International Journal of Biological, Life and Agricultural Sciences
ISSN: 2415-6612
Vol:6, No:12, 2012

Categorization and Estimation of Relative Connectivity of Genes from Meta-OFTEN Network

U. Kairov, T. Karpenyuk, E. Ramanculov, and A. Zinovyev

***Abstract*—**The most common result of analysis of high-throughput data in molecular biology represents a global list of genes, ranked accordingly to a certain score. The score can be a measure of differential expression. Recent work proposed a new method for selecting a number of genes in a ranked gene list from microarray gene expression data such that this set forms the Optimally Functionally Enriched Network (OFTEN), formed by known physical interactions between genes or their products. Here we present calculation results of relative connectivity of genes from META-OFTEN network and tentative biological interpretation of the most reproducible signal. The relative connectivity and inbetweenness values of genes from META-OFTEN network were estimated.

***Keywords*—**Microarray, META-OFTEN, gene network.

I. Introduction

MANY genome-scale studies in molecular biology deliver results in the form of a ranked list of gene names, accordingly to some scoring method. Many methods were developed for estimating a statistically justified threshold for the score value used to select a number of top-scored genes. The derived in this way gene signature are used, for example, for predicting outcome of treatment in cancer therapies [1-4]. The question is how many top-ranked genes to consider for further analysis. This question is usually approached from a statistical point of view, without considering any biological properties of top-ranked genes or how they are related to each other functionally.

Several efforts have been made in attempt to take into account physical interactions between gene products at the top of the ranked list of genes. For example, in [5] network signatures of breast cancer metastases were derived using protein-protein interaction (PPI) database in combination with differential gene expression values. Various machine learning frameworks were developed in order to include network information into the analysis of gene expression data [6,7]. Several attempts of meta-analysis of gene signatures were made for multiple cancer studies and for breast cancer in particular [8,9], finding recurrent patterns appearing in them (for example, the role of proliferation, RNA splicing, immune response genes). The method proposed in [10] is inspired by the idea of percolation in graph theory. Given a connected graph and k randomly selected nodes, one may estimate the expected size of the largest connected component formed by these genes. For many type of graphs, the typical behavior is the following: at some critical k_{crit} number of nodes, most of them start to be connected in a large connected component. This means that if the first $k << k_{crit}$ top-ranked genes form a relatively large connected component (compared to the randomly expected), their distribution on the graph of protein-protein interactions is highly non-random and they form a tightly connected functional group. In this paper we estimate the values of in betweenness and relative connectivity of genes forming META-OFTEN network.

U.Kairov is with the Kazakh National University after Al-Farabi, Almaty, Kazakhstan and with the National Center for Biotechnology, Astana, Kazakhstan (e-mail: ukairov@gmail.com).

T.Karpenyuk is with the Kazakh National University after Al-Farabi, Almaty, Kazakhstan (e-mail: Tatyana.Karpenyuk@kaznu.kz).

E.Ramanculov is with the National Center for Biotechnology, Astana, Kazakhstan (e-mail: ramanculov@biocenter.kz).

A.Zinovyev is with the Institute Curie, Paris, France; with the INSERM U900, Paris, France; with the Mines ParisTech, Fontainebleau, France (e-mail: andrei.zinovyev@curie.fr).

II. Materials and Methods

Gene signature consisting of 74 genes forming META-OFTEN network, OFTEN networks for each dataset and ranked lists of differentially expressed genes obtained from [10]. We used HPRD version 9 database without complex with id=COM_2971 as a source of protein-protein interactions in human cells. For constructing the interaction graph, we used all binary protein interaction part of the database. In addition, the protein relations inside protein complexes were used. Cytoscape software [11] and BiNoM plugin [12] were used for constructing META-OFTEN network. Enrichment analysis of META-OFTEN genes performed by using Database for Annotation, Visualization and Integrated Discovery (DAVID) [13].

III. Results and Discussion

Gene signatures extracted from the lists of differentially expressed genes using the standard statistical approach usually show much more modest overlap (not more than few percents, see [14]). For example, in our analysis of the lists of differentially expressed genes from four breast cancer datasets [1] there are six genes CCNB1, CCNB2, DTL, NEK2, UBE2S and ZWINT genes are found in at least three datasets and only two RACGAP1 and RRM2 genes found in common between the top 100 differentially expressed genes (Fig. 1).

