FUNDAMENTAL BOOLEAN NETWORK MODELLING FOR CHILDHOOD ACUTE LYMPHOBLASTIC LEUKAEMIA PATHWAYS

Lesli Chen¹, Don Kulasiri¹*, Sandhya Samarasinghe²

¹ Computational Systems Biology Laboratory, Centre for Advanced Computational Solutions (C-fACS), Lincoln University, Lincoln, Canterbury 7647, New Zealand
² Integrated Systems Modelling Group, Centre for Advanced Computational Solutions (C-fACS), Lincoln University, Lincoln, Canterbury 7647, New Zealand
* Correspondence: Don.Kulasiri@lincoln.ac.nz

RESEARCH ARTICLE

Background: A novel data-driven Boolean model, namely, the fundamental Boolean model (FBM), has been proposed to draw genetic regulatory insights into gene activation, inhibition, and protein decay, published in 2018. This novel Boolean model facilitates the analysis of the activation and inhibition pathways. However, the novel model does not handle the situation well, where genetic regulation might require more time steps to complete.

Methods: Here, we propose extending the fundamental Boolean modelling to address the issue that some gene regulations might require more time steps to complete than others. We denoted this extension model as the temporal fundamental Boolean model (TFBM) and related networks as the temporal fundamental Boolean networks (TFBNs). The leukaemia microarray datasets downloaded from the National Centre for Biotechnology Information have been adopted to demonstrate the utility of the proposed TFBM and TFBNs.

Results: We developed the TFBNs that contain 285 components and 2775 Boolean rules based on TFBM on the leukaemia microarray datasets, which are in the form of short-time series. The data contain gene expression measurements for 13 GC-sensitive children under therapy for acute lymphoblastic leukaemia, and each sample has three time points: 0 hour (before GC treatment), 6/8 hours (after GC treatment) and 24 hours (after GC treatment).

Conclusion: We conclude that the proposed TFBM unlocks their predecessor’s limitation, i.e., FBM, that could help pharmaceutical agents identify any side effects on clinic-related data. New hypotheses could be identified by analysing the extracted fundamental Boolean networks and analysing their up-regulatory and down-regulatory pathways.

Keywords: Boolean modelling; Boolean network; time series data; network inference; data-driven boolean modelling; fundamental boolean model; fundamental boolean networks; orchard cube

INTRODUCTION

A novel data-driven Boolean model, namely, the fundamental Boolean model (FBM), has been proposed to draw genetic regulatory insights into gene activation, inhibition, and protein decay, published in Chen et al. [1]. This novel Boolean model separates the activation and inhibition functions from conventional Boolean
functions, and this separation could facilitate scientists in seeking answers in such as how an amendment of one gene distresses other genes at the expression level. In this paper, we extend the FBM to take the best temporal time step as a count for the clinical data [2], which are in the form of short-time series. The biological meaning of the fundamental Boolean functions separated in terms of activation and inhibition is that a fundamental Boolean function can be regarded as a regulatory activator/inhibitor or regulatory complex function (transcription factor) of its target gene [1].

Fundamental background

The dominant belief of cellular functions mainly depends on coordinated interactions between genes, RNAs and proteins, to form the foundation of genetic regulatory networks (GRNs). Within GRNs, activators and inhibitors play an important role in controlling gene expression pattern by activating or inhibiting cellular functions [3]. Hence, discovering interconnected knowledge about gene activation and inhibition is essential for uncovering the apoptosis process mechanisms; these are crucial for cancer therapy today.

As shown in Fig. 1, an activator is a transcription factor (TF) type of protein that can increase the protein concentration through direct binding to the protein or the promoter sites of its genes to increase its genetic activities. The process is termed gene activation [4].

In contrast, an inhibitor is a repressor that decreases the protein’s concentration to reduce its genetic activities. Thus, the process is named gene inhibition [4]. Genetic inhibitors can be used as pharmaceutical agents in human, veterinary medicine, herbicides and pesticides [4,5].

Facilitated by the emergence of biotechnologies, such as Affymetrix™ microarray technology, an enormous amount of high-throughput genetic data are being generated every day, enabling reverse engineering of unknown regulatory networks, such as revealing the relationships among the functional genes in the mammalian cell cycle [6,7] and leukaemia [8–14]. However, it is evident from these examples that analysing massive datasets to understand the coordinated interactions among genes is still a significant challenge.

Boolean modelling

Different GRN models, such as ordinary differential equations (ODE) [15], neural networks [16], information theory model [17], Bayesian networks [18] and Boolean networks [19], have been proposed to reconstruct genetic regulatory networks. However, the current experimental methods are usually insufficient to identify large GRNs due to the lack of reproducibility for many genes involved in complex GRNs [20]. Even so, of the models, Boolean networks still attract much interest [21]. Boolean network models do not need information about kinetic parameters [22–27] and have explicit regulatory rules while carrying vital information [28]. However, Boolean models are still complex enough to review non-trivial behaviour among the genes, in general [29]. Boolean network modelling’s downside is that it may oversimplify biological signalling, where molecules often exist in multiple conditions with connections that are rarely binary [30].

Boolean modelling was initially presented by Kauffman et al. [31,32] in 1969, following discovering the primary gene regulatory mechanisms in bacteria [33]. A Boolean model consists of Boolean variables in either two binary states – On (1) or Off (0) as in digital circuits, denoting gene activation or inhibition. Each Boolean variable in a GRN represents a gene with its next state determined by a Boolean function.

The Boolean network’s fundamental premise is that the genes exhibit switch-like behaviour during the regulation of their functional states, ensuring the movement of a GRN from one state to another [3,22,34]. Hence, within signal processing theory, Boolean models can be transformed into electronic circuits, as shown in Fig. 2, to facilitate the study of Boolean networks’ rich dynamics [28].

By definition, the conventional Boolean network is wired in a series circuit format, as shown in Fig. 2. When there is a perturbation, the control switch could be On/Off, which turns the lamp On/Off. The resistor represents a functional rule that controls the lamp’s light intensity, i.e., controlling the lamp’s expression level. Because a conventional Boolean model only has two states, it can serve only as a series circuit, and the output of the circuit is either expressed or not expressed.

![Figure 1. Concept of gene activation and repression.](image-url)
Microarray data analysis

Affymetrix® GeneChip arrays are high-density oligonucleotide gene expression arrays mainly used in biomedical research. The DNA microarray technology, including the design of experiments to extract mRNA samples, has been applied to analyse human cancers, such as breast, prostate and leukaemia [35]. mRNA samples are hybridised using a gene chip, which contained a strand of all human genome genes, such as HG-U133 Plus 2. Raw gene data are extracted from image analysis by measuring the level of hybridisation on the chip. Complementary DNA (cDNA) microarray and oligonucleotide chips are the two approaches for manufacturing the microarrays. cDNA arrays are fabricated by robotic spotting on glass slides, and oligonucleotide arrays are developed by photolithographic chemistry and light-directed chemical synthesis on small glass plates [36].

Gene expression matrices are microarray data analysis, where rows denote genes and columns indicate samples. Microarray data analysis can be conducted by generating cell intensity (.CEL) files using the Affymetrix GeneChip Operating System Software (GCOS). The generated cell intensity files can then be converted into gene expression matrices using the R package affy [37]. An experiment may commonly involve extracting gene expression matrices at different time points with the same sample set. We will yield time series expression data if we reorganise the extracted gene expression matrices based on time points.

Time series expression data typically contain a series of \( m \) microarray expression measurements in the order of time points involving \( n \) genes. The gene expression data represent an \( m \times n \) table \( (\hat{T}) \) where \( m \) served as columns and \( n \) as rows [38]. There might be multiple samples, and each sample contains the same number of \( m \) and \( n \) but with different measurements. Combining all samples, it becomes a three-dimensional sample data space, \( \hat{S} \). Hence, the entry, \( e_{ij} \), in row \( i \) and column \( j \) of the table \( \hat{T} \) denotes the expression level of gene \( i \) in the \( j \)-th measurement of the sample, \( s \). Most data analysis is undertaken with a straightforward table, \( \hat{T} \) (matrix) such as using cell cycle analyses. However, data analysis on a three-dimensional sample data space \( (\hat{S}) \) might become more prevalent [38].

Meaningful temporal gene expression patterns can be extracted from the time series data and genes associated with each gene group. The relationships of gene groups can be modelled and depicted by GRNs, such as Boolean network modelling. Based on the availability of time points extracted from experiments, time series data can be categorised into two main groups: short time series with the number of time points fewer than eight and long-time series with the number of time points more than eight [39,40].

According to previous research [40,41], about 80% of published experimental data are short time series because the cost involved in acquiring microarray data is still high. Besides, the period of a patient’s treatment is usually either too short or fatal [42]. Even if the expense is not a concern, short time series experiments are still dominant because obtaining large quantities of biological material is prohibitive [40].

Traditional algorithms do not perform well on short time series data due to the lack of the time series’s required length [42,43]. The construction and validation of traditional models are also complicated [44]. Short time series data typically contain an enormous number of genes but only a few observations. Knowledge of the kinetic parameters and mechanical details cannot be inferred consistently from short time series data because the data are very noisy and contain various lengths of temporal observational gaps. Valuable information may be missing between the sparse observation gaps and may lead to incorrect conclusions.

Choosing the most suitable and dependable method to address a particular biological question from a specific dataset is a significant research question. One criterion is the capability to detect differentially expressed genes in terms of precision (specificity/variance) and accuracy (sensitivity/bias) [45].

Differentially expressed genes are highly dependent on the normalisation methods that alter how the correction structure from the data impacts the accuracy of cellular networks’ inference. Microarray normalisation typically involves three main steps: background correction that removes background noise from the signal intensities; data normalisation that eliminates non-biological variability between arrays and makes distributions across arrays; and summarisation, which
provide a single expression measure to each probe set in the array. The most common normalisation methods are MAS5.0, RMA (robust multichip average) and GCRMA (GeneChip RMA). MAS5.0 applies MM probes to adjust the PM probes for probe-specific non-specific binding for background correction. MAS5.0 uses a baseline array and scales all the other arrays to have the same mean intensity for normalisation and uses Tukey’s biweight function for summarisation [46]. RMA [47] applies a global correction, quantile normalisation and a median polish summarisation. The GCRMA [48] has been improved from RMA in that it uses the probe sequence information for background correction and is bias-corrected.

Leukaemia

Leukaemia is white blood cell-related disease driven by the cumulative mutations in the immature white blood cells from the bone marrow that reduce red cells, healthy white cells and platelets [49–53]. The cause of leukaemia is controversial due to multifactorial and exogenous or endogenous exposures and genetic (inherited) susceptibility [49–55]. Exposure to radiation and certain chemicals has been commonly linked to leukaemia, but evidence shows that these associations are only found in a very small minority of cases [54,55]. The redundant and unhealthy white blood cells enter the bloodstream and accumulate in organs such as the liver or spleen that could cause many problems [53,55]. For example, leukaemia presenting symptoms could be bruising or bleeding because of thrombocytopenia, pallor and fatigue from anaemia, and infection caused by neutropenia [55].

