Quantitative Biology 2022, 10(1): 94–121
https://doi.org/10.15302/J-QB-021-0280

RESEARCH ARTICLE

Fundamental Boolean network modelling for childhood acute lymphoblastic leukaemia pathways

Leshi Chen[1], Don Kulasiri[1,*], Sandhya Samarasinghe[2]

[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

Received January 20, 2021; Revised May 21, 2021; Accepted June 13, 2021

***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

Author summary: Boolean modelling has been applied in numerous areas. However, minimal effort has been put into creating activation, inhibition and protein decay networks. In the previous study, we proposed a novel concept, namely, fundamental Boolean model, to address these issues. However, it does not reflect that genetic regulation might require more time steps to complete. Hence, in this paper, we proposed extending the original model to temporal fundamental Boolean model and demonstrated the new model with the leukaemia datasets downloaded from NCBI. Our result shows that the proposed model's capacity could identify crucial signalling pathways for glucocorticoid treated childhood leukaemia patients.

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

© The Author(s) 2021. Published by Higher Education Press

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.

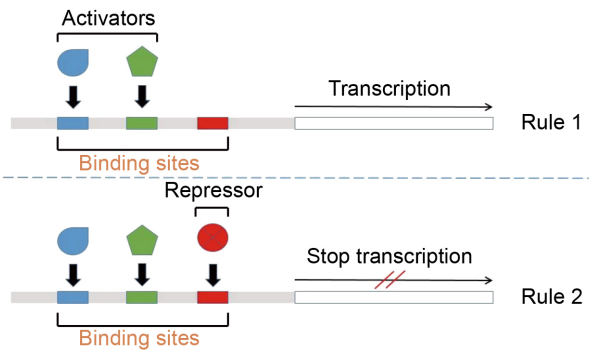

Figure 1. Concept of gene activation and repression.

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.

© The Author(s) 2021. Published by Higher Education Press

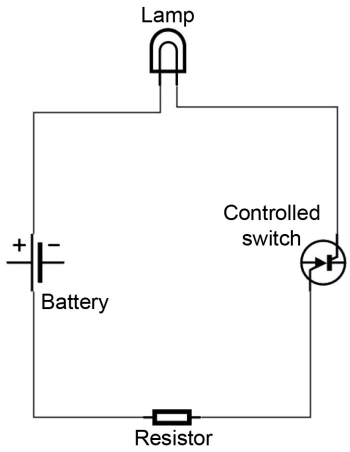

Figure 2. A simple series circuit representing a conventional Boolean model.

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 ($\check{T}$) 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, $\check{S}$. Hence, the entry, e^{s}_{ij}, in row i and column j of the table $\check{T}_s$ 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, $\check{T}_s$ (matrix) such as using cell cycle analyses. However, data analysis on a three-dimensional sample data space ($\check{S}_s$) 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

© The Author(s) 2021. Published by Higher Education Press

provides 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 controversially 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 Bcl2 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

© The Author(s) 2021. Published by Higher Education Press

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 Author(s) 2021. Published by Higher Education Press

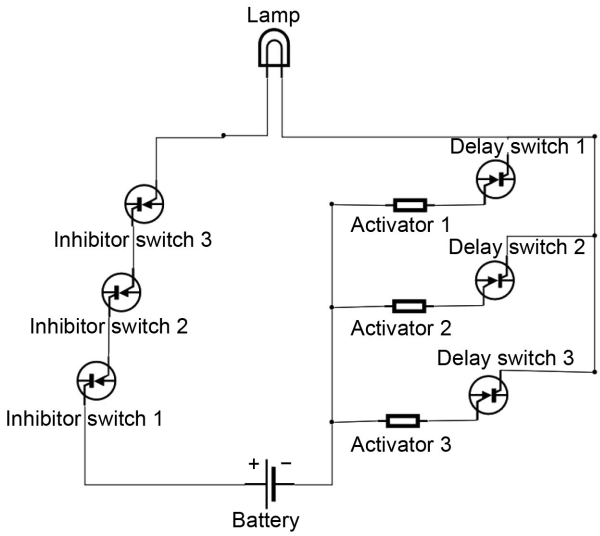

Figure 3. Series and parallel circuits representing a fundamental Boolean model.

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 (X, E_a, E_d), where the node collection, V= $\{v_1, v_2, ..., v_n\}$, corresponds to a group of states, X= $\{x_i | i = 1, ..., n\}$ of size n. Each node is a variable that is only in one of two states: *On*(1) or *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

$$F_a^i = \left\{ f_{a_j}^i | j = 1, ..., l_a(i) \right\}, f_{a_j}^i : \{0,\ 1\} \rightarrow \{-,\ 1\}; \qquad (1.a)$$

Fundamental Boolean functions of inhibition

$$F_d^i = \left\{ f_{d_k}^i | k = 1, ..., l_d(i) \right\}, f_{d_k}^i : \{0,\ 1\} \rightarrow \{-,\ 0\}, \qquad (1.b)$$

where F_a^i and F_d^i 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 *On*, the target gene i is activated, and *Off* means that the activation function does not influence the target gene i. Similarly, when the Boolean inhibition function's output is *On*, the target gene is repressed and *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 stereochemical 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:

$$C_{a_j}^i \left\lfloor f_{a_j}^i \left(A_i^j(t) \right) \right\rfloor = p\left(\sigma_i^{t+1} = 1 | A_i^j(t) = 1 \right) = \frac{p\left(A_i^j(t) = 1 \cap \sigma_i^{t+1} = 1 \right)}{p\left(A_i^j(t) = 1 \right)}; \qquad (2.a)$$

Confidence measure of inhibition:

© The Author(s) 2021. Published by Higher Education Press

$$C^i_{d_k}\left\lfloor f^i_{d_k}\left(D^k_i(t)\right)\right\rfloor = p\left(\sigma^{t+1}_i = 0 | D^k_i(t) = 1\right) = \frac{p\left(D^k_i(t) = 1 \cap \sigma^{t+1}_i = 0\right)}{p\left(D^k_i(t) = 1\right)}, \quad (2.b)$$

where σ^t_i denotes the Boolean state of gene i at time t, and σ^{t+1}_i denotes the Boolean state of gene i at time $t + 1$. $\cap$ refers to a logical *And* connector. $C^i_{a_j}$ and $C^i_{d_k}$ delegate the confidence function with the input of the fundamental Boolean functions $f^i_{a_j}$ and $f^i_{d_k}$, respectively. A^j_i and D^k_i denote the set of inputs required or the state of the gene functions, $f^i_{a_j}$ and $f^i_{d_k}$, respectively. $A^j_i(t) = 1$ or $D^k_i(t) = 1$ mean the required gene input of $f^i_{a_j}$ or $f^i_{d_k}$ 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_{decay} to fulfil the requirements of protein degradation with input from the target gene i at time t [1]:

$$f_{\text{decay}}(\sigma^t_i, \vartheta) = \neg(\tau \leqslant \vartheta) \times \sigma^t_i, \quad (3)$$

where τ represents an incremental variable presenting the number of time steps processed. τ will be reset to 0 when there is any fundamental Boolean function affecting the target gene (i) at time $t + 1$. ϑ 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_{decay} is a Boolean state of *On* at time $t + 1$ if the gene state of σ_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 σ_i of time t [1]. In this study, the tolerated period ϑ is set to one time step, *i.e.*, $\vartheta = 1$ due to the expemeriment 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 σ_i at time $t + 1$ based on the immediately previous time t as

$$\sigma^{t+1}_i = \left(f_{\text{decay}}(\sigma^t_i, \vartheta) + \bigvee_{j=1}^{l_a(i)} \left\{P\left[\!\left[C^i_{a_j}\left\lfloor f^i_{a_j}(A^j_i(t))\right\rfloor\right]\!\right]\right\}\right) \times \neg \bigvee_{k=1}^{l_d(i)} \left\{P\left[\!\left[C^i_{d_k}\left\lfloor f^i_{d_k}(D^k_i(t))\right\rfloor\right]\!\right]\right\}, \quad (4)$$

where $+$ is a logical *Or* operator and $\times$ is a logical *And* operator. The decay function $f_{\text{decay}}(\sigma^t_i, \vartheta)$ in Eq. (3) is to ensure the gene state σ_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 ϑ, a decay period. $P[\![x]\!]$ is a Boolean function that takes a uniform distributed random number, μ, and outputs a value of *On* if μ < x and *Off* otherwise. $\mathrm{V}\{x\}$ denotes the logical connective function of *Or*, *i.e.*, $\mathrm{V}_{j=1}^{l_a(i)}\{F^i_a\} = P\left[\!\left[C^i_{a_1}\left(f^i_{a_1}\right)\right]\!\right] + P\left[\!\left[C^i_{a_2}\left(f^i_{a_2}\right)\right]\!\right] + \cdots + P\left[\!\left[C^i_{a_{l_a(i)}}\left(f^i_{a_{l_a(i)}}\right)\right]\!\right]$. 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. FBN 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:

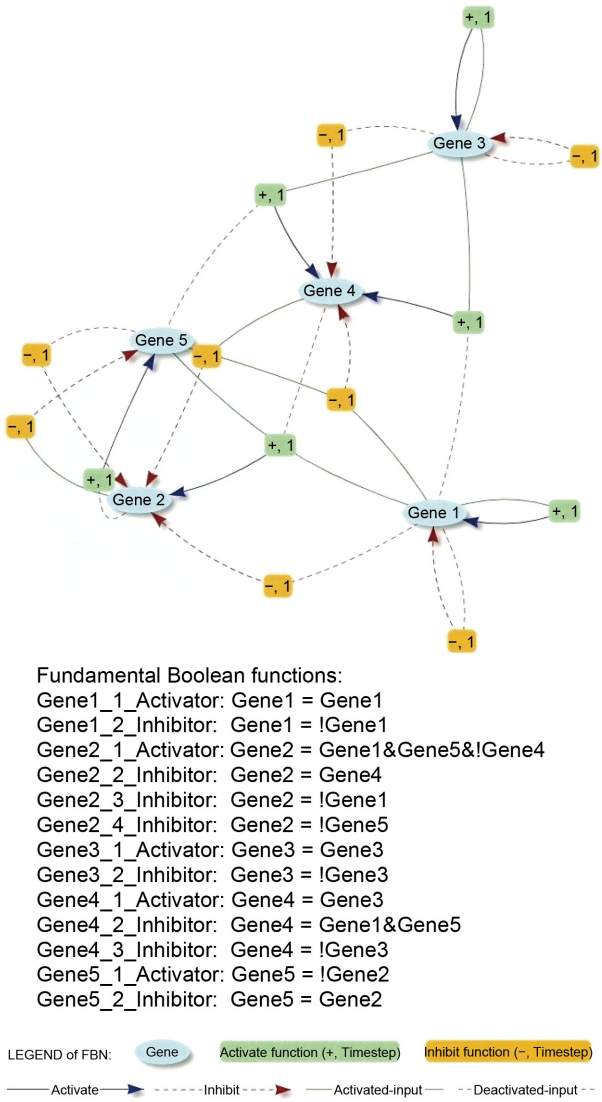

Figure 4. A general type of fundamental Boolean networks of an example FBN.

