A New Algorithm for Identification of Immunopeptides from LC-MS Data with High Sensitivity

Lin He[1], Lei Xin[1], Hieu Tran[2], Xin Chen[1], Paul Shan[1], Ming Li[2]
[1] Bioinformatics Solutions Inc., Waterloo, Canada, N2L 6J2
[2] Cheriton School of Computer Science, University of Waterloo, Waterloo, Canada, N2L 3G1

Bioinformatics Solutions Inc.

Overview

Purpose: To develop an automated software suite to accelerate the identification of immunopeptides.
Methods: AI-based *de novo* sequencing and database search were integrated for peptide identification.
Results: A software suite, PEAKS X, was provided for immunopeptidomics with high sensitivity and accuracy.

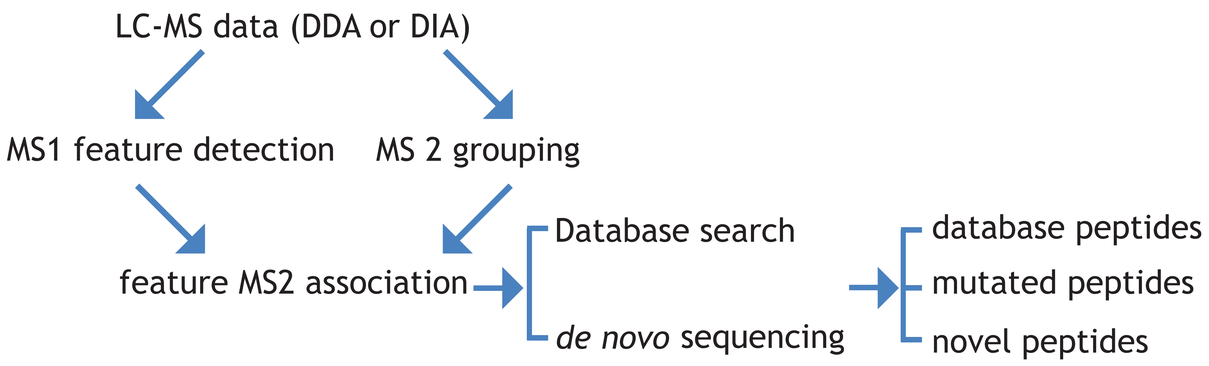
Introduction

Personalized immunotherapy, in ideal cases, should depend on the neoantigens present on the cancer cell surface, of one person, one tumor, and one time. A few research groups reported direct identification of mutated peptides isolated from human leucocyte antigens (HLA) by LC-MS. Until recently, MS technologies were not sensitive enough to do this. The key challenges include diverse C-termini of HLA-peptides, lack of sequence library for spliced peptides, no peptide *de novo* sequencing algorithms for data independent acquisition (DIA) method, etc. We have introduced deep learning into peptide *de novo* sequencing [1]. Here, a new algorithm was proposed to combine *de novo* sequencing and database search for the identification of immnuopeptides from LC-MS data with both DDA and DIA approach.

Methods

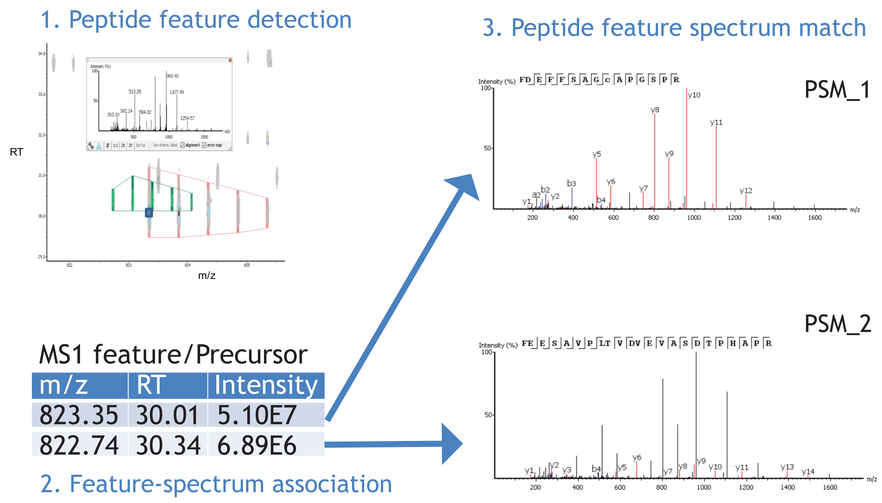
1. Workflow: A peptide feature-based workflow, shown in Figure 1, works for both DDA and DIA data.

Figure 1. Workflow of identification



2. Chimera spectra for DDA: A feature-based approach was implemented to handle chimeric tandem spectra for co-fragmented peptides, which significantly increases the efficiency of peptide identification.

Figure 2. Peptide feature-Spectrum matches for chimeric spectra



3. *De Novo* Sequencing: Since *de novo* sequencing is a completely unbiased peptide identification workflow, improvements to *de novo* sequencing is much more exploitable and useful for peptide identification. Deep learning technology was used for peptide *de novo* sequencing, yielding much higher accuracy and sensitivity [1].

