RLIMS-P Annotation guidelines

This document describes some curation guidelines for annotation in RLIMS-P.

The curation task is divided into 2 subtasks:

i) Phosphorylation information

- Provide information about **substrates**, **kinases**, and **sites** at the **abstract level and with experimental information**.

  Curation abstract level: As mentioned before curation at abstract level means that if the abstract mentions in different ways that protein X is phosphorylated by kinase Y, you only select one of the instances for annotation.

  e.g. information in sentence 7 and 10 present redundant information only validate 1

RLIMS-P Annotation ?				
No.	Kinase	Substrate	Site	Sentence
1		beta-catenin	Ser-552, Ser-665	7
2		beta-catenin	Ser-552, Ser-675	10
3		beta-catenin mutant	Ser-552, Ser-665	11

7 FSH promoted the PKA -dependent , PI3K-independent phosphorylation of beta-catenin on Ser552 and Ser665.

11 Transduction with an Ad-phospho-beta-catenin mutant (Ser552/665/Asp) enhanced Lhcgr mRNA expression in FSH-treated cells greater than 3-fold.

Experimental information: RLIMS-P will identify all mentions of phosphorylation including those describing general statements or statements of previous works, but you should only annotate the experimentally supported ones.

e.g. No experimental information in the document.

Glycogen synthase kinase-3beta (GSK-3B) *phosphorylates* ***tau*** *protein, and increased GSK-3B expression has been associated with neurofibrillary tangles."* PMID: 21443865.

RLIMS-P Annotation ?						
No.	Kinase	Substrate	Site	Sentence	Comment	Validation
1	glycogen synthase kinase-3beta (gsk-3b)	tau protein		3	no experimental information	

In the example above, although the entities extracted by RLIMS-P are correct, for curation purposes the document is no suited as the document is about a study of GSK3B and MAPT polymorphisms in Alzheimer disease, and not the phosphorylation. In this case ignore the validation but add a comment "no experimental information".

e.g. Documents with experimental information should be validated (see green check in validation in figure below).

PMID:23382206

Here we show , in the developing chick neural tube, that phosphorylation of Sox9 on S64 and S181 facilitates its SUMOylation , and the phosphorylated forms of Sox9 are essential for trunk neural crest delamination .

Text Evidence

PubMed Information				
23382206	2013 Feb 19	Liu JA, Wu MH, Yan CH, Chau BK, So H...	Proc Natl Acad Sci U S A	Full Text

RLIMS-P Annotation ?						
No.	Kinase	Substrate	Site	Sentence	Comment	Validation
1		sox9	Ser-181, Ser-64	5		✓ ✗

Add Annotation

Kinase information: report kinase only when it is mentioned along with a substrate in the abstract. For example, if the abstract has a sentence of the type 'Y is a tyrosine kinase and here we show that it phosphorylates X', we **would annotate** Y as a kinase, and X as a substrate or site in substrate.
Do not include in the general statements

e.g. we would annotate the **kinase** (in green) in these examples because the sentence (or abstract) describes a **substrate** for it (in blue) or the site (red):

*"Autoradiography showed that **TTP** was phosphorylated in vitro by **GSK3b, PKA, PKB, PKC**, but not Cdc2, in addition to **p42, p38**, and **JNK**."* PMID: 18071886.

*"Furthermore , we demonstrate that **PKC** can directly phosphorylate **S413** in vitro."* PMID: 23909401.

However, if the abstract has a sentence of the type protein 'Y is a tyrosine kinase' without any mention of substrate in the abstract, we **would not annotate** the kinase Y. However

e.g. we **would not annotate** kinase in these examples because no substrates are mentioned (we show a sentence but it is true for the abstract):

*"Here we review evidence that inhibition of **glycogen synthase kinase-3 (GSK3)** ameliorates cognitive deficits in a wide variety of animal models of CNS diseases, including Alzheimer's disease, Fragile X syndrome, Down syndrome, Parkinson's disease, spinocerebellar ataxia type 1, traumatic brain injury, and others."* PMID: 23916593.

*"**cGMP-dependent protein kinase II** (cGKII; encoded by PRKG2) is a serine/threonine kinase that is critical for skeletal growth in mammals; in mice, cGKII deficiency results in dwarfism."* PMID: 18551195.

What is a kinase or substrate in RLIMS-P?

The kinase or substrates could be **individual proteins**, **protein complexe**s or **protein family**. All should be validated, however not all of these can be normalized (see later).

Example protein complex:

'Protein X is phosphorylated by Z complex.'
Both X and Z will be annotated by RLIMS-P, X as substrate and Z as kinase.

e.g. in this case CDK1-cyclin B is annotated as **kinase**.

"Here , we report that human Crm1 is phosphorylated at serine 391 in mitosis by CDK1-cyclin-B" PMID: 23729730.

RLIMS-P Annotation ?

No.	Kinase	Substrate	Site	Sentence	Comment	Validation
1	cdk1-cyclin-b	human crm1	Ser-391	6		✓ ✗

Add Annotation

Example protein family:

Src kinases constitute a group of proteins, and in the following example would be annotated as **substrate** by RLIMS-P.

*"PAG, the phosphoprotein associated with glycosphingolipid-enriched microdomains (GEM), negatively regulates Src family kinases by recruiting C-terminal Src kinase (Csk) to the membrane, where **Csk** phosphorylates the inhibitory tyrosine of the **Src kinases**."* PMID: 18085663.

RLIMS-P Annotation ?

No.	Kinase	Substrate	Site	Sentence	Comment	Validation
1	csk	src kinases	Tyr	2		✓ ✗

Add Annotation

Same with PKC example below, there are many PKC (alpha, beta, gamma) and in this case PKC would be annotated as **kinase:**

*"Furthermore , we demonstrate that **PKC** can directly phosphorylate **S413** in vitro."* PMID: 23909401.

The site is the phosphorylation site in the substrate and could be a residue number (protein X is phosphorylated at position 123); an amino acid with or without number (protein X is phosphorylated on serine residues. Ser-123 is phosphorylated by Z kinase) or a region, motif or domain (the phosphorylation site is located on the N-terrminal region.)
Examples of site that are annotated (in red)

*"Functionally , we show that phosphorylation of **CREB** on **S270/271** during mitosis correlated with reduced CREB chromatin occupancy ."*

*"PKC collaborate to induce the phosphorylation of **Serine 413** on **PIP5K1B** resulting in decreased kinase activity and reduced PtdIns (4,5) P2 synthesis ."*

*"…**tyrosine** phosphorylation of **PKCdelta** in response to sublethal levels of H2O2 ."*

ii) The second task is for **individual protein mentions** link kinase and substrate mentions to **UniprotKB accessions** whenever possible.
RLIMS-P suggests in many cases some UniProtKB entries that potentially correspond to the mentions in the abstract. Your task is to select the one(s) that corresponds to the protein(s) in question. In some cases there will be redundant entries (many UniProtKB entries for same protein and species). Please pick only one. You can add new information for protein and accession if it hasn't been suggested by RLIMS-P.
In cases of a protein complex or protein family mention is the kinase or substrate, there is no need to normalize, but if the information in the document allows you to determine which specific proteins in the family or complex, then you should try to normalize them.