OmixAI SINCE 2025

# PROTEOMIC ANALYSIS REPORT

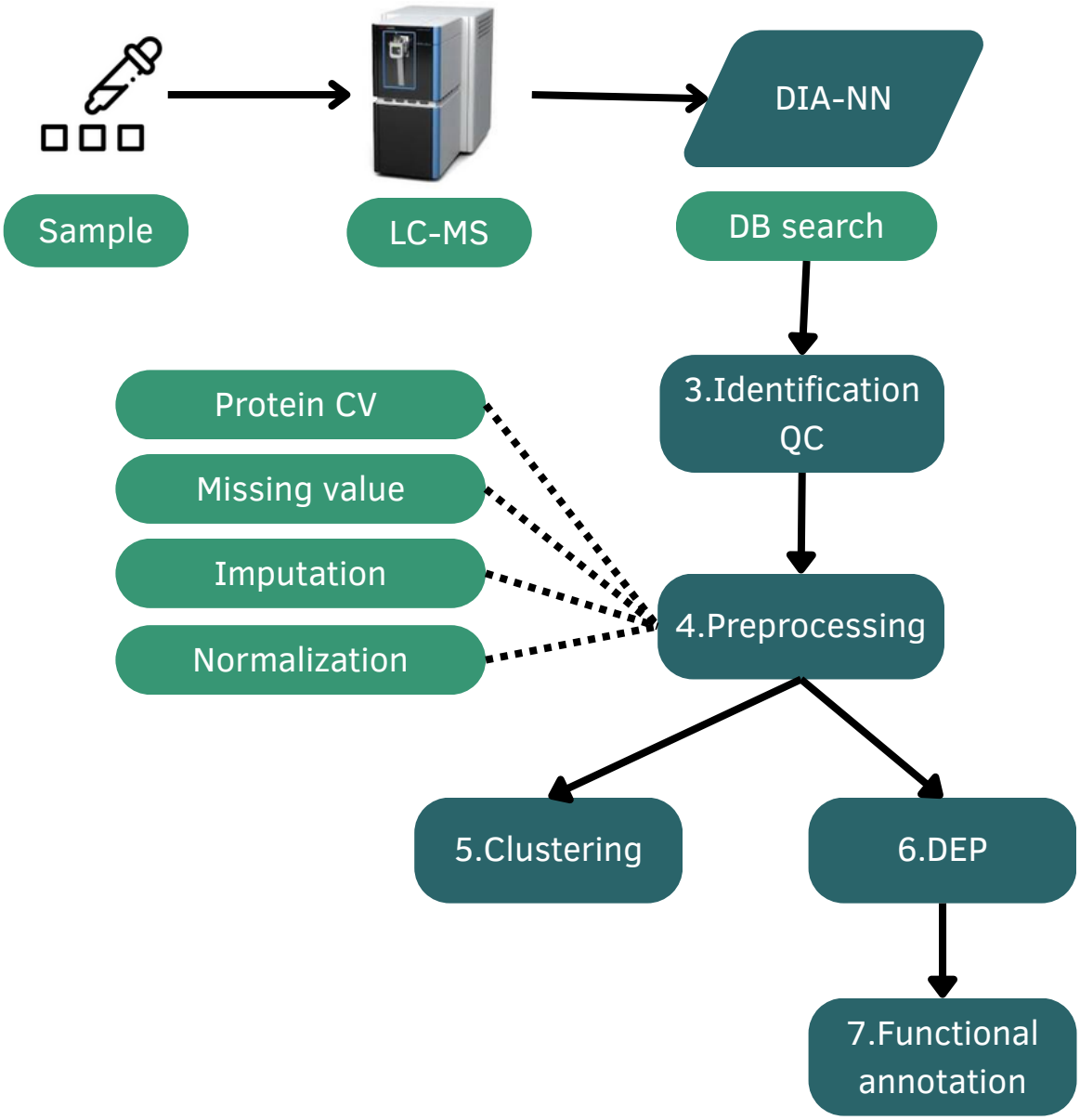
| | |
|---|---|
| **Project Name** | SEER Plasma |
| **Analysis Type** | Global proteome |
| **Species** | Human |
| **Client** | 이유민 |
| **Date** | 2026-01-07 |

www.omixai.com
omixai@omixai.com
070-7721-0553

# Table of Contents

# 1. Introduction

- Samples were acquired on an **Astral Orbitrap**.
- Peptide and protein identification and quantification were performed using **DIA-NN**.
- Downstream analyses were conducted according to the workflow described below.
- Detailed sample information is provided in the next chapter.



## 1-1. Output directory file description

All experimental results are housed within the <out_dir> directory.

<out_dir>
├── final_report.pdf [Final analysis report file]
├── 3. Identifcation
│ ├──3-1_precursorID.png [identified precursors/sample]
│ ├──3-2_proteinID.png [identified proteins/sample]
│ ├──3-3_geneID.png [identified genes/sample]
│ ├──gene_fig.png [identified genes/sample]
├── 4. quality_control
│ ├──4-1_cv.png [protein quantification CV values with groups]
│ ├──4-2_missing_value.png [missing values/sample]
│ ├──4-3_normalization.png [before/after expression density plot]
│ ├──4-3_sample_dist_box.png [before/after expresion boxplot/group]
├── 5. clustering
│ ├──4-1_pca.png [pca based on expression values]
│ ├──4-2_hierarchical_clustering.png [clustering plot]
├── 6. dep
│ ├── 6-1.dep_summary
│ ├──DE_summary.csv [no.of DEPs / groups]
│ ├──DEP_#treatment_vs_#control.csv [DEPs /groups]
│ ├──#treatment_vs_#control.csv [total gene list/groups]
│ ├── 6-2_volcano
│ ├──#treatment_vs_#control.png [volcano plots/groups]
│ ├── 6-3. heatmap
│ ├──dep_heat_map.png [dep foldchange heatmap]
│ ├──#target_precursor_heatmap.png [if applicable]
├── 7.functional_annotation
│ ├──functional_annotaion_bar.png and bubble.png

# 2. Sample Overview

• Total number of samples: 8

## 2-1. Sample Information

• Sample: the basic unit of data (e.g. one run).

• Group: the defined categories used as the comparison units in DEP.

• N(samples): the count of individual smaples within a specific sample groups.

| Group | N(samples) | Sample |
|---|---|---|
| Test1 | 4 | SEER_Std_1, SEER_Std_2, SEER_Std_3, SEER_Std_4 |
| Test2 | 4 | SEER_Std_5, SEER_Std_6, SEER_Std_7, SEER_Std_8 |

## 2-2. Contrast Information

Comparison units for the DEP analysis:

• Control: baseline or reference condition.

• Treatment: the experimental intervention being tested.

• Multigroup: a higher-level categorization used to cluster similar comparison groups.

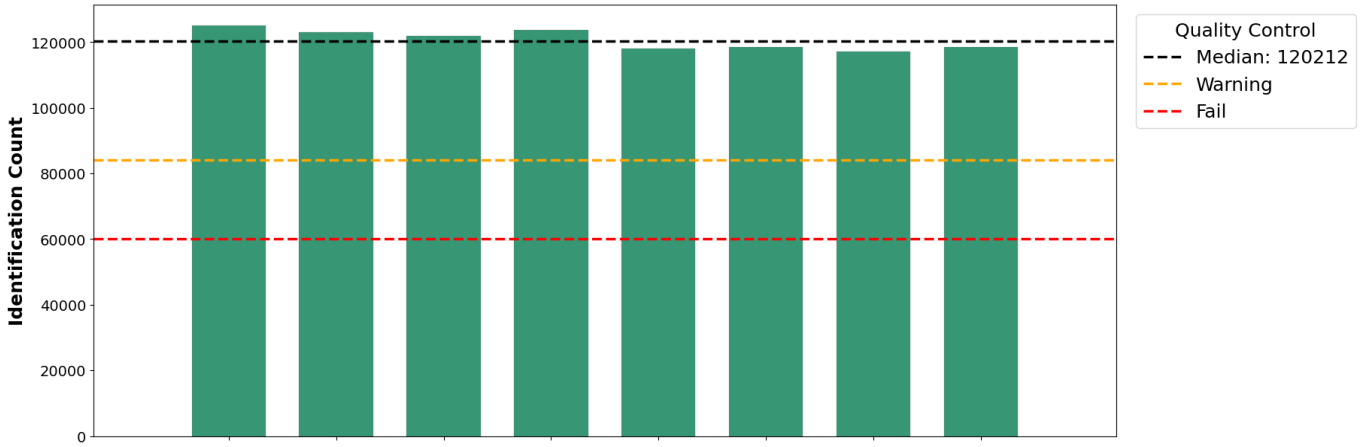
• Target: client-defined subset of analytes(e.g. genes).

| Control | Treatment | Target |
|---|---|---|
| Test1 | Test2 | FAT3,LMOD1 |

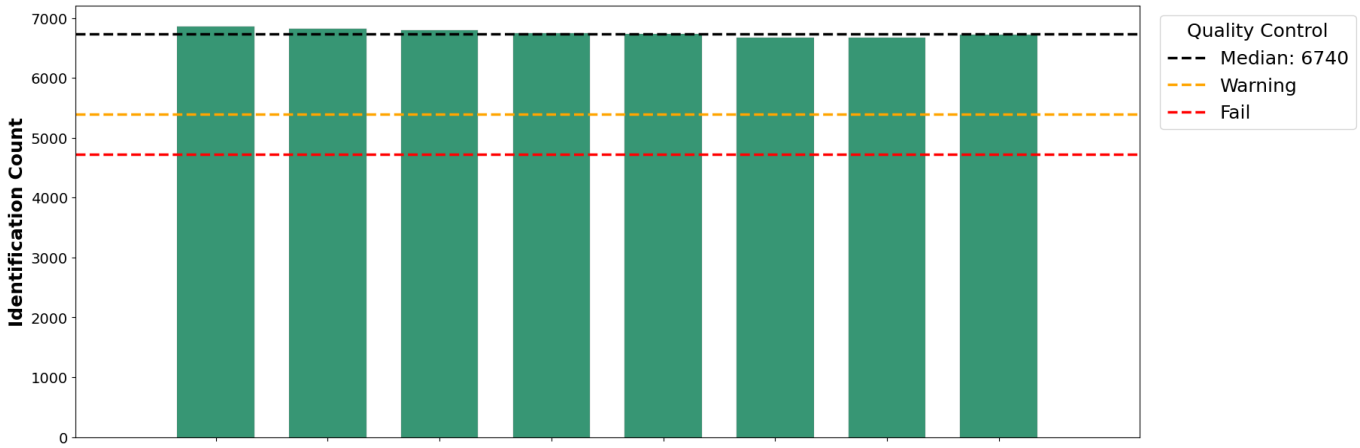
# 3. Identification

The total number of identified features (e.g. protein) represents a composite metric reflecting the cumulative efficiency of the analytical workflow, including sample loading precision and the intensity profiles of both MS1 and MS2 spectra. Maintaining these identification counts within established thresholds serves as a primary benchmark for technical consistency and ensures the high-fidelity nature of the acquired data.