© The Author(s) 2021. Published by Higher Education Press

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.

```
Discovered attractors via fundamental Boolean model
Genes are encoded in the following order:
CycD Rb E2F CycE CycA p27 Cdc20 Cdh1 UbcH10 CycB

Attractor 1 is a simple attractor consisting of 1 state(s)

| --< - - - - - - - - - - |
v                         ^
0 1 0 0 0 1 0 1 0 0        |
|                         |
v                         ^
| - - - - - - - - - - >-- |


Attractor 2 is a complex attractor consisting of 7 state(s):

| --< - - - - - - - - - - |
v                         ^
1 0 0 0 0 0 1 1 1 0        |
|                         |
1 0 1 0 0 0 0 1 1 0        |
|                         |
1 0 1 1 0 0 0 1 0 0        |
|                         |
1 0 1 1 1 0 0 1 0 0        |
|                         |
1 0 0 1 1 0 0 0 0 0        |
|                         |
1 0 0 0 1 0 0 0 1 1        |
|                         |
1 0 0 0 1 0 1 0 1 1        |
|                         |
v                         ^
| - - - - - - - - - - >-- |
```

Figure 5. Synchronous attractors of the cell cycle fundamental Boolean model, published in [1]. Reproduced with permission.

Temporal fundamental Boolean modelling

The original FBM we defined in Eq. (4) provides a mechanism to calculate gene state σ_i 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 $t - 1, \ldots, t - m$ ($1 \leqslant m < t$). 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 $G(X, E_a, E_d, T)$, where T is the best time step in the time window during which a gene can be regulated by corresponding E_a, E_d. 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

$$\sigma_i^{t+1} = (f_{\text{decay}}(\sigma_i^t, \vartheta) + \bigvee_{j=1}^{l_a(i)} \left\{ P\left[\!\left[C_{a_j}^i \left\lfloor f_{a_j}^i (A_i^j(\mathrm{T}_i^j)) \right\rfloor \right]\!\right] \right\}) \times \neg \bigvee_{k=1}^{l_d(i)} \left\{ P\left[\!\left[C_{d_k}^i \left\lfloor f_{d_k}^i (D_i^k(\mathrm{T}_i^k)) \right\rfloor \right]\!\right] \right\}, \tag{5}$$

where T_i^j and T_i^k are the best previous time step for the activation function $f_{a_j}^i$ and inhibition function $f_{d_k}^i$ of gene i, respectively. A_i^j and D_i^k represent the set of required inputs for the gene state functions, $f_{a_j}^i$ and $f_{d_k}^i$ at the best previous time step T_i^j and T_i^k, respectively.

To calculate the best previous time step T_i^j or T_i^k 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 $\tilde{\mathrm{A}}_i$ for the activation function $f_{a_j}^i$, and a perfect vector $\ddot{\mathrm{A}}_i$, which

© The Author(s) 2021. Published by Higher Education Press

contains the best value for all measurements:

$$\tilde{A}_i = \begin{matrix} \text{measurement}_{i_1}^{t} & \text{measurement}_{i_1}^{t-1} & \dots & \text{measurement}_{i_1}^{t-m} \\ \text{measurement}_{i_2}^{t} & \text{measurement}_{i_2}^{t-1} & \dots & \text{measurement}_{i_2}^{t-m} \\ \dots & \dots & \dots & \dots \\ \text{measurement}_{i_e}^{t} & \text{measurement}_{i_e}^{t-1} & \dots & \text{measurement}_{i_e}^{t-m} \end{matrix}$$

where e denoted as a measurement of the matrix $\tilde{A}_i$. The perfect vector $\ddot{A}_i$ 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 $f_{a_j}^{i}$ can be calculated based on the shortest distance between measurements on time step t to $t-m$ and the perfect vector $\ddot{A}$:

$$T_i^j = \min\left(\text{dist}_t^{t-m}\left(\tilde{A}_i,\ \ddot{A}_i\right)\right).$$

T_i^j 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].

If we define the measurement matrix as:

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

© The Author(s) 2021. Published by Higher Education Press

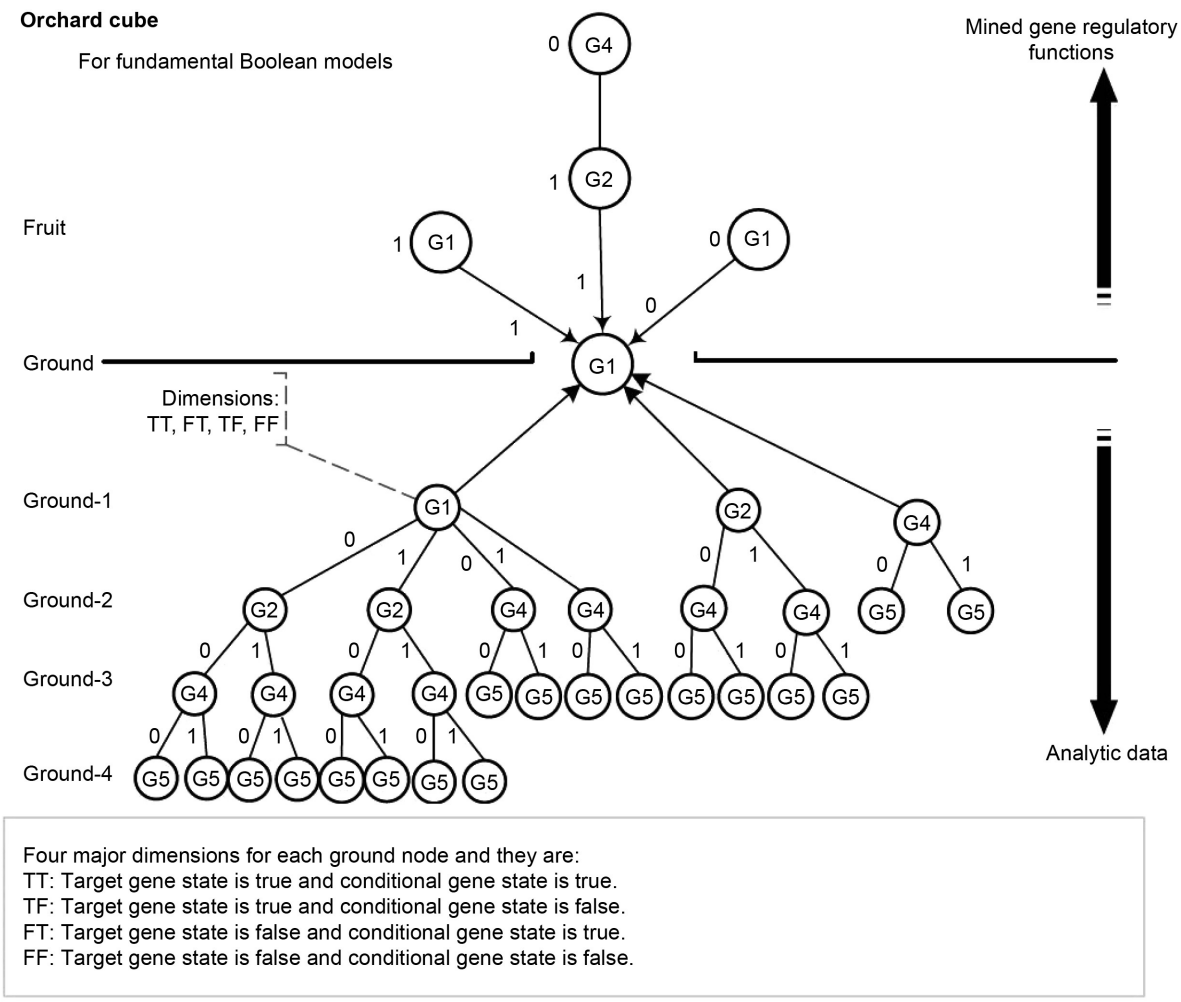

Figure 6. Illustration of an orchard cube, published in [1]. Reproduced with permission.

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.

• FBNNet_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.

• FBNNet_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.

• FBNNet_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.

• FBNNet_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

© The Author(s) 2021. Published by Higher Education Press

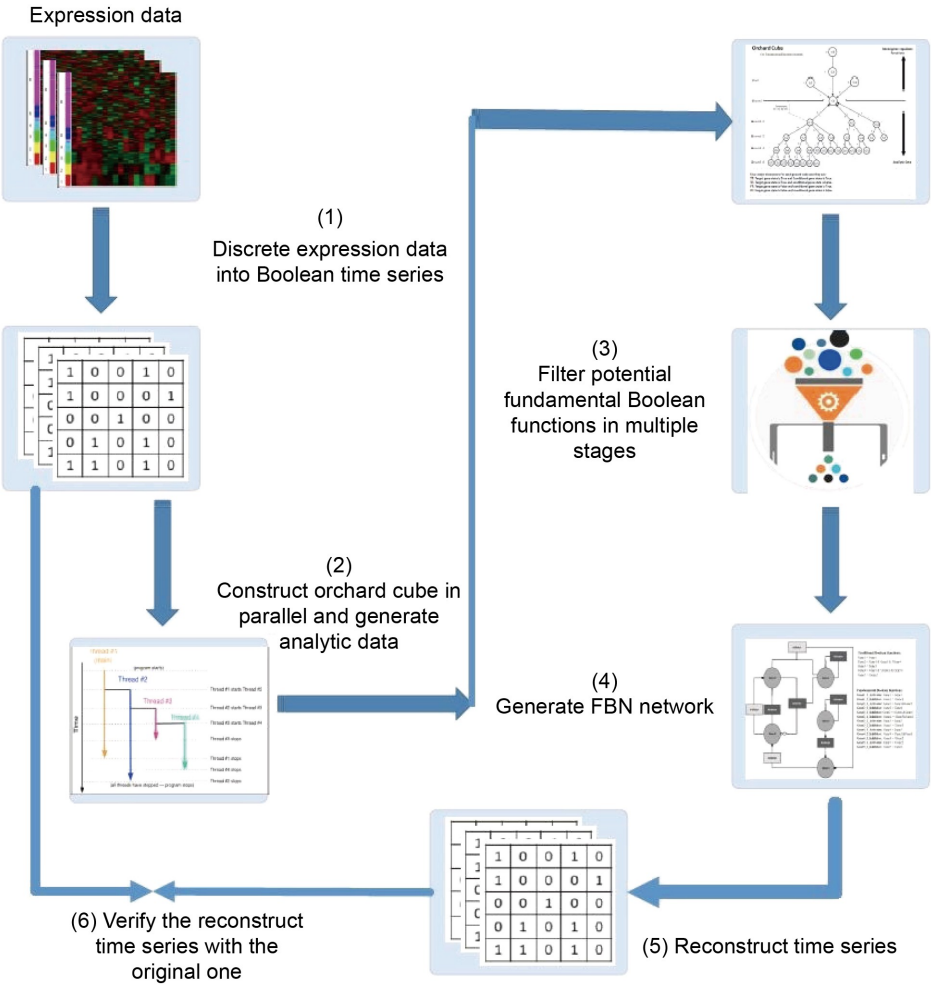

Figure 7. Schematic diagram of FBN modelling and network inferences [1]. Reproduced with permission.

