Modeling Patterns of microRNA:mRNA Regulation Through Utilization of Cryptographic Algorithms

Harry C. Shaw
NASA/Goddard Space Flight Center
Telecommunication Networks & Technology Branch
Greenbelt, MD, USA
Harry.c.shaw@nasa.gov

***Abstract*—The properties of microRNA mediated messenger RNA regulation (miRNA:mRNA) of expression are explored by use of an experimental cryptographic representation of miRNA and mRNA structures. The cryptographic approach allows differentiation of identical miRNA:mRNA bonding sequences. Hamming coding, Huffman coding and Rivest, Shamir, Adleman (RSA) encryption techniques are incorporated in the modeling process. Regulation is evaluated as a function of vector projections of mRNA sequences onto a miRNA. The model has a calibration function based upon the use of vector projections onto miRNA complementary and anti-complimentary sequences.**

Keywords-miRNA, mRNA; secondary structure; RSA algorithm; cryptography; hash code

I. INTRODUCTION

A model of microRNA:mRNA regulation and bonding from the standpoint of a cryptographic problem is being developed. miRNA:mRNA bonding exhibits a wide range of regulation. A single miRNA can regulate expression of many mRNAs and a single mRNA can be regulated by many miRNA. The mechanisms that determine specificity are not fully understood. The methodology of the work described in this paper is to treat miRNA:mRNA interaction as a short hash code authentication problem wherein multiple messages with different meanings may generate similar or identical codes. The underlying concept is that miRNA sequences can be treated as short messages. These messages appear to be identical, but are not. In fact, they are uniquely authenticated for a given mRNA target under conditions and conformations that are not totally apparent or understood. The challenge of this experimental modeling approach is that no assumptions are made about the seed and target sequences such as assumptions about evolutionary conservation of sequences, length of seed sequence, number and location of Watson-Crick pairs within seed:target locus, or any of the other commonly used (and often well-validated) assumptions. The goal is to determine if miRNA:mRNA regulation can be modeled within the context of a cryptographic approach.

The paper is structured as follows: A description of the motivation for the research and the potential benefits is provided as well as the challenges in analyzing microRNA (miRNA) mediated regulation. This is followed by a section summarizing miRNA regulation pathway. This is followed by a section summarizing the steps in coding and modeling of the miRNA:mRNA problem. This is followed by a detailed discussion of the coding and modeling of the miRNA:mRNA problem and the resulting output data. The paper concludes with a discussion of model validation, future work and conclusions.

A. Potential benefits to using cryptography to analyze problems in molecular biology

The process of encryption and decryption can be used to accommodate uncertainty and lack of complete information with regards to biochemical pathways. Unknown information that would be essential in creating an accurate physics-based model can be accommodated in a cryptographic model. The cryptographic process allows for wide differentiation between objects that appear to be similar or identical. In the case of microRNA, two similar sequences of 22 bases may have drastically different patterns of post-transcriptional regulation of mRNA expression. By using techniques from coding theory, identical genomic or proteomic sequences can be differentiated by specific coding for secondary structure, tertiary structure or other differentiating characteristics. The cryptographic process permits a hierarchy of coding such that an iterative modeling process can be performed.

B. Advantages of using a cryptography-based process

One major advantage of this approach is that a properly operating authentication process incorporates all of the behavior of the system, including behavior that is currently not well understood or modeled. This includes lack of accurate knowledge of all relevant steric and electrostatic forces, concentration dependencies, and understanding of all relevant molecular interactions (protein-protein, nucleic acid-nucleic acid, protein-nucleic acid, etc.). In this paper, a specific implementation of the protocol will be described that only includes the miRNA to mRNA interaction. New information about a process can be incorporated by expanding the model. The extensible nature of the model will allow for adding details such as RNA Induced Silencing Complex (RISC)-miRNA interaction coding as a future enhancement.

