Deciphering breast cancer metastasis using protein networks

Han-Yu Chuang[1*], Eunjung Lee[2,3*], Vagisha Sharma[4], Yu-Tsueng Liu[5], Doheon Lee[3] and Trey Ideker[1,2+]

UCSD, [1]Bioinformatics Program, [2]Dept. of Bioengineering, [4]Dept. of Computer Science and Engineering, and [5]Moores Cancer Center, La Jolla, CA 92093

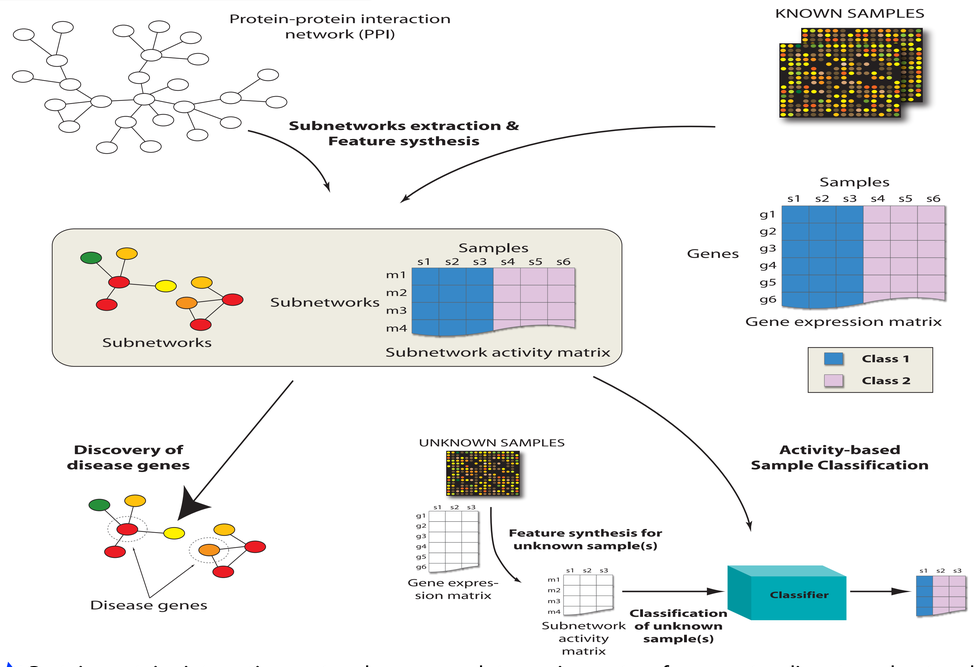
[3]Korea Advanced Institute of Science and Technology, Daejeon 305-701, Korea

[*]These authors contributed equally to this work [+]To whom correspondence should be addressed. Emai: trey@bioeng.ucsd.edu

Introduction

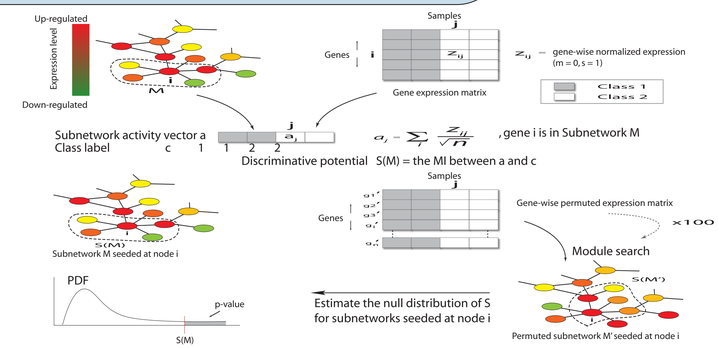
Mapping the pathways that give rise to metastasis is one of the key challenges of breast cancer research. Recently, several large-scale studies have shed light on this problem through analysis of gene expression profiles to identify markers correlated with metastasis. However, each study identifies a different set of marker genes, and it remains unclear how these genes interrelate within a larger functional network. Here, we apply a protein-network-based approach that identifies markers not as individual genes but as subnetworks extracted from protein interaction databases. The resulting subnetworks identify new putative cancer genes and provide novel hypotheses for pathways involved in tumor progression. Although genes with known breast cancer mutations are typically not detected through analysis of differential expression, they play a central role in the protein network by interconnecting many expression-responsive genes. Beyond suggesting new pathways, we further demonstrate the accuracy of subnetwork markers in the classification of metastatic versus non-metastatic tumors.

Research Design



- Protein-protein interaction networks are used to assign sets of genes to discrete subnetworks. (see Subnetwork Extraction)
- Gene expression profiles of tissue samples drawn from each type of cancer (i.e., metastatic or non-metastatic) are transformed into a "subnetwork activity matrix".
- Subnetworks and the activity matrix are then used to identify disease genes and also used to train a classifier for predicting new unknown samples.
- Datasets: 57,235 interactions among 11,203 proteins and the expression profiles of 295 breast cancer patients previously reported by van de Vijver *et al*, *NEJM* 2002.