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,

© The Author(s) 2021. Published by Higher Education Press

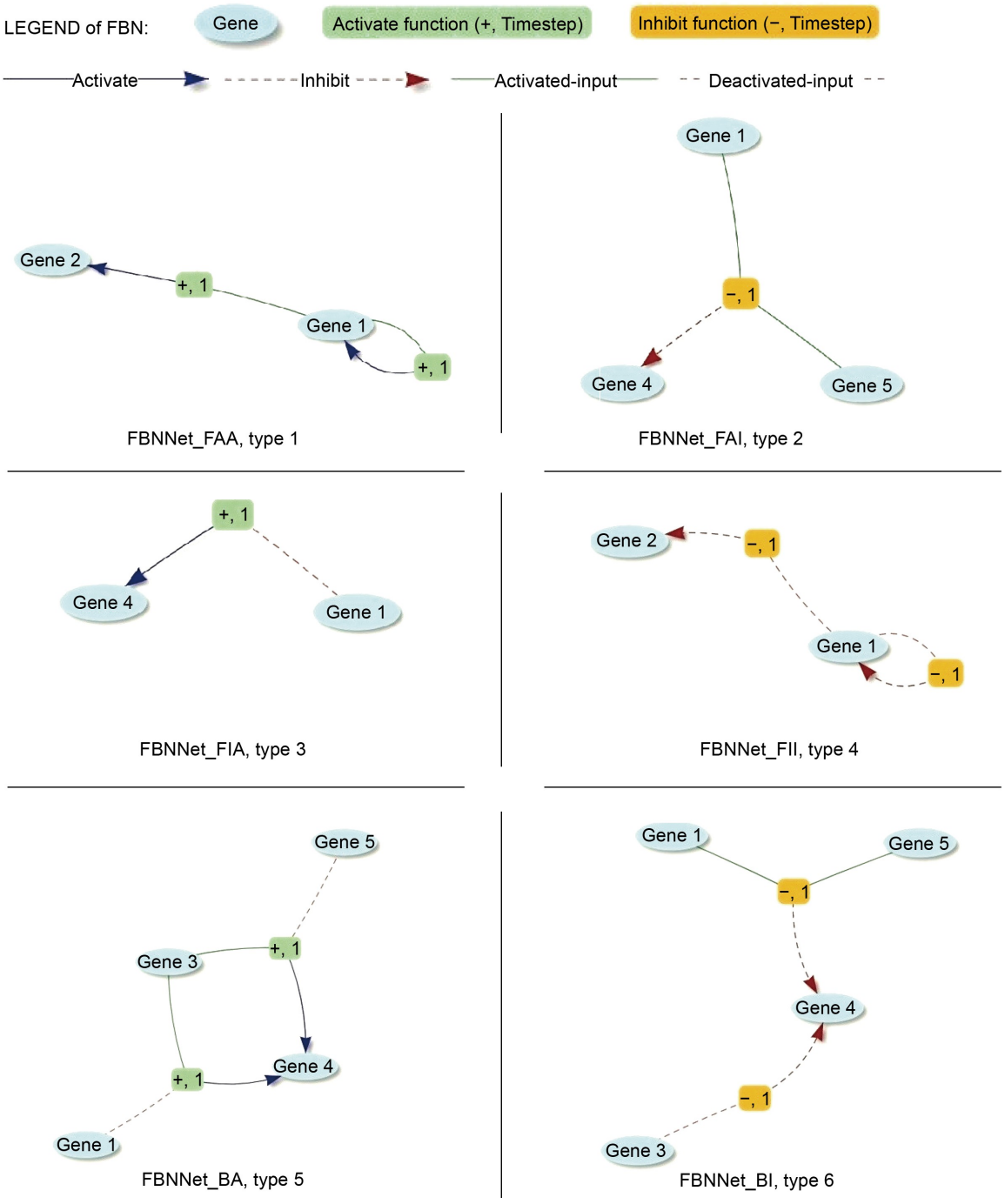

Figure 8. The subnetwork types of fundamental Boolean networks.

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 LGAL-S3 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

© The Author(s) 2021. Published by Higher Education Press

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 either in 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, BTNL9, 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. BTNL9 (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. BTNL9 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 BTNL9 were found to inhibit their target genes when activated. The target inhibited genes LOC100505650 (uncharacterized [*Homo sapiens*]), CENPU (transcription regulation), SQLE (membrane and transmembrane), BCAT1 (transferase), and PRPS2 (ATP-binding, transferase and metal-binding) may be the target responsible genes to be suppressed, of the GC treatment.

© The Author(s) 2021. Published by Higher Education Press

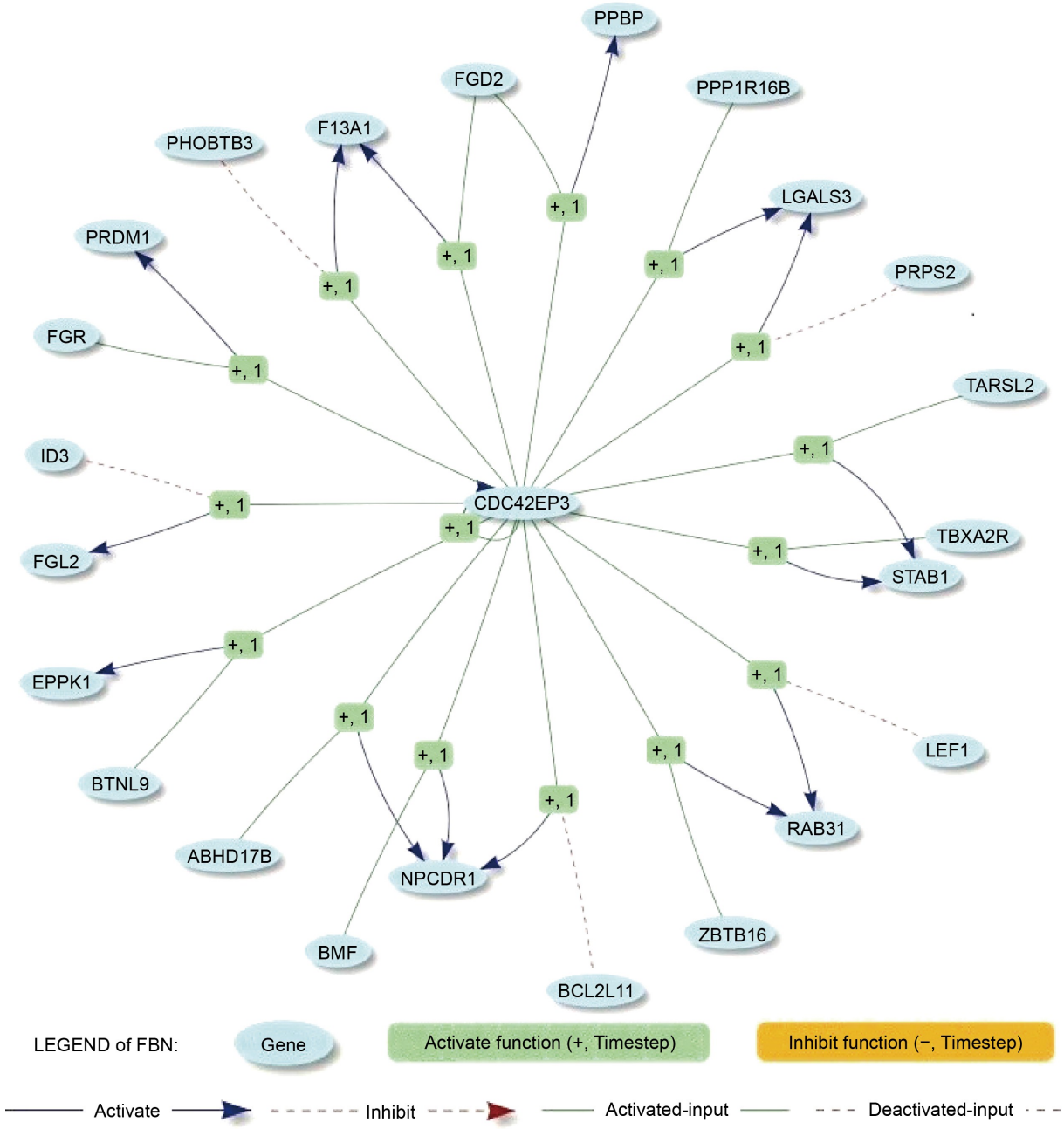

Figure 9. Type 1 networks of CDC42EP3.

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.

© The Author(s) 2021. Published by Higher Education Press

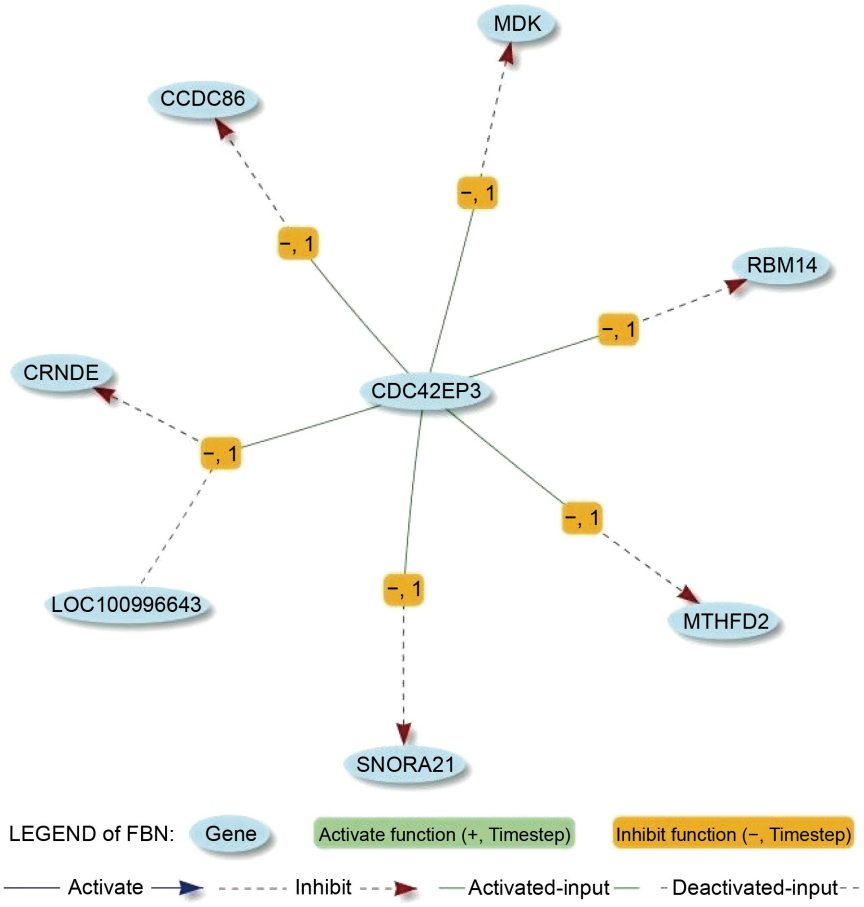

Figure 10. Type 2 networks of CDC42EP3.

