



Decoding the Role of Epigenetics in Breast Cancer Using Formal Modeling and Machine-Learning Methods

Ayesha Asim, Yusra Sajid Kiani, Muhammad Tariq Saeed and Ishrat Jabeen *

School of Interdisciplinary Engineering and Sciences (SINES), National University of Sciences and Technology, (NUST), Islamabad, Pakistan

OPEN ACCESS

Edited by:

Adil Mardinoglu,
King's College London,
United Kingdom

Reviewed by:

Fu Hui,
Tianjin University of Traditional
Chinese Medicine, China
Han Jin,
Science for Life Laboratory
(SciLifeLab), Sweden

*Correspondence:

Ishrat Jabeen
ishrat.jabeen@sines.nust.edu.pk

Specialty section:

This article was submitted to
Molecular Diagnostics and
Therapeutics,
a section of the journal
Frontiers in Molecular Biosciences.

Received: 24 February 2022

Accepted: 25 May 2022

Published: 11 July 2022

Citation:

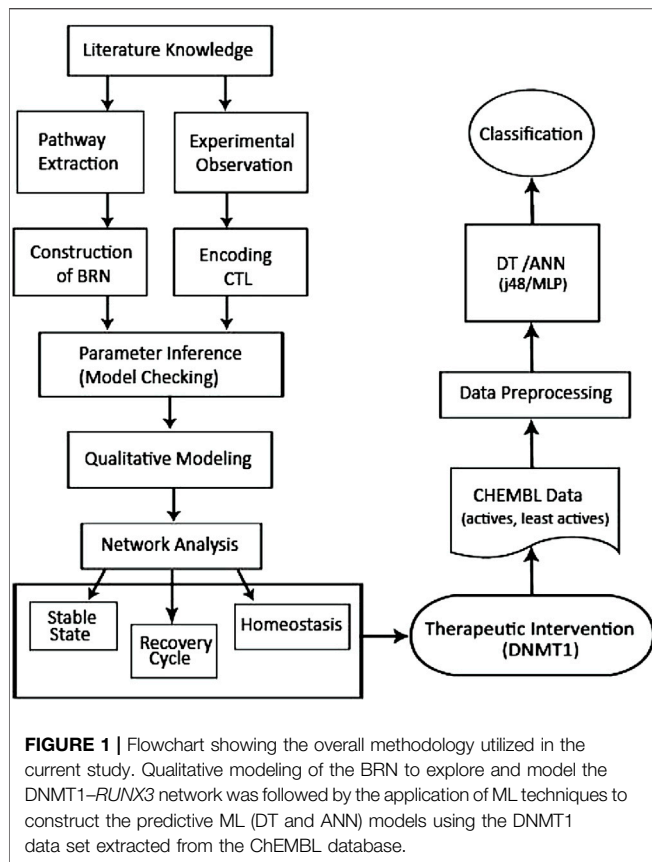
Asim A, Kiani YS, Saeed MT and
Jabeen I (2022) Decoding the Role of
Epigenetics in Breast Cancer Using
Formal Modeling and Machine-
Learning Methods.
Front. Mol. Biosci. 9:882738.
doi: 10.3389/fmolb.2022.882738

Breast carcinogenesis is known to be instigated by genetic and epigenetic modifications impacting multiple cellular signaling cascades, thus making its prevention and treatments a challenging endeavor. However, epigenetic modification, particularly DNA methylation-mediated silencing of key TSGs, is a hallmark of cancer progression. One such tumor suppressor gene (TSG) *RUNX3* (Runt-related transcription factor 3) has been a new insight in breast cancer known to be suppressed due to local promoter hypermethylation mediated by DNA methyltransferase 1 (DNMT1). However, the precise mechanism of epigenetic-influenced silencing of the *RUNX3* signaling resulting in cancer invasion and metastasis remains inadequately characterized. In this study, a biological regulatory network (BRN) has been designed to model the dynamics of the DNMT1–*RUNX3* network augmented by other regulators such as p21, c-myc, and p53. For this purpose, the René Thomas qualitative modeling was applied to compute the unknown parameters and the subsequent trajectories signified important behaviors of the DNMT1–*RUNX3* network (i.e., recovery cycle, homeostasis, and bifurcation state). As a result, the biological system was observed to invade cancer metastasis due to persistent activation of oncogene c-myc accompanied by consistent downregulation of TSG *RUNX3*. Conversely, homeostasis was achieved in the absence of c-myc and activated TSG *RUNX3*. Furthermore, DNMT1 was endorsed as a potential epigenetic drug target to be subjected to the implementation of machine-learning techniques for the classification of the active and inactive DNMT1 modulators. The best-performing ML model successfully classified the active and least-active DNMT1 inhibitors exhibiting 97% classification accuracy. Collectively, this study reveals the underlined epigenetic events responsible for *RUNX3*-implicated breast cancer metastasis along with the classification of DNMT1 modulators that can potentially drive the perception of epigenetic-based tumor therapy.

Keywords: *RUNX3* signaling pathway, Dnmt1, machine learning, qualitative modeling, c-myc, SMBioNet

1 INTRODUCTION

Breast cancer is one of the frequently diagnosed lethal malignancies affecting millions of women worldwide. The risk factors include both genetic and epigenetic abnormalities whereby the latter provides early prognostic biomarkers for breast cancer therapeutics (Chen, 2012; Sun et al., 2017). Furthermore, DNA methylation is the most prominent epigenetic marker that frequently influences the expression or silencing of genes involved in vital cellular activities such as cell proliferation,



programmed cell death, and cell differentiation (Chen, 2012). DNA methylation is mediated by DNA methyltransferases (DNMTs), in which the DNMT1 isomer is explicitly responsible for the maintenance of methylation status during replication (S-phase) (Pouliot et al., 2015). Briefly, it transfers the methyl group (–CH₃) on cytosine in CpG dinucleotide regions by keeping the CpG islands non-methylated that are preferably located in the proximal promoter region of a gene (Tang et al., 2009). On the contrary, the hypermethylation of promoter CpG islands of TSGs due to the overexpression of DNMT1 is a popular mechanism of gene silencing and a hallmark of cancer (Subramaniam et al., 2014; Pouliot et al., 2015).

DNMT1-regulated hypermethylation of the TSG *RUNX3* promoter region is a new insight and an early event in breast cancer predominantly in TNBC (triple-negative breast cancer) (Widschwendter and Jones, 2002; Jung et al., 2007; Jiang et al., 2008; Subramaniam et al., 2009; Shin et al., 2016). Generally, *RUNX3* is known to elicit its tumor-suppressive ability through major cancer signaling pathways including TGF- β , Wnt/ β -catenin, and KRAS (Chen, 2012). *RUNX3* either influences the downstream target of tumor suppressor signaling pathways or acts as an antagonist for oncogenic pathways to exert its antitumor activity. However, despite the evident role of *RUNX3* in breast cancer and the relation of DNMT1 with *RUNX3* gene, no study has explored the underlying epigenetic molecular events by which DNMT1 silences *RUNX3* to promote breast cancer metastasis.

Furthermore, due to the pivotal implication of DNMT1 in various tumors (Li et al., 2019) and the reversible nature of its methylation activity, targeting DNMT1 through small modulators is a promising pharmacological intervention to revive the suppressed TSGs (such as *RUNX3*). To date, two DNMT inhibitors (i.e., azacytidine and decitabine) have been approved by FDA for the treatment of myeloid malignancies (Gnyszka, Jastrzębski, and Flis 2013). However, the potential side effects associated with these drugs limit their use in high-grade malignancies. Therefore, the development of more potent DNMT1 inhibitors acquiring potential binding pocket features is highly desirable to restore the suppressed TSGs (*RUNX3*) in cancer therapeutics. Previously, several *in vivo* and computational studies have targeted DNMT1 for therapeutic purposes (Yoo 2011; Yoo, Kim, and Robertson 2012; Mirza et al., 2013; Krishna et al., 2017); however, its impact on downstream target genes, particularly *RUNX3*, requires further elucidation.

Therefore, in the present study, a BRN of the DNMT1–*RUNX3* signaling was developed to provide an insight into the epigenetic-mediated silencing of *RUNX3* leading to cancer development and metastasis. For this purpose, qualitative modeling by the René Thomas formalism was applied in a SMBioNet tool utilizing the existing wet-laboratory data in the form of computation tree logic (CTL) for parameter estimation and verification through the model checking technique (Figure 1). The model trajectories were explored to identify the paths involved in the overexpression of DNMT1, activation of oncogene (c-myc), and suppression of tumor suppressor genes (*RUNX3*, p21, and p53) leading to cancer invasion or recovery (homeostasis). In addition, DNMT1 was advocated as a potential drug target and favored to be subjected to ML approaches to classify diverse DNMT1 modulators. Decision tree (DT) and neural network classifiers were built utilizing the data set of DNMT1 from the ChEMBL database to identify potential two-dimensional binding features essential for the modulation of DNMT1 activity (Figure 1).

2 MATERIALS AND METHODS

The overall methodology used in the current study is provided in Figure 1.

2.1 Qualitative Modeling Framework

Definition 1. Biological Regulatory Network (BRN)

A BRN is a labeled, directed graph $G = (V, E)$ in which biological regulators are denoted as a set of nodes or vertices V and their interactions as a set of edges $E \subseteq V \times V$. Each edge is labeled by a pair of elements (τ, σ) , where τ is the threshold at which a gene U starts regulating a gene V , and σ represents the type of interactions $\sigma = +$ (activation) and $-$ (inhibition) between the nodes (Bernot et al., 2004).

