UPPSALA
UNIVERSITET

Expression of Spleen tyrosine kinase (Syk) in epithelial tumor cell lines – a role in tumor progression

Zuobai Wang

Degree project in biology, Master of science (2 years), 2010
Examensarbete i biologi 45 hp till masterexamen, 2010
Biology Education Centre, Uppsala University, and Department of Microbiology, Tumor and Cell biology at Karolinska Institute
Supervisors: Ingemar Ernberg and Fu Chen

Summary

Epstein-Barr virus (EBV) is a gamma-herpes virus that was discovered in 1964 by Sir Anthony Epstein and his colleagues. EBV infects almost 90% of the human population, resulting in a lifelong coexistence. The virus is also associated with a wide variety of lymphoid and epithelial malignancies, which seem to be rare, adverse consequences of latent virus infections. In this latent infection, viral proteins such as LMP1 (latent membrane protein 1) and LMP2A (latent membrane protein 2A) can be expressed, and they both confer some tumor-related properties.

In B cells, EBV viral protein LMP2A is involved in controlling latent infection. It blocks B cell receptor (BCR)-mediated signaling by carrying some protein binding motifs of the BCR and also provides a weak survival signal for the B cells. These motifs are binding sites for some protein tyrosine kinase, e.g. spleen tyrosine kinase (Syk).

There are two isoforms of Syk in human proteome – one is the full length isoform (referred to as Syk (L)) with 635 amino acid residues while the other is its 23 amino acid residues shorter isoform (referred to as Syk(S)). Syk is found to act as a potential tumor suppressor in epithelial carcinoma cells such as breast carcinoma. Although Syk-mediated suppression of tumorigenesis is well accepted, the detailed suppressing mechanisms are still under study. It is, then, worthy to explore which isoform plays the key role in inhibiting the invasion of epithelial carcinoma cell line and study the isoform expression pattern in different cells as well as the factor regulating the isoform expression.

To explore the isoform expression in several selected cell lines, Western blot and reverse transcriptase PCR (RT-PCR) were employed. Based on Western blot, Syk from the transfected DNA was slightly shorter in size than the full length endogenous isoform. RT-PCR analyses indicated that Syk expression in HeLa cells was at an undetectable level; Syk (L) was the only isoform in Raji cells; while both isoforms of Syk were detected but at different expression levels in 5637 cells (bladder tumor cells) and Rko cells (Bukitt's lymphoma cells).

In order to study the regulation of Syk expression at the transcription level, bioinformatics was used. Results indicated that there was one CpG island in the promoter region of *Syk* gene. Methylation at the cytosine residues in CpG islands, inhibiting gene expression, is a profound epigenetic way. Therefore, epigenetic modification could regulate the *Syk* expression. The PROMO version 3.0, a search tool using a database called TRANSFAC v 8.3, revealed that 21 putative transcription factors might bind at the promoter region.

In LMP2A positive epithelial cells, Syk is known to interact with the LMP2A protein. Therefore, co-immunoprecipitation was initiated to further confirm the interaction between Syk and LMP2A in epithelial cells. However, the results are not yet fully conclusive, and more experiments are needed.

Contents

1. Introduction

1.1. Epstein-Barr virus and its contribution to tumorigenesis

Epstein-Barr virus (EBV), also known as human herpes virus 4 (HHV-4), is a human specific gamma-herpes virus that permanently colonizes nearly 90% of the human population (Thompson and Kurzrock, 2004). It was discovered by Epstein and his colleagues in 1964 (Bornkamm, 2009; Epstein et al., 1964). Its genome is about 170 kb, double stranded DNA and it is maintained in the cells as circular episomes (Adams and Lindahl, 1975). A harmless life-long coexistence with EBV is found in the majority of infected individuals, indicating a balance between the virus and host immune system. The virus is strongly believed to be associated with various tumours such as B-cell malignancies (Burkitt's lymphoma, immunoblastic lymphoma, and Hodgkin's disease (HD)), epithelial cell malignancies (nasopharyngeal carcinoma (NPC) and gastric adenocarcinomas, breast carcinoma), AIDS-related lymphoma and so on (Thompson and Kurzrock, 2004).