Leukaemia has been categorised into two main groups: childhood and adult. Childhood leukaemia can be divided further into two subtypes: acute or chronic. Most childhood leukaemia is acute. Acute childhood leukaemia can be divided into two groups: acute lymphoblastic leukaemia (ALL), in which lymphocytic cells are affected; and acute myelogenous leukaemia (AML), in which granulocytic cells are affected.

Apoptosis is a programmed cell death (PCD) process, also named ordered cellular suicide, which may happen in a multicellular organism as a controlled mechanism to maintain the balance of cell multiplication [56–58]. Introducing apoptosis in the aberrant white blood cells is common to stop cumulative mutations [56]. The apoptosis process in cells involves multiple biochemical events that lead to characteristic cell changes, such as cell shrinkage, blebbing, chromatin condensation, nuclear fragmentation, chromosomal DNA fragmentation and death [56]. Therefore, drugs like glucocorticoids (GCs) are commonly applied in chemotherapy. Glucocorticoids, a family of steroid hormones, contain synthetic products like dexamethasone (Dex) and prednisolone (PRD). Mainly, Dex is applied as an alternative to the natural human glucocorticoid cortisol [59]. GCs are essential steroid types of drugs commonly used to induce apoptosis in the malignant cells of childhood ALL during chemotherapy [59–61]. However, prolonged use of chemotherapy to introduce apoptosis may result in severe short-term or long-term side effects, such as osteoporosis, hypertension, psychosis, Cushing’s syndrome and leucopenia [41,58,62,63]. GCs enter into the leukaemia cell via a functional glucocorticoid receptor (GR), i.e., NR3C1 [64] ligand-activated transcription factor that exerts a pivotal role in inducing apoptosis in malignant lymphoid cells. The steroids are located in the cytosolic compartment in the absence of ligands [59]. When GRs bind with ligands on their high-affinity site in the carboxy-terminal portion, the glucocorticoid receptors (GRs) translocate to the nucleus and then bound with other transcription factors to regulate specific sets of genes [59]. However, GR alone is not sufficient for producing apoptosis. Accumulating evidence suggests that many leukaemic cells, which contain abundant quantities of normal GRs, are still unaffected by glucocorticoid-evoked apoptosis. For example, the steroid ligands could be blocked from passage through the plasma membrane and are destroyed biochemically-conjugated with GRs [59]. Besides, the resistant cells may have genetically or phenotypically altered the response systems to GCs to resist their lethal effects such as critical reductions in the quantity of one or more transcription factors, development of a dominant-negative form of such a factor or improper post-translational modifications of GRs or an interactive element [59]. The changes that affect the general pathways of apoptosis are: the alterations in the balance of pro- and anti-apoptotic members of the Bcl family of proteins; the loss of or inactivating mutations in caspases or other lethal proteases; the changes in one or more critical protease substrates rendering them, and the alternation in specific genes’ abilities to be regulated by ligand-driven GRs [59].

Currently, the transactivation or transrepression of target genes caused by GCs is still not well understood; primarily, the clinical effects of GCs are poorly understood [65]. For example, glucocorticoid (GC) resistance mechanisms in the clinical setting remain largely unresolved because the findings from the cell line model of GC resistance in ALL almost invariably exhibiting altered GR function are incongruous with those using specimens derived directly from a leukaemia patient [66]. Besides, GC signalling exerts a wide range of physiological actions because of the broad distribution. The activities include positive regulation of
metabolism in the liver, adipose tissue or the induction of apoptosis and cell cycle arrest, and anti-inflammatory effects in the immune compartment [64]. Another example is PFKFB2 (6-phosphofructo-2-kinase/fructose-2,6-biphosphatase-2) is widely believed as a crucial regulator of glycolysis that is induced more than 4-fold in all three T-ALL cases as well as in the T-ALL cell line CCRF-CEM [67]; however, Carlet et al. [67] suggested that the GC response gene, PFKFB2 (6-phosphofructo-2-kinase/fructose-2,6-biphosphatase-2) is not a critical upstream regulator of the anti-leukaemic effects of GCs.

To understand these drug-related genetic problems, scientists attempt to reconstruct the dynamics represented by time and the discrete state transition systems to gain insights into cell systems’ functioning [68–71]. These dynamics can be used to simulate the perturbations of new drugs in silico to reduce the potential risks of applying drugs to humans. Two common research issues are emerging for GCs: GC regulated genes and the glucocorticoid receptor gene network. Signalling pathways and gene networks can be inferred from gene expression data grouped in a time series format. The concept of Boolean modelling has been applied to the signalling pathways and gene network analysis from time series data.

**Fundamental Boolean modelling**

The hypotheses of conventional Boolean models do not deliver an intuitive technique to separate the individual activation and inhibition pathways [1]. The processes of gene activation and inhibition are the two fundamental processes of genetic regulation. For example, activation may result in substantial drug regulatory effects, such as modifications in the metabolism of in vivo substances and vitamins [72]. Likewise, inhibition may result in crucial clinical drug interactions formed by a wide range of drugs [72]. Inhibition can be classified into two groups: reversible inhibitors that can be easily inverted by dilution or dialysis since the interactions of this group are non-covalent with the enzyme surface and irreversible inhibitors that usually endure even during complete protein breakdown due to their sturdy covalent bonds on the enzyme surface [1,4].

Base on the theory of an enzyme reaction exposed to the action of a reversible inhibitor, the degree of inhibition may be modelled as the decreased rate of reaction divided by the uninhibited reaction rate [1,4]:

\[ i = \frac{V_o - V}{V_o} \]

where \( V \) and \( V_o \) represent the rates of the inhibited and uninhibited reactions, respectively [4]. The degree of inhibition \( i \) may present uncertainty into the target gene if the value of \( i \) is lower than 1. Similarly, enzyme activation contains the same concept as a reversible type of reaction. Hence, the degree of inhibition can be upgraded to the degree of the enzyme reaction, encapsulating the inhibition and activation degrees. For that reason, we could redefine the degree of the enzyme reaction to a conditional probability measure to represent the propensity rate of an enzyme reaction towards the target gene [1]. A conditional probability measure is the probability of an event that occurs given another event has happened. If the conditional probability measure is 1, the inhibitor is irreversible; otherwise, it is reversible [1].

Conventional Boolean models do not consider the reversible and irreversible behaviour of enzyme reactions. In biology, the disappearance of an activator does not preclude the emergence of an inhibitor because the proteins transcripted by a pre-activated gene might be still in the status of activation. The way we judge whether a gene activates or inhibits is based solely on the concentration rate of the proteins produced by the gene. Therefore, there are logical reasons to separate the general Boolean function into the domains of gene activation and inhibition [1].

At present, the leading emerging biological network inference methods to recognise functional modules are motivated either by the definition of gene regulatory networks or functional networks in which an edge indicates a functional relationship, and this is also a subset of entities that describe, explain or predict a biological process or phenotype [73]. Minimal effort has been made to construct activation, inhibition, and protein decay networks that could specify the direct functions of a gene or its synthesised protein as an activator or an inhibitor. To overcome the limitation of current conventional Boolean modelling, Chen et al. [1] proposed a novel Boolean model, denoted as fundamental Boolean modelling (FBM), to draw insights into gene activation, inhibition, and protein decay. The FBM can serve as both a series and a parallel circuit, as shown in Fig. 3.

The delay switch represents a gene decay function that might take a few time steps to turn the expression of its target gene entirely off if no inhibitors and activators are present. However, if any inhibitor exists (one with the inhibitor switched on), the target gene (the lamp) will be turned off immediately regardless of the presence of activators. The model shown in Fig. 3 is still a Boolean model as the series circuit shown in Fig. 2 because it has
the same Boolean output, i.e., expressed or not expressed. However, it wires the subfunctions of activation as a parallel circuit and the inhibition subfunction as a series circuit. One of the chief advantages is that we can split the fundamental Boolean network into an up-regulation network and a down-regulation network by removing all inhibitor or activator circuits.

Base on the concept of FBM, Chen et al. [1] extended the original definition of Boolean network modelling as a graph \( G(\mathbb{X}, E_a, E_d) \), where the node collection, \( \mathbb{V} = \{v_1, v_2, ..., v_n\} \), corresponds to a group of states, \( \mathbb{X} = \{x_i | i = 1, ..., n\} \) of size \( n \). Each node is a variable that is only in one of two states: \( \text{On}(1) \) or \( \text{Off}(0) \). The general edge set, \( E \), commonly found in traditional Boolean modelling, is divided into two sets of fundamental Boolean functions, \( E_a \) and \( E_d \), based on their regulatory functions, i.e., activation and inhibition, rather than a single function, as in all conventional Boolean models. The direction of the edges represents the propagation of their effectiveness on the target node, such as the signal flow between signalling molecules, genes or protein regulation. This graph, thus, conceptualised by Chen et al. [1] as a new type of Boolean network modelling, namely the fundamental Boolean network (FBN).

The two sets of fundamental Boolean functions, \( E_a \) and \( E_d \), are modelled as:

Fundamental Boolean functions of activation

\[
F_a = \left\{ f_a^i | j = 1, ..., l_a(i) \right\}, f_a^i : \{0, 1\} \rightarrow \{-, 1\}; \quad (1.a)
\]

Fundamental Boolean functions of inhibition

\[
F_d = \left\{ f_d^i | k = 1, ..., l_d(i) \right\}, f_d^i : \{0, 1\} \rightarrow \{-, 0\}, \quad (1.b)
\]

where \( F_a \) and \( F_d \) denote a set of fundamental Boolean activation and inhibition functions of gene \( i \), respectively. Notably, \(-\), here, refers to that the output of the function does not affect the target gene \( i \). \( l_a(i) \) symbolises the total number of fundamental Boolean functions activating the target gene \( i \). \( l_d(i) \) symbolises the total number of fundamental Boolean functions deactivating the target gene \( i \). When the Boolean activation function’s output is \( \text{On} \), the target gene \( i \) is activated, and \( \text{Off} \) means that the activation function does not influence the target gene \( i \). Similarly, when the Boolean inhibition function’s output is \( \text{On} \), the target gene is repressed and \( \text{Off} \) means that the inhibition function does not affect the target gene. The definition of the two types of Boolean functions set out the novelty of the proposed Boolean modelling [1].

