## Annotate the text using unified names

“Span” indicates the span (i.e. the set of characters to select) of an entity. It refers to a selection of consecutive characters of the entity. Annotations in **Blue** denote disease and in **Green** denote organism.

1. Annotation is correct for both the span and type

Finally, the researchers report that injection of PKHB1 reduced the **tumor** burden in a mouse model of CLL. [PMC4348493]

“tumor” is annotated as Disease, which is correct both for span and type.

1. Annotation type is correct but the span is wrong

Finally, the researchers report that injection of PKHB1 reduced the **tumor** **burden** in a mouse model of CLL. [PMC4348493]

“tumor burden” is annotated as Disease. The correct annotation should be “tumor” and the type should be Disease. The annotation is longer than the expected entity “tumor”. Therefore, it has wrong span but correct type.

Finally, the researchers report that injection of PKHB1 reduced the **tum**or burden in a mouse model of CLL. [PMC4348493]

“tum” is annotated as Disease. The correct annotation should be “tumor” and the type should be Disease. The annotation is shorter than the expected entity “tumor”. Therefore, the annotation has wrong span but correct type.

1. The span is correct but the type is wrong

Finally, the researchers report that injection of PKHB1 reduced the **tumor** burden in a mouse model of CLL. [PMC4348493]

“Tumor” is annotated as Organism. The correct annotation should be “tumor” and the type should be Disease. Therefore, the annotation has wrong type but correct span.

1. Both the span and type are wrong

Finally, the researchers report that injection of PKHB1 reduced the **tumor burden** in a mouse model of CLL. [PMC4348493]

“tumor burden” is annotated as Organism. The correct annotation should be “tumor” and the type should be Disease. Therefore, the annotation has wrong type and wrong span.

1. Missing entity (false negative)

Finally, the researchers report that injection of PKHB1 reduced the tumor burden in a mouse model of CLL. [PMC4348493]

“tumor” is missing from the annotation. Therefore, it’s a missing annotation.

Gene-Disease Relationship annotation:

1. The relationship is correct
   1. Both of the gene and disease entities are correct
   2. The relationship exists between the entities.
2. The relationship is wrong
   1. One or both of the entities in the pre-annotated relationship have wrong type
      1. Refer to the entity annotation to check whether the type is correct
   2. Entities are correct but relationship doesn’t exist
3. The relationship is ambiguous:
   1. Both entities have correct type, but the relationship is ambiguous
4. In the current phase, it is not necessary to annotate missing relationship

## Tag scheme for annotations

1. Tags for indicating wrong/correct annotations:

|  |  |
| --- | --- |
| **Category** | **Tag** |
| Wrong type | WT |
| Wrong span | WS |
| Missing | MIS |
| Correct | CRT |
| Table 1 | |

1. Tags for entity:

|  |  |
| --- | --- |
| **Name** | **Tag** |
| Gene/Protein | GP |
| Organism | OG |
| Disease | DS |
| Table 2 | |

1. Tags for gene-disease relationship:

|  |  |
| --- | --- |
| **Category** | **Tag** |
| Correct relationship | YGD |
| Wrong relationsjip | NGD |
| Ambiguous | AMB |
| Table 3 | |

1. Special tag:

|  |  |
| --- | --- |
| **Special Tag** | **Tag** |
| All | ALL |
| Table 4 | |

## Usage of annotation tags

In order to indicate both the wrong correct tags. We suggest to use following scheme to report wrong/correct/missing annotations.

1. Annotation is correct for both the span and type

|  |  |
| --- | --- |
| **Type** | **Tag** |
| Gene/Protein | CRT\_GP |
| Organism | CRT\_OG |
| Disease | CRT\_DS |
| Table 5 | |

1. Annotation type is correct but the span is wrong

|  |  |
| --- | --- |
| **Type** | **Tag** |
| Gene/Protein | WS\_GP |
| Organism | WS\_OG |
| Disease | WS\_DS |
| Table 6 | |

1. The span is correct but the type is wrong

In order to record the wrong annotation type, we need to use underscore to indicate the wrong type. For example, [WT\_GP] means the wrong annotation type is Gene/Protein. The correct type can be indicated using an additional tag as shown below. If the annotation is a false positive, then we don't need to provide the correct type.

