Bayesian method for filtering out mRNA turnover rate bias from siRNA knockdown measurements

Abstract

siRNA performance prediction calculations for a given siRNA may be divided into two broad categories: functions of the siRNA's sequence, hereafter referred to as "intrinsic" properties of the siRNA, and functions of the target mRNA, hereafter referred to as "extrinsic" properties of the siRNA. When training a statistical or machine learning model to select high potency siRNAs from a pool of candidates, we consider both intrinsic and extrinsic effects. However, some extrinsic effects bias the observed knockdown in a way that dilutes the ability of an siRNA selection algorithm to learn from the intrinsic effects of siRNAs of known performance. Therefore a means of filtering out this bias is desirable. Here we present one such bias filtering strategy; a Bayesian method that removes the extrinsic effect of target mRNA turnover rate, which is known to have a small, but non-negligible, correlation with knockdown performance. Application of the method leads to improved training sets for siRNA prediction algorithms considering the intrinsic properties of candidate siRNAs.

Introduction

Small interfering RNA (siRNA) molecules bind specifically to messenger RNA (mRNA) molecules, triggering the destruction of targeted mRNAs by promoting the degradation of the mRNAs by ribonucleases [1]. When the sequence of the siRNA is chosen to complement a target mRNA, a particular gene may be silenced by preventing translation of the mRNA. By observing effects on cells after silencing a gene using an siRNA, the function of the gene can be inferred.

Reynolds [2] observed that some siRNA sequences for a given target mRNA produce better gene knockdown than other siRNA sequences designed for the same target. This has led to the creation of design strategies for siRNA sequence selection. For example, Khvorova [3] reports that low stability of the 5' end of the antisense strand aids knockdown. Building on this work, Huesken [4] and Wang [5] developed machine-learning models to predict siRNA performance based on calculated features of a candidate siRNA design such as GC content and folding energy.

These features may be intrinsic to the siRNA, i.e., functions of the siRNA's sequence, or extrinsic to the siRNA, i.e., functions of the mRNA being targeted. We hereafter refer to these features as "intrinsic properties" and "extrinsic properties". Note that extrinsic properties of siRNAs may reflect differences between mRNAs targeted, as well as differences in the regions of a particular mRNA where, for example, secondary structure effects may be active.

An algorithm trying to learn the correlation between one or more intrinsic properties and siRNA performance may be confounded by the effects of extrinsic properties. Here we describe one such confounding effect found in the literature: mRNA turnover rate [6]. We then demonstrate a Bayesian method for removing the bias it generates from siRNA performance measurements, thereby producing estimates of the intrinsic performance of siRNA designs to which correlation and machine learning algorithms can be applied.

mRNA Turnover Rate

Larsson [6] reports that high mRNA turnover (decay) rate correlates with low siRNA knockdown potency. To facilitate this analysis, ThermoFisher Scientific provided Larsson and his team with RT-qPCR knockdown measurements for 2622 Silencer siRNAs (Supplementary Table 1). These results were then compared, by gene, to mRNA turnover rates which they measured by examining a microarray time-series of transcriptionally inhibited cells. The proceeding analysis uses both data sets.

Method

Stochastic Model

We begin by relating measured knockdown K with intrinsic potency P and extrinsic effects f(T) for each siRNA i as follows:

$$K_i = P_i - f(T_i)$$

T_i in the function is the measured turnover rate of the mRNA targeted by siRNA i.

To promote simplicity, we then model f(T) as second-order polynomial:

$$K_i = P_i - \alpha \cdot T_i - \beta \cdot T_i^2$$

Finally, we treat measured knockdown as a random variable:

$$K_i \sim Normal(P_i - \alpha \cdot T_i - \beta \cdot T_i^2)$$

We now have a probabilistic model relating measured knockdown to both intrinsic and extrinsic siRNA potency, where the extrinsic potency is the potency affected by mRNA turnover rate.

Bayes' Theorem and Gibbs Sampling

Bayes' theorem provides the means for relating the measured data to the model parameters:

$$P(P_i, \alpha, \beta | K_i, T_i) = \frac{P(K_i, T_i | P_i, \alpha, \beta) \cdot P(P_i, \alpha, \beta)}{P(K_i, T_i)}$$

We maximize the likelihood function using Gibbs sampling to find the parameters that best explain the data, and then combine that result with the priors to compute the maximum *a posteriori* (MAP)

estimation of the model parameters. In this way we obtain estimates of siRNAs' intrinsic potencies. Gibbs sampling is conducted using OpenBUGS software [7].

