**Molecular pathways, pathway modelling and networks and network extension with drug related database information**

In this tutorial we will use Cytoscape and information from two different chemical compound databases for network extension to find possible drug targets within a pathway based network.

**Preparation:**

1. **Install Cytoscape**

Cytoscape is a free open source software for network creation and analysis. Download is available here: <https://cytoscape.org/>

1. **Download the linksets**

For this tutorial, you will also need 2 linksets:

DrugBank 4.2: <https://ndownloader.figshare.com/files/21623682?private_link=32aae0822ffdd1f5660b>

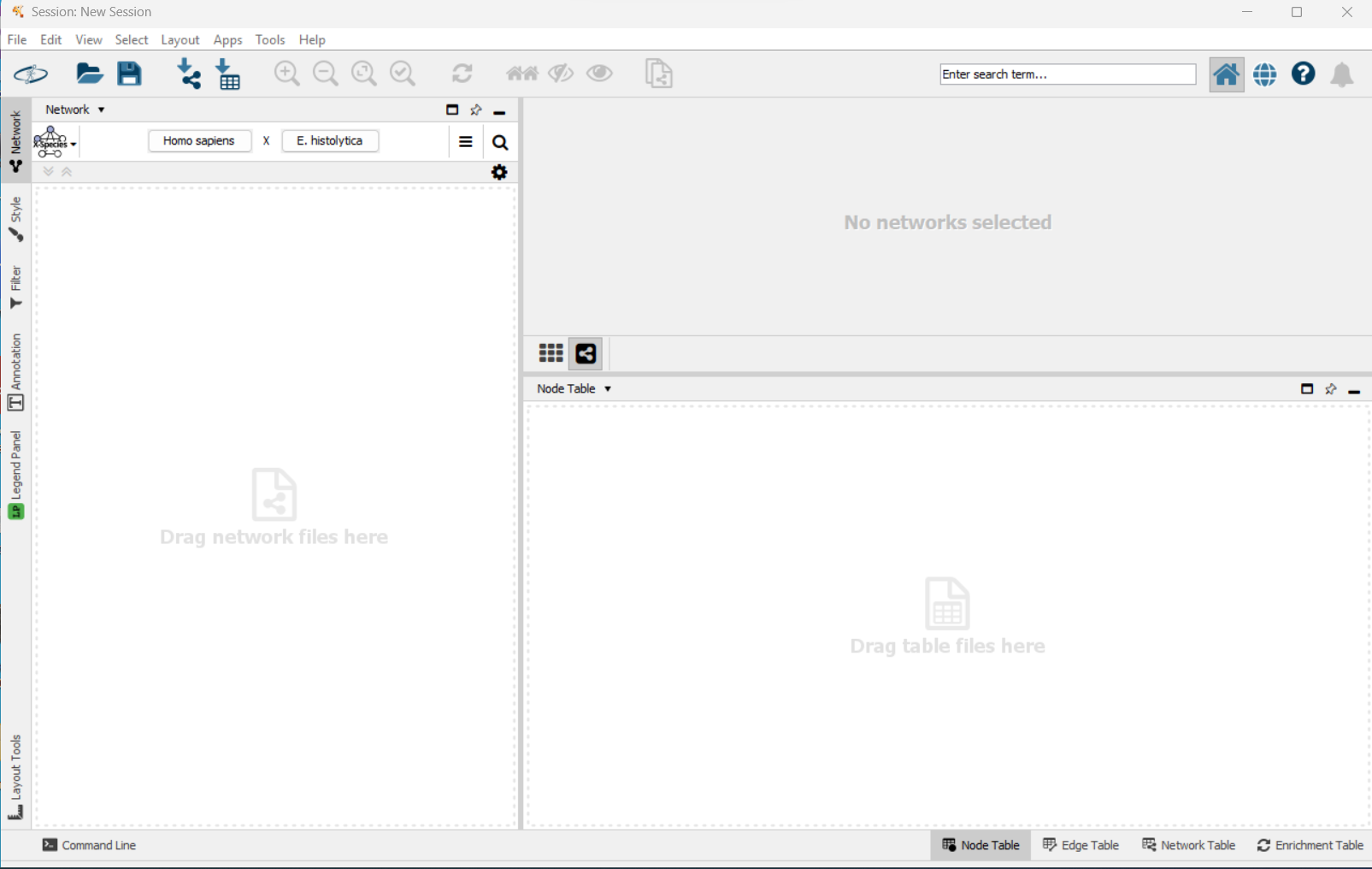
ChEMBL release 23: <https://ndownloader.figshare.com/files/21623691?private_link=6cf358aaaaf5adeecce9>