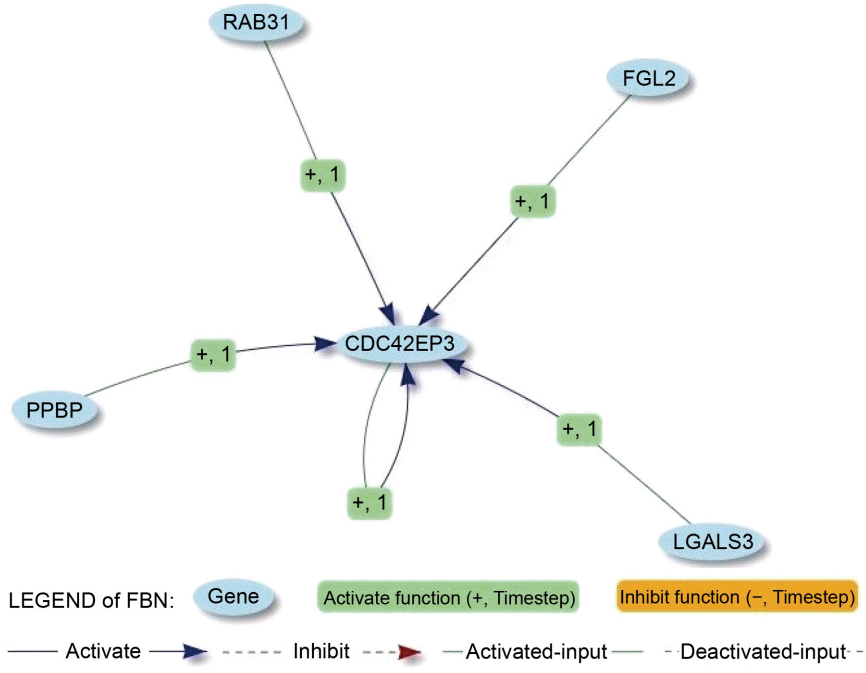

Figure 11. Type 5 networks of CDC42EP3.

© The Author(s) 2021. Published by Higher Education Press

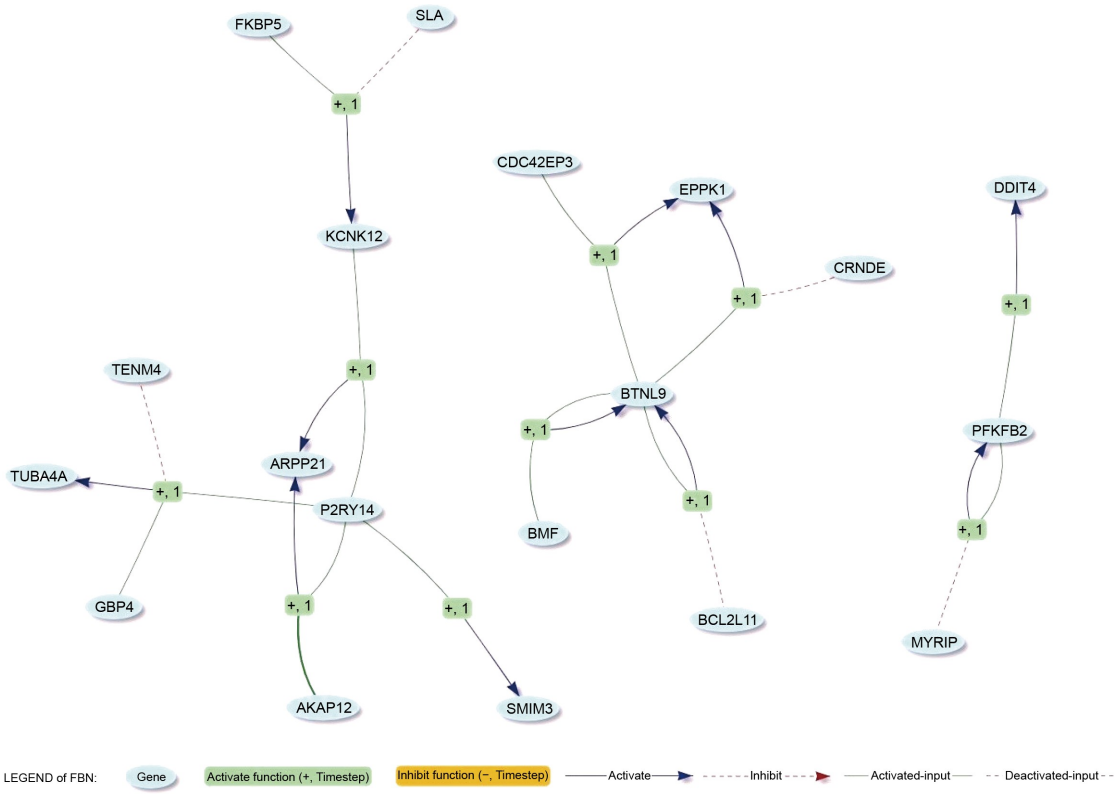

Figure 12. Type 1 subnetworks of genes PFKFB2, BTNL9, FKBP5 and P2RY14.

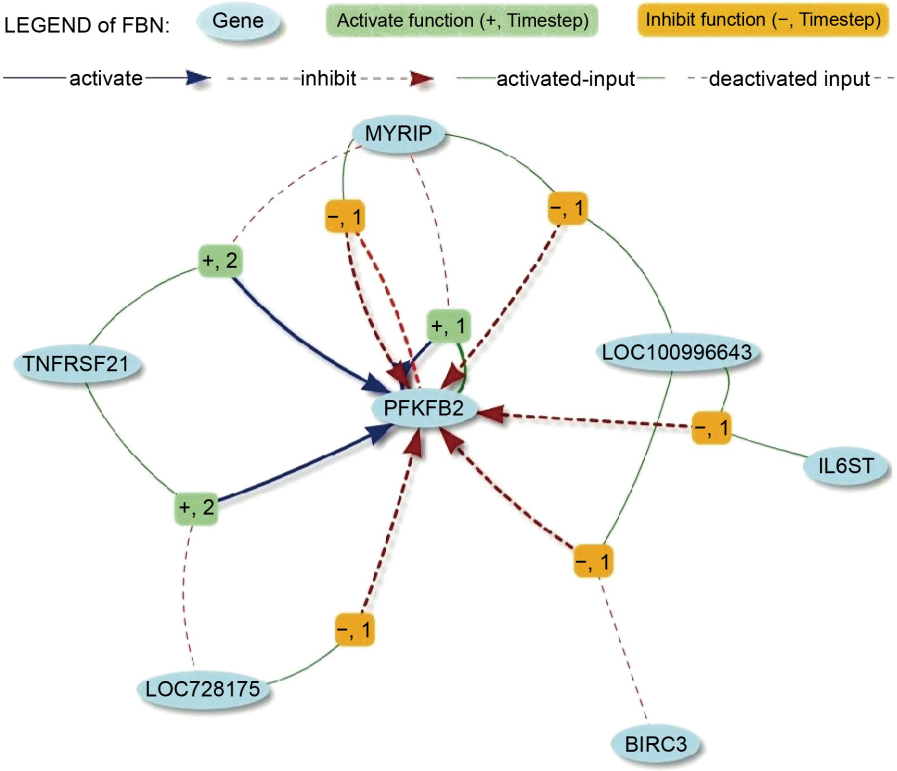

Figure 13. PFKFB2 backward regulatory networks.

© The Author(s) 2021. Published by Higher Education Press

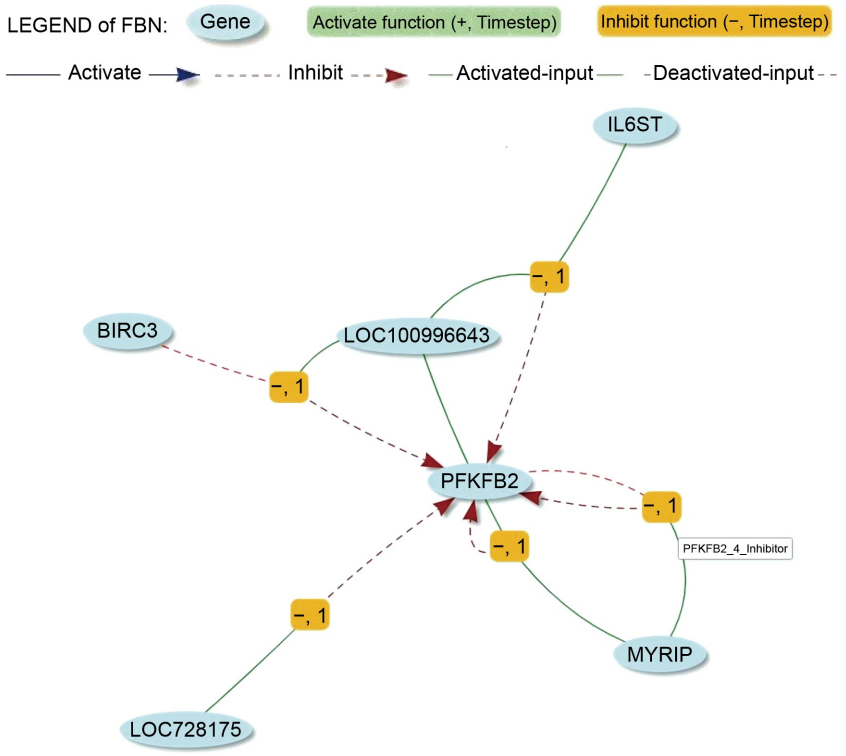

Figure 14. PFKFB2 backward regulatory networks under the original FBM modelling.

