APAF Analytical Services Report

Macquarie University t/a Australian Proteome Analysis Facility ("APAF")
ABN: 90 952 801 237
APAF, Australian Proteome Analysis Facility
Level 4, 4 Wally's Walk, Macquarie University, Sydney NSW 2109
Ph: +61 2 9850 6201 Fax: +61 2 9850 8313
Email: info.apaf@mq.edu.au Website: www.mq.edu.au/research/APAF

Final Report

Label Free Quantification by LC MS/MS Analysis

Report number:	MS-R30001-1
Report date:	11th March 2026
Project number:	30001
Client name:	Thiri Winn
Client organisation:	APAF
Client address:	Level 4, 4 Wally's Walk Macquarie University, NSW, 2109, AUSTRALIA
Client contact number:	(02) 9850 6201
Client email:	info.apaf@mq.edu.au
Date sample(s) received:	1st March 20XX
Number of samples:	6
Project leader:	Thiri Winn
Authorised by:	Muhammad A. Zenaidee, PhD
APAF email:	info.apaf@mq.edu.au
Attachments	Yes (xx.pg_matrix.xlsx)

The results apply to the sample(s) as received.

This report may only be reproduced in its entirety.

As per APAF Terms and Conditions, **samples will be retained for a period of thirty (30) days** and **testing records will be accessible for a period of three (3) years** from the date of reporting results unless other arrangements have been made; refer to Clause 11.1 (sample retention) and Clause 10.3 (test records) for conditions that apply.

Acknowledgment: To comply with our NCRIS (National Research Infrastructure for Australia) operating grant, we request that any publication arising from access to the facility acknowledge the contribution of APAF staff and include the statement "***This study/project/research used NCRIS-enabled Australian Proteome Analysis Facility (APAF) infrastructure***".

Accreditation Number: 20344
Accredited for compliance with ISO/IEC 17025 - Testing

SAMPLE DETAILS

- Sample 1 (APAF Sample code: S0000001)
- Sample 2 (APAF Sample code: S0000002)
- Sample 3 (APAF Sample code: S0000003)
- Sample 4 (APAF Sample code: S0000004)
- Sample 5 (APAF Sample code: S0000005)
- Sample 6 (APAF Sample code: S0000006)

METHOD DETAILS

- SOP used: MS-080_Cell and Tissue lysis and digestion protocol for DDA and DIA MS_V4

In brief:

- Each sample was resuspended in x µL of digestion solution (1% SDC, 100 mM TEAB)
- Protein concentrations were determined
- X µg of each sample was taken for digestion
- Samples were cleaned up by SDB stage tip.
- Cleaned digested samples were dried completely.
- Each sample was resuspended in x µL of loading buffer (0.1 % formic acid)
- X µL (x µg) of each sample was taken for LC MS/MS analysis

DATA ACQUISITION DETAILS

- SOP used: MS-101_HFX Exploris Data Acquisition_V1

Instrument:

Mass spectrometer:	Orbitrap Exploris (Thermo Fisher Scientific)
NanoLC system:	Ultimate 3000 UHPLC System (Thermo Fisher Scientific)
Trap column:	300 µm × 5 mm, C18 PepMap 100, 5µm, 100 Å (Thermo)
Analytical column:	Dr Maisch Reprosil-Pur 120 C18-AQ 1.9µm, 75µm × 30 cm (self-packed)
Loading buffer:	0.1% formic acid
Mobile phase A:	0.1% formic acid
Mobile phase B:	80% acetonitrile, 0.1% formic acid

LC method:

- Sample was injected onto a reverse-phase peptide trap for pre-concentration and desalted with loading buffer.
- Peptides were eluted from the analytical column using a linear gradient of mobile phase B (2.5 - 37.5%) at a flow rate of 300 nL/min over a 60 min period.

MS method (Data Independent Acquisition):

- The column eluent was directed into the ionization source of the mass spectrometer operating in positive ion mode.

Full Scan	
Scan Range (m/z)	350-1450
Orbitrap Resolution	60000
AGC Target	Standard
Maximum Injection Time Mode	Auto
DIA	
Precursor Mass Range (m/z)	*20 DIA ranges between 350 - 1450*
DIA Window Type	Auto
Multiplex Ions	False
Isolation Windows (m/z)	*Fixed or Variable*

DIA Window Mode	m/z Range
Collision Energy Type	Normalized
HCD Collision Energy (%)	27
Orbitrap Resolution	30000
Scan Range Mode	Auto
AGC Target	Standard
Maximum Injection Time Mode	Auto

DATA PROCESSING DETAILS

Software used	DIA-NN
Version	1.9.1
Precursor ion generation	
Enzyme Name	Trypsin/P
Database	uniprotkb_homo_sapiens.fasta
Missed Cleavages	1
Modification(s)	N-term M excision; Ox (M); C carbamidomethylation
Peptide length range	7 - 30
Precursor charge range	1 – 4
Precursor m/z range	300 - 1800
Fragment ion m/z range	200 - 1800
Algorithm	
Unrelated runs	checked
Protein inference	Protein names from FASTA
Neutral network classifier	Single-pass mode
Quantification strategy	Quant UMS (high precision)
Cross-run normalisation	RT-dependent
Library generation	IDs, RT & IM profiling
Output	
Precursor FDR (%)	1

Data analysis:

Results were analysed using APAF's in-house statistical pipeline, (MultiScholar), which includes normalisation, differential expression testing, and data visualisation.

