Interactome and Pathway Analysis of EBNA Transformed Human B-Cell

CHEN MENG

KTH Computer Science and Communication

Master of Science Thesis
Stockholm, Sweden 2011
Interactome and Pathway Analysis of EBNA Transformed Human B-Cell

CHEN MENG

Master's Thesis in Biomedical Engineering (30 ECTS credits) at the Computational and Systems Biology Master Programme
Royal Institute of Technology year 2011
Supervisor at CSC was Erik Aurell
Examiner was Anders Lansner

TRITA-CSC-E 2011:113
ISRN-KTH/CSC/E--11/113--SE
ISSN-1653-5715

Royal Institute of Technology
School of Computer Science and Communication

KTH CSC
SE-100 44 Stockholm, Sweden
URL: www.kth.se/csc
Abstract

Epstein-barr virus (EBV) is a member of human herpesvirus family which infects most human populations without obvious symptom, that is, latent infection. EBV establishes the latent infection through protein interactions between human and viral encoded proteins.

EBV encoded nuclear antigens (EBNA) are important viral gene products which are essential for maintaining EBV infection phenotype, such as proliferation and anti-apoptosis. Through interacting with EBNA binding proteins in host cell, EBNAs modulate both viral and cellular transcription.

In this project, I started from identified EBNA binding proteins and performed interactome and pathway analysis. A global map of pathways regulated by EBNAs is reconstructed. In addition, a simplified qualitative model is implemented. Results show that most EBNAs (except EBNA-6) have anti-apoptosis effects and that the Wnt pathway is essential for cell proliferation.
Chapter 1

Introduction

1.1 Project Goal and Outline

A multi-cellular life form requires both intracellular and inter-cellular cooperation of signaling pathways to successfully realize a biological function. In order to organize the behavior of a cell properly, cells recruit highly elaborated pathways to control their manner of response to varied kinds of external and internal stimuli.

Epstein-Barr virus (EBV), a human herpesvirus which infects most adult in the population of human, has developed a complex system which allow it to escape the supervising by the human immune system and to interfere with several cellular processes, including apoptosis, proliferation, differentiation, etc. EBV latent infection is established through the interaction of viral proteins and host cellular components, which further regulates signal transduction pathways to maintain the presence of the EBV episome and the EBV latent phenotype.

In this project, I explore the interactions of EBV encoded nuclear antigens (EBNAs) and their interacting cellular proteins, to determine the role of EBNAs in B-cell transformation and to avoid a global point of view toward mechanism of B-cell immortalization. There are four steps two achieve this goal:

1. determining the proteins interacting with EBNAs directly (Distance 1 protein, D1 protein) and proteins interacting with D1 proteins (Distance 2 protein, D2 protein) through manual curation and database exploration;

2. finding pathways extensively affected by D1 and D2 proteins via enrichment analysis and manual curation;
3. reconstructing a network connecting these pathways and translating this network into a Boolean logical model;

4. analyzing the network property and dynamic property of the reconstructed system.

1.2 EBV

Epstein-Bar virus (EBV) is a member of human herpesvirus family which preferentially infects B lymphocyte, and also epithelial cells, with lower efficiency. Over 90% of the adult population is infected by EBV without obvious symptoms. In EBV transformed lymphoblastoid cell lines (LCL), cells contain multi-copies of circular viral episome (DNA exist independent of host chromosomes) and express six EBV nuclear antigens (EBNAs 1-6), three latent membrane proteins (LMPs 1, 2a and 2b), EpsteinBarr encoded RNAs (EBER) and BARTs. This program is the type III latency. In some other latency types, less genes are expressed, e.g. in Burkitt’s lymphoma (BL) only requires expression of EBNA-1, EBERs as well as BARTs (latency I). In the establishment of Hodgin lymphoma (HL), EBNA-1, LMPs, EBERs and BARTs are expressed (Latency II) [1]. In this study, I mainly focus on LCL, more precisely, on the role of EBNAs in the LCL establishment.

As shown in figure [1.1] the 170kb circulized EBV chromosome contains one replication origin site (oriP) and the two promoters, namely, Wp and Qp. All EBNAs are transcribed either from Wp (from Qp in latency I and II) with different splicing manner. EBNA-5 transcripts contain a highly variable number of the W-repeats. The transcription of all LMPs are controlled by EBNA-2. LMP-1 and LMP-2b are from the same bidirectional promoter, while LMP-2a is transcribed from a separate one [2].

LMP-1 is an important transforming protein which could constitutive activate tumour necrosis factor receptor (TNFR) and subsequently activate other pathways including NF-κB pathway and MAPK pathway, etc. LMP-2a and LMP-2b are less important and mainly target on B-cell receptors (BCR) and its downstream pathways [3]. The following part gives more instruction about EBNAs and their interaction proteins.
EBNAs and Their Interacting Proteins

1.3.1 EBNA-1

EBNA-1 was the first identified EBV gene product (Reedman and Klein) and it may be the only EBNA expressed in type I latency, for example, in the Burkitt’s lymphomas (BLs). EBNA-1 binds both viral genome and host cell genome in a sequence dependent manner to induce the replication of viral genome [4].

Also some host cellular components cooperate with EBNA-1 to help it to implement expression regulation function, e.g. ORC (human origin recognition complex), which protein has been shown involved in the viral genome replication [5]; RPA (8 replicative single-strand DNA binding protein), helps host factors to access to oriP and works as early step of latent origin activation [6]; BRD4 (a trans-active protein preferentially...
attached the family of repeat (FR) elements, which is regulated by EBNA-1. EBNA interacts with TAF-1 and NAP-1, proteins involved in replication and transcription of other compacted templates, so it seems that these interactions could regulate EBNA-1’s replication function. Specifically, TAF is considered a down-regulator since the expression of TAF negatively relates to EBNA-1 trans-active activity [7]. EBNA-1 deficient in binding with EBP2 fail to maintain oriP plasmid activity while has no influence on transient oriP plasmid replication. EBP2-EBNA-1 interaction is supposed to play an vital role in stable partition of EBV episomes [8]. P32/Tap, a protein with function believe in RNA processing and binding to hyaluronic acid. Some literature data suggest that intracellular hyaluronic acid is able to affect cell proliferation, hence, this interaction is postulated to help the maintenance of episomal viral DNA in proliferating cells [9].

In addition to the trans-activation function, it is believed that EBNA-1 also involves in some intracellular pathways. HAUSP, an ubiquitin-specific protease, and CK2 are found to interact with EBNA-1; both of proteins disrupt PML nuclear bodies in independent mechanisms [10]. Except in regulating PML, EBNA-1 was suggested to interfere with p53 pathway based on the fact that HAUSP also involves in p53 pathway [11]. P33/gC1Q-R, a homotrimeric protein that is one of putative C1q binding cell surface expressed protein. C1q plays a role in recognition of the signal that triggers activation of classical pathways. In lymphocytes and some other cell type, such as leukocytes, dendritics cells, endothelial cells et al. it is capable to induce cell specific biological response [12].

By considering the finding that EBNA-1 forms complex with nuclear transporter karyopherin α 1 (NPI1) and karyopherin α2 (Rch1), a function involved in localization regulatory system of EBNA1 is also proposed [13].

1.3.2 EBNA-2

EBNA-2 is unable to bind DNA directly. However, it works as a trans-activator by recruiting other cellular factors [2]. It plays an important role in the regulation of notch signaling pathway by interacting with RBP-Jκ. RBP-Jκ is a transcriptional repressor in normal condition, while when Notch pathway is activated, Notch intracellular domain (NICD) is cleaved and transported into nucleus, where it binds to RBP-Jκ and some other co-activators, including MAML1, P300 and PCAF, to convert RBP-Jκ to a transcriptional activator. EBNA-2 has similar effects to NICD and could be viewed as an viral counterpart of NICD. Correspondingly, EBNA-2 interacts with several intracellular co-activators, e.g. P300/CBP and PCAF, proteins that participate in Notch pathway and p100, a protein that serves as a bridge to associate STAT5/6 and RNA polymerase
II to facilitate the trans-activation [14, 15].

Apart from trans-activation property, the involvement of EBNA2 in RNA processing is supported by the fact that it interacts with DP103, a member of DEAD box family and SMN [16, 17, 18]. EBNA-2 could also interact with hSNF5 which forms a SNF-SWI complex to participate in gene regulation through the alternation of nucleosome configuration, so Wu (1996) postulated that EBNA-2 engages hSHF-SWI complex to change chromatin conformation thereby strengthen RBP-Jκ-EBNA2-polymerase II transcription complex [19]. Another PPI between EBNA-2 and NUR77, suggests EBNA-2 protect latent EBV infected cell from NUR77-mediated apoptosis [20].

