Beyond the Genome: Identification of Protein Sequence and PTMs with Quantum-Si's Next-Generation Protein Sequencer™ Platinum®

Richard Broadhead, Khanh D. Q. Nguyen, Sharmin Sultana, Alexa Andrzejewski, Marla Charron, Cassie Simonides, Kenneth Skinner, Michael Meyer, Abde Ali Kagalwalla, Ben Moree, and Brian Reed

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

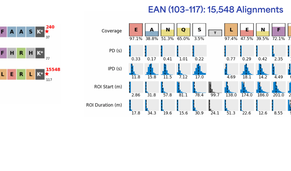
INTRODUCTION

Protein sequencing is a groundbreaking advancement in proteomics that augments genomics and transcriptomics research by providing crucial insights into the functional proteins encoded by the genome. Protein sequencing offers a more complete understanding of cellular processes and disease mechanisms by detecting changes at the protein level, such as post-translational modifications (PTMs), which cannot be captured by genomics data alone. Next-Generation Protein Sequencing™ (NGPS) on Platinum® enables researchers to identify and characterize proteins with single-molecule resolution in a simple workflow and on a benchtop instrument.

To demonstrate the versatility of Platinum and the use of kinetic signatures, we sequenced various types of samples, including mixtures of recombinant proteins, peptides with PTMs, proteins immunoprecipitated from human se¬rum, and proteins isolated from human serum via fractionation with SDS-PAGE. First, we **sequenced a mixture of ten recombinant proteins: HSA, VIME, IL6, PDL1, APOE4, FGF2, AKT1, CDNF, IL4, and H4.** The resulting peptides generated distinct kinetic signatures aligned to their respective sequences, highlighting the efficacy of Quantum-Si's sequencing platform in analyzing multi-protein mixtures at reduced input concentrations. Next, we demonstrated the power of Platinum to detect PTMs on the basis of kinetic changes by **detecting citrullination and dimethylation of arginine**–two PTMs that play key roles in disease states such as cardiovascular disease, autoimmune disease, and cancer.

Finally, we developed software based on a statistical inference method to identify proteins from sequencing data without prior knowledge via mapping to a large reference panel consisting of a subset of the human proteome. This software can also map to user-specified panels, enabling protein identification tailored to specific biological pathways. To demonstrate this capability, we **isolated proteins from human serum via immunoprecipitation or SDS-PAGE and correctly identified them from the sequencing data with high confidence** using an 8,000-protein reference panel.

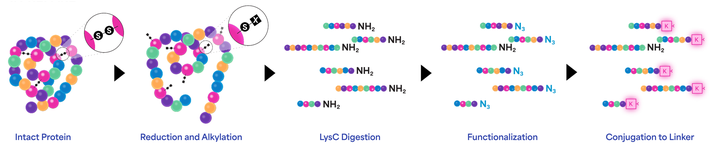
As we envision future enhancements to our system, we anticipate an expanded coverage of the proteome, thereby unraveling even more information often overlooked with genomics data.



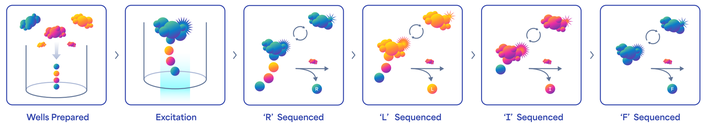
METHODS

- Proteins are reduced, alkylated, and digested with LysC.
- Peptides are functionalized, conjugated, and immobilized on the surface of Quantum-Si's semiconductor chip.
- Fluorescently labeled N-terminal amino acid (NAA) recognizers and aminopeptidases are added to the semiconductor chip.
- Fluorescence lifetime, intensity, and kinetic properties of NAA binding events generate a unique kinetic signature.
- Kinetic signatures are analyzed to align reads to reference peptides and compute false discovery rate (FDR).

