Dr Tomasz Kurowski
t.j.kurowski@cranfield.ac.uk

Cytoscape practical – Part II

In cases where there are few measured interactions, text mining can be a useful mechanism for inferring network data. The Agilent Literature Search plug-in for Cytoscape provides a flexible, interactive platform for mining text and assessing the results in a network context. Here we shall explore the use of this plug-in.

This plug-in searches public literature repositories such as PubMed for articles matching user-specified queries, and then builds a network based on putative associations suggested in the text of the articles. Putative associations are sentences of two or more gene or protein names, and verbs that suggest interaction such as "catalyzes", "is repressed by", or "regulates". In this tutorial you will:

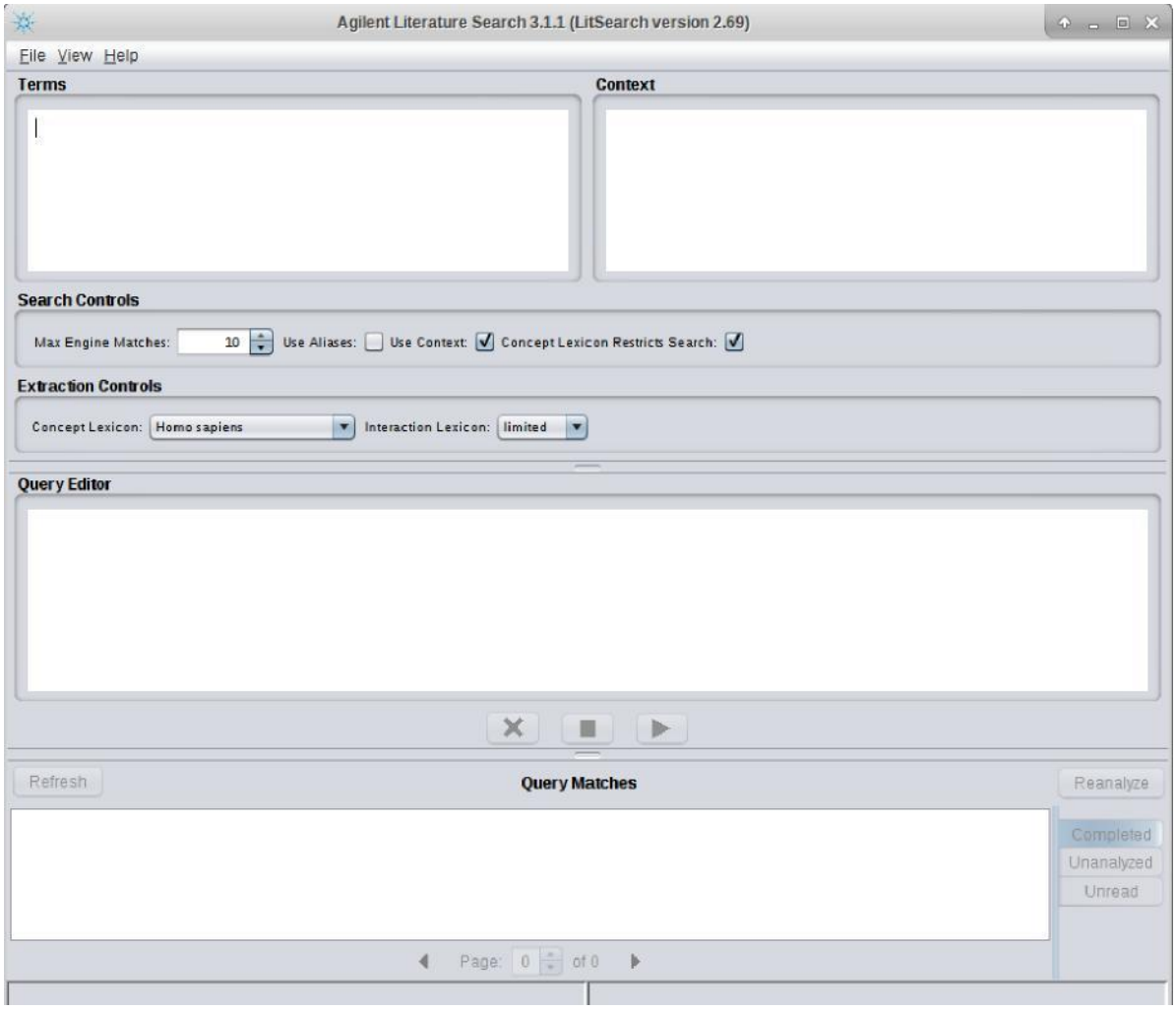
- Learn how to apply the Agilent Literature Search plug-in to build a network of putative molecular associations from a set of literature search terms
- Explore the associations generated by the app and learn how to remove any that you judge to be in error.
- Learn how to refine your search by using context information.

