QUANTUM SI™

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

Khanh D. Q. Nguyen, Mathivanan Chinnaraj,
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

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.
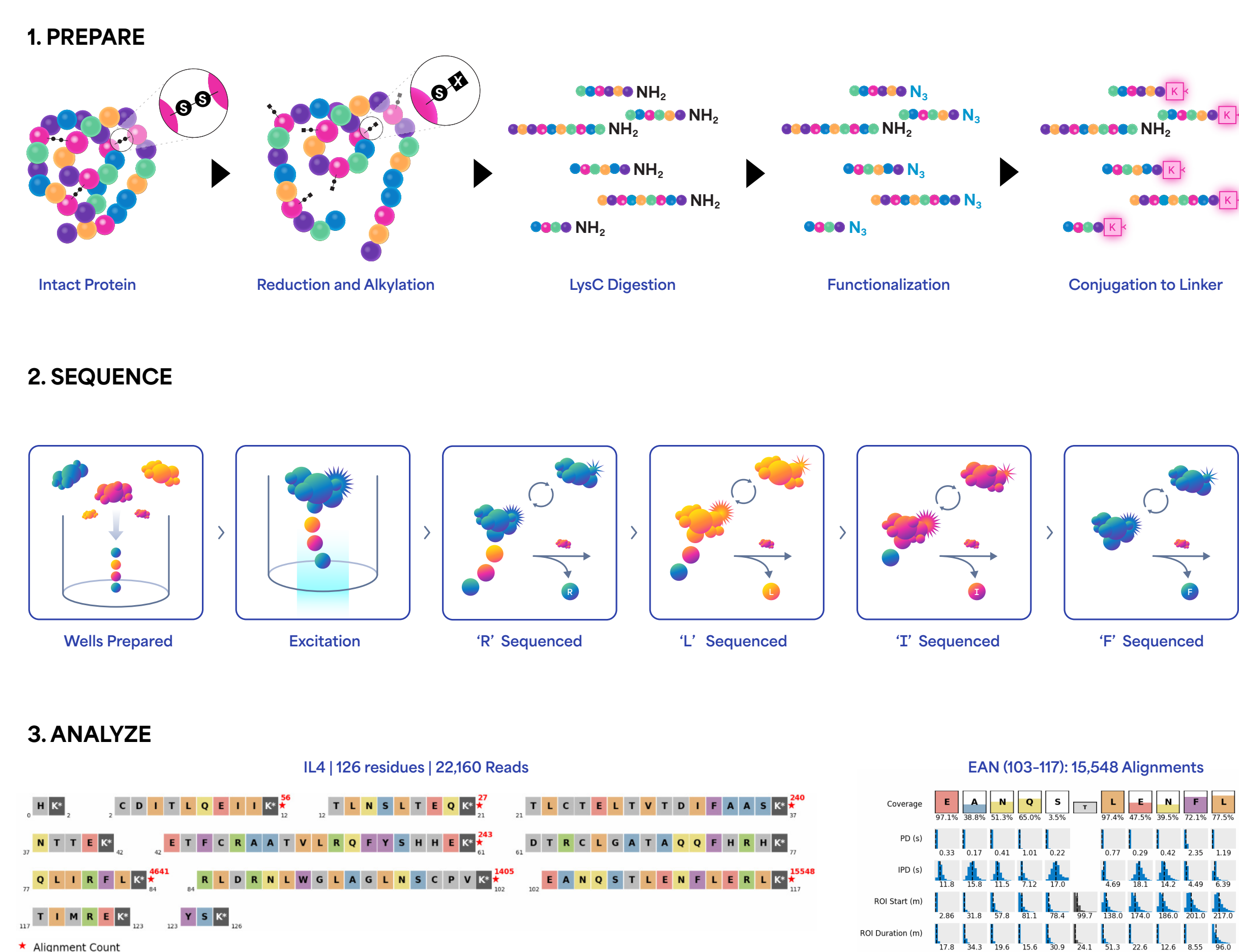
To demonstrate the versatility of Platinum and the use of kinetic signatures, we sequenced various types of samples, including protein variants with single-amino-acid changes, mixtures of recombinant proteins, peptides with PTMs, proteins immunoprecipitated from human serum, and proteins isolated from human serum via fractionation with SDS-PAGE. First, 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. Next, we **successfully discriminated three variants of SARS-CoV-2 spike proteins: Alpha, Delta, and Omicron** using two single amino acid substitutions among these variants.

