Call: HORIZON-HLTH-2024-DISEASE-08-20
Topic: Pandemic preparedness and response: Host-pathogen interactions of infectious diseases with epidemic potential
Funding Scheme: HORIZON Research and Innovation Actions (RIA)

Deliverable No. 2.2
Host surfaceome

Grant Agreement no.: 101191739

Project Title: **Comparative signature of Marburg virus cell activation as a blueprint for the identification of antiviral targets against newly emerging viruses**

Contractual Submission Date: 30/06/2026

Actual Submission Date: 30/06/2026

Responsible partner: P1: Helmholtz-Zentrum für Infektionsforschung GmbH (HZI)

Funded by
the European Union

Grant agreement no.	101191739
Project full title	**COMBINE – Comparative signature of Marburg virus cell activation as a blueprint for the identification of antiviral targets against newly emerging viruses**

Deliverable number	**D2.2**
Deliverable title	**Host surfaceome**
Type[1]	R
Dissemination level[2]	PU
Work package number	WP2
Author(s)	Karsten Cirksena (HZI)
COMBINE reviewers	Lothar Jänsch, Christian Sieben (HZI)
Keywords	Host factors, Receptors, Surface proteins, Tissue-specific

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Health and Digital Executive Agency (HaDEA).

Neither the European Union nor the granting authority can be held responsible for them.

[1] **Type**: Use one of the following codes (in consistence with the Description of the Action):
R: Document, report (excluding the periodic and final reports)
DEM: Demonstrator, pilot, prototype, plan designs
DEC: Websites, patents filing, press & media actions, videos, etc.

[2] **Dissemination level**: Use one of the following codes (in consistence with the Description of the Action)
PU: Public, fully open, e.g. web
SEN: Sensitive, limited under conditions of the Grant Agreement

Table of Contents

Summary 4
1 Introduction 4
2 Description of Activities 4
2.1 Cell models and experimental design 4
2.2 Surfaceome enrichment and LC-MS/MS analysis 5
3 Results 5
3.1 Surface proteins identified across different MARV host cells 5
3.2 N-glyco sites discovered on MARV host factors, lectins and PS-binding proteins 7
4 Conclusion 8
5 References 9
6 Annexes 10
6.1 Annex 1 – Dataset metadata 10
6.2 Annex 2 – Representative dataset preview 11

Summary

This deliverable reports a comparative surfaceome analysis of **Huh7, THP-1, THP-1-M0 and THP-1-M2 cells**. The dataset identified **5,656 surface/glycoproteins** and revealed cell-type-specific expression of MARV-relevant host factors, lectins and PS-binding proteins. In addition, **29 potential N-glycosylation sites** were detected on 11 candidate proteins, providing an additional annotation layer for downstream COMBINE validation studies.

1 Introduction

WP2 (*Identification of the signature of MARV–host cell interaction*) aims to define viral and host-cell factors involved in Marburg virus (MARV) attachment, entry and cellular activation by liquid chromatography–tandem mass spectrometry (LC-MS/MS) -based proteomics. This includes the identification of tissue-specific MARV post-translational modification (PTM) profiles, analysis of MARV-induced host signalling networks, as well as the characterization of tissue-specific glycosylated host surface proteins (surfaceomes) and identification of host proteins involved in MARV–host cell attachment and entry.

Within this framework, Deliverable 2.2 focuses on defining the surfaceome of MARV-relevant host cells. The objective is to generate a tissue- and cell-type-specific host-cell surfaceome resource that supports the identification of candidate attachment factors, entry receptors or co-receptors involved in MARV–host cell interaction. The present deliverable reports the surfaceome analysis of liver- and macrophage-lineage cell models and provides protein-level identifications together with glycosite assignments as a basis for subsequent biological prioritisation and integration with other COMBINE datasets.