Important: Public literature repositories such as PubMed are always changing. The illustrations shown here are based on the public literature databases as they existed when the document was last updated. At another time, after more papers have been published, the composition of the databases will change, and so will the search results. This means your output will not look exactly the same as what is shown in the handouts.

Basic operation

1. Start up Cytoscape. Under the **Apps** menu, select **Agilent Literature Search**. You may need to first install it using the App Manager. The Agilent Literature Search Agreement will appear if this is the first time you have used it. Check the "Don't show this dialog again" box and click the Accept button.

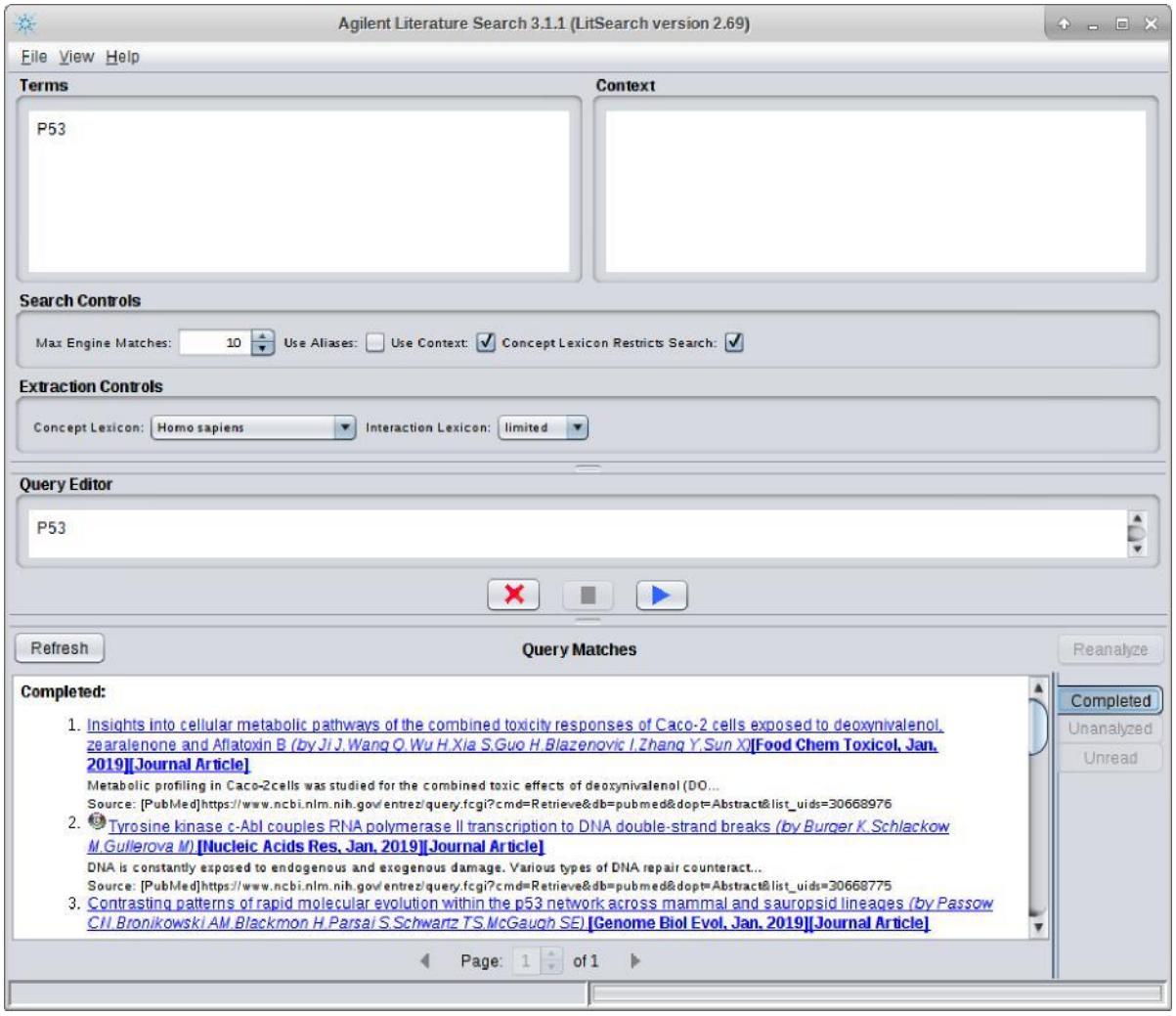
2. The following window should appear.