## 3-1. Identified precursors per sample




## 3-2. Identified proteins per sample



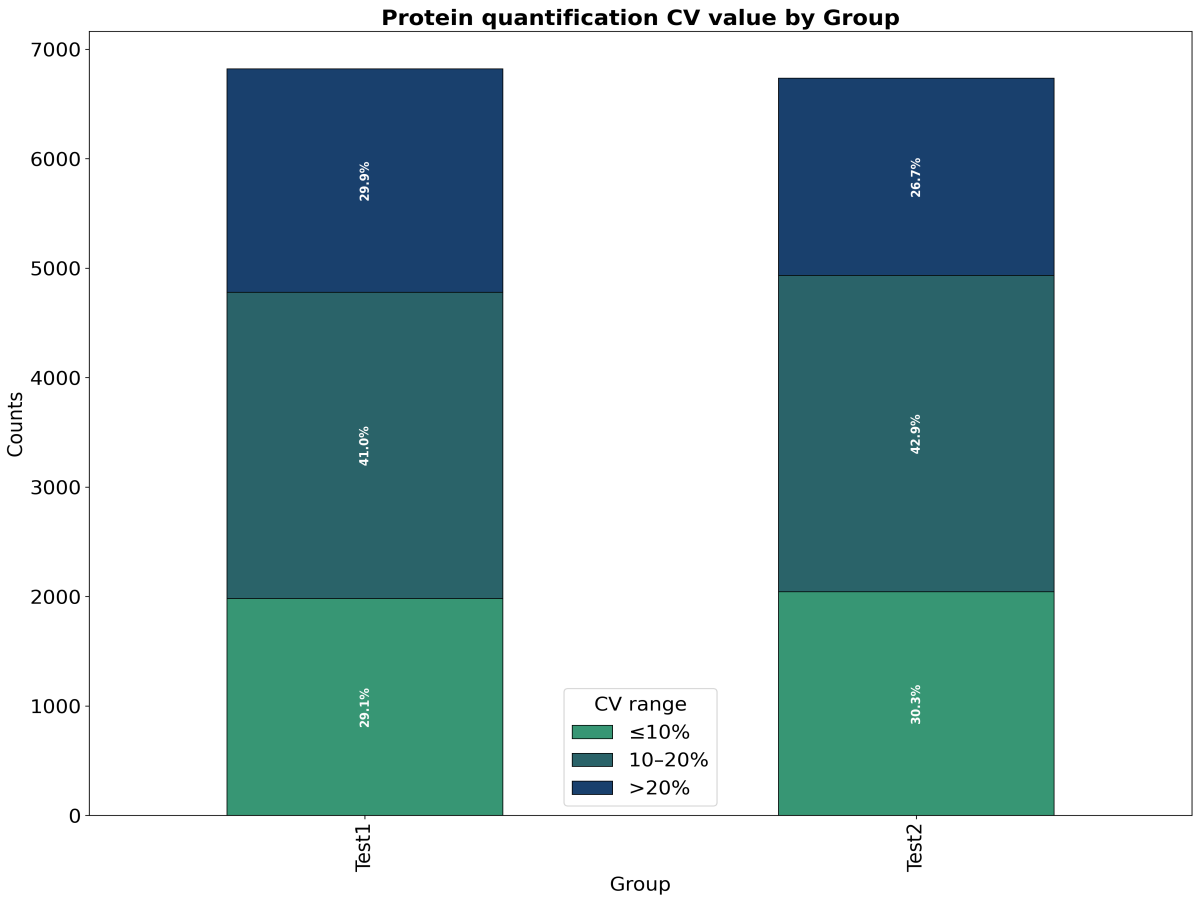
# 4. Preprocessing

## 4-1. CV

**Preprocessing and Quantitative Quality Control:**

Preprocessing is a critical step to ensure data integrity for quantitative analysis. When a group (condition) contains two or more replicates, a **low and stable Coefficient of Variation (CV)** within the group is indicative of high-quality experimental execution.

For groups with excessively high CV values, we recommend **data refinement or sample exclusion** based on QC metrics to maintain statistical robustness. Please note that CV calculation is not applicable for groups with only a single sample.



## 4-2. Missing Value Analysis

**Origin of Missing Values:**

Missing values typically occur due to **differential identification counts** between samples or when quantitative signals are **not detected during LC-MS/MS acquisition** (e.g., values below the limit of detection). While inherent in large-scale profiling, these are managed through rigorous filtering and normalization to ensure statistical reliability.

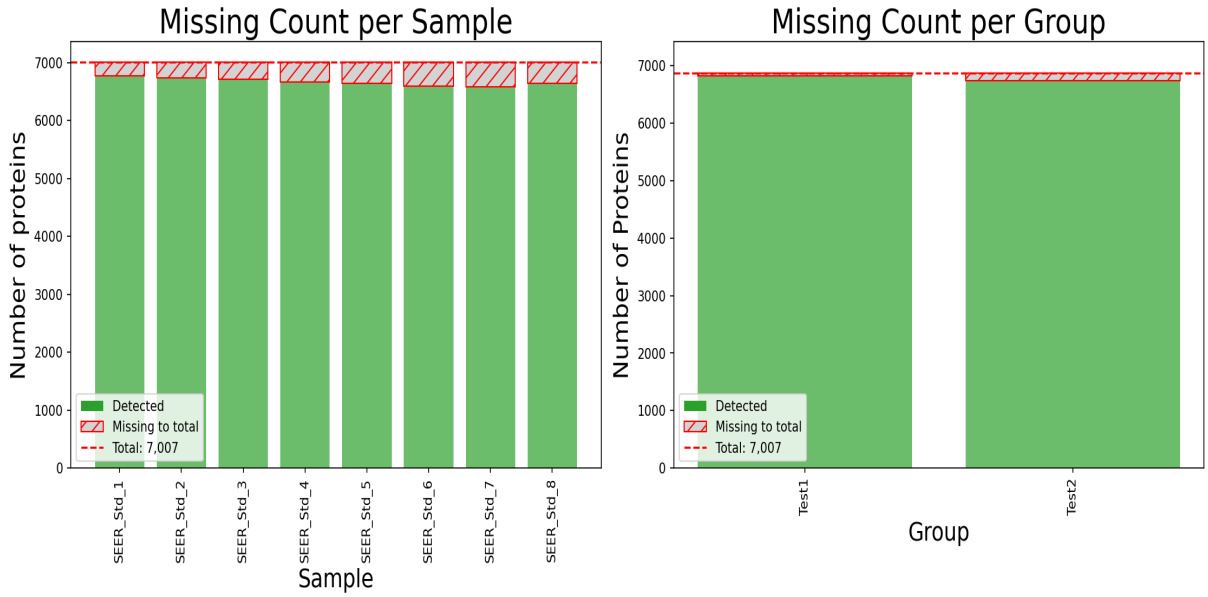
| Data | Quality Grade | Missing Rate |
|---|---|---|
| Original | Excellent | 4.8% |
| Final | Complete | 0% (imputed) |

**Imputation Strategy:**

Our pipeline supports four imputation modes: **'qrilc'** (left-censored/low abundance), **'hybrid'** and **'none'**. Selecting the optimal method is essential to minimize statistical bias and maintain the biological integrity of the quantitative dataset.