Particular attention is given to known or proposed MARV host factors, lectins, phosphatidylserine (PS)-binding proteins and experimentally assigned N-glycosylation sites, as these features may be relevant for MARV attachment and entry.

2 Description of Activities

2.1 Cell models and experimental design

Surfaceome analyses were performed for Huh7, THP-1, THP-1-M0 and THP-1-M2 cells. Huh7 cells represent a liver-cell model, while THP-1-derived cells represent macrophage-lineage conditions relevant to MARV host-cell interaction (**Table 1**). MUTZ-3 cells were not included because stable culture under the required conditions was not feasible. In general, cells were cultured in 10 cm^2 wells (6-wells) and multiple wells per condition were pooled (**Table 1**) to have enough starting material for glycoprotein enrichment. THP1 cells were differentiated into M0-like macrophages by incubation with PMA for 24 hours. THP1-M0 cells were further polarized into M2-like macrophages (M2a subtype) by incubation with IL4 and IL13 for 24 hours. Successful differentiation/polarization was verified by LC-MS data (2.2).

Table 1 – Overview of analysed cell models and LC-MS/MS acquisition design. *For each cell model, the number of pooled 10 cm² wells per biological replicate, polarization condition, biological interpretation, number of independent biological replicates and number of technical LC-MS/MS replicates are shown.*

Cell line	**10 cm² wells** (1 replicate)	**Polarization**	**As a model for**	**Biological replicates** (independent samples)	**Technical replicates** (LC-MS runs)
Huh7	4	-	Hepatocytes	3	2
THP1	2	-	Macrophage precursor / Monocytes	5	2
	2	M0	Macrophages	5	2
	2	M2	Anti-inflammatory Macrophages	5	2

2.2 Surfaceome enrichment and LC-MS/MS analysis

Cells were harvested and lysed in an SDS-based lysis buffer. Glycans of surface proteins/glycoproteins were oxidized by sodium periodate and bound to **hydrazide beads** (**Figure 1**) catalyzed by addition of imidazole. Non-glycosylated peptides were released upon a **trypsin** digest (fraction 1) while glycosylated peptides remained at the beads. N-glycosidic bonds of glycosylated peptides were cleaved by a **PNGaseF** digest, resulting in a second fraction of de-glycosylated peptides with a deamidation at the former glycosite. Both fractions were analyzed by **DDA LC-MS/MS on an Orbitrap Exploris** platform equipped with an UltiMate 3000 HPLC systems and a µPac Neo low load column, and applying a 100 min gradient of acetonitrile. Raw data were processed using the **deep-learning–assisted CHIMERYS algorithm** (*Proteome Discoverer* feature). Deamidation was included in the search to identify the glycosylation sites within the enriched proteins. Due to the high requirements for sample preparation, experiments on Huh7 and THP1-derived cells were performed separately followed by data processing in one batch.

