Chapter 4

Models of RNA Interaction from Experimental Datasets: Framework of Resilience

William Seffens

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.69452

Abstract

Resilience is a network property of systems responding under stress, which for biomedicine correlates to chronic or acute insults. Current need exists for models and algorithms to study whole transcriptome differences between tissues and disease states to understand resilience. Goal of this effort is to interpret cellular transcription in a dynamic system biology framework of RNA molecules forming an information structure with regulatory properties acting on individual transcripts. We develop and evaluate a bioinformatics framework based on information theory that utilizes RNA expression data to create a whole transcriptome model of interaction that could lead to the discovery of new biological control mechanisms. This addresses a fundamental question as to why transcription yields such a small fraction of protein products. We focus on a transformative concept that individual transcripts collectively form an "information cloud" of sequence words, which for some genes may have significant regulatory impact. Extending the concept of *cis*- and *trans*-regulation, we propose to search for RNAs that are modulated by interactions with the transcriptome cloud and calling such examples *nebula* regulation. This framework has implications as a paradigm change for RNA regulation and provides a deeper understanding of nucleotide sequence structure and -omic language meaning.

Keywords: transcriptome, RNA, diffusion, secondary structure, resilience, information theory

1. Introduction

The concept of resilience is receiving increasing attention in chronic stress-related disease conditions. Resilience has been shown in clinical studies to play a protective role in patients

with chronic disease conditions including osteoarthritis, breast and ovarian cancer, diabetes, and cardiovascular disease. The purpose of this study is to explore the relationships between RNA-RNA interactions and to devise a related measure of resilience from network properties of the whole transcriptome.

1.1. RNA physiology

At various levels, RNA is processed by alternate mechanisms [1], suggesting a biological framework that supports important system network features such as resilience. Trafficking of RNAs is essential for cellular function and homeostasis, but only recently it has become possible to visualize molecular events *in vivo*. Analysis of RNA motion within the cell nucleus has been particularly intriguing as they have revealed an unanticipated degree of dynamics within the organelle [2]. Single-molecule RNA imaging methods have revealed that the intranuclear and cytoplasmic trafficking occurs largely by energy-independent mechanisms and is driven by diffusion. RNA molecules undergo constrained diffusion, largely limited by the spatial constraint imposed by chromatin and chromatin-binding proteins if in the nucleus as demonstrated in numerous studies. In the cell, transcripts move by a stop-and-go mechanism, where free diffusion is interrupted by random association with cellular structures [3]. The ability and mode of motion of RNAs has implications for how they find nuclear targets on chromatin or cellular sub-compartments and how macromolecular complexes are assembled *in vivo*. Most importantly, the dynamic nature of RNAs is emerging as a means to control physiological cellular responses and pathways [4]. For example, unexpectedly complicated nuclear egress and nuclear import of small RNAs is more common than previously appreciated [5].

Much attention has been focused on noncoding RNAs and their physiological/pathological implications [6]. This focus in RNA research is ultimately directed toward understanding the regulation of protein-coding gene networks, but ncRNAs also form well-orchestrated regulatory interaction networks [7]. For example, computational prediction of miRNA target sites suggests a widespread network of miRNA-lncRNA interaction [8]. Others suggest the possibility of widespread interaction networks involving competitive endogenous RNAs (ceRNAs) where ncRNAs could modulate regulatory RNA by binding and titration of binding sites on protein coding messengers [9]. Cellular uptake and trafficking of RNA could be widespread [10]. As the number of experiments increases rapidly, and transcriptional units are better annotated, databases indexing RNA properties and function will become essential tools to understand physiologic processes in the transcriptome.

1.2. Biological-omic information theory

Much of bioinformaticians sequence analyses focuses on methodologies based on string alignment algorithms. However, such approaches fail to discover genomic aspects of systemic nature regarding dynamics or resilience. An alternative framework is based on alignment-free methods of genome analysis, where global properties of genomes are investigated [11]. A key concept of informational analysis is that of probability distributions. A genomic, or in our case transcriptomic, distribution associates to discrete values defined on transcripts, the number of

http://dx.doi.org/10.5772/intechopen.69452

times these values occur in a given transcriptome. The general concept of discrete probability distribution, called information source, was the starting point of information theory developed by Shannon [12]. Links between information theory and biology emerged from Shannon's Ph. D. thesis, titled "An Algebra for Theoretical Genetics" (1940), where the notion of information entropy was introduced [13]. For example, distributions of codons have shown characteristic properties that are linked to biological meanings, such as secondary structure free energy [14]. Other approaches based on the recurrence of genomic elements and on correlation structures in DNA sequences use mutual information, which plays a central role in the mathematical analysis of message transmission. Dictionary-based methodologies analyze sequences through properties of collections of words. Dictionaries are concepts from formal language theory, probability, and information theory that provide new perspective which may uncover the physiology of internal transcriptome structures.

2. Methods

We formally define the transcriptome as an information structure, and then construct several simple models as examples. The most realistic model is used to examine real datasets of partitioned RNAs for validation of framework.

2.1. Transcriptome information theory structure

RNA sequence is abstractly represented as a string over the nucleotide alphabet $\mathfrak{R} = \{A, C, G, U\}$. This can be extended to modified nucleotides with an extended alphabet $\mathfrak{R} \cong \{A, C, G, U, N\}$, such that symbol N represents a modified nucleotide. W_k denotes a set of alphabet letters of length k, called k-mers and $\mathfrak{W}$ denotes the set of all possible nonempty strings over the alphabet $\mathfrak{R}$. Given a transcript string $S = s1, s2, \ldots$ to sn, of length n, $S[i, j]$ with $1 \leq i \leq j \leq n$ is the substring of S from position i to position j (included). The length of S is $|S| = n$. Substrings of S of length k are called k-words or simply words of S. In the following, the entire transcriptome is denoted by W based on k-mer dictionaries and entropies, which are aimed at defining and computing informational indexes for representative sets of transcriptomes. We assume that the complexity of a transcriptome increases with its distance from randomness, as identified by suitable comparison between transcriptomes of the same length. This framework provides clues about the appropriate k length to consider for analysis of transcriptome properties.

2.2. Spatial transcriptome information cloud (STIC) model construction

We hypothesized that miRNA localization in cellular compartments is an emergent property from Brownian motion interactions of a cloud of RNA sequences and RNA-binding proteins that can be analyzed in W [15]. There k-mer words of miRNA functional size were added to a dictionary from sliding windows of transcript sequences S. A prediction from this cloud model is that anomalous diffusion can occur if random-walk transcripts interact with their surrounding scaffold as a stochastic semantic cloud, and if the cloud relaxation time is a longer time frame than transit [16]. We showed that RNAs with sequences similar to the whole transcriptome exhibit modified or enhanced transport compared to RNA sequences without

similar sequences [17]. Thus, RNAs were found to partition into different cellular compartments based on a semantic similarity of word compositions within W_k. We determine the frequency of all *k*-mers in the transcriptome as a matrix composed of RNA sequences and their word copy levels. For each transcript, we count the number of *k*-mer words in common within W_k or dictionary as a semantic similarity measure to the transcriptome, and we can also able to compare such counts to randomized W_k sequence words.

Model assumptions are: (1) RNA diffuses away from point of transcription creating a cloud of *k*-mer sequences. (2) All RNAs comprise the transcriptome, and each transcript is affected by local RNAs with effective interaction windows of some sequence word length *k* nucleotides (nt). We assume significant *k*-mer word size to be 3–22 nt, which is equal to the functional miRNA size at the high side, and down to below the size of the "core" sequence [18]. (3) The diffusion rate of individual RNAs depends on degree of (a) sequence similarity and (b) reverse complementarity of RNA words at that location in the STIC (**Figure 1**). (4) Cloud dictionaries (collection of transcriptome word sets in W_k) change as function of distance from transcriptome site and cell state. (5) The cloud affects anomalous RNA diffusion that can give rise to an emergent and patterned behavior in the cell [19].

