Deep learning to assist the identification of neoantigens

Dr. Mathias Wilhelm

Dr. Daniel Zolg, Mr. Michael Graber, Mr. Siegfried Gessulat, Mr. Tobias Schmidt, Prof. Bernhard Kuster

Technical University of Munich, Germany

The mass spectrometric identification of tumor specific HLA class I peptides as candidates for immunotherapy is making rapid progress over the last years. However, the yield of such peptides is impaired by the quality of the immunoaffinity purification and technical limitations of the chromatography and mass spectrometric analysis. Especially the latter poses specific statistical challenges because current methods for peptide identification fail to confidently differentiate correct from incorrect matches due to the nature of the short non-tryptic HLA peptides. Within the ProteomeTools project (www.ProteomeTools.org), ~240,000 such HLA class I and II peptides were synthesized and systematically characterized. In conjunction with >500,000 synthetic tryptic peptides these synthetic standards were used to train our deep neural network Prosit, which is now able to prediction the expected tandem mass spectrum of any peptide sequence with very high accuracy (R>0.97). These predictions can now be used to aid the process of HLA peptide identification by comparing the measured and expected tandem mass spectra. This allows the very efficient separation of correct and incorrect matches and thus increases the number of confidently (1% FDR) identified HLA class I peptides by ~2 fold, in turn significantly boosting the chances of finding disease/patient-specific HLA markers and candidates for immunotherapy.

www.premc.org/conferences
pemed2018@premc.org