Integrated assignment
Part2

Background:
Stomach cancer or gastric cancer is cancer developing from the lining of the stomach. Early symptoms may include heartburn, upper abdominal pain, nausea and loss of appetite. Later signs and symptoms may include weight loss, yellow skin, vomiting, difficulty swallowing and blood in the stool. The most common cause is infection by the bacteria Helicobacter pylori, which accounts for more than 60% of cases. Certain types of *H. pylori* have greater risks than others. Other common causes include eating pickled vegetables and smoking.

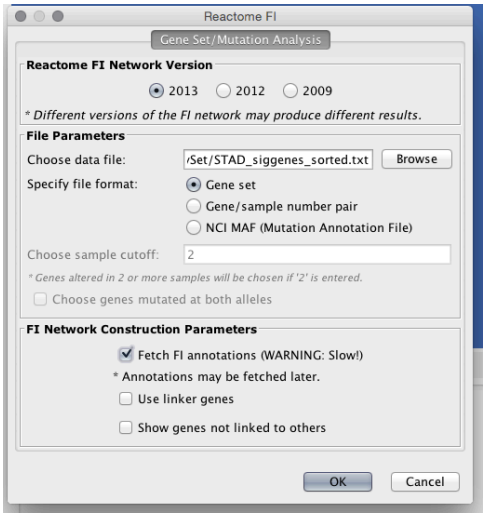
MutSig – is a mutation signal-processing tool created by Broad Institute. It estimates the significance of the gene mutation rate based on abundances of the mutations, clustering of the mutations in hotspots and conservation of the mutated positions.

The gene list for this assignment is the output from MutSig run based on Stomach Adenocarcinoma somatic mutations found in ~300 samples. It is publicly available through TCGA portal.

Files provided: STAD_MutSig.txt

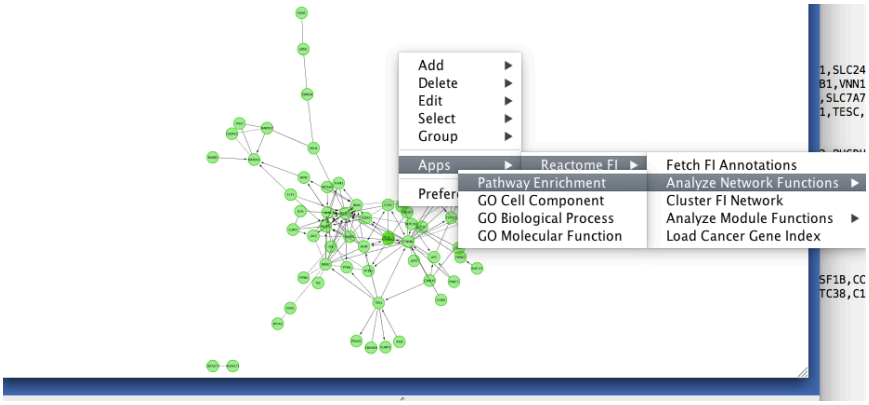
Goal: familiarize yourself with ReactomeFIViz, learn how to do a pathway and network-based analysis.

1. Open Cytoscape;
2. Choose Apps -> Reactome FI -> Gene set/mutation analysis
3. Upload GastricCancer_mutsig.txt and build a network without linkers:



4. How many genes (nodes) and interactions (edges) are in the network (Hint: check control panel)?

5. Try different layouts (for example: Layout -> yFiles Layout -> Circular). Which do you like the most?
6. Run Pathway annotation (Hint: right click on the network panel and follow the path showed below)



