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 split 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 analysis (ORA) [8]; (2) Gene set enrichment analysis (GSEA) [9]; (3) Parametric analysis of gene set enrichment (PAGE) [10]; (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 therapeutic 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 (GSEMs), GSEA (Nom p-value < 0.05), PAGE (FDR<0.16), and GAGE (FDR<0.7), 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 GSP 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 in (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 correlate 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, chemotherapeutic drugs may promote CSCs. 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 anti-tumor 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 cancer treatment by anti-tumor drugs.