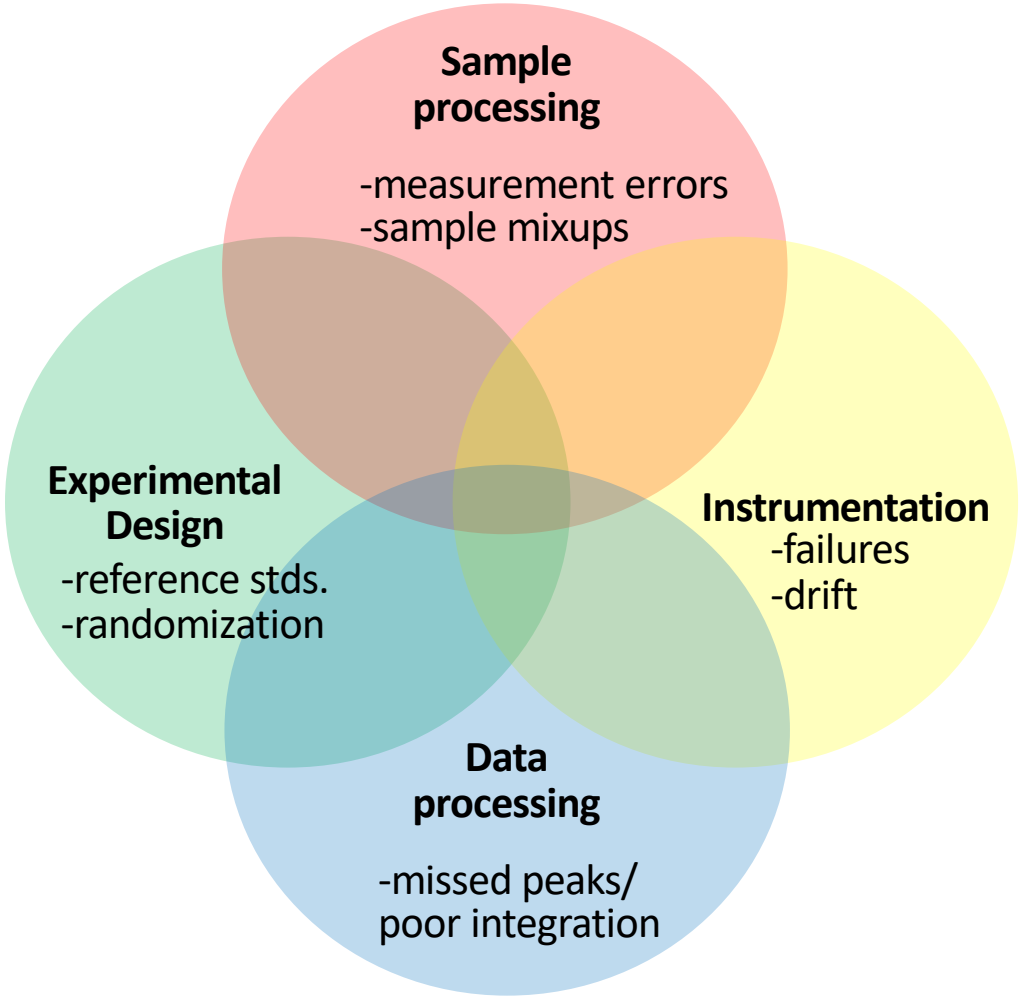
QC in Untargeted Metabolomics

Ping-Ching Hsu
Clary Clish
Dan Bearden

Introduction

- The purpose of quality control (QC) is to monitor the performance of metabolomics workflows against standards to detect problems and inform corrective actions
- Why do we need QC in untargeted metabolomics?
- Metabolomics is a complicated process; variability and problems may come from a number of sources, individually or in combination



Outline

- General QC practices for untargeted metabolomics
  - Study design & QC practices used during data acquisition
  - QC practices used during data processing

- Real life examples
  - Replicates in LC-MS
  - QC in larger studies of human disease
  - Use of test materials (e.g. NIST Standard Reference Materials; SRMs)

Study design and QC during data acquisition

- Column conditioning
  - SOPs for preparing LC columns and evaluating performance
- Randomization of sample analysis order
  - Mitigate systematic bias
- Pooled samples
  - Regular, repeated measures of a representative sample
  - “Real time” review during analysis of large sample numbers
- Blanks
  - Identification of system contaminants and batch-to-batch carryover of biological sample
- Replicates (technical and process)
  - Evaluation of reproducibility
- Internal standards
  - “Real time” review during analysis of large sample numbers
  - Acceptance criteria and triggering repeats
- Reference samples
  - Metabolite standards, long term reference samples, Standard Reference Materials (SRMs)

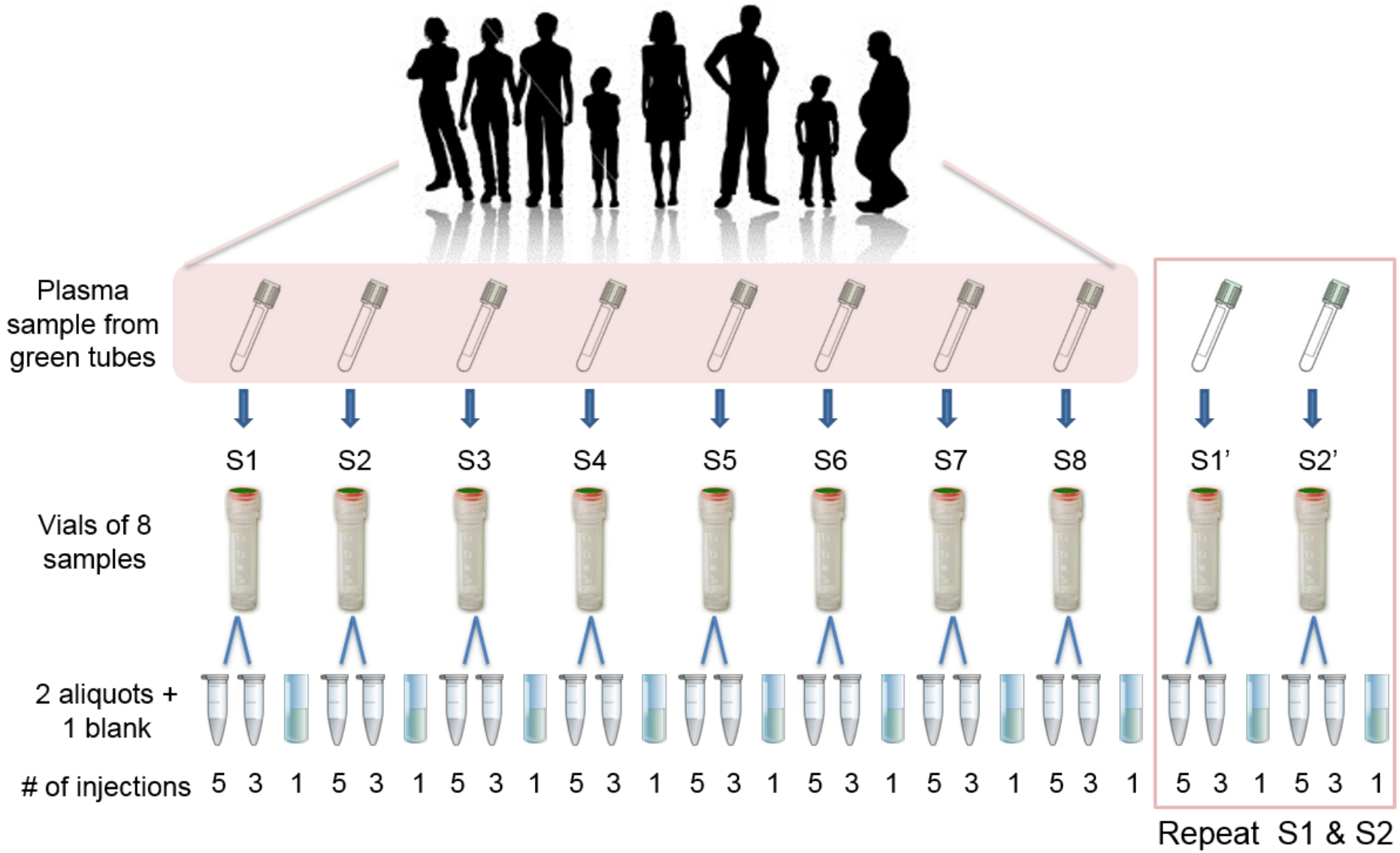
QC during data processing

- Pooled QC samples
  - Overlay of raw data (e.g. TIC) among pooled QCs
  - Evaluation of coefficients of variation for every metabolite
- Review of internal standards among all samples
- Principal component analysis
  - Identification of obvious outliers
  - Confirmation of clustered pooled QCs, replicates, and/or reference samples
  - Batch effects
- Correlation of replicate samples
- Manual review of peaks
  - Confirmation of accurate peak integration (mainly “knowns”)
- Peak filtering and data reduction
  - Redundant ion features, features with many missing values, features above a CV threshold, …

