Chemotext Tutorial (2017)

<http://chemotext.mml.unc.edu/>

Chemotext is a publicly-available Webserver that mines published literature in PubMed in the form of Medical Subject Headings (MeSH) terms. The goal of Chemotext is to establish text-based drug-target-disease relationships in order to identify novel drug repurposing candidates. In principle, Chemotext is an extension of Swanson’s ABC paradigm, wherein “A” terms are disease-related MeSH terms, “B” terms are so-called “target” MeSH terms, *i.e.* proteins and pathways, and “C” terms are chemical (drug) MeSH terms.

I.) Find Connected Terms

In this module, every MeSH term that occurs in the same article as a query term is returned, and the total number of co-occurring terms and the associated article counts are reported.

We will perform query in this module and explore its functionalities.

*Please note: The Chemotext database is quite large (~19 million articles and 78 million relationships), so some queries may take time. Please allow the server some time to run! If the wheel is spinning, Chemotext is working…*

A) Performing Queries and Filtering Results

1) First, **click on the “Find Connected Terms”** tab and ensure that it is highlighted in white.



2) **Enter “Kinase”** in the search bar. Note the auto-complete of MeSH terms in the search bar.



3) **Hit “Search”.** All MeSH terms that co-occur in the same article as “Kinase” are returned.



A total of 7,537 unique MeSH terms co-occur in the same articles as “Kinase”. *Please note: exact article counts will change due to constant updates to the Chemotext database.*

It can be seen that “Adenosine Triphosphate” has the most “Kinase” co-occurrences with 1865 articles in which both were mentioned. Kinases hydrolyze ATP, so it is not surprising that the two terms have a high degree of co-occurrence. Chemotext captures this relationship.

4) The default table limit for results is 10 terms. But this table can be expanded. **In the “Set Table Limit” search bar, set the count to 1000**.



When the table limit is expanded, other associated MeSH terms can be investigated. Above, we see that “Imatinib”, an FDA-approved kinase inhibitor, has 20 article co-occurrences with “Kinase”.

5) Next, the full list of co-occurring terms can be filtered for specific categories of MeSH subtypes. Scroll down the page. **In the “Filter” search bar, select “Chemicals”**, which will include all associate C-terms.



6) **Click on “Filter”**

****

A total of 2,811 unique MeSH C-terms co-occur in the same articles as “Kinase”. It can be seen that “Magnesium” has the most “Kinase” co-occurrences with 822 articles in which both were mentioned. Magnesium makes bonds with ATP when bound in the kinase active site, so it is not surprising that the two terms have a high degree of co-occurrence. Chemotext again captures this relationship.

It should be noted that in the MeSH term ontology, “Adenosine Triphosphate” is categorized as a B-term, so it does not appear among the chemicals.

*The user is encouraged to again expand the table and find the drug imatinib.*

The user can filter the original list of co-occurring MeSH terms by any number of categories provided in the filter bar. Also, the terms can be filtered by publication dates, as well.

B) Linking to Articles and Downloading Results

7) With Chemotext, the article in which both terms co-occur can be directly accessed. **Click on the “822” article count for “Magnesium”.**

Chemotext will provide the PMIDs or actual article title (when available) for all 822 associated articles.



8) From this page, an article of interest in PubMed can be directly accessed. **Click on the link for the article entitled “TRPM7 Channel Is Regulated by Magnesium Nucleotides via its Kinase Domain”**



This article illustrates how Chemotext can retrieve articles that relate two terms, *i.e.*, “Magnesium” and “Kinase”.

9) In Chemotext, the associated MeSH terms and article counts can be downloaded. Return the Chemotext page. **Click on “View CSV”.** Then all entries can be copied and saved in Excel or Notepad. Please note that pop-up blockers must inactivated on the Chemotext page, otherwise the CSV functionality does not work as intended.



10) In Chemotext, the associated MeSH terms, article counts, and PMIDs can be downloaded. Return the Chemotext page. **Click on “CSV with pmids”.** Then all entries can be copied and saved in Excel or Notepad.

II.) Find Shared Terms

In this module, two query terms are input, and co-occurring terms and the article counts that are shared between the queries are returned.

This type of search outputs the associated counts of co-occurrence for three instances: (1) when all three terms (query 1, query 2, and co-occurring term) are present in the same article, (2) when the term co-occurs only in articles with query 1, and (3) when the term co-occurs only in articles with query 2.