**Imputation strategy applied here** : HYBRID

**Distribution of missing values (pre-imputation)**:

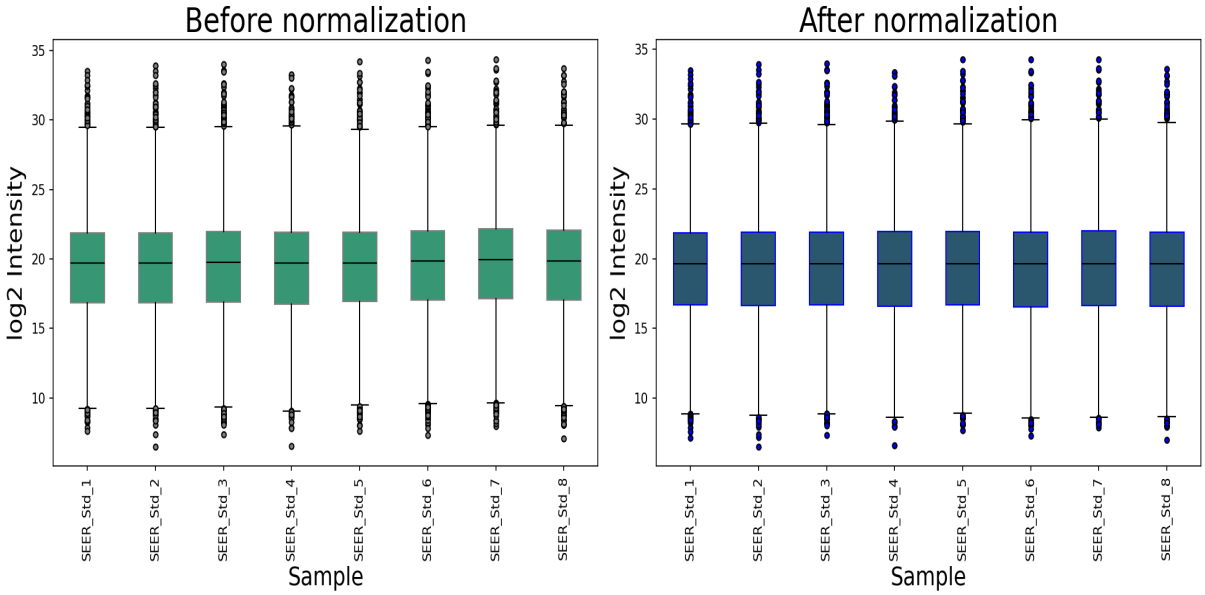


## 4-3. Normalization

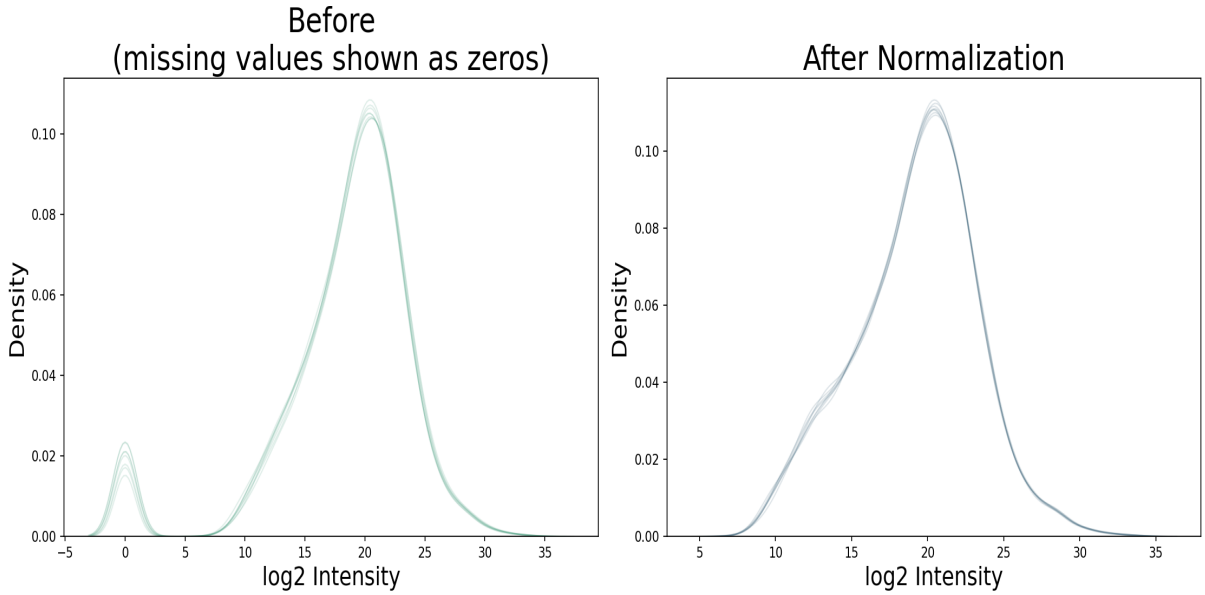
**Normalization Strategy :**

**Median** (aligning sample medians for robust scaling), **Quantile** (harmonizing global distributions to ensure high comparability), and **None** (preserving raw intensities).

**Normalization strategy applied here** : Median

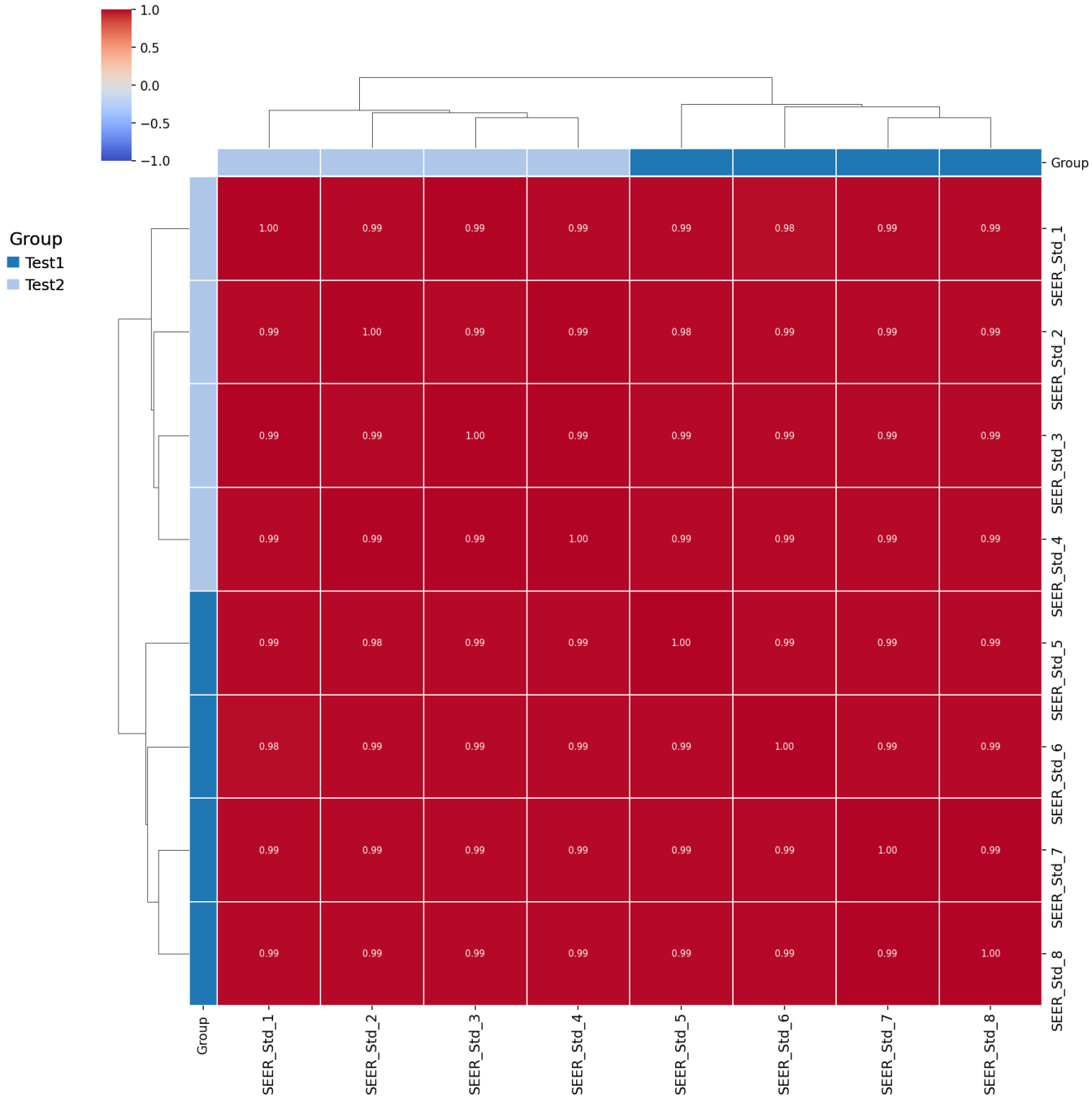



Density plots show the overall distribution shape of intensity values across all proteins for each sample.



## 4-4. Correlation Matrix between samples

A **sample-to-sample correlation** heatmap was used as a quality control measure to evaluate **data consistency** across samples. High correlation coefficients among samples from the same experimental condition indicate reliable and reproducible measurements, whereas reduced correlations may suggest technical artifacts or sample-specific issues.



# 5. Clustering

## 5-1. PCA

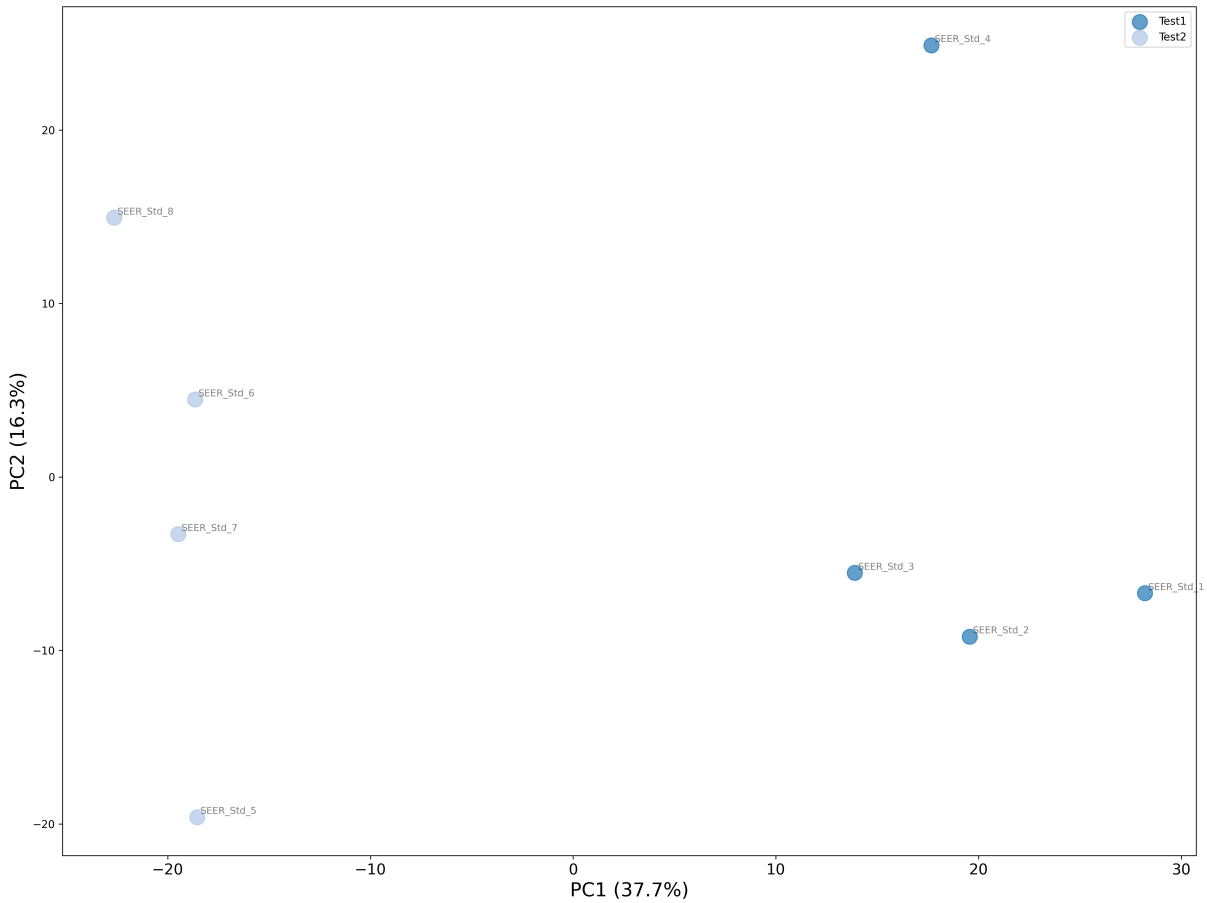
| Analysis Method | PCA |
|---|---|
| PC1 Variance Explained | 37.7% |
| PC2 Variance Explained | 16.3% |
| Total (PC1+PC2) | 54.0% |