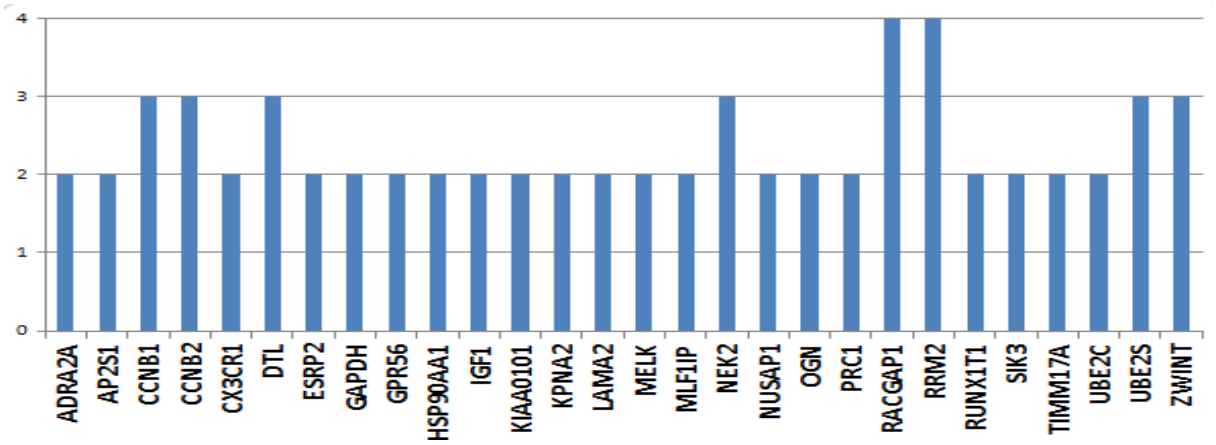

Fig. 1 Most variable genes from the list of top 100 differentially expressed genes between four breast cancer datasets

TABLE I
INBETWEENNESS VALUES OF META-OFTEN GENES

gene	inbetweenness	subnetwork	global	ratio
CDK1	0,336	18	133	0,135
HSP90AA1	0,211	7	159	0,044
BIRC5	0,189	2	15	0,133
NR3C1	0,069	6	99	0,06
KIF11	0,058	2	9	0,222
AURKA	0,052	4	27	0,148
CDC20	0,037	5	25	0,2
CCNB2	0,034	2	5	0,4
RUNX1T1	0,034	3	26	0,115
PTTG1	0,021	2	7	0,288
CCNA2	0,021	2	30	0,067
PCNA	0,018	4	85	0,047
TPX2	0,017	2	5	0,4
MARS	0,017	2	14	0,143
LMNB1	0,017	2	19	0,105
ITPR1	0,017	3	31	0,097
MCM2	0,017	2	36	0,055
TOP2A	0,017	2	52	0,038
SPTAN1	0,017	2	65	0,03
HSPA4	0,013	2	28	0,071
KPNA2	0,012	3	38	0,079
TXN	0,006	2	10	0,2
PRPF40A	0,003	4	87	0,045
CD74	0,003	5	13	0,385
CCNB1	0,001	4	37	0,108
PECAM1	0,001	2	24	0,0833
GAPDH	3,81E-04	2	43	0,046
TXNIP	3,17E-04	2	8	0,25

Nodes of the META-OFTEN network are organized into several functional subgroups, represented by network modules and containing many genes known to be implicated in tumorigenesis. The most evident components of META-OFTEN network described in [1]. Nodes of the META-OFTEN network can be ranked with respect to their role in forming the structure of the graph (Table I). We calculated values of inbetweenness and ratio of the connectivity for the genes from META-OFTEN network. For example, genes CDC20, CCNB2, TPX2, CD74, PTTG1 have the highest relative connectivity. Realtive connectivity is the ratio of the connectivity in the network and the global connectivity in the global PPI network, as in [15]. Subnetwork values shows how many edges has every node in the META-OFTEN network. CDK1, HSP90AA1 and NR3C1 have highest number of edges. CDK1, HSP90AA1, BIRC5 have the highest inbetweenness values. Genes with highest inbetweenness values can be classified as “routers” or “bottlenecks” as in [16].

We used the Database for Annotation, Visualization and Integrated Discovery (DAVID) [13] to reveal biological meaning of genes forming META-OFTEN network from four independent breast cancer datasets. Results for selected the most significant top 10 categories are illustrated in Table II.

