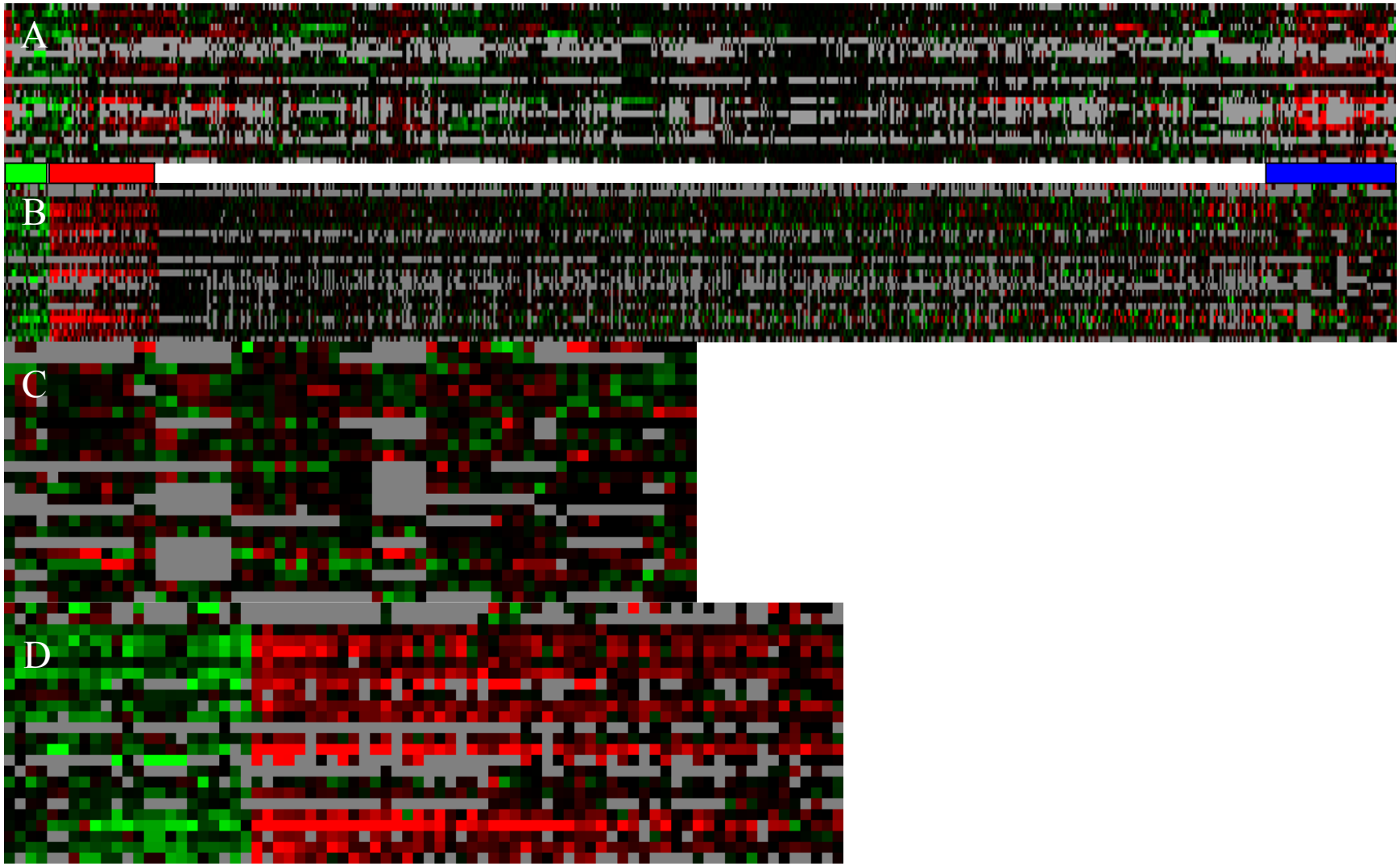
A
B
C
D

Supplementary Figure 1 - Clustering genes only vs. clustering both genes and samples (biclustering)

This figure shows the expression of the genes (rows) that were selected as Module 29 when using the Module Map algorithm. Red represents induced, green represents repressed, black represents unchanged relative to the median and gray represents missing values. When looking at all samples in their original random ordering (A), it is very hard to see a coherent pattern. A coherent pattern is a sample (column) that has most of the genes with a consistent expression; either induced (red) or repressed (green).

After grouping the samples by expression coherence (B), it is possible to see samples where most of the genes are repressed (below the green rectangle), and samples where most of the genes are induced (red rectangle). We can also see a group of samples which contains genes that are expressed but not in a coherent manner, so that for a given sample some genes are induced and some are repressed (below the blue rectangle). A large group of samples (middle part of figure B) are unchanging (black) - expressed at a level very close to the median, neither induced nor repressed.

Standard clustering algorithms attempt to find a group of genes that behave in a consistent manner in all samples - the genes must be consistently induced, repressed, or unchanging in each of the samples in the group (all 700 in this case). The presence of samples where the genes are both induced and repressed (below the blue rectangle in B, seen in higher magnification in C) would thus prevent standard clustering algorithms from working - no clusters will be discovered in this group of genes if we demand that all samples will be consistent.

Using biclustering algorithms, such as Module Map, we can discover genes behaving in a coherent manner in some of the samples (red & green rectangles in B, higher magnification in D). Focusing on these samples allows us to discover statistically significant consistent behavior, which can have biological meaning (see the discussion of Module 29 in the text).