In terms of regulation of viral genome, PU.1 together with RBP-Jκ are employed by EBNA-2 as well as EBNA-6 to activate the expression of viral latent membrane protein LMP-1 [21]. The interaction between EBNA6 and PU.1 facilitates the EBNA-2 dependent expression of LMP-1, which seems essential to supply adequate LMP-1 for cell growth [22].

For a long time, the identification of EBNA-2 target genes is hampered by the fact that almost all viral genes are affected by EBNA-2. As a consequence, it is hard to distinguish which genes are direct targets of EBNA-2 [2]. At present, Kaiser et al. discovered c-myc gene as an EBNA-2 target by applying an EBNA-2 induced expression system. A more comprehensive study on EBNA-2 target genes employing high throughput microarray method has been conducted. As the result, 311 activated and 239 repressed genes were identified in BJABK3 and KB41K3 cell lines. These results also confirmed the regulatory role of EBNA-2 in the Notch signal pathway [23].

1.3.3 EBNA-5

EBNA-5, alternatively named as EBNA leader protein (EBNA-LP), is one of the first expressed EBNA in EBV transformed of B-cell. It plays a role in the initiation of immortalization of EBV-tranformed cells. The length of EBNA-5 peptide is highly variable resulting from different number of W-repeats. The coexpression of EBNA-5 and EBNA-2 results in increasing transcriptional activity of EBNA-2.

The idea that EBNA-5 protects cell from apoptosis is supported by the fact that EBNA-5 associates with heat shock protein 70 family proteins Hsp72/73, p14ARF, Ste1/S3a and MDM2. Among these proteins, p14ARF and MDM2 are the upstream regulators of p53 pathway, an important cancer related pathway inducing growth arrest and apoptosis. The interaction between EBNA-5 and p14ARF are believed to take place.
through elimination of p14ARF-p53 complexes [18]. Whereas MDM2 is able to function as a bridge to connect EBNA-5 and p53 protein, the formation of EBNA3-MDM2-p53 complex represses the trans-activating function of p53, resulting in down-regulation of p53 target genes, including p21 [21]. It is possible that these mechanisms contribute to the survival of infected cells in the condition where p53 highly expressed. EBNA-5 is thought to modulate p53 through heat shock protein 70 family proteins Hsp72/73 as well [25] [26]. Based on the fact that over-expression of Hsp72 strongly up-regulates EBNA-5-EBNA-2 co-activated gene, LMP-1, rather than themselves, researchers (Peng et al. 2007) proposed a model that Hsp72 cooperates with EBNA-5 to shuttle from repressors from EBNA-1-enhanced promoters. This model corresponding to the finding that EBNA-5 interacts with SP100 and HDAC4, proteins that bind to DNA strands and function as co-repressors [27]. Fte-1/S3a is a tumor related protein which interacts with the transcriptional factor CHOP and apoptosis regulator poly (ADP-ribose) polymerase (PARP), a protein holds cell in undifferentiated state and malignant phenotype [28].

EBNA-5 could interact with DNA-PKCs and affect their ability to interact with other proteins, including proteins participated in transcription, DNA repair, p53 pathway. It could interact with HA95, a nuclear protein homologous to AKAP95 as well. AKAP95 is critical in the chromosome condensation during mitosis. It also has been shown that HA95 is involved in the inhibition of EBNA-5 and EBNA-2 trans-activation [18]. HSP27 is another protein associated with EBNA-5, it primarily locates in cytoplasm and is transported into nucleus in response to insult-induced stress, so it is proposed that Hsp27 affects EBNA-5’s function on cell growth according to cell stress [29].

Other proteins that interact with EBNA-5 are prolyl 4-hydroxylase alpha1 subunit, α- and β- tubulin, the function of these interactions are still waiting to be illustrated.

1.3.4 EBNA-3 Family

EBNA-3 family consists of EBNA-3 (EBNA-3a), EBNA-4 (EBNA3b) and EBNA-6 (EBNA-3c) where genes are tandemly arranged in the the location opposite of oriP in the circular viral genome. They are different splicing forms of transcripts encoded by the gene originating from the Cp promoter. All EBNA-3 family members could interact with RBP-Jκ, RBP-2N (an isoform of RBP-Jκ in B-cell), and CHK2. As described before, RBP-Jκ is a main intracellular partner of EBNA-2 which is involved in trans-activation of the Notch signaling downstream genes. It is believed that EBNA-3 family proteins serve as negative regulators of EBNA-2. This point is supported by the PPIs between EBNA-6 and with CtBP, HDAC1, mSIN3a and NCoR, all of which are function as co-repressor in
Notch signaling pathway and other pathways. Apart from these transcriptional repressors, EBNA-6 could interact with P300, an important co-activator also interacting with EBNA-2, so it is possible that EBNA-3 family play a role in the fine-tune of the function of EBNA-2 [2]. Another cellular protein interacts with all EBNA-3 family members is CHK2. Several studies have shown that this binding contributes to the block of G2/M transition by targeting the ATM/ATR signaling pathway [30, 31].

Recently, Nikitin et al. claimed that EBNA-6 is a prerequisite factor to attenuate the EBV-induced DNA damage which process equips B-cell with growth repressive effect. The growth repressive effect is important for B-cell immortalization [32]. In addition to immortalization, EBNA-6 promotes cells enter into the cell cycle via binding MRS18-2 and Prothymosin alpha. MRS18-2 serves as a bridge to connect EBNA-6 and Rb. The formation of EBNA6-MRS18-Rb further results in an increase of free E2F1, a nucleus protein could promote cell come into cell cycle [33]. Prothymosin alpha is a cellular protein previously identified to be important for cell division and proliferation [34].

There are several proteins that interact with both EBNA-2 and EBNA-6, including RBP-Jκ, p300, DP103, SMN.

Most PPIs of EBNA-3 support the opinion that EBNA-3 promotes cycle proliferation and survival. For instance, it is possible that EBNA-3 induces trans-activaton of Aryl Hydrocarbon signal transduction pathway by interacting with aryl hydrocarbon receptor (AhR) and XAP2 so as to regulate proliferation and survival [35]. Furthermore, EBNA-3 is found to interact with vitamin D receptor (VDR). It has been shown that VDR regulated genes involved in proliferation and promoting cell growth. By considering the observation that EBNA-3 inhibits vitamin D3-induced growth arrest and apoptosis and other discoveries, it was recently been proposed that the binding of EBNA-3 could down-regulate the activation of VDR dependent genes, which leads to keep cell from apoptosis [36].

Other ascertained cell partners of EBNA-3 are TCP1 which is thought to help nascent EBNA-3 folding and UK/UPRT which promote the synthesis of UTP by tuning the Salvage pathway [18].
1.4 Biological Networks

From the requirement of systems biology, biological networks, e.g. gene regulatory network (GRN), protein protein interaction network (PPIN), signal transduction network (STN), metabolic network (MBN), receive increasingly more attention in biological sciences. Among them, PPIN is one of the most well studied networks. Proteins are direct executors of information encoded in genes which implement most cellular functions, including performing catalysis, transporting or storing other molecules, providing mechanical strength, conferring immunity, transmitting signals, and controlling growth and development (chapter 9 of reference [37]). However, a monomeric protein, that is, a single protein molecule, rarely acquires its biological function. Proteins are converted to activated forms either through binding with partner or through undergoing some post-translational modifications (PTM) processes, such as phosphorylation, methylation, acetylation, ubiquitylation, etc. Some proteins even show distinct functions when it binds different partners or experiences different PTMs.

