Biomarker discovery by automatic annotation of N-glycan species in MALDI-TOF-TOF spectra

Chuan-Yih Yu†, Anoop M. Mayampurath†, Yehia Mechref‡, Milos Novotny‡, Haixu Tang†
†School of Informatics and Computing, ‡Department of Chemistry, Indiana University, Bloomington IN 47405

Overview

We present Multi N-Glycan, a software can annotate different N-glycans by examining mass spectra automatically.

- User- or Pre-defined[1] N-glycan list will be tested for each potential isotopic envelope.
- Each envelope will have a correlation score in overlap and non-overlap scenarios.
- Multiple groups of spectra can be used to discover the significant different glycans between groups.

Software Link: http://mendel.informatics.indiana.edu/~chuyu/MultiNGlycan/

Introduction

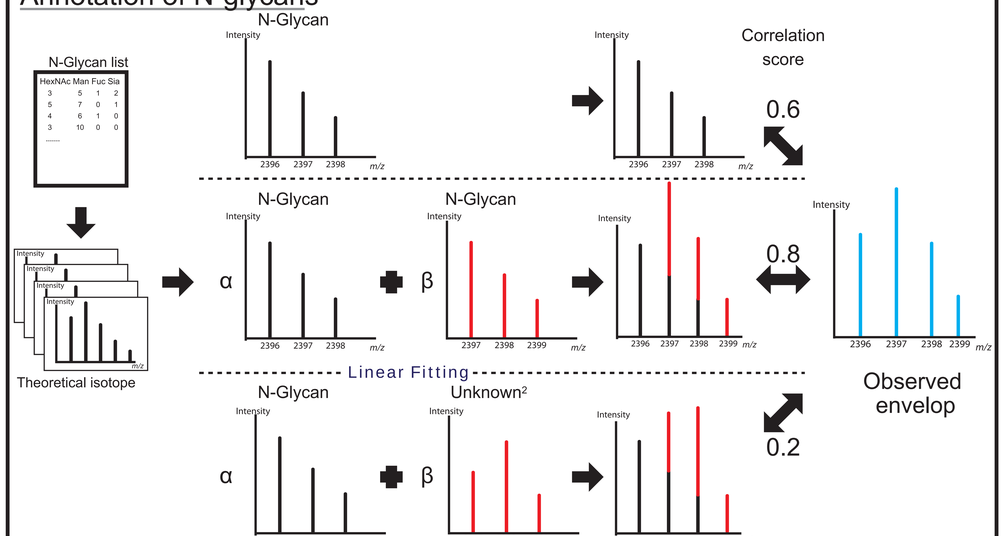
Mass Spectrometry (MS) is a high-throughput and high sensitive technology for the analysis of complex biological samples. The glycoproteins are more complex than proteins with other post-translational modifications, because of the microheterogeneity and diverse tree structure of the glycosylations. As a result, glycans are often cleaved from the glycoproteins and then analyzed separately by MS (e.g. MALDI-TOF-TOF). We combined machine learning, statistical and signal processing techniques to automatically annotate MALDI-TOF-TOF spectra of N-glycans. Our tool can detect the overlapping isotopic envelopes resulted from N-glycans of similar masses, and can also identify N-glycans of significantly different abundances across multiple spectra. Our program can be used in the biomarker discovery by comparing N-glycan profiles from two (or more) groups of samples (e.g. disease vs. normal).

Conclusions

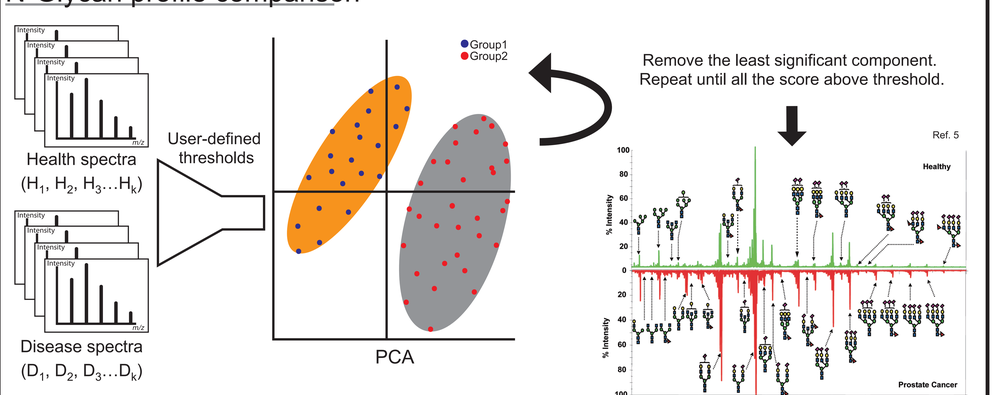
Our tool, Multi N-Glycan, can rapidly annotate and compare N-Glycan in MALDI-TOF spectra. It uses traditional biomarker discovery protocol with pre-annotated spectra; and comes out a high-quality and reliable result for preliminary screening.

Methods

Annotation of N-glycans



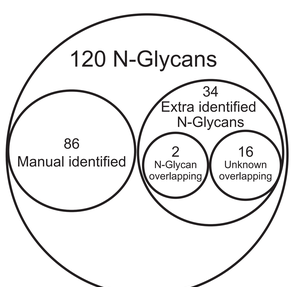
N-Glycan profile comparison

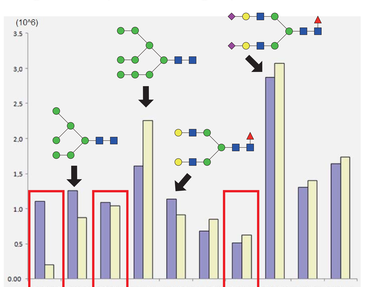


Results

In a MALDI-TOF-TOF experiment, we only can identify 86 N-glycans in manual inspection, but the software can identify an extra 34 N-glycans. Among these 120 N-glycans, 40 N-glycans received fitting scores above 0.7; 16 N-glycans were detected to have the isotopic envelope overlapping with that of another unknown molecule; and two N-glycans were detected to have overlapping isotopic envelopes with each other.(Buttom left)

We also compare our result with previous study[3]. Three out of ten N-glycans have the similar trend. Four out of seven the rest N-glycans found structures in the CFG database. These glycans might be considerable potential biomarkers for in addition investigation. (Buttom right)





Future work

We would like to extend this approach on O-glycan profiling, and also perform *de novo* glycan sequencing on reported glycan if the data contain MS/MS spectrum. This can help to differentiate microheterogeneity of the glycosylation.

References

1. Krambeck, F.J. and M.J. Betenbaugh, A mathematical model of N-linked glycosylation. Biotechnology & Bioengineering, 2005
2. Rockwood, A., S. Van Orden, and R. Smith, Rapid Calculation of Isotope Distributions. Analytical Chemistry,1995
3. Zhiqun T., et al., Identification of N-Glycan Serum Markers Associated with Hepatocellular Carcinoma from Mass Spectrometry Data. Journal of Proteome Research, 2009
4. Hastie, T., et al., 'Gene shaving' as a method for identifying distinct sets of genes with similar expression patterns. Genome Biology, 2000
5. Zuzana K., et al., Alterations in the serum glycome due to metastatic prostate cancer. Journal of Proteome Research, 2007

Acknowledgements

This work is funded by NCI/NIH (grant number 1 U01 A128535-01).