3. In the **Terms** window, enter P53. The term "P53" should appear in the **Query Editor**, and the forward arrow just below should turn blue toindicate it is available. Click on the forward arrow to begin searching.

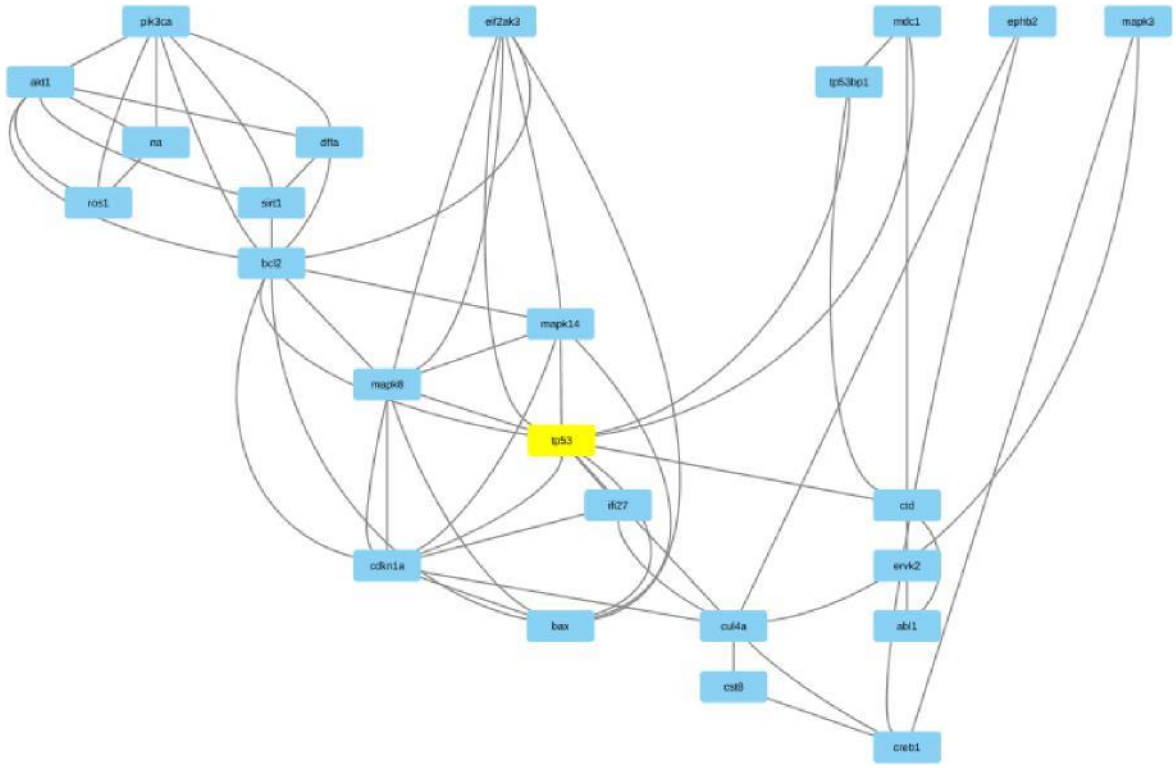
4. After a brief interval, the search results should appear in two places:

