MSProt2025_Team 2 – Final Report & Interpretation

1. Background of dataset and study

The study *"Proteomic Landscape of Pattern Triggered Immunity in the Arabidopsis Leaf Apoplast"* (preprint DOI: 10.1101/2025.02.06.636724, dataset ID: PXD060654) investigated the proteomic changes in the apoplast of *Arabidopsis thaliana* leaves during pattern-triggered immunity (PTI). PTI is initiated when plants recognize conserved microbial features, such as the flg22 (22 amino acid bacterial flagellin epitope) peptide derived from bacterial flagellin, which was used in this study. To analyse these changes, the researchers extracted apoplastic washing fluid (AWF) from four individual *A. Thaliana* plants leaves treated with flg22. These were compared to four individual plants leaves used as mock-treated control, which were treated with the same buffer (dimethyl sulfoxide) but without flg22, and which served as negative controls. Apoplast, the space between cell membranes and cell wall, has secreted proteins, peptides and specialized metabolites making it ideal to study PTI with MS. The researchers also ensured minimal cytoplasmic contamination with enzymatic assays and immunoblots. The proteins in the AWF were then subjected to LC-MS/MS for identification and quantification. The study aimed to characterize the proteomic alterations in the apoplast during PTI and identify proteins that are differentially abundant upon flg22 treatment.

Protein identification and quantification was performed by using Proteome Discoverer software. False discovery rate (FDR) control was applied and at least one peptide should have mapped with the protein ID. Differentially expressed proteins were identified using t tests and Benjamini- Hochberg adjustments with Log2FC >=1. Gene Ontology (GO) enrichment analysis determined biological processes overrepresented among proteins that increased or decreased in abundance during PTI. Known extracellular vesicle (EV) marker proteins were mapped and quantified to explore their enrichment under immune activation.

2. Key results of original publication

Principal Component Analysis (PCA) was used to examine the distribution of variation between flg22-treated and mock- treated samples (**Fig. 1A**). The first two components, Dim1 and Dim2, explained 27.9% and 25.9% of the total variance, together explaining more than 50% of the total variation in the data. The PCA plot indicated distinct profiles between treatment conditions but revealed that two samples (1 and 5) were far from the rest within their condition and could be biological or technical outliers. A total of 108 proteins were significantly enriched in flg22-treated samples (**Fig. 1B**), classified into six groups: (1) Receptor-Like Kinases/Proteins, (2) redox proteins, (3) hydrolytic enzymes, (4) peptides, (5) extracellular vesicles, and (6) others. The heatmap (**Fig. 1B**) shows the abundance of enriched proteins, ranked by their $\log_2$ fold change.

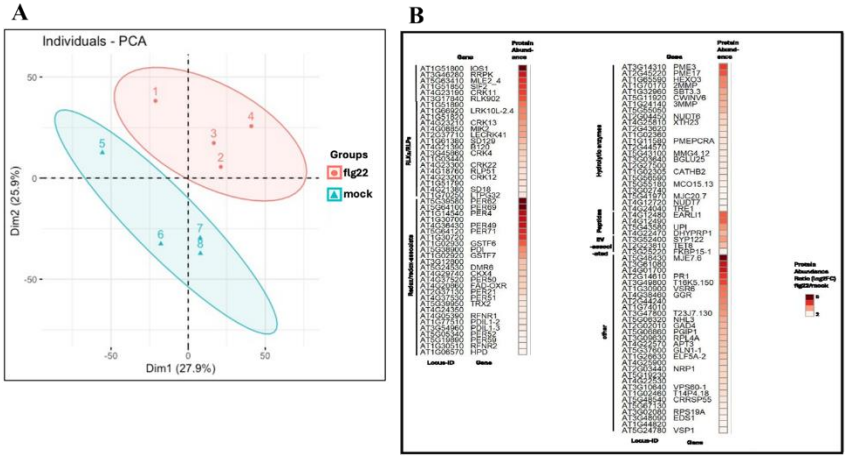


Figure 1. A) Principal Component Analysis (PCA) of flg22 (1-4) and mock (5-8) samples. **B)** Heatmap presents flg22-enriched protein abundances on a red scale. A total of 108 significantly enriched (Log2FC > 1, $p < 0.05$) proteins identified and manually categorized into key functional groups.

