A Draft so rough that you can't put a differentiable structure on it...

1 Abstract

Idea: We build a mathematical model and associated simulation which gives evidence to suggest that miRNAs have a greater effect in developing organism than in mature ones, and furthermore, shows that miRNAs aid in tumor suppression

2 Introduction

Mirco RNAs (miRNAs) are nonoding RNA molecules which are noncoding that are believed to provide translational regulation of messenger RNAs (mRNAs). There is, however, some dispute over the effects and mechanism by which miRNAs affect mRNAs, see [7]. The study of miRNAs is further complicated by the fact that there exists no consensus on when miRNAs are most active. Evidence exists to suggest that miRNA expression guides organism development [2]. In addition, miRNAs have been associated with tumor supression, see [6] and the references therein. For example, in [14] the authors found connections between the expression specific miRNAs and T cell llymphoblastic leukemia in mice. Additionally, in [5] the authors showed that miR-3151, when expressed in cultured cancer cells, reduced apoptosis.

Say more about the proposed mechanism by which mRNAs work 3' UTR binding, cleavage etc. Possibly say more about the p53 let7 etc

With the continued increase in computing power there is interest in mathematical models which describe what are called gene regulatory networks (GRNs), see [9]. Given the cost and length of time required to study miRNAs in a laboratory setting mathematical simulation has increasingly been seen as a way to potentially describe and give motivation for various hypotheses [15].

In an effort to build a mathematical model of GRNs researchers have been forced to find a compromise between models which attempt to describe the biology very accurately but are inherently quite complex and those which simplify their approach in an effort to find broader themes which may remain true in a variety of different contexts. A variety of approaches based on the use of chemical master equations can be found in [8, 13].

The typical approach, see [15], is to use graphs consisting of vertices or nodes and edges. One then associates to each node a function, and each function produces an output in the set $\{0, 1\}$. One then has to consider what class of functions are appropriate. Give the desire to have an output in the set $\{0, 1\}$ it is natural to consider truth tables, as in [1]. However the size of the set of functions is prohibitively large for numerical simulation and too general for mathematical analysis. For this reason many authors choose instead to work with threshold functions, see [4, 3], and [11].

This study builds off of the work of [10] which analyzes random Boolean networks miRNA implemented as feed-forward loops, to examine the relationship between the presence of miRNAs and the stability of the network. Furthermore, the authors consider Boolean networks under the influence of intrinsic cellular noise. This work shows the trend that an increase in miRNA produces, on average, a more stable system under a particular metric. The metric, called Average Maximal Transition Probability measures the probability that a given initial configuration of the network will eventually transition to an attractor.

Where previous work has represented genes, protein, and RNA with identical nodes in a directed graph which updates according to various rules. We have shifted to a paradigm that is based on two operators. Transcription, denoted $\mathcal{C}$, and translation, denoted $\mathcal{L}$. This means that our model is less of a simplification of the biology then the existing directed graph models, but is not so complex that we are limited to simulation as our only means of understanding the model and its implications. In what follows we give both analytic and simulation based results of how our model suggests that miRNAs increase stability in gene regulatory networks. In particular, we show that our system always has an absorbing fixed point corresponding to no transcriptional or translational activity, and that miRNAs increase the rate at which the system transitions to this state.

Include a diagram of the transition from transcription to translation with the effect of miRNAs according to the model. Also give an overview of the model possibly in the case of an example GRN

3 The Transcription Operator, $\mathcal{C}$

The transcription operator, denoted $\mathcal{C}$, is effectively a threshold function applied to the vector of "transcription factors" (dim $= nMess$), and maps to the space of "genes" (dim $= nRNA$). Though this differs from the update function found in [10], that original function is contained in our set of candidate functions for $\mathcal{C}$. Hence, the analysis below is applicable to functions of the previous type as well as of this new form.

3.1 Previous form

The previous form was defined in the following way. Let $B = (b_{ij})$ be an $nRNA \times nRNA$ matrix taking values in $\{-1, 0, 1\}$ with $b_{ii} = 0$ for all i. If we let B_i be the ith row of the matrix B, then the transcription operator is as follows:

$$x_{t+1}^i = \begin{cases} 0 & \text{if } (B_i \cdot x_t) < 0 \\ 1 & \text{if } (B_i \cdot x_t) > 0 \\ x_t^i & \text{if } (B_i \cdot x_t) = 0 \end{cases}$$

This function was extended to have a stochastic delay as well. The stochastic delay version, here called $\widetilde{\mathcal{C}}$, had this form:

$$x_{t+1}^i = \begin{cases} (B_i \cdot x_{t+1}) & \text{if } X_i = 1 \\ x_t^i & \text{if } X_i = 0 \end{cases}$$

where $X = (X_1, \ldots, X_{nRNA})$ are a set of i.i.d Bernoulli random variables. Our current version extends $\mathcal{C}$, but includes stochastic noise rather than stochastic delay.

3.2 Current form

The fundamental form of $\mathcal{C}$ remains essentially the same, except that it is a mapping from $\{0,1\}^{nMess} \to \{0,1\}^{nRNA}$. That is, it represents the effect of the transcription factors (of which there are $nMess$) on the transcribed RNAs (of which there are $nMess + nMicro = nRNA$). To build our network, we currently specify four parameters: $\deg_l$, $\deg_h$, p_{nz} and pmr. These correspond to the lowest possible indegree for each

gene, the highest possible indegree for each gene, the probability of an element to be the nonzero given that there are at least $\deg_l$ nonzero elements , and the ratio of $+1$'s to -1' s in the matrix B. In short, this means that any row has a total number of nonzero elements distributed as $\deg_l + Bin\left(deg_h - deg_l, p_{nz}\right)$ which has $+1$ and -1 values in the ratio pmr. This rule is fairly straightforward and is identical to the form used in [10] where $deg_l = 0,\ \deg_h = 9,\ pmr = 1$. We also add stochastic noise to our system using two additional parameters, p_0 and p_1, which are the probability for a node to ignore the update rule and take the values 0 and 1 respectively. Clearly this means that with probability $1 - (p_0 + p_1)$ a gene will be updated according to the update rule of $\widetilde{\mathcal{C}}$. Our expression for the operation of $\widetilde{\mathcal{C}}$ is then:

$$x_{t+1}^i = \begin{cases} 0 & \text{if } X_t^i = 0 \\ 1 & \text{if } X_t^i = 1 \\ B_i \cdot x_t^i & \text{if } X_t^i = 2 \end{cases}$$

where $X_t = \left(X_t^1, \ldots, X_t^{nRNA}\right)$ are i.i.d random variables which take on the values $\{0, 1, 2\}$ with probabilities $\{p_0, p_1, 1 - (p_0 + p_1)\}$ respectively. As is clear from the notation, X_t is updated at each time step. With a bit more work, the Markov transition matrix for $\widetilde{C}$ can still be determined as well.

There is a second deviation from the previous form as well. We do not impose the restriction that $B_{ii} \equiv 0$, and reduce the number of cases to:

$$x_{t+1}^i = \begin{cases} 0 & (B_i \cdot x_t) \leq 0 \\ 1 & (B_i \cdot x_t) > 0 \end{cases}$$

This deviation is not actually very significant. Within the set of admissible matrices B are the set of matrices for which $B_{ii} \equiv 1$.

