

Gilbert S. Omenn, Lydie Lane, Christopher M. Overall, Fernando J. Corrales, Jochen M. Schwenk, Young-Ki Paik, Jennifer E. Van Eyk, Siqi Liu, Michael Snyder, Mark S. Baker, and Eric W. Deutsch
Ann Arbor, Vancouver, Geneva, Madrid, Stockholm, Seoul, Los Angeles, Shenzhen, Stanford, Sydney, Seattle

Summary: The Human Proteome Project (HPP) annually reports on progress throughout the field in credibly identifying and characterizing the human protein parts list and making proteomics an integral part of multiomics studies in medicine and the life sciences. NeXtProt release 2018-01-17 contains 17 470 PE1 proteins, 89% of all neXtProt predicted PE1-4 proteins, up from 17 008 in release 2017-01 (Table 1/ Fig. 1). Conversely, the number of neXtProt PE2,3,4 missing proteins has been reduced from 2949 to 2579 to 2186 over the past two years. Of the PE1 proteins, 16 092 are based on mass spectrometry results, and 1378 on other protein studies, notably protein-protein interaction (Fig. 2). PeptideAtlas has 15 798 canonical proteins, up 625 over the past year, including 269 from SUMOylation studies (Fig. 3). The largest reason for missing proteins is low abundance. Meanwhile, the Human Protein Atlas has released its Cell Atlas, Pathology Atlas, updated Tissue Atlas, and recommendations on antibody validation. Finally, multiplex organ-specific popular proteins targeted proteomics has been applied to various diseases.

PE Level	Feb 2012	Sept 2013	Oct 2014	April 2016	Jan 2017	Jan 2018
1: Evidence at protein level	13,975	15,646	16,491	16,518	17,008	17,470 ^a
2: Evidence at transcript level	5205	3570	2647	2290	1939	1660
3: Inferred from homology	218	187	214	565	563	452
4: Predicted	88	87	87	94	77	74
5: <i>Uncertain or dubious</i>	622	638	616	588	572	574
Human PeptideAtlas canonical proteins	12,509	13,377	14,928	14,629	15,173	15,798

^aPercent of predicted proteins classified as PE1 by neXtProt = PE1/PE1 + 2 + 3 + 4 = 89%. ^bMissing Proteins PE 2 + 3 + 4 = 2186, down from 2579 in neXtProt v2017-01. 'More stringent guidelines imposed in 2016' caused adjustments between 2014 and 2016.

Table 1. neXtProt Protein Existence Evidence Levels from 2012 to 2018 Showing Progress in Identifying PE1 Proteins and PeptideAtlas Canonical Proteins