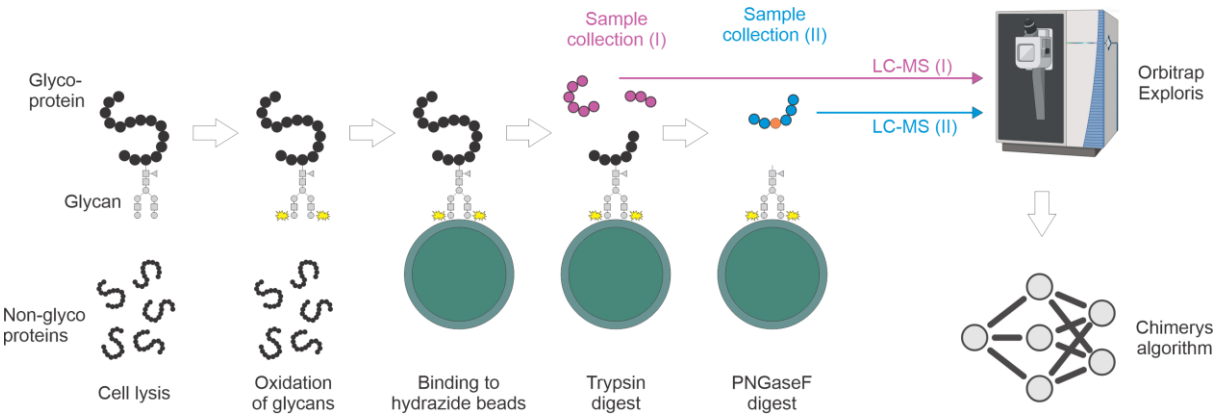


Figure 1 – Workflow for surfaceome analysis*. Oxidized glycoproteins are enriched by hydrazide beads and sequentially released as (de-glycosylated) peptides by trypsin and PNGaseF digests. The resulting two fractions are measured separately on the Orbitrap Exploris platform and recorded spectra are searched with CHIMERYS, a deep-learning–assisted algorithm.*

3 Results

3.1 Surface proteins identified across different MARV host cells

The list of identified proteins was filtered for high-confidence identifications and representative proteins (“Master”), while common lab contaminants were removed. The resulting noise-reduced dataset comprises 5,656 surface-/ glyco-proteins, of which 4,242 (75 %) were detected in all analyzed cell types (**Figure 2**). 741 proteins were found to be exclusively expressed in THP1-derived cells in contrast to only 61 proteins in Huh7 cells.

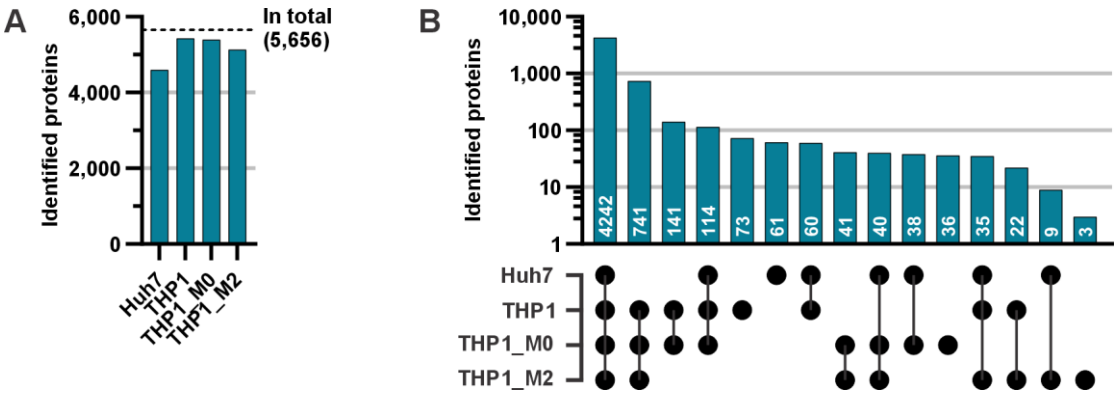


Figure 2 – Number and overlap of proteins identified in the surfaceomes of different MARV host-cell models. *All numbers were extracted after filtering for high-confidence master proteins and removal of common laboratory contaminants. (**A**) Number of identified surface/glycoproteins per analysed cell type. (**B**) UpSet plot showing shared and cell-type-specific protein identifications across Huh7, THP-1, THP-1-M0 and THP-1-M2 cells.*

To focus on proteins potentially relevant for viral uptake, the list of identified proteins was further filtered for known host factors as well as all proteins annotated as lectins (including C-type lectins / CLECs) or phosphatidylserine-binding (PS-receptors) (**Table 2**). In total, the surfaceome of Huh7 and THP1-derived cells comprised 143 Lectins, 26 PS-receptors (**Figure 3A**) and the eight MARV host factors CD209, AXL, MERTK, FOLR1, ASGR1, TYRO3, HAVCR1 and NPC1 (**Figure 3B**). Hierarchical clustering of protein abundances revealed that overall expression of lectins and PS-receptors was highly specific for the 2 tissue representative cell lines (**Figure 3C**). Furthermore, high expression of all detected host factors (except CD209) became visible in Huh7 cells and also comprised HAVCR1 (TIM1) which was not detectable at all in THP1-derived cell. For THP1-derived cells high expression was restricted to tyrosine kinases AXL and MERTK and especially the C-type lectin CD209. Of note, expression of these heavily increased upon polarization to M2-like macrophages and was barely detectable in undifferentiated THP1 cells.

