



List of Dataset Used in the Analysis

- GSE17648
- GSE21304
- GSE22867
- GSE26126
- GSE26990
- GSE27284

Using six independent cancer datasets (5 solid tumors including breast, colorectal, glioblastoma, lung, and prostate, and 1 myeloma) profiled on Illumina Human Methylation27 BeadArray, we identified genes harboring aberrant promoter methylation between normal and cancer samples using an empirical Bayes method. Each gene was assigned a p-value and change in % methylation indicating the level of hyper- or hypo-methylation. The gene list in each study was then subjected to LRPath enrichment analysis using GO terms, Biocarta, EHMN metabolic, KEGG, and Panther Pathways.

To identify the GO terms and pathway concepts commonly altered across various types of cancer, concepts with p-value less than $1e-3$ in at least half of the studies were subjected to the clustering analysis using Euclidean distance matrix and the ward linkage method. The resulting files were opened and viewed using Java TreeView. In Java TreeView, the user is able to flexibly work with the data, including options for searching, filtering, adjusting images, and exporting.

Clustering analysis performed on the results of LRPath identified a tight cluster of 82 concepts that show consistent aberrant DNA methylation across the various cancer studies. In the heatmap, red indicates hypomethylation and green indicates hypermethylation in cancer. From the clustering tree for the columns (studies), we also see that there appears to be one outlier cancer study, GSE26126.