Due to the cooperating property of proteins, they interact with each other to form a large PPIN. Along with the accumulation of results from protein protein interaction experiments, there is an increasing demand to reconstruct PPIN. The most traditional
method creating PPIN is literature-based, that is, manual or computationally mining data from published papers. These PPI data are mainly supported by classical physiological detection, such as coimmunoprecipitation, chromatography, yeast two hybridization. At present, lots of PPI databases are based on this method, for example, HPRD, BIND, DIP, MINT, IntAct. \[39\]. As the development of computational science, PPI can be predicted as well, the reliability of computational prediction is becoming increasingly better because of the progress in prediction algorithm. This is the so-called "in silico" experiment. To date, databases, such as HPID, Pridictome, STRING, PIPs, FunCoup, etc, combine manually curated data and computational data to get a more comprehensive PPI network. Recently, PPI network construction was also benefited from the advance of high throughput technology, e.g. protein microarray, mass spectrometry (MS), etc, which make it possible that measure PPI in a proteomic scale. However, the result of high throughput method remains noisy because only a small part of interactions are valid, which makes it is hard to estimate the rate of false positive and false negative rates \[37\]. In total, every PPI construction method has its advantages and disadvantages, the increasingly more precise PPI data will result from the integration of a variety of methods.

The network study, including PPIN, enjoyed a fast development over last decades. Some concepts are usually used in the network study: (concepts based on reference \[38\])

**Digree (k)** means how many links a node has to other nodes.

**Digree distribution P(k)** denotes the probability that a selected node has exactly $k$ links.

**Shortest path and mean path length ($l_{AB}$)** Given two nodes in a network, many paths is able to connect them, the path with smallest number of edges between them is the *shortest path*. In the whole network, *mean path length* is the mean of shortest paths over all pairs of nodes

**Clustering coefficient ($C_i$)** measures the number of triangles that go through a node. It is calculated as $C_i = 2n_I/k(k-1)$. In this equation, for a node $I$, $n_I$ is the number of links between all neighbours of the node $I$.

It has been shown that many of these networks follow a scale-free topology shown in figure 1.2. The remarkable feature of scale-free network is majority of nodes have few links, while a small number of nodes, which are called hubs, are highly connected. Further study shows these hubs are not tend to connect with each other, conversely, they preferentially link to nodes with few links. This kind of topology exhibit both topology robustness and functional robustness against random failure. Moreover, *small-world effect,*
a phenomenon widely exist from neural networks to the World Wide Web, is observed in scale free network, it means that any two nodes in the network are able to be linked with a path containing only few nodes. The small-world property hints that local perturbation of a single node could be spread to the whole network very quickly. (Chapter 4 in [37])

Except these basic properties of a network, more advanced analysis is explored as well. For example, the most important node in a pathway tend to be a hub nodes, so the detection of links between hubs could give a possible links between pathway. Another example is based on the fact that a large complex involved in a process always consists of several protein molecules, which proteins are inter-connected with each other. So the identification of hyper-linked subnet is able to discover a protein comlpex. Algorithms are implemented to unravel such properties, such as MCODE [40].

![Figure 1.3](image_url) – The pathways regulatory circuit involved in cancer development of a mammalian cell. Reproduced from Hanahan and Weinberg (2000) [41]

It is true that PPI network analysis plays an important role in understanding biological process globally, nevertheless, simple PPI network (PPIN) has shortcoming as well. For instance, most PPI networks are undirected, as a consequence, it is hard to know how a signal is transduced in PPI network. In this case, signal transduction network (STN) as shown in figure 1.3 should be applied. Cells response to an external or internal stimuli through signal transduction pathways, all pathways cross-talk with others rather than being independent. Different pathways and their cross-talking form STN. Signal transduction network which mainly consists of protein protein interactions, equips some
distinct features compare with PPI. The most significant difference is that STN is direct
graph, whereas PPI is not direct. Directed interaction further makes the network has
input and output layers, upstream nodes and downstream nodes. From this point of view,
STN is more informative compared with PPIN, as it is a higher level network comparing
with PPIN. In general, an STN starts from a ligand binding receptor, ends with a tran-
scriptional factor (TF) or metabolic enzyme and there are signal transduction cascades
between them. The signal transduction cascade mainly consists of reversible reactions
between reactors such as phosphorylation, complex formation, etc. As a consequence, the
reactors in this process are rather more stable than in metabolic network which mostly
involves mass flow, that is, the synthesis and degradation of metabolites. Secondly, not
all PPIs are followed by chemical reactions and in turn, many components in STN, e.g.
hormones, small nucleotides and ions, etc, are not present in PPIN, so they cannot be
mirrored with each other. Owing to the informative in STN, the in silicon experiments
is harder to be applied, so most pathway databases are based on literature mining, e.g.
KEGG, REACTOME, PID, etc. Currently, high throughput experiments also offers a
power tool to integrate pathways \[42\].

1.5 Grapical Representation of Signaling Network

Along with accumulation of knowledge in molecular biology and the development of sys-
tems biology, biological networks are becoming more complicated. Hence, there is an
increasing requirement of a standard method for graphical presentation of molecular in-
teraction maps, especially in dealing with large bio-regulatory network.

The graph could be classified in different ways. Kohn (2000) classified biological net-
work diagram into either "heuristic" or "explicit". "Heuristic" graph contains detailed
knowledge of all possible reaction paths is not required and some degree of uncertainty is
allowed, whereas "explicit" diagrams meaning that the diagrams are totally unambiguous
and suitable for simulation \[43\].

Researchers are able to benefit from the standard graphical network representation
in at least three aspects. First, it helps people to integrate existing knowledge. This
is important to get a global point of view on biological systems. Second, visualization
is a good way to store, explain and exchange ideas. A graphical representation is more
inspiring compared to a list of reactions. The last one is a requirement to simulations.
During modeling of biological network, the most common starting point is to reconstruct
the graphical network in order to guide the modeling process.
Despite great advantages, there are also difficulties to do the graphical representation. First, there are large number of molecules and interactions in networks. Second, graphical representation is also hampered by the complexity of bio-molecular interactions. It is undoubtedly that the transcription and translation are extensively regulated by proteins. Even in the post-translation stage, there are several kinds of post-translational modifications (PTMs) such as phosphorylation, acetylation, and ubiquitylation. Except the above events, proteins form multi-molecular complex to fulfill their function. We also should keep in mind that there could be some important mechanisms are yet unknown. Often their interactions have diverse effects on function. So the potential number of modification-multimerization combinations is extremely large. Third, the incompleteness and uncertainty of knowledge. Especially in large networks contain several pathways, it is known that there is cross-talking between different pathways. And different reactions are ascertained to different accuracy.

To solve these problems, additional rules are introduced with different objectives. For example, Kohn firstly introduced a standard biological network representation method and applied it on cell cycle and DNA repair systems and EGFR pathway [44, 43]. However, there are some disadvantages in Kohn’s network graph, for instance, it does not describe the time sequence of a biological process. More importantly, it cannot illustrate different roles of the same molecule involved in multi-reactions. To resolve this drawback, Kitano proposed a human-readable diagrammatic representation method, that is, a state transition-based diagram [45]. Different from entity-relationship diagram, state transition-based diagram means that the reactant and product in a reaction are the same entity but within different states, for example, phosphorylated, involved in large complex, DNA translated into RNA, etc., whereas protein interactions corresponding to the modification of state transition, they could be inhibiting or facilitating. The most outstanding feature of Kitano’s graphical map is that it is supported by CellDesigner, a software could be used to edit Kitano’s diagrammatic network maps. As the requirement of rule-based modeling, Faeder et al. further revised Kitano’s diagram with an additional definition for each reaction and species [46]. This makes the process diagram more constrained and easier to be translated into mathematical model and machine-readable format.

We can image that the graphic could be more complicated with the increasing requirement of representational precision, so we should make a tradeoff between comprehensibility and precision. In the study under discussion, I used Kitano’s diagram for two reasons. First, this map is a global network, different reactions are studied in different
Figure 1.4 – comparison of Kitano’s and Faeder’s graphical representation [46]. The left graph is a set of signal transduction reactions represented by Kitano’s convention. Right one Faeder’s map toward same reactions. Both of them are state-transition based map and unambiguity, while Faeder’s one is more informative. The corresponding BioNetGen Language (BNGL) definition is given for every species in the graph, which could be easily used in rule-based modeling.

details, a precise reaction mechanism is not always available, some reactions are even based on putative. Second is that CellDesigner is used easily to edit this graph.

1.6 Logic Model and Network Analysis

Mathematics is the language of science! Along with the increasing complexity of network, it becomes impossible to understand its dynamic intuitively, mathematics then is recruited as a powerful tool to understand the behaviour of complex system. One important application of that is constructing models for studying a system. To date various kinds of models, from Boolean models to ODE models and to non-deterministic ones, are implemented for different purposes. However, there is not one type of model which could simply be viewed as the best, rather, each has its pros and cons. For example, an ODE model could be employed to study a system quantitatively, while it requires a large number of variables to be decided and becomes computationally intensive in a large system. In this project, there are several tens nodes and edges in the system, many reactions between nodes are omitted and for most reactions, the only supplied information is repressing or activating of two proteins. Hence, a Boolean logical model is applied in this work.

