Study of miRNA Let-7 Involvement in Breast Cancer through the Light of Graph Theory and Canalizing Function

Antara Sengupta
MCKV Institute of Engineering
243.G.T. Road(North), Liluah
Howrah-711204, West Bengal, India

Camellia Ray
Applied Statistics Unit, Indian Statistical Institute, Calcutta
700108, West Bengal, India

Pabitra Pal Choudhury
Applied Statistics Unit, Indian Statistical Institute, Calcutta
700108, West Bengal, India

ABSTRACT
miRNAs are small, non-coding RNAs and of 21-25 nucleotide in size. They negatively regulate protein-coding mRNAs at the post-transcriptional level. miRNAs play very important roles in various biological processes like cell growth, differentiation, and development. Abnormal expressions of miRNAs can lead various diseases including cancer and therefore miRNAs are considered as strong diagnostic and therapeutic candidates for the treatment of human diseases. The let-7 "lethal-7" is a microRNA and has active participation in Breast cancer. Not only that but also let-7 affects self-renewal ability of cancer cells. So it may be novel idea to make cancer therapy by targeting let-7 in breast cancer patients. In this paper we have tried to study the Gene Regulatory Network (GRN) of Let-7 in Breast Cancer through the light of Graph theory and Canalizing function.

General Terms
Cancer Genomics, Mathematical Modeling, Graph Theory.

Keywords

1. INTRODUCTION
Breast cancer is actually breast tumor which develops from epithelial cells that develop neoplasia in breast tissue [19] and so have carcinoma. Some specific genes are there which are responsible for breast cancer. Although lots of reasons are there, but role of some specific miRNAs are unavoidable [14][16][17]. Recently, it has been proposed that the presence of genetic variations in microRNA genes, their biogenesis pathway and target binding sites affect the miRNA processing machinery and targeting, and have a significant genetic effect. Although like other types of cancers breast cancer also develops stepwise but it majorly depends on the age and genetic susceptibility [18].

The let-7 family consists of 12 miRNAs. Those are let-7-a1, let-7-a2, let-7-a3, let-7-b, let-7-c, let-7-d, let-7-e, let-7-f1, let-7-f2, let-7-g, let-7-h and miR-98 [1, 2]. A thorough biological study has investigated the expression of let-7 in multiple breast cancer cell lines [3]. Further research investigated that Let-7 has very reduced expression in BT-IC, but the level of let-7 can increase with differentiation[4] and when the expression of Let-7 lost at that time the restoration of its expression may help in cancer therapy [17]. So, greater understanding of the functionality of miRNA Let-7 can make it easier to fight or prevent many cancers[13] as miRNAs are now have been made on of the leading causes of cancers[14].

Gene Regulatory Networks (GRNs) are nothing but interactions between large numbers of genes and their regulators which have been mapped onto graphical interpretations that are used to visualize the regulatory relationships [15].

Boolean networks describe the state of genes with binary (ON/OFF) variables, which has dynamic behavior of each variable and is controlled by Boolean function [9]. Although Boolean networks allow the analysis of the dynamics of the gene regulatory networks, they ignore the effect of genes at intermediate levels. Boolean networks can be used as models for genetic control in cells, where each gene represents the node, and each node has two states ON (activation) or OFF (inhibition) as during regulation of functional states the cell exhibits switch-like behavior, which ensures the movement of cell from one state to another [10]. The Boolean networks have ability to contain very large number of nodes, and for multiple Transcription Factors, the combined interaction can be like logic gates.

In Boolean network canalizing functions are mainly used for representation of the nodes because of variety of reasons. It has been seen in [5, 6] that Boolean networks with canalizing rules show ordered behavior. In nested canalizing functions variables in a given order dominate the function and hence they have been proposed as a frame work for network modeling [7]. Boolean network is used to represent the gene network but sometimes it becomes very much chaotic, canalizing function can be used to reduce the chaotic behavior of the Boolean network in dynamical system and also to bring the stability in the gene regulatory network [8].

Mathematical modeling of Gene Regulatory network is not a new concept. Several mathematical models are there, each of which are developed for some specific purposes. It is to be noted that no such quantitative approaches have been yet applied, but several biological studies have been made which proves that miRNA Let-7 is actively participating in Breast Cancer. According to the cancer stem cell hypothesis, T-ICs are responsible for the initiation, progression, metastasis and resistance to therapy. Moreover studies have shown that often silence of LET-7 in cells exhibits a mesenchymal phenotype and represent cancers in more advanced stages [12]. Based on those studies and hypothesis, in this paper we have tried to build a stable structure which can study the involvement of miRNA Let-7 in breast Cancer in different stages through GRN.
In the following in section 2 we enumerate or methodology and results. Next section 3 contains the conclusion and further research.

2. METHOD AND RESULTS

The proposed study will go through the following step by step procedures:-

Firstly, collection of dataset i.e., Gene Regulatory Network (GRN), secondly, conversion of GRN in to Directed Acyclic Graph (DAG), thirdly, directed hyper graph representation of the Gene Regulatory Network with cooperative interactions, and finally, representation of gene regulatory network through the light of canalizing function.