7. What is the most significant pathway? How many genes from our set belong to this pathway?

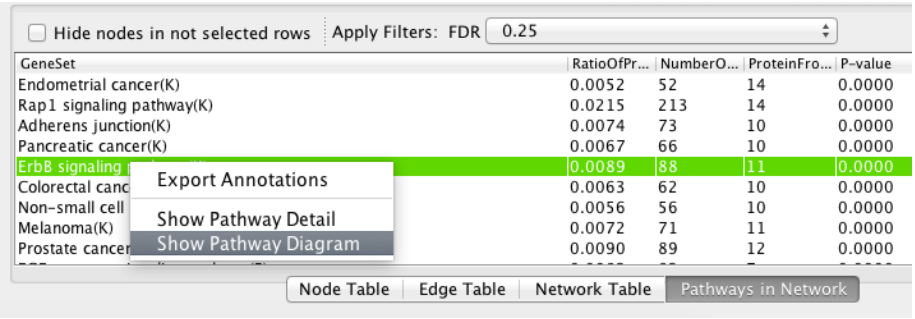
8. Run enrichment analysis for GO:Molecular Function annotations. What is the most significant molecular function for this set?

9. Click on “ErbB signaling” pathway. Genes in the network that belong to this pathway will be highlighted with yellow. By what knowledgebase this pathway was curated? How many genes does this pathway contain?

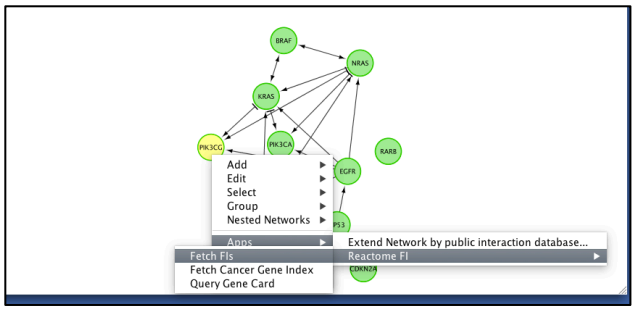
10. Hide all genes that DO NOT belong to “ErbB signaling” (Hint: Select -> Node -> Invert node selection, Select -> Node -> Hide selected nodes) and create an ErbB-specific subnetwork. Have a close look at the genes.

11. What type of interaction is between ERRB2 and PIK3CG (Hint: highlight the queried interaction, go to Table Panel and open Edge Table tab. Find FI annotation)

12. Highlight “ErbB signaling” in the Table Panel, right click and choose “Show diagram pathway”. A new browser page with the KEGG diagram will appear. Genes of interest will be highlighted with red.



13. How many interactive partners does PIK3CG have in gastric cancer-related network? How many in the whole FI network? (Hint: see a screenshot below):

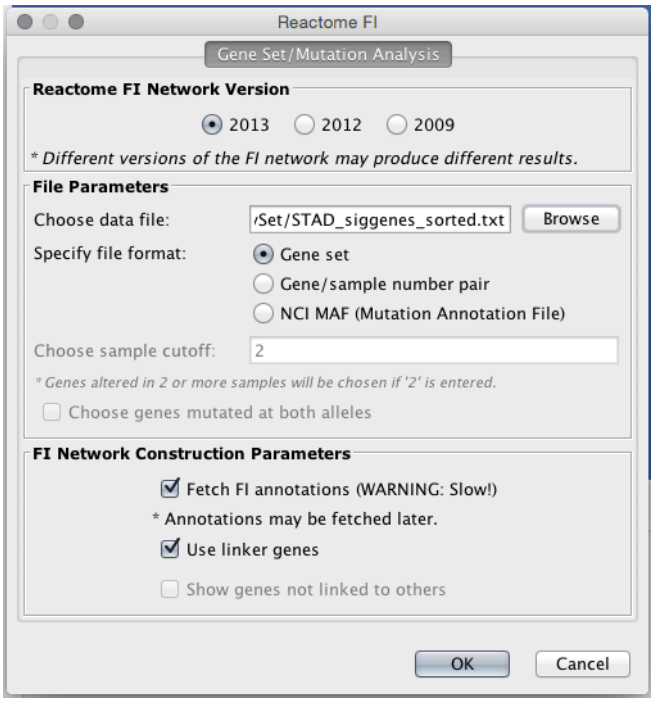


14. Run network clustering (Hint: Right click on the network panel, choose Apps -> reactome FI -> Cluster FI Network). Nodes colored with the same color belong to the same module. Modules are mutually exclusive meaning that a gene can belong only to one module. How many modules does this network have? Do all modules have similar size? How many nodes does the biggest module have?