Table 2 – Filters used to select candidate MARV-relevant surface proteins. *Gene names and annotations of public protein databases used to filter known MARV host factors, lectins and PS-binding proteins. Host factors were filtered based on published literature (known host factors listed). Lectins and PS-receptors were filtered using protein annotations from different databases (databases in parentheses, entries in angle brackets) or strings within the gene symbol.*

Subgroup	**Filter**
Host factors	ASGR1; AXL; CD209; CLEC10A; CLEC4M; FOLR1; HAVCR1; MERTK; NPC1; TYRO3
Lectins / CLECs	>**Lectin**< (Uniprot Keyword) >**REACTOME_C_TYPE_LECTIN_RECEPTORS_CLRS**< (GSEA) >**C-type_lectin-like/link_sf**< (Interpro) >**C-type_lectin-like**< (Interpro) >**C-type_lectin_CS**< (Interpro) "**CLEC**" (string inside gene name)
PS-receptors	>**Phosphatidylserine Binding**< (Gene Ontology, Molecular Function)

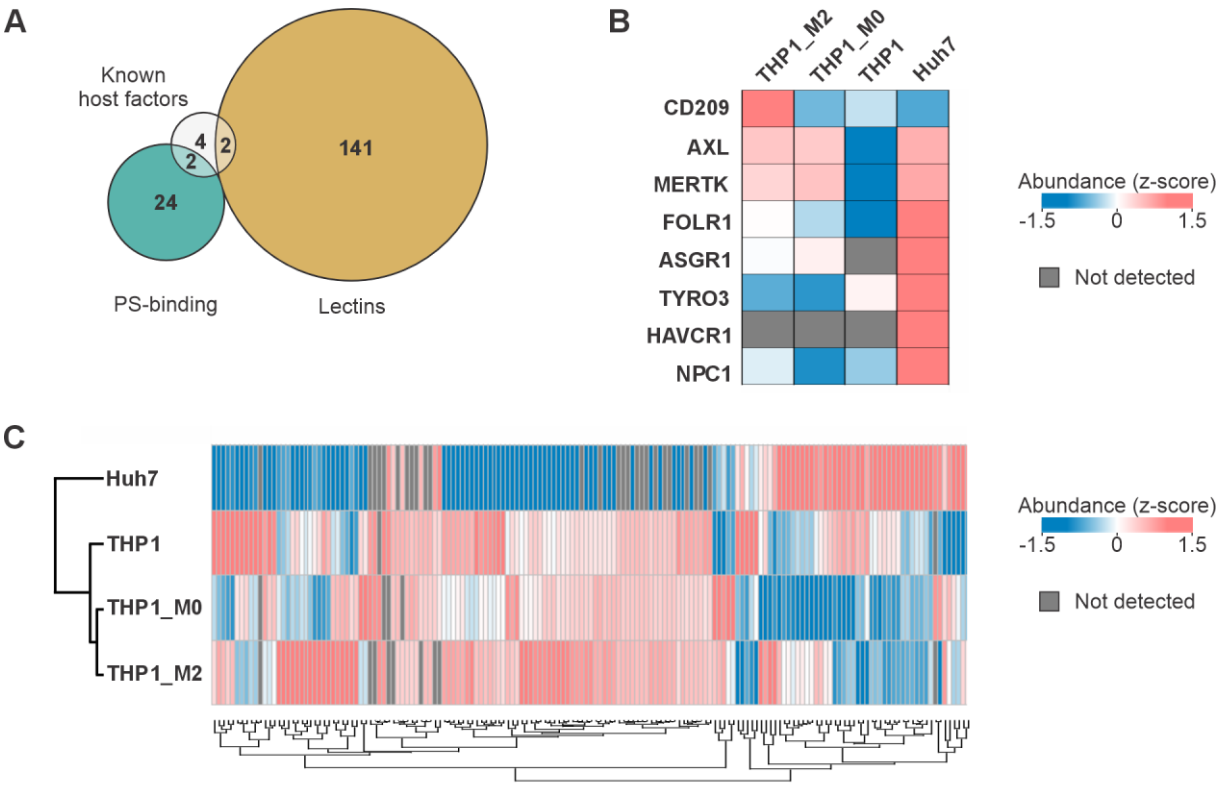


Figure 3 – Host factors detected in Huh7 and THP1-derived cells as detected by surfaceome analysis. *(**A**) Number of MARV host factors, Lectins and PS-receptors detected in Huh7 or THP1-derived cells. (**B-C**) Abundance of MARV host factors (B) as well as lectins and PS-receptors (C). Intensities are shown z-transformed upon hierarchical clustering (blue/red = low/high abundance; grey = not detected).*

The comparative surfaceome dataset provides a cell-type-specific inventory of surface/glycoproteins in liver- and macrophage-lineage models relevant to COMBINE. Such surfaceome information is needed to identify host-cell proteins that may contribute to MARV attachment, entry and tissue specificity. The filtered candidate list of known host factors, lectins and PS receptors therefore provides a focused set of proteins that can be cross-checked against MARV-binding/interactome data, infection susceptibility and downstream validation assays.

3.2 N-glyco sites discovered on MARV host factors, lectins and PS-binding proteins

The list of host factors, lectins and PS-receptors (**Figure 3**) was further filtered for proteins, in which deamidations were detected, as side products of N-glycan removal by PNGaseF. In total, we identified 11 proteins with 29 potential N-glycosylation sites, of which 7 have not been described so far according to entries in the Uniprot database (**Figure 4**).