**Key Observations:**

• Group separation patterns

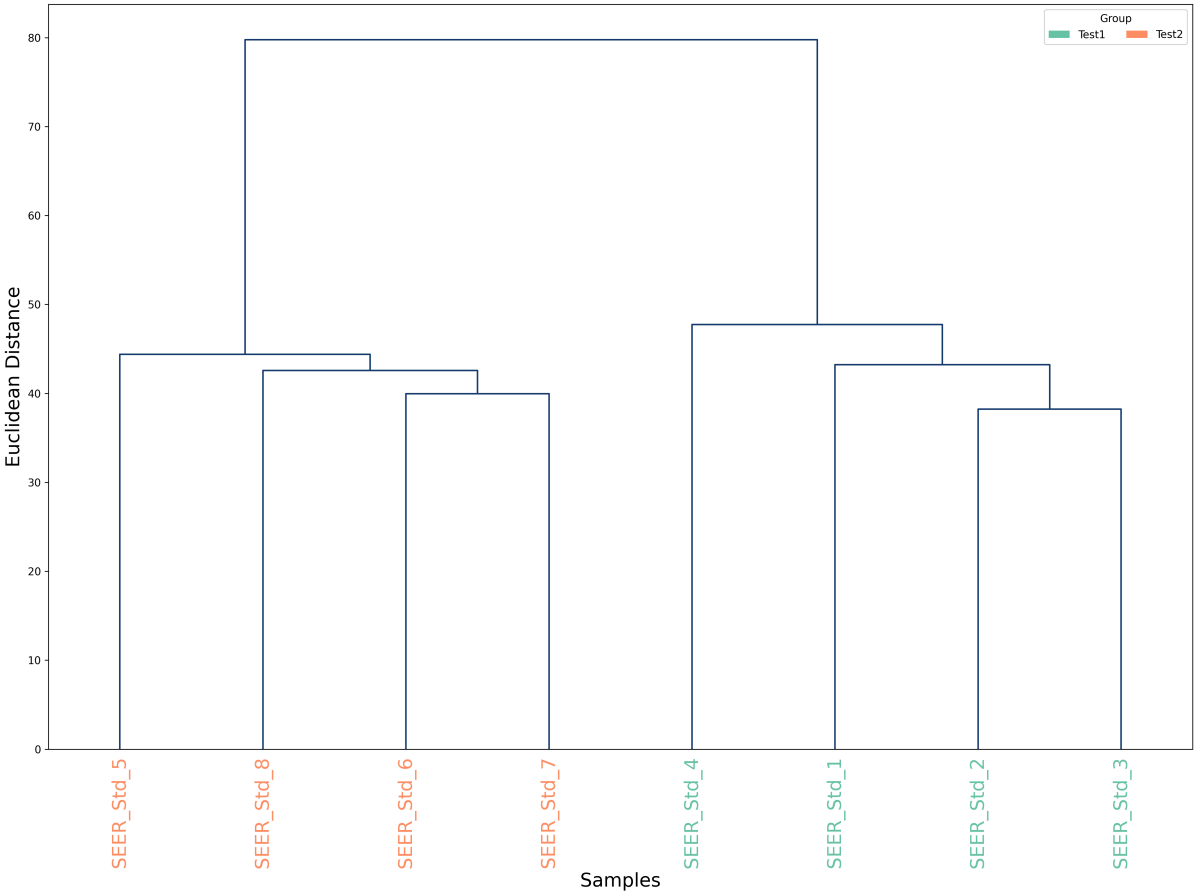
• Batch effect impact (before/after correction)

• Outlier sample identification



## 5-2. Hierarchical Clustering

Hierarchical clustering was performed using the Ward's method based on Euclidean distance metrics. This analysis provides insights into the similarity relationships among samples based on their global protein expression profiles.

