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 leukocyte 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 immunopeptides from LC-MS data with both DDA and DIA approaches.

Methods
Since de novo sequencing is a completely unbiased peptide identification workflow, improvement of 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].

Results
1. Increased peptide identifications: By integrating the improved AI-based de novo sequencing into database search, PEAKS X gives higher peptide identification than current published methods. A comparison of peptide identification results for HLA peptidomes of glioblastoma multiforme (GBM) dataset (PXD008127 [2]) was conducted among the publication [2] and PEAKS X. PEAKS X identified a total of 3648 peptide sequences of soluble HLA-bound peptides in patient 11-002-V1* plasma sample, 73% more than analyzed with current database. A test on the public DDA dataset (MSV00080527 [3]) shows that PEAKS X identified 454 peptide sequences by de novo only with 80% ALC for allele HLA-A*01:01. The sequence logos show that the peptides identified by de novo only and the peptides in IEDB database with strong binding have similar patterns.

<table>
<thead>
<tr>
<th>ALC (%)</th>
<th># peptides</th>
<th># in IEDB</th>
</tr>
</thead>
<tbody>
<tr>
<td>80</td>
<td>454</td>
<td>8</td>
</tr>
<tr>
<td>95</td>
<td>33</td>
<td>2</td>
</tr>
</tbody>
</table>

2. Confident peptides found by de novo but not in the database: PEAKS X introduces an advanced AI-based de novo sequencing algorithm, which significantly improves the ability to identify peptides that are not included in current database. A test on the public DDA dataset (MSV00080527 [3]) shows that PEAKS X identified 454 peptide sequences by de novo only with 80% ALC for allele HLA-A*01:01. The sequence logos show that the peptides identified by de novo only and the peptides in IEDB database with strong binding have similar patterns.

3. DIA for HLA peptides: In addition to DDA data analysis, PEAKS X also supports de novo sequencing as well as direct database search without spectral library for the identification of immunopeptides from LC-MS data with DIA approach. A DIA dataset (PXD001904 [4]) was tested to compare the performance of PEAKS X against other methods. In total, PEAKS X identified 662 peptide sequences for Jurkat alleles (HLA-A03, -B07, -B35). Among those 304 high-confidence peptides, which were identified by at least two software tools. 90.1% (274/304) were reported by PEAKS X, 26% and 19.4% more than OpenSWATH (195/304) and a third party tool (215/304), respectively.

4. Chimera spectra for DDA: Feature-based approach was implemented to handle chimeric tandem spectra for co-fragmented peptides, which significantly increases the efficiency of peptide identification. Figure 6 shows a chimeric scan with two features associated and both were identified with PEAKS X.

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 USA. 2017, 114:8247-8252