PaIntDB: Visualizing Protein-protein interaction networks in *Pseudomonas aeruginosa*

CPSC 547 Project Pitch
Javier J. Castillo-Arnemann
October 8, 2019

Background

- All biological processes are the result of complex interactions between many genes and gene products.

- Traditionally, biology has taken a reductionist approach, characterizing the function of individual genes or a small group of genes, due to experimental limitations.

- Next-Generation sequencing methods and other high-throughput technologies allow biologists to study every gene at the same time.

- The new challenge is analyzing and interpreting the huge datasets that results from these experiments.

Background: PaIntDB

- **P**seudomonas **Int**eraction **D**ata**B**ase

- Protein-protein and protein-metabolite interactions in *Pseudomonas aeruginosa* strains PAO1 and PA14. (157,427 interactions)

- *P. aeruginosa* is a multi-drug resistant pathogen involved in cystic fibrosis and other diseases. Antibiotic resistance has gotten worse and will continue to do so.

- Helps visualize and interpret RNASeq Differentially Expressed genes, TnSeq phenotypically important genes, or any kind of gene list.

PaIntDB pipeline

1. Run experiment (gene knockouts, antibiotic treatment, temperature...)
2. Perform RNASeq/TnSeq.
3. Perform statistical analyses to determine genes of interest.
4. **Analyze and interpret list of genes of interest.**

PaIntDB

Input:

List of genes with optional expression data. (usually >1000 genes)

gene	baseMean	log2FoldChange	pvalue
PA1560	3486.61813214782	2.88944102878766	1.83855837198106E-83
PA1559	6153.24412240954	3.08847395382963	2.28370592381295E-82
PA2358	808.695598527297	3.01960522454999	2.71654461450156E-62
PA4775	1790.76162936442	2.19818907841521	2.02984086854949E-60
PA2655	645.450521682679	3.54173280007369	9.42164711279003E-59
PA4774	4271.83860737805	2.65376155361687	3.90484062480077E-57
PA4776	867.685586500117	1.91332193364473	3.54925102992402E-51
PA2357	92.5374717459836	3.13145859125876	1.19422356220102E-40
PA3554	3516.91767956598	2.0675401596055	9.06709986293868E-38
PA3559	2180.85840438509	2.11035091327087	7.90645152368141E-34
PA4773	3537.67629309535	2.28550685614544	1.047021212805E-32

Output:

Network showing interactions between these genes.
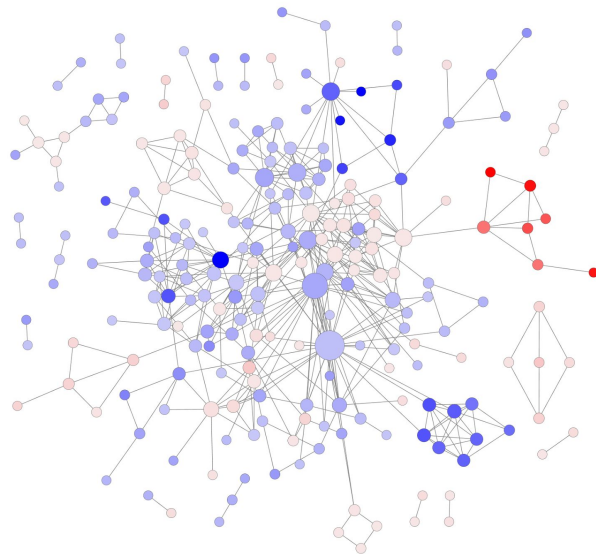
Attributes

Attribute	Type	Channel
Location	Categorical	Luminance, hue, position?
Type	Categorical	Luminance, shape
GOterm	Categorical	Luminance, hue
SourceOfInterest	Categorical	Luminance, hue
Degree	Quantitative	Node size
Log2FoldChange	Quantitative (divergent)	Hue

Prototype, screenshots and progress

Issues

Hairball effect:
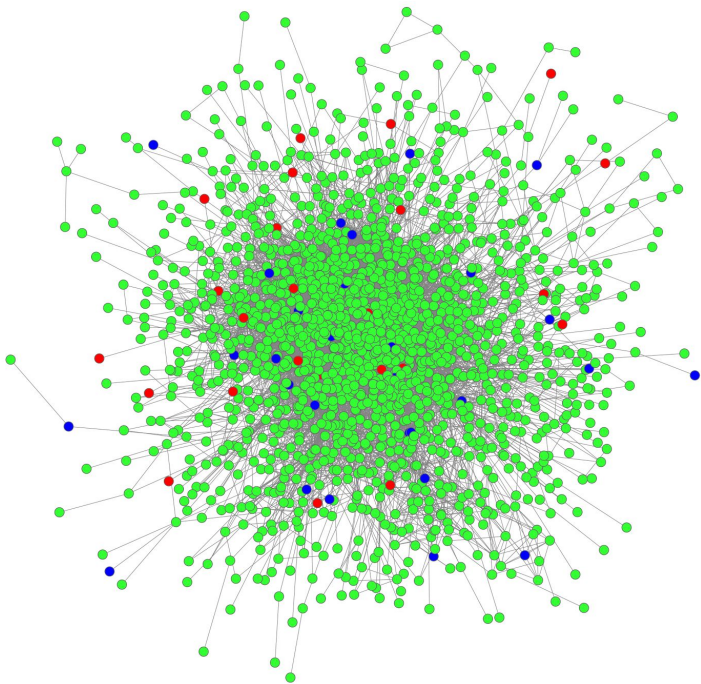
One solution:

Generate sub-networks out of functional enrichment.

tRNA processing genes
sub-network
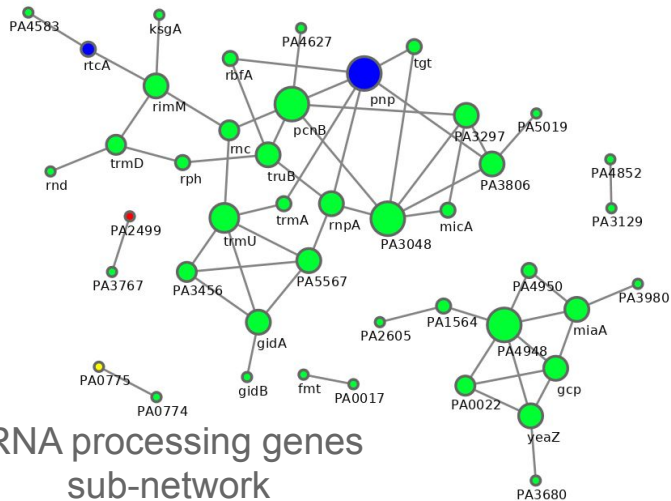
Project Goals

- Finish stylesheets to map node attributes to respective channels.

- Implement GO term enrichment and add results as attributes.

- Query the database to show all the information associated with the selected node(s).

- Implement node clustering and expand on-click for node-link views.