

DGAP Protocol

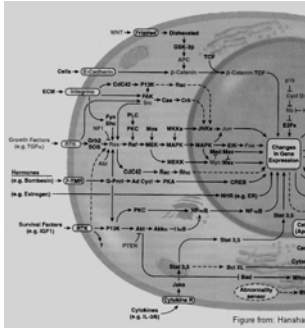
Introduction to Relevance Networks

Atul Butte, MD

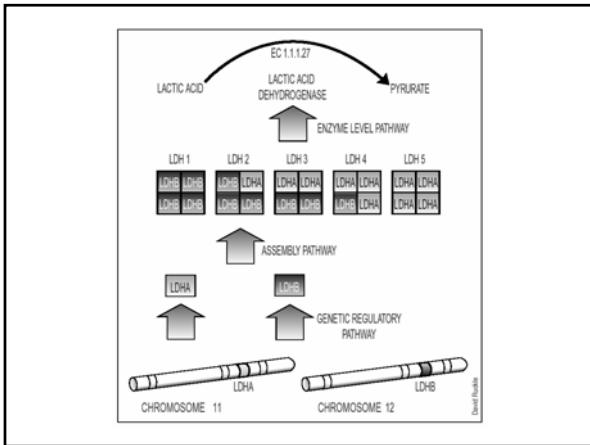
November 18, 2003

Relevance Networks

- How can we find gene regulatory networks or physiological regulatory networks with little or no *a priori* knowledge (unsupervised learning)
- How can we link microarray measurements to clinical measurements?
- Relevance Networks are an approach to analyze these data sets



Butte AJ, Kohane IS, Unsupervised Knowledge Discovery in Medical Databases Using Relevance Networks, Symposia AMIA, 1999.
 Butte and Kohane, Children's Hospital, Patent Pending.





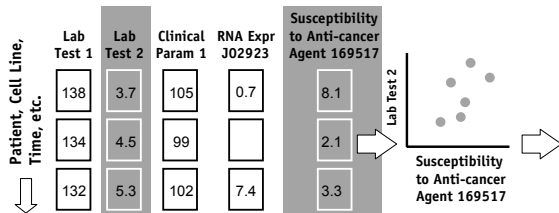
Construction of Relevance Networks 1

- Patients and cell lines are analyzed as cases
- Clinical parameters, laboratory tests, RNA expression, and susceptibility to anti-cancer agents are all example features of those cases
- Features include functional genomics as well as phenotypic

Cases Patient, Cell Line, Time, etc.	Features				Susceptibility to Anti-cancer Agent 169517
	Lab Test 1	Lab Test 2	Clinical Param 1	RNA Expr J02923	
138	3.7	105	0.7	8.1	
134	4.5	99		2.1	
132	5.3	102	7.4	3.3	

Construction of Relevance Networks 2

- For all pairs of features, we take overlapping values over the cases and make a scatter plot of values



Construction of Relevance Networks 3

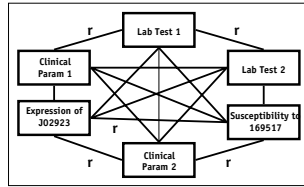
- Perform a pairwise comparison between all features
- For each scatter plot, we compute how similar the two features are (dissimilarity measure)
 - Fit a linear model and store correlation coefficient r



- Every feature is completely connected to every other feature by a dissimilarity measure (e.g. a linear model)

Construction of Relevance Networks 4

- Choose a threshold r to split the network
- Drop links with r under threshold
- Breaks the completely connected network into islands where connections are stronger than threshold
- Islands are what we call “relevance networks”
- Display graphically, with thick lines representing strongest links



Software available at www.chip.org/relnet

Butte AJ, Tamayo P, Slonim D, Golub TR, Kohane IS. Proceeding of the National Academy of Science, 97:12182, 2000.

