SUPPLEMENTAL DATA

An advanced strategy for comprehensive profiling of ADP-ribosylation sites using mass spectrometry-based proteomics

Ivo A. Hendriks[1,2], Sara C. Larsen[1,2], and Michael L. Nielsen[1,*]

[1]Proteomics program, Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Blegdamsvej 3B, 2200 Copenhagen, Denmark

[2]These authors contributed equally

*Correspondence to: michael.lund.nielsen@cpr.ku.dk

SUPPLEMENTAL DATA INVENTORY

The Supplemental Data contain:

- Five (5) supplemental figures and legends
- Seven (7) supplemental tables and legends
- Supplemental references

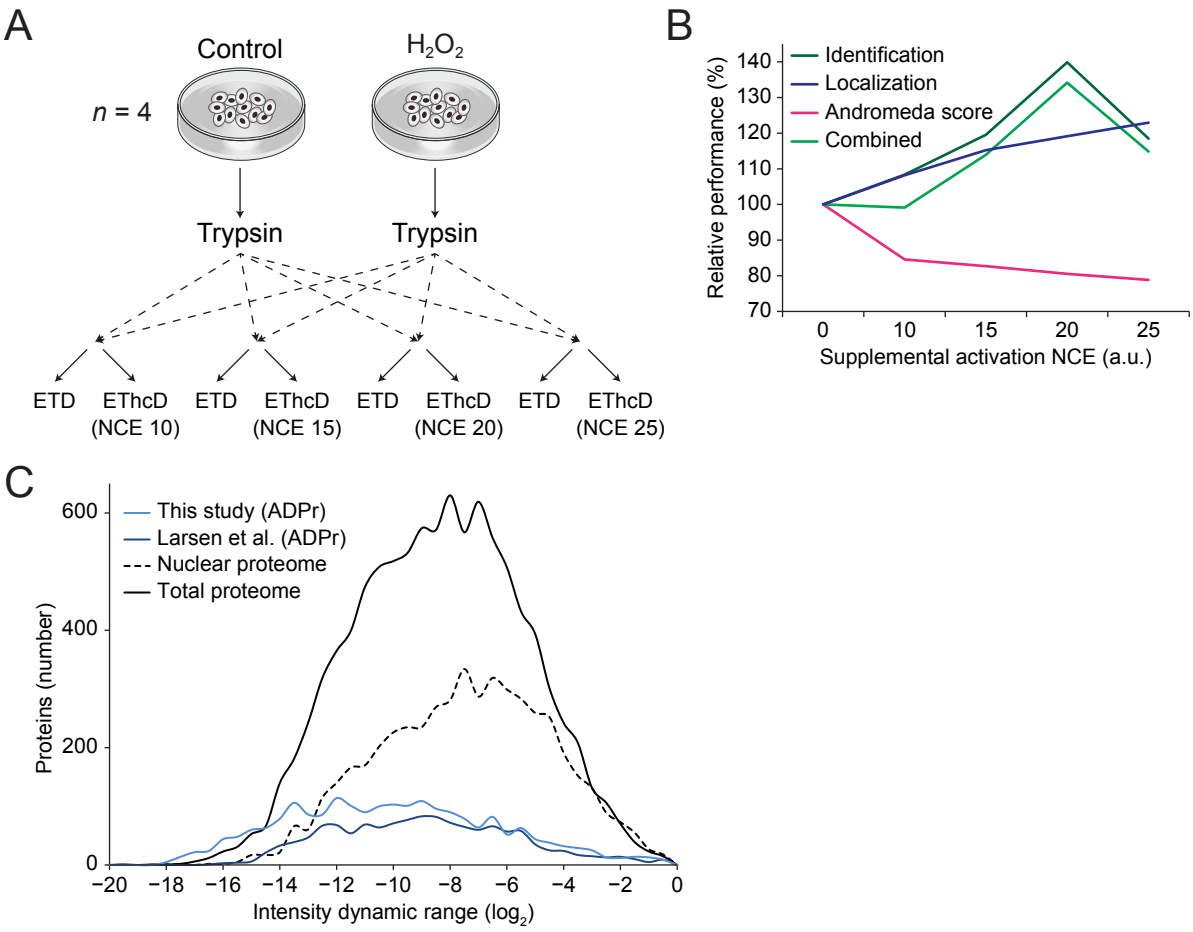
A
Control
H2O2
n = 4
Trypsin
Trypsin
ETD
EThcD
(NCE 10)
ETD
EThcD
(NCE 15)
ETD
EThcD
(NCE 20)
ETD
EThcD
(NCE 25)
B
Identification
Localization
Andromeda score
Combined
Relative performance (%)
140
130
120
110
100
90
80
70
0
10
15
20
25
Supplemental activation NCE (a.u.)
C
This study (ADPr)
Larsen et al. (ADPr)
Nuclear proteome
Total proteome
Proteins (number)
600
400
200
0
−20
−18
−16
−14
−12
−10
−8
−6
−4
−2
0
Intensity dynamic range (log2)