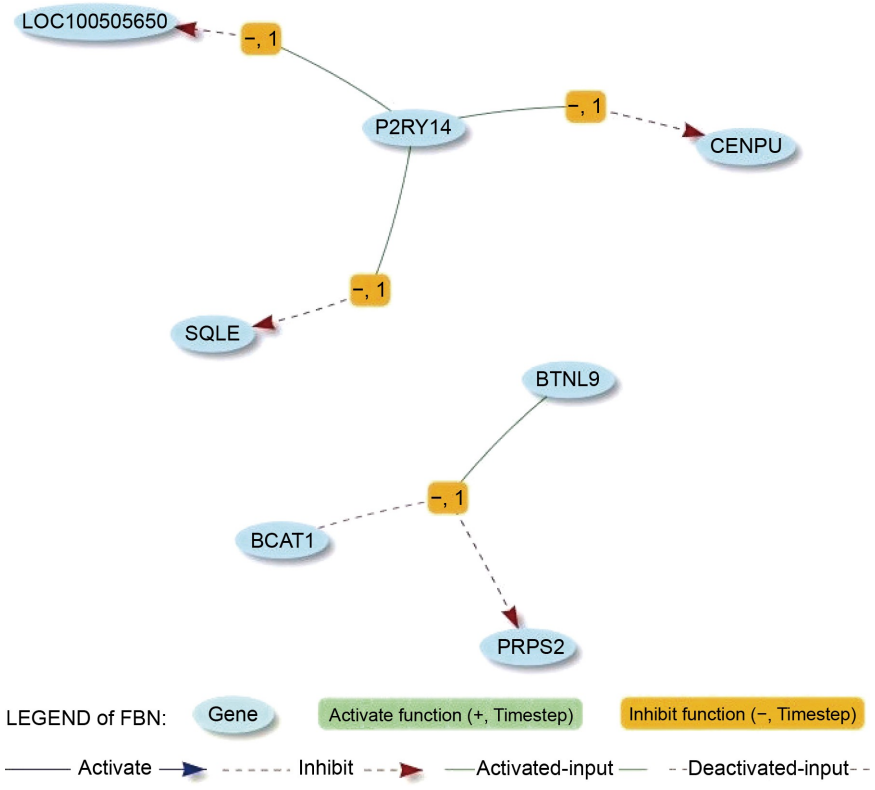

Figure 15. Type 2 subnetworks of P2PY14 and BTNL9.

© The Author(s) 2021. Published by Higher Education Press

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 22 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 BCl-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 TFBN 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

© The Author(s) 2021. Published by Higher Education Press

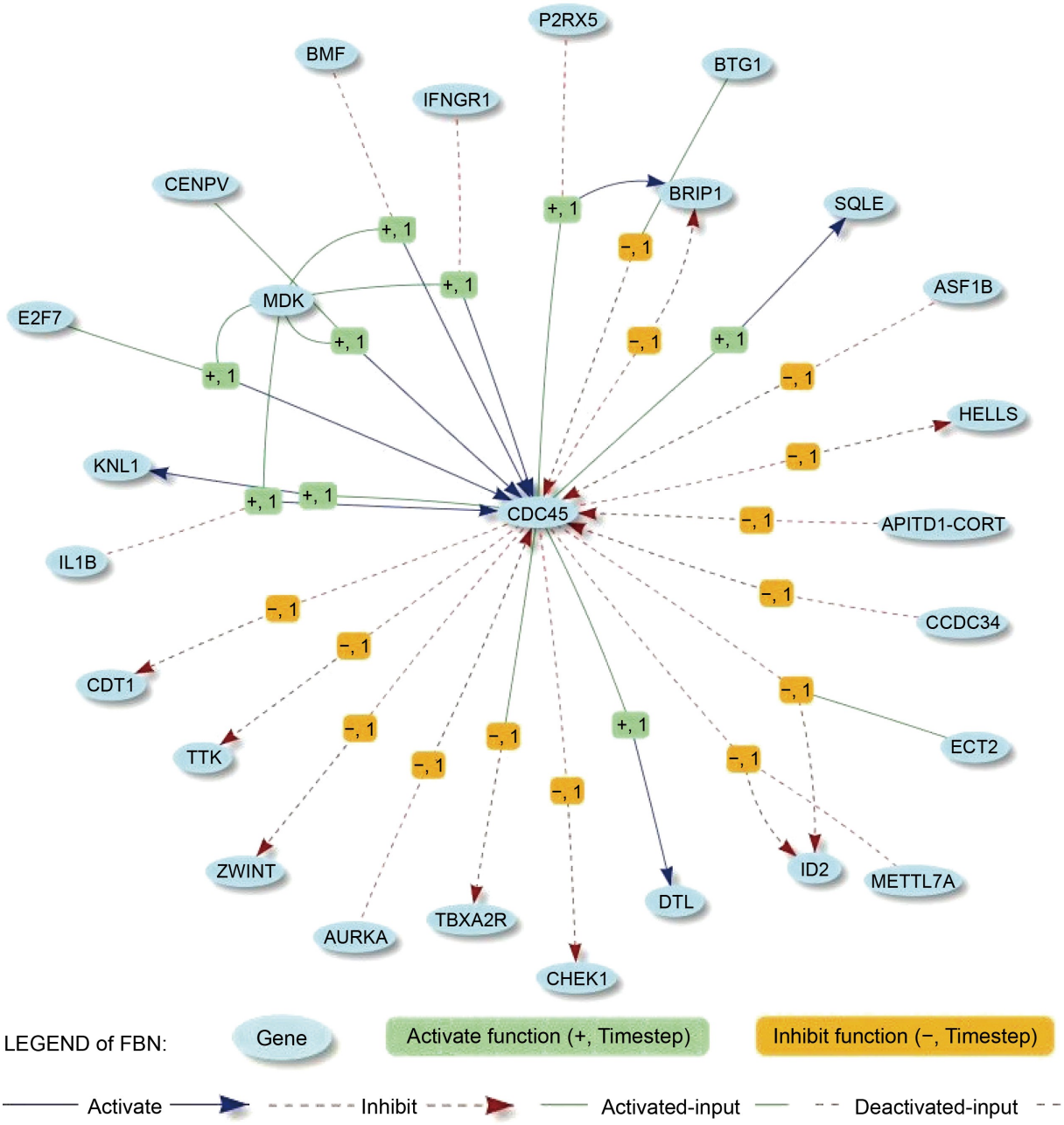

Figure 16. Networks of CDC45.

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

© The Author(s) 2021. Published by Higher Education Press

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 FBNNet 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 modelling (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 FBNNet (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 assession 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

© The Author(s) 2021. Published by Higher Education Press

extracted differentially expressed time series data.

This study selected R software as the leading platform to analyse genetic networks. R (Available from the website of r-project) 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 $\geqslant 1$, 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 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

Table 1 GC-regulated differentially expressed genes, M>=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	
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::gcrma	117 (93)	153 (110)	156 (59)	121 (97)	427 (220)	446 (346)	1420 (665)

© The Author(s) 2021. Published by Higher Education Press

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 to cluster the common genes in a group of functional clusters to better interpret the genes' functionality, as shown in Table 3.

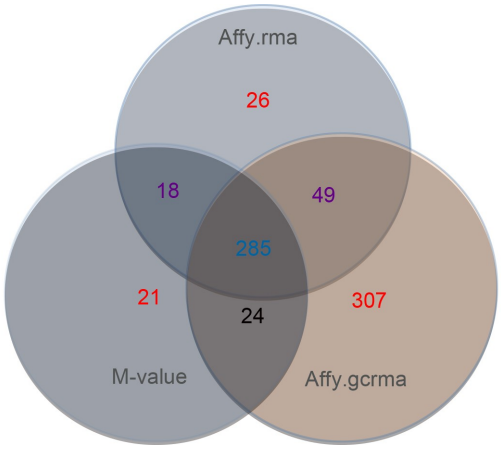

Figure 17. Venn diagram of the unique glucocorticoid-induced and differentially-expressed genes across all time points.

Table 2 Common genes

Common results	Genes
Common genes (285)	MS4A4A, CLN8, SNX10, RAB31, SERPINA1, LILRA1, LILRB2, PFKFB2, SOCS1, LGALS3, SIK1, SLA, FCGR3B, FGL2, BTNL9, RBMS3, DPEP1, MNDA, FKBP5, DDIT4, WFS1, S100A11, ZBTB16, P2RY14, GSN, EPPK1, ZFP36L2, FCER1G, FGR, IRAK3, PPBP, MYRIP, PIK3IP1, KIF26A, NUF2, ARPP21, ABHD17B, ASPM, PBK, KIF20A, CEP55, PDE4B, HMMR, CDKN3, FUS, AURKA, CENPF, POU4F1, PRR11, KIF14, CENPE, CENPA, NEK2, TTK, KIF11, GIMAP4, DLGAP5, PTTG1, UBE2C, CDC20, MKI67, CCNB2, BIRC5, FAM72C, GBP4, KIF23, NPCDR1, CCNL1, LOC728175, CLEC2B, BCL10, TOP2A, TENM4, CCNB1, SNORA21, MDM2, DEPDC1, DEPDC1B, GIMAP7, TBXA2R, STAB1, MIR4683, SMIM3, SLC22A23, TMEM100, LOC100130872, SESN1, TMEM2, IFNGR1, BIRC3, P2RX5, CDC42EP3, METTL7A, IL18RAP, CCR1, SNX29P2, LOC100505650, RHOBTB3, BCL2L11, SNTB2, MS4A1, SOS1, TNFSF8, RNASET2, CD53, LY96, SCML4, LOC285097, ITGAM, DEFA1, ISG20, DFNA5, MTSS1, IL6ST, MIR6845, NEAT1, KLF9, TXNIP, IL18R1, HBB, CELF2, HBG1, S100A8, MPV17L, NEDD9, FGD2, BTG1, TUBA4A, RPS6KA2, DENND3, IL27RA, MS4A7, SMAP2, RASSF4, BMF, ELL2, TARSL2, SEMA4D, ITGB2-AS1, PPP1R16B, CPM, ITPKB, IL1B, MCM10, CDCA3, UBE2T, TCF19, BRIP1, KIF18A, KCNK12, HELLS, E2F8, NUSAP1, ECT2, CENPN, SHCBP1, POLQ, KIF15, B3GNT2, TIPIN, DSCC1, CENPU, TNFRSF21, ATAD2, HJURP, NCAPG, DTL, KIF4A, CKAP2, MSH6, MCM7, LEF1, AKAP12, SKP2, TPX2, CDT1, RRM2, BUB1, KIF2C, MDK, ANP32E, CDK1, F13A1, BYSL, HIST4H4, CCNA2, MND1, EGR1, PRPS2, SNORD3B-1, HRK, CHEK1, WDR76, IGLL1, RAG1, RCC1, POLE2, MCM4, BCAT1, STIL, PSPH, PCNA, RAD51, PLK4, TRIP13, MELK, FEN1, WDHD1, BRCA1, SMC2, WASF1, CDC45, RFC3, ZWINT, CDC6, BUB1B, MAD2L1, CCDC86, TIMELESS, OIP5, TYMS, DHFR, FOXM1, KIAA0101, TK1, FABP5, TRIB1, CKAP2L, CKS1B, CENPH, MTHFD2, NME1, PPIF, EMP1, GGH, PAICS, GINS1, AURKB, ASF1B, GINS2, CENPK, IQGAP3, ZNF367, PTP4A1, ANLN, C5orf24, SUV39H2, SELENOI, RBM14, HSP90AB1, FH, RAD51AP1, DHX9, TMEM97, NCAPH, FANCI, APITD1-CORT, RAG2, MCM5, KIF18B, CDCA5, ARRDC3, SQLE, LOC100996643, CCDC34, RMI2, CENPV, CDCA2, CENPW, PHF19, E2F7, KNL1, C4orf46, CDK6, CRNDE, SERPINB9, PRDM1, ID2, GVINP1, MIR8071-1, CCNE2, ID3, IGLC1, LYZ, IGH
Common genes from B-ALL only for 10/10 samples	CDC42EP3
Common genes from T-ALL (23/285), for 3/3 samples	FKBP5, PFKFB2, MPV17L, SCML4, BTG1, ZFP36L2, DDIT4, RHOBTB3, SLA, ISG20, TNFSF8, SNX29P2, PIK3IP1, SMAP2, SNTB2, SEMA4D, CDKN3, BCL10, PTTG1, FEN1, MIR8071-1, MCM4, IGLC1
Common genes fromB-ALL (6/10) and T-ALL (3/3)	SCML4, DDIT4, SLA, PFKFB2, CDKN3, ZFP36L2, FKBP5, SNX29P2, PIK3IP1, TNFSF8, PTTG1, MCM4, MIR8071-1
Key GC regulate/response genes reported in Schmidt's research	PFKFB2, BTNL9, SNF1LK, FKBP5, ZBTB16, KIF26A, SLA, SOCS1, DDIT4, GBP4, MGC17330, ZFP36L2, EPPK1, P2RY14, FGR, WFS1, ARPP-21, SERPINA1, GIMAP7, MYCPBP, LGALS3
Common genes for all-time span induced in the majority of 6/13	PFKFB2, BTNL9, FKBP5, P2RY14
Common genes for all-time span repressed in the majority of 6/13	CEP55, MKI67, CCNB2, BIRC5, TOP2A

