

Bioinformatics Lab Part 2

By: Ritali Jain

STRING DB NETWORKS:

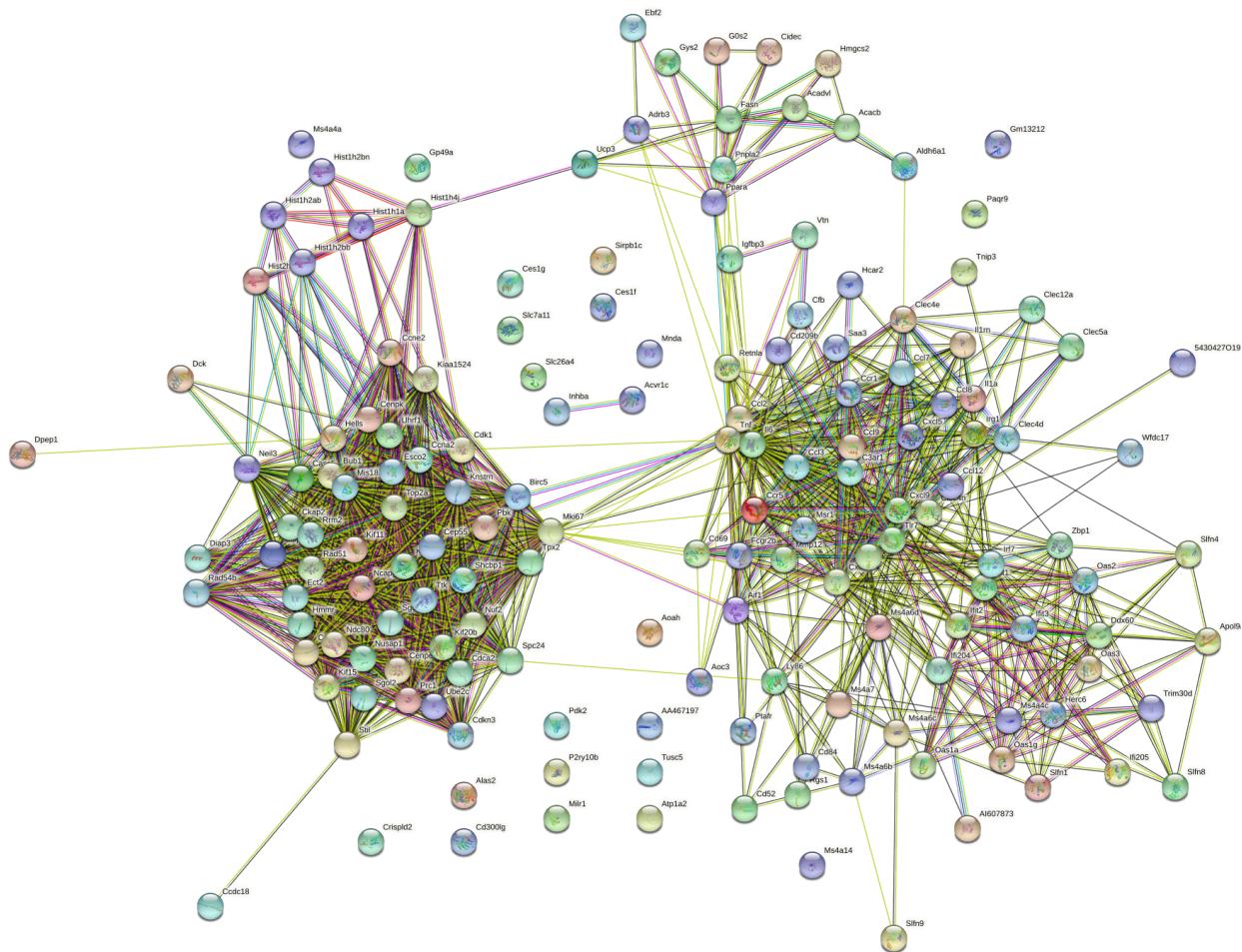


Figure 1 - LPS

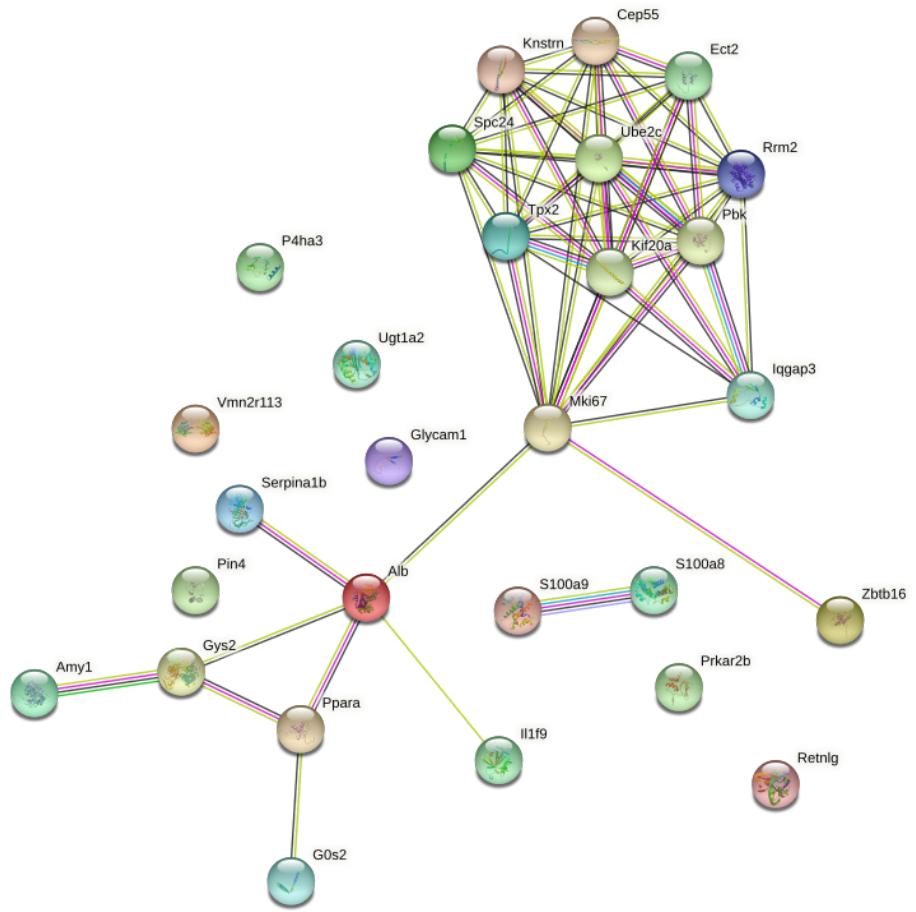


Figure 2 - BaP

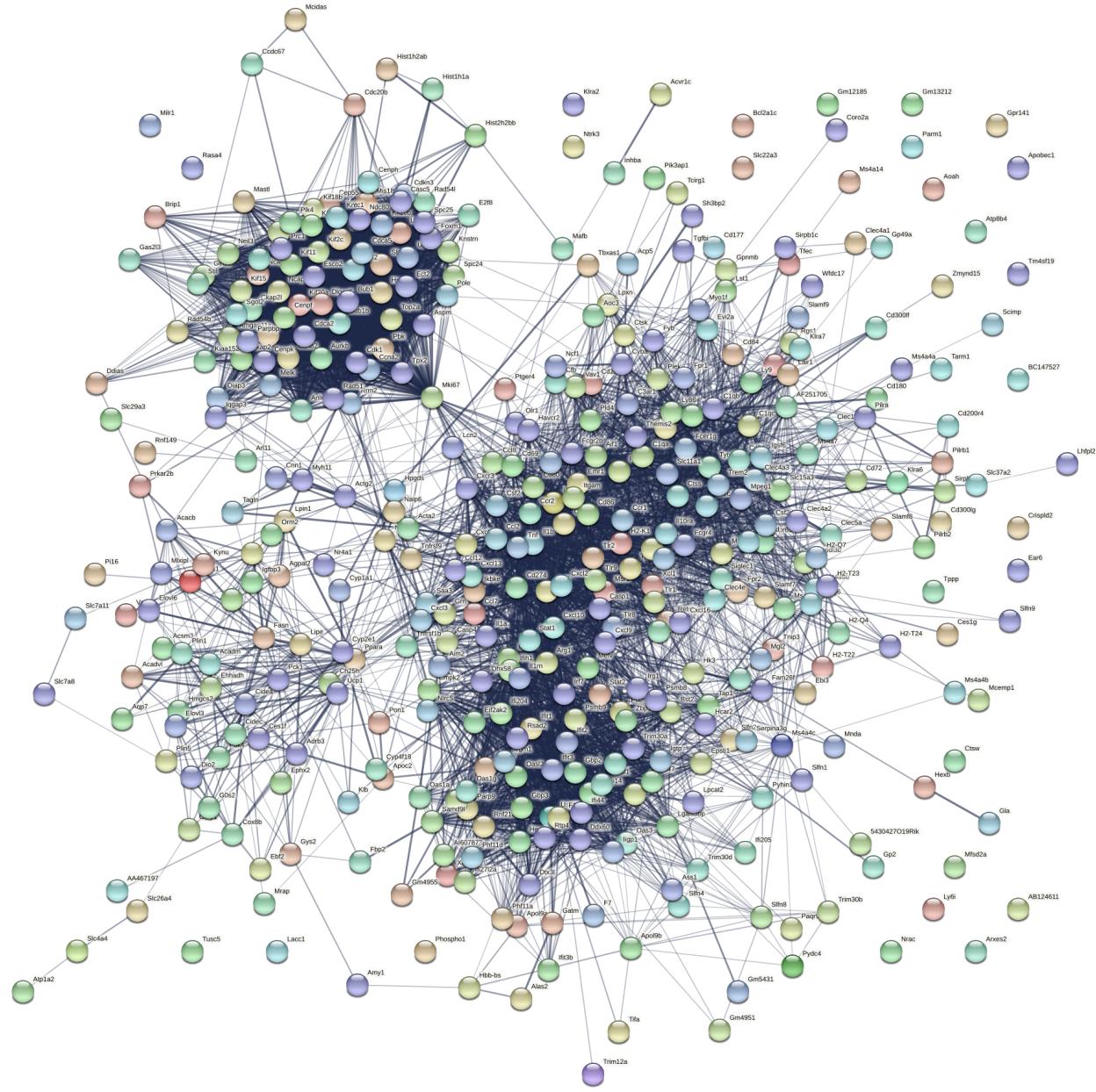


Figure 3 - LPS-BaP

1) Do all your networks look like each other or are there differences?

The networks look different from each other. Firstly, the volume of protein nodes in the networks varies greatly between the treatment groups because there is a different number of DEGs in each. Furthermore, the amount and size of subnetworks varies. For the LPS network, there seem to be two subnetworks of approximately the same size. BaP has one main network off of which proteins branch off. The LPS-BaP network has one large and one small subnetwork.

- 2) Look at some of the nodes (circles) in your networks, are all the nodes connected within the network? If some are unconnected what does this tell you about these nodes (genes)?