After infecting the host, EBV has three main phases: (1) episomal state; (2) latent state; and (3) lytic state (Bornkamm and Hammerschmidt, 2001). In the latent state, there are some latency phases – Latency 0, Latency Ⅰa, Latency Ⅰ, Latency Ⅱ, Latency Ⅱb and Latency Ⅲ. The latent infection is established in the resting memory B lymphocytes. EBV encodes six nuclear proteins (Epstein-Barr virus encoded nuclear antigens (EBNA) 1-6) and three kinds of membrane-associated protein (LMP) (LMP1, LMP2A and LMP2B); in addition, EBV also expresses several noncoding RNAs, e.g. EBV-encoded RNA1 (EBER1), EBER2, BARF1, etc. (Bornkamm, 2009). In different latency phases, different combinations of proteins and non-encoding RNAs are expressed (Table 1).

During a viral life cycle of infection, persistence and replication, viral gene products frequently recruit and interact with host cell proteins to disrupt the normal cellular signalling pathway by mimicking certain growth factors, transcription factors, and antiapoptotic factors (Thompson and Kurzrock, 2004).

Table 1. Gene expression from latent Epstein-Barr virus. (Adapted from Ingemar Ernberg, unpublished data, with permission)

Latency phase	Genes expressed	Cell types or tumors
Latency 0	EBER1& 2	Blood of healthy EBV carrier
Latency Ⅰa	EBER1&2, LMP2A	Blood of healthy EBV carrier
Latency Ⅰ	EBER1&2, EBNA1	Burkitt's lymphoma
Latency Ⅱ	EBER1&2, EBNA1, LMP1, LMP2a&2b, BARF0	Nasopharyngeal carcinoma, Hodgkin's disease, T-cell lymphoma
Latency Ⅱb	EBER1&2, EBNA1&2, BARF0	Gastric carcinoma
Latency III	EBER1&2, EBNA1-6, LMP1, LMP2a&2b	Immunoblastic lymphoma, post-transplant & AIDS-lymphoma, lymphoblastoid cell line (LCL)

1.2. Epstein-Barr viral protein latent membrane protein 2A

Both latent membrane protein 2A (LMP2A) and latent membrane protein 2B (LMP2B) are encoded by the LMP2 gene starting from two different promoters (Longnecker and Miller, 1996; Thompson and Kurzrock, 2004). LMP2A is 497 AA in length, including a 119 AA N-terminal intracellular domain (absent in LMP2B), 12 hydrophobic transmembrane domains of at least 16 AA, and a 27 AA C-terminal intracellular domain (Figure 1).

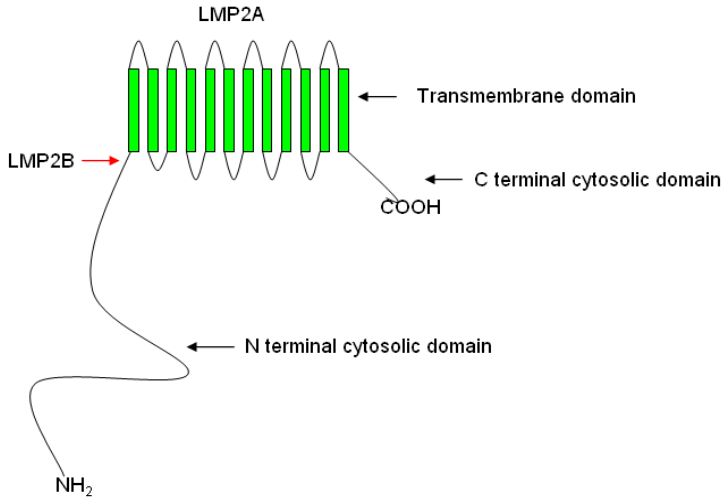


Figure 1. Predicted structure of LMP2A (Latent membrane protein 2A) and LMP2B (Latent membrane protein 2B) proteins. LMP2A contain three main parts – N terminal cytosolic domain, transmembrane domain, and C terminal cytosolic domain. The red arrow indicates the start site of LMP2B. N terminal part of LMP2B does not contain the cytosolic domain.