FIGURE S1

Figure S1. (A) Experimental design for initial experiments performed to evaluate different levels of EThcD supplemental activation (SA) energy. Cell cultures were performed in quadruplicate; *n*=4, and all cells were either left untreated or treated with H_2O_2 to induce ADP-ribosylation. All samples were prepared through the trypsin workflow, and final samples were split in halves and measured back-to-back as outlined. (B) Schematic overview of various attributes as profiled across the different SA energies. Values represent the number of total ADP-ribosylated peptide identifications, the average ADP-ribose localization probabilities, and the average Andromeda scores, in each instance normalized between the EThcD half of the sample and the ETD half of the sample. The combined score was derived by multiplication of the three other values. (C) Dynamic depth of sequencing graph depicting the range of protein intensity values determined in this study, plotted against the number of proteins identified, and compared to various other studies[1-2]. Protein intensity values were derived from all peak intensities of all corresponding peptides. All studies were normalized to have the most-abundant proteins align at the right side of the plot.

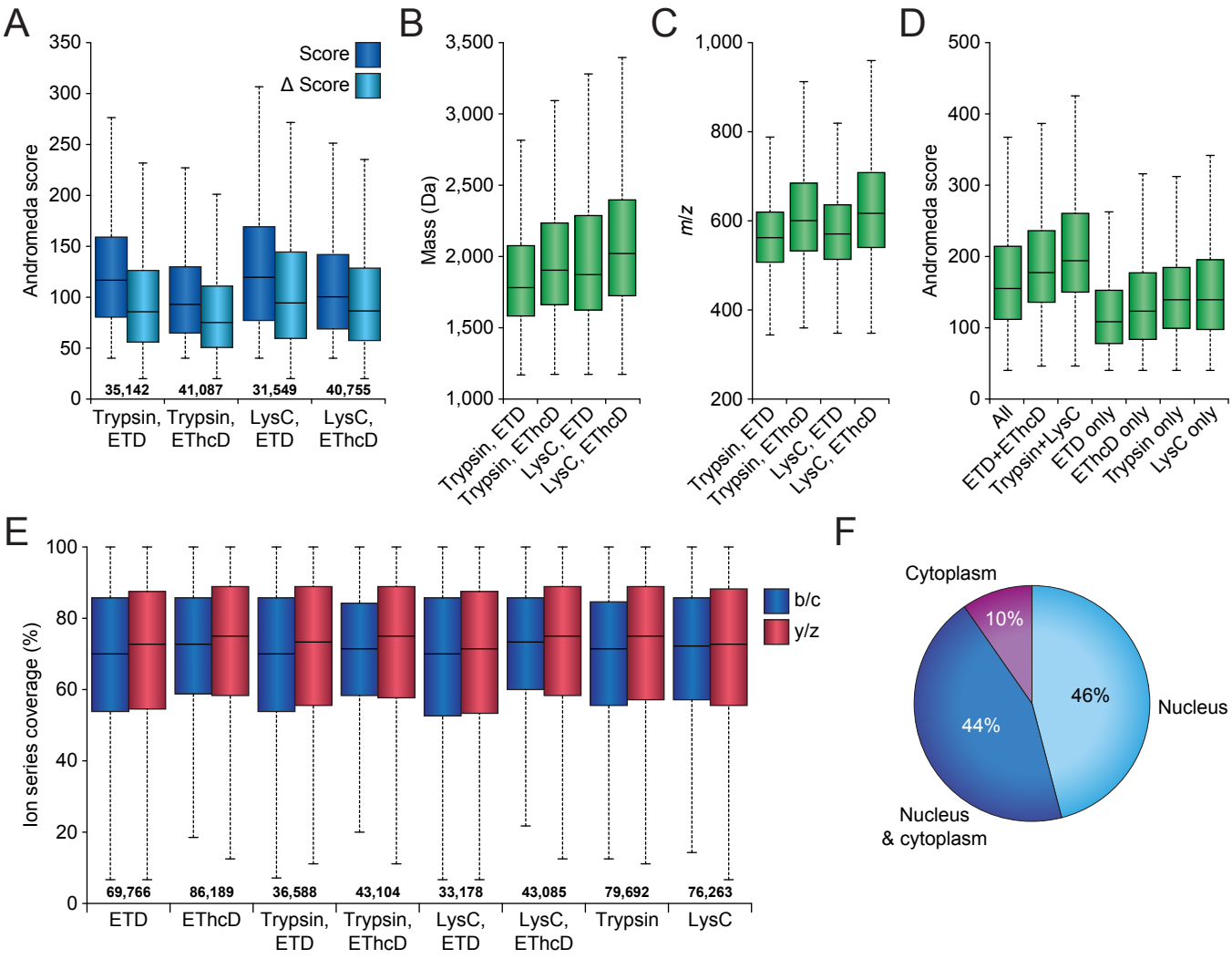
A
Andromeda score
Score
Δ Score
35,142
41,087
31,549
40,755
Trypsin, ETD
Trypsin, EThcD
LysC, ETD
LysC, EThcD
B
Mass (Da)
C
m/z
D
Andromeda score
All
ETD+EThcD
Trypsin+LysC
ETD only
EThcD only
Trypsin only
LysC only
E
Ion series coverage (%)
b/c
y/z
69,766
86,189
36,588
43,104
33,178
43,085
79,692
76,263
ETD
EThcD
Trypsin, ETD
Trypsin, EThcD
LysC, ETD
LysC, EThcD
Trypsin
LysC
F
Cytoplasm
10%
46%
Nucleus
44%
Nucleus & cytoplasm

