Functional discrimination between cancer stem cells types: epithelial-mesenchymal transition and proliferation

Chueh-Lin Hsu[1], Feng-Hsiang Chung[1,2], Chih-Hao Chen[1] and Hoong-Chien Lee[1-4]

[1]Institute of Systems Biology and Bioinformatics, National Central University, Zhongli, Taiwan 32001
[2]Center for Dynamical Biomarkers and Translational Medicine, National Central University, Zhongli, Taiwan 32001
[3]Department of Physics, Chung Yuan Christian University, Zhongli, Taiwan 32001
[4]Cathay Medical Research Institute, Cathay General Hospital, Taipei, Taiwan

Background and Motivations

Disappointing results of standard treatments for preventing cancer relapses, include chemotherapy and radiotherapy, have recently been attributed to the stem cell-like properties of cancer cells[1-3]. The introduction and advancement of high-throughput gene expression profiling, through technologies such as microarray and next-generation sequencing, affords biologists an unprecedented means to discriminate between cancer cells and cancer stem cells (CSCs) for discovering novel therapeutic approaches[4, 5]. The vast amount of gene expression data that have been collected at public repositories in the last few years make excellent materials for such a study.

Proposed Approaches

We collected available gene expression CSCs data sets of multiple cancer types from the Gene Expression Omnibus (GEO) database [6] and used a variety of qualitatively different methods to cluster the data sets to establish functional characteristics of cancer specific CSCs. Data sets were sift by Principle Component Analysis (PCA). Fourteen CSC and four control data sets were used for the study. Methods used include: (1) Standard t-tests for selecting differentially expressed genes (DEGs), followed by identification of functional terms, as defined by Gene ontology (GO) [7], via overrepresentation; (2) Gene set enrichment analysis (GSEA) [8]; (3) Parametric analysis of gene set enrichment (PAGE) [9]; (4) Generally applicable gene-set enrichment (GAGE) [10]; (5) A statistical method respecting molecular heterogeneity, Weighting Arrays By Error (WABE)[11], to identify DEGs, followed by GO analysis. We used the clustering results to query the Connectivity Map (CMAP) database [12] to search therapeutics drugs.

Results and Conclusions

Cancer types represented in the fourteen CSCs data sets used in this study are: breast, glioma, colon, lung, ovarian, and prostate; while those represented in the four non-CSCs data sets are: colon adenoma, embryonic stem cell, induced pluripotent stem cell and TGF-beta treated lung adenocarcinoma.

There was no significant common intersection of genes among the 14 DEG sets culled from the 14 data sets using the standard t-test method. The three gene set-based methods (GSMs), GSEA (NOM p-value < 0.05), PAGE (FDR< 5e-18), and GAGE (FDR< 5e-7), yielded similar clustering results; all made the same division of the 14 CSC data sets into two types (Figure 1). One, the proliferation type, included the glioma and lung CSCs, was highly enriched in genes involved in proliferation (but not EMT) functions, and the other, the EMT type, included the breast, colon, prostate and ovarian CSCs, was highly enriched in genes involved in EMT (but not proliferation) functions (Figure 2).

We queried the CMAP with the GSM gene sets to construct lists of drugs with statistically significant high GSEA scores (Figure 3). About 18% of drugs in both lists constructed from the proliferation and EMT types were anti-tumor drugs. The list for the EMT type was rich ($p < 0.05$ in Fisher's exact test) in "promoting" drugs, or drugs whose genomic profile correlate with genomic change from cancer to CSC, while the list for the proliferation type was rich in "reversing" drugs, drugs whose genomic profile correlated with CSC-to-cancer change. A high proportion of the promoting drugs were observed to be drugs used for chemotherapy. This implies that when administered to EMT-type CSCs chemo-drugs may promote CSC. Conversely, a majority of anti-tumor drugs are predicted to reduce CSC when administered to proliferation-type CSCs (Figure 4).

GO analysis of the CSC data sets by WABE showed that functions related to cell cycle processes were up-regulated in proliferation-type CSCs and down-regulated in EMT-type CSCs. Since many antitumor agents were designed for restraining cell cycle, our result suggests that such drugs are therapeutically ineffective for EMT type CSCs (Figure 5).