Boolean logical calculation has lots of subclasses, one of them, Propositional Calculus, is described here. The definition and description are given by Mounia Lalmas in 1998 [47]. In logical L, a vocabulary is connected by logical connectives. For example, a vocabulary composed of propositions \( P=p, q, r, s, \ldots \) and logical connectives are negation (\( \neg \)),
Table 1.1 – Semantics of negation, conjunction, disjunction, implication and equivalence

<table>
<thead>
<tr>
<th></th>
<th>p</th>
<th>q</th>
<th>¬q</th>
<th>p∧q</th>
<th>p∨q</th>
<th>p→q</th>
<th>p↔q</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

conjunction(∧), disjunction(∨), implication(→) The logical $L$, then, defines a formal language by its syntax and semantics. Syntax and semantics are defined as [47]:

Syntax “The syntax formally specifies a set of well-formed formulae (wffs) or sentences of the logical $L$:

1. if $p$ is a proposition in $P$, then $p \in L$,
2. if $\phi \in L$ and $\psi \in L$, then $\neg \phi \in L$, $\phi \land \psi \in L$, $\phi \lor \psi \in L$, $\phi \rightarrow \psi \in L$ and $\phi \leftrightarrow \psi \in L$,
3. no other formula belongs to $L$”

Semantics “Any formula in $L$ has an intended meaning called a semantic value.” Table 1.1 is a truth table that describes the semantics attached to $\neg$, $\land$, $\lor$, $\rightarrow$ and $\leftrightarrow$ in Classical Logic. (table 1.1 reproduced from [47])
Chapter 2

Methods

2.1 D1 Protein and D2 Protein Selection

D1 proteins are ascertained by a study of literature, mainly based on reference [48]. Databases such as OMIM, UniProt are also considered. D2 proteins which interact with D1 proteins are derived from I2D PPI database. ("http://ophid.utoronto.ca/ophidv2.201/", Version 1.8 is used). I2D database contains both manual curated and predicted PPIs of six model species, that is, human, fly, mouse, yeast, rat and worm [49]. Therefore, in order to get reliable human PPIs, PPIs other than human’s are removed. Finally, 146,312 out of 681,404 PPIs are left. Cytoscape ("http://www.cytoscape.org") is used to visualize the D1 and D2 proteins interaction networks. MCODE ("http://baderlab.org/Software/MCODE") is utilized to identify hyperlinked sub-networks.

2.2 Pathway Selection

I explored the D1 and D2 proteins in several pathway databases, including REACTOME ("http://www.reactome.org/ReactomeGWT/entrypoint.html"), Biocarta ("http://www.biocarta.com"), PID("http://pid.nci.nih.gov") and KEGG ("http://www.genome.jp/kegg") [50] [51]. Two kinds of pathways are selected for further consideration. First ones are pathways in which D1 and D2 proteins are enriched. This study is different from enrichment analysis of differential expressed gene since differential expressed genes are effects of certain causes, it is a bottom-up approach searching for possible reason. However, this case uses a top-down method, that is, we have causes firstly (D1 or D2 proteins), and then predict the effect of these causes. So another class of pathways, the hub nodes of which are directly regulated by D1 or D2 proteins, are selected for further examination as well. Figure 2.1 gives an example for each class of
pathway.

Figure 2.1 – An illustration for pathway selection. The left diagram is an example of D1 and D2 proteins enriched pathway. The right one exemplifies a "hub-regulated pathway”, that is, a hub gene (β-Catenin in this case) is directly regulated by D1 or D2 proteins (shown in red). Both classes of pathways are further considered.

2.3 Network Representation

The D1 and D2 proteins are mapped into pathways in public pathway database, mainly based on KEGG (http://www.genome.jp/kegg/). Other pathway databases, including REACTOME (http://www.reactome.org/ReactomeGWT/entrypoint.html), biocarta (http://www.biocarta.com/) and PID (http://pid.nci.nih.gov/) are also explored [50, 51]. Protein databases such as OMIM, UniProt are taken into consideration as well. The network is edited by CellDesigner according to Kitano’s graph notation, which is shown in figure 2.2 [52].

The Kitano’s state transition map is used to represent the global regulatory map. The notation for the graph are show in figure 2.2.

2.4 Logical Model Implementation

Firstly I reduced the state-transition model into an entity relation model. Only a small number of nodes which are considered important and well supported is retained.

Classical Boolean model is applied in this work, that is, each node is either in 0 or in 1 state at a time point. The state of a node is decided by its upstream nodes. For example, \( G_1, G_2, G_3 \) are upstream nodes of node \( N \), shown as following,
Figure 2.2 – The legend for Kitano’s graphical representation. Selection from Kitano et al. 2005 [45]

\[
\begin{align*}
\text{Then,} & \\
S(N_{(t+1)}) &= F(S(G_1(t)), S(G_2(t)), \ldots, S(G_n(t)))
\end{align*}
\]

where \( S \) is the state of a node, 0 or 1; \( F \) denotes the updating rule for nodes \( N \), which is a logical function. The initial state for every node is assigned and all nodes’ states are updated simultaneously in each time step according to their specific rules \( F \).
Chapter 3

Results

3.1 D1 and D2 Proteins

Table 3.1 shows proteins interacting with EBNAs, most of them are validated by experiments such as yeast two-hybrid system, GST pull down or immunoprecipitation. High throughput methods, such as mass spectrometry, are also employed to detect the EBNA-interacting proteins after immunoprecipitation. A study exploring EBNA-5 binding proteins with mass spectrometry was conducted and showed 147 novel putative partners [53]. Figure 3.1A shows the interactome between EBNAs and their direct host partners. However, only proteins ascertained by low throughput method are considered in the analysis of D2 proteins with aim to get more reliable results. Finally, 1619 D2 proteins is suggested by I2D PPI database, their interaction network is shown in 3.1B.

<table>
<thead>
<tr>
<th>EBNAs</th>
<th>EBNAs binding proteins (D1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EBNA-1</td>
<td>Karyopherins α1, Importin-1 (NPI-1), RPA, EBP2, NAP-1, SET (TAF-1), pp32, TAP/p32, USP7 (HAUSP), CK2, p33/gC1Q-R, BRD4, CDC6</td>
</tr>
<tr>
<td>EBNA-2</td>
<td>RBP-Jκ, SPI1 (PU.1), SNF5, DP103, SMN, SND1 (p100), NUR77, p300/CBP, BS69</td>
</tr>
<tr>
<td>EBNA-5</td>
<td>Hsp27, Hsp70 (Hsp72), Hsc70 (Hsp73), HAX-1, HA95, p1e-1/S3a4ARF, Fte-1/S3a, MDMX/MDM4, HDAC4, SP100, prolyl-4-hydroxylase α1</td>
</tr>
<tr>
<td>EBNA-3,4,6</td>
<td>RBP-jκ, RBP-2N, CHK2/CDS1</td>
</tr>
<tr>
<td>EBNA-3</td>
<td>XAP-1, TCP-1, AhR, UK/UPRT</td>
</tr>
<tr>
<td>EBNA-6</td>
<td>SPI1 (PU.1), CtBP, HDAC1, HDAC2, mSIN3a, NCoR, p300, prothymosin α, DP103, SMN1, S18-2</td>
</tr>
</tbody>
</table>

Sub-network detection showed that several protein complexes are influenced by EBNAs
or EBNA D1 proteins as shown in figure 3.1 C to H. Such complexes include origin recognition complex (ORC), RNA polymerase complex, histone deacetylation complex.

### 3.2 Network Reconstruction

**Notch pathway** (shown in block A, figure 3.2) Notch pathway is a highly conserved pathway existing in most multi-cellular organism, which is important for cell fate decision during development. There are four groups of Notch receptors, namely, Notch-1, -2, -3 and -4. Their intracellular domain (NICD) is cleaved by proteolysis in response to external signals and subsequently transported into the nucleus where it converts RBP-Jκ, the main nuclear effector of NICD, from a repressor to an activator. The target genes of RBP-Jκ include cellular genes, such as Hes1, Hes5, Hey1, cyclin D1, CDC25A, BCL-2, Myc and viral genes, for example, EBNA-6, LMP-1, LMP-2A, etc.

