Distinguishing primary breast cancers and their lymph node metastases by knowledge-based multidimensional scaling analysis

Harri Lähdesmäki, Xishan Hao, Baocun Sun, Limei Hu, Olli Yli-Harja, Gregory N. Fuller, Ilya Shmulevich, Wei Zhang

Department of Pathology, Cancer Genomics Core Lab, the University of Texas M. D. Anderson Cancer Center, Houston, Texas (HL, LH, GNF, IS, WZ); Institute of Signal Processing, Tampere University of Technology, Tampere, Finland (HL, OY-H); Tianjin Medical University Cancer Hospital (XH, BS), Tianjin, China.

Purpose of the study: The goal is to combine gene expression profiling with knowledge-based multidimensional scaling analysis to identify key biological pathways or processes that distinguish between primary breast cancers and their lymph node metastases.

Experimental Procedures: We first acquired gene expression profiles of 9 matched primary tumors and metastases using cDNA microarrays generated by our Genomics Core Laboratory. Some of the gene expression results were validated by a tissue microarray study using 100 paired samples. As a clustering tool, we applied the multidimensional scaling (MDS) algorithm to the preprocessed data. We first performed unsupervised MDS using all of the 2,300 genes. Then, we selected 280 genes that showed significant changes in expression in at least three pairs of matched tumors. This 280 informative gene set was used in the second MDS analysis. Further, we separate the 280 genes into six functional groups based on prior knowledge of those genes: cell cycle, apoptosis, metabolism, cell adhesion and migration, signal transduction, and transcriptional factor and DNA binding molecules. Each of the six functional gene sets was then used to perform knowledge-based MDS analyses.

Summary of results: Unsupervised MDS clustering methods using all the 2,300 genes did not separate primary and metastases. MDS analysis using 280 informative gene set revealed that primary tumors were quite tightly clustered whereas the metastases samples were relatively heterogeneous. MDS analysis showed that different functional gene sets vary in their ability to separate primary tumors and their metastases. The best separations were found with gene sets of "cell adhesion and migration," "metabolism," "signal transduction," and "transcription factor and DNA binding molecule." In contrast, cell cycle separated the two groups less well and apoptosis regulatory genes did not separate the two groups at all. Results from the MDS analysis suggest that alteration in apoptosis regulation is not the key event leading to lymph node metastasis, whereas the alteration in cell migration and metabolism may be crucial. This is consistent with the prior knowledge regarding the biological difference between primary tumors and their metastases.

Conclusions: Incorporation of prior biological knowledge into MDS analysis may help reveal important biological mechanisms from gene expression profiles. Our MDS analysis showed that lymph node metastases represent a hetero-

geneous group of tumors with major alterations in cell adhesion and migration, metabolism, signal transduction, and transcription regulation.