

Supplementary Data

The Human Proteoform Atlas: A FAIR community resource for experimentally derived proteoforms

Michael A. R. Hollas¹, Matthew T. Robey¹, Ryan T. Fellers¹, Richard D. LeDuc¹, Paul M. Thomas¹ and Neil L. Kelleher¹

¹ Departments of Molecular Biosciences, Chemistry, and the Chemistry of Life Processes Institute, Northwestern University, Evanston, IL, 60208, USA

Table S1. Datasets currently contained in the Human Proteoform Atlas.

PI	Dataset Title	Year	Enhanced?	Proteins	Proteoforms	PMIDs	Refs
C. Borges	Gamma-carboxylation and fragmentation of osteocalcin in human serum defined by mass spectrometry	2015		1	13	25855755	(1)
C. Hirtz	Identification of multiple proteoforms biomarkers on clinical samples by routine Top-Down approaches	2018		3	7	29900270	(2)
C. Hendrickson	Identification and Characterization of Human Proteoforms by Top-Down LC-21 Tesla FT-ICR Mass Spectrometry	2017		591	805	27936753	(3)
D. Strasser	Immunoaffinity Targeted Mass Spectrometry Analysis of Human Plasma Samples Reveals an Imbalance of Active and Inactive CXCL10 in Primary Sjögren's Syndrome Disease Patients	2020		1	2	32870689	(4)
D. Nedelkov	Mass Spectrometric Immunoassay for the qualitative and quantitative analysis of the cytokine Macrophage Migration Inhibitory Factor (MIF)	2014		1	3	25328446	(5)
D. Nedelkov	The association of plasma cystatin C proteoforms with diabetic chronic kidney disease	2016		1	5	27019641	(6)
D. Nedelkov	Quantitative mass spectrometric immunoassay for the chemokine RANTES and its variants	2015		1	11	25549571	(7)
D. Nedelkov	The Association of Human Apolipoprotein C-III Sialylation Proteoforms with Plasma Triglycerides	2015		3	7	26633899	(8)
D. Nedelkov	Development of quantitative mass spectrometric immunoassay for Serum Amyloid A	2016		3	15	27308834	(9)
F. Benavente	A chemometric approach for characterization of serum transthyretin in familial amyloidotic polyneuropathy type I (FAP-I) by electrospray ionization-ion mobility mass spectrometry	2018		1	2	29426546	(10)

F. Becher	Top-Down and Bottom-Up Proteomics of Circulating S100A8/S100A9 in Plasma of Septic Shock Patients	2020		2	4	31913637	(11)
H. Yassine	Posttranslational modifications of apolipoprotein A-II proteoforms in type 2 diabetes	2017		1	3	27578111	(12)
I. Messina	Proteomic investigation of whole saliva in Wilson's disease	2015		21	42	26254010	(13)
J. Brodbelt	High-Throughput Analysis of Intact Human Proteins Using UVPD and HCD on an Orbitrap Mass Spectrometer	2017	*	299	1166	28412815	(14)
L. Sun	Toward a Universal Sample Preparation Method for Denaturing Top-Down Proteomics of Complex Proteomes	2020		252	543	32419461	(15)
L. Smith	Constructing Human Proteoform Families Using Intact-Mass and Top-Down Proteomics with a Multi-Protease Global Post-Translational Modification Discovery Database	2020	*	418	2301	31479276	(16)
L. Smith	Construction of Human Proteoform Families from 21 Tesla Fourier Transform Ion Cyclotron Resonance Mass Spectrometry Top-Down Proteomic Data	2020	*	374	1666	33074679	(17)
N. Kelleher	Haematopoietic proteoforms (unpublished)	2021	*	1724	27660	-	-
N. Kelleher	Quantitation and Identification of Thousands of Human Proteoforms below 30 kDa	2016		1274	5936	26795204	(18)
N. Kelleher	Evaluation of the compact high-field orbitrap for top-down proteomics of human cells	2012		670	672	22746247	(19)
N. Kelleher	Integrated Bottom-Up and Top-Down Proteomics of Patient-Derived Breast Tumor Xenografts	2016		725	3747	26503891	(20)
N. Kelleher	Advancing top-down analysis of the human proteome using a benchtop quadrupole-Orbitrap mass spectrometer	2017	*	454	1855	28152595	(21)
N. Young	Quantitative Mass Spectrometry Reveals that Intact Histone H1 Phosphorylations are Variant Specific and Exhibit Single Molecule Hierarchical Dependence	2016		3	15	26209608	(22)
P. Reaven	Disialylated apolipoprotein C-III proteoform is associated with improved lipids in prediabetes and type 2 diabetes	2016		1	5	26945091	(23)
P. Lescuyer	Identification of hemoglobin variants by top-down mass spectrometry using selected diagnostic product ions	2015		1	5	25753013	(24)
R. Oliva	Characterization of Human Sperm Protamine Proteoforms through a Combination of Top-Down and Bottom-Up Mass Spectrometry Approaches	2020		2	24	31703166	(25)
S. Wu	Deep Intact Proteoform Characterization in Human Cell Lysate Using High-pH and Low-pH Reversed-Phase Liquid Chromatography	2019		658	2726	31755044	(26)
T. Cabras	RP-HPLC-ESI-IT Mass Spectrometry Reveals Significant Variations of the Human Salivary Protein Profile Associated with Predominantly Antibody Deficiencies	2020	*	29	250	31916122	(27)