 ISBN: 978-1-61208-260-8

II. CHALLENGES IN ANALYZING MICRORNA MEDIATED REGULATION

It is known that sequence information alone is insufficient to predict miRNA:mRNA regulation. A single miRNA can regulate numerous mRNA. A single mRNA can be regulated by numerous miRNA. The presence of the seed-target matches does not guarantee downregulation of mRNA. There are structural considerations such as single or double strandedness of the miRNA, and protein bonding domain interaction with miRNA are factors in successful miRNA translational regulation. Both cis- and trans-regulatory effects exist in miRNA regulation as well as cooperative bonding effects on multiple miRNAs amplifying the regulation of a mRNA target. Location of the target within the 3' Untranslated Region (UTR) of the mRNA is a factor, but the target sequence can appear with an open reading frame, and appearance of the target within the 3' UTR is not a necessary and sufficient condition for mRNA regulation.

A. miRNA Regulation Pathway

The miRNA gene is transcribed by RNA Polymerase II to produce a primary miRNA transcript (pri-miRNA). Due to the charged nature of RNA, it is never found free and is always complexed with a protein. The pri-miRNA transcript is processed by a complex of DROSHA and DGC8 proteins to a ds-miRNA hairpin pre-miRNA transcript. The pre-miRNA transcript is transported from the nucleus to the cytoplasm by Exportin 5 (XPO5) that can only bind pre-miRNA in the presence of the RAN-GTPase cofactor [1]. TRBP recruits the RNAse III DICER complex [2] and DICER cleaves the transcript into a ds-miRNA guide and passenger strand. The guide strand complexes with Ago2 and other proteins required for miRNA silencing, e.g., GW182) to form a RNA Induced Silencing Complex. RISC and mRNA targets that associate with sufficient complementarity and energetically favorable structures downregulate (or upregulate [3]) mRNA translational expression .

B. miRNA:mRNA coding and modeling process

Each miRNA and mRNA sequence is coded into a 15-bit vectors that represent the nucleotide sequence and assignment of secondary structure. There are 22 rows; one row for every nucleotide in the sequence. The 15 bit binary coded vectors are compressed into two sets of integer vectors (one integer equaling the sum of the 7 most significant bits and the other equaling the sum of the 8 least significant bits) for each miRNA, mRNA, a fully complementary miRNA sequence, e.g., let7d_comp) and a fully anti-complementary, e.g., let7d_anti). The following basic steps are performed:

- Each vector is encrypted using a Rivest, Shamir and Adleman (RSA algorithm) key pair (e,n).
- The encrypted vectors are decrypted with the correct private key and 12 other off-nominal keys in a process called detuning.
- For each 7-bit integer vector and each 8-bit integer vector, a linear vector projection is performed on mRNA vector to the miRNA vectors, the miRNA_comp vectors and miRNA _anti vectors.
- The projections of mRNA vectors onto the miRNA vector are compared to the projection of the mRNA vectors onto the miRNA_comp vector and the mRNA vector onto the miRNA_anti vector
- A regulation of expression score is computed on the basis of these comparisons.

III. MATERIALS AND METHODS

A. MATLAB Prototype

A prototype model built on the MATLAB platform was developed for this research. In this particular model, only nucleic acid interactions at the primary and secondary structure levels are captured. The prototype utilizes the RSA algorithm for encryption. Although the algorithm was not intended for this application, it performs well enough to demonstrate the principles behind the operation of the model. Future versions of the model will utilize a new key pairing algorithm that has been optimized for this application.

B. Coding of RNA bases

A binary representation of the four RNA bases is generated from a series of four 11-bit vectors. Each base is represented by a (15, 11) Hamming code. A (15, 11) hamming code codes a 11 bit data field in 15 bits and provides 4 check bits.

