

# Identification of genes with common differential expression pattern across different metastatic microarray data

Lam C. Tsoi\*, and W. Jim Zheng

Department of Biostatistics, Bioinformatics, and Epidemiology,  
Medical University of South Carolina, Charleston, USA

\*E-mail: tsoi@musc.edu

## Abstract

Metastasis involves the spreading of the cancer cells beyond their origin, and it often associates with poor prognosis. Different types of cancers could acquire the ability to become metastatic, and we hypothesize that we could utilize this information to identify the genes, as well as their corresponding functions and pathway that are involved in the metastatic behaviors and are conserved among different cancers. Gene expression data from 8 different cancer types were analyzed in this study, and for each cancer type we identified the genes that are differentially expressed between metastatic condition (lymph node or distant organ metastasis) and primary condition by using the pre-defined false discovery rate (FDR) in SAM (Significance Analysis of Microarrays). By investigating the genes that are present across all the data, we identified 4799 and 4726 genes that are differentially expressed in at least one cancer type by using two different pre-defined FDR: 0.1 and 0.55, respectively. Two methods were used to identify genes that are consistently differentially expressed in the metastatic versus non-metastatic conditions: genes that have conserved pattern of differential expression were identified and the genes that have common differentially expressed pattern are clustered. By using the first method, we identified genes that are involved in the cyclin-dependent kinase pathway, and interestingly, genes that encoding heat shock proteins and chaperonins are conservely upregulated in the metastatic cases. For the second approach, hierarchical clustering with specified distance matrix of the differential expression pattern was used. The enriched motif for the clusters, as well as those from the genes from the first approach were also studied, and the binding sites for the transcription factors ETS, GABP, and STAT were enriched consistently in each approach. Our results match and support previous studies: GABP was found to be cooperating with other transcription factor to induce gene that involves in metastasis (Jiang P. et al., 2002); STAT1 was negatively associated with tumor angiogenesis and metastasis (Huang S. et al., 2002); overexpression of ETS-1 was significantly related to lung metastasis (Sato T. and Miwa a., 2002). We also investigate the differences between the metastatic and primary cells in the system level by looking for the enrichment of clustered genes in the signaling pathway. One cluster from the result is enriched with the genes involved in the actin cytoskeleton pathway, and this includes functions relating to adherens junction, lamellipodia, and filopodia formation. The meta-analysis of the gene expression data in this study provides a better understanding about the genes that are involved consistently in different types of cancer metastasis, and this could provide valuable information for us to study the mechanism of the progression of cancer, as well as the mutations for the metastasis to occur.

**References:**

Huang, S., et al. 2002. Stat1 negatively regulates angiogenesis, tumorigenicity and metastasis of tumor cells. *Oncogene* (16):2504-12.

Jiang P., et al. 2002. Cloning and characterization of the human heparanase-1 (HPR1) gene promoter: role of GA-binding protein and Sp1 in regulating HPR1 basal promoter activity. *J. Biol. Chem.* (11): 8989-98.

Sato T., and A. Miwa. 2002. Ets-1 and integrin beta3 for lung metastasis from colorectal cancer. *APMIS*. 110(4):347-53.