Next, 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. Additionally, 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.

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).
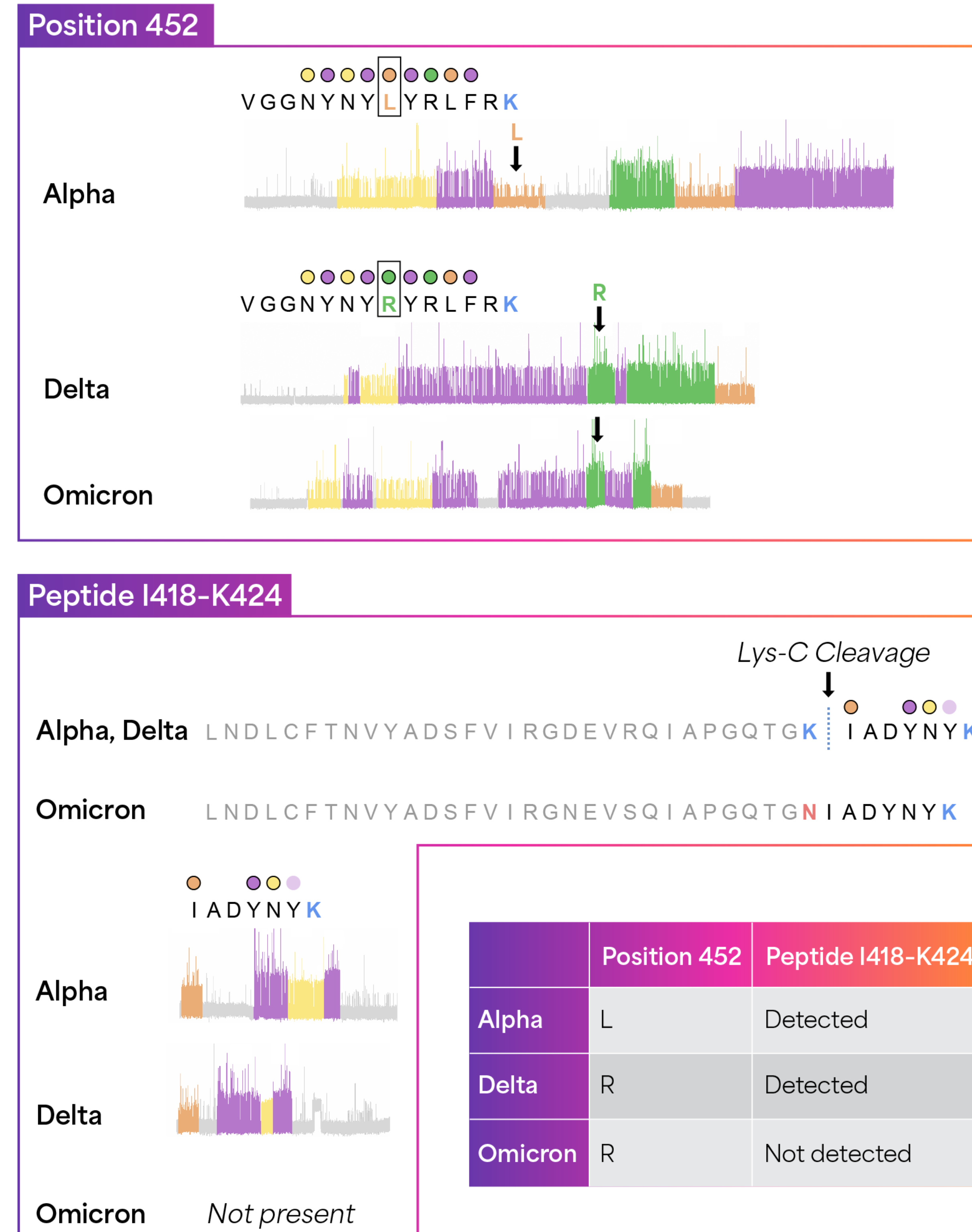