Figure 3. Deep learning system

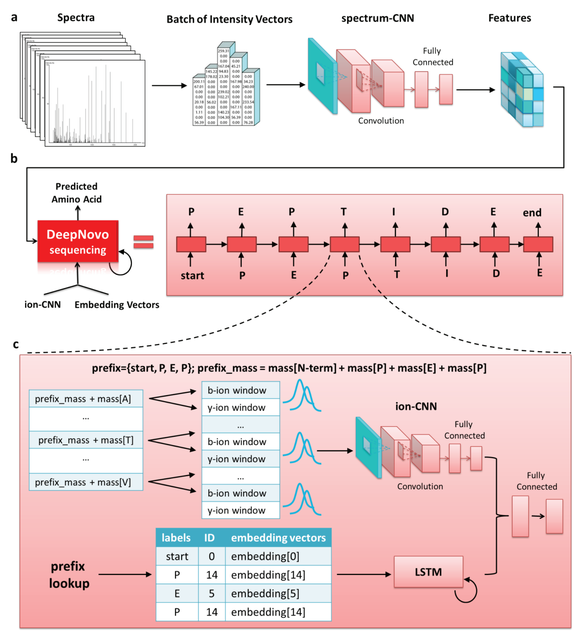
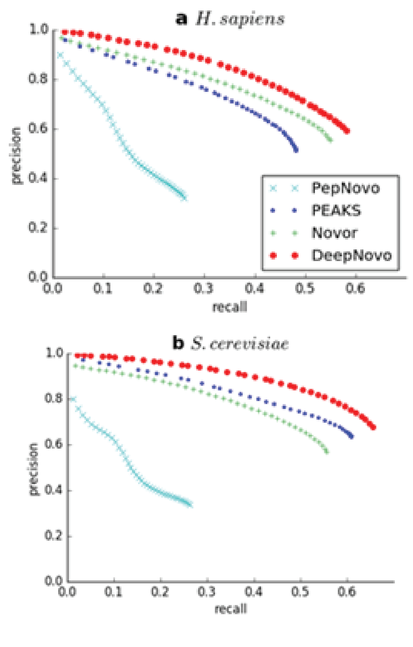
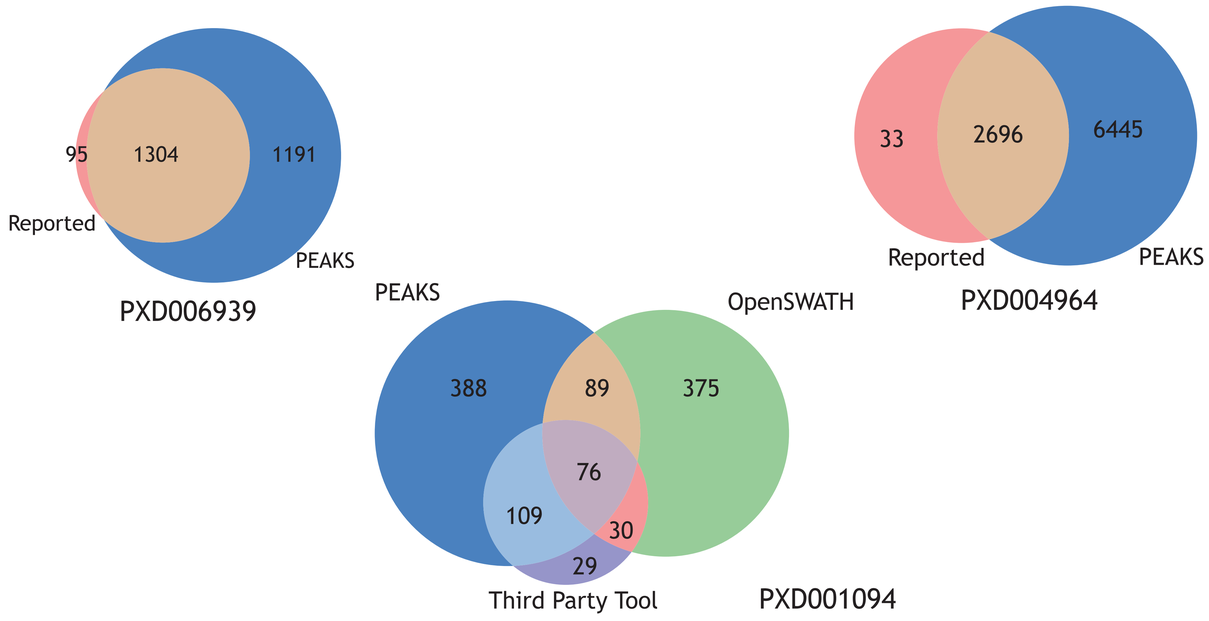


Figure 4. Performance of deep learning system



Results

Two public DDA data sets (PXD006939 [2] and PXD004964 [3]) and on DIA data set (PXD001094 [4]) were tested with the proposed algorithm.



Conclusion

An AI-based data analysis workflow would provide a novel solution for immunopeptidomics with high sensitivity and accuracy.

Reference

[1]. Tran, N.H., Zhang, X., Xin, L., Shan, B., Li, M. Proc Natl Acad Sci U S A. 2017, 114:8247-8252
[2]. Abelin, J. G., Keskin, D. B., Sarkizova, S., Hartigan, C. R., et al., Immunity, 2017, 46:315-326
[3]. Secher, A., Kelstrup, C. D., Conde-Frieboes, K. W., et al., Nat Commun, 2016, 7:11436
[4]. Caron, E., Espona, L., Kowalewski, D.J., Schuster, H., et al., Elife, 2015, 4:07661