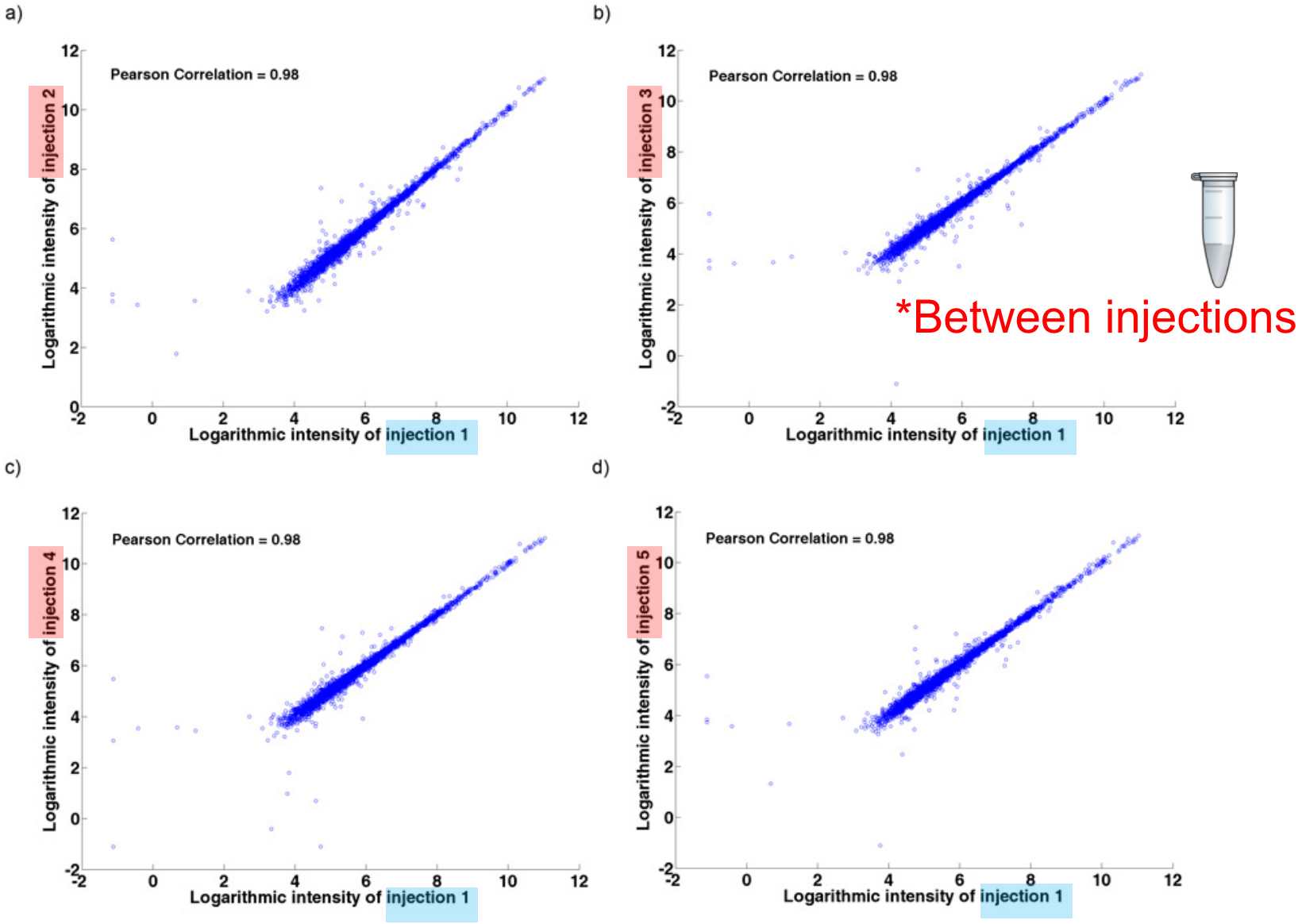
Example 1:
Replicates in LC-MS

Ping-Ching Hsu

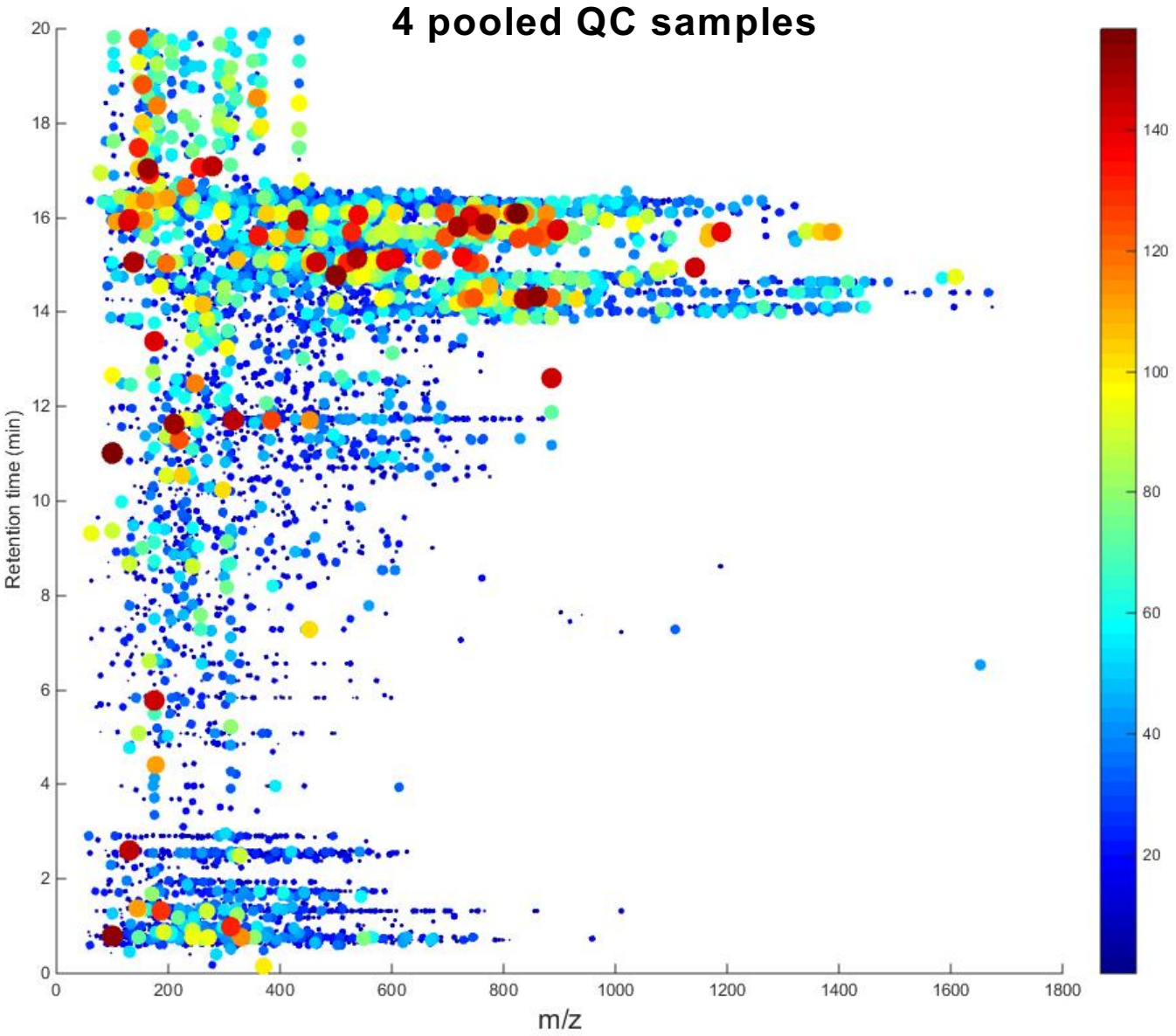
Experimental design used to test the reproducibility of UPLC-QTOF-MS



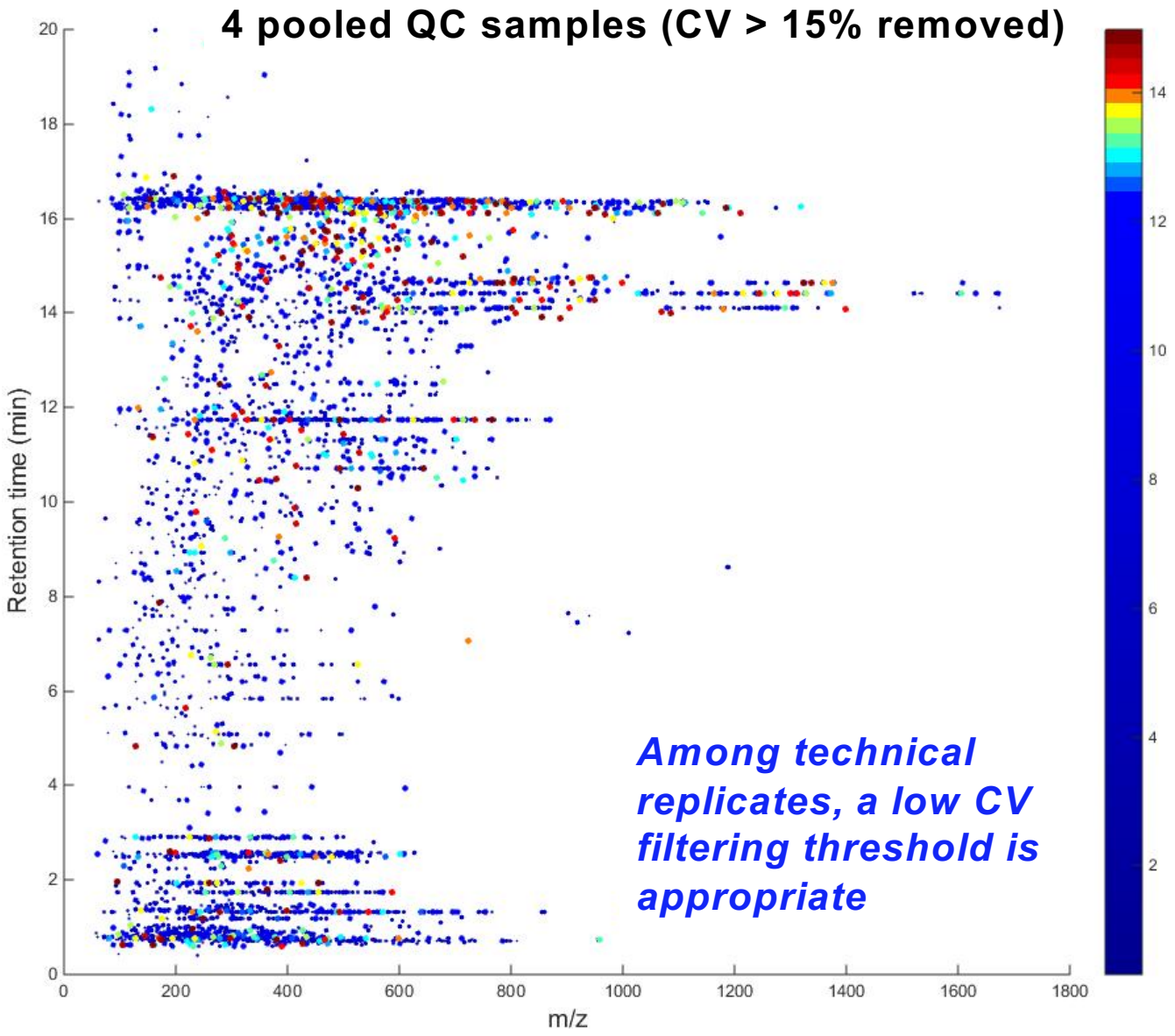
Evaluation of the variation due to the measurement noise



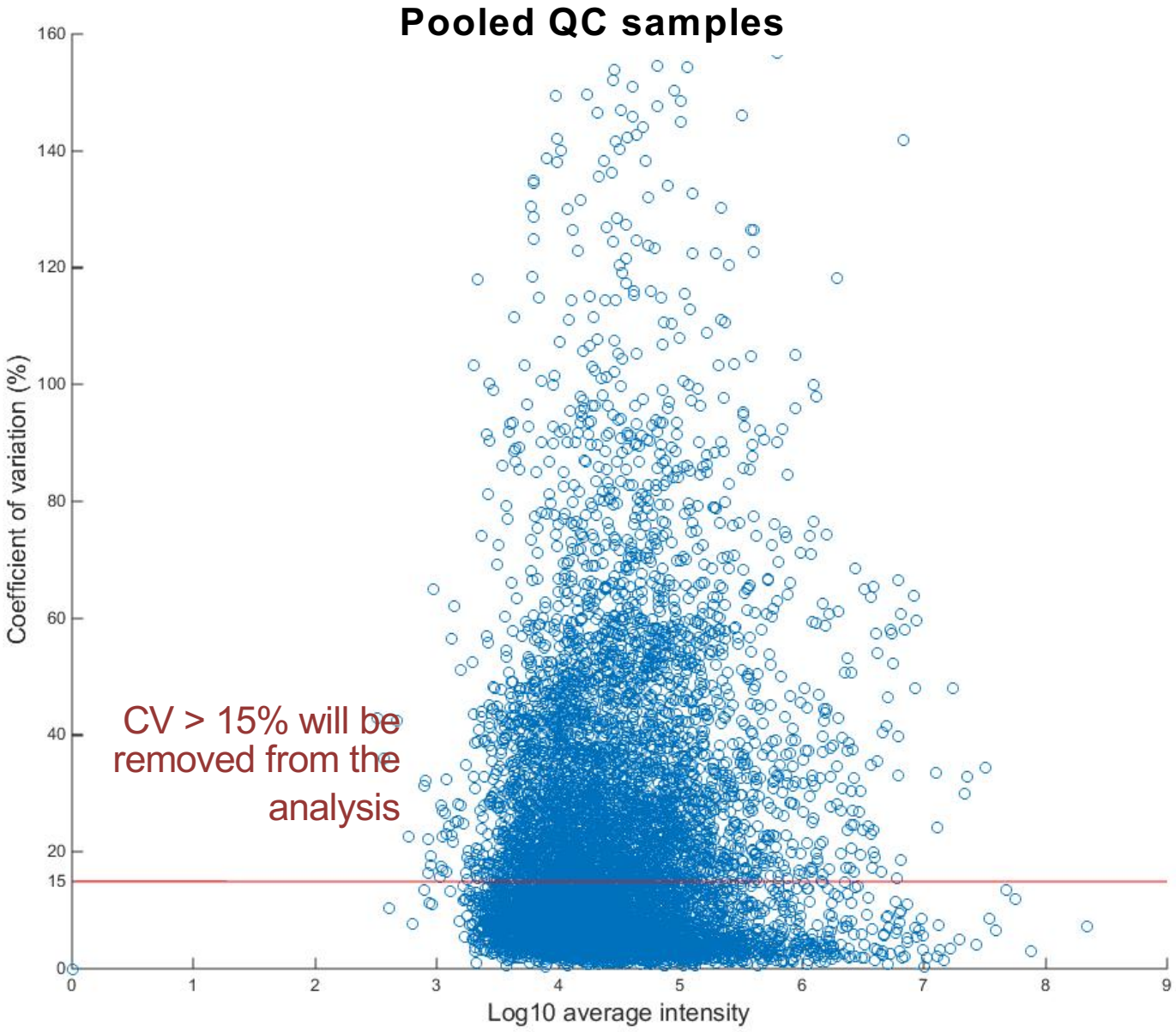
Distribution of the estimated measurement error in CV(%)