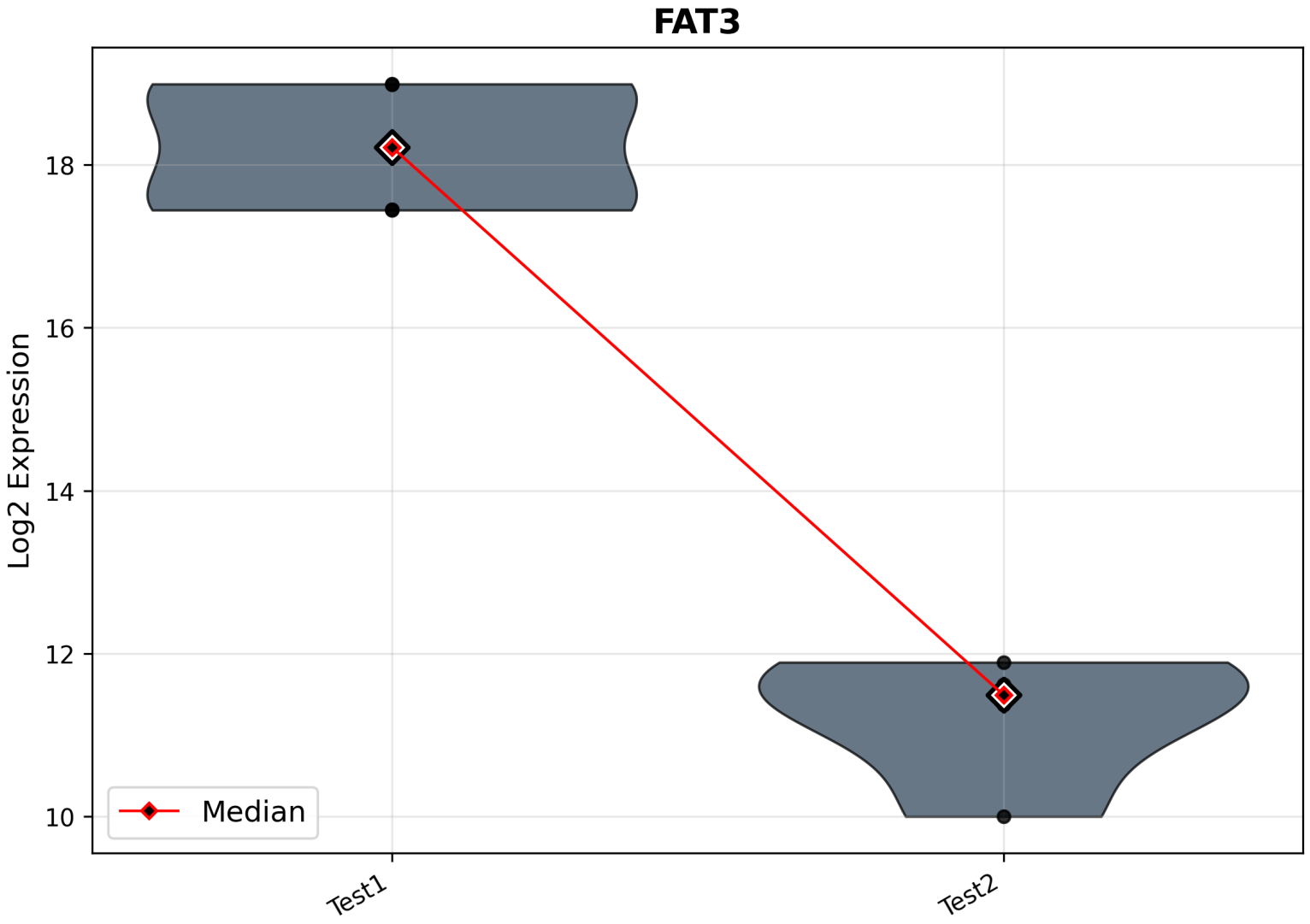


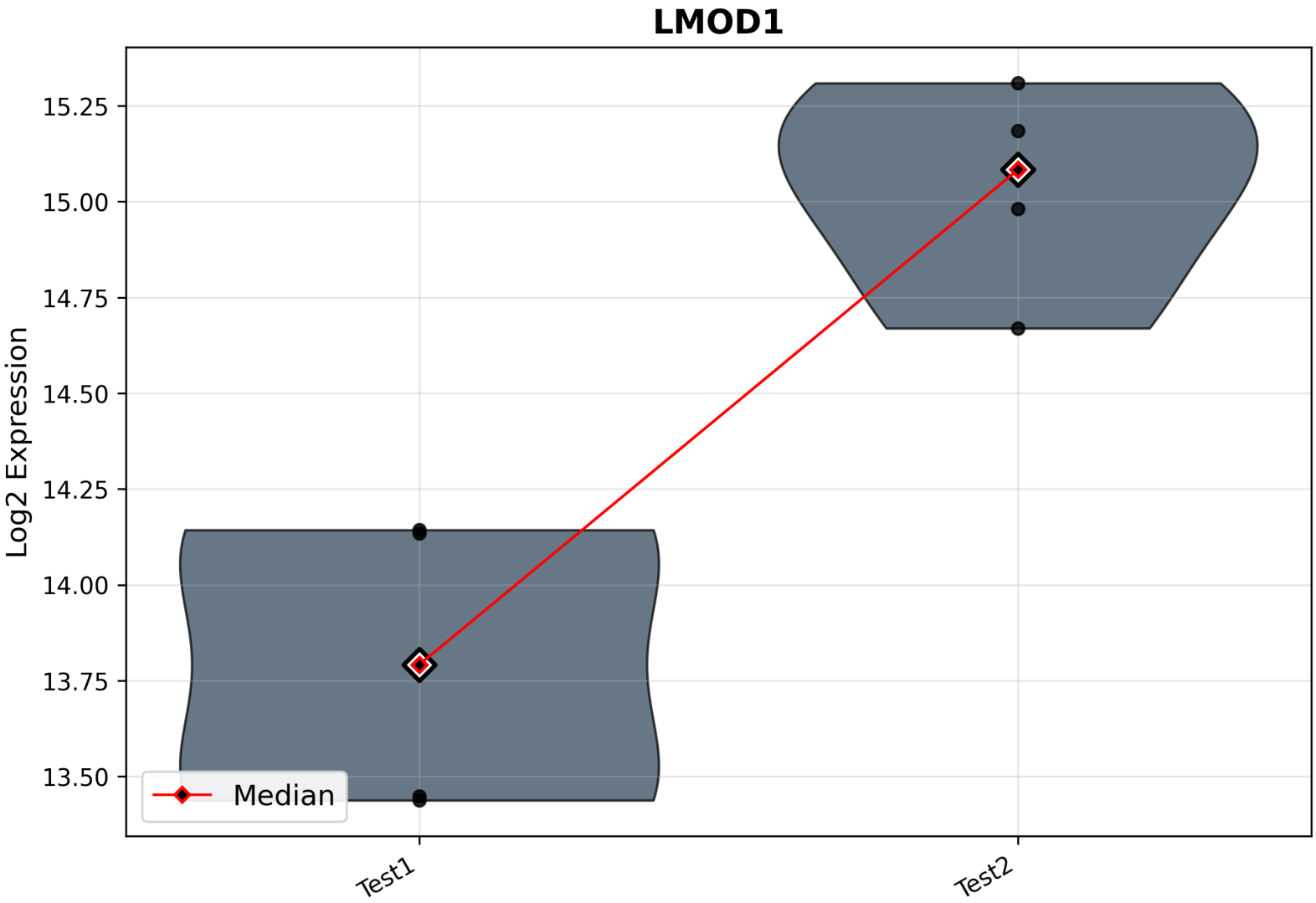
## 5-3. Target Violin plot

The target violin plot illustrates how the quantitative values of the client-specified target proteins differ across groups.
By visualizing the density of the quantitative values of all precursors associated with each target protein within each group, this plot enables comparison of changes in target protein abundance between groups.

The x-axis represents the groups,
and the y-axis shows the log2-transformed quantitative values of the precursors associated with the target proteins.

**From the next page onward, precursor-based violin plots for each target protein are presented, with one plot shown per page.**


FAT3
Log2 Expression
18
16
14
12
10
Median
Test1
Test2

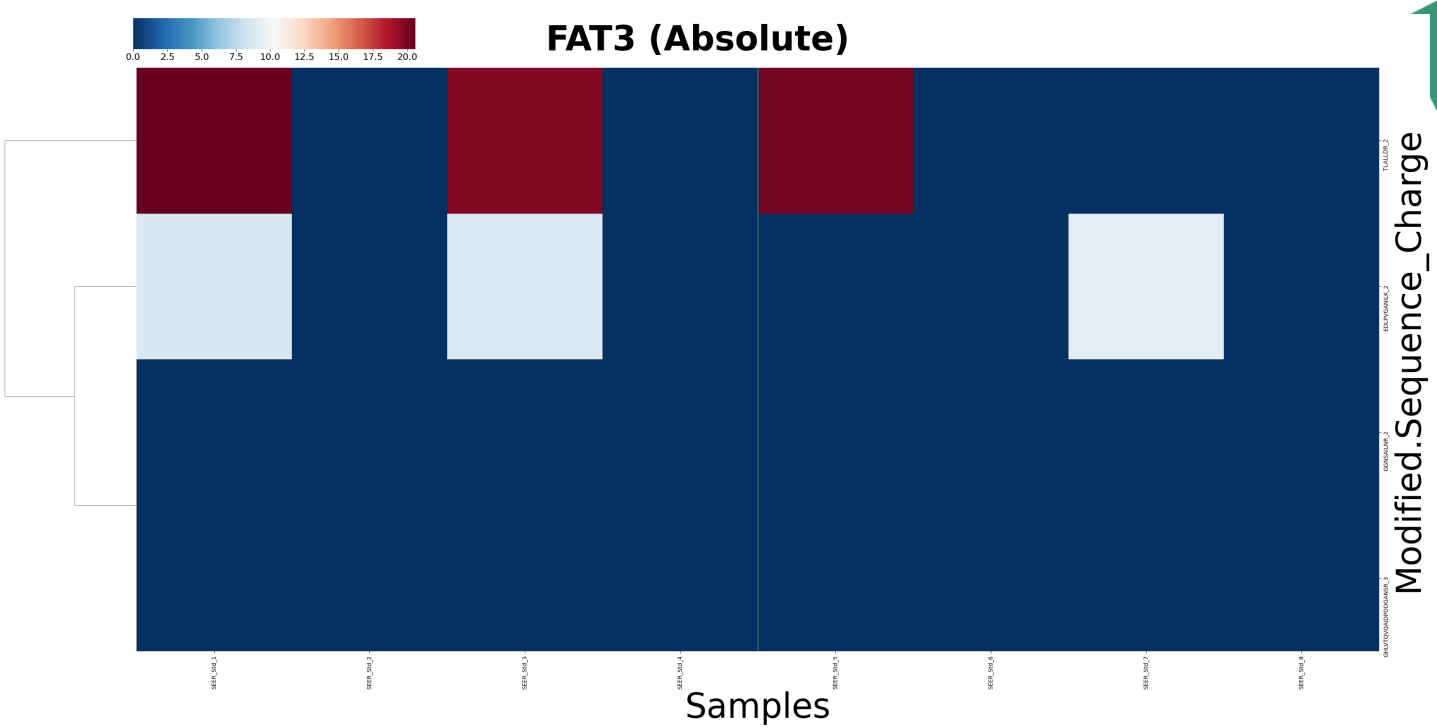

LMOD1
Log2 Expression
15.25
15.00
14.75
14.50
14.25
14.00
13.75
13.50
Median
Test1
Test2

