

A Novel Approach for Integrating and Assessing Computational Tools to Detect Transcription Factor Binding Sites

K. Dempsey*, D. Quest**°, D. Bastola*, H. Ali**°. *College of Information Science & Technology, University of Nebraska at Omaha, Omaha, NE 68182; °Department of Pathology & Microbiology, University of Nebraska Medical Center, Omaha, NE 68198

Background and Objective: Transcription factor binding sites (TFBS) regulate the expression of genes in the cell. Their discovery remains one of the most challenging problems in molecular biology. To solve this problem, traditional techniques have been supplemented by the development of computational prediction methods. Today over 100 algorithms exist for detecting TFBS in various problem domains, yet it remains unclear how to assess their performance. Comparing these tools across datasets is nearly impossible due to a lack of standards for running programs and reporting results. This work proposes a novel method for standardizing runtime procedures and assessing tool performance.

Methods: A standard reporting format was developed for each tool to compare the methods. We developed a pipeline framework to integrate, score, and evaluate TFBS identification for each method.

Results: We evaluated 12 computational methods for detecting TFBS and obtained statistics that describe their performance. These results allowed us to rank each method by performance on a range of datasets.

Discussion and Conclusions: Computational detection of TFBS remains a challenging problem. Sensitivity and specificity remain low for the algorithms regardless of dataset. Our method for comparing tools exposes a unique view into the behavior of TFBS detection and enables the improvement of methods in the future.

Acknowledgement: This project was supported by the NIH grant number P20 RR016469 from the INBRE Program of the National Center for Research Resources.