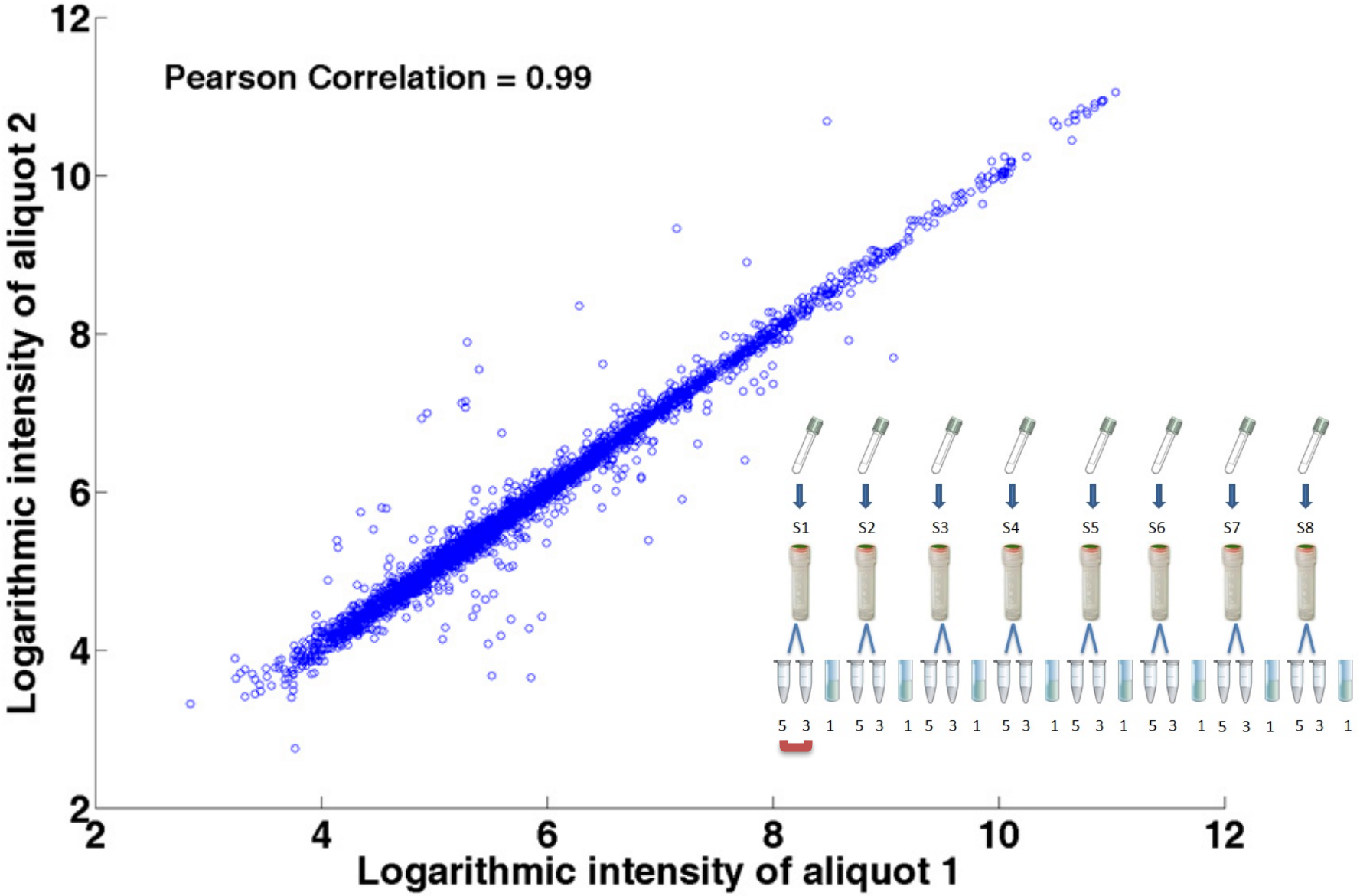
Filtering out features based on a CV threshold



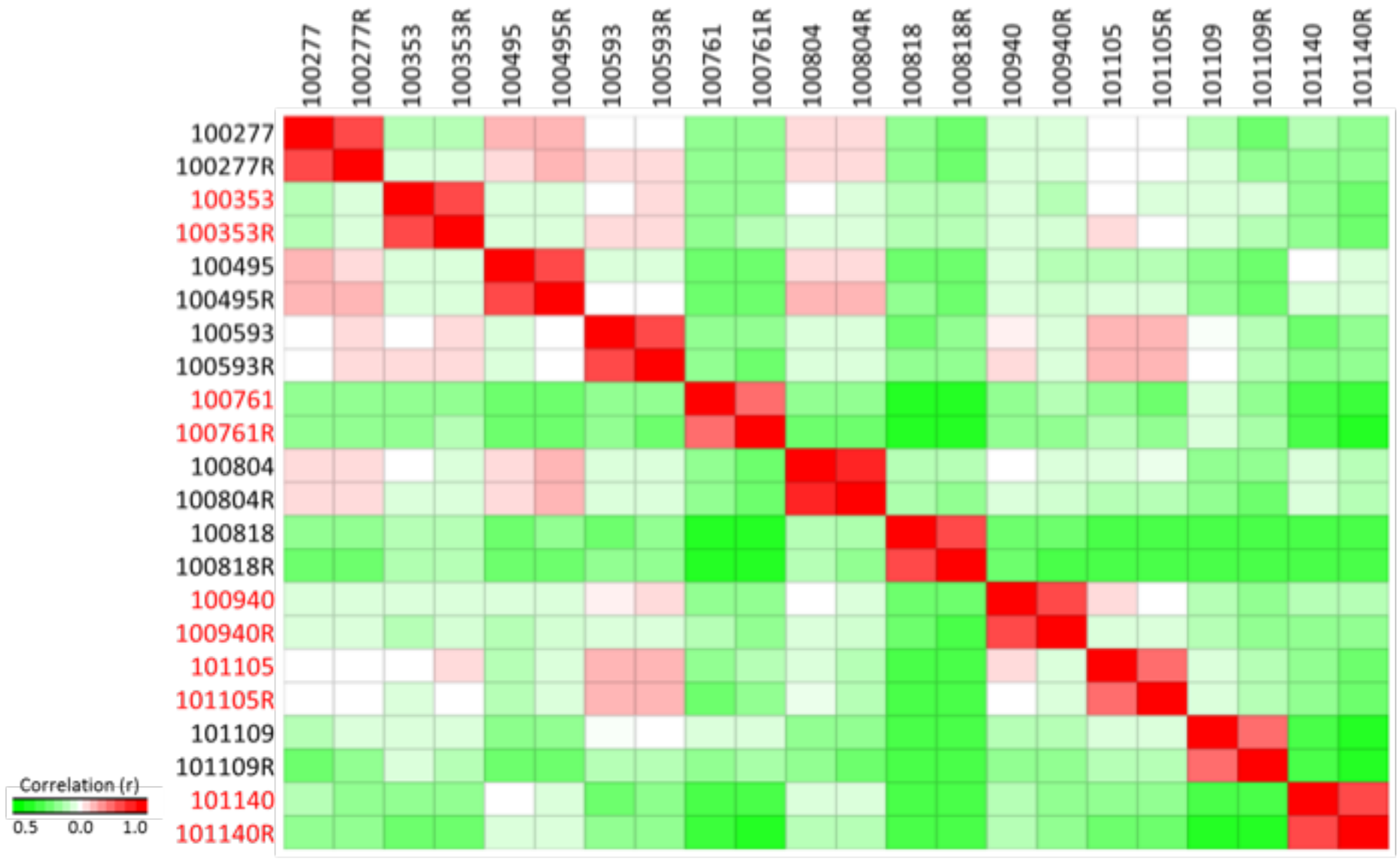
Scatter plot of the estimated measurement noise in CV (%)



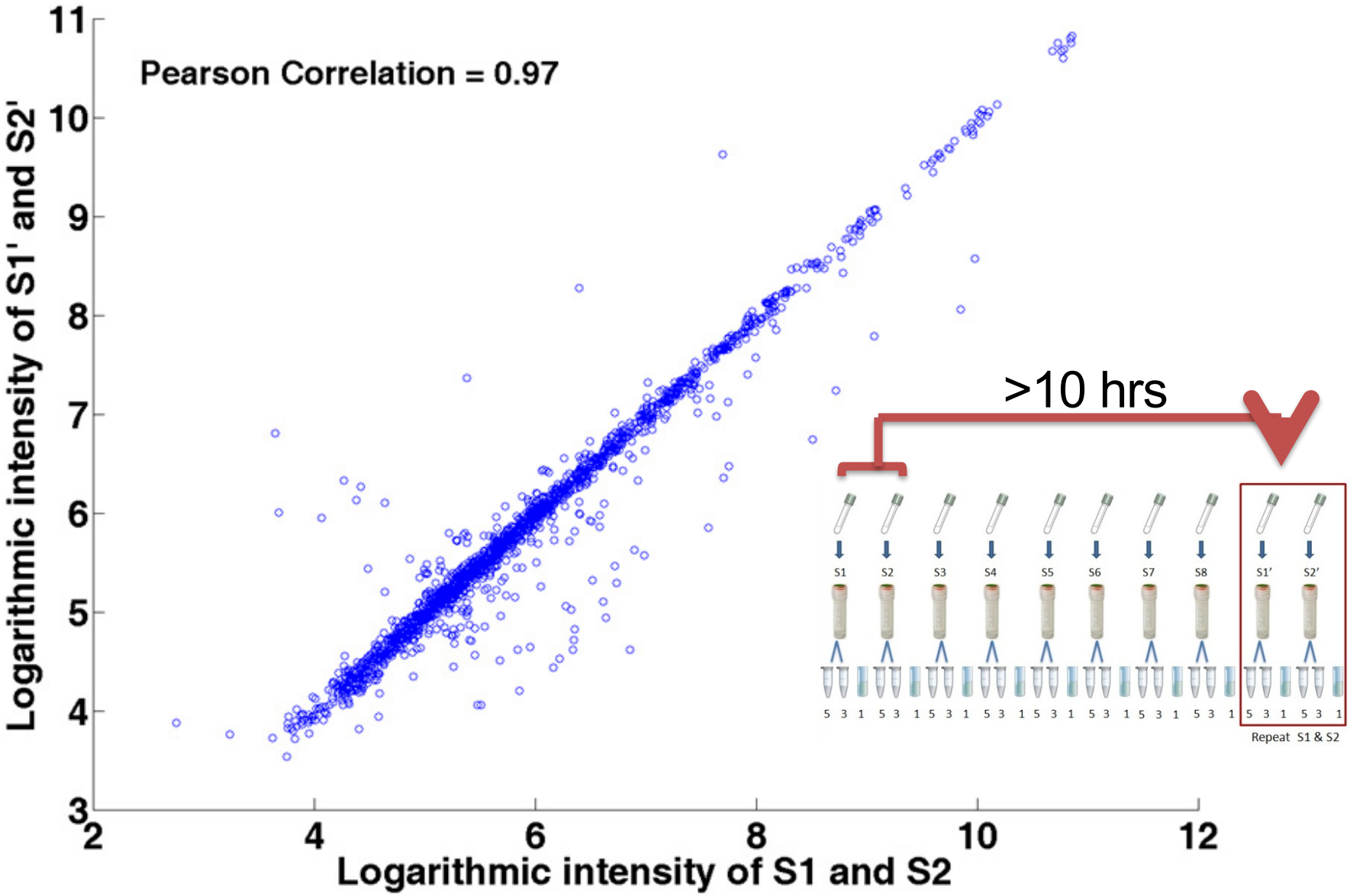
Evaluation of the variation due to sample preparation



Correlation of replicates

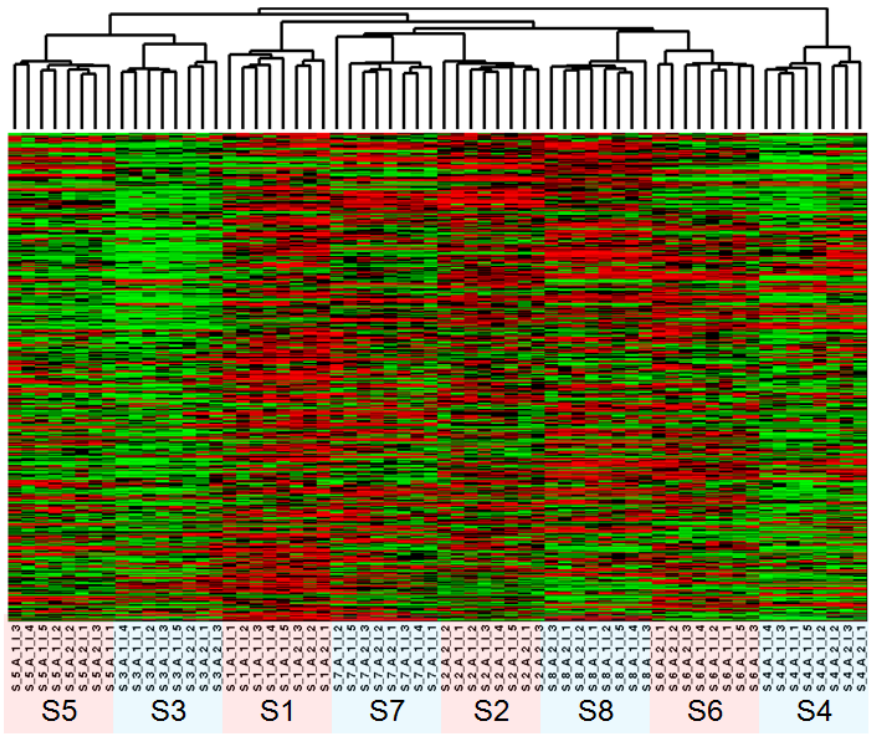


Evaluation of the effects on run time, measurement error & sample preparation variation

