Last updated 08/28/2013

RLIMS-P Website Help Document

Table of Contents

URL: http://annotation.dbi.udel.edu/text_mining/rlimsp2/

Introduction

RLIMS-P (Figure 1) is a rule-based text-mining program specifically designed to extract protein phosphorylation information on protein kinases, substrates and phosphorylation sites from biomedical literature (Hu *et al.*, 2005). **RLIMS-P** currently works on PubMed abstracts, but it will be extended to open access full text articles soon. **RLIMS-P** allows users to quickly find the relevant literature for phosphorylated proteins and their kinases, thereby facilitating the study of kinase-substrate networks.

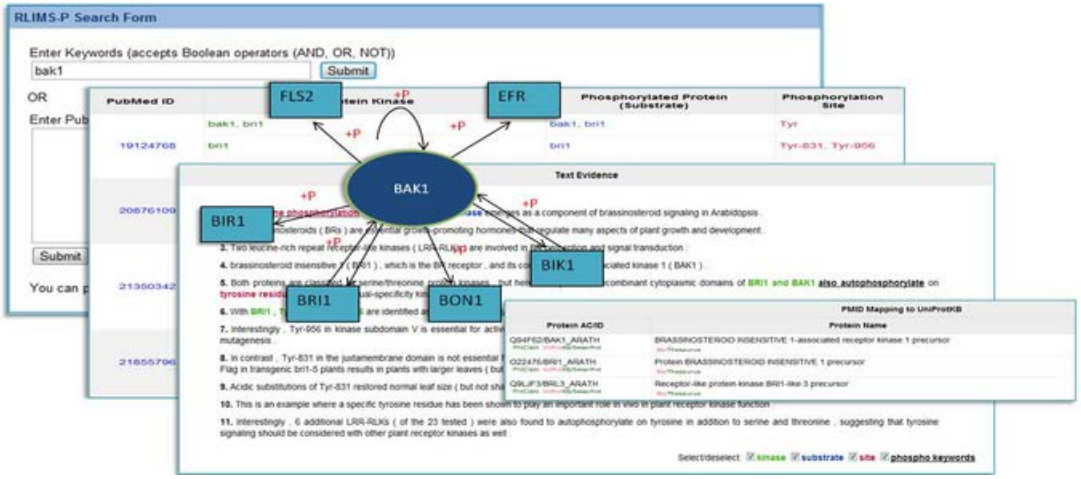


Figure 1: RLIMS-P overview.

RLIMS-P architecture

The RLIMS-P website consists of two parts: **1**) a back-end database and **2**) a web interface (Figure 1). Phosphorylation information is first extracted from Medline abstracts by RLIMS-P version 2.0, and then processed and stored in the database for easy and fast later retrieval. The web interface enables users to search for phosphorylation information using keywords or a list of PMIDs. The results (kinase, substrate, site) are displayed in sortable tables, which are downloadable for further research.

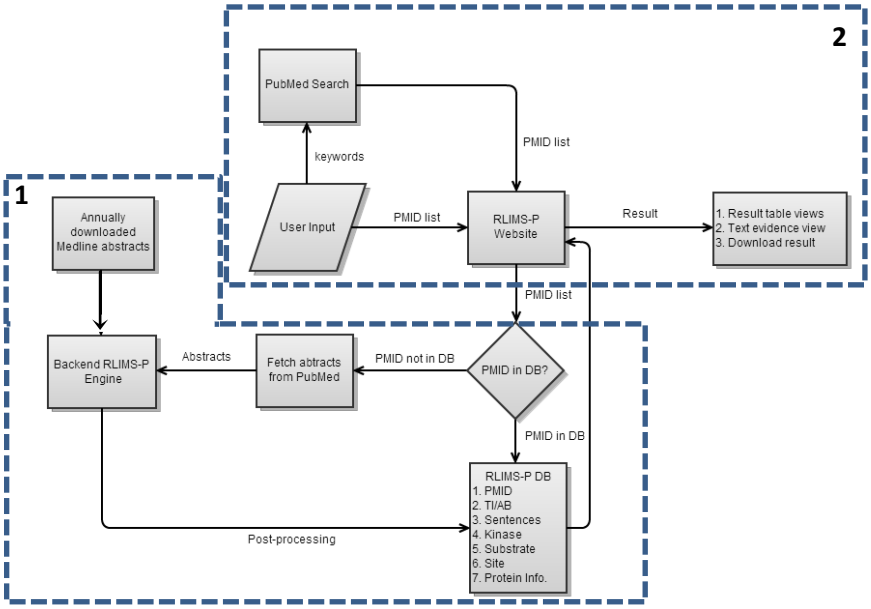


Figure 2: RLIMS-P system architecture

RLIMS-P interface

Login

To edit and export curated RLIMS-P results, users need to login (Figure 3, 1). In order to login for the first time, users need to sign up (Figure 3, 2) by entering their e-mail, name and affiliation (Figure 3, 3). Once logged in, the heading will appear as in Figure 4.

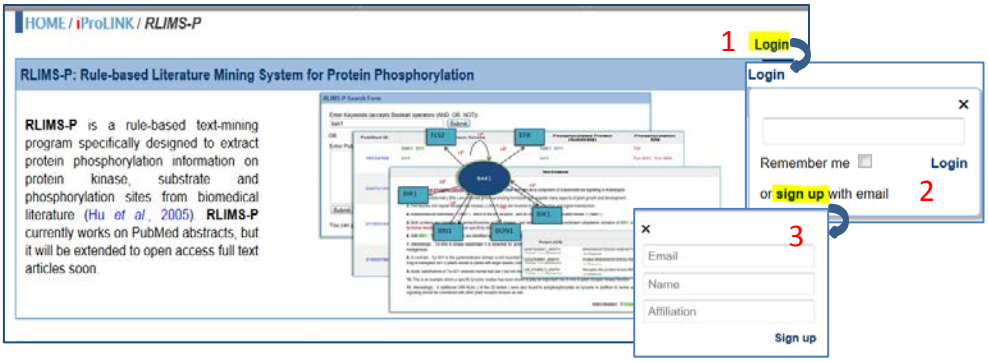


Figure 3: Login screen

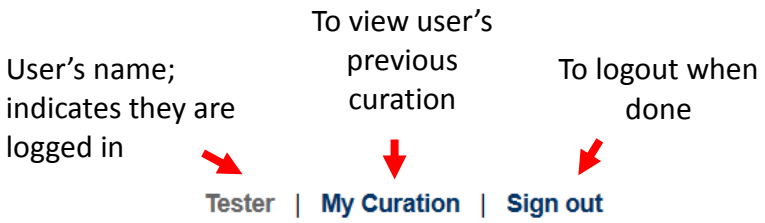


Figure 4: Appearance of heading after login

Input page

In short
Input: Keywords or Pubmed IDs (IDs)

The new website allows the input of keywords or phrases (Figure 5, **1**) that can be combined using Boolean operators. This input is used to query PubMed documents, and relevant documents are retrieved for processing by RLIMS-P. Alternatively a list of PMIDs (Figure 5, **2**), delimited by comma, space, or new line can be entered. Users can enter up to 200 PMIDs per run. In both cases, clicking on Submit Query will retrieve the results, and clicking on Reset will empty the query box, so that a different query can be entered.

