

## HPP Scientific Terms, Definitions & Abbreviations\*

\*Definitions of terms commonly used by HPP researchers, based on past literature or consensus reached by the HPP community (e.g. HPP publications, HPP guidelines, neXtProt, PeptideAtlas, Human Protein Atlas, GPMdb, ProteomeXchange), (Updated: Aug 30, 2018).

Term	Definition	Additional Information, Web links and References
<b>HPP</b>	The Human Proteome Project (HPP) is an international project organized by the Human Proteome Organization (HUPO) that aims to map, annotate, and functionally characterize the entire human proteome in a systematic way using mass spectrometry complemented by antibody and affinity-based techniques and many other protein methods. The HPP extends and is a direct counterpart to the Human Genome Project. HPP annotation of the human genome gene products adds significant value and insights about human biology. The HPP is composed of two complementary initiatives: The Chromosome-centric HPP (C-HPP) and Biology/Disease HPP (B/D-HPP). The former focuses on the completion of the “parts list” for proteins and their proteoforms whereas the latter aims to make proteomics an integral part of multi-omics research throughout the life sciences and biomedical research communities. Both initiatives are supported by 4 resource pillars: (i) mass spectrometry (MS), (ii) affinity reagents (Ab), (iii) knowledge base (Kb), and (iv) pathology.	<p><a href="http://www.hupo.org">www.hupo.org</a></p> <p>Note: Completion of the HPP will generate a protein-based map of the molecular architecture of human cells and the human body, enhance our understanding of human biology at the cellular level and lay a foundation for the development of novel diagnostic, prognostic, therapeutic, and preventive medical applications. The HPP is governed by the HPP Executive Committee.</p> <p>Legrain P, Aebersold R, Archakov A et al., The human proteome project: current state and future direction. <i>Mol Cell Proteomics</i> 2011 Jul 10(7):M111 009993. doi: 10.1074/mcp.M111.009993.</p> <p>Omenn GS, Lane L, Overall CM, Corrales FJ, Schwenk JM, Paik YK, van Eyk JE, Liu S, Snyder M, Baker MS, Deutsch EW. Progress on identifying and characterizing the Human Proteome: 2018 Metrics from the HUPO Human Proteome Project. <i>J Proteome Res</i> 2018, doi: 10.1021/acs.jproteome.8b00441.</p>
<b>C-HPP</b>	The Chromosome-Centric HPP (C-HPP) is an international collaborative initiative of the HPP that aims to map, annotate, and characterize the human proteome on a chromosome-by-chromosome basis. The 25 international teams from 20 countries use various proteomics technologies to study how the proteome is encoded in Chr 1 – 22, X, Y, and mitochondrial DNA. Currently, major foci of the C-HPP are to map all remaining missing proteins (PE2,3,4 proteins in neXtProt 2018-1-17 = 2,186) and characterize 1,260 PE1 proteins with no function annotated in neXtProt 2018-1-17 (uPE1).	<p><a href="http://www.c-hpp.org">www.c-hpp.org</a></p> <p>Note: The initial goal of C-HPP is to identify at least one representative protein with three post-translational modifications (PTMs) (phosphoryl, -glycosyl-, acetyl-) and alternative splicing isoform encoded by each of c.a. 20,300 human protein-encoding genes with their tissue localization and quantitative studies using MS and/or antibody reagents.</p> <p>Paik YK, Jeong SK, Omenn GS et al., The Chromosome-Centric Human Proteome Project for cataloging proteins encoded in the genome. <i>Nat Biotechnol</i> 2012 Mar 7;30(3):221-3. doi: 10.1038/nbt.2152.</p> <p>Paik YK, Omenn GS, Hancock WS et al., Advances in the Chromosome-Centric Human Proteome Project: looking to the future. <i>Expert Rev Proteomics</i> 2017 Dec 14(12):1059-1071. doi: 10.1080/14789450.2017.1394189.</p>