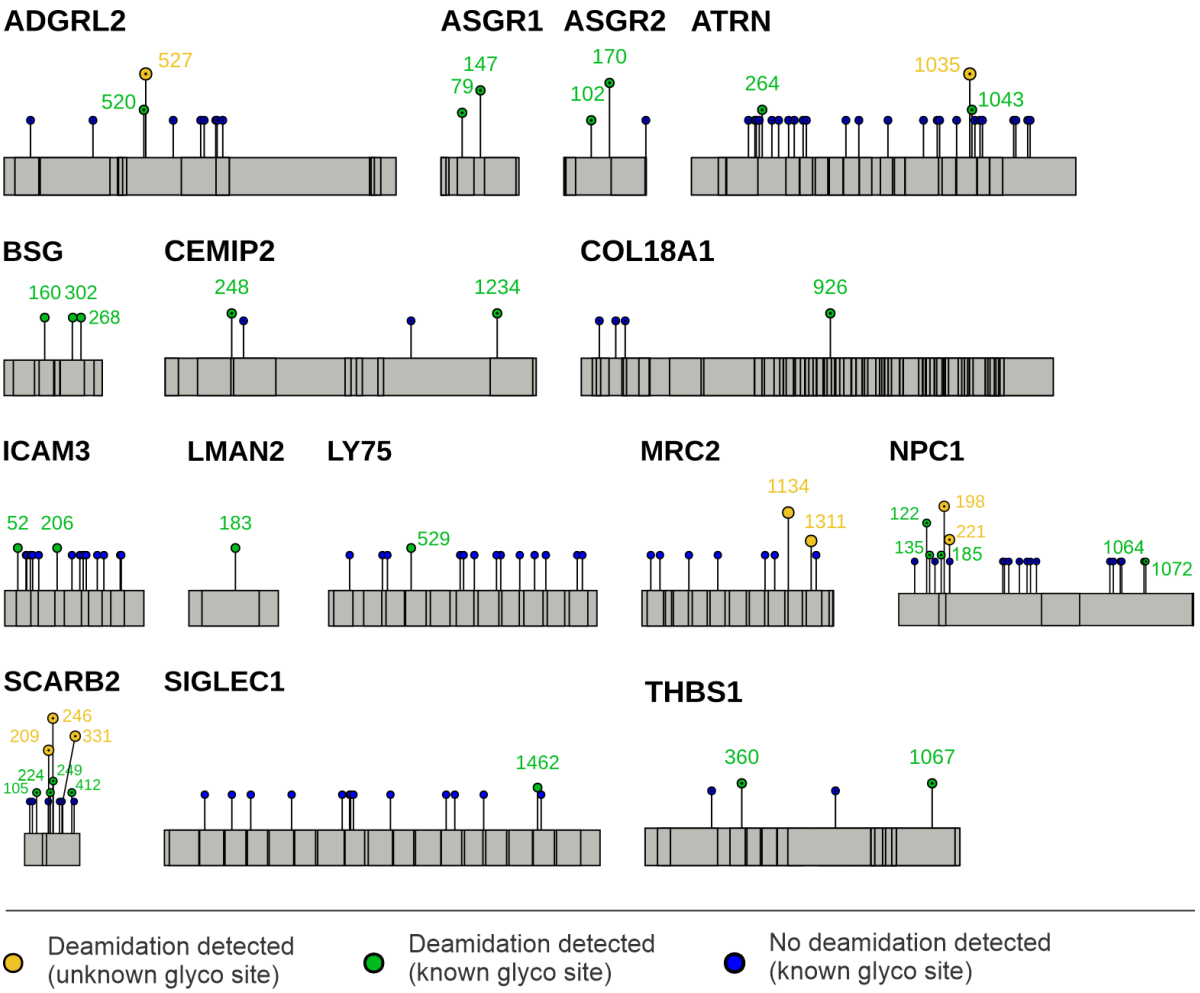


Figure 4 – N-glycosylation sites detected on candidate lectins and phosphatidylserine-binding proteins. *Protein schemes indicate annotated and experimentally detected N-glycosylation sites. Yellow circles indicate PNGaseF-dependent deamidation sites detected in this study that are not annotated as glycosites in UniProt; green circles indicate detected deamidation sites corresponding to known UniProt-annotated glycosites; blue circles indicate UniProt-annotated glycosites not detected in the present dataset.*

The N-glycosite data provide site-level information for a subset of the identified candidate surface proteins. According to the COMBINE proposal, host-cell surface glycosylation may contribute to viral entry and tissue specificity. Therefore, these glycosite assignments complement the comparative surface-protein lists and provide additional information for candidates that may later be evaluated together with MARV-binding, interactome and entry-dynamics data.

4 Conclusion

The surfaceome analysis generated a **cell-type-specific inventory of surface/glycoproteins** for Huh7 and THP1-derived cell models. This dataset is directly relevant for WP2 because it provides a focused resource to compare liver- and macrophage-lineage surface protein repertoires and to identify candidate host factors potentially involved in MARV attachment, entry or tissue specificity.

Filtering for known host factors, lectins and PS receptors revealed **distinct cell-type-specific expression patterns**. Huh7 cells showed high abundance of several known or proposed MARV host factors, including HAVCR1/TIM1, whereas THP1-derived cells showed stronger representation of selected macrophage-associated candidates such as AXL, MERTK and CD209, especially after M2-like polarization. These differences indicate that MARV-relevant host-cell interaction may rely on **cell-type-**

specific attachment and entry routes, providing a useful basis for selecting candidates that can be cross-checked against MARV-binding/interactome data, infection susceptibility and downstream validation assays.

