BGK2007 - Blok 2.5

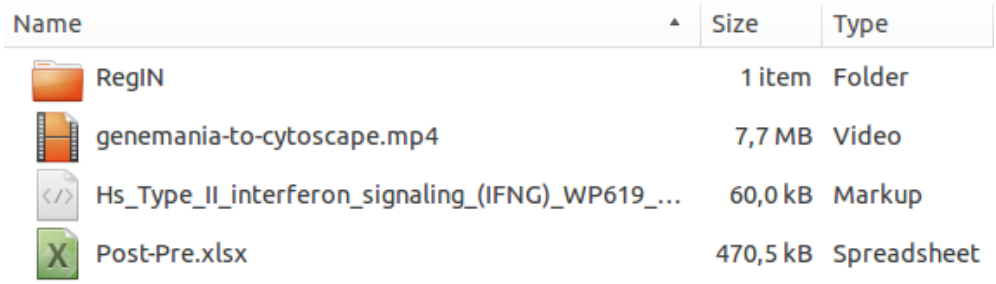
Practical Session 9: Network Biology
Date: April 29th, 2015

In this session you will go through some basic applications in network biology. You will learn

1. Where to find biological interaction data
2. How to use Cytoscape
3. How to visualize experimental data on a network
4. How to open a pathway as a network
5. How to extend a network with regulatory information

Preparation:

Go to Eleum and download the **network-biology-practical.zip** file and save it in the Downloads folder and unzip it (right click → extract all). This folder will now contain all the files you require to perform the practical. The folder should contain the following files:



Part 1: Learn how to use Cytoscape

We will go through this part together with you to show you how Cytoscape works.

Cytoscape (www.cytoscape.org) is a popular and commonly used open source tool for network biology. The program has lots of different functionalities making it possible to perform a variety of network analysis, from the basics to advanced analysis. Since it is as open community project, many different researchers and developers contribute to Cytoscape. This is demonstrated by the wide variety of Cytoscape apps.
For the following two parts you will need different Cytoscape functionality that is explained in one of their tutorials online. You don't need to go through the tutorials when you followed the introduction session.

http://opentutorials.cgl.ucsf.edu/index.php/Tutorial:Introduction_to_Cytoscape_3.1-part2

Part 2: Using networks to analyze transcriptomics datasets

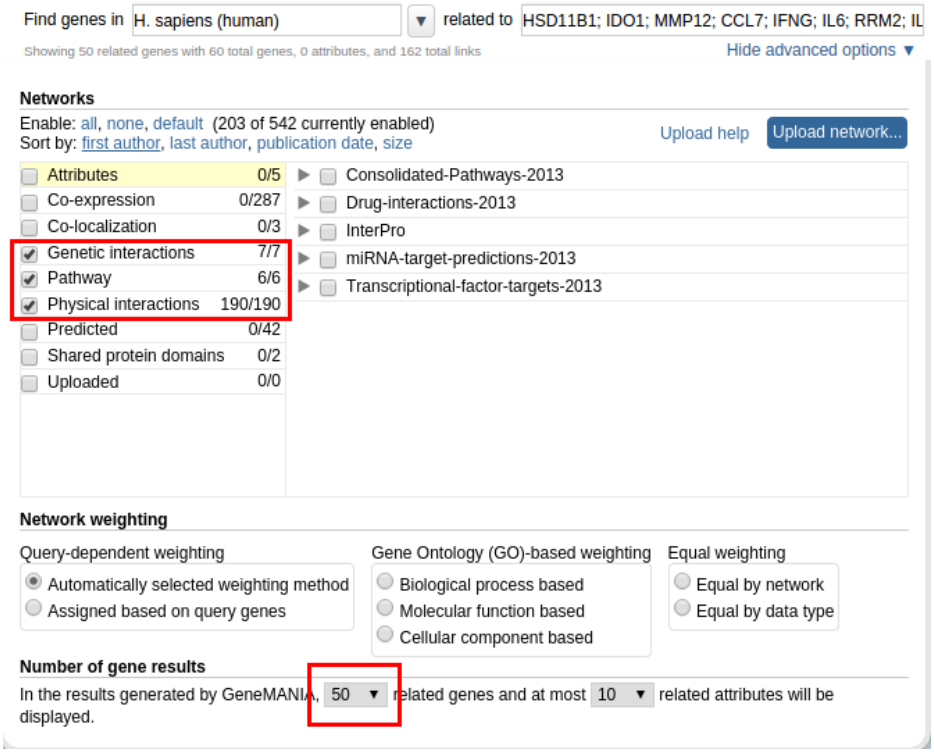
n this session we will continue using the dataset (**Post-Pre.xlsx**) that you analyzed in previous practicals on gene expression data and pathway analysis. You can find the file in the previously downloaded *network-biology-practical folder* (see preparation!).

We will use of the statistical results of the dataset from the paper "Gene Expression Patterns Induced by HPV-16 L1 Virus-Like Particles in Leukocytes from Vaccine Recipients" by García-Piñeres *et al* published in The Journal of Immunology in 2008 (see paper on EleUM or find it in Pubmed). We compare gene expression before and after vaccination.