|  |  |  |
| --- | --- | --- |
| **Wrong Type** | **Correct Type** | **Tag** |
| Gene/Protein | Organism | [WT\_GP][OG] |
| Gene/Protein | Disease | [WT\_GP][DS] |
| Gene/Protein | None | [WT\_GP] |
|  |  |  |
| Organism | Gene/Protein | [WT\_OG][GP] |
| Organism | Disease | [WT\_OG][DS] |
| Organism | None | [WT\_OG] |
|  |  |  |
| Disease | Gene/Protein | [WT\_DS][GP] |
| Disease | Organism | [WT\_DS][OG] |
| Disease | None | [WT\_DS] |
| Table 7 | | |

1. Both the span and type are wrong

Refer to Table 7. If the type is wrong, the span is not important. Therefore, we don't need to record whether the span is right ot not. Use the scheme in Table 7.

1. Missing entity (false negative)

|  |  |
| --- | --- |
| **Type** | **Tag** |
| Gene/Protein | MIS\_GP |
| Organism | MIS\_OG |
| Disease | MIS\_DS |
| Table 8 | |

1. Usage of the special tag

The special tag [ALL] is used when the current annotation can be applied to the same annotations in the full text. For example, if all the pre-annotations of “tumor” are correctly tagged as Disease with the right span in one article, then we can use the combination of [CRT\_DS][ALL] to indicate all the same pre-annotations o “tumor” are correct. Therefore, we can skip the same pre-annotations after it.

1. Annotation of gene-disease relationship

If the pre-annotation of the relationship is correct, use tag **YGD** from Table 3.

If the pre-annotation of the relationship is wrong, use tag **NGD** from Table 3:

* One or both the entities have wrong type
* Both entities have correct type, but no relationship

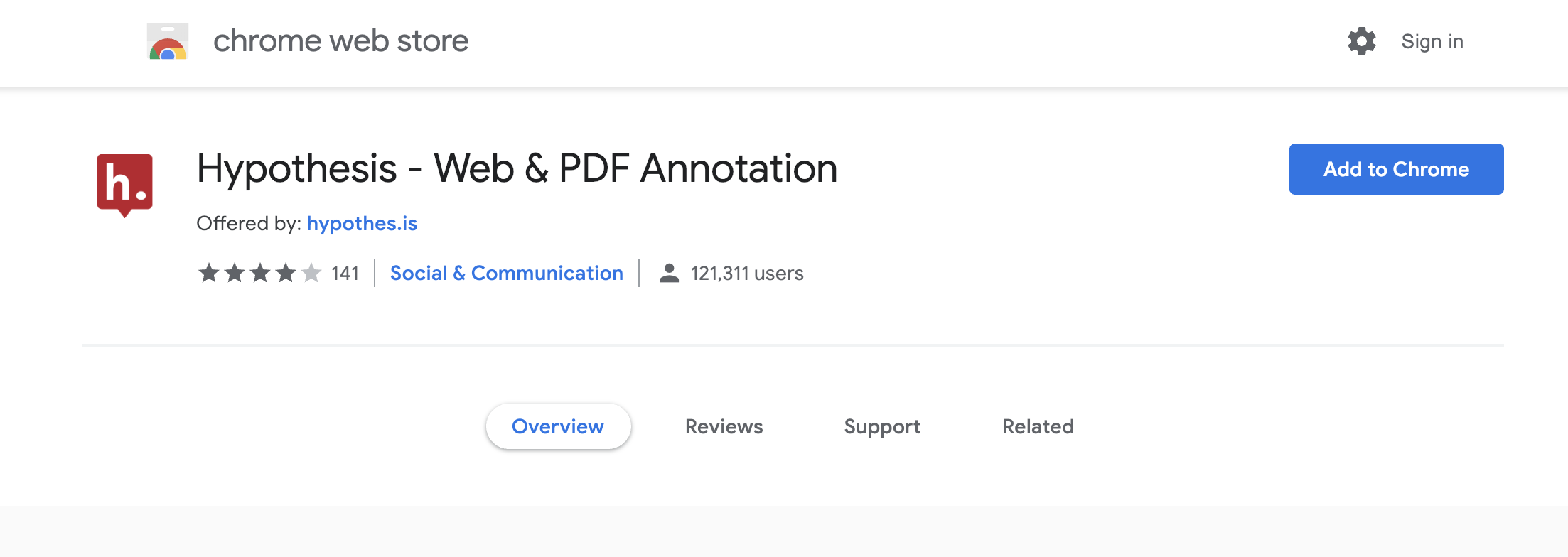
If the pre-annotation of the relationship is vague/ambiguous, use tag **AMB** from Table 3:

* Both entities have correct type, but the relationship is ambiguous

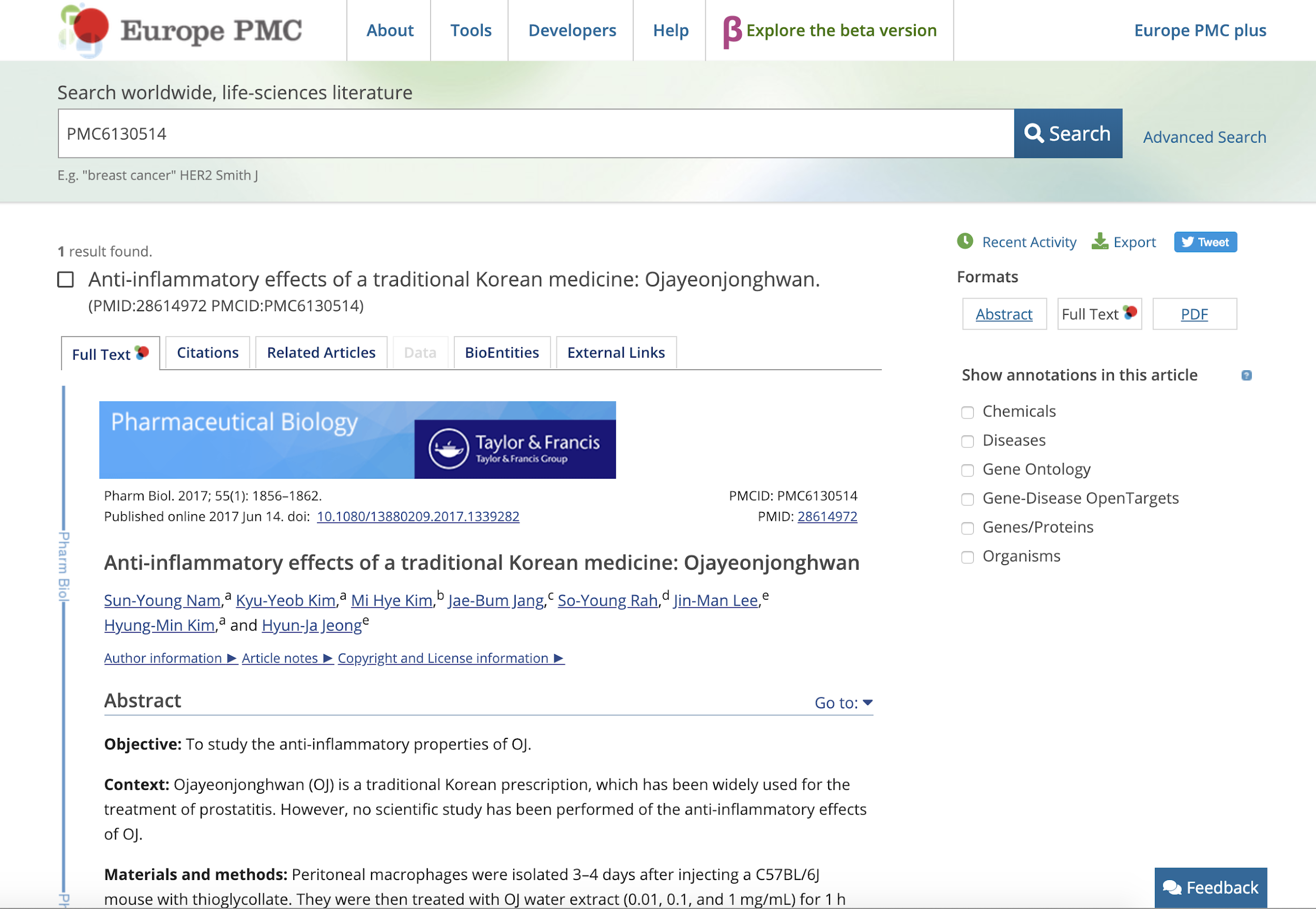
## How to use the interface

The following examples illustrate how to use the Hypothes.is plug-in. Chrome must be installed as the current plug-in only support Chrome. The screenshots may differ from the aforementioned tagging scheme, therefore please refer to tagging scheme for annotation.