Several nodes are not connected within the networks. These unconnected nodes/genes do not have a functional relationship or protein interaction with the other nodes in the network.

- 3) Look at your networks, is it one entire big network or is there multiple small network clusters? If there are any smaller clusters, click on those nodes within the cluster and look at them. What do all the nodes in the smaller cluster (s) have in common?

The networks tend to have multiple small network clusters.

The nodes in the larger cluster for the LPS-BaP network seem to pertain to immune and viral biological processes or functions. For the smaller cluster in the LPS-BaP network and the cluster on the left of the LPS network, the nodes seemed to pertain to mitosis or other chromosome-related activity. The nodes in the cluster on the right of the LPS network pertain to inflammatory response.

- 4) Choose any node of interest to you within your networks, give the name of the gene in your node and some information you found from it in String-db. From the information you found from this gene, why do you think this gene is being affected (depending on what network you are looking at) from the treatment itself.

Gene: Interleukin-1 alpha

Identifier: ENSMUSP00000028882, IL1a

Information: Produced by activated macrophages, IL-1 stimulates thymocyte proliferation by inducing IL-2 release, B-cell maturation and proliferation, and fibroblast growth factor activity. IL-1 proteins are involved in the inflammatory response, being identified as endogenous pyrogens, and are reported to stimulate the release of prostaglandin and collagenase from synovial cells

I am looking at this gene from the LPS network. I think this gene is being affected by the LPS treatment because LPS induces pulmonary inflammation, and this gene is involved in the inflammatory response. Its functions include processes which are a part of inflammation.

CYTOSCAPE NETWORKS:

