TEAM 4 Project Report

Tiialotta Tohmola, Hao Li, Oluwatosin Abe

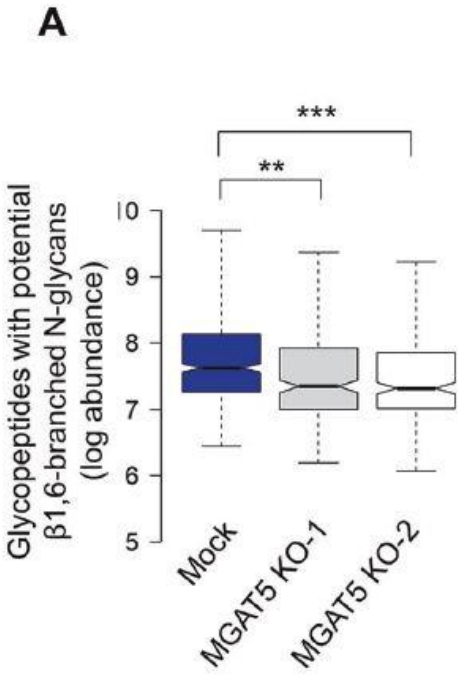
30.5.2025

1. Project Background

Due to our team members' background in head and neck cancers and glycoproteomics, we decided to choose an article focusing on PD-L1 N-glycans in immunotherapy responses in patients with recurrent/metastatic HNSCC (https://www.nature.com/articles/s41389-024-00532-3#Sec25). N-glycosylation is one of the most common post-translational modifications of proteins and often altered in different states of stress, including cancer. The article aimed to recognize a biomarker for patients with a positive response to immunotherapy with head and neck squamous cell carcinoma. By focusing on the key regulator of N-glycan synthesis, MGAT5, known to dysregulate in cancer progression, they tried to recognize if the expression of it changes between wild type and knock out cells or if its' protein substrates are linked to specific protein pathways responsible of the different immunotherapy responses.

2. Summary of the Original Study

The authors observed elevated levels of MGAT5 expression in HNSCC tumors (Fig 1). Global analysis of cellular N-glycoproteins using LC-MS/MS was performed and found significant decreases of MGAT5 in the knockout cells. The article found 163 potential protein substrates of MGAT5 where the fold change between the wild type and knock out samples were over two. These substrate proteins were analyzed with gene ontology (GO) pathway analysis and found to be linked to for example T-cell proliferation and activation pathways. PD-L1 was found as one of the MGAT5 substrates and MGAT5-positive tumors had a better response to the anti-PD-1 therapy than those with MGAT5-negative tumors.



3. Team Findings and Comparative analysis

The data with an identifier code PXD045506 was uploaded from PRIDE to R-studio and the file was unzipped. The mzID files were converted to PSMs (Fig 2) and a lot of low score PSMs (from 20-40) were excluded from the dataset. Ranks higher than 1 were also excluded from the dataset (n=5). Total number of filtered PSMs were 87451 (Fig 3).

