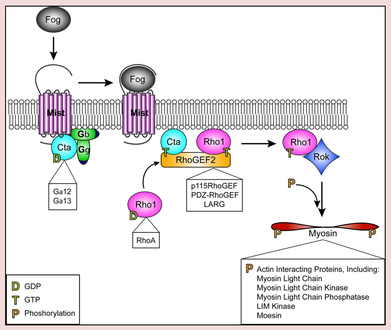
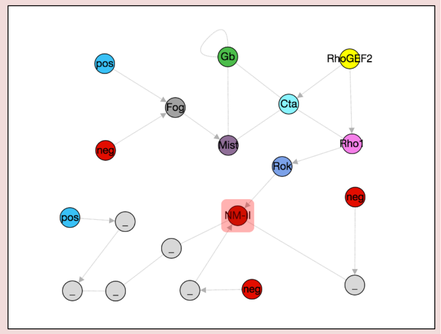
Exploration and Identification of Novel Fog Signaling Pathway Members via Application of Pathfinding Algorithms

Miriam Bern, Wyatt Gormley, Elaine Kushkowski, Kathy Thompson, Logan Tibbetts, Anna Ritz

Bio 331: Computational Systems Biology

Computational Investigation of the Fog Signaling Pathway





In order to respond to an internal or external stimulus, cells produce and utilize a variety of proteins. More proteins are then recruited in turn to coordinate and effect the necessary changes in the cell. The sequence of proteins running from stimulus to effector are known collectively as a Signaling Pathway.

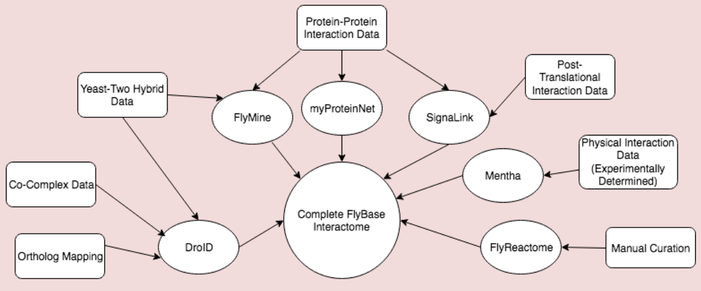
The Fog Signaling Pathway (above, left) is essential for the process of Apical Constriction. Apical Constriction is a critically important part of Gastrulation, and disruptions to this pathway during development are pathogenic.

Experimental investigation of these pathways can be time consuming and expensive. Graphs may be used to represent protein-protein interactions. Applying pathfinding algorithms to these networks is a way to find new proteins for investigation and improve the efficiency and quality of research.

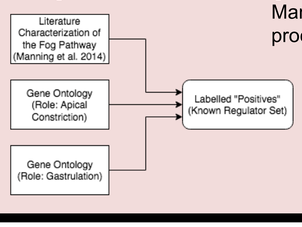
The example graph (above, right) describes a hypothetical graph representing the relationships of the proteins in the Fog pathway. By evaluating the attributes and relationships of known members in the pathway we are better able to determine which uninvestigated proteins are the best candidates for further investigation.

Driving Question: Can we find previously unassociated members of the Fog pathway by using a computational approach?