Data filtering and normalisation:

Data filtering and normalization Peptide and protein quantification was performed using a series of filtering and aggregation steps to ensure high-confidence identifications and accurate quantification. Initially, low-confidence peptide identifications were removed using q-value and proteotypic peptide filtering, retaining only peptides with a q-value threshold of 0.01 and only those that are proteotypic. Precursor ion intensities were then aggregated into peptide-level quantification values, combining all charge states of the same peptide sequence while keeping different modified peptides (peptidoforms) as separate entries.

Intensity threshold filtering was applied to remove peptides where a significant proportion of samples fell below a specified log-intensity threshold. By default, peptides were excluded if 50% or more of the samples were below the lowest 1st percentile of intensity values. Proteins were required to have a minimum of 2 peptides, which could include different modifications of the same sequence, to be considered for further analysis.

Sample Quality Control and Filtering Samples with poor performance, indicated by an unusually low number of identified peptides, were identified and removed to ensure data quality and reliable downstream analysis. Such samples may be

compromised due to issues in sample preparation, instrument performance, or other technical factors. Including these low-quality samples in downstream analysis could introduce bias and reduce the statistical power to detect true biological differences. Samples with fewer than 500 peptides were excluded from further analysis. This minimum peptide threshold ensures that all retained samples have sufficient protein coverage for reliable quantification and statistical testing. Of 57 initial samples, 57 samples were retained after filtering. The following samples were removed due to insufficient peptide counts: c().

Peptides appearing in only one replicate across all groups were also excluded to maintain measurement reproducibility. Protein-level quantification was performed using the IQ algorithm, implementing the maxLFQ algorithm to generate protein abundance values. Peptide intensities were aggregated into protein-level quantification without normalization at this stage, resulting in a protein quantification matrix with log2-transformed intensity values for each protein across all samples.

FASTA Reference and Protein Accession Cleanup Protein identifications were validated against a standard UniProt FASTA database, providing comprehensive protein metadata including evidence levels, reviewed status, gene names, and isoform information. The analysis used a mixed-species FASTA database containing proteins from multiple organisms. The primary organism was identified as Homo sapiens. Proteins from non-target organisms were filtered at the import stage. This metadata was used to select the highest-quality protein identifications during accession cleanup.

Protein groups containing multiple accessions (separated by ';' delimiters) were resolved by selecting the best representative protein based on protein evidence levels, sequence length, and database annotation quality. Quantitative values for proteins within the same group were aggregated using the 'mean' method, reducing 4564 protein groups to 3354 unique protein identifications. After quality control filtering, 3063 proteins remained for normalization.

Using the protein intensity matrix, zero values were replaced with NA denoting these are missing values. Protein filtering was then applied based on missing value patterns and intensity thresholds to ensure robust statistical analysis. Proteins were excluded if they had 42.857% or more missing values or values below the 1st percentile of intensity values within any individual experimental group. Additionally, proteins were excluded if they failed this threshold in more than 77.778% of experimental groups. Additionally, proteins with intensity values below the 1st percentile across all samples were excluded from further analysis as likely noise or background signal. The data matrix was then log (base 2) transformed and between-sample normalization were performed using the 'cyclicloess' method from the 'limma' R package (Ritchie et al., 2015).

To remove batch effects from biological data, the remove unwanted variation (RUVIII-C) method was used. The method relies on having a set of endogenous negative control proteins, which are proteins with little or no changes in protein abundances between different samples. For this study, an automatic optimization process was run, which identified that using 15.0%% of features as empirical negative controls (457 proteins) provided the best separation of biological groups. Based on this, 3 unwanted components were selected for removal. Following RUV-III batch correction and subsequent missing value cleanup, 3029 proteins were retained for differential expression analysis.

At each stage of the data normalization, the samples were checked for batch effects using principal component analysis (PCA) plot, density boxplots, relative log-expression (RLE) plots and the distribution of the Pearson correlation between replicate samples within groups.

