Large scale, deep and unbiased plasma proteomics profiling a sub-study of a multi-cancer cohort enabling biomarker discovery

[1]Bruce Wilcox, [1]Kavya Swaminathan, [1]Preston Williams, [1]Jared Deyarmin, [1]Brian Young, [1]Chinmay Belthangady, [1]Manway Liu, [1]Mi Yang, [2]Xiaoyan Zhao, [2]Xiaoju Zhang, [2]Theo Platt and [1]Philip Ma

20th Human Proteome Organization World Congress - HUPO ReConnect 2021

[1]PrognomiQ Inc.
[2] Seer Inc.

Deep proteome coverage is key to PrognomiQ's multi-omics approach

New technologies and sample availability address historical challenges to scale proteomics

Challenges

- Access to large well collected, annotated sample cohorts for specific clinical questions
- Technical challenges associated with plasma proteomics - reproducibility, throughput and depth of coverage have limited the translation to the clinic
- Reproducible measurement and integration of multi-omic datasets providing novel insights into cancer biology

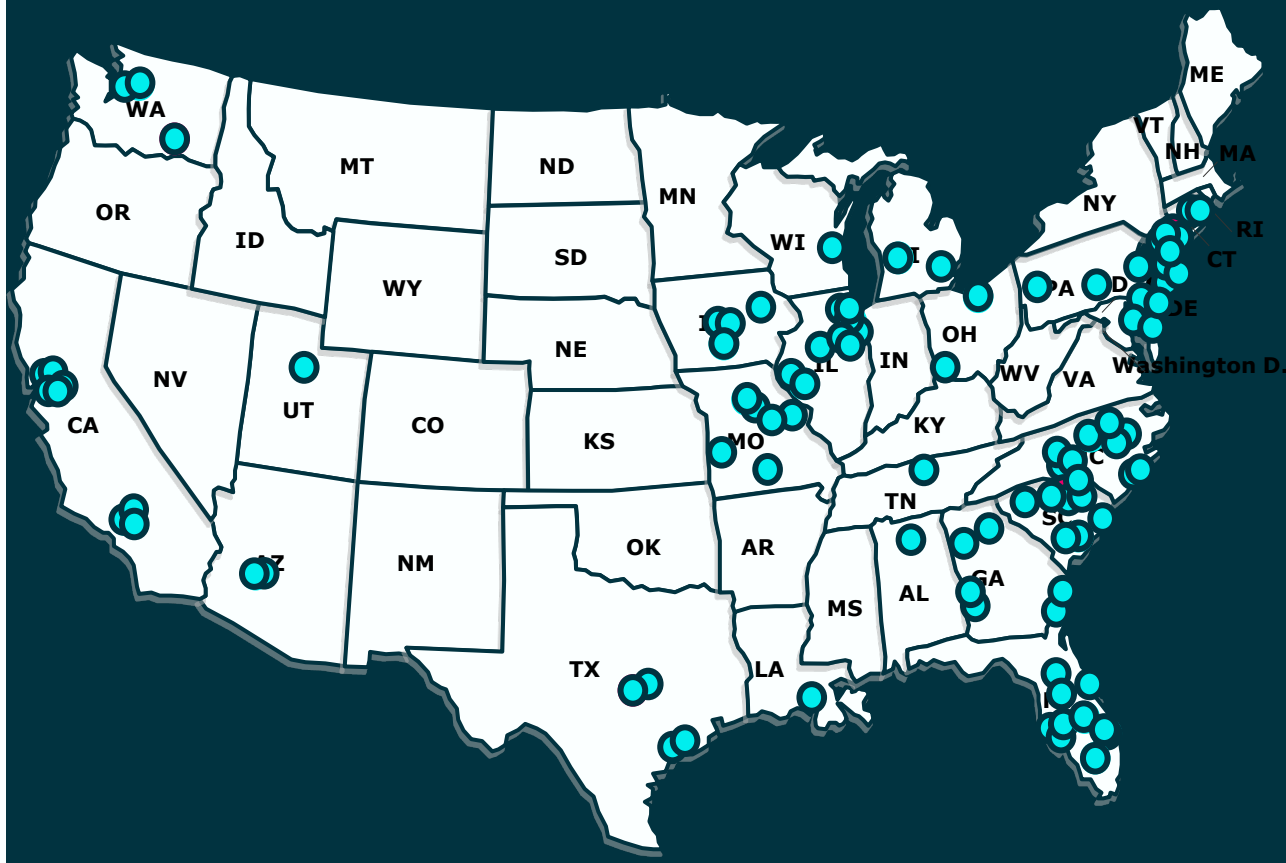
Multi-omics Strategy

- Well-defined disease biobank with multiple sample types optimized for the multi-omic measurements
- Development and optimization of novel proteomics technologies to increase proteome coverage and throughput without compromising reproducibility.
- Unbiased multi-omics platform deploying state-of-the-art instrumentation and advanced machine learning analysis to transform complex early disease detection