The essential biological philosophies behind the fundamental Boolean functions are that a fundamental Boolean function can be treated as a simple transition rule. The rule takes a minimum required essential gene states as the input and then governs their regulation effects on the target gene [1]. In general, a fundamental Boolean function is an atomic function that cannot be separated any further. Hereafter, we can treat the concept of fundamental Boolean functions as conditions that constrain gene activity, a delegation of stereochimical reactions, and a transcription factor complex moulded by the transcription factor to proteins or protein to protein bindings [1].

The output of the proposed fundamental Boolean functions is only associated with the potential effectiveness of gene regulation on the target gene. For that reason, there is a need to calculate the level of confidence by what percentage we can trust the regulatory functions in affecting the target gene [1]. As stated previously, the degree of enzyme reaction can be substituted by the conditional probability that an enzyme reaction can influence the target gene. Hereafter, the concept of conditional probability can be used to measure the confidence of the proposed functions. The following formulae, called confidence measures, model the conditional probability of each fundamental Boolean function [1].

Confidence measure of activation:

\[
C_a\left[ f_a^i \left( A^i(t) \right) \right] = \frac{p \left( \sigma^{t+1} = 1 | A^i(t) = 1 \right)}{p \left( A^i(t) = 1 \right)}; \quad (2.a)
\]

Confidence measure of inhibition:
where \( \sigma_i^t \) denotes the Boolean state of gene \( i \) at time \( t \), and \( \sigma_i^{t+1} \) denotes the Boolean state of gene \( i \) at time \( t + 1 \). \( \wedge \) refers to a logical And connector. \( C_{d_i} \) and \( C_{a_i} \) delegate the confidence function with the input of the fundamental Boolean functions \( f_d \) and \( f_a \), respectively. \( A_{d_i} ^t \) and \( D_{d_i} ^t \) denote the set of inputs required or the state of the gene functions, \( f_{d_i} \) and \( f_{a_i} \), respectively. \( A_{d_i} ^t (t) = 1 \) or \( D_{d_i} ^t (t) = 1 \) mean the required gene input of \( f_{d_i} \) or \( f_{a_i} \) at time \( t \) is satisfied with the conditions of affecting the target gene, \( i \) [1].

There are various debates about mRNA/protein decay times in Boolean models. The decay time is the time that allows a gene to remain in the On state when there are no activators or inhibitors [1]. Albert [74] assumed that this decay might occur in two time steps. To capture the characteristics of protein decay, we induced a function \( f_{\text{decay}} \) to fulfill the requirements of protein degradation with input from the target gene \( i \) at time \( t \) [1]:

\[
f_{\text{decay}} (\sigma_i^t, \theta) = - (\tau \leq \theta) \times \sigma_i^t, \tag{3}
\]

where \( \tau \) represents an incremental variable presenting the number of time steps processed. \( \tau \) will be reset to 0 when there is any fundamental Boolean function affecting the target gene \( i \) at time \( t + 1 \). \( \theta \) delegates the decay period to reflect that the attenuation or enhancement of the mRNA expression requires time. \( \neg \) represents a negation operator that changes a Boolean function from On to Off or vice versa. \( \times \) is a logical And operator [1]. The output of the decay function \( f_{\text{decay}} \) is a Boolean state of On at time \( t + 1 \) if the gene state of \( \sigma_i \) of time \( t \) is On within the endured period or Off at time \( t + 1 \) when the tolerated period is expired regardless of the gene state of \( \sigma_i \) of time \( t \) [1]. In this study, the tolerated period \( \theta \) is set to one time step, i.e., \( \theta = 1 \) due to the experiment data are short time series [1].

By combining Eqs. (1.a), (1.b), (2.a), (2.b) and (3), Chen et al. [1] defined the novel Boolean model (FBM) to calculate a gene state \( \sigma_i \) at time \( t + 1 \) based on the immediately previous time \( t \) as

\[
\sigma_i^{t+1} = \left( f_{\text{decay}} (\sigma_i^t, \theta) + \bigvee_{j=1}^{l=1} \left[ P\left[ C_{d_i} \left[ f_{d_i}^j \left( A_{d_i}^j (t) \right) \right] \right] \right] \right) \times \left( \bigwedge_{k=1}^{l=1} \left[ P\left[ C_{a_i} \left[ f_{a_i}^k \left( D_{a_i}^k (t) \right) \right] \right] \right] \right), \tag{4}
\]

where \( + \) is a logical Or operator and \( \times \) is a logical And operator. The decay function \( f_{\text{decay}} (\sigma_i^t, \theta) \) in Eq. (3) is to ensure the gene state \( \sigma_i \) at time \( t + 1 \) depends on the prestate of the gene at time \( t \) if no activators are present at time \( t \) and they are still tolerated by the parameter \( \theta \), a decay period. \( P[x] \) is a Boolean function that takes a uniform distributed random number, \( \mu \), and outputs a value of On if \( \mu < x \) and Off otherwise. \( V \{ x \} \) denotes the logical connective function of On, i.e., \( V \{ x \} = P\left[ C_{a_i} \left( f_{a_i} \right) \right] + P\left[ C_{a_i} \left( f_{a_i} \right) \right] + \cdots + P\left[ C_{a_i} \left( f_{a_i} \right) \right] \). The FBM defines rules on how a gene’s state can be transited from \( t \) to \( t + 1 \) based on its activation \( (E_a) \) and inhibition \( (E_d) \) functions. FBM shows how genes can be regulated via their activation \( (E_a) \) and inhibition \( (E_d) \) functions in a graph. Figure 4 shows an example of a fundamental Boolean network that includes up-regulation (activation regulation) and down-regulation (inhibition regulation) networks with a regulatory time step of 1:
Figure 4 also shows the fundamental Boolean functions used to construct the example FBN. Gene 1 can be activated among the fundamental Boolean functions if its previous state was promoted, but it can also be inhibited if its previous state was deactivated; Gene 2 can be activated if the previous states of Gene 1 and Gene 5 were activated and Gene 4 was deactivated. The inhibitors of Gene 2 show that Gene 2 can be inhibited by Gene 4 when its previous state was activated. Gene 2 can also be inhibited by Gene 1 if its previous state was deactivated. Gene 5 is also an inhibitor of Gene 2 if Gene 5’s previous state was deactivated.

To proof, the concept of FBM, Chen et al. [1] conducted an experiment to demonstrate that, under synchronous Boolean model schema, FBM can produce the same result as the traditional Boolean modelling. For example, Chen et al. applied the proposed FBM with the mammalian cell cycle and yielded the same attractors as that have been reported by [6,75], as shown in Fig. 5. Attractors refer to the recurrent cycles of the states [75] and are of particular interest in Boolean modelling. Once a network reaches an attractor, it is entrapped in a cycle that repeats until an external perturbation happens to change some of the production of the essential genes of the attractor to let the network come out from entrapment. In the study of [1], this outcome confirmed that FBM and FBN are novel extensions of the traditional Boolean modelling and networks.

**Temporal fundamental Boolean modelling**

The original FBM we defined in Eq. (4) provides a mechanism to calculate gene state $\sigma_t$ at time $t+1$ based on the immediately before time $t$; however, some gene regulations might require more time steps to complete than others. Silvescu and Honavar [38] extended the traditional Boolean network modelling to temporal Boolean network modelling that transforms the Boolean Networks from a Markov(1) to Markov($T$) model, where $T$ is the length of the time window during [38]. Silvescu and Honavar explained that a gene state at $t+1$ should not only depend on the inputs immediately before $t$ but also $t − 1, ⋯ , t − m$ ($1 ≤ m < t$), $m$ here refers to the maximum temporal decrement value [38].

A similar but not identical to the concept of Silvescu and Honavar [38], we propose to extend the original fundamental Boolean network modelling to temporal fundamental Boolean network modelling as graph $G(X, E_a, E_i, T)$, where $T$ is the best time step in the time window during which a gene can be regulated by corresponding $E_a, E_i$. We denoted this extension model as the temporal fundamental Boolean model (TFBM) and its associated network as the temporal fundamental Boolean network (TFBN). Unlike the temporal Boolean network modelling proposed by Silvescu and Honavar [38], in which a gene’s state depends on several previous time steps, TFBM’s concept is that a gene at time $t + 1$ only depends on the best previous time step that has the best statistical measurements on its activation and inhibition functions. Statistical measurements are used to mine the fundamental Boolean functions from time series data such as confidence measure, confidence counter measure, and conditional causality test measure as discussed in [1]. By extending the Eq. (4), TFBM can be defined as

$$
\sigma^{t+1}_i = (f^{\text{decay}}(\sigma^t_i, \theta) + \sqrt{\lambda(t)} \left\{ \prod_{j=1}^{\lambda(t)} \left[ C_{\alpha(j)} \left( f^{\alpha(j)}_a (A^{(j)}(T^j_i)) \right) \right] \right\} \\
\times \sqrt{\lambda(t)} \left\{ \prod_{k=1}^{\lambda(t)} \left[ C_{\alpha(k)} \left( f^{\alpha(k)}_i (D^{(k)}(T^k_i)) \right) \right] \right\}, \quad (5)
$$

where $T^j_i$ and $T^k_i$ are the best previous time step for the activation function $f^{\alpha(j)}_a$ and inhibition function $f^{\alpha(k)}_i$ of gene $i$, respectively. $A^{(j)}_i$ and $D^{(k)}_i$ represent the set of required inputs for the gene state functions, $f^{\alpha(j)}_a$ and $f^{\alpha(k)}_i$ at the best previous time step $T^j_i$ and $T^k_i$, respectively.

To calculate the best previous time step $T^j_i$ or $T^k_i$ for the extended model, we need to calculate all previous’s measurement matrix (input genes’ statistical states) that could derive the target gene $i$ at $t + 1$, up to $t − m$ level. Let us define a measurement matrix as $\hat{A}_i$ for the activation function $f^{\alpha(j)}_a$, and a perfect vector $\hat{A}_i$, which
contains the best value for all measurements:

$$
\tilde{A}_i = \begin{pmatrix}
\text{measurement}_{i}^{t_0} & \text{measurement}_{i}^{t_1} & \cdots & \text{measurement}_{i}^{t_m} \\
\text{measurement}_{i}^{t_0} & \text{measurement}_{i}^{t_1} & \cdots & \text{measurement}_{i}^{t_m} \\
\vdots & \vdots & \ddots & \vdots \\
\text{measurement}_{i}^{t_0} & \text{measurement}_{i}^{t_1} & \cdots & \text{measurement}_{i}^{t_m}
\end{pmatrix}
$$

where $e$ denoted as a measurement of the matrix $\tilde{A}_i$. The perfect vector $\tilde{A}_i$ is a one column measurement matrix that contains the best value of all measurements. For example, the best value for the confidence measure is 1, and the confidence counter measure is 0.

