Protein Barcoding and Next-Generation Protein Sequencing™ for Multiplexed Protein Selection, Analysis, and Tracking

Poster P01.24

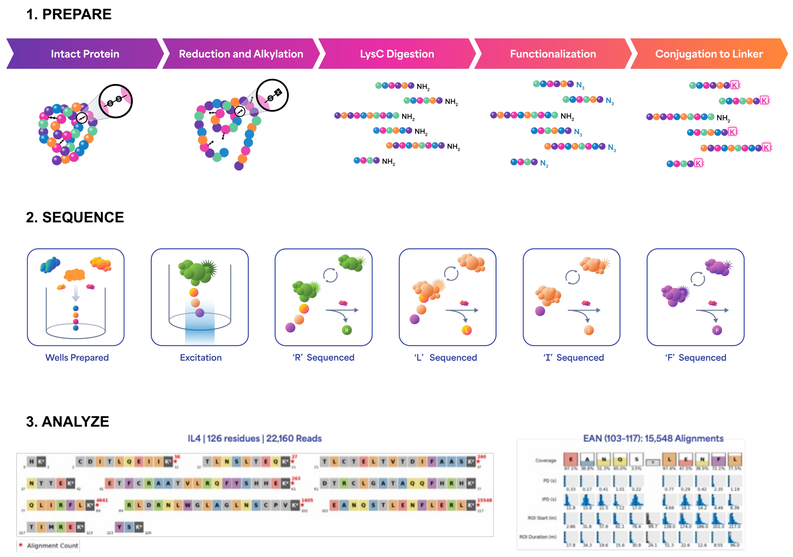
Presenter: Joel McDade,[1]
Mathivanan Chinnaraj,[1] Haidong Huang,[1] Sebastian Hutchinson,[1] Michael Meyer,[1] Douglas Pike,[1] Marco Ribezzi,[1] Sharmin Sultana,[1] Ghada H. Mansour,[2] Derrek Ocampo,[1] Fengling Ding,[1] Meredith L. Carpenter,[1] Ilya Chorny,[1] and John Vieceli[1]

1. Quantum-Si Incorporated, 29 Business Park Drive Branford, CT 06405, USA
2. Laboratoire de Biochimie, Chimie Biologie et Innovation, ESPCI Paris, Université PSL, Paris, France

Abstract

Protein barcoding has emerged as a powerful tool for the multiplexed identification and characterization of proteins, providing a mechanism for precise tracking of protein affinity, location, and expression. In this study, we describe the development of a protein barcoding workflow for use with single-molecule Next-Generation Protein Sequencing™ (NGPS™) on the benchtop Platinum® and Platinum® Pro instruments. We present data on the validation of eight peptide barcodes, each designed to minimize detection bias and maximize sensitivity across various experimental conditions. We have also optimized the design of expression constructs to decrease both the hands-on time and input requirements of the workflow. In this workflow, affinity-tagged proteins are expressed with unique peptide barcodes. Following experimental selection or treatments, the proteins are purified, and the peptide barcodes are cleaved and sequenced on the Platinum instrument. We demonstrate that we can detect barcodes at 400 fmol of sample input concentration within the eight-plex mixture, and at 50 fmol of sample input for individual barcodes. We also show the capacity of this barcoding approach to achieve a ten-fold dynamic range, underscoring its sensitivity in recovering variants with low abundance. Through the combination of protein barcoding and NGPS, we lay the groundwork for future studies aimed at characterizing protein interactions and improving targeted drug delivery strategies.

About Next-Generation Protein Sequencing



NGPS on the Platinum instruments:

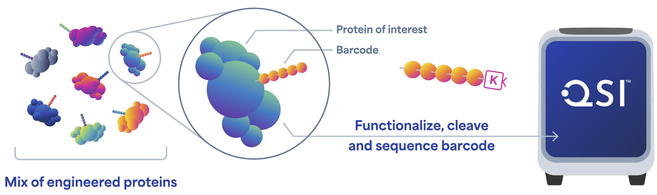
- Proteins are reduced, alkylated, and digested with LysC.
- Peptides are functionalized, conjugated, and immobilized on the surface of a proprietary semiconductor chip.
- Fluorescently labeled N-terminal amino acid (NAA) recognizers and aminopeptidases are added to the semiconductor chip.
- Fluorescent 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).

References

1. Chinnaraj et al. bioRxiv 2024.12.31.630920; doi: https://doi.org/10.1101/2024.12.31.630920

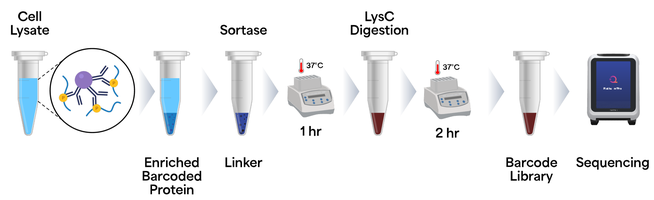
Scan for Publications

Protein Barcoding Workflow



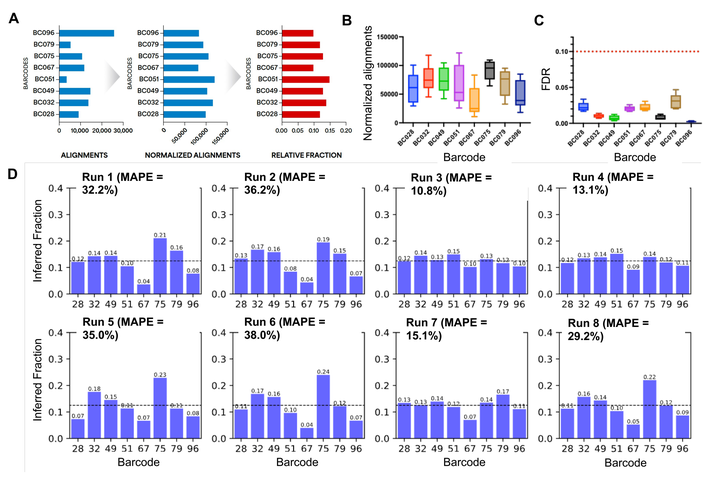
Barcoded protein enrichment and sequencing workflow

