Beyond the Genome: Advancing our Understanding of the Proteome with Next-Generation Protein Sequencing™

Richard Broadhead, Khanh D. Q. Nguyen, Mathivanan Chinnaraj, Abde Ali Kagalwalla, Ben Moree, Ilya Chorny, John Vieceli, and Brian Reed

Quantum-Si Incorporated, 29 Business Park Drive, Branford, CT 06405, USA

Scan for publications

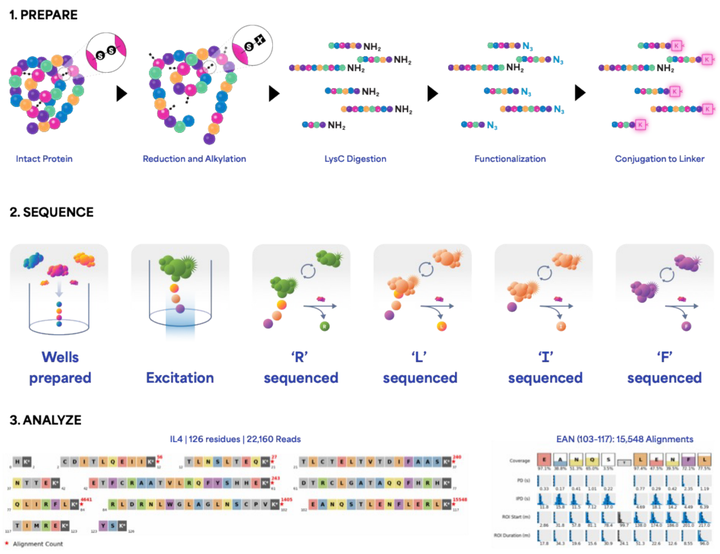
INTRODUCTION

Protein information is a crucial component of multiomics workflows, offering insights that integrate genetic, transcriptomic, and proteomic data for a more comprehensive understanding of biological systems. Multiomics studies would greatly benefit from methods that enable direct protein detection and characterization with exceptional sensitivity. **In this context, we present the unique capabilities of Next-Generation Protein Sequencing™ (NGPS™) on Quantum-Si's Platinum® and Platinum® Pro instruments, which utilize single-molecule protein sequencing to achieve detection of unknown proteins and protein variants.** Individual peptides are digested from proteins, immobilized on a semiconductor chip, and probed by dye-labeled N-terminal amino acid (NAA) recognizers, followed by subsequent aminopeptidase cleavage to expose each NAA in the peptide for recognition. Through the recording and analysis of fluorescent intensity, lifetime, and binding kinetics of each NAA binding event, we can successfully identify proteins, protein variants, and PTMs. **In this study, we demonstrate the application of NGPS to identify unknown proteins, detect proteoform-specific variation, and enable multiplexed protein analysis via the use of protein barcodes.** In light of these capabilities, we envision NGPS as a complementary and valuable addition to multiomics workflows.

METHODS

- Proteins are reduced, alkylated, and digested with LysC.
- Peptides are functionalized, conjugated, and immobilized on the surface of a proprietary semiconductor chip.
- Fluorescently labeled N-terminal amino acid (NAA) recognizers and aminopeptidases are added to the semiconductor chip.
- Fluorescent intensity and duration of each NAA binding event generates a unique kinetic signature.
- Kinetic signatures are analyzed to align reads to reference peptides and compute false discovery rate (FDR).

Platinum® Pro

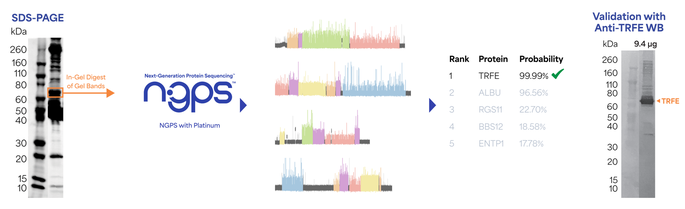


RESULTS

Identification of Unknown Proteins from Human Serum

Human serum was run on SDS-PAGE, and individual bands were excised, followed by in-gel digestion, library preparation, and sequencing using NGPS. **Three proteins, alipoprotein A-I, albumin, and transferrin were correctly identified** without prior knowledge using the Protein Inference analysis workflow. The results were validated with Western blot.