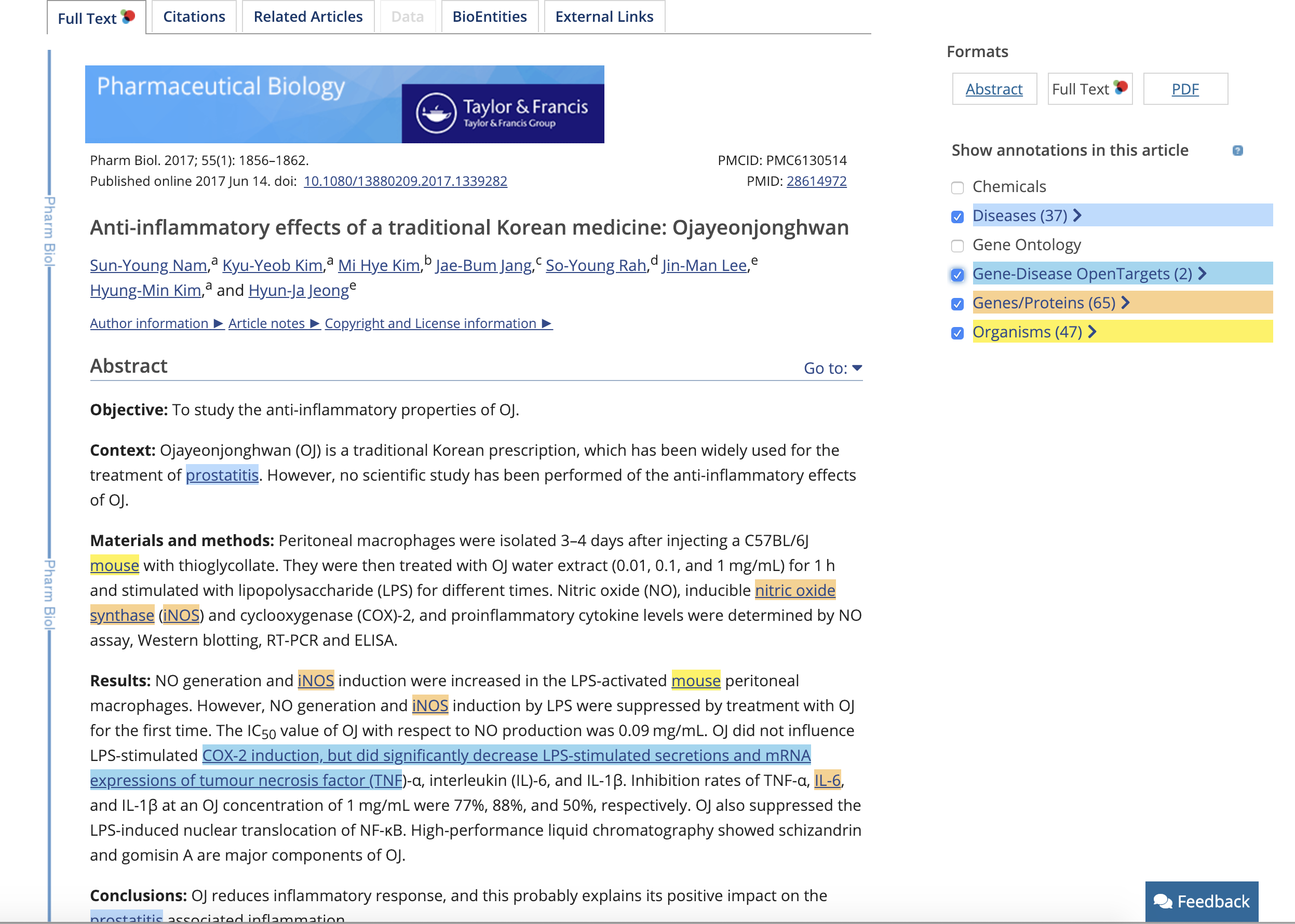
1. Annotators create Hypothes.is account.
2. An invitation of joining the annotation group will be sent to all annotators.
3. Install Hypothes.is plug-in in Chrome app store (Add to Chrome)



