CURATION GUIDE

InnateDB Project Leader: David Lynn

Submission System Development: Calvin Chan

Curation Team: Carol Chen, Ana Sribnaia, Raymond Lo,

Previous Curation Team Members: Misbah Naseer, Melissa Yau, Giselle Ring, Kathleen Wee

Previous Assistant Curators: Patrick Taylor, Tracee Wee, Tom Yang.

Further help and tips are available on the PI2 wiki
https://www.pathogenomics.ca/wiki/index.php/Curators_Group

Last Modified: July 19, 2011

Table of Contents

TABLE OF CONTENTS **2**

CHAPTER 1: A GUIDE TO THE INNATEDB SUBMISSION SYSTEM **4**

1.1 Logging In **4**

1.2 Searching Interactions **5**

1.3 Adding a New Interaction **7**
1.3.1 Interaction 8
1.3.1.1 Interaction type 8
1.3.1.2 Full name 10
1.3.1.3 Comments 10
1.3.2 Participant 10
1.3.2.1 Molecule Type 10
1.3.2.2 Species (Human & Mouse Only!) 10
1.3.2.3 Molecule 11
1.3.2.4 Biological Role 13
1.3.3 Evidence 14
1.3.3.1 PubMed ID 14
1.3.3.2 Interaction Detection Method 15
1.3.3.6 Host System and Host Organism 16
1.3.3.7 Cell Status 17
1.3.3.8 Cell Line 17
1.3.3.9 Cell Type 17
1.3.3.10 Tissue Type 18
1.3.3.11 Subcellular localization 18
1.3.3.12 Participant Identification Method and Experimental Role 20
1.3.3.13 Accession Number 23
1.3.3.14 Comments 23

1.4 Editing an Interaction **25**
1.4.1 Editing a Curated Interaction 25
1.4.2 Editing a Public Interaction 26

1.5 Deleting an interaction **27**
1.5.1 Deleting a Curated interaction 28
1.5.2 Deleting a Public interaction 28

1.6 Annotating Innate Immune Genes **29**
1.6.1 Adding annotation for a gene 30
1.6.2 Editing/Deleting an annotation 32

CHAPTER 2: CURATION RELATED ISSUES **35**

2.1 Confirming Species **35**

2.2 Recording Subcellular Localization for a Gene **35**

2.3 Using Pathogenomics Wiki Site **36**
2.3.1 Guidelines for submitting interactions 36

2.3.2 Track Curation Progress 37
2.3.3 Record immune genes and their function 38
2.3.4 Curation Tips 39

CHAPTER 3: IMEX CURATIONS BY INNATEDB 39

3.1 The International Molecular Exchange Consortium 39

3.2 Curating for IMEx 39
3.2.1 PMID 40
3.2.2 Experiment(s) 40
3.2.3 Interaction(s) 41
3.2.4 Participants 41

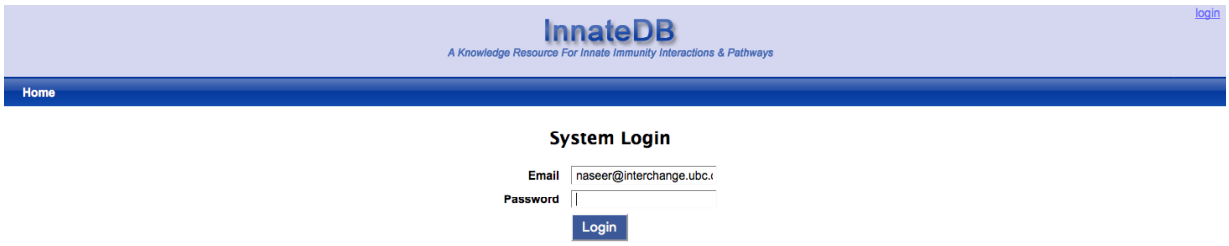
3.3 Useful Links 42
3.3.1 IMEx 42
3.3.2 IntAct 42
3.3.3 Cross References 43

Chapter 1: A Guide To The InnateDB Submission System

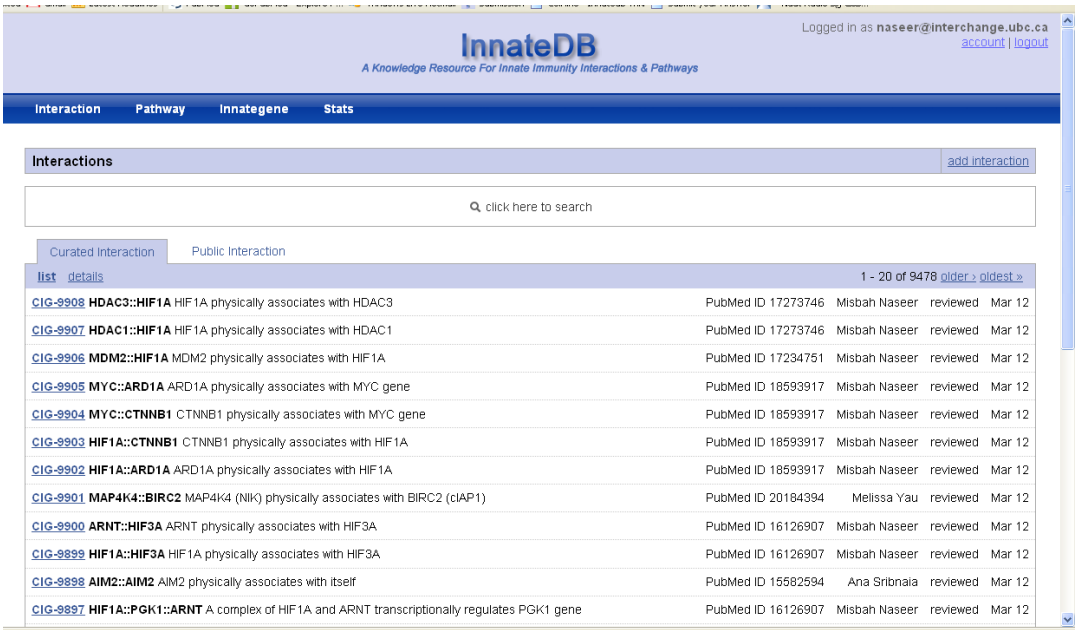
http://www.innatedb.com/dashboard/

1.1 Logging In

A user email address and password are required to access the submission system.



First Page allows users to view & review submitted interactions:

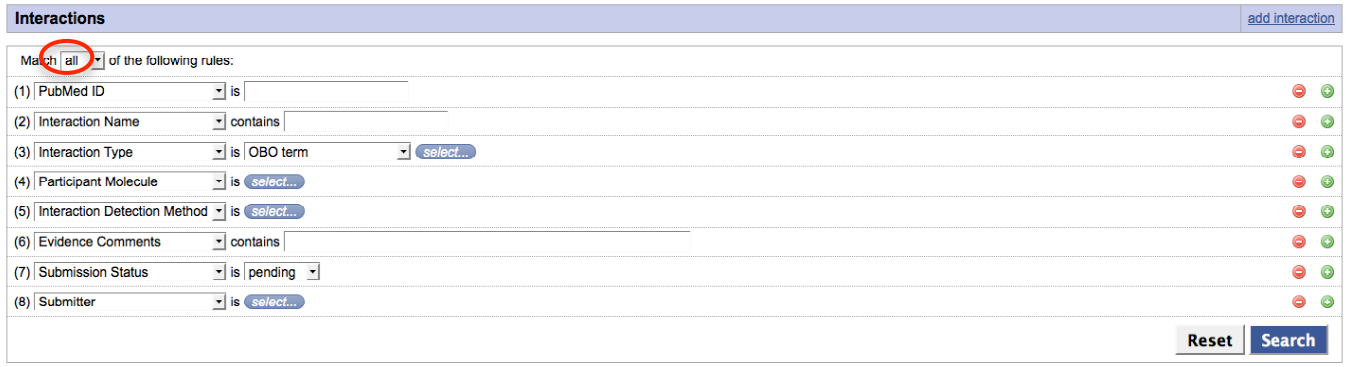


1.2 *Searching Interactions*

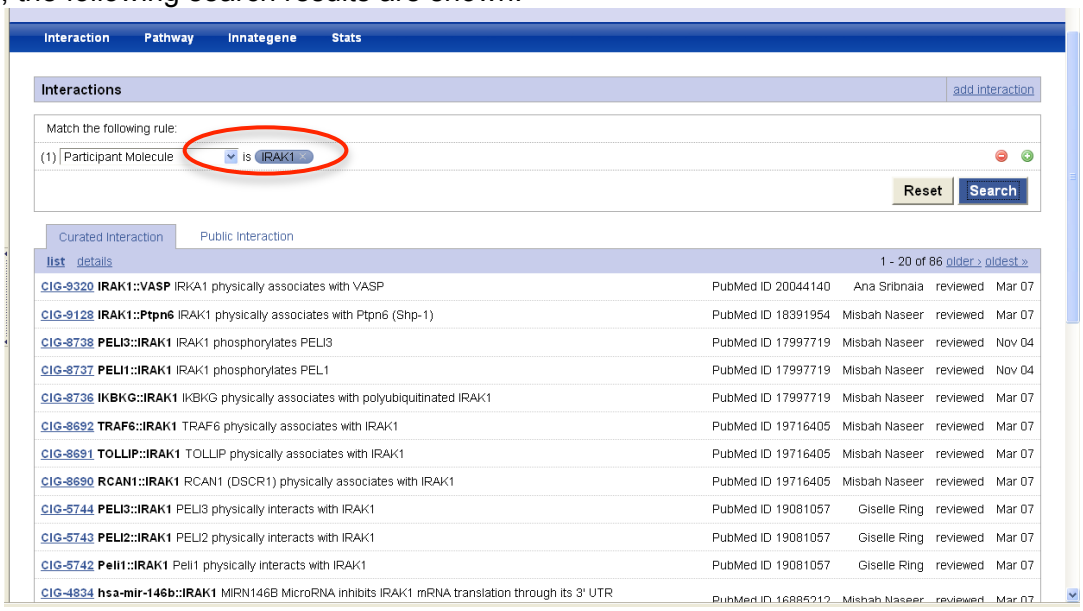
The submission system allows users to search single or multiple criteria at a time. Searchable fields include:

- PubMed ID
- Interaction Name
- Interaction Type
- Participant Molecule
- Interaction Detection Method
- Evidence Comments
- Submission Status
- Submitter

To add multiple search criteria, click the "+" icon. The user can search for interactions matching **all** criteria or **any** of the criteria selected. To take out a search criteria, click on the "-" icon.



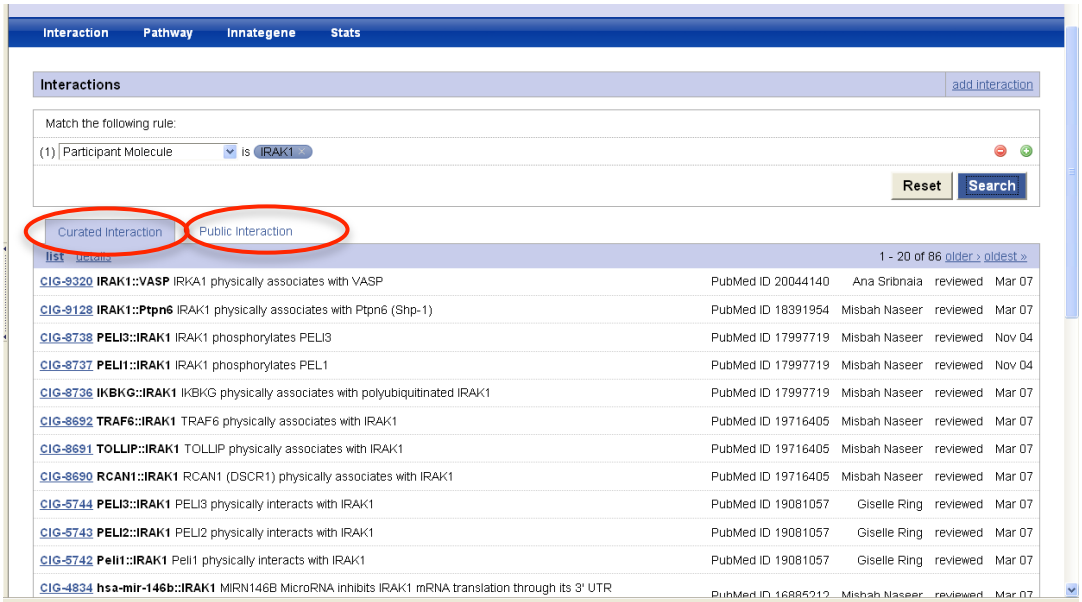
Click the "Search" button after entering the search criteria. For example, if participant molecule= IRAK1, the following search results are shown:



Results are categorized into:

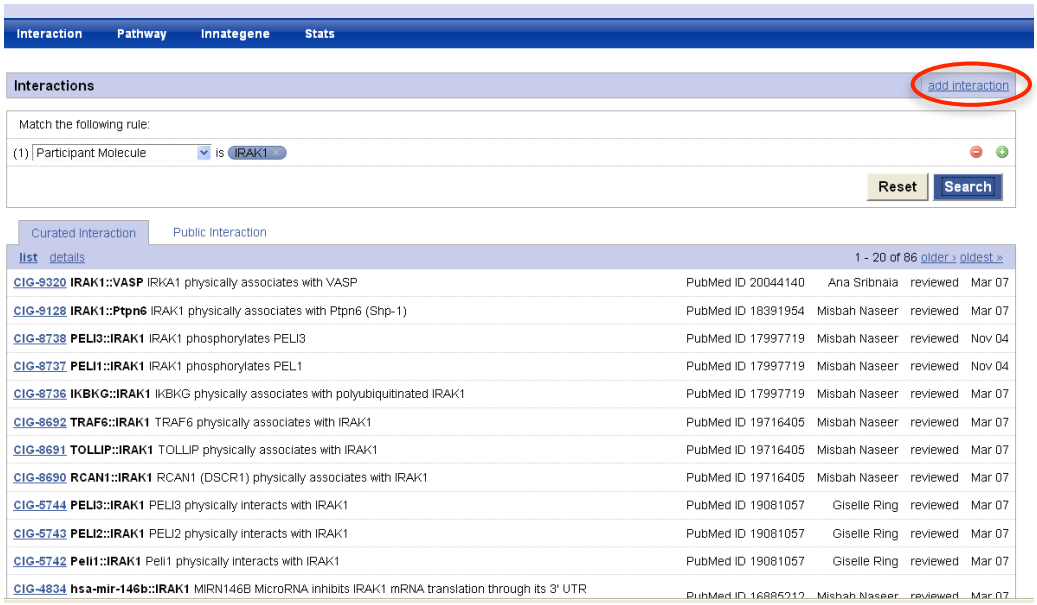
- **“Curated interactions”** which refer to interactions manually curated by the InnateDB team

- **“Public Interactions”** which refer to all interactions displayed on the main public site (www.innatedb.com)



1.3 *Adding a New Interaction*

Click “Add Interaction” in the top right-hand corner of page to begin submitting a new interaction.



The submission page is broken into 3 main sections – Interaction, Participant & Evidence.

THE PSI-MI Controlled Vocabulary (CV) is a hierarchical ontology of terms used to describe molecular interactions.

It can be browsed here http://www.ebi.ac.uk/ontology-lookup/browse.do?ontName=MI

If a required term does not exist in the CV - it can be requested here http://sourceforge.net/tracker/?atid=511101&group_id=65472&func=browse

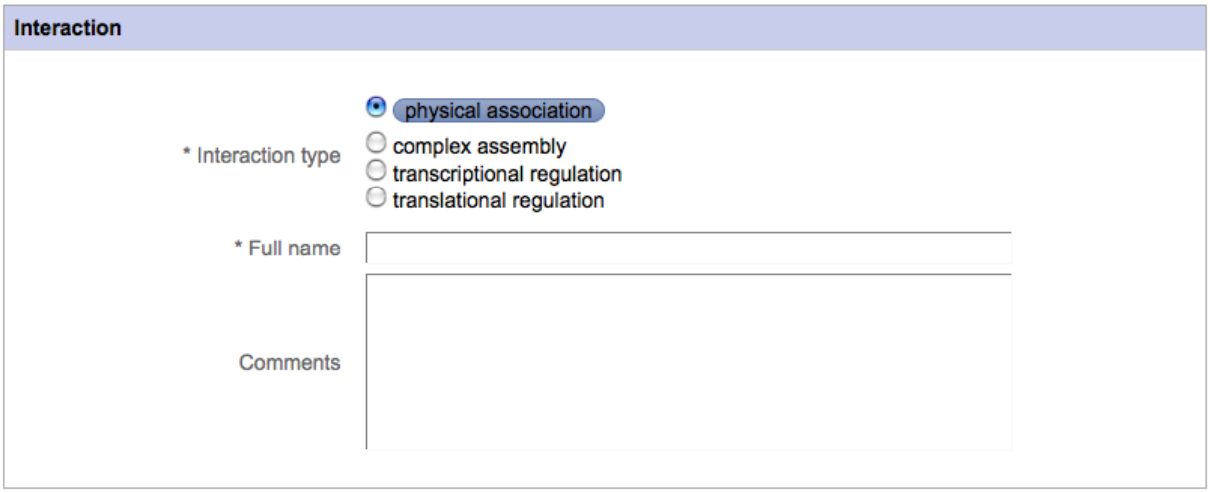
The PSI MI format is a data exchange format for molecular interactions.

