A COMPREHENSIVE STRATEGY FOR BUILDING AND EVALUATING PLASMA PROTEOMICS DATA-DEPENDENT ANALYSIS DERIVED SPECTRAL LIBRARIES WITH ZenoTOF 7600

Jimmy Yi Zeng, Hao Qian, Ruby Karimjee, Yuntao Hu, Mark Maruspini, Jessica Chan, Megan Mora , Benjamin Ta, Ehdieh Khaledian , Chi-Hung Lin, Robert Zawada, Joon-Yong Lee, Purva Ranjan, Chinmay Belthangady, Philip Ma, Bruce Wilcox
PrognomiQ, Inc., San Mateo, CA

INTRODUCTION

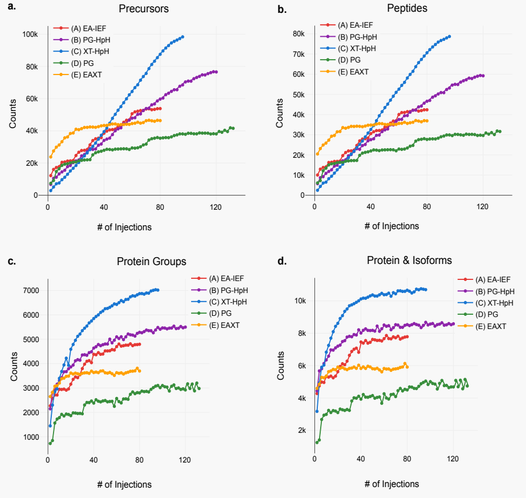
- Deep, unbiased proteomic analysis has been made possible by recent advances in sample preparation (ie, Seer's Proteograph™ Product Suite) and improved mass spectrometry instrument sensitivity and speed (ie, Sciex ZenoTOF 7600)
- These technologies enable the quantification of thousands of proteins from human plasma at the necessary throughput and reproducibility for large-scale biomarker studies
- Detection of thousands of proteins in plasma with data independent acquisition/sequential window acquisition of all theoretical mass spectra (DIA/SWATH) is predicated upon a high-quality spectral library, which provides sufficient depth and breadth to leverage the latest sample preparation techniques
- Here, we compared multiple comprehensive and robust strategies to build a human plasma spectral library that contains over 10,000 proteins and 110,000 precursors with Spectronaut 16.2

OBJECTIVE

- To investigate and compare the efficiency of multiple sample preparation methods for constructing an information-dependant acquisition (IDA) human plasma library
- To build a comprehensive human plasma IDA spectral library for deep and high-throughput biomarker discovery studies

RESULTS

FIGURE 2. Comparison of strategies for spectral library creation.



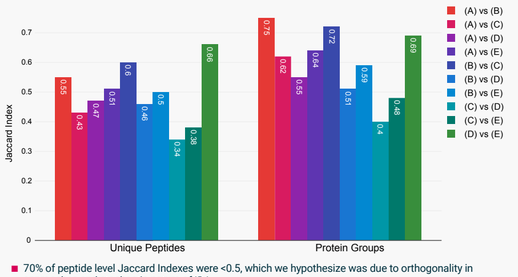
- The (D) PG strategy was the least efficient approach while the (C) XT-HpH strategy provided the largest library within 100 injections

TABLE 2. Evaluation of spectral library quality.

Samples Sets	(A) EA-IEF	(B) PG-HpH	(C) XT-HpH	(D) PG	(E) EAXT
Precursors	54,038	76,811	98,157	41,314	46,438
Precursors filtered to >6 transitions	49,546	71,283	92,191	39,511	43,549
Protein groups	4,737	5,473	7,015	2,919	3,697
Proteins filtered to minimum 2 unique proteins	3,779	4,586	5,846	2,566	3,037

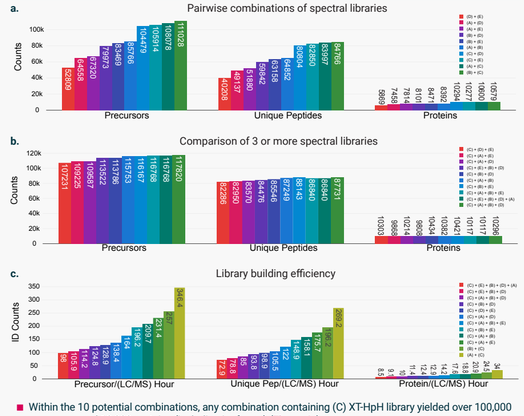
- All libraries maintained >90% of precursors after a 6 transitions threshold was applied, indicating high quality MS2 spectra
- Only the (A) EA-IEF spectral library maintained <80% of protein groups when filtered to remove 1-hit wonders

FIGURE 3. Pairwise Jaccard Index comparison of spectral libraries.



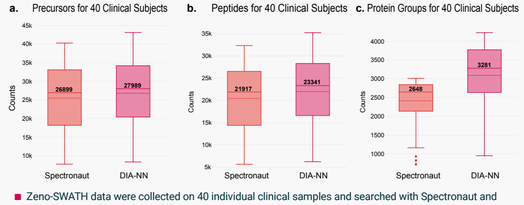
- 70% of peptide level Jaccard Indexes were <0.5, which we hypothesize was due to orthogonality in approaches and stochastic nature of IDA
- The highest Jaccard Index of 0.66 (D vs E) was likely due to utilizing the same biological samples

FIGURE 4. XT-HpH was the most efficient approach to building deep plasma based spectral libraries.