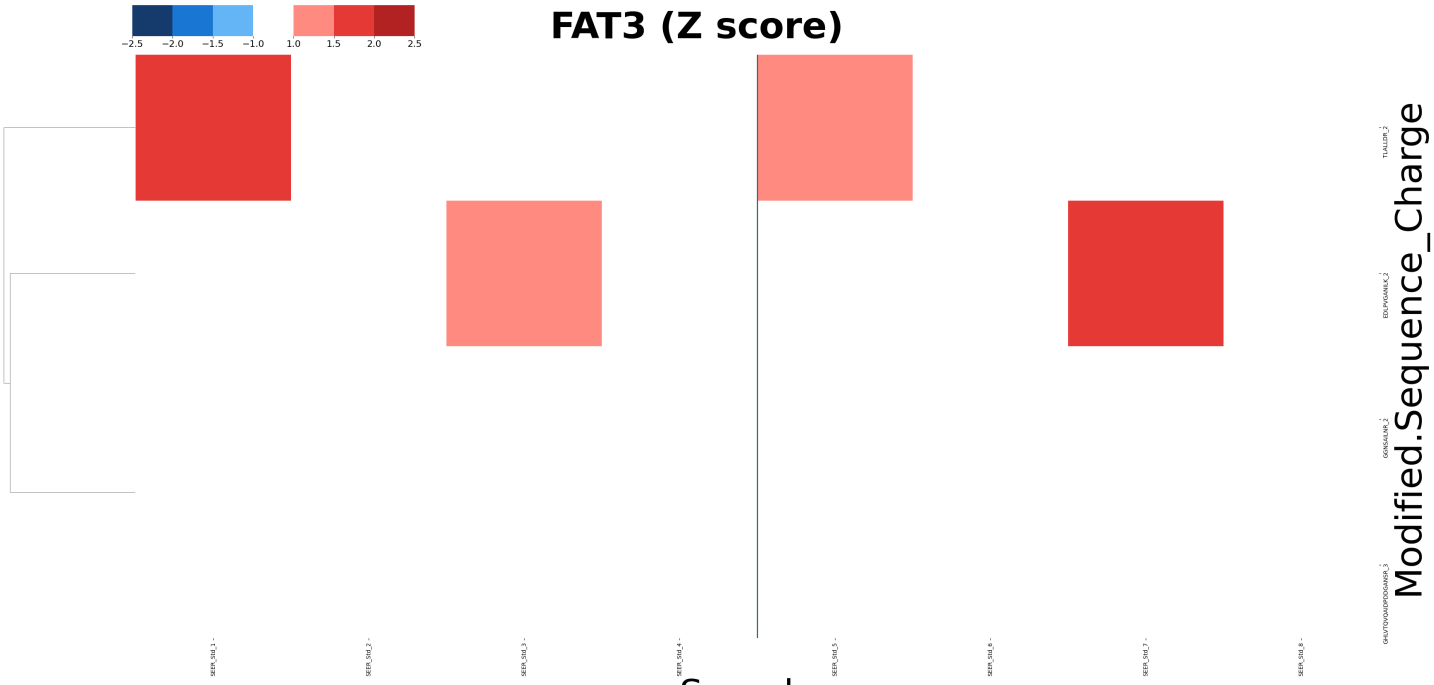
## 5-4. Target Precursor Heatmap

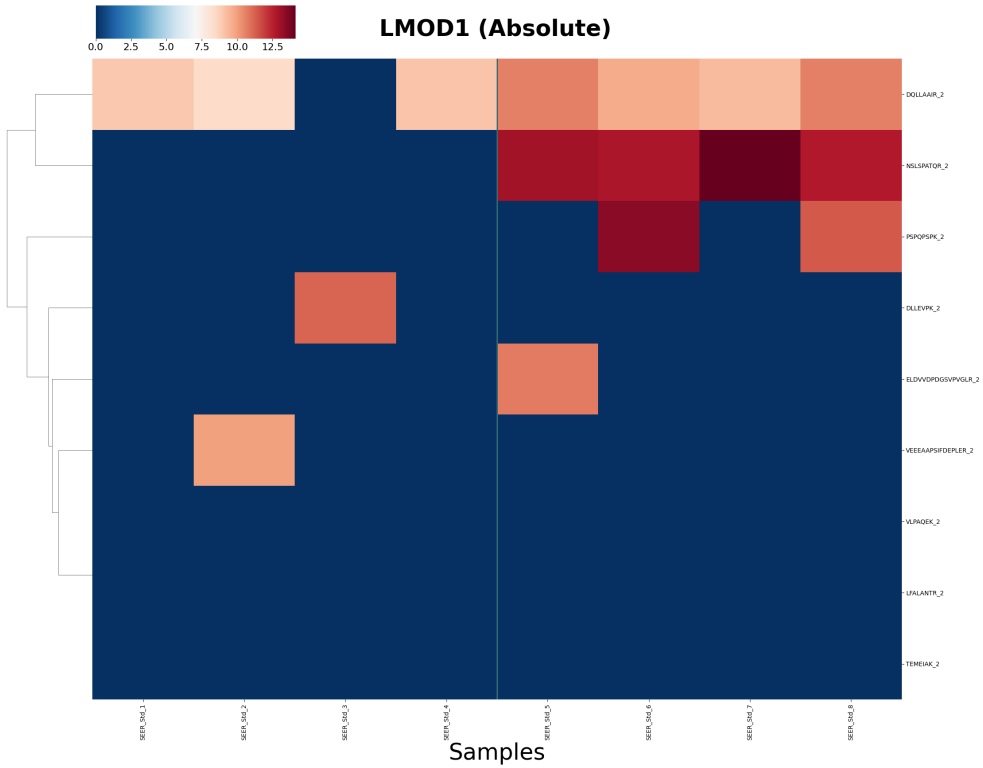
The target precursor heatmap visualizes the quantitative expression patterns of **all precursors associated with client-specified target proteins** when predefined targets are provided.
For each target protein, the quantitative values of all corresponding precursors are displayed as heatmaps to enable detailed comparison of expression patterns across groups.

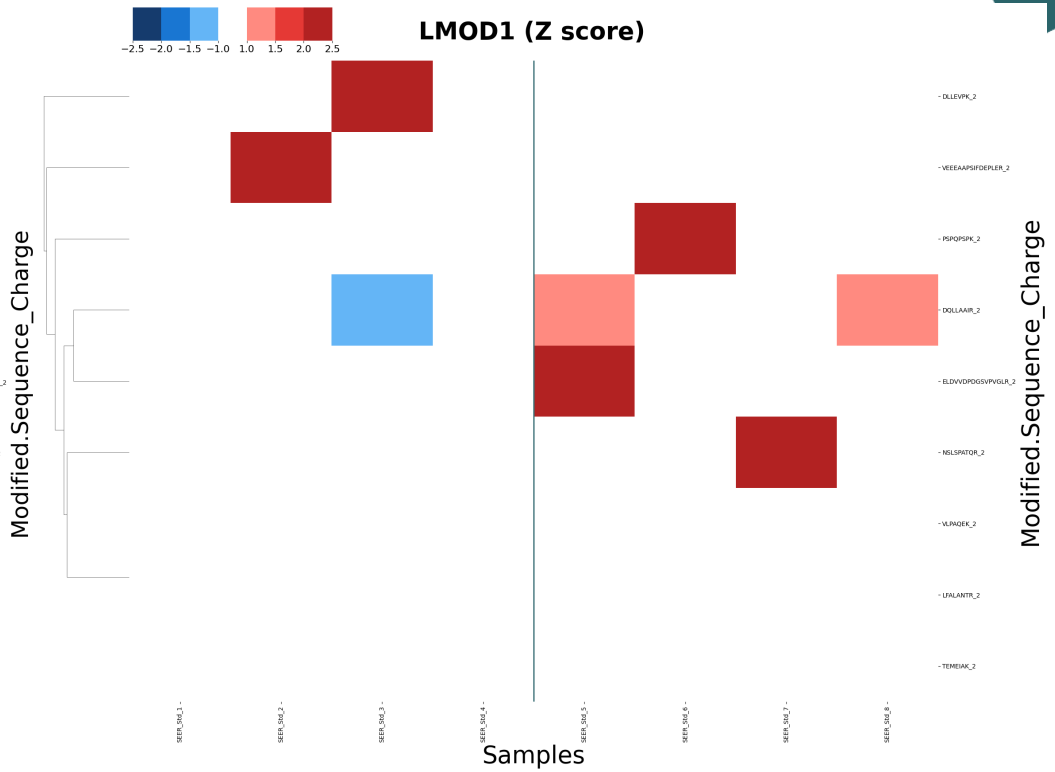
For each target, **two heatmaps** are presented: the first heatmap shows the **absolute quantitative values** of the precursors, while the second heatmap displays the same data after **z-score normalization**, allowing comparison of relative expression patterns independent of absolute abundance differences.

From the next page onward, heatmaps are presented on a **target-by-target basis**, with **one target displayed per page**, ensuring clear and focused visualization of precursor-level expression patterns.


FAT3 (Absolute)
0.0 2.5 5.0 7.5 10.0 12.5 15.0 17.5 20.0
Modified.Sequence_Charge
Samples


FAT3 (Z score)
−2.5 −2.0 −1.5 −1.0 1.0 1.5 2.0 2.5
Modified.Sequence_Charge
Samples


LMOD1 (Absolute)
0.0 2.5 5.0 7.5 10.0 12.5
DQLLAAIR_2
NSLSPATQR_2
PSPQPSPK_2
DLLEVPK_2
ELDVVDPDGSVPVGLR_2
VEEEAAPSIFDEPLER_2
VLPAQEK_2
LFALANTR_2
TEMEIAK_2
SEER_Std_1
SEER_Std_2
SEER_Std_3
SEER_Std_4
SEER_Std_5
SEER_Std_6
SEER_Std_7
SEER_Std_8
Samples


LMOD1 (Z score)
−2.5 −2.0 −1.5 −1.0 1.0 1.5 2.0 2.5
Modified.Sequence_Charge
DLLEVPK_2
VEEEAAPSIFDEPLER_2
PSPQPSPK_2
DQLLAAIR_2
ELDVVDPDGSVPVGLR_2
NSLSPATQR_2
VLPAQEK_2
LFALANTR_2
TEMEIAK_2
SEER_Std_1
SEER_Std_2
SEER_Std_3
SEER_Std_4
SEER_Std_5
SEER_Std_6
SEER_Std_7
SEER_Std_8
Samples
Modified.Sequence_Charge

# 6. DEP

Differentially Expressed Proteins (DEPs) refer to proteins that show statistically significant changes in abundance between different experimental conditions (e.g., Treatment vs. Control).

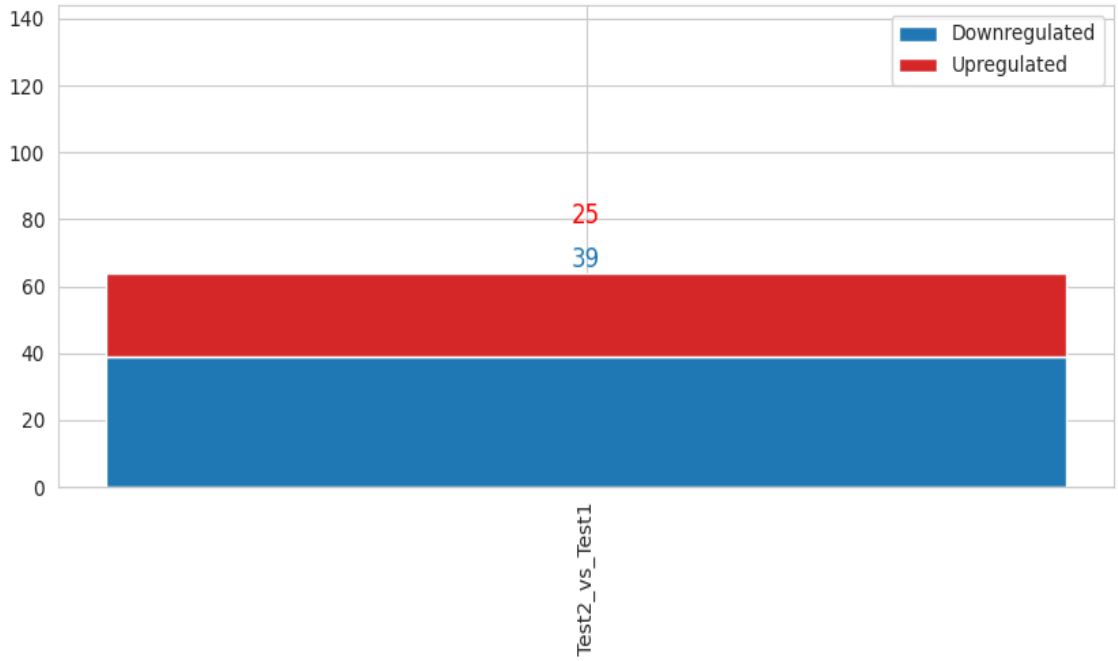
.