1. PREPARE

Intact Protein → Reduction and Alkylation → LysC Digestion → Functionalization → Conjugation to Linker

2. SEQUENCE

Wells Prepared → Excitation → 'W' Sequenced → 'L' Sequenced → 'T' Sequenced → 'F' Sequenced

3. ANALYZE

RESULTS

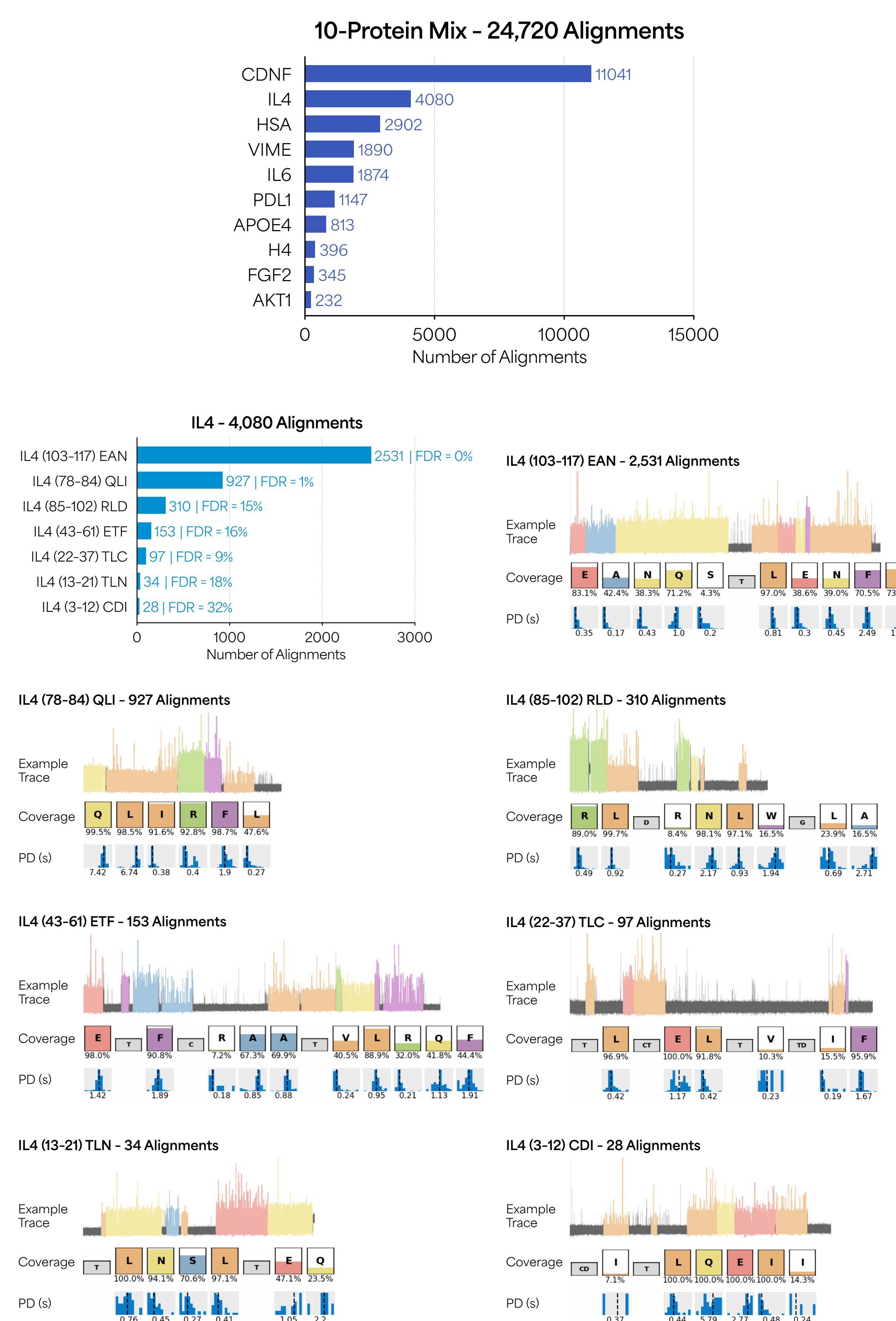### Detection