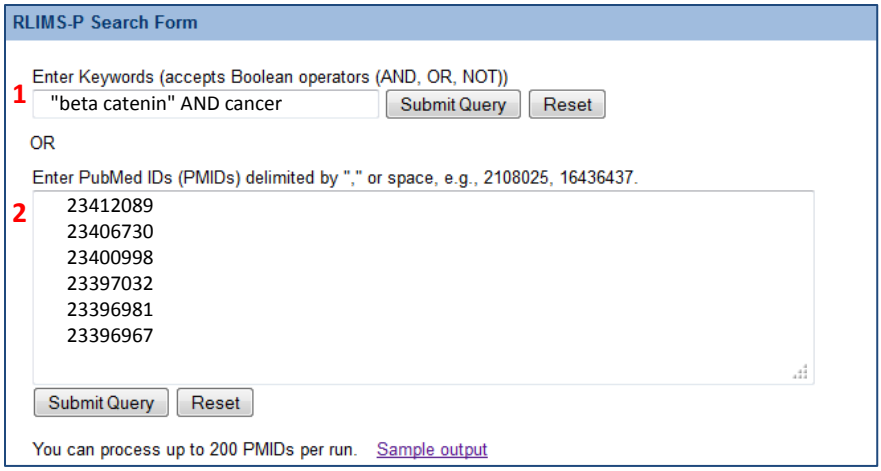


Figure 5: RLIMS-P input interface

Results Page

In short
RLIMS-P Results: statistics, summary table with different views (kinase, substrate, PMID), access to text evidence, and table saving

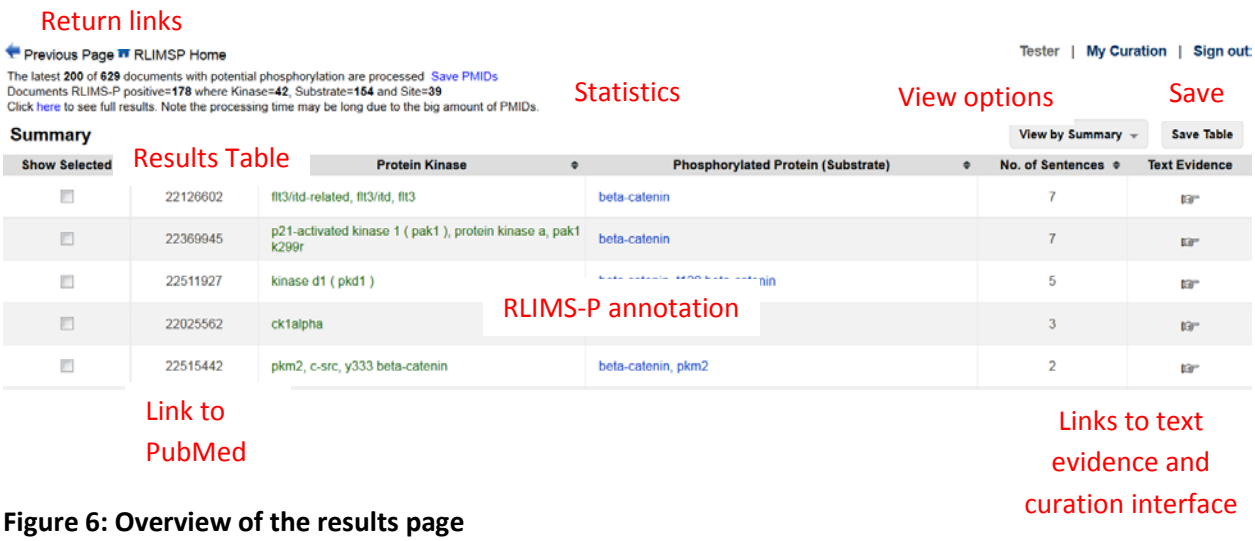


Figure 6: Overview of the results page

Overview: The results page contains the search statistics and the results table (Figure 6). Users can customize their view of the information in the table and download their results from this page.

RLIMS-P Statistics: The new RLIMS-P results page presents detailed statistics on the documents with potential phosphorylation information (those containing a phosphorylation-related trigger word) and those with phosphorylation information according to RLIMS-P processing (Figure 7). For convenience, only the results for the latest 200 PMIDs are shown for a keyword search, but the user can choose to access the full result set.

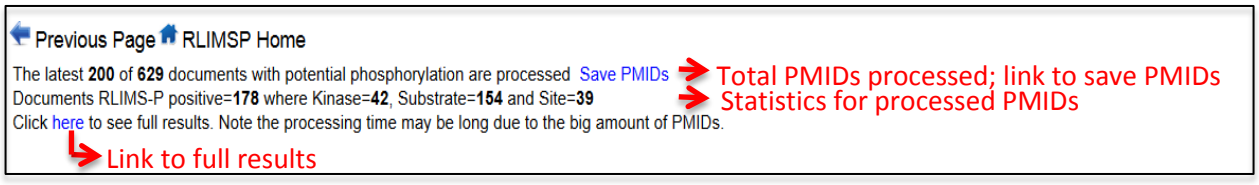


Figure 7: RLIMS-P Statistics

Results table:

(i) Columns in the results table: The results table contains the following columns. Note that the order and appearance of these columns will vary depending on a variety of user-settable options (see below).

Show Selected: Allows the user to select which annotation lines to display. Annotation lines are selected by clicking on the corresponding check boxes and then on "Show Selected" (Figure 8). annotation lines by clicking on the corresponding check boxes and then on Show Selected.

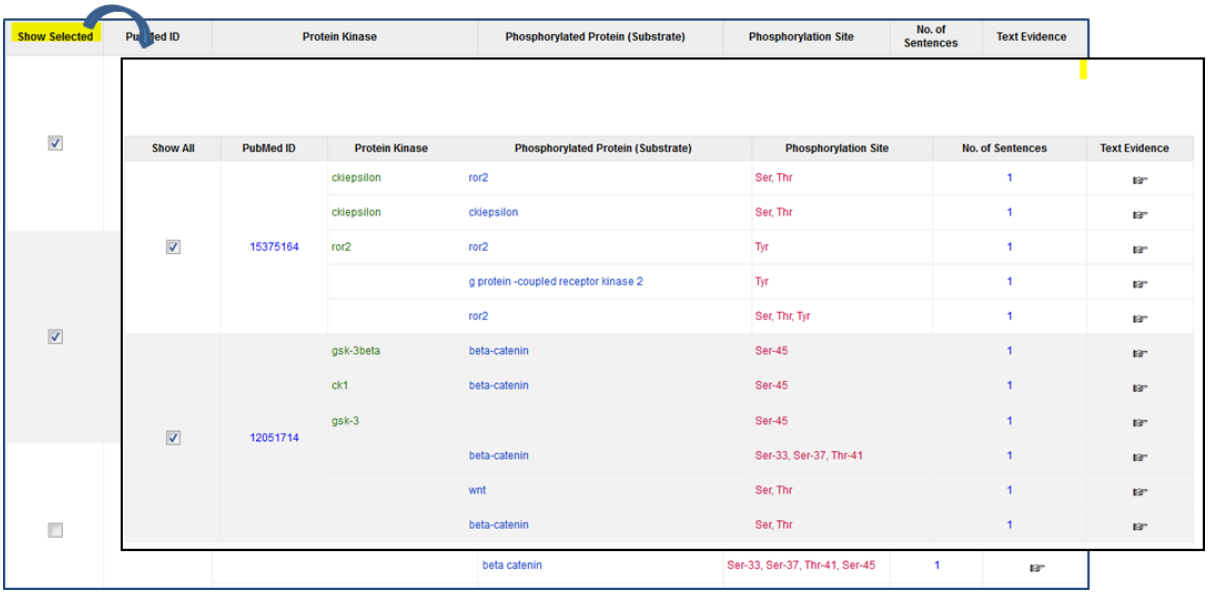


Figure 8: Displaying desired annotation lines using the Show Selected column

PubMed ID: Displays the PubMed ID of each RLIMS-P positive document. Clicking on the ID will link to the PubMed abstract.

Protein Kinase: Kinases identified by RLIMS-P are shown in green.

Phosphorylated Protein (Substrate): Proteins determined by RLIMS-P to be phosphorylated substrates are shown in blue.

Phosphorylation Site: Phosphorylation sites identified by RLIMS-P are shown in red. This column is not included in the Summary view (see below).

No. of Sentences: Indicates the number of sentences from the document that contain evidence for that line of annotation. Clicking on the number links to the Text Evidence/Curation page for that line of annotation (Figure 9).

