HPP Scientific Terms, Definitions & Abbreviations* (Approved by HPP EC)

**Proposed definitions of terms commonly used by HPP researchers, based on past literature or consensus reached by HPP community (e.g., EC Minutes, HPP guidelines, neXtProt, PeptideAtlas, GPMdb databases).(Last updated: July 21, 2018)*

Terms	Definition	Additional Information, Web links and References
HPP	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 and affinity-based techniques. The HPP acts as a direct counterpart to the human genome project, adding significant value and insights into human biology. The HPP is composed of two complementary initiatives: the Chromosome 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 while 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 technology pillars: (i) mass spectrometry (MS), (ii) affinity reagents (Ab), and (iii) knowledgebase (Kb).	Note: Completion of the HPP will generate a protein-based map of the molecular architecture of 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. www.hupo.org 1. Legrain P, Aebersold R, Archakov A et al., The human proteome project: current state and future direction. Mol Cell Proteomics 2011 Jul 10(7):M111 009993. doi: 10.1074/mcp.M111.009993. PubMed PMID: 21742803.
C-HPP	The Chromosome-Centric HPP (C-HPP) is an international collaborative project 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 focus of the C-HPP is to map all remaining missing proteins (2168, neXtProt 2018-1-17) and characterize 2000 uPE1 (uncharacterized PE1, 2018-1-17).	Note: The initial goal of C-HPP is to identify at least one representative protein with three PTMs (phosphoryl, -glycosyl-, acetyl-) and alternative splicing isoform encoded by each of c.a. 20,300 human genes with their tissue localization and quantitative studies using mass spectrometry (MS) and/or antibody reagents. www.c-hpp.org 1. 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. PubMed PMID: 22398612. 2. 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. PubMed PMID: 29039980.
B/D-HPP	The Biology/Disease-based HPP (B/D-HPP) is an international collaborative project that deals with 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 22 initiatives that integrate about 50 multinational research groups.	Note: The goals of B/D-HPP are to assemble publicly accessible prioritized panels of proteins relevant to different human biology or diseases (*e.g.*, cardiovascular, cerebral, hepatic, renal, pulmonary, and intestinal systems) and organ/systems (e.g. mitochondria). Additionally, it aims to develop standardized methods for proteome detection and quantification and to promote translation of discoveries into clinical settings. Finally, to provide means to intergrate proteomics throughout the broader disease and biology sciences. www.hupo.org/B/D-HPP Van Eyk JE, Corrales FJ, Aebersold R et al., Highlights of the Biology and Disease-driven Human Proteome Project, 2015-2016. J Proteome Res 2016 Nov 4;15(11):3979-3987. doi: 10.1021/acs.jproteome.6b00444. PubMed PMID: 27573249.

PE	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 and neXtProt as follows. •PE1: evidence at the protein level (identification by mass spectrometry (MS) according to HPP guidelines, validated antibody (Ab)-based detection, or other characterization). •PE2: evidence at the transcript level (detection by RNAseq or presence of expressed sequence tag). •PE3: inferred by gene homology (assigned membership of a defined protein family). •PE4: predicted protein (not yet assigned membership of a defined protein family). •PE5: uncertain (dubious sequences such as erroneous translation products or pseudogenes). Note: 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 human proteome.	Note: www.nextprot.org/about/protein-existence www.uniprot.org/help/protein_existence https://hupo.org/Guidelines 1. Deutsch EW, Overall CM, Van Eyk JE et al., Human Proteome Project Mass Spectrometry Data Interpretation Guidelines 2.1. J Proteome Res 2016 Nov 4;15(11):3961-3970. doi: 10.1021/acs.jproteome.6b00392. PubMed PMID: 27490519. 2. 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. PubMed PMID: 24364385.
Missing Proteins	Missing proteins (MPs) are defined as those gene-encoded proteins with less confident mass spectrometry evidence or that are inadequately annotated. To refine the search for and annotation of the so-called missing proteins, the HPP also defines this term as the sum of proteins within the neXtProt categories PE2 – PE-4, inclusive.	Note: To refine their search for and annotation of the so-called missing proteins, the HPP defines the term 'missing proteins' as the sum of proteins within neXtProt PE2 – PE-4 categories inclusive, as defined above. https://www.nextprot.org/proteins/search?mode=advanced&queryId=NXQ_00204 http://www.missingproteins.org/protein/web/ 1. 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. PubMed PMID: 22398612. 2. 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. PubMed PMID: 24364385.
Proteoforms	Alternative forms of the same gene product produced after protein post-translational modification/s.	http://repository.topdownproteomics.org/ 1. Smith LM, Kelleher NL,Consortium for Top Down P, Proteoform: a single term describing protein complexity. Nat Methods 2013 Mar 10(3):186-7. doi: 10.1038/nmeth.2369. PubMed PMID: 23443629. 2. 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. PubMed PMID: 29397739.
uPE1 Protein	Unannotated PE1 proteins (uPE1s) include PE1 proteins currently lacking any functional characterization (experimental or predicted) in UniProtKB, Swiss-Prot. or neXtProt. Some of these proteins can have associated gene ontology (GO) terms, and biological process (BP) and molecular function (MF) categories that are not indicative of a cellular function.	https://tinyurl.com/upe1proteins 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. PubMed PMID: 29039980.
Dark Proteome	The dark proteome is a colloquial term that includes missing proteins (PE2 – PE4), uncertain proteins (PE5), uPE1 proteins, smORF (small proteins), and any proteoforms translated by long non-coding RNAs or uncharacterized transcripts including those arising from non-coding regions of DNA and/or novel alternative splicing.	http://darkproteome.ws/ (UniProt) Perdigao N, Heinrich J, Stolte C et al., Unexpected features of the dark proteome. Proc Natl Acad Sci U S A 2015 Dec 29;112(52):15898-903. doi: 10.1073/pnas.1508380112. PubMed PMID:

