Annotation Guideline

# Purpose

Creating training data for machine learning. The task to be solved is Causal Relation Extraction between genes/proteins in the biomedical text.

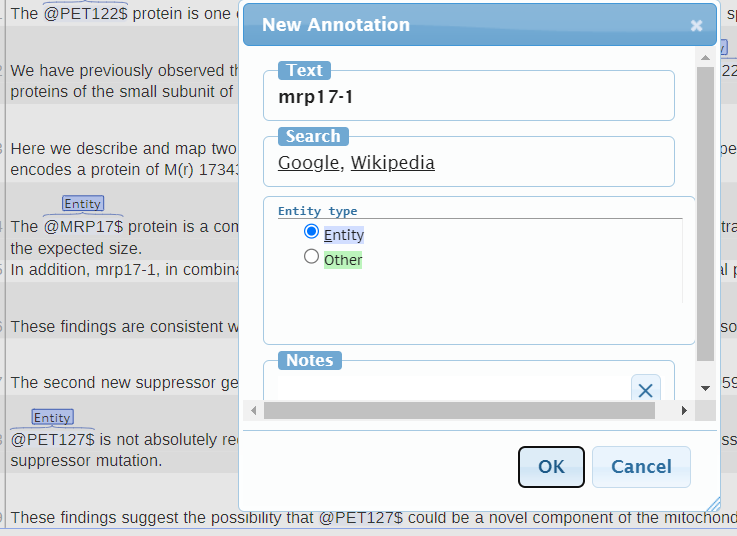
# Annotation Tool

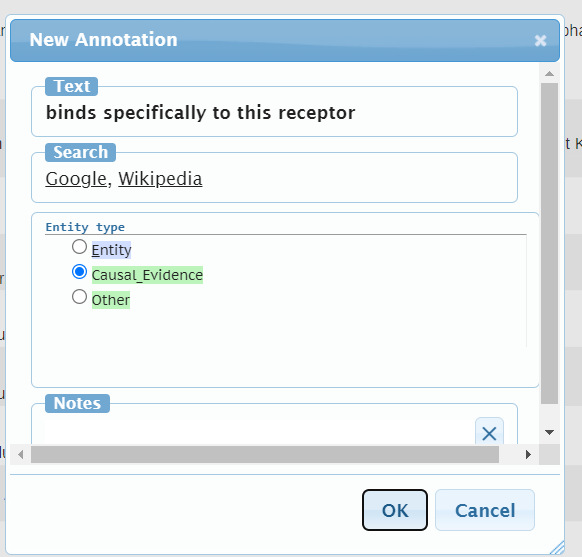
We use BRAT annotation tool. How to install: <http://brat.nlplab.org/installation.html>   
We recommend to use the **standalone server** for a quick and easy installation. We will provide the data, put the data into the brat data folder, and you can access and annotate them via your browser. Please note that BRAT only runs on UNIX-like systems. To use it on Windows, the easiest way is to install Linux for Windows or other similar things where you can run Linux on windows.   
**Example on our server**: http://10.25.181.158:8001/index.xhtml#/biogrid\_ann/1279374  
Point the cursor to the top-right page and log in with *username/password = user/user* to start annotating. *We will provide short training for the annotator on how to use the tool to annotate, if necessary.*