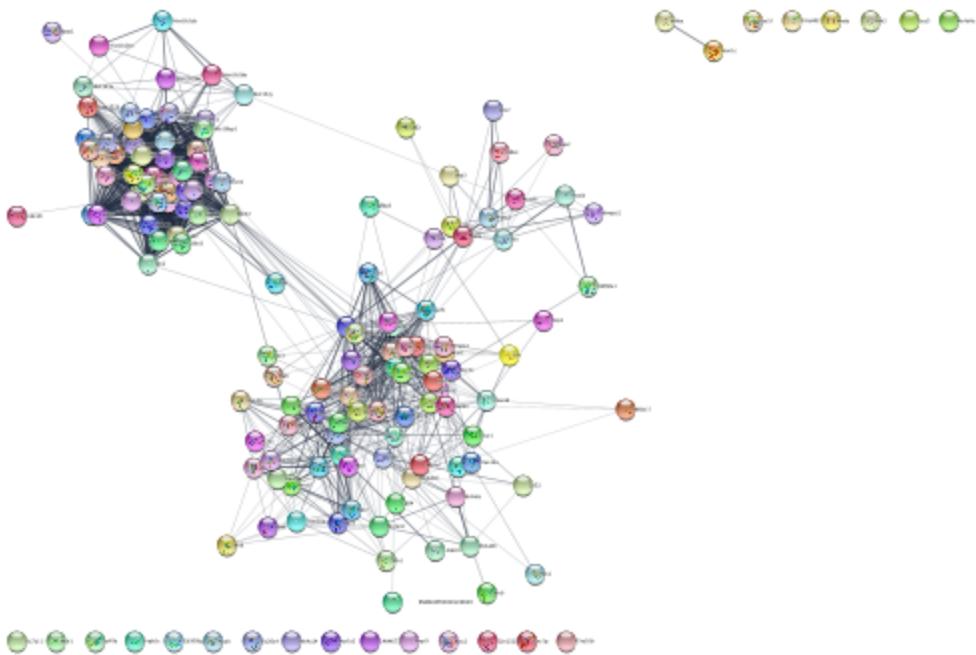


Figure 4 - LPS



Figure 5 - BaP

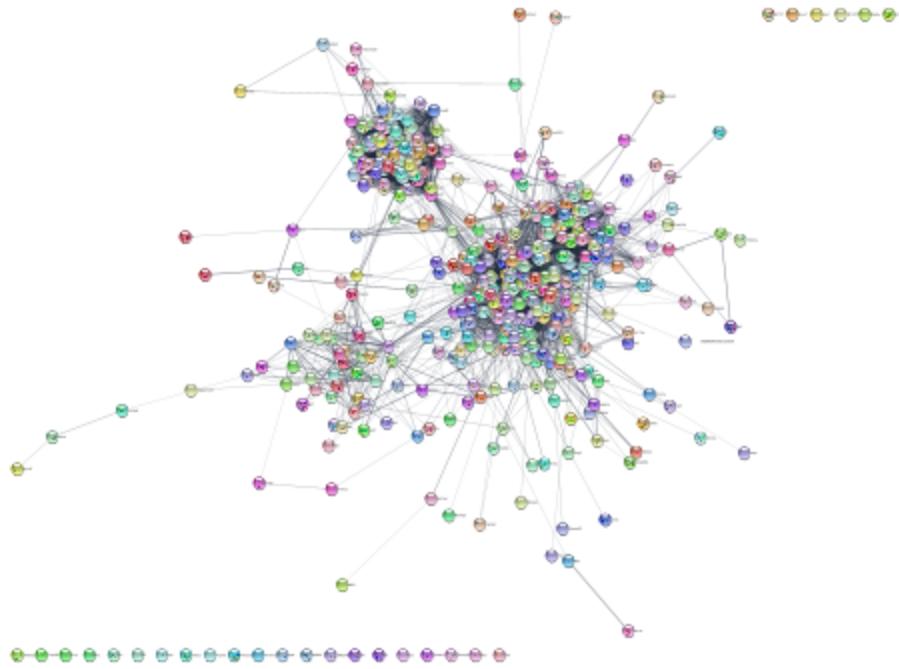


Figure 6 - LPS-BaP

5) Top terms in GO descriptions, sorted by cluster frequency

For the LPS GO network, some of the top terms that are unspecific or not helpful are cellular process, biological regulation, regulation of biological process, and regulation of cellular process. Out of the top ten terms, the one that seems the most informative is primary metabolic process. The cluster frequency of this term is 42/144, or 29.1%. This term indicates that the presence of LPS causes changes in anabolic and catabolic pathways.

For the BaP network, some of the top terms that are unspecific or not helpful are cellular process. Out of the top ten terms, the one that seems the most informative is M phase of mitotic cell cycle. The cluster frequency of this term is 4/29, or 13.7%. This term indicates that the presence of BaP causes changes in cell division and replication, which makes sense because BaP is known to be carcinogenic and causing cell proliferation.

For the LPS-BaP network, some of the top terms that are unspecific or not helpful are cellular process and biological regulation. Out of the top ten terms, the one that seems the most informative is immune system process. The cluster frequency of this term is 700/28959, or 2.4%. This term indicates that the presence of LPS-BaP interaction is related to a change in a process pertaining to the immune system. The top 10 terms are not too specific so it is harder to make a conclusion for this term.

6) Top terms in GO descriptions, sorted by Corrected p-value

For the LPS network, the 3 most informative terms are M phase of mitotic cell cycle, chromosome segregation, and chemotaxis. These seem relevant to the LPS condition because the presence of LPS might affect the process of mitosis and DNA replication, as well as the movement of cells towards certain chemicals. This would make sense for the inflammatory response that is supposed to be connected to LPS.

For the BaP network, the 3 most informative terms are regulation of mitotic spindle organization, regulation of integrin biosynthetic process, and germ-line stem cell division. These seem relevant to the BaP condition because the presence of BaP is known to be cancerous on its own, which would mean that this treatment should cause changes in mitosis and stem cell division. This is reflected by the specific and relevant terms in this analysis.

For the LPS-BaP network, the 3 most informative terms are M phase of mitotic cell cycle, organelle fission, and chemotaxis. These terms pertain to the LPS-BaP condition because the interaction of LPS and BaP would likely impact the same biological processes that LPS and BaP were impacting before, but this time, possibly in different ways.

