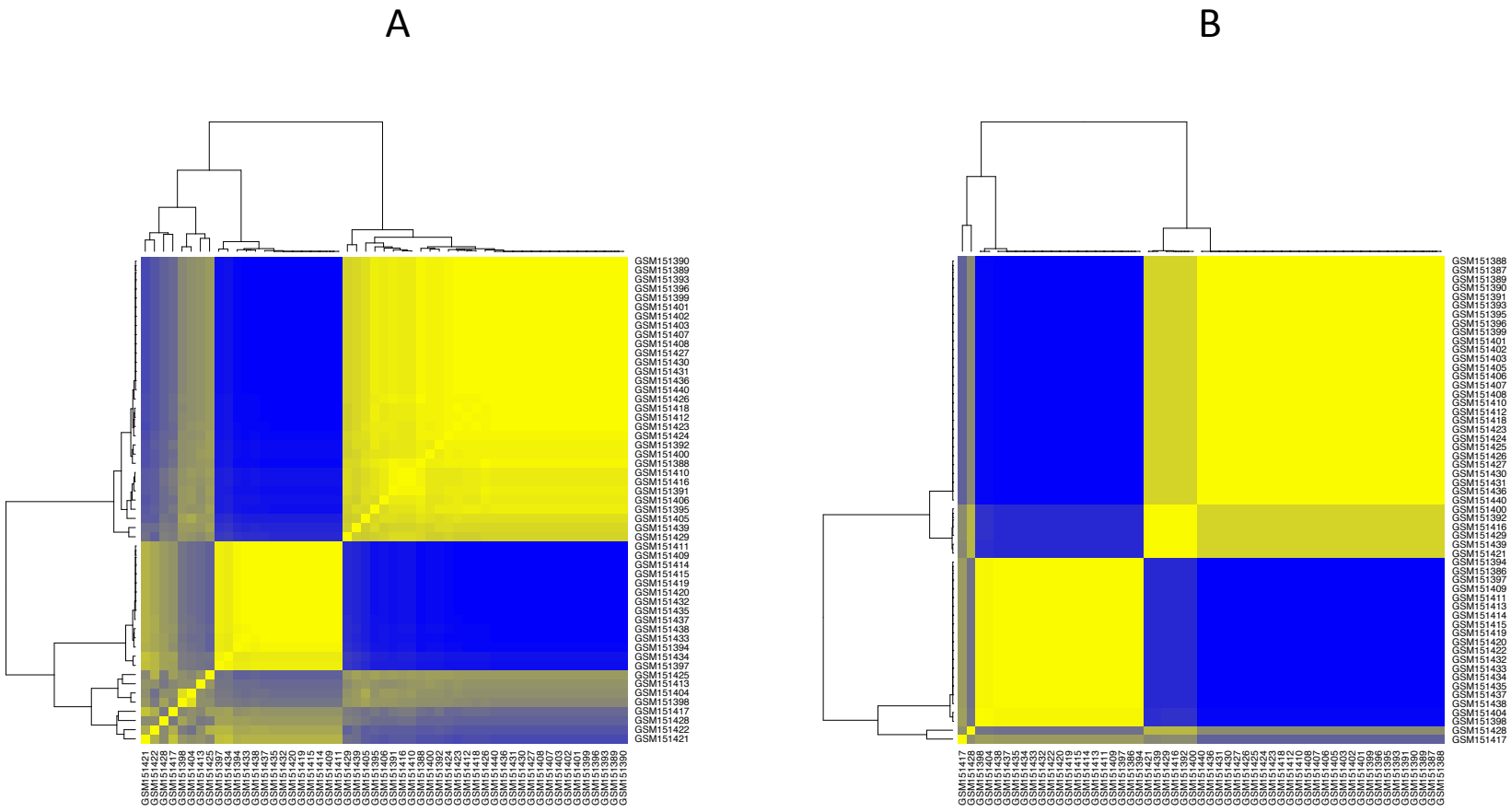

Supplementary Figure 1. Bootstrapping cluster analysis of the derivation cohort with *k* = 2, and 200-fold re-sampling. Hierarchical clustering was used, with euclidean distance and Ward's method for agglomeration. Color map values range from pure blue (the samples are in the same branch 0% of the time) to pure yellow (the samples are in the same branch 100% of the time). (A) Result using the initial gene set derived from Genbank. (B) Results following the gene enrichment stages. Analysis carried out in R using the ClassDiscovery package.