Proteomics feasibility study for a multi-cancer multi-omics study

- Our mission is to build better tools for physicians treating a multitude of cancers using multi-omics
- We have created a multi-cancer sample repository:
  - 1,000's of samples
  - >1,000 cancer subjects
  - Optimal sample type per 'omic
- Every 'omics technology in our platform was empirically selected through a series of feasibility studies
- We report here on a proteomics feasibility study of 212 samples performed using Proteograph™ Product Suite with a multi-nanoparticle (NPs) enrichment technology[1] and LC-IM-MS/MS analysis

[1]Blume et al. Nat. Comm. (2020)



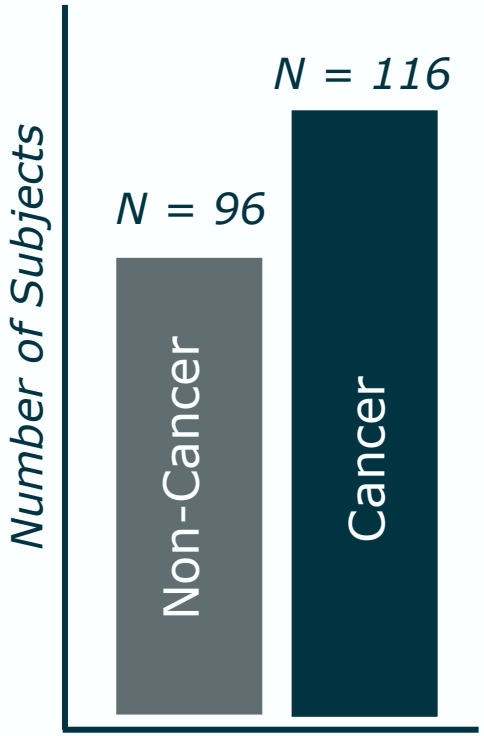
Enabling deep, untargeted, rapid proteomic biomarker studies

Evaluation of the Proteograph and timsTOF Pro to improve plasma proteome coverage

