

Formula 1: Position-specific weight matrix (PSWM) scores

$$score = \sum_{i=1}^l \log\left(\frac{(n_{if} + 1)}{(N_i + 4)} * (1/4)\right)$$

where l : length of the matrix; n_{if} : the observed number of base ' f ' (A, C, G or T) at the i -th position in the training; N_i : sum of the observed number of all bases at i -th position.

Formula 2: Ratio of the score to the maximum score

$$ratio = \frac{score - \min(score)}{\max(score) - \min(score)}$$

where $\min(score)$: minimum value of each PSWM, $\max(score)$: maximum value of each PSWM.

Formula 3: Hypergeometric distribution

$$\sum_{x=i}^{Np} \frac{\binom{Np}{x} \cdot \binom{N(1-p)}{n-x}}{\binom{N}{x}}$$

where N : total number of promoters, p : expected frequency of the promoters associated with a given PSWM (number of associated promoters/ N), n : total number of promoters belonging to a dataset of concern, and i : number of the promoters associated with a given PSWM in a dataset.