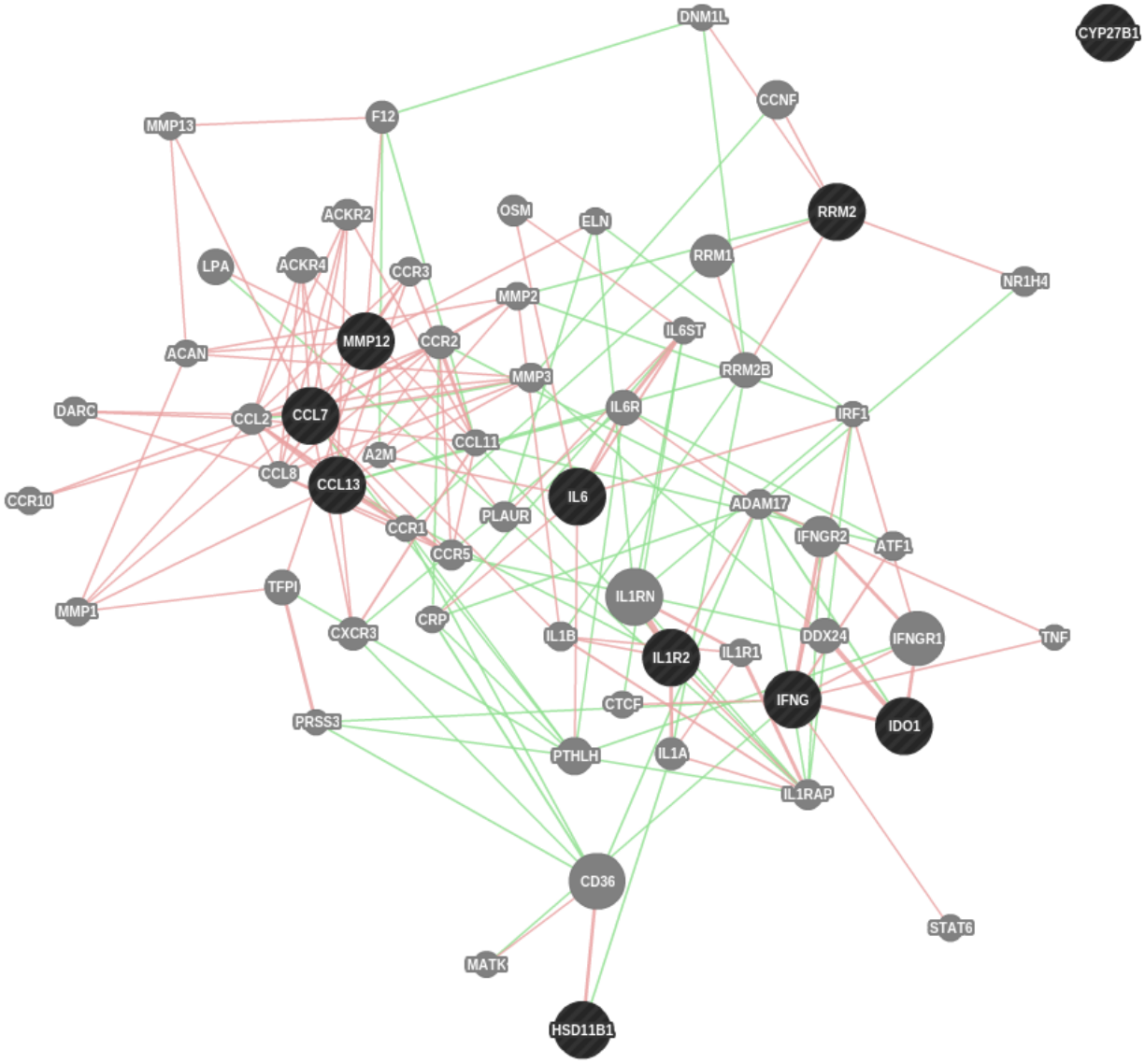
2a) Select the top 10 genes that are significantly up-regulated and create a network
Tip: in Excel go to Data → Sort → Sort by logfc_ex from largest to smallest). Make sure they are significant (p-value < 0.05).

ENSG_ID	logFC	P.Value	external_gene_id
ENSG00000117594	3.0912129947	4.78E-06	HSD11B1
ENSG00000131203	2.5377004216	0.0001980429	IDO1
ENSG00000110347	2.3434771652	0.0021228694	MMP12
ENSG00000108688	2.3215796167	0.0013875429	CCL7
ENSG00000111537	2.2025575463	9.19E-06	IFNG
ENSG00000136244	2.1844767992	6.16E-05	IL6
ENSG00000171848	2.1098896644	0.0007871284	RRM2
ENSG00000115590	1.9133764955	0.0023539543	IL1R2
ENSG00000181374	1.8166074422	0.011714916	CCL13
ENSG00000111012	1.7673872689	0.0026100875	CYP27B1

In the next step you will further investigate how the top 10 up-regulated genes are connected with each other. You will use GeneMania (www.genemania.org) to find the interactions between the genes.
Copy the gene symbols of the genes (external_gene_id) in answer 2a and use them as the query on GeneMania. Click on 'Show advanced options'. Select 'Genetic interactions, Pathway and Physical Interactions'. Make sure that you increase the number of related genes to 50.