3.3 Stochastic part

This section might not quite make sense as the values of S and D vary from one time step to the next so the decomposition is time dependent

Denote by C_t the transition matrix of the operator $\widetilde{\mathcal{C}}$. With stochastic noise added, it makes sense to break C_t into two parts: S_t and D_t. We represent the deterministic part of the operator with D_t and S_t represents the stochastic part. Naturally, $C_t = (p_0 + p_1)\, S_t + (1 - p_0 - p_1)\, D_t$. I think the line below is incorrect. The size of S depends on the values of X, for instance, when we change the location of the value 2 this changes the S matrix and in those rows we should have 0. Similarly matrix D will be determined by the values in X. The stochastic portion, S_t is merely a rank-1 $nRNA \times nMess$ matrix of the form (where $n \equiv nRNA$):

$$S_t = \begin{bmatrix} p_0^n & p_0^n & \cdots & p_0^n \\ p_0^{n-1}p_1 & p_0^{n-1}p_1 & \cdots & p_0^{n-1}p_1 \\ \vdots & & & \vdots \\ \\ p_1^n & p_1^n & \cdots & p_1^n \end{bmatrix} \tag{1}$$

It is particularly uninteresting, but later on we may want to change its form (e.g. with a stencil).

3.4 Deterministic part

3.4.1 By recursion

Note that one initial simplification of the D_t comes from the fact that all input vector with n non-zero elements are permutations of one another, so we can define an the same equivalence relation on them as was done for $\mathcal{L}$. The rows of D_t are indistinguishable in the sense that they are all determined in the same way by deg_l, deg_h, p_{nz}, and pmr. For this reason we will consider just the first row and calculate the probability that the first row of $D_t(x) \leq 0$. These calculations will then apply to the other rows. Write $d_t = \left(d_t^0, d_t^1 \ldots, d_t^{mMess-1}\right)$ for the values of this row.

It's true that in what follows it's easier to number from 0 to $nMess - 1$, we should consider number mRNAs from 0 for the rest of the paper

Only for the special case of $deg_l = 0$, $deg_h = nMess$ are the values d_t^i i.i.d. Consequently, we proceed inductively. The following notation will be used: given a vector x with n non-zero values, let P_n^k indicate the probability that $d_t \cdot x_t = k$, $|k| \leq n$ where $x_t = (x_t^0, \ldots, x_t^{n-1})$. For simplicity of the notation we suppress the dependence of P_n^k on t and B, but the reader should be aware that we could write $P_n^k = P_n^k(t, B)$. This probability can then be expressed recursively. We will consider the probabilities conditioned on the fact that $x_{n-1} = 1$. Now we have

$$P_n^k = P_{n-1}^k \cdot \mathbb{P}\left(d_t^n = 0 \middle| \sum_{j=0}^{n-1} d_t^j x_t^j = k\right) + P_{n-1}^{k+1} \cdot \mathbb{P}\left(d_t^n = -1 \middle| \sum_{j=0}^{n-1} d_t^j x_t^j = k+1\right) + P_{n-1}^{k-1} \cdot \mathbb{P}\left(d_t^n = +1 \middle| \sum_{j=0}^{n-1} d_t^j x_t^j = k-1\right) \tag{2}$$

To simplify the exposition we add some notation. We write $z_t(n,k) = \mathbb{P}\left(d_t^n = 0 \middle| \sum_{j=0}^{n-1} d_t^j x_t^j = k\right)$. Note that the reflection principal proves that $z_t(n,k) = z_t(n,-k)$. For convenience write $q = \mathbb{P}\left(d_t^j = +1 \middle| d_t^j \neq 0\right) = \frac{pmr}{1+pmr}$. The formula above becomes:

$$P_n^k = P_{n-1}^k z_t(n-1,k) + P_{n-1}^{k+1} \bar{z}_t(n-1,k+1)\,\bar{q} + P_{n-1}^{k-1} \bar{z}_t(n-1,k-1)\,q \tag{3}$$

where we reuse out previous convention that $\bar{z} = 1 - z$. The boundary conditions which we have are naturally:

$$P_0^0 = 1,\ P_m^k = 0 \text{ for } k, m \text{ where } |k| > m \tag{4}$$

To give the reader a feel for the kind of calculations that we're doing we present a simplified case of the problem of calculating P_n^k in the proposition below.

Proposition 3.1. *Assume $deg_l = 0$ and $deg_h = nMess$, the we have a simplified formula for P_n^k*

$$P_n^k = \frac{\sum_{r \leq n-k} M_{p,q,r}}{\sum_{k \leq n} \sum_{r \leq n-k} M_{p,q,r}} \tag{5}$$

where

$$M_{p,q,r} = \binom{p+q+r-1}{r}\left(\frac{p-q}{p+q}\right)\binom{p+q}{p} \tag{6}$$

The goal is to calculate P_n^k for $k \leq n$ where $(n,k) \in \mathbb{Z}^2$ for some fixed time t. Write $d_t = (d_t^0, \ldots, d_t^{n-1}, d_t^i) \in \{0, \pm 1\}$, and $P_n^k = \mathbb{P}(d_t \cdot x_t = k)$. The assumptions about deg_l and deg_h imply that the d_t^i are i.i.d. Partition the d_t^i according to their values and assign $p = \{\#d_t^i | d_t^i = 1\}$, $q = \{\#d_t^i | d_t^i = -1\}$, and $r = \{\#d_t^i | d_t^i = 0\}$. These all depend on t but because we only consider fixed t we omit this in our notation. It is then the case that $k = p - q$, and $n = p + q + r$. Let $s_t^j = \sum_{i=1}^{j} d_t^i$. Then $s_t^j - s_t^{j-1} = d_t^j$, and $s_t^n = k$. The sequence $(s_t^1, \ldots s_t^n)$ is called a "path". We now consider paths from $(0,0) \in \mathbb{Z}^2$ to (n,k) where $s_t^i \geq 0$. These will correspond to all of the ways in which we can have a value of P_n^k which is nonzero.

Let

$$N_{p,q,r} = N_{n,k} = \binom{p+q+r}{r}\binom{p+q}{p} \tag{7}$$

$$= \binom{p+q+r}{r}\binom{p+q}{q}. \tag{8}$$

The function $N_{n,k}$ counts the number of paths from the point $(0,0)$ to (n,k) when there are r pauses or instances of $d_t^i = 0$. Again, $N_{n,k}$ depends on t but we suppress this portion of the notation. We wish to count just the paths for which $s_t^1 > 0, \ldots, s_t^n > 0$. There are as many allowable paths as there are paths from the point $(1,1)$ to (n,k) which neither touch nor cross the x-axis. Using the reflection principal this is equal to

$$N_{n-1,k-1} - N_{n-1,x+1} = \binom{p+q+r}{r}\binom{p+q-1}{p-q} - \binom{p+q+r-1}{r}\binom{p+q-1}{p} \tag{9}$$

$$= \binom{p+q+r-1}{r}\left(\frac{p-q}{p+q}\right)\binom{p+q}{p} \tag{10}$$

Denote the last line by $M_{p,q,r}$. We may then write

$$P_n^k = \frac{\sum_{r \leq n-k} M_{p,q,r}}{\sum_{k \leq n}\sum_{r \leq n-k} M_p, q, r} \tag{11}$$

We sum over $0 \leq r \leq n-k$ as this is the range of possible pauses that the sequence may have, and $k \leq n$ as this is the range of possible values that $d_t \cdot x_t$ can take.

