

The Multi-Node Topological Overlap Measure For Gene Neighborhood Analysis

Ai Li*, Steve Horvath*[‡]

*Human Genetics, David Geffen School of Medicine
Gonda Center, University of California, Los Angeles, CA 90095

[‡]Biostatistics, School of Public Health, UCLA

Correspondence: shorvath@mednet.ucla.edu

Abstract—Defining the neighborhood of an initial set of nodes is an important task in network analysis. For example, we show that the neighborhood of an initial set of brain cancer related genes is highly enriched with other cancer genes as well. It remains an active area of research to define a biologically meaningful concept of neighborhood in gene or protein networks. In gene networks, genes with high topological overlap have been found to have an increased chance of being part of the same biological pathway. Since our main interest lies in gene and protein networks, we propose a generalization of the topological overlap matrix to define the neighborhood of a set of genes. A pair of nodes in a network is said to have high topological overlap if they are both strongly connected to the same group of nodes. We generalize the standard pairwise topological overlap measure to multiple nodes. The resulting neighborhoods are comprised of tightly interconnected nodes. We provide empirical evidence that a neighborhood surrounding an initial set of 2 nodes can be far more informative than the neighborhood of a single node. Using a cancer- and a yeast network applications, we provide empirical evidence that the multiTOM approach yields biologically meaningful results and compares favorably to alternative approaches. Our approach is implemented in the freely available multiTOM software package.

I. INTRODUCTION

Network methods are a valuable tool for realizing the goals of systems biology: they may enhance the design of effective drugs and our understanding of the mechanisms behind an organism’s ability to react to its environment. For example, networks are a natural framework for understanding protein-protein interactions [1], [2], [3], [4], [5].

Finding sets of highly interconnected nodes is an important task in network analysis. For example, clusters of interconnected nodes (modules) may correspond to important biological pathways [6], [7], [8], [9]. But we will not discuss different module detection methods in this paper. Instead, the main focus of this paper is a more fundamental screening task: how to define the neighborhood of an initial set of nodes (genes) in a network. Intuitively speaking, a neighborhood is comprised of nodes that are highly connected to a given set of genes. Thus neighborhood analysis facilitates a guilt-by-association screening strategy for finding genes that interact with a given set of biologically interesting genes.

To define a neighborhood of a initial gene set, one can make use of a similarity measure. For example, when dealing with gene expression microarray data, it is natural to

use the correlation coefficient to measure pairwise gene co-expression similarity [10], [11]. But here we consider the setting of an undirected networks that can be represented by a symmetric adjacency matrix A that encodes pairwise connection strengths. Specifically, the adjacency (connection strength) between nodes i and j is given by $a_{ij} \in [0, 1]$.

In our applications, we will make use of weighted gene co-expression network [12] where the adjacency is defined as a power of the absolute value of the Pearson correlation coefficient: $a_{ij} = |\text{cor}(x_i, x_j)|^\beta$. A simple (naive) approach for defining a neighborhood of node i is to choose the nodes with highest adjacencies a_{ij} with node i . In an unweighted network, this amounts to choosing the directly connected neighbors of node i . In weighted gene co-expression networks, this naive approach amounts to choosing genes based on the absolute values of their correlations with a given gene expression profile x_i .

Since spurious or weak connections in the adjacency matrix may lead to ‘noisy’ neighborhoods, several groups have made use of a pairwise topological overlap measure to measure node similarity [9], [12]. The topological overlap of two nodes reflects their similarity in terms of the commonality of the nodes they connect to. Ravasz *et al.*(2002) used a pairwise topological overlap measure as input of a clustering method to define modules in metabolic networks. In their study of a *E. coli* metabolic network, they show that two substrates having a higher overlap are more likely to belong to the same functional class than substrates having a lower topological overlap. Roughly speaking, the topological overlap measure serves as a filter that decreases the effect of spurious or weak connections. The topological overlap matrix and our proposed generalization to multiple nodes will be described in the following.

II. METHODS

A. Multi-node Topological Overlap Measure

1) *Multi-node Topological Overlap Measure for Unweighted Networks*: We briefly review the pairwise topological overlap measure for an undirected and unweighted network. In this case, the network (adjacency matrix) $A = [a_{ij}]$ contains binary entries. By convention, the diagonal elements of A are assumed to be zero. In the method supplement of [9],

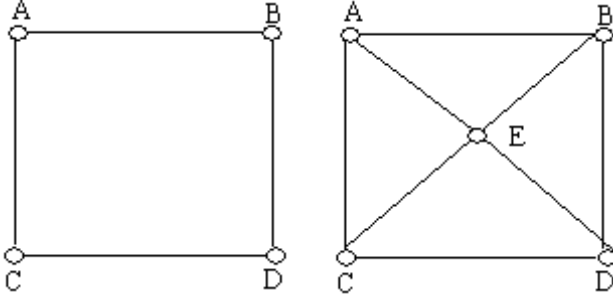


Fig. 1. Computing the MTOM measure for simple examples. 1) The MTOM value for nodes A,B,C and D in the picture on the left side is $\frac{0+4}{0+6} = 0.667$; 2) The MTOM for nodes A,B,C and D in the picture on the right side is $\frac{1+4}{1+6} = 0.714$.