Hence, the best previous time step value of the activation function $f^0_i$ can be calculated based on the shortest distance between measurements on time step $t$ to $t - m$ and the perfect vector $\tilde{A}_i$:

$$
T'_{i} = \min \{ \text{dist}_{t-m}(\tilde{A}_i, \tilde{A}_i) \}.
$$

$T'_{i}$ illustrates a simple method to find the best previous time step value where dist() is a Euclidean distance function. However, it is unnecessary to set $m$ to the value of the total previous time points less one as the biological reaction might only need a few time steps to complete. It might be common to set the maximum decrement value ($m$) to two or three because about 80% of time series data are short time series data in which the sparse gap between each time step might not support the hypothesis that a regulation process of a gene might take more than two or three time points to complete.

TFBM may handle short time series better because it evaluates more time points than the initially proposed model [1]. Furthermore, it reflects the reality that most biochemical reactions are asynchronous since each gene may be updated in different timescales. For example, a gene could be regulated by an activation function at $t - 2$, and an inhibition function could regulate other genes at $t - 1$.

### Fundamental Boolean network inference

There are two main steps required to infer fundamental Boolean networks. The initial step is to construct a cube type database to store all critical pre-computed measures, and the second step is to search for the best Boolean functions from the cube [1]. Hence, the network inference process is separated from constructing the cube and identifying the Boolean rules from the cube. The separation between the network extraction and construction of the cube enables further development of scalable methods to infer genetic networks effectively and efficiently because a cube has comparatively fewer updates, although it can be consistently enhanced by feeding it more time series data [1].

A data cube is a data abstraction providing a mechanism to analyse aggregated data from multiple dimensions. A data cube can also be regarded as a collection of identical 2-D tables stacked one upon the other. For example, many standard genetic time series data are multi-dimensional and involve the three main dimensions of genes, time steps, and samples. Researching multi-dimensional data could entrap performance bottlenecks [1].

To mine the fundamental Boolean networks, Chen et al. [1] extended the data mining technique of bottom-up computation (BUC) to a prefix tree type of cube, namely, Orchard cube, as shown in Fig. 6. BUC is an algorithm designed to compute sparse cubes from the Apex cuboid downward [76].

Every branch or link of a tree above ground is referred to as a regulatory function. Each node on a branch contains possible regulatory functions. Due to the regulatory functions being the information we are searching for, we call them fruit. The gene nodes on the ground are named seeds. The training data are called fertilisers as they aid the trees to grow more significant (more confident and, hence, more satisfied with the functions). This cube can distribute the computational costs to multiple computing nodes in a cloud computing environment because each branch can be calculated independently. Moreover, the pre-computing cube can persist in any distributed database, so inferring networks from the cube is straightforward [1].

As shown in Fig. 6, the mined result of G1 has two activation functions and one inhibition functions: The G1 can be activated by itself if its previous state was activated; G1 can also be activated by G2 and G4 if G2 was activated and G4 was inhibited; G1 can be inhibited if its previous state was deactivated. Figure 7 presents a schematic diagram of the fundamental Boolean network inferences [1].

### Network types of fundamental Boolean model

The novel Boolean model, i.e., the fundamental Boolean model and the related Boolean network, provides a mechanism to intuitively analyse the activation, inhibition, and protein decay pathways. We outline the main subtypes of the novel Boolean networks that could be applied to investigate the drug-related gene
regulations because a novel drug’s inhibition pathways can be exposed intuitively through an investigation of activation or inhibition related downstream cascade networks. These subtypes of fundamental Boolean networks are the principal characteristics that differ FBM from other traditional Boolean networks. There are six subnetwork types derived by the novel fundamental Boolean modelling and networks, as shown in Fig. 8.

• FBNet_FAA (type 1): the input genes are up-regulated, and their target genes are up-regulated, denoted as the forwarding regulatory pathway of type 1. The subnetwork type 1 shown in Fig. 8 presents an example to answer the question of that if Gene 1 is an input gene and is activated, then what the up-regulation network should look like, as a downstream effect. In this case, Gene 1 activates Gene 2.

• FBNet_FAI (type 2): the input genes are up-regulated, and their target genes are down-regulated, denoted as the forwarding regulatory pathway of type 2. The subnetwork type 2 shown in Fig. 8 presents an example to answer the question of that if Gene 1 is an input gene, and is activated, then what the down-regulation network should look like, as a downstream effect. In this illustration, the activated Gene 1 and Gene 5 inhibit Gene 4.

• FBNet_FIA (type 3): the input genes are down-regulated, and their target genes are up-regulated, denoted as the forwarding regulatory pathway of type 3. The subnetwork type 3 shown in Fig. 8 presents an example to answer the question of that if Gene 1 is an input gene and is inhibited, then what the up-regulation network should look like, as a downstream effect. In this case, the inhibited Gene 1 activates Gene 4.

• FBNet_FII (type 4): the input genes are down-regulated, and their target genes are down-regulated, denoted as the forwarding regulatory pathway of type 4. The subnetwork type 4 shown in Fig. 8 presents an example to answer the question of that if Gene 1 is an input gene and is inhibited, then what the down-regulation network should look like, as a downstream effect.
In this example, inhibited Gene 1 continually down-regulates itself, and it also deactivates Gene 2.

- FBNNet_BA (type 5): the backward regulatory pathway of activation. i.e., the networks that drive a target gene to be activated denoted as type 5. The subnetwork type 5 shown in Fig. 8 presents an example to answer the question of that if the Gene 4 is the target gene and is activated, then what causes the target gene to be activated as an upstream effect. In this case, Gene 4 can be activated by either Gene 3 & not Gene 1 or Gene 3 & not Gene 5.

- FBNNet_BI (type 6): the backward regulatory pathway of inhibition. i.e., the networks that drive a target gene to be inhibited denoted as type 6. The subnetwork type 6 shown in Fig. 8 presents an example to answer the question of that if the Gene 4 is the target gene and is inhibited, then what causes the target gene to be inhibited, as an upstream effect. In this case, Gene 4 is deactivated by either the transcription factor type component made from Gene 5 & Gene 1 or not Gene 3.

**RESULTS**

The networks (TFBN) extracted from the constructed Orchard cube contain 285 components and 2775 Boolean rules, separated by activation and inhibition. Appendix C shows the common annotated genes, and Appendix D shows the complete fundamental Boolean networks. For brevity, we only discuss a few genes in the following subsections. As discussed in [1], the fundamental Boolean model splits the Boolean functions into gene activation and inhibition domains. Hence, we started to explore the gene networks by filtering the extracted TFBN (see Appendix D) and then plotted their network graphs within the six types (We skipped the general type as it contains too many nodes).

As shown in Table 2, gene CDC42EP3, clustered as a membrane-type gene (Table 3), was the only one that highly expressed in all B-All samples, and genes SCML4, DDIT4, SLA, PFKFB2, CDKN3, ZFP36L2, FKBP5, SNX29P2 (pseudogene), PIK3IP1, TNFSF8,
PTTG1, MCM4, MIR8071-1 (non-coding RNA) were differentially regulated in all T-ALL samples and the majority of B-All samples (6/13). The 11 coding genes (excludes the non-functional genes SNX29P2 and MIR8071-1) may suggest that the T-ALL samples may be more sensitive to GC treatment than the B-ALL samples. Genes PFKFB2, BTNL9, SNF1LK, FKBP5, ZBTB16, KIF26A, SLA, SOCS1, DDIT4, GBP4, MGC17330, ZFP36L2, EPPK1, P2RY14, FGR, WFS1, ARPP-21, SERPINA1, GIMAP7, MYCPBP and LGALS3 are the key GC-regulated genes reported in [2] but also appeared in the common gene list. Notably, as shown in Table 3, 72 common genes belong to the cell cycle, and 50 of them also belong to cell division. Gene CDC45 belongs to the two classes: signal and cell cycle, which might indicate the gene CDC45 is the bridge type connector between signal and cell cycle classes. Indeed, CDC45 is an essential protein.
required to initiate DNA replication [77]. Hence, the 285 common differentially expressed genes across three different normalisation methods do encapsulate important genes. Additionally, the genes ABHD17B, BCL10, CPM, EGR1, ELL2, KCNK12, PFKFB2, RASSF4, SNTB2, ZFP36L2, identified especially by TFBN, have rules of two time steps. These genes confirmed that some genes require more time steps to complete their biochemical reactions, and the proposed TFBN did extract these regulations.

**Networks of CDC42EP3**

The product of CDC42EP3 belongs to BORG family proteins, and overexpression of Cdc42EP3 in fibroblasts can enhance the formation of pseudopodia and F-actin-containing structures [78]. However, BORG proteins’ role in the tumour microenvironment is still unclear [78]. Therefore, we interpret the fundamental Boolean networks of CDC42EP3 extracted from this experiment.

As shown in Table 2, gene CDC42EP3 was highly expressed in all B-All type samples and was a common gene at 0–24 h and 6–24 h. Hence, we are interested in finding out what activated this gene and the consequences of the activation. The sub-networks type 1, type 2, and type 5 networks should address this question.

**Figure 9** shows that activated CDC42EP3 up-regulates genes EPPK1, F13A1, FGL2, LGALS3, NPCDR1, PPBP, PRDM1, RAB31 and STAB1. Five out of nine genes were documented in the B-ALL gene network list from Chaiboonchoe [41]: EPPK1, FGL2, LGALS3, PRDM1 and STAB1. In addition, PPBP was documented in the list of T-ALL gene networks list from [41]. The remaining three genes, F13A1, NPCDR1 and RAB31, are new findings that had not been reported before that could be up-regulated by CDC42EP3 in either B-All samples or T-All samples.

**Figure 10** shows activated CDC42EP3 down-regulates genes CCDC86, CRNDE, MDK, MTHFD2, RBM14, and SNORA21. These genes have not been reported in previous studies. This could be the reason that down-regulation may be more challenging to detect than gene induction [2]. However, with the fundamental Boolean model’s facilitation, identifying down-regulation networks is straightforward as well as identifying up-regulation networks.

**Figure 11** shows the genes that activate CDC42EP3: RAB31, PPBP, LGALS3, and FGL2. LGALS3 and FGL2 have been documented in B-ALL gene networks, and PPBP has been documented on T-ALL gene networks [41]. RAB31 was a new finding that has not yet been reported to regulate CDC42EP3 in previous studies. Thus, new gene regulations and potential side effects could have been identified through types 1, 2 and 5 regulatory networks.

**Networks of the four genes induced across all periods**

Genes PFKFB2, BTNLT9, FKBP5 and P2RY14 have been reported in Table 2 that have been induced across the three time spans of 0–6 h, 0–24 h and 6–24 h.

