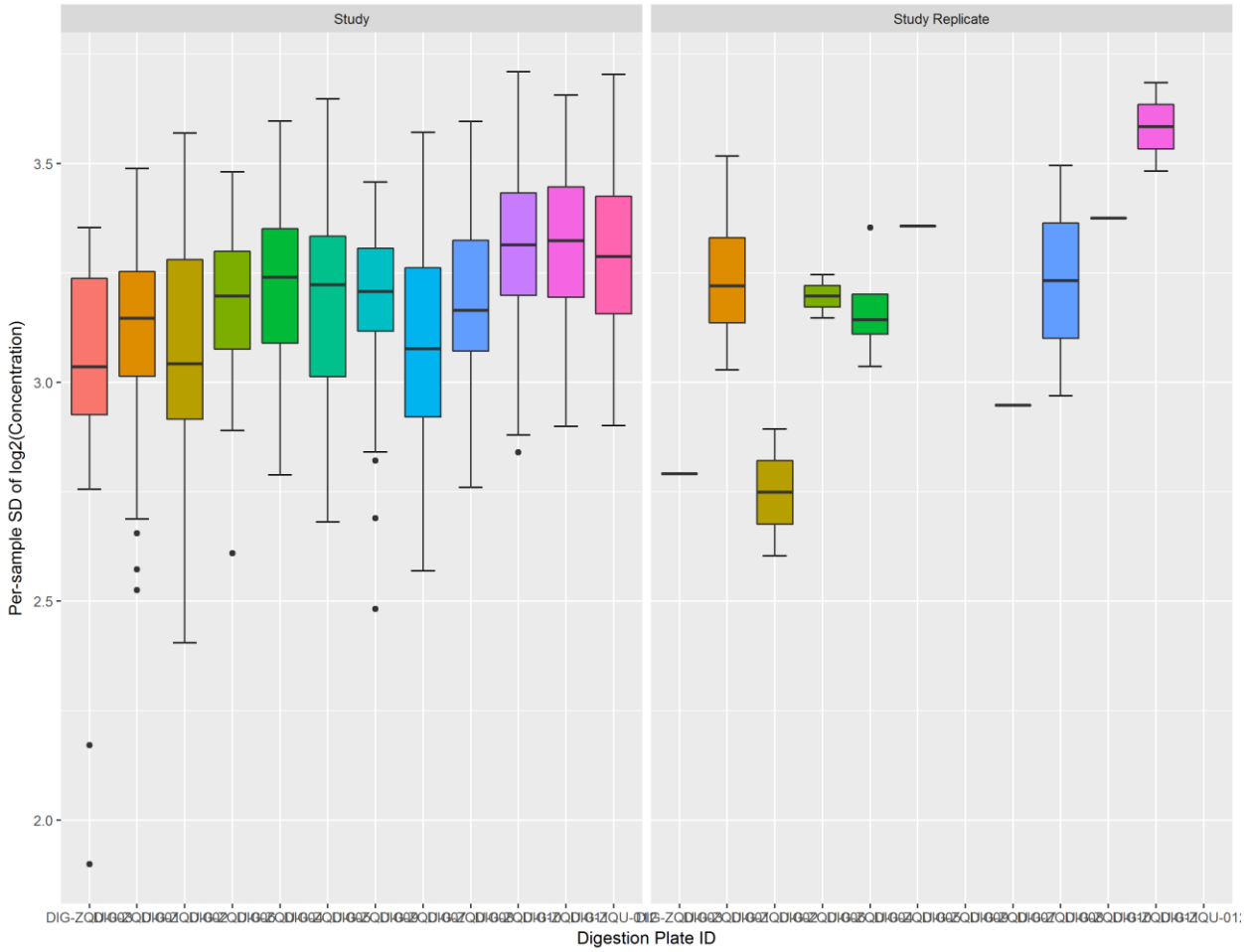

Figure 8: Per-sample standard deviation of absolute intensity as a function of digestion plate ID.

A principal component analysis (PCA) was applied to the doubly-centered log-LH. Double centering is the operation after which both sample and peptide absolute log-intensity averages are 0. To avoid issues with missing values and for consistency with the relative quantitation data set, a probabilistic PCA was used. The number of principal components to retain was decided as 2 via a scree plot. The components were then visually inspected to assess the presence of any patterns in the PPCA.

Because no outlier samples were identified as outliers when evaluating per-sample detection rates, absolute log-intensity averages, absolute log-intensity variability, or sample absolute intensity distributions, a new iteration excluding those samples was not performed to re-assess previously identified patterns in the PCA.

When absolute-quantitation data were analyzed in a VCA, digestion plate ID and sample injection order were the only SPVs with statistically significant effects on variance components.

Similarly, a VPA of absolute-quantitated peptides identified large effects of run order, digestion plate, and instrument ID on both principal components and individual peptide levels. Because digestion plates were only processed on a single instrument, any normalization for digestion plate ID would also correct for any inter-instrument differences as well. **Figure 9** shows the percentage of variance in each principal component of absolute-quantitated intensities explained by each SPV.

