QUANTUM SI™

Beyond the Genome: Unraveling Protein Variability with Quantum-Si's Next-Generation Protein Sequencing™ Technology

John Vieceli, Khanh D. Q. Nguyen, Mathivanan Chinnaraj, Michael Meyer, Abde Ali Kagalwalla, Ben Moree

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

INTRODUCTION

Next-generation protein sequencing is a transformational tool for protein science to unlock new insights into the function of proteins in health and diseases. Quantum-Si's Platinum® technology brings the insights of protein sequencing to every lab with a space-friendly benchtop instrument, a simple end-to-end workflow, and single-molecule resolution that enables detection of protein variants and modifications.

To demonstrate the versatility of this core methodology and its kinetic principles across a variety of proteins, we **sequenced three variants of SARS-CoV-2 spike proteins: Alpha, Delta, and Omicron**. Using Platinum, we successfully discriminated two single amino acid substitutions among these variants, effectively distinguishing them from one another.

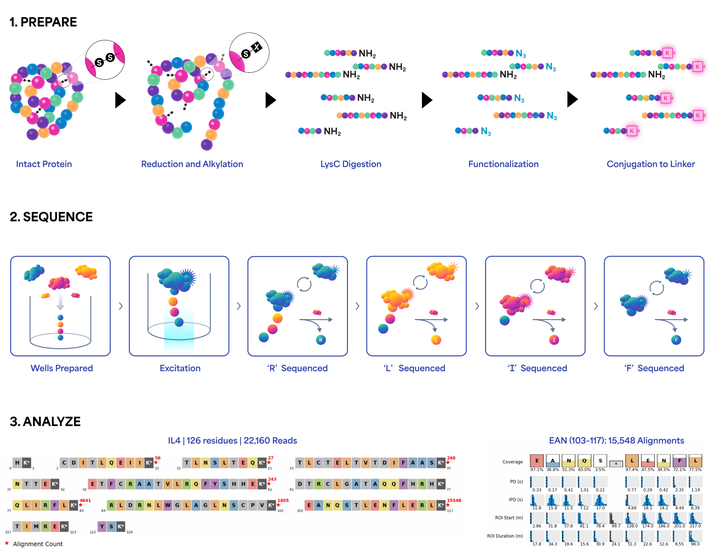
Additionally, Platinum was utilized to **successfully discern a single amino acid substitution at the 12th position of a peptide in ubiquitin**, showcasing the sensitivity of the system in the discovery of mutations deep into peptides.

Finally, recent developments in our analysis algorithms have allowed us to **sequence proteins without prior knowledge of their identities**, enabling protein inferences tailored to specific biological pathways.

As we envision future enhancements to our system, we anticipate an expanded coverage of the proteome, coupled with heightened precision in distinguishing protein variants down to the amino-acid level.

METHODS

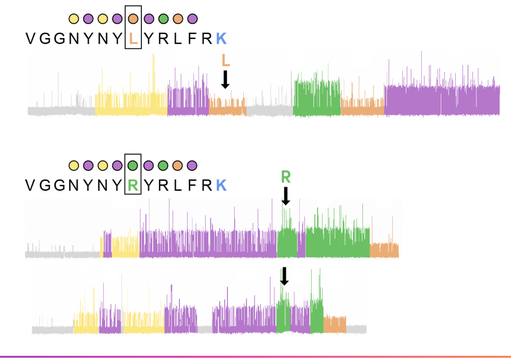
- Proteins are reduced, alkylated, and digested with LysC.
- Peptides are functionalized, conjugated, and immobilized on the surface of a proprietary semiconductor chip.
- Fluorescently labeled N-terminal amino acid (NAA) recognizers and aminopeptidases are added to the semiconductor chip.
- Fluorescent intensity and duration of each NAA binding event generates a unique kinetic signature.
- Kinetic signatures are converted into amino acid calls to identify peptides and proteins.