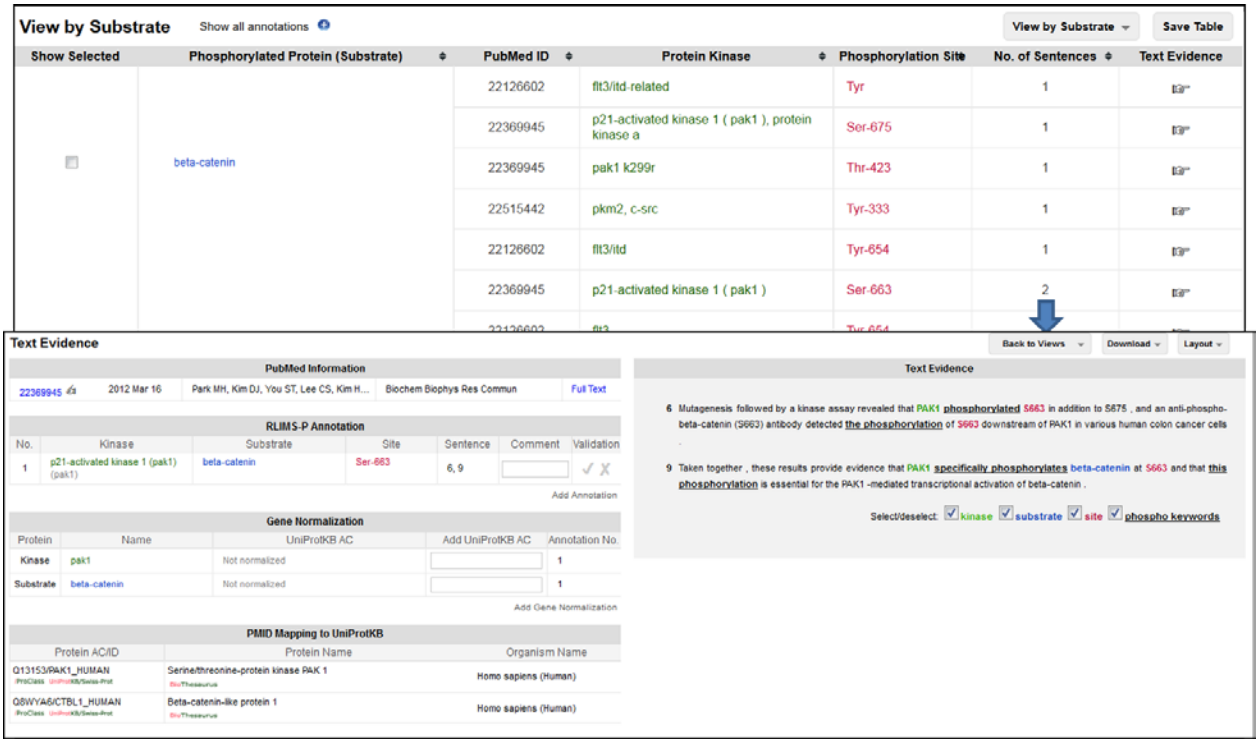


Figure 9: Clicking on the number of sentences displays the Text Evidence page for that line of annotation

Text Evidence/Curation: Clicking on the ☞ icon provides access to the RLIMS-P text evidence and editing/curation page for the entire title and abstract of the document indicated on that line of annotation (see Text Evidence/Curation section below).

(ii) Column sorting: Each column in the results table can be sorted based on ascending or descending numerical or alphabetical by clicking on the arrows next to the column headings (Figure 10).

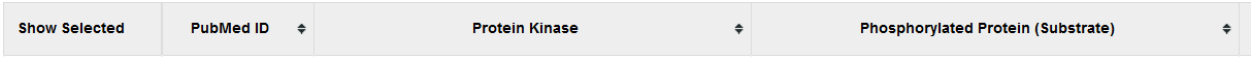


Figure 10: Arrows next to the column headings can be used to sort the results

(iii) View Options: Users can organize the display in the results table according to their interests using the "View by" menu (Figure 6, View options).

View by Summary: The default view of the RLIMS-P results table is the Summary view, in which all of the kinases and phosphorylated substrates identified in a particular document are summarized in a single line of annotation (Figure 11). Documents containing kinase, substrate and site information are listed first. Phosphorylation site information is not presented in this view.

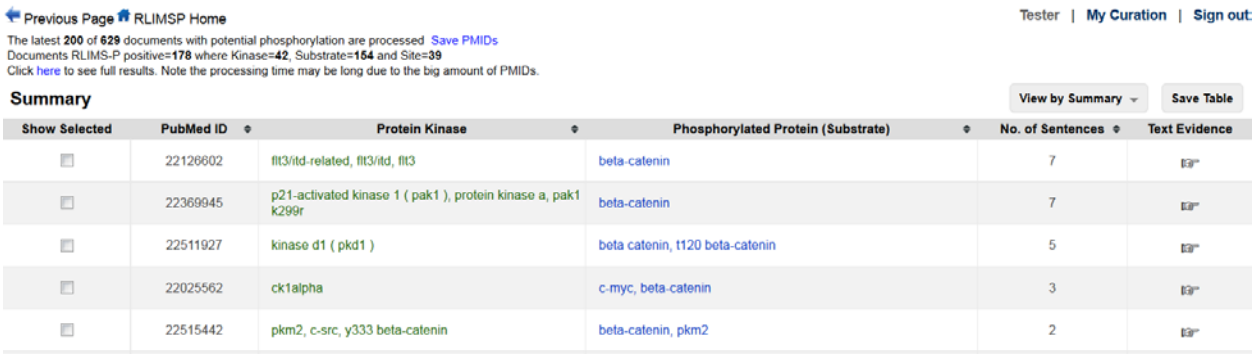
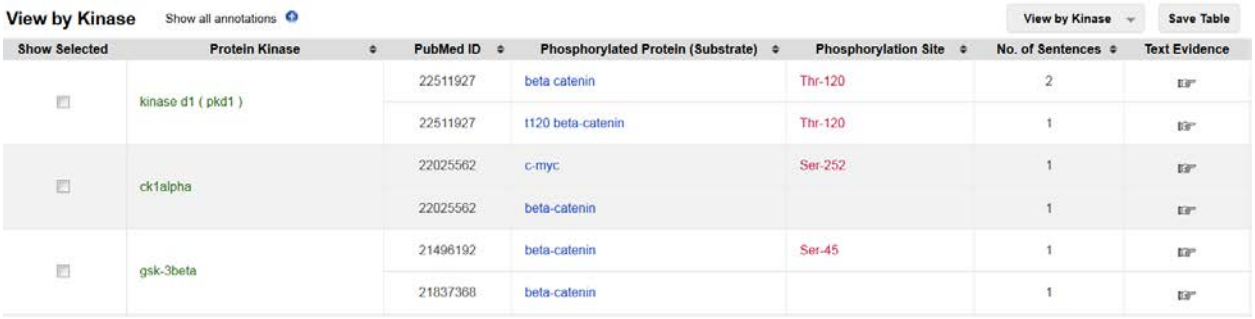


Figure 11: View by Summary

View by PMID: This view is document-centric, grouping together all of the annotation lines for a particular document (Figure 12). Unlike the Summary view, each line of annotation in the PMID view consists of a single kinase, its substrate, and the corresponding phosphorylation site(s). Columns are left blank if kinase, substrate, and/or site information was not obtained from the document.

Figure 12: View by PMID

View by Kinase: The view is kinase-centric, grouping together information for each unique kinase mentioned in the document set (Figure 13). Note that we are still working on improving the standardization of protein names, and therefore in some cases, if a kinase is referred to by multiple names, all mentions of the kinase will not be collected into a single group. Within a group, each line of annotation shows an individual substrate of the kinase, the phosphorylation site(s), and the PMID for the document containing the evidence for that annotation. Substrate and/or site columns will be left blank if that information was not obtained from a particular document.