Importantly, within the THP1-derived macrophage-lineage models, relevant host-factor expression was mainly observed in **M2-like macrophages**. This suggests that **M2-like macrophages may represent the more permissive macrophage polarization state for MARV infection**, whereas M1-like macrophages have been reported to be non-permissive for EBOV (Rogers et al., 2019; Wanninger *et al*., 2022). This observation may be relevant for the whole consortium, as it supports prioritizing **M2-like macrophages** for future MARV binding, infection, interactome and validation experiments.

The detected PNGaseF-dependent deamidation sites add **site-level N-glycosylation information** for a subset of candidate surface proteins. This complements the comparative surface-protein lists and provides an additional annotation layer for candidates that may later be evaluated in the context of MARV binding and entry.

One deviation from the original plan is that **MUTZ-3 cells were not included**, because stable culture under the required experimental conditions was not feasible. In addition, the current analysis focuses on **protein identification and N-glycosite assignment**; O-glycan-focused analysis and deeper biological prioritization are planned as subsequent data-analysis steps applying algorithms being part of available *MS Fragger* or *Peaks* software packages.

5 References

Rogers KJ, Brunton B, Mallinger L, et al. IL-4/IL-13 polarization of macrophages enhances Ebola virus glycoprotein-dependent infection. *PLoS Negl Trop Dis*. 2019;13(12):e0007819. Published 2019 Dec 11. doi:10.1371/journal.pntd.0007819

Wanninger TG, Millian DE, Saldarriaga OA, et al. Macrophage infection, activation, and histopathological findings in ebolavirus infection. *Front Cell Infect Microbiol*. 2022;12:1023557. Published 2022 Oct 12. doi:10.3389/fcimb.2022.1023557

6 Annexes

A representative excerpt of processed data (database-search results) for D2.2 is provided as an annex to this report. The dataset contains protein-level identifications and PNGaseF-dependent N-glycosite assignments from Huh7, THP-1, THP-1-M0 and THP-1-M2 surfaceome samples. **The data will be made publicly available in an appropriate repository in conjunction with the corresponding scientific publication.**

6.1 Annex 1 – Dataset metadata

Metadata field	**Description**
Dataset name	COMBINE D2.2 host surfaceome dataset
Dataset type	Processed LC-MS/MS database-search results
Cell models	Huh7, THP-1, THP-1-M0 and THP-1-M2
Sample type	Surfaceome-enriched glycoprotein samples
Sample preparation	Hydrazide bead-based enrichment followed by sequential trypsin and PNGaseF digestion
LC-MS/MS platform	Orbitrap Exploris operated in DDA mode
Data-processing software	CHIMERYS / Proteome Discoverer
Dimensions	518 columns, 8,604 proteins
Main data content	Protein-level identifications, abundance information and PNGaseF-dependent N-glycosite assignments
Main identifier fields	Protein accession, gene name and protein description
File format	txt file, comma-separated values
Data availability	Full dataset to be made publicly available in conjunction with the corresponding scientific publication

6.2 Annex 2 – Representative dataset preview

Because of the size of the full dataset, only a representative preview with selected columns is shown below.