the topological overlap matrix $T = [t_{ij}]$ is defined as follows

$$t_{ij} = \begin{cases} \frac{l_{ij}+a_{ij}}{\min\{k_i, k_j\}+1-a_{ij}} = \frac{l_{ij}+a_{ij}}{\min\{k_i-a_{ij}, k_j-a_{ij}\}+1} & \text{if } i \neq j \\ 1 & \text{if } i = j. \end{cases} \quad (1)$$

where, $l_{ij} = \sum_{u \neq i, j} a_{iu}a_{uj}$, $k_i = \sum_{u \neq i} a_{iu}$.¹ A straightforward proof that $0 \leq t_{ij} \leq 1$ can be found in [12].

Our multi-node TOM is motivated by the observation that formula (1) can be expressed as

$$t_{ij} = \begin{cases} \frac{|N(i,j)|+a_{ij}}{\min\{|N^{(-j)}(i)|, |N^{(-i)}(j)|\}+1} & \text{if } i \neq j \\ 1 & \text{if } i = j. \end{cases} \quad (2)$$

where $N(i)$ denotes the set of neighbors of i excluding i itself, $N(i, j)$ denotes the set of common neighbors excluding i and j , $N^{(-j)}(i)$ denotes the set of neighbors of i excluding i and j and $|\cdot|$ denotes the number of elements (cardinality) in its argument. Thus $|N(i, j)|$ equals the number of common neighbors that nodes i and j share.

In light of formula (2), it is natural to define the MTOM of three different nodes i, j, k as follows

$$t_{ijk} = \frac{|N(i, j, k)|+a_{ij}+a_{ik}+a_{jk}}{\min\{|N^{(-k)}(i, j)|, |N^{(-j)}(i, k)|, |N^{(-i)}(j, k)|\} + \binom{3}{2}} \quad (3)$$

where $N(i, j, k)$ denotes the set of common neighbors shared by i, j, k excluding i, j, k themselves, $N^{(-k)}(i, j)$ denotes the set of common neighbors shared by i, j excluding k and $\binom{3}{2}$ is the combinatorial number indicating that there are maximally $\binom{3}{2}$ pair-wise connections

It is easy to define topological overlap measure for four or more nodes according to the same rule.

A simple example illustrating the MTOM computation is presented in Figure 1.

¹The definition given in the main article of [9] is slightly different $(l_{ij} + a_{ij}) / \min\{k_i, k_j\}$. But according to the methods supplement and a personal communication, the authors propose eq. (1).

2) *MTOM for a Weighted Networks*: To compute MTOM, one can make use of the following formulas that express the cardinalities of sets in terms of the adjacency matrix:

$$|N(i, j, k)| = \sum_{u \neq i, j, k} a_{iu}a_{ju}a_{ku} \quad (4)$$

$$|N^{(-k)}(i, j)| = \sum_{u \neq i, j} a_{iu}a_{ju} - a_{ik}a_{jk}. \quad (5)$$

Note that these formulas do not require that a_{ij} take on binary values. Since the formulas remain mathematically meaningful if $a_{ij} \in [0, 1]$, they allow us to generalize MTOM to weighted networks as well.

B. MTOM-based Neighborhoods

Given an initial seed neighborhood, we consider two basic approaches for defining a neighborhood based on the concept of multi-node topological overlap. The first and recommended approach is to build the neighborhood recursively. The second is a computationally faster, non-recursive approach.

The MTOM based neighborhood analysis requires as input an initial seed neighborhood comprised of at least one node and the requested size of the neighborhood $S \geq 1$.

1) Recursive approach

- a) For each node outside of the neighborhood set, compute its MTOM value with the nodes of the current version of the neighborhood set.
- b) Add the node with the highest MTOM value to the neighborhood set.
- c) Repeat steps a) and b) until the neighborhood size S is reached.

2) Non-recursive approach

- a) For each node outside of the neighborhood set, compute its MTOM value with the nodes the current version of the neighborhood set.
- b) Choose the S nodes with the highest MTOM values as neighborhood.

Since the recursive approach leads to neighborhoods with high MTOM values, we focus on it below.

C. Local Permutations to Choose a Neighborhood Size S

An important issue is how to choose the size S of a neighborhood. In the applications below, we show that neighborhood sizes of 20 lead to biologically meaningful results. But more formally, we propose a permutation test for determining S . The permutation test is based on comparing the MTOM values of the observed network neighborhood to those of permuted versions of the network. We find that global (whole network) permutations can noise up the module structure of the network and thus lead to very large neighborhoods comprised of possibly thousands of nodes. This motivated us to define a *local* permutation scheme, see Fig. 2. For a given node in the initial neighborhood set, it permutes the adjacencies with all

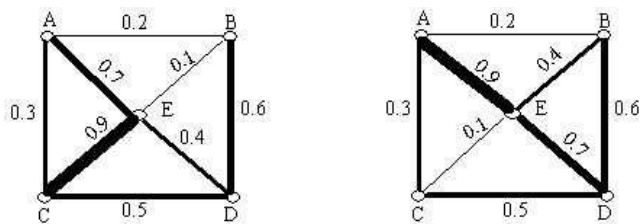


Fig. 2. Illustrating a local permutation centered around node E. The left panel shows a weighted network and the right panel shows a corresponding local permutation. Note that only the adjacencies (connection strengths) involving node E are permuted while the remaining adjacencies remain unpermuted.