Details of the XML Schema etc can be found here http://www.psidev.info/index.php?q=node/60

A PSI 2.5 XML validator can be found here http://www.ebi.ac.uk/intact/validator/start.xhtml

1.3.1 Interaction

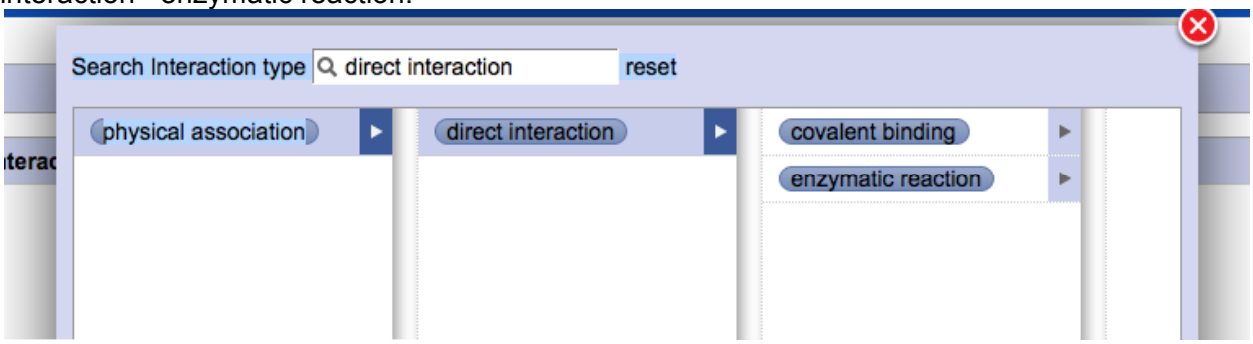
1.3.1.1 Interaction type



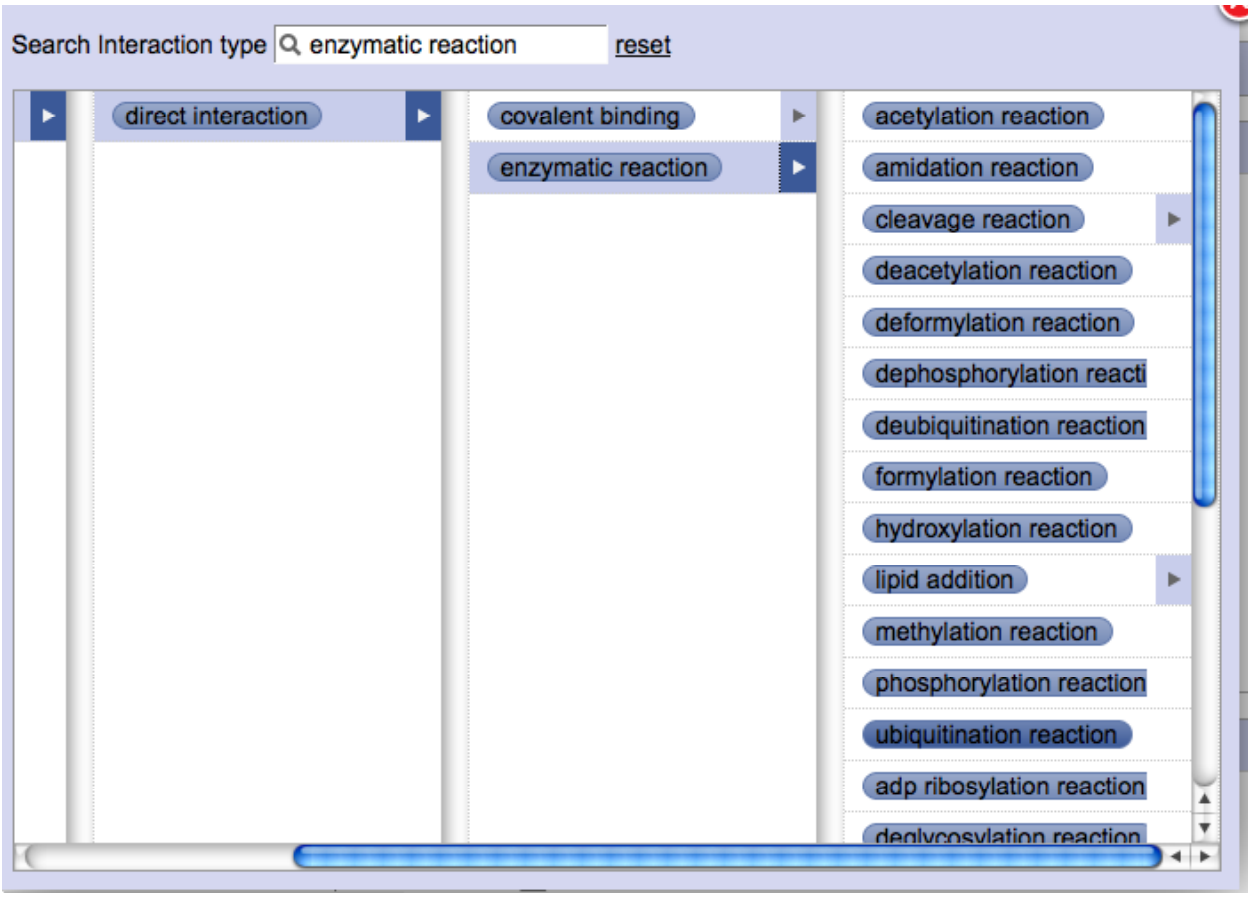
Select the appropriate interaction type from the list.
"**Physical Association**" is the most general term used to describe an interaction and should only be used when no other information is available (This is frequently the case).

"Direct Interaction" is the term used to describe an experiment in which the number of interactors equals 2 **highly purified** molecules and the interaction occurs in vitro, such that no host proteins may interfere.

For **biochemical reactions**, such as *phosphorylation*, *ubiquitination, cleavage* etc, click: "Physical Association". A search window will pop-up. The user may either manually enter the term in the search box or click on the arrow beside physical association and expand to direct interaction >enzymatic reaction.



From the list of enzymatic reactions, select the appropriate term for the interaction.



Commonly used enzymatic reactions are:

- cleavage reaction
- phosphorylation reaction
- dephosphorylation reaction
- ubiquitination reaction

"**Complex Assembly**" This should no longer be used as it is not a valid PSI MI term. Use Physical Association instead.

"**Transcriptional Regulation**" – remember we only describe direct, experimentally validated interactions in InnateDB. Experiments where a gene up/down regulates another gene should not be entered unless a direct interaction has been confirmed i.e. a Transcription Factor is shown to bind the promoter region of gene (via ChIP or EMSA) and up/down regulation is observed in an assay e.g. luciferase assay. When both physical interaction and gene regulation are experimentally proven, select "transcriptional regulation" as interaction type. The physical interaction evidence is described in the evidence section for the interaction with the transcriptional regulation assay is described in the comments. If no up/down regulation observed then the interaction type is just "Physical Association".

"**Translational Regulation**" – rarely used. See "Transcriptional Regulation". Example interaction of RNA x with gene y leads to increased/decreased translation of protein. Examples of this are very rare.

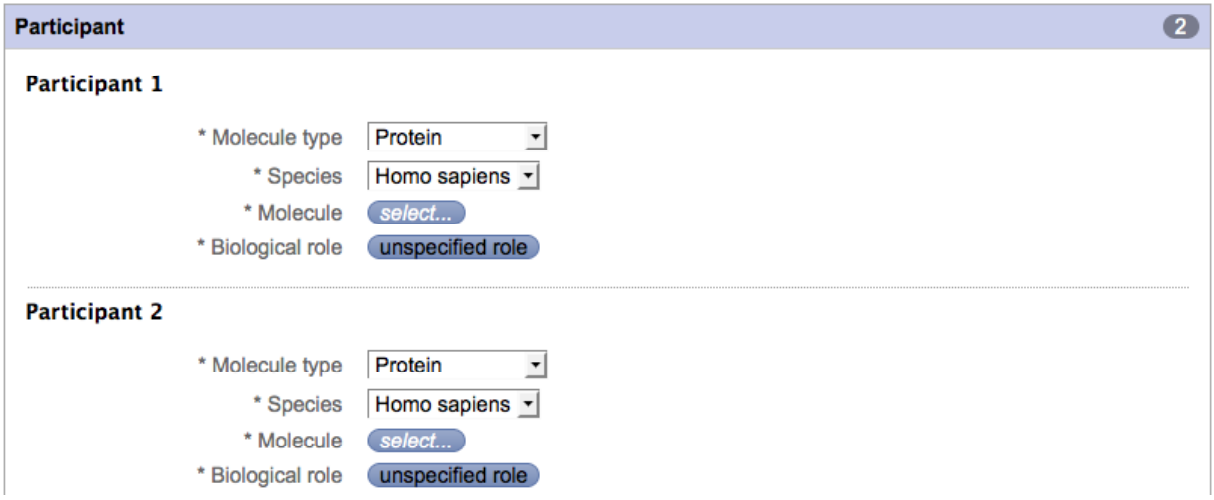
1.3.1.2 Full name

Provide one sentence describing the interaction between the interactors and any specific conditions applying to the interaction. The format of the sentence should agree with the interaction type selected e.g. IRAK4 phosphorylates IRAK1; TLR4 physically associates with LY96; ITCH autoubiquitinates itself in the presence of UBE2L3; IL1 stimulation leads to the complex formation of IRAK1, TRAF6 and MAP3K3. Use HGNC (HUGO Gene Nomenclature Committee) symbols for human participants (UPPER CASE) and Mouse Genome Informatics (MGI) symbols (Title case) for mouse participants.

1.3.1.3 Comments

If information applicable to **all** possible evidences is available, record it in the comments field e.g. a common name for a complex etc. Usually this can be left blank.

1.3.2 Participant



If there are > 2 participants in an interaction click “add new participant”. This can be done as many times as required.

1.3.2.1 Molecule Type

Select the correct molecule type of the participant from the list (Protein, DNA, RNA). The default selection is Protein. Different participants may have different molecule types e.g. a protein (e.g. transcription factor) may interact with a gene.
Note: the ability to describe interactions between a complex and another molecule type will be added later in InnateDB development.

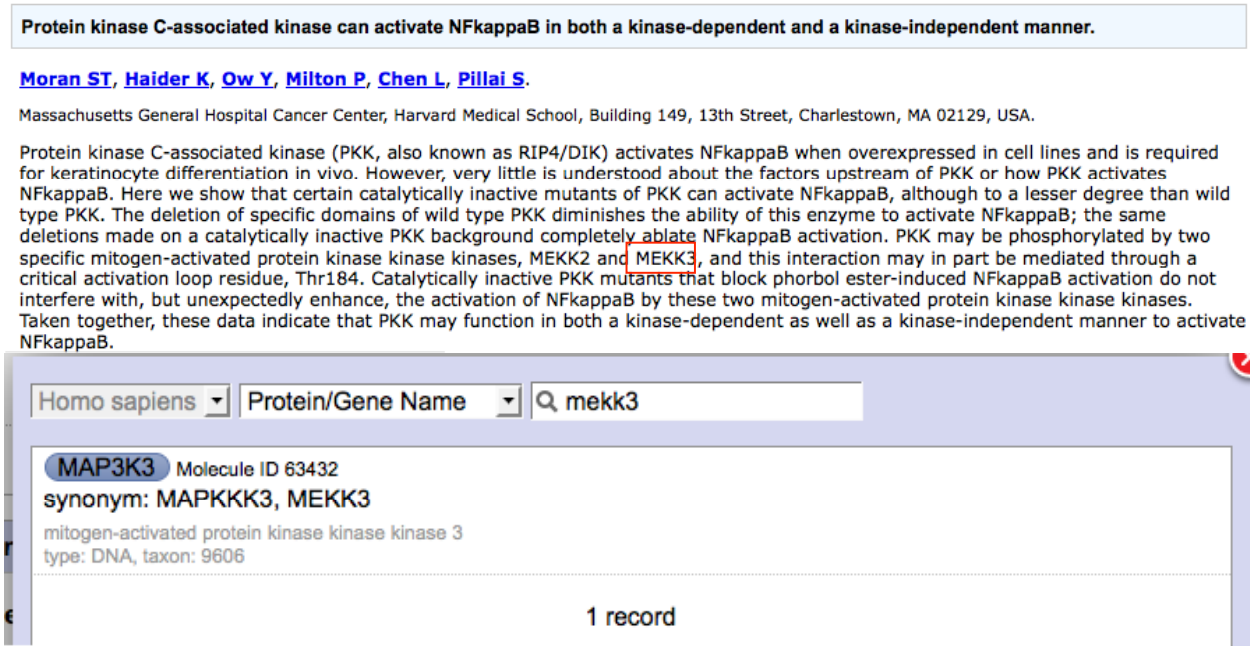
1.3.2.2 Species (Human & Mouse Only!)

Select the species of the participant from the list provided. InnateDB only includes interactions involving human/mouse molecules. The different participants may be of different species e.g. an experiment that shows a human protein interacts with a mouse protein. If no information about

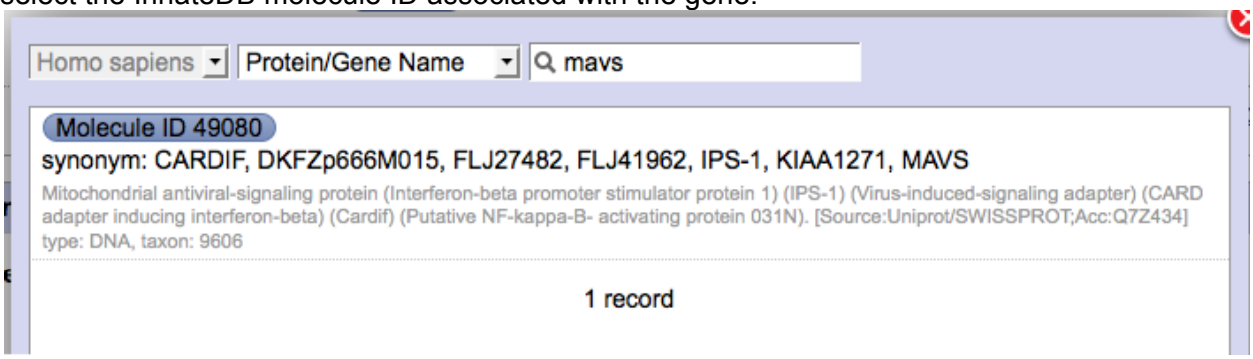
species can be gathered from the paper and references, contact the author of the paper to ensure the species.

1.3.2.3 Molecule

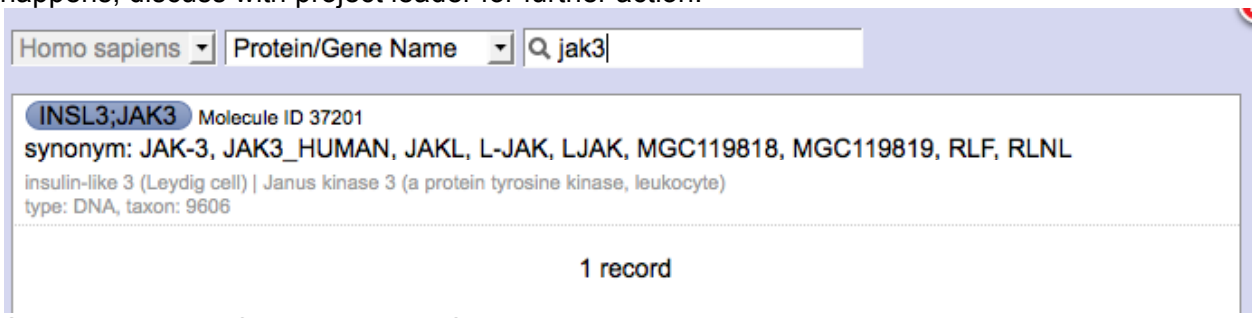
Click “select” button. Enter the symbol for the participant in the text field and hit the ENTER key. Note: Often genes are more commonly known by another name (synonym) so you will need to watch out for this and ensure you are entering the correct name. If a synonym is entered (e.g. MEKK3), the search will provide HGNC symbol for gene(s) with the entered synonym (e.g. MAP3K3).



