Classification of cirrhotic patient samples based on imaging MS of multiplexed N-glycan markers in biofluids

Nico Verbeeck[a], Danielle A. Scott[b], Alice Ly[a], Stephen Castellino[b], Richard R. Drake[b,c], Peggi M. Angel[b,c] , Anand S. Mehta[b,c] and Marc Claesen[a]

[a] Aspect Analytics NV, Gent, Belgium;
[b] Glycopath Inc. Charleston, SC;
[c] Medical University of South Carolina, Charleston, SC

Introduction

Alterations in glycosylation patterns of immunoglobulin G (IgG) in patient serum and plasma have been reported for different diseases including cancer, rheumatoid arthritis, and liver disease. These changes in N-glycan signatures have been proposed for diagnostic use, however, determining protein-specific N-glycosylation changes using traditional methods are laborious.

Glycopath recently developed the GlycoTyper assay which is a streamlined antibody capture slide array approach to directly profile N-glycans of captured serum glycoproteins including IgG. This method needs only a few microliters of serum and utilizes a simplified processing protocol that requires no purification or sugar modifications prior to analysis. In this method, antibody captured glycoproteins are treated with peptide N-glycosidase F (PNGase F) to release N-glycans for detection by MALDI imaging mass spectrometry (IMS). We recently successfully utilized this novel method to examine the glycosylation of total IgG, as well as IgG1, IgG2, IgG3 and IgG4 in patients with fibrosis and cirrhosis [1].

In this work, we present a proof-of-concept data analysis workflow for the N-glycans detected by our protein immunocapture method, consisting of a system for IMS data and metadata inventorization and analysis, automated extraction of IMS spot data and machine learning methods for exploration and classification of the resulting N-glycan IMS data.

N-Glycans IMS data collection

The GlycoTyper workflow is shown in Fig. 1 below. Capture antibodies (IgG, IgG1-4) were spotted onto nitrocellulose-coated slides, dried overnight, then rinsed in buffers to remove unbound protein. After adding patient serum, slides were incubated at room temperature for 2 hours in a humidity chamber and washed to remove residual salt. PNGaseF (0.1µg/µL) was applied using an automated sprayer and incubated overnight at 37°C in a humidity chamber. 7mg/ml CHCA matrix was applied using the same automated sprayer. Data were acquired using a solariX 7T FT-ICR in positive ion mode over m/z 600–4000.

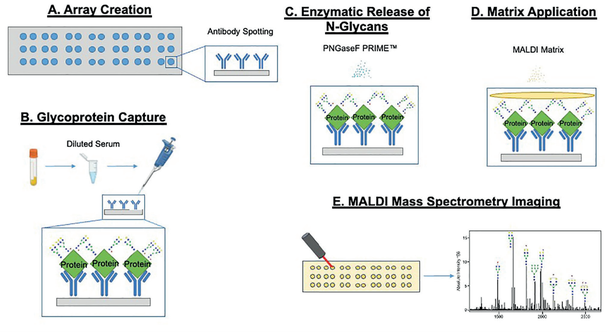
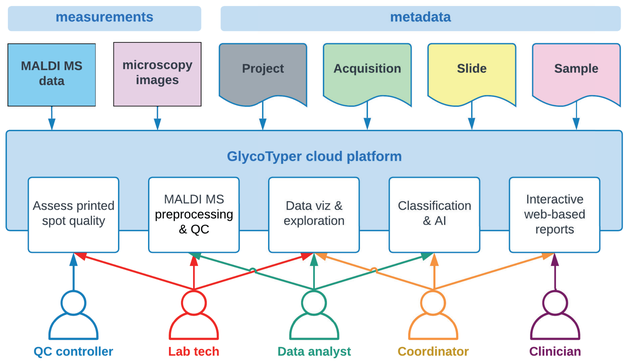


Fig 1: GlycoTyper workflow for the glycan analysis of antibody captured glycoproteins. (A) Creation of antibody array using desired antibodies immobilized on nitrocellulose coated slides. (B) Antibody array is incubated with diluted serum. (C) Slides are washed after protein capture using a MS compatible detergent before glycan are released by application of a thin coating of recombinant PNGase F. (D, E) MALDI matrix is applied and slide is imaged using IMS (D, E). Image adapted from Scott et al. [1]

GlycoTyper cloud platform

In order bring the GlycoTyper workflow to a high-throughput setting, a dedicated software pipeline is required to automate IMS data extraction, combine MS measurements with patient metadata, and provide data-driven QC and robust machine learning methods to ensure generalizability and reproducibility of the results.

