
Combining computational and experimental approaches to comprehend the "how" and "why" of bacterial transcriptional regulation

Peter Freddolino^{*†1}

¹University of Michigan – United States

Abstract

Recent advances in high-throughput sequencing technology have yielded a huge increase in our knowledge of genomic sequences, but DNA sequence information remains meaningless without corresponding functional insight. It is only through a synthesis of computational approaches and highthroughput experiments that any meaningful headway can be made in the task of moving from genome sequence information to functional information at the scales of modern biology. We have recently launched two such initiatives, aimed at completely mapping the transcriptional regulatory logic and functional proteome of *Escherichia coli*. Using a broadly applicable non-specific method for mapping genome-wide protein occupancy, we have begun to identify the binding motifs, functions, and condition-dependent behavior of many cryptic *E. coli* transcription factors. In the process, we have also identified the presence of heterochromatin-like silenced regions on bacterial chromosomes, which we have found play a key role in regulating stress-response and virulence genes across several bacterial species. Combination of our findings with multiscale simulations of protein-DNA interactions promise to provide a unifying model for the interplay of bacterial chromosomal structure and gene expression. To address the problem of assigning functions to poorly annotated proteins without suitably close homologs for sequence-based annotation methods to be effective, we have recently developed a hybrid pipeline combining structural prediction/alignment, sequence alignment, and protein-protein interaction information to obtain combined structure predictions and functional annotations for entire proteomes. We find that our inclusion of structural information makes our workflow unusually strong in performance on difficult targets with limited sequence identity to annotated proteins. Application of our methods at the scale of entire proteomes yields a rich new source of information to seed detailed investigation of the functions of many previously mysterious protein-coding genes.

*Speaker

†Corresponding author: petefred@umich.edu