C. Coding of Secondary Structure Dictionary

RNA folding is necessary for proper functioning of biological activities [4]. The model requires that a secondary structure code be applied to every sequence (mRNA or miRNA), even if the secondary structure under evaluation is modeled as linear, single stranded. The secondary structure coding allows two identical sequences to be differentiated. Therefore the secondary sequence coding provides for an authentication capability. The prototype model has 5 structural categories (classes of symbols) each with a probability mass function that is user-defined. Table I summarizes the categories. The location of each base within the sequence has a probability of being in a given structural category at its location in the nucleotide sequence. The base at that location is assigned a value from the Huffman code dictionary. The dictionary is applied via a language that provides shorthand for performing the secondary structure coding. There are 409 codes assigned to each category.

TABLE I. SECONDARY STRUCTURE CATEGORIES

Alpha	Description
X	Double strand, unpaired bulge
P	Double strand, Watson-Crick pair
W	Wobble pair
L	Loop
S	Ss, unpaired

 ISBN: 978-1-61208-260-8

D. Secondary Structure Coding

A Huffman code generator produces a set of 2K x 15 bit secondary structure code dictionary to be applied as a hash code to each base in the RNA sequence. A Huffman code dictionary, which associates each data symbol with a codeword, has the property that no codeword in the dictionary is a prefix of any other codeword in the dictionary. The statistical frequency of a given secondary structure can be correlated to its Huffman code. Let *N* equal a finite field of a secondary structure space partitioned into sets of five members, with q = 409:

$$N_k = \{X_i, P_i, W_i, L_i, S_i\},\ 1 \le i \le q,\ 1 \le k \le 22, \quad (1)$$

$$X = \{SS_1^x, SS_2^x, \ldots, SS_q^x\}, \quad (2)$$

$$P = \{SS_1^p, SS_2^p, \ldots, SS_q^p\}, \quad (3)$$

$$W = \{SS_1^w, SS_2^w, \ldots, SS_q^w\}, \quad (4)$$

$$L = \{SS_1^l, SS_2^l, \ldots, SS_q^l\}, \quad (5)$$

$$S = \{SS_1^s, SS_2^s, \ldots, SS_q^s\} \quad (6)$$

Every base in the sequence falls into one of the above sets. The hamming distance, d, between nearest neighbors in any given set is:

$$d(SS_i^x, SS_{i+1}^x) = d(SS_i^p, SS_{i+1}^p) = 3 \quad (7)$$

$$d(SS_i^w, SS_{i+1}^w) = d(SS_i^s, SS_{i+1}^s) = 3 \quad (8)$$

$$d(SS_1^l, SS_{i+1}^l) = 6 \quad (9)$$

RNA structures have been experimentally shown to have a diverse free energy landscape and structural diversity [5]. The use of a space of Hamming codes to represent the multiplicity of energy and mechanical constraints on each molecular structure permits a dynamic modeling approach in which a progression of codes can be modeled to represent the diversity in the free energy states and structure. It remains to be seen if this approach is better than a conventional molecular simulations that rely on two-body additive force field equations or assumptions that involve finding the lowest energy conformation of a RNA sequence.

The use of Hamming distances as a metric to evaluate RNA secondary structures and folding has been utilized other modeling approaches [6], however the taxonomy of the secondary structures being utilized in this approach is very different as well as the method of developing the secondary structure codes.