3.4.2 By direct calculation

We consider the calculation of P_n^k with general choices of our parameters for fixed t. Without loss of generality let $x_t^i = 1,\ i = 1, \ldots, n_1,\ x_t^j = 0,\ j = n+1, \ldots, n$. We wish to calculate $d_t \cdot x_t$. Given d having n_2 nonzero elements, we have

$$\mathbb{P}\left(k \text{ nonzero values in first } n_1 \text{indices} \middle| n_2 \text{ indicies with nonzero values}\right) = \frac{\binom{n_1}{k}\binom{n-n_1}{n_2-k}}{\binom{n}{n_2}} \tag{12}$$

We also have

$$\mathbb{P}\left(\#\left(+1\right)=j,\#\left(-1\right)=k-j\right)=\binom{k}{j}q^{j}\bar{q}^{k-j} \tag{13}$$

where $q=\frac{pmr}{1+pmr},\ \bar{q}=1-q$. Hence, given k nonzero elements in some dx , we have

$$\mathbb{P}_{dx}^{j}=\binom{k}{\frac{k+j}{2}}q^{\frac{k+j}{2}}\bar{q}^{\frac{k-j}{2}} \tag{14}$$

with the constraint that $k+j$ is even. Finally, we have

$$\mathbb{P}\left(n_2\text{ indicies with nonzero values in }d\right)=\binom{\deg_h-\deg_l}{n_2-\deg_l}p_q^{n_2-\deg_l}\bar{p}_{\bar{q}}^{\deg_h-n_2} \tag{15}$$

The composition of all of these components can be done, but recall that we only want to know $\sum_{k=1}^{n}P_n^k$ and $\sum_{k=0}^{n}P_n^{-k}$ (as those are the probabilities that $\widetilde{\mathcal{C}}x=1,\ 0$, respectively).

4 The Translation operator, $\mathcal{L}$

In our model, translation depends only on the state of the RNAs. Naturally, mRNA must be present in order to be translated, while the presence of one or more miRNAs has the opportunity to inhibit that translation event. The miRNAs may inhibit any combination of different mRNAs during a single operation. As a result, the presence of a particular mRNA does not guarantee its translation. However, the absence of an mRNA should guarantee the absence of protein (increases). Hence, the operator $\mathcal{L}$ can be considered as follows.

Translation maps the set of all RNAs, denoted as $nRNA=nMess{+}nMicro$ (where $nMess$ is the number of mRNAs and $nMicro$ is the number of miRNAs), to the set of all protein products, which necessarily has the dimension $nMess$. Furthermore, the presence of a given mRNA does not guarantee its translation, but absence of the mRNA implies absence of protein product. The possibility of an mRNA going untranslated increases as more miRNAs which can target a given mRNA are added to the system. We consider the translation of a lone mRNA and the independent effects of each miRNA as a Bernoulli trial.

Let ρ_i denote the probability of mRNA i (denoted R_i) being "seen" by the translational apparatus - i.e. the ribosome. Let $\zeta_{i;j}$ denote the "ability" of miRNA j (denoted r_j) to inhibit that translation. Finally, denote

$$x_t = \begin{bmatrix} R_1 \\ R_2 \\ \vdots \\ R_{nMess} \\ \hline r_1 \\ r_2 \\ \vdots \\ r_{nMicro} \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n_1} \\ \hline x_{n_1+1} \\ x_{n_1+2} \\ \vdots \\ x_{n_1+n_2} \end{bmatrix} \tag{16}$$

as the disjoint union of the messenger and micro RNAs. Then $\mathcal{L} : \{0,1\}^{nRNA} \to \{0,1\}^{nMess}$ has the following distribution:

$$\mathbb{P}(x_{t+1}^i = 1 \,|\, x_t^i) = \rho_i x_t^i \prod_{j=nMess+1}^{nRNA} \left(1 - \zeta_{i;j} x_t^j\right) \tag{17}$$

This naturally indicates the requirement of the presence of R_i, a "successful trial" on R_i, and a combined set of "failures" for all miRNAs which may affect translation of R_i. This operator is not only sensible but also quite tractable. We will write out the complete Markov transition matrix for this operator as well as its spectrum.

4.1 Markov transition matrix of $\mathcal{L}$

We order the $\{0,1\}^{nMess}$-valued vectors first by the number of non-zero values it contains and subsequently in reverse lexicographic order. For $\{0,1\}^3$, that is the ordering $\{000, 100, 010, 001, 110, 101, 011, 111\}$. We also denote the quantity above, $\mathbb{P}\left(x_{t+1}^i = 1 \,|\, x_t^i\right)$ by q_i, and $\mathbb{P}(x_{t+1}^i = 0 \,|\, x_t^i) = 1 - q_i = \bar{q}_i$. Below we give the transition matrix L, for this 3 dimensional case to demonstrate the pattern in the operator $\mathcal{L}$:

$$L|_{\{0,1\}^{nMess}} = \begin{bmatrix} 1 & \bar{q}_1 & \bar{q}_2 & \bar{q}_3 & \bar{q}_1\bar{q}_2 & \bar{q}_1\bar{q}_3 & \bar{q}_2\bar{q}_3 & \bar{q}_1\bar{q}_2\bar{q}_3 \\ 0 & q_1 & 0 & 0 & q_1\bar{q}_2 & q_1\bar{q}_3 & 0 & q_1\bar{q}_2\bar{q}_3 \\ \vdots & \ddots & q_2 & 0 & \bar{q}_1 q_2 & 0 & q_2\bar{q}_3 & \bar{q}_1 q_2\bar{q}_3 \\ \vdots & & \ddots & q_3 & 0 & \bar{q}_1 q_3 & \bar{q}_2 q_3 & \bar{q}_1\bar{q}_2 q_3 \\ \vdots & & & \ddots & q_1 q_2 & 0 & 0 & q_1 q_2\bar{q}_3 \\ \vdots & & & & \ddots & q_1 q_3 & 0 & q_1\bar{q}_2 q_3 \\ \vdots & & & & & \ddots & q_2 q_3 & \bar{q}_1 q_2 q_3 \\ 0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & q_1 q_2 q_3 \end{bmatrix} \tag{18}$$

Through this instructive example, we can see a few key points - first, that L is upper-triangular suggests that miRNAs are "rank-reducing" - this is unsurprising, as miRNAs serve as translation inhibitors. While some miRNAs may have "enhancing" effects, this is really a two-step process; the miRNA must inhibit an inhibitor.

Not surprisingly, then, the stationary vector π such that $L\pi = \pi$ is the vector $\pi = [1,0,0,0,0,0,0,0] = e_1$. Let A_i denote the set of non-zero indices of state i where $0 \leq i \leq 7$, and corresponds to the ordering given above. Furthermore, we denote the probability to transition from state i to state j is given by $[l_{ij}]$. Then

$$L = [l_{ij}] = \begin{cases} \prod_{l \in A_i \cap A_j} q_l \prod_{k \in A_i \setminus A_j} \bar{q}_k & A_j \subset A_i \\ 0 & otherwise \end{cases} \tag{19}$$

and the spectrum of $\mathcal{L}$, $\sigma(\mathcal{L}) = \sigma_1 \cup \{1\} \setminus \{0\}$, with $\sigma_1 = \bigcup_{V \in 2^S} \prod_{i \in V} q_i$ where $S = \{1, 2, ...nMess\}$. In simpler terms, it is the collection of products of elements in the power set of $\{q_1, q_2 ... q_{nMess}\}$. The eigenvectors have a similar simple description with $v_i = \sum_{A_j \subset A_i} (-1)^{|A_j|} e_j$.

4.1.1 Dimensions of L

Russ did you type what you meant below? The above matrix is $2^{nMess} \times 2^{nMess}$. What is the non-square form? Either way we don't need the paragraph below do we?

It is important to remember that L is generally expressed as a $2^{nMess} \times 2^{nRNA}$ matrix. In the case above, the values q have the elements $x_{nMess+j} = r_j$ embedded in their value. It is a very straightforward exercise to extend this to the full (non-square) form; columns with various miRNAs absent ($= 0$) simply are replicas of their corresponding elements above.