LMP2A is important for the maintenance of the latency in infected B lymphocytes (Longnecker and Miller, 1996; Miller et al., 1994). After stimulation by antigen, the B cell receptor (BCR) crosslinks with the immunoglobulin receptors - Igα and Igβ. Both of these immunoglobulin receptors contain the immunoreceptor tyrosine-based activation motif (ITAM) and become tyrosine-phosphorylated in the ITAM domain after antigen-stimulation and crosslinking with the BCR. The protein tyrosine kinases (PTK), such as spleen tyrosine kinase (Syk) and sarcoma tyrosine kinase (Src), bind to the phosphorylated immunoglobulin receptors via the interaction between their SH2 domains and the phosphorylated ITAM motif in the receptor. Then, these protein tyrosine kinases become phosphorylated according to autophosphorylation mechanism or by other protein tyrosine kinases. Phosphorylated PTKs are activated and could regulate downstream signalling pathways that contribute to lytic progression. The LMP2A has an ITAM motif in the N-terminal domain. In the LMP2A aggregates, the ITAM-like motif could be autophosphorylated. By constitutively mimicking the activated immunoglobulin receptors due to its phosphorylated ITAM-like motif, LMP2A aggregates compete with the immunoglobulin receptors for protein tyrosine kinases to block the reactivation of virus to enter the lytic phase, thus maintaining the latent phase (Figure 2).

The LMP2A mRNA is detected both in peripheral B cells of healthy EBV carriers and in patients with EBV-associated diseases, such as Burkitt's lymphoma, Hodgkin's disease, nasopharyngeal carcinoma, etc (Pang et al., 2009; Thompson and Kurzrock, 2004).

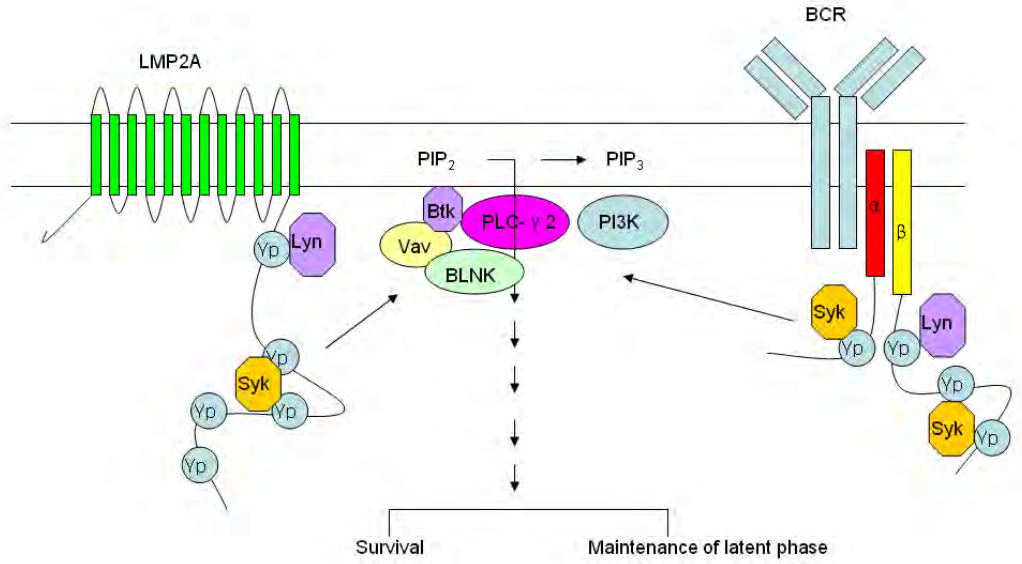


Figure 2. B cell Receptor (BCR) -related signal transduction and LMP2A-mediated interference of BCR-related signal transduction. Under stimulation by certain antigen, BCR crosslinks with the immunoglobulin receptor Igα and Igβ, both of which contain ITAM that becomes tyrosine-phosphorylated. The SH2-containing protein tyrosine kinases (PTK) such as Src and Syk, bind to the phosphorylated ITAM via interaction between the SH2 domain and phosphorylated ITAM, and become autophosphorylated or phosphorylated by other PTKs to induce the downstream pathways that contribute to the lytic phase progression. LMP2A has the ITAM motif in the N-terminal domain. After ITAM-phosphorylation, LMP2A aggregates compete with the BCR and its crosslinked Ig receptor for PTK. By doing this, LMP2A blocks the signalling transduction that leads to lytic progression and thus maintains the latent phase.