   - Under **Query Matches**, there should appear a numbered list of articles labeled **Results**, as shown below. A slider at the right side of the window allows you to scroll through the list of selected articles. Each article should be listed along with a URL, and a hyperlink for jumping directly to that URL.



Since the number of matches was set to ten, up to ten articles will be displayed. In general, these are the most recent articles that match your search terms. If you bring up a web browser and search PubMed on the same search term (P53) you should see exactly the same matches.

- A network should appear on the Cytoscape canvas, showing interactions inferred from sentences in the selected articles similar to that shown below. The network may vary according to the article hits found in the database and the algorithm applied.



In this case, the canvas shows tp53 connected directly to eleven nodes, with a number of additional nodes without a direct connection with tp53. How did these other nodes and their interactions get into a network generated by a search on P53? This is a consequence of the two-step process used by the Literature Search app:

First, it selects articles according to your query terms. In general, this retrieves articles based on comparing the search term to the abstract and keywords.

- Second, it scans the complete text of the articles for sentences describing putative interactions between genes or proteins. It then adds all putative interactions to the network, whether or not they involve any of the search terms. Why? If a putative interaction appeared in an article that relates to the search terms, then it's expected that it relates to the search terms in some way.

- In consequence, when you perform a literature search, you will typically get a network with many genes or proteins that you did not search for specifically. In some cases, the genes or proteins that you did search for might not appear in your network. This is not a bug or a usage error: it simply reflects what putative interactions were found in the articles retrieved. You can usually avoid this by increasing the number of matches.

Validating, refining, and saving your search results

You can explore the sentences that were selected as evidence of interaction between these nodes. Go to your Cytoscape canvas and click on some edge with your right mouse button. A menu should appear listing the interaction and listing a sub-menu labelled Evidence from Literature. Click on an entry, and you should see the actual sentence that was presented in this interaction.

In addition to exploring individual sentences, you can explore the evidence from entire articles, as follows:

1 Return to the **Query Matches** section of the **Agilent Literature Search window**.

2 Right-click on the first match.

3 A pop-up menu should appear with the option **Delete Match**. Click to remove the match to the first article, along with any interactions supported by that article only. After deleting the match, any nodes containing only edges that are supported by only that article will be deleted.

4 If the article has a small Cytoscape logo next to the title, the right-click menu will also show the option "Highlight Match". Selecting this option will highlight the matches derived from this article on the Cytoscape canvas. The nodes should turn yellow, and the edges between them should turn red.

5 Under the **File** menu of the **Agilent Literature Search** window, you will see options labelled **Load Search Results** and **Save Search Results**. If you want to save a set of search results for later analysis, these options will allow you to do so.

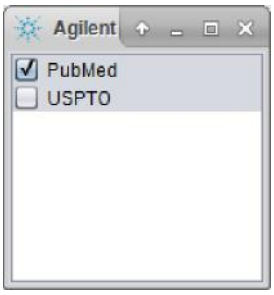
Refining your search

Under the **Agilent Literature Search** search term window, there are a number of basic search controls, as described here.