V. Sanz-Nebot	On-line immunoaffinity solid-phase extraction capillary electrophoresis mass spectrometry using Fab' antibody fragments for the analysis of serum transthyretin	2017	1	7	28501163	(28)
V. Sanz-Nebot	Comparison of capillary electrophoresis and capillary liquid chromatography coupled to mass spectrometry for the analysis of transthyretin in human serum	2016	1	14	27052822	(29)
V. Sanz-Nebot	Analysis of serum transthyretin by on-line immunoaffinity solid-phase extraction capillary electrophoresis mass spectrometry using magnetic beads	2016	1	6	26842820	(30)
Y. Ge	(Skeletal muscle) Dissecting human skeletal muscle troponin proteoforms by top-down mass spectrometry	2015	4	6	25613324	(31)
Y. Ge	(Vastus lateralis) Dissecting human skeletal muscle troponin proteoforms by top-down mass spectrometry	2015	7	12	25613324	(31)
Y. Ge	(Tibialis anterior) Dissecting human skeletal muscle troponin proteoforms by top-down mass spectrometry	2015	7	10	25613324	(31)
Y. Zhu	Sensitive Top-Down Proteomics Analysis of a Low Number of Mammalian Cells Using a Nanodroplet Sample Processing Platform	2020	1655	2735	32374172	(32)

REFERENCES

1. Rehder, D.S., Gundberg, C.M., Booth, S.L. and Borges, C.R. (2015) Gamma-carboxylation and fragmentation of osteocalcin in human serum defined by mass spectrometry. *Mol Cell Proteomics*, **14**, 1546-1555.
2. Vialaret, J., Schmit, P.O., Lehmann, S., Gabelle, A., Wood, J., Bern, M., Paape, R., Suckau, D., Kruppa, G. and Hirtz, C. (2018) Identification of multiple proteoforms biomarkers on clinical samples by routine Top-Down approaches. *Data Brief*, **18**, 1013-1021.
3. Anderson, L.C., DeHart, C.J., Kaiser, N.K., Fellers, R.T., Smith, D.F., Greer, J.B., LeDuc, R.D., Blakney, G.T., Thomas, P.M., Kelleher, N.L. *et al.* (2017) Identification and Characterization of Human Proteoforms by Top-Down LC-21 Tesla FT-ICR Mass Spectrometry. *J Proteome Res*, **16**, 1087-1096.
4. Schmidt, A., Farine, H., Keller, M.P., Sebastian, A., Kozera, L., Welford, R.W.D. and Strasser, D.S. (2020) Immunoaffinity Targeted Mass Spectrometry Analysis of Human Plasma Samples Reveals an Imbalance of Active and Inactive CXCL10 in Primary Sjögren's Syndrome Disease Patients. *J Proteome Res*, **19**, 4196-4209.
5. Sherma, N.D., Borges, C.R., Trenchevska, O., Jarvis, J.W., Rehder, D.S., Oran, P.E., Nelson, R.W. and Nedelkov, D. (2014) Mass Spectrometric Immunoassay for the qualitative and quantitative analysis of the cytokine Macrophage Migration Inhibitory Factor (MIF). *Proteome Sci*, **12**, 52.
6. Yassine, H.N., Trenchevska, O., Dong, Z., Bashawri, Y., Koska, J., Reaven, P.D., Nelson, R.W. and Nedelkov, D. (2016) The association of plasma cystatin C proteoforms with diabetic chronic kidney disease. *Proteome Sci*, **14**, 7.
7. Trenchevska, O., Sherma, N.D., Oran, P.E., Reaven, P.D., Nelson, R.W. and Nedelkov, D. (2015) Quantitative mass spectrometric immunoassay for the chemokine RANTES and its variants. *J Proteomics*, **116**, 15-23.
8. Yassine, H.N., Trenchevska, O., Ramrakhiani, A., Parekh, A., Koska, J., Walker, R.W., Billheimer, D., Reaven, P.D., Yen, F.T., Nelson, R.W. *et al.* (2015) The Association of Human Apolipoprotein C-III Sialylation Proteoforms with Plasma Triglycerides. *PLoS One*, **10**, e0144138.