		26578815.
neXt-MP50	A specific two-year C-HPP initiative that aims to accelerate the validation of the existence of 50 currently missing proteins per chromosome team.	https://www.nextprot.org/proteins/search?mode=advanced&queryId=NXQ_00204 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. PubMed PMID: 29039980
neXt-CP50	A specific C-HPP initiative that aims to characterize some cellular function/s of 50 uPE1 proteins within 3 years by >14 working groups.	https://tinyurl.com/upe1proteins 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. PubMed PMID: 29191025.
ProteomeXchange	The ProteomeXchange database was built to globally coordinate the submission of mass spectrometry proteomics data to the main existing proteomics repositories and to encourage optimal data dissemination. It includes PRIDE, PeptideAtlas, MassIVE, jPOST, and iProX	http://www.proteomexchange.org/ Vizcaino JA, Deutsch EW, Wang R et al., ProteomeXchange provides globally coordinated proteomics data submission and dissemination. Nat Biotechnol 2014 Mar 32(3):223-6. doi: 10.1038/nbt.2839. PubMed PMID: 24727771.
PeptideAtlas	The PeptideAtlas is a public data repository that accepts submissions of proteomics mass spectrometry (MS) datasets from laboratories all over the world, reprocesses them within the Trans-Proteomic Pipeline suite of software tools, collates all datasets into a global view of the human proteome as observed in MS datasets, and makes all the results publicly available to the community.	http://www.peptideatlas.org/ Desiere F, Deutsch EW, King NL et al., The PeptideAtlas project. Nucleic Acids Res 2006 Jan 1;34(Database issue):D655-8. doi: 10.1093/nar/gkj040. PubMed PMID: 16381952.
SRMAtlas	The 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 used for validation of evidence of detection of missing proteins under the SRMAtlas.	http://www.srmatlas.org/ 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. PubMed PMID: 27453469.
neXtProt	The neXtProt is an online knowledge platform of human protein validation, function, subcellular location, expression, interactions, and role in diseases. Information in the neXtProt is obtained from the UniProtKB/Swiss-Prot and from high-throughput studies from various sources, with an emphasis on proteomics. The neXtProt also 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.	www.nextprot.org 1. Lane L, Argoud-Puy G, Britan A et al., neXtProt: a knowledge platform for human proteins. Nucleic Acids Res 2012 Jan 40(Database issue):D76-83. doi: 10.1093/nar/gkr1179. PubMed PMID: 22139911. 2. Gaudet P, Argoud-Puy G, Cusin I et al., neXtProt: organizing protein knowledge in the context of human proteome projects. J Proteome Res 2013 Jan 4;12(1):293-8. doi: 10.1021/pr300830v. PubMed PMID: 23205526. 3. Gaudet P, Michel PA, Zahn-Zabal M et al., The neXtProt knowledgebase on human proteins: 2017 update. Nucleic Acids Res 2017 Jan 4;45(D1):D177-D182. doi: 10.1093/nar/gkw1062. PubMed PMID: 27899619.
MissingProteinPedia	The MissingProteinPedia is a protein data and information-sharing Web system that collates data about current missing PE2-4 proteins, as defined above. It allows unpublished, preliminary, or proprietary data (e.g., antibody, MS, cell biological and genetic	http://www.missingproteins.org/protein/web/ Baker MS, Ahn SB, Mohamedali A et al., Accelerating the search for the missing proteins in the

	studies) to be shared with collaborators through a protected interface.	human proteome. Nat Commun 2017 Jan 24;8:14271. doi: 10.1038/ncomms14271. PubMed PMID: 28117396.
Popular Proteins	Popular Proteins is a BD-HPP initiative to define the most-cited proteins based on health and diseases as found in PubMed.	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. http://tinyurl.com/proteinpurpose http://pubpular.net 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. PubMed PMID: 27356587. 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. PubMed PMID:
HPA	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 using the integration of various omics technologies, including antibody-based imaging, mass spectrometry-based proteomics, transcriptomics, and systems biology.	https://www.proteinatlas.org/ Thul PJ, Akesson L, Wiking M et al., A subcellular map of the human proteome. Science 2017 May 26;356(6340). doi: 10.1126/science.aal3321. PubMed PMID: 29505266.