Special case 1*:* If a HGNC symbol has not been assigned to a gene, search the gene and select the InnateDB molecule ID associated with the gene.

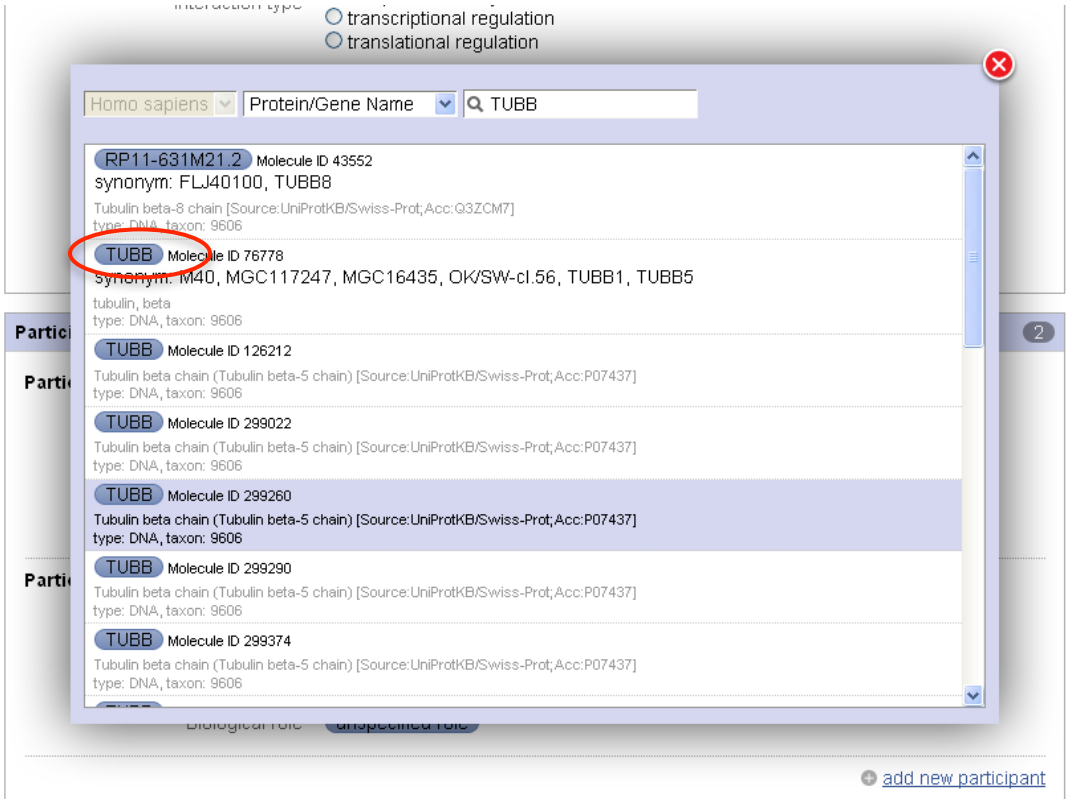


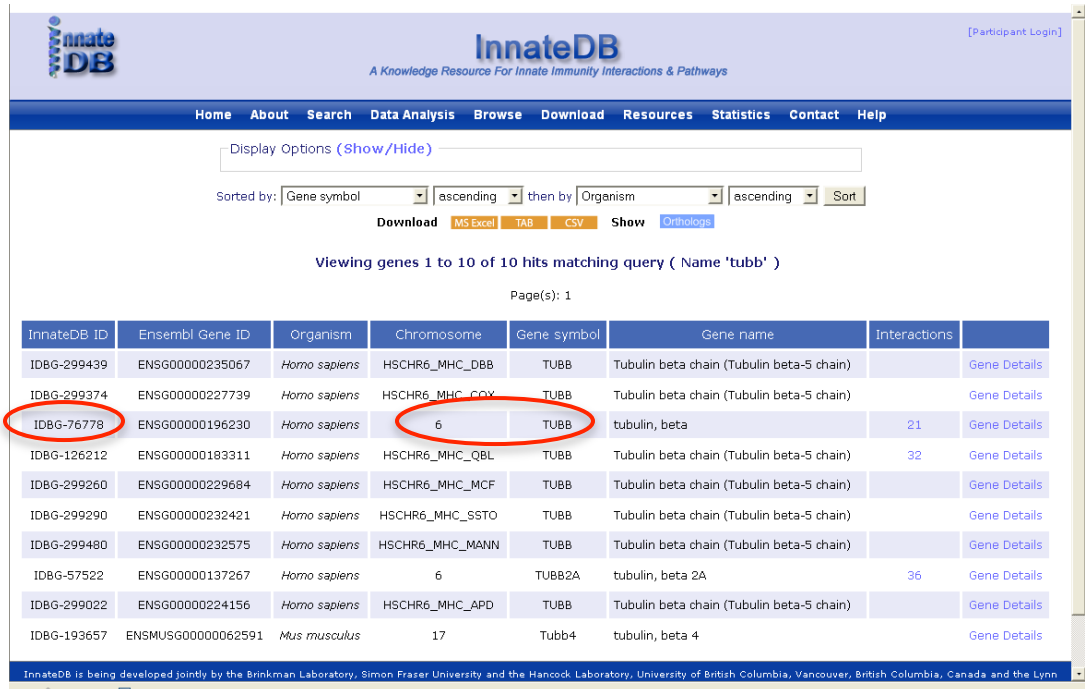
Special case 2: Sometimes 2 HGNC IDs are assigned to the same InnateDB gene. If this happens, discuss with project leader for further action.



Special case 3: Since 2009, ENSEMBL has included nine haplotypic regions, mainly in the MHC region of chromosome 6 in its gene database. Consequently, searches for a gene may generate several results with identical HGNC IDs but different InnateDB molecule IDs.

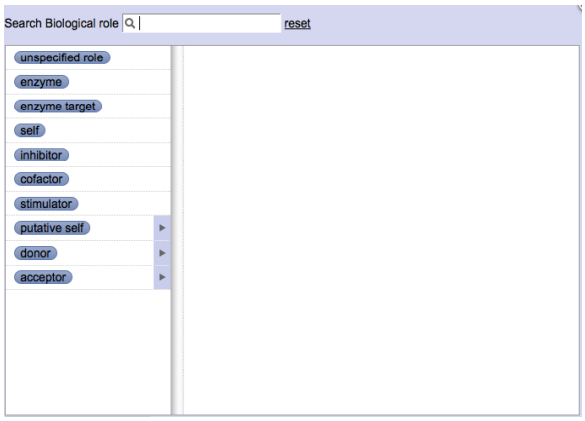
If this happens, search the gene on the main site (www.innatedb.com). Note the InnateDB molecule ID of the gene which is located on Chromosome 6 (not the MHC regions). Use this molecule ID for submitting interactions for the gene.



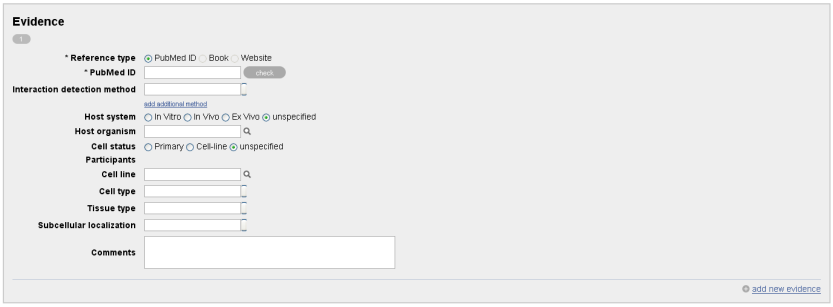


1.3.2.4 Biological Role

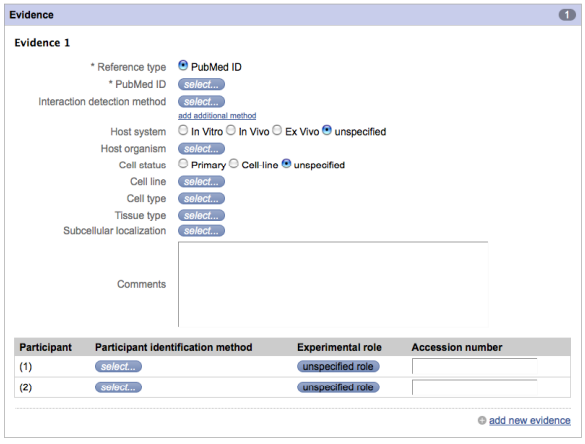
Select an appropriate term for the biological role for the interactor in context of the interaction e.g. *unspecified* for physical interactions, *enzyme and enzyme target* for phosphorylation, ubiquitination, dephosphorylation etc.



1.3.3 Evidence

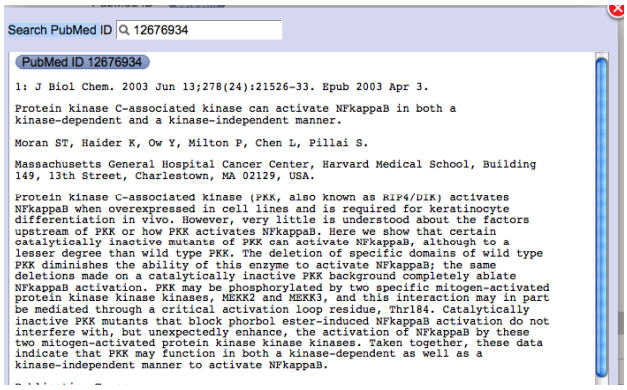


Multiple evidences are possible for an interaction; either from the same PubMed ID where different experiments prove the same interaction or where two or more papers describe evidence of an interaction. Click “+ add new evidence” to add additional evidence either with the same PubMed ID or with a different one.



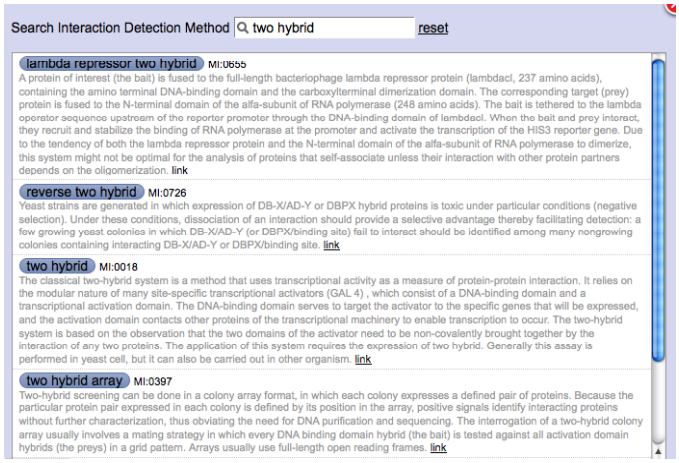
1.3.3.1 PubMed ID

Enter the PubMed ID (PMID) for the publication in which the interaction and its evidence is described. The search box will display the abstract for the entered PMID. Ensure that the correct paper is retrieved.

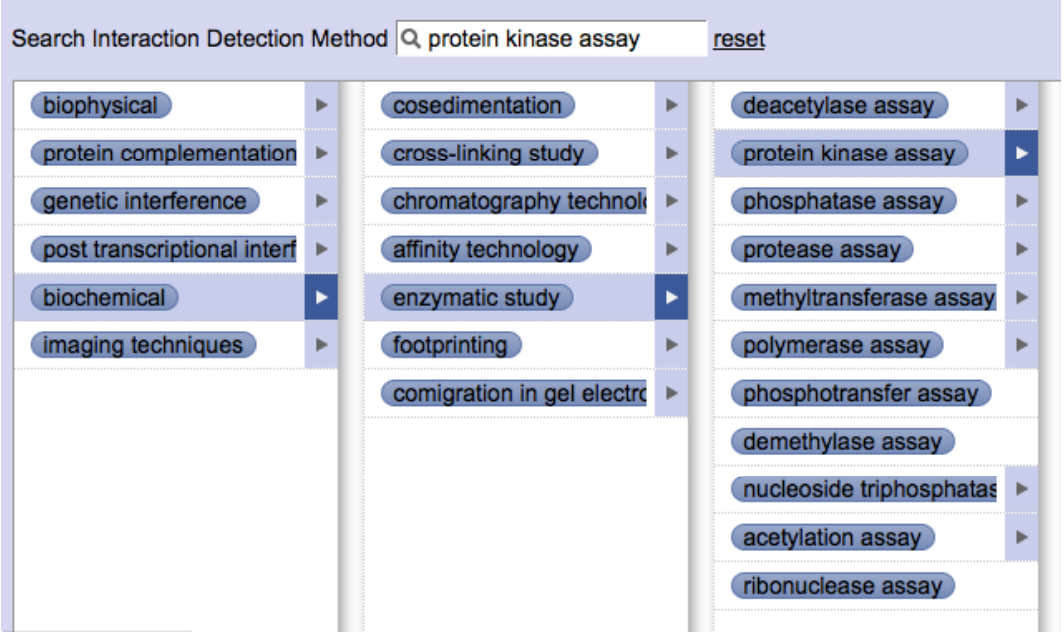


1.3.3.2 Interaction Detection Method

This is the experimental method used to detect the interaction, usually the basis for evidence of an interaction. Interaction detection method terms are controlled PSI MI (http://www.ebi.ac.uk/ontology-lookup/browse.do?ontName=MI) terms for Molecular Interactions. These terms are searchable by typing a term (or part of a term) into the box and hitting the ENTER key. Once search terms are entered, a selection of controlled vocabulary terms appear in a drop down menu, select the interaction term which best fits the experiment performed.



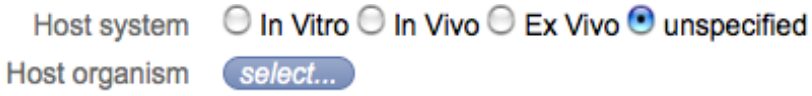
Alternatively, different PSI MI terms listed in the search menu can be expanded until the appropriate term is found.



Following are some PSI MI terms used for common detection methods:

- *coimmunopercipiation/ anti-tag coimmunoprecipitation:* coimmunoprecipitation
- *protease assay:* cleavage reactions
- *enzymatic study:* luciferase assay, ubiquitination/conjugation assay
- *protein kinase assay/ in-gel kinase assay*: phosphorylation reaction
- *phosphatase assay*: dephosphorylation reaction
- *pull down: GST pull down assay*

1.3.3.6 Host System and Host Organism



Host System:

In Vitro – experiments performed in a cell-free system; also used for experiments involving immortalized and commercially sold cell lines.

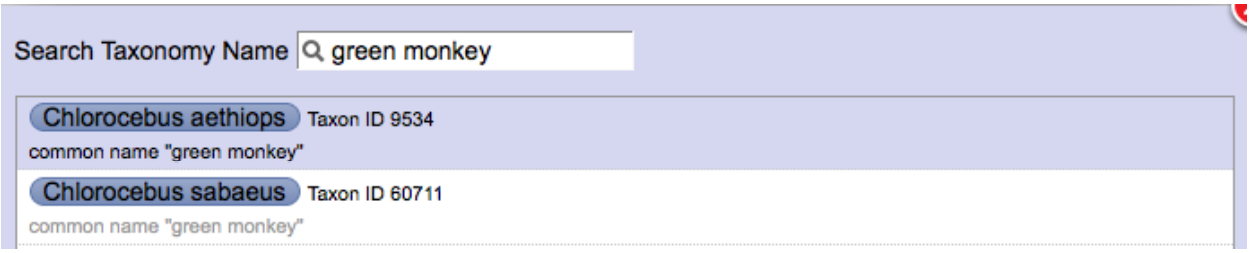
In Vivo – experiments performed in an organism or with cells extracted from an organism which have not been subject to any treatment.

Ex Vivo – experiments performed on cells extracted from a living organism that have been subjected to some form of treatment e.g. TNF, LPS stimulation , also cells derived/cultured from living cells e.g. monocytes cultured from PBMCs.

Unspecified – Experiments performed in a foreign system such as yeast two hybrid experiment.

Host Organism:

The host system is the species where the interaction was shown to take place. This is NOT to be confused with the species of the participant molecules (human and mouse only). For example yeast two-hybrid would be "yeast" and HEK293 cells would be "human". This is a searchable field with controlled vocabulary (NCBI Taxonomy homepage: http://www.ncbi.nlm.nih.gov/Taxonomy/), thus the host system species must be selected from the drop menu of controlled vocabulary which appears in the search box. Any species can be added here e.g. African green monkey, yeast. Sometimes there are several possible options for a common name – make sure you choose the correct species based on the scientific name.



1.3.3.7 Cell Status

Primary - Cells were taken directly from a living organism, which is not immortalized. Cells may be cultured following isolation e.g. monocytes derived from peripheral blood mononuclear cells (PBMCs), bone-marrow derived macrophages

Cell-line - Cells which are grown under controlled conditions e.g. HeLa.

