Another 100 genes

Tom Röschinger[1], Grace Solini[1], Anika [2], Stephen Quake[2], and Rob Phillips[1, 3, +]

[1] *Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA 91125, USA*
[3] *Department of Physics, California Institute of Technology, Pasadena, CA 91125, USA*
[+] *Correspondence: phillips@pboc.caltech.edu*

1 Abstract

2 Introduction

It has been more than sixty years since Jacob and Monod (TR: cite) shaped the way we think about transcriptional regulation in procaryotes, yet, although more than x (TR: look up number) whole genome sequences have been stored in the NIH database (TR: cite), we have yet to obtain a full understanding of how all the genes of a single organism are regulated. Even in the case of one of biology's best studied model organism, *Escherichia Coli*, about two thirds of the genes lack any regulatory annotation (TR: add section to supp with details). For other prokaryotic model organisms the numbers are similar, while higher order model organisms as yeast (TR: write full name) and *C. elegans* have close to no regulatory annotations, given the arguably more complex nature of gene regulation in eukaryotes. Understanding how genes are regulated is required to understand how an organism adapts its physiology on short time scales to environmental stresses, as well as phylogenetic adaption on long time scales. In addition, gene regulation networks and their building blocks, such as transcription factor binding sites and RNA polymerase (RNAP) promoters, are key elements in the design of synthetic gene circuits.

With its ever increasing availability, Next Gen Sequencing (NGS) is primed to be the method of choice to discover transcription factor and RNAP binding sites *in vivo* with nucleotide resolution in high throughput, allowing us to build thermodynamic models for any promoter architecture and predict the change in gene expression due to mutations in binding sites (TR: cite kinney, belliveau, barnes, ireland. Do I need to mention other methods such as chip seq, atac seq etc and why they don't suffice?). Here, we build on previous work developing the method called Reg-Seq (TR: cite ireland), showing crucial improvements to the method and present the regulatory architecture of one hundred genes (TR: continue describing the genes studied and what we find about their architecture)

3 Methods

4 Results

4.1 Improved Method and summary of cloning results

4.2 Transcription Factor identification

4.3 Growth Conditions

4.4 Gold Standard genes

4.5 Ethanol iModulon

4.6 DNA damage repair iModulon

4.7 Antitoxin/Antibiotic genes

4.8 other y-ome genes

5 Discussion

- discuss how to scale to 1000 genes

6 To do list

- Write Introduction
- Collect references from reg-seq paper and new references
- write paragraphs about genes chosen
-

Supplemental Information for: Whatever the title will be

Tom Röschinger[1], Grace Solini[1], Anika [2], Stephen Quake[2], and Rob Phillips[1, 3, +]

[1] *Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA 91125, USA*
[3] *Department of Physics, California Institute of Technology, Pasadena, CA 91125, USA*
[+] *Correspondence: phillips@pboc.caltech.edu*

S1 Reporter Sequence design