A Highly Scaled Proteomic Discovery Study for Prostate Cancer Diagnostic Signatures Using Proteograph™ Workflow with Trapped Ion Mobility Mass Spectrometry

Flory M[1], Chang M[1], Cartier J[1], Lange J[1], Kopp R[1], Liss M[2], Leach R[2], Albracht B[2], Dai J[3], Krawitzky M[4], Elgierari E[5], Chu J[5], Cuevas J[5], Pease P[5], Mahoney M[5], Ferdosi S[5], Benz R[5], Siddiqui A[5], Zamanighomi M[5], Guturu H[5], Hornburg D[5], and Batzoglou S[5]

[1]Oregon Health and Science University, Knight Cancer Institute, Cancer Early Detection Advanced Research Center, Portland, OR USA
[2]University of Texas Health Science Center at San Antonio Health, San Antonio, TX USA, [3]Fred Hutchinson Cancer Research Center, Seattle, WA USA
[4]Bruker Daltonics, Billerica, MA USA, [5]Seer Inc., Redwood City, CA USA

Prostate Cancer Diagnostic Goal

- **PROBLEM:** Current standard-of-care prostate cancer diagnosis using PSA blood testing exhibits low specificity for cancer and leads to unwarranted prostate biopsy
- **GOAL:** Our study seeks to discover new proteomic PSA reflex markers in blood with increased specificity for total and high-grade prostate cancer
- **APPROACH:** We leverage a novel multi nanoparticle-based, deep and scalable proteomics platform to investigate ~900 patient serum samples and controls

The Technology

Proteograph™ Product Suite

Automated multi nanoparticle (NPs)-based protein enrichment and processing to purified, quantitated, reconstituted peptides

Protein corona formation upon incubation of nanoparticles with fluid biopsy input

Peptides

Proteograph™ Workflow
- 250 ul serum input and 5 NPs panel
- Proteograph SP100 automation system
- Multiple Proteograph workflow controls
- 16 samples/8 hours/instrument

BRUKER timsTOF Pro

Mass Spectrometry
- dia-PASEF, primary acquisition method
- DDA PASEF, for spectral library generation

Ideal Cohort and Rigorous Design

Total of >900 Clinical Patient Serum Specimens

~70% for Training Set | **~30% for Test Set**

- **Recruitment:** Elevated PSA and/or abnormal digital rectal exam
- **Sourcing:** SABOR Repository and prior study cohort
- **Clinical Course:** All men biopsied with deep clinical data; classes:
  - ➢ No Cancer Detected
  - ➢ Lower-Grade Cancer Detected
  - ➢ Higher-Grade Cancer Detected
- **Specimen Collection:** All specimens collected pre-biopsy and before any treatment, unified and consistent EDRN protocol

UT Health San Antonio

The PRoBE Biomarker Discovery Study Framework

Prospective-Specimen-Collection, Retrospective-Blinded-Evaluation

- **Sufficient, ideal specimen set**
- **Specimen set division into training and test subsets**
- **Blinding, randomization**
- **Systematic evaluation of marker performance**

References:
- Prostate cancer diagnostics: Litwin and Tan, 2017, *JAMA* **317**:2532.
- PRoBE framework: Pepe and Feng, 2011, *Clin. Chem.* **57**:1093
- Seer Proteograph: Blume *et al.,* 2020, *Nat. Comm.* **11**:3662.
- Bruker timsTOF, dia-PASEF: Meier *et al.,* 2020, *Nat. Meth.* **17**:1229.
- DIA-NN: Demichev *et al.* 2020, *Nat. Meth.* **17**:41.
- MSFragger: Kong *et al.* 2017, *Nat. Meth.* **14**:513.