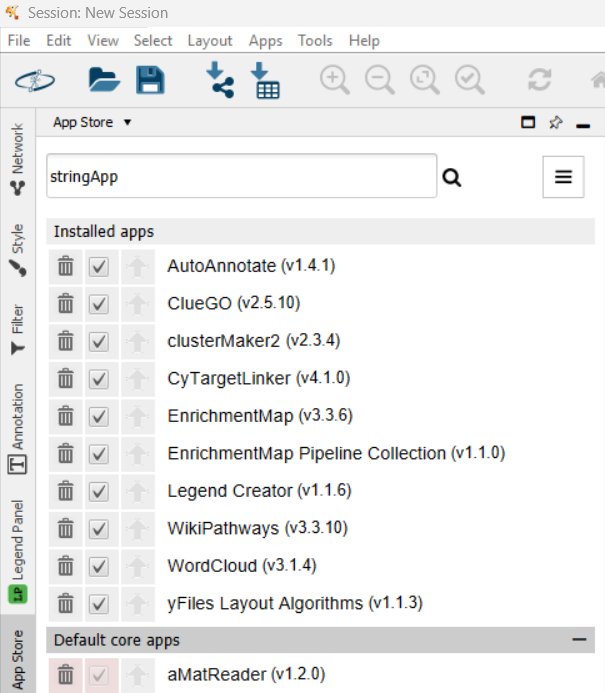
More linksets are available here: <https://cytargetlinker.github.io/pages/linksets>

Please download, unzip, and put them in your workshop folder where you can find them easily.

1. **Open Cytoscape and install the required apps**

The current version is 3.10.2, but other versions should work, too if they are not too old.



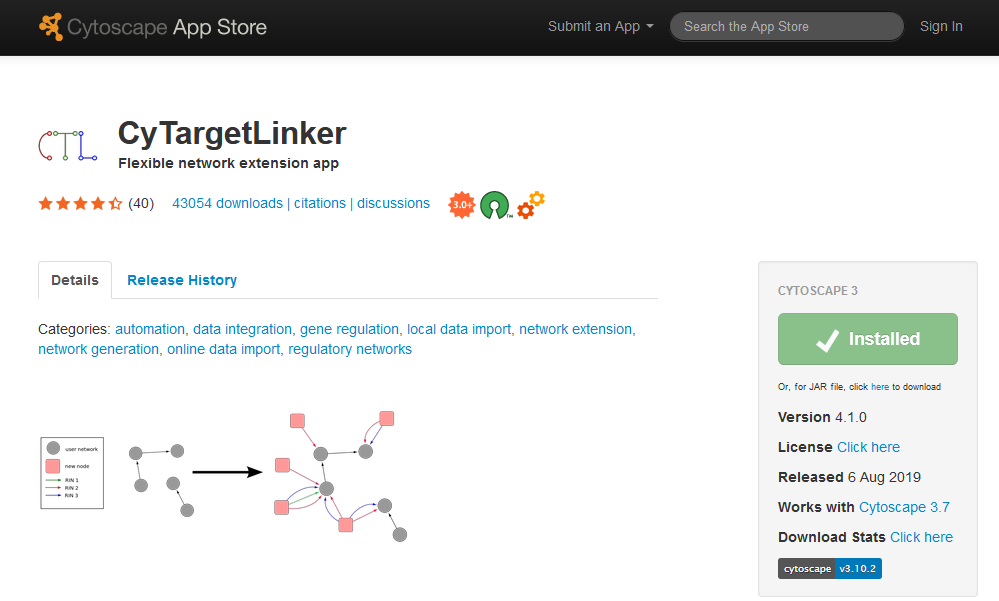


For this tutorial, we will need 2 apps: WikiPathways and CyTargetLinker.

Install them in Cytoscape via “Apps” – “App Store” – “Show App Store”.

<= This is an example picture of the App store with more apps than you need for this tutorial!