4.2 Simplification of $\mathcal{L}$ - distinguishable elements

There is a single unattractive component of L under our current set of constraints - our mRNAs are currently indistinguishable: that is, R_i is in no way "more important" than R_j. No past study that has been conducted by this group has *ever* considered any precedence of mRNAs (or miRNAs, for that matter), and hence, we can lower the complexity of the system. Since all of the analyses & statistics that have been considered have been entirely invariant under permutations of our mRNAs, we can define an equivalence relation amongst all states which have the same number of non-zero elements; e.g. in the 3-dimensional example above, $e_2 \equiv e_3 \equiv e_4$

The equivalence relation needs to be spelled out a bit more

and $e_5 \equiv e_6 \equiv e_7$. Our eigenvalues reduce then to the average of the equivalent eigenvalues, or

$$\bigcup_{k=1}^{nMess} \frac{\sum_{A_j \in \binom{S}{k}} \prod_{i \in A_j} q_i}{\binom{nMess}{k}}$$

while eigenvectors reduce to $v_k = \sum_{l=1}^{k} (-1)^l e_l$. This equivalence is particularly useful because our $nMess$-element set of "translated" transcription factors also has this same equivalence relation. Since we have defined $\mathcal{L}$ as a reducing operator (it can only turn 1's into 0's, not 0's into $1's$), we can neatly express $\mathcal{L}$ in either form. However, the transcription operator, $\mathcal{C}$, which is not expressed as a reducing operator, is slightly less simple (though not all that bad). Also, each of the elements q depends on the presence of each miRNA r_j, which is not static. Although the space of "phenotypes" is of dimension $nMess$, we need to consider the independent effects of $\mathcal{L} \circ \mathcal{C}$ on both the sets R and r, namely mRNA and miRNA in order to fully elucidate

the function of miRNAs on this system.

4.2.1 Complete transition matrix under equivalence Needs to be edited, partially edited version below

The matrix L derived above is correct provided we have distinctly identified the states of each miRNA. However, we want to determine the entire transition matrix under *all* possible network configurations. Hence, we now consider the expansion of the elements $q_i, \bar{q}_i$ above under the equivalence specified. It is clear that miRNAs are distinct from mRNAs in this case, so we separate our domain into the following spaces under the equivalence. If $A = \{0,1\}^{nMess}$ represent the state of the mRNAs, and $B = \{0,1\}^{nMicro}$ represents the state of the miRNAs, the domain of $\mathcal{L}$ is then specified as the product $A \otimes B$. This result is trivial, but useful when one considers the indistinguishability of mRNAs and miRNAs in this space (do we need a proof of this)? Let $X = \{0,1,2 \dots nMess\}$ and $Y = \{0,1,2 \dots nMicro\}$ represent the number of nonzero mRNAs and miRNAs, respectively.

For given $k, m \in X$ and $l \in Y$, we have the sets $S_1 = \binom{A}{k}, S_2 = \binom{A}{m}, s_1 = \binom{B}{l}$, the set of elements of A with k, m nonzero elements, and the set of elements of B with l nonzero elements. The quantity of interest initially appears to be equal to:

$$P\left(\mathcal{L}\left(m,l\right)=k\right) = \sum_{\sigma_k \in S_1} \sum_{\sigma_l \in s_1} \sum_{\sigma_m \in S_2} P\left(\mathcal{L}\left(\sigma_m, \sigma_l\right) = \sigma_k\right) \tag{20}$$

the probability of transition from $(m,l) \in X \otimes Y \to (k) \in X$. However, those probabilities are for a specific value of $\zeta = \{\zeta_{i;j}\},\ i = 1 \dots nMess,\ j = 1 \dots nMicro$. In fact, the final quantity we want to obtain is:

$$\mathbb{E}_\zeta\left[P\left(\mathcal{L}\left(m,l\right)=k\right)\right] = \sum_{\zeta} \sum_{\sigma_k \in S_1} \sum_{\sigma_l \in s_1} \sum_{\sigma_m \in S_2} P\left(\mathcal{L}\left(\sigma_m,\sigma_l\right)=\sigma_k \middle| \zeta\right) P\left(\sigma_m, \sigma_l, \sigma_k \middle| m,l,k\right) P\left(\zeta\right) \tag{21}$$

It is also clear that for any σ_k, $P\left(\mathcal{L}\left(\sigma_m,\sigma_l\right)=\sigma_k\right) \equiv 0$ if $\sigma_m - \sigma_k \notin A$. Two simplifications can be made. First, the action of $\mathcal{L}$ on a given mRNA is independent of its action on all other mRNAs. So, in general,

$$P\left(\mathcal{L}\left(\cdot,\sigma_j\right) \middle| \left(\cdot\right)_i = 1\right) = \rho_i \prod_j \left(1 - \zeta_{i;j}\sigma_j\left(i\right)\right) \tag{22}$$

Second, in 43 we may apply any permutation σon the j's without changing the value on the left-hand side (this is regardless of mRNAs/miRNAs being indistinguishable or not):

$$\rho_i \prod_j^{nMicro} \left(1 - \zeta_{i;j}\sigma_l\left(j\right)\right) = \rho_i \prod_j^{nMicro} \left(1 - \zeta_{i;\sigma(j)}\sigma_l\left(j\right)\right) \tag{23}$$

$$= \rho_i \prod_{j=1}^{l} \left(1 - \zeta_{i;\sigma(j)}\right) \tag{24}$$

As we have the admissibility requirement $\sigma_m - \sigma_k \in A$, one can extend σ so that the remaining $m-k$ elements of σ_m which are equal to 1 occupy the $k+1^{st}$ through the m^{th} elements of our binary vector. Hence, we can rewrite 42 as:

$$\mathbb{E}_\zeta\left[P\left(\mathcal{L}\left(m,l\right)=k\right)\right] = \sum_{\sigma_k\in S_1}\sum_{\sigma_l\in s_1}\sum_{\sigma_m\in S_2}\sum_{\zeta} P\left(\mathcal{L}\left(\sigma_m,\sigma_l\right)=\sigma_k\middle|\zeta\right)P\left(\zeta\right)P\left(\sigma_m,\sigma_l,\sigma_k\middle|m,l,k\right)$$

$$\sum_{\zeta} P\left(\mathcal{L}\left(\sigma_m,\sigma_l\right)=\sigma_k\middle|\zeta\right)P\left(\zeta\right) = \prod_{\sigma_k(i)=1}\rho_i\sigma_m\left(i\right)\prod_{j=1}^{l}\left(1-\zeta_{i;\sigma(j)}\right)\prod_{\sigma_k(i)=0}\left(1-\rho_i\sigma_m\left(i\right)\prod_{j=1}^{l}\left(1-\zeta_{i;\sigma(j)}\right)\right)P\left(\zeta_1,\zeta_2\right.$$

$$= \prod_{\sigma_k(i)=1}\rho_i\sigma_m\left(i\right)\prod_{j=1}^{l}\left(1-\zeta_{i;j}\right)\prod_{\sigma_k(i)=0}\left(1-\rho_i\sigma_m\left(i\right)\prod_{j=1}^{l}\left(1-\zeta_{i;j}\right)\right)P\left(\zeta_i\right)P\left(\zeta_i\right)..._l$$

where the second statement uses the fact that each miRNA is independent so that the columns of ζ are iid, namely $\zeta_{i;j}\overset{d}{=}\zeta_i$. This organization implies that $P\left(\zeta=\langle\zeta_{i;j}\rangle\right)=P\left(\zeta=\left\langle\zeta_{i;\sigma(j)}\right\rangle\right)$.