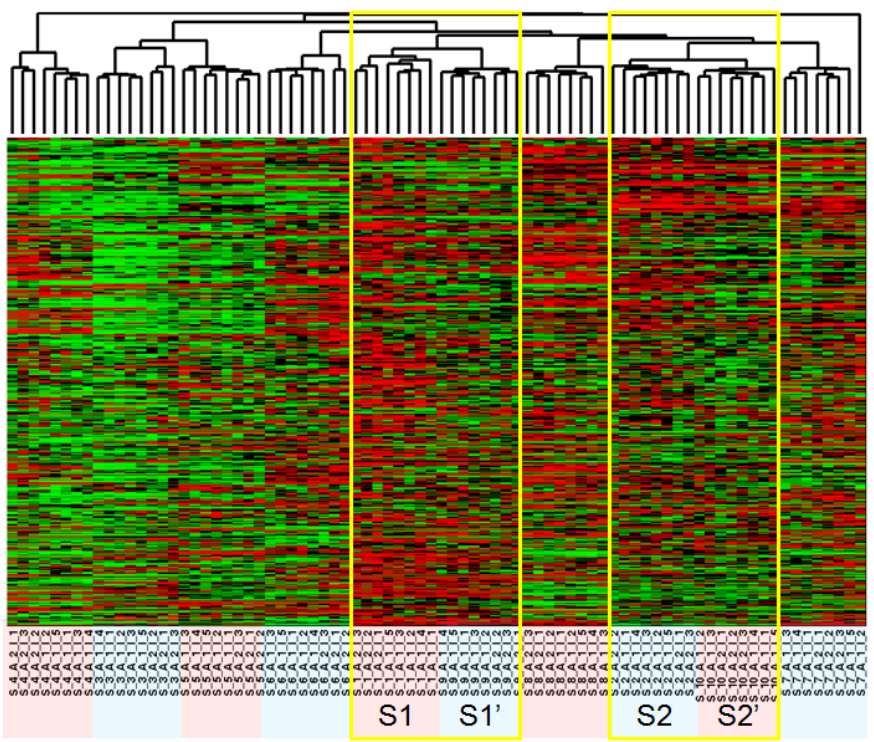


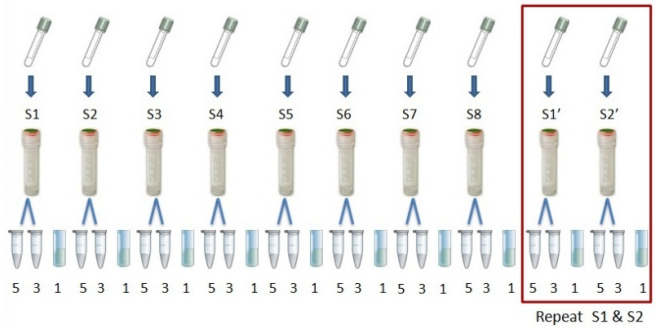
Hierarchical clustering of all metabolites with and without analytical replicates

a)



b)

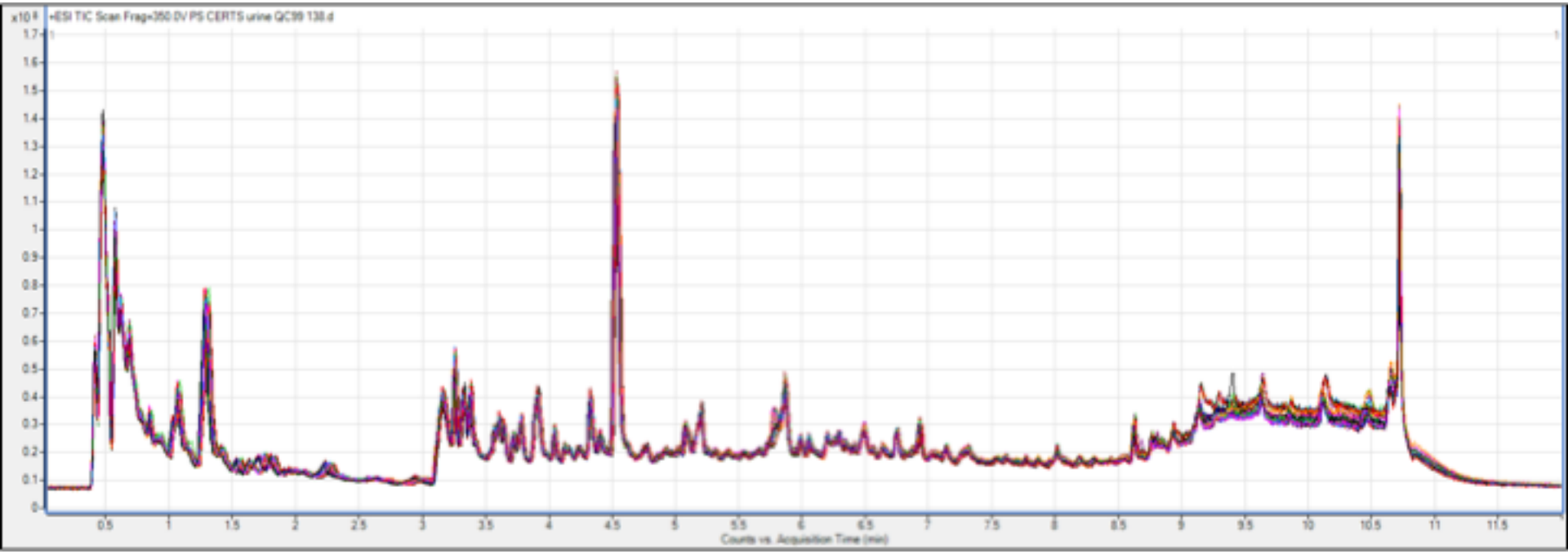




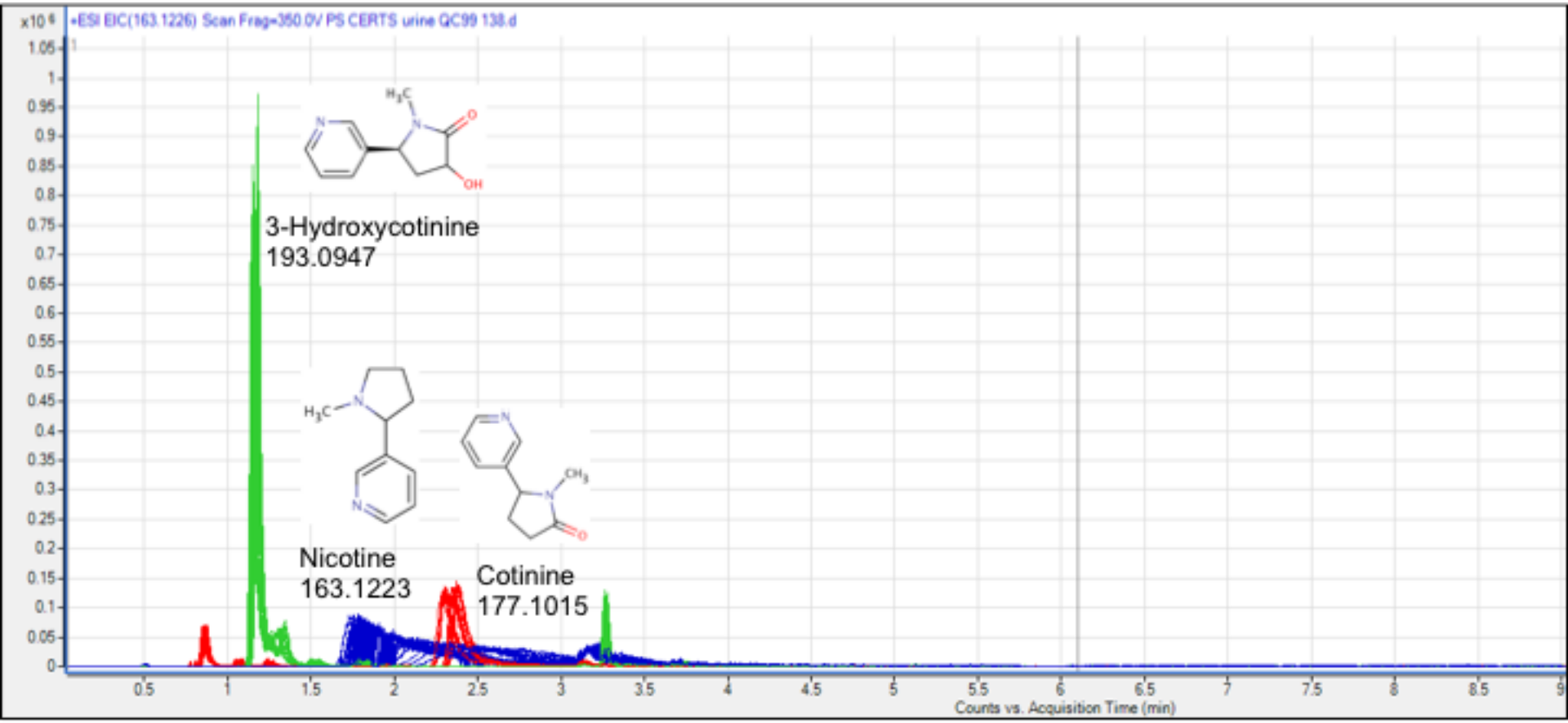
Summary of the measured variation in human plasma samples

	Mean of CV (%)	Median of CV (%)
Technical variation	7.2-8.8	5.7-7.2
Experimental variation	7.2-12.3	4.6-8.7
Biological	22.0-22.3	17.2-18.2

Visual QC: 31 TICs overlaid



Overlay of 3 nicotine metabolites among 31 QCs



Filtered out from the analysis →

	Mass	CV(%)
Nicotine	163.1226	22.33
Cotinine	177.1023	2.91
3-hydroxycotinine	193.0974	3.28

Summary Example 1

- It's important to pilot methods using technical and processing replicates in order to understand analytical performance
  - This type of pilot may be done before engaging in large studies
- Inclusion of technical and/or processing replicates may be feasible for smaller studies
- Measurement variation and sample preparation variation are generally low when samples are measured consecutively
  - Therefore, a low CV threshold may be applied to filter signals from replicate data

Example 2:
QC in larger studies of human disease

Clary Clish

Challenges associated with applying nontargeted methods to discover early indicators of disease in humans

- Metabolic dysregulation may be very modest early in disease
  - E.g. metabolite levels may differ by only 10% between incident cases and controls
  - Large sample numbers are needed for statistical power
- Funds tend to need to be applied to increase biological ”n’s” rather than cover cost of technical “r’s”
  - Replicates are generally out
- It’s often necessary to analyze samples over multiple LC columns and over periods of months
  - Risk of complications due to batch effects are high
- Data must be standardized across batches
- Small differences in measured retention times and MS mass calibration complicates “aligning” nontargeted features among batches

QC approach for large, nontargeted LC-MS-based studies

Reference mixtures analyzed before and after to assure system performance

Internal standard(s) added in first step of sample extraction

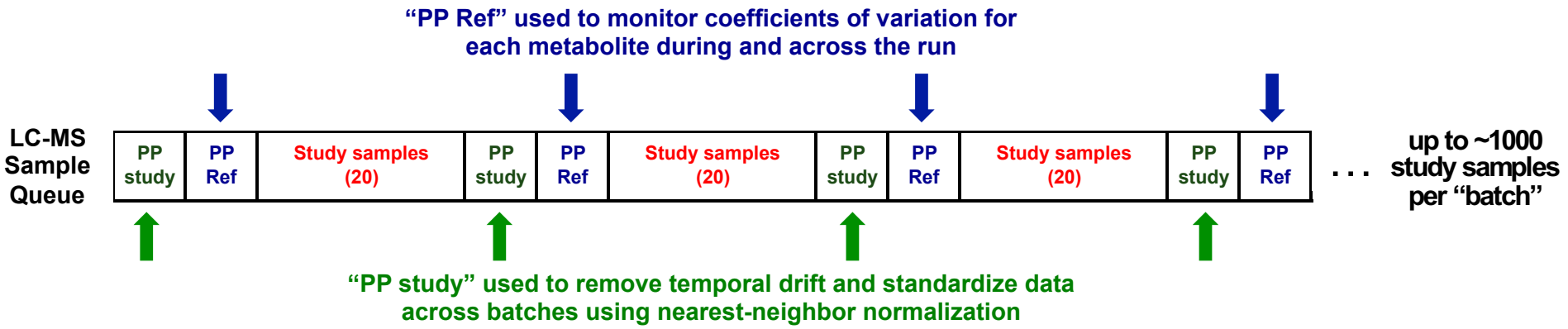
- monitored during analyses
- may be used to standardize data