Results in Table II provides information about biological themes in Gene Ontology terms enriched in a gene list from our META-OFTEN network relative to all annotated genes in human genome. The most evident genes from META-OFTEN network related to cell cycle processes (PRC1, NEK2, C13ORF15, AURKA, CEP55, PTTG1, CCNA2, TXNIP, CDK1, KIF11, TPX2, CENPF, BIRC5, CDC20, MCM2, TACC3, CDKN3, RACGAP1, GAS7, MCM6, CCNB1, CCNB2, MAD2L1, PSMC2, CKS2, BUB1B, KPNA2). These results highlighted that the most reproducible biological signal in four breast cancer is the cell cycle process.

In this paper we presented results of relative connectivity of genes from META-OFTEN network and approach for tentative biological interpretation of the most reproducible signal.

TABLE II
CATEGORIZATION OF GENES FROM META-OFTEN NETWORK

Category	Term	Count	%	PValue	Genes
GOTERM_BP_ALL	GO:0007049~cell cycle	27	36	1,98E-15	PRC1, NEK2, C13ORF15, AURKA, CEP55, PTTG1, CCNA2, TXNIP, CDK1, KIF11, TPX2, CENPF, BIRC5, CDC20, MCM2, TACC3, CDKN3, RACGAP1, GAS7, MCM6, CCNB1, CCNB2, MAD2L1, PSMC2, CKS2, BUB1B, KPNA2
GOTERM_BP_ALL	GO:0022402~cell cycle process	23	31	2,37E-14	CDK1, KIF11, PRC1, NEK2, TPX2, CENPF, AURKA, CDC20, BIRC5, PTTG1, CEP55, RACGAP1, TACC3, CDKN3, GAS7, CCNB1, CCNB2, MAD2L1, PSMC2, CKS2, BUB1B, KPNA2, CCNA2
GOTERM_BP_ALL	GO:0000279~M phase	19	25,6	2,63E-14	CDK1, KIF11, PRC1, NEK2, TPX2, CENPF, AURKA, CDC20, BIRC5, PTTG1, CEP55, TACC3, CCNB1, CCNB2, MAD2L1, CKS2, BUB1B, KPNA2, CCNA2
GOTERM_BP_ALL	GO:0022403~cell cycle phase	20	27	1,07E-13	CDK1, KIF11, PRC1, NEK2, TPX2, CENPF, AURKA, CDC20, BIRC5, PTTG1, CEP55, TACC3, CDKN3, CCNB1, CCNB2, MAD2L1, CKS2, BUB1B, KPNA2, CCNA2
GOTERM_BP_ALL	GO:0000278~mitotic cell cycle	19	25,6	1,98E-13	CDK1, KIF11, PRC1, NEK2, TPX2, CENPF, AURKA, CDC20, BIRC5, PTTG1, CEP55, CDKN3, CCNB1, CCNB2, MAD2L1, PSMC2, BUB1B, KPNA2, CCNA2
GOTERM_BP_ALL	GO:0006996~organelle organization	30	40,5	1,89E-12	CAV1, PRC1, NEK2, AURKA, NR3C1, CEP55, PTTG1, PTK2B, HSPA4, CCNA2, TOP2A, CDK1, UBE2A, KIF11, HSP90AA1, CYCS, TPX2, CENPF, BIRC5, CDC20, MCM2, TACC3, RACGAP1, GAS7, CCNB1, CCNB2, MAD2L1, CKS2, BUB1B, SMARCA4
GOTERM_BP_ALL	GO:0000087~M phase of mitotic cell cycle	15	20,2	4,63E-12	CDK1, KIF11, NEK2, TPX2, CENPF, AURKA, CDC20, BIRC5, PTTG1, CEP55, CCNB1, CCNB2, MAD2L1, BUB1B, CCNA2
GOTERM_BP_ALL	GO:0048285~organelle fission	15	20,2	6,24E-12	CDK1, KIF11, NEK2, TPX2, CENPF, AURKA, CDC20, BIRC5, PTTG1, CEP55, CCNB1, CCNB2, MAD2L1, BUB1B, CCNA2
GOTERM_BP_ALL	GO:0051301~cell division	16	21,6	1,42E-11	CDK1, KIF11, PRC1, NEK2, CENPF, CDC20, BIRC5, PTTG1, CEP55, RACGAP1, CCNB1, CCNB2, MAD2L1, CKS2, BUB1B, CCNA2
GOTERM_BP_ALL	GO:0016043~cellular component organization	33	44,5	8,30E-08	CAV1, PRC1, NEK2, AURKA, NR3C1, CEP55, PTTG1, HLA-DMA, CD74, PTK2B, HSPA4, CCNA2, TOP2A, CDK1, UBE2A, HSP90AA1, KIF11, CYCS, TPX2, CENPF, BIRC5, CDC20, MCM2, TACC3, RACGAP1, GAS7, CCNB1, CCNB2, MAD2L1, RRM2, CKS2, BUB1B, SMARCA4
KEGG_PATHWAY	hsa04110:Cell cycle	11	14,8	1,35E-07	CCNB1, CDK1, MAD2L1, CCNB2, PCNA, BUB1B, CDC20, MCM2, PTTG1, CCNA2, MCM6