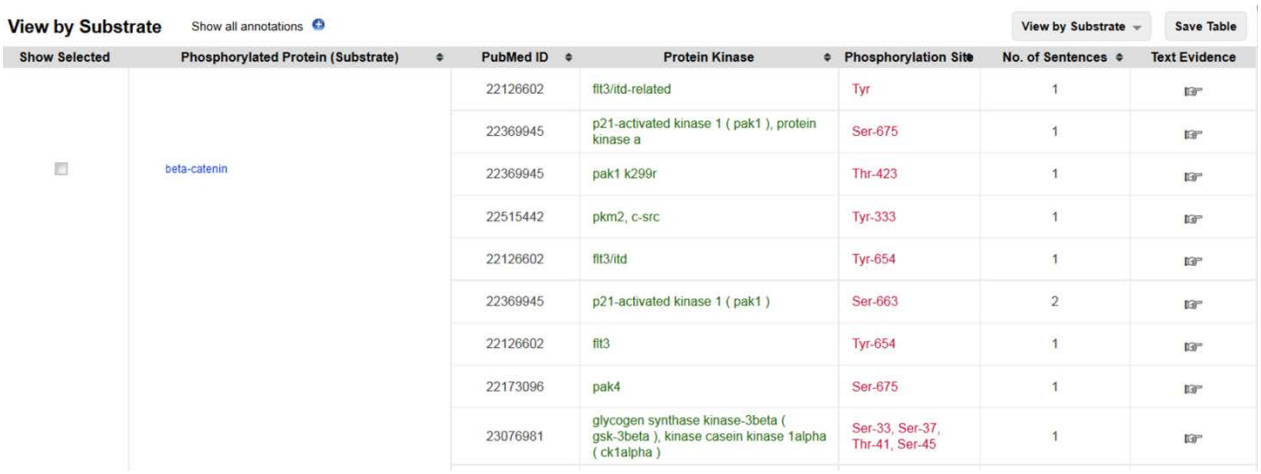


View by Kinase Show all annotations View by Kinase Save Table

Show Selected	Protein Kinase	PubMed ID	Phosphorylated Protein (Substrate)	Phosphorylation Site	No. of Sentences	Text Evidence
☐	kinase d1 (pkd1)	22511927	beta catenin	Thr-120	2	☞
		22511927	t120 beta-catenin	Thr-120	1	☞
☐	ck1alpha	22025562	c-myc	Ser-252	1	☞
		22025562	beta-catenin		1	☞
☐	gsk-3beta	21496192	beta-catenin	Ser-45	1	☞
		21837368	beta-catenin		1	☞

Figure 13: View by Kinase

View By Substrate: This view is substrate-centric, grouping together information for each unique substrate mentioned the document set (Figure 14). Note that we are still working on improving the standardization of protein names, and therefore in some cases, if a substrate is referred to by multiple names, all mentions of the substrate will not be collected into a single group. Within a group, each line of annotation shows an individual kinase for the substrate, the phosphorylation site(s), and the PMID for the document containing the evidence for that annotation. Kinase and/or site columns will be left blank if that information was not obtained from a particular document.



View by Substrate Show all annotations View by Substrate Save Table

Show Selected	Phosphorylated Protein (Substrate)	PubMed ID	Protein Kinase	Phosphorylation Site	No. of Sentences	Text Evidence
☐	beta-catenin	22126602	flt3/itd-related	Tyr	1	☞
		22369945	p21-activated kinase 1 (pak1), protein kinase a	Ser-675	1	☞
		22369945	pak1 k299r	Thr-423	1	☞
		22515442	pkm2, c-src	Tyr-333	1	☞
		22126602	flt3/itd	Tyr-654	1	☞
		22369945	p21-activated kinase 1 (pak1)	Ser-663	2	☞
		22126602	flt3	Tyr-654	1	☞
		22173096	pak4	Ser-675	1	☞
		23076981	glycogen synthase kinase-3beta (gsk-3beta), kinase casein kinase 1alpha (ck1alpha)	Ser-33, Ser-37, Thr-41, Ser-45	1	☞

Figure 14: View by Substrate

(iv) Expanded RLIMS-P results: Clicking on the  icon will expand the table to show the full RLIMS-P results, including information that is partially redundant (as is the case for the annotation in the dashed box in Figure 15). This option is available for all of the table arrangements described in the previous section.

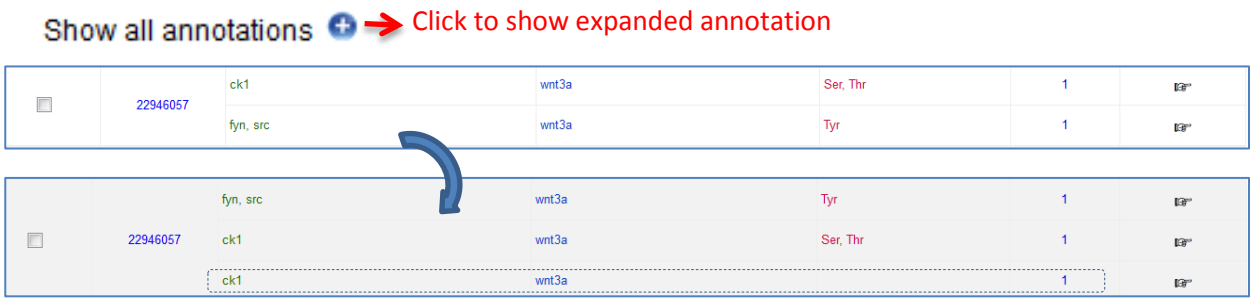


Figure 15: Expanded annotation

(v) Downloading the results table: Clicking on the Download button will create a comma-delimited file containing the PMID, kinase, substrate, and site information and the associated evidence sentences (Figure 16). The order of the information in the file will vary depending on the view from which is was downloaded.

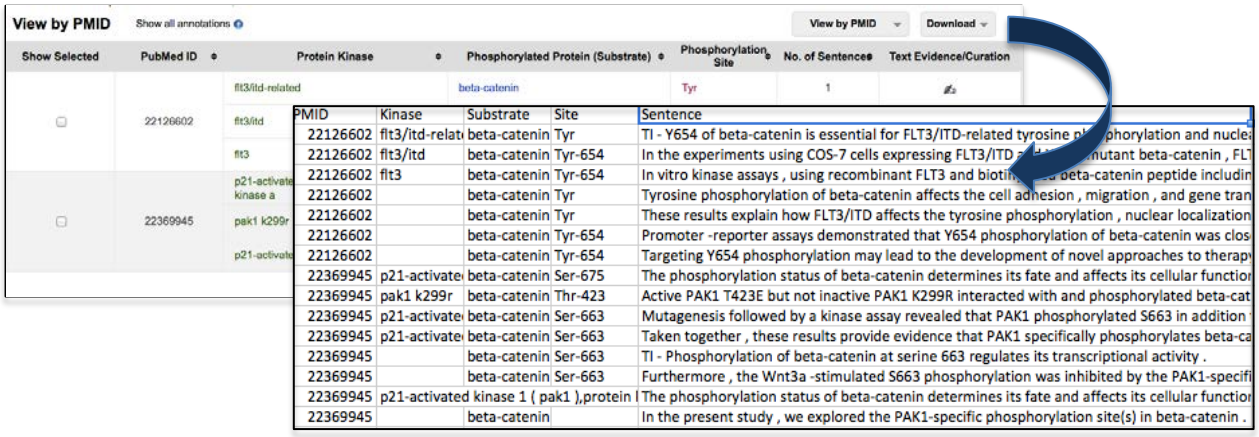


Figure 16: Downloading the results table

Text Evidence/Curation Page

In short
RLIMS-P Text Evidence/Curation page: statistics, text evidence, curation interface (including PubMed information, RLIMS-P annotation, gene normalization, and PMID mapping to UniProtKB), and download options

Accessing the Text Evidence/Curation Page: Clicking on the ☞ icon in the results table provides access to the RLIMS-P text evidence and editing/curation page for the entire title and abstract of the document indicated on that line of annotation (Figure 6). Clicking on the number in the "No. of Sentences" column of the results table provides access to the text evidence and editing/curation page for that line of annotation only (Figure 6, Figure 9).

Returning to the results table: From the Text Evidence/Curation page, users can return to the Results page by choosing a viewing option (Summary, PMID, Kinase, or Substrate) from the Back to Views menu (Figure 17).

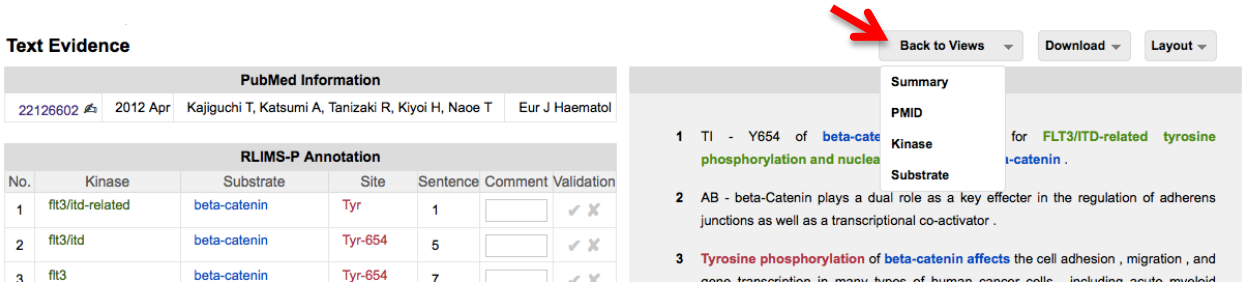


Figure 17: The Back to Views menu on the Text Evidence/Curation page

RLIMS-P Statistics: Like the Results page, the Text Evidence/Curation page displays the RLIMS-P statistics for the current query (Figure 7).