As discussed in the background section, GC induced apoptosis by influencing hormone metabolism. BTNLT9 (butyrophilin like 9), as shown in Fig. 12, was found to be highly expressed in B-ALL samples, indicating that its pathway could only affect B-All type patients. BTNLT9 activated EPPK1, and they were turned on by the overexpression of gene BMF and the underexpression of BCL2L11. BMF and BCL2L11 both belonged to the BCL2L11 family. As documented in [2], genes LDHA, GPR65, MAP2K3, GZMA, MYC, NR3C1 and BCL2L11 were the top candidate genes.

**Figure 13** shows the backward regulatory networks of PFKFB2. The gene PFKFB2 can be enhanced by the gene TNFRSF21, which is in the clusters of signal and immunity (Table 3), in two time steps (The unique features of TFBN). PFKFB2 can be inhibited by IL6ST, a transmembrane type of gene with multiple functions such as signal (Table 3) in one time step. The functionality of LOC100996643, which is pseudogene, is not clear, and hence it is not important in this study. We conclude that the critical gene PFKFB2 could be activated by the signal gene TNFRSF21 in two time steps and inhibited by the other signal gene IL6ST in one time step. Hence, the mechanism of how the gene PFKFB2 to be regulated can be revealed by the novel Boolean model networks TFBN.

**Figure 14** shows the backward regulatory networks of PFKFB2 under the original fundamental Boolean modelling where the previous time step is fixed at one time step. However, the genes TNFRSF21 and MYRIP, shown in Fig. 13 that can regulate PFKFB2 in two time steps, now disappeared under the original fundamental Boolean modelling. Hence, with the same short time-series data, the temporal fundamental Boolean network modelling can uncover more genetic regulation rules that might require more time steps to complete their biological reactions than the original FBM.

As shown in Fig. 15, among the four induced genes, only P2RY14 and BTNLT9 were found to inhibit their target genes when activated. The target inhibited genes LOC100505650 (uncharacterized [Homo sapiens]), CENPU (transcription regulation), SQLE (membrane and transmembrane), BCA1 (transferase), and PRPS2 (ATP-binding, transferase and metal-binding) may be the target responsible genes to be suppressed, of the GC treatment.
Networks of CDC45

As mentioned previously, the gene CDC45 is a connector between signal and cell cycle groups. Figure 16 shows the network of CDC45 that contains forward and backward regulations.

As shown in Fig. 16, CDC45 can be inhibited by genes ASF1B (transcription regulation), CCDC34, AURKA (cell cycle), BTG1 and APITD1-CORT (cell cycle), and activated by genes IFNGR1 (receptor, glycoprotein, membrane and signal), MDK (signal), CENPV (cell cycle), IL1B, BMF (BCL2L11) and E2F7 (cell cycle and transcription regulation). The downstream of CDC45 is that CDC45 inhibits CDT1 (cell cycle), TTK (transferase), HELLS (cell cycle, cell division and ATP-binding), CHEK1 (cell cycle, P53 pathway), TBXA2R (membrane), ZWINT (cell cycle), METTL7A (membrane, signal), ID2 (transcription regulation) and BRIP1 (ATP and metal-binding) and activates DTL (Ubl conjugation pathway, membrane), KNL1 (cell cycle and cell division), SQLE (membrane) and BRIP1 if the gene P2RX5 is inhibited.
Figure 10. Type 2 networks of CDC42EP3.

Figure 11. Type 5 networks of CDC42EP3.
Figure 12. Type 1 subnetworks of genes PFKFB2, BTN9, FKBP5 and P2RY14.

Figure 13. PFKFB2 backward regulatory networks.
Figure 14. PFKFB2 backward regulatory networks under the original FBM modelling.

Figure 15. Type 2 subnetworks of P2RY14 and BTNL9.
DISCUSSION

As shown in Fig. 9, five out of nine genes are associated with B-All samples, which meant that the gene CDC42EP3 is mainly associated with B-ALL type of patients. F13A1 encoded a protein (glycoprotein) for coagulation Factor XIII A chain, the last zymogen to become activated in the blood coagulation cascade [79]. Diseases associated with F13A1 included Factor Xiii (8), a subunit deficiency of Factor Xiii (8) [79]; NPCDR1, nasopharyngeal carcinoma, RNA gene and the diseases associated with NPCDR1, including nasopharyngeal carcinoma [79]; RAB31, a member of RAS oncogene family, was associated with diseases including estrogen-receptor-positive breast cancer [79]. Hence, we suspected that the up-regulation of the three genes F13A1, NPCDR1 and RAB31 could cause side effects under GC induced apoptosis. To inhibit them, we may consider disabling their conditional genes, such as turning on RHOBTB3 and turning off FGD2, to prevent F13A1 from being activated.

As discussed in Fig.12, the BCL2L11 and Bcl-2 rheostat were proven to induce GC that led to cell death. The target gene of the activated BTNL9 was EPPK1 (its related pathway was cytoskeleton remodelling neurofilaments). EPPK1 could be associated with leukaemia healing because EPPK1 can accelerate keratinocyte migration during wound healing. Gene FKBP5 activated KCNK12 while the GC essential gene SLA was inhibited. The stimulated purinergic receptor (P2RY14) activated TUBA4A, which was connected to the diseases of amyotrophic lateral sclerosis 2 with or without frontotemporal dementia, robinow syndrome and autosomal dominant 3 [80]. The activated TUBA4A could be the side effect of GC-related treatment, but under two conditions, gene TENM4 must be inhibited, and gene GBP4 must be activated.

Moreover, the GC related gene PFKFB2 [2,67] activated the critical gene, DDIT4 (DNA damage-inducible transcript 4), an essential candidate for GC-induced apoptosis. With the inhibition of MYRIP, PFKFB2 is self-activated, indicating that the GC treatment turned PFKFB2 on by inhibiting gene MYRIP, which coded the myosin VIIA and Rab interacting proteins. The MYRIP-related pathway was through peptide hormone metabolism. Hence, we disagree with the suggestion of Carlet et al. [67] that the PFKFB2 is not a critical upstream regulator of the anti-leukaemic effects of GCs and suggest that PFKFB2 is a critical upstream regulator of GC.

As illustrated in Fig. 16, BMF and BCL2L11 belong to the BCI-2 family and perform as a central regulator of the intrinsic apoptotic cascade and mediates cell apoptosis [81]. Hence the inhibition of BMF may trigger the GC related apoptosis. The TBFN of CDC45, as shown in Fig. 16, then provides insights into how the mechanism of BMF works. When BMF is inhibited, it will activate the key gene CDC45 and then kick off the intrinsic apoptotic cascade by activating DTL, KNL1, SQLE and BRIP1, to mediate cell apoptosis. Notably, the three genes KNL1 [82], SQLE [83] and BRIP1 [84] are associated with the apoptosis of the cancer cell. However, the activation of DTL could be a serious side effect of GC related treatment because the overexpression of DTL is significantly up-regulated in cancer tissues than in normal tissues [85]. Furthermore, Cui et al. [85] pointed out that higher DTL expression owned a lower survival rate.

In summary, we pointed out some potential side effects and discussed some new findings. These could be useful for pharmaceutical agents as well. New hypotheses could be identified by analysing the extracted fundamental Boolean networks and analysing their up-regulatory and down-regulatory pathways. The subnetwork types show that the fundamental Boolean networks could easily split the up-regulation systems and down-regulation without applying other tools or previous knowledge about networks. We can also find what causes the genes (induced and repressed) to be activated or inhibited by finding their backward regulation. For example, we pointed out three genes, F13A1, NPCDR1 and RAB31, which could be side effects of GC introduced apoptosis, as activated by CDC42EP3. Besides, we disagreed with the suggestion of Carlet et al. [67] that the PFKFB2 is not a critical upstream regulator of the anti-leukaemic effects of GCs and concluded that PFKFB2 is still a critical upstream regulator of GC based on this study. Moreover, we discovered insight into the role of CDC45 and found that if it is activated, it will trigger cell apoptosis by activating the three apoptosis-related genes KNL1, SQLE and BRIP1. However, the activation of CDC45 also activates the gene DTL that could be a serious side effect of GC related treatment because higher DTL expression owned a lower survival rate.

CONCLUSIONS

The previous study [1] investigated the physiognomies of enzyme activation, enzyme inhibition and protein decay, then proposed a novel data-driven Boolean modelling, namely, the FBM and FBN, to draw insights into gene regulatory networks. The FBM separates the activation and inhibition functions from conventional Boolean functions, and this separation could facilitate scientists in seeking answers in such as how an amendment of one gene distresses other genes at the
expression level. The previous study also proposed a new data-driven method to infer FBNs. The novel method comprises two different parts: the first part was to construct an orchard cube to persist all pre-computed measures for all potential fundamental Boolean functions; the second part was to infer FBNs from the orchard cube constructed by filtering each tree’s underground part, based on different criteria [1].

This paper extended the FBM to temporal fundamental Boolean modelling (TFBM) to address dependencies among the state transition of genes that could span more than one unit of time. During the study, we produced the temporal fundamental Boolean networks (TFBNs) based on TFBM, as shown in Appendix D, on the childhood acute lymphoblastic leukaemia data, which were produced in clinic settings. The networks may be useful for pharmaceutical agents to identify any side effects when applying GC induced apoptosis on children. For example, the genes ABHD17B, BCL10, CPM, EGR1, ELL2, KCNK12, PFKFB2, RASSF4, SNTB2, ZFP36L2 have been identified by the TFBNs that have rules of two time steps. These genes confirmed that some genes

![Figure 16. Networks of CDC45.](image)
require more time steps to complete their biochemical reactions, and the proposed TFBM did extract these regulations. Hence, the proposed TFBM unlocks their predecessor’s limitation, i.e., FBM.

The traditional Boolean attractor study is not included in this paper because searching attractors in the 285 common genes are not feasible. The current version of the R package FBNNNet cannot handle searching attractors with a large gene set under TFBM and hence reserved for future improvement. The findings reported in this paper are experimental hypotheses only due to the limitation of time series data availability, i.e., short time series data. Any insight gained from the modelling effort must be proved experimentally before any medical applications. This study demonstrates the availability of constructing large GRNs from clinic data of short time series using fundamental Boolean modelling [1].

The proposed concept of fundamental Boolean modeling (FBM and TFBM) and related networks (FBNs and TFBNs) are novel, and hence, they do need further research on how to apply them to clinic data. In this paper, we only discuss a small set of critical genes for leukaemia, and hence it requires further analysis of the networks attached in Appendix D. Moreover, the unpublished R package FBNNNet (1.0 and 2.0) are developed as a prototype specifically for this study, and hence, it requires further work to make it publishable.