We model the RNA sequence word content of the transcriptome cloud as a function of distance from transcription site at the chromosome. RNA molecule diffusion in nuclear compartments would lead to cytoplasmic and extracellular localization of RNA if the transcript half-life is greater than its transit time. Calculations at arbitrary transit distances could be determined

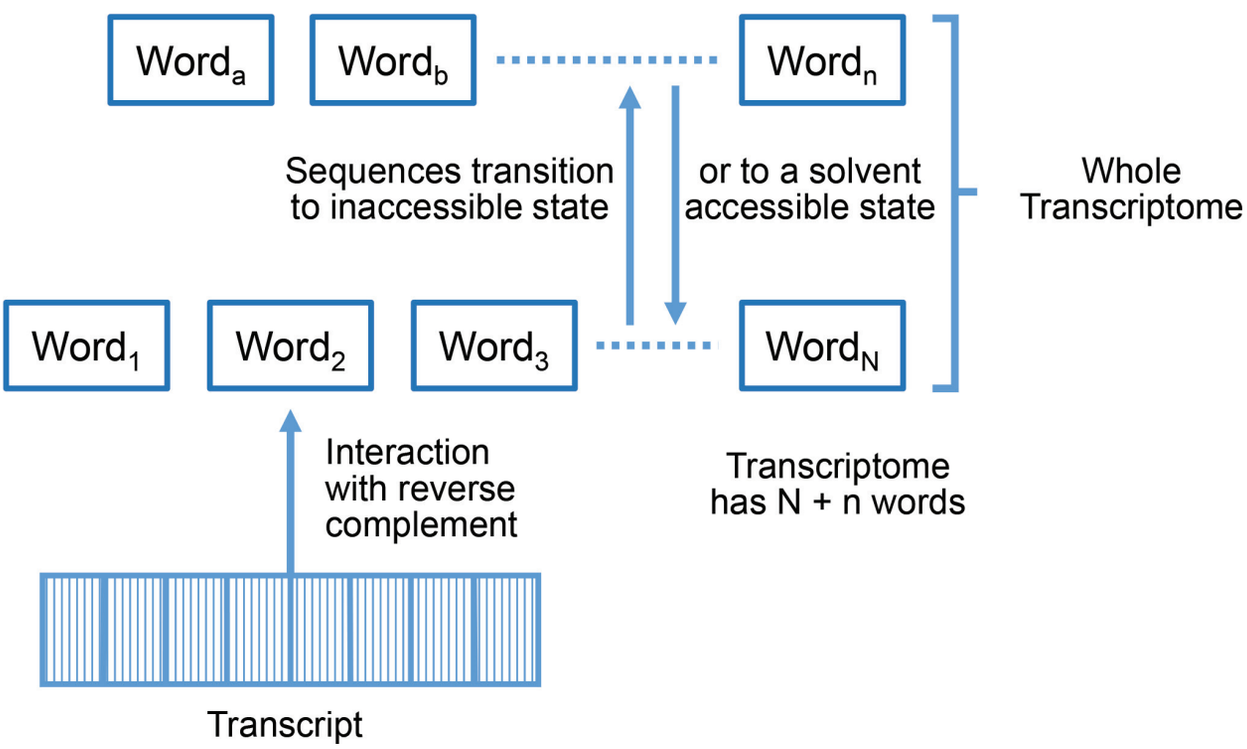

Figure 1. *K*-mer words classified as solvent-accessible or inaccessible. Note that in this framework, *k*-mer words are generated from a sliding window and not from a contiguous word segments as could be interpreted by the figures adjacent to word blocks. Transcriptome is a dictionary of $N + n$ words and their associated frequencies. Interaction of transcript with transcriptome affects diffusion from solvent-accessible bases (words). Transcript itself is a part of the transcriptome.

from this model with a large set of partial differential equations modeling RNA mobility, but was described as computationally prohibitive [15] for any realistic sized transcriptome. Each sequence S would be dynamically modeled with a neighborhood sized number of RNA-RNA interactions. Instead, we pursue a thermodynamic approach based on the Fokker-Planck equation to quantify stochastic processes in liquid medium [20]. RNA interactions are assumed to function over sequence lengths encompassing small single-stranded and solvent-accessible regions.

Assuming smallest word in the cloud is 3 nt long, corresponding to the lower limit of size for a seed sequence in miRNA [18]. The upper limit for word size is set at 22 nt, corresponding to the size of a typical mature miRNA. Again, this is the same as the miRNA response elements (MRE) size in the simpler related ceRNA hypothesis by Salmena [9]. Instead, we determine the frequency of all words in the transcriptome as a matrix composed of RNA sequences and copy levels from RNA-seq datasets. For each transcript, count number of words in common with the cloud dictionary as a similarity measure to the transcriptome (tCount), and also count reverse complement words (rcCount) for RNA-RNA interactions. These raw counts can be multiplied by the frequencies of repetitive words in W to yield tWord = tCount * tFreq and rcWord = rcCount * rcFreq. As shown by Seffens [17], miRNAs with greater similarity to the transcriptome, i.e., greater tCount and tWord, are suggested to diffuse differentially based on spatial partitioning. In addition, greater intramolecular RNA-RNA interaction would be expected to hinder diffusion. This work proposes a general RNA sequence function that combines the influence of similarity with native (NAM) and reverse complementarity (RCM) measures as a cloud interaction function: $\mathfrak{C}$[W, NAM, RCM], such that cloud interactions increase with RCM, and decrease with NAM. Transcripts with low $\mathfrak{C}$ would have "ideal solution" diffusion coefficients and found in cytoplasmic compartment, and those with greater $\mathfrak{C}$ would be slowed by RNA-RNA interactions and hence enriched in nuclear or perinuclear compartments.

2.2.1. Accessibility of an RNA sequence word

For each component word of a transcript, determine whether it is expected to be in a single-stranded and solvent-accessible state (state "A"), or double-stranded or buried within the RNA molecule and is inaccessible (state "I"). For model calculations and preliminary studies (Model-1 W^1 discussed later), we assume all words are accessible in state "A," and the transcriptome is uniform within the cell (i.e., ignore distance r from transcription site). Construct a matrix $W_k(T, f_A, f_I, r)$ for each word size k, and populate the respective matrix with the component words of the transcriptome from RNA-seq reads such that S is the actual word sequence, f_A is the frequency or number of accessible words of that sequence, and f_I is the frequency or number of inaccessible words. Matrix W_k then contains information of all transcript sub-sequences and is a representation of the spatial transcriptome information cloud in some volume elements of the cell fraction. Let the diffusion coefficient for a transcript be described [24]as D_{RNA} [21]. Then the effect of interaction of that transcript with the cloud would yield

$$D_{RNA} = D_{RNA}{}^{ideal} - \mathfrak{C}\left[W_k\left(S, f_A, f_I, r\right), S\right] = D_{RNA}{}^{ideal} - \text{RCM} + \text{NAM} \quad (1)$$

where $\mathfrak{C}$ is the cloud interaction term for molecule RNA exhibiting probabilities of RNA-RNA interactions as a function of the STIC represented by matrix W_k at some position r in the cell.

For RNA expression data from the whole cell, r is ignored. In experiments from purified nuclei, r ranges from 0 to the radius of the cell's nucleus, r_N. In experiments derived from the cytosol, r ranges from r_N to the cellular membrane radius r_C. Experimental data from extracellular vesicles will have $r > r_C$ and RNA half-life becomes important to consider as a factor. As a first approximation for the $\mathfrak{C}$ function, we assume that the deviation from ideal D_{RNA} scales as the number of reverse complement words (rcCount) in common with transcriptome W_k and is measured by difference to the number tCount of words in common with W_k, which normalizes for transcript size. We could also compare to ranCount, number of words in common with a randomized W_k. Putting together, we have

$$\mathfrak{C}[W_k, S] = \alpha \text{ rcCount}/4^k \tag{2}$$

to normalize number of words, or alternately,

$$\mathfrak{C}[W_k, S] = \alpha(\text{rcCount} * \text{rcFreq} - \text{tCount} * \text{tFreq}) = \text{RCM for } \alpha = 1.0 \tag{3}$$

where α is a scaling factor and is dimensionless. The reverse complement measure (RCM), which factors rcCount word frequencies by rcFreq, then subtracting the count of words in common with transcriptome (tCount) by the corresponding tFreq, is one of several possible measures for correlation to measured compartmentalization of individual transcripts from RNA-expression datasets. The content of $W_k(r)$ changes as a function of r due to changing concentrations of transcripts in the cell. Boundary condition on whole cell measurement from microarray or RNA-seq experiments would be

$$W_k = \int_0^{r_C} W_k(r)\, dr \tag{4}$$

assuming no export from the cell. If there are no reverse complement words in common between transcript S and W_k, then $\mathfrak{C}$ $[W_k, S]$ is zero and the diffusion of that molecule is ideal. As a first approximation for the $\mathfrak{C}$ function, we assume that the deviation from ideal D_{RNA} scales as the number (rcCount) of reverse complement words in common with W_k and is compared by a difference to the number tCount of words in common with W_k to normalize for transcript size. We could also compare to ranCount, number of words in common with a randomized W_k. Reverse complement measure (RCM) factors word frequencies to assess transcript-cloud interactions that correlate to measured compartmentalization of individual transcripts [17].