## 6-1. DEP Result

DEPs were identified using a dual-threshold approach to ensure both biological relevance and statistical stringency:

**1) Fold Change:** A minimum **2-fold difference (|log2FC| ≥ 1)** was required. Values ≥ 1 indicate up-regulation (red), while values ≤ -1 indicate down-regulation (blue).
**2) Statistical Significance:** A **p-value < 0.05** was applied to confirm that the observed variations were statistically significant and not due to random chance.

The bar plot illustrates the distribution of these DEPs across each experimental comparison.



## 6-2. DEP Volcano

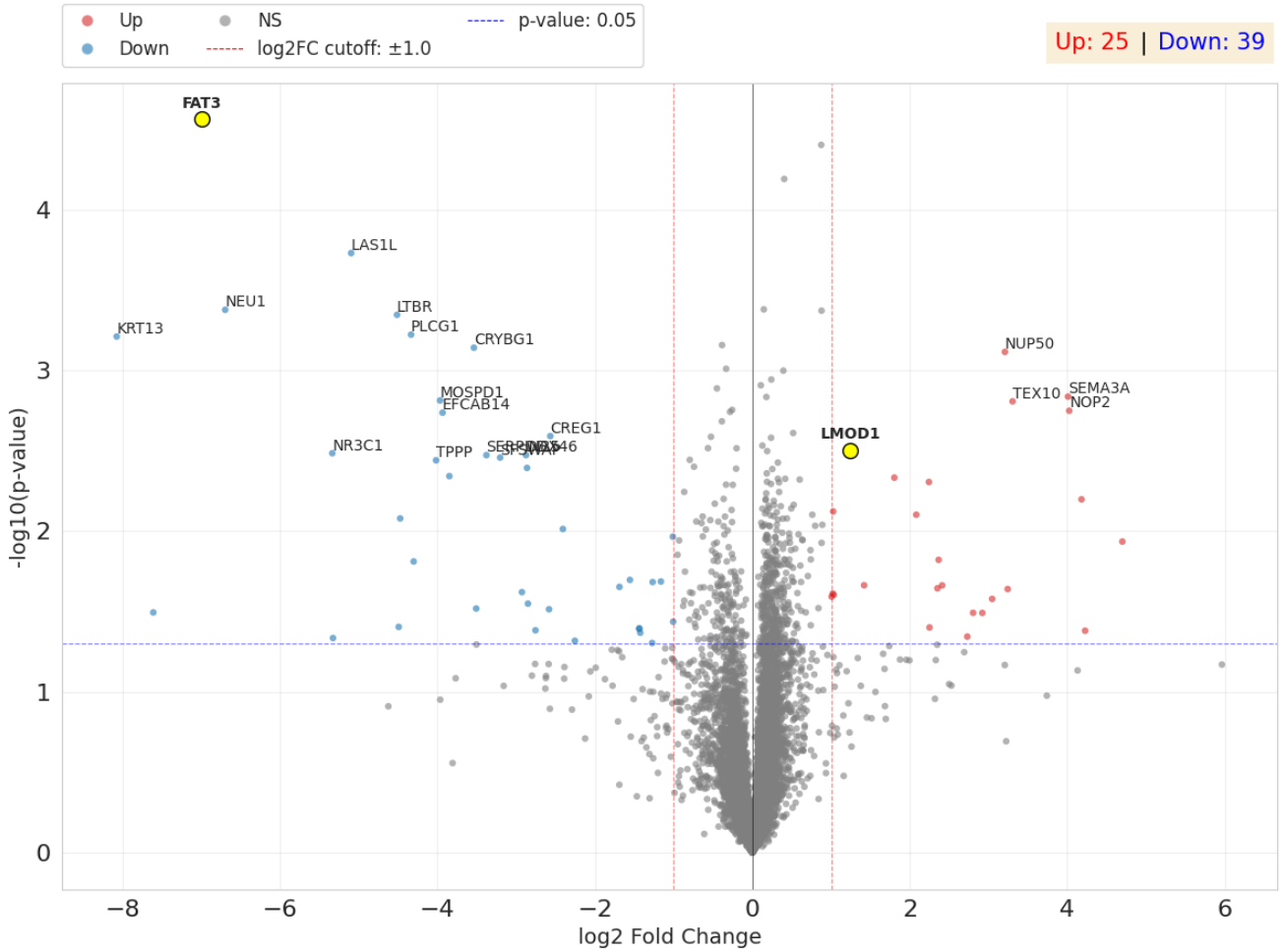
The DEP volcano plot illustrates the results of differential expression analysis between **treatment and control** groups.
In this plot, the x-axis represents the log2 fold change, while the y-axis represents the –log10 transformed p-value, allowing simultaneous assessment of expression magnitude and statistical significance.

For each comparison, a volcano plot of **treatment versus control** is generated. Among the identified **DEPs**, the **top 20 DEPs ranked by absolute fold change** are annotated directly on the plot. In addition, a summary table is provided, listing the names of these DEPs along with their corresponding fold change and statistical values.

When client-specified target proteins are present among the DEPs, they are **highlighted in yellow** in the volcano plot for visual emphasis.

From the next page onward, **one treatment-versus-control volcano plot is presented per page**.

## DE_volcano_Default_Test2_vs_Test1

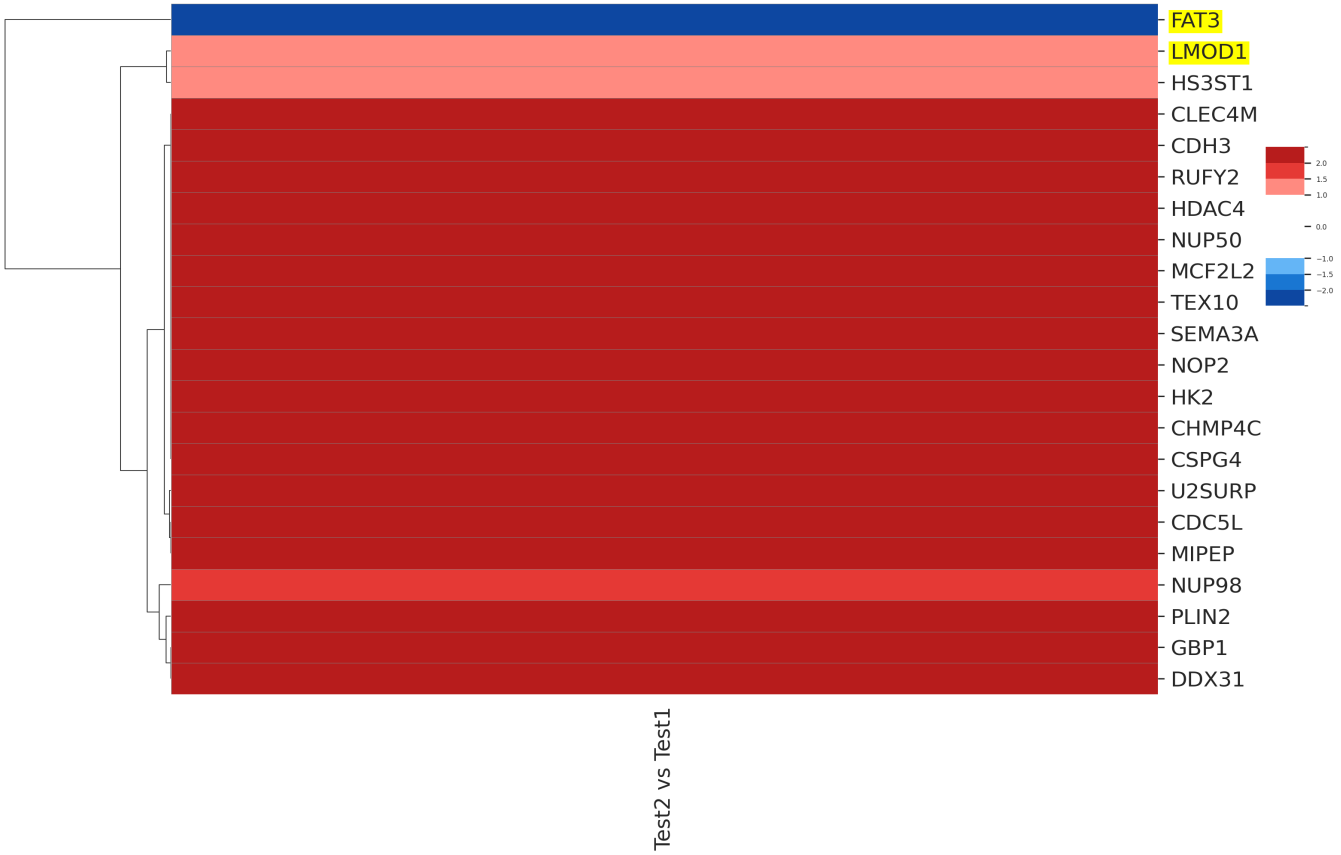



| Top20 | log2FC | pvalue | Significance |
|---|---|---|---|
| FAT3 | -7.0 | 2.7e-05 | Down |
| LAS1L | -5.098 | 0.000186 | Down |
| NEU1 | -6.698 | 0.00042 | Down |
| LTBR | -4.515 | 0.000451 | Down |
| PLCG1 | -4.337 | 0.000599 | Down |
| KRT13 | -8.077 | 0.000615 | Down |
| CRYBG1 | -3.539 | 0.000723 | Down |
| NUP50 | 3.208 | 0.000767 | Up |
| SEMA3A | 4.01 | 0.001456 | Up |
| MOSPD1 | -3.969 | 0.001538 | Down |
| TEX10 | 3.306 | 0.001558 | Up |
| NOP2 | 4.026 | 0.001783 | Up |
| EFCAB14 | -3.937 | 0.001832 | Down |
| CREG1 | -2.568 | 0.002563 | Down |
| LMOD1 | 1.245 | 0.003146 | Up |
| NR3C1 | -5.338 | 0.003276 | Down |
| DDX46 | -2.877 | 0.003367 | Down |
| SERPINB5 | -3.38 | 0.003368 | Down |
| SFSWAP | -3.205 | 0.003486 | Down |
| TPPP | -4.019 | 0.003626 | Down |

## 6-3. DEP Heatmap

The DEP heatmap visualizes the quantitative expression patterns of proteins identified as **DEPs** across different groups.
In this heatmap, **red** indicates upregulated expression, **blue** indicates downregulated expression, and **white** represents proteins that are not significant (NS) and therefore not classified as DEPs in the corresponding comparison.