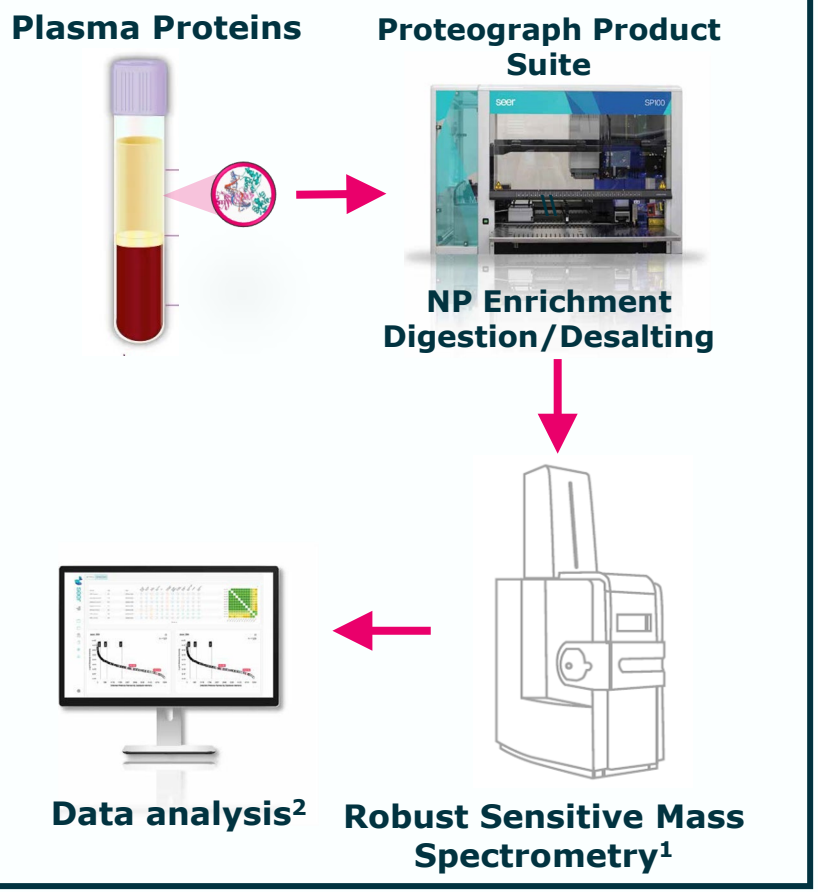
Integrative multi-omics

- Genomics
- Transcriptomics
- Proteomics
- Metabolomics

Study Subjects



Proteomics Platform

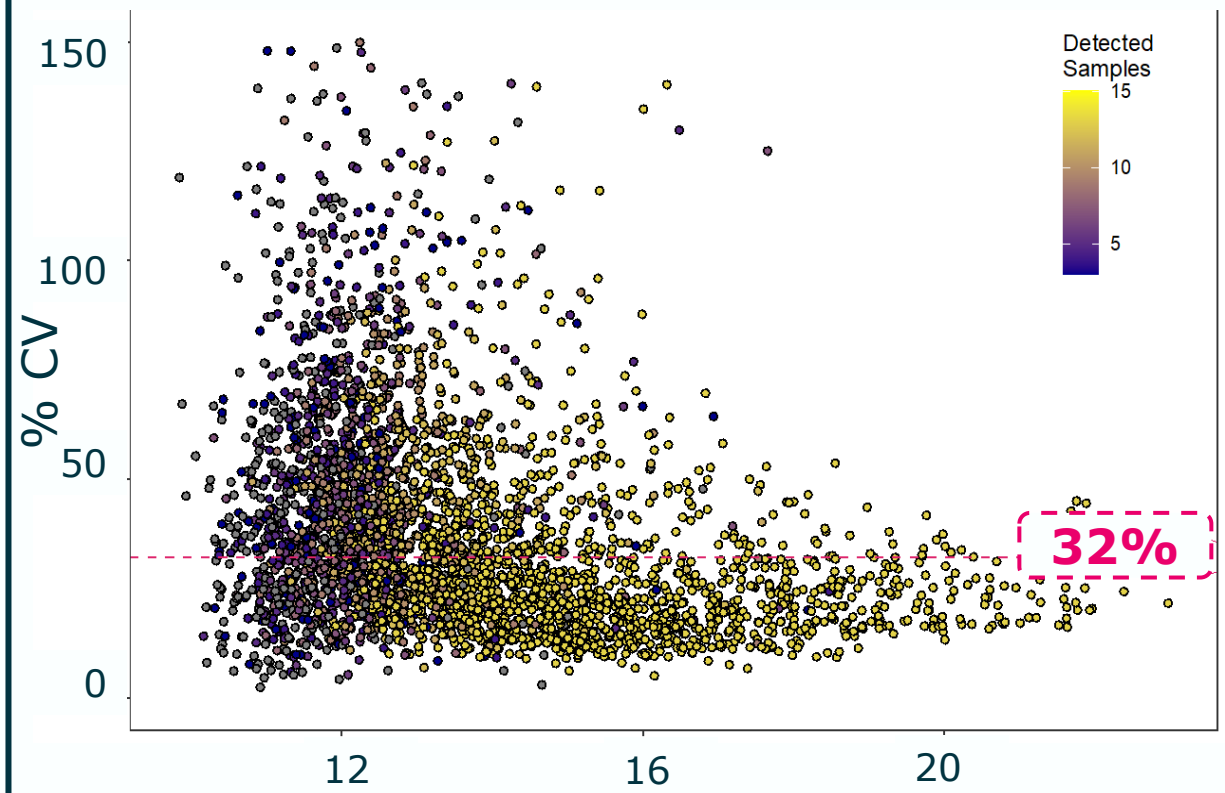


[1]60min DDA-PASEF runs
[2]MaxQuant search parameters: 0.1% peptide/protein FDR search, default timsTOF parameters searched against complete UniProt SwissProt human proteome database with contaminants (50% reversed decoys)

