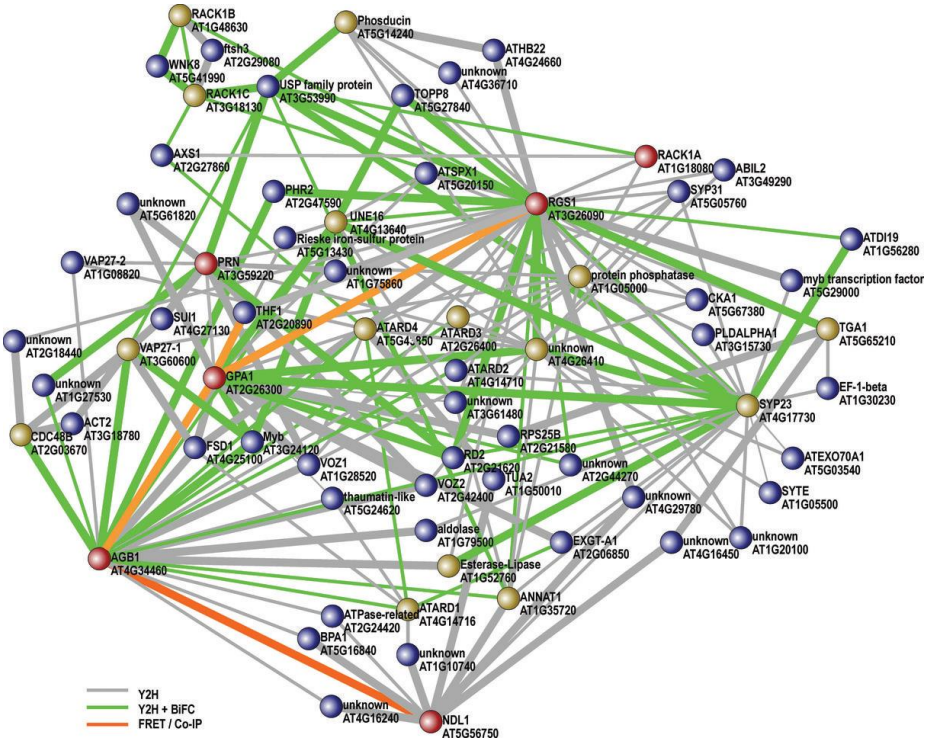
PROJECT 2: **Automated mining for protein-protein interactions in signaling networks**

Pathogens shed signals that are perceived by cells and consequently interpreted as “danger”. Cells react by preparing a defense against the pathogen ingress as well as alerting other cells of this danger. In this cellular signaling pathway, there is a central protein called RGS1 that communicates with partnering proteins to bring this action about. Communication is the result of direct physical contact between RGS1 and its partners; the message conveyed is encoded in the conformation of RGS1 when it interacts.

The JonesLab used two established techniques to “fish” for these RGS1-partner interactions. One technique that we used to fish for protein-protein interactions revealed only first-order relationships meaning that the two proteins directly interact but second and third order relationships are missed. The other technique that we used simply identified all the proteins in a protein complex with no information about the first or second order interaction relationships.

An example of our network of RGS1 proteins is shown, a so-called RGS1 interactome. The individual proteins are the nodes (Figure, blue and gold spheres) and their interaction relationships to other nodes are illustrated as edges (Figure, green, orange and greay lines). The confidence of predicted networks is influenced by the weight of the supporting data and often this is represented by differences in the size and color of the nodes and thickness and color of the edges.

Our RGS1 interactome is supported entirely by the limited experimental data that the JonesLab obtained, but the interactome community has archived a large number of protein-protein interactions into public databases. Therefore, in order to fully maximize the totality of this knowledge, we need an automated method to reach into the public databases for interactions to query what other proteins relevant to the RGS1 interactome need to be considered. Each iteration of query sequence increases the data pool exponentially therefore such a program would need to apply working criteria to focus in on highly- supported new interactions with the goal of assembling signaling complexes with defined relationships to each of the components.



When successful, this program would be used by thousands of cell biologists. A well-formed interactome is not required; even a simple two-node, one-edge relationship would greatly benefit from this automated mining program. Therefore, we intend to publish this new method and each member of the team will be co-author.

For this project, two postdoctoral associates as well as Dr. Jones will meet the team regularly and provide feedback.