This is the first large-scale study to meta-analyze CSC gene expression data, to functionally discriminate between the two cancer stemness types, EMT and proliferation, and to discuss implication of this discrimination on therapeutic effect of CSC treatment by antitumor drugs.

Figure

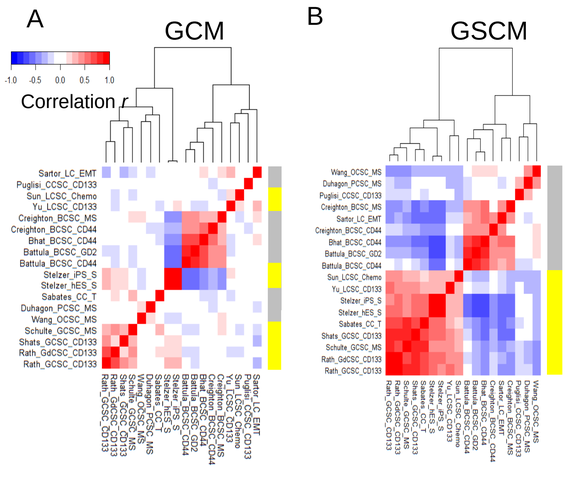

Figure 1. Comparison of gene correlation matrix (GCM) and gene-set correlation matrix (GSCM) within CSCs and control studies. (2a) GCM was produced using top-100 highest variance genes obtained from comparing all datasets with gene expression ratio (CSCs verse non-CSCs samples). The color bar represents two clusters identified in GSCM (gray and yellow). (2b) GSCM was produced using 152 significantly enriched/depleted gene-sets (nominal *p*< 0.05 by GSEA algorithm) with normalized enrichment scores (NES).

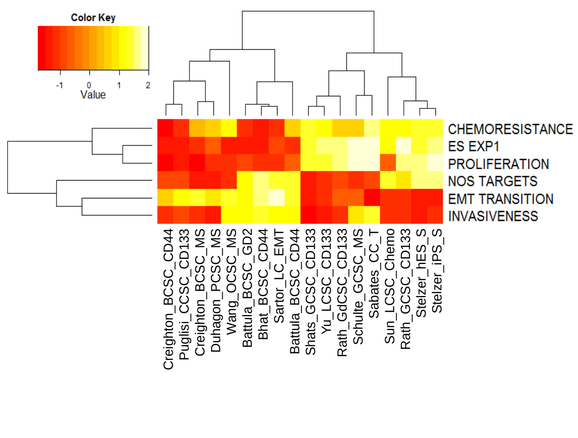

Figure 2. Hierarchical clustering using suggestive gene-sets related to cancer/stem cell signatures. By using the known gene-sets related to cancer and stem cell signatures, hierarchical clustering result showed that two important cancer signatures, proliferation and EMT transition.