- There is a pull-down menu to select an organism ("**Concept Lexicon**").
- There is a threshold on the maximum number of matches per search engine ("**Max Engine Matches**"). Out of courtesy to the public search engines, always try to use a low threshold! If you are experimenting with the use of the plug-in, or have just started a new line of analysis, always start with a small number of matches and increase that gradually as needed. The maximum allowable is 1000.
- Under **Extraction Controls**, there is a menu labeled **Interaction Lexicon** with a choice of ***limited*** and ***relaxed***. This controls the set of

verbs that identify putative protein interaction sentences**: *limited*** selects a high-confidence set of verbs (such as "activate", "methylate" and "cleave"), while ***relaxed*** selects a more permissive set (including "join", "augment", and "induce"). Repeat your search on P53 with the **Interaction Lexicon** set to ***relaxed***. How has the network changed? Can you identify any new edges? Compare the sentences associated with the old and the new edges.

- Under the **View** menu of the **Agilent Literature Search** window, you will find the option **Engine Selections**. When you click on **Engine Selections**, you should see the following menu:



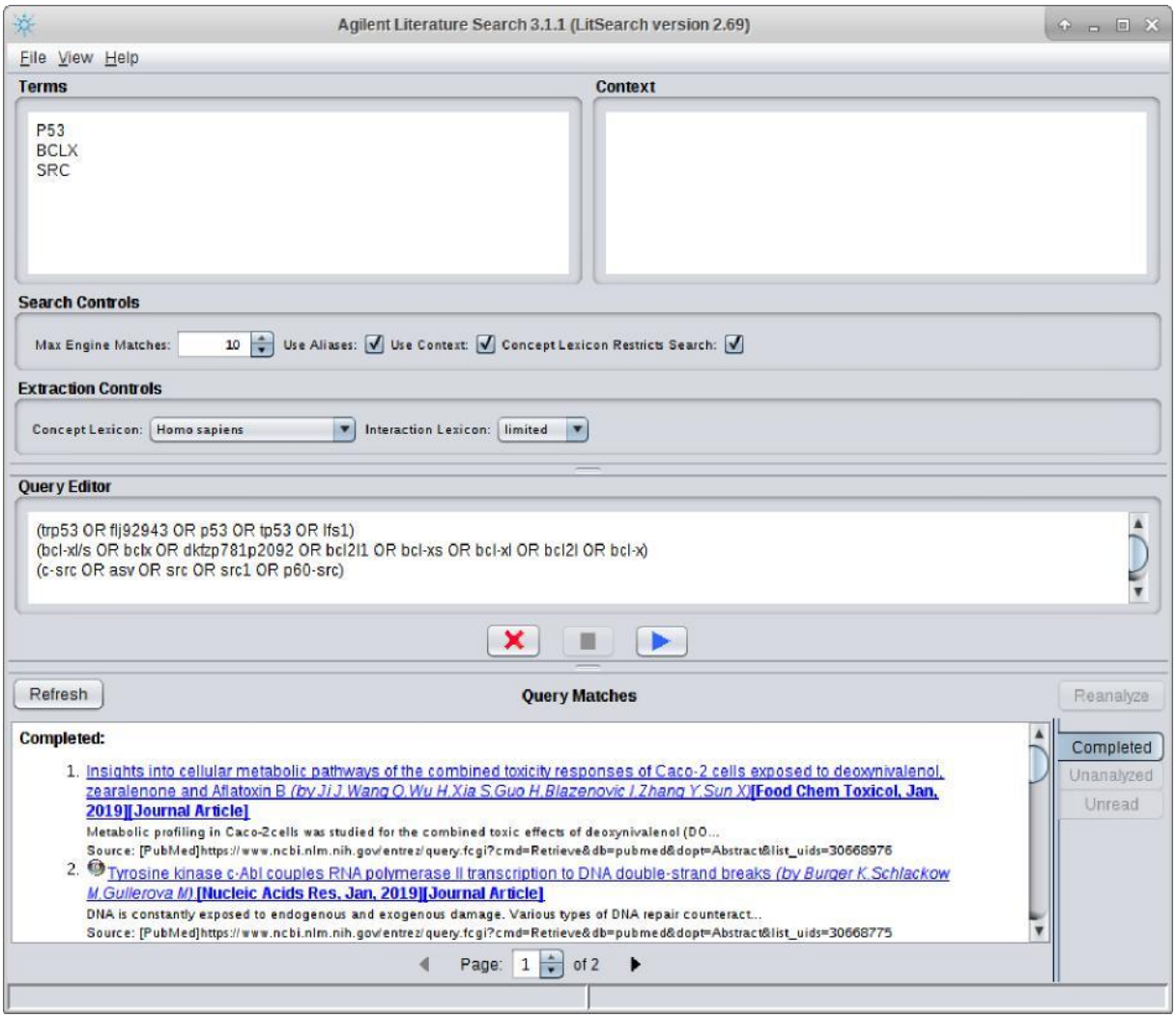
Repeat your query with USPTO selected in addition to PubMed. How does the network change? Return to this menu, and turn off querying USPTO for the moment.