Pooled study sample: analyzed every 20 study samples

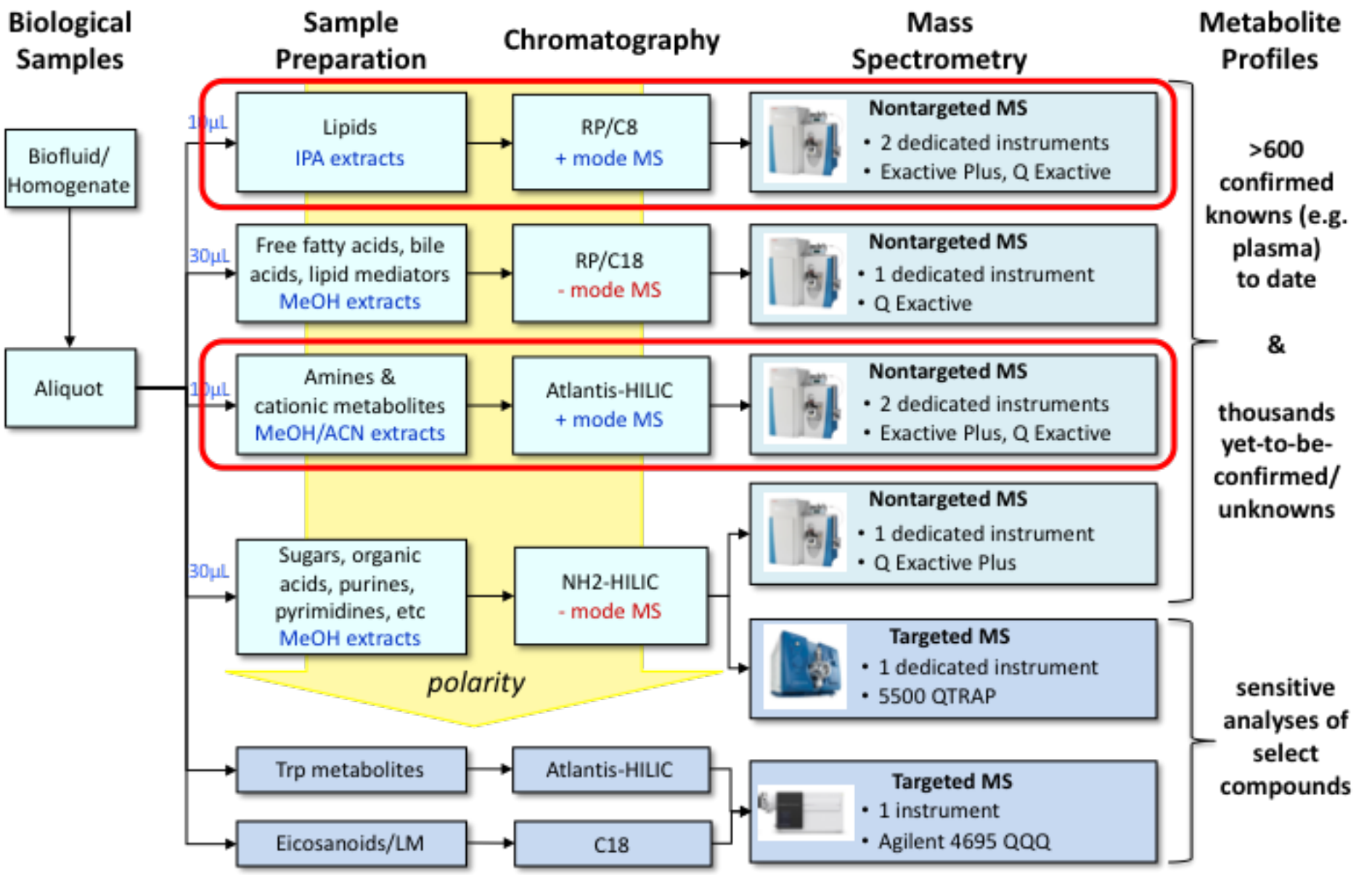
- used to standardize data across datasets

Second pooled reference sample, analyzed every 20 study samples

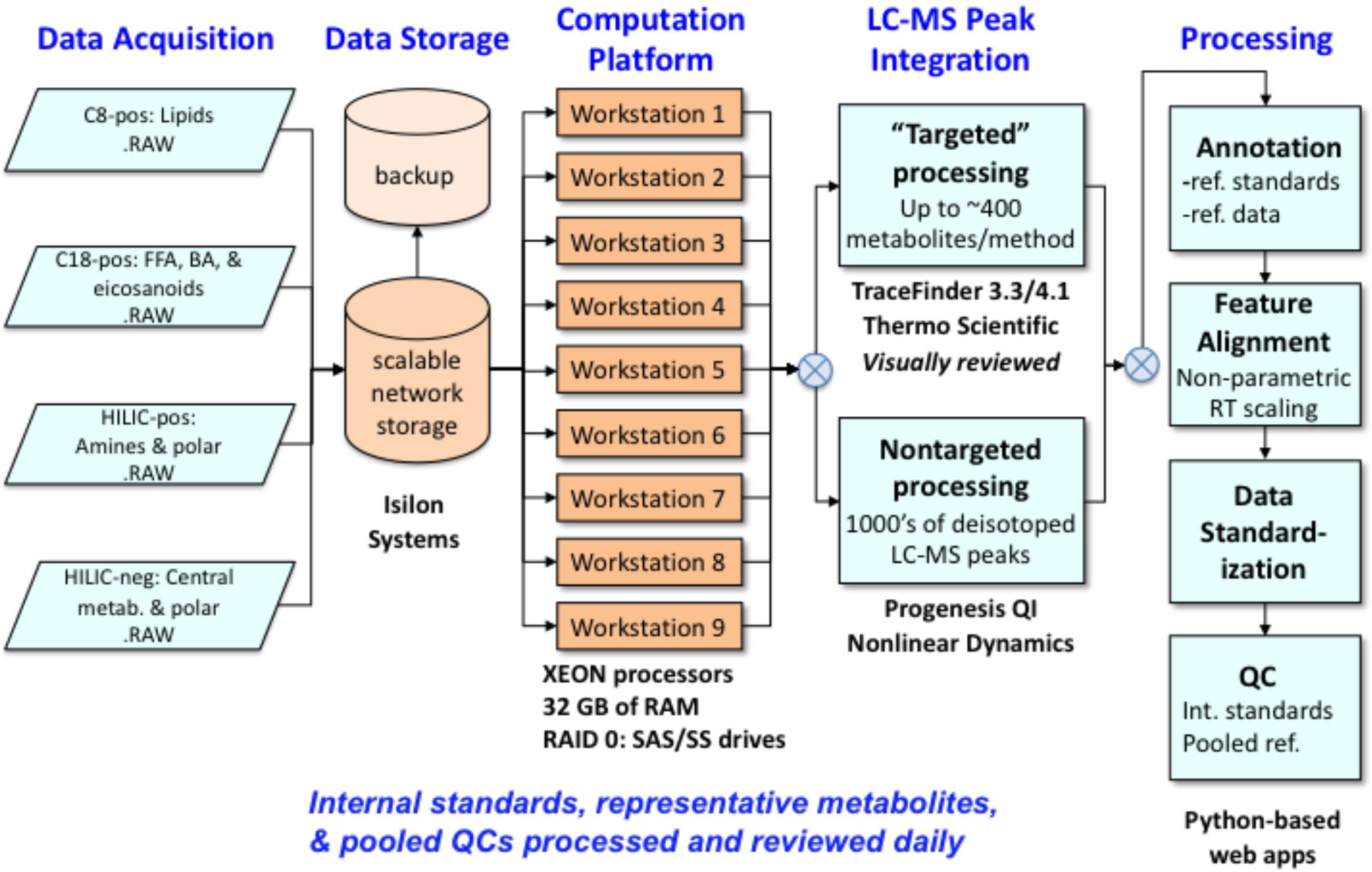
- used to assess: overall reproducibility & impact of standardization procedures
- we typically use the pooled study sample