# Target Data

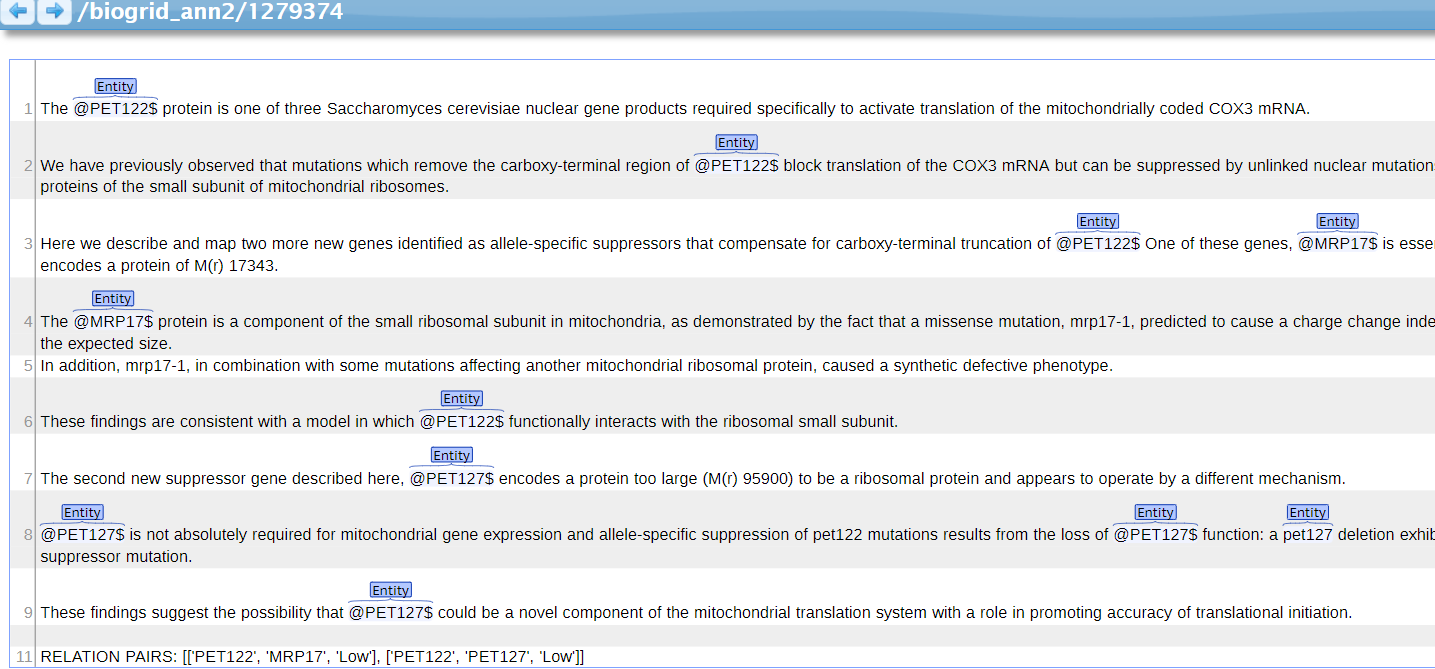
The target data is 250 abstracts from Pubmed biomedical articles used as the sources for the curation of the BioGRID (thebiogrid.org) database. The Pubmed IDs were collected from BioGRID, and the abstract is then crawled using Pubmed API. For the complete data see <http://10.25.181.158:8001/index.xhtml#/biogrid_ann2>

# Annotation Rule and How-to

1. **Entity Annotation**
   1. Tag all of Gene (Gene Product or Protein?) appearing in the text as Entity.Select word/phrase (highlight it with your cursor) that you want to tag as entity, release the cursor and choose the ‘Entity’ entity type in the dialog that appears.  
      
   2. Some of the entity has been annotated in the data; in this case, you can **edit** (e.g., fix the span of the entity) or **remove** (un-annotated them as entity) them, or just leave them as they are (assuming they are correct), according to your judgment. Double click on the already-tagged entity to edit/remove them.
   3. Even if a gene/gene product/protein does not have relation with others as written in the text, still tag all of them as Entity.
2. **Causal Evidence Annotations**
   1. If a gene/word/phrase/ in a sentence acts as causal evidence, tag the gene/word/phrase as a *Causal\_Evidence* tag as in the example below. If the sentence itself is the causal evidence, try to find specific genes/words/phrase in the sentence to be tagged with *Causal\_Evidence* tag. Put some explanation in Notes if necessary.



1. **Relation Annotation: annotate the relation between entities.** 
   1. **First, check the last line of each document, i.e.**, the line that starts with “RELATION PAIRS”. This line contains the gene pairs having a relation, as extracted automatically from the BioGRID database. This can be used as a reference when annotating the relation between genes. For example, in this screenshot below, one of the relation pairs is between PET122 and MRP17 genes, ['PET122', 'MRP17', 'Low'], so your task is to choose/to annotate which gene pairs in which sentence in the document represent those relation pairs. Of course, this is only for reference, and as an expert, you are allowed to choose and annotate the relation by yourself if you think this reference is not quite correct.