<b>B/D-HPP</b>	The Biology/Disease HPP (B/D-HPP) is an international collaborative initiative of the HPP that focuses on mapping, annotating, and characterizing the proteome using proteomics technologies in relation to human biology and/or diseases. The B/D-HPP provides a framework for the coordination of 19 initiatives. A popular proteins strategy has been developed to stimulate use of targeted proteomics by each of the B/D initiatives and throughout the life sciences and biomedical community. Combining C-HPP, B/D-HPP, and Resource Pillar, the HPP has 50 international teams.	<p><a href="http://www.hupo.org/B/D-HPP">www.hupo.org/B/D-HPP</a></p> <p>Note: The goals of B/D-HPP are to conduct experimental studies of specific organs and biofluids in health and disease and to assemble publicly accessible prioritized panels of proteins relevant to biological processes, organs (e.g., cardiovascular, cerebral, hepatic, renal, pulmonary, and intestinal systems) and organelles (e.g. mitochondria). More broadly, it aims to develop standardized methods for protein detection and quantification by proteomics to promote translation into clinical settings.</p> <p>Van Eyk JE, Corrales FJ, Aebersold R et al., Highlights of the Biology and Disease-driven Human Proteome Project, 2015-2016. <i>J Proteome Res</i> 2016 Nov 4;15(11):3979-3987. doi: 10.1021/acs.jproteome.6b00444.</p>
<b>PE</b>	Protein existence (PE) levels indicate the degree of evidence for the existence of a human protein based on curated information. The levels PE1 to PE5 are assigned by UniProtKB/SwissProt and neXtProt as follows.	<p>Note: <a href="http://www.nextprot.org/about/protein-existence">www.nextprot.org/about/protein-existence</a>; <a href="http://www.uniprot.org/help/protein_existence">www.uniprot.org/help/protein_existence</a> <a href="https://hupo.org/Guidelines">https://hupo.org/Guidelines</a></p>

<p><b>PE (cont.)</b></p>	<ul style="list-style-type: none"> <li>•PE1: evidence at the protein level (identified by mass spectrometry (MS) according to HPP guidelines, or curated from multiple other experimental protein methods).</li> <li>•PE2: evidence at the transcript level (detection by RNAseq or presence of expressed sequence tag).</li> <li>•PE3: inferred by gene homology (assigned membership of a defined protein family).</li> <li>•PE4: predicted protein (not yet assigned membership of a defined protein family).</li> <li>•PE5: uncertain or dubious sequences (such as erroneous translation products or pseudogenes). In 2013, the HPP excluded PE5 entries from the search for missing proteins.</li> </ul>	<p>The HPP publishes annual HPP Metrics for worldwide progress identifying and characterizing the Human Proteome, based on the PE levels, e.g., Lane et al and Omenn et al (see HPP box above).</p> <p>Lane L, Bairoch A, Beavis RC et al., Metrics for the Human Proteome Project 2013-2014 and strategies for finding missing proteins. J Proteome Res 2014 Jan 3;13(1):15-20. doi: 10.1021/pr401144x.</p> <p>Deutsch EW, Overall CM, Van Eyk JE et al., Human Proteome Project Mass Spectrometry Data Interpretation Guidelines v2.1. J Proteome Res 2016 Nov 4;15(11):3961-3970. doi: 10.1021/acs.jproteome.6b00392.</p>
<p><b>uPE1 Proteins</b></p>	<p>Uncharacterized PE1 proteins (uPE1s) devoid of any functional annotation in neXtProt or annotated only with broad Gene Ontology Molecular Function/Biological Process terms not linked to any specific function. As of neXtProt release 2018-01-17, there were 1,260 uPE1 proteins.</p>	<p>The current list of uPE1 proteins can be retrieved at:  <a href="https://www.nextprot.org/proteins/search?mode=advanced&amp;queryId=NXQ_00022">https://www.nextprot.org/proteins/search?mode=advanced&amp;queryId=NXQ_00022</a></p> <p>* Currently neXtProt excludes these 11 broad Gene Ontology (GO) function terms: GO:0005509, calcium ion binding; GO:0008270, zinc ion binding; GO:0005515, protein binding; GO:0042802, identical protein binding; GO:0051260, protein homooligomerization; GO:0005524, ATP-binding; GO:0000287 magnesium-binding; GO:0003676 nucleic acid binding; GO:0003824 catalytic activity; GO:0007165 signal transduction; GO:0035556 intracellular signal transduction.</p> <p>In the next neXtProt release this query will be refined by adding three terms: GO:0046914 transition metal ion binding; GO:0046872 metal ion binding; and GO:0035556 intracellular signal transduction.</p> <p>Paik YK, Omenn GS, Hancock WS et al., Advances in the Chromosome-Centric Human Proteome Project: looking to the future. Expert Rev Proteomics 2017 Dec;14(12):1059-1071. doi: 10.1080/14789450.2017.1394189.</p>