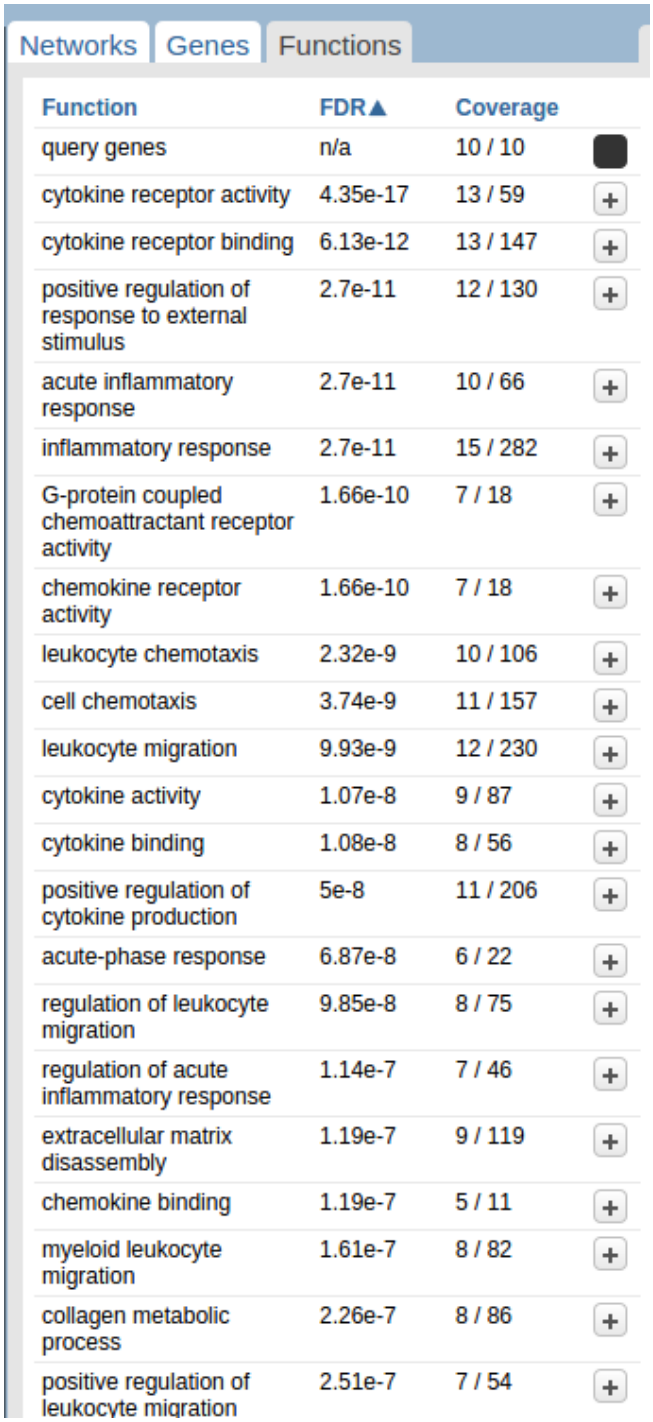


b) Make a screenshot of the network. Are all nodes connected?



No, CYP27B1 is not connected with the rest of the nodes.

c) In GeneMania you can also check in which biological processes the genes in the network are involved. Open the "Functions" tab and study the terms over-represented in the network. In what processes are the genes in this network involved?