The same trick can be applied to σ_m and σ_k simultaneously. Since

$$P\left(\sigma_m,\sigma_k,\sigma_l\middle|m,l,k\right) = P\left(\sigma_m\middle|m\right)P\left(\sigma_k\middle|k\right)P\left(\sigma_l\middle|l\right)$$
$$= \binom{nMess}{m}^{-1}\binom{nMess}{k}^{-1}\binom{nMicro}{l}^{-1} \tag{28}$$

is independent of the values of σ_m,σ_k, we can apply a permutation $\sigma$to σ_m,σ_k so that the first m elements of $\sigma\left(\sigma_m\right)$ equal 1, and, for admissible σ_k, the first k elements of $\sigma\left(\sigma_k\right)$ equal 1. Also the indistinguishability (*not* the independence) of the mRNA gives us the simplification $\rho_i=\rho$. Putting this together,

$$\sum_{\zeta} P\left(\mathcal{L}\left(\sigma_m,\sigma_l\right)=\sigma_k\middle|\zeta\right)P\left(\zeta\right) = \prod_{\sigma_k(i)=1}\rho_i\sigma_m\left(i\right)\prod_{j=1}^{l}\left(1-\zeta_{i;j}\right)\prod_{\sigma_k(i)=0}\left(1-\rho_i\sigma_m\left(i\right)\prod_{j=1}^{l}\left(1-\zeta_{i;j}\right)\right)\left[P\left(\zeta_i\right)\right]^l \tag{29}$$

$$= \prod_{i=1}^{k}\rho\prod_{j=1}^{l}\left(1-\zeta_{\sigma(i);j}\right)\cdot\prod_{i=k+1}^{m}\left(1-\rho\prod_{j=1}^{l}\left(1-\zeta_{\sigma(i);j}\right)\right)\left[P\left(\zeta_i\right)\right]^l \tag{30}$$

Since σ_m *must* transition to an admissible σ_k, we get

$$P\left(\sigma_m,\sigma_k\middle|m,\ k,\ \sigma_k\text{ admissible}\right)=\left[\binom{nMess}{m}\binom{m}{k}\right]^{-1} \tag{31}$$

which makes sense: we sum over $\binom{nMess}{m}$ values of σ_m, for which there are $\binom{m}{k}$ admissible transition states each. There remains one more important simplification to make, which is namely to show that the column represented by $\zeta_i\overset{d}{=}\zeta_{\sigma(i)}$: that the ζ_i are identically distributed (but not necessarily independent). This result follows entirely from the distribution that we choose for ζ_i. Using three parameters, $mdeg_h$, $mdeg_l$, p_{eff}, we select $k\sim Binom\left(mdeg_h-mdeg_l,p_{eff}\right)+mdeg_l$. k unique indices in $\{1\ldots nMess\}$ are selected uniformly and each ascribed a value $\zeta_i=\frac{1-defect}{k}$, where $defect$ is a parameter indicating the "multiplicity" of the effects of a given mRNA. That is, the mean number of mRNAs affected in a given step by a single miRNA is $1-defect$; $defect$ is specified to be in $(0,1)$ but we plan to implement a means to include $defect>1$ in a sensible way (i.e. if $defect=-\left(k+1\right)$, then the probability of a miRNA neutralizing its target is 1, but k is random, so we get potential values of $\zeta_i>1$).

Returning to the equations above,

$$\sum_{\zeta} P\left(\mathcal{L}\left(\sigma_m, \sigma_l\right) = \sigma_k \middle| \zeta\right) P\left(\zeta\right) = \int \left[\prod_{i=1}^{k} \rho \prod_{j=1}^{l} \left(1 - \zeta_{i;j}\right)\right] \left[\prod_{i=k+1}^{m} \left(1 - \rho \prod_{j=1}^{l} \left(1 - \zeta_{i;j}\right)\right)\right] d\zeta_{\cdot}^{l}$$

where $\zeta_{\cdot}$ corresponds to any *column* of the $mRNA \times miRNA$ matrix ζ (makes sense as $l \in [0, nMicro]$). This integral form is useful for the following reason: for $defect \in (0,1)$, we have $\zeta_{i;j} \in (0,1)$. The function

$$f\left(\zeta; m, l, k\right) = \left[\prod_{i=1}^{k} \rho \prod_{j=1}^{l} \left(1 - \zeta_{i;j}\right)\right] \left[\prod_{i=k+1}^{m} \left(1 - \rho \prod_{j=1}^{l} \left(1 - \zeta_{i;j}\right)\right)\right]$$

is continuous in the matrix ζ. As a probability measure, we naturally have $\int d\zeta_{\cdot}^{l} = 1$, so, applying the mean value theorem to f, $\exists z \in [0,1]$ so that

$$\begin{aligned}
\int f\left(\zeta; m, l, k\right) d\zeta_{\cdot}^{l} &= f\left(z; m, l, k\right) && (32)\\
&= \left[\prod_{i=1}^{k} \rho \prod_{j=1}^{l} \left(1 - z\right)\right] \left[\prod_{i=k+1}^{m} \left(1 - \rho \prod_{j=1}^{l} \left(1 - z\right)\right)\right] && (33)\\
&= \left[\prod_{i=1}^{k} \rho \left(1 - z\right)^{l}\right] \left[\prod_{i=k+1}^{m} \left(1 - \rho \left(1 - z\right)^{l}\right)\right] && (34)\\
&= q^k \bar{q}^{m-k} && (35)
\end{aligned}$$

where $q = \rho\left(1 - z\right)^{l} \in [0,1]$. From the calculation in 52, this quantity is multiplied by the number of admissible values for σ_k uniformly, giving us $\binom{m}{k} q^k \bar{q}^{m-k}$ so the binomial theorem gives us, for any m and $\forall l \in [0, nMicro]\,,\ \rho \in [0,1]\,,\ 0 \leq mdeg_l \leq mdeg_h \leq nMess,\ defect \in [0,1]\,,$

$$\begin{aligned}
\sum_{k=0}^{m} \int \left[\prod_{i=1}^{k} \rho \prod_{j=1}^{l} \left(1 - \zeta_{i;j}\right)\right] \left[\prod_{i=k+1}^{m} \left(1 - \rho \prod_{j=1}^{l} \left(1 - \zeta_{i;j}\right)\right)\right] d\zeta_{\cdot}^{l} &= \sum_{k=0}^{m} \binom{m}{k} q^k \bar{q}^{m-k} && (36)\\
&= 1 && (37)
\end{aligned}$$

That is, $P\left(\#nz\ \mathcal{L}\left(x\right) = k \middle| \#nz\ x = m\right)$ is a valid probability distribution, and represents the probability of transitioning from m nonzero elts to k nonzero elts with l active miRNAs.

4.2.2 Figuring out how to calculate values in $\mathcal{L}$

Consider the decomposition of $\mathcal{L} = M^l \circ R$, where R is the "mRNA recognition" (or the ρ operator), and M is the "miRNA interference" operator for exactly *one* miRNA. This design is analogous to the formulation directly preceding this section, but has the potential to be computationally manageable. R is merely the

binomial matrix$R_{ij} = \binom{j}{i}\rho^i\rho^{j-i}$. When $l = 1$, we have:

$$\int \prod_{i=1}^{k}(1-\zeta_i)\prod_{j=k+1}^{m}\zeta_j d\zeta$$

4.2.3 Complete transition matrix under equivalence This is the one I'm editing

The matrix L derived above is correct provided we have distinctly identified the states of each miRNA. However, we want to determine the transition matrix averaged over all possible network configurations for any given state. Hence, we now consider the expansion of the elements $q_i, \bar{q}_i$ above under the equivalence specified above. We consider the action of $\mathcal{L}$ at a fixed time t. It is clear that miRNAs are distinct from mRNAs in this case, so we separate our domain into the following spaces under the equivalence. Denote the set of possible states of the mRNA by $A = \{0,1\}^{nMess}$, and the set of possible states of the miRNA by $B = \{0,1\}^{nMicro}$. Let $\widetilde{A}$ be the space of elements of A under our equivalence relation. Let $\varphi_A : A \to \widetilde{A}$ be the map defined for $a \in A$ by $\varphi(a) = (\#$ number of nonzero elements of $a)$; define φ_B analogously. The domain of $\mathcal{L}$ is then $\widetilde{A} \times \widetilde{B}$.