of a Single Amino Acid Variant at the 12th Position of a Ubiquitin Peptide
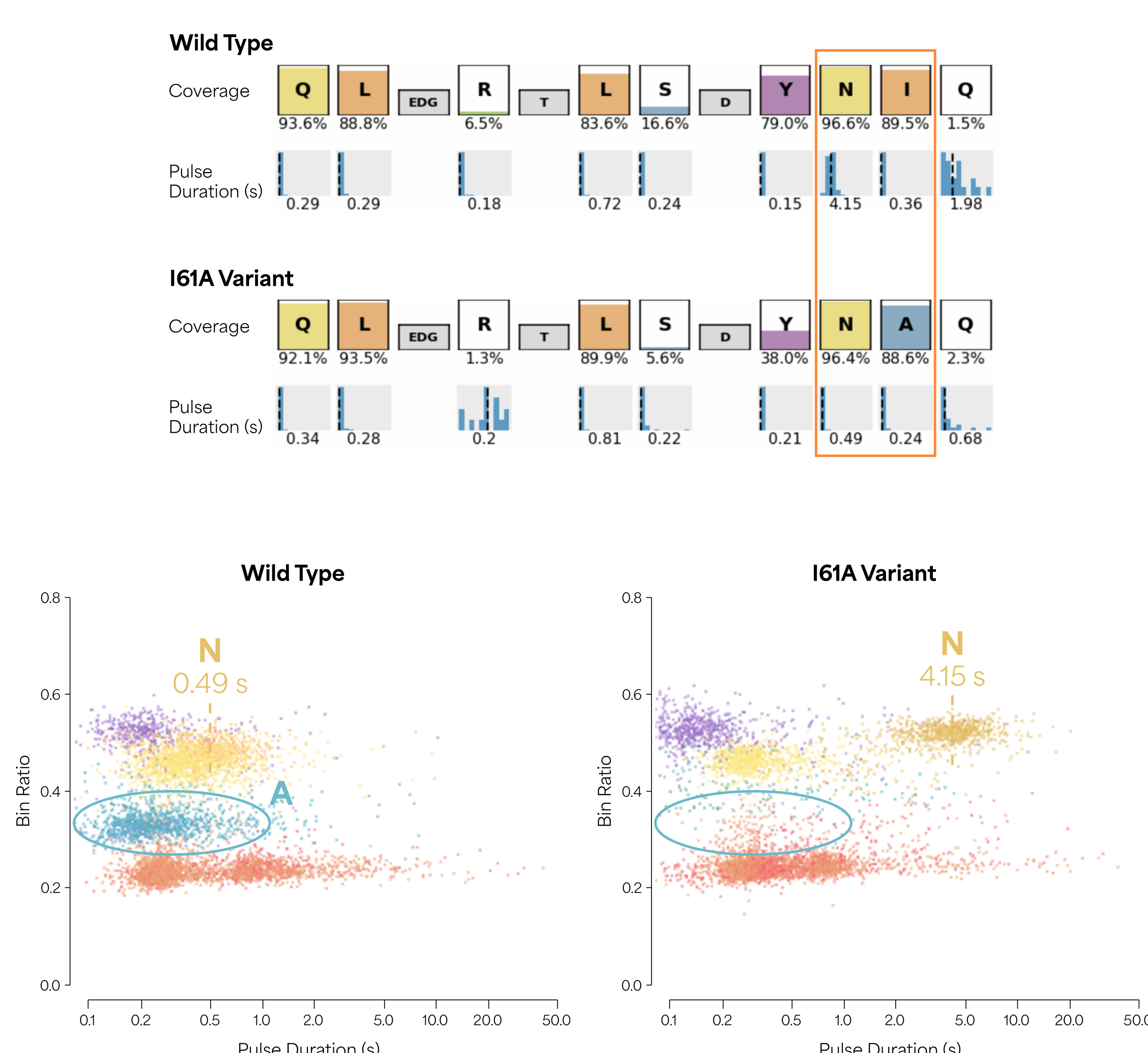
Distinction of Variants of SARS-CoV-2 Virus

Kinetic signatures enable distinction of three major variants of Covid spike proteins through just two single-amino-acid substitutions.