FIGURE S2

Figure S2. (A) Boxplot graph visualizing distribution of the Andromeda score and delta score for ADP-ribosylation PSMs, as measured across the four experimental conditions. Whiskers; 1.5× interquartile range (IQR), box limits; 3rd and 1st quartiles, center bar; median. Numbers set above the horizontal axis correspond to the number of identified ADP-ribosylation PSMs. (B) As **A**, but for peptide mass. (C) As **A**, but for precursor *m*/*z*. (D) As **A**, but for ADP-ribosylation sites, and additionally for subsets of sites identified through multiple experimental conditions, or exclusive to certain experimental conditions. (E) Ion series coverage analysis, displaying the fraction of the maximum observable ion series for which peaks were detected and assigned, divided between C-terminal (b/c) and N-terminal (y/z) fragment series. Whiskers; 1.5× interquartile range (IQR), box limits; 3rd and 1st quartiles, center bar; median. Numbers set above the horizontal axis correspond to the number of identified ADP-ribosylation PSMs. (F) Pie-chart visualization of the subcellular localization of proteins identified to be ADP-ribosylated.

Rank	Motif	Score	Foreground matches	Foreground size	Background matches	Background size	Fold increase
1	....KSG....	615.3	190	6,245	298	46,045	4.70
2	....KS.P...	320.5	145	6,055	262	45,747	4.18
3	....KS.G...	321.0	153	5,910	293	45,485	4.02
4	....KSS....	316.5	207	5,757	528	45,192	3.08
5	....KS.A...	317.0	118	5,550	267	44,664	3.56
6	....RSG....	615.3	171	5,432	319	44,397	4.38
7	....KS.....	307.7	949	5,261	4,135	44,078	1.92
8	....RSS....	314.0	144	4,312	631	39,943	2.11
9	....GSG....	317.8	128	4,168	198	39,312	6.10
10	.....SG....	307.7	631	4,040	1,918	39,114	3.19
11	....RS.....	307.7	645	3,409	4,565	37,196	1.54
12	.....SS....	307.7	542	2,764	4,043	32,631	1.58
13	....GS.....	307.7	351	2,222	1,983	28,588	2.28
14	.....SA....	307.7	284	1,871	1,685	26,605	2.40
15	.....S.G...	307.7	201	1,587	1,248	24,920	2.53
16	.....S.P...	307.7	181	1,386	1,431	23,672	2.16
17	....AS.....	13.6	171	1,205	1,722	22,241	1.83
18	....VS.....	14.3	134	1,034	1,296	20,519	2.05
19	....LS.....	8.2	167	900	2,295	19,223	1.55
20	.....S.K...	6.7	158	733	2,456	16,928	1.49

FIGURE S3

Figure S3. Motif-X analysis of the sequence context surrounding serine ADP-ribosylation sites. Foreground, *n*=6,245 ADP-ribosylation serine-centered sites, background, *n*=46,045 serine-centered sites from the same proteins. The minimum required sequences for a motif were set at 100, and the $p<1*10^{-6}$ cut-off was used.