It was been shown that EBV regulates the Notch pathway extensively through cellular-viral protein interactions. Experiments have shown that EBNA-2 and all EBNA-3 family members are able to interact with RBP-Jκ, nevertheless, the mechanism of EBV modulation toward Notch is far from clear on account of both the complexity of the Notch signaling pathway and the uncertainty of EBV encoded genes involved in Notch-modulating. For instance, the entry of NICD into nucleus involves multi-step proteolytic cleavages, which have different function and signaling consequence [54]. Based on the fact that EBNA-6 interacts with RBP-Jκ and its co-repressors, such as CtBP, HDAC, NCoR and mSIN3a etc., it is proposed that EBNA-3 family play roles in the fine-tune of the Notch signaling. The interaction of other EBNA-3 family members with these proteins are left to be examined to support this point of view. In addition to EBNAs, it is also reported that other EBV gene products are engaged in Notch pathway, for example, Anderson and Longnecker suggested that Notch is involved in the EBNA-2-independent expression of LMP-2A [55]. Kusano and Traub reported that RK-BART0, an EBV encoded protein interacting with Notch-4, activates LMP-1 expression through modulating Notch pathway [56].

Regardless of these uncertainties, it is considered in the model that EBNA-2 acts as an activator of RBP-Jκ target genes, especially in regulating anti-apoptotic genes. On the other hand, EBNA-3 family serves as antagonist of EBNA-2 in this process. This simplified model is supported by the recent study about the functional differences between NICDs and EBNA-2, which shown that EBNA-2 is more potent in activating anti-apoptotic genes, such as BCL-2 and Myc [57].
Figure 3.1 – Interactome analysis of EBNAs, D1 and D2 proteins. A network of host-EBNAs interactions. There are 213 nodes and 423 edges. Big red nodes are EBNAs. Rest nodes are proteins directly interacting with EBNAs. Yellow ones are considered in this study, most of them are validated by reliable experimental methods. Light grey ones are not further considered. (There are lot of grey node in EBNA-5)
binding protein since the database used, VirusMint, includes a high throughput experiment in detecting EBNA-5 binding proteins [58, 53]. B An interactome network of D1 and D2 proteins which consists of 1619 nodes and 15652 edges. Pink and grey nodes are D1 and D2 proteins respectively. C to H Hyperlinked subnetworks derived from A and B which probably form complexes in a biological process. They are identified by MCODE analysis [40]. C to G are from the host-EBNA network, A. Network H is a RNA polymerase complex, which is detected from D1-D2 network, that is B. C is origin recognition complex. E is a histone deacetylation complex.

c-Myc  

*c-Myc gene* is a proto-oncogene which is deregulated in most cancers and involved in many biological processes, including cell growth control, cell cycle and apoptosis. It is suggested that c-myc regulates cell growth and cell cycle in an independent manner. The detailed roles of *c-myc* gene in these processes are largely unknown. Therefore, I include it based on the view that c-myc regulates cell cycle in this network. Evidence also suggested that onco-protein mutated Ras takes part in the cell cycle regulation function of c-myc. While there are at least three different mechanisms were postulated to explain the cooperation in this process. One is that mutated Ras protect c-myc from degradation and enable it to induce adequate proteins involved in S phase entry. However, a more sophisticated model is that c-myc not only trans-activates its target genes, but changes the chromatin structure of them as well, so as to make it accessible for other distinct signals, for example, Ras-signal [59]. Due to this uncertainty role of Ras, only c-myc is considered in this network. In fact, in some culture system, it has been shown that ectopic expression of myc in is able to induce G1/S transition of quiescent cells. The validated targets of *c-myc* include CDC25a, E2F2 and cyclin D2, all of which are important cell cycle regulators [60, 61, 62].

**Rb-E2F pathway and cellular proliferation** (shown in block H, figure 3.2) The family of Rb proteins, including Rb, p107 and p130, are key regulators of cell cycle involved both in E2F dependent and in independent regulation. On the one hand, they form complexes with different subgroups of E2F proteins. In quiescent cell, Rb binds the trans-active E2Fs, namely, E2F1, E2F2 and E2F3s, to repress E2F target genes, which generally regulate S phase entry. While p107 and p130 interact with the trans-repressive E2Fs, including E2F4 and E2F5. The complexing blocks the transcription of S phase genes. Studies have shown that E2F4 and either p107 or p130 are detected in E2F-responsive promoters, whereas relative rare promoters are associate with E2F1/2/3a and Rb [63]. During the late G1 phase, the G1 cyclin, cyclin D1, associates with CDK4 and CDK6 to initiate the phosphorylation of Rb family proteins. Phosphorylation of Rb inhibits its binding to E2F proteins [64].
As a consequence, E2F1/2/3a bind their responsive promoters to induce the expression of S phase protein, including ORC, MCM, cyclin E, CDK2, cyclin A, etc. Among them, cyclin E is able to form complex with CDK2, which in turn further activates the trans-activation function of E2F1/2/3a through phosphorylating Rb. This positive feedback loop is important for the fully activating of E2F1/2/3a function in late G1 phase.

During the cell cycle checkpoint or the end of S phase, CDK2, CDK4 and CDK6 are inactivated by their inhibitors referred as CKIs, including p27, p21, p15, p16 [64]. The mechanism that CKIs inactivating CDKs has known as following: CDK-cyclin complex is phosphorylated by CDK-activating kinase (CAK) to be activated, for example, CDC25A is required to activate CDK2 during G1/S transition, CDC25B and CDC25C activates CDK1-cyclin B during entry into mitosis. However, in the presence of CKIs, CDKs will associate with CKI, this reaction blocks their activation by CAK [65]. In the model under discussion, since CAK activation is an automatic process, CAKs are not represented in the network in order to make it easier to read.

On the other hand, in addition to E2F dependent pathway, Rbs also regulate G1/S transition via E2F independent manner. For example, p130 and p107 are able to inhibit CDK2. The inhibition mechanism is similar to the CDK inhibitors (p21, p27, etc.) discussed above [63]. Recently, the ability of pRb to stabilize p27 is reported, the stabilization of p27 results in the inhibition of CDK activity and G1/S progression. Other pRb-mediated mechanisms contributing to the G1/S inhibition include inducing the formation of PML complex, suppressing Ras signaling, etc. [63]. But its effects in these pathways remain unknown, therefore, all these E2F independent mechanisms are not included in the model.

As was mentioned earlier, targets of trans-active E2Fs include cyclin A, cyclin E, CDK1, CDK2, ORC, MCM, p14ARF, APAF-1 and so on [66]. These target genes are classified into three distinct functional groups. The first group is composed of cyclin A, cyclin E and CDK2, etc., which are engaged in cell cycle regulation as discussed above.
Among them, cyclin A is employed to control the G2/M transition and DNA repairment. ORC and MCM are the second group which lead to the replication of host chromosomes and viral episomes. EBNA-1 is able to recruit ORC into the viral episome so as to initiate the replication of viral genome. The stable maintenance of the viral episome in B-cell requires stable partition in every cell cycle, this process is mediated by EBP2 in the model. The third group of E2Fs targets consists of APAF-1 and p14ARF, proteins link the Rb-E2F pathway with apoptosis pathways [66].

AhR, S18-2 and EBNA-3  (shown in block G, figure 3.2) Aryl Hydrocarbon Receptor (AhR) is a ligand activated nuclear receptor which interacts with EBNA-3. It primarily exists in cytoplasm in a complex containing Hsp90 and XAP-2. In response to activating ligands, such as TCDD, XAP-2, it is dissociated from the cytoplasmic complex. The rest Hsp90-AhR complex is, then, transported into nucleus and interacts with nuclear partner ARNT. Finally, the ligand-AhR-ARNT complex serves as a sequence specific transcriptional factor, which preferentially recognizes and interacts with the Xenobiotic response element (XRE) enhancer. It has been shown that AhR and XAP-2 are able to bind EBNA-3. XAP-2 presents in nucleus in the condition of EBNA-3 highly expressed, while the subcellular location of AhR yet not been elucidated in this condition. Hence, it is possible that the translocation of XAP-2 leads to that AhR enters into nucleus, and EBNA-3 further regulates AhR activity here [67].