MATERIALS AND METHODS

Before the experiment conducted by Schmidt et al. (2006), the investigations had led to some conflicting hypotheses, which have not yet been tested in a clinical setting [2]. Hereafter, Schmidt et al. generated 13 comparative wholegenome expression profiles (purified at three time points) using lymphoblasts from 13 GC-sensitive children under therapy for ALL [2]. Consequently, a substantially complete list of GC-regulated candidate genes in clinical settings and experimental systems has been generated to immediately analyse any gene for its potential significance to GC-induced apoptosis [2]. Schmidt’s study identified a small number of novel candidate genes (22 genes); however, this study was inconsistent with most model-based hypotheses [2].

The data generated by Schmidt et al. (2006) are short time series, but they were still valuable. Researchers have continually analysed the data and proposed novel hypotheses or questions, such as the study in [41,86,87], in which more novel glucocorticoid-regulated genes were identified through inferred GC-regulation networks. The newly identified genes may pave the way to develop new chemotherapy drugs that have fewer side effects. However, the models Chaiboonchoe et al. [86] applied were based on emerging clustering methods, such as self-organising maps (SOMs), emergent self-organising maps (ESOM), the short time series expression miner (STEM) and fuzzy clustering by local approximation of membership (FLAME).

Dataset and pre-processes

Raw data (GEO accession code: GSE2677), provided by Schmidt et al. [2], had been downloaded from the website of the NCBI (National Centre for Biotechnology Information). The data contained gene expression measurements for 13 samples, and each sample has three time points: 0 hour (before GC treatment), 6/8 hours (after GC treatment) and 24 hours (after GC treatment). Among the 13 samples, three samples from T-ALL patients, and the rest were B-ALL patients.

The common pre-process of analysing gene expression data involves data normalisation, selection of differentially expressed genes, and data discretisation. In the previous sections, we briefly discuss the methods of normalisation. In general, gene expression has two main basic patterns: underexpression (down-regulation) and overexpression (up-regulation). Overexpressed genes have higher expression values when two samples are compared, for example, cancer (target) and healthy. On the other hand, underexpressed genes have lower expression values in the target than in reference samples [88]. Commonly, a gene with more than twofold changes is considered significant or differentially expressed.

Data discretisation is a process of converting continuous data attribute values into a finite set of intervals such as 0 or 1 in Boolean modelling, with minimal loss of information [89]. Different emergent discretisation methods have been developed to address different needs [89]. Example of the traditional approaches for data discretisation are equal-width [90], equal-frequency [90,91], K-means [92,93], graph-theoretic based discretisation [94] and decision trees [95].

Different methods have advantages and disadvantages under different conditions, such as K-means performing poorly when the clusters are of different shapes, sizes, and density [93]. The well-developed R package BoolNet [96] provides a function, namely binarizeTimeSeries, to convert continual time series data into Boolean time series data using K-means clustering, edge detection, or scan statistics. K-means clustering is better and accurate than edge detection [97]. Wheeler [98] compared Kulldorff’s spatial scan statistic, K function, Cuzick and Edward’s method and the kernel intensity function to test for significant local clusters in childhood leukaemia in Ohio and concluded the spatial scan statistic in SaTScan found no significant clusters but others did. Hence, in this study, we applied the function binarizeTimeSeries with K-mean to discretise the
extracted differentially expressed time series data.

This study selected R software as the leading platform to analyse genetic networks. R (http://www.r-project.org) is an open-source platform for statistical computing, developed by Ross Ihaka and Robert Gentleman at the University of Auckland in New Zealand. We applied the following leading tools to conduct the computational experiments described in this paper:

- FBNNET, unpublished R package [1], version 2.0.0. A package implemented explicitly for the FBM and the temporal fundamental Boolean model.
- BoolNET, A R package for analysing conventional Boolean networks [96].
- RMA (robust multi-array average) is a method that converts probe level data (CEL files) into a gene expression measure.
- GCRMA (GeneChip RMA) is an improvement from RMA that uses the probe sequence information for background correction and is bias-corrected (Wu & Irizarry, 2004).
- Gene annotation via DAVID bioinformatic database [99,100].

Inferring genetic networks from the whole genome is usually very time consuming and very difficult to achieve. The biologist or pharmacist needs a small subgroup of differentially expressed genes for further experiments. The original dataset contains more than 30,000 genes/proteins; hence, we need to identify the significantly expressed genes and use these genes to construct FBN networks. Therefore, we only choose a few critical genes from the inferred fundamental Boolean networks to discuss.

Differentially genes

Differentially expressed genes are calculated based on comparing two different time points (e.g., the gene expression at 0 h and 6/8 h) as log-ratio (base two), namely, fold change or cutoff threshold. The M-values were denoted as regulation values, and the E-values were indicated as normalised expression values. The cutoff point represented the threshold for fold changes. For example, a cutoff of 1 means M values of \( \geq 1 \), representing a two-fold regulation; a cutoff of 0.7 means a 1.4 fold regulation; a cutoff of 2 means a four-fold regulation. The majority value means the fold regulation happens in at least a specific number of samples out of total samples, such as the fold regulation happens in at least six out of 13 samples (6/13) documented in [2].

The data contain 39 CET files, created by Affymetrix DNA microarray image analysis software, from the GenBank. The original file names have been renamed with the format of type-sample number-time such as B-ALL-13-0h.cel. First, the data have been normalised using RMA and GCRMA, respectively. GCRMA is a method to convert background adjusted probe intensities to expression measures using the same normalisation and summarization approaches as RMA and is bias-corrected [48]. It is straightforward to normalise the raw data using affy.gcrma, provided by the Bioconductor project (http://www.bioconductor.org) via R. Secondly, we computed the differential expressed genes for the 13 samples (three T-All samples and 10 B-All samples) with the criteria of the cutoff at 0.7; the majority were six out of 13 samples, using three different methods:

- Re-analyse the M-values provided by Schmidt et al. [2].
- Normalise the data using RMA.
- Normalise the data using GCRMA.

The original research focuses on a small gene set and less significant (less than 2 fold changes). Although calculating differentially expressed genes based on fold changes has been criticised for its propensity to variation or unreliability because the method does not consider the variability of inter-experiment noise and outliers [41], it is a common method to produce differentially expressed genes such as applied in the study of Schmidt et al. [2]. This study focuses on reanalysing the data with TFBN modelling with differentially expressed genes based on fold changes, and the reliability of any insight gained from this study still requires further research to verify the network result.

Table 1 presents the results of differentially expressed genes that were identified by the three different methods. Re-analysed M-values figures were identical to the study conducted by [41] because we used the same criteria to select the differentially expressed genes.

Interestingly, the results from RMA under R version

|表 1 GC-regulated differentially expressed genes, M\(\geq\)0.7, 6/13 (6 out of 13 samples) |
|---|---|---|---|
| | 0–6/8 h | 6/8–24 h | 0–24 h | Total (unique) |
| | Activated | Inhibited | Activated | Inhibited | Activated | Inhibited | Activated | Inhibited |  |
| Re-analysis M-values | 58 (44) | 66 (52) | 63 (24) | 61 (49) | 212 (105) | 258 (193) | 718 (348) |
| Analysis with R 3.6.2 affy::rma | 59 (46) | 78 (62) | 61 (22) | 66 (52) | 223 (115) | 268 (202) | 755 (378) |
| Analysis with R 3.6.2 affy::germa | 117 (93) | 153 (110) | 156 (59) | 121 (97) | 427 (220) | 446 (346) | 1420 (665) |
3.6.2 were different from the re-analysis M-value. The M-value was provided by [2], who also applied RMA (supplied by an old version of R in 2006) to compute the M-values. Thus, the results from GCRMA contained more differentially expressed genes than the other two methods. Figure 17 visualises the common and different figures using a Venn diagram.

We were interested in the genes set in the intersections of the results with the three methods. Hence, we constructed the FBM cube and extracted the fundamental Boolean networks based on the most common genes, i.e., 285 genes. Table 2 lists the common genes.

We applied the DAVID bioinformatics database (https://david.ncifcrf.gov) to cluster the common genes in a group of functional clusters to better interpret the genes’ functionality, as shown in Table 3.

Figure 17. Venn diagram of the unique glucocorticoid-induced and differentially-expressed genes across all time points.