- Under the **Search Controls** section of the window, there is a button labeled **Use Aliases**. Click on this button, and in the **Query Editor** you should see your search term of "p53" change to "(p53 OR trp53 OR tp53)". This is a very useful option, because gene names have many aliases. The only time when it is not valuable is when you believe that the aliases really identify two distinct macromolecules. In such cases, you can still edit your query under the **Query Editor** to remove any alias you do not wish to use.

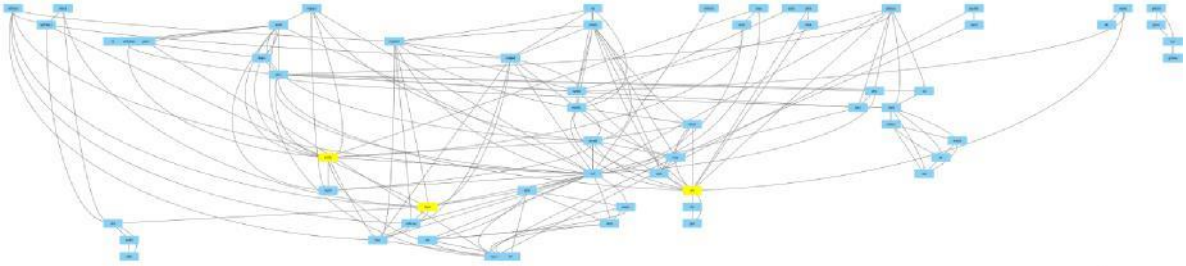
- Repeat the search using aliases. Did your network change?

- In the **Query Editor**, modify the query so that it reads "(p53 OR tp53)" and repeat your search. How did the network change this time?

- You can specify multiple search terms. Under **Terms**, under P53, add the oncogenes BCLX and SRC. Note that each term should be on a separate line. Run the search by clicking on the blue forward arrow. Your query window should appear as follows:



