

## System requirements for BioCreative IV IAT task

**Prepared by:** Cecilia Arighi, Sherry Matis and Phoebe Roberts; with input from the User Advisory Group (<http://www.biocreative.org/#committee>)

We are asking for web-based text mining systems that allow the user to modify text mining results. Based on User Advisory Group (UAG) discussions, we have come up with a list of desired functionalities that already exist in working tools (see below). Note that the examples are supplied to clarify the intent of the functionality and are not meant to be comprehensive.

### Mandatory functionalities:

- System should be compatible with the most commonly used web browsers: Firefox, Chrome, Safari and Explorer.
- System should highlight entities and relationships (if applicable) relevant to the annotation task. We encourage color coding of entity types and links to relevant data sources when appropriate. E.g. Reflect (<http://reflect.embl.de/>) and PubTator (<http://www.ncbi.nlm.nih.gov/CBBresearch/Lu/Demo/PubTator/>) for entities. eFIP (<http://proteininformationresource.org/pirwww/iprolink/eFIP.shtml>) as example of entities and relations color coded.

The screenshot shows a web browser window with the Reflect interface. The main text area contains a paragraph about caldesmon: "cdc2, caldesmon. cdc2 and caldesmon are shown here to localize in membrane ruffles in motile cells. These results show that it promotes cell migration." Below the text, there are several colored boxes highlighting entities: "cdc2" (green), "caldesmon" (red), and "ATP" (blue). A sidebar on the right shows a search for "caldesmon" with results from Wikipedia and other sources. The interface includes a search bar, a list of entities, and a table of results.

Reflect

The screenshot shows the PubTator interface. At the top, there is a search bar and a "Go back" button. Below that, the PubMed ID (PMID:23244119) and the title of the article are displayed: "Menopausal Status Modifies Breast Cancer Risk Associated with ESR1 PvuII and XbaI Polymorphisms in Asian Women: a HuGE Review and Meta-analysis." The abstract text is shown below the title. At the bottom, there is a table of entities with columns for Entity type, Entity mention, Concept ID, Nomenclature, and Delete. The table lists several entities, including Breast Cancer, ESR1, Women, and Human, with their corresponding Concept IDs and Nomenclatures.

PubTator

PMID 10611223      Select/deselect:  kinase  substrate  site  interactant  impact  phospho/PP1

0. p21-activated kinase 1 phosphorylates the death agonist bad and protects cells from apoptosis .

1. Bad is a critical regulatory component of the intrinsic cell death machinery that exerts its death-promoting effect upon heterodimerization with the antiapoptotic proteins Bcl-2 and Bcl-x(L) .

2. Growth factors promote cell survival through phosphorylation of Bad , resulting in its dissociation from Bcl-2 and Bcl-x(L) and its association with 14-3-3tau .

3. Survival of interleukin 3 ( IL-3 ) -dependent FL5.12 lymphoid progenitor cells is attenuated upon treatment with the Rho GTPase-inactivating toxin B from Clostridium difficile .

4. p21-activated kinase 1 ( PAK1 ) is activated by IL-3 in FL5.12 cells , and this activation is reduced by the phosphatidylinositol 3-kinase inhibitor LY294002 .

5. Overexpression of a constitutively active PAK mutant (PAK1-T423E) promoted cell survival of FL5.12 and NIH 3T3 cells , while overexpression of the autoinhibitory domain of PAK ( amino acids 83 to 149 ) enhanced apoptosis .

6. PAK phosphorylates Bad in vitro and in vivo on Ser112 and Ser136 , resulting in a markedly reduced interaction between Bad and Bcl-2 or Bcl-x(L) and the increased association of Bad with 14-3-3tau .

7. Our findings indicate that PAK inhibits the proapoptotic effects of Bad by direct phosphorylation and that PAK may play an important role in cell survival pathways .

eFIP

- c. User should be able to edit the text mining results by correcting errors or adding missing information, and should be able to export the corrected data.
- Ability to review curation decisions and revise if necessary. Systems such as PubTator (see snapshot below and previous example) and eFIP (once logged in) offer this functionality.