9. Trenchevska, O., Yassine, H.N., Borges, C.R., Nelson, R.W. and Nedelkov, D. (2016) Development of quantitative mass spectrometric immunoassay for serum amyloid A. *Biomarkers*, **21**, 743-751.
10. Pont, L., Sanz-Nebot, V., Vilaseca, M., Jaumot, J., Tauler, R. and Benavente, F. (2018) A chemometric approach for characterization of serum transthyretin in familial amyloidotic polyneuropathy type I (FAP-I) by electrospray ionization-ion mobility mass spectrometry. *Talanta*, **181**, 87-94.
11. Dubois, C., Payen, D., Simon, S., Junot, C., Fenaille, F., Morel, N. and Becher, F. (2020) Top-Down and Bottom-Up Proteomics of Circulating S100A8/S100A9 in Plasma of Septic Shock Patients. *J Proteome Res*, **19**, 914-925.
12. Azizkhanian, I., Trenchevska, O., Bashawri, Y., Hu, J., Koska, J., Reaven, P.D., Nelson, R.W., Nedelkov, D. and Yassine, H.N. (2016) Posttranslational modifications of apolipoprotein A-II proteoforms in type 2 diabetes. *J Clin Lipidol*, **10**, 808-815.
13. Cabras, T., Sanna, M., Manconi, B., Fanni, D., Demelia, L., Sorbello, O., Iavarone, F., Castagnola, M., Faa, G. and Messana, I. (2015) Proteomic investigation of whole saliva in Wilson's disease. *J Proteomics*, **128**, 154-163.
14. Cleland, T.P., DeHart, C.J., Fellers, R.T., VanNispen, A.J., Greer, J.B., LeDuc, R.D., Parker, W.R., Thomas, P.M., Kelleher, N.L. and Brodbelt, J.S. (2017) High-Throughput Analysis of Intact Human Proteins Using UVPD and HCD on an Orbitrap Mass Spectrometer. *J Proteome Res*, **16**, 2072-2079.
15. Yang, Z., Shen, X., Chen, D. and Sun, L. (2020) Toward a Universal Sample Preparation Method for Denaturing Top-Down Proteomics of Complex Proteomes. *J Proteome Res*, **19**, 3315-3325.
16. Dai, Y., Buxton, K.E., Schaffer, L.V., Miller, R.M., Millikin, R.J., Scalf, M., Frey, B.L., Shortreed, M.R. and Smith, L.M. (2019) Constructing Human Proteoform Families Using Intact-Mass and Top-Down Proteomics with a Multi-Protease Global Post-Translational Modification Discovery Database. *J Proteome Res*, **18**, 3671-3680.
17. Schaffer, L.V., Anderson, L.C., Butcher, D.S., Shortreed, M.R., Miller, R.M., Pavelec, C. and Smith, L.M. (2021) Construction of Human Proteoform Families from 21 Tesla Fourier Transform Ion Cyclotron Resonance Mass Spectrometry Top-Down Proteomic Data. *J Proteome Res*, **20**, 317-325.
18. Durbin, K.R., Fornelli, L., Fellers, R.T., Doubleday, P.F., Narita, M. and Kelleher, N.L. (2016) Quantitation and Identification of Thousands of Human Proteoforms below 30 kDa. *J Proteome Res*, **15**, 976-982.
19. Ahlf, D.R., Compton, P.D., Tran, J.C., Early, B.P., Thomas, P.M. and Kelleher, N.L. (2012) Evaluation of the compact high-field orbitrap for top-down proteomics of human cells. *J Proteome Res*, **11**, 4308-4314.
20. Ntai, I., LeDuc, R.D., Fellers, R.T., Erdmann-Gilmore, P., Davies, S.R., Rumsey, J., Early, B.P., Thomas, P.M., Li, S., Compton, P.D. *et al.* (2016) Integrated Bottom-Up and Top-Down Proteomics of Patient-Derived Breast Tumor Xenografts. *Mol Cell Proteomics*, **15**, 45-56.
21. Fornelli, L., Durbin, K.R., Fellers, R.T., Early, B.P., Greer, J.B., LeDuc, R.D., Compton, P.D. and Kelleher, N.L. (2017) Advancing Top-down Analysis of the Human Proteome Using a Benchtop Quadrupole-Orbitrap Mass Spectrometer. *J Proteome Res*, **16**, 609-618.
22. Chen, Y., Hoover, M.E., Dang, X., Shomo, A.A., Guan, X., Marshall, A.G., Freitas, M.A. and Young, N.L. (2016) Quantitative Mass Spectrometry Reveals that Intact Histone H1 Phosphorylations are Variant Specific and Exhibit Single Molecule Hierarchical Dependence. *Mol Cell Proteomics*, **15**, 818-833.
23. Koska, J., Yassine, H., Trenchevska, O., Sinari, S., Schwenke, D.C., Yen, F.T., Billheimer, D., Nelson, R.W., Nedelkov, D. and Reaven, P.D. (2016) Disialylated apolipoprotein C-III proteoform is associated with improved lipids in prediabetes and type 2 diabetes. *J Lipid Res*, **57**, 894-905.
24. Coelho Graça, D., Hartmer, R., Jabs, W., Beris, P., Clerici, L., Stoermer, C., Samii, K., Hochstrasser, D., Tsybin, Y.O., Scherl, A. *et al.* (2015) Identification of hemoglobin variants by top-down mass spectrometry using selected diagnostic product ions. *Anal Bioanal Chem*, **407**, 2837-2845.
25. Soler-Ventura, A., Gay, M., Jodar, M., Vilanova, M., Castillo, J., Arauz-Garofalo, G., Villarreal, L., Ballescà, J.L., Vilaseca, M. and Oliva, R. (2020) Characterization of Human Sperm Protamine