2.2.2. Words that are solvent-accessible

The above model treatment assumed that all RNA sequences are available for reverse complementarity interactions. RNAs except for miRNAs typically have regions that are solvent-inaccessible and/or double-stranded, preventing intramolecular interactions [22]. mRNAs have more secondary structures or intra-strand base pairing than expected by chance [23]. We have determined the secondary structure of all RefSeq transcripts to predict single-stranded regions using RNAfold [24], while others have used RNA structure predictors (RNAplfold in

Refs. [25, 26]) in a pooling predictor using machine learning [27]. Additionally, nucleotide solvent-accessibility in RNA structures could be estimated by the neural network method of Singh [22] using models of window size 3 nt, which could be expanded to 5–9 nt windows for k length. Alternatively, accessible surface area can be calculated by a publically available program NACCESS [28] to refine the STIC transcriptome words to those populated from single-stranded regions only, along with confidence measures. Solvent accessibility estimates for each transcript word partition the frequency entries in the transcriptome matrix $W_k(S, f_A, f_I, r)$ by reducing f_A in the amount that f_I increases. Shifts of f_A to f_I could be caused by RNA-binding factors (RNA or protein) that cover a word in the transcript or the word in the transcriptome, or indirectly by binding to some other region of the RNA causing a *cis*-type of structural alteration leading to solvent inaccessibility (**Figure 2**). Transcriptome cloud or *nebula* regulation is introduced here and is proposed to occur as an indirect result of some factor that changes f_A/f_I for some word that then alters a different interacting transcript's diffusion coefficient. Conversely, f_I to f_A shifts could be caused by the release of binding factors or conformational change leading to exposure of the particular word and to nearby target RNAs. Dictionaries with this dynamic accounting of the transcriptome are labeled T, instead of the simpler W word matrix.

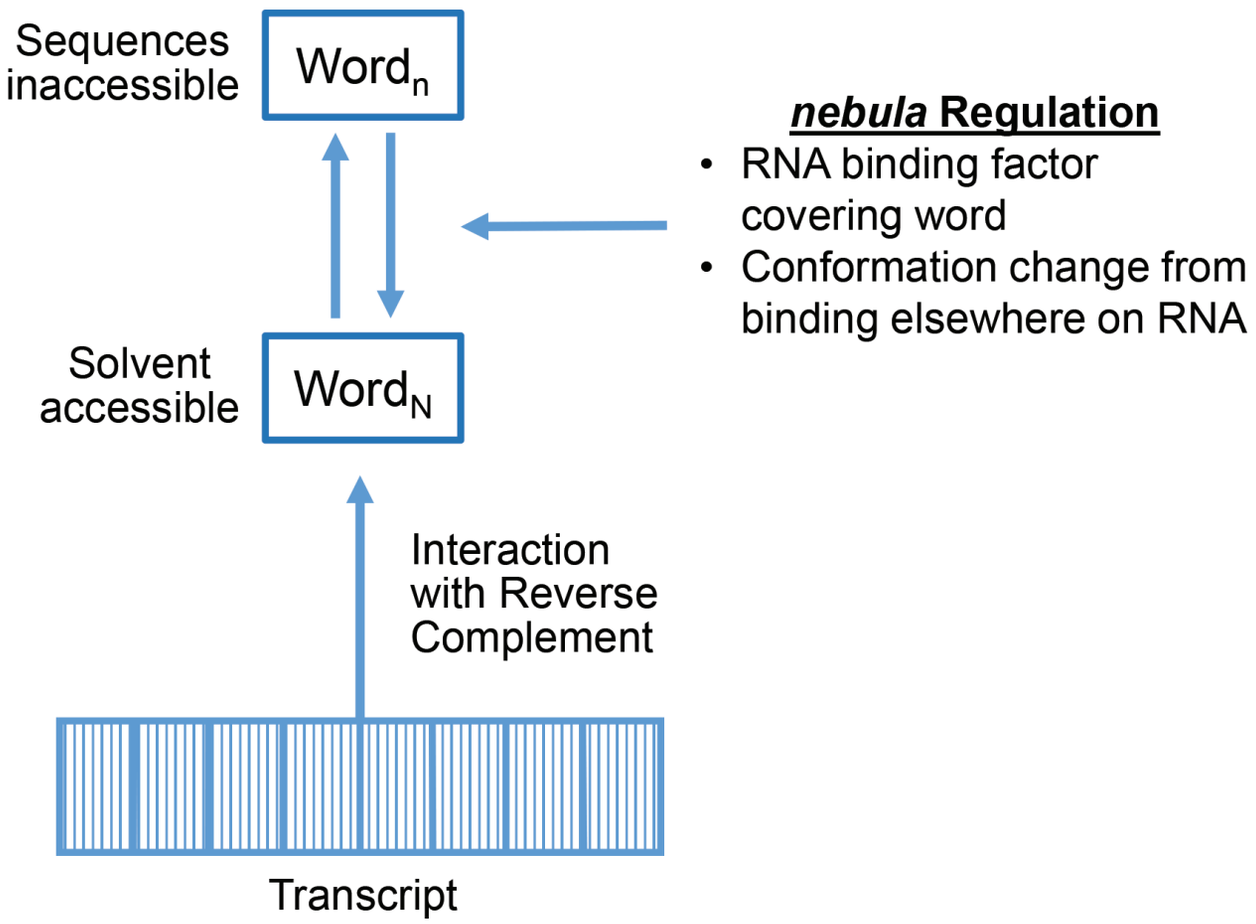

Figure 2. Words classified as solvent-accessible or -inaccessible in dynamic transitions. Interaction of RNA binding-proteins and other RNA affects balance of solvent-accessible words in transcriptome and transcript.

2.2.3. Genome-wide profiling of in vivo *RNA structure*

Recent transcriptome-wide RNA structure profiling through application of structure-probing enzymes or chemicals combined with high-throughput sequencing promises *in vivo* RNA structural information availability. Resultant datasets provide opportunity to investigate RNA structural information on global scale for STIC development. The analysis of high-throughput RNA structure profiling data requires considerable computational effort and currently dataset are not readily available. StructureFold processes and performs analysis of raw high-throughput RNA structure profiling data [29], incorporating wet-bench structural information from chemical probes and ribonucleases to restrain RNA structure prediction via RNAstructure and ViennaRNA package algorithms. StructureFold is deployed via the Galaxy platform. Alternatively, structure-seq is a recent quantitative and high-throughput method that provides genome-wide information on RNA structure with single-nucleotide resolution [30]. The methodology can perform both *in vitro* and *in vivo* RNA structure-function determinations, with insights to RNA regulation of gene expression and RNA processing. Implementation of structure-seq begins with chemical RNA structure probing under single-hit kinetics conditions. Modified RNA is then subjected to reverse transcription using random hexamer primers, then reverse transcription executed until it is blocked by chemically modified residues. Resultant cDNAs are amplified by adapter-based polymerase chain reaction (PCR) which are subjected to high-throughput sequencing, subsequently allowing retrieval of structural information on a genome-wide scale. A single structure-seq experiment can provide information on tens of thousands of RNA structures in a matter of weeks. Ding et al. [31] used RNAstructure calculated for each of thousands of mRNAs a positive predictive value (PPV), which they use to compare relative frequencies of base pairing *in vivo* constrained RNA structures to *in silico* predicted RNA structure. They found that most mRNAs do not fold *in vivo* as to *in silico*-predicted structures, as evident from a broad PPV distribution. Interestingly, mRNAs of cold and metal ion stress-response genes folded *in vivo* significantly different from their unconstrained *in silico* predictions. These stressors are known to affect RNA structure and thermo-stability like melting temperature T_{m}. Instead, genes involved with basic biological functions such as gene expression, protein maturation or processing, and peptide metabolic processes show little change in their *in vivo*-constrained and *in silico*-predicted RNA secondary structures. Ding speculated mRNAs related to cell maintenance and showing high PPV may have evolved to resist large conformational changes in order to maintain homeostasis, an idea suggesting RNA resilience. This bias may be detectable in our transcriptome model W. As genome-wide profiling of *in vivo* RNA structure datasets becomes easily available [32], the information can be incorporated into the STIC model by adding custom transform and load tools for each dataset.

2.3. Simplest models of transcriptomes

We consider simplest models of the transcriptome to examine some limits on the model parameters and functions. Component of a transcriptome are listed in **Table 1**.

2.3.1. Transcriptome model 0 (W^0)

Simplest model considers a spatially uniform transcriptome composed of ribosomes, tRNA, and a minor number of mRNAs. Here, the distribution of transcripts is assumed uniform and the transcriptome model lacks spatial elements. Further, consider a transcriptome where

Transcriptome construction		
Class	**Biotype**	**Network system function**
ncRNA	rRNA	Most abundant; constant character
ncRNA	tRNA	Most frequent; constant character
Protein coding	mRNA	Act as gene; bigger than coded protein
ncRNA	miRNA	Block some mRNA; smallest func. RNA
Processed transcripts	lncRNA	Block or regulate some other RNA
Transcriptome $= \Sigma\, W^{\text{Biotype}}$		

Notes: Sequence words summed into dictionary over all transcript types.