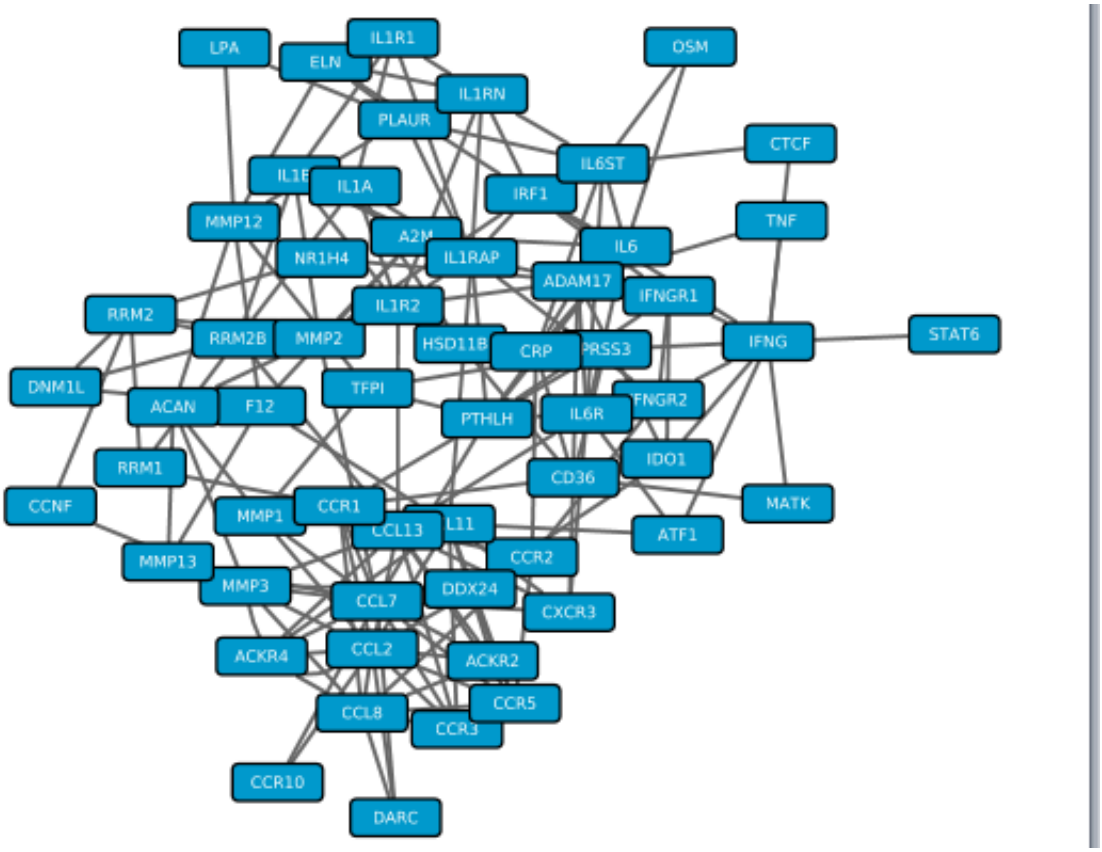


Networks | Genes | Functions

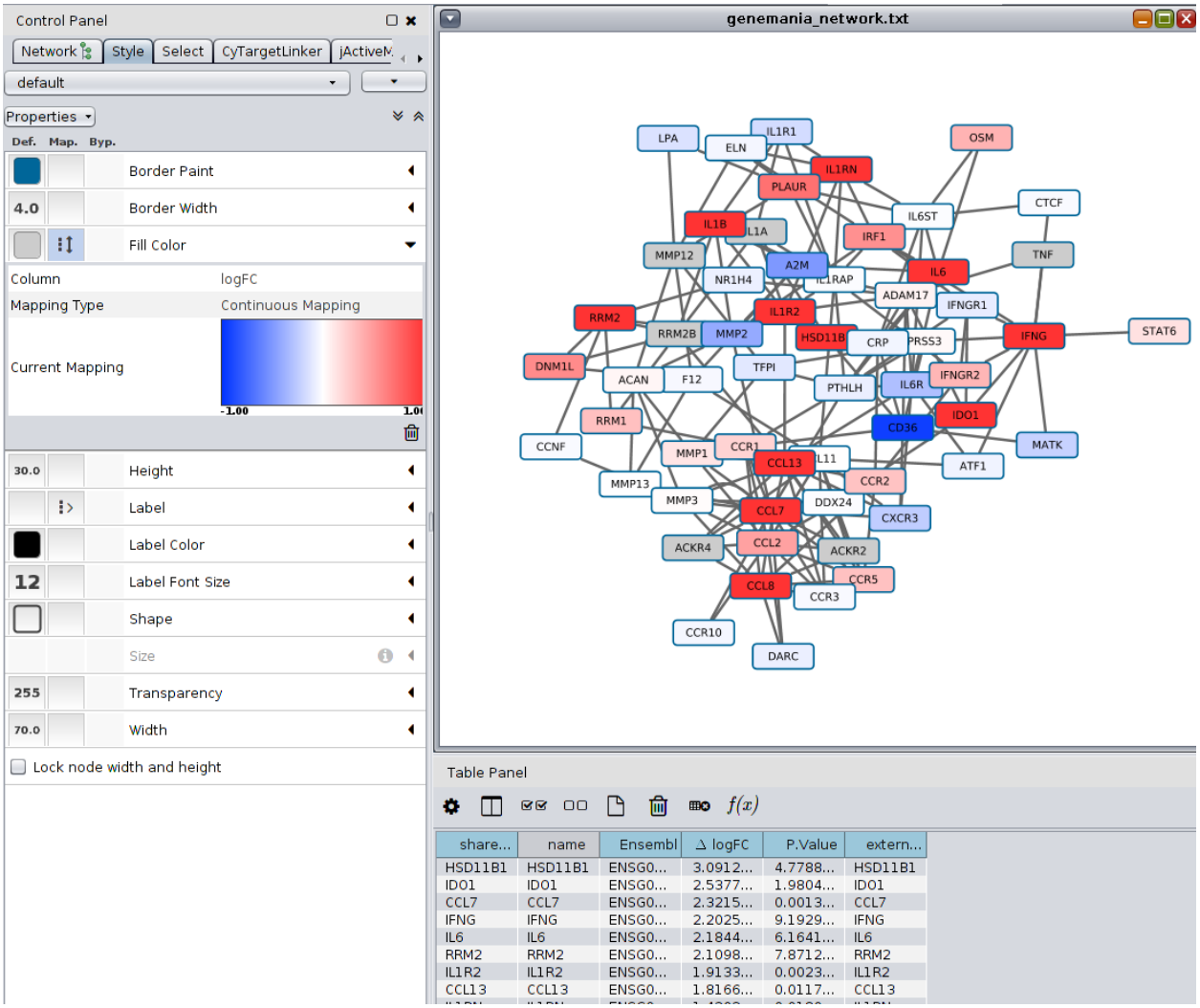
Function	FDR▲	Coverage
query genes	n/a	10 / 10
cytokine receptor activity	4.35e-17	13 / 59
cytokine receptor binding	6.13e-12	13 / 147
positive regulation of response to external stimulus	2.7e-11	12 / 130
acute inflammatory response	2.7e-11	10 / 66
inflammatory response	2.7e-11	15 / 282
G-protein coupled chemoattractant receptor activity	1.66e-10	7 / 18
chemokine receptor activity	1.66e-10	7 / 18
leukocyte chemotaxis	2.32e-9	10 / 106
cell chemotaxis	3.74e-9	11 / 157
leukocyte migration	9.93e-9	12 / 230
cytokine activity	1.07e-8	9 / 87
cytokine binding	1.08e-8	8 / 56
positive regulation of cytokine production	5e-8	11 / 206
acute-phase response	6.87e-8	6 / 22
regulation of leukocyte migration	9.85e-8	8 / 75
regulation of acute inflammatory response	1.14e-7	7 / 46
extracellular matrix disassembly	1.19e-7	9 / 119
chemokine binding	1.19e-7	5 / 11
myeloid leukocyte migration	1.61e-7	8 / 82
collagen metabolic process	2.26e-7	8 / 86
positive regulation of leukocyte migration	2.51e-7	7 / 54

d) Open network in Cytoscape and make a screenshot. How many nodes and edges are in the network?