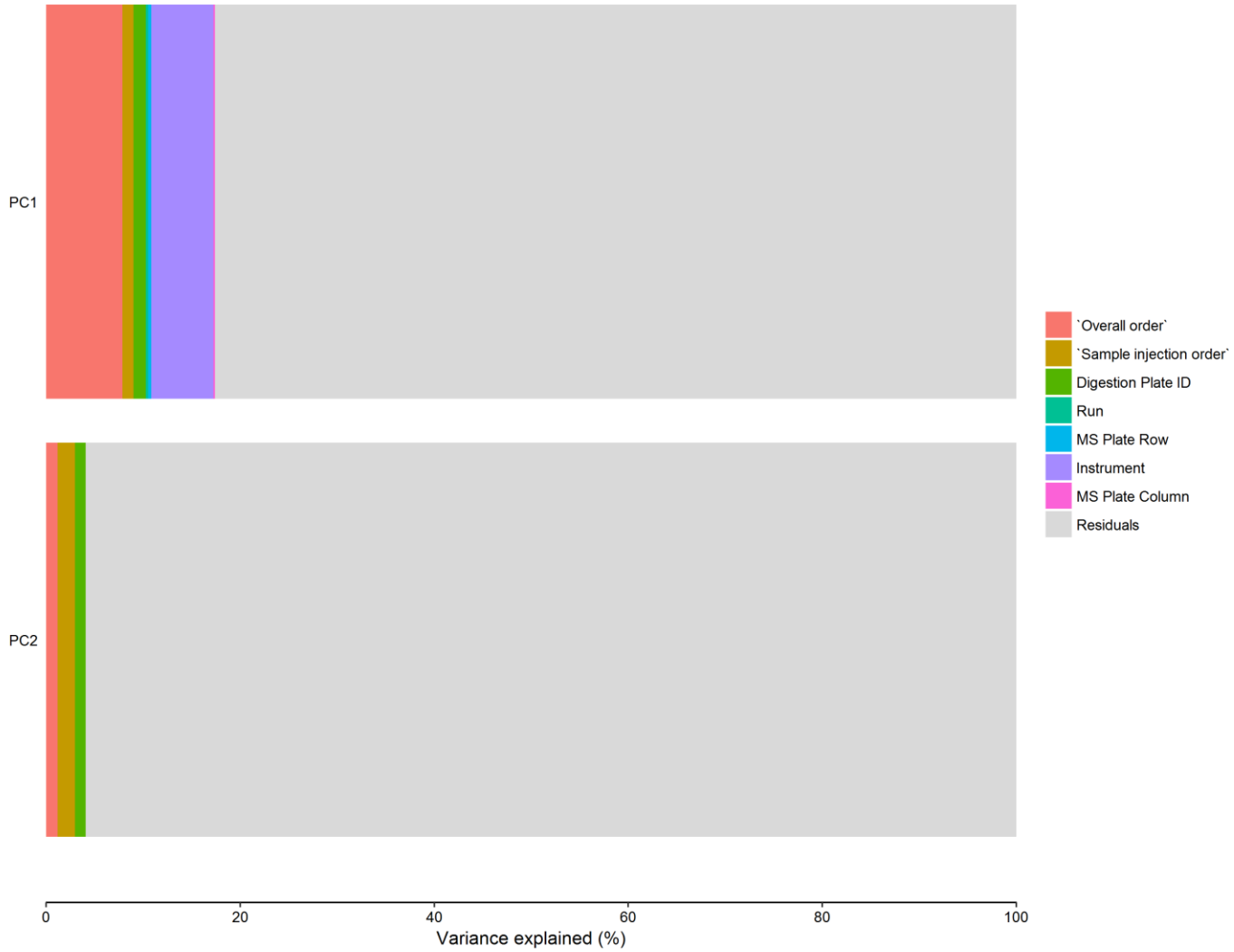

Figure 9: Variance of absolute-quantitated components explained by each SPV.

Data Transformation for normalization and expression analysis

For LH values that were replaced by 0, this replacement was considered as a flag and not as a value to use for analysis. The original peak integration value was used for analysis.

Normalization

The Instrument, MS Plate, Run, and Injection Order were determined to have residual variance in the data. The MS plate is confounded with Run and Instrument. Injection order is a function of each run. Normalization that takes both plates and run order into account should also account for systematic differences between Instruments and Runs.

The following normalization was applied to both the absolute and relative quantitative datasets separately. The absolute quantitative data represents 5 protein transitions. The relative quant data represents 502 transitions. The PQC samples were replicated 24 times across the plates. Additionally, there are 20 Subject samples with replicates randomized into different blocks.