Search for the app names: **CyTargetLinker** and **WikiPathways**– the website with the online app store will open – select the correct one and click “Install”. For more layout options, we also recommend **yFiles Layout Algorithms**.

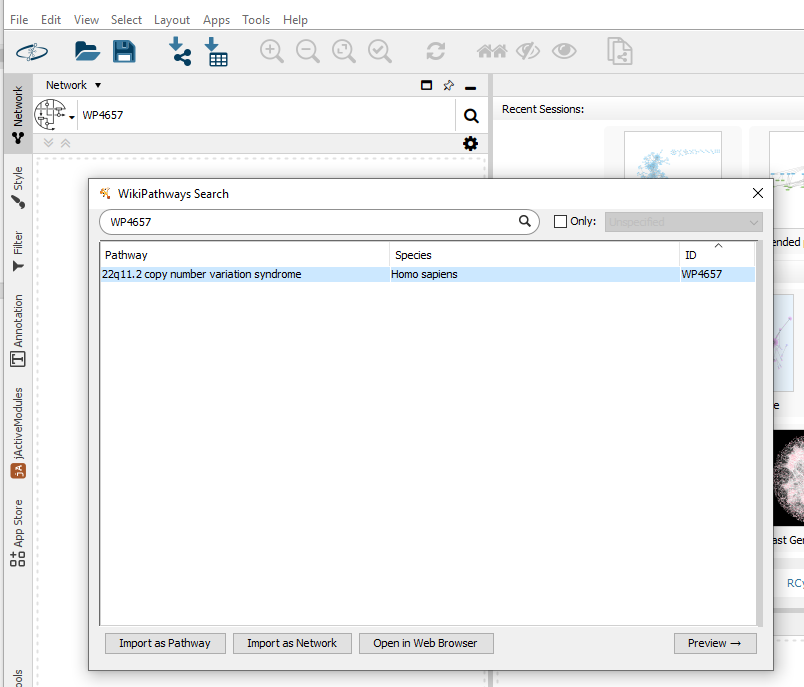


The new apps should become visible immediately in the app list.

# **Identifying drug targets from DrugBank within a molecular pathway**

In the first assignment, we are using a molecular pathway from WikiPathways and extend it with drug-target information from DrugBank.

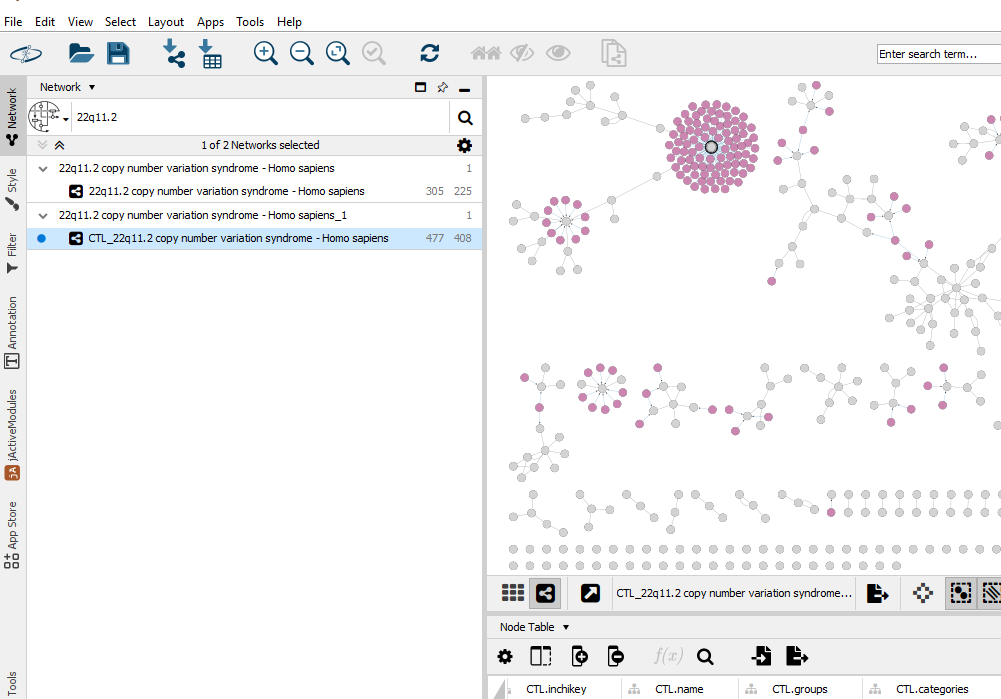
1. Use WikiPathways app to import the 22q11.2 copy number variation pathway WP4657 as a network
   1. Go to Network, select the WikiPathways query and search the 22q11.2 copy number variation pathway, either by text or by using the WikiPathways identifier - WP4657 in *Homo sapiens*.
   2. Now you can choose whether you want to import the pathway as pathway or as network. Try both and compare the two networks. What is different between the two?



