

Input: training set TD with G genes and n samples

Pre-processing step: Rank-order all G genes by applying Cox Proportional Hazards Regression to each individual gene. Let $x_1, x_2, \dots, x_G$ be the ordered list of genes, sorted in descending order of log likelihood. Let $maxNvar$ denote the user-specified size of the BMA window (maximum 30).

Parameters: $nbest$ and p , where p is the total number of genes to be processed such that $maxNvar < p \leq G$.

1. Initially, start with the $maxNvar$ top ranked genes ($x_1, x_2, \dots, x_{maxNvar}$), and apply the traditional BMA algorithm for survival analysis (Volinsky et al., 1997). Let $toBeProcessed$ be an ordered list of genes with ranks $(maxNvar + 1)$ to p . Initially, $toBeProcessed \leftarrow (x_{maxNvar+1}, x_{32}, \dots, x_p)$.
2. Repeat until all p genes are processed
 - a. Remove all genes i with $\Pr(b_i \neq 0 \mid TD) < 1\%$.
 - b. *Adaptive threshold step:* If all genes have $\Pr(b_i \neq 0 \mid TD) \geq 1\%$, determine the minimum $\Pr(b_i \neq 0 \mid TD)$, $minProbne0$, among the $maxNvar$ genes in the current BMA window. Remove all genes with $\Pr(b_i \neq 0 \mid TD) < (minProbne0 + 1)\%$.
 - c. Let $removedGenes$ be the set of genes removed, and suppose q genes are removed.
 - d. Replace the q removed genes with the q -next-up genes from $toBeProcessed$. Update $toBeProcessed \leftarrow toBeProcessed - q\text{-next-up}$.
 - e. Apply the traditional BMA algorithm for survival analysis.

Output: selected models and their posterior probabilities, selected genes and their corresponding posterior probabilities ($\Pr(b_i \neq 0 \mid TD)$), maximum-likelihood estimates of the regression parameters in each model.