E.g. Pilot study: 2000 human plasma samples from TOPMed

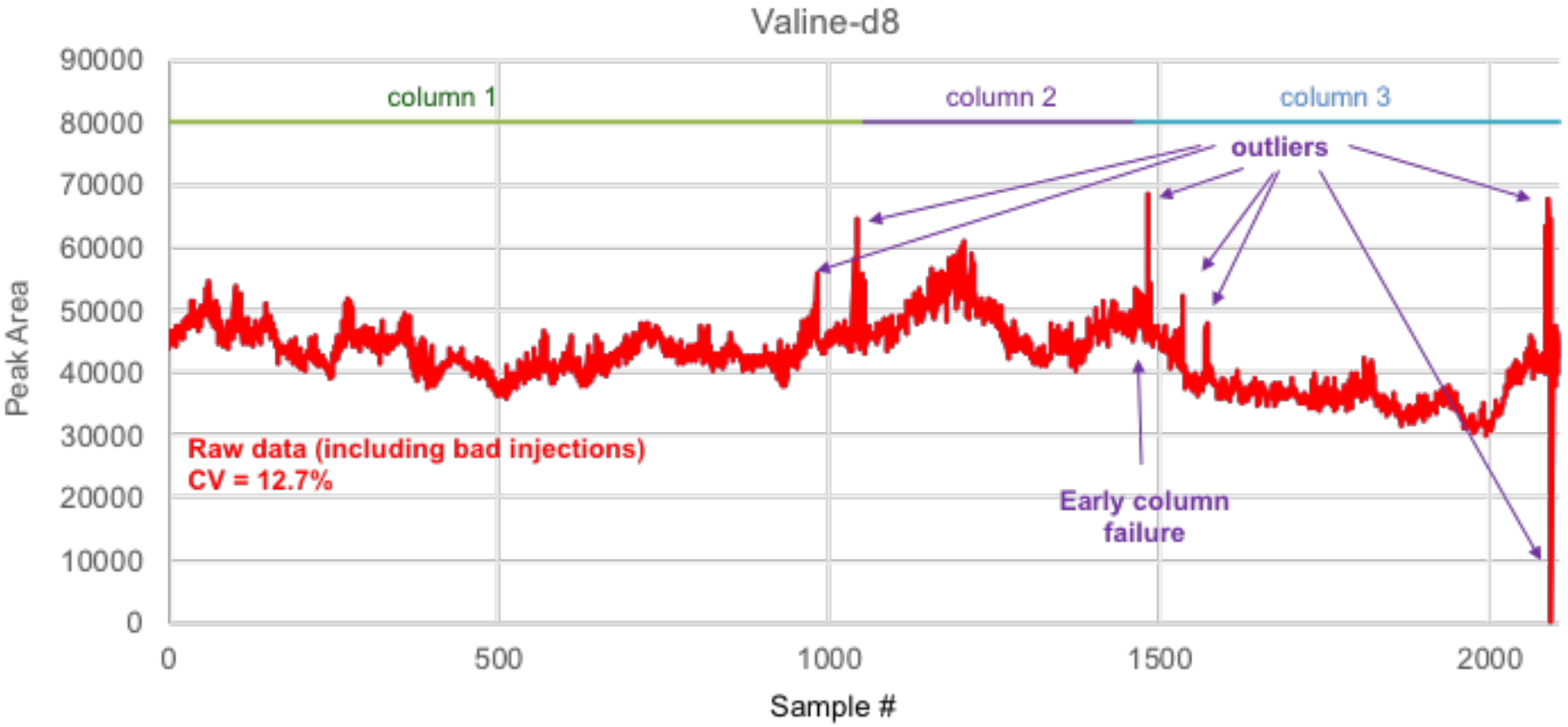


Nontargeted LC-MS metabolomics data processing workflow



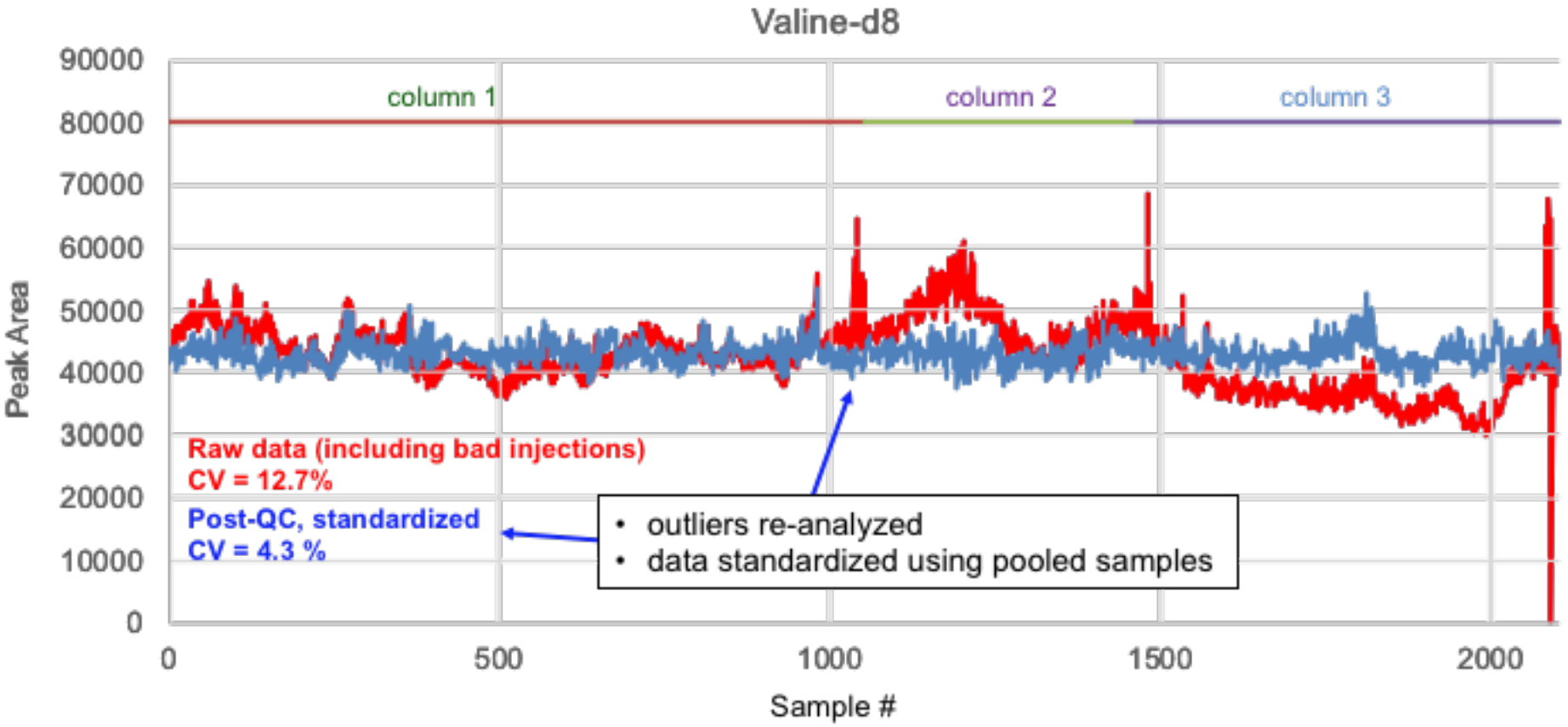
- Analysis plan:
  - 1000 samples/column x 2
  - 10% pooled QC samples
- Analyze samples nearly continuously for 1.5-2 months/method

Problems illuminated by QC:
- Injection problems
- Instrument noise and drift
- Failure during second HIL-pos column
- Samples flagged for re-analysis



- Analysis plan:
  - 1000 samples/column x 2
  - 10% pooled QC samples
- Analyze samples nearly continuously for 1.5-2 months/method

Problems illuminated by QC:
- Injection problems
- Instrument noise and drift
- Failure during second HIL-pos column
- Samples flagged for re-analysis



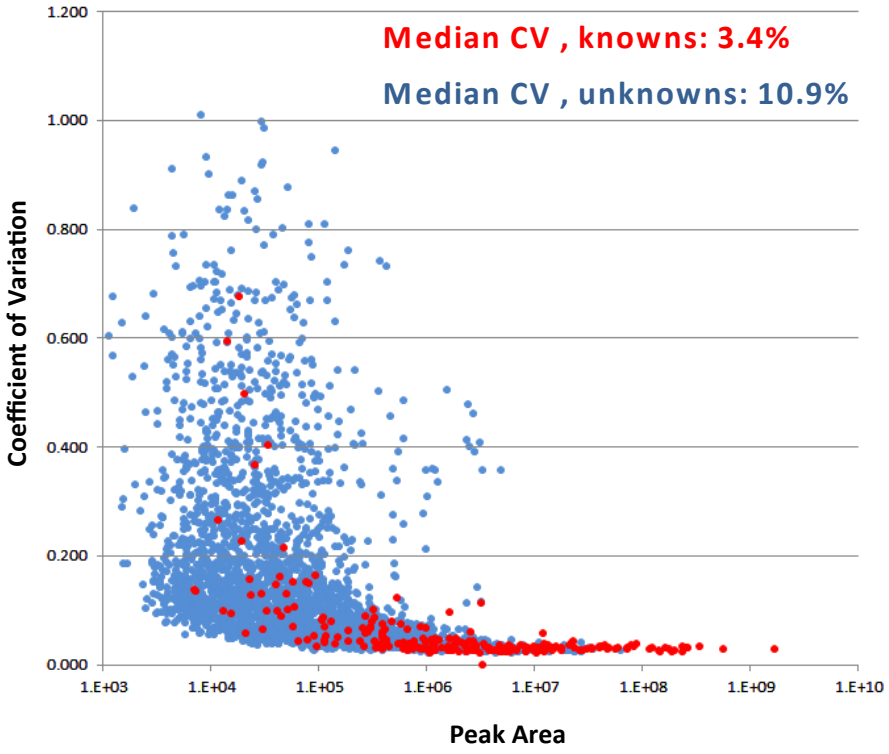
Evaluating reproducibility of pooled QC samples: Pilot study of 2000 TOPMed plasma samples

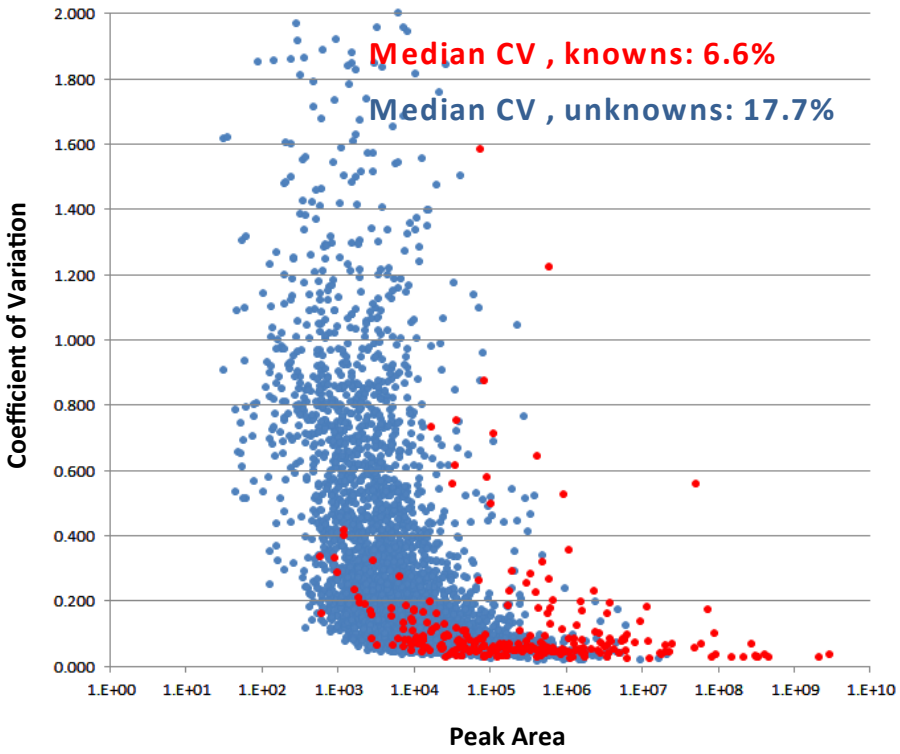
C8-pos

- Nontargeted analysis of lipids
- 2 columns; ~1.5 months
- 228 lipids of known ID
- 2662 unknowns aligned between two columns
- n = 98 pooled QC samples

HILIC-pos

- Nontargeted analysis of polar metabolites
- 3 columns; 2 months
- 253 confirmed knowns
- 3966 unknowns aligned across three columns
- n = 104 pooled QC samples





Do nontargeted methods really measure thousands of unique metabolites in a single analysis?

- No
- Why all the peaks then?
  - Metabolites may form multiple, different ion adducts in the MS ionization source, e.g. $[M+H]^+$, $[M+Na]^+$, $[M+K]^+$, $[M+NH_4]^+$, etc.
  - Molecules may fragment during the ionization process to yield additional product ions
  - dimers, trimers, etc. may form in the MS ionization source
  - many contaminants from both solvents and consumables are measured
  - some data processing algorithms do not “de-isotope” the data (e.g. ^{13}C isotopologue peaks)
  - noise
- Data may be “cleaned” by evaluating correlations among co-eluting peaks and selecting the dominant ion (e.g. $[M+H]^+$)
- However, a multiplicity of ions can sometimes be helpful for ID

Summary Example 2

- It's generally cost prohibitive to analyze replicates of biological samples in large studies
- Periodic analysis of pooled samples enables both standardization of data between batches and evaluation of measurement reproducibility for all signals
- Daily monitoring of QC data is essential for early detection of problems
- See posters P-349 example of application to a 7000+ sample study and poster P-318 for details on the processing workflow

Example 3:
Use of test materials

Dan Bearden

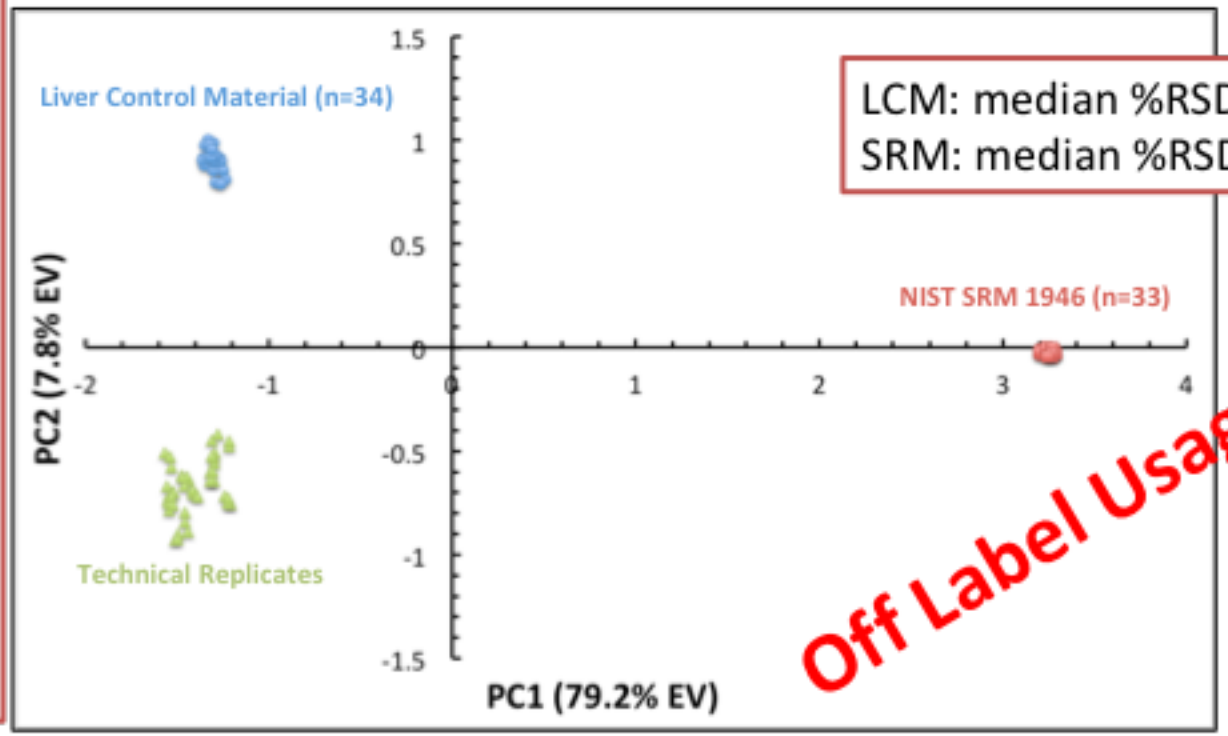
Examples of use of test materials

Liver QC sample PCA score plot
700 MHz NMR, Bligh-Dyer extractions

Standard Reference Material® 1946

Lake Superior Fish Tissue

This Standard Reference Material (SRM) is a frozen fish tissue homogenate that was prepared from adult lake trout fillets (Salvelinus namaycush namaycush) collected near the Apostle Islands in Lake Superior (U.S./Canada), and is intended primarily for use in evaluating analytical methods for the determination of … (PCB) congeners, chlorinated pesticides, … (PBDE) congeners, … (PFOS), fatty acids (including omega-3 fatty acids), extractable fat, methylmercury, total mercury, proximates, α-HBCD, and selected trace elements in fish tissue and similar matrices. All of the constituents … are naturally present in the fish tissue homogenate.



LCM: median %RSD = 7.1%
SRM: median %RSD = 8.9%

Off Label Usage!!

Casu F, Watson AM, Yost J, Leffler JW, Gaylord GT, Barrows FT, Sandifer PA, Denson MR, Bearden DW. Metabolomics Analysis of Effects of Commercial Soy-based Protein Products in Red Drum (*Sciaenops ocellatus*). *Journal of Proteome Research* 2017; 16:2481-2494.