- Proteoforms through a Combination of Top-Down and Bottom-Up Mass Spectrometry Approaches. *J Proteome Res*, **19**, 221-237.
26. Yu, D., Wang, Z., Cupp-Sutton, K.A., Liu, X. and Wu, S. (2019) Deep Intact Proteoform Characterization in Human Cell Lysate Using High-pH and Low-pH Reversed-Phase Liquid Chromatography. *J Am Soc Mass Spectrom*, **30**, 2502-2513.
 27. Contini, C., Firinu, D., Serrao, S., Manconi, B., Olianias, A., Cinetto, F., Cossu, F., Castagnola, M., Messina, I., Del Giacco, S. *et al.* (2020) RP-HPLC-ESI-IT Mass Spectrometry Reveals Significant Variations of the Human Salivary Protein Profile Associated with Predominantly Antibody Deficiencies. *J Clin Immunol*, **40**, 329-339.
 28. Pont, L., Benavente, F., Barbosa, J. and Sanz-Nebot, V. (2017) On-line immunoaffinity solid-phase extraction capillary electrophoresis mass spectrometry using Fab´antibody fragments for the analysis of serum transthyretin. *Talanta*, **170**, 224-232.
 29. Pont, L., Poturcu, K., Benavente, F., Barbosa, J. and Sanz-Nebot, V. (2016) Comparison of capillary electrophoresis and capillary liquid chromatography coupled to mass spectrometry for the analysis of transthyretin in human serum. *J Chromatogr A*, **1444**, 145-153.
 30. Peró-Gascón, R., Pont, L., Benavente, F., Barbosa, J. and Sanz-Nebot, V. (2016) Analysis of serum transthyretin by on-line immunoaffinity solid-phase extraction capillary electrophoresis mass spectrometry using magnetic beads. *Electrophoresis*, **37**, 1220-1231.
 31. Chen, Y.C., Sumandea, M.P., Larsson, L., Moss, R.L. and Ge, Y. (2015) Dissecting human skeletal muscle troponin proteoforms by top-down mass spectrometry. *J Muscle Res Cell Motil*, **36**, 169-181.
 32. Zhou, M., Uwugiaren, N., Williams, S.M., Moore, R.J., Zhao, R., Goodlett, D., Dapic, I., Paša-Tolić, L. and Zhu, Y. (2020) Sensitive Top-Down Proteomics Analysis of a Low Number of Mammalian Cells Using a Nanodroplet Sample Processing Platform. *Anal Chem*, **92**, 7087-7095.