E. Encryption, Decryption and Detuning

The Asymmetric Encryption Algorithm was developed by Rivest, Shamir and Adleman from MIT to facilitate development public key infrastructure based security. Originally a classified algorithm, it passed into the public domain in September 2000 [7]. The RSA algorithm is a convenient mechanism for generating public/private key pairs for evaluation of secondary structure hash codes. In this model three sets of RSA public/private key pairs are used to eliminate any possible bias that might be introduced by a single set of keys. The vectors representing each sequence will be decrypted with the corresponding private key and with 12 other prime numbers in a process called detuning. The operation of decryption with keys of successively greater distances from the primary decryption key should be traceable to thermodynamic states of greater energy, i.e., lower stability and lower probability of occurrence) for a given sequence. For encryption key E = 5, the decryption keys of 53, 47, 43, 41, 37, 31, 29, 23, 19, 17, 13, 11, 7 are used. The key pair of (5, 299) & (53,299) satisfies the RSA algorithm; the other decryption keys are for detuning. Similarly, for E = (1657, 24811) & D = (73, 24811), D = 73, 1559, 1483, 1367, 1013, 1009, 691, 601, 509, 443, 439, 349, 337 are used. For E = (89, 5629) & D = (233, 5629), D = 233, 229, 227, 223, 157, 137, 109, 101, 97, 61, 53, 47, 43 are used. The vector is encrypted with three RSA key pairs, in this case (5, 53), (1657, 24811) and (89, 5629).

The data is formed by the set of error vectors of the decrypted vector projected onto the reference vector. The RSA algorithm uses a public key (e, x) and a private key (d, x). There exist two prime numbers, p and q, $p \neq q$. The following relationships apply:

$$x = p * q, \quad (10)$$

$$\phi(x) = (p-1) * (q-1), \quad (11)$$

$$e < \phi(x) \text{ and relatively prime to } \phi(x), \quad (12)$$

$$de = 1 \bmod \phi(x)\ . \quad (13)$$

From the calculated key pairs, additional *d's* that do not meet the $de=1 \bmod \phi(x)$ are utilized as detuning keys. The resulting encrypted vector is decrypted with the decryption key specified by the algorithm and 12 additional keys that are different from the decryption key. The decryption key pairs for (5, 53) are (53, 299), (47,299), (43, 299), (41,299), (37,299), (31,299), (29,299), (23,299), (19,299), (17,299), (13,299), (11,299), and (7,299). The decryption key pairs

 ISBN: 978-1-61208-260-8

for (1657,24811) are (73,24811), (1559,24811), (1483,24811), (1367,24811), (1013,24811), (1009, 24811), (691, 24811), (601,24811), (509,24811), (443,24811), (439, 24811), (349,24811) and (337,24811). The decryption key pairs for (89, 5629) are (233,5629), (229, 5629), (227,5629), (223,5629),(157,5629), (137,5629), (109,5629), (101,5629), (97,5629), (61, 5629), (53,5629), (47,5629) and (43, 5629). The data is formed by the set of error vectors described in the following paragraph.

The working hypothesis is that the (d_m, x), (m=1,2,..,13) combination represents a projection of miRNA:mRNA with successively lower probabilities of occurrence as d_m diverges from d_1, e.g., higher minimum free energy for a successful regulation event (either up or down regulation). Each decryption key pair is associated with a probability mass function (pmf), P:

$$P = \{p_1,p_2,\ldots p_m\} = \{ 2^{-1},2^{-2},\ldots,2^{-13}\},\ m=1,..,13 \quad (14)$$

$$\sum_{m}^{13} p_m \cong 1 \ . \quad (15)$$

F. Projection of vector column spaces and generation of error vectors

Each sequence is now represented by a series of decrypted column vectors. The modeling of miRNA association to a mRNA is simulated by the projection of the column space of a decrypted mRNA onto the column space of the decrypted miRNA under evaluation. It is the error vector of the projections that form the data, not the projection vectors. Each representation of a sequence consists of a column space of two vectors, one derived from the 7-bit portion and one derived from the 8-bit portion. Fig. 1 details the vector projection algorithm. Each mRNA:miRNA simulation has 6 sets of error vectors calculated from the error vectors of projections of:

a) 7-bit mRNA on miRNA for all 3 (e,x), (d_m,,x) combinations

b) 8-bit mRNA on miRNA for all 3 (e,x), (d_m,,x) combinations

c) 7-bit mRNA on miRNA complementary sequence for all 3 (e,x), (d_m,,x) combinations

d) 8-bit mRNA on miRNA complementary sequence for all 3 (e,x), (d_m,,x) combinations

e) 7-bit mRNA on miRNA anti-complementary sequence for all 3 (e,x), (d_m,,x) combinations

f) 8-bit mRNA on miRNA anti-complementary sequence for all 3 (e,x), (d_m,,x) combinations