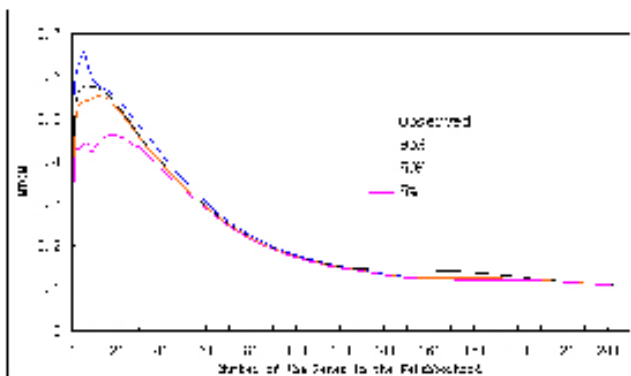


Fig. 3. Comparing the MTOM values of the observed network with locally permuted versions to determine the neighborhood size S . Local permutations were applied to each node in the initial set. Next, the recursive approach was used to define neighborhoods of different sizes (x-axis). This procedure was repeated 100 replicate times and the 5%, 50% and 95% quantiles of the MTOM values (y-axis) were determined as a function of the neighborhood sizes (x-axis). Note that for small values of $S \leq 10$ the MTOM values of the observed network are larger than those of the permuted networks.

other nodes while keeping the remaining adjacencies intact. Comparing the MTOM values of the observed network with locally permuted versions allows one to determine the size of the neighborhood as described in Figure 3.

III. APPLICATIONS

A. Neighborhood Analysis in a Yeast Protein-Protein Physical Interaction Network

The proposed MTOM based neighborhood analysis can be used for weighted and unweighted networks. Here we illustrate its use for unweighted networks by studying the protein-protein physical interaction network of yeast. The protein identifiers of the open reading frames (ORF) were obtained from the Saccharomyces Genome Database (SGD) and the yeast protein-protein interactions (PPI) were retrieved from the Munich Information Center for Protein Sequences (MIPS) [13]. We focused on the largest connected component comprised of 3858 proteins with 7196 pairwise interactions. To compare different neighborhood analysis approaches, we studied the neighborhoods of subsets of 101 cell cycle genes found in the Kyoto Encyclopedia of Genes and Genomes (KEGG). We considered each of the 101 genes as initial protein and

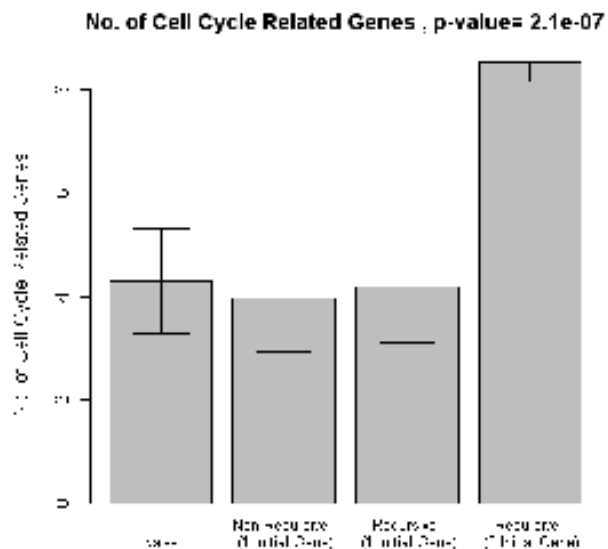


Fig. 4. Comparing the number of cell cycle proteins C (y-axis) in neighborhoods constructed in different ways. The recursive approach involving an initial neighborhood of 2 cell cycle related genes performs better than approaches based on an initial set comprised of a single gene. We also report the p-values of the Kruskal-Wallis rank sum test, which is a standard non-parametric multi-group comparison test.

identified its 10 closest neighbors using the MTOM based approaches in Section II-B. As metric for judging the biological meaning of size $S = 10$ neighborhoods, we determined the number C of cell cycle related genes. Overall, we find that neighborhood analysis leads to neighborhoods with high numbers of cell cycle genes C . However, C is significantly correlated with the network connectivity k of the initial protein (Spearman correlation $r = 0.36$, $p\text{-value} \leq 0.001$). This led us to focus the analysis on the 50 most highly connected ‘hub’ proteins as source of initial seed neighborhoods. A major advantage of the MTOM screening approach consists in being able to input initial neighborhoods comprised of more than 1 node. Figure 4 shows that an initial seed neighborhood composed of 2 cell-cycle proteins leads to far better results than using a single gene as input. In this application, the recursive and the non-recursive MTOM neighborhood analysis involving a *single* protein do not lead to better results than the naive approach of building a neighborhood on the basis of direct connections (adjacency=1) with the initial protein. The results that 2 initial proteins lead to informative neighborhoods is highly dependent on the topological overlap of the proteins in the initial seed neighborhood as can be seen from Figure 5. Note that pairs of proteins that result in neighborhoods with high C (x-axis) are likely to be comprised of proteins with high topological overlap (y-axis).