In this step, we need to download the network from GeneMania and open it in Cytoscape. We prepared a video that will go through all the steps that you can follow. The **video** is located in the *network-biology-practical folder* → ***genemania-to-cytoscape.mp4***).



e) Visualize the gene expression data on the network (use a continuous mapper for the logFC for fill color of the nodes). You will see all the significantly up-regulated genes but what about the closest neighbours that were added by GeneMania? Are they up- or down-regulated or not changed between before and after vaccination?
Hint: Import Table → Post-Pre.xlsx → map ENSG_ID column in the file to Ensembl column in the network table → Use continuous mapping for logFC for node fill color.



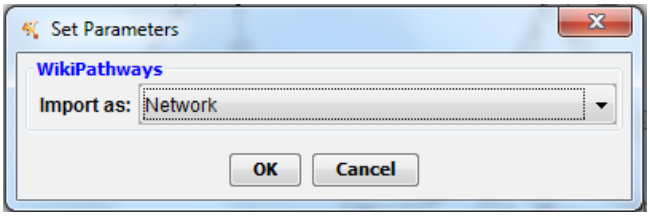
You can clearly see the 10 top up-regulated genes (in red). Most of the closest neighbors are not changed but there are some that are up- (CCL8, PLAUR) or down-regulated (CD36, A2M). There is no gene that is significantly down-regulated (logFC < -1) in the neighborhood of the up-regulated genes. (Tip: also check the table panel and sort by logFC)

Part 3: Biological pathways as networks

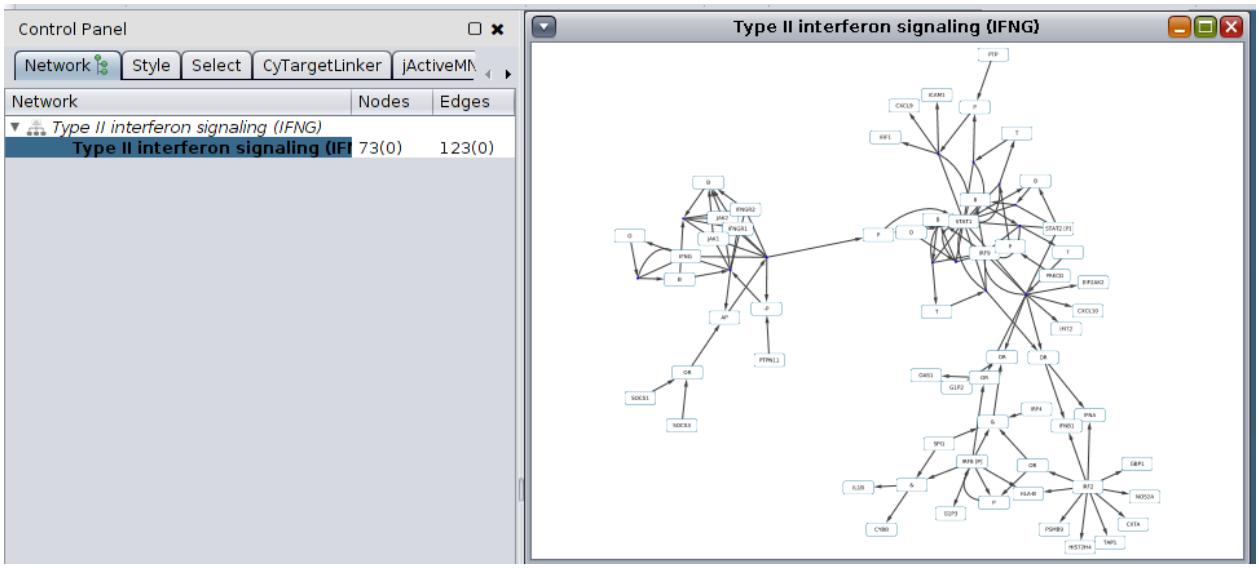
In this part, we are going to open a pathway from WikiPathways as a network and extend it with regulatory interactions. We will use two different Cytoscape apps: WikiPathways and CyTargetLinker. In the computer rooms both apps are pre-installed.
In the pathway analysis part two practical sessions ago, the pathway with the highest Z-Score was the "Type II interferon signaling (IFNG)" pathway. You can find it in the downloaded *network-biology-practical folder:*
Hs_Type_II_interferon_signaling_(IFNG)_WP619_71168.gpml.

In Cytoscape:

1. File → Import → Network → File...
2. Select Hs_Type_II_interferon_signaling_(IFNG)_WP619_71168.gpml
3. Make sure you select "Network" in the first properties dialog:



a) How many nodes and edges are present in the pathway?