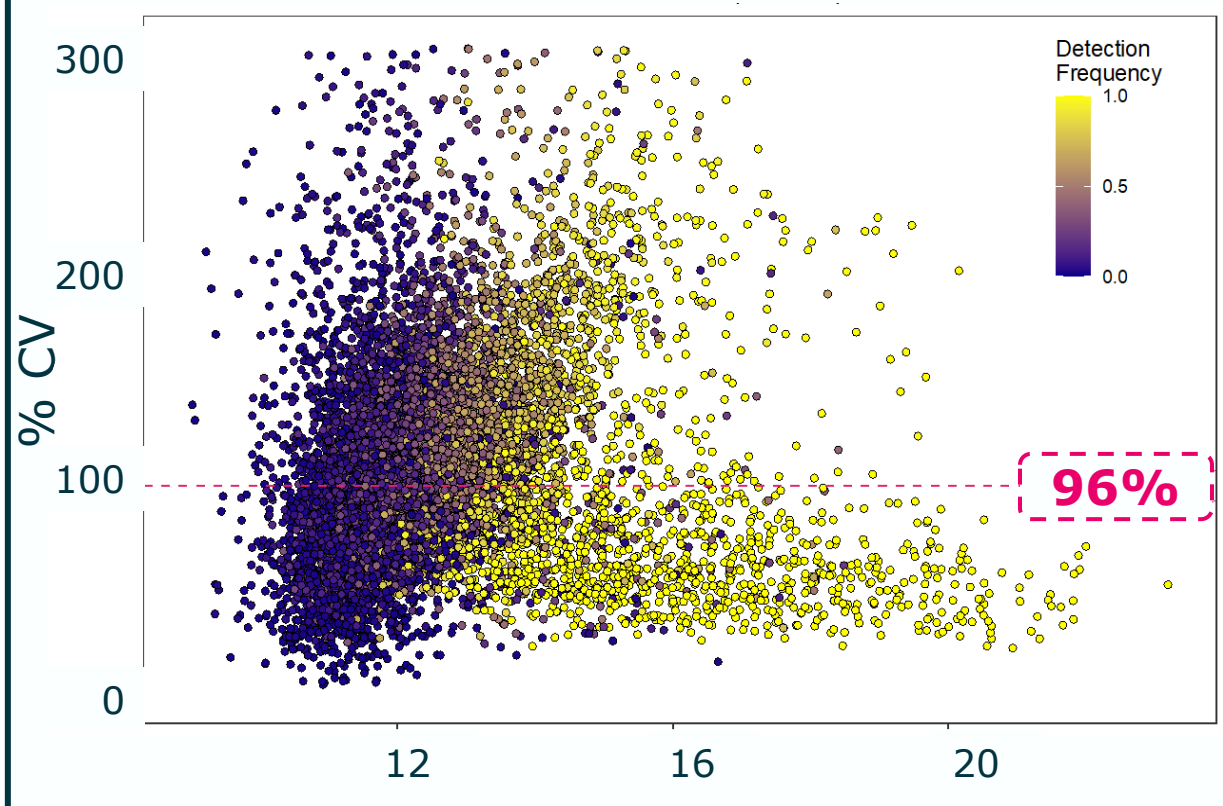
Technical variability is lower than detected biological variability

Median CV of 32% across all technical controls with biological variability of 96%

Precision of Plasma Control Protein Measurements as a Function of Intensity



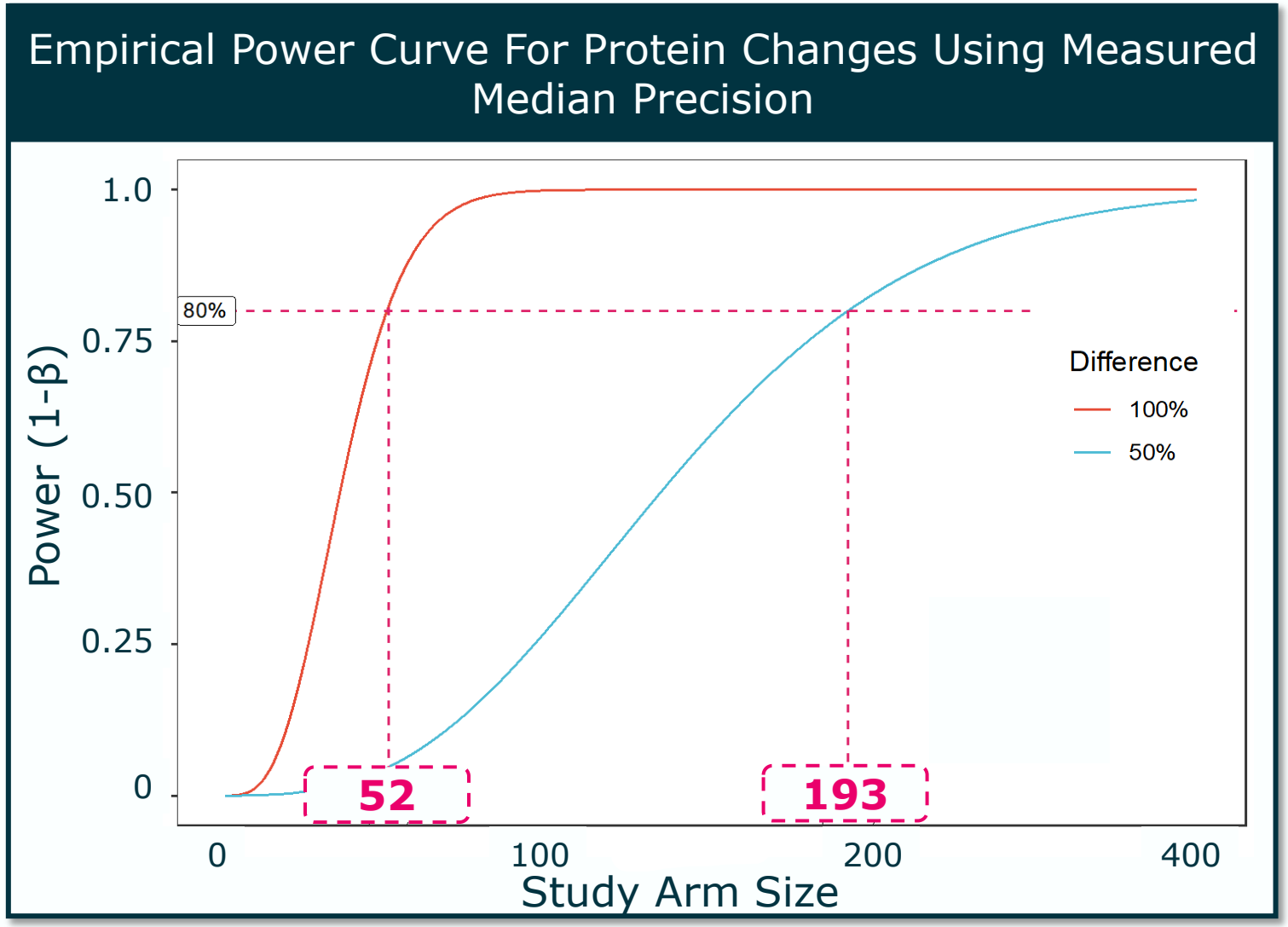
Precision of Subject Sample Protein Measurements as a Function of Intensity