B. Neighborhood Analysis in a Brain Cancer Network

To illustrate the use of neighborhood analysis in a weighted network, we study here the weighted brain cancer gene co-expression network described in [12]. The network was built on the basis of 55 brain cancer samples whose expression

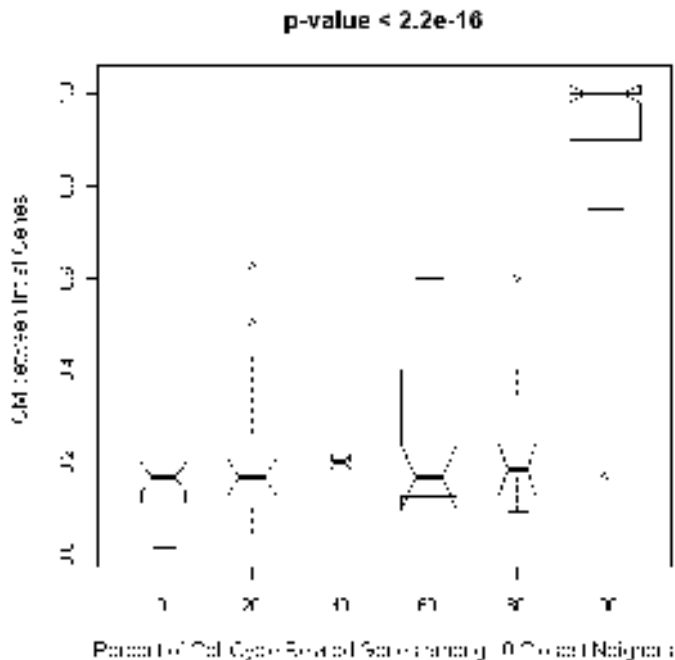


Fig. 5. Boxplots for visualizing the distribution of the topological overlap (y-axis) of initial protein pairs that lead to neighborhoods with a high percentage of cell cycle genes (x-axis). A boxplot consists of the most extreme values (the whiskers) in the data set (maximum and minimum values), the lower and upper quartiles (lower and upper boundary of the box), and the median value (horizontal line inside the notch). A notch is drawn in each side of the box. If the notches of two plots do not overlap the two medians differ significantly.

profiles were measured using Affymetrix HG-U133A microarrays. The details of the gene co-expression network construction are given in [12]. Briefly, the absolute value of the Pearson correlation coefficient is calculated for all pair-wise comparisons of gene-expression values across all microarray samples. The correlation matrix is then transformed into a matrix of connection strengths using the power function ($a_{ij} = |cor(x_i, x_j)|^6$) resulting in a weighted network. The use of weighted networks represents an improvement over unweighted networks based on dichotomizing the correlation matrix, because a) the continuous nature of the gene co-expression information is preserved and b) the results of weighted network analyses are highly robust with respect to the choice of the power $\beta = 6$. Since probesets (nodes) with low variance may represent noise, we focused the analysis on the 15000 most varying probes (highest variance).

Table I reports the recursive MTOM based neighborhood of size $S = 20$ of the patient survival time (TTS). We find that that it is highly enriched in cancer- and neuron related genes. Out of the 20 probe sets 11 probe sets are related to neuron cells and 10 probe sets are related to cancers. Note that several of the probe sets in Table I correspond to the same genes but the correlations between gene expression values and TTS for the different probe sets of a gene vary greatly. In contrast, a standard approach which simply selects a neighborhood on the

Probe Name	Gene Name	References	Correlation
208464_at	GRIA4*	[14]	0.62
221623_at	BCAN*	[15]	0.406
91920_at	BCAN*	[15]	0.226
219107_at	BCAN*	[15]	0.212
216476_at	LOC115131		0.142
222301_at	CROC4*	[16]	0.093
212655_at	BDG29		0.297
213768_s_at	ASCL1**	[17]	0.233
209988_s_at	ASCL1**	[17]	0.191
209987_s_at	ASCL1**	[17]	0.129
212265_at	QKI*	[18]	0.163
218902_at	NOTCH1**	[19]	-0.023
202981_x_at	SIAH1+	[20]	0.085
221776_s_at	BRD7+	[21]	0.068
212615_at	FLJ12178		0.123
212616_at	FLJ12178		0.126
213891_s_at	TCF4+	[22]	0.091
201310_s_at	C5orf13**	[23]	0.072
214239_x_at	RNF110+	[24]	0.076
213551_x_at	RNF110+	[24]	0.075

TABLE I

NEIGHBORHOOD OF SURVIVAL TIME BASED ON THE RECURSIVE MTOM APPROACH. CANCER RELATED GENES ARE MARKED BY "*" AND NEURON CELL RELATED ONES MARKED BY "**". THE LAST COLUMN LISTS THE CORRELATION BETWEEN THE GENES AND TTS.

basis of the absolute values of the correlations between gene expression profile and survival time, leads to a neighborhood with fewer cancer- and neuron related genes (Table II). Out of the 20 most highly correlated probe sets in Table II only 4 probe sets are related to neuron cells and 6 probe sets are related to cancer.