1. PREPARE



2. SEQUENCE

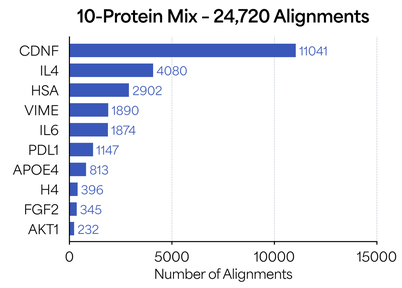


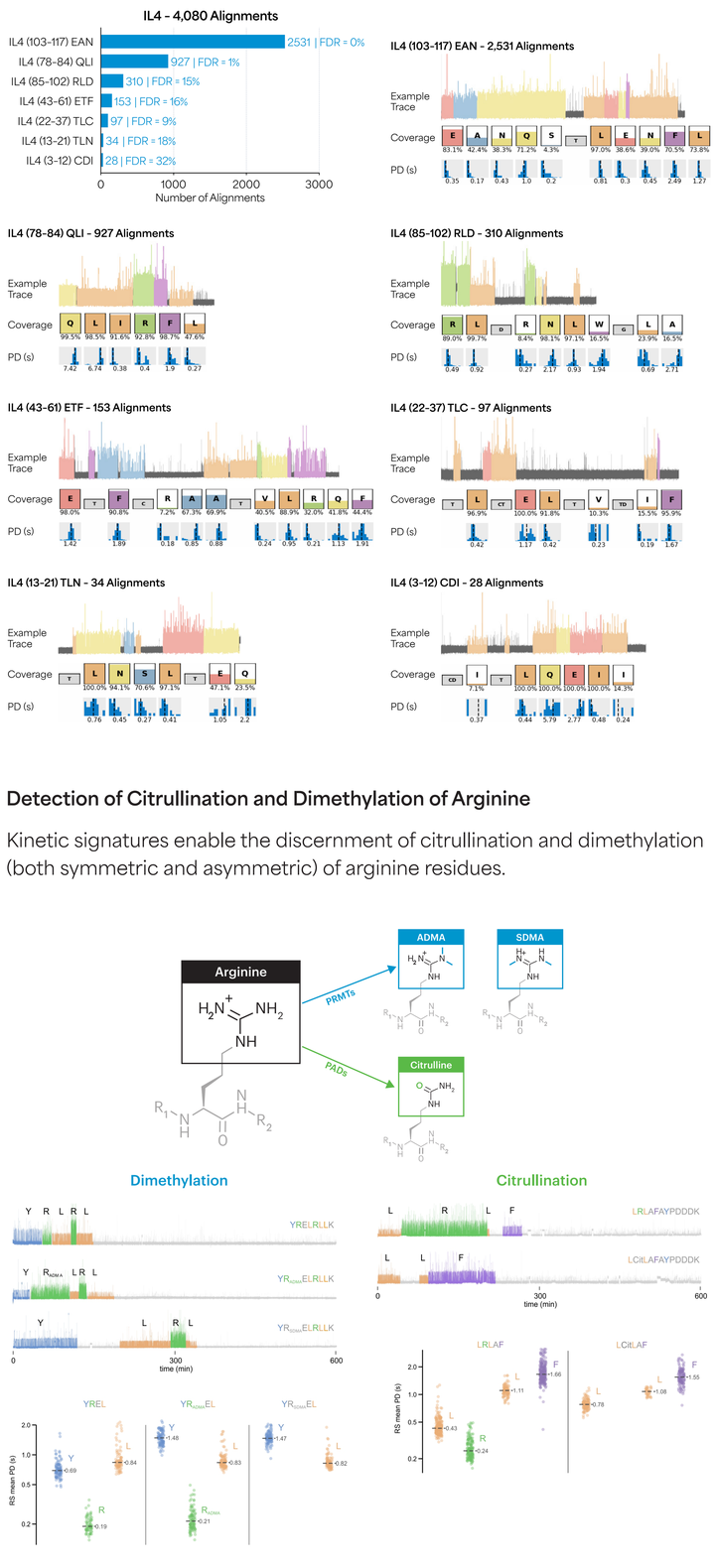
3. ANALYZE

RESULTS

Accurate Identification of a Mixture of 10 Proteins

A mixture of 10 recombinant proteins of various molecular weights and biological functions was successfully sequenced with Platinum.