Although the function of LMP2A in Burkitt's lymphoma is well recognized, its role in oncogenesis in EBV-derived epithelial cell tumors such as nasopharyngeal carcinoma is still under study (Thompson and Kurzrock, 2004). Michael D. Allen and his colleagues in 2005 found that expression of LMP2A and LMP2B could enhance the capacity of squamous epithelial cell to spread and migrate on extracellular matrix (Allen et al., 2005), suggesting a role of LMP2A in controlling epithelial migration and invasion. In addition, LMP2A was also shown to affect epithelial cell growth and differentiation (Scholle et al., 2000). The possible signaling pathway employed by LMP2A to regulate epithelial cell migration and invasion was likely part of the PI3-kinase-Akt pathway, but not the mitogen-activated protein kinase (MAPK) pathway since no activation of the MAPK was seen (Scholle et al., 2000). E-cadherin, a marker of epithelial cells that is involved in attaching the epithelial cells to the extracellular matrix (ECM), was found to be down regulated in the LMP2A positive cells (Scholle et al., 2000); however, the RNA levels of integrin α6β4 regulating both cell adhesion to the ECM and metastasis (Wilhelmsen et al., 2006) were up-regulated in LMP2A expressing epithelial cells (Pegtel et al., 2005).

Interaction of Syk with LMP2A expressing epithelial cells has been recently studied. It was found that Syk could be involved in LMP2A-mediated regulation of epithelial cell migration, contributing to migration and invasion of tumors (Lu et al., 2006).

1.3. Spleen tyrosine kinase

Spleen tyrosine kinase (Syk) is a nonreceptor cytosolic protein tyrosine kinase (Sada et al., 2001) that is widely expressed in hematopoietic cells (Duta et al., 2006) as well as in many non-hematopoietic cells such as epithelial cells, e.g. low invasive breast carcinoma cell (Coopman et al., 2000) and endothelial cells, e.g. human umbilical vein endothelial cells (HUVECs) (Inatome et al., 2001; Yanagi et al., 2001).

Based on NCBI Gene database, the *Syk* locus is in chromosome 9, q22.2, starting from 93564012 bp to 93660832 bp with a length of about 96 kb (Figure 3). There are four different RNA transcripts – variant 1, variant 2, variant 3 and variant 4. Variant 1 and variant 2 encode full length Syk isoform (Syk(L), sometimes also referred to as Syk(A)) and short isoform of Syk (Syk(S) or Syk(B)), respectively. The translation start site resides in exon 2 at 93606181 bp. The difference between the two encoding RNA transcripts is the inclusion or skipping of exon 7, which is 69 bp in length (Figure 3). In other words, an alternative splicing mechanism likely plays a key role in producing two different isoforms of Syk protein. On the contrary, both variant 3 and variant 4 are noncoding RNAs that do not produce functional proteins.