We set the priors for alpha and beta to Normal(0, 1) to reflect our complete absence of knowledge about their distributions. We also set the priors for Pi for all siRNAs i in the validation set to Normal(0, 1). However, these last priors are "semi-informed"; for each Pi such a distribution implies 100% intrinsic knockdown. During Gibbs sampling, such priors for Pi force the computations to converge such that Pi is generally greater than Ki, which fits with our biological intuition.

We describe the stochastic model and the data in the OpenBUGS language (see the Supplementary Code) and simulate using a "burn in" period of 2000 iterations and a total analysis of 12000 iterations. Figure 1 shows convergence of the simulation onto constant values for intrinsic potencies for siRNAs one through four in the data set during program execution.
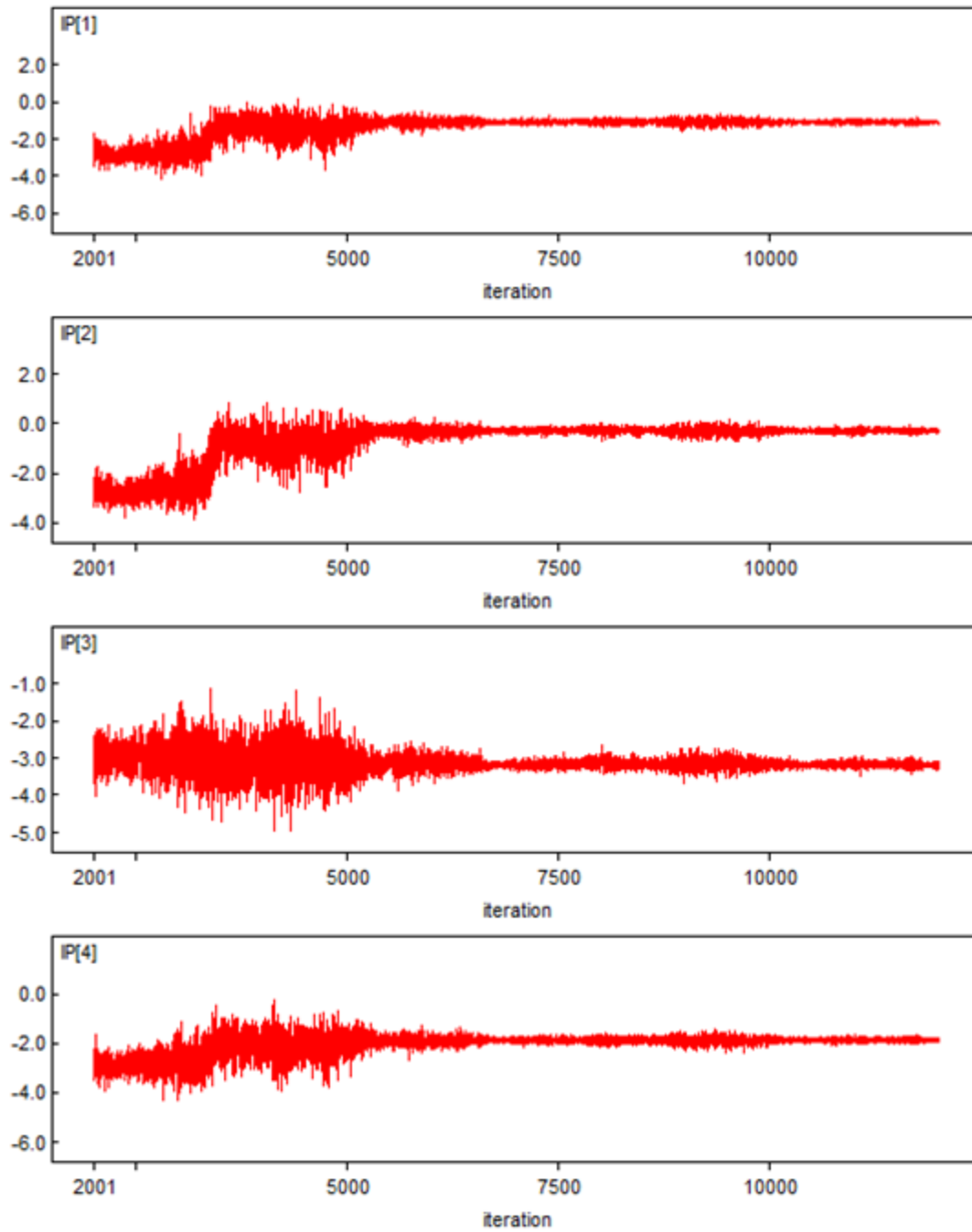
Figure 1: Convergence of Estimate of Intrinsic siRNA Potencies