A) All calculations were performed in log2 units. "Raw" refers to pre-normalized data. "Norm" refers to the post-normalized data.
B) Using the Raw data, the mean value was calculated across all samples for each transition. Samples with hemolysis were excluded from the mean. These values were used in (D) to normalize plates
C) For each run, a regression of transition values against run order was performed. To account for plate differences, each plate had a separate offset in the regression. After regression, the residual represented the variance not attributed to run order.
    a. Samples with hemolysis were not included in the regression but were corrected
    b. To maximize robustness, only transitions within 2 standard deviations of the mean were included in the regression
    c. Run 2 was split into two separate runs before and after instrument reset.
D) The plate average was reset to be the same as the original global average (B). Samples with hemolysis were corrected but not included in calculating the plate mean.

Table 4 represents the quantitative results of normalization. Simca-P 15 (Umetrics) was used to generate Partial-Least-Squared (PLS) Models utilizing the full matrix of transitions to predict the block effect variables. The values in the table represent 7-fold cross validation R^2 values. Red Italic values were not statistically significant. As previously demonstrated, the raw relative quantitative data had a stronger correlation with the block effects than the absolute quantitative data. In all cases, the normalization removed the residual block effects in the data.

Table 4: Model Cross-Validation R^2 Outcomes of Normalization

PCA/PLS: Outcome	Raw Quant	Norm Quant	Raw Relative	Norm Relative
Instrument (class)	0.313	*-0.0106*	0.936	*-0.325*
Injection-Number (continuous)	*0.00911*	*-0.00757*	0.706	*-0.456*
Run (class)	0.226	*-0.0111*	0.729	*-0.407*

PCA (**Figures 10a-d)** was performed on the raw and normalized data for both the absolute and relative quantitation data sets. In general, the characteristics of the PCA plots were maintained pre and post

normalization. The replicated pairs were in close proximity and the PQC samples clustered around the center. Samples that were outliers remained outliers. In the relative quantitation pre-normalized plot, the PQC samples were tightly clustered with limited dispersion on PC1. Post-normalization, the dispersion of the PQC samples looked much more like the absolute quantitation projections. There was considerably more variance due to the block effects in the relative quantitation than the absolute quantitation data and its removal brought more qualitative parity between the relative and absolute quantitation datasets.

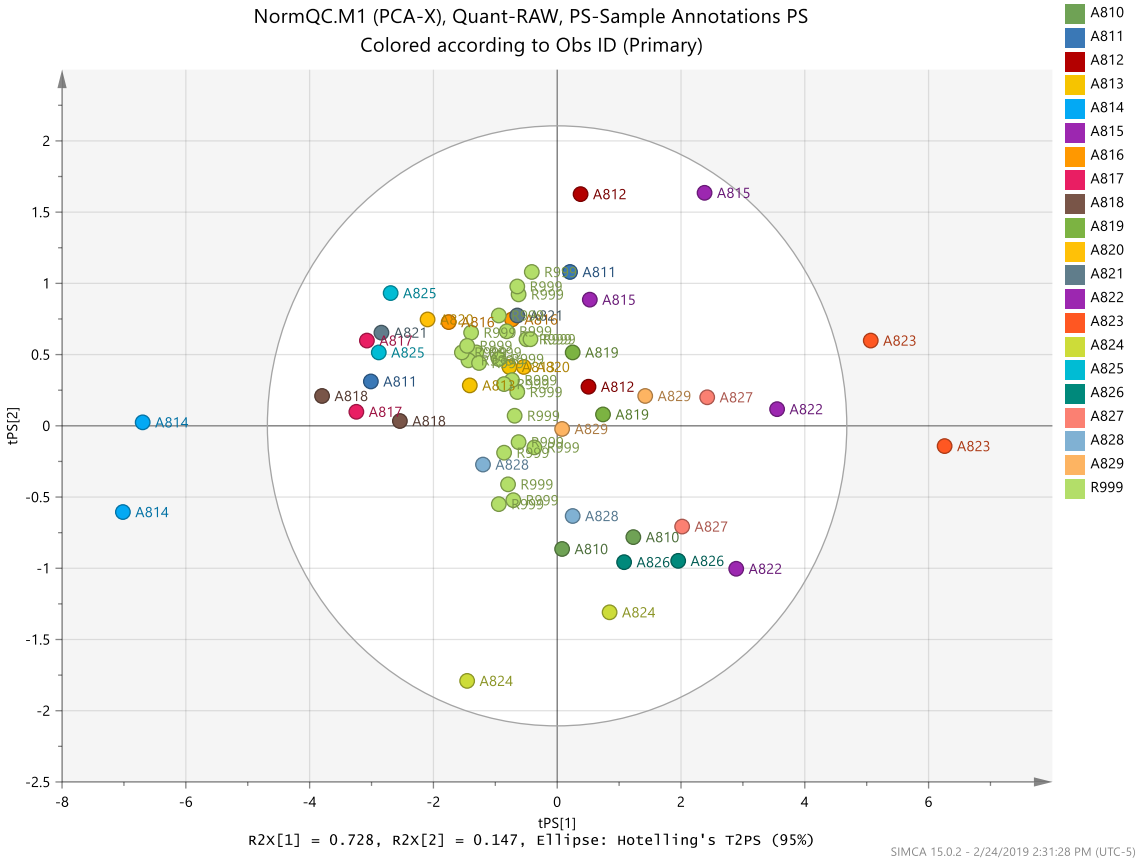

Figure 10a: PCA Absolute Quantitation Pre-Normalization. R999 represents the 24 PQC samples. A810-A829 represents the 20 pairs of replicated subject samples.