<p><b>Missing Proteins</b></p>	<p>Missing proteins (MPs) are defined as those protein entries that belong to categories PE2,3,4 in neXtProt. They correspond to confidently predicted proteins that lack sufficient experimental data from mass spectrometry or other direct protein methods to qualify as PE1.</p> <p>The annual HPP Metrics papers in the special issues of the Journal of Proteome Research assess progress on identifying MPs and strategies needed to enrich and detect MPs.</p>	<p>The current list of MPs, based on neXtProt curation, can be retrieved at:  <a href="https://www.nextprot.org/proteins/search?mode=advanced&amp;queryId=NXQ_00204">https://www.nextprot.org/proteins/search?mode=advanced&amp;queryId=NXQ_00204</a>.</p> <p>Paik YK, Jeong SK, Omenn GS et al., The Chromosome-Centric Human Proteome Project for cataloging proteins encoded in the genome. Nat Biotechnol 2012 Mar 7;30(3):221-3. doi: 10.1038/nbt.2152.</p> <p>Lane L, Bairoch A, Beavis RC et al., Metrics for the Human Proteome Project 2013-2014 and strategies for finding missing proteins. J Proteome Res 2014;13(1):15-20. doi: 10.1021/pr401144x.</p> <p>Baker, MS, et al. Accelerating the search for the missing proteins in the human proteome. Nat Commun 2017; 8, 1471 doi: 10.1028/ncomms14271.</p>
<p><b>HPP Guidelines</b></p>	<p>The HPP Mass Spectrometry Data Interpretation Guidelines version 2.1.0 (“the Guidelines”) provide a set of expectations for data interpretation of MS data that is contributed to the HPP. There are broadly two sections, one that applies to all datasets, including data deposition requirements and false discovery rate thresholds, and a second that provides enhanced expectations for evidence of detections of missing proteins or translation products not currently listed in neXtProt.</p>	<p><a href="https://hupo.org/HPP-Data-Interpretation-Guidelines">https://hupo.org/HPP-Data-Interpretation-Guidelines</a></p> <p>Deutsch EW, Overall CM, Van Eyk JE, Baker MS, Paik YK, Weintraub ST, Lane L, Martens L, Vandenbrouck Y, Kusebauch U, Hancock WS, Hermjakob H, Aebersold R, Moritz RL, Omenn GS, Human Proteome Project Mass Spectrometry Data Interpretation Guidelines 2.1., J Proteome Res. 2016 Nov 4;15(11):3961-3970.</p>

<b>Dark Proteome</b>	The dark proteome is a colloquial term that includes missing proteins (PE2 – PE4), uncertain/dubious predicted proteins (PE5), uPE1 proteins, smORF (small proteins), and any proteins translated by long non-coding RNAs or uncharacterized transcripts including those arising from non-coding regions of DNA and/or novel alternative splicing.	
<b>Proteoforms</b>	Alternative protein products from the same gene resulting from genomic sequence alterations, alternative splicing, RNA editing, post-translational modifications of amino acid side chains, and proteolytic processing events.	<p><a href="http://repository.topdownproteomics.org/">http://repository.topdownproteomics.org/</a> Smith LM, Kelleher NL, Consortium for Top Down Proteomics. Proteoform: a single term describing protein complexity. Nat Methods 2013 Mar 10(3):186-7. doi: 10.1038/nmeth.2369.</p> <p>LeDuc RD, Schwammle V, Shortreed MR, et al., ProForma: A Standard Proteoform Notation. J Proteome Res 2018 Mar 2;17(3):1321-1325. doi: 10.1021/acs.jproteome.7b00851.</p> <p>Aebersold, R, Agar, JN, et al. How Many Proteoforms are there? Nature Chemical Biology 2018; 14:206-214. doi: <a href="https://doi.org/10.1038/nchembio.2576">10.1038/nchembio.2576</a></p>