Protein of Interest – Affinity Tag – GS Linker – LysC Cleavage – Barcode – Sortase Tag – His Tag



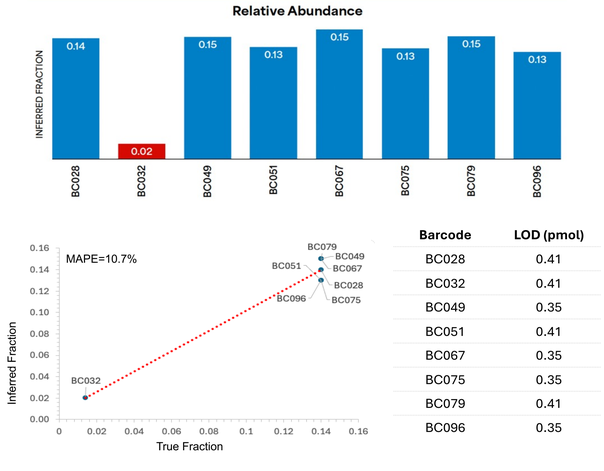
- Barcoding construct designed to minimize impact on protein function — entire tag (including barcode) is 35 amino acids
- Workflow takes less than six hours, with less than one hour of hands-on time

Normalization of Barcodes



- Eight barcodes were designed for maximum compatibility with NGPS sequencing and normalization factors were derived from over 25 sequencing runs
- When applied, mean absolute percent error (MAPE) is ~25% and false discovery rate (FDR) is below 0.1, indicating high accuracy and reproducibility

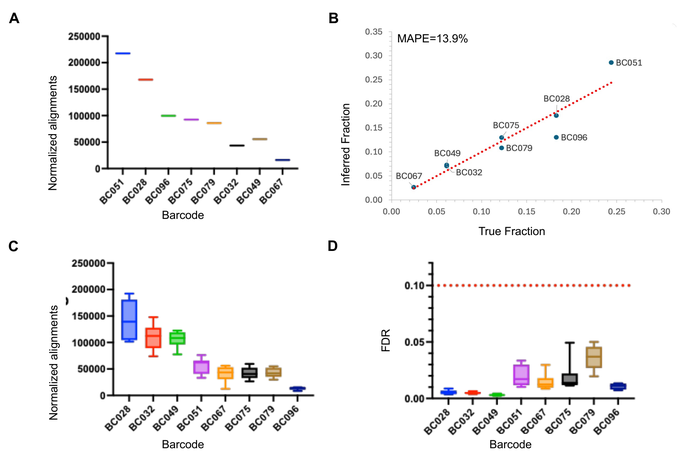
Limit of Detection Below 400 fmol in Eight-Plex Mixture



Barcode	LOD (pmol)
BC028	0.41
BC032	0.41
BC049	0.35
BC051	0.41
BC067	0.35
BC075	0.35
BC079	0.41
BC096	0.35

- 10-fold range titrations were performed, with each barcode tested at decreased input
- LODs were ~400 fmol in eight-plex mixture (above) and 50 fmol for individual barcodes (not shown)

Ten-Fold Dynamic Range

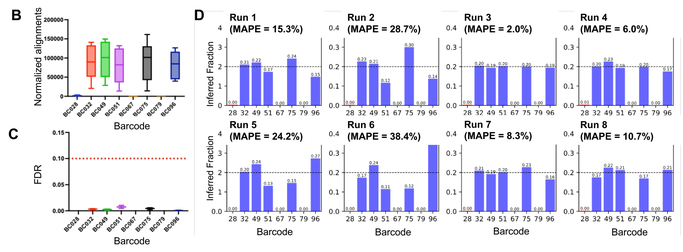


- In eight-plex mixture, titration from 1x down to 0.1x produces linear dynamic range (R^2 of 0.9) with low MAPE and FDR below 0.1
- Accurate relative input amount was also recovered from unknown titrations using normalized, inferred abundances

Recovery of a Mixture of Five Proteins

A

Barcode ID	Protein	UniProt ID	AA Length	MW (kDa)
BC032	IFNg	P01579	206	23.69
BC049	PTEN	P60484	443	51.43
BC051	TAU441	P10636	481	50.34
BC075	UCHL1	P09936	263	29.21
BC096	p53	P46037	433	47.90



- Equimolar mixture of 5 full-length, barcoded proteins (listed in table) produced positive identification for all five barcodes
- Eight runs with multiple lots of reagents/kits and two operators produced MAPE average of 16.7%, indicating high reproducibility

Summary

Objective: To describe the development and validation of a novel protein barcoding workflow for use with NGPS on the Platinum/Platinum Pro instrument

Key findings:

- **Validation and normalization**: carefully designed and normalized barcodes improve linearity and minimize bias
- **Limit of detection:** Lower limit of detection was found to be 400 fmol within an 8-plex mixture and 50 fmol for a single barcode
- **Broad dynamic range:** Barcoding method achieves a 10-fold dynamic range, emphasizing its sensitivity in detecting low-abundance variants

Conclusion: The protein barcoding workflow — now supported by the Barcoding Kit — enables multiplexed protein selection, analysis, and tracking, with applications ranging from protein engineering to nucleic acid therapy development