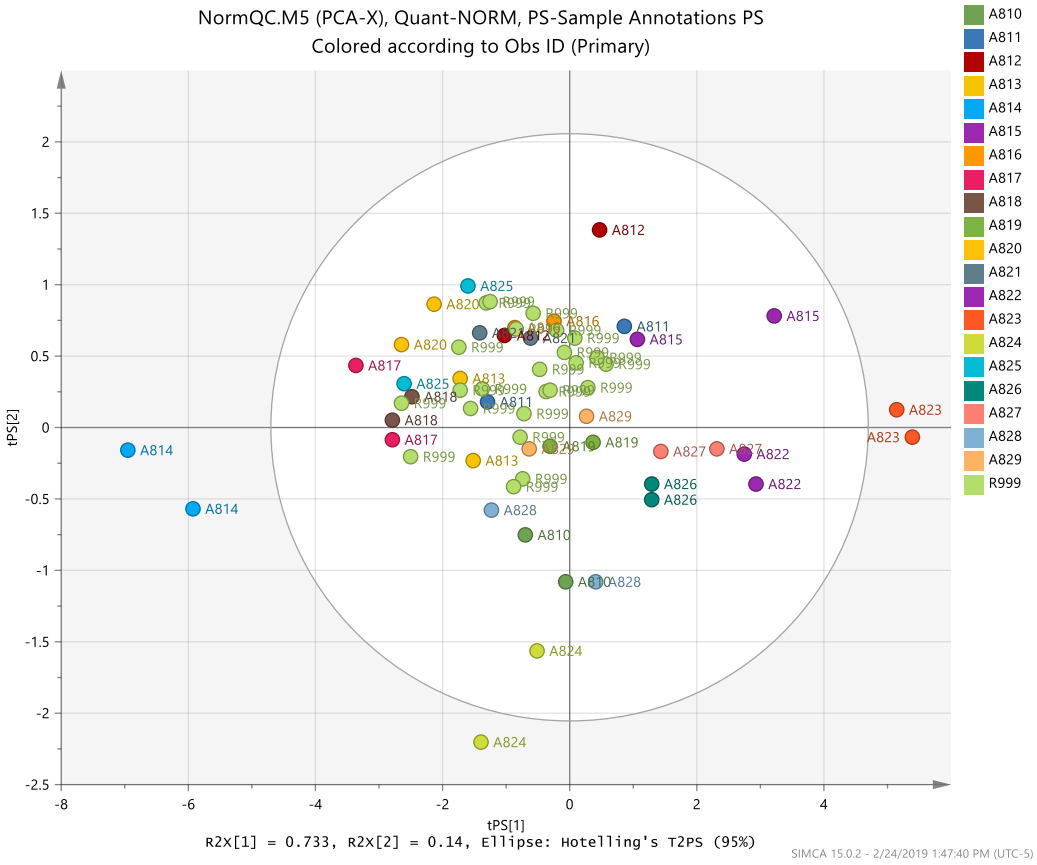

Figure 10b: PCA Absolute Quantitation Post-Normalization.