© The Author(s) 2021. Published by Higher Education Press

Table 3 Common genes in functional clusters

Functional cluster	Related common genes
Cell cycle	CDK6, ASPM, AURKA, MAD2L1, APITD1-CORT, BUB1, FOXM1, TIPIN, SMC2, ZWINT, CDCA3, NUF2, ECT2, KNL1, CDK1, CDCA2, CDC6, CKS1B, NEDD9, CENPV, MCM7, TPX2, CCNB1, PTP4A1, CDC20, TXNIP, FANCI, BRCA1, CDT1, E2F7, NUSAP1, CENPA, NCAPG, DLGAP5, CKAP2, KIF2C, HELLS, BIRC5, NCAPH, CENPE, KIF11, CDCA5, CHEK1, MCM4, CENPF, CCNA2, CEP55, SIK1, DSCC1, CCNB2, MKI67, CCNE2, TIMELESS, UBE2C, NEK2, OIP5, RCC1, SUV39H2, MELK, CDKN3, RASSF4, HJURP, KIF18B, BUB1B, AURKB, CENPW, MCM5, PTTG1, KIF23, ANLN, CDC45, E2F8
Cell division	CDK6, ASPM, AURKA, MAD2L1, APITD1-CORT, BUB1, TIPIN, SMC2, ZWINT, CDCA3, NUF2, ECT2, KNL1, CDK1, CDCA2, CDC6, CKS1B, NEDD9, CENPV, TPX2, CCNB1, CDC20, NUSAP1, CENPA, NCAPG, KIF2C, HELLS, BIRC5, NCAPH, CENPE, KIF11, CDCA5, CENPF, CCNA2, CEP55, CCNB2, CCNE2, TIMELESS, UBE2C, NEK2, OIP5, RCC1, KIF18B, BUB1B, AURKB, CENPW, MCM5, PTTG1, KIF23, ANLN
p53 signaling pathway	CDK6, MDM2, CDK1, SESN1, CCNB1, RRM2, CHEK1, CCNB2, CCNE2
Membrane	CPM, TNFSF8, FGD2, RAB31, STAB1, IL6ST, BCL2L11, SLC22A23, MS4A1, TNFRSF21, GBP4, CLN8, HBB, MS4A4A, FCGR3B, HRK, EMP1, ARRDC3, CLEC2B, TBXA2R, PTP4A1, SNTB2, BRCA1, LILRA1, SQLE, B3GNT2, CD53, FCER1G, BCL10, P2RX5, SELENOI, BTNL9, AKAP12, MS4A7, RAG2, CDC42EP3, LILRB2, SMIM3, SNX10, METTL7A, PIK3IP1, CCR1, SEMA4D, IL18R1, IL27RA, IFNGR1, MELK, DFNA5, TMEM97, WFS1, DTL, TYMS, PPP1R16B, MPV17L, DPEP1, TMEM2, ITGAM, P2RY14, TENM4, IL18RAP, KCNK12, S100A8, FGR, TMEM100
Transmembrane	CPM, TNFSF8, STAB1, IL6ST, SLC22A23, MS4A1, TNFRSF21, CLN8, HBB, MS4A4A, FCGR3B, HRK, EMP1, CLEC2B, TBXA2R, BRCA1, LILRA1, SQLE, B3GNT2, CD53, FCER1G, P2RX5, SELENOI, BTNL9, MS4A7, RAG2, LILRB2, SMIM3, METTL7A, PIK3IP1, CCR1, SEMA4D, IL18R1, IL27RA, IFNGR1, TMEM97, WFS1, MPV17L, TMEM2, ITGAM, P2RY14, TENM4, IL18RAP, KCNK12, TMEM100
Cell membrane	CPM, FGD2, IL6ST, MS4A1, TNFRSF21, FCGR3B, ARRDC3, TBXA2R, PTP4A1, CD53, FCER1G, PIK3IP1, CCR1, SEMA4D, IFNGR1, MELK, DFNA5, TMEM97, PPP1R16B, DPEP1, P2RY14, TENM4, S100A8, FGR, TMEM100
Glycoprotein	CPM, TNFSF8, HSP90AB1, STAB1, IL6ST, SLC22A23, FGL2, TNFRSF21, HBB, FCGR3B, EMP1, MCM7, CLEC2B, SERPINA1, TBXA2R, RNASET2, LILRA1, B3GNT2, CD53, P2RX5, HMMR, BTNL9, GGH, LILRB2, F13A1, PIK3IP1, CCR1, SEMA4D, IL18R1, IL27RA, IFNGR1, WFS1, DPEP1, TMEM2, ITGAM, KIF18A, P2RY14, LY96, TENM4, IL18RAP, KCNK12
Receptor	STAB1, IL6ST, TNFRSF21, MS4A4A, FCGR3B, TBXA2R, LILRA1, FCER1G, P2RX5, HMMR, MS4A7, LILRB2, CCR1, SEMA4D, IL18R1, IL27RA, IFNGR1, ITGAM, P2RY14, IL18RAP
ATP-binding	CDK6, NME1, KIF14, PRPS2, AURKA, HSP90AB1, KIF20A, BUB1, RPS6KA2, PAICS, SMC2, RHOBTB3, PFKFB2, BRIP1, CDK1, CDC6, MCM7, RAD51, MSH6, KIF2C, PBK, KIF26A, TRIP13, HELLS, CENPE, KIF11, TTK, KIF4A, KIF15, ITPKB, TOP2A, CHEK1, MCM4, SIK1, PLK4, MKI67, UBE2C, TK1, POLQ, NEK2, MELK, ATAD2, KIF18B, BUB1B, TARSL2, AURKB, IRAK3, KIF18A, UBE2T, MCM5, KIF23, DHX9, FGR
Transferase	CDK6, BIRC3, NME1, PRPS2, AURKA, SKP2, BUB1, RPS6KA2, BCAT1, PFKFB2, CDK1, CKS1B, POLE2, BRCA1, B3GNT2, PBK, PRDM1, SELENOI, TTK, ITPKB, CHEK1, F13A1, SIK1, PLK4, METTL7A, UBE2C, TK1, POLQ, NEK2, SUV39H2, MELK, CDKN3, TYMS, BUB1B, AURKB, IRAK3, UBE2T, FGR
Transcription regulation	CCNL1, CENPU, RBM14, KLF9, FOXM1, SCML4, LEF1, PHF19, ZBTB16, EGR1, ID2, POU4F1, MNDA, DEPDC1, TXNIP, BRCA1, E2F7, TCF19, ID3, HELLS, PRDM1, BIRC5, ELL2, TIMELESS, ZNF367, SUV39H2, ATAD2, ASF1B, E2F8
Ubl conjugation pathway	BIRC3, SKP2, MDM2, CDCA3, ZBTB16, CDC20, BRCA1, RAG1, UBE2C, SOCS1, DTL, UBE2T
PI3K-Akt signaling pathway	CDK6, HSP90AB1, MDM2, BCL2L11, SOS1, BRCA1, DDIT4, CCNE2
SH3 domain	NEDD9, SLA, FGR
Metal-binding	CPM, FUS, FGD2, BIRC3, NME1, KLF9, PRPS2, MCM10, MDM2, RPS6KA2, HBB, GSN, ISG20, BRIP1, PSPH, PHF19, ZBTB16, EGR1, BRCA1, TCF19, PRDM1, BIRC5, RAG1, SELENOI, ZFP36L2, SMAP2, RRM2, TOP2A, RAG2, FEN1, F13A1, SIK1, TK1, PDE4B, ZNF367, NEK2, OIP5, SUV39H2, HBG1, DPEP1, AURKB, ITGAM, S100A11, MYRIP, S100A8
Immunity	TNFRSF21, ISG20, LILRA1, BCL10, LGALS3, PRDM1, LILRB2, LY96, S100A8, FGR
Signal	CPM, STAB1, IL6ST, FGL2, TNFRSF21, GSN, FCGR3B, SERPINA1, RNASET2, LILRA1, FCER1G, PPBP, BTNL9, ABHD17B, IGLL1, GGH, LILRB2, METTL7A, PIK3IP1, DEFA1, SEMA4D, IL18R1, IL27RA, IFNGR1, DFNA5, DPEP1, IRAK3, ITGAM, LYZ, LY96, MDK, IL18RAP, CDC45
Transit peptide	FH, MTHFD2, PPIF
Immunoregulation	FKBP5
Pseudogene	SNX29P2, LOC100996643

© The Author(s) 2021. Published by Higher Education Press

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

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

(B) Model terminology abbreviations

AR	Accurate rate
ER	Error rate
FBM	Fundamental Boolean model
FBN	Fundamental Boolean network
TFBN	Temporal Fundamental Boolean network
MMR	Mismatched rate
PMR	Perfect matched rate

(C) Boolean function notations

&	Logical *And* connector
|	Logical *Or* connector
!	Logical negation symbol

(D) Boolean model notations

$\neg$	A negation operator that turns a Boolean function from TRUE to FALSE or vice versa
$\times, \cap$	Logical *And* operator
+	Logical *Or* operator
τ	An incremental variable presenting the number of time steps processed
ϑ	A decay period reflects the fact that the attenuation or enhancement of the expression of mRNA requires time
σ_i^t	Boolean state of gene σ_i at time t
σ_i^{t+1}	Boolean state of gene σ_i at time t +1
$l_a(i)$	Total number of fundamental Boolean functions
$l_d(i)$	Total number of fundamental Boolean functions
$f_{a_k}^i$	A fundamental Boolean function of the activation
$f_{d_k}^i$	A fundamental Boolean function of inhibition
$P[[x]]$	A Boolean function that takes a uniform distributed random number, μ, and an output of 1 if μ < x and 0 otherwise.
$V\{x\}$	Logical connective function of *Or*
T_i^j	The best previous time step value for the activation function $f_{a_j}^i$
T_i^k	The best previous time step value for the inhibition function $f_{d_k}^i$
A_i^j	The set of required inputs for the gene state functions, $f_{a_j}^i$
D_i^k	The set of required inputs for the gene state functions, $f_{d_k}^i$

(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.