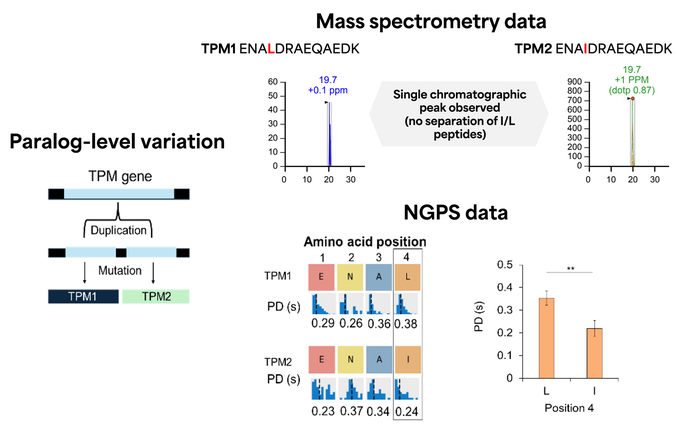
Below, the 75-kDa band was identified to be TRFE with 99.99% confidence.



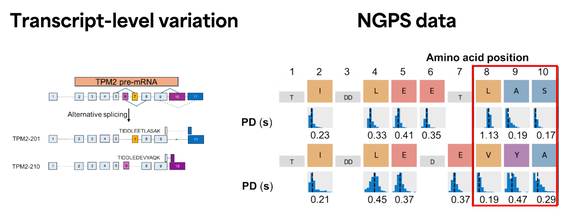
High-resolution Detection of Proteoforms Arising from Genomic, Transcriptomic, and Post-translational Variation

- Tropomyosin is implicated in multiple genetic disorders, including cancer and heart disease
- Despite high amino acid sequence identity, TPM variants have differential impact on disease phenotypes
- NGPS was able to discern peptides that unambiguously match to unique TPM proteoforms to enable targeted research

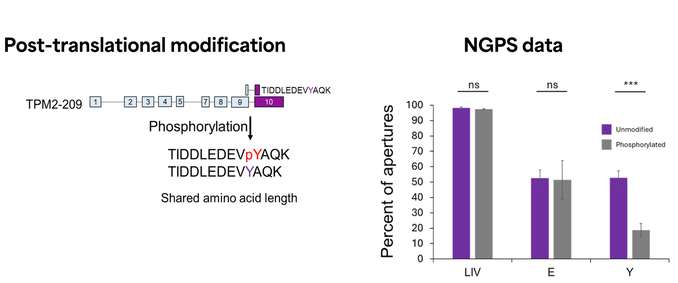
NGPS Distinguishes Isobaric Variants



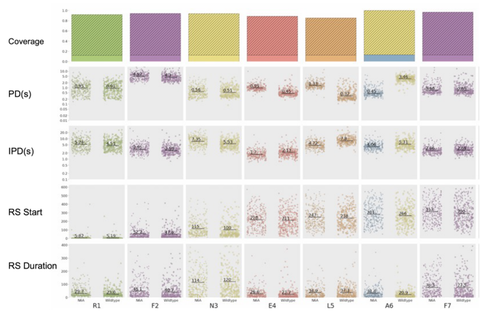
NGPS Discerns Variants Arising from Alternative Splicing



NGPS Can Detect Post-translational Modifications



Visualization of Variants with ProteoVue™



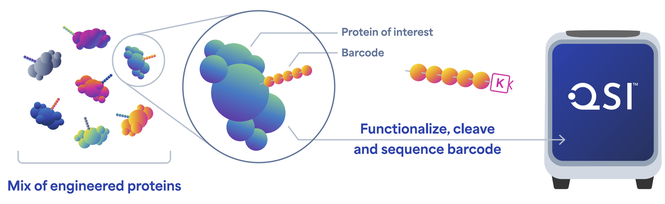
- **Variant Differentiation:** ProteoVue effectively distinguishes between kinetic profiles of single amino acid variants (SAAVs).
- **Population Quantification:** The workflow accurately estimates relative single-amino acid variant (SAAV) abundances to within a factor of 10, enabling precise population differentiation.