AhR involves in several pathways. Sears et al. suggested two roles of AhR. First, the absence of AhR triggers the increase of reactive oxidative species (ROS) which results in E2F1-mediated induction of apoptosis. Second, AhR is capable of repressing E2F1-dependent trans-activation of APAF-1 and p73 through forming complex with E2F1 [68]. Therefore, both proliferation and apoptosis are repressed by AhR. It is also suggested that AhR is able to down-regulate ESR1/2 [69]. By considering the fact that both ESR1 and ESR2 are down-regulated in LCL cell lines, it implies AhR is activated in these cells. Of course, it is also possible that interaction of EBNA-3 with AhR differentially affects its targets [35].

AhR is also involved in non-canonical Rb-E2F pathway where it binds E2F1 to inhibit E2F1-induced apoptosis. In the model, it is assumed that EBNA-3 family binding AhR to repress its activity in order to induce G1/S transition, correspondingly, there is a byproduct that the level of ROS becomes higher and proapoptotic genes, including APAF-1 and p73, etc. are activated subsequently. As a consequence, additional mechanisms are required to protect cell from the E2F1 induced apoptosis. Another protein considered in
the model is MRS18-2, which is co-localized with EBNA-6 in nucleus

**G2/S transition** (shown in block L, figure 3.2) CDC2-cyclin B kinase is the key effector in G2/M transition, which is inactivated by Wee1 and Myt1 in G2 phase and is activated by CDC25C to enter M phase. During the G2/M checkpoint, there are two alternative mechanisms involved in inactivating CDC2-cyclin B complex depending on whether DNA damage is detected. First is the faster one that CHK kinases inactivate CDC25C by phosphorylation, which make it unable to activate CDC2. Second is the slower cascade, which requires p53 pathway. P53 is phosphorylated in response to DNA damage signal. The phosphorylation releases p53 from the inactivation of MDM2. Then, p53 target genes serve as CDC2-cyclin B inhibitor through different ways. For example, 14-3-3σ binds to phosphorylated CDC2-cyclin B and export it into cytoplasm where it cannot activate CDC2; GADD45 binds and dissociates the CDC-cyclin complex; and p21, together with SET, are able to inactivate CDK1 activity.

Evidence suggested that both slow and rapid cascades involved in G2/M arrest are regulated by EBNA. On the one hand, proteins of EBNA-3 family target the G2/M checkpoint by interfering with CHK2, which results in the promotion of G2/M transition [30, 70]. On the other hand, EBNA-1 interacts with SET, an CDC2-cyclin B modulator in cell cycle. By cooperating with p21, SET specifically inhibits the activity of CDC2-cyclin B but has no effects on other cylin-CDK complex. The specific inhibitory effect of SET resides in its acidic terminal, and the p21 binding domain is mapped to two distinct regions, that is, aa. 81-180 and 181-277 [71].

**Wnt pathway** (shown in block B, figure 3.2) Wnt pathway is known to play a role in cancer and early development. β-Catenin is the main intracellular effector. In the absence of Wnt signal, cytoplasmic β-Catenin is phosphorylated by a destruction complex which consists of CK1α, GSK-3β, Anxin and APC. The phosphorylated β-Catenin then is targeted by proteasomal for degradation. When Wnt activating signals bind frizzled (Fz) and LRP co-receptors, dishevelled (Dsh) is inactivated which further leads to the inactivation of GSK by recruiting GSK binding protein (GBP). The inactivation of GSK holds the GSK destruction complex on the inner side of cell membrane. As a consequence, cytoplasmic β-Catenin is stabilized and subsequently transported into the nucleus to activate the trans-activation function of Tcf/Lef by directly trans-activating or displacing repressors of Tcf/Lef. The mechanism of β-Catenin activating Tcf/Lef remains under debate [72]. The targets of Tcf/Lef include c-myc, cyclin D1 and c-Jun.
A possible role of EBNA-1 in the Wnt pathway is introduced by the finding that EBNA-1 interacts with CK2 and by the fact that CK2 is an essential actor in the Wnt pathway. Gao et al. proposed that CK2 phosphorylates β-Catenin at Thr393, leading to its stabilization and accumulation [73]. However, it is worth to know that EBNA-1 interacts with CK2 through its β-subunits, while Catenin phosphorylation is mediated by its α-subunit.

**TGF-β pathway** (shown in block C, figure 3.2) The TGF-β pathway is involved in several important cell functions such as proliferation, differentiation, apoptosis and cellular homeostasis. The main intracellular partners of TGF receptor are R-Smads, which transfer signal from cell membrane into nucleus. In the model under discussion, only few target genes are considered. For example, in response to the TGF-β signal, Smad complexes combine with co-activators, such as p300 and Sp1, to induce the expression of p15, p16, p21 and p27. In parallel with trans-activation ability, it performs trans-repression function through cooperating with co-repressors. For example, a co-repressor complex containing Smad3, E2F4/5, p107 and DP1 exists in cytoplasm in general condition. In response to TGF-β signal, it is translocated into nucleus in where it binds to Smad4 to repress the expression of c-myc. Since c-myc is able to repress the expression of p15, the inactivation of c-myc relieves the Myc-mediated repression of p15 [74].

**NF-κB pathway** (shown in block D, figure 3.2) One important function of IκB kinase/NF-κB (IKK/NF-κB) signaling pathways is promoting cell survival through trans-activation of anti-apoptotic genes. The mammalian NF-κB consists of five members: NF-κB1 (p105 and p50), NF-κB (p100 and p52), c-Rel, RelB and RelA (p65) [75]. In resting state, NF-κBs form dimmers, which are maintained in cytoplasm by interacting with IκB. The canonical activation of NF-κB is through IKK and IκB. As a result of activating signals, IKK complex is activated by signal activated receptors and then phosphorylates IκB, which reaction leads to the ubiquitination and degradation of IκB. As a consequence, the NF-κB is no longer held by IκB in cytoplasm and is transported into nucleus where it serves as a trans-activator to induce target genes, such as anti-cancer gene Bcl-2 [76].

Phillips et al. linked NF-κB with the E2F1 pathway, that is, E2F1 represses the NF-κB pathway through down-regulation of TRAF2, an upstream regulator of IKK, to potentiate the apoptosis [77]. This effect is also included in the model. It has long been known that LMP-1 is able to activate both canonical and non-canonical NF-κB pathway, whereas recent studies showed that EBNA-1 represses p65 NF-κB pathway by inhibiting the phosphorylation of IKK, probably through regulating the expression of a cellular
genes involved in this process [78].

**P53 pathway** (shown in block I, figure 3.2) P53 is one of most important tumor suppressor protein which takes part in processes both in nucleus and in cytoplasm. The tetramer form of p53 binds DNA in a sequence specific manner in nucleus. The DNA binding ability of p53 is mediated by its core DNA binding domain (aa. 94-292). Except this domain, there is also a sequence nonspecific DNA binding region residing in the C-terminal lysine and arginine rich sequence. Other domains in p53 peptide include trans-activation domain (aa. 1-70) and oligomerization region (aa. 320-360).

In nucleus, p53 is a transcription factor that modulates the expression of genes involved in cell cycle (such as cyclin B, GADD45, 14-3-3σ, etc.), apoptosis (Bcl-2, Puma, APAF-1, Bax, etc.) and other cell processes. In keeping with the critical role of p53, there is an elaborated regulation system of p53 in cell. MDM2, the most well studied p53 negative regulator, regulates p53 in at least two mechanisms. First, it physically associates with p53 so as to inhibit its trans-activation function. Second, p53 is ubiquitinated by MDM2, and the polyubiquitinated p53 will be degraded by proteasomes. MDM2 is trans-activated by p53, in turn, increased MDM2 inhibits the function of p53. The negative feedback loop plays an important role in p53 regulation. MDM2 is also regulated by other factors, such as p14ARF, which inhibits MDM2 function via forming complex with it, and MDMX/MDM4, a homolog of MDM2.

It is certain that EBNAs are involved in the regulation of p53 pathway. Kashuba et al. proved that EBNA-5, MDM2 and p53 forms complex which stabilizes p53 but represses its trans-activation function. In addition to MDM2, p14ARF regulates p53 pathway as well. It binds to MDM2 so that MDM2 is unable to degrade p53. Other regulators, such as MDMX, are also targeted by EBNAs, while functions of these interactions remain to be defined. hence, MDMX is not present in the model.