For given $k, m \in \widetilde{A}$ and $l \in \widetilde{B}$, define the sets

$$S_1 = \ a \in A \mid \varphi_A(a) = k \ , \tag{38}$$

$$S_2 = \ a \in A \mid \varphi_A(a) = m \ , \tag{39}$$

$$S_3 = \ b \in B \mid \varphi_B(b) = l \ . \tag{40}$$

the set of elements of A with k, m nonzero elements, and the set of elements of B with l nonzero elements. The quantity of interest initially appears to be equal to:

$$\mathbb{P}\left(\mathcal{L}\left(m,l\right) = k\right) = \sum_{s_1 \in S_1}\sum_{s_3 \in S_3}\sum_{s_2 \in S_2}\mathbb{P}\left(\mathcal{L}\left(s_1, s_3\right) = s_2\right) \tag{41}$$

the probability of transition from $(m,l) \in \widetilde{A} \times \widetilde{B} \to k \in \widetilde{A}$. However, those probabilities are for a specific value of $\zeta = (\zeta_{i;j})\,,\ i = 1 \ldots nMess,\ j = 1 \ldots nMicro$. In fact, the final quantity we want to obtain is:

$$\mathbb{E}_\zeta\left[\mathbb{P}\left(\mathcal{L}\left(m,l\right) = k\right)\right] = \sum_{\zeta}\sum_{s_1 \in S_1}\sum_{s_3 \in S_3}\sum_{s_2 \in S_2}\mathbb{P}\left(\mathcal{L}\left(m,l\right) = k \middle| \zeta\right)\mathbb{P}\left(s_1, s_3, s_2 \middle| m, l, k\right)\mathbb{P}\left(\zeta\right). \tag{42}$$

It is also clear that for any k, $\mathbb{P}\left(\mathcal{L}\left(m,l\right) = k\right) = 0$ if $m < k$. Two simplifications can be made. First, the action of $\mathcal{L}$ on a given mRNA is independent of its action on all other mRNAs since it occurs for a particular time t. So, in general,

$$P\left(\mathcal{L}\left(\cdot, \sigma_j\right) \middle| \left(\cdot\right)_i = 1\right) = \rho_i \prod_j \left(1 - \zeta_{i;j}\sigma_j\left(i\right)\right) \tag{43}$$

Second, in 43 we may apply any permutation σon the j's without changing the value on the left-hand side (this is regardless of mRNAs/miRNAs being indistinguishable or not): I don't think this makes sense, one would need something like the probability that given j miRNAs a given gene i will be hit. Start here next

time

$$\rho_i \prod_{j}^{nMicro} \left(1-\zeta_{i;j}\sigma_l\left(j\right)\right) = \rho_i \prod_{j}^{nMicro} \left(1-\zeta_{i;\sigma(j)}\sigma_l\left(j\right)\right) \quad (44)$$

$$= \rho_i \prod_{j=1}^{l} \left(1-\zeta_{i;\sigma(j)}\right) \quad (45)$$

As we have the admissibility requirement $\sigma_m - \sigma_k \in A$, one can extend σ so that the remaining $m-k$ elements of σ_m which are equal to 1 occupy the $k+1^{st}$ through the m^{th} elements of our binary vector. Hence, we can rewrite 42 as:

$$\mathbb{E}_\zeta\left[P\left(\mathcal{L}\left(m,l\right)=k\right)\right] = \sum_{\sigma_k\in S_1}\sum_{\sigma_l\in s_1}\sum_{\sigma_m\in S_2}\sum_{\zeta} P\left(\mathcal{L}\left(\sigma_m,\sigma_l\right)=\sigma_k\middle|\zeta\right)P\left(\zeta\right)P\left(\sigma_m,\sigma_l,\sigma_k\middle|m,l,k\right)$$

$$\sum_{\zeta} P\left(\mathcal{L}\left(\sigma_m,\sigma_l\right)=\sigma_k\middle|\zeta\right)P\left(\zeta\right) = \prod_{\sigma_k(i)=1}\rho_i\sigma_m\left(i\right)\prod_{j=1}^{l}\left(1-\zeta_{i;\sigma(j)}\right)\prod_{\sigma_k(i)=0}\left(1-\rho_i\sigma_m\left(i\right)\prod_{j=1}^{l}\left(1-\zeta_{i;\sigma(j)}\right)\right)P\left(\zeta_1,\zeta_2\right.$$

$$= \prod_{\sigma_k(i)=1}\rho_i\sigma_m\left(i\right)\prod_{j=1}^{l}\left(1-\zeta_{i;j}\right)\prod_{\sigma_k(i)=0}\left(1-\rho_i\sigma_m\left(i\right)\prod_{j=1}^{l}\left(1-\zeta_{i;j}\right)\right)P\left(\zeta_i\right)P\left(\zeta_i\right)..._{l}$$

where the second statement uses the fact that each miRNA is independent so that the columns of ζ are iid, namely $\zeta_{i;j} \overset{d}{=} \zeta_i$. This organization implies that $P\left(\zeta=\langle\zeta_{i;j}\rangle\right)=P\left(\zeta=\left\langle\zeta_{i;\sigma(j)}\right\rangle\right)$.

The same trick can be applied to σ_m and σ_k simultaneously. Since

$$P\left(\sigma_m,\sigma_k,\sigma_l\middle|m,l,k\right) = P\left(\sigma_m\middle|m\right)P\left(\sigma_k\middle|k\right)P\left(\sigma_l\middle|l\right)$$

$$= \binom{nMess}{m}^{-1}\binom{nMess}{k}^{-1}\binom{nMicro}{l}^{-1} \quad (49)$$

is independent of the values of σ_m,σ_k, we can apply a permutation $\sigma$to σ_m,σ_k so that the first m elements of $\sigma\left(\sigma_m\right)$ equal 1, and, for admissible σ_k, the first k elements of $\sigma\left(\sigma_k\right)$ equal 1. Also the indistinguishability (*not* the independence) of the mRNA gives us the simplification $\rho_i=\rho$. Putting this together,

$$\sum_{\zeta} P\left(\mathcal{L}\left(\sigma_m,\sigma_l\right)=\sigma_k\middle|\zeta\right)P\left(\zeta\right) = \prod_{\sigma_k(i)=1}\rho_i\sigma_m\left(i\right)\prod_{j=1}^{l}\left(1-\zeta_{i;j}\right)\prod_{\sigma_k(i)=0}\left(1-\rho_i\sigma_m\left(i\right)\prod_{j=1}^{l}\left(1-\zeta_{i;j}\right)\right)\left[P\left(\zeta_i\right)\right]^l \quad (50)$$

$$= \prod_{i=1}^{k}\rho\prod_{j=1}^{l}\left(1-\zeta_{\sigma(i);j}\right)\cdot\prod_{i=k+1}^{m}\left(1-\rho\prod_{j=1}^{l}\left(1-\zeta_{\sigma(i);j}\right)\right)\left[P\left(\zeta_i\right)\right]^l \quad (51)$$

Since σ_m *must* transition to an admissible σ_k, we get

$$P\left(\sigma_m,\sigma_k\middle|m,\ k,\ \sigma_k \text{ admissible}\right) = \left[\binom{nMess}{m}\binom{m}{k}\right]^{-1} \quad (52)$$

which makes sense: we sum over $\binom{nMess}{m}$ values of σ_m, for which there are $\binom{m}{k}$ admissible transition states each. There remains one more important simplification to make, which is namely to show that the column represented by $\zeta_i \overset{d}{=} \zeta_{\sigma(i)}$: that the ζ_i are identically distributed (but not necessarily independent). This result follows entirely from the distribution that we choose for ζ_i. Using three parameters, $mdeg_h,\ mdeg_l,\ p_{eff}$, we select $k \sim Binom\left(mdeg_h - mdeg_l, p_{eff}\right) + mdeg_l$. k unique indices in $\{1 \dots nMess\}$ are selected uniformly and each ascribed a value $\zeta_i = \frac{1-defect}{k}$, where $defect$ is a parameter indicating the "multiplicity" of the effects of a given mRNA. That is, the mean number of mRNAs affected in a given step by a single miRNA is $1 - defect$; $defect$ is specified to be in $(0,1)$ but we plan to implement a means to include $defect > 1$ in a sensible way (i.e. if $defect = -(k+1)$, then the probability of a miRNA neutralizing its target is 1, but k is random, so we get potential values of $\zeta_i > 1$).