Unspecified - when primary cells or cell line were not used or not indicated in the experiment.

1.3.3.8 Cell Line

Enter the name of the cell line in which the interaction was found to occur. Where possible use the ATCC cell line names found at **http://www.atcc.org/**. Other details such as what cell type, tissue and species a cell line is derived from can also be looked up here. See **https://www.pathogenomics.ca/wiki/index.php/Curators_Group** for cell line details that have already been looked up and for information about standards used to classify a cell line not listed on the page.

Example: Caco 2: cell type: epithelial cell; Tissue type: colorectal adenocarcinoma cell line; *Homo sapiens*

1.3.3.9 Cell Type

Enter the distinct morphological or functional form of cell e.g. macrophage, epithelial etc. Cell line names should not be entered here. The cell type of a cell line should be entered. Cell type terms are OBO controlled terms and can be searched similarly to interaction detection method.

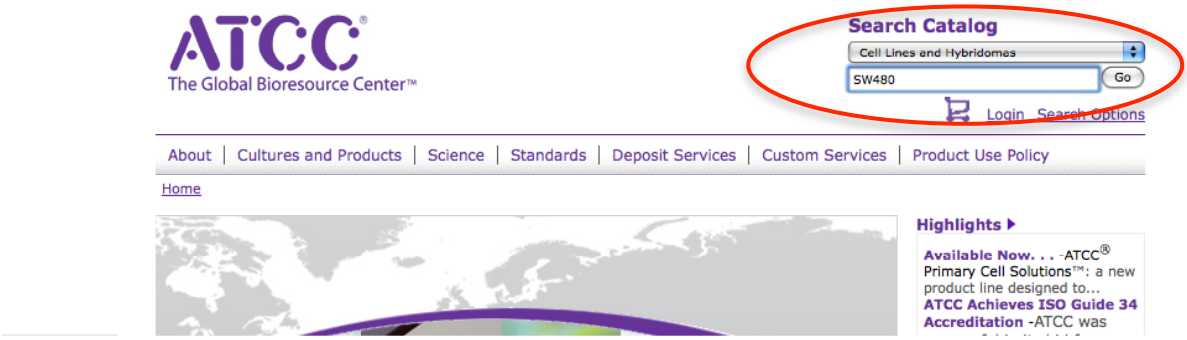
1.3.3.10 Tissue Type

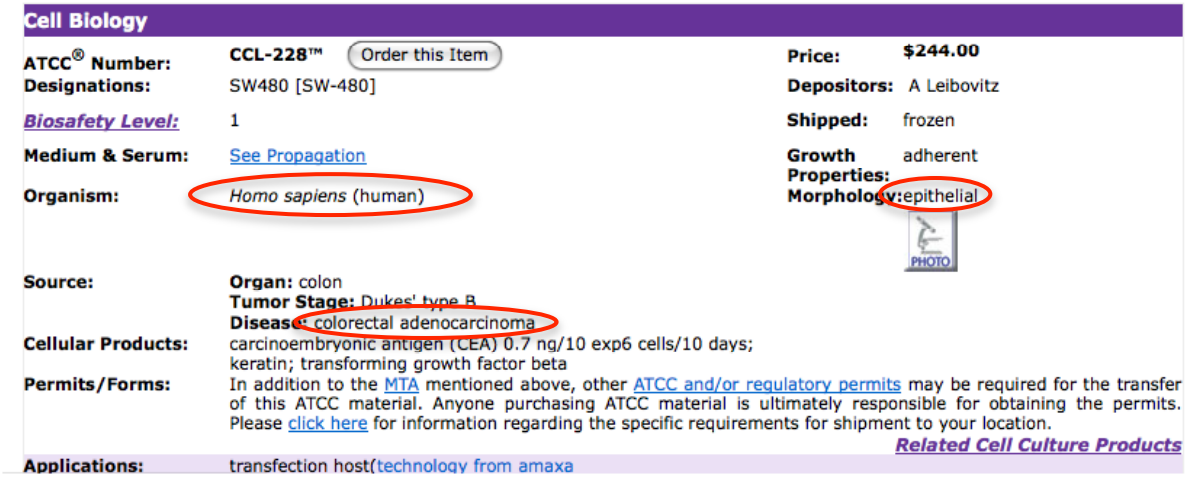
The tissue in which the cells were derived from e.g. lung, heart, brain etc. Tissue terms are OBO controlled terms and can be searched similarly to interaction detection method.

Example

To determine the cell type, tissue type and species of the cell line, search the ATCC cultures to find most of the cell line descriptions.

For example if a paper mentions SW480 cells:



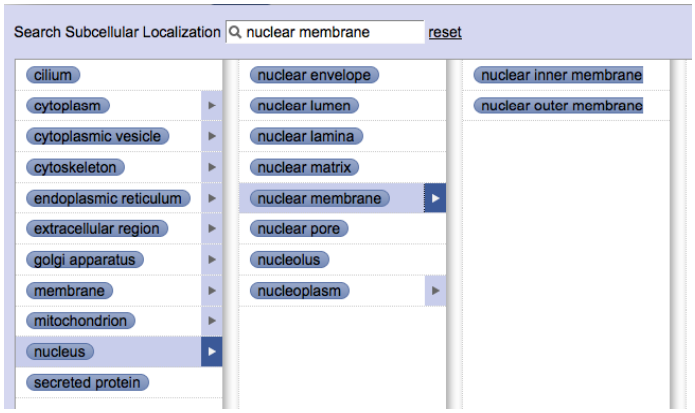


Hence, the following information will be entered:

Cell line: SW480
Cell type: epithelial cell
Tissue type: colorectal adenocarcinoma cell line

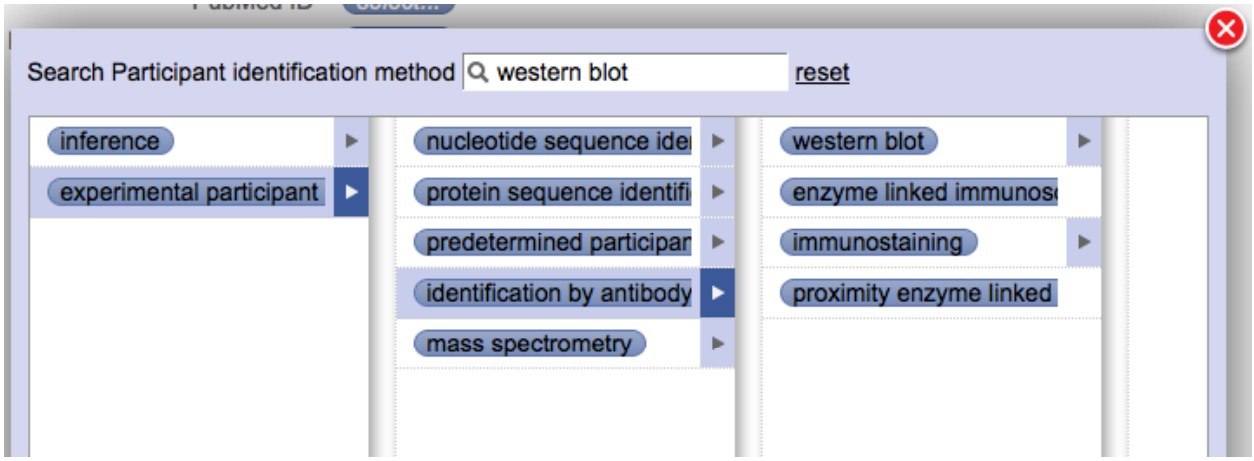
1.3.3.11 Subcellular localization

If specified, this is where the location within the cell in which the **interaction** is observed and NOT the subcellular localization of where a protein is normally located. When inputting subcellular localization, only controlled terms are allowed and thus this is a searchable field. Once search terms are entered, a selection of controlled vocabulary terms appear in the search window, select the term which best fits the subcellular localization. Alternatively, the given list in the search box can be expanded to the appropriate term.

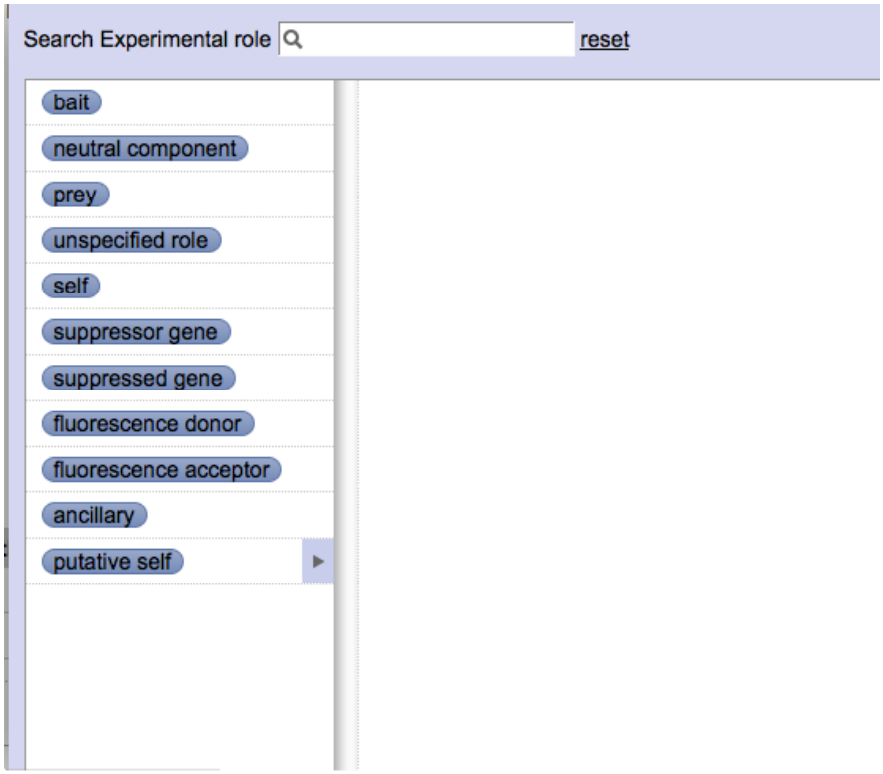

Search Subcellular Localization
nuclear membrane
reset
cilium
cytoplasm
cytoplasmic vesicle
cytoskeleton
endoplasmic reticulum
extracellular region
golgi apparatus
membrane
mitochondrion
nucleus
secreted protein
nuclear envelope
nuclear lumen
nuclear lamina
nuclear matrix
nuclear membrane
nuclear pore
nucleolus
nucleoplasm
nuclear inner membrane
nuclear outer membrane

1.3.3.12 Participant Identification Method and Experimental Role

The method used to identify or determine the participant in the interaction detection experiment is entered in this field. The participant identification method is usually found in the Materials and Methods sections of papers describing interactions and there may more than one used. If more than one method is used, add the one which is more specific in identifying the participant.



Experimental Role explains the role of each interaction participant in the experiment demonstrating the interaction.



bait – The participant was used as a “bait” to find the other participant(s). For example, in a GST pulldown experiment, the GST fusion protein is the bait since it is used to detect other interacting proteins. This is usually the stationary/immobilized participant.

prey – The participant that interacts with the bait; the interactor that is detected because of its interaction with the bait.

ancillary – This is the participant(s) in a complex which links other components of a complex together.

neutral component – An interaction participant that has a neutral role in the interaction; for example in a non-screening yeast two hybrid, neither component is the bait or prey, but instead come together to demonstrate an interaction.

suppressor gene – The gene which suppresses another gene.

suppressed gene – The gene which is suppressed by a suppressor gene.

fluorescence donor – In experiments where fluorescence energy transfer is used (such as in FRET), this is the participant which is the source of fluorescence energy.

fluorescence acceptor – In experiments where fluorescence energy transfer is used (such as in FRET), this is the participant which receives fluorescence energy from the fluorescence donor.

fluorescence acceptor donor pair – Comprised of a fluorescence acceptor and donor.

self – When there is only one participant i.e. an interaction with itself (dimer formation, auto-phosphorylation).

unspecified role – Used when the role of the participant is unknown. This option is rarely used.

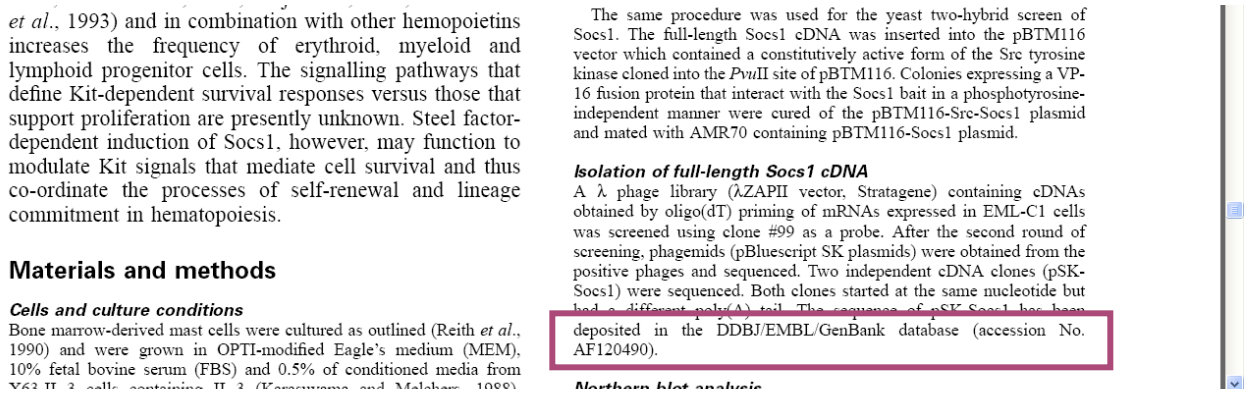
In general, there are many methods which may be used in an article to detect the participant in the interaction, however here are some common ones:

Interaction Detection Method	Expt. Role	Participant Identification Method
Yeast two hybrid	• **Bait**	Plasmid verified by nucleotide sequencing
	• **Prey**	Positive clones verified by nucleotide sequencing
Coimmunoprecipitation	• **Bait**	- Western blot (mono/polyclonal if specified) or - Plasmid was verified by nucleotide sequencing
	• **Prey**	Western blot (mono/polyclonal if specified)
Anti Tag Coimmunoprecipitation	• **Bait**	Western blot (mono/polyclonal if specified) or tag western blot
	• **Prey**	- Anti tag western blot or - Western blot (mono/polyclonal if specified)
Pull down	• **Bait**	Predetermined participant
	• **Prey**	- Western blot (mono/polyclonal if specified) or - Autoradiography if used in vitro translated 35S-labelled protein
Chromatin Immunopercipitation Assays	• **Neutral Component**	- Primer specific PCR (gene)
	• **Neutral Component**	- Identification by antibody (protein)
Electrophoretic Mobility Shift Assay	• **Neutral Component**	- Autoradiography (gene)
	• **Neutral Component**	- Identification by antibody (protein)

1.3.3.13 Accession Number

A unique identifier given to a biological polymer sequence (DNA, protein) when it is submitted to a sequence database. If an accession number for a participant(s) is given in the paper please record it as it can be used to identify the exact variant of a gene or protein used in the experiment. Many different identifiers could be named in articles; the more common identifiers are GenBank, Swiss-Prot, RefSeq and are usually found in material and methods section.

****TIP: It is best to verify the accession number by going to NCBI to make sure that it is the protein you are interested in and it also verifies the species of the protein.**



et al., 1993) and in combination with other hemopoietins increases the frequency of erythroid, myeloid and lymphoid progenitor cells. The signalling pathways that define Kit-dependent survival responses versus those that support proliferation are presently unknown. Steel factor-dependent induction of Socs1, however, may function to modulate Kit signals that mediate cell survival and thus co-ordinate the processes of self-renewal and lineage commitment in hematopoiesis.

Materials and methods

Cells and culture conditions

Bone marrow-derived mast cells were cultured as outlined (Reith *et al*., 1990) and were grown in OPTI-modified Eagle's medium (MEM), 10% fetal bovine serum (FBS) and 0.5% of conditioned media from

The same procedure was used for the yeast two-hybrid screen of Socs1. The full-length Socs1 cDNA was inserted into the pBTM116 vector which contained a constitutively active form of the Src tyrosine kinase cloned into the *Pvu*II site of pBTM116. Colonies expressing a VP-16 fusion protein that interact with the Socs1 bait in a phosphotyrosine-independent manner were cured of the pBTM116-Src-Socs1 plasmid and mated with AMR70 containing pBTM116-Socs1 plasmid.