RESULTS

Sample preparation, data acquisition, data analysis, data processing and reporting were performed (XX/XX/XXXX-XX/XX/XXXX):

Differential abundance analysis of proteins was performed using the adjusted abundance matrix for comparing each pair of consensus clusters. The 'limma' R package (Ritchie et al., 2015) was used. A linear model for comparing each pair of time points was fitted using the formula '~ 0 + group' with the 'lmFit' function, and p-values were calculated using the 'eBayes' function. A trended and robust empirical Bayes analysis was performed. The false discovery rate correction was applied to the moderated p-values by calculating the q-values (Storey, 2002). The standard empirical Bayes method was used to test for any deviation from zero log-fold-change. Significant differentially expressed proteins were defined as those with q-values less than 0.05.

The number of significant differentially expressed proteins is summarized in Figure 1. For the comparison F1_Treatment_S0_vs_F1_Control_S0, there were 401 upregulated and 463 downregulated proteins.

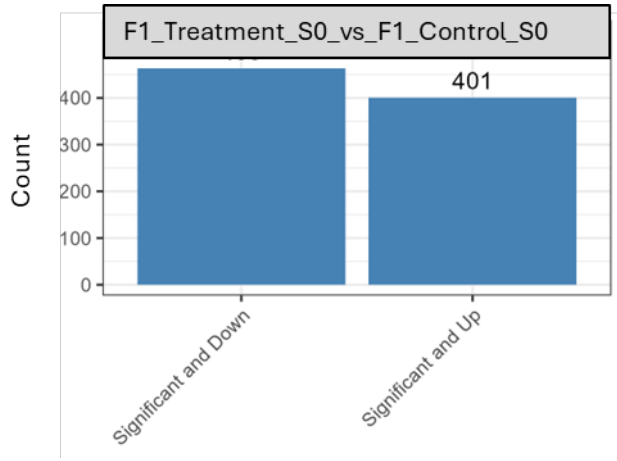

Figure 1: Number of significant differentially expressed proteins for F1_Treatment_S0_vs_F1_Control_S0

Volcano plots of differentially expressed proteins across all groups are shown in Figure 2, with a threshold of q-value < 0.05 and no log fold-change threshold. Full details of all proteins are provided in the Supplementary results table, DE_proteins_results.xlsx within the Publication_tables folder.

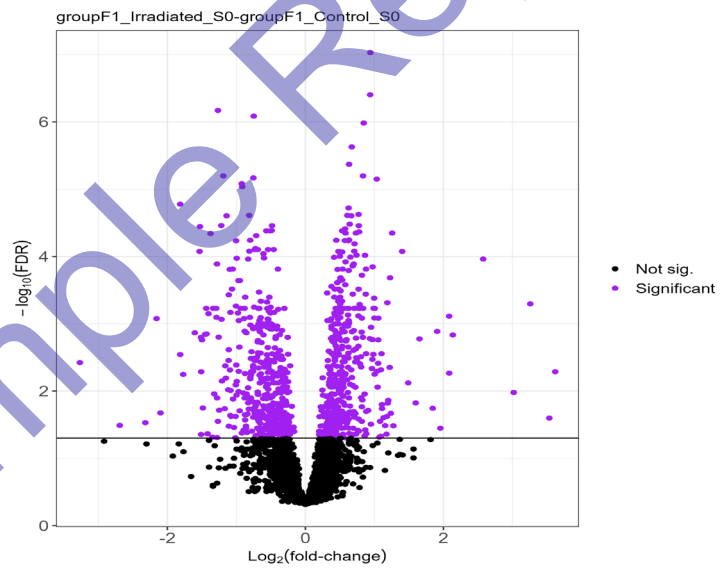

Figure 2: Volcano plots of differentially expressed proteins for F1_Treatment_S0_vs_F1_Control_S0.

Gene Ontology (GO) enrichment analysis was performed on the differentially expressed proteins to identify significantly enriched biological processes, molecular functions, and cellular components. The analysis was conducted using the following parameters:

1. No fold change cutoff was applied to filter differentially expressed proteins.
2. An adjusted FDR cutoff of 0.05 was used to identify significantly enriched GO terms
3. The background protein set consisted of all identified proteins from the species under study (NCBI Taxonomy ID: 10090) that passed the quality control filtering steps.

The enrichment analysis was performed using the clusterProfiler R package (Yu et al., 2012). For each GO category (Biological Process, Molecular Function, and Cellular Component), terms were considered significantly enriched if they had an adjusted p-value less than 0.05 after multiple testing correction using the Benjamini-Hochberg method.The complete list of enriched GO terms with statistics is provided in the supplementary file 'Pathway_enrichment_results.xlsx' within the Publication_tables folder. A list of proteins (xx.pg_matrix.xlsx) is also enclosed with the report.

OPINIONS AND INTERPRETATIONS

Interpretation and/or detailed discussions may be required to fully understand the results presented to you. APAF is committed to assist our clients/collaborators to maximise the value from their results through these consultations.
It should be noted that if these results are to be incorporated into a publication, then APAF will be pleased to supply further details/methodology as required by the publishing journal.