Figure

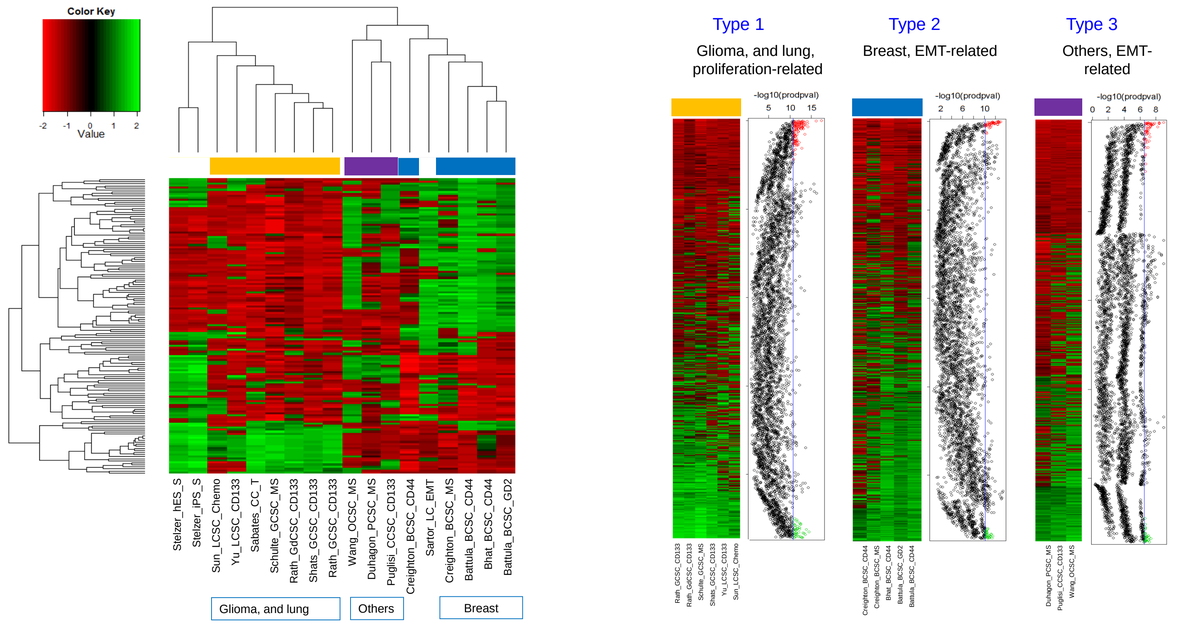

Figure 3. Identification of specific signatures for each types of CSC. We ranked gene-sets by product of nominal p-values for each CSC types, and selected top-5% of gene-sets.

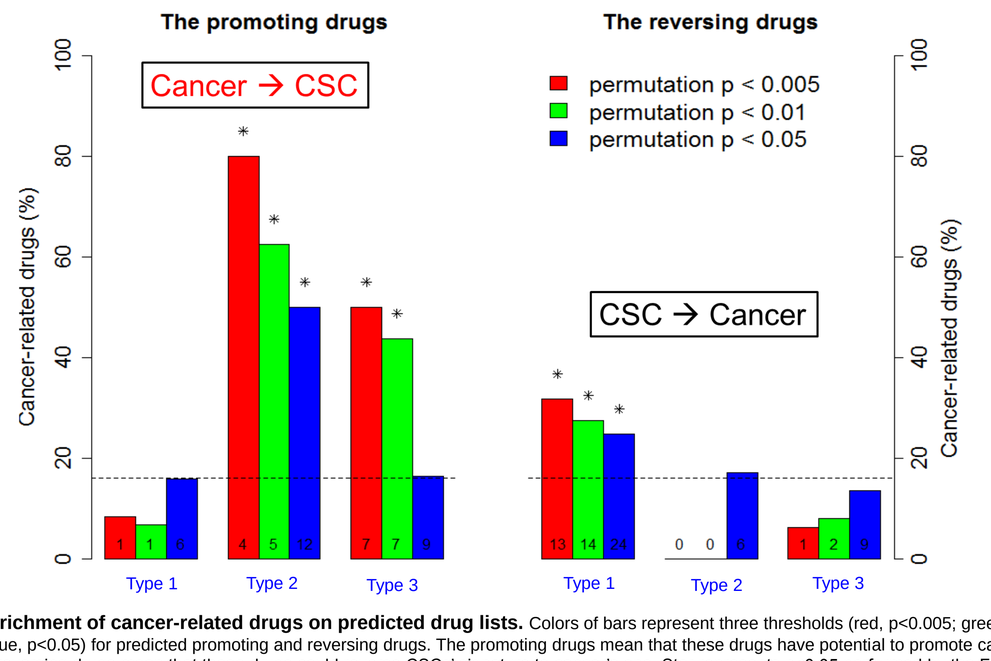

Figure 4. Enrichment of cancer-related drugs on predicted drug lists. Colors of bars represent three thresholds (red, p<0.005; green, p<0.01; and blue, p<0.05) for predicted promoting and reversing drugs. The promoting drugs mean that these drugs have potential to promote cancers to CSCs. The reversing drugs mean that these drugs could reverse CSCs' signature to cancer's one. Star represents p<0.05 performed by the Fisher's exact test comparing with reference background (~18% of cancer-related drugs according to TTD database).