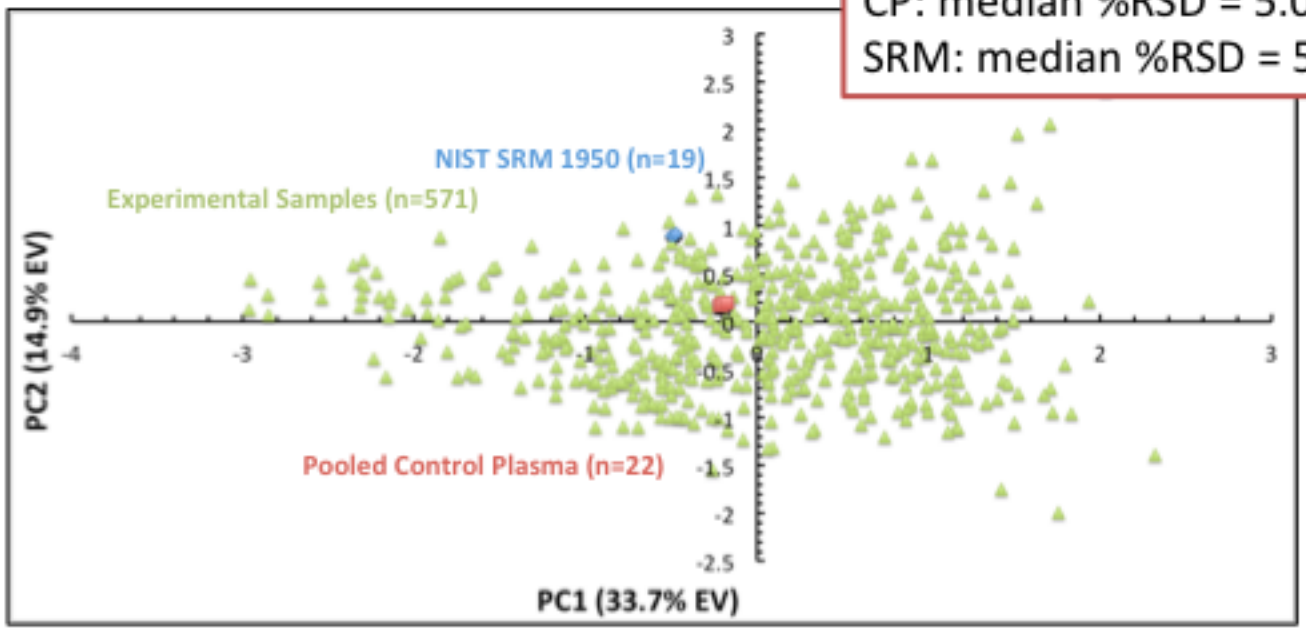
Examples of use of test materials

Plasma QC sample PCA score plot
700 MHz NMR, Filtered Plasma

Standard Reference Material® 1950

Metabolites in Human Plasma

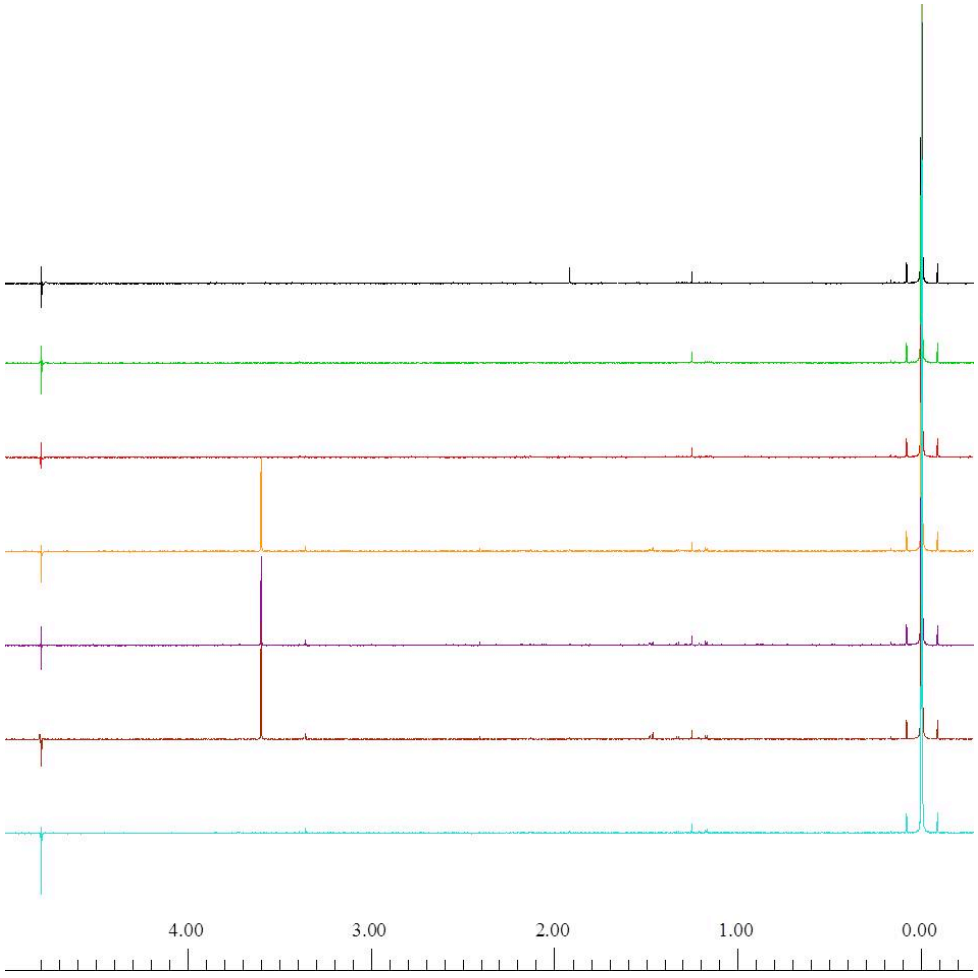
This Standard Reference Material (SRM) is intended primarily for validation of methods for determining metabolites such as fatty acids, electrolytes, vitamins, hormones, and amino acids in human plasma and similar materials. This SRM can also be used for comparison of measurement technologies used in metabolomic studies and for quality assurance when assigning values to in-house reference materials. The SRM is intended to represent normal human plasma.



CP: median %RSD = 5.0%
SRM: median %RSD = 5.1%

Casu F, Watson AM, Yost J, Leffler JW, Gaylord GT, Barrows FT, Sandifer PA, Denson MR, Bearden DW. Metabolomics Analysis of Effects of Commercial Soy-based Protein Products in Red Drum (*Sciaenops ocellatus*). *Journal of Proteome Research* 2017; 16:2481-2494.

Purposeful control of systematic errors: Blank sample analysis

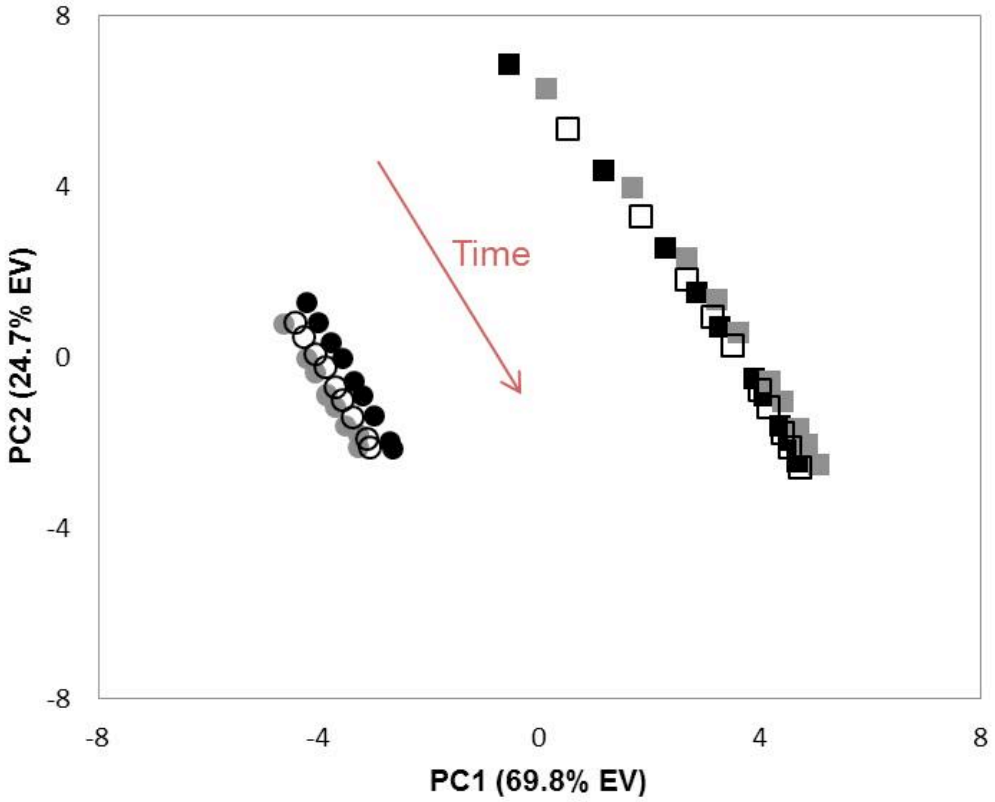


- ^{1}H NMR spectra of six extraction blanks (top six) and the NMR buffer (bottom)
- The arrow indicates a contaminant from one brand of bead beating tubes
- This peak was excluded from statistical analyses

Schock TB, Duke J, Goodson A, Weldon D, Brunson J, Leffler J, Bearden DW. Evaluation of Pacific White Shrimp (Litopenaeus vannamei) Health During a Superintensive Aquaculture Growout using NMR-based Metabolomics. *PLoS ONE* 2013; 8:e59521.

Method development – Sample stability

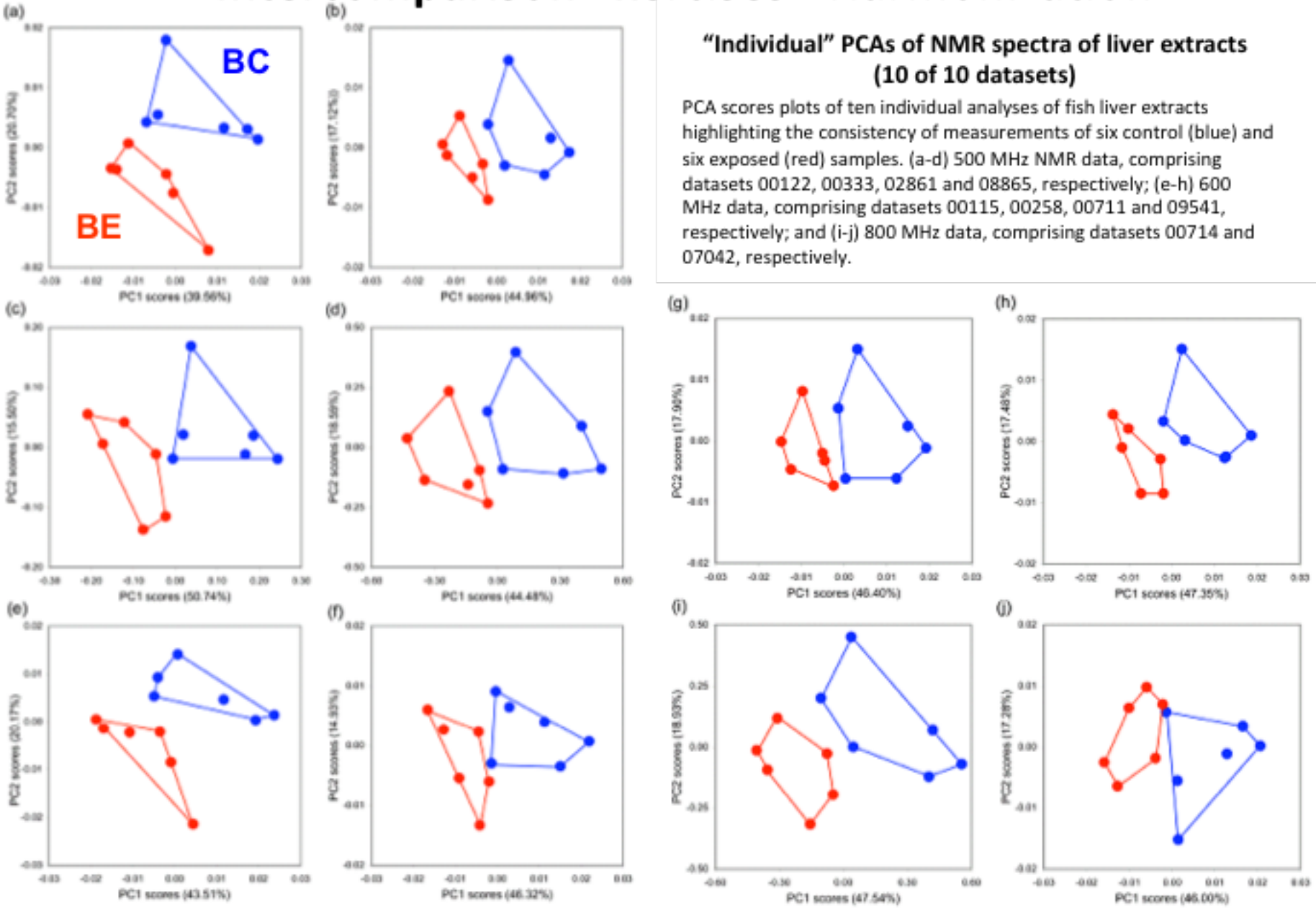
Repeated NMR analysis of extracted sample



- PCA scores plot of worm control material extracted by a chloroform/methanol/water protocol (circles) and a D_2O buffer method (squares) in triplicate (1: black, 2: gray, 3: open)
- Samples were repeatedly analyzed over a four day period
- The arrow shows the direction of metabolic changes with time from unstable samples

Schock T, Strickland S, Steele EJ, Bearden DW. Effects of Heat-treatment on the Stability and Composition of Metabolomic Extracts from the Earthworm *Eisenia fetida. Metabolomics* 2016; 12:1-7.

