

(a) Simulated y_{sg} ordered by true cluster assignments.

(b) Simulated y_{sg} ordered by estimated cluster assignments.

Figure 4.5: (a): Observations ordered according to the simulation truth. (b): Observations ordered according to estimated cluster membership indicators a posteriori. In both panels, rows represent samples while columns represent proteins.

4.5 Lung cancer dataset

4.5.1 The data

The real dataset analyzed in this section consist of protein profiles coming from an RPPA experiment on lung cancer. The data contains 233 proteins pre-selected according to their biological relevance to the study of this type of cancer. The data contains samples from 687 patients and 124 cell lines. The objective is to identify similarities with

4.5.2 Simulation

We describe here the results of successive application of POE model from section 4.2 followed by NoB-LoC from section D.3.

We start by analysing the results of directly applying the NoB-LC

model to the original lung data (without running POE first), which is illustrated in Figure 4.6 (a).

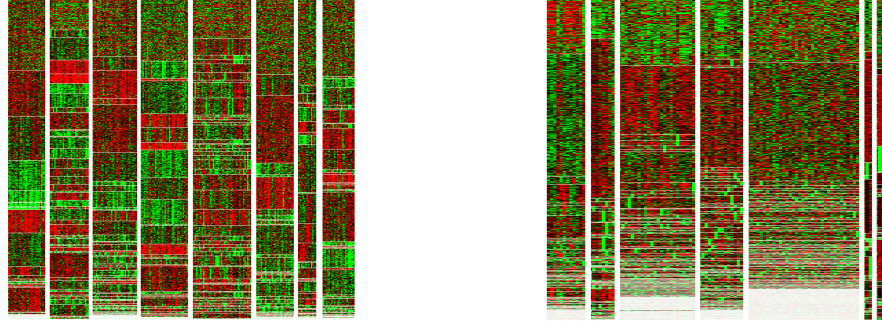


Figure 4.6: Observed protein expression arranged according to posterior estimated cluster structure obtained by NoB-Loc. Only active proteins are displayed. Panel (a) shows the result of application of the NobLoc model on the original data and panel (b) shows the results after the application of POE

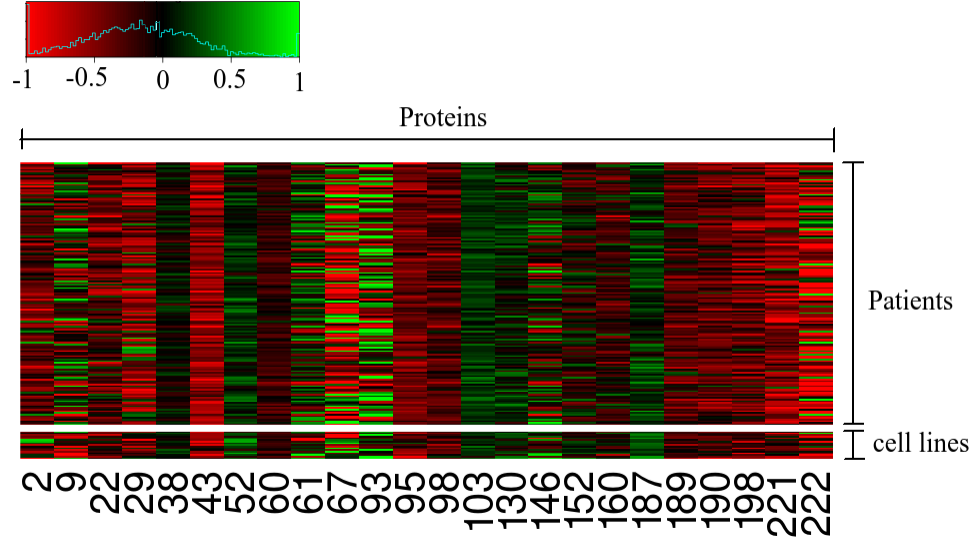


Figure 4.7: Protein expressions within one of the samples/proteins blocks of Figure 4.6 (b). The cell lines and patients exhibit very similar profiles when considering the subset of proteins that were clustered together by the model.