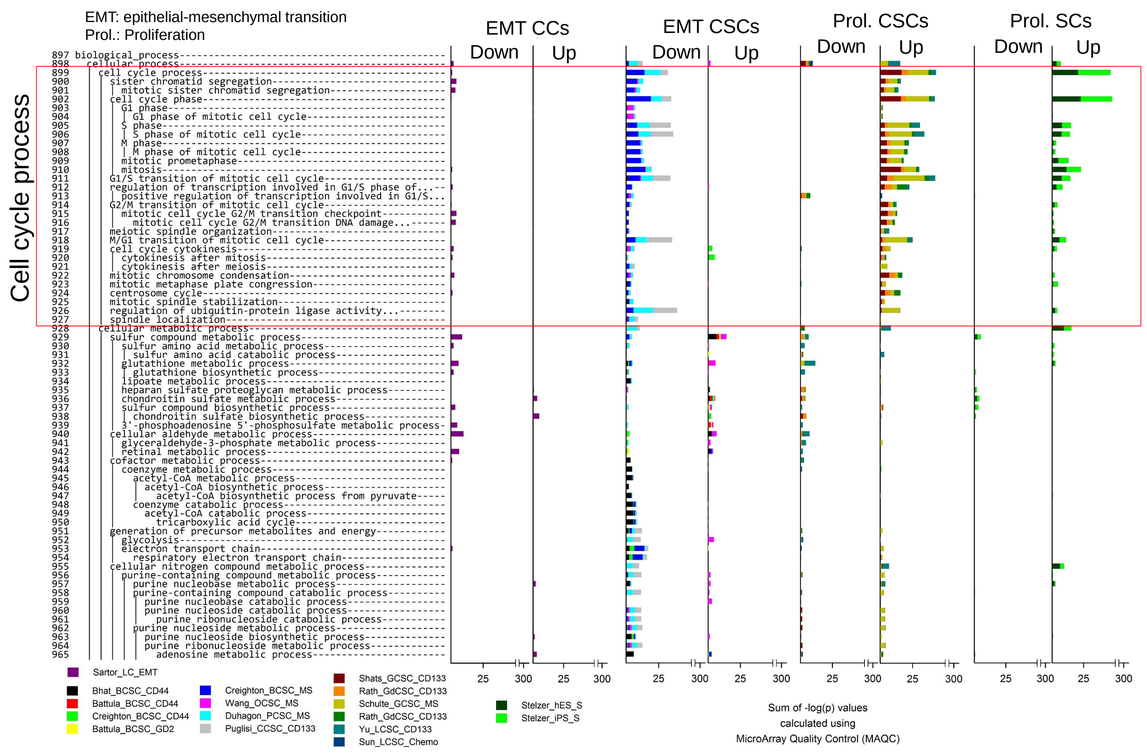

Figure 5. The comparison of enrichment ratios of GO with cell cycle process in EMT CSCs and Proliferation CSCs, Controlling for EMT CCs and Stem cells.

References

1. Diehn, M. and M.F. Clarke, Cancer stem cells and radiotherapy: new insights into tumor radioresistance. Journal of the National Cancer Institute, 2006. 98(24): p. 1755-1757.
2. Morrison, R., et al., Targeting the mechanisms of resistance to chemotherapy and radiotherapy with the cancer stem cell hypothesis. Journal of oncology, 2010. 2011.
3. Bao, S., et al., Glioma stem cells promote radioresistance by preferential activation of the DNA damage response. Nature, 2006. 444(7120): p. 756-760.
4. Ponti, D., et al., Isolation and in vitro propagation of tumorigenic breast cancer cells with stem/progenitor cell properties. Cancer research, 2005. 65(13): p. 5506-5511.
5. Al-Hajj, M., et al., Prospective identification of tumorigenic breast cancer cells. Proceedings of the National Academy of Sciences, 2003. 100(7): p. 3983-3988.
6. Edgar, R., M. Domrachev, and A.E. Lash, Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic acids research, 2002. 30(1): p. 207-210.
7. Ashburner, M., et al., Gene Ontology: tool for the unification of biology. Nature genetics, 2000. 25(1): p. 25-29.
8. Subramanian, A., et al., Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences of the United States of America, 2005. 102(43): p. 15545-15550.
9. Kim, S.-Y. and D.J. Volsky, PAGE: parametric analysis of gene set enrichment. BMC bioinformatics, 2005. 6(1): p. 144.
10. Luo, W., et al., GAGE: generally applicable gene set enrichment for pathway analysis. BMC bioinformatics, 2009. 10(1): p. 161.
11. Chen, C.-H., et al., Method designed to respect molecular heterogeneity for replicated analysis of dfferential expression can substantially correct present interpretations of 80–90% of existing data(submitted). 2013.
12. Lamb, J., et al., The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. Science, 2006. 313(5795): p. 1929-1935.