<b>neXt-MP50</b>	A specific two-year C-HPP initiative, announced in September 2016, that aims to accelerate the identification and validation of the existence of 50 currently missing proteins per chromosome team while incorporating progress from the entire international proteomics community.	<p>Paik YK, Omenn GS, Hancock WS et al., Advances in the Chromosome-Centric Human Proteome Project: looking to the future. Expert Rev Proteomics 2017 Dec 14(12):1059-1071. doi: 10.1080/14789450.2017.1394189.</p> <p>Omenn GS, Lane L, Lundberg EK, Overall CM, Deutsch EW. Progress on the HUPO Draft Human Proteome: 2017 Metrics of the Human Proteome Project. J Proteome Res. 2017 Dec 1;16(12):4281-4287. doi: 10.1021/acs.jproteome.7b00375. Epub 2017 Oct 9.</p>
<b>neXt-CP50</b>	A specific C-HPP initiative, announced in September 2017, that aims to characterize some cellular function/s of 50 uPE1 proteins within 3 years by >14 C-HPP working groups.	<p>Paik YK, Overall CM, Deutsch EW et al., Progress and Future Direction of Chromosome-Centric Human Proteome Project. J Proteome Res 2017 Dec 1;16(12):4253-4258. doi: 10.1021/acs.jproteome.7b00734.</p>
<b>Popular Proteins</b>	Popular Proteins is a BD-HPP initiative to define the most-cited proteins based on health and diseases as found in PubMed and thereby stimulate wide use of targeted proteomics in the life sciences/biomedical community	<p><a href="http://tinyurl.com/proteinpurpose">http://tinyurl.com/proteinpurpose</a> <a href="http://pubpular.net">http://pubpular.net</a></p> <p>Note: There are two algorithms which can be used to assist B/D-HPP initiatives and those with particular interest in a disease, state or organ. The development of mass spectrometry-based assays to allow easier and accurate quantification across all fields of science for those proteins which are currently most studied. New expansion will be in popular PTMs and identification of proteins and specific amino acid residues which are most cited in PubMed for particular PTMs.</p> <p>1. Lam MP, Venkatraman V, Xing Y et al., Data-Driven Approach To Determine Popular Proteins for Targeted Proteomics Translation of Six Organ Systems. J Proteome Res 2016 Nov 4;15(11):4126-4134. doi: 10.1021/acs.jproteome.6b00095.</p> <p>2. Yu KH, Lee TM, Wang CS et al., Systematic Protein Prioritization for Targeted Proteomics Studies through Literature Mining. J Proteome Res 2018 Apr 6;17(4):1383-1396. doi: 10.1021/acs.jproteome.7b00772.</p>

<p><b>HPA</b></p>	<p>The Human Protein Atlas (HPA) is a Swedish-based program started in 2003 with the aim to map all the human proteins in cells, tissues and organs primarily using antibodies for immunohistochemistry and immunofluorescence. The HPA has been expanded to incorporate transcriptomes and systems biology. The Human Protein Atlas has provided the Affinity-based Resource Pillar for the HPP from the outset.</p>	<p><a href="https://www.proteinatlas.org/">https://www.proteinatlas.org/</a></p> <p>Uhlen et al, Tissue-based map of the human proteome. <i>Science</i>. 2015 347(6220):1260419. doi:10.1126/science.1260419.</p> <p>Thul PJ, Akesson L, Wiking M et al., A subcellular map of the human proteome. <i>Science</i> 2017 May 26;356(6340). doi: 10.1126/science.aal3321.</p>
<p><b>ProteomeXchange</b></p>	<p>ProteomeXchange is a consortium of proteomics data repositories that support public submission of data, stimulated by the HPP and built by the founding members PRIDE at the European Bioinformatics Institute and PeptideAtlas at the Institute for Systems Biology. ProteomeXchange aims to coordinate globally the submission of mass spectrometry proteomics data to the main existing proteomics repositories and to facilitate optimal dataset dissemination and access. It includes PRIDE, PeptideAtlas, MassIVE, jPOST, iProX, and Panorama Public.</p>	<p><a href="http://www.proteomexchange.org/">http://www.proteomexchange.org/</a></p> <p>Vizcaino JA, Deutsch EW, Wang R et al., ProteomeXchange provides globally coordinated proteomics data submission and dissemination. <i>Nat Biotechnol</i> 2014 Mar 32(3):223-6. doi: 10.1038/nbt.2839.</p> <p>Deutsch EW, Csordas A, Sun Z, et al., The ProteomeXchange consortium in 2017: supporting the cultural change in proteomics public data deposition, <i>Nucleic Acids Res</i>. 2017 Jan 4;45(D1):D1100-D1106. doi: 10.1093/nar/gkw936. Epub 2016 Oct 18.</p>