© The Author(s) 2021. Published by Higher Education Press

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

1. Chen, L., Kulasiri, D. and Samarasinghe, S. (2018) A novel data-driven Boolean model for genetic regulatory networks. Front. Physiol., 9, 1328
2. Schmidt, S., Rainer, J., Riml, S., Ploner, C., Jesacher, S., Achmüller, C., Presul, E., Skvortsov, S., Crazzolara, R., Fiegl, M., *et al*. (2006) Identification of glucocorticoid-response genes in children with acute lymphoblastic leukemia. Blood, 107, 2061−2069
3. Shmulevich, I., Dougherty, E. R., Kim, S. and Zhang, W. (2002a) Probabilistic Boolean networks: a rule-based uncertainty model for gene regulatory networks. Bioinformatics, 18, 261−274
4. Saboury, A. A. (2009) Enzyme inhibition and activation: A general theory. J. Iran. Chem. Soc., 6, 219−229
5. Fontes, R., Ribeiro, J. M. and Sillero, A. (2000) Inhibition and activation of enzymes. The effect of a modifier on the reaction rate and on kinetic parameters. Acta Biochim. Pol., 47, 233−257
6. Fauré, A., Naldi, A., Chaouiya, C. and Thieffry, D. (2006) Dynamical analysis of a generic Boolean model for the control of the mammalian cell cycle. Bioinformatics, 22, e124−e131
7. Ruz, G. A., Goles, E., Montalva, M. and Fogel, G. B. (2014) Dynamical and topological robustness of the mammalian cell cycle network: a reverse engineering approach. Biosystems, 115, 23−32
8. Hwang, W. and Lee, D. (2010) Cell signaling dynamics analysis in leukemia with switching Boolean networks. Comput. Syst. Biol., 13, 168−175
9. Saadatpour, A., Wang, R. S., Liao, A., Liu, X., Loughran, T. P., Albert, I. and Albert, R. (2011) Dynamical and structural analysis of a T cell survival network identifies novel candidate therapeutic targets for large granular lymphocyte leukemia. PLOS Comput. Biol., 7, e1002267
10. Zañudo, J. G. and Albert, R. (2013) An effective network reduction approach to find the dynamical repertoire of discrete dynamic networks. Chaos, 23, 025111
11. Saadatpour, A., Albert, R. and Reluga, T. C. (2013) A reduction method for Boolean network models proven to conserve attractors. SIAM J. Appl. Dyn. Syst., 12, 1997−2011
12. Campbell, C. and Albert, R. (2014) Stabilization of perturbed Boolean network attractors through compensatory interactions. BMC Syst. Biol., 8, 53
13. Saez-Rodriguez, J., Simeoni, L., Lindquist, J. A., Hemenway, R., Bommhardt, U., Arndt, B., Haus, U. U., Weismantel, R., Gilles, E. D., Klamt, S., *et al*. (2007) A logical model provides insights into T cell receptor signaling. PLOS Comput. Biol., 3, e163
14. Wittmann, D. M., Krumsiek, J., Saez-Rodriguez, J., Lauffenburger, D. A., Klamt, S. and Theis, F. J. (2009) Transforming Boolean models to continuous models: methodology and application to T-cell receptor signaling. BMC Syst. Biol., 3, 98
15. Polyanin, A. D. and Zaitsev, V. F. (2003) Handbook of Exact

 © The Author(s) 2021. Published by Higher Education Press

Solutions for Ordinary Differential Equations (2nd ed.). Boca Raton: Chapman & Hall/CRC Press
16. Ling, H., Samarasinghe, S. and Kulasiri, D. (2013) Novel recurrent neural network for modelling biological networks: oscillatory p53 interaction dynamics. Biosystems, 114, 191−205
17. Wang, Z., Huang, D., Meng, H. and Tang, C. (2013) A new fast algorithm for solving the minimum spanning tree problem based on DNA molecules computation. Biosystems, 114, 1−7
18. Kim, S. Y., Imoto, S. and Miyano, S. (2003) Inferring gene networks from time series microarray data using dynamic Bayesian networks. Brief. Bioinform., 4, 228−235
19. Akutsu, T., Miyano, S. and Kuhara, S. (1999) Identification of genetic networks from a small number of gene expression patterns under the Boolean network model. Pac. Symp. Biocomput., 1999, 17−28
20. Liu, F., Zhang, S. W., Guo, W. F., Wei, Z. G. and Chen, L. (2016) Inference of gene regulatory network based on local Bayesian networks. PLOS Comput. Biol., 12, e1005024
21. Wu, H. C., Zhang, L. and Chan, S. C. (2014) Reconstruction of gene regulatory networks from short time series high throughput data: Review and New Findings. In:19th International Conference on Digital Signal Processing (DSP), HongKong IEEE
22. Tušek, A. and Kurtanjek, Ž. (2012) Mathematical modelling of gene regulatory networks. Applied Biological Engineering–Principles and Practice, Naik, G. R. (ed.) Vol. 5. London
23. Liang, S., Fuhrman, S. and Somogyi, R. (1998) Reveal, a general reverse engineering algorithm for inference of genetic network architectures. Pac. Symp. Biocomput., 1998, 18−29
24. Abou-Jaoudé, W., Traynard, P., Monteiro, P. T., Saez-Rodriguez, J., Helikar, T., Thieffry, D. and Chaouiya, C. (2016) Logical modeling and dynamical analysis of cellular networks. Front. Genet., 7, 94
25. Barberis, M., Todd, R. G. and van der Zee, L. (2017) Advances and challenges in logical modeling of cell cycle regulation: perspective for multi-scale, integrative yeast cell models. FEMS Yeast Res., 17, fow103
26. Traynard, P., Tobalina, L., Eduati, F., Calzone, L. and Saez-Rodriguez, J. (2017) Logic modeling in quantitative systems pharmacology. CPT Pharmacometrics Syst. Pharmacol., 6, 499−511
27. Wang, R. S., Saadatpour, A. and Albert, R. (2012) Boolean modeling in systems biology: an overview of methodology and applications. Phys. Biol., 9, 055001
28. Xiao, Y. (2009) A tutorial on analysis and simulation of Boolean gene regulatory network models. Curr. Genomics, 10, 511−525
29. Samuelsson, B. (2006) Dynamics in random Boolean networks. In: Department of Theoretical Physics. Lund: Lund University
30. Silverbush, D., Grosskurth, S., Wang, D., Powell, F., Gottgens, B., Dry, J. and Fisher, J. (2017) Cell-specific computational modeling of the PIM pathway in acute myeloid leukemia. Cancer Res., 77, 827−838
31. Kauffman, S. A. (1969) Metabolic stability and epigenesis in randomly constructed genetic nets. J. Theor. Biol., 22, 437−467
32. Kauffman, S., Peterson, C., Samuelsson, B. and Troein, C. (2003) Random Boolean network models and the yeast transcriptional network. Proc. Natl. Acad. Sci. USA, 100, 14796−14799
33. Jacob, F. and Monod, J. (1961) Genetic regulatory mechanisms in the synthesis of proteins. J. Mol. Biol., 3, 318−356
34. Shmulevich, I. and Dougherty, E. (2005) Modeling genetic regulatory networks with probabilistic Boolean networks. New York: Hindawi
35. Russo, G., Zegar, C. and Giordano, A. (2003) Advantages and limitations of microarray technology in human cancer. Oncogene, 22, 6497−6507
36. Taub, F. E., DeLeo, J. M. and Thompson, E. B. (1983) Sequential comparative hybridizations analyzed by computerized image processing can identify and quantitate regulated RNAs. DNA, 2, 309−327
37. Gautier, L., Cope, L., Bolstad, B. M. and Irizarry, R. A. (2004) affy–analysis of Affymetrix GeneChip data at the probe level. Bioinformatics, 20, 307−315
38. Silvescu, A. and Honavar, V. (2001) Temporal Boolean network models of genetic networks and their inference from gene expression time series. Complex Syst., 13, 61−78
39. Ernst, J. and Bar-Joseph, Z. (2006) STEM: a tool for the analysis of short time series gene expression data. BMC Bioinformatics, 7, 191
40. Ernst, J., Nau, G. J. and Bar-Joseph, Z. (2005) Clustering short time series gene expression data. Bioinformatics, 21, i159−i168
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
42. Wang, Z., Yang, F., Ho, D. W., Swift, S., Tucker, A. and Liu, X. (2008) Stochastic dynamic modeling of short gene expression time-series data. IEEE Trans. Nanobioscience, 7, 44−55
43. Tchagang, A. B., Phan, S., Famili, F., Shearer, H., Fobert, P., Huang, Y., Zou, J., Huang, D., Cutler, A., Liu, Z., *et al*. (2012) Mining biological information from 3D short time-series gene expression data: the OPTricluster algorithm. BMC Bioinformatics, 13, 54
44. Bockmayr, A. (2009) Logic-based modeling in systems biology. In: Logic Programming and Nonmonotonic Reasoning, Erdem, E., Lin, F. and Schaub, T. (eds.). Heidelberg: Springer
45. Irizarry, R. A., Wu, Z. and Jaffee, H. A. (2006) Comparison of Affymetrix GeneChip expression measures. Bioinformatics, 22, 789−794
46. Affymetrix, Inc. (2002) Statistical algorithms description document. Part number 701137 Rev 3
47. Irizarry, R. A., Hobbs, B., Collin, F., Beazer-Barclay, Y. D., Antonellis, K. J., Scherf, U. and Speed, T. P. (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics, 4, 249−264
48. Wu, Z., Irizarry, R. A., Gentleman, R., Martinez-Murillo, F. and Spencer, F. (2004) A Model-Based Background Adjustment for Oligonucleotide Expression Arrays. J. Am. Stat. Assoc., 99,