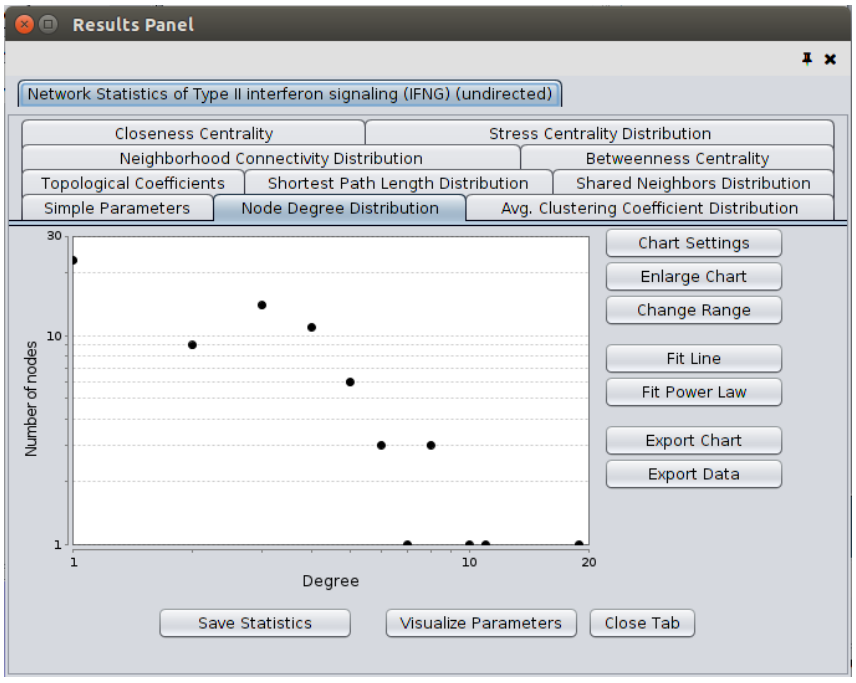


73 nodes and 123 edges.

b) Look at the node degree distribution. What does the node degree tell you?

Hint: Check lecture for explanation of node degree.

Use the NetworkAnalyzer to calculate the different network properties that were discussed during the lecture. Go to Tools → NetworkAnalyzer → Analyzer Network. Treat the network as undirected.

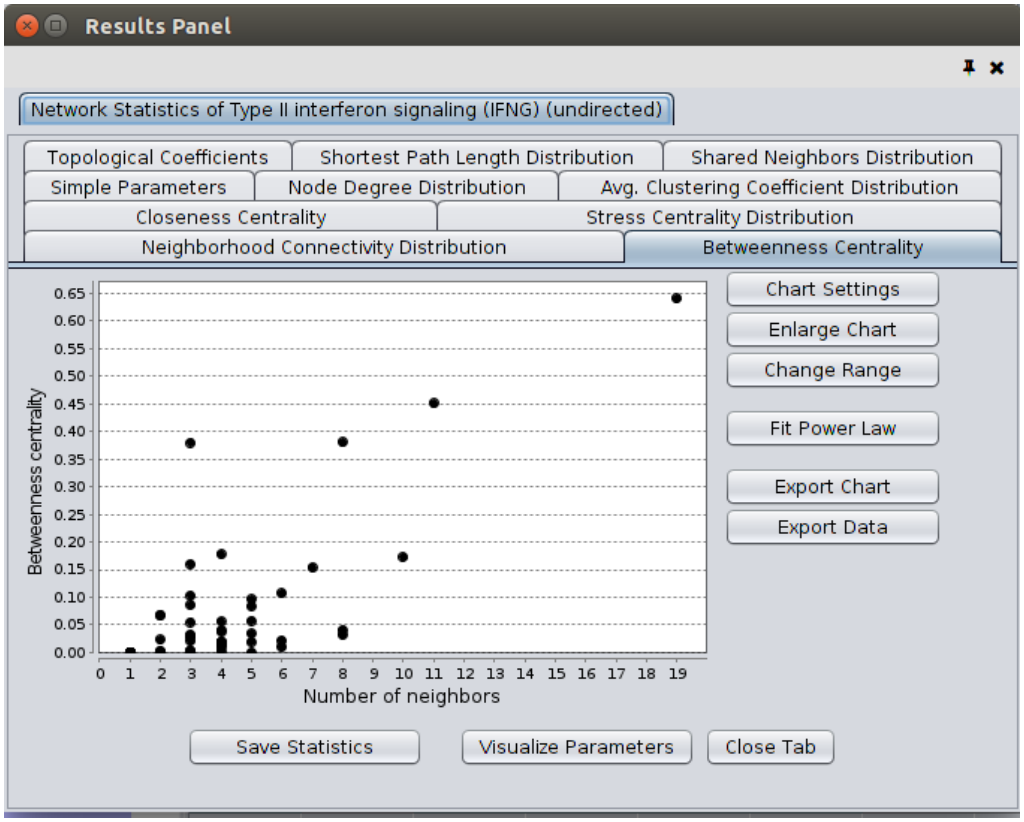


Node degree = number of neighbours

> 25 nodes have only one neighbour. ~3 nodes have > 10 neighbours. Overall a highly connected network.

c) When you look at the betweenness centrality distribution, can you see a correlation of number of neighbours and betweenness? Is the betweenness higher for nodes with a lot of neighbours? Why or why not?