Apart from the trans-regulation role in nucleus, p53 is also engaged in the cytoplasm processes to mediate apoptosis. It has been shown that monoubiquitination of P53 stimulates the nuclear export of p53. In cytoplasm, p53 could be deubiquitylated by mitochondrial HAUSP/USP7, thus generating the apoptotically active non-ubiquitylated p53 [79]. In nucleus, function of HAUSP-p53 association is activating the trans-activation function of p53. The association takes place between the C-terminal regulatory region (residues 351-382) of p53 and N-terminal region (residues 53-208) of HAUSP [80]. In the model under dicussion, only the first role is included since the trans-regulatory role of
HAUSP is still not clear.

Cellular stress signals interrupt the inhibition of p53 by MDM2, allowing p53 to accumulate both in the nucleus and in the cytoplasm. In cytoplasm, p53 is sequestered by anti-apoptotic Bcl2 proteins such as Bcl-XL. In addition, PUMA, another target of nuclear p53, functions to disrupt the Bcl-XL-p53 interaction. The released p53 from this complex triggers MOMP and apoptosis through interacting with, for example, Bax [79].

**Figure 3.3** – Simplified network for modeling. In the top level, expression of 5 EBNAs (EBNA-1, EBNA-2, EBNA-3, EBNA-4, EBNA-6) are set as input. In the bottom, 4 cellular process (genome replication, G2/M transition, viral genome partion and apoptosis) are as output.

**Other pathway or effectors** The NK-mediated apoptosis is shown in block I, figure 3.2. In this pathway, SET, pp32 are multi-functional proteins interacting with EBNAs. In the natural killer cell mediated apoptosis, SET complex cleaved by GzmA, a protein released from the cytotoxic T cell. The cleaved complex further relieves NM23-H1, which induces the apoptosis via its DNase ability. In addition, PP32, together with its hommolg April and PHAPIII, promote the formation of apoptosome (cytochrome c and APAF-1), while prothymosin-α antagonize this process. The trans-regulatory functions of SET,
pp32 and prothymosin-α are suggested by the fact that all of them are present in the INHAT complex, which complex binds histone to regulate histone modification and gene transcription [81].

It has been discussed above that the repression of AhR results in the elevation of ROS in cell. Evidence showed that Hsp27 prevent cell from ROS induced polyglutamine toxicity, probably through activating glutatonce [82, 83]. So it is possible that EBNAs prevent ROS induced apoptosis through promoting the Hsp27’s anti-ROS ability. At the same time, Hsp27 represses apoptosis through two mechanisms. First, the oligomers of Hsp27 negatively regulate cytochrome c induced cell death [81]. Second, the phosphorylation of Hsp27 by MAPKAPK2/3 accelerates the accumulation of Hsp27 dimers rather than Hsp27 oligomers, the dimers subsequently disrupt the transduction of the Fas-DAXX-Ask1 pathway, which protects cell from apoptosis [83].

3.3 Translate Into Logical Model

The detailed graph shown in the previous part is simplified to implement in Boolean logic model. The results is shown in figure 3.3. The simplified network consists of 60 nodes and 100 links. EBNAs could be expressed or repressed in the model, the output layer contains four events, namely, viral genome replication, host genome replication, G2/M transition and apoptosis, as shown in figure 3.3.

Each node in the model is given an initial state, either 0 or 1. Then, the state is updated according to rules that is defined as logical functions of previous states of their governors in once time step. (see section 2). Different combination of EBNAs are examined to explore the dynamic of the model.

**EBNA-1 is essential for viral genome replication and partition** EBNA-1 is important for viral genome replication. Because the proteins related to genome replication, such as RPA, ORC, are only interact with EBNA-1, so viral genome replication is achieved as long as the presence of EBNA-1. Therefore, EBNA-1 is an important factor for viral replication and keeping the EBV transformed phenotype.

**Minimum set of EBNAs induce proliferation** EBNA-1, -3 and -6 promote cell proliferation as shown in the figure 3.4a. However, the dynamic of genome replication and G2/M transition exhibit oscillation patterns (1 denotes event happening, 0 denote event repressed). It seems that proliferation is able to initiate but in a low efficiency in
Figure 3.4 – 3.4a The dynamics in the condition that EBNA-1, EBNA-3 and EBNA-6 are expressed. Four biological processes are presented in the graph, that is, genome partition (left top), G2/M transition (right top), genome replication (left bottom) and apoptosis (right bottom). Value 1 corresponds to a process tending to occur, whereas 0 denotes process is inhibited. In the right top and left bottom of plot 3.4a, the dynamic of G2/M transition and genome replication exhibit oscillation patterns. Plot left top shows genome partition properly happened. While apoptosis is not repressed as shown in right bottom. 3.4b The dynamics in the condition that EBNA-1, EBNA-3, EBNA-5 and EBNA-6 are expressed. Compare with left plots, EBNA-5 is expressed which results in the vanish of oscillation in genome replication and G2/M transition.

**EBNA-5 promote proliferation** While EBNA-5 is present, the oscillation of proliferation (genome partition and G2/M transition) is vanished as shown in figure 3.4b. This result means the proliferation is promoted in the presence of EBNA-5. In order to find out which proteins are responsible for this change, differential state matrix is derived from subtraction of state matrix of EBNA-5 absent and present as shown in figure 3.5. By analyzing this state matrix, a feedback loop, which contributes to the oscillation in the presence of EBNA-1, -3, and -6, is extracted from the model as shown in figure 3.6.

In this feedback loop, EBNA play different roles to modulate its dynamic. For example, EBNA-1 is important to activate β-Catenin, probably through interacting with CK2. While EBNA-3 and EBNA-6 are mainly achieve their functions by affecting E2F1. EBNA-3 represses AhR to release E2F1. Similarly, EBNA-6 releases E2F1 by binding S18-2. Eventually, the feed back loop keeps oscillating when EBNA-1, -3 and -6 are the expressed. While in the condition of EBNA-5 expressed, EBNA-5 inhibits the function of p53 through associating with MDM2. Subsequently, Siah-1 is no longer to be activated.
Figure 3.5 – Heatmap derived from state matrix of EBNA-1, -3, -6 subtracts EBNA-1, -3, -5, -6 one. It is used to detect the influence of EBNA-5. Pink dots represent proteins highly expressed when EBNA-5 exist, while relative darker blue nodes mean repressed proteins. The first change related to cell cycle is MDM2 in this model.

Figure 3.6 – The feedback loop resulting in oscillation of cell cycle in figure 3.4a. Red nodes are EBNAs, yellow ones are EBNA-binding proteins, other proteins in the model is represented in blue. EBNA-1, -3, -5, -6 play distinct roles in the regulation of the feedback loop.

Finally, β-Catenin is always hyper-activated and oscillation is interrupted.

**EBNA-1, -2, -3, -4, -5 have anti-apoptosis effect** Most EBNAs also have anti-apoptosis effects. There are several pathways result in apoptosis in the model, including ROS, p73, killer Bcl-2, etc. In order to maintain the latency infection, all of apoptosis inducers have to be repressed. So there are more than one mechanisms involved in the apoptosis repressing. Different pathways are repressed by different EBNAs. The mimic of expression and deletion analysis is performed in this study to explore the role for each EBNA in apoptosis inhibition, results are shown in the table 3.2. The × means corresponding apoptotic contributor is inactivated. The √ corresponds to a process is activated. The grey cells in the table highlight results different from control groups, that is, column NO in expression analysis and column all in deletion analysis. The results
suggest that all other EBNAs except EBNA-6 have anti-apoptosis effects, so the loss of any of them would lead to the failure of transformation.