Gene Symbol	Accession	Protein Group FDR Confidence: Combined	Protein FDR Confidence: Combined	Master	Contaminant	Abundances (Grouped): Huh7, PNGaseF	Abundances (Grouped): Huh7, Trypsin	Abundances (Grouped): THP1, PNGaseF	Abundances (Grouped): THP1, Trypsin	Abundances (Grouped): THP1_M0, PNGaseF	Abundances (Grouped): THP1_M0, Trypsin	Abundances (Grouped): THP1_M2, PNGaseF	Abundances (Grouped): THP1_M2, Trypsin	Deamidated (N) Positions
MYH9	P35579	High	High	Master Protein	FALSE		17618.23726		587200.7418	2532.821452	700683.2258		1517973.605	
LRP1	Q07954	High	High	Master Protein	FALSE	5599.744504	1519968.167	1971.942383	186220.4526		654114.7208		864668.8511	Deamidated [N2472; N2475; N2486; N2502; N2815; N3333; N3477; N3488; N3662; N3788; N3839; N4364; N4383]
AHNAK	Q09666	High	High	Master Protein	FALSE	3627.209064	2670.378246	2902.528646	277903.9404	4699.845703	136075.7612		382550.1749	
IGF2R	P11717	High	High	Master Protein	FALSE		907046.0807	5350.833976	7142625.813	10880.82845	3488258.516	9178.683763	6056711.114	Deamidated [N112; N123; N435; N543; N581; N747; N793; N871; N1656; N2136; N2141; N2144]
APOB	P04114	High	High	Master Protein	FALSE	1471962.516	1414603.784	2090.46582	10168.93221	3050.644091	9629.403619	2014.262695		Deamidated [N1523; N2982; N3411; N3465; N3895]
FLNA	P21333	High	High	Master Protein	FALSE		12650.84499		664771.9393		793810.2252		2149538.702	
GLG1	Q92896	High	High	Master Protein	FALSE		2171252.544		2012164.663		1186799.887		1502195.421	
HYOU1	Q9Y4L1	High	High	Master Protein	FALSE	311526.4598	6892843.828	38911.65359	16686989.66	98731.85069	6702189.256	75987.4991	10662699.07	Deamidated [N544; N596; N830; N862; N869; N922; N931]
DYNC1H1	Q14204	High	High	Master Protein	FALSE	3487.709961	489121.0787		131681.2123		48038.80563		115934.3437	
HSP90B1	P14625	High	High	Master Protein	FALSE		4303504.292	5520.050549	13144587.99	40715.99446	6646972.711	12796.19108	11387790.2	Deamidated [N62; N217; N228; N337; N445]
TLN1	Q9Y490	High	High	Master Protein	FALSE	2785788.072	247682.0092	1282.74837	207099.5039	16195.55722	326888.3159	18507.61979	771863.2512	
ITGB2	P05107	High	High	Master Protein	FALSE	14255.25952	15878.46481	6519.490679	9738872.884	28954.31121	8883976.608	21538.2952	14498217.23	Deamidated [N92; N212; N213; N215; N642]
PRKDC	P78527	High	High	Master Protein	FALSE		13258.76866	3409.010782	350060.2261	4466.070693	240491.7978	2713.729869	386747.5777	
NUP210	Q8TEM1	High	High	Master Protein	FALSE	176469.7866	456600.2726	19137.02758	7430364.312	845276.5748	2018961.04	644682.9819	2294963.551	Deamidated [N337; N405; N926]
PLEC	Q15149	High	High	Master Protein	FALSE		185202.3503		367528.384	2998.055339	158814.1946		209712.3839	
RPN1	P04843	High	High	Master Protein	FALSE	6855.761068	5032798.292	6632.121642	11413966.87	18321.28906	7097455.76	14229.74267	8751607.952	
FASN	P49327	High	High	Master Protein	FALSE	976.7114992	135338.1935	301.3542277	167944.8112		110865.4186		179068.644	
UGGT1	Q9NYU2	High	High	Master Protein	FALSE	952.1630421	102654.3454	3100.835612	781673.8252		317152.7026		554557.7068	
PECAM1	P16284	High	High	Master Protein	FALSE	8681.318987	374582.542	5543.787109	1702948.504	4523.852214	3208987.398	3586.032933	3272700.619	
FN1	P02751	High	High	Master Protein	FALSE	2884.398688	1477029.542	38409.5022	1209638.402	19681.80314	496859.5044	6224.111378	1816304.271	Deamidated [N528; N1007]

Funded by
the European Union