Funding:
- Knight/CEDAR Project Proposal Award
- M. J. Murdock Charitable Trust Grant

Analytics Overview

Database Searching → **Candidate Discovery** → **Signature Evaluation**

- Employ Proteograph™ Analysis Suite (PAS), DIA-NN module
- Data inspection including spectral library strategy assessment
- Cross-validate potential tools including classifier
- Probe for markers with increased disease specificity at high sensitivity
- Leverage reserved/untouched data from test set
- Implement chosen analytic approach to validate signature performances

Spectral Library Assessment

Description of Various Approaches

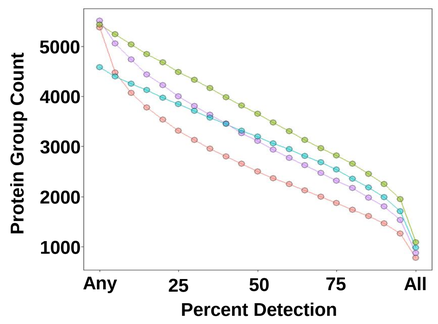
"In-Silico Library Filtered by Peptide or Protein" — Spectra culled from DIA-NN library-free search of entire Training Set

"Offline Fractionation" — Subset Training Set serum samples chosen systematically (n=75) → Offline HpH-RP peptide fractionation by NP → 240 LC-MS DDA PASEF mode injections

"Long Gradient DDA" — Subset Training Set serum samples chosen systematically (n=75) → Supplemental Proteograph Workflow → 25 long-gradient LC-MS DDA PASEF mode injections

Spectral Library	Protein Groups	Peptides	Fragment Ions
Fractionation	6258	69079	1867008
Long DDA	2365	21251	728942
All Combined	6354	74158	2082962

Training Set Proteomic Depth Across Database Search Strategies



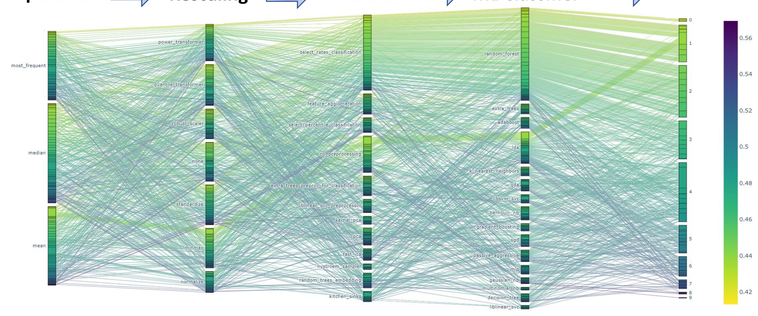
Fractional detection of proteins in training set using different search strategies. All dia-PASEF (DIA) database searching done using DIA-NN in Seer PAS. DDA PASEF database searching for spectral library generation from deep fraction analysis done using MSFragger. *In-silico* spectral libraries culled from DIA-NN training set library-free database searches.

- *In-silico* spectral library using spectra filtered according to peptides identified in this study
- *In-silico* spectral library using spectra filtered according to proteins identified in this study
- Traditional spectral library generated using DDA PASEF analysis of fractionated sample subset
- Library-free analysis for comparison (not using any spectral library)

ML Assessment for Signature Detection

INPUT: ✓ NP-protein (or peptide) features + intensities ✓ Specimen metadata ✓ Time allotment

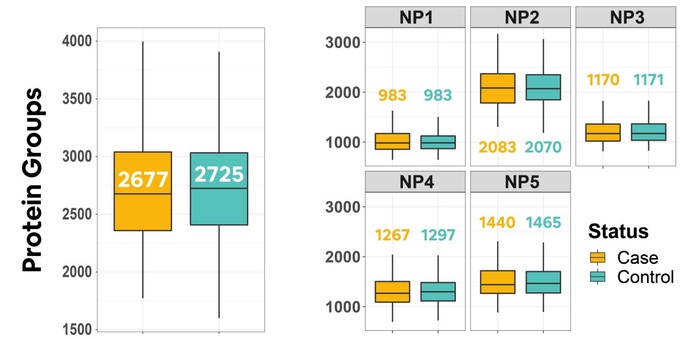
Imputation → **Rescaling** → **Features** → **ML Classifier** → **Cost**



