

**Post Genomics – Fall 2025**  
**Homework 5**  
**Due: Thursday, October 2, 2025**

**Part 1. Protein-Protein Interaction Networks. (Non-Programming) (60 points)**

1.1 Use the CSV (AML\_gene.csv), to determine, create a list, of the unique genes within this dataset. Submit these genes into String-DB for the Protein-Protein Interaction Network:

[https://string-db.org/cgi/input?sessionId=byv6kGAGpNla&input\\_page\\_active\\_form=multiple\\_identifiers](https://string-db.org/cgi/input?sessionId=byv6kGAGpNla&input_page_active_form=multiple_identifiers)

1.2 Under the Analysis Tab,

1.2.1 Take a screen shot of the Network Stats

number of nodes: 1171	expected number of edges: 4436
number of edges: 4819	PPI enrichment p-value: 7.22e-09
average node degree: 8.23	your network has significantly more interactions than expected (what does that mean?)
avg. local clustering coefficient: 0.336	

1.2.2 For each Gene Ontology terms (Biological Process, Molecular Function, and Cellular Component) what is the top strength term? What is the count of network for each of these terms?

For Biological process:

> Biological Process (Gene Ontology)					
GO-term	description	count in network	strength	signal	false discovery rate
GO:0006909	Phagocytosis	28 of 173	0.43	0.4	0.0131
GO:0007156	Homophilic cell adhesion via plasma membrane adhesion molecu...	27 of 169	0.43	0.37	0.0193

The maximum strength is **0.43**, the **biological processes** with this strength are **Phagocytosis** and **Homophilic Cell Adhesion** with count in network **28 of 173** and **27 of 169** respectively.

For molecular function:

Molecular Function (Gene Ontology)					
GO-term	description	count in network	strength	signal	false discovery rate
GO:0005509	Calcium ion binding	78 of 717	0.26	0.42	0.0052

The maximum strength is **1.23**, and the **molecular function** is **Calcium Ion Bonding** with a count in network **78 of 717**.

For Cellular Component:

> Cellular Component (Gene Ontology)					
GO-term	description	count in network	strength	signal	false discovery rate
GO:0005592	Collagen type XI trimer	5 of 5	1.23	0.47	0.0146

The maximum strength is **1.23**, and the **Cellular Component** is **Collagen type XI trimer** with a count in network **5 of 5**.