Layout: Using the Layout menu, users can switch between a two-column (Figure 18) and one-column (Figure 19) layout of the Text Evidence/Curation page.

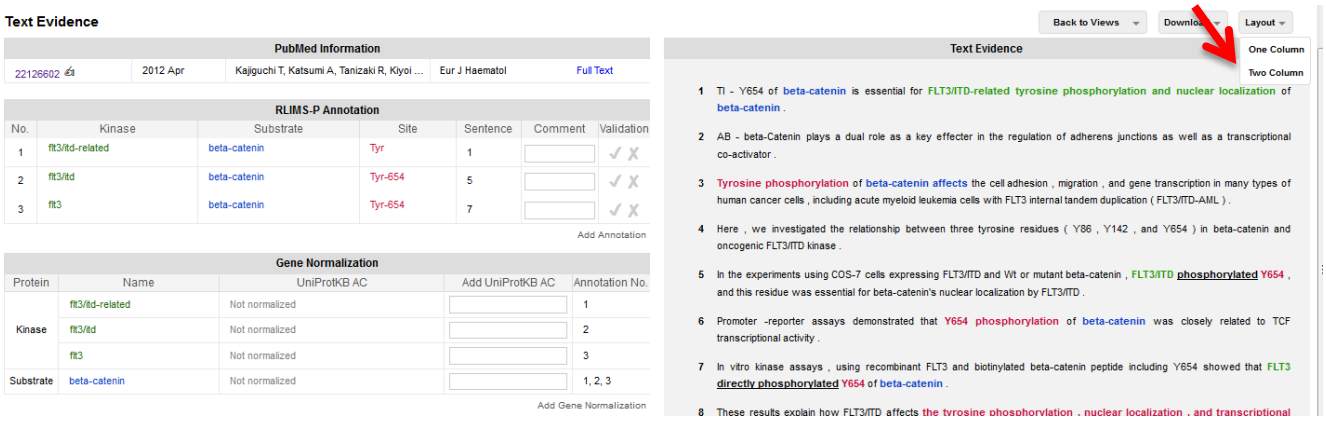


Figure 18: Two-column layout of the Text Evidence/Curation page

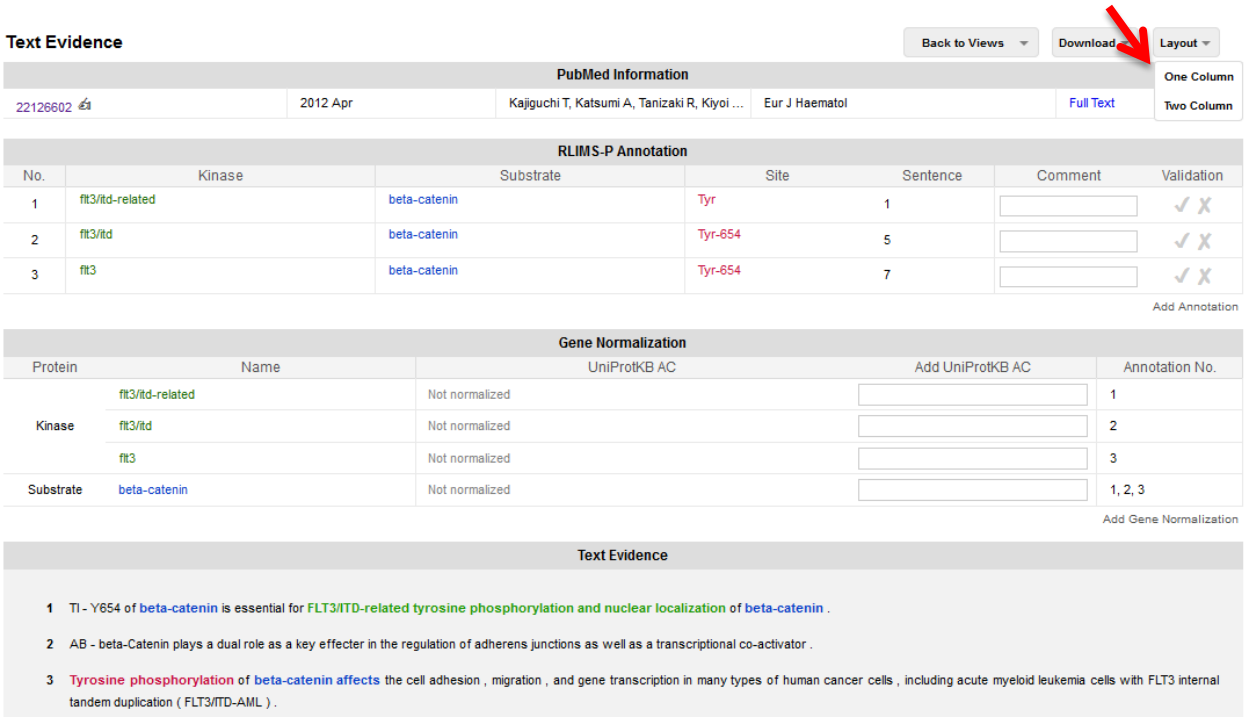


Figure 19: One-column layout of the Text Evidence/Curation page

Text Evidence: When accessed via the ☞ icon in the results table, the Text Evidence section displays each sentence of the title and abstract with phosphorylation-related information highlighted. By default, kinase (green), substrate (blue), site (red), and phospho-keyword (black, underlined) evidence is highlighted (Figure 20). Users can customize the highlighting using the check boxes provided. When accessed via the number in the No. of Sentences column in the results table, the Text Evidence section will display only the sentences containing evidence for that line of annotation (Figure 9).