Definition 2. Qualitative State

The state of the BRN is a configuration of the expression level of all the nodes at a particular time instant. A state is n-tuple $S =$

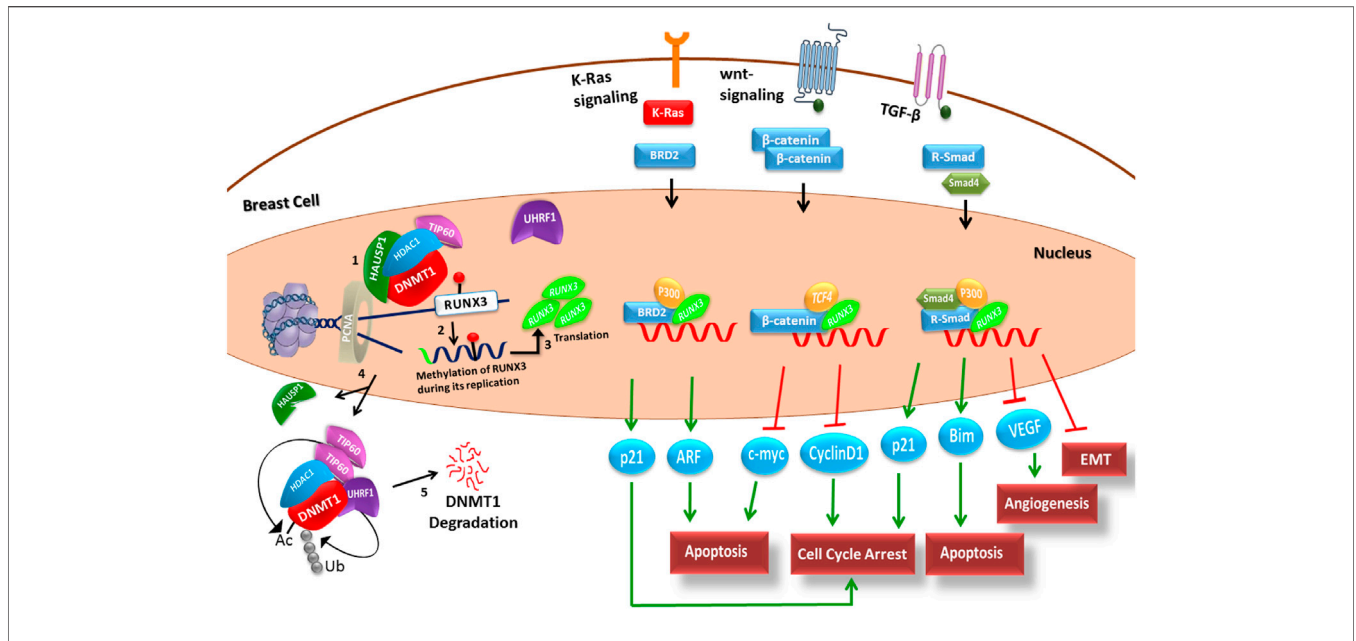


FIGURE 2 | Schematic knowledge-based network is presented to illustrate the stepwise process of replication and methylation (by DNMT1) of TSG *RUNX3* and its implication in different cancer pathways. Step 1: At the replication fork, UHRF1 recognizes hemimethylated DNA and recruits other proteins including DNMT1, Tip60, HAUSP1, HDAC1, and PCNA to make a macroprotein complex. Step 2: DNMT1 in complex with these regulators transfers a methyl group through the base flip mechanism onto the nascent daughter strand. The green color highlighted in the daughter strand depicts the hypomethylated promoter region of the nascent *RUNX3* gene. The red-headed lollipop structure here mimics the normal methylation status of *RUNX3* gene. Step 3: Transcription machinery then successfully identifies the promoter region to translate the functional *RUNX3* protein, which acts as a tumor suppressor and combats the cancerous environment through the regulation of major signaling pathways including TGF-beta, Wnt/β, and KRAS pathways. TSG *RUNX3* exerts its antitumor activity by regulating the transcription of significant target genes including p21, c-myc, and p53 (oval blue structures at the bottom). Steps 4 and 5: The entire macroprotein complex after performing its function undergoes stepwise proteasomal degradation in the late S-phase of cell cycle. PCNA = proliferating cell nuclear antigen; DNMT1 = DNA methyltransferase 1; HAUSP1 = herpesvirus-associated ubiquitin-specific protease; HDAC1 = histone deacetylase 1; Tip60 = histone acetyltransferase; UHRF1 = ubiquitin-like, containing PHD and RING finger domains 1; Ac = acetylation; *RUNX3* = Runt-related transcription factor 3; TGF-β = transforming growth factor-beta; VEGF = vascular endothelial growth factor; EMT = epithelial-mesenchymal transition; and TSG = tumor suppressor gene.

$\{s_{v1}, \dots, s_{vn}\}, \forall s_{vi} \in \delta_{vi}$, where s_{vi} is the expression level of v_i (Bernet et al., 2004).

Definition 3. Resources

Each v_i in a state of the BRN is controlled by its precursors G_{v_i} , called the set of resources ω_{v_i} .

Let $G = (V, E)$, at level s_{v_j}, ω_{v_i} , defined as:
 $\omega_{v_j} = v_i \in G_{v_j}^- | (s_{v_i} \geq \tau_{v_i,v_j} \text{ and } \sigma_{v_i,v_j} = +)$ or $(s_{v_i} < \tau_{v_i,v_j} \text{ and } \sigma_{v_i,v_j} = -)$ (Bernet et al., 2004).

Definition 4. Parameters

Parameters of a biological regulatory network are indexed by its set of resources. It is a cartesian product of each variable's element and resource (Saeed et al., 2016).

Definition 5. Betweenness Centrality

Network analysis techniques such as graph theory-based approaches can be applied to further analyze the state graph by sorting it based on betweenness centrality (Freeman 2019). The betweenness centrality measures the extent to which a single vertex or node is more central/connected than all other nodes. For a particular node, the centrality metric is measured with the

number of shortest paths that pass through it, whereas the betweenness would be high for a particular node if it appears in many shortest paths (Golbeck 2015). Likewise, the qualitative states with higher betweenness centrality value are more likely to occur in the system, implying that these entities are frequently expressed in biological phenomena. Moreover, the central or more connected nodes in the system might represent a potential therapeutic target (Golbeck 2015; Saeed et al., 2016).

2.2 Qualitative Modeling and Parameter Estimation

René Thomas introduced a graph theory-based approach for qualitative modeling of dynamic biological regulatory networks (Thomas 1978). In this method, each BRN (Definition 1) is modeled as a weighted, directed graph, which consists of a set of nodes and edges. Nodes represent a biological entity (i.e., genes or proteins), whereas edge represents the relationship of activation or inhibition between the nodes.

Herein, the role of the epigenetic (DNMT1)-mediated *RUNX3* silencing in cancer development and progression was examined by regulating key oncogenes and TSGs. The literature-driven information was used to generate the BRN, which signifies important entities involved in the DNMT1-*RUNX3* signaling

and the relationship (activation and inhibition) among the chosen entities. The unknown parameters were inferred by encoding wet-laboratory biological observations as propositional calculus or more precisely computation tree logic (CTL) and verified through the model checking technique as previously reported by Saeed et al. (2018). Briefly, model checking is an automatic technique based on the exhaustive exploration of the entire state space of a biological system, which therefore allows the analysis and cross-verification of a large number of possible outcomes of a network (Bernot et al., 2004). All compatible combinations of parameters (Definition 4) were generated and evaluated by a model checker using a SMBioNet tool, and the model parameters that violate laboratory data were eliminated.

Briefly, SMBioNet (Khalis et al., 2009) is a qualitative framework-based tool that calls New Symbolic Model Verifier (NuSMV) (Cimatti et al., 2002) as a model checker. NuSMV works by considering a Model M of the BRN and its property φ , which exhaustively explores M to verify φ . The SMBioNet engages in this principle to identify logical parameters of the models that comply more with the known observations. Once the computational verification with laboratory data was completed, the resultant model and all its trajectories were further analyzed to understand how the systematic evolution of the DNMT1–RUNX3 system takes place with time. The network was explored using the concept of betweenness centrality to underline important trajectories of the dynamic biological system (i.e., homeostasis, bifurcation state, and recovery cycle). Furthermore, the paths involved in the activation of oncogenes, suppression of tumor suppressor genes, cancer invasion, and recovery were also identified. This study also highlighted a qualitative bifurcating at which the system can invade tumorigenesis or normal homeostasis depending upon the successive changes in the expression level of TSG and oncogenes.

2.3 Data Set Compilation or Collection

A total of 738 DNMT1 inhibitors were collected from the ChEMBL database (target ID ChEMBL1993). Only those compounds were extracted for which the biological activity was estimated experimentally as the inhibitory potency (IC_{50} value). Initially, the removal of duplicates, small fragments ($MW < 200$), and inconsistent activity values (IC_{50}) was performed, which was followed by the exploration and manual correction of stereoisomers. The data preprocessing resulted in a final data set of 242 DNMT1 inhibitors with IC_{50} values in the range of 0.01–1,600 μM (Section 2, Supplementary Tables S1, S2).

Briefly, our data set contains compounds of diverse origin that include natural, synthetic (nucleoside/non-nucleoside), and FDA-approved drugs against DNMT1, thus incorporating all major scaffolds of DNMT1 inhibitors available to date. The shortlisted DNMT1 inhibitors were used to build machine-learning (DT and ANN) models. Therefore, a diverse subset selection approach was utilized to divide the data set into a training set (80%) and a test set (20%) for model building and validation, respectively. Briefly, a diverse subset splits the data into two sections based on chemical diversity calculated as a function of distance between molecular descriptors (Koutsoukas

et al., 2013). The absolute biological activity values were converted into binary numbers on the basis of an activity threshold value ($IC_{50} \leq 10 \mu M$) such that 1 represents active and 0 indicates least-active class of DNMT1 inhibitors. This binarization of DNMT1 data was supported by a histogram plot provided in Supplementary Figure S1.