Hint: Check lecture for explanation of node betweenness and relationship to node degree.



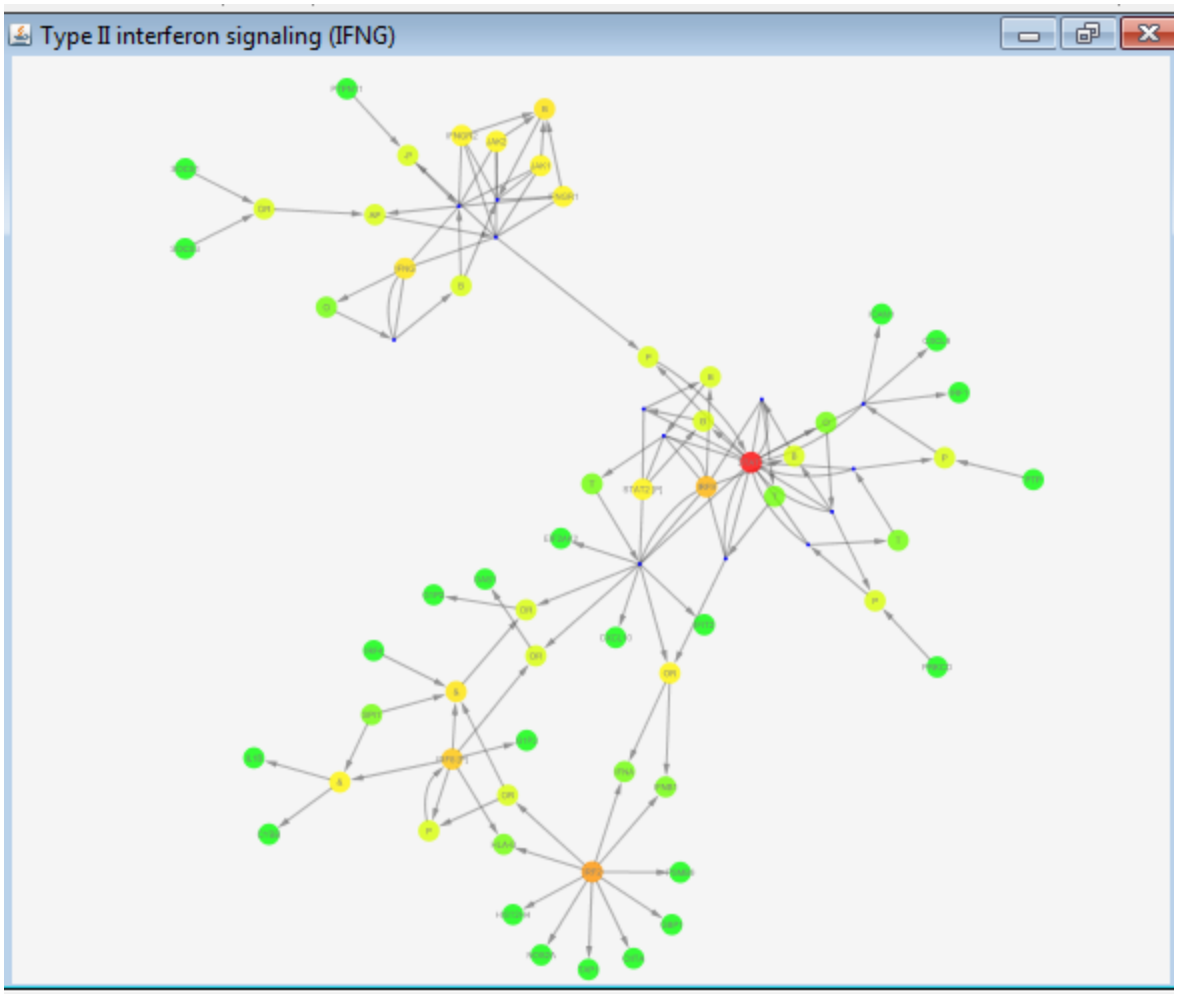
Check table of nodes to see the details of which nodes have a high node degree and high betweenness. The betweenness of a node is not directly linked with the number of neighbors. You see nodes with 3 neighbors having a higher betweenness of 0.38. The betweenness states how much information goes through this node. If a node with two connections links two subnetworks to each other it will have a very high betweenness.

Degree	BetweennessCentral...
19	0.64083401
11	0.45083545
10	0.17286869
8	0.38048166
8	0.03225526
8	0.04106947
7	0.15387883
6	0.10730308
6	0.02190541
6	0.00988958
5	0.08418101
5	5.8685E-4
5	0.09610113
5	0.03435055
5	0.01980356
5	0.0577943
4	0.05627282
4	0.17785321
4	0.03956703
4	0.03896714
4	0.01036368
4	0.00288645

d) Can you find hub proteins in the network?

Visualize the node degree as node color gradient on the nodes in the NetworkAnalyzer "Visualize Parameters" dialog.