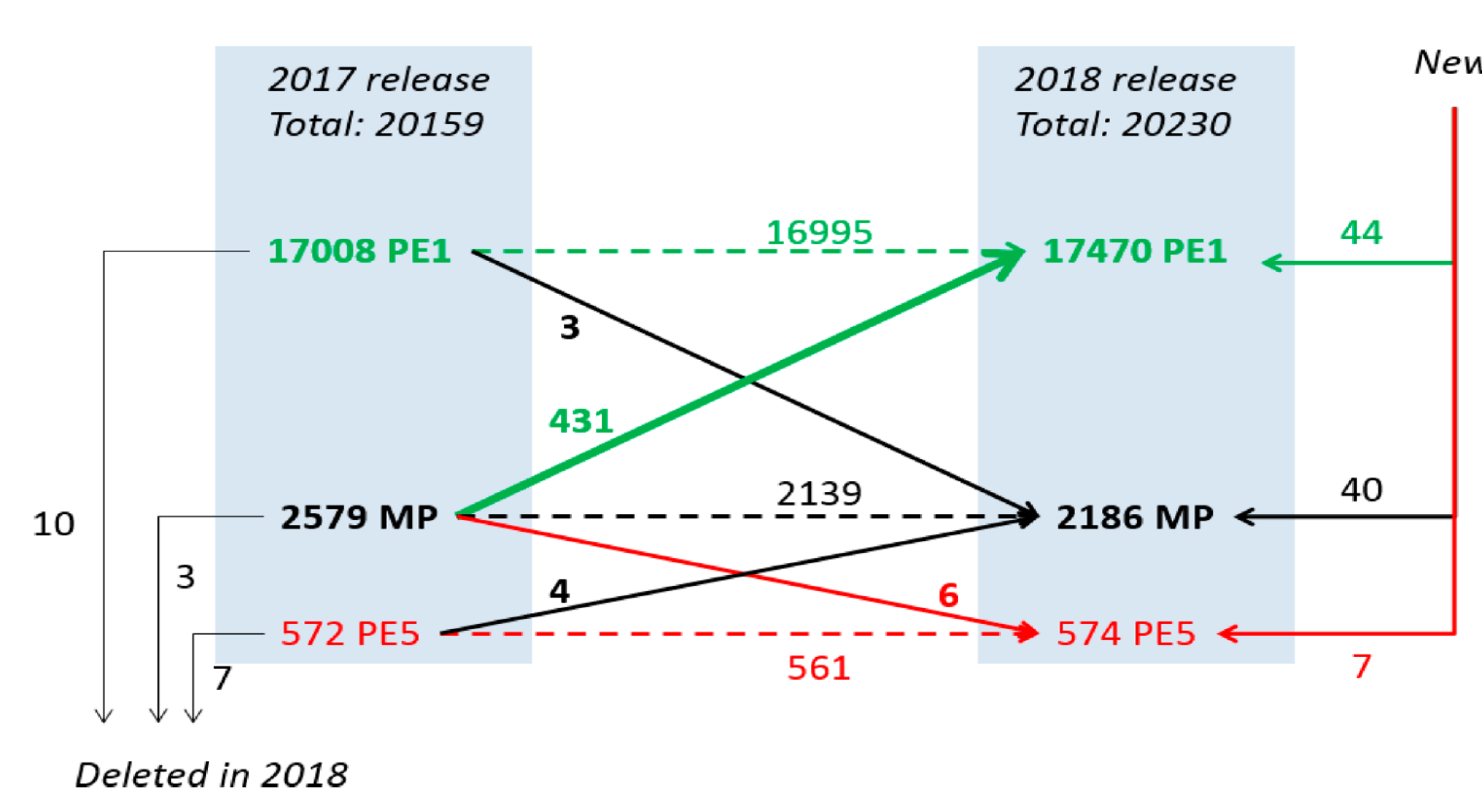


Fig 1. This flowchart depicts the changes in neXtProt PE1-5 categories from release 2017-01-23 to release 2018-01-17. There are 431 missing proteins promoted to PE1 and 44 new SwissProt proteins added as PE1 proteins, while 3 PE1 proteins were demoted to PE2,3,4 MPs and 10 PE1 proteins were deleted altogether.

Omenn et al. *J. Proteome Res.* 2018, DOI: 10.1021/acs.jproteome.8b00441. Many other articles in the sixth Annual HPP Special Issue of JPR will be online in September.

Fig. 2. Pie chart shows the nature of the evidence data for PE1 proteins, as well as the numbers in other categories as of neXtProt release 2018-01-17. There are 16 092 PE1 proteins identified with MS data compliant with HPP Guidelines, of which 98% are canonical in PeptideAtlas. There are 1378 additional PE1 proteins, identified by protein-protein interactions (530), disease mutations (176), PTMs or proteolytic processing (170), Edman sequencing (99), 3D structures (75), Ab-based techniques (58). PE2,3,4 missing proteins are divided into those with no MS data (1798) and those with insufficient or unconfirmed MS data.