**Table 3.2** – Result for expression and deletion analysis of EBNA in apoptosis

<table>
<thead>
<tr>
<th>apoptosis inducer</th>
<th>apoptosis</th>
<th>expression analysis</th>
<th>deletion analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>No 1 2 3 5 6 all 1 2 5 3 6</td>
<td></td>
</tr>
<tr>
<td>ROS</td>
<td>×</td>
<td>× × × × ✓</td>
<td>✓ x x x ✓ x</td>
</tr>
<tr>
<td>p73</td>
<td>×</td>
<td>× × × x x</td>
<td>x x x ✓ x x</td>
</tr>
<tr>
<td>killer Bcl2</td>
<td>✓</td>
<td>✓ x × ✓ ✓</td>
<td>✓ x ✓ ✓ x x</td>
</tr>
<tr>
<td>cyt-c</td>
<td>×</td>
<td>× × × × ×</td>
<td>x x × ✓ x x</td>
</tr>
<tr>
<td>hsp70</td>
<td>x</td>
<td>x x x x x</td>
<td>x x x ✓ x x</td>
</tr>
<tr>
<td>NM23-H1</td>
<td>✓</td>
<td>✓ ✓ ✓ ✓ ✓</td>
<td>✓ ✓ x x x x</td>
</tr>
<tr>
<td>VDR</td>
<td>✓</td>
<td>✓ ✓ ✓ ✓ ✓</td>
<td>✓ ✓ x x ✓ x</td>
</tr>
<tr>
<td>apoptosis</td>
<td>✓</td>
<td>✓ ✓ ✓ ✓ ✓</td>
<td>✓ ✓ ✓ ✓ ✓ x</td>
</tr>
</tbody>
</table>
Chapter 4

Discussion

The trans-regulatory roles of EBNAs have been shown important for EBV latent infection and are studied for a long time. The trans-regulatory effects are established by EBNA-host interactions. In this study, I explored the possible roles of EBNAs in several pathways and qualified their effects by mathematical model.

4.1 Experimental Support and Prediction

The results of this study is supported by experimental results, for example, EBNA-5 promote cell proliferation. Also, it is have been suggested for a long time that EBNA-3 and EBNA-6 are essential for transformation [85]. In the model of this study, β-Catenin plays a central role in maintaining cell proliferation, its activation or repression directly results in the cell cycle promotion or inhibition. Although RBP-Jκ is able to activate genes essential for cell proliferation, its activity is repressed by EBNA-3 family. As a consequence, the β-Catenin undertakes the responsibility of activation of proliferation related genes, such as Cyclin D and c-myc.

In the model under discussion, EBNA modulating CK2 leads to the hyper-activation of β-Catenin. However, in the real case, it could be activated either in transcriptional level or protein level, this can be a subject for the further experiments.

4.2 Manual Curation Is Insufficient

People dreamed earlier that cancer would be conquered when we finish the human genomic project (HGP), since we supposed genes for cancer would be identified and corresponding treatments could be developed afterward. However, unfortunately, HGP brought out more questions than answers on account of that we underestimated the complexity of
biology. Now we have known that from gene to mRNA to protein and to protein interaction, each step is extensively regulated by various mechanisms. Hence, it is impossible to understand the signal regulation from a single resource.

In order to execute their diverse functions, proteins undergo different protein-protein interaction processes, either to form stable complexes, such as co-repressors, co-activators, or to be post-translational modified, such as phosphorylation, histone-acetylation and ubiquitination. One important feature for co-repressors, co-activators and post-translational modification (PTM) proteins is that they are involved in many pathways and their roles are different among them. As a consequence, the knowledge of involvement of such proteins is not enough to understand their contribution. Unfortunately, most current knowledge about these proteins is only whether they are involved or not involved in a process, or in some better case, their subsequent phenotype is observed.

At present, lots of identified EBNA-binding proteins belong to co-factors or PTM proteins, for example, pp32, p300/CBP, CHK2 (phosphorylation proteins), HDACs (histone-deacetylation protein) and CtBP, NCoR, mSIN3a (co-repressor). The multi-role phenomenon of EBNA-binding proteins makes it especially troublesome to understand the
influence of EBNAs merely by the means of PPI data. For example, p300 is involved in
several pathways including notch, p53, p73, cell cycle, TGF-β etc. CtBP also at least
takes part in RB, ATM, Notch and Wnt pathways.

Recently, the high-throughput technologies, which are capable of recognizing variable
variations in a global wide, provide a powerful tool to analyze EBNAs’ effects on global
scale. Most current efforts in this area are focusing on transcriptomic level, for exam-
ple, Hertle et al. (2009) identified 129 down-regulated and 167 up-regulated genes in
EBNA-3 deletion experiment by microarray technology [86]. EBNA-2, EBNA-4, EBNA-
6 target genes is also screened by microarray analysis [23, 87]. A higher level study, that
is PPI between human and viral proteins, is also conducted via high throughput yeast
two-hybridization and mass spectrometry [88, 53]. While the high throughput study on
the PTM level is lagged behind transcriptional study, we are still in a very early stage in
this area. Recently, Choudhary et al. improved the mass spectrometry-based acetylation
profiling and applied it into a global acetylation study [89]. By considering the fact that
some EBNAs interact with HDACs, we can expect this technology will benefit EBNA
study in the near future.

In conclusion, due to complexity of biological processes, integrating data from different
sources will play an increasingly important role in biological study in the future, EBV
study is not an exception as well.
Acknowledgements

I would like to express my gratitude to all those who gave me help in the last two years. I want to thank the Department of Computational Biology, CSC, KTH for giving me permission to this great master program. All classmates and teachers in RB35 provides me a pleasant atmosphere.

I am deeply indebted to my academic supervisor, associate professor Elena Kashuba in Karalinska Institute, who gave me the chance to do this project and carefully revised my thesis. Also, I want to thank my local supervisor, professor Erik Aurell, who read and improve my thesis and gave me great help in difficult time, and examiner, professor Anders Lansner, who approve and grade my thesis.

Especially, I would like to give my special thanks to my parents, thanks for them giving financial support in this two years.
# List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>A schematic representation describing B95-8 viral genome</td>
<td>3</td>
</tr>
<tr>
<td>1.2</td>
<td>Comparison of random network and scale-free network</td>
<td>8</td>
</tr>
<tr>
<td>1.3</td>
<td>The global pathways regulatory circuit</td>
<td>10</td>
</tr>
<tr>
<td>1.4</td>
<td>Comparison of Kitano’s and faeder’s graphical representation</td>
<td>13</td>
</tr>
<tr>
<td>2.1</td>
<td>Illustration for pathway selection</td>
<td>16</td>
</tr>
<tr>
<td>2.2</td>
<td>The legend for Kitano’s graphical representation</td>
<td>17</td>
</tr>
<tr>
<td>3.1</td>
<td>Interactome analysis of EBNAs, D1 and D2 proteins</td>
<td>20</td>
</tr>
<tr>
<td>3.2</td>
<td>A global map of pathways influenced by EBNA</td>
<td>22</td>
</tr>
<tr>
<td>3.3</td>
<td>Simplified network for modeling</td>
<td>23</td>
</tr>
<tr>
<td>3.4</td>
<td>The dynamics in the condition that EBNA-1, EBNA-3 and EBNA-6 are expressed</td>
<td>30</td>
</tr>
<tr>
<td>3.5</td>
<td>Heatmap</td>
<td>31</td>
</tr>
<tr>
<td>3.6</td>
<td>The feedback loop resulting in oscillation of cell cycle</td>
<td>31</td>
</tr>
<tr>
<td>4.1</td>
<td>Signal regulation in different levels</td>
<td>34</td>
</tr>
</tbody>
</table>
List of Tables

1.1 Semantics of negation, conjunction, disjunction, implication and equivalence 14

3.1 EBNAs and their validated binding proteins . . . . . . . . . . . . . . . . 18
3.2 Result for expression and deletion analysis of EBNA in apoptosis . . . . 32
Bibliography


atm/atr signaling effector chk2 is targeted by epstein-barr virus nuclear antigen 3c
to release the g2/m cell cycle block,” *Journal of Virology*, vol. 81, no. 12, pp. 6718–

atm/chk2-mediated dna damage-responsive signaling pathway suppresses epstein-
barr virus transformation of primary human b cells,” *Cell Host and Mirobe*, vol. 8,

[33] E. Kashuba, M. Yurchenko, S. P. Yenamandra, B. Snopok, M. Isaguliants, L. Szekely,
and G. Klein, “Elv-encoded ebna-6 binds and targets mrs18-2 to the nucleus, result-
ing in the disruption of prb-e2f1 complexes,” *Proceedings of the National Academy

[34] M. A. Cotter and E. S. Robertson, “Modulation of histone acetyltransferase activity
through interaction of epstein-barr nuclear antigen 3c with prothymosin alpha,”

[35] E. V. Kashuba, K. Gradin, M. Isaguliants, L. Szekely, L. Poellinger, G. Klein, and
A. Kazlauskas, “Regulation of transactivation function of the aryl hydrocarbon re-
ceptor by the epstein-barr virus-encoded ebna-3 protein,” *The Journal of Biological

G. Klein, and E. Kashuba, “Epstein-barr virus encoded ebna-3 binds to vitamin d
receptor and blocks activation of its target genes,” *Cellular and Molecular Life


43