2.4 Calculation and Selection of 2D Chemical Descriptors

All 2D MOE descriptors (2019.01) (ULC 2018) by excluding energy-related descriptors were calculated for the training set. A total of 158 2D descriptors were computed through MOE after which the redundant, null, and constant value descriptors were excluded from the final set. The descriptors with negligible relevancy and weightage were also removed from the final set of descriptors to improve the overall predictive ability of ML models. Consequently, the selected descriptors were provided as an input to WEKA (3.9.3) for the construction of a DT classifier that was further used for shortlisting of the most relevant and decision-making attributes.

2.5 Machine-Learning Approaches

Herein, the DT and ANN classification models were built using the training set and the models were validated using the 20% test set compounds.

2.5.1 Decision Tree

The C4.5 variant of the J48 algorithm implemented in WEKA was used to build the univariate tree of training set attributes. The J48 algorithm works by splitting the data into smaller subsets based on features that will produce the most uniform child node at each step (Yosipof, Guedes, and García-Sosa 2018). The process is repeated iteratively until no more splits can be made or data are uniformly classified into terminal nodes. Tree parameters were tuned to improve the overall performance of the model and limit the overfitting of data. Therefore, the lowest number of confidence factor was used to incur more pruning and a minimum of one instance per leaf was set for the splitting rule, with a 10-fold cross-validation approach.

2.5.2 Artificial Neural Network

Artificial neural networks (ANNs) are nonparametric human nervous system-inspired computational models, which process complex input information to produce the output. A multilayer perceptron (MLP) function of WEKA (3.9.3) was utilized to build the ANN model of DNMT1 inhibitors. The MLP network typically consists of at least three layers: one input, one output, and one hidden layer (can be more than one). The perceptron calculates a linear combination of inputs and their weights to compute a sum, and the output is calculated through an activation function (most often the sigmoid) (Jiao et al., 2020). Moreover, MLP uses back-propagation to find the optimized input weights and builds hidden layer(s) for the classification of nonlinear data (Frank, Hall, and Witten 2017). Generally, MLP normalizes each attribute by default, to improve the performance of the network. Overall, model training was performed using a

10-fold cross-validation and the parameters were optimized on the basis of number and nodes of hidden layer by constructing several perceptron models. The default values of momentum, learning rate and training time were used in order to build the best-fit neural network.

2.5.3 Model Performance Validation

The DT and ANN models were trained using the training set (80%), and the classification was validated using the test set (20%). The performance of the machine-learning model was further evaluated using statistical parameters including accuracy (Eq. 1), sensitivity (Eq. 2), specificity (Eq. 3), F-measure (Eq. 6), and MCC measure (Eq. 7). Accuracy is the percentage of correctly classified active and least-active compounds, whereas specificity is the percentage of true least-active predicted compounds, and sensitivity is the proportion of true active compounds predicted from our model. In addition, F-score, also known as balanced accuracy, measures the precision (how many compounds are correctly classified) and robustness (resistance to errors) of ML models. Likewise, Matthew's correlation coefficient (MCC) measures the difference between actual and predicted values, and it is only high (near to 1) if all the classes are predicted in good proportion.

Overall accuracy:

$$\frac{TP + TN}{TP + FP + TN + FN} \quad (1)$$

Sensitivity (true-positive rate):

$$Tp = \frac{TP}{TP + FN} \quad (2)$$

Specificity (true-negative rate):

$$Tn = \frac{TN}{TN + FP} \quad (3)$$

Precision:

$$\frac{TP}{TP + FP} \quad (4)$$

Recall:

$$\frac{TP}{TP + FN} \quad (5)$$

F-measure:

$$2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (6)$$

Mathew's correlation coefficient:

$$C = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (7)$$

3 RESULTS

A schematic model depicting the stepwise process of replication and methylation of TSG *RUNX3* by DNMT1 and its subsequent

influence on various cancer signaling pathways was established using the literature-driven information. Qualitative modeling was then performed to unfold the cellular events involved in the epigenetic, that is, DNMT1-mediated silencing of *RUNX3* that leads to cancer invasion. The comprehensive literature-driven pathway of DNMT1–*RUNX3* is elaborated in the **Supplementary Material Section S3, Figure 2**. Briefly, a series of stepwise events at the replication fork led by macroprotein complexes (DNMT1, UHRF1, HAUSP, Tip60, HDAC1, and PCNA) ensures the methylation of the newly synthesized daughter strand during the S-phase (**Figure 2**). As a result, the nascent *RUNX3* strand acquires the normal status of methylation, whereas the promoter region remains hypomethylated to facilitate active transcription of *RUNX3* gene. Subsequently, *RUNX3* (transcription factor) functions as a tumor suppressor protein to combat cancer initiation and metastasis through various signaling pathways. The TSG *RUNX3* elicits an antitumor activity by regulating the transcription of target genes (p21, c-myc, etc.) of the major cancer signaling cascades, such as transforming growth factor-beta (TGF-β), Wnt/β-catenin, and mitogenic KRAS pathway as explained in **Figure 2**.

From literature-driven pathways described in **Figure 2**, a qualitative biological regulatory network was generated based on preferred entities. DNMT1, *RUNX3*, p21, c-myc, p53, and MDM2 nodes were selected due to their well-established functionality in the biological system (**Supplementary Material Section S3**). As a result, the BRN (**Figure 3A**) composed of total six nodes and nine interactions exhibiting all the significant activation and inhibition relationships was obtained. Initially, negative feedback loops were observed from an inhibitory set of genes required by the system to generate the stable states. According to the interaction graph (**Figure 3A**), *RUNX3* transactivates p21 to maintain the concentration of DNMT1 through a negative feedback loop. In addition, *RUNX3* inhibits the onset of oncogene c-myc that might upregulate the expression level of DNMT1 in a positive manner. Interestingly, p53 also prevents upregulation of DNMT1 through the activation of p21 and inhibition of oncogene c-myc. Furthermore, the proposed network characterized oscillatory behavior of two regulatory loops involving DNMT1 and *RUNX3*: 1) a negative feedback loop between p21 and DNMT1 through *RUNX3* and 2) a positive feedback loop between c-myc and DNMT1 through *RUNX3* (**Figure 3A**). It is notable that the representation of the BRN as a weighted, directed graph (**Figure 3A**) was obtained using GENOTECH through the implementation of discrete formalism. Therefore, the interaction graph utilized the qualitative data only, that is, type of interaction (activation +, or inhibition –) among the nodes (genes), and the threshold value for each interaction. However, modeling the dynamic behavior of such complex systems that include both positive and negative loops requires the computation of accurate logical parameters. These parameters were generated in the SMBioNet software using the known experimental observations encoded as CTL formulas reported in **Figure 3B**. The first CTL observation searches for a state with high expression of DNMT1 and oncogenes leading to tumor invasion. The later CTL formula seeks for the stable state or

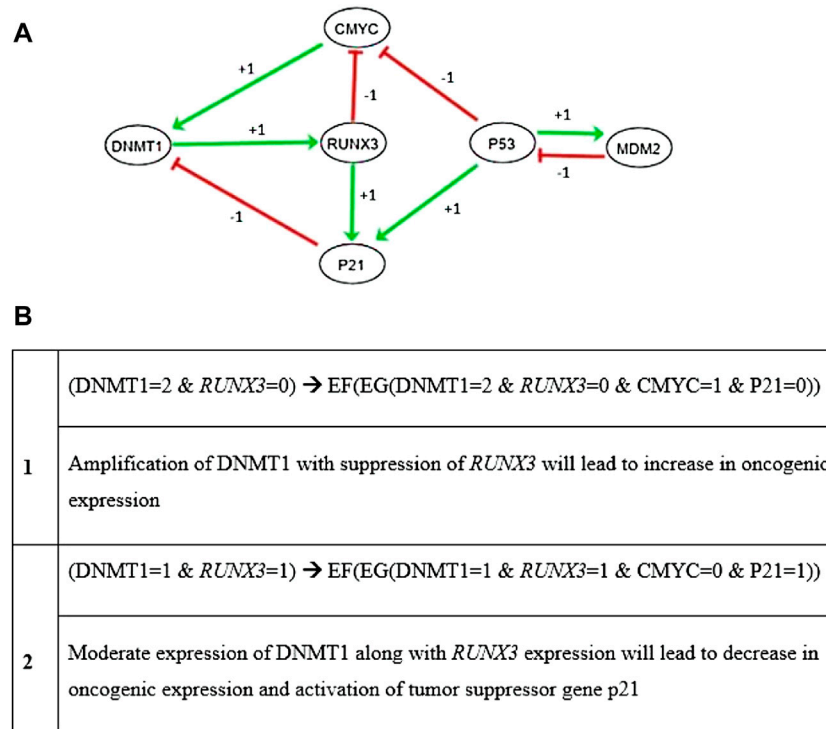


FIGURE 3 | (A) DNMT1–RUNX3 interaction graph (BRN) was generated by utilizing the preferred entities to discover all the important activation (+) and inhibition (–) relationships among them. The network demonstrates predominantly two oscillatory behaviors in DNMT1–RUNX3 graph. The first one illustrates the RUNX3-stimulated onset of p21, which in turn positively regulates DNMT1 (shown with red arrows). The other loop exhibits the inhibition of c-myc by RUNX3, which also instigates the onset of DNMT1 (shown with green pointed arrows). In addition, p53 can also be seen regulating DNMT1 through the activation of p21 and inhibition of c-myc signals. **(B)** The second part of the figure demonstrates two CTL observations utilized by SMBioNet for the estimation of parameters that were later used to generate the state graph of the dynamic model. According to CTL formulas, the overexpression of DNMT1 is responsible for hypermethylation at the promoter region of RUNX3 and ultimately its suppression, which is associated with many cancer types including breast cancer. Each circle/node represents a gene, and the arrows among them show the type of interaction they hold. Activation is denoted with green pointed arrows, and blunt red arrows represent inhibition whereby the weight of the arrows depicts threshold values of interactions.

homeostasis that has normal expression level of tumor suppressor genes (**Figure 3B**). These observations were coded as input in SMBioNet to generate the preferred set of logical parameters.