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Common genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Common results</td>
<td>Genes</td>
</tr>
<tr>
<td>Common genes (285)</td>
<td>MS4A4A, CLN8, SNX10, RAB31, SERPINA1, LILRA1, LILRB2, PFKFB2, SOCS1, LGALS3, SIK1, SLA, FCGR3B, FGIL2, BTN19, RBMS3, DPEP1, MNDA, FKB5P, DDT4, WFS1, S100A1I, ZBTB16, P2RY14, GSN, EPKK1, ZFPE6L2, FCER1G, FGR, IRAK3, PBFB, MYRIP, PIK3IP1, KIF26A, NUFB2, ARPP21, ABHD17B, ASPM, PKB, KIF20A, CEP55, PDE4B, HMMR, CDKN3, FUS, AURKA, CENPE, POU4F1, PRR11, KIF14, CENPC, CENPA, NEK2, TTK, KIF11, GIMAP4, DLGAP5, PTG1, UBE2C, CDC20, MKI67, CCNB2, BIRC5, FAM72C, GBP4, KIF23, NPCDR1, CCNL1, LOC285097, CLEC28B, CBL10, TOP2A, TENM4, CCNB1, SNORA21, MDM2, DEPDC1, DEPDC1B, GIMAP7, TBX2A3R, STAB1, MIR4683, SMIM3, SLC22A3, TMEM100, LOC100130872, SESN1, TMEM2, IFNNG1, BIRC3, P2RX5, CDC42E3P, METTL7A, IL18RAP, WFS1, S100A11, ZBTB16, P2RY14, GSN, EPPK1, ZFP36L2, FCER1G, FGR, IRAK3, PPBP, MYRIP, PIK3IP1, KIF26A, NUFB2, ARPP21, ABHD17B, ASPM, PKB, KIF20A, CEP55, PDE4B, HMMR, CDKN3, FUS, AURKA, CENPE, POU4F1, PRR11, KIF14, CENPC, CENPA, NEK2, TTK, KIF11, GIMAP4, DLGAP5, PTG1, UBE2C, CDC20, MKI67, CCNB2, BIRC5, FAM72C, GBP4, KIF23, NPCDR1, CCNL1, LOC285097, ITGAM, DEFA1, ISG20, DFN5, MST5, IL6ST, MR84584, NEAT1, KLF9, TXNIP, IL18R1, HBB, CELFL2, HBGI, S100A8, MPV17L, NEDD9, FG2D, BTG1, TUBA4A, RPS6KA2, DENND3, IL12RA, MS4A4T, SMAP2, RASSF4, BMF, ELL2, TARSL2, SEMA4D, ITGB2-AS1, PP1R16B, CPM, ITPKB, IL1B, MCM10, CDC4A, UBE2T, TCF19, BRIP1, KIF18A, KCNK12, HELL5, E2F8, NUSAP1, ECT2, CENPN, SHCBP1, POLQ, KIF15, B3GNT2, TIPIN, DSCC1, CENPU, TNFRSF21, ATAD2, HJUPR, NCPP, ATG1, KIF4A, CKBAP, MSH6, MCM7, LEFI, AKAP12, SKP2, TPX2, CDT1, RRM2, BUB1, KIF2C, MDK, ANP32E, CDK1, F13A1, BYSL, HIV4H4, CCNA2, MND1, EGR1, PRPS2, SNORD3B-1, HRK, CHER1, WDR76, IGL1L, RAG1, RCC1, POL2, MCM4, BCA1, STIL, PSHP, PCNA, RAD51, PLK4, TRIP13, MELK, FEN1, WDD1H, BRCAL, SM2, WASF1, CDC45, RFC3, WINT, CDC6, BUB1B, MAD2L1, CDC28, TIMELESS, OIIP5, TUMS, DHRX, FOXM1, KIAA0101, TK1, FABP5, TRIB1, KAPA, CXS1B, CENPH, MTHFD2, NME1, PP1F, GGHP, PAICS, GINS1, AURKB, ASFB1, GINS2, CENPK, IQGAP3, ZNF367, PTP4A1, AMLN, C5orf24, SUV39H2, SELENOM1, RBM14, HSP90AB1, FH, RAD51A, DHX9, TMEM97, NCAHR, FANCI, APTID1-CORT, RAG2, MCM5, KIF18A, CDC4A, ARRD3C, SQLE, LOC100996643, CDC34, RIM2, CENPC, CDC2A, CENPW, PHF19, E2F7, KNL1, C4orf46, CDK6, CRNDE, SERPINB9, PRDM1, ID2, GVINP1, MIR8071-1, CCNE2, ID3, IGLC1, LYZ, IGH</td>
</tr>
<tr>
<td>Common genes from B-ALL only for 10/10 samples</td>
<td>CDC42EP3</td>
</tr>
<tr>
<td>Common genes from T-ALL (23/285), for 3/3 samples</td>
<td>FKB5P, PFKFB2, MPV17L, SCML4, BTG1, ZFPE6L2, DDT4, RHOBTB3, SLA, ISG20, TNFSF8, SNX29P2, PIK3IP1, SMAP2, SNTB2, SEMA4D, CDKN3, BCL10, PTG1, FEN1, MIR8071-1, MCM4, IGLC1</td>
</tr>
<tr>
<td>Common genes from B-ALL (6/10) and T-ALL (3/3)</td>
<td>SCML4, DDT4, SLA, PFKFB2, CDKN3, ZFPE6L2, FKB5P, SNX29P2, PIK3IP1, TNFSF8, PTG1, MCM4, MIR8071-1</td>
</tr>
<tr>
<td>Key GC regulate/response genes reported in Schmidt’s research</td>
<td>PFKFB2, BTNL9, SNF1LK, FKB5P, ZBTB16, KIF26A, SLA, SOCS1, DDT4, GNB4, MGC17330, ZFPE6L2, EPKK1, P2RY14, FGR, WFS1, ARPP21, SERPINA1, GIMAP7, MYCPBP, LGALS3</td>
</tr>
<tr>
<td>Common genes for all-time span induced in the majority of 6/13</td>
<td>PFKFB2, BTNL9, FKB5P, P2RY14</td>
</tr>
<tr>
<td>Common genes for all-time span repressed in the majority of 6/13</td>
<td>CEP55, MKI67, CCNB2, BIRC5, TOP2A</td>
</tr>
</tbody>
</table>
### Table 3  Common genes in functional clusters

<table>
<thead>
<tr>
<th>Functional cluster</th>
<th>Related common genes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cell cycle</strong></td>
<td>CDK6, ASPM, AURKA, MAD2L1, APITD1-CORT, BUB1, FOXM1, TIPIN, SMC2, ZWINT, CDCA3, NUF2, ECT2, KNL1, CDK1, CDCA2, CD6, CKS1B, NEDD9, CENPV, MCM7, TPX2, CCNB1, PTP4A1, CDC20, TXNIP, FANC1, BRCA1, CDT1, E2F7, NUSAP1, CENPA, NCAF, DLAG5, CKAP2, KIF2C, HELLs, BIRC5, NCAF, CENPE, KIF11, CDCA5, CHEK1, MCM4, CENPF, CCNA2, CEPP5, SIK1, DSCC1, CCNB2, MKI67, CCNE2, TIMELESS, UBE2C, NEK2, OIP5, RCC1, SUV39H2, MLK, CDKN3, RASSF4, HJURP, KIF18B, BUB1B, AURKB, CENPE, MCM5, PTG1, KIF23, ANLN, CDC45, E2F8</td>
</tr>
<tr>
<td><strong>Cell division</strong></td>
<td>CDK6, ASPM, AURKA, MAD2L1, APITD1-CORT, BUB1, TIPIN, SMC2, ZWINT, CDCA3, NUF2, ECT2, KNL1, CDK1, CDCA2, CD6, CKS1B, NEDD9, CENPV, TPX2, CCNB1, CDC20, NUSAP1, CENPA, NCAF, KIF2C, HELLs, BIRC5, NCAF, CENPE, KIF11, CDCA5, CHEK1, MCM4, CCNA2, CEPP5, CCNB2, CCNE2, TIMELESS, UBE2C, NEK2, OIP5, RCC1, KIF18B, BUB1B, AURKB, CENPE, MCM5, PTG1, KIF23, ANLN</td>
</tr>
<tr>
<td><strong>p53 signaling pathway</strong></td>
<td>CDK6, MD2M, CDK1, SESN1, CCNB1, RRM2, CHEK1, CCNB2, NEK2</td>
</tr>
<tr>
<td><strong>Membrane</strong></td>
<td>CPM, TNFSF8, FGD2, RAB31, STAB1, IL6ST, BCL2L11, SLC22A23, MS4A1, TNFRSF21, GBP4, CLN8, HBB, MS4A4A, FCGR3B, HRK, EMP1, ADRD3C, CLEC2B, TXB2A2R, PTP4A1, SNTB2, BRCA1, LILRA1, SQLE, B3GNT2, CD3, FCDR, CLECR1G, BCL10, P2RX5, SELEN01, BTN79, MS4A7, RAG2, LILR2, SMIM3, SNX10, METTL7A, PIK3P1, CCR1, SEAMA4D, IL18R1, IL27RA, IFNGR1, MLK, DFNA5, TMEM97, WFS1, DTLY, TYNPS, PPI116B, MPV17L, DPEP1, TPME, ITGAM, PTG14, TENM4, IL18RAP, KCNK12, S100A8, FGR, TMEM100</td>
</tr>
<tr>
<td><strong>Transmembrane</strong></td>
<td>CPM, TNFSF8, STAB1, IL6ST, SLC22A23, MS4A1, TNFRSF21, CLN8, HBB, MS4A4A, FCGR3B, HRK, EMP1, CLEC2B, TXB2A2R, BRCA1, LILRA1, SQLE, B3GNT2, CD3, FCDR, CLECR1G, BCL10, P2RX5, SELEN01, BTN79, MS4A7, RAG2, LILR2, SMIM3, METTL7A, PIK3P1, CCR1, SEAMA4D, IL18R1, IL27RA, IFNGR1, MLK, DFNA5, TMEM97, WFS1, MPV17L, TLM2, ITGAM, PTG14, TMEM4, IL18RAP, KCNK12, S100A8, FGR, TMEM100</td>
</tr>
<tr>
<td><strong>Cell membrane</strong></td>
<td>CPM, FGD2, IL6ST, MS4A1, TNFRSF21, FCGR3B, ADRD3C, TXB2A2R, PTP4A1, CD3, FCDR, PIK3P1, CCR1, SEAMA4D, IFNGR1, MLK, DFNA5, TMEM97, PPI116B, DPEP1, PTG14, TENM4, S100A8, FGR, TMEM100</td>
</tr>
<tr>
<td><strong>Glycoprotein</strong></td>
<td>CPM, TNFSF8, HSP90AB1, STAB1, IL6ST, SLC22A23, FGL2, TNFRSF21, HBB, MS4A4A, FCGR3B, HRK, EMP1, CLEC2B, TXB2A2R, BRCA1, LILRA1, SQLE, B3GNT2, CD3, FCDR, CLECR1G, BCL10, P2RX5, SELEN01, BTN79, GGG, LILR2, F13A1, PIK3P1, CCR1, SEAMA4D, IL18R1, IL27RA, IFNGR1, MLK, DFNA5, TMEM97, WFS1, MPV17L, TLM2, ITGAM, PTG14, TMEM4, IL18RAP, KCNK12, S100A8, FGR, TMEM100</td>
</tr>
<tr>
<td><strong>Receptor</strong></td>
<td>STAB1, IL6ST, TNFRSF21, MS4A4A, FCGR3B, TXB2A2R, LILRA1, FCR1G, P2RX5, HMMR, MS4A7, LILR2, CCR1, SEAMA4D, IL18R1, IL27RA, IFNGR1, MLK, DFNA5, TMEM97, PPI116B, DPEP1, PTG14, TENM4, S100A8, FGR, TMEM100</td>
</tr>
<tr>
<td><strong>ATP-binding</strong></td>
<td>CDK6, NME1, KIF14, PRPS2, AURKA, HSP90AB1, KIF20A, BUB1A, RPS6K2A, PAI5CS, SMC2, RHOBTB3, PFKFB2, BRIP1, CDK1, CDC6, MCM7, RAD51, MS6H, KIF2C, PBK, KIF26A, TRIP3, HELLS, CENPE, KIF11, TTK, KIF4A, KIF15, ITPKB, TOP2A, CHEK1, MCM4, SIK1, PLK4, MKI67, UBE2C, TK1, PO1Q, NEK2, MLK, ATAD2, KIF18B, BUB1B, TARS2L, AURKB, IRAK3, KIF18A, UBE2T, MCM5, KIF23, DIXH, FGR</td>
</tr>
<tr>
<td><strong>Transerase</strong></td>
<td>CDK6, BIRC3, NME1, PRPS2, AURKA, SKP2, BUB1, RPS6K2A, BCA1, PFKFB2, CD1, KCS18, POLE2, BRCA1, B3GNT2, PBK, PRDM1, SELEN01, TTK, ITPKB, CHEK1, F13A1, SIK1, PLK4, METTL7A, UBE2C, TK1, PO1Q, NEK2, SUV39H2, MLK, CDKN3, TYS, BUB1B, AURKB, IRAK3, UBE2T, FGR</td>
</tr>
<tr>
<td><strong>Transcription regulation</strong></td>
<td>CCNL1, CENPU, RBM14, KLF9, FOXM1, SCML4, LEF1, PHF19, ZBTB16, EGR1, ID2, POU4F1, MNDA, DEPC1, TXNIP, BRCA1, E2F7, TCF19, ID3, HELLs, PRDM1, BIRC5, ELL2, TIMELESS, ZNF367, SUV39H2, ATAD2, ASF1B, E2F8</td>
</tr>
<tr>
<td><strong>Ub conjugation pathway</strong></td>
<td>BIRC3, SKP2, MD2M, CDCA3, ZBTB16, CDC20, BRCA1, RAG1, UBE2C, SOCS1, DTLY, UBE2T</td>
</tr>
<tr>
<td><strong>PI3K-Akt signaling pathway</strong></td>
<td>CDK6, HSP90AB1, MD2M, BCL2L11, SOS1, BRCA1, DDIT4, CCNE2</td>
</tr>
<tr>
<td><strong>SH3 domain</strong></td>
<td>NEDD9, SLA, FGR</td>
</tr>
<tr>
<td><strong>Metal-binding</strong></td>
<td>CPM, FUS, FGD2, BIRC3, NME1, KLF9, PRPS2, MCM10, MD2M, RPS6K2A, HBB, GSN, ISG20, BRIIP1, PSHI, PHF19, ZBTB16, EGR1, BRCA1, TCF19, PRDM1, BIRC5, RAG1, SELEN01, FZP362, SMAP2, RRM2, TOP2A, RAG2, FENI, F13A1, SIK1, TK1, PDE4B, ZNF367, NEK2, OIP5, SUV39H2, HBGI, DPEP1, AURKB, ITGAM, S100A8, MYRIP, S100A8</td>
</tr>
<tr>
<td><strong>Immunity</strong></td>
<td>TNFRSF21, ISG20, LILR1A, BCL10, LGALS3, PRDM1, LILR2B, LY96, S100A8, FGR</td>
</tr>
<tr>
<td><strong>Signal</strong></td>
<td>CPM, STAB1, IL6ST, FGL2, TNFRSF21, GSN, FCGR3B, SERPIN1, RNASET2, LILRA1, FCR1G, PRPS2, RRM2, TOP2A, RAG2, FENI, F13A1, SIK1, TK1, PDE4B, ZNF367, NEK2, OIP5, SUV39H2, HBGI, DPEP1, AURKB, ITGAM, S100A8, MYRIP, S100A8</td>
</tr>
<tr>
<td><strong>Transit peptide</strong></td>
<td>FH, MTHFD2, PP1F</td>
</tr>
<tr>
<td><strong>Immunoregulation</strong></td>
<td>FKB5</td>
</tr>
<tr>
<td><strong>Pseudogene</strong></td>
<td>SNX29P2, LOC100996643</td>
</tr>
</tbody>
</table>
Model construction