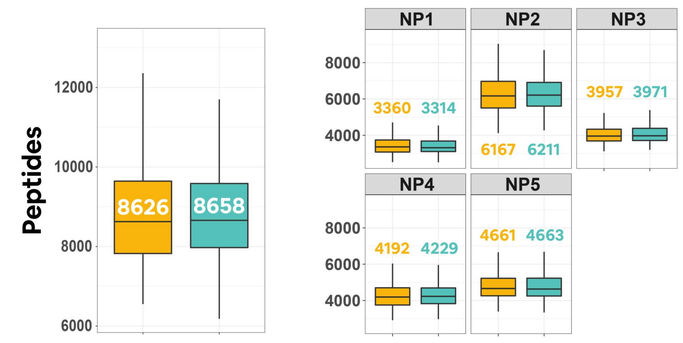
Automated assessment of tools for multiple data analytic steps including machine learning classifier selection. Optimal paths exhibit smaller "cost" value and bias toward yellow end of color spectrum.

Training Set Data

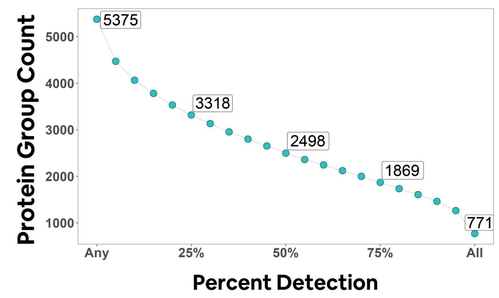
Protein-Level Coverage Across Study Arms and Nanoparticles



Peptide-Level Coverage Across Study Arms and Nanoparticles



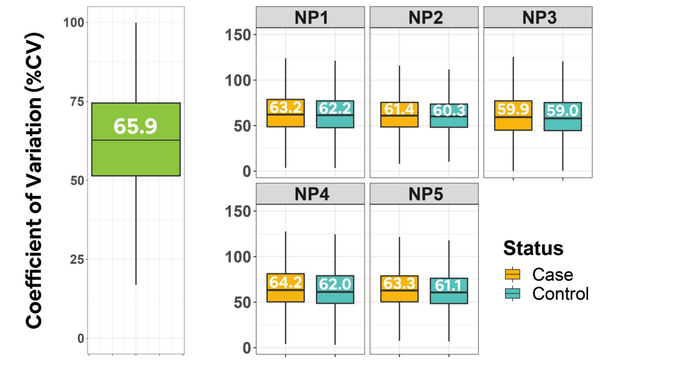
Percent Protein Detection Across Training Set



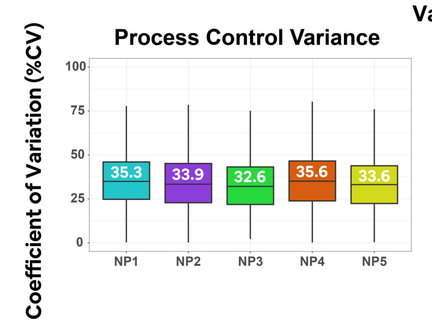
Aggregate count of detected protein groups in fraction of samples across entire set. All dia-PASEF (DIA) database searching done in Library-free mode using DIA-NN in Seer PAS.

Measurement Consistency and Variance Sources

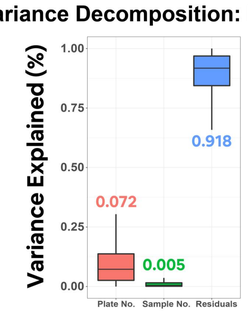
Coefficient of Variation for Protein Measurements



Process Control Variance



Variance Decomposition: NP2



Work in Progress: Analytics with spectral library(ies) implemented for database searching, peptide-level feature analysis, contextualization with rich available clinical metadata including PSA assay measurements. Validation to be performed thereafter.