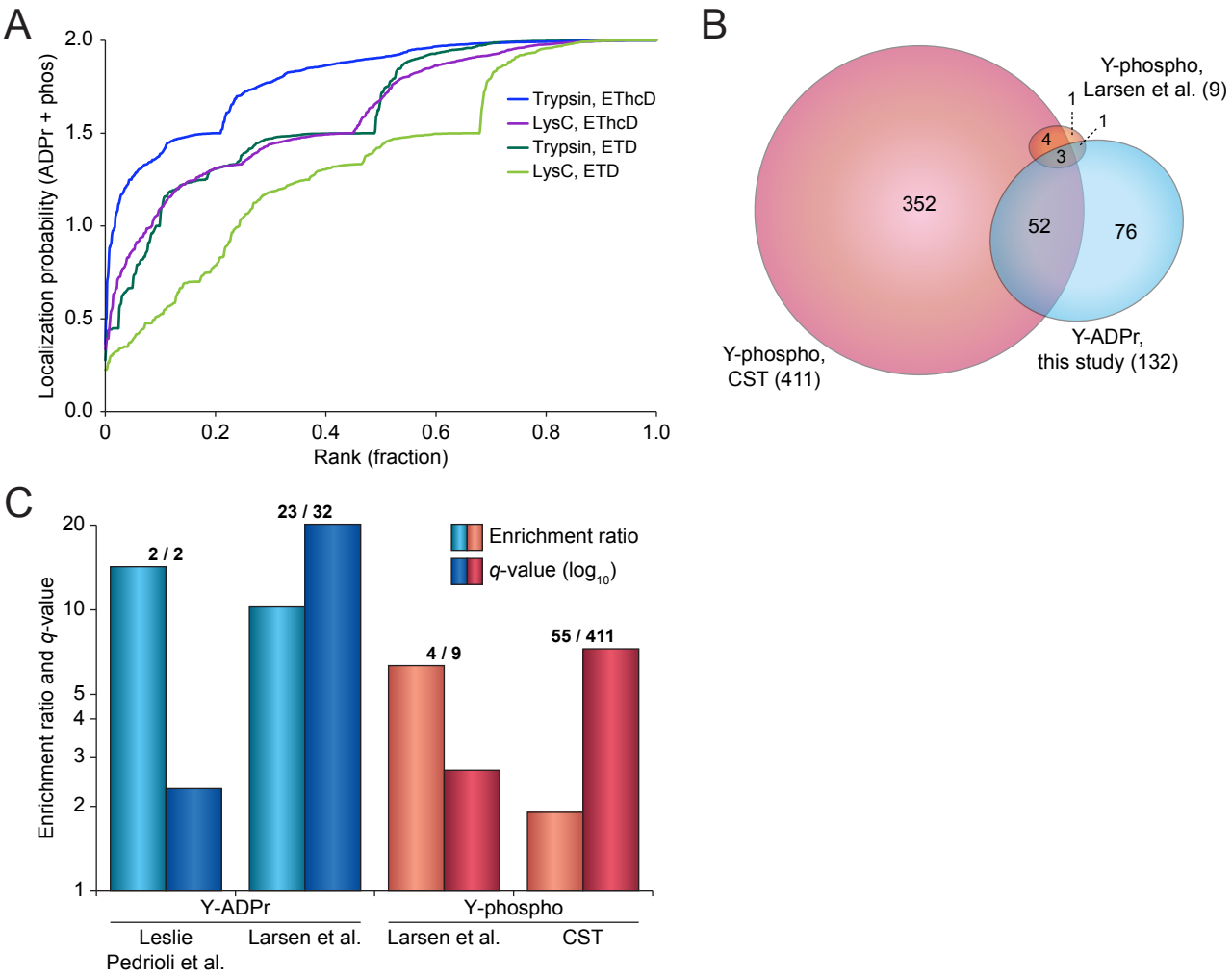
A
Localization probability (ADPr + phos)
2.0
1.5
1.0
0.5
0.0
0
0.2
0.4
0.6
0.8
1.0
Rank (fraction)
Trypsin, EThcD
LysC, EThcD
Trypsin, ETD
LysC, ETD
B
Y-phospho,
Larsen et al. (9)
1
1
4
3
352
52
76
Y-phospho,
CST (411)
Y-ADPr,
this study (132)
C
Enrichment ratio and q-value
20
10
5
4
3
2
1
2 / 2
23 / 32
4 / 9
55 / 411
Enrichment ratio
q-value (log10)
Y-ADPr
Y-phospho
Leslie
Pedrioli et al.
Larsen et al.
Larsen et al.
CST

