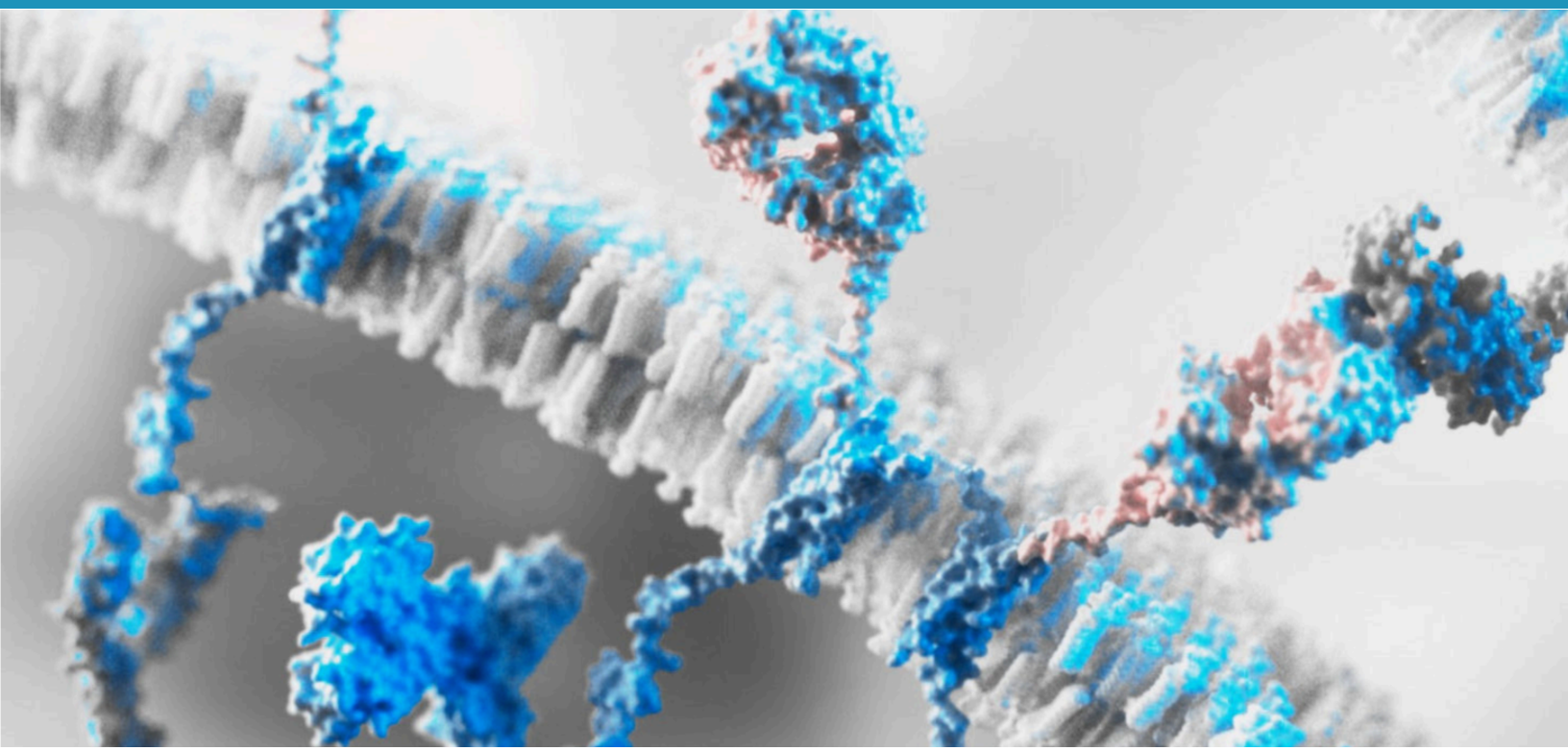
Bioinformatics Solutions Inc.



Complete workflow for immunopeptidome analysis in PEAKS 11

Qing Zhang, MSc, Software Systems Manager

Bioinformatics Solutions Inc., Waterloo, Canada

Abstract:

Immunopeptidome is an attractive avenue for peptide-based vaccine and cancer immunotherapy development. De novo-based search could help to identify potential non-canonical peptides from MS2 spectra directly without protein sequence databases. So, combining normal database search and de novo search would be a promising approach for Immunopeptidomics. Here we demonstrated a complete workflow for immunopeptidomes analysis in newly released PEAKS 11 for both DDA and DIA data.

Inroduction:

Mass spectrometry-based workflow are essential for discovering immunopeptides, especially for tumor-specific antigens (TSA) or tumor-associated antigens (TAA). Mass spectrometry data analysis usually rely on priori protein sequences to do database (DB) searching. This works for canonical peptides, but non-conical peptides are not included in standard protein sequence databases and are often missed.