1. Use the “network” version for further analysis.
2. Extend the pathway network with CyTargetLinker: Apps – CyTargetLinker – Extend Network
3. Select User Network: 22q11.2 copy number variation syndrome – Hs = the network you want to extend.
4. Select your network attribute: Ensembl – this is the column that contains identifiers that are recognized by CTL and the selected linkset.
5. Select Link Sets – Browse to the folder that contains the linksets. You won’t see the linkset files there, but CTL will find them if you give the correct folder.
6. Select direction – Linksets contain usually source-target information which in this case are drugs (=chemical compounds) <-> proteins(=drug targets). If both is selected, CTL will add drugs to proteins, and proteins to chemical compounds but you could select only one direction. In our case, the Ensembl column only contains Ensembl (=protein/geneproduct) identifiers so automatically only drugs will be added to proteins.

|  |  |
| --- | --- |
|  |  |

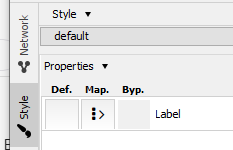
1. The next step asks which linksets should be used. Check drugbank4-2.xgmml.
2. When you hit “OK” the extension should start and the network should look like this:



**Q1: How many drugs have been added?**

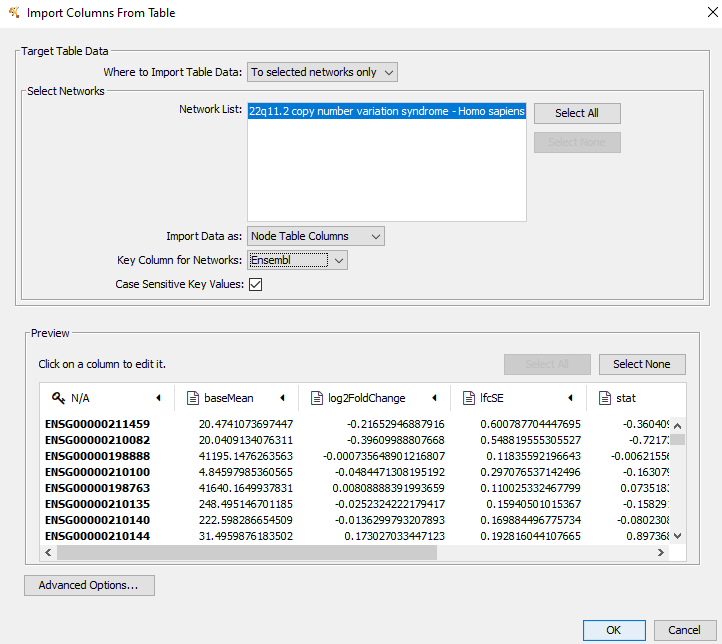
**Q2: Which proteins are the “hubs”? (= the proteins which can be targeted by most drugs?)**