3. Our analysis and findings

We performed the analysis on the mzID file provided in the PRIDE by the authors and also performed peptide spectrum matching in MaxQuant using the raw files provided in the PRIDE. We also used the cRAP contaminant fasta that was used in the original publication to resemble the analysis as much as possible.

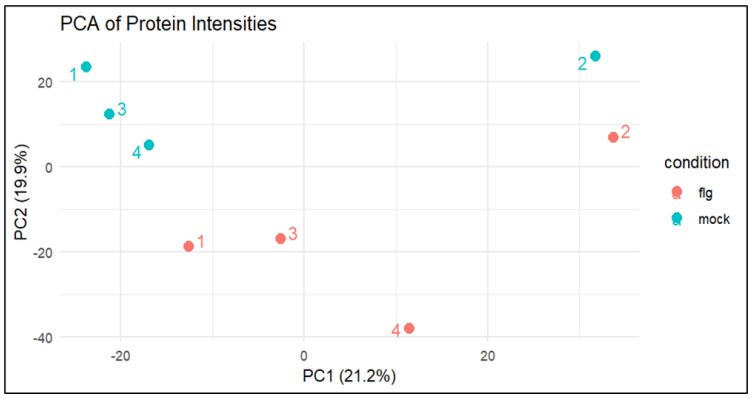
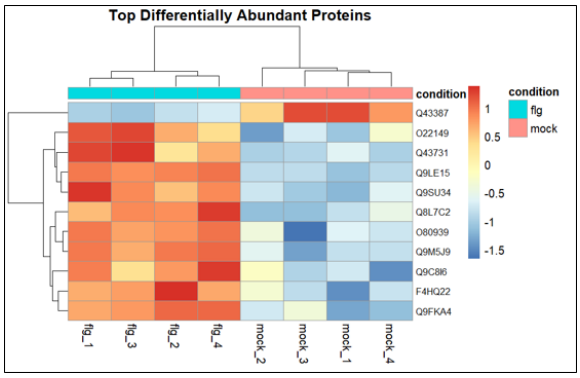


Figure 2: PCA plot of protein intensities comparing the different replicates in the mock vs flg22 conditions

A



B

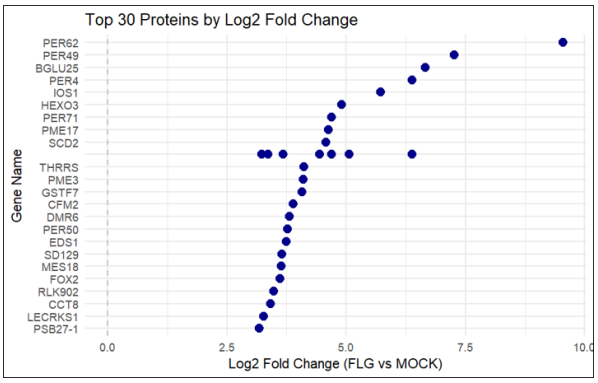


Figure 3. A) DIfferentially expressed proteins obtained in the flg22 vs mock condition using Limma analysis **B)** Top 30 differentially expressed protein with Log(fold change)>1

Analysis	Original Finding (From the mzID file/paper)	Our Finding (From our analysis of raw data in MaxQuant)
Number of peptides	21107	23964
Number of proteins	3288	3399
Differentially expressed proteins	108	164 (11)

We have analysed the log fold change differences from two different files that were obtained from MaxQuant, the evidence file and the protein groups file. From the evidence file we compared the flg22 vs mock intensities in the replicates. We constructed PCA plot (**Fig. 2**), which resemble the result obtained in the original publication (**Fig. 1A**) with one replicate in each condition being distant from other replicates in the group. We did median normalization, knn imputation, and differential expression analysis using Limma. We found only 11 significantly differentially expressed proteins (**Fig. 3A**). Such a low value could be because of our differences in the analysis methods and there were a lot of missing values that were imputed by knn. From the protein groups file we used the MaxQuant generated normalized intensities and q values, and we found 164 differentially expressed proteins with Log2fold change >1 (**Fig.**

3B). Thus, the method we chose to identify these differentially expressed proteins produced varied results. However, it was reassuring that the some of the proteins identified by our analysis were also top hits in the original publication. We thus need more details in the publications about their analysis to be able to replicate their findings.