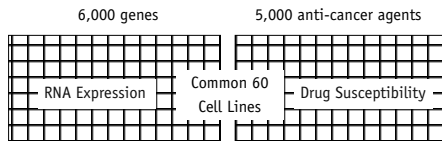
NCI 60 data set

- NCI 60 is set of 60 cancer cell lines against which the National Cancer Institute has tested over 50,000 chemicals to find anti-cancer agents
- In collaboration, NCI provided us with re-validated data for 5,084 agents
- Concentration of each agent causing 50% growth inhibition compared to control (GI50) for each cell line in NCI 60
- Data was provided as $-\log_{10}(\text{GI50})$
 - Lower number means higher GI50, or less sensitivity to an agent
- Unfortunately, few of these anti-cancer agents have documented mechanisms of action (or even a name listed)

Weinstein, et al. Science; 258: 447 (1992).
<http://www.dtp.nci.gov>

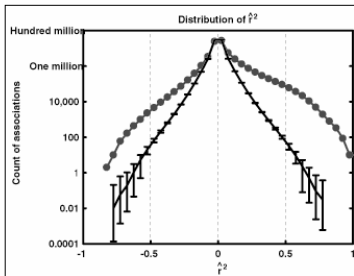
RNA expression data set

- RNA expression in NCI 68 cell lines was determined using Affymetrix HU6000 arrays in collaboration with Whitehead Institute
 - 5,223 known genes
 - 1,193 expressed sequence tags
- The RNA expression data set and Anti-cancer susceptibility data set were merged, using the 60 cell lines the two tables had in common



Distribution of r^2

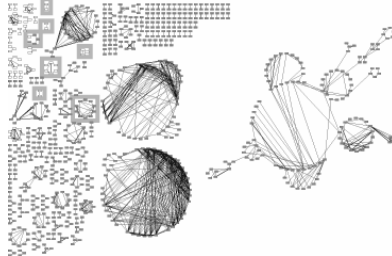
- 11,692 features
- 68,345,586 total associations
 - 22 M between genes
 - 12 M between agents
 - 33 M between a gene and an agent



Entropy →

Genes and Anti-Cancer Agents

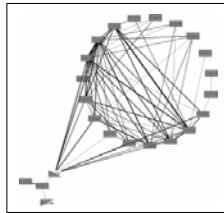
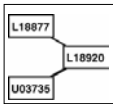
- Threshold r^2 was 0.8
- 202 networks
- 834 features out of 11,692 (7.1%)
- 1,222 links out of 68,345,586 (.0018%)
- Only one link between a gene and anti-cancer agent



Butte AJ, Tamayo P, Slonim D, Golub TR, Kohane IS. Proceeding of the National Academy of Science, 97:12182, 2000.

Taxonomy of Network Links

- Identity / Synonymy
 - L10838 and D28423: SRP20
 - M19267 and Z24727: Tropomyosin alpha chain
- Functional Similarity
 - Melanoma-associated antigens 2, 3 and 12
 - Caldesmon 1 and alternative splicing products 3 and 4
- Derivation
 - 295500, 374028, 606497, 606499, 610456, 610457, and 610459 are all camptothecin derivatives
- Biological Relationship
 - Keratin 8 linked to keratin 18



Genes and Anti-Cancer Agents

- Elevated levels of J02923 (lymphocyte cytosolic protein-1, LCP1, L-plastin, pp65) is associated with increased sensitivity to agent 624044
- Agent 624044 is 4-Thiazolidinecarboxylic acid, 3-[[6-[2-oxo-2-(phenylthio)ethyl]-3-cyclohexen-1-yl]acetyl]-2 thioxo-, methyl ester, [1R-[1a(R*),6a]]- (9CI)
- LCP1 is an actin-binding protein involved in leukocyte adhesion
- LCP1 is thought to play a role in tumorigenicity
- LCP1 is expressed in most human cancer cell lines
- Other thiazolidine carboxylic acid derivatives are known to inhibit tumor cell growth