Table 1. Transcriptome construction from different RNA biotypes. Transcriptome model W^1 is a subset of 8 human rRNA and tRNA types.

ribosomes and tRNAs are all "A" for adenosine, and the mRNAs are also all "A." We can construct transcriptome model W^0 using transcript values found in **Table 1** from Seffens [17], with sizes n and respective abundances. The number of nonunique words (of size k) would be $n - k + 1$ for each transcript. Using sizes and respective abundance values in **Table 1** [17], gives 2.8×10^{10} words, composed of all "A"s. Assume mRNAs are 2 kb "A"s, then there are no reverse complement interactions, so RCM should be zero, while NAM would be maximal (RCM and NAM defined in Section 2.2). The diffusion coefficient for these mRNAs would be ideal since there are no base-pairing interactions within W^0 and transcript similarity to the transcriptome is maximal. Now assume the mRNAs are all "U"s, they now strongly interact with the majority of W^0. RCM becomes $2000 \times 2.8 \times 10^{10}$ or 5.6×10^{13} for each mRNA transcript (while NAM would be zero). This would be the expected maximal value for RCM with mRNAs of 2 kb size, yielding diffusion coefficients smaller than ideal values. These RNAs would exhibit larger intramolecular RNA-RNA interactions, and they do not look (NAM = 0) and behave ($D_{\text{RNA}} \ll D_{\text{RNA}}{}^{\text{ideal}}$) as the rest of the transcriptomes.

2.3.2. Transcriptome model 0-R ($W^{0\text{-}R}$)

Now assume that the transcriptome W^0 is composed of completely random sequences of *A*, *C*, *G*, *U*s-labeled $W^{0\text{-}R}$. How many of the 2.8×10^{10} words of length k composed of four different letters would be unique (not identical) in the model? Combination of all possible k-mer words would be $4^{22} = 1.76 \times 10^{13}$ since there are four possible nucleotide letters at each of the k positions. Since there are about 1000 times more possible combinations than there are $k = 22$ words, we could assume that all 22-mer words are unique. Smaller values of k will result in repeats or duplicate words increasing frequency values in W_k. These calculations give an expected value for RCM based on no biases in the sequences.

2.3.3. Transcriptome model 1 (W^1)

The next more realistic model is composed of eight real human RNA transcripts comprising a simple representation of the transcriptome in a cell (**Table 1** and in Ref. [17]). It is assembled

from four of the most prevalent human tRNAs with lengths of $n = 71–73$ nt, and four of the major subunits of the eukaryotic ribosome with sizes from $n = 121$ to 5034 nt, with the total number of nucleotides N being the sum of the nucleotides in each transcript, or $N = 7470$ nt. Then the frequency of words with length k that are contained in each transcript is a subset of the number of possible k-mer words which is $n – k + 1$. In Model-1 labeled as W^1, for each word length from $k = 3$ to 22, word count was calculated along with the sum of the frequencies of those words extracted from the simple eight RNA transcripts. The intermediate output from program TIC-generator (for transcriptome information cloud generator, described in Ref. [17]) listed all k-mer words contained in each transcript, together with their frequency of occurrence. These lists from the eight rRNA and tRNA transcripts were combined, and then duplicate words resolved to form dictionary $W_k{}^1$. With the total possible number of words of length $k = 4^k$, the fraction of all the words actually present in $W_k{}^1$ decreased for increasing word size [17]. It is interesting that the peak in unique and total duplicate (blue diamonds in Ref. [*17*]) words is maximal at the same size as the miRNA "seed" sequence as defined in Ref. [18].

2.3.4. Randomized transcriptome of Model-1 ($W^{1\text{-}R}$)

We ran TIC-generator with shuffled-sequence transcripts labeled Model 1-R. Base composition of Model-1 transcriptome is 1341 "A," 2320 "C," 2519 "G," and 1291 "T," or 18% "A," 31% "C," 34% "G," and 17% "T." Using a random letter generator, we assembled four random transcriptomes with the same transcript length for the eight sequences and equal Model-1 base composition. We examined mostly word lengths k of 7 and 8 in preliminary studies shown below.

2.4. Real model validation

As a validation of this transcriptome model framework, we utilized the simple transcriptome model version (simple model W^1) that used real highly expressed genes, and for comparison separately, randomized sequences of that transcriptome ($W^{1\text{-}R}$). This simple realistic model is composed of only eight real human RNA transcripts as a basic representation of the transcriptome in a cell. Experimental validation of the basic model transcriptome for k-mers considered various trial functions of semantic word similarity and reverse complementarity, which were calculated using published data sets. For example, trial functions evaluated include tWord for transcriptome words in common with target multiplied by respective word frequency in W_k. A total of seven RNA studies, with data sources grouped into high and low study parameter sets, were statistically analyzed by mean values and t-test calculated as two-tail t-test under two-sample equal variance assumption models (**Table 2**). Validation for the STIC model examined various functions of reverse complementarity using these published data sets. Here, we assume that appearance in exosomes or microparticles requires greater mobility and hence larger diffusion coefficients than cytoplasmic or nuclear RNAs [17]. Several functions tested include tWord for transcriptome words in common with target multiplied by word frequency in the transcriptome, rcWord (reverse complement k-mer words in common times frequency), RCM = rcWord – tWord, reverse complement count (RCC) measure = tCount – rcCount, Z-RCC as a z-score of RCC compared to four randomized transcriptomes Model 1-R, Z-RCM as the z-score of RCM, RCC-Ran which subtracts the value computed from 1-R and finally (RCC-Ran)/Len which

is normalized for sequence length. The first five studies examined miRNA, while the Chen [33] and Friedel [34] studies measured mRNAs. Description of data sources that were grouped into high and low study parameter sets, with mean values and t-tests calculated detailed in sections below.

2.4.1. Model 1 validation with miRNA from exosome datasets

The Villarroya-Beltri [35] work reports on microarray datasets of exosome and cellular fractions from activated and resting human T lymphocyte cultures. They differentially assessed whether RNAs are specifically enriched within exosomes by performing microarray analysis of activation-induced variations in mRNA and miRNA profiles from primary T lymphoblast and their secreted exosomes. Data found in their supplementary data and also data publicly available at gene expression omnibus as Gene Expression Omnibus (GEO) Series accession number GSE50972 were used for **Table 2**. They showed that for most cases, miRNAs modulated upon activation are differentially found in cells and exosomes for either upregulated or downregulated miRNAs. This suggests that mRNA and miRNA loading into exosomes is not a simple passive process. Specific miRNAs were more highly expressed in exosomes than found in the cells, and in most cases this difference is preserved under cellular resting or activated conditions. Similarly, most miRNAs that are preferentially found in cells than in exosomes also keep this tendency regardless of the activation state of the cell. As such, Villarroya-Beltri classified some miRNAs as specifically sorted into exosomes (labeled EXOmiRNAs), whereas others are specifically retained in cells (as CLmiRNAs). We calculated tCount and rcCount as a count (**Table 2** in Ref. [17]), and tWord and rcWord, the latter which factor the expression level of that word. Other measures compared counts and words to a randomized transcriptome (RAN). We used a word size $k = 7$ roughly equal to the miRNA seed sequence length [17]. Values of rcWord (mean 10.31) were lower than tWord (mean 12.45), and hence RCM and RCC were more negative for exosomes compared to cytoplasmic miRNAs. This supports the STIC model since exosome transcripts must diffuse further than cytoplasmic (CL) RNA, so avoid reverse complementarity. In summary, all trial measures calculated from this dataset showed significant support for the transcriptome model except for Z-RCM.