The modes for each of the predicted intrinsic potencies stand out clearly in Figure 2 for estimates one through ten.
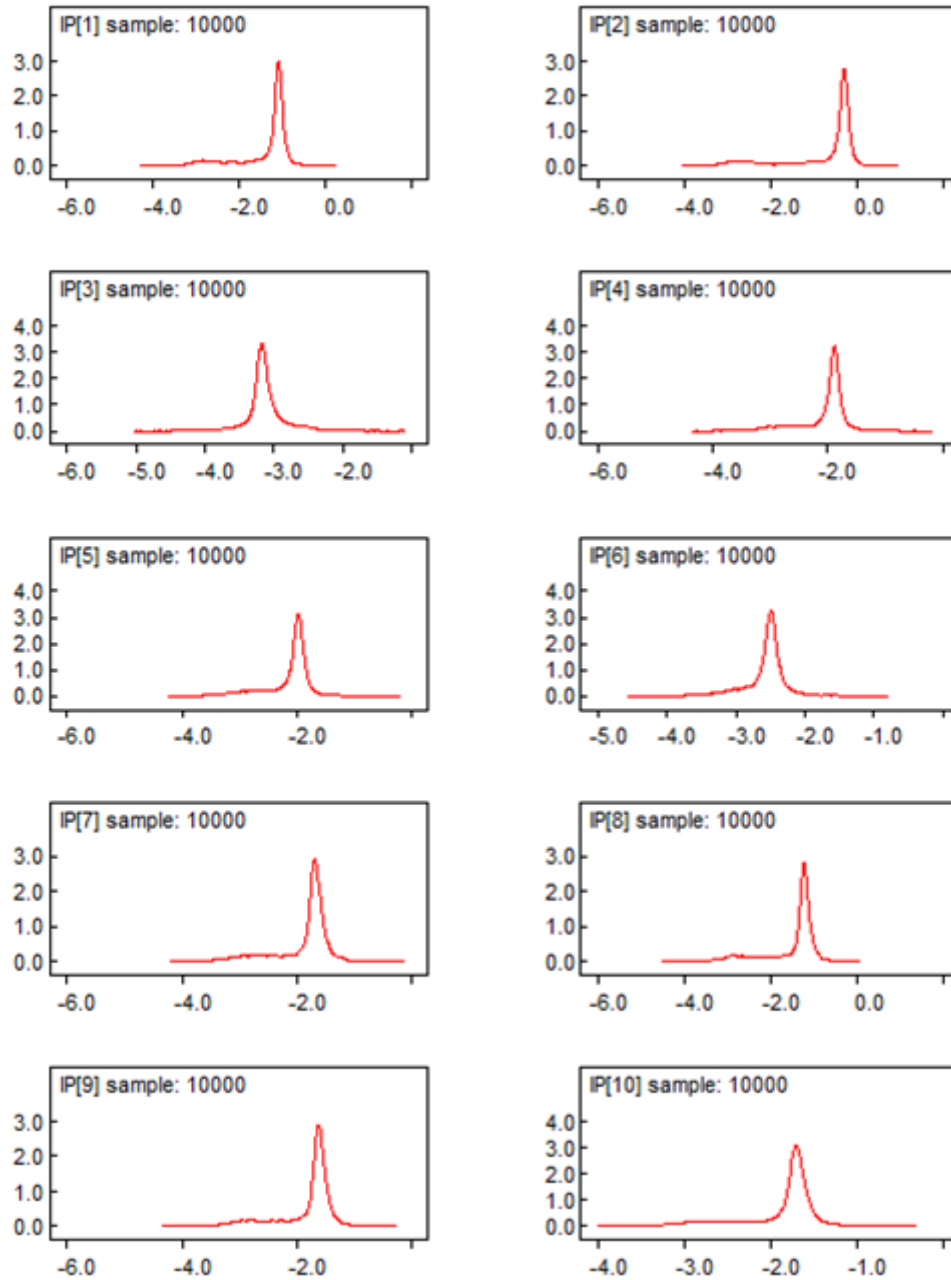
Figure 2: Density of Estimates of Intrinsic siRNA Potencies

