Finding drug targets using network extension

In this tutorial we will use two different network extension workflows to find possible drug targets within a network.

**Preparation:**

1. **Download the linksets**

For this, you need 2 linksets from <https://cytargetlinker.github.io/pages/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>

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

1. **Install the required apps in Cytoscape**

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

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

Search for the app names – the website with the online app store will open – select the correct one and click “Install”.

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

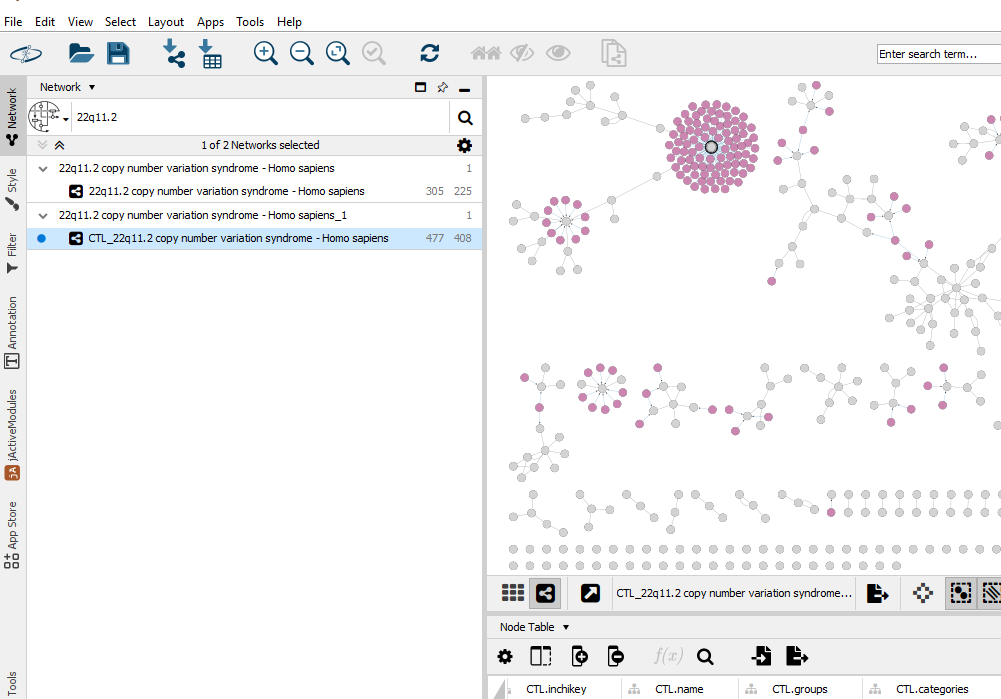
# Extension with DrugBank

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

1. Open Cytoscape
2. Use WikiPathways app to import the 22q11.2 copy number variation pathway WP4657 as a network
3. Extend the pathway network with CyTargetLinker: Apps – CyTargetLinker – Extend Network
4. Select User Network: 22q11.2 copy number variation syndrome – Hs = the network you want to extend.
5. Select your network attribute: Ensembl – this is the column that contains identifiers that are recognized by CTL and the selected linkset.
6. Select Link Sets – Browse to the folder that contains the linksets. You won’t see them there, but CTL will find them if you give the correct folder.
7. 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:



Feel free to modify the visualization of the network! CTL addon applies an automatic visualization that is not optimal for further investigation. You can also upload the DEG\_F.csv table and investigate if there are highly differentially expressed nodes here that can be targeted.