2.1 Collection of Dataset i.e., Gene Regulatory Network

A Gene Regulatory Network is being collected, where the GRN is clearly demonstrating the involvement of the miRNA Let-7 with several genes in various stages like Cell Cycle, Cell Division, DNA Replication, Angiogenesis, Apoptosis and Cell Proliferation of Breast cancer. The GRN is introduced by Barh D. et al [9] (see Figure 1).

![Figure 1: GRN of miRNA Let-7 with different target genes.](image)

2.2 Conversion of GRN in to Directed Acyclic Graph (DAG)

To model a genetic regulatory network the most suitable and significant way is to view it as a directed graph(G) which is defined as a tuple (V,E), with V a set of vertices and E a set of edges. A directed edge is a tuple (i, j) of vertices, where i denotes the head and j the tail of the edge. The vertices of a directed graph correspond to genes or other elements in the regulatory system, while the edges denote interactions among the genes. Moreover a directed edge can be defined as a tuple (i, j, s), with s equal to + or - , which can be indicated whether i is activated or inhibited by j.

As per the graph (see Figure 2),

\[
V=\{1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,P1,P2,P3,P4,P5,P6\} \\
E=\{(1,2,+), (1,3,-), (1,5,+), (1,7,-), (1,8,+), (1,9,-), (1,10,-), (1,11,+), (1,12,-), (1,13,+), (2,6,+), (3,4,+), (4,6,+), (4,5,+), (5,6,+), (6,6,+), (7,4,+), (8,3,+), (9,3,+), (10,2,+), (10,1,+), (11,1,+), (12,15,+), (13,14,+), (14,15,+), (15,6,+))\}
\]

![Figure 2: Directed Acyclic Graph Representation of the GRN.](image)

In this concern it is to be noted that as per GRN some nodes are being represented for the set of multiple nodes and specifications of those nodes in the graph will be found in the supplementary.

2.3 Directed hypergraph representation of the Gene Regulatory Network with cooperative Interactions

A hyper graph H may be defined as \( H = (X, E) \) where \( X \) is a set of elements called nodes or vertices, and \( E \) is a set of non-empty subsets of \( X \) called as hyper edges or edges. Therefore, \( E \) is a subset of \( \mathcal{P}(X) \) – {Ø}, where \( \mathcal{P}(X) \) is the power set of \( X \). While graph edges are pairs of nodes, hyper edges are arbitrary sets of nodes, and can contain an arbitrary number of nodes. Here in this paper hyper graphs can be used to deal with situations in which miRNAs/genes cooperatively regulates the expression of genes. The edges are then defined by (I,J,S) where J represents a list of regulating genes and S a corresponding list of signs indicating their regulatory influence.

Here in this paper from the graph1 we can find that,

\[
X=\{1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,P1,P2,P3,P4,P5,P6\} \\
E=\{(1,2,5,6,7,8,9,10,11,12,13,14,15,\cdots\cdots\cdots\cdots\cdots\},(2,3,4,5,6,7,8,9,10,11,12,13,14,15,\cdots\cdots\cdots\cdots\cdots\},(3,4,5,6,7,8,9,10,11,12,13,14,15,\cdots\cdots\cdots\cdots\cdots\},(4,5,6,7,8,9,10,11,12,13,14,15,\cdots\cdots\cdots\cdots\cdots\},(5,6,7,8,9,10,11,12,13,14,15,\cdots\cdots\cdots\cdots\cdots\},(6,7,8,9,10,11,12,13,14,15,\cdots\cdots\cdots\cdots\cdots\},(7,8,9,10,11,12,13,14,15,\cdots\cdots\cdots\cdots\cdots\},(8,9,10,11,12,13,14,15,\cdots\cdots\cdots\cdots\cdots\},(9,10,11,12,13,14,15,\cdots\cdots\cdots\cdots\cdots\},(10,11,12,13,14,15,\cdots\cdots\cdots\cdots\cdots\},(11,12,13,14,15,\cdots\cdots\cdots\cdots\cdots\},(12,13,14,15,\cdots\cdots\cdots\cdots\cdots\},(13,14,15,\cdots\cdots\cdots\cdots\cdots\},(14,15,\cdots\cdots\cdots\cdots\cdots\},(15,\cdots\cdots\cdots\cdots\cdots\})
\]
2.4 Representation of Gene Regulatory Network Through the Light of Canalizing Function

The idea of canalizing function was given by Kauffman and Waddington. Canalizing function is a kind of Boolean function where the input of one variable can generate the output of the function. For example, if a function $f = x_1 \cup x_2 \cup x_3$ or $f = x_1 \cap x_2 \cap x_3$, then the input of one variable can only give the output of the function $f$.

The figure 01 shows genes responsible for causing cancer at different stages. When the Gene Regulatory Network (GRN) is represented as a Boolean network each of the genes in the GRN is considered as a node. As shown in the figure there are multiple genes which are responsible for causing cancer in any particular stage. All the nodes that are responsible for causing cancer is expressed in terms of a Boolean canalizing function. Some of the nodes can also be represented in the form of non canalizing function, but since canalizing function gives more stability to the network, so instead of using non canalizing function for representing the nodes, here only canalizing functions has been used.