Position 452

Alpha: VGGNYNYL YRLFRK

Delta: VGGNYNYR YRLFRK

Omicron

Peptide I418-K424

Lys-C Cleavage

Alpha, Delta: LNDLCFTNVYADSFVIRGDEVRQIAPGQTGK IADYNYK

Omicron: LNDLCFTNVYADSFVIRGNEVSQIAPGQTGN IADYNYK

Alpha: IADYNYK

Delta

Omicron: *Not present*

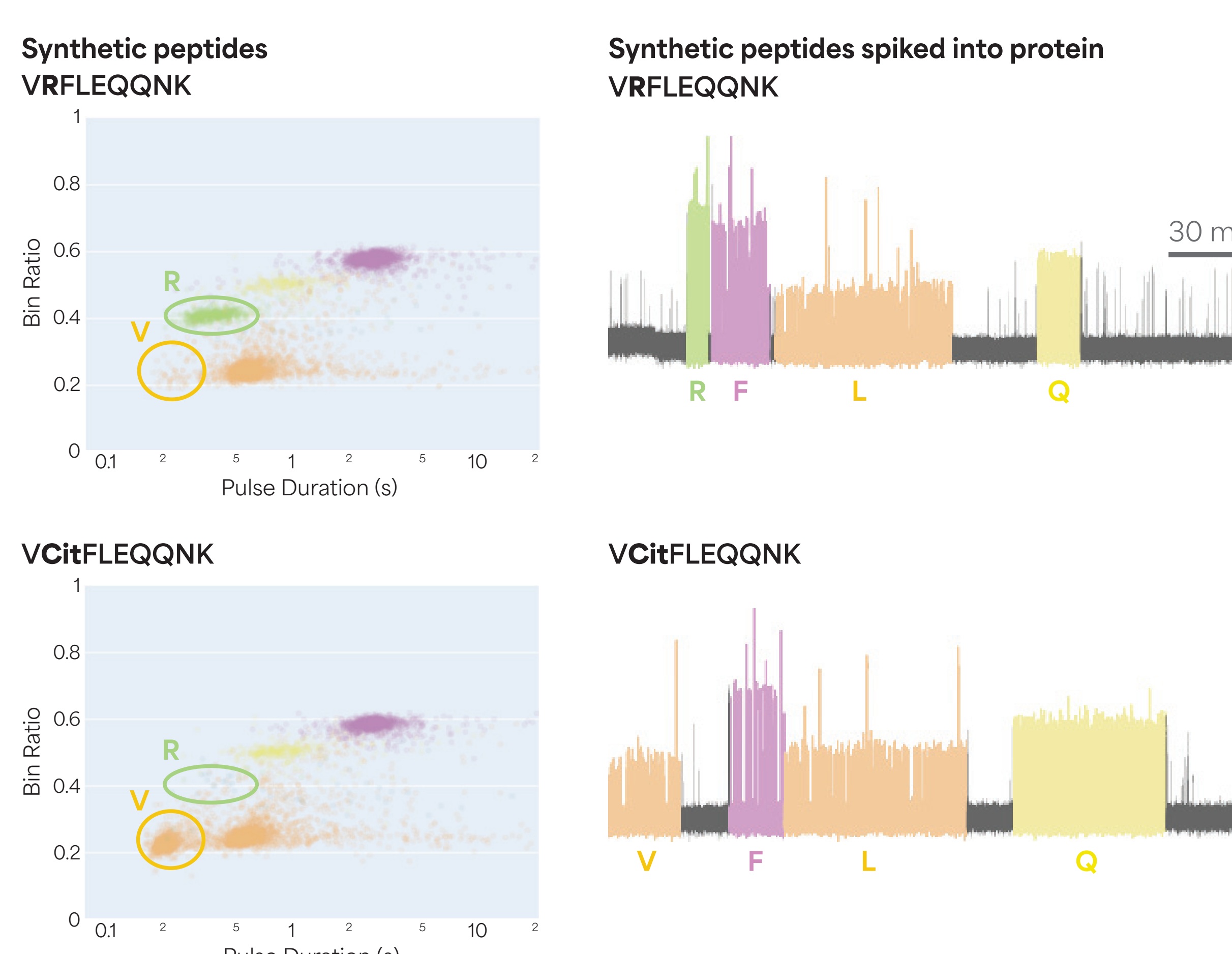|	Position 452	Peptide I418-K424
Alpha	L	Detected
Delta	R	Detected
Omicron	R	Not detected

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.