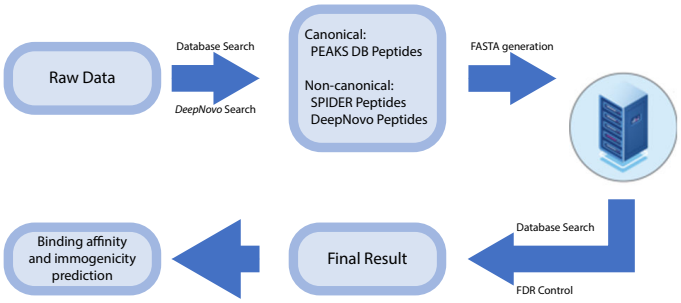


Fig 1. Schematic workflow for DeepNovo peptidome workflow.

De novo-based searching could be an alternative for non-canonical discovery since it does not require previous protein sequence identifications. But well-established "target-decoy" approach is designed for DB search, not for de novo peptides. Here, we demonstrated an approach to control FDR for both DB search and de novo sequencing results. We applied our workflow for both DDA and more complicated DIA data to show sensitivity of our peptidome workflow.

Methods:

A dataset from published paper was used [1]. For demonstrating this workflow, we selected DDA and DIA data from human cell line RA957 as demo samples. DDA data were processed in our peptidome workflow in PEAKS 11, named DeepNovo Peptidome, with default setting for orbitrap data. DIA data were processed in PEAKS 11 streamlined DIA identification workflow with comparable parameters from published work [2].

Results:

DeepNovo Peptidome workflow is a complete workflow combing DB, SPIDER and DeepNovo search (Fig.1). All results are controlled and filtered by FDR using the target-decoy fusion method.

DDA Data Analysis:

We processed previously published data in DeepNovo Peptidome workflow with similar search parameter but using a stricter FDR control (0.1% peptide FDR). We identified almost double the number of peptides (28813 vs 14789), including 10% of peptides found by DeepNovo. The peptide sequence overlap shows our workflow covers almost 93% of reported sequences (Fig.2a). We further checked reported peptide length distribution and untargeted motif deconvolution. Similar length distribution (Fig.2b) and motifs (data not shown here) validate our results further.

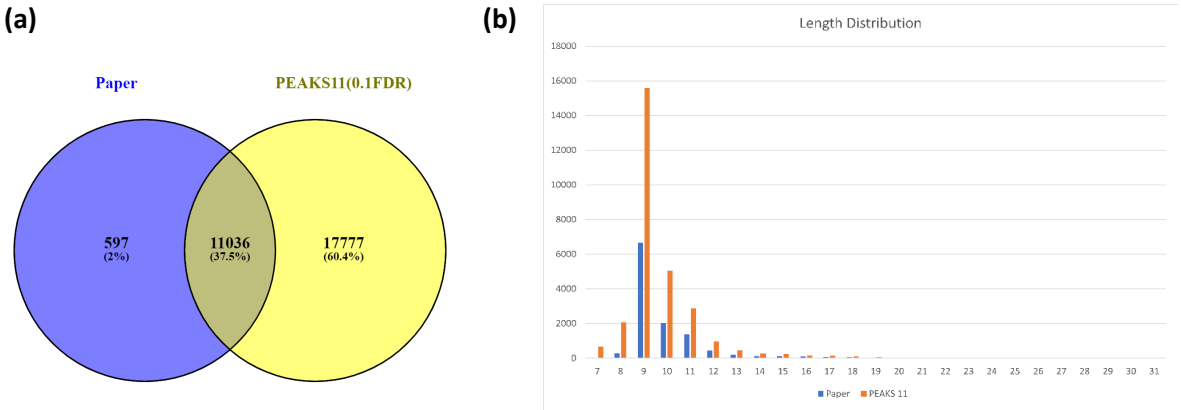


Fig 2. DDA data analysis result (a) Venn Diagram between Paper reported and Peptidome workflow from PEAKS 11 (b) Length distribution of paper and PEAKS11 result.

We processed DIA data from the paper and compared our results with newly published AlphaPeptDeep algorithms [2]. In our streamlined DIA workflow, some Deep Learning based re-scoring and predictions were used to boost the accuracy and sensitivity in the DIA analysis. We reported comparable peptide identification numbers (36039) with those published (36947). Also, the sequence overlap showed each algorithm identifies significantly different sequences (Fig.3a). Further analysis of peptide length distribution (Fig.3b) and motif (Fig.4) show similarity between the two datasets.