The miRNA complementary sequence and anti-complementary sequence used in this paper is shown in table IV.

$$c_n = \sqrt{E_n^2 + d_m^{\ 2}}, n = 1,2,3 \quad (16)$$

$$\sin\theta_n = E_n / c_n, n = 1,2 \quad \sin\rho = E_3 / c_3 \quad (17)$$

An example of the origin of the variables is shown in table III. E_1, E_2, and E_3 are the value of the error projection for mRNA GATM at the 22nd base position in (starting at position 289 referenced to the 5'end) when projected onto let-7d miRNA, let-7d miRNA complement and let-7d miRNA anticomplement at the same position. The projection is for the encryption key pair (89,5629) and the decryption key pair (43,5629). This leads to geometric relationship shown in Fig. 2.

G. Significance of the relationship between θ and ϕ

θ represents the angular distance between the mRNA:miRNA projection to a perfectly complimentary sequence projected on the miRNA (let-7d comp:let-7d). The smaller the angular distance the greater the similarity between the state of basepairing (bp) in mRNA:miRNA bp and the corresponding mRNA_comp:miRNA. ϕ represents the angular distance between the mRNA:miRNA projection to a perfectly anti-complimentary sequence projected on the miRNA (let-7d anti:let-7d). The smaller the angular distance the greater the similarity between the error vector projection in mRNA:miRNA and the corresponding mRNA_anti:miRNA. It is postulated that there exists an optimum θ and ϕ that represent a maximization of the probability of stabilizing the sequences such that regulation (either up or down) is maximized. The scoring criteria is to retain all scores that satisfy $\sin\theta < r*\sin\phi$, where $r \leq 1$. For this paper, r = 1. The following scoring logic applies:

$$s_{j,m} = [(\sin\theta_{j,m})/(\sin\phi_{j,m})]*p_m \ , \quad (18)$$

where j = {1,…,22} nucleotides and m = {1,…13} decryption key pairs.

IV. SCORING

The evaluation of six mRNAs against a single miRNA yields a dataset of up to 20,592 values of ϕ and θ. Large datasets such as this are good candidates for single value decomposition (SVD) analysis is used to extract the principal values and score the data. SVD is widely used to identify the significant elements in large data sets [8]. It is widely used in analysis of gene expression. The entire set for each mRNA:miRNA combination is combined into matrix of the form $X = USV^T$. The score is the product of the most significant members of *U* and *S*. The software allows the user to bound the maximum and minimum scores. Widening the scoring range expands the predicted levels of up or down regulation. The scoring range in table II was (-0.5, +0.5). The score is a direct correlation to the predicted level of post-transcriptional regulation, B_t, by:

$$Score_t = (U_{1,1} * S_{1,1}),\ B_t = 2^{Score_j} \quad (19)$$

 ISBN: 978-1-61208-260-8

There are two scores, t = 1 and t = 2 representing the 7-bit and 8-bit integer scores respectively. The two scores provide an upper and lower bound of modeled regulation. The scoring range can span upregulation and downregulation predicted outcomes.

A. *mirSVR Scoring.*

The mirSVR algorithm is a powerful scoring system with a high degree of "ground truth" in its scoring methodology [9]. mirSVR scoring has been calibrated to be linear with log expression change. The mirSVR algorithm was trained on data from mRNA expression data from a panel of microRNA transfection experiments. Target sites are represented by miRNA:mRNA features, and local and global contextual features such as the composition of AU flanking sequences around the presumed mRNA target site. mirSVR uses features derived from the miRanda-predicted miRNA:mRNA sites, the extent of 3' binding and local features such as AU composition flanking the target site and miRSVR defined secondary structure accessibility score. Other global features of the mirSVR scoring include length of UTR, relative position of target site from UTR ends, and conservation level of the block containing the target site. The downregulation scores from mirSVR correlate linearly with the extent of downregulation. Genes with multiple target sites can be scored by addition of the individual target scores.