[1]median normalization based on features common to all samples for a particle
[2]intensities are natural log transformed prior to normalization

Measurement reproducibility enables detection of expected fold changes

Providing desired statistical power for multi-cancer biomarker studies



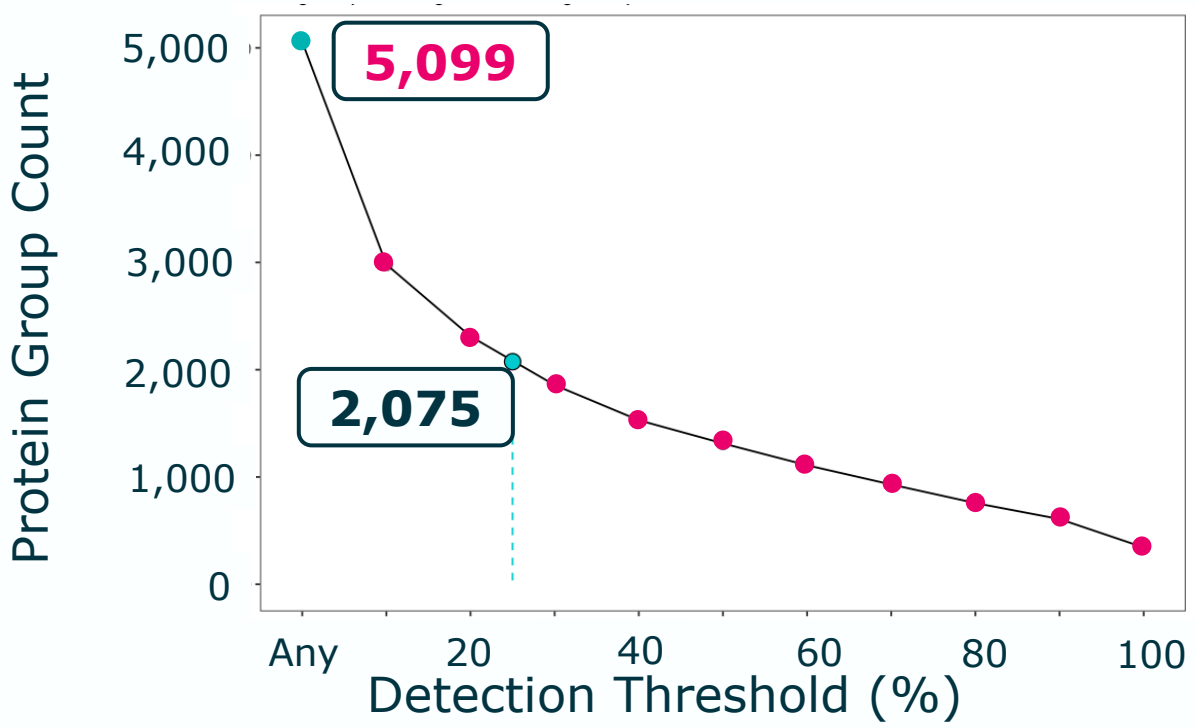
[1]Using median precision of 96% and Bonferroni correction assuming 2,000 proteins