Returning to the equations above,

$$\sum_{\zeta} P\left(\mathcal{L}\left(\sigma_m, \sigma_l\right) = \sigma_k \middle| \zeta\right) P\left(\zeta\right) = \int \left[\prod_{i=1}^{k} \rho \prod_{j=1}^{l} \left(1 - \zeta_{i;j}\right)\right] \left[\prod_{i=k+1}^{m} \left(1 - \rho \prod_{j=1}^{l} \left(1 - \zeta_{i;j}\right)\right)\right] d\zeta_{\cdot}^{l}$$

where $\zeta_{\cdot}$ corresponds to any *column* of the $mRNA \times miRNA$ matrix ζ (makes sense as $l \in [0, nMicro]$). This integral form is useful for the following reason: for $defect \in (0,1)$, we have $\zeta_{i;j} \in (0,1)$. The function

$$f\left(\zeta; m, l, k\right) = \left[\prod_{i=1}^{k} \rho \prod_{j=1}^{l} \left(1 - \zeta_{i;j}\right)\right] \left[\prod_{i=k+1}^{m} \left(1 - \rho \prod_{j=1}^{l} \left(1 - \zeta_{i;j}\right)\right)\right]$$

is continuous in the matrix ζ. As a probability measure, we naturally have $\int d\zeta_{\cdot}^{l} = 1$, so, applying the mean value theorem to f, $\exists z \in [0,1]$ so that

$$\begin{aligned}
\int f\left(\zeta; m, l, k\right) d\zeta_{\cdot}^{l} &= f\left(z; m, l, k\right) && (53)\\
&= \left[\prod_{i=1}^{k} \rho \prod_{j=1}^{l} \left(1 - z\right)\right] \left[\prod_{i=k+1}^{m} \left(1 - \rho \prod_{j=1}^{l} \left(1 - z\right)\right)\right] && (54)\\
&= \left[\prod_{i=1}^{k} \rho \left(1 - z\right)^{l}\right] \left[\prod_{i=k+1}^{m} \left(1 - \rho \left(1 - z\right)^{l}\right)\right] && (55)\\
&= q^k \bar{q}^{m-k} && (56)
\end{aligned}$$

where $q = \rho\left(1 - z\right)^{l} \in [0,1]$. From the calculation in 52, this quantity is multiplied by the number of admissible values for σ_k uniformly, giving us $\binom{m}{k} q^k \bar{q}^{m-k}$ so the binomial theorem gives us, for any m and $\forall l \in [0, nMicro],\ \rho \in [0,1],\ 0 \leq mdeg_l \leq mdeg_h \leq nMess,\ defect \in [0,1]$,

$$\begin{aligned}
\sum_{k=0}^{m} \int \left[\prod_{i=1}^{k} \rho \prod_{j=1}^{l} \left(1 - \zeta_{i;j}\right)\right] \left[\prod_{i=k+1}^{m} \left(1 - \rho \prod_{j=1}^{l} \left(1 - \zeta_{i;j}\right)\right)\right] d\zeta_{\cdot}^{l} &= \sum_{k=0}^{m} \binom{m}{k} q^k \bar{q}^{m-k} && (57)\\
&= 1 && (58)
\end{aligned}$$

That is, $P\left(\#nz\ \mathcal{L}(x) = k \middle| \#nz\ x = m\right)$ is a valid probability distribution, and represents the probability of transitioning from m nonzero elts to k nonzero elts with l active miRNAs.

4.2.4 Figuring out how to calculate values in $\mathcal{L}$

Consider the decomposition of $M^l \circ R$, where R is the "mRNA recognition" (or the ρ operator), and M is the "miRNA interference" operator for exactly *one* miRNA. This design is analogous to the formulation directly preceding this section, but has the potential to be computationally manageable. R is merely the binomial matrix$R_{ij} = \binom{j}{i}\rho^i\rho^{j-i}$. When $l = 1$, we have:

$$\int \prod_{i=1}^{k}(1-\zeta_i)\prod_{j=k+1}^{m}\zeta_j d\zeta$$

5 Explicit calculation of $\mathcal{L} \circ \mathcal{C}$

5.1 Order of operations

Because we are interested in "phenotype" only, we consider $\mathcal{L} \circ \mathcal{C}$ and not $\mathcal{C} \circ \mathcal{L}$. Consequently, the number of nonzero mRNAs and not the number of nonzero miRNAs is of interest. Although $\mathcal{C}$ considers mRNAs and miRNAs as identical, $\mathcal{L}$ does not. In fact, the form of $\mathcal{L}$ as the matrix $\mathbb{P}$ explicitly depends on the values for the miRNAs $(x_j,\ j = nMess + 1 \ldots nRNA$. Importantly, the composition $\mathcal{L} \circ \mathcal{C}$ is a mapping from $2^{nMess} \to 2^{nMess}$, while its converse is $2^{nRNA} \to 2^{nRNA}$As the expression of $\mathcal{L}$ indicates, the derived matrix $\mathbb{P}$ depends on the values of $q_i = \rho_i x_i \prod_{j=nMess+1}^{nRNA}(1-\zeta_{i;j}x_j)$. The values of $\rho_i \equiv \rho$ as well as the distribution of $\zeta_{i;j}$ are parameters, but the values of the miRNAs, x_j, are not.

5.1.1 Lifetimes of RNA vectors

Incomplete, also this section should come after both operators are completely described.

The state $\overrightarrow{0} = \{0, 0...0\}$ is an absorbing state for $\mathcal{L}$ as well as for $\mathcal{C}$. This notion is no surprise and is in fact desirable - we expect any system to eventually "die" at some time. Calculating lifetimes for $\mathcal{L}$ is very straightforward if r is static, and so it will be omitted at this time. Lifetime calculation is extremely important, but needs to be evaluated for a non-constant miRNA states. This will be returned to in the complete analysis of $\mathcal{L} \circ \mathcal{C}$. Finite lifetimes are brought up here because this point is a key deviation from [10]. In this paper, any gene in state 0 whose inputs were all equal to 0 would revert to state 1. This construct allowed for non-trivial absorbing states. However, it does not seem entirely biologically applicable. Nonetheless, it is important to note that that additional constraint does not prevent one from resolving the entire transition matrix for both operators.

The start of the Lifetime Calculation

Write T_{ij} for the transition probability associated with transitioning from state i to state j under the operation of $\mathcal{L} \circ \mathcal{C}$. The probabilities associated to the transition to the state $\overrightarrow{0} = \{0, \ldots, 0\}$ is given by the

first column of T. Let X_T be a random variable which represents the "lifetime" of the system. Explicitly

$$X_T = \inf\{k \,|\, T^k = \overrightarrow{0}\}. \tag{59}$$

We wish to consider the expectation of X with and without miRNAs in the system. We have

$$\mathbb{E}(X_T) = \sum_{k=0}^{\infty} k\mathbb{P}(X_T = k) \tag{60}$$

$$= \sum_{k=1}^{\infty} k(\widetilde{T}^k). \tag{61}$$

where we denote by $\widetilde{T}$ the matrix T where we have set the first column of T to be 0. The point of this is two-fold. First, it allows us to count the number of ways that a state can transition to $\overrightarrow{0}$ without counting situations where we transition to $\overrightarrow{0}$ and then to another state and then back to $\overrightarrow{0}$. The other effect is that setting that column to 0 removes the eigenvector of T with eigenvalue 1. This ensures that the potentially infinite sum converges. Thus, one can view $\mathbb{E}(X)$ as a function of T. Since transition matrices of size $nMess \times nMess$ form an algebra we may write $\mathbb{E}(X) = f(T)$ where

$$f(T) = \int_{\partial B(\sigma(T),0)} \frac{f(z)}{z - T} dz. \tag{62}$$

We integrate over a ball who radius is chosen so that the spectrum of is contained within it.