**Q3: Do you recognize any drugs for 22q or schizophrenia treatment in here?**

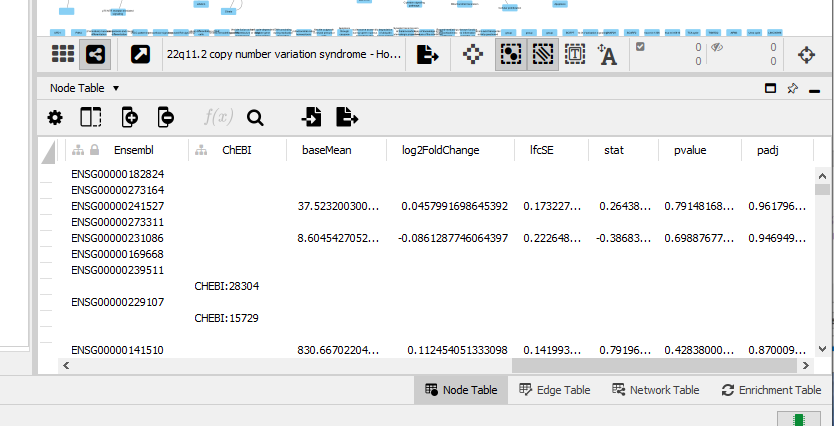


Now we need to work on the visualization of the network! The CTL addon applies an automatic visualization that shows the type of extension (chemical compounds = pink, intial network nodes = grey) which is not optimal for further investigation. You can change that by changing the style back to default and apply own visualization, e.g. by showing the expression levels of differentially expressed genes.

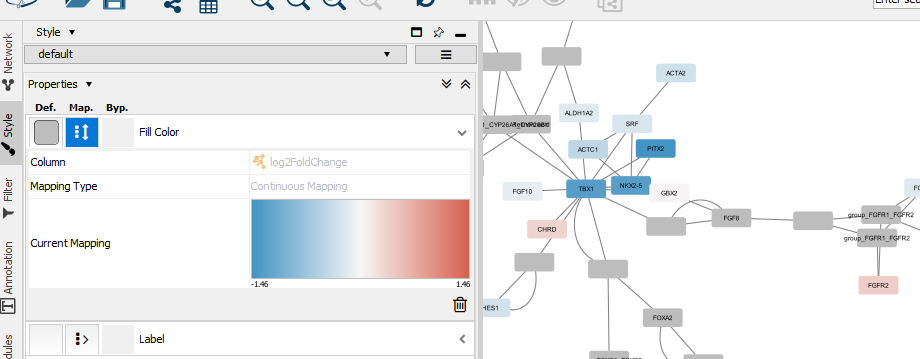
1. Import the DEG.csv table by and investigate if there are highly differentially expressed nodes here that can be prioritized for further investigation.
   1. “File” – “Import” – “Table from file”
   2. Navigate to the **DEG.csv** file
   3. Make sure to add the table “To selected networks only” and select your actual network with the CTL extension.
   4. The key column should be Ensembl to ensure matching identifiers in the table and the network.

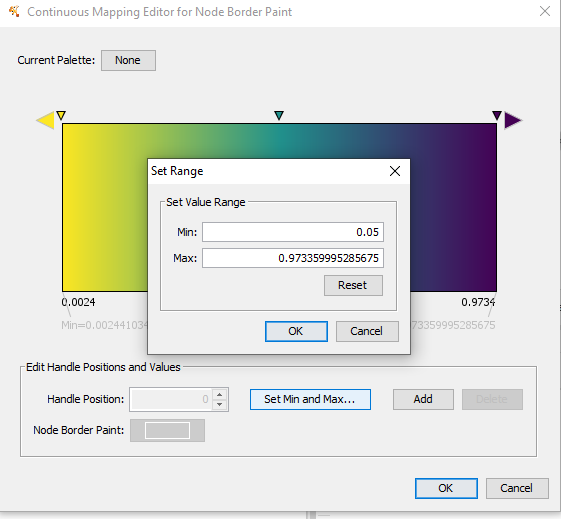


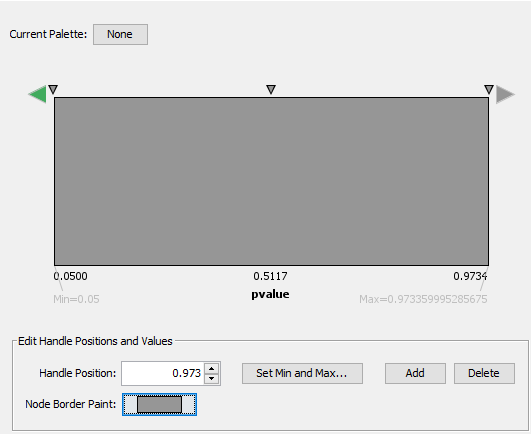
* 1. If the import was successful, additional columns with the expression data should show up on the right side of the node table.