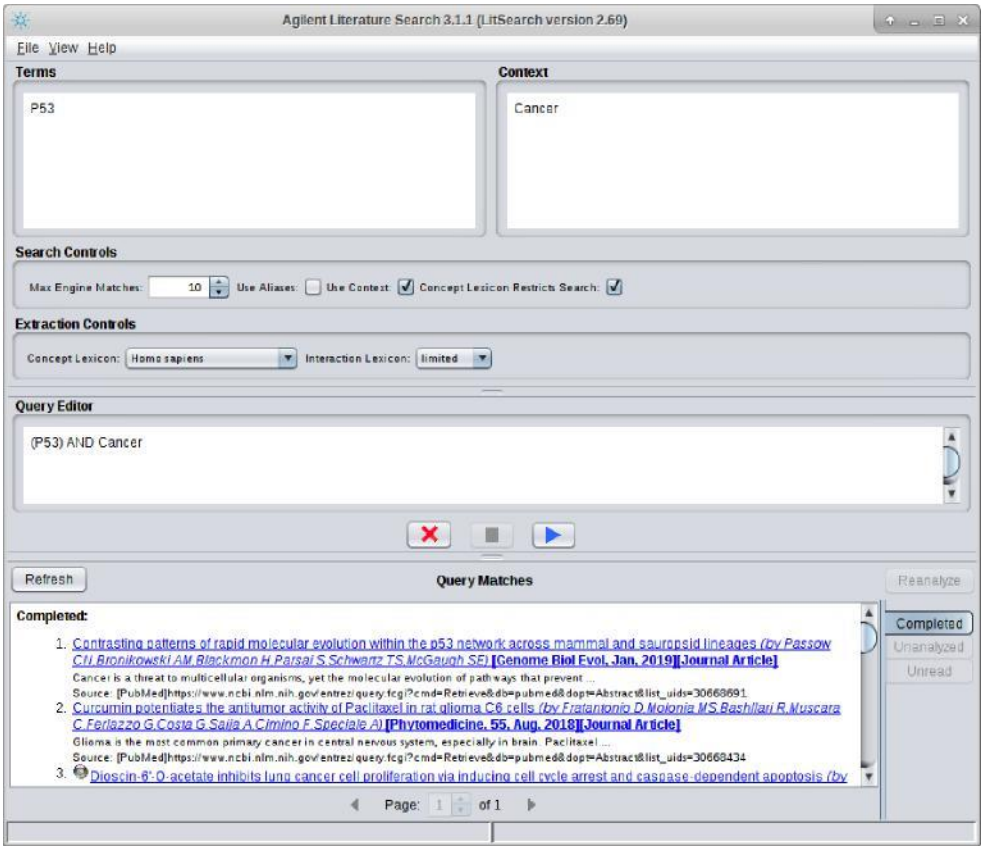
When performing a complex query, you should see up to ten matches per search term returned, and a larger network on the canvas such as the one shown:



- Specifying a context provides a valuable way to refine your search. This can yield a network that is more specific to your biological question, and potentially just as large.

1. Set your list of terms to P53.
2. In the **Context** window, enter "Cancer".
3. Click on the **Use Context** button in the **Search Controls** section.

   Your search window should appear as follows:



Notice that in your query window, your query specifies P53 AND Cancer.

In other words, the context acts as a filter.

1. Experiment with adding some additional search and context terms (one per line) to see how the query changes. Note that you can also enter or modify your query under the Query Editor.

2. Perform a search on P53 AND Cancer. You should get back ten query matches and the corresponding network.

3. Enter a new search on P53, this time with the context "dna repair". **Note: when a search term consists of two or more consecutive words, the term should be put in quotation marks.** This should produce a different set of ten articles and a different network.

If you still have time remaining, explore the features of Cytoscape further – try using the Agilent Literature Search plug-in to investigate genes of interest to you. Try the other layout options etc. Additionally, use this time to explore some of the other interaction networks resources such as: NDEx, IntAct, STRING, STITCH, etc.

Cytoscape.js

You can display networks generated in Cytoscape inside web applications by using a library called Cytoscape.js. The easiest way to do this is to download the `cytoscape.min.js` file from its website (https://js.cytoscape.org/) or GitHub (I have also uploaded it on Canvas) and include it in the `<head>` tag of an HTML page like so:

```
<!doctype html>
<html>
<head>
    <meta charset="utf-8">
    <script src="cytoscape.min.js"></script>
</head>
<body>
</body>
</html>
```

Inside of the page body, you also need to create a container tag for your Cytoscape output:

```
    <div id="cy" style="width:800px; height:800px"></div>
```

You can then interact with it (either in a <script> tag directly or inside of an included script.js file, like you have done before), like so:

```
const cy = cytoscape({
    container: document.getElementById('cy'),
    elements: [
      { data: { id: 'nodefrom' } },
      { data: { id: 'nodeto' } },
      { data: { id: 'edge', source: 'nodefrom', target: 'nodeto'} }
    ],
    layout: {
      name: 'grid',
      rows: 1
    },
    style: [
        {
          selector: 'node',
          style: {
            'label': 'data(id)'
          }
        }
    ]
});
```

As a bonus task, try to connect to the microbial growth REST API you have developed earlier in the week, and display a network showing the relationship between authors and experiments in the data. This could make the use of your junction table (`experiments_authors` / `authorship`).