2.4.2. Model 1 validation with nuclear-enriched miRNAs

Park et al. [36] study compared microarray analysis of cytoplasmic and in this case nuclear fractions of hct116 colon cancer cells. They identified various miRNAs that exist in isolated nuclei from miRNA profiles correlated between cytoplasmic and nuclear fractions from multiple microarray analyses. Nuclear confinement of the mature form of miRNAs was validated by controlling reverse transcriptase RT-PCR conditions excluding the presence of precipitate forms of miRNA (e.g., as pri-miRNA or pre-miRNA). They found that elevated levels of representative miRNAs in purified nuclei support the idea that significant numbers of mature miRNAs survive not only in the cytoplasm but also in the nucleus. We sorted their data by N/C ratio and *partitioned these data into two groups: $N/C > 0.47$, which was nuclear-enriched (45 samples), and $N/C < 0.47$, which was preferentially found in the cytoplasm (33 samples). We found that tCount was 4.02 for nuclear-enriched, and 5.00 for cytoplasmic, with a t-test p-value of 0.116 between the groups; while tWord was 4.73 for nuclear and 10.58 for cytoplasmic

Source experiment	N	tWord	rcWord	RCM	RCC	Z-RCC	Z-RCM	RCC-Ran	(RCC-Ran)/Len
Villarroya-Beltri									
EXO-CL resting	75	4×10^{-7}**	2×10^{-5}**	0.08*	0.023**	0.029**	0.603	0.04**	0.038**
EXO-CL activated	67	4×10^{-7}**	1×10^{-5}**	0.206	0.008 **	0.033**	0.503	0.032**	0.028**
Park paper									
N/C > 0.471 nuclear	43	0.024**	0.021**	0.62	0.76	0.77	0.31	0.41	0.42
Huang paper									
Top-low rcmm	100	0.522	0.02**	0.002 **	0.042 **	0.83	0.16	0.078*	0.072*
Cheng paper									
Top-low	50	0.128	0.002 **	0.035 **	0.002 **	0.062*	0.25	0.132	nc
Guduric-Fuchs paper									
Ratio EV/cell top-low	10	0.093*	0.39	0.3	0.075*	0.046**	0.178	0.03**	nc
EV RPMM top-low	10	0.79	0.973	0.736	0.96	0.268	0.816	0.306	nc
Chen paper									
Perinuclear-cell	6	0.62	0.76	0.24	0.14	0.15	0.18	0.076*	0.095*
Friedel									
mRNA half-life	15	0.017**	0.025**	0.86	nc	nc	0.44	nc	nc

Notes: Double-asterisk cells have significance below 0.05, while single-asterisk cells have significance below 0.10 but above 0.05. Cells with "nc" were not calculated from randomized transcriptome.

Table 2. t-Tests of case studies with STIC model parameters.

miRNAs, with a significant t-test *p*-value of 0.023 between nuclear and cytoplasmic groups. We also found nuclear-enriched miRNAs have higher rcWord values compared to cytoplasmic miRNA (*p*-value = 0.021 in **Table 2**), suggesting those transcripts have greater potential to interact with other transcriptome RNAs and hence may have lower than expected diffusion coefficients. The other evaluated measures did not show significance between groups.

2.4.3. Model 1 validation with additional RNA studies

Huang et al. [37] study utilized RNA-seq with exosomes from human plasma. We found that the top 100 abundant miRNAs in exosomes had tCount (mean 4.80) and tWord (mean 6.72) measures compared to those lower 100 with low "rcmm" reads (mean 4.64 and 7.41, respectively). In support of the STIC model, exosome transcripts have more similarity to the simple model transcriptome. Exosome abundant miRNAs had negative RCM (mean −0.87) and RCC (mean −0.27) measures compared to those with low rcmm reads (mean 1.37 and 0.55, respectively). The most significant trial function was RCM (*p*-value = 0.002) followed by rcWord (*p*-value = 0.02) measure. From these data, we find similarity that exosome transcripts have less reverse complementarity to the simple Model-1 transcriptome. Again, these results are supported by Cheng et al. [38] study of exosomes in human blood. From 50 most abundant miRNAs in exosome samples labeled "Plasma UC Exo," we find mean tCount and tWord

values of 4.56 and 6.00 compared to 5.58 and 8.80, respectively, for low abundance transcripts. This set of exosome miRNAs had RCM and RCC values of −1.54 and 3.8 compared to 0.36 and 5.8 for low abundance transcripts, again supporting the STIC model. Several of the trial functions in **Table 2** were significant measures for data sets in that study.

Pursuing in-depth understanding of the mechanism supporting selective exportation of miRNAs to extracellular vesicles (EVs), Guduric-Fuchs [39] employed next generation sequencing to discriminate global expression patterns of small RNAs in HEK293T cells and the EVs that they released. Enrichment of overexpressed miRNA in EVs was measured by RT-qPCR in HEK293T cells, mesenchymal stem cells, macrophages, and immune cells. We sorted data from Guduric-Fuchs by EV/cell ratio, then compared the top 10 (exosome-enriched) and bottom (cytoplasmic enriched) miRNAs by evaluating the measures listed in **Table 2**. Only trial functions Z-RCC and RCC-RAN were significant from this dataset. Overall from using EV/cell in various measures examined across the studies, tWord and tCount (from Ref. [17]), along with their difference (tW–tC), have values that progress from lower for nuclear, higher for cytoplasmic, and highest for exosomal miRNAs. Therefore, we consider under transitivity, EXO > CL > NUC for these transcriptome measures of similarity. This supports the notion that miRNAs with sequence similarity to the overall transcriptome can random-walk furthest from their points of transcription if the secretion mechanism requires a great distance to travel. These conclusions on trial functions are most significant with the tCount measure, with a p-value close to zero for the Villarroya-Beltri study, and 0.016 for the Guduric-Fuchs study, while the Park study showed little difference (p-value = 0.122) for tCount between nuclear and cytoplasmic enrichment.

2.4.4. Word count normalization from RNA-seq datasets

Normalization is a crucial step in the analysis of RNA-seq data and has a strong impact on the detection of differentially expressed genes sought to validate the STIC model. Several normalization strategies have been proposed to correct for between-sample distributional differences in read counts, such as differences in total counts (i.e., sequencing depths), and within-sample gene-specific effects, such as gene length or GC-content effects [40]. Global-scaling normalization adjusts gene-level counts by a single factor per sample, such as the per-sample total read count, or reads per kilobase of exon model per million mapped reads (RPKM), or some housekeeping gene count. Statistical corrections by a quantile per-sample count distribution or other robust summaries obtained by relating each sample to a reference sample (e.g., trimmed mean of M values (TMM) and methods of Anders and Huber [41]). Although there have been efforts to systematically compare normalization methods [42], this important aspect of RNA-seq analysis is still not fully resolved. When data arise from complex experiments as in Section 2 above, involving cell fractionation, low-input RNA or different batches and read lengths, there may be more to correct for than differences in sequencing depth, referred to as unknown nuisance technical variation error. One methodology correction is the addition of spike-in controls within the normalization procedure [43]. Control designs have been successfully employed in microarray normalization, for miRNA and mRNA arrays [44]. Negative controls in the normalization procedure test the assumption that the majority of genes are not differentially expressed between study conditions. This assumption can be violated when a

global shift in expression occurs between conditions, such that control-based normalization may be necessary for technical variation, and a global mean read for global differences in RNA levels.

3. Spatial and temporal localization

We follow with a description of possible experimental data sets for populating transcriptome model in W. RNA-seq data sets would be the preferred source for fine structure of word contents, but microarray expression data could also be used for overall population of W.

3.1. Spatial localization by RNA imaging

The only method that provides insight into both the level and localization in single cells is *in situ* hybridization (ISH), which has increased considerably in importance in RNA research. ISH along with multiplex RNA profiling (MERFISH) can be used to measure the degree of associations among transcripts. Numerous RNA species have been identified, counted, and localized in single cells using MERFISH, a single-molecule imaging approach that uses combinatorial labeling and sequential imaging with an encoding scheme capable of detection and/or correction of errors. This multiplexed measurement of individual RNAs can be used to measure the gene expression profile and noise, along with covariation in expression among different genes, and spatial distribution of RNAs within single cells.

3.1.1. Localization of small RNAs

For miRNAs, ISH is exceptionally challenging because of miRNA features such as small size, sequence similarity among various miRNA family members, and low tissue-specific or development-specific expression levels. Standard ISH protocols can be modified to improve miRNA detection [45]. Locked nucleic acid (LNA/DNA) probes have great utility in miRNA detection because of short hybridization time, high efficiency, discriminatory power, and high melting temperature of the miRNA/probe complex [46]. Minimal length of LNA/DNA probes was found to be 12 nt with probes usually containing 30% LNA nucleotides [46]. A mixture of 2′-OMe RNA and LNA modifications in a 2:1 ratio resulted in improved specificity and stability of the probe/RNA duplex in comparison to LNA/DNA probes [47]. Experiment specificity was found to be further improved by lengthening the probe length to 19 nt [48].

3.1.2. Localization by MERFISH

Chen et al. [33] used array-synthesized oligopools as templates to make encoding probes in the MERFISH protocol. An oligopaint approach developed by Beliveau et al. [49] can generate a large number of oligonucleotide probes to label chromosome DNA. Inspired by this approach, Chen et al. [33] designed a two-step labeling scheme to encode and read out cellular RNAs. They labeled a target set of cellular RNAs with a set of encoding probes, each probe comprising a RNA targeting sequence and two flanking readout sequences. Four readout sequences were assigned to each target RNA species based on error-correction optimized code words.

They identified these readout sequences with complementary FISH probes via rounds of hybridization and imaging; each round using a different readout probe. To increase the signal-to-background ratio, each cellular RNA is labeled with ∼192 encoding probes.