10-Protein Mix - 24,720 Alignments

Protein	Number of Alignments
CDNF	11041
IL4	4080
HSA	2902
VIME	1890
IL6	1874
PDL1	1147
APOE4	813
H4	396
FGF2	345
AKT1	232

IL4 - 4,080 Alignments

Peptide	Alignments	FDR
IL4 (103-117) EAN	2531	FDR = 0%
IL4 (78-84) QLI	927	FDR = 1%
IL4 (85-102) RLD	310	FDR = 15%
IL4 (43-61) ETF	153	FDR = 16%
IL4 (22-37) TLC	97	FDR = 9%
IL4 (13-21) TLN	34	FDR = 18%
IL4 (3-12) CDI	28	FDR = 32%

IL4 (103-117) EAN - 2,531 Alignments

IL4 (78-84) QLI - 927 Alignments

IL4 (85-102) RLD - 310 Alignments

IL4 (43-61) ETF - 153 Alignments

IL4 (22-37) TLC - 97 Alignments

IL4 (13-21) TLN - 34 Alignments

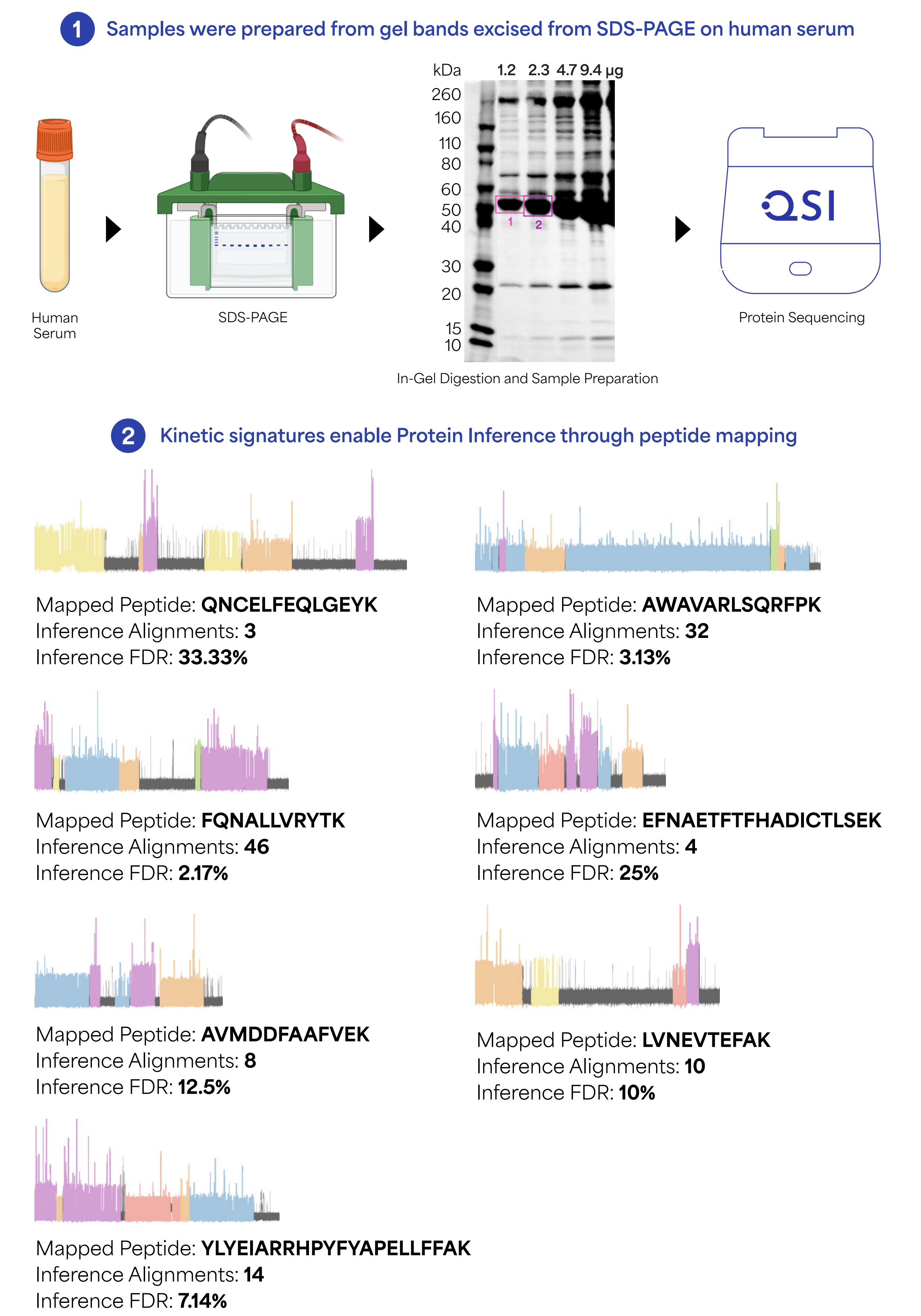IL4 (3-12) CDI - 28 Alignments

Detection of Arginine Citrullination from Peptides Derived from Vimentin

Citrullination led to the recognition of the first V residue, while the recognition of the second R residue was eliminated.

Synthetic peptides
VRFLEQQNK

Synthetic peptides spiked into protein
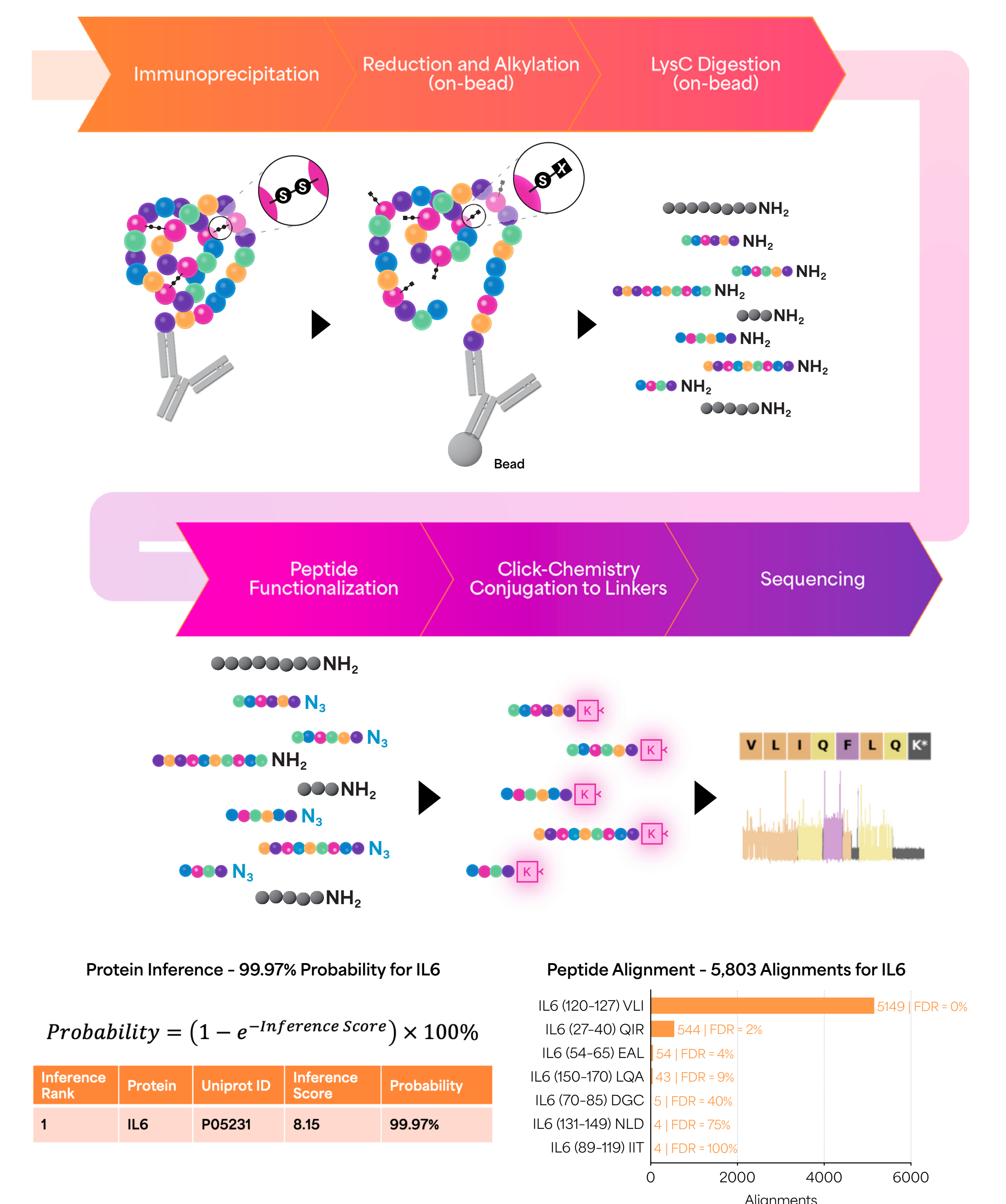VRFLEQQNK