PubTator

Concept View  Mention View [Add bio-relation annotation to the table below.](#)

Entity type	Entity mention	Concept ID	Nomenclature	Delete
Disease	Breast Cancer breast cancer	D001943	MESH	Delete
Gene	CIIs	1154	NCBI Gene	Delete
Gene	ESR1	2099	NCBI Gene	Delete
Species	Human women Women	9606	NCBI Taxonomy	Delete

Save Annotation Results    Save & Export Annotation Results

Add new annotation

Delete annotation

eFIP

Evidence Information [in CSV format](#)

Phospho-protein	Phospho-site	Interactant	Impact	Confidence	Sentence	Acceptance
Bad		14-3-3tau	resulting in - assoc	***	2	Yes No
Not yet evaluated *		Bcl-2 and Bcl-x(L)	resulting in - disoc	*	2	Yes No
Not yet evaluated *		Bcl-2	reduced - interacts	***	6	Yes No
Not yet evaluated *		14-3-3tau	increased - associ	***	6	Yes No
Not yet evaluated *						

Additional evidence provided by the bio-curator

Phospho-protein	Phospho-site	Interactant	Impact	Sentence

Provide additional evidence

Sentence number:  Phospho-protein:  Phospho-site:   
Interactant:  Impact:

Edit annotation

Add new annotation

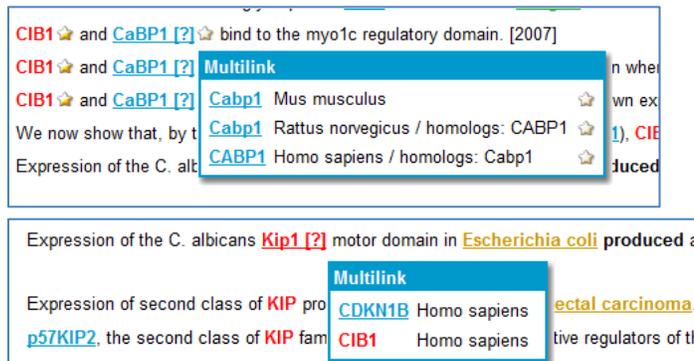
- d. Use standard input and output formats; if possible support more than one format type:  
*Input:* For biomedical literature: document ID (e.g. PubMed IDs, or PubMed central IDs) and/or document text (e.g. text, html, pdf or XML). For other types of input use adopted standards when possible, which should be consulted with user community.

*Output:* Export the results at least in tab-delimited and XML formats. We also encourage teams to adopt the BioC format (XML-based) described in the interoperability Track I (<http://www.biocreative.org/tasks/biocreative-iv/track-1-interoperability/>)

**Other strongly desired functionalities:**

- a. Full-text processing. Minimally to be able to process PubMed Central open access articles. This was one of the top selected features requested by UAG. However, we know that some curation activities are abstract-based, and whether to enable full text processing will be left to your discretion.
- b. Interactive disambiguation of domain entities. E.g. in iHOP (see example <http://www.ihop-net.org/UniPub/iHOP/gs/95608.html?IN=1>) a question mark is added next to the entity for disambiguation of both names and species to select correct one.

Example:



- c. Ability to filter/sort the results according to different criteria; rank results based on what is more relevant to the user. E.g., when searching for a gene in GeneView (<http://bc3.informatik.hu-berlin.de>) you can sort results by relevancy, mentions of SNPs, or date of publication. In addition, you can filter and display only those that contain certain information (such as drugs, PPIs, histone modification, etc).

The screenshot shows the GeneView WBI interface. At the top, there is a search bar with 'ENTREZ.25' and a search button. Below the search bar are various filters: 'Sort order' (set to 'desc'), 'Date of publication', 'Page size' (set to '20'), and 'Only results with genes from list' (set to 'colorectal cancer'). There are also checkboxes for 'Only results with' and 'Fulltext', 'Gene', 'SNPs', 'Chemicals', 'Stages', 'Histone mod.', and 'PPVs'. A box labeled 'Filtering/sorting options' highlights these filter controls.

Below the filters is a table of search results. The table has columns for 'Options', 'Entry', 'Date', 'Genes', and 'SNPs'. The results are numbered 1 through 9, each with a title, authors, journal information, and publication date.