Use of Protein Barcodes for Multiplexed Protein Analysis

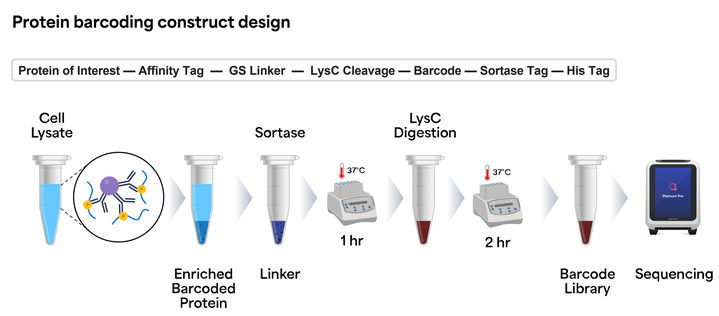
Protein barcodes are short peptides that are highly visible and distinguishable with NGPS on Platinum. When expressed with proteins of interest, they can facilitate multiplexing of functional expression of proteins.

These barcodes can be used for multiple applications:

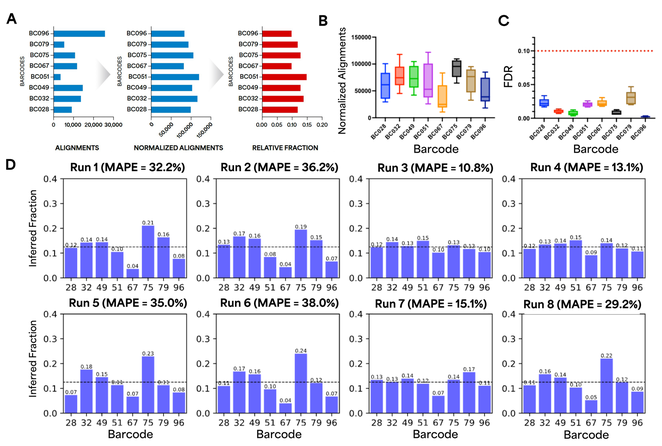
- Directly correlate multiple protein functions to sequence at once to increase throughput.
- Study protein trafficking of proteins from various organelles.
- Identify and characterize protein-protein interactions.
- Screen and characterize proteins with different properties and functions (e.g., LNP/AAV delivery).



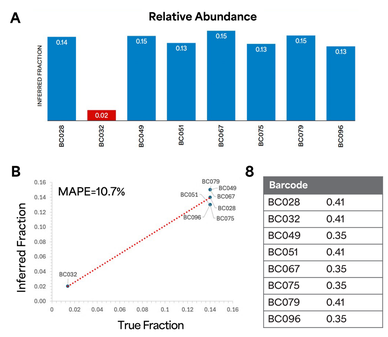
Validation of an Eight-plex Barcode Mixture



- Eight barcodes designed for maximum sensitivity and differentiation via NGPS
- Now available in Barcoding Kit from Quantum-Si

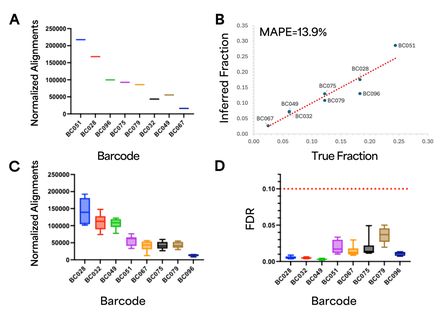


Limit of Detection Down to 400 fmol for Individual Barcodes



Tenfold Dynamic Range Enables Relative Quantitation

- Dynamic range enables use of protein barcodes for relative quantitation within biological contexts



SUMMARY

In this study, we:

- **Isolated proteins from human serum via immunoprecipitation or SDS-PAGE and correctly identified them** from the sequencing data with high confidence
- Demonstrated the power of NGPS to **detect protein variation arising at the genetic, transcriptomic, and post-translational level**
- Showcased **the development and validation of protein barcodes** to streamline protein engineering, antibody engineering, and enzyme engineering

Our technology enables the detection and characterization of proteins and peptides in complex mixtures and biofluids, as well as the detection of critical protein variants implicated in human pathophysiology. **This level of detail is crucial for understanding how genomic and transcriptomic findings translate to cellular phenotypes.** Integrating proteomic data with these other multiomic data types enhances researchers' ability to map complex biological pathways and identify novel biomarkers.

REFERENCES

- Brian D. Reed et al, *Science* 2022, 378 (6166) 186-192
- Sittipongpittaya et al. bioRxiv 2024.11.04.621980; doi: https://doi.org/10.1101/2024.11.04.621980
- Chinnaraj et al. bioRxiv 2024.12.17.629036; https://www.biorxiv.org/content/10.1101/2024.12.17.629036v1
- Chinnaraj et al. bioRxiv 2024.12.31.630920; https://doi.org/10.1101/2024.12.31.630920