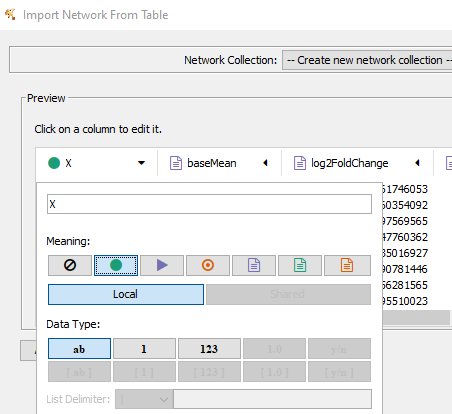
**Q1: How many drugs have been added?**

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

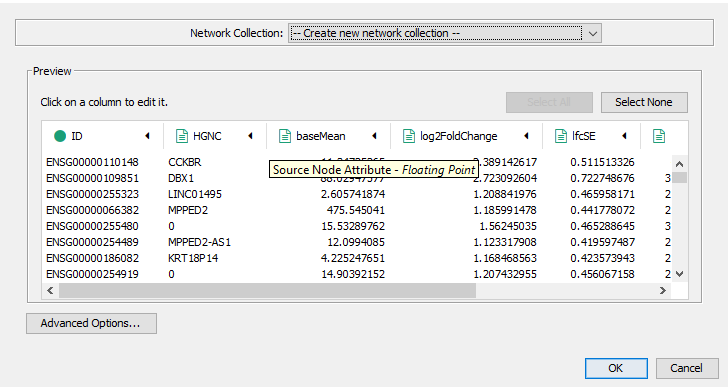
**Q3: Do you recognize any drugs for 22q or schizophrenia treatment in here?** (Hint: play with sorting and filtering the node table)

# Extension with ChEMBL

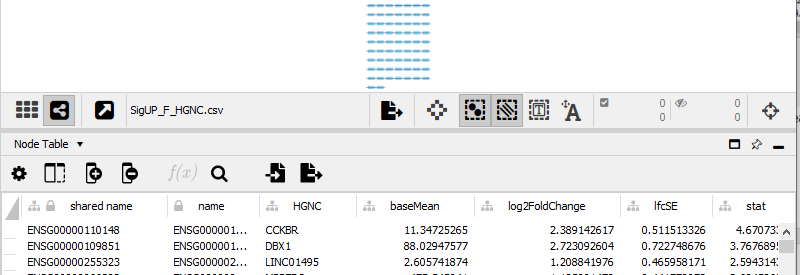
1. Import a new network based on the upregulated genes: “File” – “Import” – “Network from file” – navigate to the file SigUP\_F\_HGNC.csv.

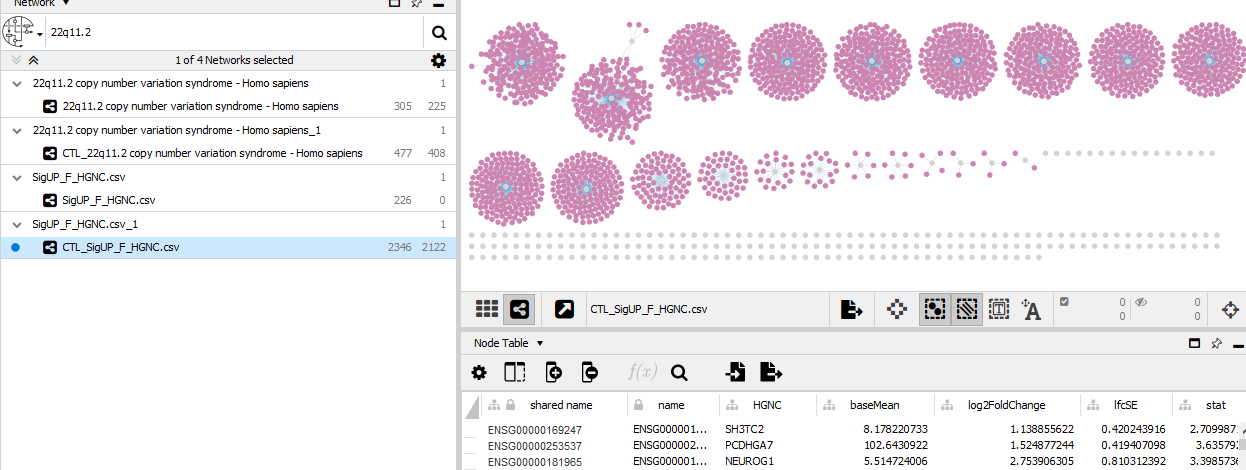


1. Name the first column ID as the column header must not be empty and select the green dot (= source node). Data type = string (= ab)
2. Select source node attribute (the green text icon) for all the other nodes. HGNC should be string, the others floating point (1.0) data type.



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



1. If you like you can STRINGify the nodes in order to find protein-protein interactions and to form a network.
2. Extend the pathway network with CyTargetLinker: Apps – CyTargetLinker – Extend Network
3. Select the correct network (SigUP\_F\_HGNC.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.
4. The extended network should look like this: 

Again, feel free to modify the visualization of the network! CTL addon applies an automatic visualization that is not optimal for further investigation. You could e.g. visualize the logFC and significance on the protein nodes.

The ChEMBL linkset is quite large, the chemical names are not included. You could export the node table, do identifier mapping from ChEMBL to chemical name and re-import the table in order to get human readable chemical compound names.

With >2000 nodes this network may cause already memory problems with older laptops!