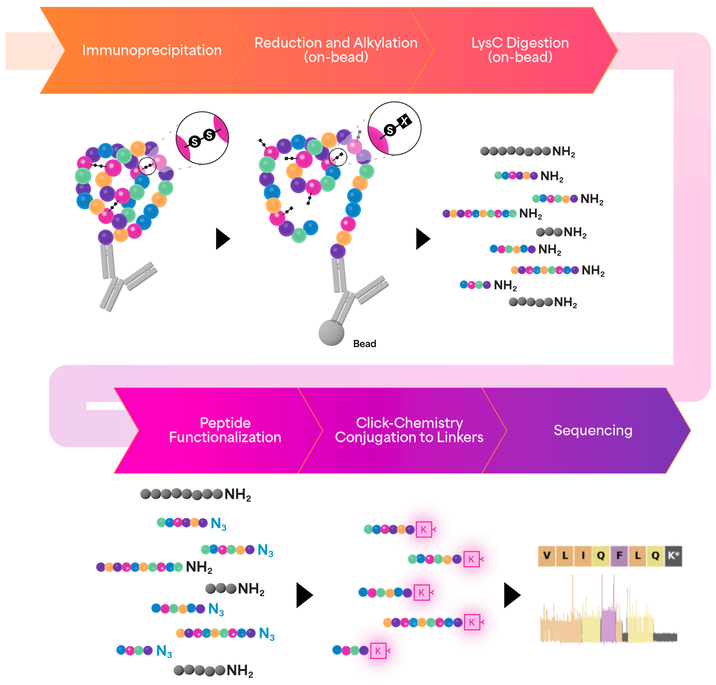


Detection of Citrullination and Dimethylation of Arginine

Kinetic signatures enable the discernment of citrullination and dimethylation (both symmetric and asymmetric) of arginine residues.

Accurate Identification of IL6 Immunoprecipitated from Human Serum

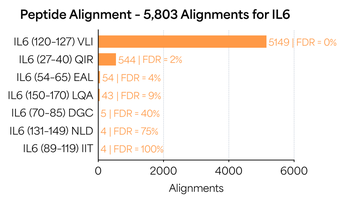
IL6 immunoprecipitated from human serum was correctly identified as the top protein against an 7,921-protein reference panel.



Protein Inference - 99.97% Probability for IL6

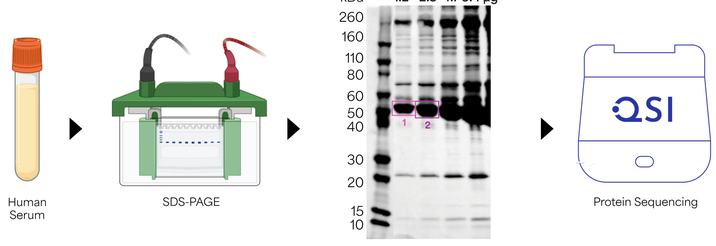
$$Probability = \left(1 - e^{-Inference\ Score}\right) \times 100\%$$

Inference Rank	Protein	Uniprot ID	Inference Score	Probability
1	IL6	P05231	8.15	99.97%



Accurate Identification of HSA from Human Serum Extracted by SDS-PAGE

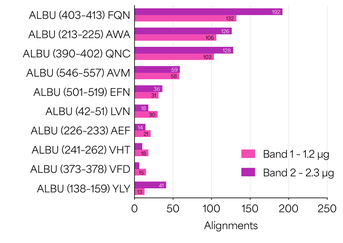
The new protein inference tool enables accurate identification of HSA from extracted SDS-PAGE gel band of human serum.



Protein Inference - 99.99% Probability for ALBU

$$Probability = \left(1 - e^{-Inference\ Score}\right) \times 100\%$$

Band	Protein	Uniprot ID	Inference Score	Probability
1 - 1.2 µg	ALBU	P02768	8.85	99.99%
2 - 2.3 µg	ALBU	P02768	9.64	99.99%



REFERENCE

Brian D. Reed et al, Science 2022, 378 (6166) 186-192.

For research use only.
Not for use in diagnostic procedures.

TRADEMARKS/LICENSING