B. *Cryptographic Scoring*

The cryptographic model provides a range of two scores. In this comparison, a let-7d sequence and mRNA sequences (from microRNA.org, with modifications to fit the 22 base sequence length model requirement) are programmed as single stranded RNA molecules without any internal Watson-Crick or G:U wobble pairing. The error vectors for the mRNA projected onto the let-7d vector are scored as previously described. Table II summarizes some early results. The scores for multiple target sites can also be summed to provide total mRNA regulation change.

TABLE II. SCORING EXAMPLES

mRNA	Test sequence	start position from 5' end	mirSVR	Cryptographic high score	Cryptographic low score
C14ORF28	UUUUUUAUUUAUAUGUACCUCA	261	-0.3232	-0.21855	-0.04203
C14ORF28	UGUAUUUCUUUGCCCUACCUCA	423	-0.859	-0.16676	-0.13331
C14ORF28	ACAAUGGAACUUACCUACCUCA	1596	-0.7795	-0.05975	-0.0286
C14ORF28	AAAAAAACAUUUUUCUACCUCU	1676	-1.0895	-0.13636	0.031029
DNA2	CUAUCCUCCCUUACUAUCCUCC	894	-1.2921	-0.07659	-0.05211
HMGA2	UAAAAUUUUUAUUUCUACCUCA	8	-0.3281	0.111411	-0.31562
HMGA2	CAACGUUCGAUUUUCUACCUCA	1244	-0.994	0.083544	0.070764
HMGA2	CACUACUCAAAUUACUACCUCU	1604	-0.0989	0.014682	-0.0218
HMGA2	UACCCUCCAAGUCUGUACCUCA	1655	-0.2446	0.134442	0.05026
HMGA2	GACUUUGCAAAGACCUACCUCC	2213	-0.002	0.411992	0.118015
HMGA2	GUUUCAAAGGCCACAUACCUCU	2507	-0.8811	0.040069	0.001315
HMGA2	AUCAAAACACACUACUACCUCU	2526	-0.1048	0.156676	-0.24195
SMARCAD1	UCUUAAGUCCCAGUAUACCUCA	1498	-0.6971	0.005753	-0.08738
FIGNL2	UCAUGUGUUAAAUACUACCUCC	2227	-0.1242	-0.1562	-0.0833
CTPS2	GCCUAGGUGGGCACCUACCUCA	1315	-0.9344	0.031346	-0.021
FIGN	CAAAACCCAUACUACUACCUCA	635	-0.1957	-0.32631	-0.09044
FIGN	UUGUGAUUUGUACAGUACCUCA	724	-0.3286	-0.21063	0.057325

V. VALIDATION PLAN AND FUTURE WORK

The concepts in this paper require a high degree of validation. The goal is to team with a partner with the laboratory capabilities in running miRNA:mRNA regulation assays to:

- Validate the results with through carefully selected and executed assays.
- Evaluate all aspects of the coding and scoring algorithms.
- Determine the limits of the model to provide an accurate up or down regulation assessment.
- Increase the fidelity of the model by adding ribonucleoprotein-RNA interactions and RISC-RNA interactions.
- Development of a rule set to facilitate automated coding and analysis of miRNA:mRNA model interactions.

VI. CONCLUSIONS

A methodology for using tools from cryptography and information theory to analyze the regulation of mRNA:mRNA interactions has been presented. The work in progress preliminary model results support the concept of operations such that additional work will be conducted to optimize the model and validate the model results.

ACKNOWLEDGMENT

This paper is developed from a dissertation submitted to The George Washington University in partial fulfillment of the requirements for the Ph.D. degree.