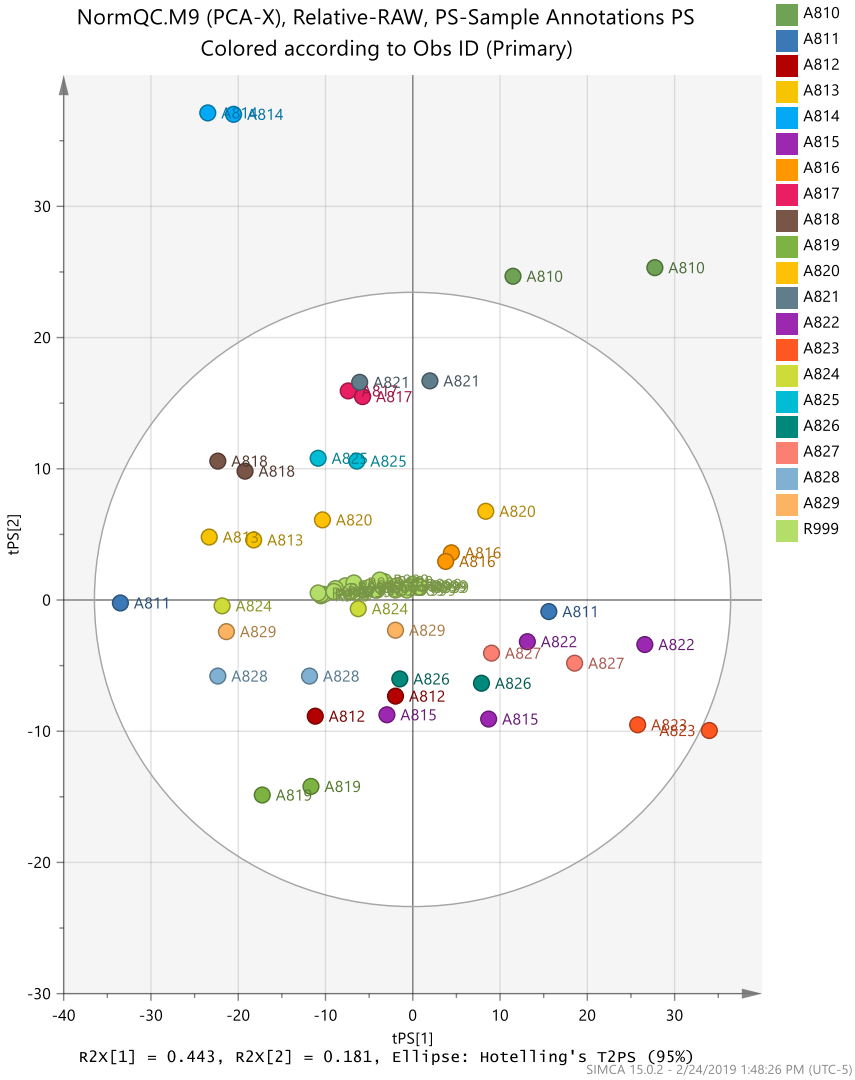

Figure 10c: PCA Relative Quantitation Pre-Normalization.

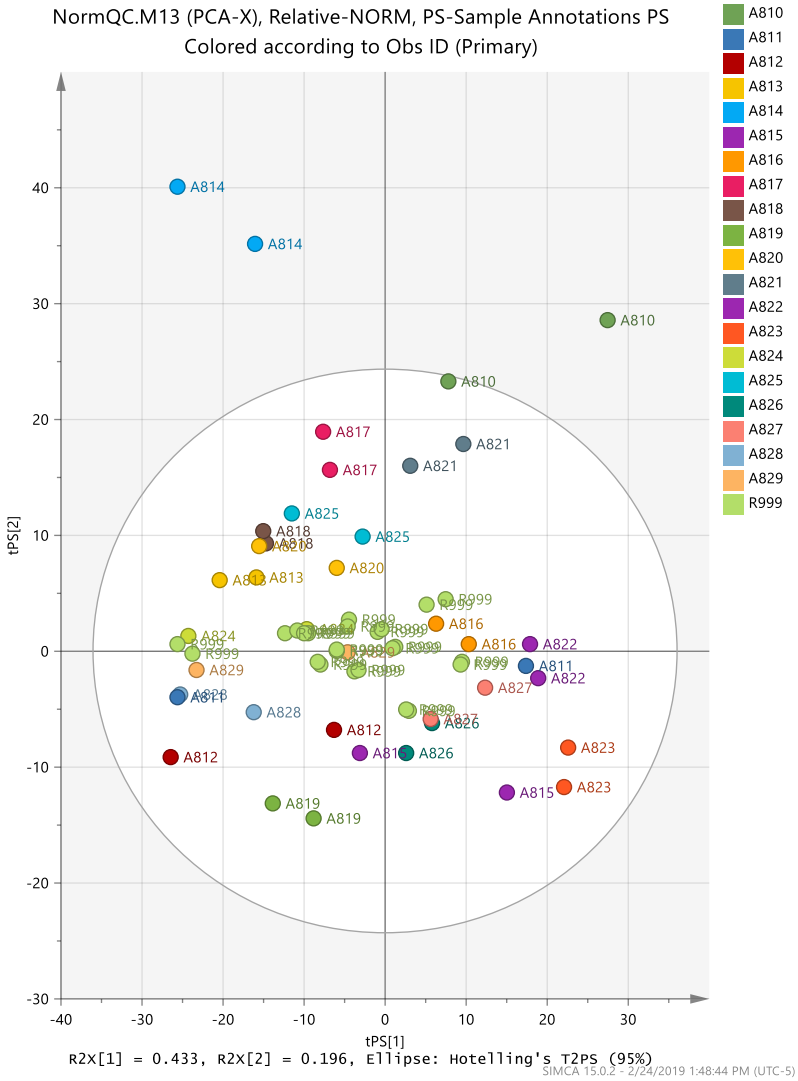

Figure 10d: PCA Relative Quantitation Post-Normalization.

Comparisons were made between pre- and post-normalization sets regarding the reproducibility of the PQC samples and the replicated subject samples. For the PQC samples, the average Z value across transitions was tabulated. For the replicated pairs, the root mean square differences were tabulated (Table 5). In general, normalization improved the average reproducibility of the data but the gains were small. The large sample size for this project contributed to the block effects, which were small in magnitude, being statistically significant. Since all the timepoints for a subject were run within half a plate, the run order block effects would be minimal and shifting plate means would have no effect on tabulating changes from baseline. The normalization may have minor impact on cross sectional comparisons of sample groups that span plates and runs. Both the raw and normalized data have been included in the uploaded dataset.