© The Author(s) 2021. Published by Higher Education Press

909−917
49. Weinberg, R. (2007) The biology of cancer. New York: Garland Science
50. Hornberg, J. J., Bruggeman, F. J., Westerhoff, H. V. and Lankelma, J. (2006) Cancer: a systems biology disease. Biosystems, 83, 81−90
51. Hanahan, D. and Weinberg, R. A. (2000) The hallmarks of cancer. Cell, 100, 57−70
52. Martinez, J. D., *et al*. (2003) Molecular Biology of Cancer. Canada: John Wiley & Sons, Inc
53. Banjar, H., Adelson, D., Brown, F. and Chaudhri, N. (2017) Intelligent techniques using molecular data analysis in leukaemia: An opportunity for personalized medicine support system. BioMed Res. Int., 2017, 3587309
54. Inaba, H., Greaves, M. and Mullighan, C. G. (2013) Acute lymphoblastic leukaemia. Lancet, 381, 1943−1955
55. Hunger, S. P. and Mullighan, C. G. (2015) Acute lymphoblastic leukemia in children. N. Engl. J. Med., 373, 1541−1552
56. Green, D. R. (2011) Means to an End: Apoptosis and Other Cell Death Mechanisms. Cold Spring Harbor Laboratory Press
57. Lakna. Difference Between Apoptosis and Necrosis. (2017) Available from the website of PEDIAA
58. Schmidt, S., Rainer, J., Ploner, C., Presul, E., Riml, S. and Kofler, R. (2004) Glucocorticoid-induced apoptosis and glucocorticoid resistance: molecular mechanisms and clinical relevance. Cell Death Differ., 11, S45−S55
59. Thompson, E. B. and Johnson, B. H. (2003) Regulation of a distinctive set of genes in glucocorticoid-evoked apoptosis in CEM human lymphoid cells. Recent Prog. Horm. Res., 58, 175−197
60. Planey, S. L., Abrams, M. T., Robertson, N. M. and Litwack, G. (2003) Role of apical caspases and glucocorticoid-regulated genes in glucocorticoid-induced apoptosis of pre-B leukemic cells. Cancer Res, 63, 172−178
61. Schmidt, S., Irving, J. A., Minto, L., Matheson, E., Nicholson, L., Ploner, A., Parson, W., Kofler, A., Amort, M., Erdel, M., *et al*. (2006) Glucocorticoid resistance in two key models of acute lymphoblastic leukemia occurs at the level of the glucocorticoid receptor. FASEB J., 20, 2600−2602
62. Smith, L. K. and Cidlowski, J. A. (2010) Glucocorticoid-induced apoptosis of healthy and malignant lymphocytes. Prog. Brain Res., 182, 1−30
63. Rhen, T. and Cidlowski, J. A. (2005) Antiinflammatory action of glucocorticoids–new mechanisms for old drugs. N. Engl. J. Med., 353, 1711−1723
64. Rainer, J., Lelong, J., Bindreither, D., Mantinger, C., Ploner, C., Geley, S. and Kofler, R. (2012) Research resource: transcriptional response to glucocorticoids in childhood acute lymphoblastic leukemia. Mol. Endocrinol., 26, 178−193
65. Yoshida, N. L., Miyashita, T., U, M., Yamada, M., Reed, J. C., Sugita, Y. and Oshida, T. (2002) Analysis of gene expression patterns during glucocorticoid-induced apoptosis using oligonucleotide arrays. Biochem. Biophys. Res. Commun., 293, 1254−1261
66. Bachmann, P. S., Gorman, R., Papa, R. A., Bardell, J. E., Ford, J., Kees, U. R., Marshall, G. M. and Lock, R. B. (2007) Divergent mechanisms of glucocorticoid resistance in experimental models of pediatric acute lymphoblastic leukemia. Cancer Res., 67, 4482−4490
67. Carlet, M., Janjetovic, K., Rainer, J., Schmidt, S., Panzer-Grümayer, R., Mann, G., Prelog, M., Meister, B., Ploner, C. and Kofler, R. (2010) Expression, regulation and function of phosphofructo-kinase/fructose-biphosphatases (PFKFBs) in glucocorticoid-induced apoptosis of acute lymphoblastic leukemia cells. BMC Cancer, 10, 638
68. Lee, W. P. and Tzou, W. S. (2009) Computational methods for discovering gene networks from expression data. Brief. Bioinform., 10, 408−423
69. Ay, A. and Arnosti, D. N. (2011) Mathematical modeling of gene expression: a guide for the perplexed biologist. Crit. Rev. Biochem. Mol. Biol., 46, 137−151
70. Wang, Y., Zhang, X. S. and Chen, L. (2011) Computational systems biology: integration of sequence, structure, network, and dynamics. BMC Syst. Biol., 5, S1
71. Hood, L. (2013) Systems biology and p4 medicine: past, present, and future. Rambam Maimonides Med. J., 4, e0012
72. Barry, M. and Feely, J. (1990) Enzyme induction and inhibition. Pharmacol. Ther., 48, 71−94
73. Lazzarini, N., Widera, P., Williamson, S., Heer, R., Krasnogor, N. and Bacardit, J. (2016) Functional networks inference from rule-based machine learning models. BioData Min., 9, 28
74. Albert, R. (2004) Boolean modeling of genetic regulatory networks. Lect. Notes Phys, 650, 459−481
75. Hopfensitz, M., Müssel, C., Maucher, M. and Kestler, H. A. (2013) Attractors in Boolean networks: a tutorial. Comput. Stat., 28, 19−36
76. Han, J., Kamber, M. and Pei, J. (2012) Data Mining-concepts and Techniques, 3rd ed. Morgan Kaufmann Publishers
77. RefSeq. (2020) Available from the website of NIH
78. Calvo, F., Ranftl, R., Hooper, S., Farrugia, A. J., Moeendarbary, E., Bruckbauer, A., Batista, F., Charras, G. and Sahai, E. (2015) Cdc42EP3/BORG2 and septin network enables mechanotransduction and the emergence of cancer-associated fibroblasts. Cell Rep., 13, 2699−2714
79. Genecards. org. F13A1 Gene. (2020) Available from the website of GeneCards
80. Genecards. org. TUBA4A. (2020) Available from the website of GeneCards
81. Zhang, H., Duan, J., Qu, Y., Deng, T., Liu, R., Zhang, L., Bai, M., Li, J., Ning, T., Ge, S., *et al*. (2016) Onco-miR-24 regulates cell growth and apoptosis by targeting BCL2L11 in gastric cancer. Protein Cell, 7, 141−151
82. Bai, T. , Zhao, Y. , Liu, Y. , Cai, B. , Dong, N. and Li, B. (2019) Effect of KNL1 on the proliferation and apoptosis of colorectal cancer cells. Technol. Cancer Res. Treat., 18, 153303381985-8668
83. Zhang, R., Deng, Y., Lv, Q., Xing, Q., Pan, Y., Liang, J., Jiang, M., Wei, Y., Shi, D., Xie, B., *et al*. (2020) *SQLE* promotes

 © The Author(s) 2021. Published by Higher Education Press

differentiation and apoptosis of bovine skeletal muscle-derived mesenchymal stem cells. Cell. Reprogram., 22, 22−29

84. Zou, W., Ma, X., Hua, W., Chen, B., Huang, Y., Wang, D. and Cai, G. (2016) BRIP1 inhibits the tumorigenic properties of cervical cancer by regulating RhoA GTPase activity. Oncol. Lett., 11, 551−558
85. Cui, H., Wang, Q., Lei, Z., Feng, M., Zhao, Z., Wang, Y. and Wei, G. (2019) DTL promotes cancer progression by PDCD4 ubiquitin-dependent degradation. J. Exp. Clin. Cancer Res., 38, 350
86. Chaiboonchoe, A., Samarasinghe, S., Kulasiri, D. and Salehi-Ashtiani, K. (2014) Integrated analysis of gene network in childhood leukemia from microarray and pathway databases. BioMed Res. Int., 2014, 278748
87. Chaiboonchoe, A. , Samarasinghe, S. and Kulasiri, D. (2009) Using emergent clustering methods to analyse short time series gene expression data from childhood leukemia treated with glucocorticoids. In: 18th World IMACS/MODSIM Congress, pp. 13–17. Cairns, Australia
88. Berrar, D. P. , Dubitzky, W. and Granzow, M. (2003) Introduction to microarray data analysis. In: A Practical Approach to Microarray Data Analysis. Kluwer Academic Publishers
89. Li, Y., Jann, T. and Vera-Licona, P. (2019) Benchmarking time-series data discretization on inference methods. Bioinformatics, 35, 3102−3109
90. Catlett, J. (1991) On changing continuous attributes into ordered discrete attributes. In: EWSL'91: Proceedings of the 5th European Conference on European Working Session on Learning, pp. 164–178. Heidelberg: Springer
91. Jiang, S. -Y. , Li, X. , Zheng, Q. and Wang, L-X. (2009) Approximate equal frequency discretization method. In: 2009 WRI Global Congress on Intelligent Systems, pp. 514–518, Xiamen, China
92. MacQueen, J. (1967) Some methods for classification and analysis of multivariate observations. In: Proc. of the fifth Berkeley Symposium on Mathematical Statistics and Probability. Berkeley: University of California Press
93. Gupta, A., Mehrotra, K. and Mohan, C. K. (2009) A clustering based discretization for supervised learning. Statis. & Prob. Lett. 80, 816−824
94. Dimitrova, E. S., Licona, M. P., McGee, J. and Laubenbacher, R. (2010) Discretization of time series data. J. Comput. Biol., 17, 853−868
95. Schmidberger, G. and Frank, E. (2005) Unsupervised discretization using tree-based density estimation. In: Knowledge Discovery in Databases: Pkdd, 2005, Jorge, A.M., Torgo, L., Brazdil, P., Camacho, R., and Gama, J. (eds). Heidelberg: Springer
96. Müssel, C., Hopfensitz, M. and Kestler, H. A. (2010) BoolNet–an R package for generation, reconstruction and analysis of Boolean networks. Bioinformatics, 26, 1378−1380
97. Win, H. M. L. and Htwe, N. A. A. (2014) Comparison between edge detection and K-means clustering methods for image segmentation and merging. Inter. J. Scient. Engin. Technol. Res., 3, 3012−3017
98. Wheeler, D. C. (2007) A comparison of spatial clustering and cluster detection techniques for childhood leukemia incidence in Ohio, 1996−2003. Int. J. Health Geogr., 6, 13
99. Huang, W., Sherman, B. T. and Lempicki, R. A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc., 4, 44−57
100. Huang, W., Sherman, B. T. and Lempicki, R. A. (2009) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res., 37, 1−13

© The Author(s) 2021. Published by Higher Education Press