Predicting Glycan Structure from Tandem Mass Spectrometry via Deep Learning

Description

Glycans constitute the most complicated post-translational modification, modulating protein activity in health and disease. However, structural annotation from tandem mass spectrometry (MS/MS) data bottlenecks in glycomics, preventing high-throughput endeavors and relegating glycomics to a few experts. Trained on a newly curated set of 500,000 annotated MS/MS spectra, the authors present CandyCrunch, a dilated residual neural network predicting glycan structure from raw liquid chromatography–MS/MS data in seconds (top-1 accuracy: 90.3%). They developed an open-access Python-based workflow of raw data conversion and prediction, followed by automated curation and fragment annotation, with predictions recapitulating and extending expert annotation. The authors demonstrate that this can be used for *de novo* annotation, diagnostic fragment identification and high-throughput glycomics. For maximum impact, this entire pipeline is tightly interlaced with a glycowork platform and can be easily tested at:

https://colab.research.google.com/github/BojarLab/CandyCrunch/blob/main/CandyCrunch.ipynb.

The authors envisage to democratize structural glycomics and elucidating the biological roles of glycans.
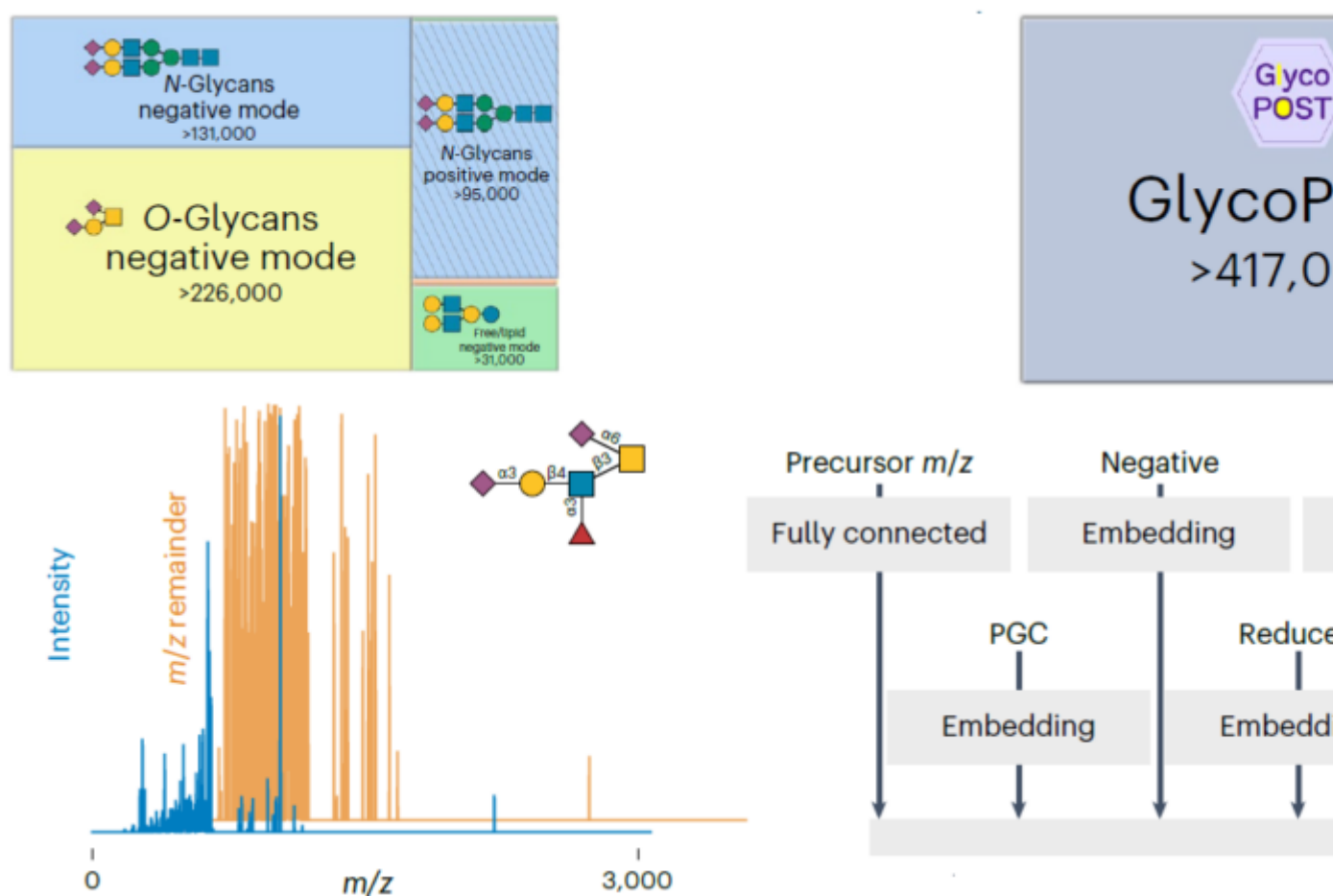
Figure: Predicting glycan structure via deep learning. a,b, Overview of the curated dataset of glycomics LC–MS/MS by glycan class and source .

Category

1. News