Intercomparison Exercises - Harmonization

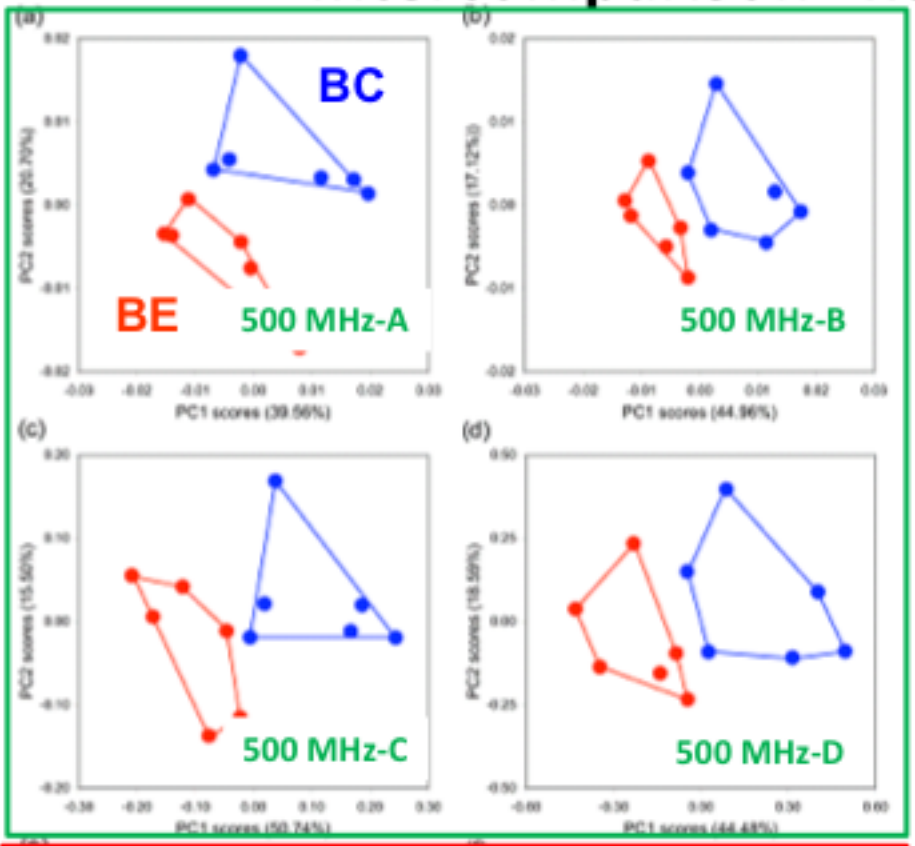


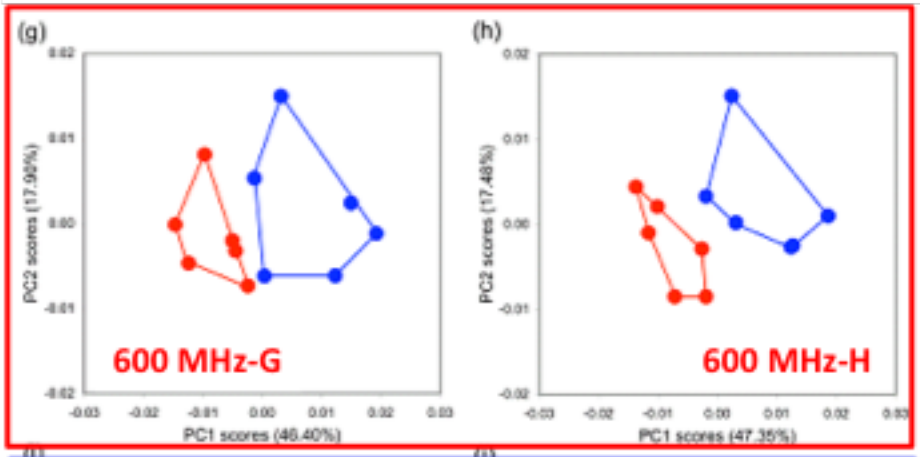
"Individual" PCAs of NMR spectra of liver extracts (10 of 10 datasets)

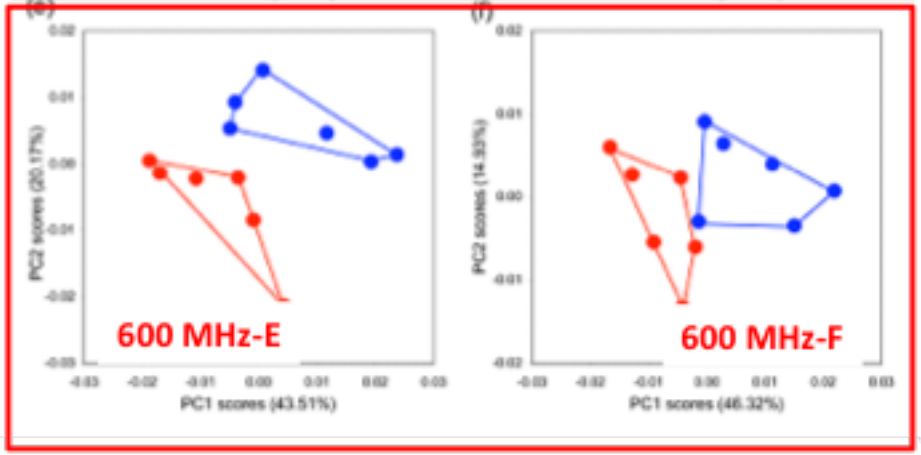
PCA scores plots of ten individual analyses of fish liver extracts highlighting the consistency of measurements of six control (blue) and six exposed (red) samples. (a-d) 500 MHz NMR data, comprising datasets 00122, 00333, 02861 and 08865, respectively; (e-h) 600 MHz data, comprising datasets 00115, 00258, 00711 and 09541, respectively; and (i-j) 800 MHz data, comprising datasets 00714 and 07042, respectively.

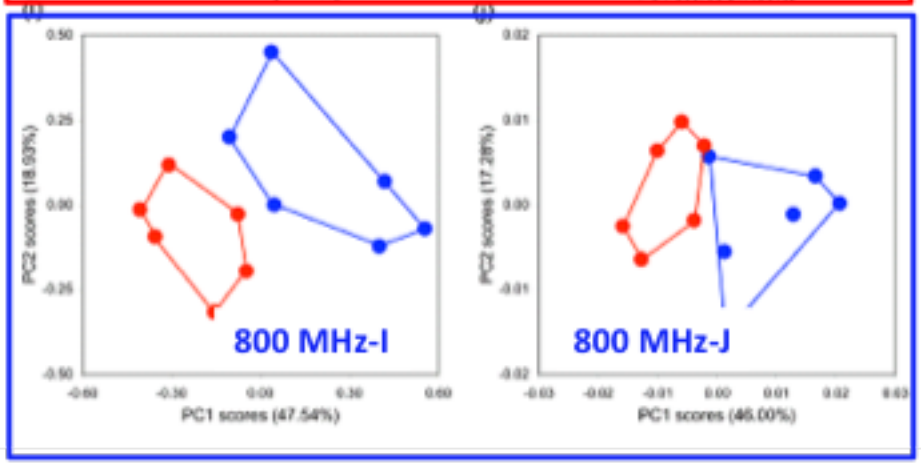
Viant MR et al. International NMR-Based Environmental Metabolomics Intercomparison Exercise. *Environmental Science & Technology* 2009; 43:219-225.

Intercomparison Exercises - Harmonization







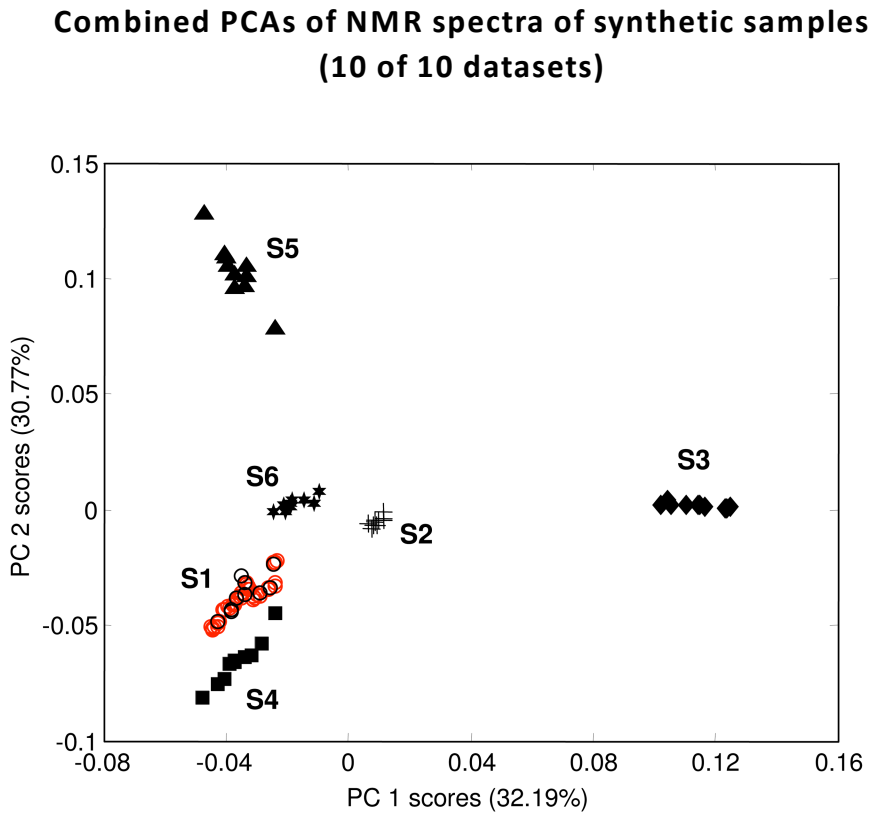


"Individual" PCAs of NMR spectra of liver extracts (10 of 10 datasets)

PCA scores plots of ten individual analyses of fish liver extracts highlighting the consistency of measurements of six control (blue) and six exposed (red) samples. (a-d) 500 MHz NMR data, comprising datasets 00122, 00333, 02861 and 08865, respectively; (e-h) 600 MHz data, comprising datasets 00115, 00258, 00711 and 09541, respectively; and (i-j) 800 MHz data, comprising datasets 00714 and 07042, respectively.

Viant MR et al. International NMR-Based Environmental Metabolomics Intercomparison Exercise. *Environmental Science & Technology* 2009; 43:219-225.

Intercomparison Exercises - Harmonization



- PCA scores plot of the combined 500, 600 and 800 MHz NMR spectra of synthetic mixtures from all 10 analyses by the participating laboratories
- The initial PCA model was built using the unique samples (one replicate of S1 and S2-S6, each n=10, giving a total of 60 spectra)
- The additional 46 technical replicates of S1 (S1b-S1f for eight of the datasets, and only S1d-S1f for datasets 00115 and 00122) were then applied to this model, with symbols in red. The overlap of the red and black circles verifies the high consistency of the NMR measurements.

Viant MR et al. International NMR-Based Environmental Metabolomics Intercomparison Exercise. *Environmental Science & Technology* 2009; 43:219-225.

Summary Example 3

- Test materials, such as NIST Standard Reference Materials (SRM), provide a means to evaluate reproducibility of method within a study, over time, and among different labs

- Analyses of blanks can reveal contaminants that may interfere with metabolite measurements

Concluding remarks

- Untargeted metabolomics presents unique challenges for QC
  - Methods measure both knowns and unknowns
  - Internal standards are not immediately available for unknowns (by definition)
  - A single metabolite can give rise to a number of redundant features
  - It is often difficult to distinguish contaminants from actual metabolites

- Untargeted metabolomics QC procedures are often customized for specific analytical techniques and experimental designs, but there are common elements:
  - Randomization of sample analysis order to avoid systematic bias
  - Internal standards for real time and post-acquisition QC
  - Pooled reference standards, also for for real time and post-acquisition QC
  - Inclusion of reference samples, such as NIST SRM