After the differentially expressed genes have been extracted, we constructed the TFBM Orchard cube based on the genes using the package FBNNet. Constructing an Orchard cube based on the training data was similar to the way discussed in [1] for the cell cycle data. Only difference from the experiments conducted on cell cycle data was maxK value, which was 4 for cell cycle data, whereas we choose maxK = 3 for this study, where maxK defined the maximum level of the combinations of genes that could affect the target gene. The main reason is that the actual number of combinations of genes is very small, and in fact, only 0.396% of rules have input genes of 3, identified in this experiment result. Hence, to reduce the complexity of computation, we adopted maxK = 3 in this study. The maximum temporal value (m), which is a value that indicates how many times the steps are required to complete the regulation process of a fundamental Boolean function, is 2 in this study because the data we used is short time series and only contains three time points.

APPENDIX

(A) General terminology abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Term</th>
</tr>
</thead>
<tbody>
<tr>
<td>APC</td>
<td>Anaphase-promoting complex</td>
</tr>
<tr>
<td>BUC</td>
<td>Bottom-up computation</td>
</tr>
<tr>
<td>Cdk</td>
<td>Cyclin-dependent kinases (such as Cdk1, Cdk2)</td>
</tr>
<tr>
<td>CKI</td>
<td>Cyclin-dependent kinase inhibitor</td>
</tr>
<tr>
<td>CycD</td>
<td>Cdk4/6-Cyclin D complex</td>
</tr>
<tr>
<td>CycE</td>
<td>Cdk2/Cyclin E complex</td>
</tr>
<tr>
<td>CycA</td>
<td>Cdk2/Cyclin A complex</td>
</tr>
<tr>
<td>Cdc20</td>
<td>Cell-division cycle protein 20</td>
</tr>
<tr>
<td>Cdh1</td>
<td>Epithelial cadherin (E-cadherin), a classical member of the cadherin superfamily</td>
</tr>
<tr>
<td>CycB</td>
<td>Cdk1/Cycline B complex</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>E2F</td>
<td>A family of transcription factors (TF) that act as transcriptional regulators of G1–S transition</td>
</tr>
<tr>
<td>GRNs</td>
<td>Genetic regulatory networks</td>
</tr>
<tr>
<td>GF</td>
<td>Growth factor</td>
</tr>
<tr>
<td>NP</td>
<td>Nondeterministic polynomial, a computational complexity class</td>
</tr>
<tr>
<td>NP-hard</td>
<td>A class of problems in computational complexity (nondeterministic polynomial acceptable problems)</td>
</tr>
<tr>
<td>p27</td>
<td>A member of Kip/Cip family, a group of CKIs</td>
</tr>
<tr>
<td>R</td>
<td>A statistical script programming language</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>Rb</td>
<td>Retinoblastoma protein</td>
</tr>
<tr>
<td>SI</td>
<td>Supplementary information</td>
</tr>
<tr>
<td>UbcH10</td>
<td>Cancer-related E2 ubiquitin-conjugating enzyme1</td>
</tr>
<tr>
<td>GCs</td>
<td>Glucocorticoids</td>
</tr>
<tr>
<td>GC</td>
<td>Glucocorticoid</td>
</tr>
<tr>
<td>GRs</td>
<td>Glucocorticoid receptors</td>
</tr>
<tr>
<td>GR</td>
<td>Glucocorticoid receptor</td>
</tr>
</tbody>
</table>

(B) Model terminology abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Term</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR</td>
<td>Accurate rate</td>
</tr>
<tr>
<td>ER</td>
<td>Error rate</td>
</tr>
<tr>
<td>FBM</td>
<td>Fundamental Boolean model</td>
</tr>
<tr>
<td>FBN</td>
<td>Fundamental Boolean network</td>
</tr>
<tr>
<td>TFBN</td>
<td>Temporal Fundamental Boolean network</td>
</tr>
<tr>
<td>MMR</td>
<td>Mismatched rate</td>
</tr>
<tr>
<td>PMR</td>
<td>Perfect matched rate</td>
</tr>
</tbody>
</table>

(C) Boolean function notations

<table>
<thead>
<tr>
<th>Notation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>¬</td>
<td>A negation operator that turns a Boolean function from TRUE to FALSE or vice versa</td>
</tr>
<tr>
<td>⊗</td>
<td>Logical And operator</td>
</tr>
<tr>
<td>+</td>
<td>Logical Or operator</td>
</tr>
<tr>
<td>τ</td>
<td>An incremental variable presenting the number of time steps processed</td>
</tr>
<tr>
<td>θ</td>
<td>A decay period reflects the fact that the attenuation or enhancement of the expression of mRNA requires time</td>
</tr>
<tr>
<td>σᵣ</td>
<td>Boolean state of gene σᵣ at time t</td>
</tr>
<tr>
<td>σᵣ₊₁</td>
<td>Boolean state of gene σᵣ at time t+1</td>
</tr>
<tr>
<td>Lᵣ(ᵣ)</td>
<td>Total number of fundamental Boolean functions</td>
</tr>
<tr>
<td>fᵣ⁺</td>
<td>A fundamental Boolean function of the activation</td>
</tr>
<tr>
<td>fᵣ⁻</td>
<td>A fundamental Boolean function of inhibition</td>
</tr>
<tr>
<td>P[ x ]</td>
<td>A Boolean function that takes a uniform distributed random number, μ, and an output of 1 if μ &lt; x and 0 otherwise.</td>
</tr>
<tr>
<td>V[ x ]</td>
<td>Logical connective function of Or</td>
</tr>
<tr>
<td>Tᵣ⁺</td>
<td>The best previous time step value for the activation function fᵣ⁺</td>
</tr>
<tr>
<td>Tᵣ⁻</td>
<td>The best previous time step value for the inhibition function fᵣ⁻</td>
</tr>
<tr>
<td>Aᵣ⁺</td>
<td>The set of required inputs for the gene state functions, fᵣ⁺</td>
</tr>
<tr>
<td>Dᵣ⁻</td>
<td>The set of required inputs for the gene state functions, fᵣ⁻</td>
</tr>
</tbody>
</table>

(E) Fundamental Boolean network graphic types

FBNNet_ALL A general FBN type that contains up-regulatory and down-regulatory pathways. It also refers to FBNNet type 0.
FBNNet_FAA A forward FBN type that shows the up-regulatory pathways of giving activated genes. It also refers to FBNNet type 1.

FBNNet_FAi A forward FBN type that shows the down-regulatory pathways of giving activated genes. It also refers to FBNNet type 2.

FBNNet_FIA A forward FBN type that shows the up-regulatory pathways of giving inhibited genes. It also refers to FBNNet type 3.

FBNNet_FII A forward FBN type that shows the down-regulatory pathways of giving inhibited genes. It also refers to FBNNet type 4.

FBNNet_BA A backward FBN type shows the regulatory pathways that drive a target gene to be activated. It also refers to FBNNet type 5.

FBNNet_BI A backward FBN type shows the regulatory pathways that drive a target gene to be inhibited. It also refers to FBNNet type 6.

SUPPLEMENTARY MATERIALS

The supplementary materials can be found online with this article at https://doi.org/10.15302/J-QB-021-0280.

ACKNOWLEDGEMENTS

We acknowledge the Centre for Advanced Computational Solutions (C-fACS) at Lincoln University, New Zealand for facilitating this research.

COMPLIANCE WITH ETHICS GUIDELINES

The authors Leshi Chen, Don Kulasiri and Sandhya Samarasinghe declare that they have no conflict of interests or competing financial interests. All procedures performed in studies were in accordance with the ethical standards of the institution or practice at which the studies were conducted, and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

OPEN ACCESS

This article is licensed by the CC By under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

REFERENCES


41. Chaiboonchoe, A. (2010) Identification of glucocorticoid-regulated genes and inferring their network focused on the glucocorticoid receptor in childhood leukaemia, based on microarray data and pathway databases, pp. 170. Lincoln University, New Zealand


909–917


cancer cells. Technol. Cancer Res. Treat., 18, 153303381985-8668