Comparing Tables I and II provides indirect empirical evidence that in this application the MTOM neighborhood analysis leads to biologically more meaningful results than the standard approach.

A major advantage of the MTOM approach is that it allows one to input more than 1 probeset as initial neighborhood. In this application, we were interested in finding the neighborhood of five highly correlated cancer genes: TOP2A, Rac1, TPX2, EZH2 and KIF14. These cell mitosis related genes were known from the literature. Table III reports the results of recursive MTOM analysis. Out of 20 probes 13 are cancer related, which provides further empirical evidence that the MTOM approach leads to biologically meaningful results.

IV. AVAILABILITY:

A software for neighborhood analysis based on MTOM can be downloaded from the following webpage: <http://genetics.ucla.edu/labs/horvath/MTOM/>

V. CONCLUSION

This paper proposes the use of a multi-node generalization of the topological overlap measure for defining the

Probe Name	Gene Name	References	Correlation
208464.at	GRIA4*	[14]	0.62
204529.s.at	TOX		0.601
206170.at	ADRB2		0.539
216247.at	RPS20		0.537
214028.x.at	TDRD3		0.526
213447.at	IPW		0.522
207113.s.at	TNF**	[25]	0.514
218036.x.at	CGI-07		0.504
209160.at	AKR1C3 ⁺	[26]	0.5
206107.at	RGS11		0.496
209782.s.at	DBP		0.495
211653.x.at	AKR1C1 ⁺	[27]	0.494
213778.x.at	ZFP276		0.49
215119.at	MYR8*	[28]	0.486
202753.at	p44S10 ⁺	[29]	-0.481
209292.at	ID4 ⁺	[30]	0.481
204530.s.at	TOX		0.481
221974.at	SNRPN		0.481
219188.s.at	LRP16 ⁺	[31]	0.48
205630.at	CRH*	[32]	0.478

TABLE II

CORRELATION BASED NEIGHBORHOOD OF THE SURVIVAL TIME TTS . CANCER RELATED GENES ARE MARKED BY "*" AND NEURON CELL RELATED ONES MARKED BY "**". THE LAST COLUMN LISTS THE CORRELATION BETWEEN THE GENES AND TTS .

Probe Name	Gene Name	Reference
209642.at	BUB1 ⁺	[33]
218355.at	KIF4A	
222077.s.at	RACGAP1 ⁺	[34]
219918.s.at	ASPM*	[35]
207828.s.at	CENPF ⁺	[36]
202580.x.at	FOXM1 ⁺	[37]
202870.s.at	CDC20 ⁺	[38]
202095.s.at	BIRC5 ⁺	[39]
221591.s.at	FLJ10156	
218009.s.at	PRC1	
204641.at	NEK2 ⁺	[40]
209172.s.at	CENPF ⁺	[36]
209464.at	AURKB ⁺	[41]
212020.s.at	MKI67 ⁺	[42]
204962.s.at	CENPA ⁺	[43]
212023.s.at	MKI67 ⁺	[42]
204444.at	KIF11	
212949.at	BRRN1	
204026.s.at	ZWINT	
203213.at	CDC2 ⁺	[44]

TABLE III

RECURSIVE MTOM NEIGHBORHOOD OF THE FIVE MITOSIS RELATED CANCER GENES: TOP2A, RAC1, TPX2, EZH2 AND KIF14. CANCER RELATED GENES ARE MARKED BY "*" AND NEURON CELL RELATED ONES MARKED BY "**". THE LAST COLUMN LISTS THE CORRELATION BETWEEN THE GENES AND TTS .

neighborhood of an initial seed neighborhood. To define the size of the neighborhood, we propose a local permutation test approach. MTOM is a natural extension of the standard pairwise topological overlap measure [9]. It allows one to define the neighborhood of three or more genes. Using two different real applications we provide empirical evidence that our method leads to biologically and clinically meaningful results that can be superior to standard approaches.

In [45], we describe another generalization (GTOM) of the topological overlap measure based on the proportion of m -step neighbors shared by a pair of nodes. But this generalization is only defined for unweighted networks and it measures *pairwise* overlap. Thus, GTOM is very different from the multi-node topological overlap measure (MTOM) proposed in this paper.

It is worth emphasizing that when dealing with microarray data, one can determine the gene neighborhood of an 'idealized' gene expression profile that represents a clinical trait. For example, we study the gene neighborhood of brain cancer survival time to identify brain cancer related genes.