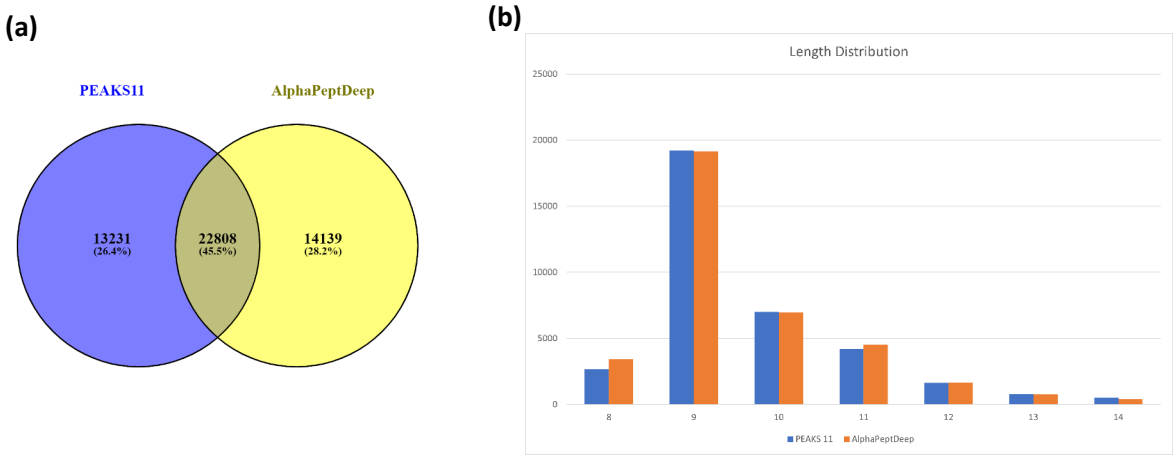


Fig 3. DIA data analysis result (a) Venn Diagram between Paper reported and Peptidome workflow from PEAKS 11 (b) Length distribution of paper and PEAKS 11 result.

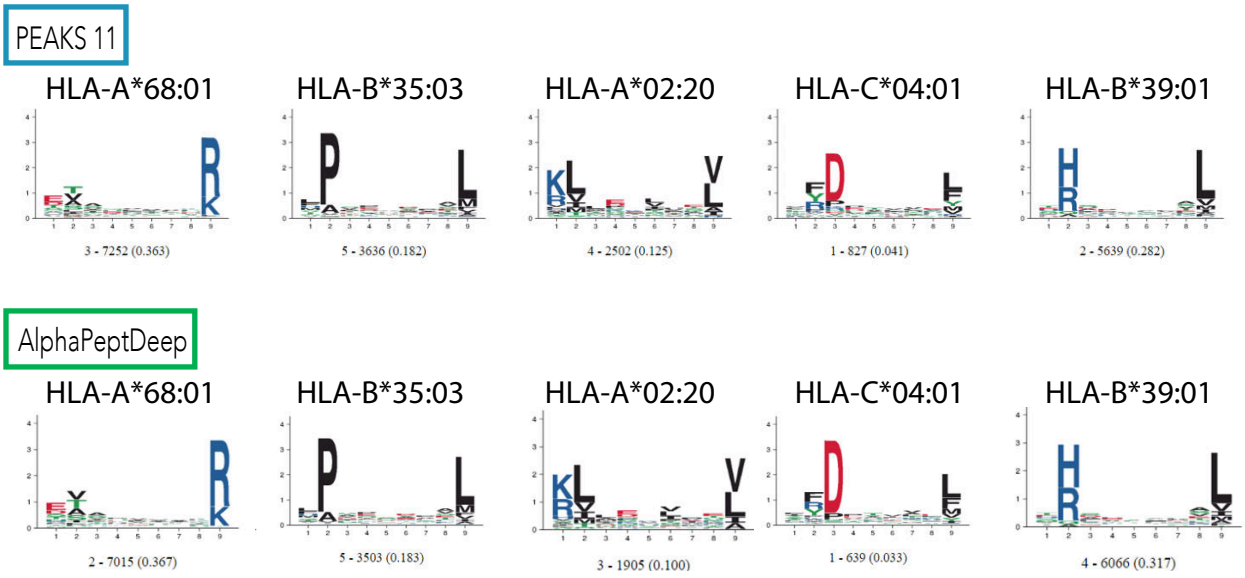


Fig 4. deconvoluted motifs from PEAKS 11 and paper

Conclusion:

DeepNovo Peptidome and streamlined DIA workflows in PEAKS 11 perform well in immunopeptidome discovery.

References:

1. Pak H. et al. Sensitive Immunopeptidomics by Leveraging Available Large-Scale Multi-HLA Spectral Libraries, Data-Independent Acquisition, and MS/MS Prediction. Mol Cell Proteomics. 20:100080 (2021).
2. Zeng, WF., Zhou, XX., Willems, S. et al. AlphaPeptDeep: a modular deep learning framework to predict peptide properties for proteomics. Nat Comm 13, 7238 (2022).

Bioinformatics Solutions, Inc.

140 Columbia St, Suite 202
Waterloo, Ontario N2L 3K8
Canada

Tel: (519) 885-8288
Fax: (519) 885-9075

sales@bioinfor.com
www.bioinfor.com