3.1 Parameters

SMBioNet computed a total of 14 sets of logical parameters that are presented as a heat map in **Figure 4**. The preferred set of logical parameters verified through model checking might reflect the probable biological trajectories in cancer invasion and recovery. According to the known biological observations, the most significant change occurs in the expression of DNMT1 and RUNX3 in different types of malignancies including breast tumor. The heat map suggests eight critical resources of DNMT1 based on the presence and absence of its activators and inhibitors. For instance, (CMYC) indicates the presence of c-myc, and (CMYC, P21) represents the presence of activator c-myc and the absence of inhibitor p21. Likewise, (CMYC, P21, and P53) shows the presence of activator c-myc whereby no inhibitor is present in the system. The ability of each entity to evolve is described as a function of the presence or absence of its resource as shown in **Figure 4**.

The trend of DNMT1 being induced in the presence of c-myc was observed in all 14 parameter sets. Therefore, a parameter set

that allows DNMT1 to achieve its maximum threshold value of “2” was selected to generate the state graph, assuming that this concentration is lethal for breast cells. The selected parameters (M6) allow all nodes to interact with others corresponding to the natural dynamic phenomenon, while maintaining their interdependencies for the activation or suppression. The source code of input models is provided in the **Supplementary Material Section S1**. The calculated parameters indicate that DNMT1 maintains a higher expression level in the presence of oncogene c-myc signaling. Conversely, an activation of an inhibitor p21 or p53 prevents the expression level of DNMT1 to exceed the normal threshold value; it is also observed that TSG RUNX3 is activated in the presence of DNMT1 signal. The expression of p53 is increased when the MDM2 inhibition signals are absent in the system. However, the collective behavior of the genes in a dynamic biological system can only be concluded by interpreting the trajectories in the state transition graph.

3.2 State Graph

On the basis of selected logical parameters (M6) inferred from SMBioNet, a state graph was generated using GENOTECH and analyzed on the Cytoscape software to highlight significant genetic

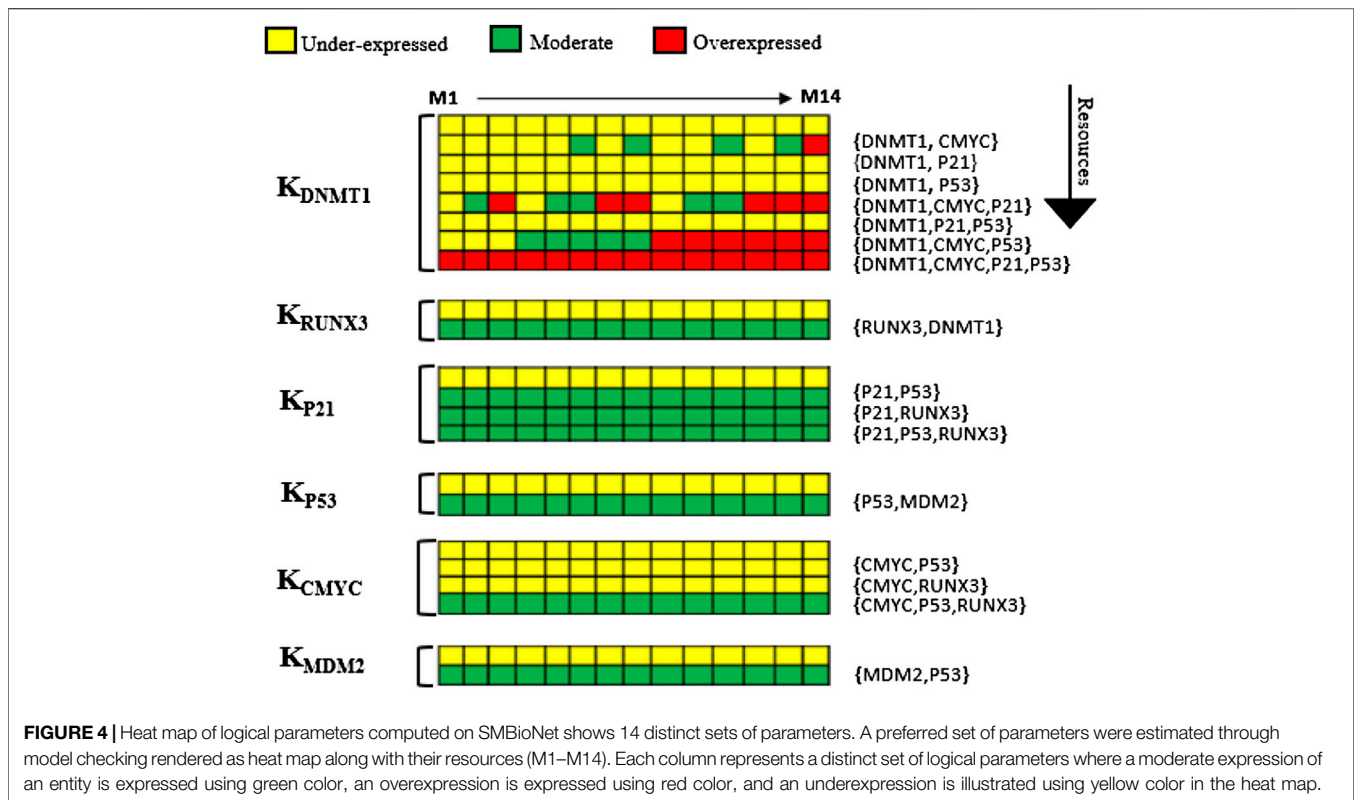


FIGURE 4 | Heat map of logical parameters computed on SMBioNet shows 14 distinct sets of parameters. A preferred set of parameters were estimated through model checking rendered as heat map along with their resources (M1–M14). Each column represents a distinct set of logical parameters where a moderate expression of an entity is expressed using green color, an overexpression is expressed using red color, and an underexpression is illustrated using yellow color in the heat map.

evolution including the recovery cycle (Figure 5). The state transition graph illustrates all the possible qualitative states exhibited by the DNMT1–RUNX3 system as shown in Figure 6 (b). It contains a total of 32 nodes and 79 edges sorted on the basis of the betweenness centrality concept of graph theory. Our generated state graph highlights the temporal evolution of the system in the form of trajectories from one qualitative state to the other. It includes homeostasis, recovery trajectories, and bifurcation state from which the biological system can evolve either to homeostasis or to tumorigenesis. Here, we highlighted two important trajectories of the qualitative model involved in tumor progression and recovery/homeostasis. According to our model, the activation of oncogene c-myc signal (c-myc = 1) introduces pathogenesis in the biological system, which is characterized by the qualitative states (0,0,1,1,0,0) and (0,1,0,1,1,0) highlighted in Figures 5 and 6C, respectively.