Future research could explore how to use the MTOM measure to define network modules.

ACKNOWLEDGMENT

The authors would like to thank our UCLA collaborators Ed McCabe, Katrina Dipple, Jun Dong, Dan Geschwind, Chi-ying Lee, Jake Lusic, Paul Mischel, Stan Nelson, Mike Oldham, Lin Wang and Wei Zhao.

REFERENCES

- [1] H. Jeong, S. P. Mason, A. L. Barabasi, and Z. N. Oltvai, "Lethality and centrality in protein networks," *Nature*, vol. 411, pp. 41, May 2001.
- [2] A. Rzhetsky and S. M. Gomez, "Birth of scale-free molecular networks and the number of distinct dna and protein domains per genome," *Bioinformatics*, vol. 17, pp. 988–996, 2001.
- [3] J. D. Han, N. Bertin, T. Hao, D.S. Goldberg, G.F. Berriz, L.V. Zhang, D. Dupuy, A.J. Walhout, M.E. Cusick, F.P. Roth, and M. Vidal, "Evidence for dynamically organized modularity in the yeast protein-protein interaction network," *Nature*, vol. 430, no. 6995, pp. 88–93, 2004.
- [4] S Y Yook, Z N Oltvai, and A L Barabasi, "Functional and topological characterization of protein interaction networks," *Proteomics*, vol. 4, pp. 928–942, 2004.
- [5] J. Berg, M. Lassig, and A. Wagner, "Structure and evolution of protein interaction networks: a statistical model for link dynamics and gene duplications.," *BMC Evol. Biol.*, vol. 5, no. 51, 2004.
- [6] N. S. Holter, M. Mitra, A. Maritan, M. Cieplak, J. R. Banavar, and N. V. Fedoroff, "Fundamental patterns underlying gene expression profiles: simplicity from complexity," *Proc Natl Acad Sci U S A*, vol. 97, no. 15, pp. 8409–14, 2000.
- [7] J. Hasty, D. McMillen, F. Isaacs, and J. J. Collins, "Computational studies of gene regulatory networks: in numero molecular biology," *Nat Rev Genet.*, vol. 2, no. 4, pp. 268–79, 2001.
- [8] S. S. Shen-Orr, R. Milo, S. Mangan, and U. Alon, "Network motifs in the transcriptional regulation network of escherichia coli," *Nat Genet.*, vol. 31, no. 1, pp. 64–8, 2002.
- [9] E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, and A. L. Barabasi, "Hierarchical organization of modularity in metabolic networks," *Science*, vol. 297, no. 5586, pp. 1551–5, 2002.
- [10] M.B. Eisen, P.T. Spellman, P.O. Brown, and D. Botstein, "Cluster analysis and display of genome-wide expression patterns," *PNAS*, vol. 95, no. 25, pp. 14863–14868, 1998.