Table 5: Reproducibility of replicated samples before and after normalization

	Average Z	
PQC Samples	Raw	Norm
Absolute	0.04	0.05
Relative	0.6	0.4
	Average RMSE	
Replicates	Raw	Norm
Absolute	0.43	0.3
Relative	0.78	0.74

6. References

Addona, T.A. et al., 2009. Multi-site assessment of the precision and reproducibility of multiple reaction monitoring-based measurements of proteins in plasma. Nat Biotechnol. (27) 633.

Geyer, P.E. et al., 2017. Revisiting biomarker discovery by plasma proteomics. Mol Sys Biol. (13) 942.

Hoffman, G. E. & Schadt, E. E., 2016. variancePartition: interpreting drivers of variation in complex gene expression studies. *BMC Bioinformatics,* Volume 17.

Kennedy, J.J. et al., 2014. Demonstrating the feasibility of large-scale development of standardized assays to quantify human proteins.Nat Methods (11) 149.

Lange, V. et al., 2008. Selected reaction monitoring for quantitative proteomics: a tutorial. Mol Syst Biol. 4:222

Stacklies, W. et al., 2007. pcaMethods - a Bioconductor package providing PCA methods for incomplete data. *Bioinformatics,* Volume 23, pp. 1164-1167.

Appendix 1

FNIH Biomarkers Consortium

Statistical Analysis Plan

Biomarkers Consortium CSF Proteomics Project

1. INTRODUCTION

The project aims to qualify candidate Alzheimer's disease (AD) biomarkers using cerebrospinal fluid (CSF) samples from Alzheimer's Disease Neuroimaging Initiative (ADNI). Previously completed stages led to the identification of several polypeptide analytes with potential diagnostic or predictive utility. The current proposal aims to explore within subject changes in the five most promising potential biomarkers over time at various stages of the illness. The five analytes under study—1) chromogranin-A (CgA); 2) neural pentraxin 2 (NPTX2); 3) neurosecretory protein VGF; 4) secretogranin-2 (SCG2); and 5) fatty acid binding protein 3 (FABP3)—are derived both from comparisons of baseline values in controls, "progressing and non-progressing" mild cognitive impairment (MCI) and Alzheimer's disease (AD) subjects, as well as published reports from longitudinal analyses.

Ultimately, our goal is to identify analytes whose abundance exhibits intra-individual trajectories that correspond to the individual's disease progression, which may lead to the discovery of novel biomarkers with utility as supportive endpoints in clinical trials of early AD.

1.1. Study Objectives

Primary

The primary objective of this study is to retrospectively investigate the longitudinal changes in the estimated abundance of the five candidate analytes (CgA, NPTX2, VGF, SCG2, and FABP3) in individuals from the ADNI cohort at various stages of the disease including cognitively normal (NL), and mildly cognitively impaired (MCI) subjects.

Secondary

The secondary objectives are to retrospectively compare the rates of longitudinal change of the five candidate proteins (CgA, NPTX2, VGF, SCG2, and FABP3) between subpopulations of the ADNI cohort defined by: a) p-Tau/Aβ status at baseline; and b) progressors vs. non-progressors.

1.2. Study Design

The participants in this study are a subset of the ADNI I, II, GO Cohorts, who have had a minimum of three longitudinal CSF collections and clinical assessments. The participants range across the disease spectrum from those diagnosed as cognitively normal (NL) to those diagnosed as MCI and AD, as defined in ADNI. They include 80 NL, 114 EMCI+MCI and 4 AD subjects; each with a minimum of three longitudinal CSF collections and clinical assessments. A total of 750 samples were distributed into 21 randomization blocks. The randomization blocks were allocated to plates (2 blocks per plate), which were run in sequence using two different mass spectrometers. Additional QC samples, (n=20) composed of replicates of 20 study samples, were interspersed across blocks. All longitudinal samples for a subject were allocated in the same block in a random order. The validity of this assumption can be tested after the data become unblinded.

Data used in the analyses described in this document are based on experiments conducted by Caprion Biosciences. These experiments involved an MRM assay developed in-house to estimate absolute abundance of the five specified analytes (CgA, NPTX2, VGF, SCG2, and FABP3). The abundance was estimated using a single transition selected out of two assayed. The rules for selecting transitions, data-processing steps including normalization and corrections for run effects, etc., are described in the Data Primer (Reference). The final normalized analysis dataset will be used to carry out the analyses. Additional data available in ADNI for these subjects, such as demographic information, clinical assessments, and biomarker results, will also be used as specified in the analyses.