3.2. RNA diffusion

Brownian effects are ubiquitous in numerous examples of soft condensed matter physics [20] in which the system can be modeled as a set of interacting degrees of freedom in contact with a heat reservoir. Brownian motion plays an important role when one infers macroscopic behaviors from mesoscopic levels of description, frequently a desire in the study of complex systems. Dynamics at the mesoscopic level is governed by a set of Langevin processes or equivalently by the corresponding *N*-particle Fokker–Planck equation. This scheme applies nonequilibrium thermodynamics to derive the kinetic equations describing the evolution of an N-particle probability distribution function [20]. One then considers a system of *N* Brownian particles diluted in a solvent, which acts as a thermal reservoir. Particle velocities are then modeled as internal thermodynamic variables and permit an analysis in the phase space of the Brownian particles. A local equilibrium hypothesis constrains the phase space level and from it one derives the thermodynamic entropy balance equation. Entropy production accounts for irreversible processes taking place in the phase space, then quantifying fluxes and forces can be done in a similar manner as in the thermodynamics of irreversible processes [20]. A general thermodynamic treatment of systems of *N* interacting Brownian motion particles as described by Fokker-Planck equations is detailed by Savel'ev et al. [16].

4. Resilience as a systems biology measure from transcriptome model

Development of a resilience measure from transcriptome RNAs could improve basic knowledge of the transcriptome and responses to stress. Transcriptome size and overall variation have been documented across cell cycle stages, tissue types, developmental stages, diurnal cycles, sexes, and environment [50]. Despite the ubiquity of transcriptome size variation, its potential to introduce systematic bias into expression profiling has been largely overlooked and this study uncovers responses of the transcriptome to stress.

4.1. Formalization of metric for resilience in biological systems using STIC metrics

Insight into structural determinants of robustness and resilience can guide the understanding of systems that go through transitions. Systems engineering research has developed methodologies to measure the functionality and complexity of engineered systems for designing and assessing system resilience. While system functions, resilience, functionality, and complexity are widely used concepts in systems engineering, there is significant diversity in definitions and no unified approach to measurement in the systems biology area [51]. One method for measuring impacts on functionality in dynamic engineered systems is based on changes in kinetic energy [52]. This metric can be applied at particular levels of abstraction and system scales, consistent with the established multiscale nature of biological systems.

4.2. Measuring complexity

A difficulty in complexity theory is the lack of a clear definition for complexity, particularly one that is measurable [53]. Underlying cause for this lack of a unified complexity definition is that there are numerous conceptual types of complexity. The first formal treatment of complexity focused on algorithmic complexity, which reflects the computation requirements for a mathematical process [54]. Senge [55] and Sterman [56] expand the scope of definition to include dynamic complexity, which is primarily characterized by difficult-to-discern and hard-to-measure cause-effect relations. A recent workable definition is that of thermodynamic depth, which essentially asserts that complexity is a "measure of how hard it is to put something together" [57]. Several variations on this approach share the commonality that complexity should disappear for both ordered and purely stochastic systems [58]. Additionally, Bar-Yam [59] defined complexity as the length of the shortest string that can represent the properties of a physical system. This string could be the result of measurements and observations over time.

An energy-based metric was proposed by Chaisson [60] measuring the energy rate density, where Φm is energy rate density, E is energy flow through a system, τ is the time frame, and m is system mass. Chaisson obtains results that correlate well with other notions of complexity, and below we add our proposed relation from this transcriptome model framework

$$\Phi m = E/\tau m \text{ or which we propose is : } \alpha(\Sigma^S \text{NAM} + \Sigma^S \text{RCM})/N \tag{5}$$

A practical difficulty in using the Φm metric is determining the appropriate mass and energy. In measuring the Φm of a transcriptome, we can use the mass of RNA production and the total energy processed by the system. Energy in this framework could be the total sum of all possible RNA-RNA interactions, which is just the count of all NAM and RCM in W as a sum of overall transcript sequences S. However, the total energy of a transcriptome does not flow just through its cell, but also exported to the extracellular space and captured from that external source of transcripts, the mass of which is difficult to measure.

While higher functionality can be associated with increased resiliency and robustness, the concepts are not synonymous. As defined by the INCOSE Resilient Systems Working Group, "Resilience is the capability of a system with specific characteristics before, during, and after a disruption to absorb the disruption, recover to an acceptable level of performance, and sustain that level for an acceptable period of time" [61]. Robustness is the ability of a system to reject disturbances without altering its state. A system is robust when it can continue functioning in the presence of internal and external challenges without fundamental changes to the original system. In relation to previous section on energy availability, robustness is the ability for a system to retain reachable states in the event of falling available energy.

4.3. Framework for measuring resilience

Instead, complexity in the presented framework can be derived from properties of W or T as in **Figure 3**. Consider a transcriptome from a cell type alpha to be represented as T_α, such that it is the sum of all RNAs, including mRNA, miRNA, lncRNA, and rRNA within the cell (**Table 1**). This set is the result of transcripts produced from the cellular DNA, ${T_\alpha}^0$, transcripts captured

http://dx.doi.org/10.5772/intechopen.69452

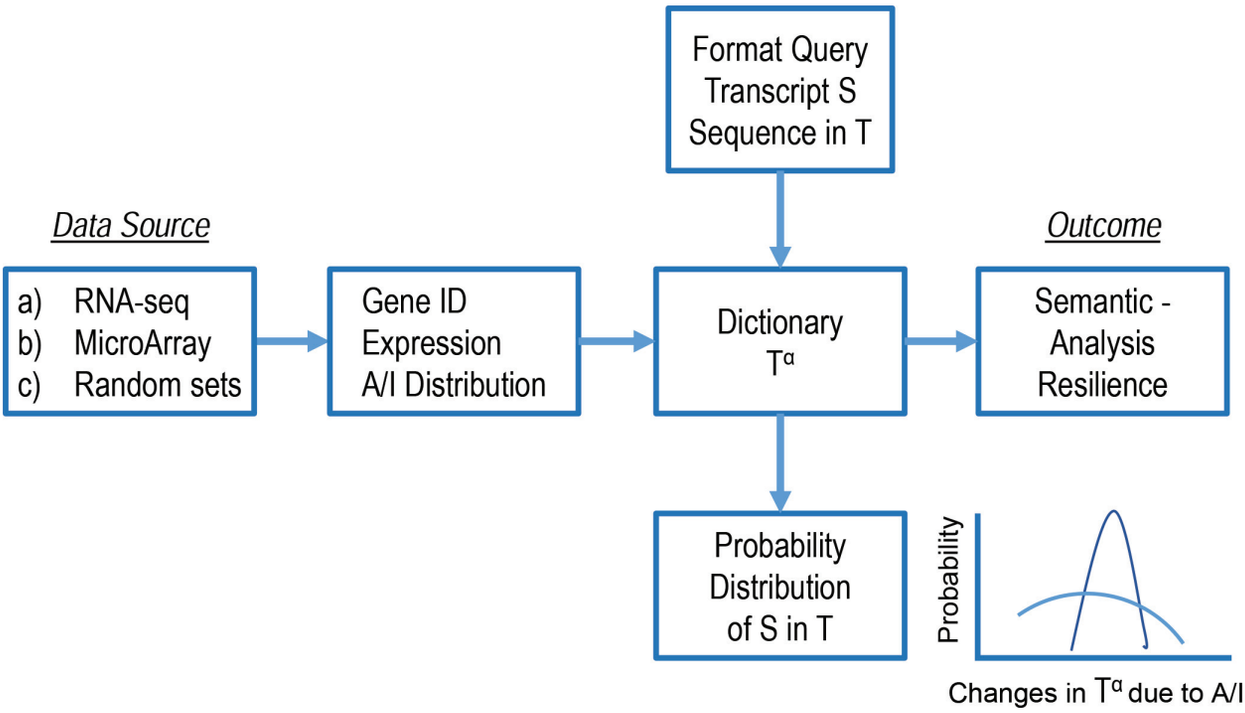

Figure 3. Framework for deriving transcriptome interactions and resilience. Data source is from RNA expression experiments using RNA-seq or microarray values, or randomized sets for controls. From input sets, the aligned gene ID and frequency of the extracted words are populated into a dictionary. Gene ID is used to calculate solvent-accessible (A) and inaccessible (I) word probabilities from full length transcripts *in silico*. The dictionary can be queried for any sequence *S* to find probability distribution of *S* in the dictionary. Changes in the transcriptome will change the distribution due to changes in A/I for affected words. Overall metrics of the dictionary measure resilience using Eq. (8) in the text.