- [11] T. R. Golub, D. K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J. P. Mesirov, I. M. L. Loh H. Coller, J. R. Downing, M. A. Caligiuri, C. D. Bloomfield, and E. S. Lander, "Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring," *SCIENCE*, vol. 286, pp. 531–537, OCTOBER 1999.
- [12] B. Zhang and S. Horvath, "general framework for weighted gene co-expression analysis," *Statistical Applications in Genetics and Molecular Biology. Free Online Journal.*, vol. 4, no. 17, 2005.
- [13] U. Guldener, M. Munsterkotter, M. Oesterheld, P. Pagel, A. Ruepp, H.W. Mewes, and V. Stumpflen, "Mpac: the mips protein interaction resource on yeast," *Nucleic Acids Res.*, vol. 34, pp. 436–441, Jan 2006.
- [14] M. Nuriya, S. Oh, and R.L. Huganir, "Phosphorylation-dependent interactions of alpha-actinin-1/igqap1 with the ampa receptor subunit glur4," *J Neurochem.*, vol. 95, no. 2, pp. 544–52, Oct 2005.
- [15] C. Brakebusch, C.I. Seidenbecher, F. Asztely, U. Rauch, H. Matthies, H. Meyer, M. Krug, T.M Bockers, X. Zhou, M.R. Kreutz, D. Montag, E.D Gundelfinger, and Fassler R., "Brevican-deficient mice display impaired hippocampal ca1 long-term potentiation but show no obvious deficits in learning and memory," *Mol Cell Biol.*, vol. 22, no. 21, pp. 7417–27, Nov 2002.
- [16] P.L. Jeffrey, A. Capes-Davis, J.M. Dunn, O. Tolhurst, G. Seeto, A.J. Hannan, and S.L. Lin, "Croc-4: a novel brain specific transcriptional activator of c-fos expressed from proliferation through to maturation of multiple neuronal cell types," *Mol Cell Neurosci.*, vol. 16, no. 3, pp. 185–96, Sep 2000.
- [17] T. Shida, M. Furuya, T. Nikaido, T. Kishimoto, K. Koda, K. Oda, Y. Nakatani, M. Miyazaki, and H. Ishikura, "Aberrant expression of human achaete-scute homologue gene 1 in the gastrointestinal neuroendocrine carcinomas," *Clin Cancer Res.*, vol. 11, no. 2 Pt 1, pp. 450–8, Jan 2005.
- [18] H.Y. Wu, M.R. Dawson, R. Reynolds, and R.J. Hardy, "Expression of qki proteins and map1b identifies actively myelinating oligodendrocytes in adult rat brain," *Mol Cell Neurosci.*, vol. 17, no. 2, pp. 292–302, Feb 2001.
- [19] A. Joutel and E. Tournier-Lasserre, "Notch signalling pathway and human diseases," *Semin Cell Dev Biol.*, vol. 9, no. 6, pp. 619–25, Dec 1998.
- [20] K. Matsuo, S. Satoh, H. Okabe, A. Nomura, T. Maeda, Y. Yamaoka, and I. Ikai, "Siah1 inactivation correlates with tumor progression in hepatocellular carcinomas," *Genes Chromosomes Cancer.*, vol. 36, no. 3, pp. 283–91, Mar 2003.
- [21] M. Zhou, C. Peng, X.M. Nie, B.C. Zhang, S.G. Zhu, Y. Yu, X.L. Li, and G.Y. Li, "[expression of brd7-interacting proteins,brd2 and brd3, in nasopharyngeal carcinoma tissues]," *Ai Zheng.*, vol. 22, no. 2, pp. 123–7, Feb 2003.
- [22] A.S. Nateri, B. Spencer-Dene, and A. Behrens, "Interaction of phosphorylated c-jun with tcf4 regulates intestinal cancer development," *Nature.*, vol. 437, no. 7056, pp. 281–5., Sep 2005.
- [23] W.S. McDonough, N.L. Tran, and Berens M.E., "Regulation of glioma cell migration by serine-phosphorylated p311," *Neoplasia.*, vol. 7, no. 9, pp. 862–72, Sep 2005.
- [24] M. Kanno, M. Hasegawa, A. Ishida, K. Isono, and M. Taniguchi, "mel-18, a polycomb group-related mammalian gene, encodes a transcriptional negative regulator with tumor suppressive activity," *EMBO J.*, vol. 14, no. 22, pp. 5672–8, Nov 1995.
- [25] A.S. Achrol, L. Pawlikowska, C.E. McCulloch, K.Y. Poon, C. Ha, J.G. Zaroff, S.C. Johnston, C. Lee, M.T. Lawton, S. Sidney, D.A. Marchuk, P.Y. Kwok, and W.L. Young, "Tumor necrosis factor-alpha-238G > A promoter polymorphism is associated with increased risk of new hemorrhage in the natural course of patients with brain arteriovenous malformations." *Stroke.*, vol. 37, no. 1, pp. 231–4, Jan 2006.
- [26] M.J. Lewis, J.P. Wiebe, and J.G. Heathcote, "Expression of progesterone metabolizing enzyme genes (akr1c1, akr1c2, akr1c3, srd5a1, srd5a2) is altered in human breast carcinoma," *BMC Cancer.*, pp. 4–27, Jun 2004.
- [27] Q. Ji, C. Aoyama, Y.D. Nien, P.I. Liu, P.K. Chen, L. Chang, F.Z. Stanczyk, and A. Stolz, "Selective loss of akr1c1 and akr1c2 in breast cancer and their potential effect on progesterone signaling," *Cancer Res.*, vol. 64, no. 20, pp. 7610–7, Oct 2004.
- [28] K.G. Patel, C. Liu, P.L. Cameron, and R.S. Cameron, "Myr 8, a novel unconventional myosin expressed during brain development associates with the protein phosphatase catalytic subunits 1alpha and 1gamma1," *J Neurosci.*, vol. 21, no. 20, pp. 7954–68, Oct 2001.