 ISBN: 978-1-61208-260-8

REFERENCES

[1] R. Yi, Y. Qin, I. G. Macara, and B. R. Cullen, "Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs", Genes & Development, vol. 17, Oct. 2003, pp.3011–3016, doi/10.1101/gad.1158803.

[2] T. P. Chendrimada, R. I. Gregory, E. Kumaraswamy, J. Norman, N. Cooch, K. Nishikura and R. Shiekhattar, "TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing", Nature, vol. 436, Aug. 2005, pp. 740-744, doi:10.1038/nature03868

[3] W. Ye, F. Qin, J. Zhang, R. Luo, and H. Chen, "Atomistic Mechanism of MicroRNA Translation Upregulation via Molecular Dynamics Simulations", PLoS One, vol. 7, Aug. 2012, ,issue 8, pp. 1-11, doi: 10.1371/ journal.pone.0043788

[4] E. Mahen, P. Y. Watson, J. W. Cottrell, and M. J. Fedor,"mRNA Secondary Structures fold Sequentially but exchange rapidly in vivo", PLoS Biology, vol. 8, Feb. 2010, issue 2, pp.1-14, doi:10.1371/journal.pbio.1000307

[5] M. Khalili, W. Kasprzak, M. H. Farris, J. Arroyo, and B. A. Shapiro, "Thermodynamic Studies of RNA Secondary Structure with Genetic Algorithm", Mitre Corporation Report 09-5182, May 2010.

[6] P. Schuster, W. Fontana, P. F. Stadler, I. L. Hofacker, "From Sequences to Shapes and Back: A Case Study in RNA Secondary Structures", Proceedings: Biological Sciences, vol. 255, Mar. 1994, no. 1344, pp. 279-284.

[7] W. Stallings, Cryptography and Network Security, 4th edition, Upper Saddle River, NJ: Pearson Prentice-Hall, 2006

[8] M. E. Wall, A. Rechtsteiner, and L. M. Rocha, "Singular value decomposition and principal component analysis", A Practical Approach to Microarray Data Analysis" D.P. Berrar, W. Dubitzky, M. Granzow, eds. pp. 91-109, LANL LA-UR-02-4001, 2003.

[9] D. Betel, A. Koppal, P. Agius, C. Sander, and C. Leslie, "Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites", Genome Biology, vol. 11:R90, Aug. 2010, pp. 1-14, doi:10.1186/gb-2010-11-8-r90

TABLE III. EXAMPLE OF THE ORIGIN OF THE DATA FOR VARIABLES, E_1, E_2, E_3, d_m

E_3 E_2 E_1

e	d1	d2	d3	d4	d5	d6	d7	d8	d9	d10	d11	d12	d13
89	233	229	227	223	157	137	109	101	97	61	53	47	43
Error vector value at base position 22 in GATM position 289 from the 5'end projected onto let-7d position 22													
	3058.9	3155.6	314.4	711.9	-2124.7	1006.4	-2633.4	-240.2	-1091.3	-147.3	43.0	-1209.9	2230.3
Error vector value for base position 22 of let-7d projected onto let-7d_comp position 22													
	2032.2	-1041.8	-593.9	-1610.5	-2840.3	-463.8	-3524.0	-1752.7	-1529.4	-2837.9	-584.7	-2693.8	-1927.1
Error vector value for base position 22 of let-7d projected onto let-7d_anti position 22													
	-1782.6	2758.5	4605.2	-1191.0	866.5	-273.8	-318.5	-650.2	2316.7	-100.2	-337.0	-889.3	-1566.9

d_{13} in (43,5629)

c_1=1542.3, c_2=459.77, c_3=730.4

TABLE IV. MIRNA CALIBRATION VECTORS

let-7d	UUGAUACGUUGGAUGAUGGAGA
let-7d comp	AACUAUGCAACCUACUACCUCU
let-7d	UUGAUACGUUGGAUGAUGGAGA
let-7d anti	CCAGCGUACCAAGCAGCAAGAG