FIGURE S4

Figure S4. (A) Combined ADP-ribosylation and phospho localization probability plotted against the ranked fraction of all peptide-spectrum-matches (PSMs) resulting from the four different approaches. Individual probabilities for ADP-ribosylation and phospho were summed, and co-modified peptides and sites were only derived when both PTMs were individually localized at over 0.9 probability. (B) Scaled Venn diagram visualizing overlap between tyrosine ADP-ribosylation sites identified in this study, and tyrosine phosphorylation identified in one previous study[1], or those identified and reported by Cell Signaling Technology (CST)[3]. (C) PTM co-targeting enrichment analysis, depicting overlap and significance of overlap between tyrosine ADP-ribosylation sites identified in this study, tyrosine ADP-ribosylation sites identified in two other studies[1, 4], and tyrosine phosphorylation identified in one previous study[1] or by Cell Signaling Technology[3]. All tyrosine residues within the ADP-ribosylated proteins were used as a background, and enrichment ratios and *q*-values were determined through Fisher Exact Testing with Benjamini-Hochberg correction. Numbers set above the bars correspond to the intersection between tyrosine ADP-ribosylation identified in this study, in relation to the total number of tyrosine PTMs reported in the other studies.

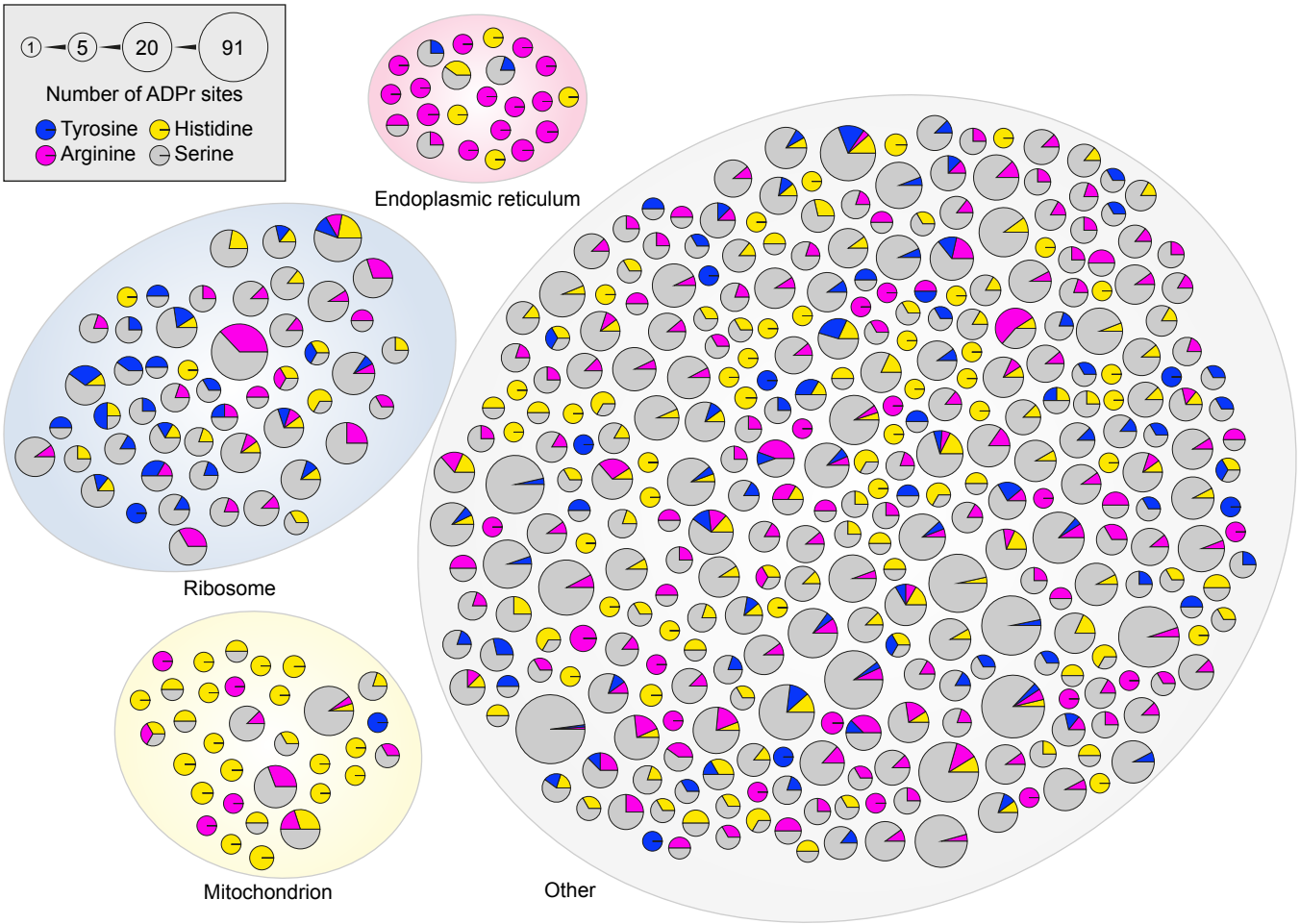
1
5
20
91
Number of ADPr sites
Tyrosine
Histidine
Arginine
Serine
Endoplasmic reticulum
Ribosome
Mitochondrion
Other

