Automated Characterization of Unknown Components in Molecular Pathway Models

Jennifer Hammelman

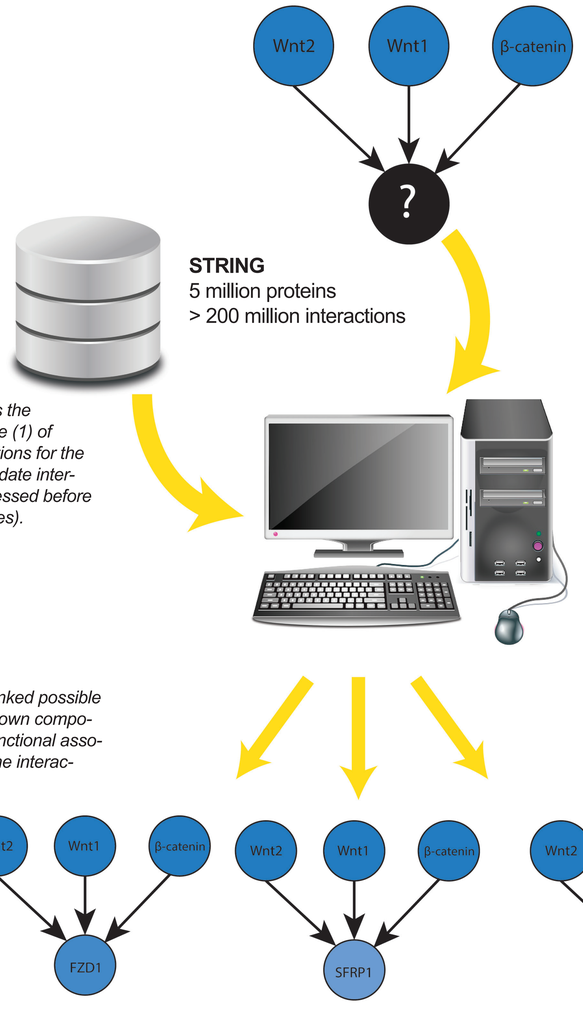
Department of Biology, Center for Regenerative and Developmental Biology, Tufts University

Introduction

A database of known effects of genetic or pharmacological perturbations on regeneration has been composed from scientific literature (2) along with a formalization of phenotype morphology (4) that serve as a platform for computationally derived models of planarian regrowth (3). The goal of this system is to identify gene network models that, when simulated in silico, explain specific morphogenetic behavior. In some cases, such discovered models include unidentified components – functional nodes whose molecular identity is as yet unknown. **My project concerned the identification of candidates for such unlabeled nodes from existing databases of protein-protein interactions based on their connections to known nodes in the proposed network.**

Algorithm

The first implemented function was a simple direct interaction query: given a STRING unique gene name, find all of the genes it regulates. This program was then expanded to allow for intermediate nodes between query gene and result, and multiple gene regulators.



The interactors for the unknown component are entered into the program. The user also specifies how many intermediate nodes are allowed between the interactors and the unknown component.

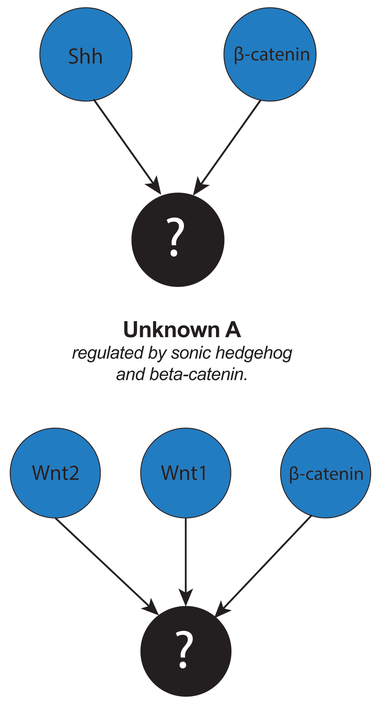
The program searches the STRING v9.1 database (1) of protein-protein interactions for the proteins that are candidate interactors. It was preprocessed before query (time ~30 minutes).

The program performs and intersection of the interactor lists and ranks the candidates by the sum of their functional association scores, as found in the STRING database.

The program returns ranked possible candidates for the unknown component, and reports the functional association and source for the interactions.

Use

An example of this query program was its use to derive potential identities for unknown components (A and B) in planarian head regeneration pathways. The query utilized human homologs for the planarian pathway genes. Potential candidates were scored by functional association.



Unknown A
regulated by sonic hedgehog and beta-catenin.

Unknown A			
Gene	Alias	Shh	β-cat
ENSPTRP00000032807	EGFR	432	966
ENSPTRP00000006874	CCDN1	511	885
ENSPTRP00000007630	CDON	758	586
ENSPTRP00000051157	BOC	732	586
ENSPTRP00000032656	GLI3	843	462

Time = 0.144s

Unknown B
regulated by Wnt1, Wnt2 and beta-catenin.

Unknown B				
Gene	Alias	Wnt2	Wnt1	β-cat
ENSPTRP00000047874	FZD1	954	928	612
ENSPTRP00000053182	SFRP1	913	909	304
ENSPTRP00000053564	NKD1	480	461	566

Time = 0.193s

Conclusion

This unknown component query tool is unique in that it provides a fast way to search for a missing-component among the extensive STRING data without need for internet connection. In the future this tool could be utilized to build an entirely independent system to propose new experiments to validate predictive models.

Acknowledgements

Thank you to Dr. Daniel Lobo and Professor Michael Levin for their guidance in this project.

References

(1) Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, Lin J, Minguez P, Bork P, von Mering C, Jensen LJ (2013) STRING v9.1: protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res 41:D808– D815.
(2) Lobo D, Beane W, Levin M (2012) Modeling Planarian Regenera- tion: A Primer for Reverse-Engineering the Worm. Plos Comput Biol 8(4):doi:10.1371/journal.pcbi.1002481.
(3) Lobo D, Levin M. (2015) Inferring regulatory networks from experimental morphological phenotypes: a computational method reverse-engineers planarian regeneration. PLoS Computational Biology, in press.
(4) Lobo D, Malone TJ, Levin M (2012) Towards a bioinformatics of patterning: a computational approach to understanding regulative morphogenesis. Biol Open 2:156–169.