from the extracellular space (EC) in the form of microparticles and exosomes ${T_{\mathrm{EC}}}^{\mathrm{IN}}$, and depletion as microparticle or exosome export to the extracellular space with ${T_{\alpha}}^{\mathrm{OUT}}$. Or,

$$T_{\alpha} = T_{\alpha}^{0} + T_{EC}^{IN} - T_{\alpha}^{OUT} \tag{6}$$

with

$${T_{\alpha}}^{\mathrm{OUT}} = T_{\alpha} * \mathfrak{F}[S,\ \mathrm{RC}, n] \approx T_{\alpha} * \ [a * f_{S}(S)/(f_{\mathrm{RC}}(S, \mathrm{RC})\ * n)] \tag{7}$$

where $\mathfrak{F}$ is a filter function with parameters *S* (transcript sequence), RC (reverse complement of transcript sequence), *n* sequence length of *S*, and "*a*" is a fitting parameter with suitable dimensions, derived from: $\mathfrak{F}\ \alpha$ NAM/(RCM * *n*) proportionality. Thus the extracellular pool is composed of transcripts with greater similarity *S*, and less reverse complementarity RC to the transcriptome of origin and also have smaller size *n*. The filter functions $f_{\mathrm{S}}(S)$ and $f_{\mathrm{RC}}(S,\mathrm{RC})$ operate on sequences *S* and RC, and essentially is a semantic selection filter on transcripts by affecting diffusion. We propose that resilience of the cell is proportional to size of the transcriptome filter $\mathfrak{F}$, then resilience $\alpha\ |\mathfrak{F}|$, where $|\mathfrak{F}| = |f_{\mathrm{S}}| + |f_{\mathrm{RC}}|$, or normalized for transcriptome size,

$$\mathrm{Resilience} = (|f_{\mathrm{S}}| + |f_{\mathrm{RC}}|)/N \tag{8}$$

such that $|f_{\mathrm{S}}|$ is sum of all similarity matches, $|f_{\mathrm{RC}}|$ is sum of all reverse complement interactions, and *N* is the total nucleotide size of the transcriptome.

5. Discussion

The concept of resilience is receiving increasing attention in chronic stress-related diseases. Resilience has been shown in clinical studies to play a protective role in patients with chronic conditions including osteoarthritis, breast and ovarian cancer, diabetes, and cardiovascular disease related to psychosocial dimensional levels. The purpose of this study is to explore the relationships between RNA-RNA interactions and to devise a measure of resilience at the cellular level.

5.1. Prospects, challenges, and limitations for resilience measure by variance in RNA-seq

Although research on empirical indicators of robustness and resilience is rudimentary, there is already a fast-growing body of engineering modeling as well as empirical work in ecology. Nonetheless, major challenges remain in developing robust procedures for assessment of the transcriptome. A goal of systems biology is to analyze large-scale multidomain networks to reveal relationships between network structures and their biological function. While generally, it is not feasible to visualize and understand whole networks, a common analysis is to partition the network into subnetworks responsible for specific biological functions. Since biological functions can be carried out by particular groups of molecules, dividing networks into naturally grouped clusters can help investigate the relationships between function and topology of system networks or reveal hidden knowledge behind them. The expression in Eq. (8) for resilience is a measure of the size of network interactions possible within a transcriptome.

5.2. Notion of the transcriptome as an information system

The body of this work considers the transcriptome as an information system modeling a dynamic system. A dynamic system is characterized by two concerns: the static structure and dynamic behavior. The structural elements of dynamic systems are those elements which may be identified from static snapshots of the problem space; while dynamic aspects involve those semantic elements of the system that exist over the time domain. While modeling the static aspects of an information system like RNA expression data, an understanding of the dynamic nature of information systems in the cell is low. Behavioral issues of large information systems are usually complex, consisting of many interactive sessions with the outside environment, tasks like coordination and collaboration among different entities. Dynamic systems can exhibit emergent properties that result from the dynamics, and which cannot be attributed to static structural factors. However, given any real world information system consisting of many multistream interactive processes, emergent properties are usually complex, without a common characteristic structure. Such emergent properties are beginning to be addressed with the transcriptome.

6. Conclusion

We show that the transcriptome can be modeled as an information system with emergent dynamic properties. The term *nebula* regulation is introduced to consider the regulatory effects

of the whole transcriptome acting locally through RNA-RNA interactions and shifts between accessible and inaccessible stretches of RNA sequence. Described as a network of interactions from semantic analysis of similarity and reverse complementarity, together with the size of a transcript, affect the diffusion of transcripts in a cell, and hence the distribution of RNAs. A measure to represent resilience is proposed as the sum of the component elements (similarity, reverse complementarity, and normalized by total nucleotides) of this transcriptome filter.

Acknowledgements

This work is supported in part by 8U54MD007588, G12MD007602, P50 HL117929, and P30 HL107238 grants from NIH/National Institute on Minority Health and Health Disparities. The content is solely the responsibility of the author and does not necessarily represent official views of the respective institutions.

Author details

William Seffens

Address all correspondence to: wseffens@msm.edu

Department of Physiology, Morehouse School of Medicine, Atlanta, GA, USA; Seftec, Inc., Atlanta, GA USA

References

[1] Tian B, Manley JL. Alternative polyadenylation of mRNA precursors. Nature Reviews Molecular Cell Biology. 2017;**18**(1):18–30

[2] Misteli T. Physiological importance of RNA and protein mobility in the cell nucleus. Histochemistry and Cell Biology. 2008;**129**(1):5–11. [Epub 2007 Nov 10]

[3] Trovato F, Tozzini V. Diffusion within the cytoplasm: A mesoscale model of interacting macromolecules. Biophysical Journal. 2014;**107**(11):2579–2591

[4] Ben-Ari Y, Brody Y, Kinor N, Mor A, Tsukamoto T, Spector D, Singer R, Shav-Tal Y. The life of an mRNA in space and time. Journal of Cell Science. 2010;**123**:1761–1774

[5] Hopper AK. Cellular dynamics of small RNAs. Critical Reviews in Biochemistry and Molecular Biology. 2006;**41**(1):3–19

[6] Jalali S, Bhartiya D, Lawani M, Sivasubbu S, Scaria V. Systematic transcriptome wide analysis of lncRNA-miRNA interactions. PLoS One. 2013;**8**:e53823

[7] Collins LJ. The RNA infrastructure: An introduction to ncRNA networks. Advances in Experimental Medicine and Biology. 2011;**722**:1–19. DOI: 10.1007/978-1-4614-0332-6_1

[8] Jeggari A, Marks DS, Larsson E. miRcode: A map of putative microRNA target sites in the long non-coding transcriptome. Bioinformatics. 2012;**28**:2062–2063. DOI: 10.1093/bioinformatics/bts344

[9] Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi P. A ceRNA hypothesis: The rosetta stone of a hidden RNA language? Cell. 2011;**146**:353–358

[10] Crooke ST, Wang S, Vickers TA, Shen W, Liang XH. Cellular uptake and trafficking of antisense oligonucleotides. Nature Biotechnology. 2017;**35**(3):230–237

[11] Bonnici V, Manca V. Informational laws of genome structures. Scientific Reports. 2016;**6**:28840. DOI: 10.1038/srep28840

[12] Shannon CE. A mathematical theory of communication. The Bell System Technical Journal. 1948;**27**:623–656

[13] Shannon CE. An algebra for theoretical genetics [PhD thesis]. Massachusetts Institute of Technology, 1940. MIT-THESES//1940–3 Online text at MIT. Contains a biography on pp. 64–65

[14] Lockhart E, Lucas M, Yoo J, Seffens W. Codon usage pattern detection in human, mouse, zebrafish and chicken genes using artificial neural networks. In: IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB); 2009; TN

[15] Wang X-Q, Abebe F, Seffens W. Dynamic system modeling the whole transcriptome in a eukaryotic cell. In: Proceedings of Dynamic Systems and Applications; 2015; Atlanta, GA. Dynamic Publishers, Inc.