The values of alpha and beta converge less quickly (Figure 3), but the simulation produces a clear mode for each variable, as shown in Figure 4.
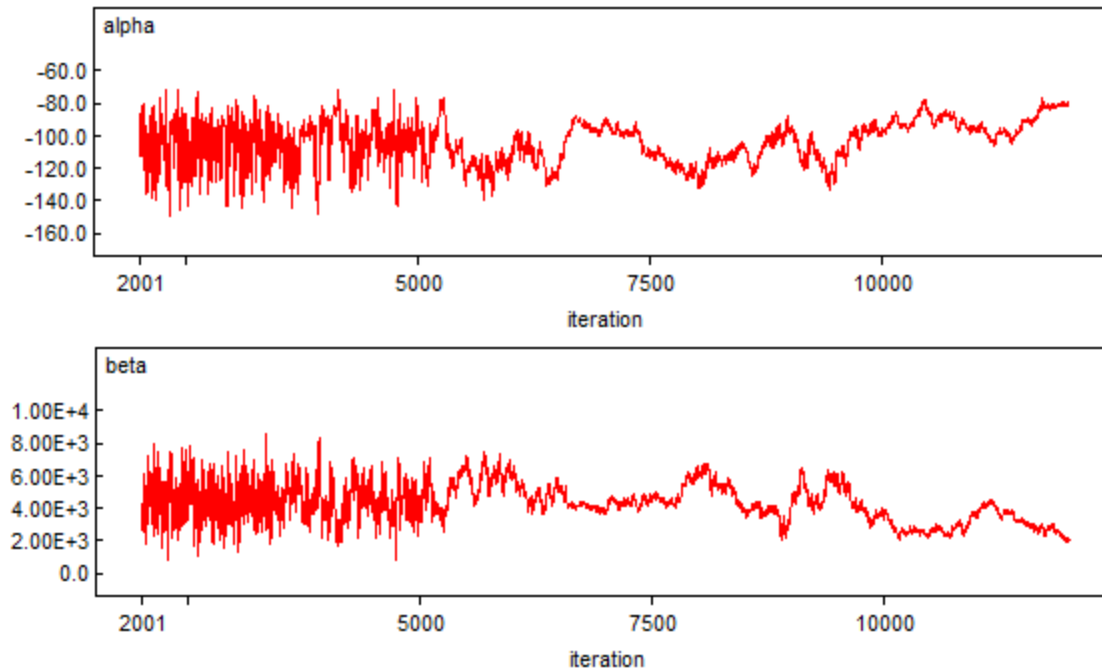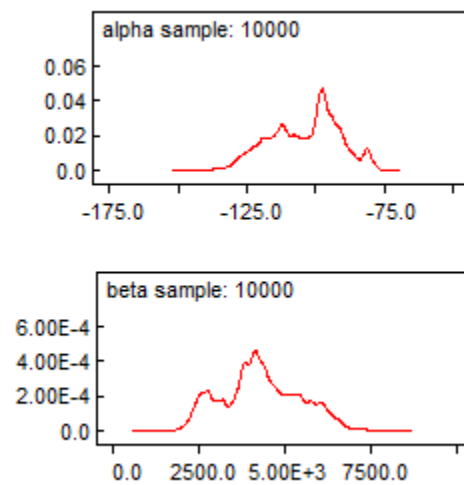
Figure 3: Convergence of Alpha and Beta