1.2.3 The format is #A of #B, what does #A represent and what does #B represent?

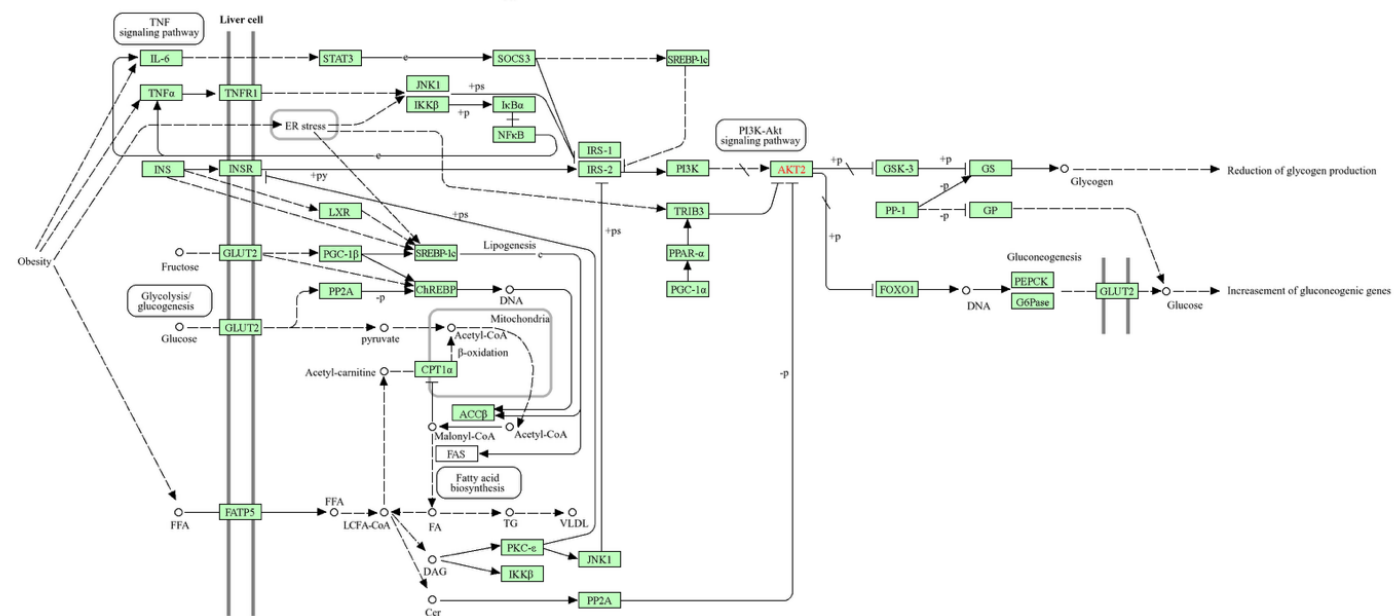
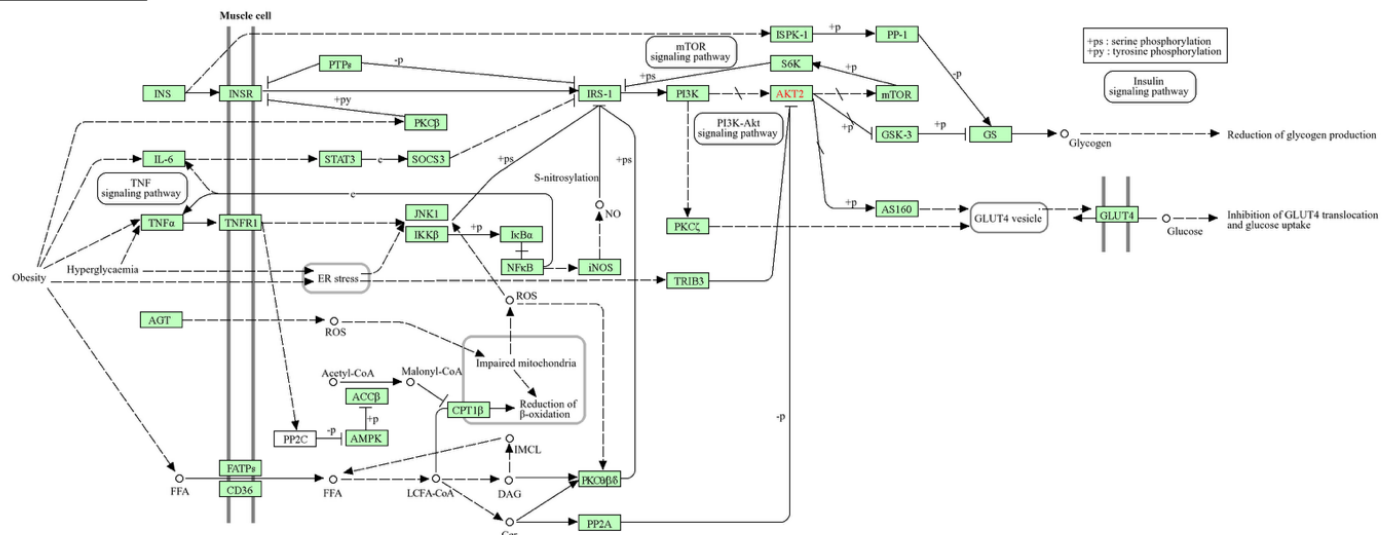
The first number (#A) indicates how many proteins in the network are annotated with a particular GO-term. The second number (#B) indicates how many proteins in total (in the network and in the background) have this term assigned.

1.2.4 What is the top strength KEGG pathway? What is its count of network? Click the link for the pathway and submit a screen shot of the pathway.

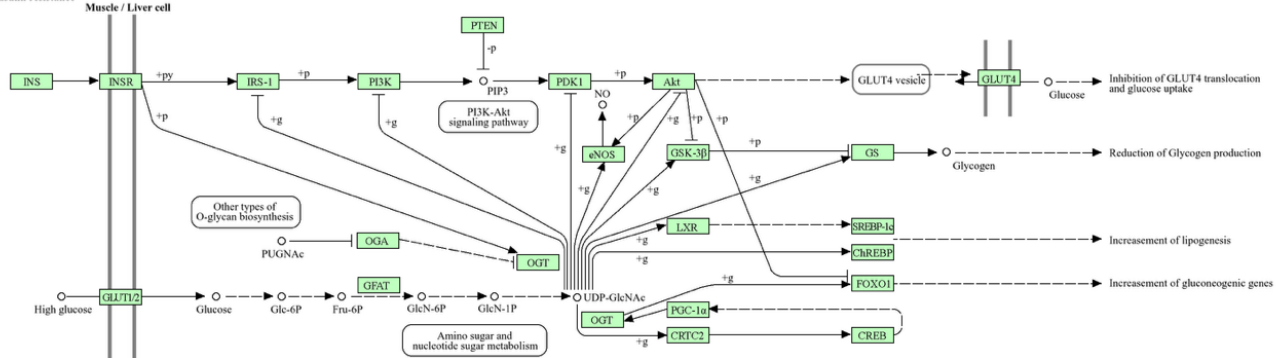
> KEGG Pathways					
pathway	description	count in network	strength	signal	false discovery rate
hsa04931	Insulin resistance	21 of 106	0.52	0.54	0.0026

The maximum strength is **0.52**, and the **KEGG pathway** is **Insulin resistance** with a count in network **21 of 106**.