Options	Entry	Date	Genes	SNPs
<a href="#">View Fulltext HTML</a> <a href="#">View Abstract</a>	1 Evaluating Translocation Gene Fusions by SNP Array Data <i>Liu, Hong et al.</i> Cancer Informatics (PMID: 2259228)	2012/12/21	8	1
<a href="#">View Fulltext HTML</a> <a href="#">View Abstract</a>	2 The Role of BCL2 Family of Apoptosis Regulator Proteins in Acute and Chronic Leukemias && <i>Tzif, Flora et al.</i> Advances in Hematology (PMID: 21941503)	2012/09/14	53	1
<a href="#">View Fulltext HTML</a> <a href="#">View Abstract</a>	3 Cytogenetic and Molecular Predictors of Outcome in Acute Lymphocytic Leukemia: Recent Developments <i>Jacobucci, Iaria et al.</i> Current Hematologic Malignancy Reports (PMID: 22528731)	2012/06/20	51	2
<a href="#">View Fulltext HTML</a> <a href="#">View Abstract</a>	4 The Role of Translation Initiation Regulation in Haematopoiesis <i>Grech, Godfrey; von Lindern, Marlene</i> Comparative and Functional Genomics (PMID: 22549283)	2012/05/09	39	1
<a href="#">View Fulltext HTML</a> <a href="#">View Abstract</a>	5 Regulation of Neuronal Cell Death by c-Abi-Hippo/MST2 Signaling Pathway <i>Liu, Weizhe et al.</i> PLoS ONE (PMID: 22590507)	2012/05/09	21	2
<a href="#">View Fulltext HTML</a> <a href="#">View Abstract</a>	6 YAP1 Recruits c-Abi to Protect Angiomotin-Like 1 from Nedd4-Mediated Degradation <i>Skouloudaki, Katsiani; Witz, Gerd</i> PLoS ONE (PMID: 22598212)	2012/04/27	27	4
<a href="#">View Fulltext HTML</a> <a href="#">View Abstract</a>	7 Role of STAT3 in Transformation and Drug Resistance in CML <i>Nair, Rajesh R. et al.</i> Frontiers in Oncology (PMID: 22549784)	2012/04/10	49	7
<a href="#">View Fulltext HTML</a> <a href="#">View Abstract</a>	8 Structural and Spectroscopic Analysis of the Kinase Inhibitor Bosutinib and an Isomer of Bosutinib Binding to the Abi Tyrosine Kinase Domain <i>Livinson, Nicholas M.; Soveri, Steven G.</i> PLoS ONE (PMID: 22492650)	2012/04/06	9	5
<a href="#">View Fulltext HTML</a> <a href="#">View Abstract</a>	9 Chronic myeloid leukemia with extreme thrombocytosis <i>Kim, Se Young et al.</i> The Korean Journal of Hematology (PMID: 22479272)	2012/03/28	2	1

**Additional Functionalities:**

- d. On/off for text mining tool, allowing manual annotation in off mode  
e.g. the Domeo Annotation tool includes both ontology matching and manual annotation by selecting the desired term and linking it to the corresponding ontology term (a list of which is suggested by the tool when the term is highlighted).

The screenshot shows the Domeo Annotation tool interface. At the top, there is a URL bar with 'http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3509163/'. Below the URL bar are buttons for 'Text Mining', 'Annotate', 'Save', and 'Open'. A box labeled 'Manual Annotation' highlights the 'Annotate' button.

The main content area displays a text snippet from a PubMed article: 'Vitamin C Induces Apoptosis in Human Colon Cancer Cell Line, HCT-8 Via the Modulation of Calcium Influx in Endoplasmic Reticulum and the Dissociation of Bad from 14-3-3β'. A box labeled 'Tools Annotation' highlights the text. Below the text is an 'Annotation Widget' with tabs for 'Qualifiers', 'Discourse Element', 'Antibody', and 'Postit'. A box labeled 'Manual Annotation' highlights the 'Qualifiers' tab.

The 'Qualifiers' tab shows a list of terms and descriptions:

Vocabulary	Term & Description
<input type="checkbox"/> PProtein Ontology	Bcl2 antagonist of cell death
<input type="checkbox"/> PProtein Ontology	Bcl2 antagonist of cell death isoform 1 phosphorylated 2
<input type="checkbox"/> PProtein Ontology	lysosome-associated membrane glycoprotein 5
<input type="checkbox"/> Chemical entities of biological interest	(S)-2-(4-(2-bromoacetamido)benzyl)DOTA
<input type="checkbox"/> NIFSTD	bad value
<input type="checkbox"/> PProtein Ontology	Bcl2 antagonist of cell death isoform 1

Below the list, there is a text input field with 'mitochondria, and 3) the expression of Bax.' and a 'Keywords' field with 'Vitamin C, Colon cancer, Endoplasmic raticulum, Apoptosis, Bad, 14-3-3β'. A box labeled 'Manual Annotation' highlights the 'Apply' button.

- e. Record time, be able to record time of curation session for each user (need log in as well)
- f. Load curation suggestions or warnings for display during curation

January, 2013

- If curator adds GO molecular function “transcription factor”, systems suggests GO cellular component (CC) “nucleus” and warns if CC “extracellular space” is entered
- g. Upload gene list or ontology term list for focused curation