1.3. Statistical Hypotheses for Trial Objectives

Statistical analyses will be carried out separately for each of the five proteins (CgA, NPTX2, VGF, SCG2, and FABP3). There are two families of hypotheses:

1. The null hypotheses for the primary family of analyses state that for a given analyte and each diagnostic group (NL, EMCI+MCI), there are no significant changes in the analyte's abundance over time (i.e., the rate of change from baseline is zero). The alternative is that the rate of change associated with the given analyte is significantly different from zero within a given diagnostic group. Both positive and negative rates of change will be examined, as the abundance of relevant analyte may increase or decrease with increasing pathology. We will also compare the rates of change from baseline in analyte abundance between sets of individuals defined by the three diagnosis groups (NL, EMCI+MCI), with the statistical null hypotheses that there is no significant difference between the rates of change for subjects in the three diagnosis categories. The alternative hypothesis is that, for a given analyte, the rate of change in one diagnostic group is significantly different from another group.

2. The statistical null hypotheses for the secondary family of analyses state that there is no significant difference between the rates of change within each of the groups—NL and EMCI+MCI—for participants classified by the following criteria:

    a. baseline p-Tau/Aβ ratio > 0.025 (positive) versus baseline p-Tau/Aβ ratio ≤ 0.025 (negative)

    b. baseline diagnosis of NL (EMCI+MCI) followed by a diagnosis of EMCI+MCI (AD) obtained within 4 years following the baseline assessment (progressors) versus baseline diagnosis of NL (EMCI+MCI) followed by the same diagnosis obtained during all clinical assessments carried out within 4 years (assuming there is at least one such assessment) following the baseline assessment (non-progressors).

    The alternative hypothesis is that the difference in rates of change from baseline in the protein's abundance between pairs of the subgroups defined above is significantly different from zero.

The primary and secondary endpoint is the change from baseline of the estimated abundances of each of the five candidate proteins. The summary measures are the estimate of the rate of change in the form of a linear slope for each disease stage (primary), and estimate of the difference between slopes for pairs of cohorts (i.e., progressors versus non-progressors and p-Tau/Aβ ratio positive versus negative) at each disease stage (secondary).

1.4. Sample Size

Sample sizes were limited to participants in the ADNI cohort with a minimum of three CSF sample collections and clinical assessments.

1.5. Randomization and Blinding

To minimize experimental effects such as run-order effects, the following randomization requirements were provided to the ADNI team providing the CSF samples, regarding the experimental design:

I. Longitudinal samples from a given subject are to be run on the same plate but randomized across the plate.

II. Stratified (by column) randomization of the subjects by disease stage and amyloid status (NC-Aβ positive, NC-Aβ negative, MCI-Aβ positive, MCI-Aβ negative, and AD) is to be carried out.

III. Samples are to be randomly assigned to run order.

Note that subsequent to these requirements, the experimental plan was changed.

The original study randomization was performed by Les Shaw's group at UPenn and was based on a sample processing workflow, which included a CSF depletion step. The 750 CSF samples were split into groups of 36 samples with all longitudinal samples from an individual contained within the same group (called run01, run02 etc.). Eleven (11) runs were planned to be depleted using one IgY14/Supermix depletion column and ten (10) runs were to be depleted using a second

depletion column. Based on further assay development, the depletion step was deemed unnecessary and was eliminated from the sample processing workflow. A modified workflow involved four runs using two 6500 QTRAP instruments instead. The group assignment and injection order described in the initial randomization scheme were preserved. The randomization groups were divided into the four runs as described below:

Caprion Run	Batch size	Instrument	UPenn Randomization Designation
Run 01	186	U	run01-run05 (depletion column 1)
Run 02	186	Z	run06-run10 (depletion column 1)
Run 03	180	U	run11 (depletion column 1) and run01-run04 (depletion column 2)
Run 04	222	Z	run05-run10 (depletion column 2)

All project team members remain blinded to the participant IDs in ADNI. Thus, experimental data cannot be linked to demographic, clinical or other biomarker data for the participants until the experimental data is uploaded to the ADNI website.

2. SUBJECT INFORMATION

Subject demographics, and baseline disease characteristics will be summarized using descriptive statistics.

2.1.1. Baseline Demographic Characteristics

Table 1 lists the demographic variables that will be summarized.

Table 1: Demographic Variables

Continuous	**Summary Type**
Age (years)	- Descriptive statistics (N, mean, standard deviation, median, and range [minimum, maximum]).
Weight (kg)	
Height (cm)	
Categorical	
Age ([18-25 years, 26-50 years, 51-64 years, and 65-74, and ≥74 years])	

Sex (Male, Female, Unknown)	- Frequency distribution with number and percentage in each category
Race (American Indian or Alaska Native, Asian Black or African American, Native Hawaiian or Other Pacific Islander, White, Other, Multiple, Not reported, Unknown)	
Ethnicity (Hispanic or Latino, Not Hispanic or Latino, Unknown, Not Reported)	
Marital status (Single, Married, Divorced, Separated, Widowed, Not reported)	
Educational level (9-12 years, College, Not Reported)	

2.1.2. Baseline Clinical Characteristics

Table 2 lists the baseline disease characteristics that will be summarized for the Analysis set.