The envisioned platform builds on Aspect Analytics' Nexus platform for IMS data management and analysis. This cloud-based platform provides scalable data storage and a flexible metadata system that can be used to inventorize and query experimental data. A web-based end-user interface allows users to interactively explore and analyze MSI data, and readily share results across teams and end-users alike.



References

[1] Scott Danielle A., Wang Mengjun, Grauzam Stephane, Pippin Sarah, Black Alyson, Angel Peggi M., Drake Richard R., Castellino Stephen, Kono Yuko, Rockey Don C., Mehta Anand S. ***GlycoFibroTyper: A Novel Method for the Glycan Analysis of IgG and the Development of a Biomarker Signature of Liver Fibrosis.*** *2022, Frontiers in Immunology, 13.*

Data analysis pipeline

Our proof-of-concept data analysis workflow is shown below. IMS data were exported using FlexImaging v5.0 to the .sqlite format, after which they were imported into our custom Python-based data processing pipeline.

Capture slide array spots were automatically detected using a combination of computer vision techniques and ROI metadata. For each of the spots, spectra were extracted, and automatically grouped per patient and glycoprotein combination using provided metadata, and stored in a measurement database.

The collected spectra were TIC normalized and realigned based on known reference peaks and the mean spectrum across all spectra. Peaks for 11 previously identified N-glycans were extracted across experiments.

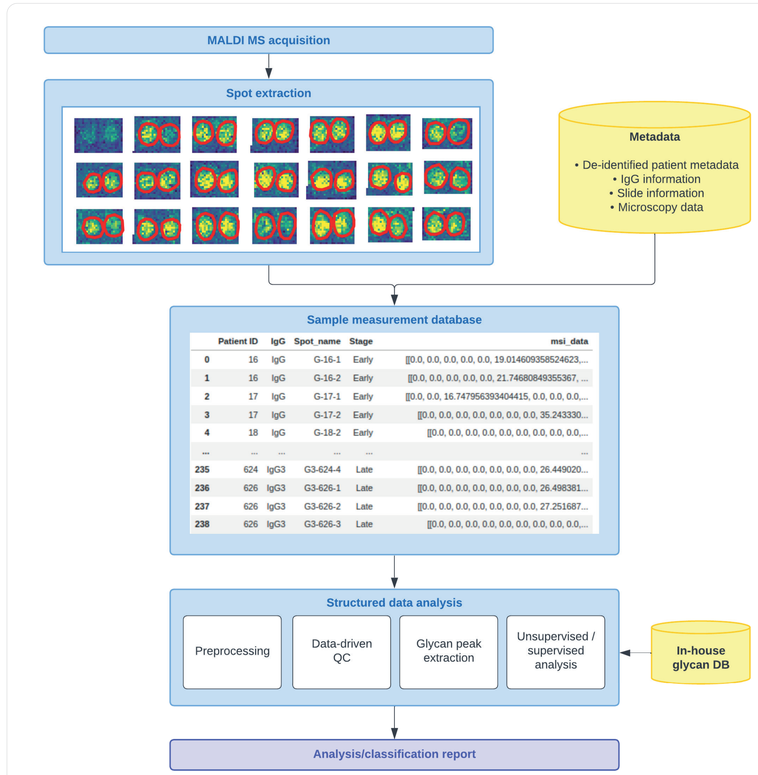


Fig 2: Proof-of-concept data analysis workflow

Preliminary data

We applied this automated workflow for a preliminary analysis of a cohort of liver disease patients, where serum from 10 late stage cirrhotic patients was compared to that of 10 early stage patients. Multiplexed arrays were created capturing each of 4 different glycoproteins IgG and IgG1-3. The multiplexed array included 2 technical replicates for the early stage samples, and 4 technical replicates for the late stage samples for each of the glycoproteins. Spectra per spot were merged by calculating the mean over the extracted ROI, resulting in a total of 240 mean spot spectra (1 spectrum per glycoprotein/patient/technical replicate combination).

