ROLE OF SYNTHETIC GENETIC INTERACTIONS IN UNDERSTANDING FUNCTIONAL INTERACTIONS AMONG PATHWAYS

S. Mohammadi G. Kollias A. Grama

Department of Computer Science Purdue University

The Pacific Symposium on Biocomputing, 2012



OUTLINE



- Models and Definitions
- Datasets

2 MOTIVATION

- Functional Similarity of Gene Pairs
- Performance Evaluation and Identifying Shortcomings

3 CONSTRUCTING KEGG CROSSTALK MAP

- Overview
- Methods
- Results and Discussions



Models and Definitions Datasets

OUTLINE



- Models and Definitions

- Functional Similarity of Gene Pairs
- Performance Evaluation and Identifying Shortcomings

- Overview
- Methods
- Results and Discussions



Models and Definitions Datasets

GENETIC INTERACTOME

Double mutants exhibit unexpected phenotypes, as compared to joint single mutations.

DEFINITION

- Negative Interactions: more sever phenotype than expected
 - Also known as aggravating or synergistic
- Positive Interactions: Less sever phenotype than expected
 - Also known as alleviating or epistatic

Most commonly used:

PHENOTYPE : Growth rate

MODEL : Multiplicative null model



Models and Definitions Datasets

ORGANIZATION OF GENETIC INTERACTIONS

DEFINITION

- Between-Pathway Model
 - Among genes participating in redundant functions
- Within-Pathway Model
 - Among genes with additive effect
- Indirect Effect
 - Among genes with distant functions that are not directly related



Models and Definitions Datasets

BETWEEN-PATHWAY MODEL (BPM)



- Bi-cliquish structure
- Have been used to:
 - Predict co-pathway membership of gene pairs
 - Extract redundant pathways



Models and Definitions Datasets

OUTLINE



- Models and Definitions
- Datasets

2 MOTIVATION

- Functional Similarity of Gene Pairs
- Performance Evaluation and Identifying Shortcomings

3 CONSTRUCTING KEGG CROSSTALK MAP

- Overview
- Methods
- Results and Discussions



Models and Definitions Datasets

THE GENETIC LANDSCAPE OF A CELL



- Baker's yeast, Saccharomyces cerevisiae
- Synthetic Genetic Array (SGA)
- 1712 query genes
 - 1378 null alleles of non-essential genes
 - 334 hypomorphic or conditional alleles of essential genes
- 3885 array strains



Adopted from Costanzo et al., 2010

Models and Definitions Datasets

FUNCTIONAL ANNOTATIONS



- KEGG Pathway Database
- Annotations for 1026 genes in the experiment
- 96 Pathways
 - 80 pathways after filtering pathways with less than 10 genes.



Functional Similarity of Gene Pairs Performance Evaluation and Identifying Shortcomings

OUTLINE

1 PRELIMINARY

- Models and Definitions
- Datasets

2 MOTIVATION

- Functional Similarity of Gene Pairs
- Performance Evaluation and Identifying Shortcomings

3 CONSTRUCTING KEGG CROSSTALK MAP

- Overview
- Methods
- Results and Discussions



Functional Similarity of Gene Pairs Performance Evaluation and Identifying Shortcomings

LOCAL NEIGHBORHOOD SIMILARITY A Predictor of Co-Pathway Membership

SIMILARITY PREDICTION METHODS

- Number of Shared Neighbors
- Ongruence Score
- Pearson Correlation of Interaction Profiles





Both v_i and v_j have three shared neighbors. However, in the first case their congruence score is almost 0.6, while in the second case it is approximately 2 (assuming a graph of size 10).



Functional Similarity of Gene Pairs Performance Evaluation and Identifying Shortcomings

OUTLINE

1 PRELIMINARY

- Models and Definitions
- Datasets

2 MOTIVATION

- Functional Similarity of Gene Pairs
- Performance Evaluation and Identifying Shortcomings

3 CONSTRUCTING KEGG CROSSTALK MAP

- Overview
- Methods
- Results and Discussions



Functional Similarity of Gene Pairs Performance Evaluation and Identifying Shortcomings

EVALUATING RANKING METHODS

Given a pathway P_A and a cut size (target set) *I*.

DEFINITION

$$P-value(X = k) = Prob(k \le X)$$

= $HGT(k|N, N_A, I)$
= $\sum_{x=k}^{min(N_A, I)} \frac{C(I, x)C(N - I, N_A - x)}{C(N, N_A)}$

X: Random variable denoting the number of true positives in a random sample, N: Total number of gene pairs, N_A : Number of gene pairs in pathway A, I: Size of target set

Functional Similarity of Gene Pairs Performance Evaluation and Identifying Shortcomings

MINIMUM HYPERGEOMETRIC (MHG) SCORE

Target size unknown:

DEFINITION

The Minimum HyperGeometric (mHG) score

$$mHG(\lambda) = min_{1 \le l \le N}HGT(b_l(\lambda); N, N_A, l)$$

where $b_i(\lambda) = \sum_{i=1}^l \lambda_i$

 λ_i is 1 if both of the genes in the *i*th ranked gene pair are members of P_A , and 0 otherwise.