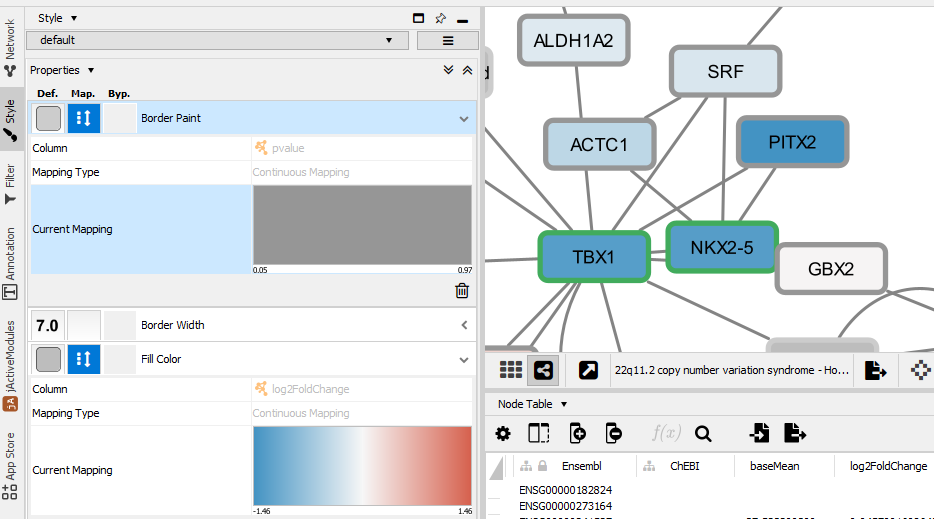
1. Change the gene node colours according to their log2FC
   1. If “Fill color” is not in your displayed list of properties you can find it by going to “Properties ↓”, “Paint” and select “Fill color”
   2. Change Def. (default) to grey
   3. Change Map. (mapping) to node:log2FoldChange and mapping type to continuous.



1. Change the node border to indicate if a DEG is significant or not.
   1. If not yet in the list of properties add “Border Paint” (also in “Paint”) and Border Width to the selection.
   2. Change the default value of Border Width to 7.
   3. For Border Paint, select pvalue for mapping and choose continuous mapping.
   4. Double click on the colorful icon to open the Editor to modify the color distribution to a clear rule that indicates: border paint = green for pvalue < 0.05, and grey for anything non-significant.
   5. Click on “Set Min and Max” and change Min to 0.05
   6. Change the colors by double-clicking on the triangles. The most left triangle indicating everything smaller than 0.05 should be green, all other triangles should be grey.



The network should look like this now:

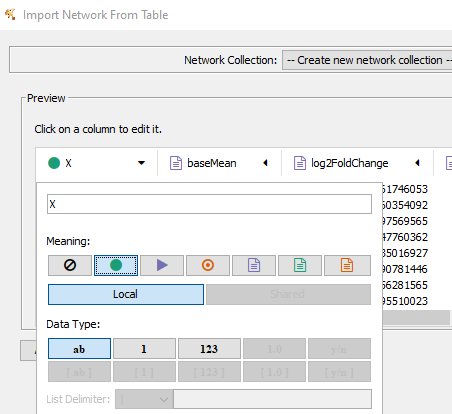


* 1. To highlight the difference between gene and drug nodes you can change the shape and color of the drug nodes.
     1. First, select “Shape” from properties in Style, choose discrete mapping and select a triangle for drugs.
     2. Second, use “Filter” to select all nodes with the CTL.type drug, in Style use bypass to select a different color.

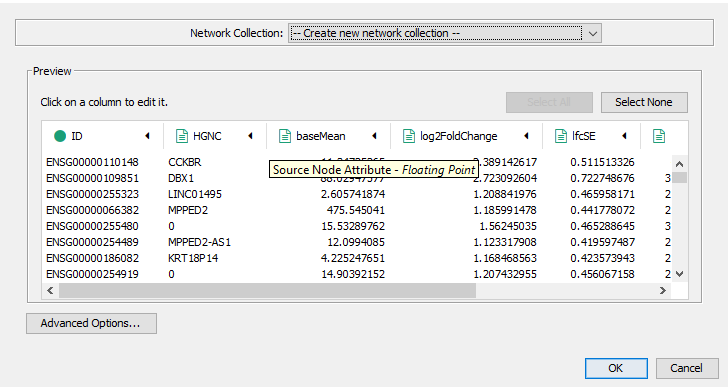
# **Identifying drug targets from ChEMBL from a list of genes**

In the second assignment, we are using a list of differentially expressed genes and extend it with information from ChEMBL database.