Isolation of full-length Socs1 cDNA

A λ phage library (λZAPII vector, Stratagene) containing cDNAs obtained by oligo(dT) priming of mRNAs expressed in EML-C1 cells was screened using clone #99 as a probe. After the second round of screening, phagemids (pBluescript SK plasmids) were obtained from the positive phages and sequenced. Two independent cDNA clones (pSK-Socs1) were sequenced. Both clones started at the same nucleotide but

deposited in the DDBJ/EMBL/GenBank database (accession No. AF120490).

Northern blot analysis

1.3.3.14 Comments

Any additional comments or clarification about the experiment can be entered here. Special conditions or treatments in the experiment must be specified. For common additional information, the format for entry into comments is:

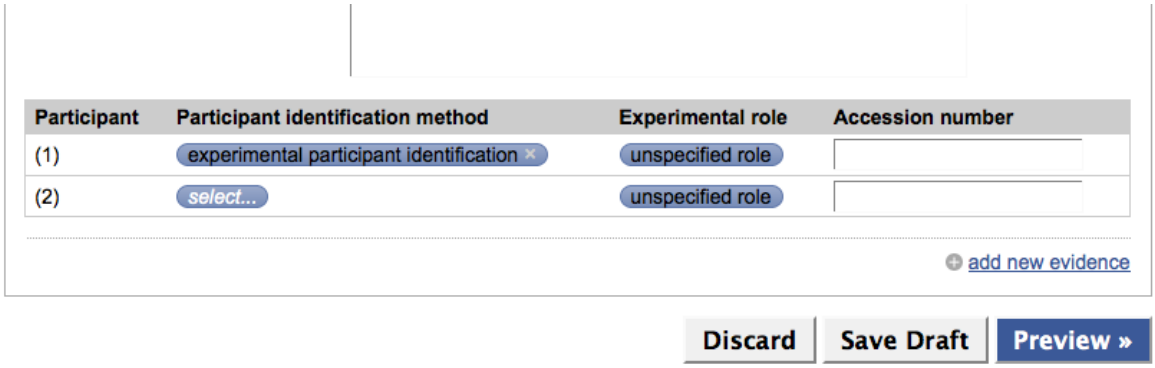
Tags: _____; Treatment:_____; interacting domain:_____; [any other information.]

Example Comments:

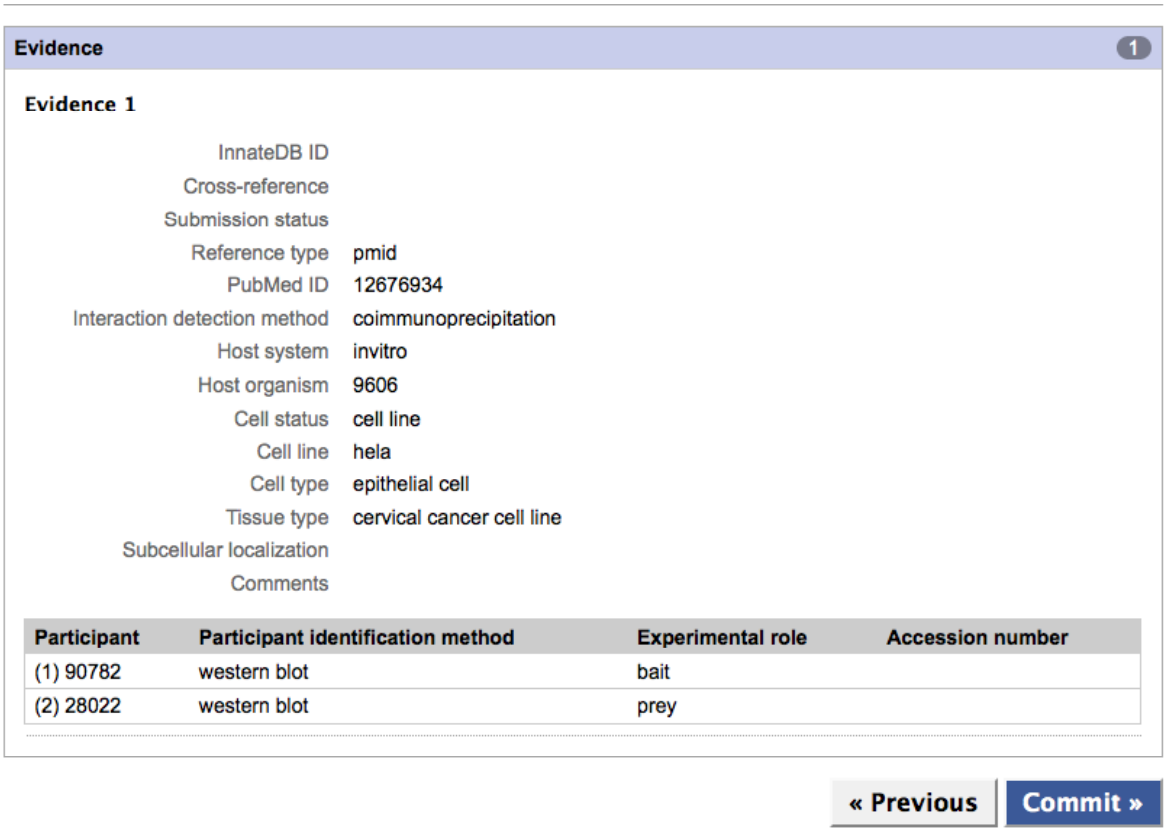
- The interaction was only present in absence of serum stimulation or very low in serum stimulated cells. Also cells treated with 200 ng/ml EGF for 10 min found to inhibit the interaction.
- Tags: CFLAR (Casper) - Flag, NFKB1 (p105) - HA;
- NFKB1 (p50) is 35S labeled
- Interaction strengthened after IL-1 beta stimulation whereas interaction between IRAK2 and AKT1 weakened after stimulation; When LY294002, inhibitor of Akt1 phosphorylation was added simultaneously with IL-1 beta treatment, interaction with IRAK2 was restored while interaction with IL1R1 weakened.

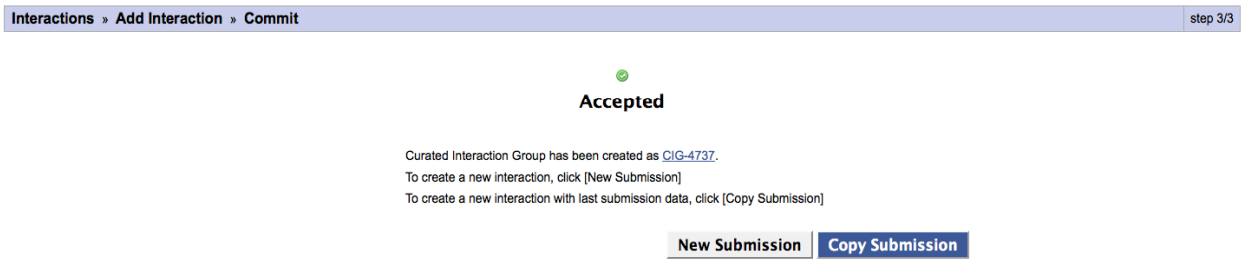
Preview and Submit

After all information has been entered, select preview to view the record and confirm the accuracy prior to submission. If interaction is not ready to be submitted, select 'Save Draft'. All information of the submission page will be saved for later access.



If changes are needed, select "previous" and make the desired changes. If no changes need to be made to the submission, click "commit".





You have now successfully submitted a new interaction. In order to submit another interaction, you can select:

New submission: a blank submission page will open OR
Copy submission: a submission page with data from your previous submission will open. This option is usually used for similar interactions in the same paper.

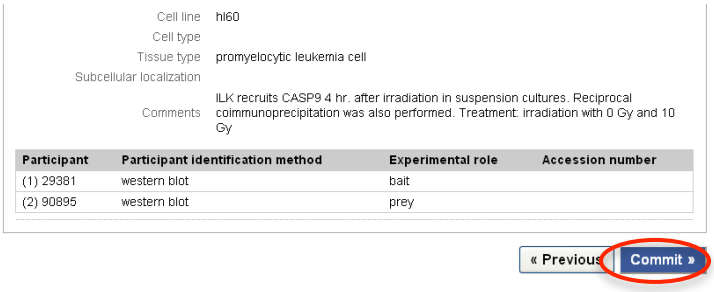
1.4 Editing an Interaction

An interaction in InnateDB may need to be edited due to misspelled words or further details may be needed to be added in the comment section after discussion among the curators. More importantly, an interaction may need to be edited due to incorrect species or missing information in uploaded interactions from other databases (MINT, BIND, etc) which were not manually curated by our own team.

1.4.1 Editing a Curated Interaction

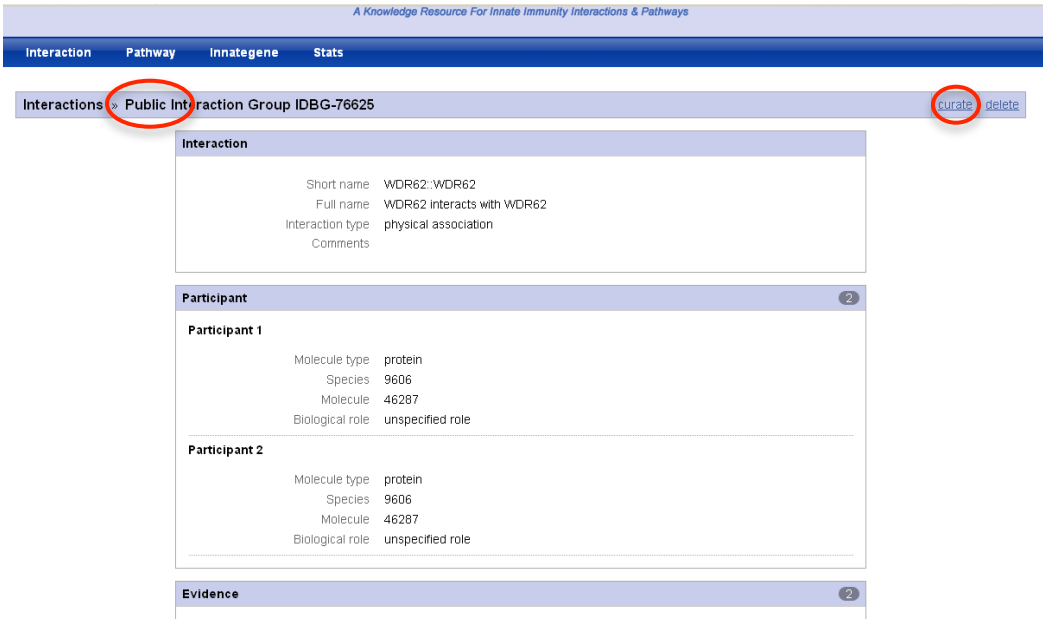
Search the desired interaction using the search criteria (see section 1.2). Click on the interaction from "**Curated Interaction"** Tab (make sure it is not the "Public Interaction" Tab). Click on "Edit" button and the make the appropriate changes. Click on "Preview" to review the interaction.

Click on “Commit” to when changes have been verified.

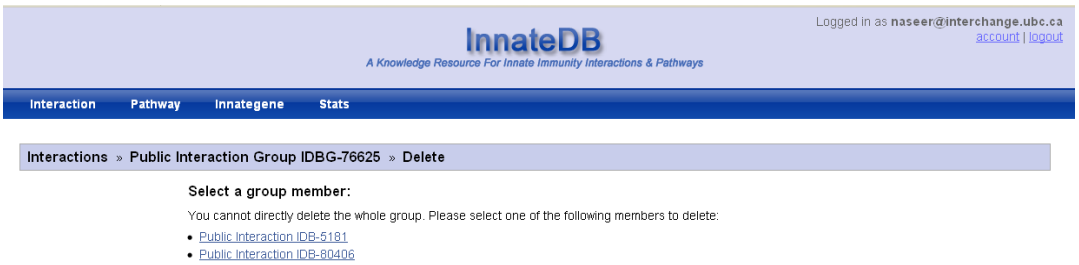


1.4.2 Editing a Public Interaction

Search the desired interaction using the search criteria (see section). Click on the interaction from Public Interaction Tab. Click on “Curate” button and the make the appropriate changes. Click on “Preview” to review the interaction.



If more than one public interaction belongs to the interaction group, you will need to click on the specific interaction you are referring to.



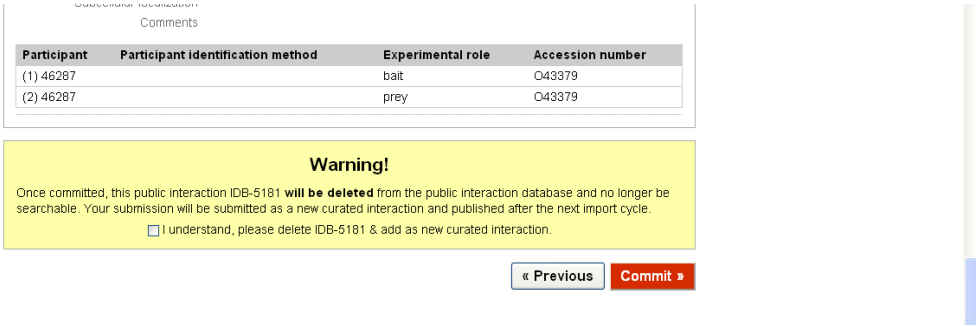
Make the appropriate changes to the interaction and click on “Preview”.

The preview page will display the following warning:

“Once committed, this public interaction IDB-XXXXX **will be deleted** from the public interaction database and no longer be searchable. Your submission will be submitted as a new curated interaction and published after the next import cycle.”

Click on the check box to confirm the deletion of the published interaction.

☐ I understand, please delete IDB-XXXXX & add as new curated interaction.



Click on “Commit” to successfully delete the published interaction and submitted a new curated interaction.

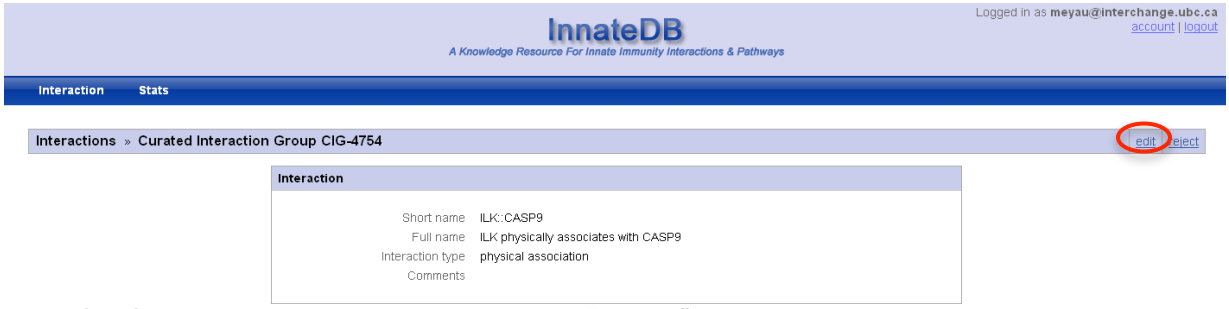
1.5 Deleting an interaction

Deleting an interaction results when published interactions uploaded into InnateDB:

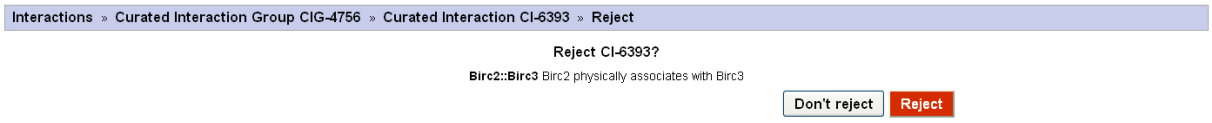
- used an interaction detection method which our curation team has decided is insufficient direct evidence to support an interaction (e.g. interaction detection method using confocal microscopy),
- when the public interaction is not found in the paper when manually curated,
- when the interaction involves other species other than human or mouse.

1.5.1 Deleting a Curated interaction

Search the desired interaction using the search criteria (see 1.2 in this chapter). Click on the interaction from the Curated Interaction Tab. Click on “Reject”. If more than one interaction belongs to the interaction group, you will need to click on the specific interaction you are referring to.