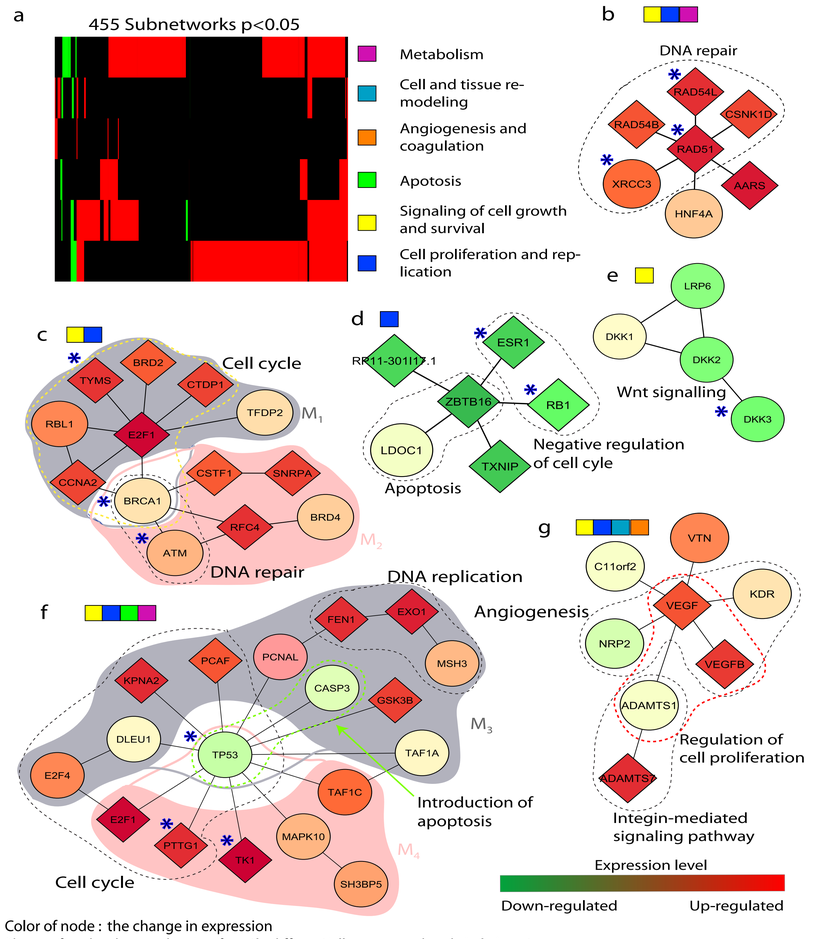
Subnetwork Extraction



For a given subnetwork M in the interaction network, the activity is a combined z-score derived from the expression of its individual genes.

After overlaying the expression vector of each gene on its corresponding protein in the interaction network, subnetworks with discriminative activities are found via a greedy search. Significant subnetworks are selected based on a null distribution estimated from 100 permuted subnetworks (seeded from the same node). The symbols g1′, g2′, …, gn′ represent genes in permuted expression matrices.

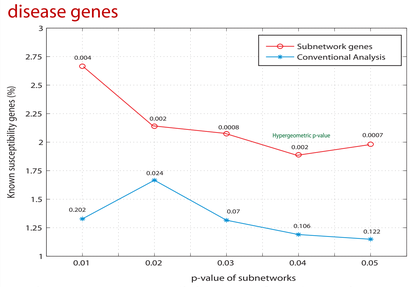
Subnetworks Enriched for Hallmarks of Cancer



Color of node : the change in expression
Shape of node : diamond is significantly differentially-expressed and circle is not.
Known breast cancer susceptibility genes are marked by a blue asterisk

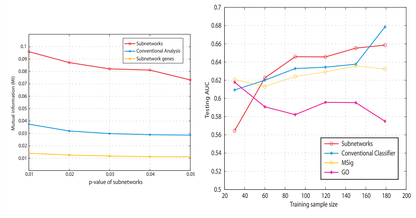
Performance

Subnetworks are informative of non-discriminative disease genes



- 59 breast cancer susceptibility genes from OMIM, de Jong MM *et al* (*J. Med. Genet.* 2002) and Lymberis *et al* (*Pharmacogenomics* 2004)
- 88/455 subnetworks contain known cancer susceptibility genes, and 13 of them have > 1 known gene.
- Disease genes that can only be detected using network information include BRCA1, BRCA1 associated RING domain 1 (BARD1), p53 (TP53), estrogen receptor ? (ESR1), nuclear receptor co-activator 3 (NCOA3), ATM, and XRCC3.
- ~94% of the subnetworks had higher activity levels in metastatic breast tumors than in non-metastatic ones.
- ~33% of subnetworks contained member genes with divergent directions of expression change.

Subnetworks demonstrate promising accuracy for sample classification



ACKNOWLEDGEMENTS

Funding for this work was provided by Unilever, PLC and the National Institute of Environmental Health Sciences (ES14811-01). TI was additionally supported by a David and Lucille Packard Fellowship, EL was partly supported by a Korean Research Foundation Grant (KRF-2005-213-C00042)), and DL was supported by the Korean National Research Laboratory Grant (2005-01450) from the Ministry of Science and Technology.