

Statistical Factor Analysis for Complex Pathway Decomposition

Jeffrey T. Chang^{1*}, Carlos M. Carvalho², Quanli Wang^{1,2}, Haige Shen², Mike West², Joseph R. Nevins¹

¹Institute for Genome Sciences & Policy

²Institute of Statistics & Decision Sciences

Duke University, Durham, NC

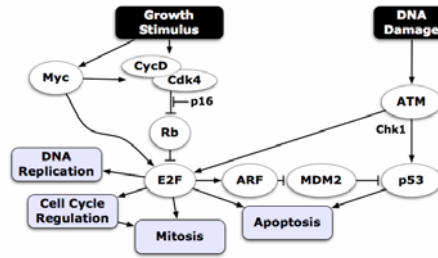
jeffrey.chang@duke.edu, carlos@isds.duke.edu

Abstract

Proper function of the Rb tumor suppressor protein is believed to be compromised in nearly all breast tumors. Therefore, the Rb pathway is an area of intense interest for cancer research. pRb controls the cell cycle transition from G1- to S-phase through its suppression of the E2F transcription factors. Disruption of the pathway disables the normal function of the cell cycle checkpoint and leads to the uncontrolled cellular proliferation observed in cancer cells. Detailed knowledge of the molecular components comprising the Rb/E2F pathway would provide valuable insight into the functioning of tumors and suggest intriguing new targets for anti-cancer drugs.

To understand the composition of the Rb/E2F pathway, we have developed a sparse latent factor model that sequentially identifies pathways from gene expression data and explores their relationships. We use an *evolutionary* approach that starts with a carefully defined initial list of Rb/E2F pathway genes and iteratively expands to associate novel genes with the pathway. The model explicitly separates signal from noise, leaving a clean pathway signature.

We previously collected microarray readings from 148 breast cancer tumors exhibiting diverse clinical phenotypes. We applied the factor analysis to the data and found 33 potential E2F-related pathways. We illustrate the potential of this decomposition with an example that relates the activities of the ER and Cyclin D pathways.



- The E2F family of transcription factors has a central role in controlling genes involved in the G1/S transition.
- E2F is linked to many functions, such as DNA replication and apoptosis.
- Activity may vary due to extracellular signals, internal conditions (cell cycle), or genetic background.

Problem: How do we organize E2F-related genes into components?

Pathways in Breast Tumors

Microarray Datasets:

148 Breast cancer tumor samples exhibiting diverse clinical pathologies and outcomes.

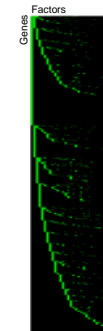
E2F ectopically expressed in various cell lines, multiple experiments.

Ectopically expressed oncogenes in human mammary epithelial cells.

Method:

Begin with 101 E2F pathway genes.

Evolve and decompose into factors based on activity observed in breast cancer.



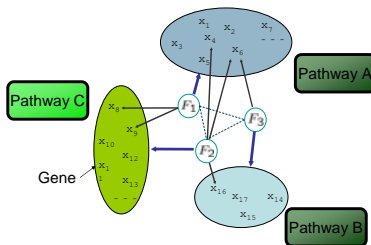
Results:

- Identified 1000 total genes.
- Decomposed into 33 factors exemplified by the activity of the following genes:

FOS	AZIN1	CCL5	IGHA1
TFF3	SCAMP5	G1P2	SPP1
---	CASP4	NR4A1	PIP
COL16A1	KRT14	NME1	
AEBP1	COL11A1	IGL@	
TEGT	---	ERBB2	
EEF1A2	MYB	ERBB1	
CKS2	HSP90A	---	
FALZ	COX5B	GGT1	
IGHA1	CDH11	FOS	

Factor Models to Identify Pathway Branches

1. Factors define the structure and relationships of pathways.



3. A Bayesian MCMC method robustly decomposes gene expression data into factors.

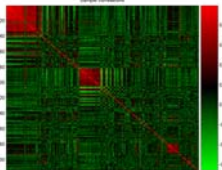
$$X = \text{gene expression profile} + \text{noise}$$

$$X_g = \sum_i \beta_{gi} F_i + \epsilon_g$$

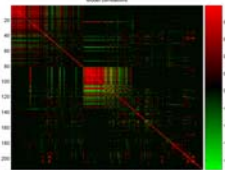
expression pattern

2. The models are able to clean up noise in the sample correlations.

Sample correlations (original data)

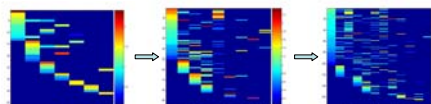


Fitted correlations in BTB'+Psi (fitted data)

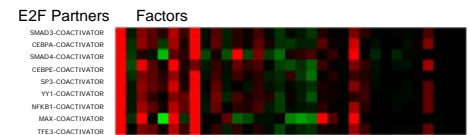


4. The method includes and evolutionary search that explores and enriches the pathways.

- Start with a core set of genes – estimate the initial factors.
- Sequentially add genes that “share” association with current pathways.
- Allow for the inclusion of additional factors creating new subpathways.

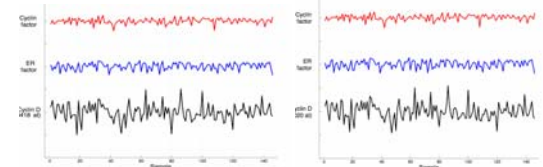


Factors are Related to E2F Co-Activators.



See Chen et al. poster for details about E2F Co-Activators.

Factors Find a Relationship Between Cyclin D and ER.



Reported in literature:

- ER-mediated upregulation of cyclin D1 seems to be a possible mechanism of maintaining cell proliferation in DCIS in case of EGFR- and HER-2/neu-negativity (Lebeau et al., 2003).
- ... cyclin D1 as a CDK-independent activator of the estrogen receptor (Zwijsen et al., 1997)

Conclusions and Future Work:

- The Rb/E2F pathway is complex and the development of incisive analytic methods is required to dissect it.
- A deeper understanding of the structure of the data can clarify methods to calibrate and sharpen the models of pathways.
- Biochemical validation of pathway activity is necessary.