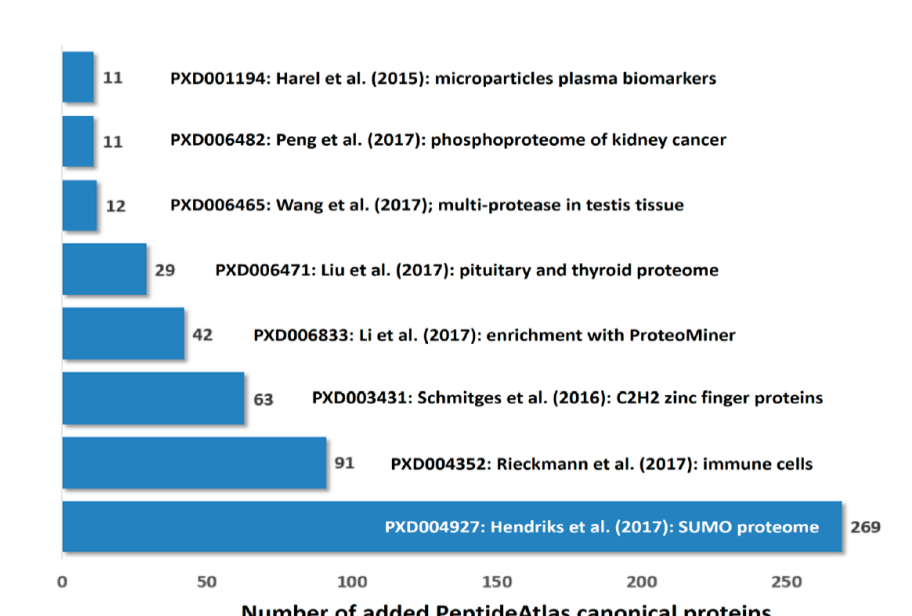
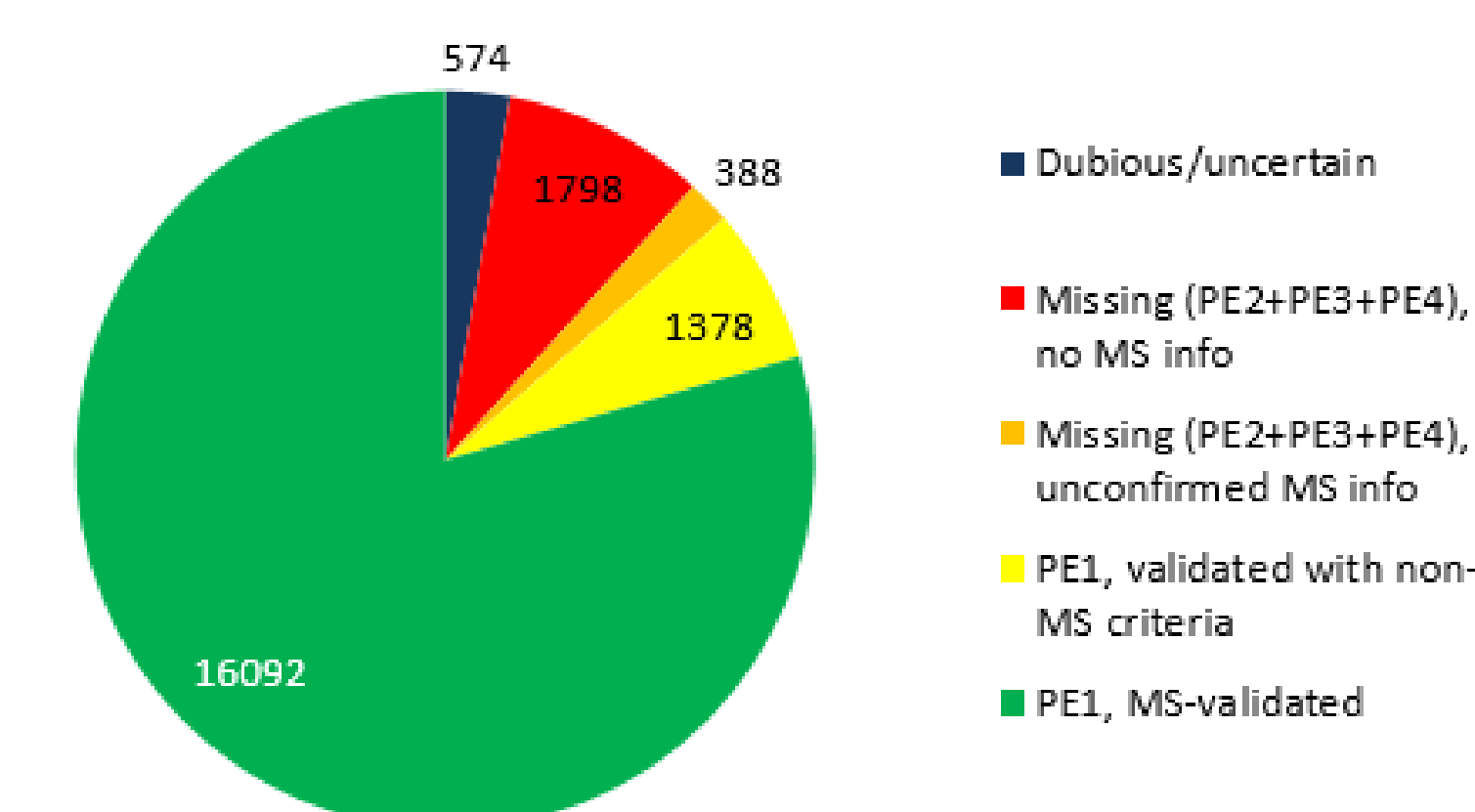


Fig. 3. These eight data sets added to PeptideAtlas in 2017 provided the evidence needed to raise the PeptideAtlas protein category to “canonical” for more than 10 proteins each. Canonical status requires two or more uniquely mapping non-nested peptides with length ≥ 9 residues with high quality spectra, not accounted for by sequence variants or isobaric PTMs in other proteins. PXD identifiers refer to ProteomeXchange.