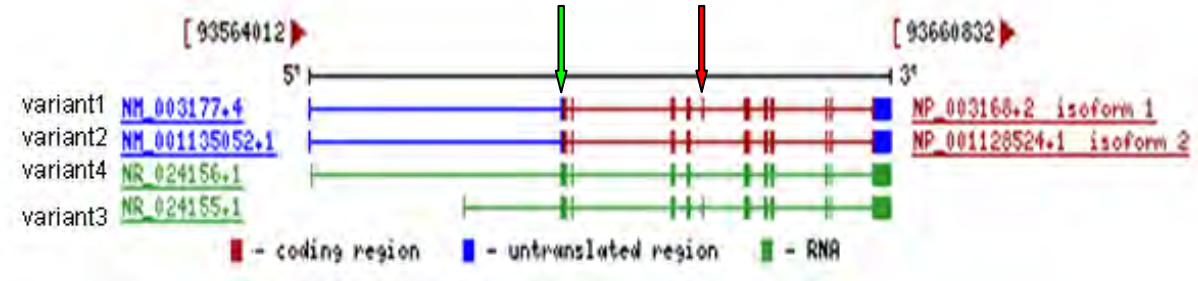


Figure 3. The *Syk* gene and its corresponding RNA transcripts. The Syk gene extends between 93564012and 936660832. Of the four RNA transcript, variants 1 and 2 are transcribed while the other two are not. The exon 7 in variant 1 was absent in variant 2. The bars indicate the exons, while the lines between two neighboring bars represent introns. The blue part in the two transcribed RNA transcripts represents the untranslated region, while the red part could be translated. (Maglott et al., 2007)(adapted from the NCBI website, public domain information on the National Library of Medicine (NLM) Web, http://www.ncbi.nlm.nih.gov/sites/entrez?Db=gene&Cmd=retrieve&dopt=full_report&list_uids=6850&log$=databasead&logdbfrom=nuccore). The green and red arrows point to the exon 2 and exon 7, respectively.

Both Syk isoforms contain N-terminal tandem Src Homology 2 (SH2) domains and C-terminal kinase domains (Figure 4) (Sada et al., 2001). The region linking the two SH2 domains is called interdomain A (IDA), while the region between the C-terminal SH2 domain and the C-terminal kinase domain is interdomain B (IDB). Syk (S) lacks a 23 AA part in IDB, which is called the DEL sequence and encoded by exon 7, as compared with Syk (L). The *Homo sapiens* DEL sequence is TWSAGGIISRIKSYSFPKPGHRK, containing 5 basic amino

acid residues (Wang et al., 2003). Among these 21 amino acid residues, there are 5 basic residues that together serve as nuclear localization signal (NLS) (Wang et al., 2005; Wang et al., 2003). Thus, in addition to existence in cytoplasm, this NLS sequence enables Syk (L) to enter the nucleus; lacking this sequence, the Syk (S) only resides in the cytoplasm.

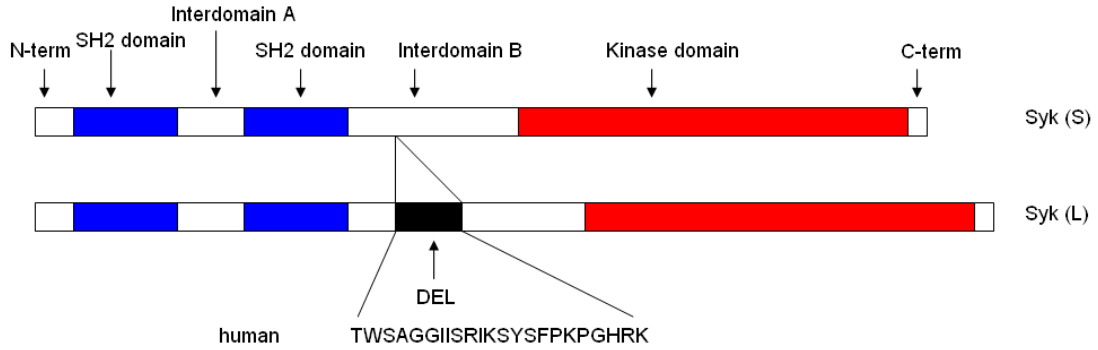


Figure 4. Schematic structure of Syk (L) and Syk (S). Both Syk (L) and Syk (S) contain the two SH2 domains at the N-termini while the catalytic kinase domain (red parts) is located in the C-termini. Interdomain A (IDA) connects the two SH2 domains (blue parts) a while interdomain B (IDB) links the C-terminal SH2 domain to the Kinase domain. In IDB, a 23 AA-long region (DEL sequence) is missing in Syk (S) as compared with the Syk (L). The sequence shown in this picture represents the DEL sequence of Syk, present only in Syk (L).

Syk has been widely studied in hematopoietic cells where it is essential for immunoreceptor signaling (Turner et al., 2000). Such a tyrosine kinase is also very critical for endothelial cell proliferation and migration since in human umbilical vein endothelial cells, lack of expression of *Syk* impaired both cell proliferation and migration (Inatome et al., 2001).