- [29] S. Ren, M.J. Smith, I.D. Louro, P. McKie-Bell, M.R. Bani, M. Wagner, B. Zochodne, D.T. Redden, W.E. Grizzle, N. Wang, D.I. Smith, R.A. Herbst, W. Bardenheuer, B. Opalka, J. Schutte, J.M. Trent, Y. Ben-David, and J.M. Ruppert, "The p44s10 locus, encoding a subunit of the proteasome regulatory particle, is amplified during progression of cutaneous malignant melanoma," *Oncogene.*, vol. 19, no. 11, pp. 1419–27, Mar 2000.
- [30] L. Yu, C. Liu, J. Vandeusen, B. Becknell, Z. Dai, YZ. Wu, A. Raval, T.H. Liu, W. Ding, C. Mao, S. Liu, L.T. Smith, S. Lee, L. Rassenti, G. Marcucci, J. Byrd, M.A. Caligiuri, and C. Plass, "Global assessment of promoter methylation in a mouse model of cancer identifies id4 as a putative tumor-suppressor gene in human leukemia," *Nat Genet.*, vol. 37, no. 3, pp. 265–74, Mar. 2005.
- [31] W.D. Han, Y.M. Mu, X.C. Lu, Z.M. Xu, X.J. Li, L. Yu, H.J. Song, M. Li, J.M. Lu, Y.L. Zhao, and C.Y. Pan, "Up-regulation of lrp16 mrna by 17beta-estradiol through activation of estrogen receptor alpha (eralpha), but not erbeta, and promotion of human breast cancer mcf-7 cell proliferation: a preliminary report," *Endocr Relat Cancer.*, vol. 10, no. 2, pp. 217–24, Jun 2003.
- [32] H.U. Rehman, "Role of crh in the pathogenesis of dementia of alzheimer's type and other dementias," *Curr Opin Investig Drugs.*, vol. 3, no. 11, pp. 1637–42, Nov 2002.
- [33] P.M. Hempen, H. Kurpad, E.S. Calhoun, S. Abraham, and S.E. Kern, "A double missense variation of the bub1 gene and a defective mitotic spindle checkpoint in the pancreatic cancer cell line hs766t," *Hum Mutat.*, vol. 21, no. 4, pp. 445, Apr 2003.
- [34] C. Yang, Y. Liu, F.C. Leskow, V.M. Weaver, and M.G. Kazanietz, "Rac-gang-dependent inhibition of breast cancer cell proliferation by beta2-chimerin," *J Biol Chem.*, vol. 280, no. 26, pp. 24363–70, Jul 2005.
- [35] A. Kumar, M. Markandaya, and S.C. Girimaji, "Primary microcephaly: microcephalin and aspm determine the size of the human brain," *J Biosci.*, vol. 27, no. 7, pp. 629–32., Dec 2002.
- [36] A. Pimkhaokham, Y. Shimada, Y. Fukuda, N. Kurihara, I. Imoto, Z.Q. Yang, M. Imamura, Y. Nakamura, T. Amagasa, and J. Inazawa, "Nonrandom chromosomal imbalances in esophageal squamous cell carcinoma cell lines: possible involvement of the atf3 and cenpf genes in the 1q32 amplicon," *Jpn J Cancer Res.*, vol. 91, no. 11, pp. 1126–33, Nov 2000.
- [37] D.R. Wonsey and M.T. Follettie, "Loss of the forkhead transcription factor foxm1 causes centrosome amplification and mitotic catastrophe," *Cancer Res.*, vol. 65, no. 12, pp. 5181–9, Jun 2005.
- [38] B. Liu, S. Hong, Z. Tang, H. Yu, and C.Z. Giam, "Htlv-i tax directly binds the cdc20-associated anaphase-promoting complex and activates it ahead of schedule," *Proc Natl Acad Sci U S A.*, vol. 102, no. 1, pp. 63–8, Jan 2005.
- [39] D. Xie, Y.X. Zeng, H.J. Wang, J.M. Wen, Y. Tao, J.S. Sham, and X.Y. Guan, "Expression of cytoplasmic and nuclear survivin in primary and secondary human glioblastoma," *Br J Cancer.*, vol. 94, no. 1, pp. 108–14, Jan 2006.
- [40] D.G. Hayward, R.B. Clarke, A.J. Faragher, M.R. Pillai, I.M. Hagan, and A.M. Fry, "The centrosomal kinase nek2 displays elevated levels of protein expression in human breast cancer," *Cancer Res.*, vol. 64, no. 20, pp. 7370–6, Oct 2004.
- [41] J.E. Jung, T.K. Kim, J.S. Lee, S.Y. Oh, S. Kwak, X. Jin, J.Y. Sohn, M.K. Song, Y.W. Sohn, S.Y. Lee, X. Pian, J.B. Lee, Y.G. Chung, Y.K. Choi, S. You, and H. Kim, "Survivin inhibits anti-growth effect of p53 activated by aurora b," *Biochem Biophys Res Commun.*, vol. 336, no. 4, pp. 1164–71, Nov 2005.
- [42] K. Revelos, C. Petraki, A. Gregorakis, A. Scorilas, P. Papanastasiou, R. Tenta, and M. Koutsilieris, "p27(kip1) and ki-67 (mib1) immunohistochemical expression in radical prostatectomy specimens of patients with clinically localized prostate cancer," *In Vivo.*, vol. 19, no. 5, pp. 911–20, Sep-Oct 2005.
- [43] T. Tomonaga, K. Matsushita, S. Yamaguchi, T. Oohashi, H. Shimada, T. Ochiai, K. Yoda, and F. Nomura, "Overexpression and mistargeting of centromere protein-a in human primary colorectal cancer," *Cancer Res.*, vol. 63, no. 13, pp. 3511–6., Jul 2003.
- [44] T. Nozoe, I. Takahashi, H. Baba, and Y. Maehara, "Relationship between intracellular localization of p34cdc2 protein and differentiation of esophageal squamous cell carcinoma," *J Cancer Res Clin Oncol.*, vol. 131, no. 3, pp. 179–83, Mar 2005.
- [45] A. Yip and S. Horvath, "Generalized topological overlap matrix and its applications in gene co-expression networks," *Biocomp 2006 Proceedings Volume.*, 2006.