Selecting High Efficacy siRNAs by Computational Models

Wei-Jie Pan¹, Chi-Wei Chen¹, Heng-Hao Liang¹, Yen-Wei Chu¹,²,³,⁴

¹Institute of Genomics and Bioinformatics, National Chung Hsing University, Taichung, Taiwan
²Biotechnology Center, National Chung Hsing University, Taichung, Taiwan
³Institute of Molecular Biology, National Chung Hsing University, Taichung, Taiwan
⁴Graduate Institute of Biotechnology, National Chung Hsing University, Taichung, Taiwan

siRNAs is an important factor which can induce RNA interference for gene silencing. However, selecting a suitable siRNA is crucial to RNA interference experiment. An siRNA with high inhibition efficacy can ensure RNA interference successfully happened. The data of siRNA efficacy have considerably grown in recent years, which provides the fundamental materials for the analyzing and selecting of high efficacy siRNAs. In this study, we selected features as inputs of our learning models by analyzing siRNA sequences and the data of siRNA inhibition efficacy, which are collected in siRecords database. To assist biologists in designing a high efficacy siRNA with a variety of experimental conditions, we divided siRecords database into subsets by three kinds of inhibition efficacy as the threshold to distinguish high and low efficacy siRNAs. Support vector machines and neural networks were used for constructing learning models in different subsets separately. The comparison of results showed the performance of support vector machines with weighted coding approach for selecting high efficacy siRNAs has the MCC of above 80% with the threshold of 90% and 70%.

Keywords: siRNA, RNA Interference, Support Vector Machines, Neural Networks