Detection of >5,000 proteins in feasibility study of 212 subjects

A median of 4 peptides per protein detected for proteins present in >25% of the samples

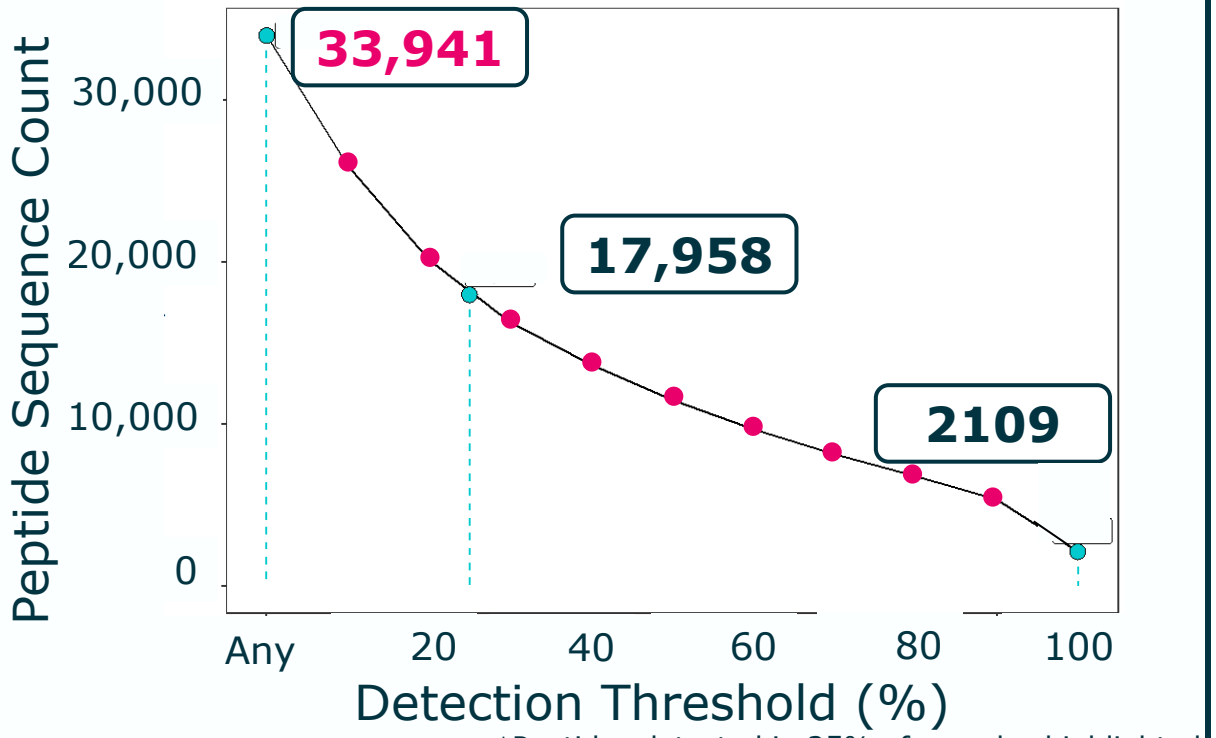
Detected Protein Groups

Percentage of Sample Which Protein is Detected



Detected Peptide Counts

Number of Peptides Detected across all Samples



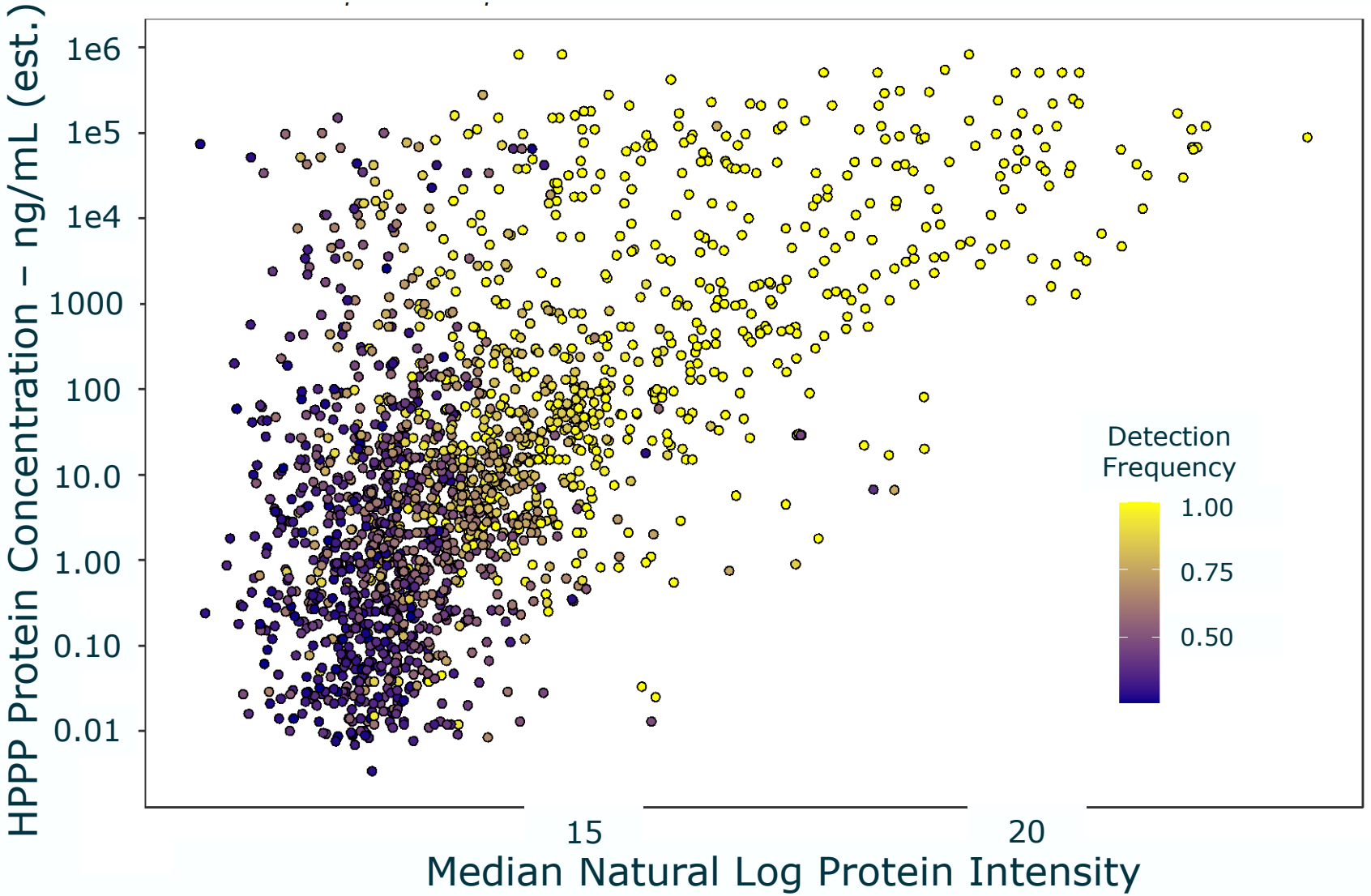
*Peptides detected in 25% of samples highlighted

[1]Keshishian, et. al, Molecular & Cellular Proteomics 2015 14:9, 2375-2393
[2]MaxQuant search parameters: 0.1% peptide/protein FDR search, default timsTOF parameters searched against complete UniProt SwissProt human proteome database with contaminants (50% reversed decoys)

Detected HPPP proteins cover 8 orders of magnitude in concentration

Increased depth of coverage is highlighted by compression of proteome dynamic range

- ~40% of 3,486 HPPP[1] proteins with estimated plasma concentrations were detected at 25% threshold
- NP-based enrichment compresses effective protein concentration and provides a rapid measurement of high and low concentration proteins
- Reproducible detection of low abundance proteins – 392 proteins with estimated concentration <10ng/mL detected in >50% of samples