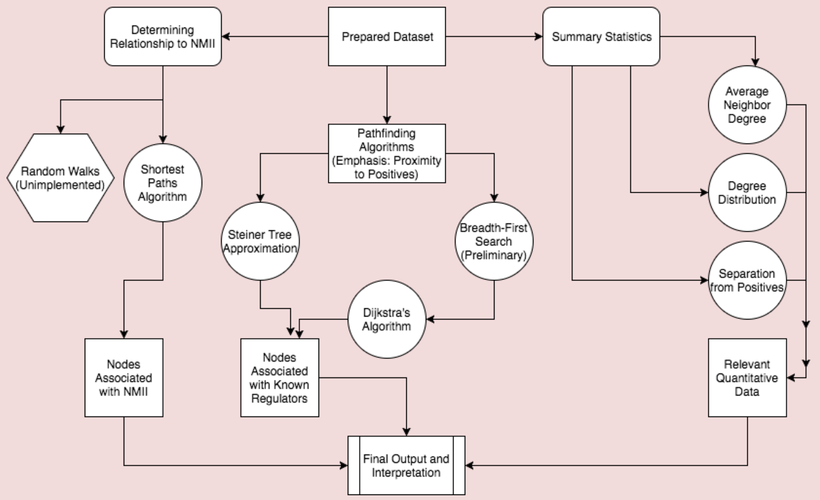
Inputs and Preprocessing



This project utilized a composite produced from six prior interactome for its analysis to increase its scope and accuracy. The above figure identifies the component interactomes and which types of data were used to produce them. A list of known regulators or "Positives" was generated from known regulators of the Fog pathway in the literature (reviewed by Manning et al. 2014), as well as genes essential to relevant processes (Apical Constriction and Gastrulation) from Gene Ontology databases. After assembly, some nodes were removed to decrease runtime. Nodes were removed if they were not present in the largest connected component or if they were greater than four edges away from the nearest known regulator.



Methods and Approach



Our primary approach focused on the identification of candidate genes based on their relationship with known regulators. In this approach, we implemented multiple pathfinding algorithms. Our implementation of a Steiner Tree approximation allowed us to determine which nodes are most relevant as intermediary connections between identified positives. In our preliminary work, a Breadth-First Search (BFS) algorithm was used to determine which nodes were proximal to the most positives - this was supplanted with an implementation of Dijkstra's algorithm for final output.

Non-Muscle Myosin II (NMII) is an integral effector during Apical Constriction – because of this, we determined it was important to produce a set of outputs with specific relevance to NMII. We did this by implementing a BFS algorithm to find nodes which were relevant to the shortest path between NMII and other positives.

Results

14-3-3epsilon	Hsc70-4	Pkc53E	Ubi-p63E	nonA-l
14-3-3zeta	Hsp83	Rbp9	drk	sbr
Act42A	Myc	Smox	dsh	sgg
Act5C	N	Smurf	fne	spi
Akt1	Nab2	Spn	gro	-
Appl	Nedd4	Stat92E	gskt	.
Cdc5	NetB	Su(dx)	mts	.
CkIIalpha	Pi3K21B	Ubi-p5E	nej	.

Table 1. List of proteins resulting from application of Dijkstra's algorithm.

CG10347	CG7164	His3:CG33839	Nup54	Spn
CG11581	Chi	Khc	Pkn	Ten-m
CG17666	CtBP	Lcp4	Rac1	Ubi-p63E
CG34168	DCTN3-p24	Lsd-2	Sgt	eIF4G2
CG34227	Drep2	NetB	Sin3A	gro

Table 2. List of proteins resulting from application of ST approximation.

14-3-3 epsilon	Acam	Act57 B	Alk	CG10347	CG17666	CG7164	Cam	CanB2	Chi	CkIIalpha	CycB
Cyp4g1	Df31	Diap2	Drep2	Hsp83	Khc	Lsd-2	Moe	N	Patj	RPA2	Rad51D
RfC3	RpA-70	Sema-1a	Sin3A	Snapin	Spn	Spps	Ten-m	Tl	Ubi-p63E	alc	cep290
cher	csw	dally	eIF4G2	ems	esc	eya	flw	p53	qkr58E-1	sgg	wbl

Table 3. List of proteins resulting from a BFS from Positives to Non-Muscle Myosin II

Discussion and Future Directions

Application of our each algorithms (Dijkstra's, ST approx., Shortest Paths) has produced a separate list for proteins of interest. Of these, we received a total list of 107 distinct genes. We would like to emphasize the 24 genes in Table 4 because they appeared in the output of multiple algorithms. Due to time constraints, we were unable to apply a Random Walk with Restarts from NMII. We think this algorithm may be valuable as an additional method of determining a protein's level of association with NMII. Results of output from our Steiner Tree algorithm are approximate and therefore variable. Running and evaluating additional iterations may help us determine the relative frequency with which proteins appear in this output.

Candidate genes/proteins identified by this project will be tested experimentally by students under the supervision of Dr. Applewhite. This data will then be used to revise our models and improve the accuracy of future iterations of this project.

14-3-3zeta	CG7164	CG10347	CG17666	CkIIalpha	Chi
Drep2	Hsp83	Khc	Pi3K21B	'Pkc53E	Spn
Ubi-p63E	Lsd-2	Ten-m	Sin3A	Ubi-p63E	eIF4G2
drk	sgg	mts	Spn	dsh	·

Table 4. List of proteins found in more than one output.

References

Manning, A.J., and Rogers, S.L. (2014). The Fog signaling pathway: Insights into signaling in morphogenesis. Developmental Biology *394*, 6–14.

Acknowledgments

Thanks Anna!