1. Map node color to → Select "Degree"
2. Apply and close the NetworkAnalyzer dialog

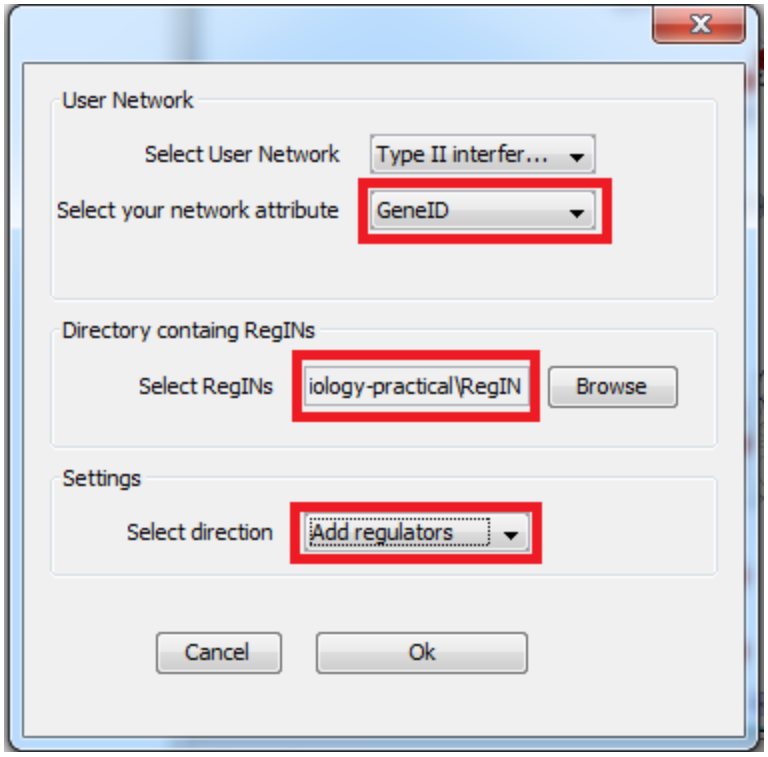


Yes, STAT1 is a hub protein with a degree of 19. Also IRF2 and IRF9 have a high degree.

e) How many drugs are targeting the genes in the pathway?

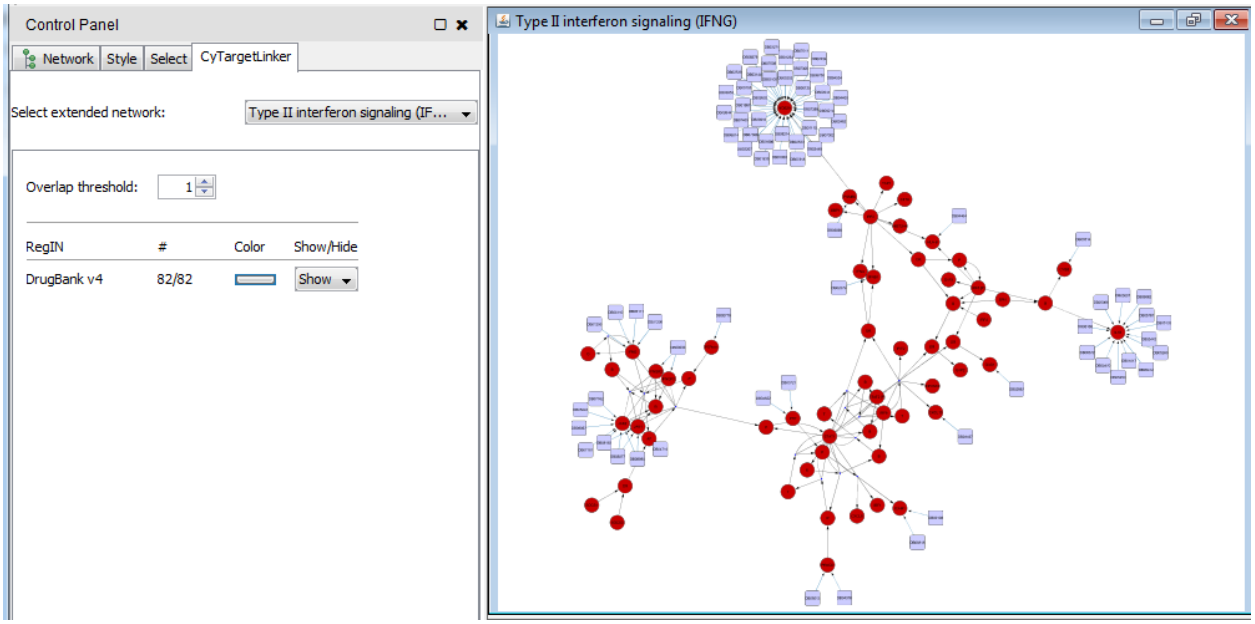
You will use a Cytoscape app, CyTargetLinker, to extend the pathway with drug-target interactions. The regulatory information is provided as a network, called Regulatory Interaction Network (RegIN). For this practical the DrugBank RegIN is stored in the downloaded zip file in the directory "RegIN".

Click on Apps → CyTargetLinker → Extend Network



- Select User Network
  - choose the network name of the pathway network
- Select your network attribute
  - choose the name of the column containing a biological identifier ("GeneID")
- Select Datasource Networks
  - Browse for the directory which contains the RegIN files (network-biology-practical/RegIN)
- Select direction "Add regulators"

82 drug-target interactions (see CyTargetLinker tab) and 77 unique drugs (select based on biologicalType=drug) were found.


Control Panel
Network
Style
Select
CyTargetLinker
Select extended network:
Type II interferon signaling (IF...
Overlap threshold:
1
RegIN
#
Color
Show/Hide
DrugBank v4
82/82
Show
Type II interferon signaling (IFNG)