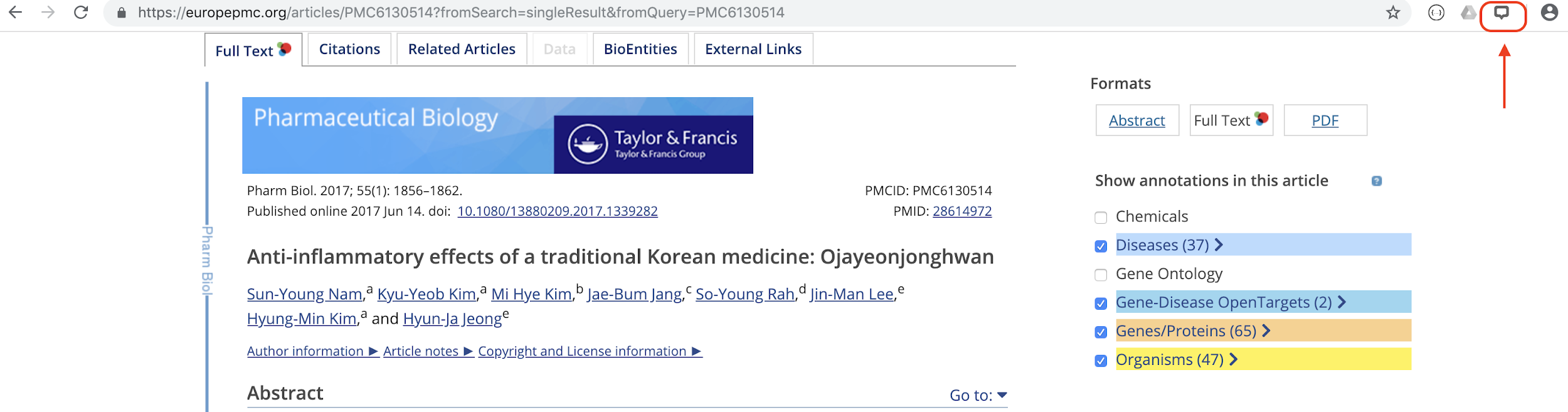
1. Open an article in EuropePMC using PMCID