7) Comparing terms across your networks, sorted by corrected p-value

Across the networks, common specific terms that appear regard mitosis, regulation, or immune response. The LPS and LPS-BaP networks had more focus on chemotaxis, inflammatory response, and processes occurring in chromosomes compared to the BaP network. The conditions with LPS are seeing responses regarding those processes, while the BaP processes are present but inhibited in the LPS-BaP network.

BINGO output						
Annotation: Curator = GO, Species or file = Mus musculus, Type = default Ontology: Curator = bingo, Type = namespace Close tab						
GO ID	GO Description	p-val	Corrected p-val	Cluster frequency	Total frequency	Genes
9611	response to wounding	1.3455E-17	1.1492E-15	22/144 15.2%	369/28959 1.2%	20302 16175 20311 19204 20210 20296 19013 21926 17084 20307...
9987	cellular process	1.6449E-15	1.3269E-15	94/144 65.2%	8929/28959 30.8%	240641 14130 20135 72415 319178 66853 20296 14373 16551 669...
48518	positive regulation of biological process	1.8771E-13	1.4345E-13	37/144 25.6%	1854/28959 6.4%	16175 14130 20135 19204 66853 20296 21926 108000 16193 1162...
6935	chemotaxis	3.3265E-13	2.3001E-11	12/144 8.3%	110/28959 0.3%	20302 20311 12768 19204 20296 20293 20308 12267 20307 56066...
42330	taxis	3.3265E-13	2.3001E-11	12/144 8.3%	110/28959 0.3%	20302 20311 12768 19204 20296 20293 20308 12267 20307 56066...
48583	regulation of response to stimulus	5.9162E-13	3.9047E-11	19/144 13.1%	432/28959 1.4%	16175 14130 19204 19013 21926 108000 16193 16181 56644 2075...
48522	positive regulation of cellular process	2.0128E-11	1.2707E-9	32/144 22.2%	1639/28959 5.6%	16175 14130 20135 66853 20296 21926 108000 16193 11629 2237...
42221	response to chemical stimulus	2.6851E-11	1.6245E-9	26/144 18.0%	1086/28959 3.7%	14130 20311 19204 623474 20296 170743 20293 12774 15945 269...
16043	cellular component organization	1.3036E-10	7.5713E-9	34/144 23.6%	1979/28959 6.8%	240641 72119 14130 20135 72415 319178 15201 68612 16551 219...
7059	chromosome segregation	1.3826E-10	7.7211E-9	8/144 5.5%	50/28959 0.1%	229841 21973 72415 12235 66977 108907 68549 11799
7610	behavior	2.2079E-10	1.1450E-8	17/144 11.8%	470/28959 1.6%	20302 20311 19204 20296 20308 20307 56066 20306 98660 16193...
9605	response to external stimulus	2.1754E-10	1.1450E-8	15/144 10.4%	345/28959 1.1%	20302 20311 19204 20296 19013 20308 20307 56066 20306 56644...
6996	organelle organization	2.7440E-10	1.3739E-8	24/144 16.6%	1026/28959 3.5%	240641 72119 21973 217653 2415 319178 15201 74107 68612 16...
7626	locomotory behavior	3.6070E-10	1.7458E-8	13/144 9.0%	247/28959 0.8%	20302 20311 19204 20296 20308 20307 56066 20306 86660 12768...
40011	locomotion	5.9273E-10	2.7763E-8	15/144 10.4%	371/28959 1.2%	20302 20311 19204 20296 21926 20308 20307 56066 20306 98660...
51707	response to other organism	6.1522E-10	2.7916E-8	13/144 9.0%	258/28959 0.8%	14130 19204 16365 21926 16193 54123 56644 12768 170743 1785...
9607	response to biotic stimulus	8.9564E-10	3.9408E-8	14/144 9.7%	322/28959 1.1%	14130 19204 16365 21926 16193 54123 56644 12768 170743 1785...
32879	regulation of localization	1.2595E-9	5.3789E-8	19/144 13.1%	677/28959 2.3%	14130 20288 66853 21926 108000 16193 23985 16181 229841 566...
51726	regulation of cell cycle	3.0046E-9	1.2465E-7	12/144 8.3%	239/28959 0.8%	72119 20555 229841 21926 12235 108907 108000 12448 11799 72...
80134	regulation of response to stress	3.7936E-9	1.5020E-7	12/144 8.3%	244/28959 0.8%	16175 14130 56644 27052 12768 19204 170743 21926 122774 1618...
31347	regulation of defense response	3.8273E-9	1.5020E-7	10/144 6.9%	148/28959 0.5%	14130 56644 27052 12768 19204 170743 21926 122774 11754 14129...
2682	regulation of immune system process	5.6900E-9	2.1742E-7	14/144 9.7%	372/28959 1.2%	20302 14130 19204 21926 108000 16193 56644 170743 14962 122...
50776	regulation of immune response	1.5329E-8	5.5791E-7	11/144 7.6%	221/28959 0.7%	14130 56644 19204 170743 14962 21926 12267 108000 16193 232...
32101	regulation of response to external stimulus	1.5369E-8	5.5791E-7	10/144 6.9%	171/28959 0.5%	14130 27052 12768 19013 21926 12267 15945 11754 14129...
51239	regulation of multicellular organismal process	1.8244E-8	6.4610E-7	22/144 15.2%	1074/28959 3.7%	20302 16175 14130 16323 20296 21926 108000 98660 16193 5664...
48584	positive regulation of response to stimulus	1.9306E-8	6.6743E-7	11/144 7.6%	226/28959 0.7%	16175 56644 19204 170743 14962 21926 12267 15945 1175...
51704	multi-organism process	5.6861E-8	1.9200E-6	13/144 9.0%	377/28959 1.3%	14130 19204 16365 21926 16193 54123 56644 12768 170743 1785...
45787	positive regulation of cell cycle	6.1623E-8	2.0336E-6	6/144 4.1%	42/28959 0.1%	229841 21926 108907 108000 11799 12534...
1817	regulation of cytokine production	7.0038E-8	2.2599E-6	9/144 6.2%	151/28959 0.5%	16175 14130 56644 170743 21926 12267 108000 16193 69165...
51716	cellular response to stimulus	7.3048E-8	2.3058E-6	17/144 11.8%	695/28959 2.3%	20302 14130 71988 19204 623474 20296 19013 21926 18140 1936...