6 Evolution Description

First Evolution Idea

Still needs to be finished and edited

Let $\{\mathcal{R}\}$ be the set of mRNAs. An mRNA $A \in \mathcal{R}$ has a coding region, A_C, and a regulatory region, A_R. Each region is defined by a numerical sequence where each base is represented by a number (A = 0, T=1, C=2,G=3). So, for instance, if $A \in \mathcal{R}$ a possible coding or regulatory region might be 01222302. If $A, B \in \mathcal{R}$ and the coding region of A is complementary to the regulatory region of B then then there is a directed edge from A to B and A promotes B. An example set of complementary sequences is 0122220 and 1033331. On the other hand if the coding region of A is the same as the regulatory region of B then there is a directed edge from A to B and A downregulates B. Otherwise there is no connection between two mRNAs.

At each time step each coding and regulatory region has a probability to mutate p. And after mutations occur we update the connections of the graph. We run the network for n time steps and measure the entropy.

Evolution is defined as gene duplication and divergence. We will choose a gene at random and duplicate it, and all of its connections. We then allow the two copies of the same gene to mutate seperately. Since the model has $\mathcal{L}$ and $\mathcal{C}$ operators we represent gene duplication as mRNA duplication and divergence. (Clarify that for technical reasons one wants to generate the sequence of genes to be duplicated first so taht they are always present and when an evolution event occurs we will simply allow the regulatory region to connect. In this way one can maintain the phase space of the experiment).

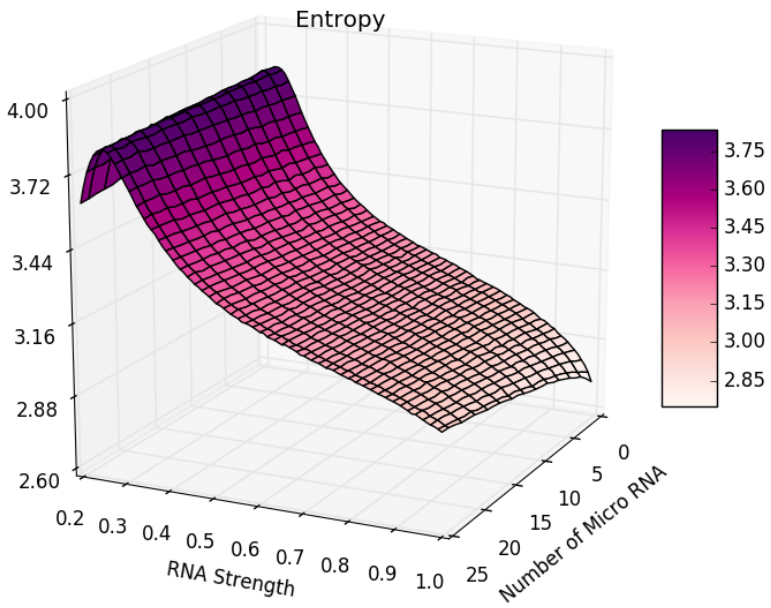

Figure 1: Example Entropy plot

After evolution we measure entropy again. Then rerun the experiment with miRNAs added. How should miRNA be defined? They could have a coding region and regulatory region which is constructed to be regulated by one mRNA and to downregulate at least one mRNA.

7 Numerical Results

These will be added once we fix the bug that is preventing RNAstrength from having an effect

8 Miscellaneous Questions:

1. Do we want exact matching or approximate as Claus had?
2. Do we want multiple coding and regulatory regions?
3. Do miRNAs effect the entropy of the evolved system?
4. Should miRNAs mutate?
5. What is the biological story that we're telling. This is key for the paper submission.

References

[1] A. Garg, K. Mohanram, A. Di Cara, and G. De Micheli *Modeling stochasticity and robustness in gene regulatory networks* Bioinformatics (2009) 25 (12) 101-109

[2] K. Cordes and D. Srivastava *MircroRNA Regulation of Cardiovascular Development* Circ Res. 2009;104:724-732

[3] M. Davidich, and S. Bornholdt *Boolean network model predicts cell cycle sequence of fission yeast* PLoS One. 2008 Feb 27; 3(2):e1672.

[4] C. Espinosa-Soto, P. Padilla-Longoria, and E. Alvarez-Buylla *A Gene Regulatory Network Model for Cell-Fate Determination during Arabidopsis thaliana Flower Development That is Robust and Recovers Experimental Gene Expression Profiles*

[5] A.K. Eisfeld, S. Schwind, R. Patel, X. Huang, R. Santhanam, C. J. Walker, J. Markowitz, K. W. Hoag, T. M. Jarvinen, B. Leffel, D. Perrotti, W. E. Carson III, G. Marcucci, C. D. Bloomfield, and A. Chapelle *Intronic miR-3151 within BAALC drives leukemogenesis by deregulating the TP53 pathway.* Sci. Signal. 7, ra36 (2014).

[6] L. Ferrarelli *Focus Issue: Noncoding RNAs in cancer* Sci. Signal. 17 Mar 2015: Vol. 8, Issue 368, pp. eg3

[7] M.R. Fabian, N. Sonenberg, and W. Filipowicz *Regulation of mRNA translation and stability by microRNAs* Annual Review of Biochemistry, (2010) Vol. 79: 351 -379 .

[8] D.T. Gillespie *Exact stochastic simulation of coupled chemical reactions* J. Phys. Chem. 1977;81:2340-2361.

[9] J. Lu, A. G. Clark *Impact of microRNA regulation on variation in human gene expression* Genome Res. 2012 Jul; 22(7): 12431254. doi: 10.1101/gr.132514.111

[10] C. Kadelka, D. Murragarra, and R. Laubenbacher *Stabilizing gene regulatory networks through feed-forward loops* Chaos **23**, 025107 (2013) doi: 10.1063/1.4808248

[11] F. Li, T. Long, Y. Lu, Q. Ouyang, and C. Tang *The yeast cell-cycle network is robustly designed* PNAS 2004 101 (14) 4781-4786

[12] L. T. MacNeil and A. J. M. Walhout *Gene regulatory networks and the role of robustness and stochasticity in the control of gene expression* Genome Res, 21:645 657, 2011.

[13] D. Shultz, E. Jacob, J. Onuchic, and P. Wolynes *Molecular level stochastic model for competence cycles in Bacillus subtilies* PNAS 2007 104 (45) 17582-17587

[14] V. R. Sanghvi, K. J. Mavrakis, J. Van der Meulen, M. Boice, A. L. Wolfe, M. Carty, P. Mohan, P. Rondou, N. D. Socci, Y. Benoit, T. Taghon, P. Van Vlierberghe, C. S. Leslie, F. Speleman, and H.G. Wendel *Characterization of a set of tumor suppressor microRNAs in T cell acute lymphoblastic leukemia* Sci. Signal. 7, ra111 (2014)

[15] Y. A. Xiao *Tutorial on Analysis and Simulation of Boolean Gene Regulatory Network Models* Current Genomics 2009;10(7):511-525. doi:10.2174/138920209789208237.