REFERENCES

[1] van't Veer L.J., Dai H., van de Vijver M.J. et al. "Gene expression profiling predicts clinical outcome of breast cancer". Nature, 415:530-6, 2002.
[2] van de Vijver M.J., van't Veer L.J. et al. "A gene-expression signature as a predictor of survival in breast cancer". N. Engl. J. Med., 347:1999-2009, 2002.
[3] Wang Y., Klijn J.G., Zhang Y., Sieuwerts A.M. et al. "Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer" Lancet, 365(9460):671-9, 2005.
[4] Cobleigh M.A., Tabesh B., Bitterman P., Baker J., Cronin M., Liu M.L., Borchik R., Mosquera J.M., Walker M.G., Shak S. "Tumor gene expression and prognosis in breast cancer patients with 10 or more positive lymph nodes" Clin. Cancer Res., 11(24 Pt 1):8623-31, 2005.
[5] Chuang H.-Y. et al. "Network-based classification of breast cancer metastasis" Mol. Syst. Biol., 3:140, 2007.
[6] Rapaport F., Zinovyev A., Dutreix M., Barillot E., Vert J.-P. "Classification of microarray data using gene networks" BMC Bioinformatics, 8:35, 2007.
[7] Foekens J. A. et al. "Multicenter validation of a gene expression-based prognostic signature in lymph node-negative primary breast cancer" J. Clin. Oncol., 24:1665–1671, 2006.
[8] Finocchiaro G. et al. "Graph-based identification of cancer signaling pathways from published gene expression signatures using PubLiME" Nucleic Acids Res., 35(7): 2343, 2007.
[9] Reyal F., van Vliet M.H., Armstrong N.J., Horlings H.M., de Visser K.E., Kok M., Teschendorff A.E., Mook S., van 't Veer L., Caldas C., Salmon R.J., van de Vijver M.J., Wessels L.F. "A comprehensive analysis of prognostic signatures reveals the high predictive capacity of the proliferation, immune response and RNA splicing modules in breast cancer" Breast Cancer Res., 10(6):R93, 2008.

[10] Kairov U., Karpenyuk T., Ramanculov E., Zinovyev A. "Network analysis of gene lists for finding reproducible prognostic breast cancer gene signatures" Bioinformation, 8(16):773-6, 2012.
[11] Cline M., Smoot M., Cerami E. et al. "Integration of biological networks and gene expression data using Cytoscape" Nature Protocols, 2:2366 – 2382, 2007.
[12] Zinovyev A. et al. "BiNoM: a Cytoscape plugin for manipulating and analyzing biological networks" Bioinformatics, 24(6):876, 2008.
[13] Huang D.W., Sherman B.T., Lempicki R.A. "Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists" Nucleic Acids Res., 37(1):1-13, 2009.
[14] Barillot E., Calzone L., Hupe P., Vert J.-P., Zinovyev A. "Computational Systems Biology of Cancer" CRC Press Inc, Chapman & Hall/CRC Mathematical & Computational Biology, 452p., 2012.
[15] Pinna G., Zinovyev A., Araujo N., Morozova N., Harel-Bellan A. "Analysis of the growth control network specific for human lung adenocarcinoma cells" Math. Model. Nat. Phenom., 7(01):337-368, 2012.
[16] Chen J., Sam L., Huang Y., Lee Y., Li J., Liu Y., Xing H.R., Lussier Y.A. "Protein interaction network underpins concordant prognosis among heterogeneous breast cancer signatures" J Biomed. Inform., 43(3): 385–396, 2010.