LINCS standardized protein information									Notes
Unique ID	LINCS Field Name	Related to	Description	Importance (1: essential; 2: desirable / recommende 3: optional)	Comments	Ontologies / references considered	Suggested terminology	Link to ontology / reference	Additional Notes (for development)
PP:1	PP_Name	canonical protein	The primary name of the protein as chosen by LINCS	1	Should be descriptive and correspond to existing protein names as much as possible; batch independent name. If there are multiple isoforms of a protein, even if each isoform does not have a unique UniProt ID, each isoform should have a unique LINCS ID.	UniProt			
PP:2	PP_LINCS_ID	canonical protein	Unique LINCS internal identifier	1	LINCS internal ID; this is a batch independent ID; canonical protein ID				
PP:3	PP_UniProt_ID	canonical protein	The UniProt accession of the specific protein, if available. If the UniProt ID of a related entity is used instead, this should be documented explicitly.	1	Uniprot accession is typically used as the primary ID	UniProt	UniProt ID	http://www.unipro	
PP:4	PP_Alternate_Name	canonical protein	List of synonymous protein names	1	Synonyms will be obtained from the UniProt database. All UniProt names (Recommended names and Alternative names) should be imported into this field.	UniProt	UniProt, Entez, ...	http://www.unipro	
PP:5	PP_Provider	batch	Vendor or lab that supplied the protein	1			Vendor name		
PP:6	PP_Provider_Catalog_ID	batch	Batch ID or catalogue number assigned to the protein by the vendor or provider	1			Vendor ID		
PP:7	PP_Batch ID	batch	Batch or lot number assigned to the protein by the vendor or provider	1	Provided by the protein vendor or provider		Vendor batch ID		
PP:8	PP_Amino_Acid_Sequen	canonical protein	If the protein is a peptide, protein fragment, or small protein, the amino acid sequence of the perturbagen should be provided	2	This information is usually provided by vendor, but not often referenced to a specific nucleotide sequence. It is more informative than the PRO Name etc.	NCBI/Protei	Protein sequence	http://www.ncbi.n	
PP:9	PP_Gene_Symbol	canonical protein	The NCBI gene name. In cases where the protein is modified (the protein sequence differs from the sequence encoded by the gene listed), it should be described in PP:18.	2	This information is very useful for queries across datasets, e.g. drug versus ligands versus expression changes.	NCBI/Gene	Gene	http://www.ncbi.n	
PP:10	PP_Gene_ID	canonical protein	Entrez Gene ID if using NCBI gene name	2		NCBI	Gene ID	http://www.ncbi.n	
PP:11	PP_Protein_Source	batch	A controlled vocabulary describing the source of the protein (e.g. chemically synthesized, recombinantly expressed in E. coli, etc.)	1		BAO	Protein preparation method	http://bioportal.bi	
PP:12	PP_Protein_Form	batch	A description of a protein's modification status (e.g, if it was mutated, post-translationally modified etc.). If a DNA vector was used to express the modified protein in a cell line, a description of the DNA vector needs to be provided (see below).	2	If available	BAO	Protein form	http://bioportal.bi	
PP:13	PP_Protein_Purity	batch	A description of a protein's level of purity (e. g., if it was partially purified, purified, unpurified, etc.)	2	This can't be a required field as this information is not always provided by the vendor.	BAO	Protein purity	http://bioportal.bi	

LINCS standardized protein information									Notes
Unique ID	LINCS Field Name	Related to	Description	Importance (1: essential; 2: desirable / recommende 3: optional)	Comments	Ontologies / references considered	Suggested terminology	Link to ontology / reference	Additional Notes (for development)
PP:14	PP_Protein_Complex	canonical protein	The description of a macromolecular complex composed of two or more polypeptide subunits, which may or may not be identical. The individual subunits should be described as being part of a protein complex (e.g. heterotrimer of alpha, beta and gamma subunits), the name of the protein complex, individual subunit protein name, protein ID, gene name and gene ID.	1, If relevant	If a protein complex is described and given a single LINCS ID, information about each subunit will have to be provided (e.g the entry will have multiple UniProt IDs associated with it....).	PRO	Protein complex: has part subunit	http://bioportal.bi	
PP:15	PP_Isoform	canonical protein	This describes whether the protein is an isoform and differs from the initial protein due to either alternative splicing, alternative initiation or ribosomal frameshifting during translation. The isoforms have different protein sequences and hence different protein IDs, though they share the same gene ID.	2	This can't be a required field as this information is not always provided by the vendor.	UniProt	Protein isoform	http://www.unipro	
PP:16	PP_Protein_Type	canonical protein	A controlled vocabulary, if one exists, specifying whether the protein is a growth factor, peptide, protein, etc.	3	The vocabulary here is important; it gets fuzzy as what is a growth factor vs. a cytokine. This can be made unambiguous by describing these concepts as a role of a protein, whether it functions as a 'ligand: growth factor' or a 'ligand: cytokine'. We agree that this seems important, but no good ontology of protein types exists and so categorizing proteins by type is a very subjective exercise now.	_PRO	Ligand: growth factor; Ligand: cytokine	http://bioportal.bi	
PP:17	PP_Source_Organism	batch	A controlled vocabulary describing the source of the protein (e.g. mouse, rabbit, horse, goat, etc.)	2		NCBITaxon	Organism	http://bioportal.bi	
PP:18	PP_Reference		Appropriate literature references can be provided.	2		PubMed	PMID	http://www.ncbi.n	
LINCS experimental protein related information (experiment specific treatment with protein/peptides)									Notes
EXP_PP: 1	PP_Protein_Concentrati	experiment	The final concentration of protein/peptide used in the assay	1		BAO	Concentration value; Concentration unit	http://bioportal.bi	
EXO_PP: 2	PP_Incubation_Time	experiment	The time of treatment with the protein/peptide target in the assay	1		BAO	Incubation time	http://bioportal.bi	