www.aspect-analytics.com | www.glycopath.com

AGENTSCHAP INNOVEREN & ONDERNEMEN · Vlaanderen is ondernemen · eit Accelerated by Health · EIT Health is supported by the EIT, a body of the European Union

Data analysis

Unsupervised explorative analysis using non-negative matrix factorization (NMF) and principal component analysis (PCA) was performed on the mean spot spectra, in order to visualize and investigate underlying trends in the data and identify potential batch effects. Figure 3 shows the first three components of the NMF analysis on the 11 selected glycan peaks. Overall we observe a clear difference in grouping between the early and late stage samples.

Figure 4 shows the distribution of intensities of the 11 selected N-glycans for IgG for early and late stage samples, showing an increase in intensity for several of the glycan peaks.

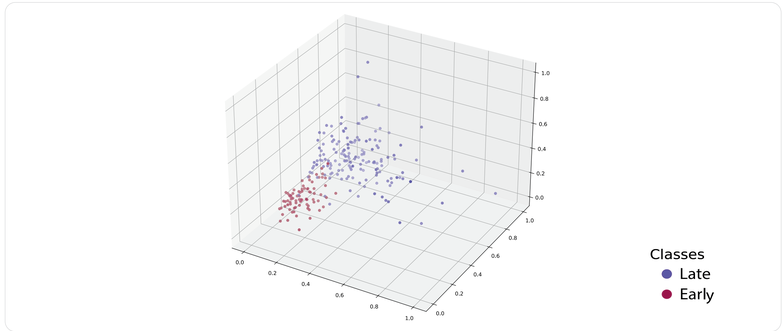


Fig 3: NMF analysis of combined liver disease data (red=early, purple=late).

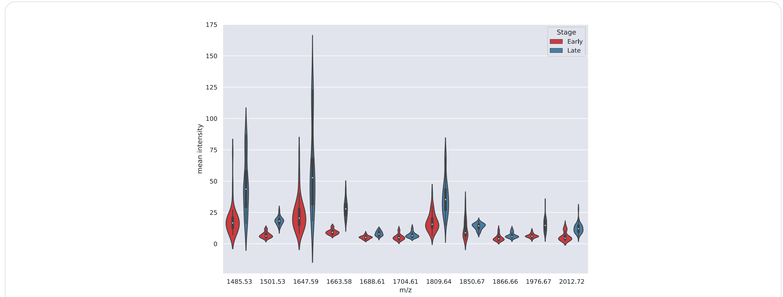


Fig. 4: intensities across spots for IgG N-glycan peaks.

A linear SVM model was made for each of the different glycoproteins separately, as well as on the combined data over all glycoproteins, resulting in a total of 6 different models. Clustered cross validation was used to measure the performance of the different models and technical replicates were used to estimate model robustness. The resulting models showed good classification results, confidently distinguishing between early and late samples.

The model combining data from the IgG and IgG 1-3 arrays achieved an **area under the ROC of 1.00** and **area under the PR curve of 1.00**, showing **classification perfectly matching the clinical assessment/state** of samples in this limited sample dataset based on the extracted N-glycan peaks. In follow-up work additional samples will be included to further establish generalizability of these results, as per Scott et al [1]. Moreover, improvements in the GlycoTyper array will futher ensure experimental reproducibility and as well as provide additional opportunities for QC

Conclusion

In this work, we present a proof-of-concept GlycoTyper workflow covering automated metadata management, spot extraction and data analysis. A linear SVM classification model on the combined N-glycan data of IgG and IgG1-3 showed very good classification on early and late liver disease patients. Further developments will embed this workflow into a cloud-based platform aimed at enabling the GlycoTyper for high-throughput screening.