3.3 Recovery

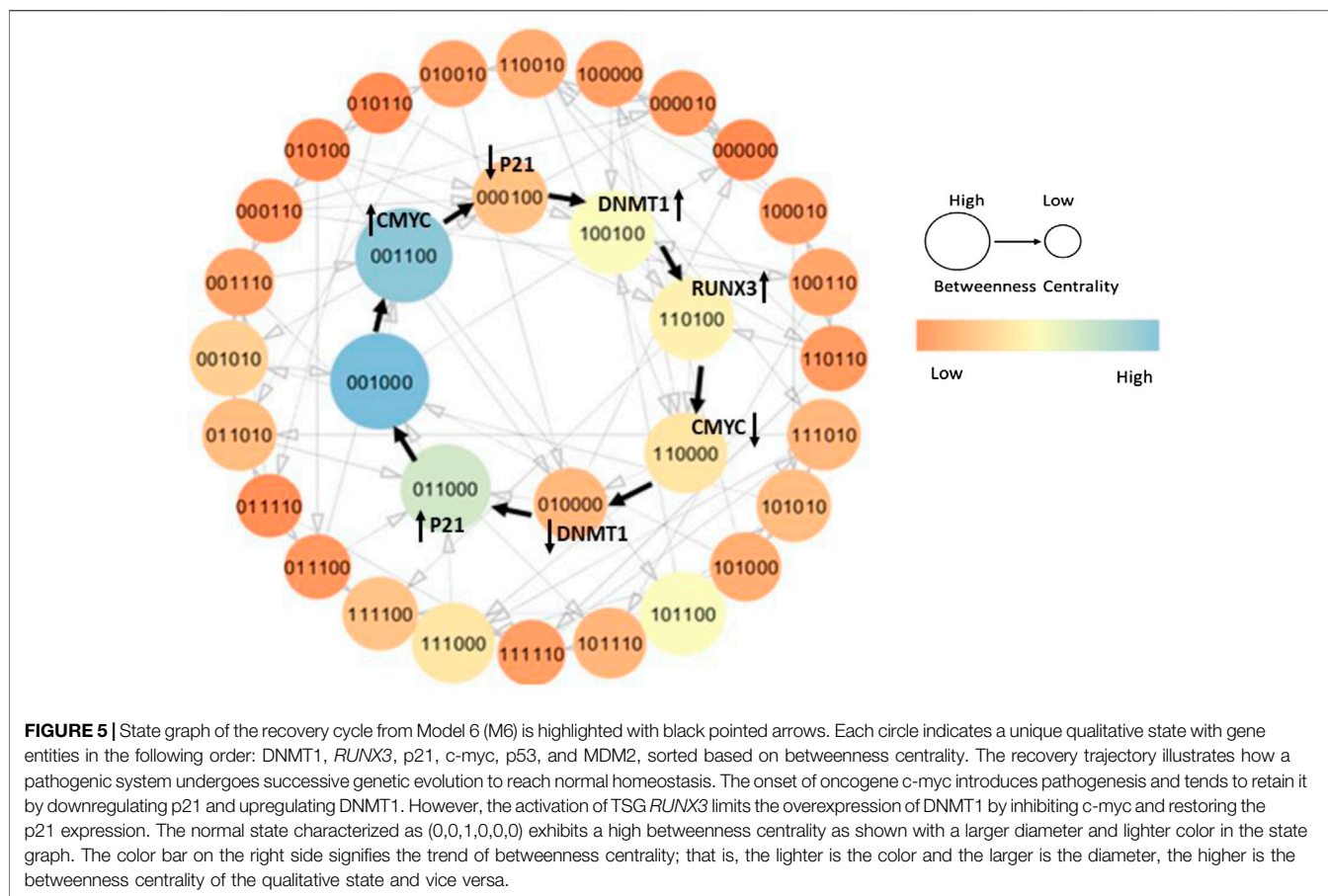
The qualitative model explains the step-by-step evolution of the system to recover from stress environment and maintain homeostasis. The onset of oncogenic c-myc (c-myc = 1) initiates the pathogenesis by inhibiting cell cycle arrest protein/TSG p21 (p21 = 0) leading to the qualitative state (0,0,0,1,0,0), which consequently activates DNMT1 (DNMT1 = 1) (1,0,0,1,0,0) in the system. It is a typical pathogenic state characterized by high expression level of DNMT1 and c-myc along with a suppressed concentration of TSGs, that is, RUNX3, p21, and p53 as shown in Figure 5.

DNMT1 activates RUNX3 by maintaining the methylation status (RUNX3 = 1) (1,1,0,1,0,0), which employs its tumor suppressor

activity and inhibits the onset of oncogene c-myc subsequently, limiting the hyperactivation of DNMT1 leading to the states (1,1,0,0,0,0) and (0,1,0,0,0,0), respectively. In addition, RUNX3 transactivates p21 (0,1,1,0,0,0) to further keep a concentration check of DNMT1 expression level. The restoration of p21 signal supervises the abnormal and uncontrolled proliferation, thus causing the system to recover from a pathogenic environment. These successive genetic changes evolve into homeostasis that is generally characterized by the presence of TSGs with a moderate concentration of DNMT1, and repressed oncogenes (0,0,1,0,0,0) as shown in Figure 5. Notably, the qualitative state (0,0,1,0,0,0) is the most central state of the system shown with lighter color and larger diameter. This suggests a high probability of the infected system to recover back to homeostasis through a series of steady states that are relatively less central (low betweenness centrality). The graph in Figure 5 with black pointed arrows describes all the states that the system exhibits to recover from a pathogenic state. Under normal circumstances, biological systems often exhibit oscillatory behavior or homeostasis during which the overall status of the system remains in a cycle of normal states. Thus, the qualitative model should expect pathogenic trajectories along with normal homeostasis behavior either as a cycle or as a closed path.

3.4 Bifurcation state

A transition bifurcation state highlighted in Figure 6 is characterized by the onset of oncogene c-myc from which the system could divert either toward pathogenic state or toward homeostasis depending on the successive genetic alterations. The suppressed level of TSG RUNX3 (right trajectory of Figure 6C)



converges the system toward pathological path in which the persistent activation of oncogene c-myc ($c\text{-myc} = 1$) with the subsequent consistent suppression of TSG *RUNX3* ($RUNX3 = 0$) augments the hyperactivation of DNMT1. Consequently, the activated c-myc also downregulates p21 and p53 to further assist the metastasis and tumor invasion (1,0,0,1,0,0) characterized by red square in **Figure 6C**. Alternatively (left trajectory of **Figure 6C**), activation of TSG *RUNX3* inhibits the c-myc signaling and restores p21 expression, which controls the concentration of DNMT1 to maintain normal homeostasis of the system. This leads to normal qualitative state (0,1,1,0,1,0) characterized by the moderate level of DNMT1, the low expression of oncogenes, and the presence of TSGs along with the oscillations of the p53–MDM2 circuit as highlighted with green square on the left side of **Figure 6C**. Moreover, the qualitative normal state (0,0,1,0,0,0) in the homeostasis trajectory is well connected as shown with lighter color and larger diameter, which explains the probability of this state to occur more in the system than any other or pathogenic state (1,0,0,1,0,0).

3.5 Machine-Learning Models

ML models were built in WEKA using the DNMT1 data set of diverse structures and FDA-approved drugs (decitabine, azacytidine) that were extracted from the ChEMBL database

(target ID ChEMBL1993). 80% of the collected DNMT1 data set was utilized for training the model, while the remaining 20% was included in the test set to validate the classification models. The number of training and test set compounds along with other details (bioactivities, activity threshold) is presented in **Section 4, Supplementary Table S3**. In order to build classification models, an activity threshold was established for both the training and test sets such that compounds with IC_{50} values $\leq 10 \mu\text{M}$ and $> 18 \mu\text{M}$ were categorized as active and least-active compounds, respectively. As a result, the training set contained 146 active (class label 1) and 68 least-active (class label 0) compounds. Initially, all the provided 2D MOE (2019.01) (ULC 2018) descriptors were calculated to train the DT model. The descriptors selected by the DT model were further utilized to build the ANN model of DNMT1 inhibitors using the MLP algorithm in WEKA. The descriptors including vsa_acc , h_logP , b_count , $radius$, $PEOE_VSA-5$, b_double , $SlogP_VSA5$, SMR_VSA2 , Q_VSA_PNEG , and $Kier3$ (Wildman and Crippen 1999) were identified to be important for model learning and differentiating between active and least-active compounds efficiently. The detail of preferred descriptors is provided in **Supplementary Table S4 of Section 4**.

3.5.1 Decision Tree

The J48 algorithm implemented in the ML software WEKA was used to build the DT classification model. Parameters of DT were