# INSULIN RESISTANCE



## O-GlcNAc regulation of Insulin resistance



### 1.3 Under the Exports Tab,

1.3.1 Download the “protein node degrees” file. How many proteins are there that have a node degree above the found average from the Network Stats?

*There are 411 proteins with the node degrees > 8.16*

1.3.2 Download the “as short tabular text output” file. Within this file, what information contributes to the Combined Score? What is the highest combine scored interaction?

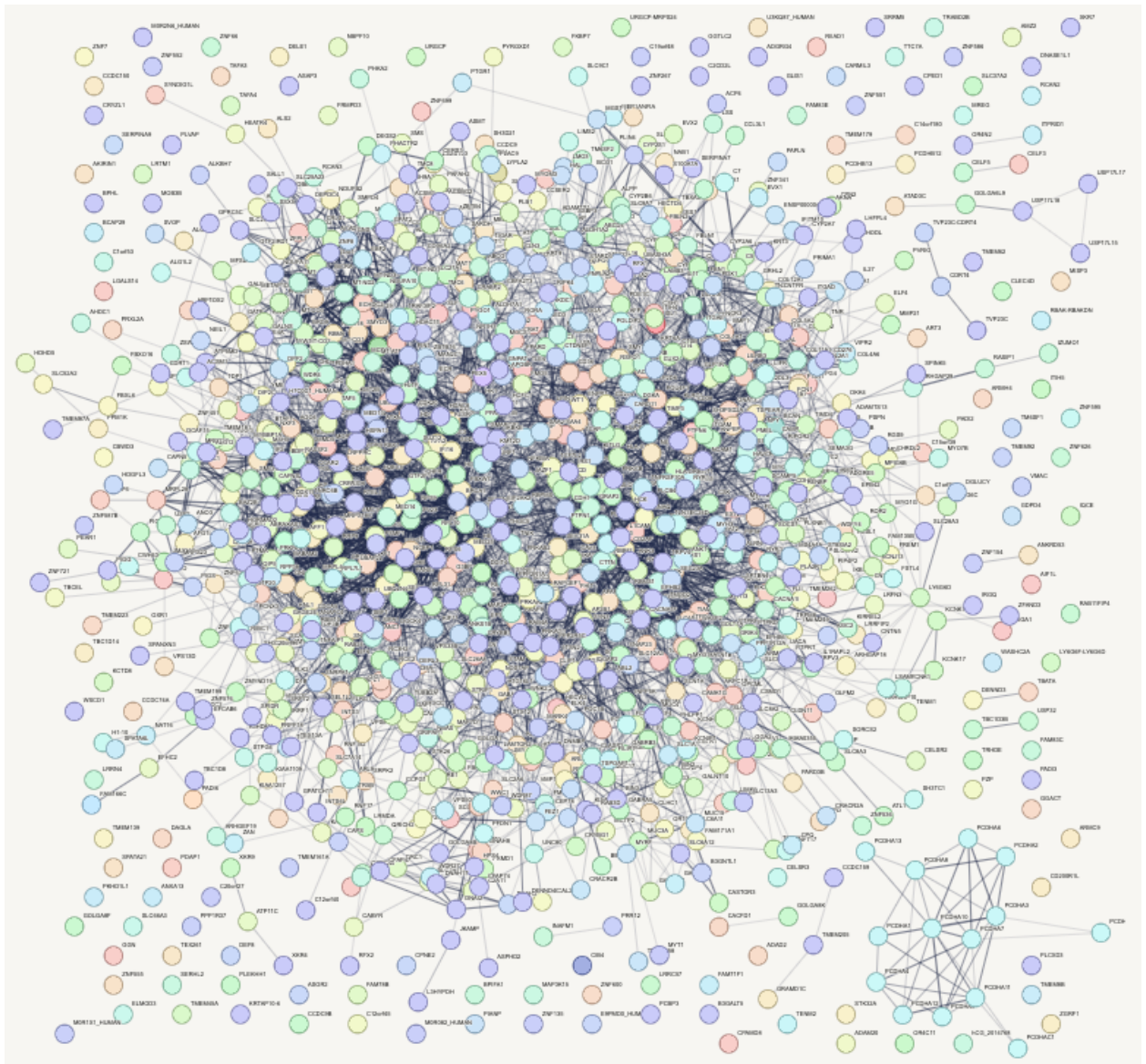
*The combined score depends on all previous rows:*