When any network of genes in the disease free condition is given, then the input for each of the genes responsible for causing cancer at different stages is canalized to either 0 or 1. If activation of any gene causes cancer then the output of the causal gene is 1, else if deactivation is responsible then output is 0. Graphs can be represented as linked list, if any causal gene is activated or deactivated then the output for corresponding cascaded genes (as seen from the linked list diagram) will also be activated or deactivated.

As seen from the figure there are 6 different stages for causing cancer namely cell cycle, cell division, cell replication, angiogenesis, apoptosis, cell proliferation. The canalizing function for each nodes and each stage is obtained and is as shown below:

$$x_1 = x_1 \cap 1; \text{where } i = (2,3,5,6,7,8,9,10,11,12,13)$$
$$x_4 = x_3 \cup 0$$
$$x_{14} = x_{13} \cup 0$$
$$x_{15} = x_{12} \cap x_{14}$$
$$p_1 = x_{10} \cup x_{11} \cup x_{15}$$
$$p_2 = x_{10} \cap 1$$
$$p_3 = x_8 \cup x_9$$
$$p_4 = x_7 \cap 1$$

For an n input canalizing function the output for at least one of the input variables is fixed. Here in the network diagram node $x_i$ inhibits all the other functions $x_j$ where $i = (2,3,5,6,7,8,9,10,11,12,13)$. So the output from $x_i$ for all these functions is 0, i.e. the output for $x_i$ here is canalized. Since $x_i$ is canalized all the nodes attached with $x_i$ is also canalized (because of cascaded nature). As a result the function for each of the stages of cancer becomes canalized. By applying backtracking method to each of the stages the original gene from where discrepancy started and cancer occurred could easily be detected. Hence mathematically by changing the value of that particular node growth of cancer can be deactivated.

Comparative result analysis for representation of GRN with canalizing and non canalizing functions is stated below.

If the gene regulatory network is represented with the help of a Boolean network and nodes are represented as a non canalizing function then some problem or error might occur. For example, considering the cell cycle (Node P1) stage in Figure 3 (see Figure 3).

![Figure 3: Directed Graph considering the cell cycle (Node P1)](image)

Directed Graph considering the cell cycle (Node P1)

If any one of the nodes $(x_{10}, x_{11}, x_{15})$ is activated then cell cycle occurs, as seen from above when canalizing function is used to represent the nodes $x_{10}, x_{11}, x_{15}$ cell cycle occurs. The nodes of the same cell cycle stage can also be represented with the help of non canalizing function as:

$$x_{10} = x_1 \oplus 1$$
$$x_{11} = x_1 \oplus 1$$
$$x_{12} = x_1 \oplus 1$$
$$x_{13} = x_1 \oplus 1$$
$$x_{14} = x_{13} \oplus 0$$
$$x_{15} = x_{12} \cup x_{12}$$

$p_1 = x_{10} \oplus x_{11} \oplus x_{15}$. If at time instant $t$ suppose $x_{1}$ does not deactivate $x_{10}$ then the graph is as follows (see figure 4).

![Figure 4: Directed Graph if at time instant $t$ suppose $x_{1}$ does not deactivate $x_{10}$](image)

Figure 4: Directed Graph if at time instant $t$ suppose $x_{1}$ does not deactivate $x_{10}$

$x_{10} = x_1 \oplus 0$, the functions for rest of the nodes remain the same.

Now $p_1 = x_{10} \oplus x_{11} \oplus x_{15} = 0 \oplus 1 \oplus 0 = 1$. Since $p_1$ becomes 0, it indicates cell cycle does not occur but if

$p_1 = x_{10} \cup x_{11} \cup x_{15} = 0 \cup 1 \cup 1 = 1$. Since $p_1$ is 1, it indicates cell cycle occurs (which is the correct result). Since using non canalizing function for representing the nodes of gene regulatory network give rise to erroneous situation, canalizing function are only used to represent the nodes of a gene regulatory network.

3. CONCLUSION AND FUTURE SCOPE

Breast Cancer generally occurs due to some complex steps. Whereas Let-7 is a member of miRNA family and actively participates in various steps of it. Numerous research works are going on to find out therapeutic procedure in micro RNA. Perfect understanding of gene regulatory system can help one step ahead to do so. Let-7 targets various key components of mitogenic and tumorigenic pathways to exert its tumor suppressor activity. Pathways include cell cycle, cell division, cell proliferation, DNA replication, angiogenesis and apoptosis. In this concern in this paper it has been tried to make an insight of miRNA Let-7 GRN through some graph
representation and Canalizing function. In near future it would be extended by making comparative study between diseased and disease free state of Gene Regulatory Network so that mathematically some indication can be obtained to get rid of this disease.

ACKNOWLEDGMENTS
The authors are indebted to their colleague Mr. Ranjeet Rout of Indian Statistical Unit, Kolkata, West Bengal, India for his kind suggestions.

REFERENCES