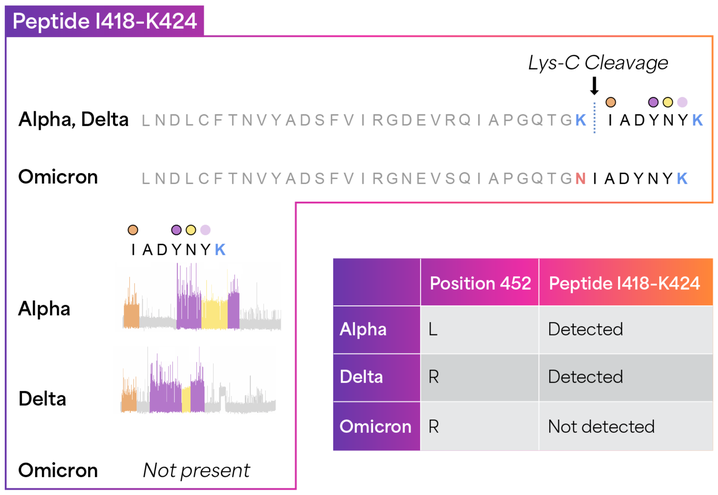
RESULTS

Distinction of Variants of SARS-CoV-2 Virus

Position 452

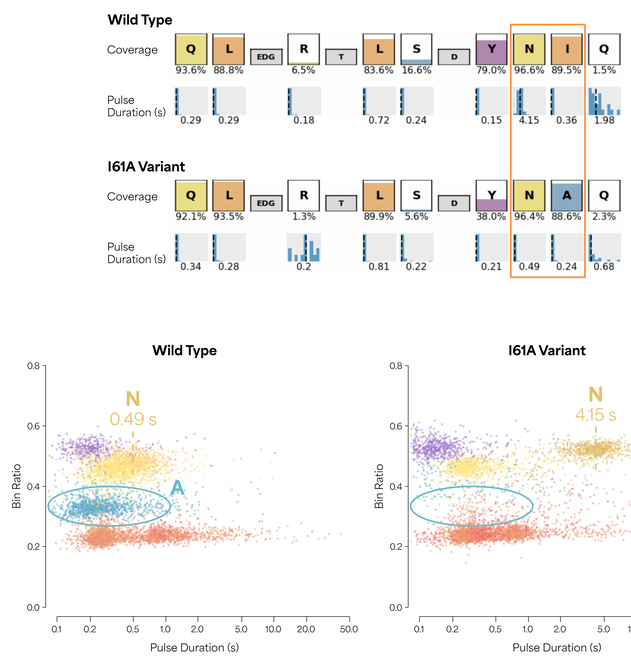


Peptide I418-K424



	Position 452	Peptide I418-K424
Alpha	L	Detected
Delta	R	Detected
Omicron	R	Not detected

Detection of a Single Amino Acid Variant at the 12th Position of a Ubiquitin Peptide



Inference of Protein from Unknown Samples

The Protein Inference workflow is designed to investigate protein samples with unknown identities, employing strict criteria for inference and ranking to ensure robust results.

In this experiment, Quantum-Si's provided reference database comprises 7,921 human proteins (10–70 kDa). Selected proteins have at least three *in silico* LysC-digested peptides with three unique, visible residues.

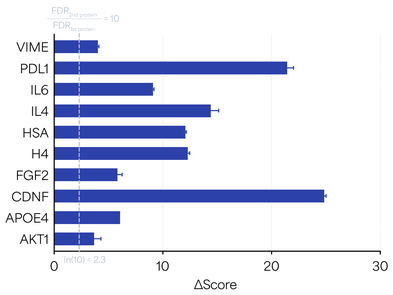
Results are presented in a table, ranking protein by their Inference Score, calculated from the FDR of all digested peptides.

$$\text{Score} = -\ln(\text{FDR}_{\text{protein}}) = -\ln(\prod \text{FDR}_{\text{all digested peptides}})$$

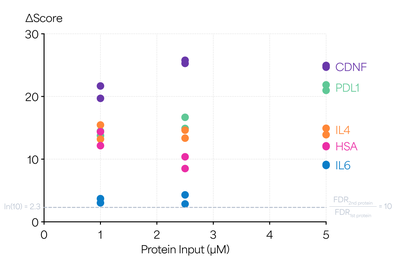
Inference Rank	Inferred Protein (Uniprot ID)	Score	No. of Peptides with < 5% FDR	No. of Peptides with 5-10% FDR	No. of Peptides with > 10% FDR	No. of Alignments
1	**IL4 (P05112)**	**26.11**	**5**	**0**	**0**	**6267**
2	PP1R7 (Q15435)	12.21	2	0	4	281
3	FEM1C (Q96JP0)	9.51	2	0	1	176
4	NUP62 (P37198)	9.22	1	1	0	552
5	CACO2 (Q13137)	8.77	1	1	0	599
6	CTBL1 (Q8WYA6)	8.16	1	1	1	123
7	NUDC1 (Q96RS6)	8.14	1	1	1	280
8	RPAP2 (Q8IXW5)	7.98	0	2	2	57
9	SIA8D (Q92187)	7.85	2	0	0	140
10	OSCP1 (Q8WVF1)	7.47	2	0	0	217

The Score Difference between the first and the second inferred protein (ΔScore) represents the level of confidence of the analysis.

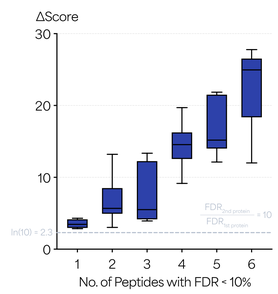
$$\Delta\text{Score} = \text{Score}_{\text{1st protein}} - \text{Score}_{\text{2nd protein}} = \ln \frac{\text{FDR}_{\text{2nd protein}}}{\text{FDR}_{\text{1st protein}}}$$



10 single proteins were successfully inferred.



Level of confidence was not heavily impacted by a 5x decrease in protein input.



Level of confidence increased with number of peptides detected in sequencing.

REFERENCE

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

TRADEMARKS/LICENSING