 ISBN: 978-1-61208-260-8

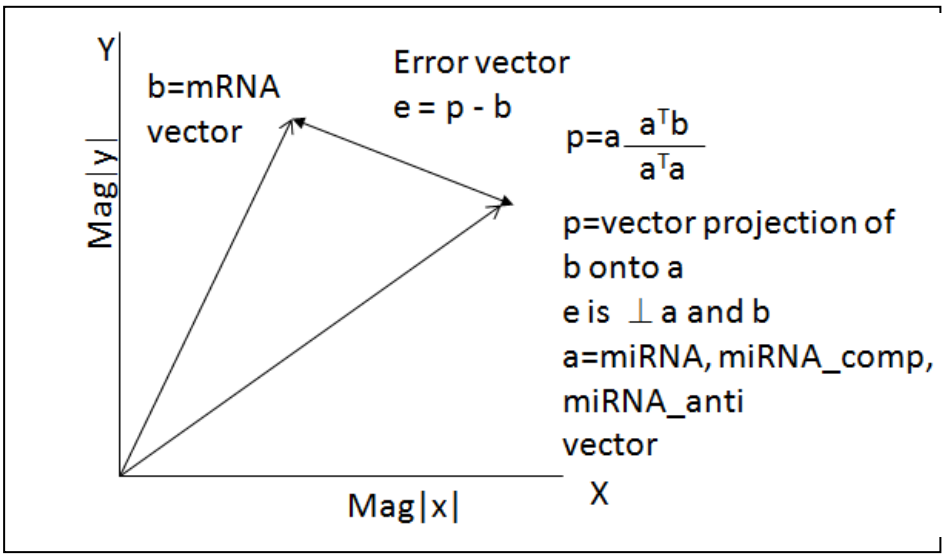

Figure 1. Vector Projection and Error Vector determination algorithms

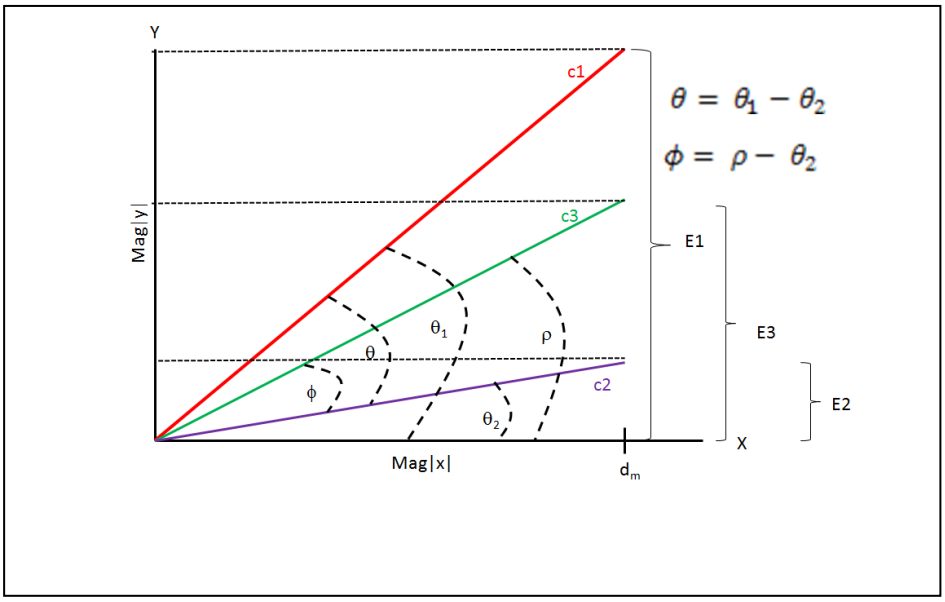

Figure 2. Geometric Relationships in (21) and (22) for example case.

 ISBN: 978-1-61208-260-8