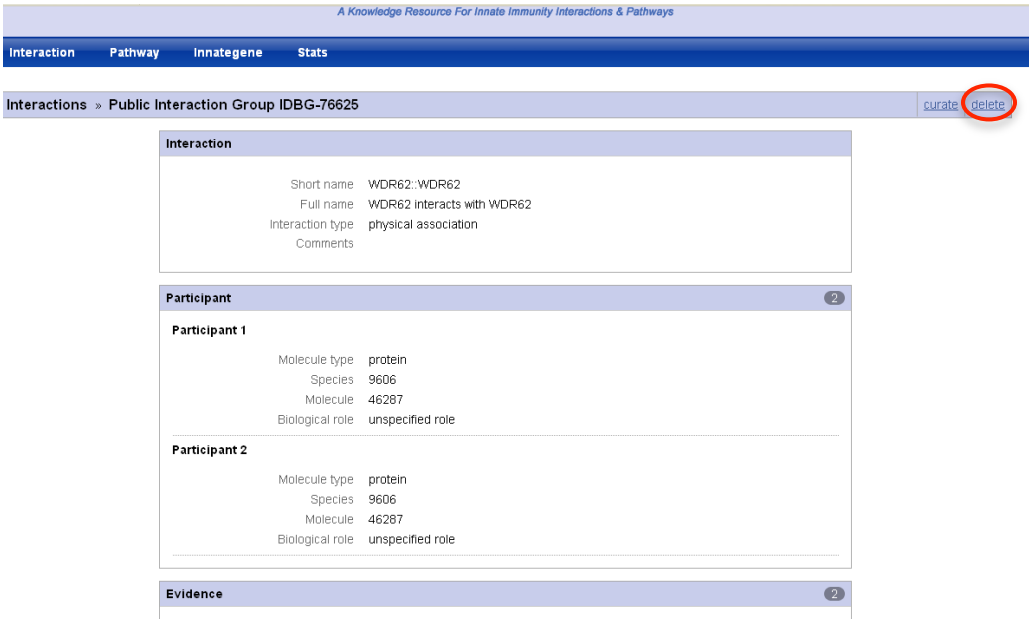


Confirm the rejection by clicking on the “Reject” button.



1.5.2 Deleting a Public interaction

Search the desired interaction using the search criteria (see 1.2 in this chapter). Click on the interaction from the Public Interaction Tab. Click on “Delete”. If more than one interaction belongs to the interaction group, you will need to click on the specific interaction you are referring to.



Confirm the rejection by clicking on the “Delete” button.

1.6 Annotating Innate Immune Genes

Aside from annotating innate immunity interactions and pathways, the InnateDB curation team has also established a project to annotate genes that have a role in the innate immune response. This has been initiated in response to the fact that Gene Ontology annotation of the innate immune response is quite limited in the numbers of genes which have been identified and in response to the fact that many users have been eager to have a defined list of innate immune genes. For innate immune gene annotation, curators employ a new tool in the InnateDB curation system to associate relevant genes with publications which provide evidence for that gene having a role in innate immunity. Along with the link to the relevant publication(s), the curators provide a one-line summary of the role similar to Entrez GeneRIFs. Such genes are also automatically associated with the Gene Ontology term "innate immune response" in InnateDB, which provides a more comprehensive list of these genes for use in the InnateDB Gene Ontology over-representation analysis tool. To date, Nearly 1000 genes have been annotated to some extent (as this is an on-going process). It should be noted that it is not the intention of InnateDB to comprehensively annotate all the roles of a given gene, but rather to provide a brief indication if a gene has a role in innate immunity.

The **Innategene** function on the main page is used to record immune genes and their function in innate immunity as described in specific scientific publications. This information can also be extracted from review articles, however the experimentally defined role of the gene/protein is preferred. This information is displayed on the gene card on the main site, under the section “InnateDB annotation.”

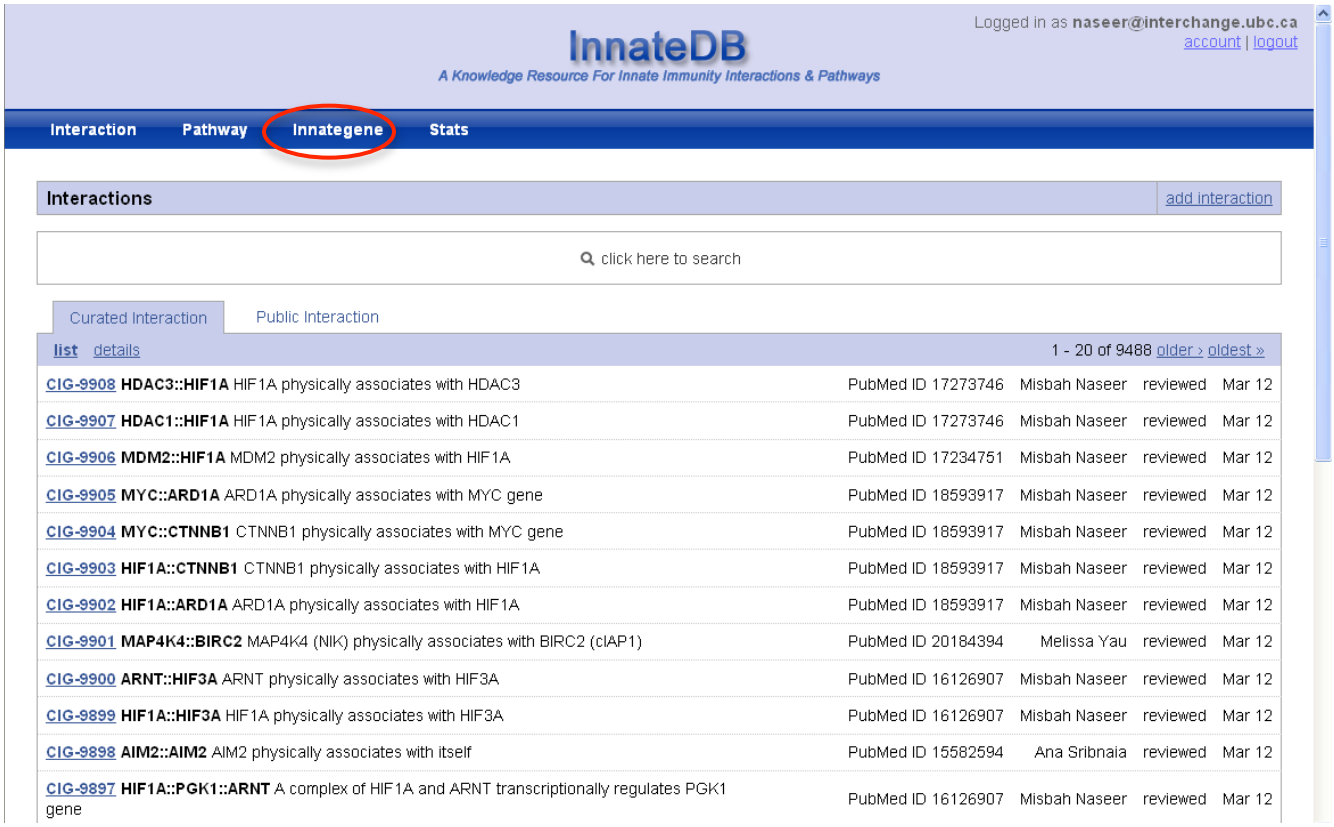
Innategenes are annotated as a part of the curation process (i.e. not restricted to innategene/curation emails) and should be captured whenever possible. The gene annotated should be the primary focus of the article, and its function should be apparent from the abstract body.

Innategene annotations should be one or two sentences, and should begin with the gene name and describe the function only as it relates to innate immunity. For example, classic innate immune receptors, such as the TLRs, may play a role in neural development, but those annotations are irrelevant to InnateDB and should not be entered.

Human and murine orthologs should be annotated together, provided that the genecard of the corresponding genes cross-references each other under the “Orthologs” section. If the gene function is only demonstrated in one species, the ortholog should be tagged with “(Demonstrated in mouse/human)”. Gene nomenclature is standardized to HGNC/MGI for human and mouse, respectively. The annotation text should adhere to Canadian English spelling (e.g. “signalling”, “defence”).

The innategene annotations are compiled in a word document on a weekly basis using the template outlined below, and reviewed internally before submitting to InnateDB.

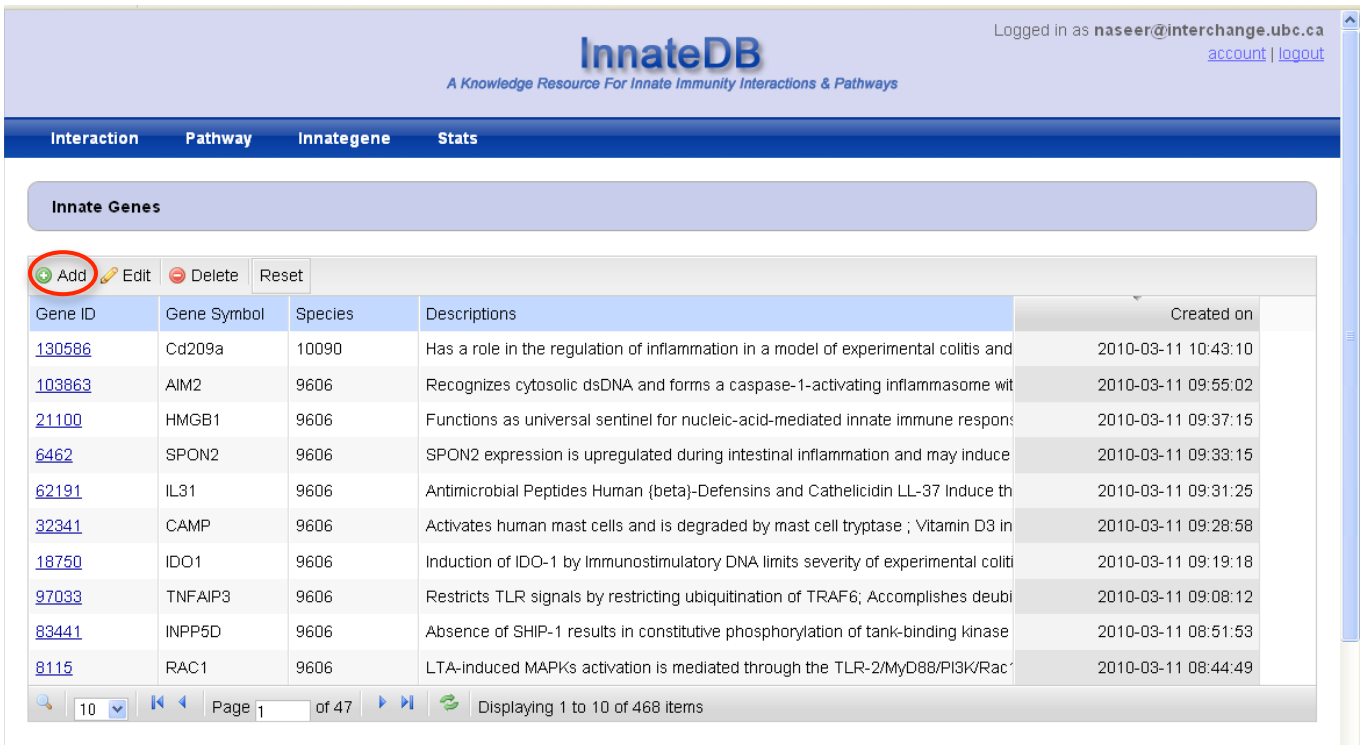
Gene Symbol	Gene (Author)	*Species*	Publication	Description
TLR2		*Homo sapiens*	21698237	TLR2 is required for rapid inflammasome activation in response to infection by cytosolic bacterial pathogens such as *Francisella novicida*. (Demonstrated in mouse)
Tlr2	TLR2	*Mus musculus*	21698237	Tlr2 is required for rapid inflammasome activation in response to infection by cytosolic bacterial pathogens such as *Francisella novicida*.



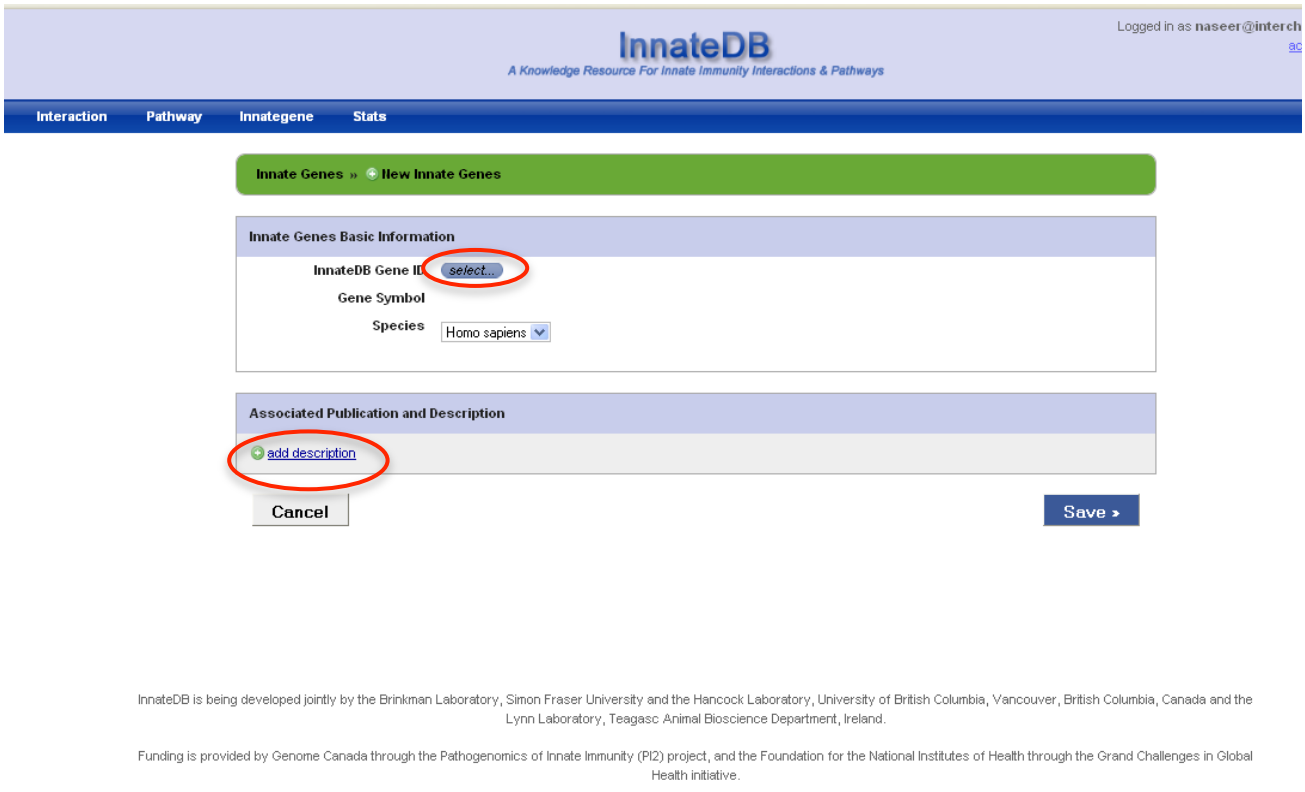
1.6.1 Adding annotation for a gene

Click **Innategene** in the top right-hand corner of page to begin submitting a new annotation.

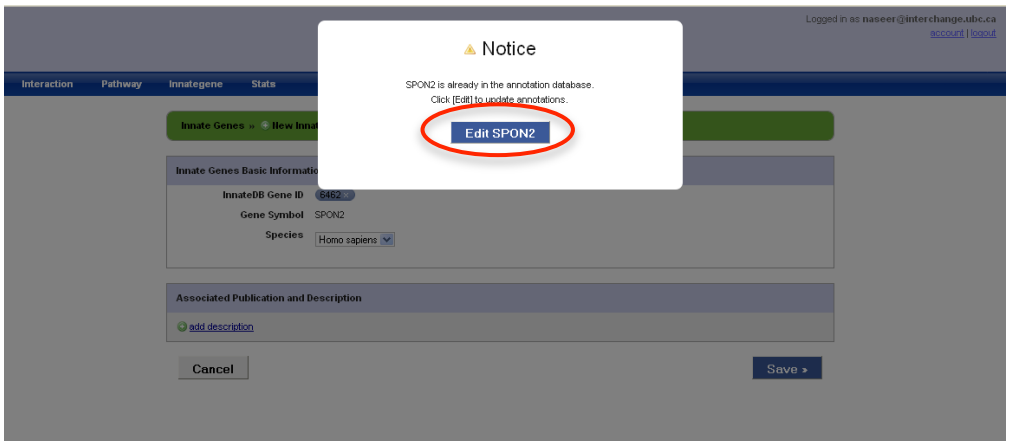
The main Innate Genes page displays the most recently annotated genes. Click **Add** icon on the main page.



On the **New Innate Genes** page, enter the gene symbol in InnateDB Gene ID field. Select the desired gene as per section 1.3.2.3.