30 min

VCitFLEQQNK

VCitFLEQQNK

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

The new protein inference tool (7,921-protein panel) enables accurate identification of HSA from extracted SDS-PAGE gel band of human serum.

1 Samples were prepared from gel bands excised from SDS-PAGE on human serum

Human Serum → SDS-PAGE → Protein Sequencing

In-Gel Digestion and Sample Preparation

2 Kinetic signatures enable Protein Inference through peptide mapping

Mapped Peptide: **QNCELFEQLGEYK**
Inference Alignments: **3**
Inference FDR: **33.33%**

Mapped Peptide: **AWAVARLSQRFPK**
Inference Alignments: **32**
Inference FDR: **3.13%**

Mapped Peptide: **FQNALLVRYTK**
Inference Alignments: **46**
Inference FDR: **2.17%**

Mapped Peptide: **EFNAETFTFHADICTLSEK**
Inference Alignments: **4**
Inference FDR: **25%**

Mapped Peptide: **AVMDDFAAFVEK**
Inference Alignments: **8**
Inference FDR: **12.5%**

Mapped Peptide: **LVNEVTEFAK**
Inference Alignments: **10**
Inference FDR: **10%**

Mapped Peptide: **YLYEIARRHPYFYAPELLFFAK**
Inference Alignments: **14**
Inference FDR: **7.14%**

3 ALBU has a probability of 99.99% to be present in the sample

Protein Inference – 99.99% Probability for ALBU

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

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%

Peptide Alignment for ALBU

	Alignments
ALBU (403-413) FQN	
ALBU (213-225) AWA	
ALBU (390-402) QNC	
ALBU (546-557) AVM	
ALBU (501-519) EFN	
ALBU (42-51) LVN	
ALBU (226-233) AEF	
ALBU (241-262) VHT	
ALBU (373-378) VFD	
ALBU (138-159) YLY	

Band 1 1 µg / Band 2 - 2.3 µg

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.

Immunoprecipitation → Reduction and Alkylation (on-bead) → LysC Digestion (on-bead)

Bead

Peptide Functionalization → Click-Chemistry Conjugation to Linkers → Sequencing

Protein Inference - 99.97% Probability for IL6

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

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

Peptide Alignment – 5,803 Alignments for IL6

Peptide	Alignments	FDR
IL6 (120-127) VLI	5149	FDR = 0%
IL6 (27-40) QIR	444	FDR = 3%
IL6 (54-65) EAL	54	FDR = 9%
IL6 (50-170) LQA	43	FDR = 9%
IL6 (70-85) DGC	5	FDR = 40%
IL6 (131-149) NLD	4	FDR = 75%
IL6 (89-119) IIT	1	FDR = 100%

REFERENCE

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

TRADEMARKS/LICENSING

All trademarks are the property of Quantum-Si, Inc. or their respective owners. For specific trademark information, see:
www.quantum-si.com/privacy-policy.

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