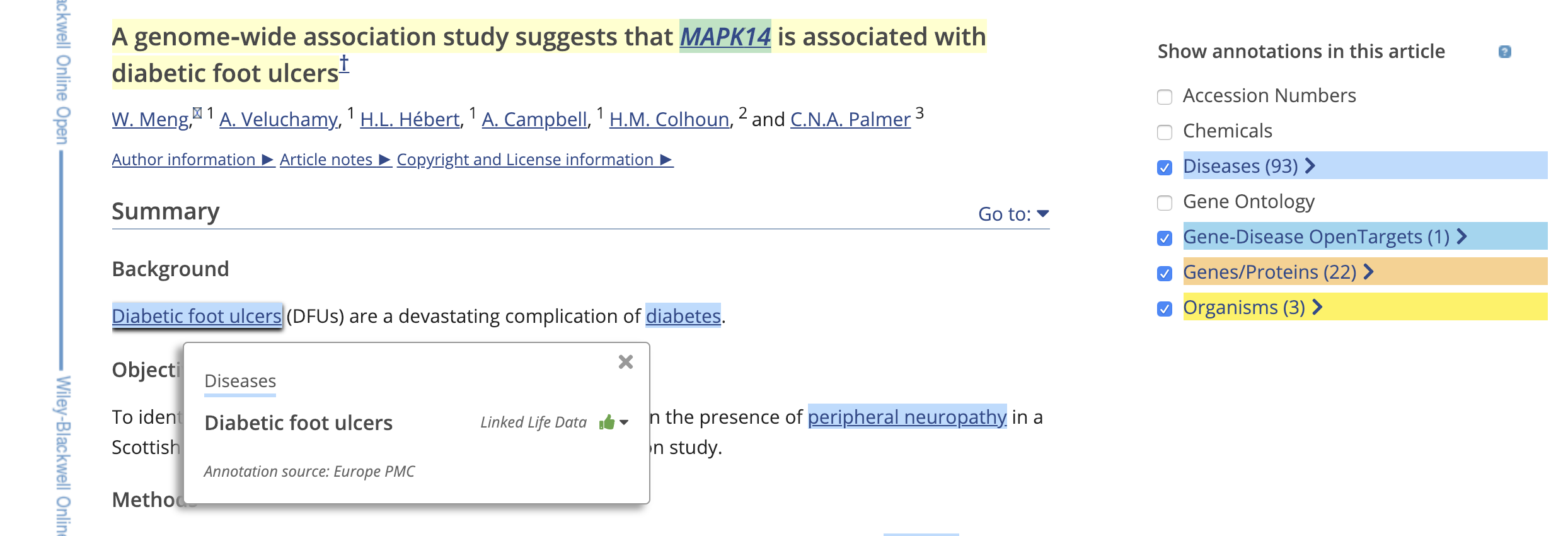
1. Select Gene/Protein, Disease, Organism and Gene-Disease OpenTargets ( if available) from the right panel to show annotations



1. Click the Hypothes.is plug-in symbol to activate Hypothes.is



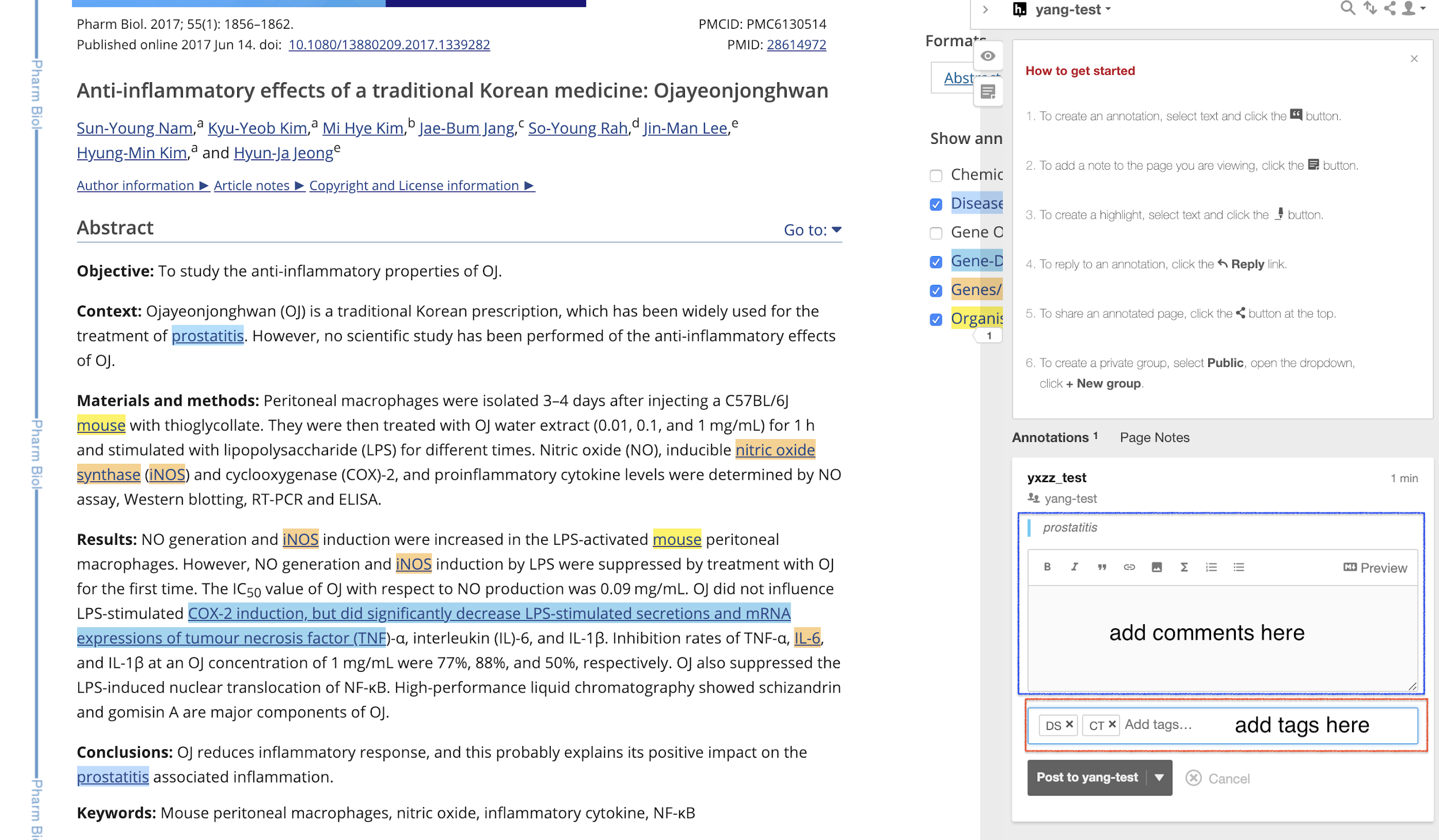
1. The pre-annotated text are highlighted in different colours. Click the highlighted text, a window will pop up to show more details e.g. the entity type and the annotated text.