Figure 4: Density of Estimates for Alpha and Beta

Results

Unlike the measured knockdown for each siRNA, the intrinsic siRNA potencies computed by the Gibbs sampling described above do not correlate with mRNA decay rate. This is demonstrated in Figure 5.
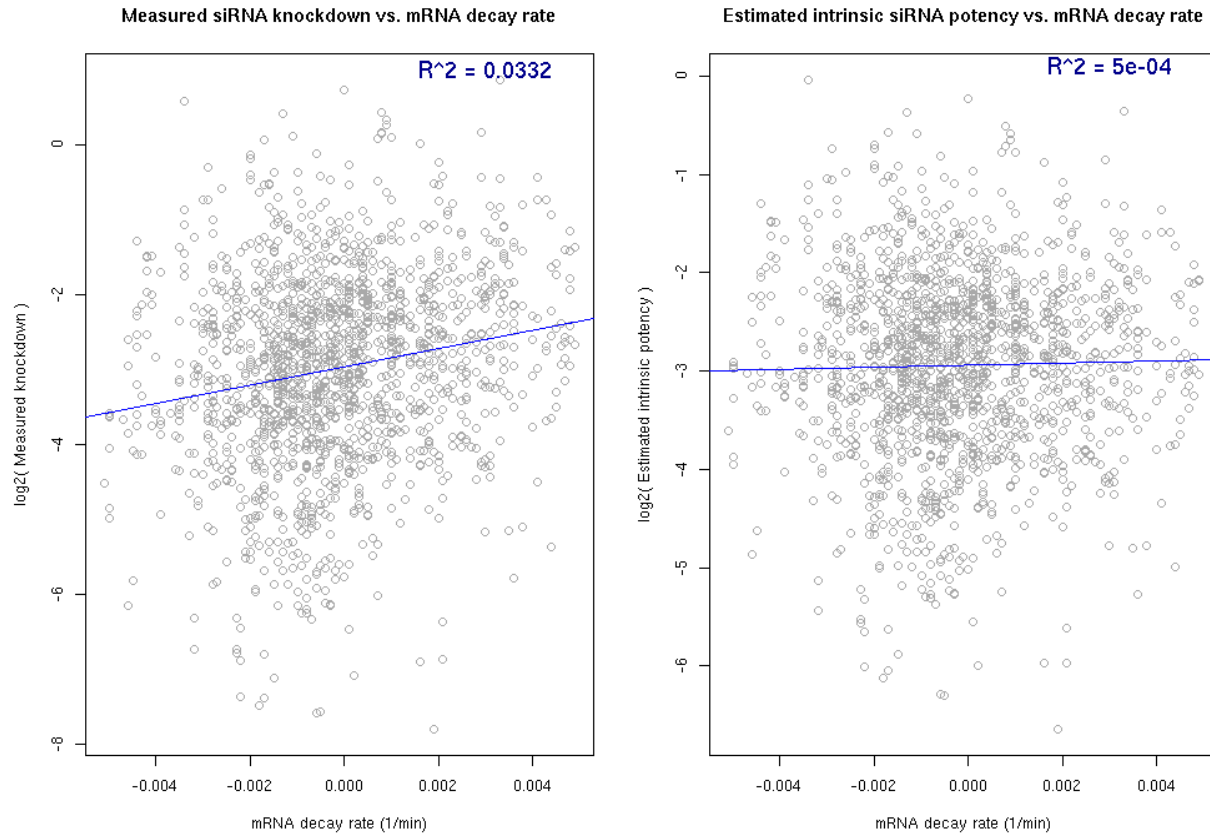
Figure 5: Measured and Intrinsic Potency vs. mRNA Decay Rate

On the left side of Figure 5 we see the measured knockdown plotted against mRNA turnover rate, per gene. A positive trend line shows that a non-negligible correlation exists between the two variables. On the right side of Figure 5 we see the estimated intrinsic potency for each siRNA plotted against the mRNA turnover rate. In this case no correlation exists; the confounding impact of mRNA decay rate has been removed.