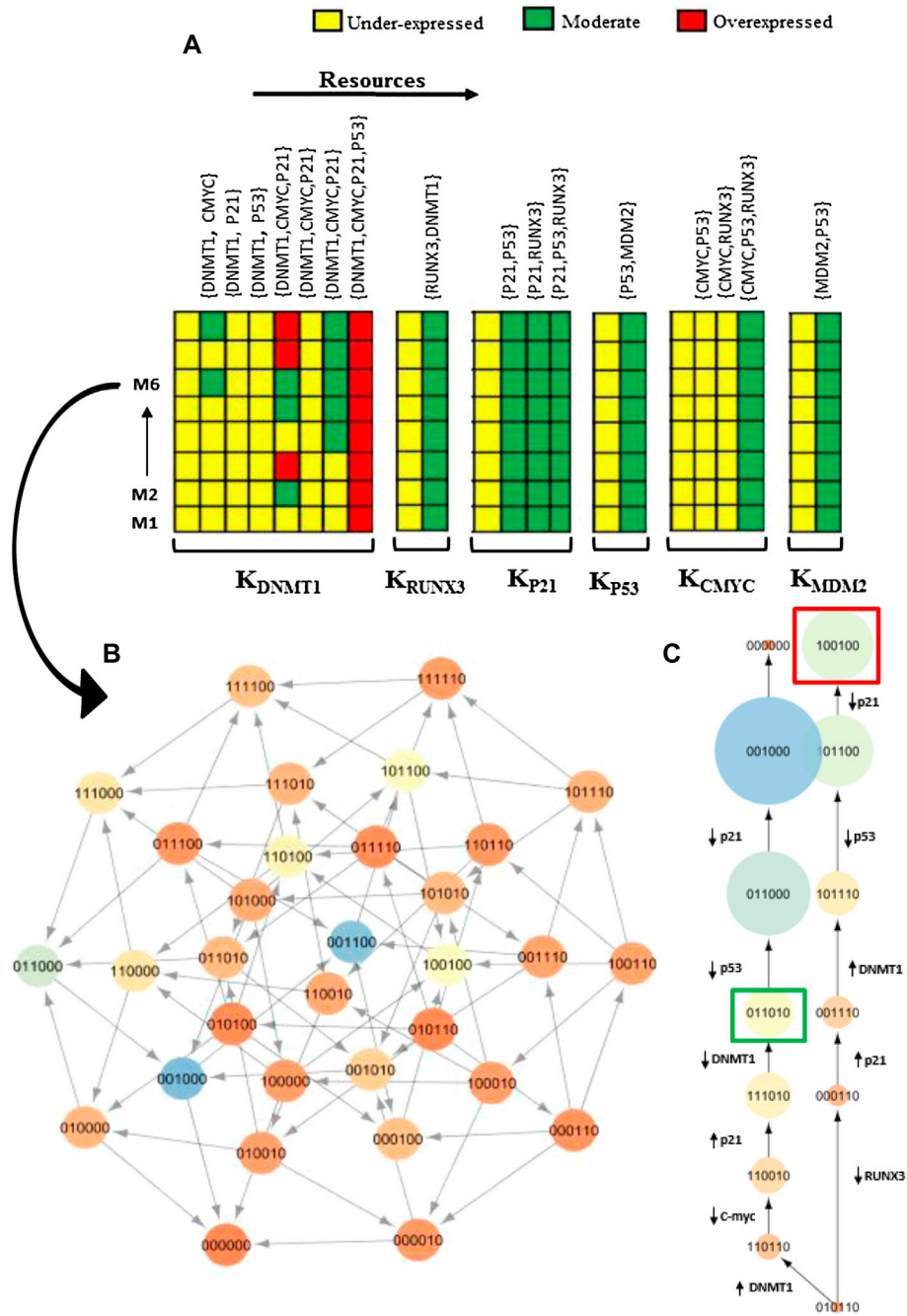


FIGURE 6 | (A) Heat map of the first 8 sets of parameters out of total 14 (refer **Figure 4**) computed by SMBioNet is displayed, whereas Model 6 (M6) parameters were used to generate the state transition graph for network analysis. **(B)** State graph of M6 is shown, which composed of 32 nodes and 79 edges sorted on the basis of betweenness centrality. Each circle represents a unique state with gene expression in the order as follows: DNMT1, RUNX3, p21, c-myc, p53, and MDM2. The generated state transition graph illustrates all the possible qualitative states of the system. Trajectories of the graph were then further analyzed to identify important genetic evolution. **(C)** A bifurcation state is highlighted characterized by the qualitative state (010110). Trajectories display distinct paths from one common transition state characterized by the onset of oncogene c-myc leading to homeostasis or pathological loop based on successive genetic changes. From the bifurcating (0,1,0,1,1,0) state, the repressed RUNX3 (0,0,0,1,1,0) converges the system toward pathological state with a successive onset of DNMT1 (1,0,1,1,1,0) and offset of p53 (1,0,0,1,0,0). Here, the qualitative state (1,0,0,1,0,0) is characterized as pathological state (highlighted in red box in the right trajectory) experienced by the system due to the consistent onset of oncogene c-myc along with persistent suppression of RUNX3. On the contrary, the system might evolve toward normal state (highlighted in green box in the left trajectory) if TSG RUNX3 gets activated causing constant inhibition of oncogene c-myc (1,1,0,0,1,0), to control the moderate expression level of DNMT1 (0,1,1,0,1,0). The normal state is characterized by the activation of RUNX3 along with the controlled expression level of DNMT1 (0,1,1,0,1,0), which is achieved by the system through continuous activation of TSG RUNX3 along with the consistent inhibition of oncogene c-myc, ultimately leading to a typical reset state (0,0,0,0,0,0).

TABLE 1 | Statistical parameters of classification models, J-48 decision tree and MLP neural network, for training data calculated from WEKA.

ML algorithm	Accuracy	Sensitivity	Specificity	Precision	F-measure	MCC
Training set (80%)						
DT	0.974	0.993	0.920	0.973	0.983	0.932
ANN	0.969	0.979	0.940	0.979	0.979	0.919
Test set (20%)						
DT	0.826	0.916	0.500	0.868	0.891	0.45
ANN	0.717	0.805	0.400	0.828	0.816	0.20

optimized, and a pruned tree was built in order to get the best accuracy and performance on the available data set. Therefore, a minimum of 1 instance per leaf was set for the splitting rule and the rest default parameters of the J48 algorithm were utilized to build an optimal DT model. A final DT model with an overall size of 23 and 12 terminal nodes was obtained using the DNMT1 data set (shown in **Section 4, Supplementary Figure S2**). The overall topology of the DT classifier is elaborated in **Section 4** of the **Supplementary Material**.

The model performance and overall predictive ability were evaluated through different statistical parameters including sensitivity (**Eq. 2**), specificity (**Eq. 3**), F-measure (**Eq. 6**), and MCC value (**Eq. 7**). The DT classifier attained accuracy, sensitivity, and specificity values of 0.974, 0.99, and 0.92, respectively, as shown in **table 1**. Overall, 97% of DNMT1 inhibitors were correctly classified into active and least-active class of compounds by the DT model. However, the ratios of correctly predicted active compounds (TPR) and the truly predicted least-active compounds (TNR) were observed to be 99% and 92%, respectively. The DT displayed an MCC value of 0.9 (**table 1**), indicating an optimal model efficiency with a strong positive correlation between the actual and predicted class labels. In addition, the F1-score was calculated to assess the balanced classification accuracy performance of the trained DT model, which turned out to be 0.98. As the DT showed an optimal performance on the training data in terms of classification and prediction, the descriptors identified by the DT model were further utilized to build the predictive ANN model.

3.5.2 Artificial Neural Network

A set of 10 selected 2D descriptors [vsa_acc, h_logP, b_count, radius, PEOE_VSA-5, b_double, SlogP_VSA5, SMR_VSA2, Q_VSA_PNEG, and Kier3 (Wildman and Crippen 1999)] were used to build the ANN model in WEKA utilizing the MLP algorithm. Several ANN models were developed exploiting different combinations of parameters in order to obtain an optimal classification model. The final ANN comprised of 10 input nodes, 1 layer of 4 hidden nodes, and two output nodes (i.e., “1” for active and “0” for least-active compounds) (**Section 4, Supplementary Figure S3**). The ANN model was optimized using a training time of 500, learning rate of 0.3, and momentum of 0.2 to improve the accuracy and speed of learning.

The classification performance of ANN was also evaluated using statistical measures including classification accuracy, F1-score, and MCC values as reported in **table 1**. The trained ANN

model attained a MCC value of 0.91, an overall classification accuracy of 0.97, and a F1-score of 0.98, thus indicating an optimal binary classification of our data set. Overall, 97% of DNMT1 data was correctly classified into active and least-active class by our learned ANN, whereby the true-positive and true-negative prediction rates were observed to be 0.98 and 0.94 implying a higher specificity and sensitivity, respectively. Both the trained DT and ANN classifiers displayed optimal statistical performance, that is, greater than 96% and a good predictive ability for the training data.

3.5.3 Model Validation

The obtained DT and ANN models were validated using the 20% test set compounds that contain the most diverse chemical structures of the DNMT1 inhibitor data set (activity range of 0.01–132 μ M). Overall, DT and ANN acquired a classification accuracy of greater than 70% for the test set whereby DT could correctly classify 83% of the data and the ANN model attained an accuracy of 73% (**Table 1**). Both the classifiers displayed an optimal value of sensitivity, that is, 0.92 and 0.80 for DT and ANN, respectively, implying a high true-positive prediction rate on the test set. Moreover, F1-score was calculated as 0.89 for the DT classifier and 0.81 for ANN indicative of the robustness and unbiased classification performance of our predictive ML classifiers. Despite the structural diversity of DNMT1 inhibitors, our models optimally retained the classification accuracy of 0.97 on the training set and 0.82 (DT) and 0.72 (ANN) on the test data. Notably, in comparison with the ANN model, the DT classifier could classify active and least-active compounds more efficiently as reported in **table 1**. The training and test sets utilized for model generation and validation in this study are shown in **Section 2, Supplementary Tables S1, S2**.

4 DISCUSSION

TNBC is the most aggressive subtype of breast cancer that lacks obvious treatment options due to the availability of limited information about definite biomarkers. Previously, aberrant epigenetic modifications have been implicated in breast cancer, highlighting *RUNX3* as a promising prognostic biomarker (Wang et al., 2014). Moreover, several former studies have reported the depleted level of *RUNX3* in breast cancer cell lines predominantly due to local hypermethylation at the proximal promoter region of TSG *RUNX3*, which is an early event in carcinogenesis (Chen, 2012; Lau et al., 2006; Chen et al., 2016). Therefore, the aberrant

DNA methylation pattern is known as the hallmark of cancer epigenetics characterized by DNMT1, also known as the maintenance methyltransferase. So far, the individual role of *RUNX3* as a TSG in breast cancer and DNMT1 as the methylation maintenance protein has been reported in various scientific studies (Jiang et al., 2008; Chen et al., 2016; Lau et al., 2006; Subramaniam et al., 2009). However, the underlined epigenetic-mediated *RUNX3*-shared signaling that instigates and disseminates the tumorigenesis remains largely elusive. This study provides an insight into the integrated DNMT1–*RUNX3* signaling by taking into account various cancer-related significant upstream and downstream regulators (such as p21, c-myc, p53, and MDM2) and presents the DNMT1–*RUNX3* signaling cascades as one consolidated network. To the best of authors' knowledge, this study is one of its kind to illustrate and model the epigenetic-mediated silencing of TSG *RUNX3* through the René Thomas framework modeling that has largely been a common practice of systems biology to investigate the dynamics of biological networks (Saeed et al., 2018).

Moreover, the model checking technique was adopted to build the interaction graph rendered as a state graph (Figure 6B) by utilizing the literature-driven information of the DNMT1–*RUNX3* signaling (Figure 2). The similar approach of the model checking has been previously applied in different former studies such as parameter estimation through formal modeling (Ahmad et al., 2012; Bibi et al., 2017), tail resorption in tadpole metamorphosis (Khalis et al., 2009), and immunity control in bacteriophage lambda (Richard, Comet, and Bernot 2006). We utilized this technique in the SMBioNet tool to exhaustively explore the model space of the DNMT1–*RUNX3* BRN for the estimation of precise parameters by encoding the existing wet-laboratory data in the form of computation tree logic (CTL) (Figure 3B). Once the computational verification with laboratory data is completed, the resultant model parameters and all its trajectories were further analyzed to understand how systematic evolution of the DNMT1–*RUNX3* system takes place with time, which might lead the system to invade cancer metastasis (Figure 6B).