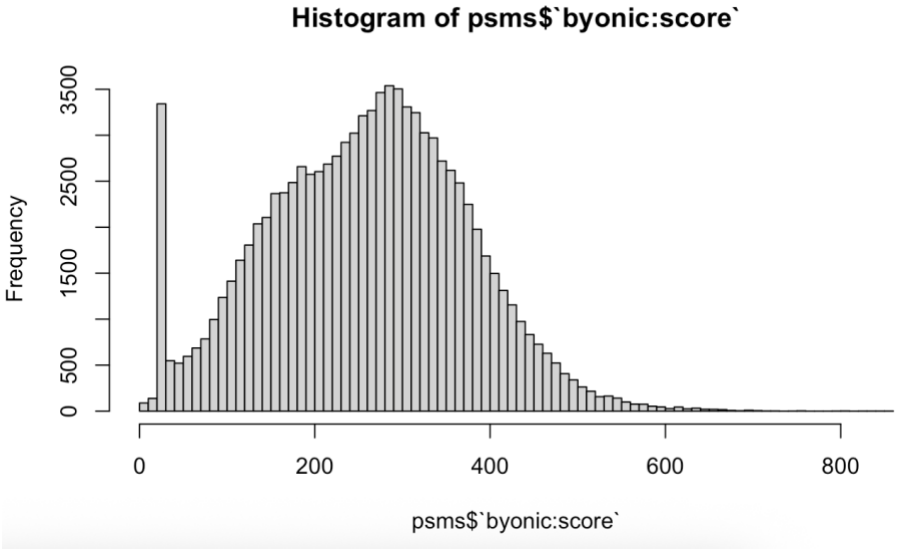


Fig 2. All PSMs plotted to histogram

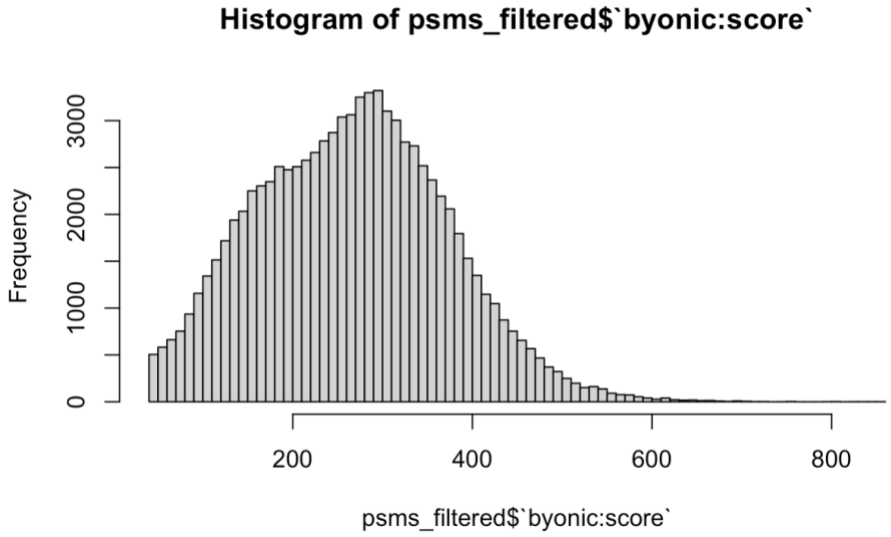


Fig 3. The filtered PSMs where low scores and ranks over 1 were removed plotted to histogram.

The raw data was uploaded to MaxQuant (Version 2.6.8.0) and produced 13 365 unique peptide sequences and 3 091 unique leading razor proteins. After MaxQuant analysis, 2073 decoy PSMs were found. The data was also uploaded to FragPipe to search for possible glycans, but unfortunately it didn't produce any mathces. With QFeature aggregation in R-studio, the data was normalized (Fig 4) and possible missing values were imputed.

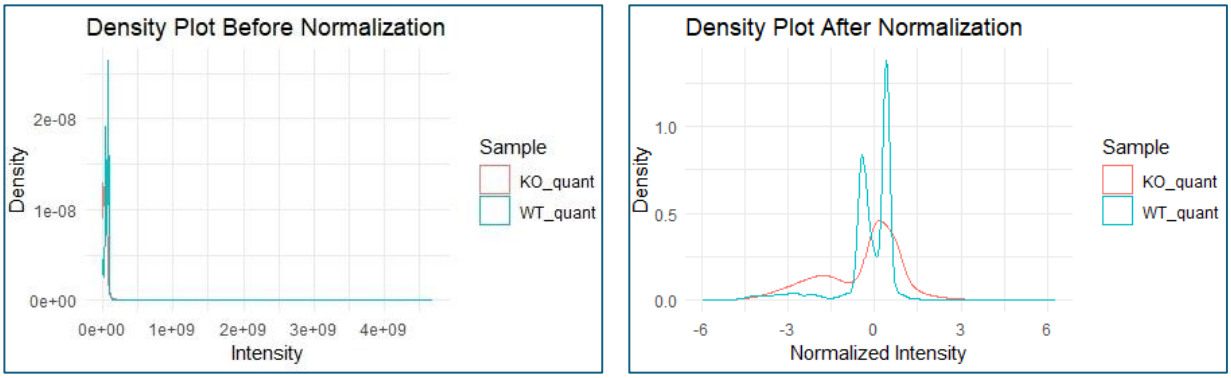


Fig 4. Density plot of the normalization

With an analysis of the fold changes inside the wild type and knock-out samples, we were able to get similar results compared to the original article (Fig 1 & Fig 5). Based on our

data, the protein expressions in the wild type cells were increased compared to the MGAT5 knock-out cells. Our results couldn't reproduce the original findings of MGAT5 and its' substrate proteins being found possible biomarkers for positive responses immunotherapy in HNSCC patients.

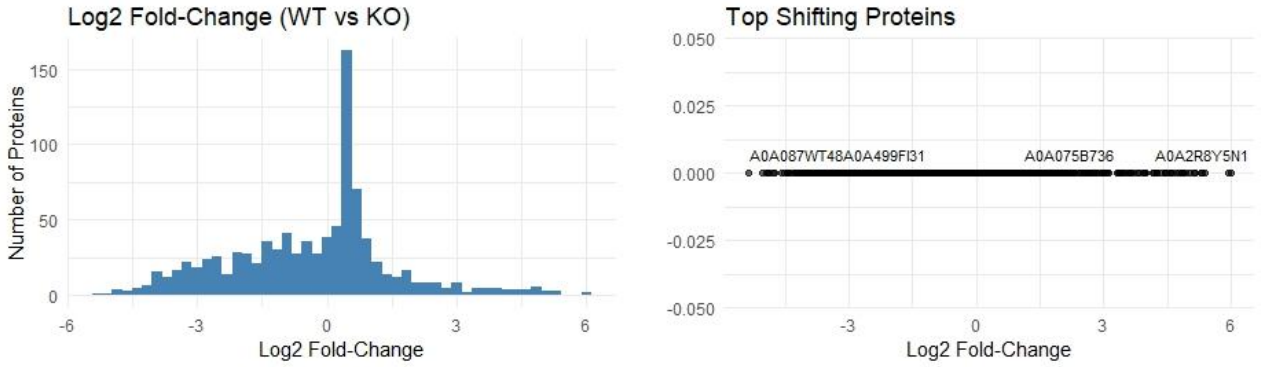


Fig 5. The logaritmic fold changes between the wild type and knock-out samples and the top shifting proteins.

The analysis of the raw files and statistical analysis faced some challenges due to following reasons: We couldn't find decoys in the original raw data downloaded from PRIDE with R-studio. Only after MaxQuant analysis, we were able to find the information about decoys that the software provided. In addition, a further statistical analysis of the data was impossible based on the number of files the authors had provided to PRIDE. Since only one raw file was uploaded for each group (wild type and knock-out) and no replicants were available, our statistical analysis was based only on the difference of protein fold changes in the samples.