15. The most interesting aspect of the network analysis is a module enrichment test (Hint: Right click on the network panel, choose Apps -> reactome FI -> Analyze Module Functions -> Pathway enrichment). Enrichment analysis will be done for each module separately (or you can determine a module size cutoff before analysis starts – usually it is a good practice to ignore very small modules). Don’t forget to adjust FDR (on Table Panel) to something more significant like 0.05.

16. Switch to tab “Pathways in Network” in Table Panel. Highlight “ErbB signaling pathway”. Are the genes from this pathway belonging to one or several modules? Why do you think it happened?

17. Re-build your subnetwork using linkers (you can save or destroy your previous work):



18. What does “linker” mean?

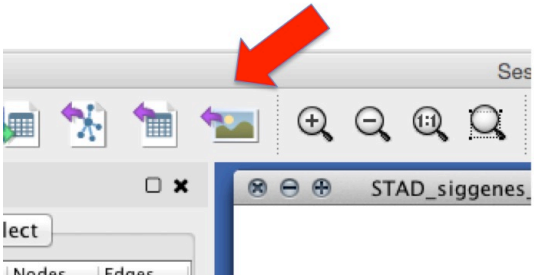
19. How many nodes and edges are in the network?

20. Run (1) pathway enrichment analysis (adjust FDR to 0.01); (2) clustering and (3) module based pathway enrichment analysis (adjust FDR to 0.01).
21. *OPTIONAL:* Does any pathway reflect the fact that gastric cancer is caused by bacterial infection (Helicobacter pylori)?

22. *OPTIONAL:* Find a quick way to calculate how many linkers are in the network.
23. Are there any modules that are not enriched in any pathways? Save genes from one of these modules in .txt file. We will use in GeneMANIA.

24. Using search box (right upper corner of the Cytoscape) find gene EP300. What module does it belong to?

25. Save your network as .jpeg or .pdf. Choose resolution 300 dpi and zoom in 300%. Open your saved image. Zoom in. Does the quality of the image suit your needs (poster, publication, .ppt presentation)?



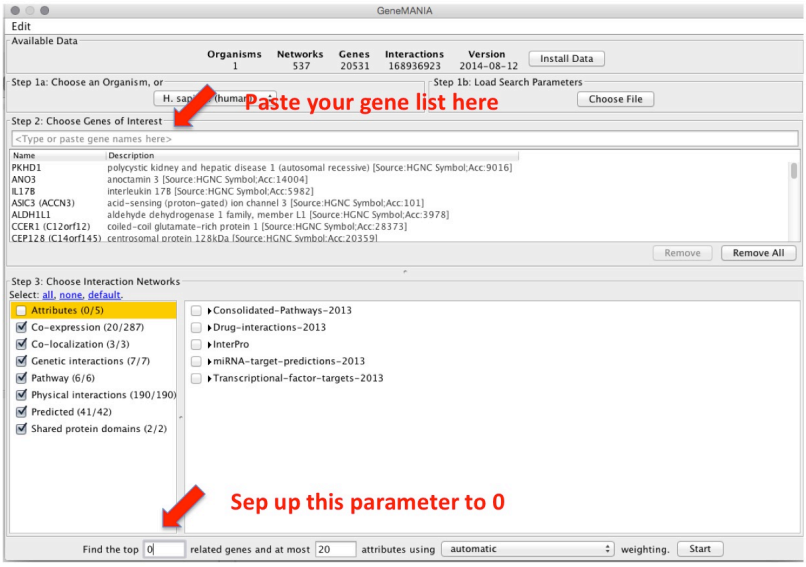
GENEMANIA

The major function of the GeneMania is a gene function prediction based on association data like protein and genetic interactions, pathways, co-expression, co-localization and protein domain similarity.