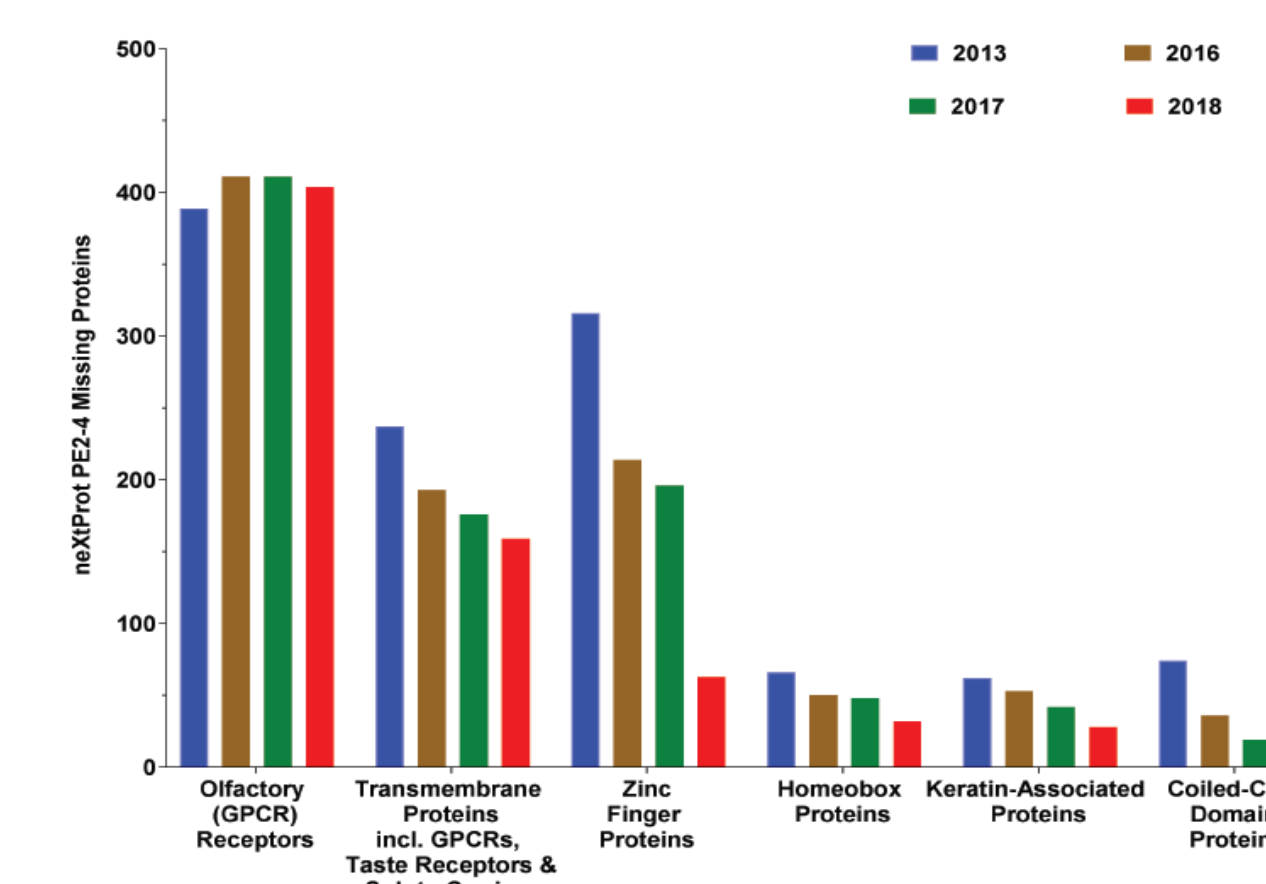


Fig. 4. Progress on identification of members of the six most numerous protein families in neXtProt Missing Protein categories PE2,3,4 from 2013 to 2018 (updated from Baker et al).