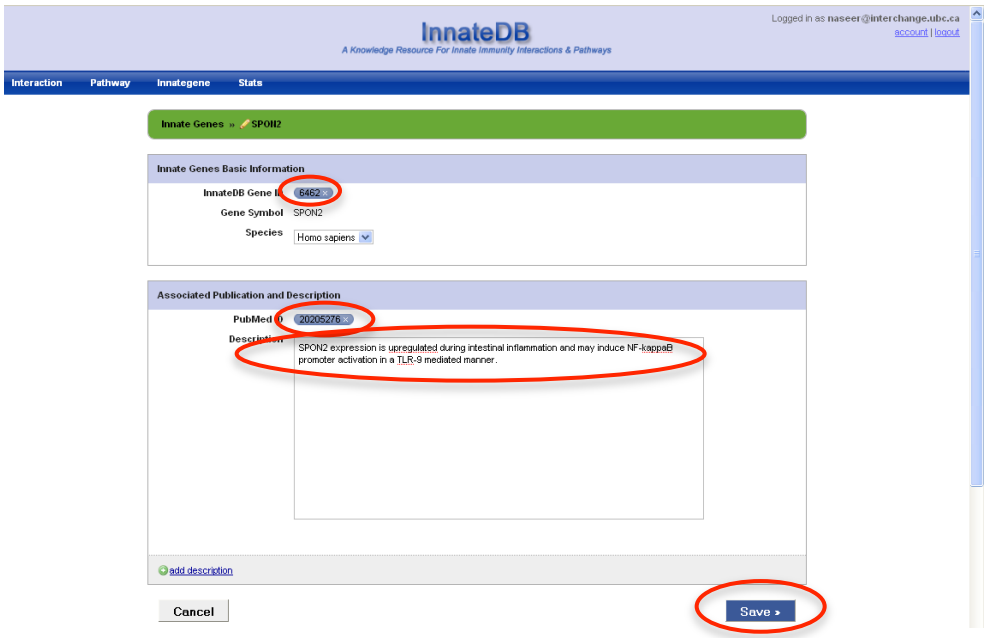
Note: When looking up a gene ID, a pop-up screen may appear notifying the user of an existing annotation for the gene. Click **EDIT [GENE SYMBOL]** button to continue.



Click **add description** to enter a new annotation.

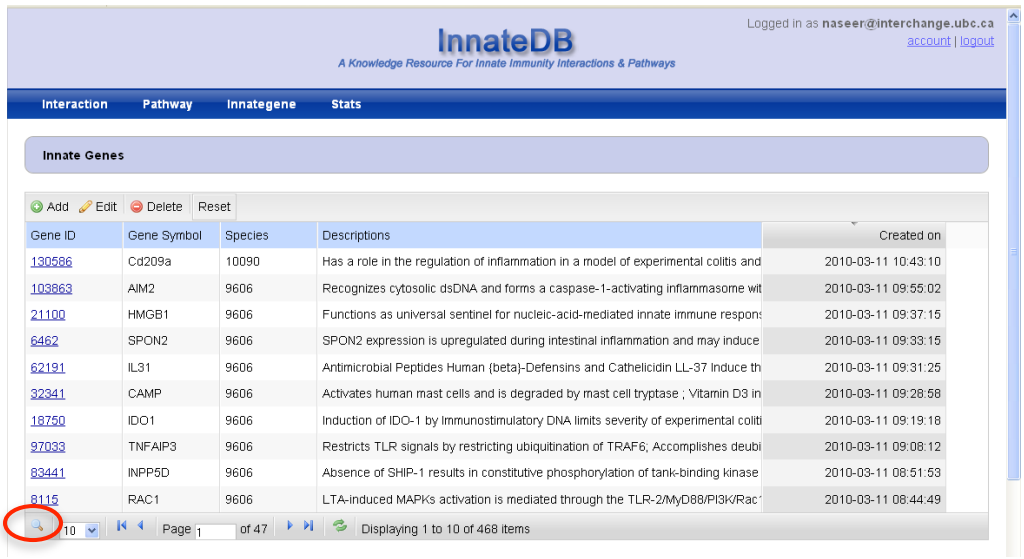
Enter the PMID of the source journal article in the **Pubmed ID** field. In the **Description** field, enter a one-sentence description of the gene in relation to its function in innate immunity. This information can usually be derived from the conclusion statement of the abstract of the source journal article.

After entering the required information, click **Save** button. The information will be updated instantly on the gene card on www.innatedb.com.

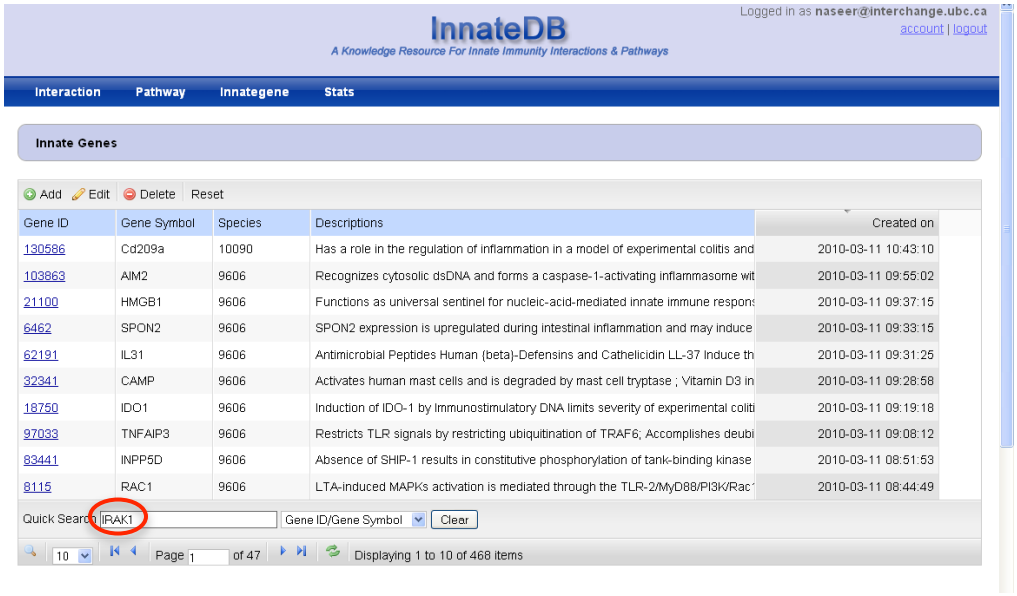


1.6.2 Editing/Deleting an annotation

Click on the magnifying glass icon in the bottom right-hand corner to search for all annotations of the gene entered in InnateDB.

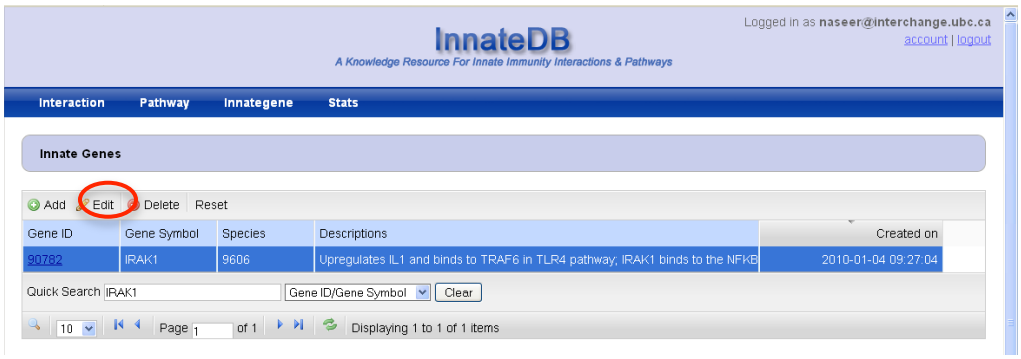


Enter the HGNC symbol for the gene and hit the **ENTER** key.

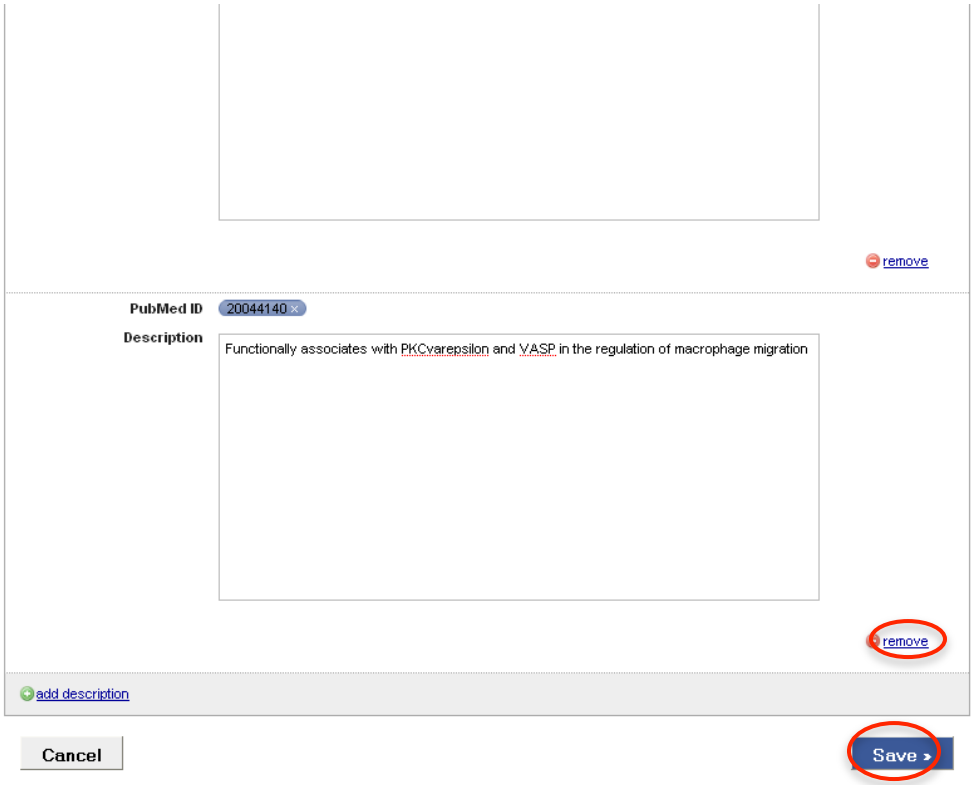


If the gene has been annotated previously, the search results page will show the entry for the desired gene.

Highlight the row displaying the gene by clicking it and click **Edit** icon.



All annotations for the selected gene will be displayed. To delete an annotation, click **remove**. To edit the text of an annotation, make the required changes on the page. Once all changes have been made, click **Save** button at the bottom of the screen.



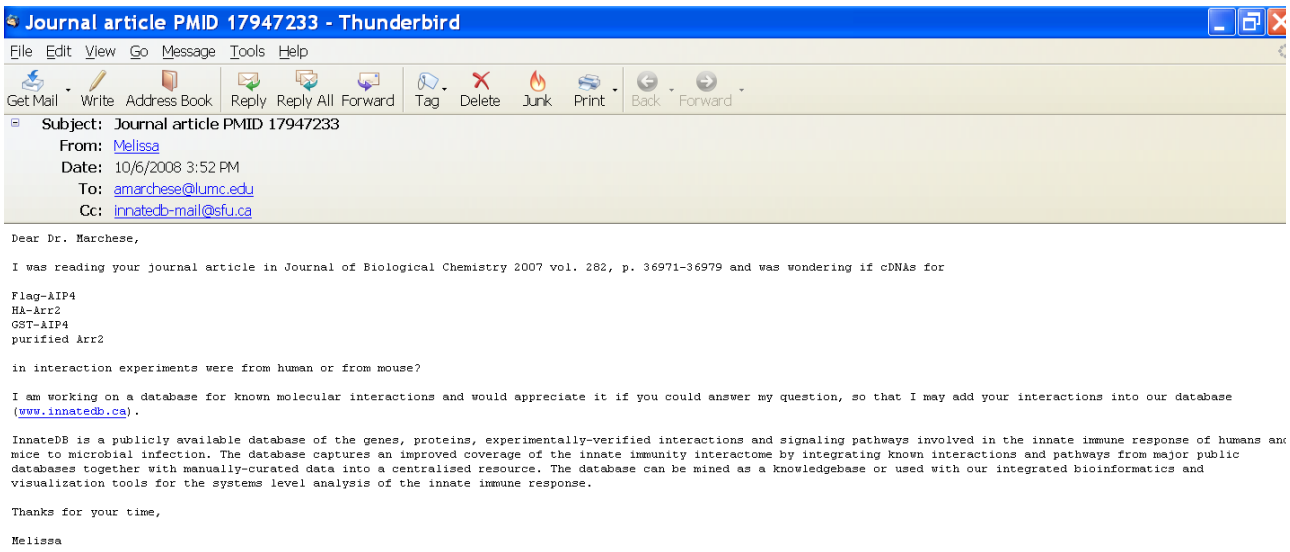
Chapter 2: Curation Related Issues

2.1 Confirming Species

If species are not specified in the scientific article, the following steps can be taken:

1) If there is an article referring to the plasmid in the Material and Methods section, look up the species from this article.
2) Contact the corresponding author in the article for species confirmation. Note: remember to copy the email to innatedb-mail@sfu.ca

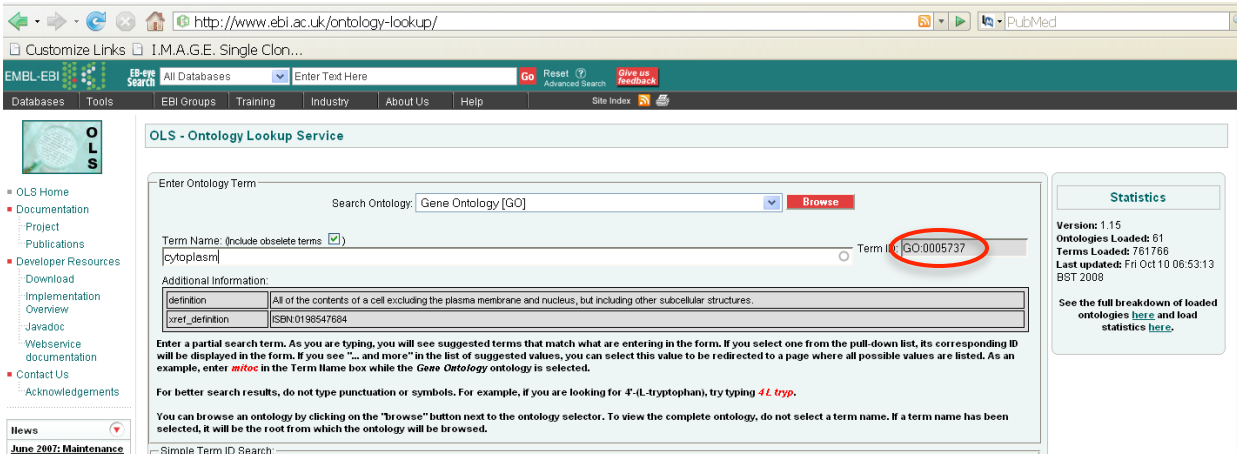
   For Example:



2.2 Recording Subcellular Localization for a Gene

If the Subcellular Localization is specified for a gene in a paper, check the gene card on InnateDB to see if the subcellular localization has been recorded. If not, then record the InnateDB gene ID and the Gene ontology term referring to the subcellular localization in an excel sheet, which will be sent to the InnateDB database developer.

The gene ontology term can be looked up at the following link:
http://www.ebi.ac.uk/ontology-lookup/



2.3 Using Pathogenomics Wiki Site

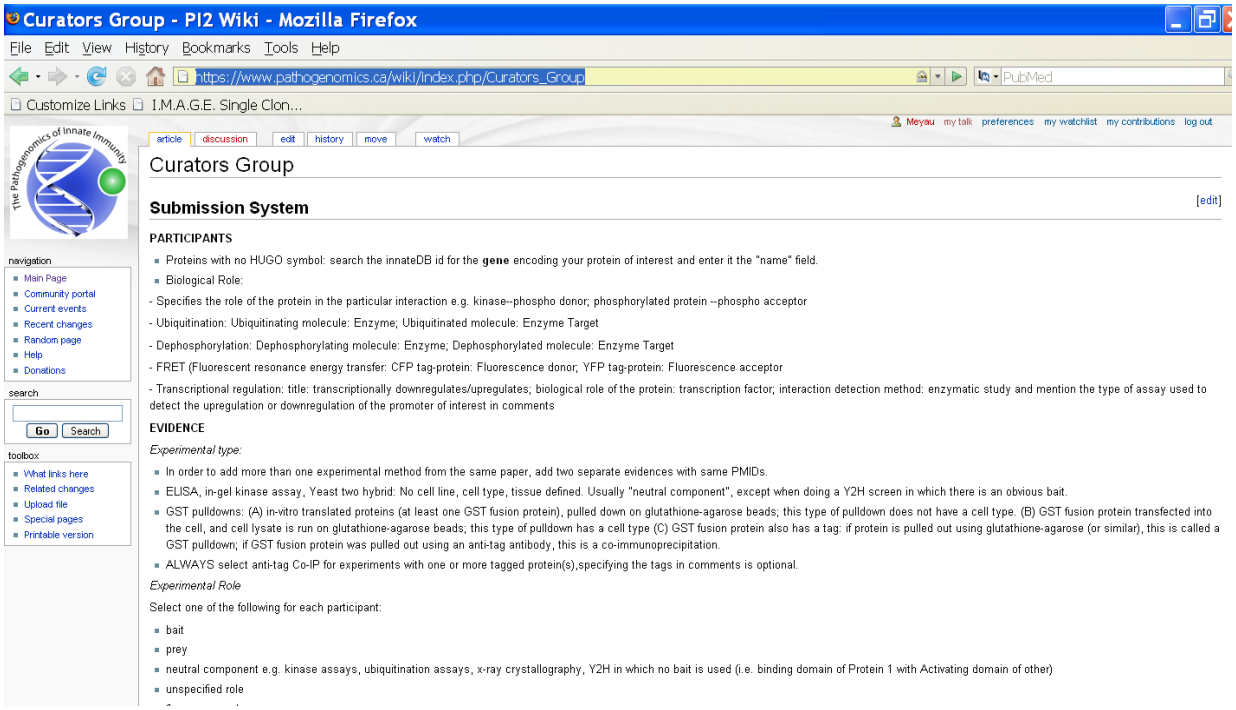
The Pathogenomics Wiki Site can be accessed at **https://www.pathogenomics.ca/wiki/index.php/Main_Page**. This site enables the curators to:

a. Guidelines for submitting interactions
b. Track curation progress (requested and curated genes)
c. Record innate immune genes and their function

2.3.1 Guidelines for submitting interactions

https://www.pathogenomics.ca/wiki/index.php/Curators_Group

To ensure consistency among curators, general rules have been outlined for submitting interactions.

