Identification of Genes that can Selectively Kill Cancer Cells Using Boolean Implications

Yang Gao[1], Subarna Sinha[2], David L. Dill[2]

1: Tsinghua University, Haidian, Beijing 100084, China; 2: Stanford University, 450 Serra Mall, Stanford, CA 94305

Motivation

- Many cancer treatment drugs do not target only cancer cells, which makes treatment less effective as well as generates serious side effects.
- Identifying drugs that only target cancer cells requires expensive and time-consuming experiments.
- Computational techniques can immensely aid in the process.
- We predict candidate genes whose altered expression can be lethal to cancer cells with a particular genetic alteration.

Background

- **Boolean Implications**[1] capture asymmetrical relations between pairs.
- Statistical test is used to test sparseness of each quadrant.
- **Synthetic lethality (SL)** arises when a combination of mutations in two genes leads to cell death, whereas a mutation in only one of these genes does not.

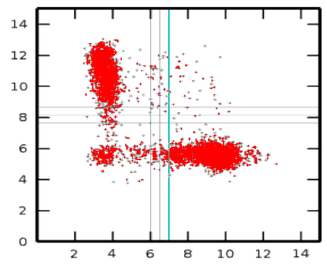

Fig. 1. Example of A high ⇒ B low. X, Y axes are expression level of gene A and B.

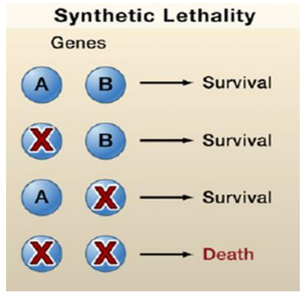

Fig. 2. Synthetic lethality [2]

- The inhibition of synthetic lethal partners of a frequently altered gene in cancer can selectively kill the cancer cell.
- HILO implication (mutual exclusion) has to exist between the alteration of two genes that have a SL interaction.
- Large cancer sequencing projects such as TCGA can be mined to find genetic alterations that are mutually exclusive with a particular genetic alteration.

Methodology and Results

Understanding Known Synthetic Lethal Relations

- Identify informative features that can distinguish SL gene pairs from the others.
- We've collected and plotted 45 known synthetic lethal relationships from literature(Fig. 3).

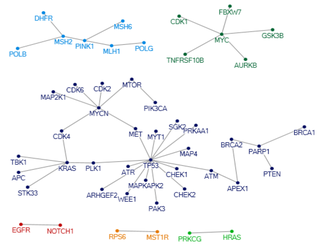

Fig. 3. 45 known SL relations.

1. Analysis using Boolean Implication

- HILO relationship means that two genes are rarely mutated together, thus exhibit mutual exclusion relation, which is expected between SL gene pairs.
- We examined the 45 relations, and found that 10 of them are HILO relationships, which shows HILO implications are useful for predicting SL relations.
- Many of the genes in the 45 pairs had low deletion frequency in the BRCA dataset and hence could not be detected using a statistical test.

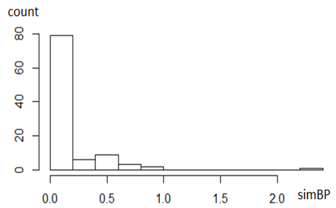

Fig. 4. Dist. of 100 similarity scores between random gene pairs with HILO relation.

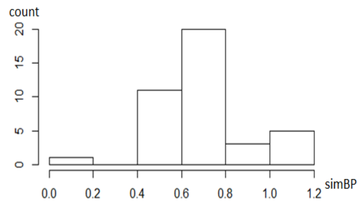

Fig. 5. Dist. of known SLs similarity scores.

2. Analysis using Gene Similarity

- We analyze these gene pairs by examining their similarity using Lin's method[3].
- Fig. 4 and Fig. 5 are distributions of similarity scores in non-SLs and SLs. We can see a great difference in the peak between the two histograms. Thus a threshold of 0.5 can be used to roughly distinguish between them.

3. Analysis using Pathways Information

- We have searched KEGG pathways that the SL gene pair belong to at the same time.
- Out of 45 known SLs, 4 pairs have more than 3 common pathways, 11 have 1~2 in common, while the rest 30 don't have those in common.

Methodology and Results

Predicting New SL Relations

- According to the analysis, we use Boolean Network and similarity to predict SLs.

Fig. 6. Flow chart of SL prediction pipeline

- Prediction of important genes in cancer

Gene	Alternations num	HILO num	Predicted SL num
TP53_M_mut	155	2051	**310**
TP53_N_mut	34	1315	**197**
TP53_FSD_mut	33	914	**139**
TP53_SS_mut	18	291	**39**
BRCA1-del	61	291	**173**
MYC-amp	290	10	**7**

Fig. 7. Prediction summary of gene TP53, BRCA1 and MYC

Conclusions and Future Work

Conclusions

- HILO implications derived from large cancer sequencing projects such as TCGA can be used to identify novel synthetic lethal relations.
- Functional buffering between SL relation can be captured by Lin's similarity measure[3].
- The prediction pipeline is generic and can be applied on the other cancer types.

Future Work

- Find more features to capture SL properties. Candidates are: 2hop physical-SL feature, gene expression data, protein-protein interaction network, etc.
- Collect experimentally verified non-SL relations, and analyze the difference compared with SL relationships.
- Develop machine learning algorithm to build a classifier.

References

[1] : D. Sahoo, D. Dill, A. Gentles, R. Tishirani and S. Plevrithis, *Boolean implication networks derived from large scale, whole genome microarray datasets*, Genome Biology, Oct 2008.
[2]: Alan Ashworth, Christopher J. Lord, and Jorge S. Reis-Filho, *Genetic Interactions in Cancer Progression and Treatment,* Cell 145, April 1, 2011
[3]: Lin D. An information-theoretic definition of similarity, Proceedings of the Fifteenth International Conference on Machine Learning. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.; 1998. p. 296-304

Acknowledgements

This investigation was supported by NIH/National Cancer Institute (NCI) grant 1U54 CA149145 (to DLD and SS) and UGVR program. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Further Info

For any further info, please contact us via:
Yang Gao: gaoyang10@mails.tsinghua.edu.cn
Subarna Sinha: subarnas@stanford.edu
David L. Dill: dill@cs.stanford.edu