We will perform query in this module and explore its functionalities.

A) Performing Queries and Filtering Results

1) First, **click on the “Find Shared Terms”** tab and ensure that it is highlighted in white.



2) **Enter “Kinase” and “Neoplasm”** in the search bar. **Hit “Search”.**

Please note that this query may take a moment to run due to the scope of the query terms.



A total of 5,404 unique MeSH terms co-occur in the same articles as both “Kinase” and “Neoplasm”.

For the term “Antineoplastic Agents”, there are 30 articles in which all three terms co-occur (instance 1); 136 articles in which “Antineoplastic Agents” and “Kinase” co-occur (instance 2); 31155 articles in which “Antineoplastic Agents” and “Neoplasm” co-occur (instance 3).

As in the previous module, the table limit can be changed.

5) Likewise, the full list of co-occurring terms can be filtered for specific categories of MeSH subtypes. Scroll down the page. **In the “Filter” search bar, select “Diseases and Indications”**, which will include all shared A-terms.



A total of 1,033 unique MeSH A-terms shared between “Kinase” and “Neoplasm”.

It should be noted that a term can be shared between the two query terms regardless of whether the associated term has co-occurred in the same article as the other two terms (instance 1).

 **Scroll through the resultant term list until “0” article counts are observed in the “Same Article” category.**



Here it can be seen that “Glioblastoma”, for instance, has yet co-occur in the same article as the two query together (instance 1); however, it has co-occurred with “Kinase” (instance 2) and with “Neoplasm” (instance 3). Therefore, this term is shared between the two query terms.

B) Linking to Articles and Downloading Results

7) With Chemotext, the article in which all three terms co-occur (instance 1) can be directly accessed. **Click on the “7” article count for “Liver Neoplams”.**

Chemotext will provide the PMIDs or actual article title (when available) for all 7 associated articles. The associated article can be accessed, as in the “Find Connected Terms Module”.



It should be noted that if the user would like to access articles associated with instance 2 and/or instance 3, this query should be repeated in the “Find Connected Terms Module”.

III.) Path Search

In this module, complete text-based A-B-C connections can be made through co-occurring MeSH terms. The name of this module - “Path Search” – indicates that these A-B-C connections can be established through several “paths,” i.e. through multiple intermediary terms or through a single intermediary term.

A) Performing Queries and Filtering Results

1) First, **click on the “Path Search”** tab and ensure that it is highlighted in white.



2) **Enter “Kinase”** in the search bar and **Choose “Diseases and Indications” as the intermediary term.**

****

It can be seen that 1170 “Disease and Indications” terms co-occur with “Kinase”. These intermediary terms are ranked by article count.

3) Any number of intermediary terms can be selected to complete the path search, *i.e.* from 1 term to all 1170 terms. For the sake of simplicity, only one intermediary term will be used to complete the path. In this example, the path will be completed by finding all associated “Chemicals”. This will form a complete A-B-C connection.

**Click “Neoplasms”** and ensure there is a check mark in the box**. Then set “Chemicals”** as the final search.



4) **Click on Finish Search** to run the query. This will find all “Chemicals” that co-occur with “Neoplasms”. Thus, these “Chemicals” will be associated with “Kinase” through the intermediary “Neoplasm” term.



In total, there 11 143 chemicals associated with “Neoplasms” ranked by article count. This means that 11 143 unique “Kinase”-“Neoplasms”-“Chemicals” connections (A-B-C).

5) **Expand the table size in the “Set Table Limit” box and find “imatinib”**



There are 255 articles in which “imatinib” and “Neoplasms” co-occur, thereby completing a “Kinase”-“Neoplasms”-“imatinib” connection. This is a well-established A-B-C connection that Chemotext is capable of retrieving.

B) Linking to Articles and Downloading Results

1) As with all previous sections, the CSV file with final terms and article counts can be accessed through the **“View CSV” tab**. Additionally, CSV file with final terms, article counts, and associated PMIDs can be accessed through the **“CSV with pmids” tab**.

2) With Chemotext, the article in which both terms co-occur can be directly accessed. Click on the **“255” article count for “imatinib”.**

Chemotext will provide the PMIDs or actual article title (when available) for all 255 associated articles.