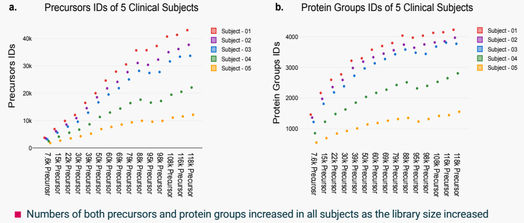
- Within the 10 potential combinations, any combination containing (C) XT-HpH library yielded over 100,000 precursors and 10,000 proteins (including isoforms) (Figure 4a)
- Higher numbers of combinations gave incremental improvement compared to trio combinations (Figure 4b)
  - ≤6% improvement in library size was observed when combining 3 or more libraries versus the pairwise combination of (B) PG-HpH + (C) XT-HpH
- Within the 10 datasets (ranked by precursor counts per LC/MS hour), it was noticeable that a combination of (A)+(C) was ~4x more efficient in spectral library generation than (A)+(B)+(C)+(D)+(E), but similar in library size (Figure 4b and 4c).

FIGURE 5. Application of maximum spectral library to 40 clinical samples.



- Zeno-SWATH data were collected on 40 individual clinical samples and searched with Spectronaut and DIA-neural network
- Similar precursor and peptide results were observed between search engines, but a significant difference at the protein group level was observed

FIGURE 6. Impact of spectral library size on Zeno-SWATH data.

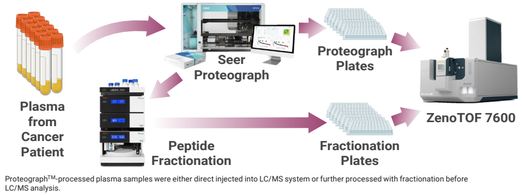


- Numbers of both precursors and protein groups increased in all subjects as the library size increased

METHODS

- **Sample preparation:**
  - A total of 260 cancer, comorbid, and healthy subjects' plasma samples were processed on the Seer Proteograph™ platform to generate digested peptide samples plates (Figure 1)
  - Four versions of Seer's kit (Early Access of Proteograph Assay, Proteograph™ workflow with Proteograph Assay, Early Access of Proteograph XT Assay, and Proteograph™ XT workflow with Proteograph XT Assay) were used to process plasma samples
  - Fractions of the Proteograph-processed samples were further pooled by nanoparticles (NP) and fractionated with either high pH reverse phase (HpH) or iso-electric focusing (IEF) fractionation methods (Table 1)
- **Liquid chromatography/mass spectrometry (LC/MS) and data analysis:**
  - All fractions and individual subject samples were reconstituted with reconstitution buffer (95% H_2O, 5% acetonitrile, 0.1% formic acid, and 0.125X iRT) and subjected to a 2-hour LC/MS IDA collection on a Waters M-class ZenoTOF 7600 (Figure 1)
  - Spectronaut (v16.2) was used to build a reliable IDA spectral library
  - Spectronaut (v16.2) and DIANN (v1.8.1) were used for SWATH data analysis

FIGURE 1. Workflow for spectral library data generation.



Proteograph™-processed plasma samples were either direct injected into LC/MS system or further processed with fractionation before LC/MS analysis.

TABLE 1. Five sample preparation strategies for building human plasma spectral libraries were compared.

Samples Sets	Seer EA[1] Kit + IEF fraction	Seer PG[2] Kit + HpH fraction	Seer EAXT[3] Kit + HpH fraction	Seer PG Kit Direct Injection	Seer EAXT[4] Kit Direct Injection
# of subjects	84	48	40	48	40
# of fractions	40	120	96	N/A[5]	N/A[5]
Hours of LC/MS	80	240	192	480	80
Abbreviation	(A) EA-IEF	(B) PG-HpH	(C) XT-HpH	(D) PG	(E) EAXT

[1]EA: Early Access of Proteograph Assay.
[2]PG: Proteograph™ workflow with Proteograph Assay.
[3]EAXT: Early Access of Proteograph XT Assay.
[4]XT: Proteograph™ XT workflow with Proteograph XT Assay.
[5]For D & E, no fractions were done. All samples generated from Proteograph™ were directly injected for LC/MS analysis.

CONCLUSIONS

- (C) XT-HpH was the most efficient strategy to generate a spectral library compared to the other 4 strategies
- Pairwise combination including (C) XT-HpH yielded >10,000 proteins and >100,000 precursors
- Combination of (A)+(C) had the greatest spectral library generation efficiency
- With the current ~118,000 precursor library, SWATH analysis of clinical subjects gave >3,000 protein groups and >28,000 precursors

DISCLOSURES
Study funded by PrognomiQ, Inc. All authors are current or former employees of PrognomiQ, Inc.

ACKNOWLEDGEMENTS
Funded by PrognomiQ, Inc (San Mateo, CA). Editorial and graphical assistance provided by Prescott Medical Communications Group (Chicago, IL).