**Understanding the toxicity effects of synthetic repressors by analysis of off-target genomic binding**

T.E. Gorochowski1, J. Zhang1 and C.A. Voigt1

1 Synthetic Biology Center, Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA.

**Keywords**: repressor; toxicity; off-target binding; gene regulation

**Abstract**

**Repressors have**

**Introduction**

* Repressors are a common regulatory component found naturally and used extensively in synthetic biology to control gene expression.
* As circuits and systems scale in size, issues arise due to the toxic effects often displayed after expression of large numbers of repressors.
* These toxicity effects are not well understood.
* We use RNA-Seq data combined with bioinformatics analysis of potential binding sites across the host genome to assess the potential for off-target binding and the functional groups of genes affected by such interference.
* Critical to make the point that binding of a repressor may not necessarily lead to a down regulation of the target gene. For example, NOR gates in synthetic biology rely on the ability to concatenate promoters. Only initiation is hindered, not the transcriptional process once it has started (only exception may be “road-blocking” repressors).

**Results**

1. Off target binding across the genome (use known experimental data on binding motifs).
   1. Characterization of binding profiles in regulatory and coding regions.
   2. Profiles of average binding and individual profiles for some genes.
   3. Statistical analysis into binding characteristics for different functional groups of genes, e.g., do essential genes have less binding sites on average to non-essential genes, are some functional groups (TFs, metabolism, etc) of genes particularly susceptible to binding by the repressors?
2. Combining binding profiles with RNA-Seq data to assess down-regulation of gene expression (subset of repressors that are included in the circuits data).
   1. Assess if there are any general trends for binding location and change in expression.
   2. Do we see additive effects (i.e., does expression of multiple repressors leads to sum of effects?)
3. Modeling of binding types and predicted impact on expression of gene and therefore toxicity profiles (Bryan’s list of difference responses by the host)
   1. Could we predict the response of a host based on the binding motif of the repressor?
4. Experimental validation (optional)
   1. Further experiments for other repressors with known binding motifs to test response.
   2. Chip-Seq type assessment of binding in genome.

**Discussion**

**Methods**

**Tables**

QacR – Sharp toxicity at transition point

LitR – Gradually toxic (non-specific binding)

TarA – Very toxic (non-specific, critical genes)

IcaRA – Sharp toxicity at transition point

SrpR – Non-toxic

AmtR – Toxic at high concentration

**Figures and captions**

**Figure 1: Overview of off-target repressor binding analysis.** Show the inputs and outputs from the analysis. It is important to highlight the experimental binding motifs. Don’t need to highlight actual scripts, but the steps in the process should be clear at a high level.

**Figure 2: Repressor binding profiles at genomic transcription start sites.** Would be nice to highlight at least 3 repressors (all would be better) with very different binding profiles across the genome. Show all sites and then zoom in on a typical one. Would also like to include averaged (moving window) type plot of the binding frequency upstream and downstream of TSS. Again, hope to see similarity in general shape, but differences in the average extent (show average +\- 75% interval).

**Figure 3: Effects of repressor binding location on gene expression.** This introduces the RNA-Seq data and tries to illustrate the region in the binding profile that actually affects expression level. This will need a more intuitive plot to show the shift in correlation between repressor expression and target gene expression for binding strengths at specific locations in the TSS region. Need to have a really clear way of showing this, all scatter plots should be placed in the SI. Idea would be to use the 3 repressors we have experimental binding data and RNA-Seq data for (AmtR, SrpR, LitR) and look for a subset of genes that see differential expression (be careful to consider additive effects, ignore to begin with). From these generate binding profiles to see if there are specific location enriched when changes in expression are observed.

**Figure 4: Toxicity effects linked to gene function of off-target binding sites.** Analysis of the GO functional categories and pathways that are impacted by strong off-target effects. If this is done for repressors with different toxicity effects can we see clustering of functional groups is present. If so then we have something really nice for predicting design constraints on using the repressors in synthetic circuits and may even be able to say something evolutionarily about essential genes and potential differences in their sequence composition around the TSS.

**Figure 5:** **Experimental validation (optional).** I am unsure if this is actually required of if we could merely extend to more circuit data that is available to see how well we can predict the toxicity of expression of various repressor combinations.

Conserved Sequences in Promoters

Cell Type Location Sequence

Prokaryotic -10

-35 TATAAT

TTGACA

Eukaryotic about -25

about -80 TATA (TATA box)

CAAT (CAAT box)

As DNA unwinds during bacterial transcription, the transcription apparatus occupies different sites in the gene during different steps of transcription.

Transcription Step Distance Covered Site

Initial complex 75 bp -55 to +20

Start of elongation 60 bp -35 to +25

During movement 30 bp 10 to +20 at start but moves as this size

**Introduction**

Repressors are pervasive DNA-binding proteins found throughout all forms of life. When binding near transcriptional initiation sites, they are capable of occluding RNA polymerase (RNAP) thereby repressing transcription of downstream genes. This ability to control gene expression has made them a common component of synthetic genetic circuits, but as these systems have grown in size, toxicity effects have been observed.

Off-target binding of repressors is not thought to generally hinder progression of RNAPs that have started transcribing, binding near transcription start sites (TSSs) on the genome can potentially influence the expression of endogenous genes. If sufficiently strong, this can have major impact on cellular physiology and impact viability. Such effects can be compounded due to the high expression levels often required for synthetic genetic circuits to robustly function and the need to simultaneously express multiple types of repressor to capture the many intermediate states within a circuit.

Here we provide a detailed analysis of 11 repressors for which binding motifs have been experimentally validated. We use this binding information to predict off-target influences that occur in the host genome and show how characteristics of this binding have a close relationship to the response of the cell to repressor over expression. From a mechanistic perspective this score is also linked to repressor binding affinity, therefore higher scores will relate to repressors that will more likely bind at a strength capable of influencing expression.

**Results**

**Predicting and classifying off-target genomic binding sites**

We developed a computational pipeline to predict the off-target binding profiles a repressor for a given genome. As input this takes a set of known repressor binding motifs and the host genome with annotated TSSs and gene names. From the binding motifs a position specific score matrix (PSSM) is calculated [?] that then allows for binding ‘scores’ to be calculated.

**Off-target repressor binding profiles**

We applied this workflow to a set of repressors that have been mined from multiple organisms and proposed as a

At any given binding site of a repressor, the probability of a binding event, and therefore expected occupancy, is governed by several key factors. Assuming a well-mixed and isotropic environment, binding probability will be directly proportional to the repressor concentration. This probability will also be modulated by other potential sites that will sequester repressors, therefore in the simplest case were all sites have equal affinity binding probability, *PB*, will be given by,

,

where *C* is the repressor concentration, and *NS* is the total number of binding sites. This ignores the intricacies of binding and potential ways in which repressors may act cooperatively (which will vary for each repressor type), but highlights that off-target binding can directly impact.

At the cellular-level a further aspect is critical in the effect observed on the cell – the essentiality of the gene. Impacts on non-essential genes will result in little if any impact on growth; however, reduced expression of critical metabolic genes will significantly hinder replication. Thus,

Essentiality, and number of genes (multiple knock-outs also likely to have a larger effect).

**Linking off-target binding characteristics to toxicity**

To do.