<p><b>PeptideAtlas</b></p>	<p>PeptideAtlas is a public data repository developed by the Institute for Systems Biology that accepts submissions of proteomics mass spectrometry datasets from laboratories all over the world, reprocesses them with the Trans-Proteomic Pipeline suite of software tools, collates all datasets into a global view of the human proteome as observed in MS datasets, informs neXtProt of HPP Guidelines-compliant findings, and makes all the results publicly available to the community.</p>	<p><a href="http://www.peptideatlas.org/">http://www.peptideatlas.org/</a></p> <p>Desiere F, Deutsch EW, King NL et al., The PeptideAtlas project. <i>Nucleic Acids Res</i> 2006 Jan 1;34(Database issue):D655-8. doi: 10.1093/nar/gkj040.</p> <p>Deutsch EW, Sun Z, Campbell D, Kusebauch U, Chu CS, Mendoza L, Shteynberg D, Omenn GS, Moritz RL. State of the Human Proteome in 2014/2015 As Viewed through PeptideAtlas: Enhancing Accuracy and Coverage through the AtlasProphet. <i>J Proteome Res</i>. 2015 Sep 4;14(9):3461-73. doi: 10.1021/acs.jproteome.5b00500</p> <p>Schwenk JM, Omenn GS, Sun Z, Campbell DS, Baker MS, Overall CM, Aebersold R, Moritz RL, Deutsch EW. The Human Plasma Proteome Draft of 2017: Building on the Human Plasma PeptideAtlas from Mass Spectrometry and Complementary Assays. <i>J Proteome Res</i>. 2017 Dec 1;16(12):4299-4310. doi: 10.1021/acs.jproteome.7b00467.</p>
<p><b>neXtProt</b></p>	<p>neXtProt is an online knowledge platform with extensive information on human proteins. It integrates data from UniProtKB/Swiss-Prot, PeptideAtlas, SRMAtlas, HPA and many other resources and offers an advanced search capacity based on the SPARQL technology and an API that allows users to programmatically extract the data stored in the resource. It also provides a peptide uniqueness checker that takes &gt;5 million validated single amino acid variants (SAAVs) into account.</p> <p>The HPP uses the PE levels assigned by neXtProt to monitor progress made collectively by the scientific community toward the complete experimental validation of the "parts list" of the human proteome.</p>	<p><a href="http://www.nextprot.org">www.nextprot.org</a>  <a href="https://www.nextprot.org/tools/peptide-uniqueness-checker">https://www.nextprot.org/tools/peptide-uniqueness-checker</a></p> <p>Lane L, Argoud-Puy G, Britan A et al., neXtProt: a knowledge platform for human proteins. <i>Nucleic Acids Res</i> 2012 Jan 40(Database issue):D76-83. doi: 10.1093/nar/gkr1179.</p> <p>Gaudet P, Argoud-Puy G, Cusin I et al., neXtProt: organizing protein knowledge in the context of human proteome projects. <i>J Proteome Res</i> 2013 Jan 4;12(1):293-8. doi: 10.1021/pr300830v.</p> <p>Gaudet P, Michel PA, Zahn-Zabal M et al., The neXtProt knowledgebase on human proteins: 2017 update. <i>Nucleic Acids Res</i> 2017 Jan 4;45(D1):D177-D182. doi: 10.1093/nar/gkw1062.</p> <p>Schaeffer M, Gateau A, Teixeira D, Michel PA, Zahn-Zabal M, Lane L. The neXtProt peptide uniqueness checker: a tool for the proteomics community. <i>Bioinformatics</i>. 2017 Nov 1;33(21):3471-3472. doi: 10.1093/bioinformatics/btx318.</p>

<b>SRMATlas</b>	SRMATlas is a compendium of the best-available SRM transitions for nearly every human protein, drawn from data generated on multiple QTOF and QQQ instruments. The data were produced by targeted proteomics and are used for validation of spectral evidence of detection of missing proteins under the SRMATlas. SRMATlas is also the base for the popular protein strategy of the B/D-HPP.	<a href="http://www.srmatlas.org/">http://www.srmatlas.org/</a> Kusebauch U, Campbell DS, Deutsch EW et al., Human SRMATlas: A Resource of Targeted Assays to Quantify the Complete Human Proteome. Cell 2016 Jul 28;166(3):766-778. doi: 10.1016/j.cell.2016.06.041.
-----------------	---	--