

PaGE 2.1 Algorithm

Let the groups be indexed by $0, 1, \dots, \ell$, where 0 corresponds to the reference group.

For each $i = 0, \dots, \ell$, let t_i be the number of replicate experiments in the i -th group.

Let $x_{g,i,k}$ be the (cleansed and normalized) intensity of gene tag g in the k -th experiment in the i -th group, when defined.

Finally, let $t_{g,i}$ be the number of replicate values for gene tag g in group i (note that $t_{g,i} = t_i$ whenever tag g has a value in each of the experiments for group i).

Let $(n_0, n_1, \dots, n_\ell)$ be the *minimum presence list*, i.e. list of minimum number of values a tag should have for each group in order to be considered.

Let s (between 0 and 1) be the desired false positive rate.

1. Eliminate from further consideration any tag g such that $t_{g,i} < n_i$ for some $i = 0, 1, \dots, \ell$.
2. For each tag g and for each, $i = 0, \dots, \ell$, let $\bar{x}_{g,i}$ be the average of the (available) values for tag g in group i .
3. If *shift* is provided, go to 5. Else, let $m = \min_{g,i} \frac{\sqrt{t_{g,i}} \bar{x}_{g,i}}{s_{g,i}}$ and $\varsigma = \max_{g,i} \frac{s_{g,i}}{\sqrt{t_{g,i}}}$ where $s_{g,i}$ is the sample standard deviation of the $x_{g,i,k}$. Let *shift* = $\max\{0, (\text{shift_par} - m)\varsigma\}$, where *shift_par* is as provided by the user (default is 3).
4. For each tag g and for each, $i = 0, \dots, \ell$ and $k = 1, \dots, t_i$, let $x_{g,i,k} = x_{g,i,k} + \text{shift}$ (if $x_{g,i,k}$ is defined) and $\bar{x}_{g,i} = \bar{x}_{g,i} + \text{shift}$.
5. Eliminate from further consideration any tag g for which $\bar{x}_{g,0} = 0$.
6. For each tag g and for each $i = 1, \dots, \ell$, let $r_{g,i} = \frac{\bar{x}_{g,i}}{\bar{x}_{g,0}}$. Then the list of ratios associated to gene tag g is $(r_{g,1}, r_{g,2}, \dots, r_{g,\ell})$.
7. For each $i = 1, \dots, \ell$, let \min_i (resp. \max_i) be the minimum (resp. maximum) of $r_{g,i}$ over all tags.
8. For each $i = 1, \dots, \ell$, let f_i be the empirical distribution of $\left(\frac{\frac{x_{g,i,k}}{\bar{x}_{g,i}} - 1}{\sqrt{t_{g,i}} - 1} + 1\right) / \left(\frac{\frac{x_{g,0,h}}{\bar{x}_{g,0}} - 1}{\sqrt{t_{g,0}} - 1} + 1\right)$, for g varying over the gene tags and k and h varying over the replicates for gene tag g in group i and 0 respectively. (If $t_{g,i} = 1$ replace the numerator by $\frac{x_{g,i,k}}{\bar{x}_{g,i}}$; similarly if $t_{g,0} = 1$.) This distribution is computed by dividing the range of the values into 100 bins.
9. For each $i = 1, \dots, \ell$, compute the smallest C_i such that

$$\sum_{j=\text{bin containing } C_i}^{\text{last bin}} f_i(j) < s$$

through a binary search. In a similar fashion compute c_i .

10. For each $i = 1, \dots, \ell$, the level cutoff list is obtained by taking all successive powers of C_i which are strictly less than \max_i and all successive powers of c_i which are strictly greater than \min_i and for which there is at least one smaller non-zero $r_{g,i}$. Let $0 < a_{i,1} < a_{i,2} < \dots < a_{i,s_i}$ be the level cutoff list thus obtained. Let $B_{i,-m_i}, B_{i,-m_i+1}, \dots, B_{i,n_i}$ denote respectively the intervals $[0, a_{i,1}), [a_{i,1}, a_{i,2}), \dots, [a_{i,s_i}, \infty)$, where $B_{i,0}$ is the interval containing 1.

11. For each g and for each $i = 1, \dots, \ell$, let j_i be such that $r_{g,i} \in B_{i,j_i}$. The pattern attached to g is then $(j_1, j_2, \dots, j_\ell)$.