



University of Zabol

Graduate school

Faculty of Science

Department of Bioinformatics

Thesis Submitted for the Degree of M.Sc (in the field of Bioinformatics)

Title

A survey on Genome-Wide distribution of Transcription Factor Binding
Sites

Supervisors

Dr. M. Rokouei

Dr. A. Ebadi Tabrizi

By

A. Dezhbord

June 2024

Abstract

The lack of production of a new class of antibiotics in recent years and the indiscriminate and unsystematic use of antibiotics in the treatment of bacterial infections in humans, animals and even plants, as well as the use of antibiotics as a growth stimulant in livestock have caused bacteria to reach a stability in antibiotic resistance. One of the computational methods to investigate the mechanism of resistance is to analyze the binding sites of transcription factors specific to antibiotic resistance genes. Since transcription binding sites bind specifically to the factor and binding transcription factors can affect the expression of genes involved in antibiotic resistance, studying them and their interaction with transcription factors is crucial to identify new targets for antibiotic development. In the upcoming study, binding sites of transcription factors in *ST131* strain of *E.Coli* bacteria in two samples including 8 antibiotic resistance genes and 1000 random genes with uniform distribution were extracted with SITECON model. By applying Fisher's exact test on the data, the number of 12 transcription factors FNR, GlpR, GNTR, HNS, IHF, LEXA, ISCR3, MARA, MetR, MLC, PHOB, TyrR were obtained with a significance level of 0.05. The binding sites of 12 statistically established transcription factors were evaluated over the entire intergenic distances in the genome. By validating the data with the CollecTF database, 4 transcription factors HNS, LEXA, IHF, Fnr overlapped with our data experimentally. Target genes for 4 considered transcription factors were extracted from the whole genome and functionally analyzed with GO and KEGG databases. This study led to the presentation of a new in silico statistical technique for the process of automatic annotation of genes.

Key words: transcription factors, gene expression, genom