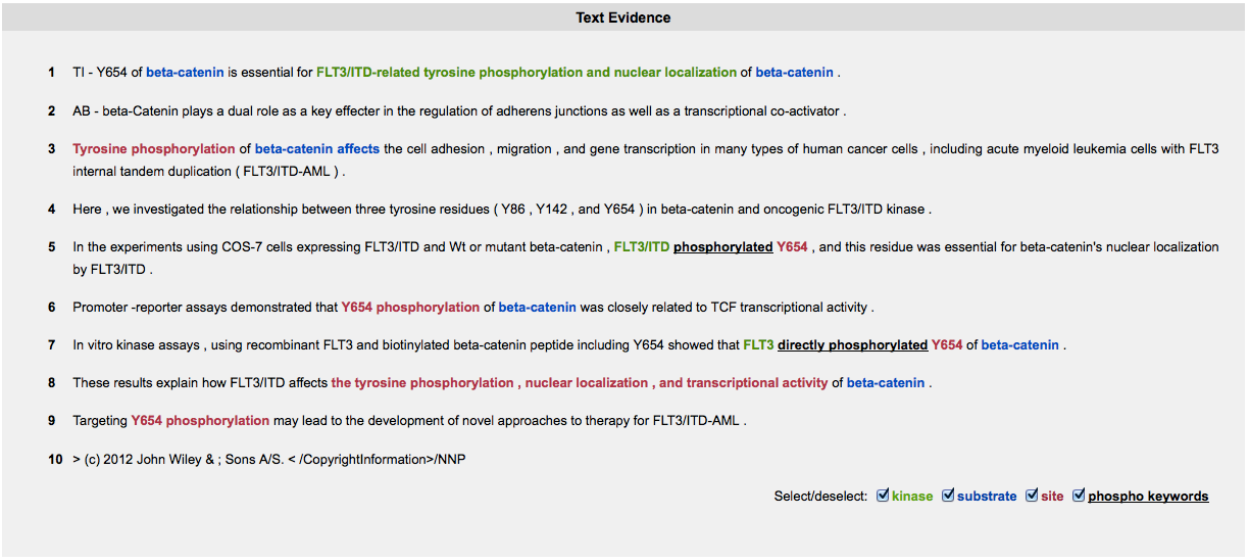
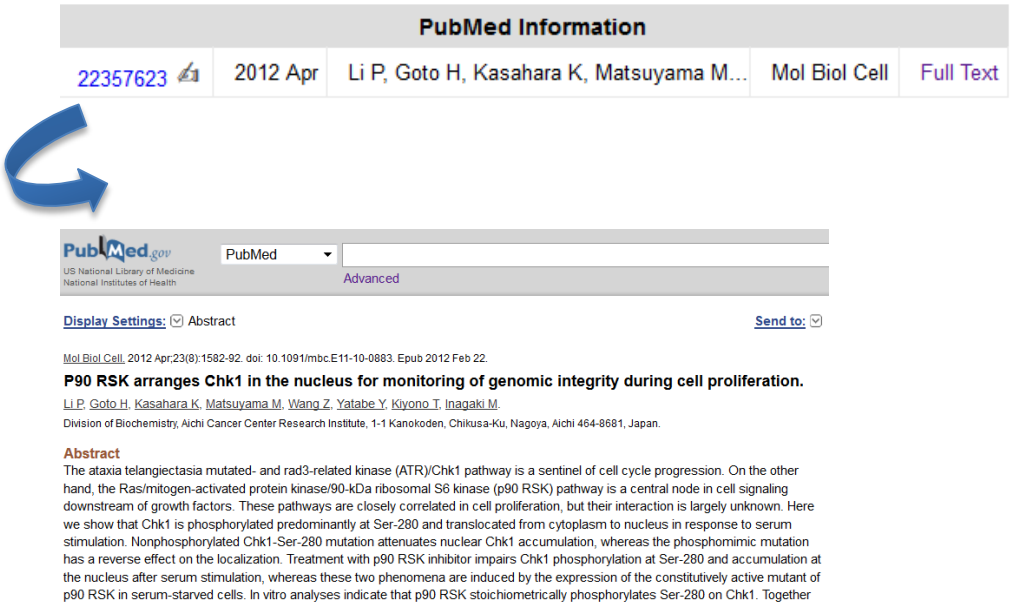


Figure 20: Text Evidence display

Curation Interface: The curation interface portion of the Text Evidence/Curation page allows the user to validate the RLIMS-P phosphorylation annotation and gene normalization. If there are any errors or omissions, the user can enter the correct information. The interface is divided into four sections: PubMed Information, RLIMS-P Annotation, Gene Normalization, and PMID Mapping to UniProt KB.

(i) PubMed Information: The PubMed information section displays the PMID, publication date, authors, and journal for the annotated document. Clicking on the PMID links to the PubMed record for the document (Figure 21A). For open access articles, the link to Full Text is also available (Figure 21B)

A



B

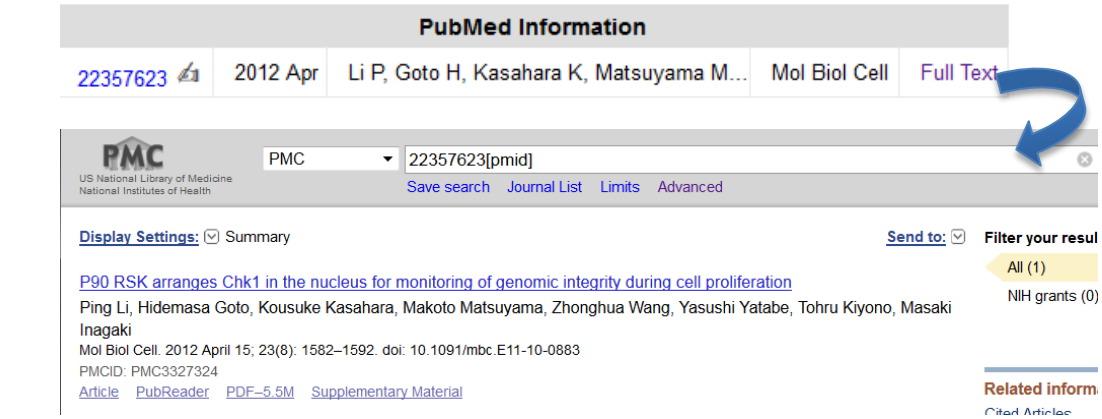


Figure 21: PubMed Information section of the curation interface

(ii) RLIMS-P Annotation: This section displays a table of the RLIMS-P phosphorylation annotation lines. The first five columns—No. (line number), Kinase, Substrate, Site, and Sentence (sentence numbers from which evidence was obtained)—present the same information as in the results table and are not modifiable by the user. The last two columns—Comment and Validation—accept user input. In the Validation column, the user can click the check mark (turns green when clicked) if the RLIMS-P annotation on that line is correct and the X (turns red when clicked) if the annotation is incorrect (Figure 22). In the Comment column, the user can enter free-text comments; for example, if the annotation is incorrect, the user can provide a brief explanation of what is wrong (Figure 23). Clicking on Add Annotation at the bottom of the table creates a User Added Annotation section, which allows the user to enter additional kinase, substrate, site, and/or text evidence information (Figure 24).

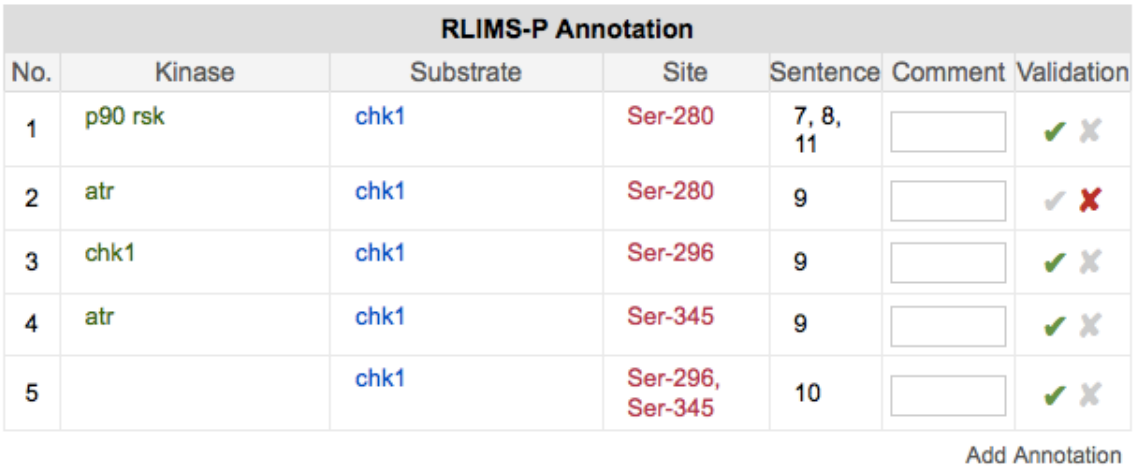


Figure 22: RLIMS-P Annotation section of the curation interface showing user validation