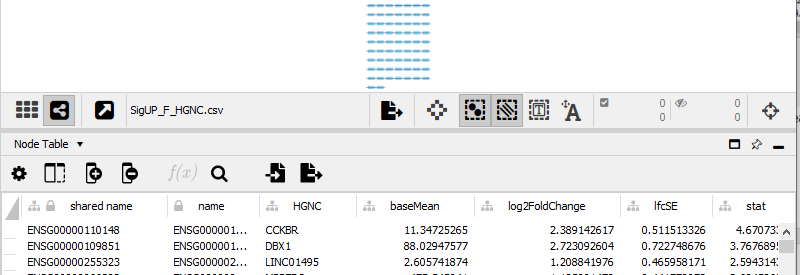
1. Import a new network based on the significantly upregulated genes: “File” – “Import” – “Network from file” – navigate to the file **SigUP.csv**.

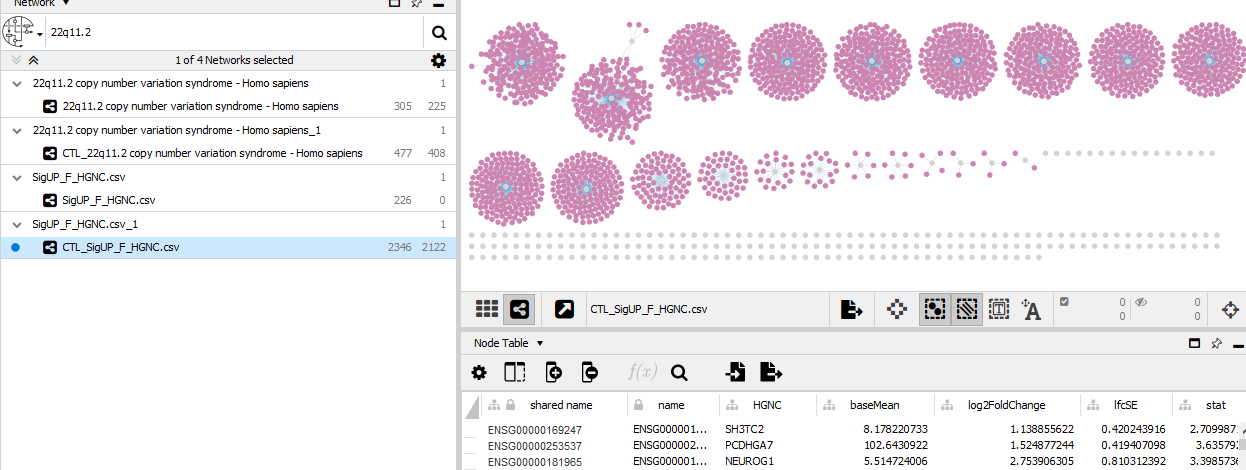


1. Name the first column ID as the column header must not be empty and select the green dot (= source node). Data type = string (= ab)
2. Select source node attribute (the green text icon) for all the other nodes. HGNC should be string, the others floating point (1.0) data type.



1. The warning “No edges will be created…” – YES
2. You will get a list of unconnected nodes (as intended)



1. For better human readability, change the node names from Ensembl IDs to HGNC. Go to Style, “Label” and select the HGNC column and mapping type PassThrough. This mapping type will take the content of the column field directly.
2. **Optional:** If you like you can STRINGify the nodes in order to find protein-protein interactions and to form a network. For this, you need the STRING app. It also adds information on average expression levels in different tissues and localization in the cell.
3. Extend the pathway network with CyTargetLinker: Apps – CyTargetLinker – Extend Network
4. Select the correct network (SigUP.csv), use shared name as attribute (ID column was automatically translated to shared name), and select your folder that contains the linksets and select there the correct (chembl\_23\_hsa…) linkset.
5. With >2000 nodes this network may cause already memory problems with some laptops!
6. The extended network should look like this: 

Again, feel free to modify the visualization of the network as described in the previous assignment!

The ChEMBL linkset is quite large, the chemical names are not included and for many of them also not available but it includes a large variety of chemical compounds with the (predicted) potential to bind to selected proteins and may qualify as new drug.