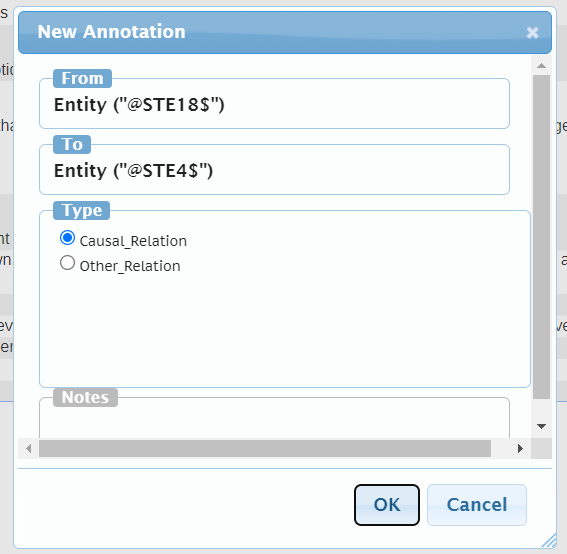


* 1. To annotate, drag and hold the cursor from one entity to another entity. You will see an arrow pointing toward the entity, indicating the relationship between them. Release the cursor and select one from 3 relationship types that appear**:** 
     1. **Causal\_Relation:** for **causal-like** relation between genes. Some keywords include, *cause, because,* *interacting/interaction, inhibit,* etc. e.g., *PKM2* and *MLC2* in the text below.

*Example: Here, we found that Aurora B phosphorylates* ***PKM2*** *but not PKM1, at T45; this phosphorylation is required for PKM2 s localization and interaction with myosin light chain 2* ***MLC2*** *in the contractile ring region of mitotic cells during cytokinesis*

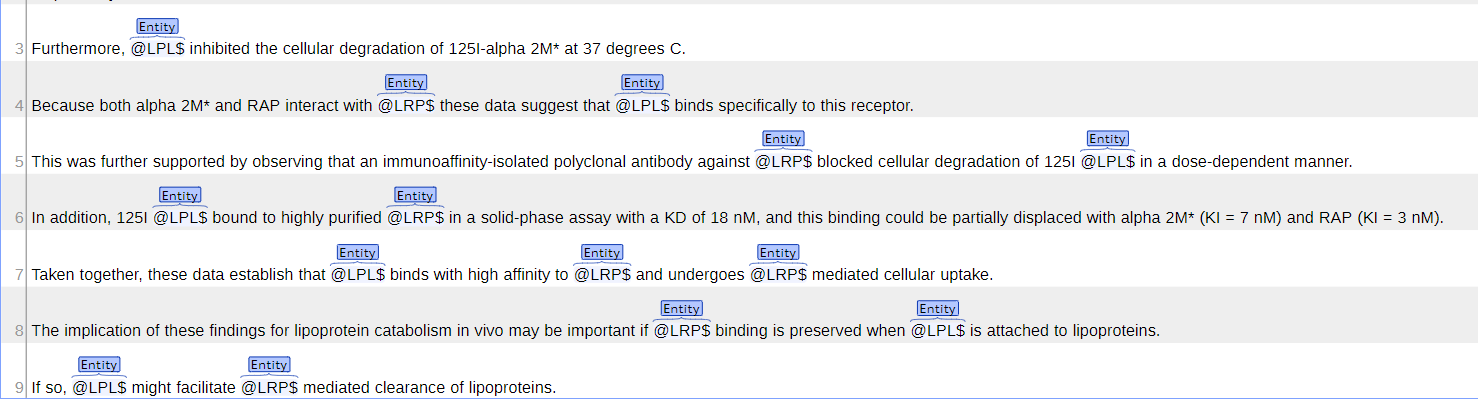
**P.S. We assume other gene pairs as non-causal unless annotated as so. (e.g.,** PKM1 and MLC2 in the example above).

* + 1. **Other relation:** For other relations (gene pair has relation but a non-causal one), please leave a note of what kind of relationship it is.



* 1. Prioritize annotating relation between two entities that appear in **one sentence** (e.g., *relation between LRP & LPL* in sentence 4 in the screenshot below)
  2. If no relation can be found between entities in a sentence, **and if you must**, a relation can be annotated between entities from different sentences. (e.g., *relation between LPL in sentence 3 and LRP in sentence 4* below)
  3. Double click on the already-tagged relationship (the arrow) to edit/remove them.

**Example: Before annotation**



**Example: After annotation**