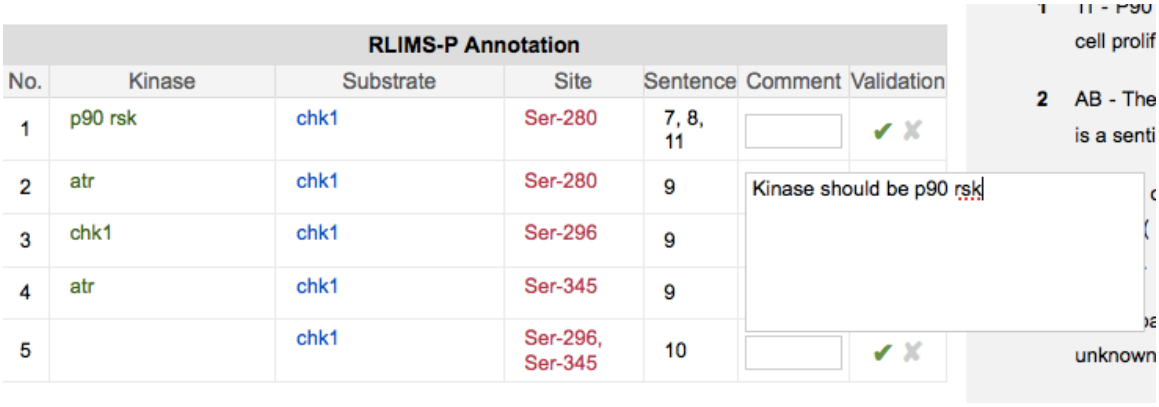
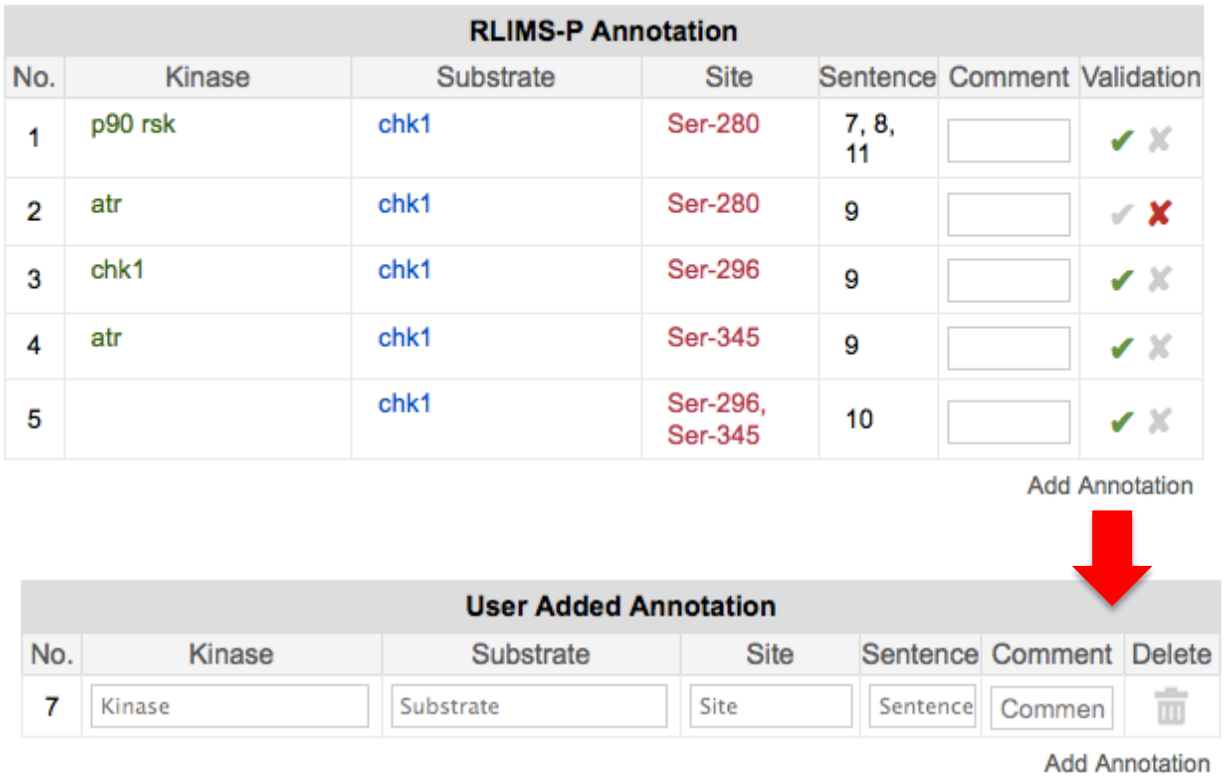


Figure 23: RLIMS-P Annotation section of the curation interface showing a user-entered comment



RLIMS-P Annotation

No.	Kinase	Substrate	Site	Sentence	Comment	Validation
1	p90 rsk	chk1	Ser-280	7, 8, 11		✔ ✘
2	atr	chk1	Ser-280	9		✔ ✘
3	chk1	chk1	Ser-296	9		✔ ✘
4	atr	chk1	Ser-345	9		✔ ✘
5		chk1	Ser-296, Ser-345	10		✔ ✘

Add Annotation

Figure 24: Clicking on the Add Annotation link below the RLIMS-P Annotation table allows users to enter new lines of annotation

(iii) Gene Normalization: The Gene Normalization section displays suggested UniProtKB accession numbers (UniProtKB ACs) for the kinase and substrate proteins mentioned in the RLIMS-P annotation (Figure 24). This normalization is done using the cross-species gene normalization tool, GenNorm (http://ikmbio.csie.ncku.edu.tw/GN/). Clicking on the UniProtKB AC links to the UniProtKB record. Users can indicate that the mapping is correct by clicking on the check mark in the UniProtKB AC box (turns the box green) or incorrect by clicking on the 'X' (turns the box red). Mousing-over any line in the table causes a search UniProt icon to appear. Clicking on this icon queries UniProtKB using the protein name as it appeared in the text. If the user identifies a UniProtKB AC that corresponds to the protein name, it can be entered in the Add UniProtKB AC column.