FIGURE S5

Figure S5. STRING network analysis of all ADP-ribosylated proteins modified on at least one histidine, arginine, or tyrosine residue. Default STRING clustering settings were used (clustering confidence >0.4), proteins not clustered were omitted from the network, and edges were hidden. Proteins were curated based on the number of ADP-ribosylation sites and the type of residues modified by ADP-ribosylation, as indicated. ADP-ribosylation sites on amino acid residues other than serine, histidine, arginine, and tyrosine, were omitted. Proteins annotated as part of the mitochondrion, ribosome, or endoplasmic reticulum, were clustered and moved out of the main network to assist visualization.

SUPPLEMENTAL TABLE LEGENDS

Supplemental Table 1

A list of all 11,265 unique modified peptides identified with localized ADP-ribosylation sites. Experiment-specific information is included.

Supplemental Table 2

A list of all 7,040 localized ADP-ribosylation sites. Fractional modification analysis and experiment-specific information are included.

Supplemental Table 3

A list of all 2,149 ADP-ribosylation target proteins. Numbers of ADP-ribosylation sites, site positions, and experiment-specific information are included.

Supplemental Table 4

An overview and comparison of 7,681 ADP-ribosylation sites identified across our study and three other ADP-ribosylation proteomics studies.

Supplemental Table 5

The complete human proteome, annotated with ADP-ribosylation target proteins identified in this study, as well as in five other ADP-ribosylation proteomics studies. Includes protein copy-numbers as derived from a deep proteome study, and all proteins are annotated with Gene Ontology, Uniprot keywords, Pfam, Interpro, and CORUM terms.

Supplemental Table 6

A collection of all statistical information related to term enrichment analyses performed throughout this study.

Supplemental Table 7

A list of all 236 unique peptides and 152 sites co-modified by ADP-ribosylation and phosphorylation. Experiment-specific information is included.

SUPPLEMENTAL REFERENCES

1. Larsen, S. C.; Hendriks, I. A.; Lyon, D.; Jensen, L. J.; Nielsen, M. L., Systems-wide Analysis of Serine ADP-Ribosylation Reveals Widespread Occurrence and Site-Specific Overlap with Phosphorylation. *Cell reports* **2018,** *24* (9), 2493-2505.e4.
2. Bekker-Jensen, D. B.; Kelstrup, C. D.; Batth, T. S.; Larsen, S. C.; Haldrup, C.; Bramsen, J. B.; Sorensen, K. D.; Hoyer, S.; Orntoft, T. F.; Andersen, C. L.; Nielsen, M. L.; Olsen, J. V., An Optimized Shotgun Strategy for the Rapid Generation of Comprehensive Human Proteomes. *Cell systems* **2017,** *4* (6), 587-599.e4.
3. Hornbeck, P. V.; Zhang, B.; Murray, B.; Kornhauser, J. M.; Latham, V.; Skrzypek, E., PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic acids research* **2015,** *43* (Database issue), D512-20.
4. Leslie Pedrioli, D. M.; Leutert, M.; Bilan, V.; Nowak, K.; Gunasekera, K.; Ferrari, E.; Imhof, R.; Malmstrom, L.; Hottiger, M. O., Comprehensive ADP-ribosylome analysis identifies tyrosine as an ADP-ribose acceptor site. *EMBO reports* **2018,** *19* (8).