Table 2: Baseline Disease Characteristics

Categorical	
Diagnosis at baseline (NL, EMCI+MCI, AD)	- Frequency distribution with number and percentage in each category
p-Tau/Aβ ratio (> 0.025, ≤ 0.025)	
ApoE4 status (Carrier, Non-carrier)	
Diagnosis at baseline for subjects with p-Tau/Aβ ratio > 0.025 and ≤ 0.025	
Diagnosis at baseline for subjects with ApoE4 carriers and non-carriers	

2.1.3. Clinical Characteristics

Table 3 lists the additional disease characteristics that will be summarized for the Analysis set.

Table 3: Disease Characteristics

Categorical	
Number of NL -> EMCI+MCI progressors	- Frequency distribution with number and percentage in each category
Number of EMCI+MCI -> AD progressors	
Number of progressors in each group with p-Tau/Aβ ratio > 0.025 and ≤ 0.025	
Number of progressors in each group who are Apoe4 carriers and non-carriers	

2.2. Disposition Information

The following information will be reported: (i) the number and percentage of subjects in the analysis population at baseline, year 1, 2, 3, … *max* after baseline clinical assessment (ii) the number and percentage of subjects who completed a clinical assessment and provided a CSF sample at baseline and at year 1, 2, 3,… *max* after baseline visit, (iii) median, range and inter-quartile range of the number of years between subjects' baseline and final CSF collection and clinical assessment.

3. ANALYSIS DEFINITIONS & SPECIFICATIONS

The full analysis set (FAS) will be defined as all participants with a minimum of three CSF sample collections and clinical assessments.

Subgroups will be defined as follows:

(i) Diagnosis at baseline (NL, EMCI+MCI)
(ii) p-Tau/Aβ ratio > 0.025 versus p-Tau/Aβ ratio ≤ 0.025
(iii) progressors vs. non-progressors (as defined in section 1.3.)

Study Day 1 or Day 1 refers to the day of the Baseline CSF assessment. All assessments at all visits will be assigned a day relative to this date. Study day or relative day for a visit is defined as: visit date - (date of Day 1) +1, if visit date is ≥ date of Day 1; visit date - date of Day 1, if visit date < date of Day 1. There is no Day 0. Baseline is defined as the first CSF and clinical assessment. Endpoint is defined as the last available postbaseline CSF and clinical assessment.

The primary objective of this study is to retrospectively investigate the longitudinal changes in the estimated abundance of the five candidate proteins (CgA, NPTX2, VGF, SCG2, and FABP3) in individuals from the ADNI cohort at various stages of Alzheimer's disease from cognitively normal (NL), mildly cognitively impaired (EMCI+MCI) to dementia (AD). The secondary objectives are to retrospectively compare the rates of longitudinal change of the five candidate proteins (CgA, NPTX2, VGF, SCG2, and FABP3) between subpopulations of the ADNI cohort defined by: a) p-Tau/Aβ status at baseline; and b) progressors versus non-progressors.

The primary analysis will employ a Linear Mixed Effects (LME) modelling methodology, with fixed effects including the baseline abundance, diagnosis at the given time point (NL, EMCI+MCI as a factor), sex, ApoE ε4 carrier status, education level, age at study entry, time point (as a continuous variable), and the time by diagnosis interaction. Subject will be included as the random factor. Other variables may be included as appropriate. The coefficients associated with each diagnostic category, and the diagnosis-by-time interaction, with the corresponding 95% confidence intervals will be estimated.

The secondary analysis will also fit an LME model to the data, with fixed effects including the baseline estimated protein abundance, p-Tau/Aβ ratio positivity (as a factor), progression status, diagnosis (NL, EMCI+MCI as a factor), sex, ApoE ε4 carrier status, education level, age at study entry, time point (as a continuous variable), and the time by disease stage interaction. Subject will be included as a random factor. Other variables may be included as appropriate. The coefficients associated with p-Tau/Aβ ratio positivity and progression status, their interaction with time, and the appropriate interactions will be estimated with the corresponding 95% confidence intervals.

Results will be presented in the form of tables and graphs. One table will contain the demographic and clinical characteristics of the analysis sample using variables and summary statistics detailed in sections 2.1.2. Additional tables will describe the model estimates and 95% confidence intervals associated with the primary and secondary endpoints. For each protein, graphs will depict the estimated linear trajectories (with 95% confidence bands) of protein abundance associated with the three diagnosis categories (NL, EMCI+MCI). A second set of graphs will show the estimated linear trajectories (with 95% confidence bands) associated with each subgroup (progressors vs. non-progressors, p-Tau/Aβ ratio positive vs. negative) and each diagnosis category (NL, EMCI+MCI). Graphs of profile plots of individual trajectories will also be depicted. All the hypothesis tests specified here are two-sided. If a p-value is ≥ 0.05, then the comparison associated with this p-value will be declared not statistically significant, otherwise if the p-value is < 0.05, then the comparison associated with this p-value will be declared statistically significant.