*(“neighborhood\_on\_chromosome, gene\_fusio, phylogenetic\_cooccurrence, homology, coexpression, experimentally\_determined\_interaction, database\_annotated, automated\_textmining”)*

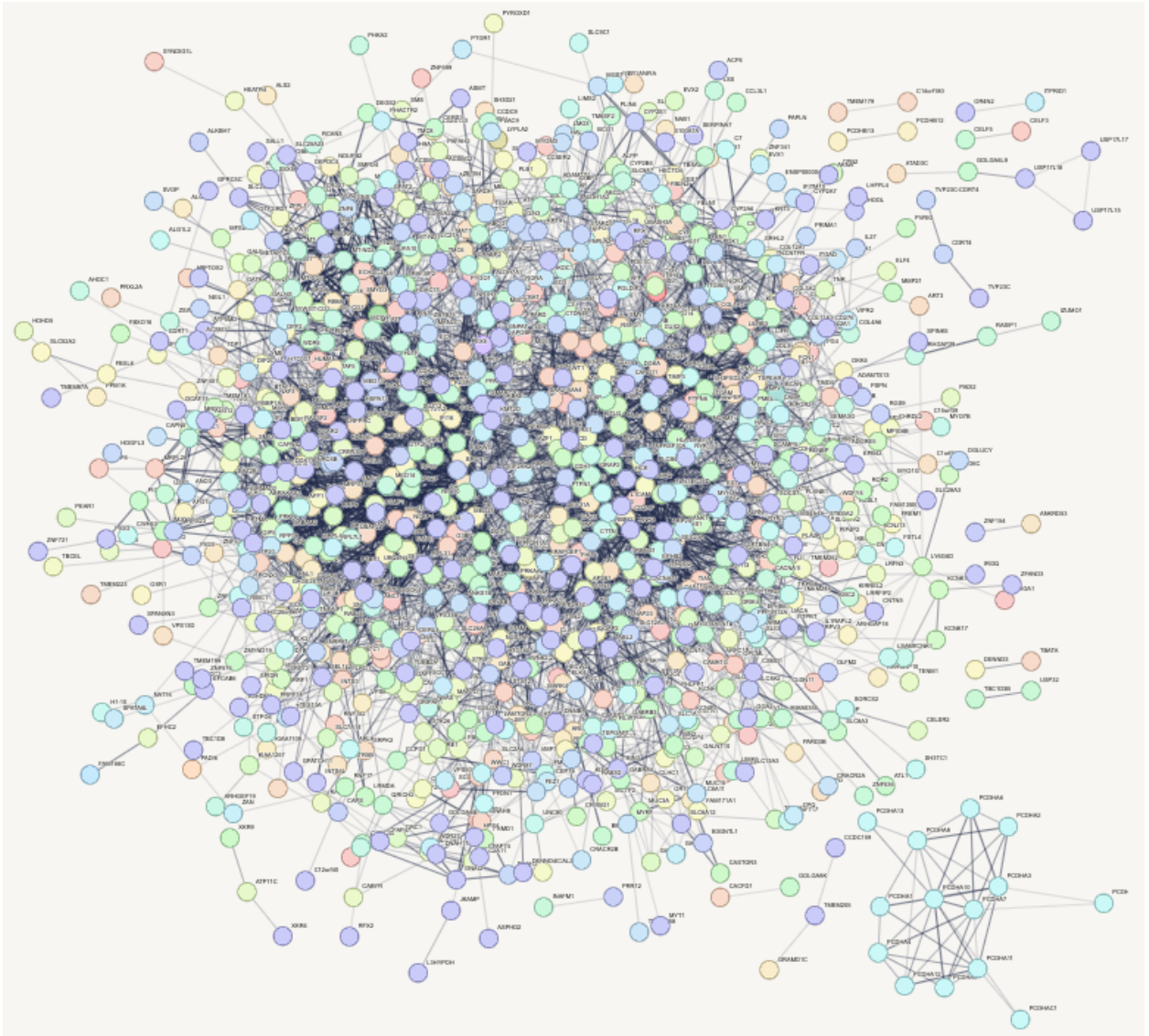
*The highest combine score is: 0.999*

## 1.4 Under the Settings Tab,

### 1.4.1 Submit a screen shot of the network as is.



1.4.2 Select the “hide disconnected nodes in the network” and update the network with this feature. Submit a screen shot of the new network.



