Poster J-2

Analysis of pleiotropy during C. elegans development and interpretation with interactome networks

Authors:
Lihua Zou (*Harvard University*)
Brian Ross (*Whitehead Institute*)
Jun Liu (*Harvard Univeristy*)
Hui Ge (*Whitehead Institute*)

Short Abstract: Pleiotropy refers to the phenomenon of a single gene controlling multiple phenotypes. We analyze phenotypic data from C. elegans development and identify pleiotropic proteins. We develop a model to estimate the importance of a given protein to the propagation of biological signals in interactome networks. We propose that pleiotropic proteins act as “information exchange centers” between complexes or pathways.

Long Abstract:

Introduction

Pleiotropy refers to the phenomenon of a single gene controlling multiple distinct and seemingly unrelated phenotypic effects. Pleiotropy reflects the fact that some genes perform multiple biological functions. Traditionally, loss-of-function phenotypes are examined for individual genes. The recent availability of high-throughput biological datasets may lead to the opportunity for systematic identification of pleiotropic genes. However, it is still not clear how effects arising from loss of a single function can be differentiated from the effects of losing multiple functions. It is also a challenge to provide mechanistic interpretations for such phenotypic complexity.

C. elegans is a free-living soil nematode extensively studied in the field of developmental biology. Because of its known genome sequence, well-characterized anatomy and the convenience of the RNA interference (RNAi) technique, C. elegans is especially amenable to genome-wide loss-of-function analysis. Sonnichsen et al performed whole-genome RNAi experiments and identified 661 genes involved in early embryogenesis [1]. A series of phenotypic descriptors were used to represent the absence and presence of cellular defects. We aim to uncover pleiotropic genes during C. elegans early development. We also investigate if a global mechanistic interpretation can be reached for the phenomenon of pleiotropy. Since many cellular events in early development may be mediated by protein-protein interactions, we examine the properties of pleiotropic proteins in interactome networks [2].

Methods

In this study, we present two novel methods to analyze high-throughput biological datasets. First, we define “pleiotropy index”, a measure for phenotypic complexity in RNAi experiments. Second, we develop a model to estimate the importance of a given protein to the propagation

of signals in a cell using interactome networks obtained by protein interaction mapping techniques such as yeast two-hybrid experiments.

I. Pleiotropy Index

Pleiotropy cannot be readily measured by counting the number of phenotypic descriptors, because some descriptors are highly correlated with one another. In other words, several descriptors may together correspond to a phenotype arising from a single function. Conventional phenome analysis assigns each gene exclusively to one phenotypic cluster, which does not serve our goal of identifying pleiotropic genes either. Also, the genes grouped together by clustering methods may not be functionally coherent.

We use pre-defined functional categories as seeds [1] to identify combinations of phenotypic descriptors, or “signatures” of defects, which are overrepresented in each functional category. We annotate the genes involved in early embryogenesis with these signatures, based on the criteria that the phenotypic descriptors of a gene cover more than half of the descriptors in a signature. Using this method, a gene can be annotated with multiple signatures, which indicates that loss-of-function of the given gene results in multiple phenotypes. We define the number of signatures of a gene as its pleiotropy index.

II. Information flow model

Interactome networks provide a global view of how proteins are connected with one another, where nodes represent proteins and edges represent interactions [2]. We develop a method to model the communication of biological signals through interacting proteins in a network. Construing the propagation of these signals as “information flow”, our method identifies proteins that are central to the transmission of information throughout the network.

Given a pair of nodes in a network, we designate one as a source node and the other as a sink node. Biological information can flow from the source node to the sink node through paths in the rest of the network. The proportion of information flow through individual nodes in the rest of the network can be calculated based on the following assumptions: 1) the proportion of flow into a node is equal to that out of a node; 2) when there are multiple parallel, non-overlapping paths between two nodes, the flow through each path is inversely proportional to the number of edges composing the path.

We repeat this calculation procedure for all pair-wise combinations of source/sink nodes in the network. The amount of information flow through an individual node is defined as the average proportion of flow through the node over all source/sink combinations. This definition is based on the assumption that there is equal amount of information exchange between any pairs of nodes in the network.

Results

We annotate genes involved in early development with phenotypic signatures representing various functional categories. Genes previously assigned to a single functional category may now be annotated with multiple signatures, yet genes bearing a common signature show significant functional enrichment. We identify pleiotropic genes involved in C. elegans early

development. Many kinases, for example, are highly pleiotropic by our index. For some genes which have been characterized by mutant analysis, the pleiotropy is supported by previous evidence from the literature. By comparing the pleiotropy index distribution with that of randomly permuted phenotypic datasets, we find that pleiotropy occurs at high frequency during early development.

We investigate whether a global mechanistic interpretation can be reached for the phenomenon of pleiotropy. In early embryonic cell divisions where zygotic transcription has not started yet, many biological processes depend on maternal proteins. Protein-protein interactions may play a major role in early embryos to mediate cellular events. We find that the pleiotropy index is positively correlated with the amount of information flow in interactome networks. This suggests that perturbation of proteins with high information flow causes pleiotropy. Highly pleiotropic proteins may act as “information exchange centers” between different complexes or pathways, which is a fundamental reason for the complexity of their loss-of-function phenotype.

[1] Sonnichsen B et al (2005). Full-genome RNAi profiling of early embryogenesis in Caenorhabditis elegans. Nature 434:462-469.
[2] Li S et al (2004). A Map of the Interactome Network of the Metazoan C. elegans. Science 303:540-543.