The complete set of DEP heatmaps can be found in the **DEP** folder. In this report, heatmaps displaying the **top 30 DEPs** ranked by expression change are shown. When client-specified target proteins are included among the DEPs, they are **highlighted in yellow** in the heatmap for clear visual emphasis.

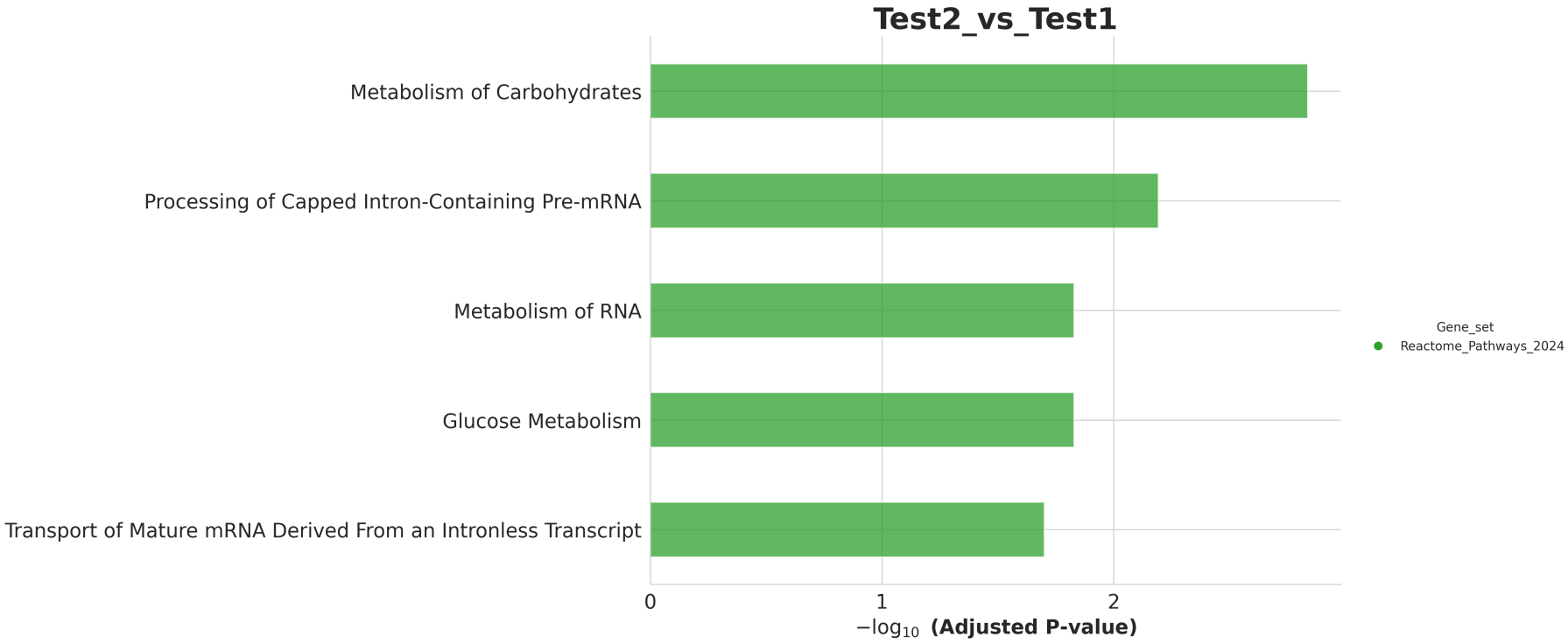


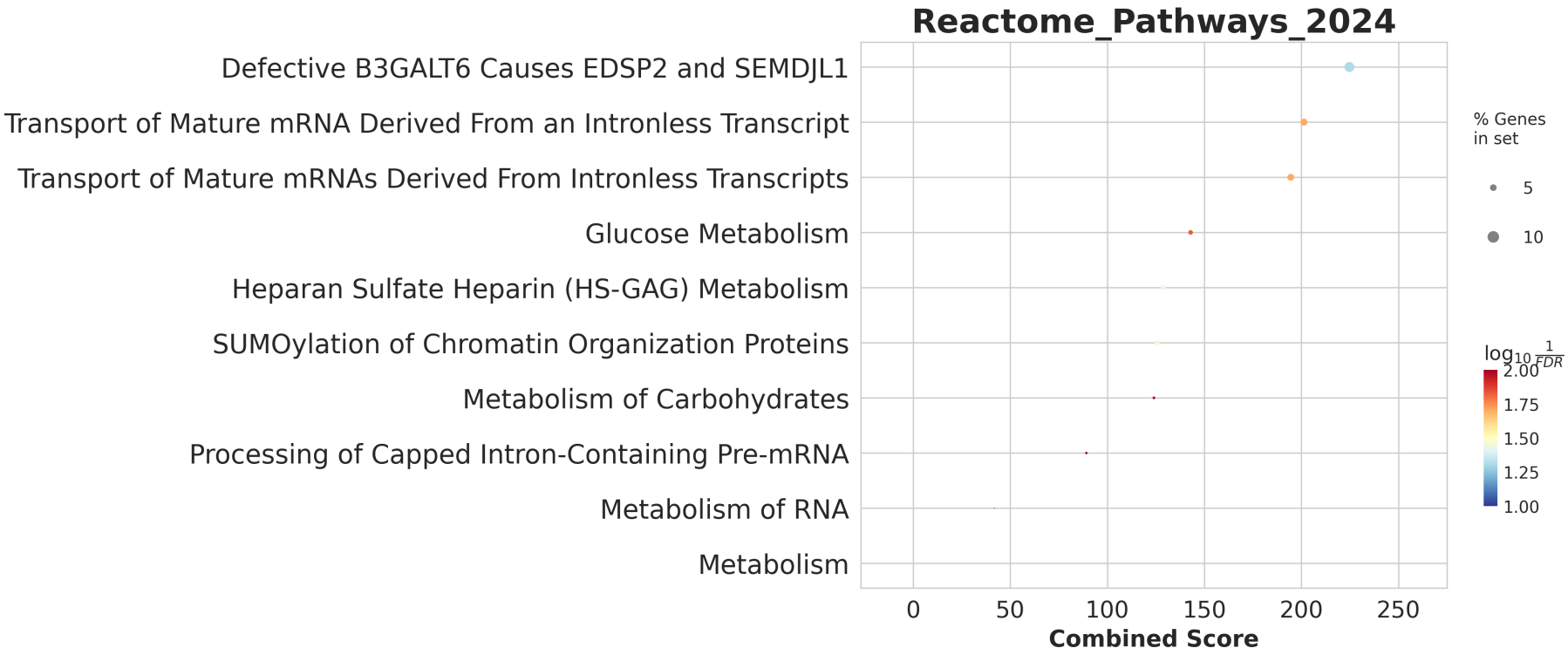
# 7. Functional annotaion

Functional annotation analysis was performed using **gene lists derived from DEPs** identified in each **treatment versus control** comparison.
For each comparison, only genes corresponding to statistically significant DEPs were used as input for over-representation analysis across multiple functional databases, including Gene Ontology, KEGG, WikiPathways, and Reactome.

The enrichment results for each DEP-derived gene list are visualized using **bar plots** and **bubble charts**, which summarize the most significantly enriched functional terms based on adjusted p-values.

For each **treatment-versus-control** comparison, the enrichment results are presented across **two pages**: the first page displays a bar plot summarizing enriched terms across databases, and the second page presents a bubble chart highlighting the most significant terms in greater detail. Each plot is shown individually, with **one plot per page**, to ensure clear and interpretable visualization.


Test2_vs_Test1
Metabolism of Carbohydrates
Processing of Capped Intron-Containing Pre-mRNA
Metabolism of RNA
Glucose Metabolism
Transport of Mature mRNA Derived From an Intronless Transcript
0
1
2
−log10 (Adjusted P-value)
Gene_set
Reactome_Pathways_2024


Reactome_Pathways_2024
Defective B3GALT6 Causes EDSP2 and SEMDJL1
Transport of Mature mRNA Derived From an Intronless Transcript
Transport of Mature mRNAs Derived From Intronless Transcripts
Glucose Metabolism
Heparan Sulfate Heparin (HS-GAG) Metabolism
SUMOylation of Chromatin Organization Proteins
Metabolism of Carbohydrates
Processing of Capped Intron-Containing Pre-mRNA
Metabolism of RNA
Metabolism
0
50
100
150
200
250
Combined Score
% Genes in set
5
10
$\log_{10}\frac{1}{FDR}$
2.00
1.75
1.50
1.25
1.00

# Notice

- This analysis report will not be disclosed externally without your consent, and **OmixAI** will keep all information strictly confidential.
- Remaining samples will be stored for **3 months** from the data of report delivery and then discarded. If you would like the samples returned, please contact us before the scheduled disposal date.
- Raw analysis data and result files will be securely retained for **3 months** from the date of report delivery and then permanently deleted. Please ensure you have saved copies to your computer before deletion.
- After report delivery, **OmixAI** can **provide tailored additional analyses upon request** following a consultation with our bioformatics specialists.

We appreciate your trust.
Please feel free to contact us with any questions or requests.