Figure 6 shows the measured knockdown plotted against the estimated intrinsic potency by siRNA, showing that a high correlation still exists between the measured and estimated results.
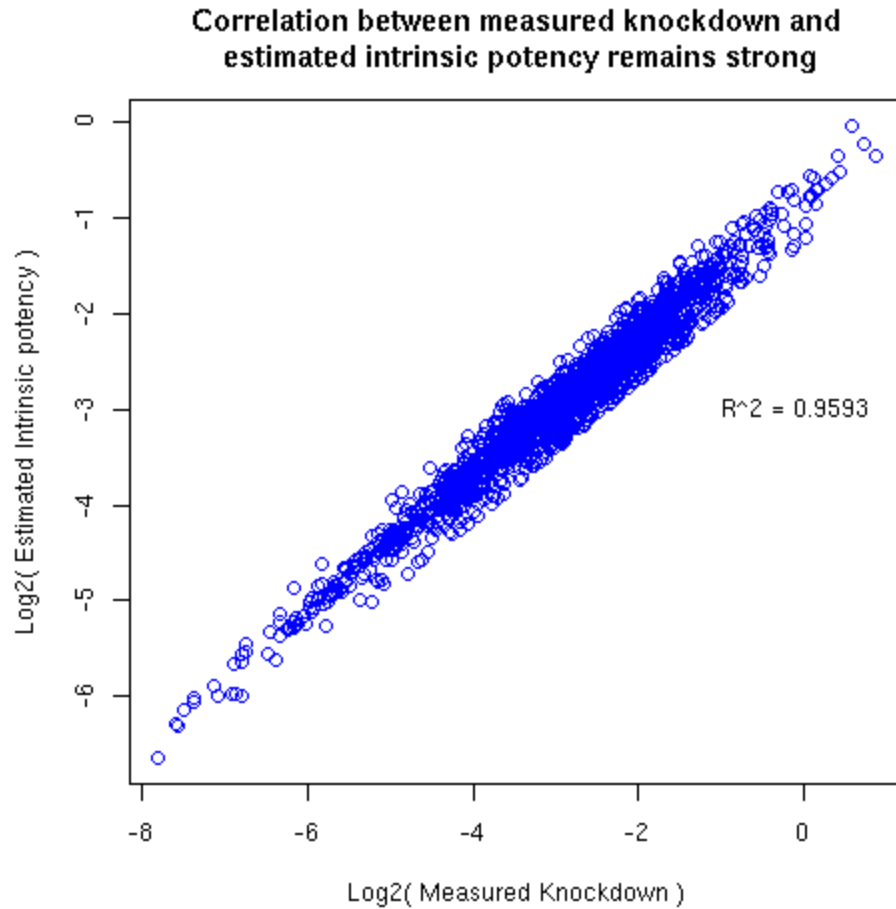
Correlation between measured knockdown and estimated intrinsic potency remains strong

$R^2 = 0.9593$

X-axis: Log2(Measured Knockdown)
Y-axis: Log2(Estimated Intrinsic potency)

Figure 6: High Correlation Remains Between Measured and Estimated siRNA Potencies

Discussion

The analysis detailed above demonstrates a method for removing the bias caused by mRNA decay rate from siRNA knockdown measurements. It can likely be extended to other measurable (or predictable) extrinsic traits such as mRNA secondary structure surrounding the siRNA binding site.

Using this filtering method provides a clearer dataset for algorithms attempting to establish the correlation between intrinsic siRNA features (such as GC content or predicted secondary structure) with siRNA performance.

In Equation 2, we originally left out the squared term. This significantly reduced this method's ability to remove the bias. From that observation, we conclude that the extreme values of mRNA turnover rate impact overall siRNA performance disproportionally when compared to average decay rates.

References

1. RNA Interference: Biology, Mechanism, and Applications. Neema Agrawal, P. V. N. Dasaradhi, Asif Mohmmed, Pawan Malhotra, Raj K. Bhatnagar, and Sunil K. Mukherjee
2. Rational siRNA design for RNA interference. Angela Reynolds, Devin Leake, Queta Boese, Stephen Scaringe, William S Marshall & Anastasia Khvorova
3. Functional siRNAs and miRNAs Exhibit Strand Bias. Anastasia Khvorova, Angela Reynolds, Sumedha D. Jayasena
4. Design of a genome-wide siRNA library using an artificial neural network. Huesken D, Lange J, Mickanin C, Weiler J, Asselbergs F, Warner J, Meloon B, Engel S, Rosenberg A, Cohen D, Labow M, Reinhardt M, Natt F, Hall J.
5. Selection of hyperfunctional siRNAs with improved potency and specificity. Wang X, Wang X, Varma RK, Beauchamp L, Magdaleno S, Sendera TJ.
6. mRNA turnover rate limits siRNA and microRNA efficacy. Erik Larsson, Chris Sander, Debora Marks
7. OpenBUGS (find reference)