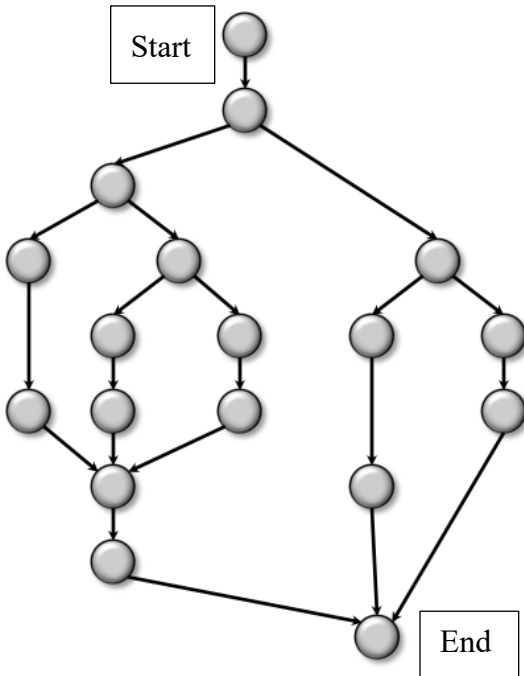
1.4.3 In your own words, what changes did you observe, if any.

The network is clearer and shows less points that leads to nothing.

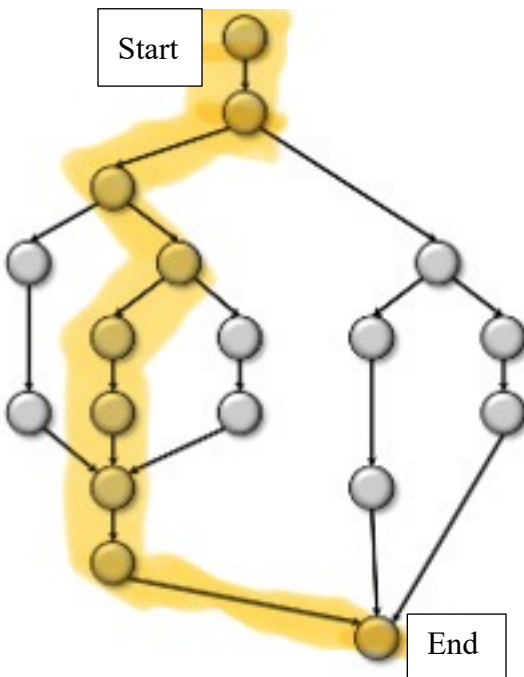


**Part 2. Node Network Practice. (Non-Programming) (40 points)**

Using the figure below, answer the following questions (assume that each edge is of length 1).



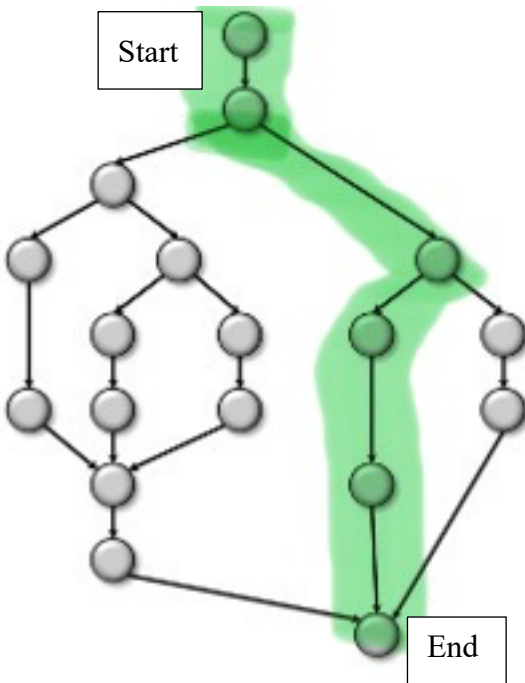
2.1 What is the longest path to get from the Start to End node? You can either: Draw by hand and submit image, edit image virtually on iPad/Computer, or give list of directionals (down, right, left, etc. )



2.2 How many edges are on the longest path?

*There are 8 edges in the longest path.*

2.3 Repeat 2.1 and 2.2 for the shortest path.



*There are 5 edges in the shortest path.*

**What to Submit:**

- 1) A single PDF with responses and screen shots from Part 1 and 2. (Your Last Name) \_HW5.pdf **\*\* Submit on Blackboard \*\***