Figure 7 - LPS enrichment analysis

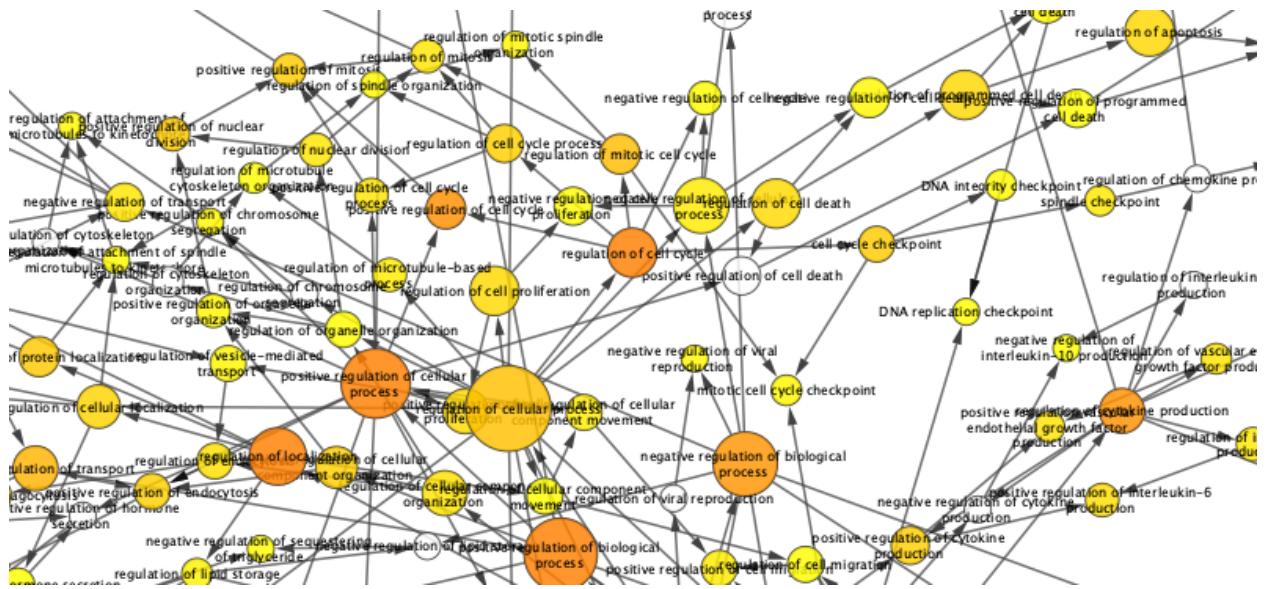


Figure 8 - LPS GO BinGO network

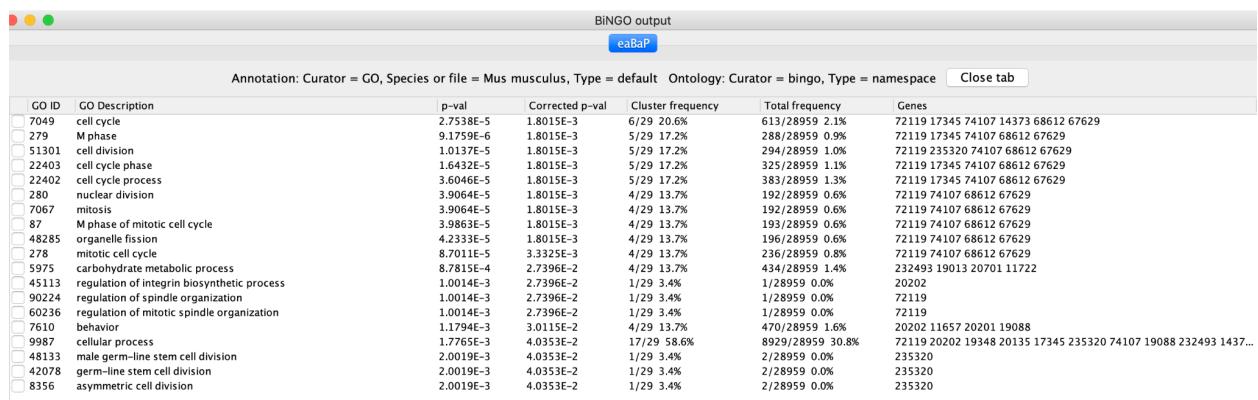


Figure 9 - BaP enrichment analysis

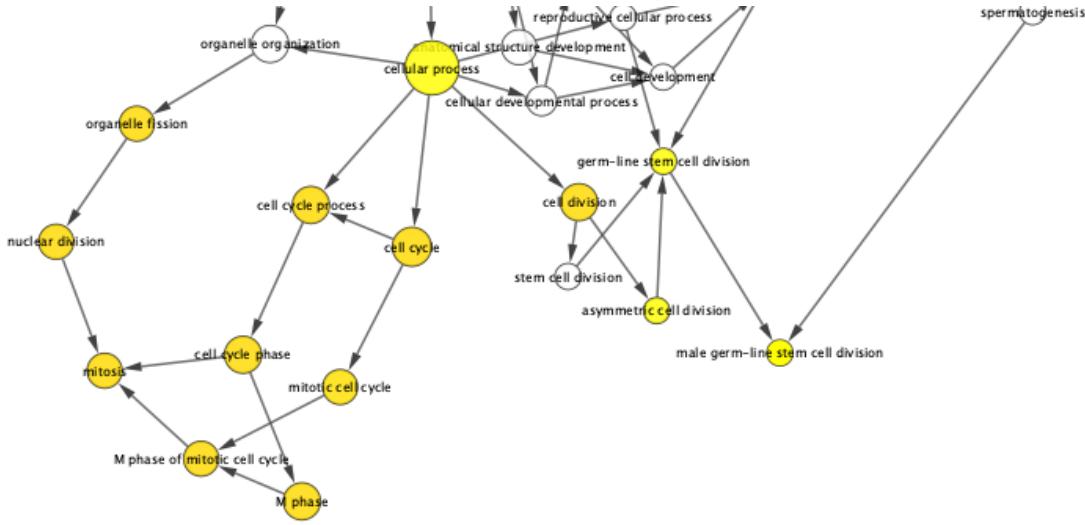


Figure 10 - BaP GO BinGO network

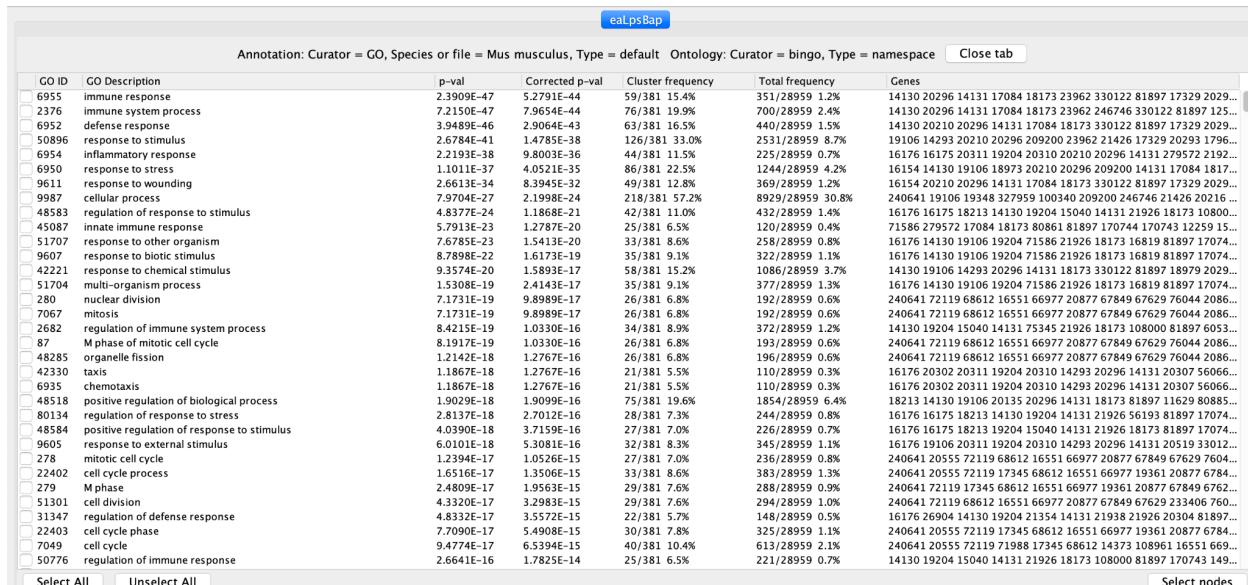


Figure 11 - LPS-BaP enrichment analysis

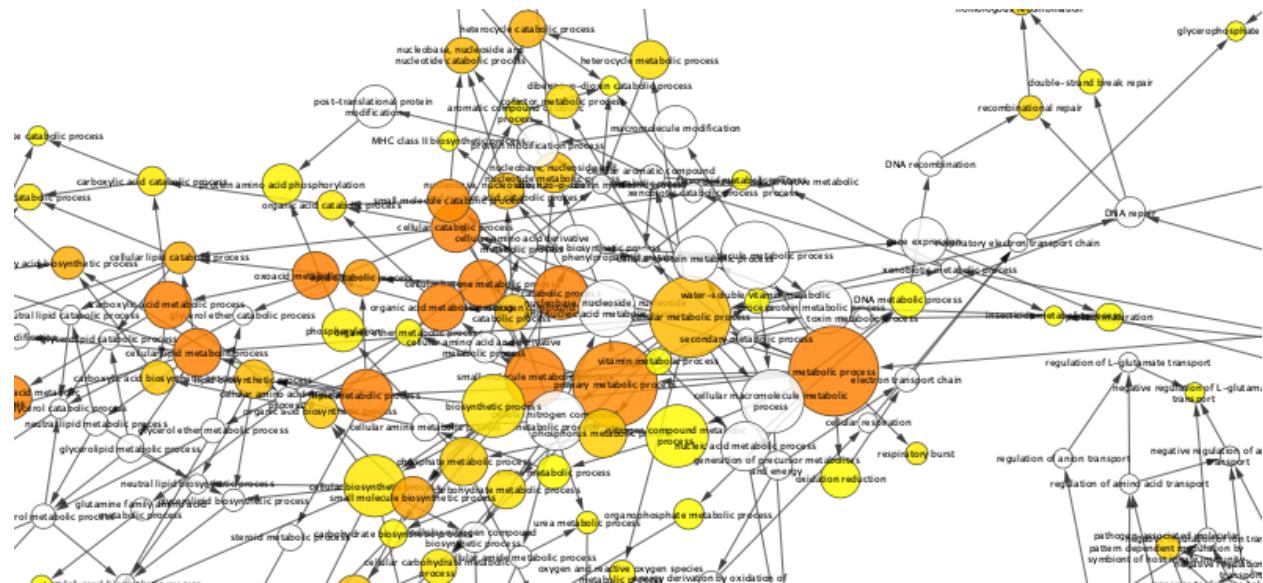


Figure 12 - LPS-BaP GO BinGO network

- 8) How does this BiNGO analysis (on sub-clusters) differ from the full network BiNGO analysis that you did? What does this BiNGO analysis tell you about the smaller cluster within your network?

These BiNGO analyses differ because they provide far more detail and specificity regarding the genes in a particular cluster, compared to the full network BiNGO analysis. For example, the smaller cluster within the LPS-BaP network showed more details, in terms of regulation of transcription in mitosis, specificity in the lungs for cells, and mentions of meiosis. This can help determine where this cluster of genes primarily acts and specifically how they are connected.

- 9) What major systems are being affected in the mice from their treatment conditions after looking at this enrichment analysis?

Based on this enrichment analysis, the inflammatory/immune response and cell division/mitosis processes are being affected by the treatment conditions.

- 10) What benefit do you think a tool like Cytoscape has for looking at datasets such as yours?

Cytoscape is beneficial because it creates visuals for complex biological datasets, creating networks, performing enrichment analysis, and simplifying gene ontology.

- 11) Do you feel like Cytoscape helped you gain some insight into your data? Would you like to use more tools like this to analyze your data?

I really liked the ability to visualize the networks of genes/proteins and the network of gene ontology. Through these visuals, Cytoscape made it really easy to see patterns and relationships between the terms we were examining, especially in the interpretation of gene ontology terms based on the size/color of the node. This helped me gain insight into the data in a way I would

not have been able to had I simply looked at text-based output. I think Cytoscape really clarified enrichment analysis. I would love to use similar tools to analyze biological data.