## **Supporting Information**

## pDeep2 Enables MS/MS Spectrum Prediction for Modified Peptides using Transfer Learning

Wen-Feng Zeng<sup>†,‡,\*</sup>, Xie-Xuan Zhou<sup>‡,¶</sup>, Wen-Jing Zhou<sup>†,‡</sup>, Hao Chi<sup>†,‡</sup>, Jianfeng Zhan<sup>‡,¶</sup>, Si-Min He<sup>†,‡</sup>

- † Key Laboratory of Intelligent Information Processing of Chinese Academy of Sciences (CAS), Institute of Computing Technology (ICT), 100190, CAS, Beijing, China.
- ‡ University of Chinese Academy of Sciences, Beijing, China.
- ¶ State Key Laboratory of Computer Architecture, ICT, 100190, CAS, Beijing, China.
- \* E-mail: zengwenfeng@ict.ac.cn

## **Supplementary Data (excel sheets):**

Supplementary Data 1: pDeep2 for common PTMs.xlsx

Note: The PCC distributions between experimental spectra of common PTMs and the spectra predicted by pDeep2.

**Supplementary Data 2:** transfer v.s. notransfer v.s. pretrained v.s. combined for PT21-PTMs.xlsx Note: We compared different methods on PT21-PTMs (21 PTMs of ProteomeTools). Here "notransfer" refers to train the model without transfer learning; pretrained refers to predict the spectra using the pretrained unmodified model; combined refers to train the model by combining modified and unmodified data.

Supplementary Data 3: comparison of different transfer learning methods.xlsx

Note: Here we compared tuning the first layer ("tune-first"), tuning the last layer ("tune-last"), tuning the first and the last layers ("tune-first-last"), tuning all layers ("tune-all"), and in-turns transfer learning. Except for "tune-last", other transfer learning methods had quite similar performance. The bad performance of "tune-last" may be because we have introduced a PTM feature for each PTM, but the weights of the input layer which connected to the PTM features could not be well trained by only tuning the last layer, showing that fine-tuning the input layer using PTM data is very necessary.

**Supplementary Data 4:** transfer learning with different data scales using PT21-PTMs.xlsx Note: We divided peptides of each modification for training and testing according to the proportion of 8:2, 5:5 and 2:8. It showed that we could obtain quite acceptable accuracies even only 20% of the data were used to train the model. It also showed that the more data we used to train the model, the more accurate the prediction we could get.

Figure S1:

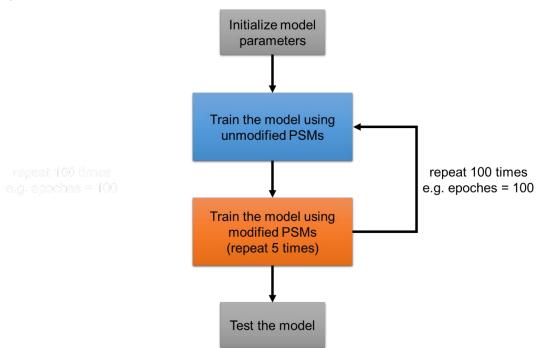


Figure S2. In-turns transfer learning. The model is firstly trained by using unmodified PSMs with 1 epoch (unmodified turn), and then trained by using modified PSMs with 5 epochs (modified turn). In the modified turn, the unmodified model is transferred to fit the modified data; and in the unmodified turn, the model is transferred back to fit the unmodified data. Repeat these two turns n times (here n = 100) to make sure the model has converged.