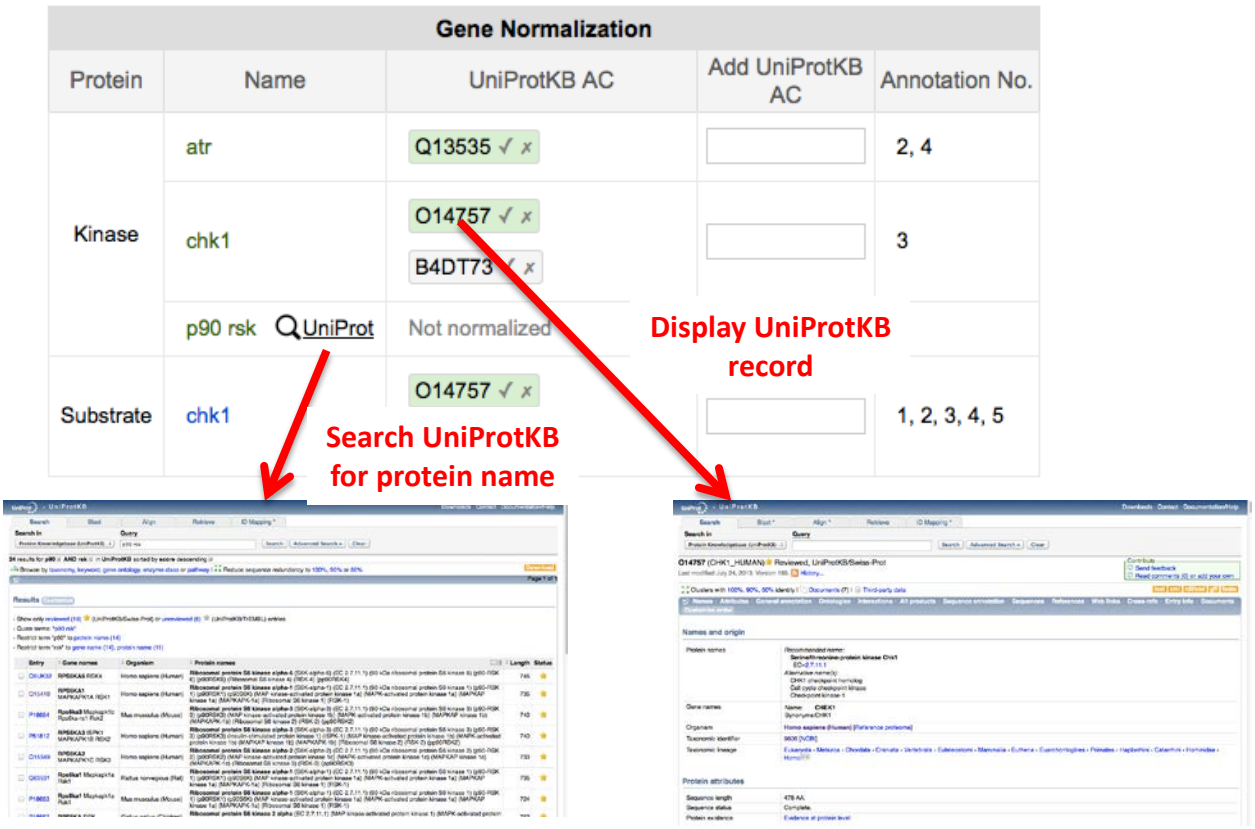


Figure 24: Gene Normalization section of the curation interface

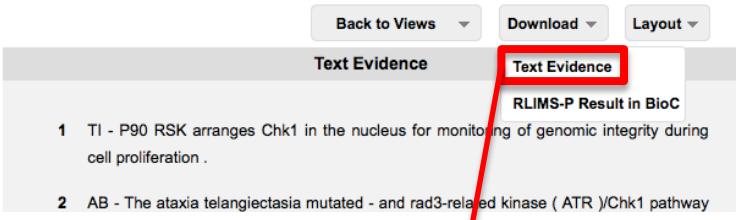
(iv) PMID Mapping to UniProt KB: The section displays a table with suggested UniProtKB ACs for the kinases and substrates obtained using a bibliography mapping service provided by the Protein Information Resource (pir.georgetown.edu). The information in this section can be used to assist in assigning UniProt KB ACs to the proteins mentioned in the RLIMS-P annotation and addition of these to the Gene normalization table. Each line provides the UniProtKB AC and ID (with links to the UniProtKB and iProClass records for the protein), the name of the protein as it appears in the UniProtKB record (with a link to Biothesaurus), and the organism name (Figure 25).

PMID Mapping to UniProtKB

Protein AC/ID	Protein Name	Organism Name
B4DDD0/B4DDD0_HUMAN iProClass UniProtKB/Swiss-Prot	cDNA FLJ59449, highly similar to Serine/threonine-protein kinase Chk1 ... BioThesaurus	Homo sapiens (Human)
B4DT73/B4DT73_HUMAN iProClass UniProtKB/Swiss-Prot	cDNA FLJ56409, highly similar to Serine/threonine-protein kinase Chk1 ... BioThesaurus	Homo sapiens (Human)
F5H7S4/F5H7S4_HUMAN iProClass UniProtKB/Swiss-Prot	Serine/threonine-protein kinase Chk1 BioThesaurus	Homo sapiens (Human)
O14757/CHK1_HUMAN iProClass UniProtKB/Swiss-Prot	Serine/threonine-protein kinase Chk1 BioThesaurus	Homo sapiens (Human)
Q15418/KS6A1_HUMAN iProClass UniProtKB/Swiss-Prot	Ribosomal protein S6 kinase alpha-1 BioThesaurus	Homo sapiens (Human)

Figure 25: PMID Mapping to UniProtKB section of the curation interface

Downloading Text Evidence/Curation: Selecting Text Evidence from the Download menu will create a comma-delimited file containing all of the information on the Text Evidence/Curation page including user-added validation and comments (Figure 26). Selecting RLIMS-P Result in BioC from the Download menu will create a file containing the RLIMS-P annotation in BioCreative format (Figure 27).



	A	B	C	D	E	F	G	H	I	J
1	###Text Evidence###									
2	#PubMed Information									
3	22357623	2012 Apr	Li P, Goto H,	Mol Biol Cell						
4	#RLIMS-P Annotation									
5	No.	Kinase	Substrate	Site	Comment	Validation	Sentence			
6	1	p90 rsk	chk1	Ser-280		Agree	Treatment with p90 RSK inhibitor impairs Chk1 phospho			
7	1	p90 rsk	chk1	Ser-280		Agree	In vitro analyses indicate that p90 RSK stoichiometrically			
8	1	p90 rsk	chk1	Ser-280		Agree	These results suggest that p90 RSK facilitates nuclear Chl			
9	2	chk1	chk1	Ser-296		Agree	Together with Chk1 phosphorylation at Ser-345 by ATR a			
10	3	atr	chk1	Ser-345		Agree	Together with Chk1 phosphorylation at Ser-345 by ATR a			
11	4	atr	chk1	Ser-280	Kinase should be p90 rsk	Disagree	Together with Chk1 phosphorylation at Ser-345 by ATR a			
12	5	N/A	chk1	Ser-280		N/A	Here we show that Chk1 is phosphorylated predominant			
13	6	N/A	chk1	Ser-296,Ser-345		Agree	In addition , Chk1 phosphorylation at Ser-345 and Ser-29			
14	#User Add Annotation									
15	Kinase	Substrate	Site	Comment	Sentence					
16	#Gene Normalization									
17	Protein	Name	UniProtKB Ac	Annotation N	Comment	Validation				
18	Kinase	atr	Q13535	3,4		Agree				
19	Kinase	chk1	O14757	2		Agree				
20	Kinase	chk1	B4DT73	2		N/A				
21	Substrate	chk1	O14757	1,2,3,4,5,6		Agree				
22	Substrate	chk1	B4DT73	1,2,3,4,5,6		N/A				
23	#PMID Mapping to UniProtKB									
24	Protein AC/Id	Protein Nam	Organism Name							
25	B4DDD0/B4D	cDNA FLJ594	Homo sapiens (Human)							
26	B4DT73/B4D	cDNA FLJ564	Homo sapiens (Human)							
27	F5H7S4/F5H	Serine/threo	Homo sapiens (Human)							
28	O14757/CHK	Serine/threo	Homo sapiens (Human)							
29	Q15418/KS6	Ribosomal p	Homo sapiens (Human)							
30										
31										

Figure 26: Downloading RLIMS-P annotation as a comma-delimited file

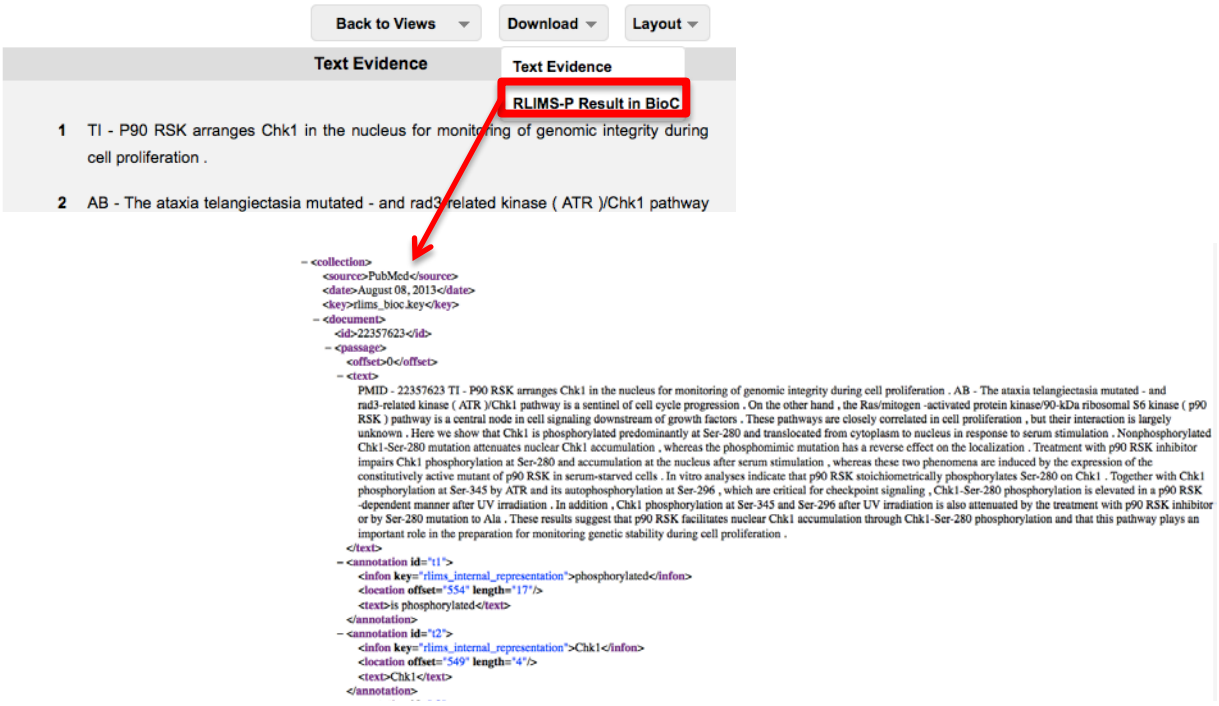


Figure 27: Downloading RLIMS-P annotation in BioC format