mHG Adjusted for Multiple Comparison



Functional Similarity of Gene Pairs Performance Evaluation and Identifying Shortcomings

PREDICTIONS ARE NOT EQUALLY ACCURATE IN DIFFERENT KEGG PATHWAYS



Overview Methods Results and Discussions

OUTLINE



- Models and Definitions
- Datasets

2 MOTIVATION

- Functional Similarity of Gene Pairs
- Performance Evaluation and Identifying Shortcomings

3 CONSTRUCTING KEGG CROSSTALK MAP

Overview

- Methods
- Results and Discussions



Overview Methods Results and Discussions

HIGHLIGHTS

BASIC IDEA		
Heterogeneous performance of co-pathway membership predictions	\Leftrightarrow	Existence of specific structure around enriched pathways

- Decomposing neighborhood of each pathway
- Inferring lethal crosstalk among pathways



Overview Methods Results and Discussions

OUTLINE



- Models and Definitions
- Datasets

2 MOTIVATION

- Functional Similarity of Gene Pairs
- Performance Evaluation and Identifying Shortcomings

3 CONSTRUCTING KEGG CROSSTALK MAP

Overview

Methods

Results and Discussions



Overview Methods Results and Discussions

MODIFIED CONGRUENCE SCORE (MCS) Evaluating Neighborhood Overlap of Gene Pairs With Respect to a Given Pathway

DEFINITION

$$\begin{array}{lll} \textit{P-value}(\textit{X}=\textit{k}_{ij}^{\textit{B}}) & = & \textit{Prob}(\textit{k}_{ij}^{\textit{B}} \leq \textit{X}) \\ & = & \textit{HGT}(\textit{k}_{ij}^{\textit{B}} | \textit{n}_{\textit{B}}, \textit{d}_{i}^{\textit{B}}, \textit{d}_{j}^{\textit{B}}) \\ & = & \sum_{\textit{x}=\textit{k}_{ij}^{\textit{B}}} \frac{\textit{C}(\textit{d}_{j}^{\textit{B}}, \textit{x})\textit{C}(\textit{n}_{\textit{B}} - \textit{d}_{j}^{\textit{B}}, \textit{d}_{i}^{\textit{B}} - \textit{x})}{\textit{C}(\textit{n}_{\textit{B}}, \textit{d}_{i}^{\textit{B}})} \end{array}$$

MCS is defined as $-log_{10}$ of the P-value.



Overview Methods Results and Discussions

MODIFIED CONGRUENCE SCORE (MCS)

EXAMPLE



A sample neighborhood configuration for v_i and v_j . Here n = 15, $D_i = 6$, $D_j = 5$, $n_B = 6$, $d_i = 3$, $d_j = 4$ and k = 2.



Overview Methods Results and Discussions

CONSTRUCTING NEIGHBORHOOD OVERLAP GRAPH FOR A GIVEN PATHWAY PAIR

DEFINITION

The neighborhood overlap graph (NOG) of a given pathway P_A with respect to pathway P_B , denoted by $H_{A \rightarrow B} = (V_H, E_H)$, is an unweighted, undirected graph defined over same vertices as P_A . In this graph, there is a link between vertices v_i and v_j if the network structure around them with respect to P_B is statistically significant.



Overview Methods Results and Discussions

PRUNING NEIGHBORHOOD OVERLAP GRAPH, FINDING COHESIVE SUBGRAPHS, AND IDENTIFYING INTERACTION PORTS



- Iterative peeling of K-shells Pruning hairy components
- Connected components in each core
- Evaluating the significance of components
 - Evaluating significance using ER random graph model



Overview Methods Results and Discussions

OUTLINE



- Overview
- Methods
- Results and Discussions



Overview Methods Results and Discussions

KEGG CROSSTALK MAP



Overview Methods Results and Discussions

INTERACTION PORT CASE STUDY CROSSTALK BETWEEN PROTEIN PROCESSING IN ER AND PROTEASOME

Summary



Mohammadi et al.

Functional Decomposition of Genetic Networks

SUMMARY

- The local neghborhood similarity gives heterogeneous performance in predicting co-pathways membership of gene pairs.
- This phenomena is due to the specific structure around enriched pathways
- Decomposing the neighborhood around each pathway sheds light on the cellular machinery.
- Future works:
 - Analysing the hierarchy of ports instead of the most significant interaction port.
 - Using our methodology to uncover dependencies among functional pathways.

FOR FURTHER READING I

M. Costanzo et al. The Genetic Landscape of a Cell Science, 425 2010.

SJ. Dixon et al. Systematic Mapping of Genetic Interaction Networks Annual review of genetics, 43, 601 2009.

D. Segre et al. Modular Epistasis in Yeast Metabolism Nature Genetics, 37, 77 2005.

C.L. Tucker and S. Fields Lethal Combinations Nature Genetics, 35, 204 2003.

