Bcrank Info

Christian Otto

To use the BCRANK (Ameur 2025) algorithm to its full potential, the DNA sequences should be placed in order of most likely to contain the consensus motif. The BCRANK algorithm begins by randomly selecting a consensus sequence. The score is then calculated by taking into consideration the number of occurrences of this consensus sequence and the rank of the DNA sequence. After this score has been calculated, all consensus sequences in a neighborhood of the initial guess are scored and the one with the highest score is kept. This finds a local optimum. The entire process is started again with a new randomly selected consensus motif. We repeat this process 100 times in hopes of finding the global optimum. (Ameur et al. 2009)

The neighborhood of the consensus sequence is represented by being able to add or remove a nucleotide on either side of the consensus sequence. The neighborhood of the consensus sequence also allows for one nucleotide to be changed to a different nucleotide. This neighborhood methodology allows for the algorithm to search for a motif of length k, but it also allows the algorithm to explore a motif up to a length of $k \pm 2$. Because the BCRANK method assumes an ordering of the DNA sequences, the scoring function checks how many times that consensus sequence appears in each of the DNA sequences, weighing the appearance of the consensus sequence higher in the top ranked DNA sequences. BCRANK then reorders the DNA sequences (to not be ranked at all) and scores them, this reordering process is repeated many times to form a distribution of unranked scores. We then compute a t-statistic for our ranked DNA motif to determine if the consensus motif is an optimum. (Ameur et al. 2009)

References

Ameur, Adam. 2025. *BCRANK: Predicting Binding Site Consensus from Ranked DNA Sequences*. https://doi.org/10.18129/B9.bioc.BCRANK.

Ameur, Adam, Alvaro Rada-Iglesias, Jan Komorowski, and Claes Wadelius. 2009. "Identification of Candidate Regulatory SNPs by Combination of Transcription-Factor-Binding Site Prediction, SNP Genotyping and haploChIP." *Nucleic Acids Research* 37 (12): e85–85. https://doi.org/10.1093/nar/gkp381.