

**YENİ NESİL DİZİLEMEDE, ÖZELLİKLE CHIP-SEQ DENEYLERİNDE
KARŞILAŞILAN VE GENOMDA BİRDEN FAZLA BÖLGEYE EŞLEŞEN DİZİLERİN
DEĞERLENDİRİLMESİ**

Mehmet Seyit Zor

Biyomühendislik Bölümü

Proje Tezi

Proje Danışmanı: Yrd. Doç. Dr. Alper YILMAZ

Yeni Nesil Dizileme teknolojilerindeki gelişmeler şüphesiz moleküler biyoloji ve genetik çalışmalarına hız kazandırmış ve geleneksel yöntemlerin modernleşmesinde bir araç olmuştur. Dizileme odaklı çalışmalardan biri olan ChIP dizileme metodu, DNA molekülü üzerinde görev yapan proteinlerin bağlanma bölgelerinin analizinde gelenekselleşmiş bir metottur. Bu metodun başarısı, laboratuvar uygulamaları sonrasında elde edilen kısa DNA dizilerinin bir referans genoma haritalanmasına bağlıdır. Haritalama sürecinde dizilerin yaklaşık %75'i genomda tek bir yere eşlenir; ancak genomda birden fazla yere eşleşen dizilerin varlığı problematik bir haldedir. Bu diziler çoğunlukla uygulama sırasında gözden çıkarılır. Biz bu projede dizi haritalama sürecine dair bilgileri, dijital araçları ve bu süreçte karşılan söz konusu çoklu eşleşmelerin geri kazanılmasına yönelik güncel yaklaşımları derledik. Yanı sıra gerçek veriler ile yaptığımız uygulamalar sonucunda birden fazla yere eşlenen dizilerin anlamlılık potansiyeli taşıdığını ve bu dizilerin geri kazanılmasında verimli kullanılabilecek değerli bilgileri ortaya koymaya çalıştık.

Anahtar Kelimeler: ChIP-Seq, yeni nesil dizileme, hizalama, çoklu eşleşme, genom, biyoinformatik

UTILIZATION OF MULTIPLE MAPPED SEQUENCES IN NEXT GENERATION SEQUENCING, ESPECIALLY IN CHIP-SEQ EXPERIMENTS.

Mehmet Seyit ZOR

Bioengineering Department

Project Thesis

Advisor: Assist. Prof. Dr. Alper Yilmaz

Advances in Next Generation Sequencing decidedly has given a momentum to molecular biology and genetics. It has been a tool for modernization of conventional techniques. A sequencing-based method CHIP-Seq is used for identification of binding patterns of DNA-associated proteins. Success of this method depends on alignment of short reads onto a reference genome. Most mappable short reads align to a single genomic location with at least the rate of %75. However existence of reads which are mapped to more than one location (multiple mapped reads) onto genome is problematic. These reads are generally excluded from the study. In this project we searched the literature and compile information about sequence alignment, digital tools and rescue approaches regarding multiple-mapped reads. Also by using real sequencing data we try to prove potential of meaningfulness of multi-reads and we try to put forth valuable information that can be used efficiently.

Key words: CHIP-Seq, next generation sequencing, alignment, multiple matches, genome, bioinformatics.