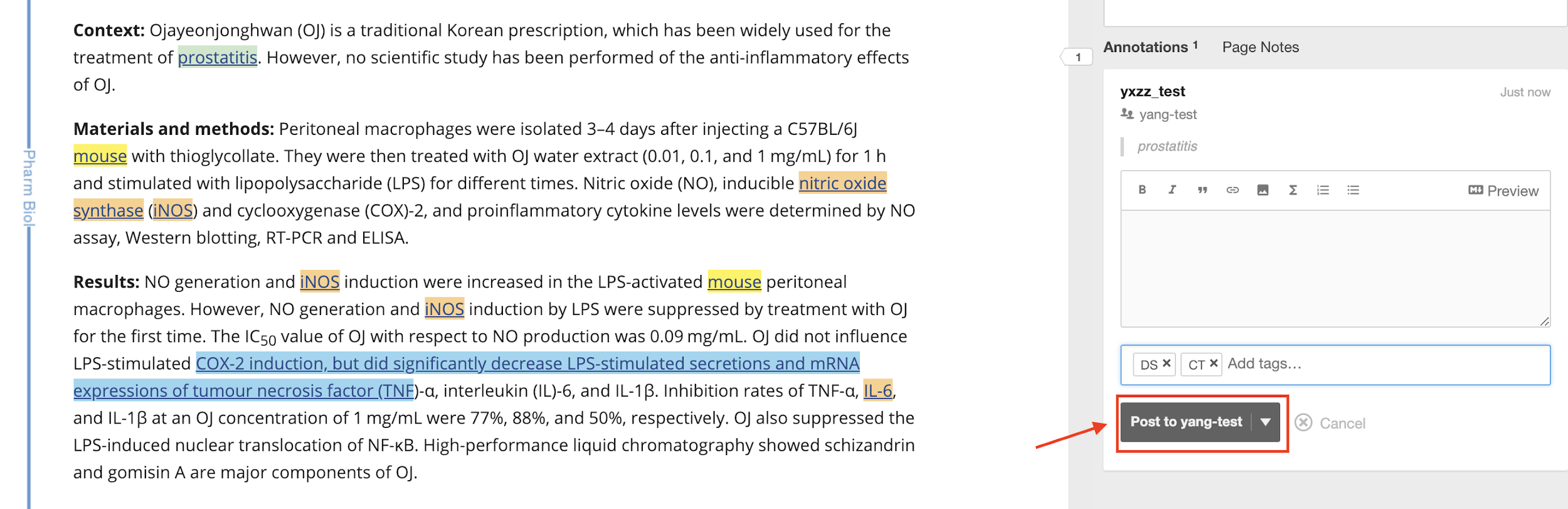
1. Use the mouse to select the entity that you would like to annotate.