[16] Savel'ev S, Marchesoni F, Taloni A, Nori F. Diffusion of interacting Brownian particles: Jamming and anomalous diffusion. Physical Review. 2006;**74**:021119

[17] Seffens W, Abebe F, Evans C, Wang X-Q. Spatial Partitioning of miRNAs is related to sequence similarity in overall transcriptome. International Journal of Molecular Sciences. 2016;**17**:830. DOI: 10.3390/ijms17060830

[18] Wang X. Composition of seed sequence is a major determinant of microRNA targeting patterns. Bioinformatics. 2014;**30**(10):1377–1383

[19] Regner B, Vucinic D, Domnisoru C, Bartol T, Hetzer M, Tartakovsky D, Sejnowski T. Anomalous diffusion of single particles in cytoplasm. Biophysical Journal. 2013;**104**: 1652–1660

[20] Mayorga M, Romero-Salazar L, Rubi J. Stochastic model for the dynamics of interacting Brownian particles. Physica. 2002;**307**:297–314

[21] Yeh I-C, Hummer G. Diffusion and electrophoretic mobility of single-stranded RNA from molecular dynamics simulations. Biophysical Journal. 2004;**86**(2):681–689

[22] Singh YH, Andrabi M, Kahali B, Ghosh C, Mizuguchi K, Kochetov A, Ahmad S. On nucleotide solvent accessibility in RNA structure. Gene. 2010;**463**:41–48

[23] Seffens W, Digby D. mRNAs have greater calculated folding free energies than shuffled or codon choice randomized sequences. Nucleic Acids Research. 1999;**27**:1578–1584

[24] Yoo J-K, Digby D, Davis A, and Seffens W. Whole transcriptome mRNA secondary structure analysis using distributed computation. In: Zhang Y-Q, Lin T, editors. Proceedings of International IEEE-Granular Computing. Atlanta, GA: Georgia State University; 2006. pp. 647–650

[25] Bernhart SH, Hofacker IL, Stadler PF. Local base pairing probabilities in large sequences. Bioinformatics. 2006;**22:**614–615

[26] Lange S, Maticzka D, Mohl M, Gagnon J, Brown C, Backofen R. Global or local? Predicting secondary structure and accessibility in mRNAs. Nucleic Acids Research. 2012;**40**(12):5215–5226

[27] Walia R, Caragea C, Lewis B, Towfic F, Terriblini M, El-Manzalawy Y, Dobbs D, Honavar V. Protein-RNA interface residue prediction using machine learning: an assessment of the state of the art. BMC Bioinformatics. 2012;**13**:89

[28] Hubbard S, Thornton JM. NACCESS. Department of Biochemistry and Molecular Biology, University College London; 1993

[29] Tang Y, Bouvier E, Kwok CK, Ding Y, Nekrutenko A, Bevilacqua PC, Assmann SM. Structure fold: Genome-wide RNA secondary structure mapping and reconstruction in vivo. Bioinformatics. 2015;**31**(16):2668–2675. DOI: 10.1093/bioinformatics/btv213

[30] Ding Y, Kwok CK, Tang Y, Bevilacqua PC, Assmann SM. Genome-wide profiling of in vivo RNA structure at single-nucleotide resolution using structure-seq. Nature Protocols. 2015;**10**(7):1050–1066. DOI: 10.1038/nprot.2015.064

[31] Ding Y, Tang Y, Kwok CK, Zhang Y, Bevilacqua PC, Assmann SM. In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. Nature. 2014;**505**(7485):696–700. DOI: 10.1038/nature12756

[32] Kwok CK, Tang Y, Assmann SM, Bevilacqua PC. The RNA structurome: Transcriptome-wide structure probing with next-generation sequencing. Trends in Biochemical Sciences. 2015;**40**(4):221–232. DOI: 10.1016/j.tibs.2015.02.005

[33] Chen KH, Boettiger AN, Moffitt JR, Wang S, Zhuang X. Spatially resolved, highly multiplexed RNA profiling in single cells. Science. 2015;**348**:aaa6090

[34] Friedel CC, Dolken L, Ruzsics Z, Koszinowski UH, Zimmer R. Conserved principles of mammalian transcriptional regulation revealed by RNA half-life. Nucleic Acids Research. 2009;**37**(17):e115

[35] Villarroya-Beltri C, Gutiérrez-Vázquez C, Sánchez-Cabo F, Pérez-Hernández D, Vázquez J, Martin-Cofreces N, Martinez-Herrera DJ, Pascual-Montano A, Mittelbrunn M, Sánchez-

Madrid F. Sumoylated hnRNPA2B1 controls the sorting of miRNAs into exosomes through binding to specific motifs. Nature Communications. 2013;**4**:2980. DOI: 10.1038/ncomms3980|

[36] Park CW, Zeng Y, Zhang X, Subramanian S, Steer C. Mature microRNAs identified in highly purified nuclei from HCT116 colon cancer cells. RNA Biology. 2010;7(5):606–614

[37] Huang X, Yuan T, Tschannen M, Sun Z, Jacob H, Du M, Liang M, Dittmar RL, Liu Y, Liang M, Kohli M, Thibodeau SN, Boardman L, Wang L. Characterization of human plasma-derived exosomal RNAs by deep sequencing. BMC Genomics. 2013;**14**:319

[38] Cheng L, Sharples RA, Scicluna BJ, Hill AF. Exosomes provide a protective and enriched source of miRNA for biomarker profiling compared to intracellular and cell-free blood. Journal of Extracellular Vesicles. 2014;**3**:23743

[39] Guduric-Fuchs J, O'Connor A, Camp B, O'Neill CL, Medina RJ, Simpson DA. Selective extracellular vesicle-mediated export of an overlapping set of microRNAs from multiple cell types. BMC Genomics. 2012;**13**:357

[40] Bullard J, Purdom E, Hansen K, Dudoit, S. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. BMC Bioinformatics. 2010;**11**:94

[41] Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biology. 2010;**11**:R106

[42] Dillies M-A, Rau A, Aubert J, Hennequet-Antier C, Jeanmougin M, Servant N, Keime C, Marot G, Castel D, Estelle J, Guernec G, Jagla B, Jouneau L, Laloe D, Gall CL, Schaeffer B, Crom SL, Guedj M, Jaffrezic F. A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. Briefings in Bioinformatics. 2013;**14**, 671–683

[43] Risso D, Ngai J, Speed TP, Dudoit S. Normalization of RNA-seq data using factor analysis of control genes or samples. Nature Biotechnology. 2014;**32**:896–902

[44] Wu D, Hu Y, Tong S, Gantier M. The use of miRNA microarrays for the analysis of cancer samples with global miRNA decrease. RNA. 2013;**19**:876–888

[45] Urbanek M, Nawrocka A, Krzyzosiak W. Small RNA detection by in situ hybridization methods. International Journal of Molecular Sciences. 2015;**16**:13259–13286

[46] Kloosterman WP, Wienholds E, de Bruijn E, Kauppinen S, Plasterk RH. In situ detection of miRNAs in animal embryos using LNA-modified oligonucleotide probes. Nature Methods. 2006;**3**:27–29

[47] Soe MJ, Moller T, Dufva M, Holmstrom K. A sensitive alternative for microRNA in situ hybridizations using probes of 2′-O-methyl RNA + LNA. Journal of Histochemistry and Cytochemistry. 2011;**59**:661–672

[48] Majlessi M, Nelson NC, Becker MM. Advantages of 2′-O-methyl oligoribonucleotide probes for detecting RNA targets. Nucleic Acids Research. 1998;**26**:2224–2229

[49] Beliveau BJ, Joyce EF, Apostolopoulos N, Yilmaz F, Fonseka CY, McCole RB, Chang Y, Li JB, Senaratne TN, Williams BR, Rouillard J-M, Wu C-t. Versatile design and synthesis platform for visualizing genomes with Oligopaint FISH probes. Proceedings of the National Academy of Sciences of the United States of America. 2012;**109**:21301–21306

[50] Coate J, Doyle J. Chromosome size: Are we getting the message?. Chromosoma. 2015;**124**: 27–43

[51] Scheffer M, Carpenter SR, Lenton TM, Bascompte J, Brock W, Dakos V, van de Koppel J, van de Leeput IA, Levin SA, van Nes EH, Pascual M, Vandermeer J. Anticipating critical transitions. Science. 2012;**338**(6105):344–348

[52] Clark J. Functionality, Complexity, and Approaches to Assessment of Resilience Under Constrained Energy and Information. Ohio: Air Force Institute of Technology, Wright-Patterson AFB; 2015. AFIT-ENV-DS-15-M-159. Accession number ADA619053

[53] Lloyd S, Pagels H. Complexity as thermodynamic depth. Annals of Physics. 1988;**188**: 186–213

[54] Corning PA. Complexity is just a word!. Technological Forecasting and Social Change. 1998;**58**:1–4

[55] Senge P. The Fifth Discipline: The Art and Practice of the Learning Organization. New York: Doubleday; 1990

[56] Sterman J. Business Dynamics: Systems Thinking and Modeling for a Complex World. Boston: McGraw-Hill, Irwin; 2000

[57] Crutchfield JP, Shalizi CR. Thermodynamic depth of causal states: When paddling around in Occam's pool shallowness is a virtue. Physical Review E. 1999;**59**(1):275–283

[58] Li W. On the relationship between complexity and entropy for Markov chains and regular languages. Complexity. 1991;**5**:381–399

[59] Bar-Yam Y. Multiscale complexity/entropy. Advances in Complex Systems. 2004;**7**:47–63

[60] Chaisson EJ. Energy rate density as a complexity metric and evolutionary driver. Complexity. 2011;**16**:27–40

[61] INCOSE. INCOSE Resilient Systems Working Group (RSWG) Charter [Internet]. 2011. Available from: URL http://www.incose.org/about/organization/pdf/RSWG_Charter.pdf

IntechOpen