

Goals: Inter-study validation of microarray data to obtain prognostic profiles.

Strategy: Two-stage Bayesian mixture modeling.

Stage1

Collect microarray data from studies that address similar questions.



Obtain the set of genes that are present on each independent array data, take base 2 logarithm and apply proper normalization on each data set.



Fit mixture models through a Markov Chain Monte Carlo (MCMC) sampling algorithm on each independent raw data matrix and obtain $poe(\hat{p}^*)$ matrices by Bayes rule:

$$\begin{array}{c} \log_2(Raw) \\ \left[\begin{array}{ccc} 5.8 & 4.2 & . \\ -1.6 & -3.1 & . \\ . & . & . \\ . & . & . \end{array} \right] \rightarrow \left[\begin{array}{ccc} 0.8 & 0.6 & . \\ -0.1 & -0.2 & . \\ . & . & . \\ . & . & . \end{array} \right] \\ \begin{array}{c} (-\infty, \infty) \end{array} \qquad \qquad \begin{array}{c} poe \\ (-1,1) \end{array} \end{array}$$



Combine data on *poe* scale.



Stage2

Build prognostic signature on combined data matrix.

(Can be generalized to any type of microarray data analyses such as discriminant analysis, gene screening, class identification and prediction, etc.)