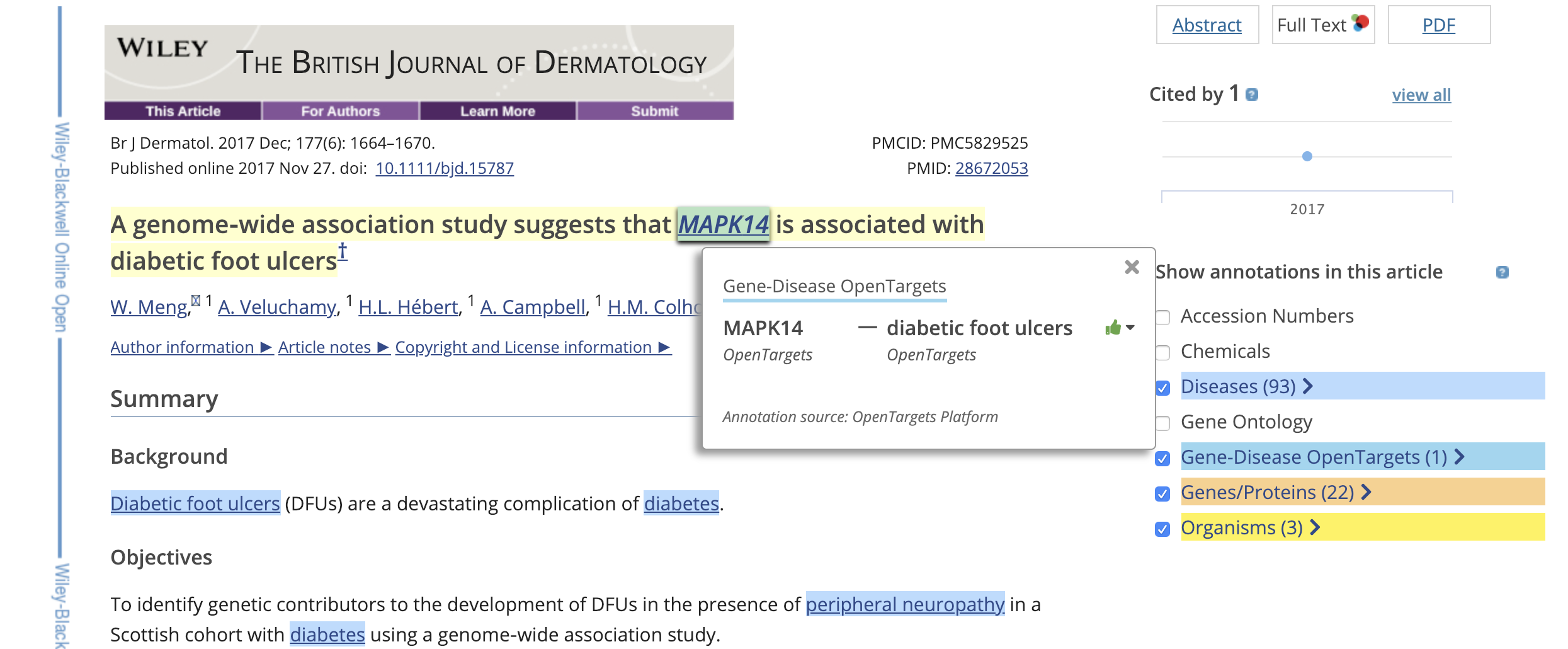
1. Click “Annotate” to annotate the select words in the pop-up panel. Add tags of the annotation in the tag box according to the Tag Scheme. If you have any comments, you can leave it in the text box.



1. To finish the annotation, click the “Post to” button to post the annotation to the correct annotation group. Then the annotation will be added to the annotation group.



1. For Gene-Disease relationship annotation, click the highlighted text, the pre-annotated relationships will appear in a pop-up window.



1. If a relationship between a gene and disease appears in the sentence, only select the part that contains the two entities using Hypothes.is. Then annotate the selected part by adding a gene-relationship tag to indicate whether it’s a correct pre-annotation or a missing relationship annotation.

If the pre-annotation is wrong, select the pre-annotation and annotate it as a wrong relation.