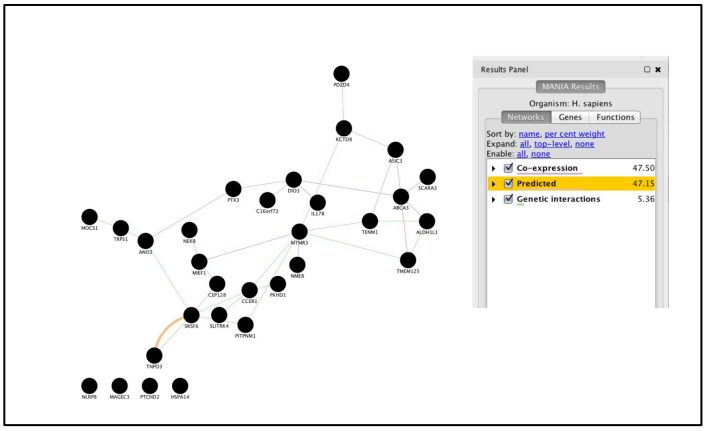
Goal: get familiar with GeneMania Cytoscape app.

Task: Functional Interaction Network (FIN) contains ~50% of all human genes. So, a good portion of the genes of interest might be excluded from the network-based analysis just because these genes have not been curated so far. MutSig predicted 162 genes to be significant in gastric cancer. Of those 132 are in FIN. Let investigate 30 excluded from FIN genes using GeneMANIA.

1. Open Cytoscape;
2. Go App -> GeneMANIA -> Search...
3. Copy top 30 genes from GasticCancer_mutsig.txt file and paste them into "Step2: Choose genes of interest" window. Usually when you are looking at a relatively large set of genes using GeneMANIA, adding related genes is not very informative function. Let us set it to 0. Press "START".

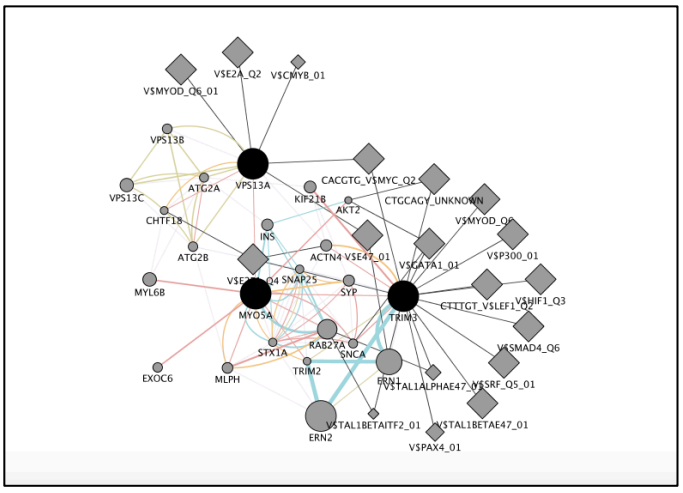


4. What kind of conclusions can you make about functional relationships between presented genes based on the network?



5. Go back to GeneMANIA input panel and add an attribute: Transcriptional-Factor-Targets-2013. Re-build a network. Go to Results Panel and find which transcriptional factor (TF) has the highest score. How many genes from our initial gene list this TF is regulating?

6. Play with other attributes. Check how these attributes are refining your knowledge about unknown set of genes.
7. Upload to GeneMANIA a saved gene list from your previous FI network-based analysis (see above paragraph 23). Use different GeneMANIA options to discover their functions and relationships (build a network without and with top related genes, add attribute(s), find a TF that regulate at least two out of three genes of interest). Save your final network as .jpeg.



HOPE WHAT YOU HAVE LEARNED TODAY
WILL BE USEFUL FOR YOUR RESEARCH.