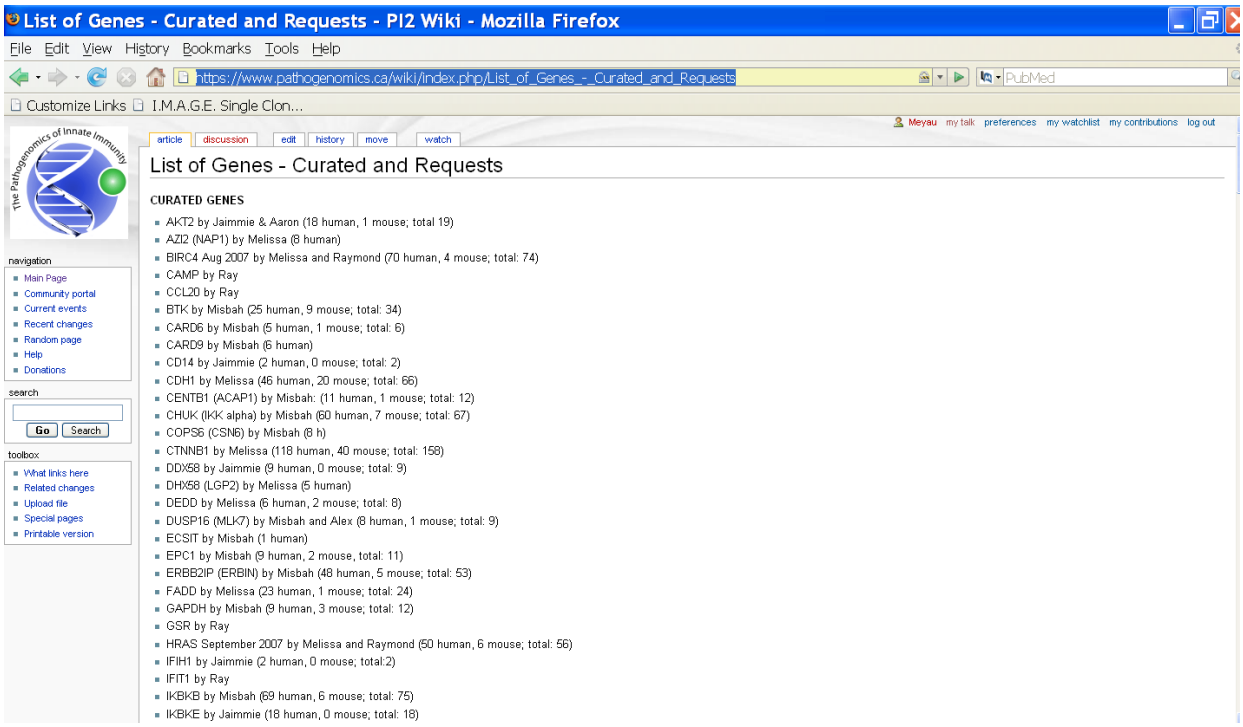


To save time and ensure consistency, tissue type, cell type and species have been recorded for curated cell lines. If a cell line is not listed, add it to the page by following the example in section 1.3.3.10 in 1.3: Adding an interaction.

2.3.2 Track Curation Progress

https://www.pathogenomics.ca/wiki/index.php/List_of_Genes_-_Curated_and_Requests

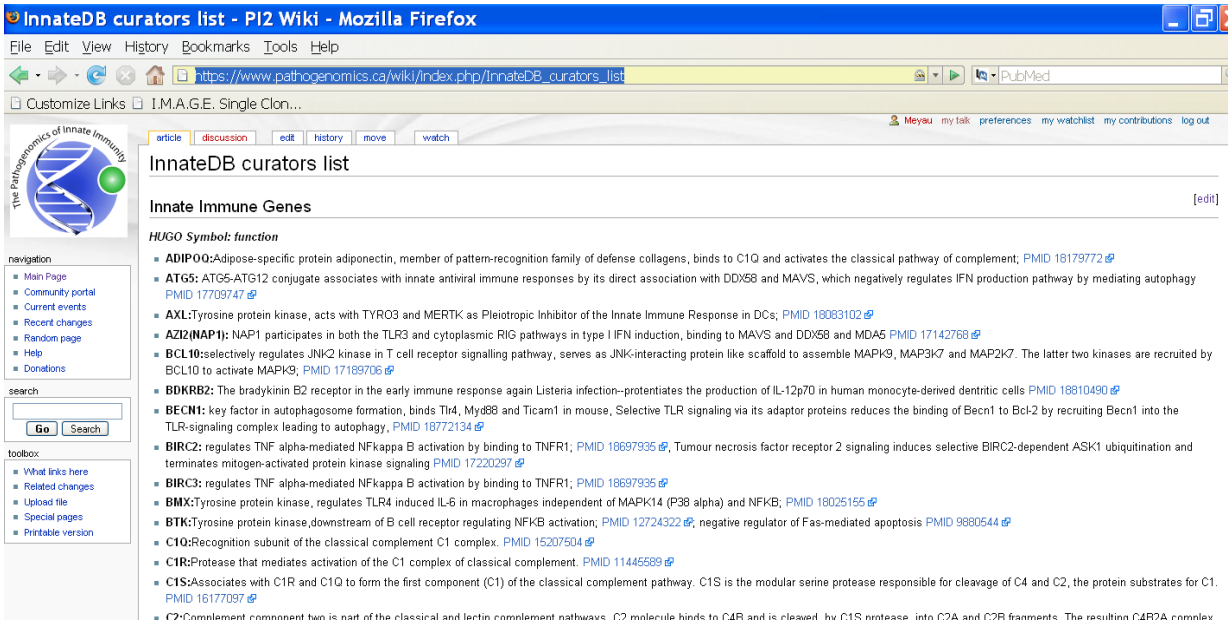
This page is used to record systematically curated genes with the number of interactions in human and mouse. And genes requested by the lab or project manager are also documented.



2.3.3 Record immune genes and their function

https://www.pathogenomics.ca/wiki/index.php/InnateDB_curators_list

This page was used previously to record immune genes and their function in specific scientific publications while curating. This is now replaced by the innategenes function in the submission system.



2.3.4 Curation Tips

Papers should be read in full
Articles sent out via innategene/curation emails should always be read to ensure all relevant interactions are being captured. Introductions and discussion sections may reference interactions or genes that are relevant for curation and innategene annotations, and it may be necessary to track those articles. [Note: Nature Immunology articles are performed as a part of InnateDB's contribution to IMEx, so those papers can be put on hold for curation until the full issue is released.]

Innate Immunity Pubcrawler queries on a semi-weekly basis
It is important that each curator set-up independent PubCrawler queries and perform dedicated searches at least once every two weeks on canonical innate immunity pathways, such as the TLR / NLR / RIG-I pathways, to ensure that InnateDB is up-to-date with current literature.

Chapter 3: IMEx Curations by InnateDB

3.1 The International Molecular Exchange Consortium

InnateDB is an active member of the International Molecular Exchange Consortium (IMEx). This organization is dedicated to developing rules for capturing protein-protein interaction data, actively curating these interactions from the scientific literature and making them available through a common website.

3.2 Curating for IMEx

InnateDB have committed to curating every issue of Nature Immunology from Sep 2010 onwards using IMEx curation standards. Since IMEx curation requires more annotation detail than InnateDB is able to capture with the current submission system, InnateDB will submit the interactions via the IntAct interaction database.

Curation is performed using a hierarchical method of PMID > Experiments > Interactions > Participant > Features, with annotations at each level. Upon submission, all interactions are thoroughly reviewed by an IntAct curator and any changes are made by the original InnateDB curator before it is accepted and released.

In addition to submitting to IntAct, all InnateDB acceptable interactions (i.e. interactions of relevance to innate immunity) and innategene annotations from Nat Immunol should be deposited into InnateDB.

This chapter will highlight the major differences between InnateDB and IMEx curation. Since interactions are submitted to and reviewed by IntAct, the official IntAct annotation rules should also be reviewed (see Useful Links).

3.2.1 PMID

Publications are curated in full
All protein-protein interactions described in the original body of the article as well as any supplementary material should be captured. As long as the protein(s) described can be mapped in UniProt, which annotates all proteins from all species, it should be annotated.

Email of corresponding author is captured
Upon the acceptance of the submitted interactions, an email will be sent to the author notifying the release of the interaction data.

Annotation details and Xrefs
As much detail describing the interaction should be recorded in the annotation fields at the Experiment, Interaction and Participant level. Some of the commonly used ones are described in Table 1, the full list and official definitions can be found in the IntAct Annotation Rules. Xref should also be used at each level, especially if an accession code is referenced in the paper (e.g. PDB deposition).

Table 1. Commonly Used IntAct Annotation Fields

PMID	
External-curation	"Curated by: InnateDB - a knowledge resource for innate immunity"
Contact-comments	Info from author correspondence (e.g. species)
Experiment	
Caution	Possible errors in experiments or on specifications (e.g. a non-specific antibody was used for WB)
Exp-modifications	Departure from standard procedures (e.g. modified cell lines)
Library-used	For Y2H or phage display
Interaction	
3d-r-factor, 3d-resolution, 3d-structure	X-ray crystallography or NMR studies
Agonist / Antagonist	Treatments on the cell that induced or inhibited the interaction, respectively
Stimulation / Inhibition	Proteins or small compounds that facilitated or inhibited the interactions, respectively
Figure legend	Reference the figure or table supporting the interaction
Kinetics	Used in isothermal titration calorimetry (can also be captured under Parameter)
Parameters	Captures Kd, Kon, Koff in SPR studies
Resulting-ptm	Post-translational modifications for enzymatic reactions

3.2.2 Experiment(s)

Redundant evidences are captured separately
IMEx curation requires the meticulous documentation of all experimental evidence supporting an interaction within a single publication. With the exception of two-hybrids, all interaction should have an accompanying figure or table reference in the publication. Evidence may be merged only if the experimental conditions, the interaction and participant detection methods, as well as the participant features are identical. A reciprocal CoIP, for example, should be entered as two separate lines of evidence.

In vitro is reserved for cell-free environments and extracellular interactions
Unlike InnateDB, where *in vitro* is used for cell-free as well as cell culture studies.

3.2.3 Interaction(s)

Exclusion of nucleic acid::protein interactions
IMEx does not accept DNA and RNA as interaction participants. These interactions should still be curated using standard InnateDB rules and deposited into InnateDB only.

Inclusion of colocalisation interactions
These interactions are typically demonstrated via confocal or fluorescent microscopy experiments at the cellular level of resolution. Marker proteins (as specified in the paper) should not be entered as a participant. Colocalisation interactions should also be Xref to the appropriate cellular component via GO.

Protein Complexes are annotated as n-ary "association" interactions
In situations where one bait captures more than two preys (e.g. pull down, LC MS/MS), the participants are entered in a single "Association" interaction. "Physical Association" designations are normally reserved for binary interactions.

3.2.4 Participants

Participants are mapped to UniProt
Participants are imported via their UniProt accession number, which designates protein and not genes. It allows the mapping to specific protein isoforms if the information is available. If the author does not specify the protein isoform used, the participants are mapped to the parent protein. In cases where a synthetic protein construct is used and is conserved across multiple species, curator may map the synthetic participant to the species matching that of its interactor(s).

Oligomer participants are imported once and assigned a stoichiometric value
Proteins that oligomerize are imported as one participant with stoichiometric value of ≥2. For example, A dimerizes and binds to B. Instead of entering A::A::B, it should be submitted as A::B with a stoichiometric ratio of 2:1.

Participant features annotates tags, binding region, mutations… etc.
The full participant feature MI terminology tree can be found on OLS (MI:0116). Each participant feature is mapped to a region on the protein (e.g. amino acids 320-400, c-c, n-n, ?-?). All tags, radiolabel, or isotope labels should be captured, and if possible, mapped to a protein region.

Deletion experiments are distilled to the narrowest possible binding region and assigned as "required to bind"; where as if the interaction is shown with a partial protein, the region is denoted as "sufficient to bind". If the region overlaps with an InterPro protein domain, it should be cross-referenced to the domain accession number.

Amino acid substitutions affecting the interaction are also captured using 3-letter codes of amino acids (e.g. Tyr34Ala) and assigned as increasing, decreasing or disrupting based on their influence.

3.3 Useful Links

3.3.1 IMEx

IMEx Consortium
http://www.imexconsortium.org/
This is the public website for the IMEx consortium where it lists the purpose of the consortium, the participating databases, and the journals that each of the database is responsible for. The official IMEx curation rules are also found here. It is a good idea to note which journals are being curated by other IMEx members to avoid duplication of curation efforts, especially for high-throughput interactome papers.

IMExCentral
https://imexcentral.org/icentralbeta/
IMExCentral is the semi-internal website that the IMEx member use to reserve PMIDs for curation, and it will return an unique IMEx accession number if that publication has not been curated by one of the IMEx members. In addition, it allows for the recommendation of papers from public users. The generation of IMEx ID is automatic through the IntAct editor so the curator should not have to manually reserve Nat Immunol articles.

IMEx Interactions via PSIQUIC
http://www.ebi.ac.uk/intact/imex/main.xhtml?query=&Search=Search
This is the search engine for all of the IMEx interactions from all the participating databases.

3.3.2 IntAct

IntAct Database
http://www.ebi.ac.uk/intact/main.xhtml
The IntAct interaction database.

IntAct Annotation Rules
http://www.ebi.ac.uk/~intact/site/doc/IntActAnnotationRules.pdf?conversationContext=2
The complete IntAct annotation manual. Keep in mind that since these interactions are submitted to IntAct, they should adhere to the IntAct annotation rules, which may be more stringent than IMEx. For example, curators are expected to capture a negative interaction if it accompanies a positive interaction evidence; conversely, IMEx does not take negative interactions at all.

Ontology Lookup Service (Molecular Interaction)
http://www.ebi.ac.uk/ontology-lookup/browse.do?ontName=MI
Nomenclature for all IMEx interaction data annotations, including participant features (MI:0116), should adhere to OLS standards. This browser offers the official definition of each MI term.

3.3.3 Cross References

Nature Immunology
http://www.nature.com/ni/index.html
Nature Immunology issues are published on a monthly basis, and was chosen by InnateDB for its focus on immunity research and the impact factor of the journal. Each issue will consist of approximately 8 original research articles for IMEx curation. Each article, including all supplementary materials, should be thoroughly read to ensure all interactions are being captured. InnateDB have curated Nature Immunology from Aug. 2010.

UniProtKB
http://www.uniprot.org/
Instead of mapping to ENSEMBL genes like InnateDB, IntAct maps their participants to UniProtKB. UniProt/TrEMBL () automatically annotates proteins from all species, while UniProt/SwissProt () entries have been manually reviewed by a curator with high quality annotations and are non-redundant. Always try to use the reviewed entries if possible.

InterPro
http://www.ebi.ac.uk/interpro/
InterPro annotates protein signatures or domains. Since each participant feature is mapped to a particular region on the protein, curators are expected to cross-reference InterPro accession number(s) if the region spans an InterPro annotated domain.

Gene Ontology
http://www.geneontology.org/
Gene Ontology is cross-referenced in colocalisation interactions to specify the cellular component that the protein complex was found in.