As a result, two important behaviors of the DNMT1–*RUNX3* system were highlighted that shows the successive genetic events of the recovery cycle (Figure 5) and a bifurcation state leading to cancer invasion (Figure 6C). The trajectory (right side of Figure 6C) from our model plotted as a state graph articulates that the onset of oncogene c-myc infects the system and anticipates the upregulation of DNMT1 by downregulating the *RUNX3* expression level. It generates a feedback loop where the consistent onset of oncogene c-myc accompanied by the persistent suppression of TSG *RUNX3* contributes toward the overproduction of DNMT1. These findings are in agreement with the previous experimental studies, which reported the transcriptional upregulation of DNMT1 in TNBC due to the amplified c-myc expression level, emphasizing the importance of DNMT1 and c-myc relation in tumorigenesis (Wu et al., 2021).

On the contrary, the activation of *RUNX3* (left side of Figure 6C) and subsequent TSGs (such as p21, p53) regulate the adequate level of DNMT1 to acquire homeostasis. In addition,

network analysis emphasized the tumor-suppressive role of *RUNX3* by demonstrating the higher probability of the infected system to recover to normal homeostasis in the presence of *RUNX3* (left side of Figure 6C) than its likelihood to invade the tumorigenesis. This is in line with the former findings that have conversely related the reactivation of *RUNX3* with a reduced potential of metastasis and invasiveness in breast cancer cells (Widschwendter and Jones, 2002). Although computational models cannot replace experiments, they are a step to demonstrate whether or not a proposed mechanism is sufficient to produce an observed phenomenon or an underlying assumption on the basis of their mathematical framework.

Moreover, the findings of our qualitative modeling suggested DNMT1 as a critical hub regulator of various oncogenic and tumor suppressor proteins that determine the ultimate status of the system. These outcomes advocate DNMT1 as a potential drug target in epigenetic cancer signaling through logic-based temporal evolution. Notably, the hypermethylation of TSGs by DNMT1 is a reversible process; therefore, *RUNX3* expression could be restored using demethylating compounds. Consequently, designing specific small modulators that inhibit DNMT1 activity in order to restore the TSG *RUNX3* could represent new clinical avenues for breast cancer therapeutics. Several pioneer studies have discussed the probable restoration of *RUNX3* and a reduced carcinogenic potential in cancer cell lines when treated with demethylating drugs (Lau et al., 2006; Jung, Park, Young Kim, et al., 2007).

To date, only two FDA-approved drugs (azacytidine and decitabine) are available that target DNMT1 in addition to other DNMTs for the treatment of myelodysplastic syndrome (Gnyszka, Jastrzębski, and Flis 2013). However, potential side effects of these drugs necessitate the design of less toxic and more specific inhibitors of DNMT1. Ever since the breakthrough of epigenetics in cancer treatment, several *in silico* studies such as pharmacophore and QSAR modeling (Yoo, Kim, and Robertson 2012; Phanus-Umporn et al., 2020) have been reported for drug discovery against DNMT1 through small modulators. However, to the best of authors' knowledge, no model is developed that could learn the diverse features of DNMT1 inhibitors. Therefore, in this study, we adopted machine-learning approaches to identify potential 2D descriptors of DNMT1 modulators by utilizing all the available chemical scaffolds of DNMT1 inhibitors (i.e., natural compounds, synthetic, nucleoside inhibitors, and FDA drugs).

ML classifiers DT and ANN were developed to classify the active and least-active compounds of the DNMT1 data set. All 10 captured features of the best-performing model (DT and ANN) along with feature descriptions are provided in Section 4, Supplementary Table S4. The trained DT identified vsa_acc, h_logP, b_count, radius, PEOE_VSA-5, b_double, SlogP_VSA5, SMR_VSA2, Q_VSA_PNEG, and Kier3 as key descriptors of demethylating compounds targeting DNMT1. However, the overall vDW surface area of hydrogen bond acceptors (vsa_acc) was reflected as the most prominent and distinguishing 2D descriptor by the DT model. This feature has also been previously identified by Hassanzadeh et al.

(2017), which further emphasizes the significance of HBA for the activity of DNMT1 modulators. Furthermore, the developed models were able to classify the training set and predict the test set with optimal accuracies (**table 1**). The final selected trained DT and ANN classifiers showed an optimal classification accuracy of 0.97 and high sensitivity and specificity along with MCC values of 0.93 and 0.92, respectively.

The subsequent screening of the diverse test set from the DT and the ANN model resulted in the classification accuracies of 0.83 and 0.72, respectively. However, the DT classifier outperformed the ANN model showing higher classification accuracy of 0.83 and sensitivity of 0.92. Notably, despite the structural diversity of the DNMT1 data set, the ML classifiers developed in the current study retained an optimal classification accuracy of 97% for the training data and greater than 70% for the test data. The relatively low predictive accuracy of ML models on the test set can be attributed to the diversity of DNMT1 inhibitors as they differ highly in terms of chemical structure, molecular weight, and other pharmacological variables such as clogP and lipophilicity. Therefore, the learned features of the training set might not be sufficient to fully predict and explain the behavior of the test set. Also, the FDA-approved drug (azacytidine) was fairly classified among most active modulators of DNMT1 by our developed model, which strengthens the classification and predictive ability of the ML models.

5 CONCLUSION

RUNX3 has been proposed as a potential biomarker in TNBC for an early prognosis, which is known to be downregulated by DNMT1. However, the precise mechanism of epigenetic-mediated silencing of TSG *RUNX3*, which results in cancer invasion and metastasis, has not yet been explored. This study deals with the formal modeling of the DNMT1–*RUNX3* signaling and the development of ML models on a diverse data set of DNMT1 modulators. First, we employed a qualitative modeling approach to provide an insight into epigenetic-inspired *RUNX3* signaling. The results revealed that the onset of oncogene *c-myc* introduces pathogenesis in the system and its consistent activation along with persistent suppression of TSG *RUNX3* hyperactivates DNMT1 leading to cancer metastasis. Conversely, the activation of *RUNX3* leads the system to acquire normal homeostasis by transactivating other TSGs such as p21 and p53. Moreover, our findings advocate DNMT1 as a potential epigenetic drug target to revive the suppressed TSG *RUNX3* in breast cancer therapeutics. Second,

predictive ML models (DT and ANN) have been developed that identify some potential 2D descriptors essential to modulate the DNMT1 activity, and the best-performing models have effectively classified the active and least-active inhibitors. The trained (DT and ANN) models acquired 97% classification accuracy on the training data set, and the subsequent screening of the test set through DT and ANN models achieved 83% and 72% predictive accuracy, respectively, emphasizing the optimal efficiency of the developed model. In general, the application of formal methods to unveil the network and model the underline genetic events responsible for DNMT1-inspired TSG *RUNX3* silencing along with ML approaches to predict the 2D attributes of hypomethylating compounds (targeting DNMT1) could present new computational avenues for the treatment of breast cancer requiring epigenetic therapy.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**; further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

AA, IJ, and MS designed the study and performed the experiments. AA, YK, and IJ analyzed the data, collected materials/data/tools, wrote the manuscript, prepared figures/tables, analyzed the results, and reviewed the draft of the manuscript. AA and MS contributed analysis tools, investigated the results, reviewed and proofread the draft of the manuscript, and provided technical support. All authors approved the results of the final draft.

FUNDING

The publication fee for this research manuscript has been funded by the National University of Sciences and Technology, NUST.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmolb.2022.882738/full#supplementary-material>

REFERENCES

- Ahmad, J., Niazi, U., Mansoor, S., Siddique, U., and Bibby, J. (2012). Formal Modeling and Analysis of the MAL-Associated Biological Regulatory Network: Insight into Cerebral Malaria. *PLOS ONE* 7 (3), e33532. doi:10.1371/JOURNAL.PONE.0033532
- Bernot, G., Comet, J.-P., Richard, A., and Guespin, J. (2004). Application of Formal Methods to Biological Regulatory Networks: Extending Thomas' Asynchronous Logical Approach with Temporal Logic. *J. Theor. Biol.* 229, 339–347. doi:10.1016/j.jtbi.2004.04.003
- Bibi, Z., Ahmad, J., Siddiq, A., Paracha, R. Z., Saeed, T., Ali, A., et al. (2017). Formal Modeling of MTOR Associated Biological Regulatory Network Reveals Novel Therapeutic Strategy for the Treatment of Cancer. *Front. Physiol.* 8 (JUN), 416. doi:10.3389/FPHYS.2017.00416/BIBTEX
- Chen, F., Liu, X., Bai, J., Pei, D., and Zheng, J. (2016). The Emerging Role of *RUNX3* in Cancer Metastasis (Review). *Oncol. Rep.* 35, 1227–1236. doi:10.3892/or.2015.4515