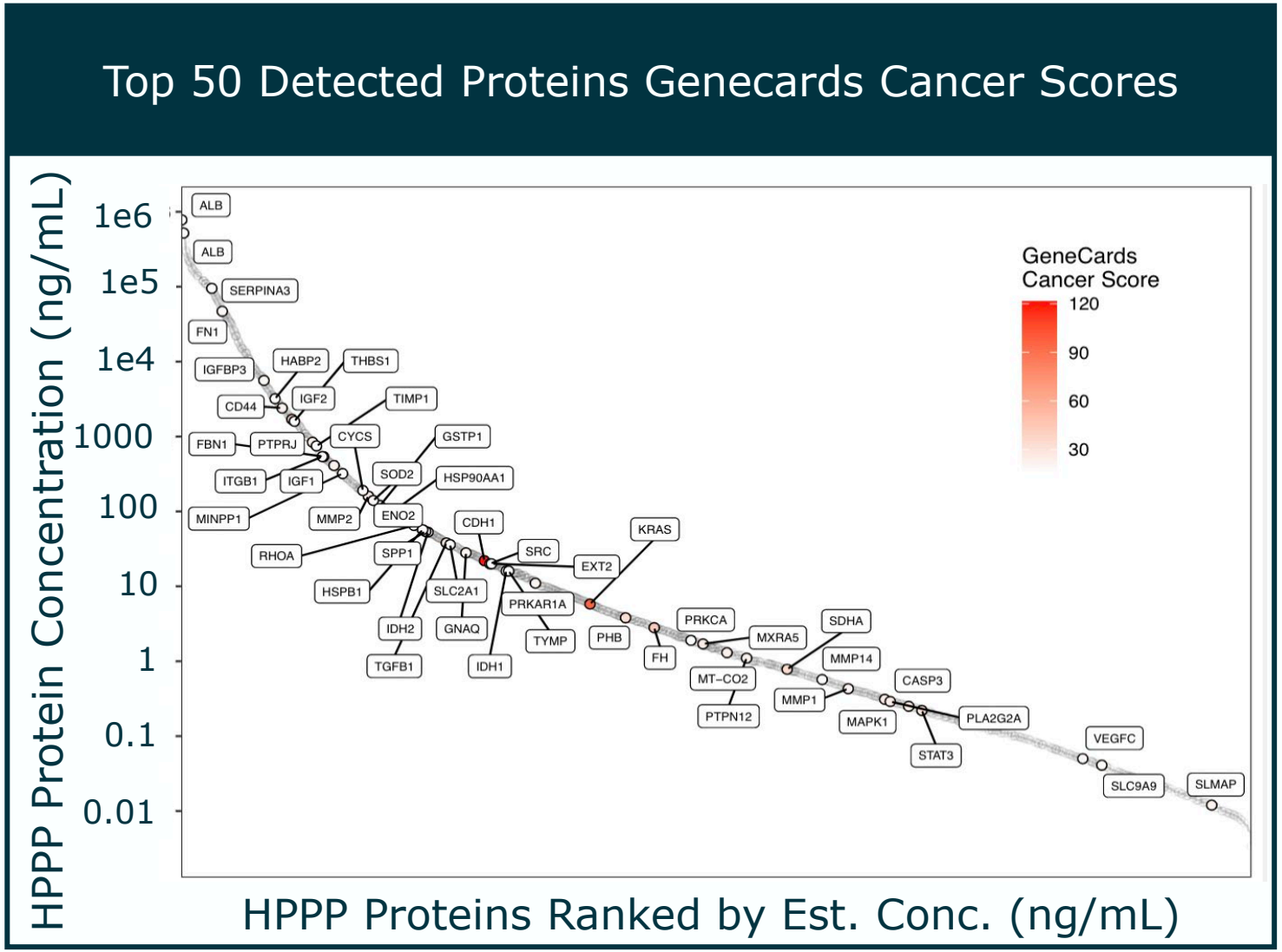


Maximum measured intensity and minimum reported concentrations for duplicates are plotted; 201 samples with full data are filtered to >=25%

1. Jochen M. et.al *Journal of Proteome Research* **2017** *16* (12), 4299-4310

Enhanced proteome coverage detects known cancer related proteins

- Measuring proteins across 8 orders of magnitude enabled detection of HPPP proteins with known correlations in Cancer
- ~40% of the top 50 detected Genecards[1] cancer proteins are known to have plasma concentrations of <10 ng/mL
- Pilot study indicates potential to detect novel cancer biomarkers that are likely to be low abundance functional proteins



All detected, matching proteins from samples plotted on HPPP curve;
GeneCards data uses score reported from matching gene id and search term 'cancer'

1- Stelzer G, et. al *The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analysis* , Current Protocols in Bioinformatics(2016), 54:1.30.1 - 1.30.33.doi: 10.1002 / cpbi.5

Conclusions and Acknowledgements

Conclusions

- The data from this feasibility study of 212 subjects, including 117 cancer subjects, demonstrate the promise of using Proteograph technology (Seer Inc.) for proteomics biomarker discovery studies.
- With Proteograph, excellent depth of coverage, reproducibility and direct detection of expected cancer relevant proteins across 8 orders of magnitude in concentration were achieved, providing a well-defined use case for large-scale discovery studies.
- Nanoparticle enrichment of proteins from plasma provides similar depth of coverage (5,099 proteins) as best of class depletion and fractionation strategies, but at a much higher throughput.
- The combination of Proteograph and Bruker timsTOF Pro dia-PASEF technology has provided a robust, sensitive and high-throughput proteomics platform to support large scale untargeted proteomics biomarker discovery studies.

Acknowledgements

1. John Blume for significant contributions in data analysis and scientific discussions