- Chen, L.-F. (2012). Tumor Suppressor Function of RUNX3 in Breast Cancer. *J. Cell. Biochem.*, a-n. doi:10.1002/jcb.24074
- Cimatti, A., Clarke, E., Giunchiglia, E., Giunchiglia, F., Pistore, M., Roveri, M., et al. (2002). "NuSMV 2: An Opensource Tool for Symbolic Model Checking," in *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 359–364. doi:10.1007/3-540-45657-0_29
- Frank, E., Hall, M. A., and Witten, I. H. (2017). The WEKA Workbench. *Data Min.*, 553–571. doi:10.1016/b978-0-12-804291-5.00024-6
- Freeman, L. C. (2019/2019). "A Set of Measures of Centrality Based on Betweenness," in *American Sociological Association Stable* (American Sociological Association Is Collaborating with JSTOR to Digit.), 40, 35–41. doi:10.2307/30335431
- Gnyszka, A., Jastrzębski, Z., and Flis, S. (2013). DNA Methyltransferase Inhibitors and Their Emerging Role in Epigenetic Therapy of Cancer. *Anticancer Res* 2996 33, 2989–2996.
- Golbeck, J. (2015). Analyzing Networks. *Introd. Soc. Media Investigation*, 221–235. doi:10.1016/B978-0-12-801656-5.00021-4
- Hassanzadeh, M., Kasymov, R., Mahernia, S., Adib, M., Emperle, M., Dukatz, M., et al. (2017). Discovery of Novel and Selective DNA Methyltransferase 1 Inhibitors by Pharmacophore and Docking-Based Virtual Screening. *ChemistrySelect* 2 (27), 8383–8392. doi:10.1002/slct.201701734
- Hong, X., Wang, Z., Wang, B., Guo, H., and Guo, H. (2014). Clinicopathological Significance and Potential Drug target of T-Cadherin in NSCLC. *Dddt* 9, 207–216. doi:10.2147/DDDT.S74259
- Jiang, Y., Tong, D., Lou, G., Zhang, Y., and Geng, J. (2008). Expression of RUNX3 Gene, Methylation Status and Clinicopathological Significance in Breast Cancer and Breast Cancer Cell Lines. *Pathobiology* 75 (4), 244–251. doi:10.1159/000132385
- Jiao, Z., Hu, P., Xu, H., and Wang, Q. (2020). Machine Learning and Deep Learning in Chemical Health and Safety: A Systematic Review of Techniques and Applications. *ACS Chem. Health Saf.* 27 (6), 316–334. doi:10.1021/acs.chas.0c00075
- Jung, Y., Park, J., Kim, T. Y., Jong, H.-S., Im, S.-A., Robertson, K. D., et al. (2007). Potential Advantages of DNA Methyltransferase 1 (DNMT1)-Targeted Inhibition for Cancer Therapy. *J. Mol. Med.* 85 (10), 1137–1148. doi:10.1007/S00109-007-0216-Z
- Khalis, Z., Comet, J.-P., Richard, A., and Bernot, G. (2009). The SMBioNet Method for Discovering Models of Gene Regulatory Networks. *Genes, Genomes Genomics* 3 (1), 15–22.
- Koutsoukas, A., Paricharak, S., Galloway, W. R. J. D., Spring, D. R., IJzerman, A. P., Glen, R. C., et al. (2013). How Diverse Are Diversity Assessment Methods? A Comparative Analysis and Benchmarking of Molecular Descriptor Space. *J. Chem. Inf. Model.* 54 (1), 230–242. doi:10.1021/ci400469u
- Krishna, S., Shukla, S., Lakra, A. D., Meeran, S. M., and Siddiqi, M. I. (2017). Identification of Potent Inhibitors of DNA Methyltransferase 1 (DNMT1) through a Pharmacophore-Based Virtual Screening Approach. *J. Mol. Graph. Model.* 75, 174–188. doi:10.1016/j.jmkgm.2017.05.014
- Lau, Q. C., Raja, E., Salto-Tellez, M., Liu, Q., Ito, K., Inoue, M., et al. (2006). RUNX3 Is Frequently Inactivated by Dual Mechanisms of Protein Mislocalization and Promoter Hypermethylation in Breast Cancer. *Cancer Res.* 66, 6512–6520. doi:10.1158/0008-5472.CAN-06-0369
- Li, F., Wan, X., Xing, J., Tan, X., Li, X., Wang, Y., et al. (2019). Deep Neural Network Classifier for Virtual Screening Inhibitors of (S)-Adenosyl-L-Methionine (SAM)-Dependent Methyltransferase Family. *Front. Chem.* 7 (MAY), 1–17. doi:10.3389/fchem.2019.00324
- Mirza, S., Sharma, G., Parshad, R., Gupta, S. D., Pandya, P., and Ralhan, R. (2013). Expression of DNA Methyltransferases in Breast Cancer Patients and to Analyze the Effect of Natural Compounds on DNA Methyltransferases and Associated Proteins. *J. Breast Cancer* 16 (1), 23–31. doi:10.4048/jbc.2013.16.1.23
- Phanus-Umporn, C., Prachayasittikul, V., Nantasenam, C., Prachayasittikul, S., and Prachayasittikul, V. (2020). QSAR-driven Rational Design of Novel Dna Methyltransferase 1 Inhibitors. *EXCLI J.* 19, 458–475. doi:10.17179/EXCLI2020-1096
- Pouliot, M. C., Labrie, Y., Diorio, C., and Durocher, F. (2015). The Role of Methylation in Breast Cancer Susceptibility and Treatment. *Anticancer Res.* 35 (9), 4569–4574. doi:10.1007/s13566-015-0216-5
- Richard, A., Comet, J.-P., and Bernot, G. (2006). "Formal Methods for Modeling Biological Regulatory Networks," in *Modern Formal Methods and Applications*, 83–122. doi:10.1007/1-4020-4223-X_5
- Saeed, M. T., Ahmad, J., Baumbach, J., Pauling, J., Shafi, A., Paracha, R. Z., et al. (2018). Parameter Estimation of Qualitative Biological Regulatory Networks on High Performance Computing Hardware. *BMC Syst. Biol.* 12 (1), 1–15. doi:10.1186/s12918-018-0670-y
- Saeed, M. T., Ahmad, J., Kanwal, S., Holowatyj, A. N., Sheikh, I. A., Zafar Paracha, R., et al. (2016). Formal Modeling and Analysis of the Hexosamine Biosynthetic Pathway: Role of O-Linked N-Acetylglucosamine Transferase in Oncogenesis and Cancer Progression. *PeerJ* 4 (9), e2348–32. doi:10.7717/peerj.2348
- Shin, E., LeeLee, Y., and Koo, J. S. (2016). Differential Expression of the Epigenetic Methylation-Related Protein DNMT1 by Breast Cancer Molecular Subtype and Stromal Histology. *J. Transl. Med.* 14 (1), 1–11. doi:10.1186/s12967-016-0840-x
- Subramaniam, D., Thombre, R., Dhar, A., and Anant, S. (2014). DNA Methyltransferases: A Novel Target for Prevention and Therapy. *Front. Oncol.* 4. doi:10.3389/fonc.2014.00080
- Subramaniam, M. M., Chan, J. Y., Yeoh, K. G., Quek, T., Ito, K., and Salto-Tellez, M. (2009). Molecular Pathology of RUNX3 in Human Carcinogenesis. *Biochimica Biophysica Acta (BBA) - Rev. Cancer* 1796, 315–331. doi:10.1016/j.bbcan.2009.07.004
- Sun, Y.-S., Zhao, Z., YangYang, Z.-N., Xu, F., LuZhu, H.-J., Zhi. Yong, Z. Y., et al. (2017). Risk Factors and Preventions of Breast Cancer. *Int. J. Biol. Sci.* 13, 1387–1397. doi:10.7150/ijbs.21635
- Tang, M., Xu, W., Wang, Q., Xiao, W., and Xu, R. (2009). Potential of DNMT and its Epigenetic Regulation for Lung Cancer Therapy. *Cg* 10, 336–352. doi:10.2174/138920209788920994
- Thomas, R. (1978). Logical Analysis of Systems Comprising Feedback Loops. *J. Theor. Biol.* doi:10.1016/0022-5193(78)90127-3
- ULC, Chemical Computing Group (2018). *Molecular Operating Environment (MOE), 2013.08.* 1010 Sherbooke St. West, Suite #910. Montreal, QC, Canada. H3A 2R7.
- Widschwendter, M., and Jones, P. A. (2002). DNA Methylation and Breast Carcinogenesis. *Oncogene* 21 (35), 5462–5482. doi:10.1038/SJ.ONC.1205606
- Wildman, S. A., and Crippen, G. M. (1999). Prediction of Physicochemical Parameters by Atomic Contributions. *J. Chem. Inf. Comput. Sci.* 39, 868–873. doi:10.1021/ci9903071
- Wu, S.-Y., Xiao, Y., WeiWei, J.-L., XuXu, X.-E., Jin, X., Hu, X., et al. (2021). MYC Suppresses STING-dependent Innate Immunity by Transcriptionally Upregulating DNMT1 in Triple-Negative Breast Cancer. *J. Immunother. Cancer* 9 (7), e002528. doi:10.1136/jitc-2021-002528
- Yoo, J., Kim, J. H., Robertson, K. D., and Medina-Franco, J. L. (2012). Molecular Modeling of Inhibitors of Human DNA Methyltransferase with a Crystal Structure. *Eighth Edi* 87, 219–247. doi:10.1016/B978-0-12-398312-1.00008-1
- Yoo, J., and Medina-Franco, J. L. (2011). Homology Modeling, Docking and Structure-Based Pharmacophore of Inhibitors of DNA Methyltransferase. *J. Comput. Aided Mol. Des.* 25, 555–567. doi:10.1007/s10822-011-9441-1
- Yosipof, A., Guedes, R. C., and García-Sosa, A. T. (2018). Data Mining and Machine Learning Models for Predicting Drug Likeness and Their Disease or Organ Category. *Front. Chem.* 6 (MAY), 1–11. doi:10.3389/fchem.2018.00162

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Asim, Kiani, Saeed and Jabeen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.