Although protein tyrosine kinases often have been described as oncoproteins, a tumor suppressor role of Syk was first found in breast carcinoma (Coopman et al., 2000; Stewart and Pietenpol, 2001). Moreover, more and more studies indicated Syk could undertake the role of a tumor suppressor in many other tumors, e.g. gastric cancer, hematopoietic cancer, etc. (Coopman and Mueller, 2006). However, although a lot of evidence supports Syk's role as a tumor suppressor, this role as a potential tumor suppressor is still under study. Recently, some paper argued that overexpression of Syk was found in some tumors, e.g. squamous cell carcinoma of the head and neck, peripheral T-cell lymphoma, etc., indicating that Syk may hold a completely different role in cancer – oncoprotein (Feldman et al., 2008; Luangdilok et al., 2007).

Coopman and his colleagues found that Syk was involved in inhibiting malignant invasion (Coopman et al., 2000). Meanwhile, Jia Le Dai's group pointed out that only the full length Syk – Syk (L) could translocate to nucleus and subsequently act as a transcriptional co-repressor for certain oncogenes (Wang et al., 2005).

Epigenetic methylation is known to be a crucial mechanism regulating gene expression. Epigenetic DNA methylation is closely related to development (Reik, 2007) and tumorigenesis (Jones and Baylin, 2002). DNA methylation occurs only at cytosine bases located 5' to a guanosine in a CpG dinucleotide (Jones and Baylin, 2002). These dinucleotides are distributed over the whole genome, but certain regions of 0.5-4kb are rich in condensed CpG sites, and therefore known as CpG islands (Jones and Baylin, 2002). Mostly, these CpG islands reside in the proximal promoter region of certain genes, containing 5' flanking DNA, exons and introns (Gardiner-Garden and Frommer, 1987). DNA methyltransferases, mostly using S-adenosyl-methionine as methyl donor, is responsible for catalyzing methyl group transferring to cytosine (Jones and Baylin, 2002). After being methylated at certain CpG sites in the CpG island near the promoter region, the expression of the corresponding gene is blocked. The inhibition mechanisms are: (1) these methylated CpG sites could either recruit certain methyl-CpG binding corepressor; and (2) after being methylated, CpG sites are not suitable for transcription factor binding any more (Jones and Baylin, 2002). Epigenetic change – hypermethylation at the Syk promoter region has been found to regulate Syk expression (Yuan et al., 2001; Yuan et al., 2006). The protein arginine N-methyltransferase 6 (PRMT6) plays a role in alternative splicing of exon 7 of Syk RNA transcript since downregulation or inactivation of PRMT6 enhanced the level of Syk (S) and thus decreased Syk(L):Syk(S) ratio (Harrison et al., 2010). Thus, alternative splicing may also regulate function of Syk. Recently, P53, a transcription factor, is found to regulate Syk expression at the transcriptional level (Vrba et al., 2008).

1.4. Computational prediction of transcription factor binding sites and CpG island

The regulation of gene transcription is critical for tissue specific expression, development, and responding gene expression under specific stimuli, such as hypoxia (Latchman, 1997). At specific developmental stages or after stimulation by certain factors, each type of cell or tissue expresses certain specific transcription factors (TF) (Kel et al., 2003). These transcription factors bind to their corresponding sequence motifs in the regulatory region of DNA (e.g. promoter region, enhancer region) (Kel et al., 2003; Latchman, 1997).

Currently, experimental techniques, such as chromatin immunoprecipitation, are widely used for searching for transcription factors at a given region. However, compared with experimental methods that find the transcription factors at the regulatory regions, computational methods seem to be cheaper, faster, and require less resources (facilities, well trained labours) although their drawback in the field of accuracy is obvious (Ben-Gal et al., 2005). Thus, computation based prediction of transcription factor binding sites are more and more used in order to guide the experimental techniques-based searching and thus save both resources and time. All the binding motifs for certain transcription factor have been experimentally verified, collected, and aligned to build a positional weight matrix (PWM) (Gershenzon et al., 2005). The PWM is widely used to search for putative binding sites of certain transcription factors at a given region (Ben-Gal et al., 2005; Gershenzon et al., 2005). The basic assumption of the PWM model is that the nucleotides at each position are independent in statistical view and therefore the joint probability of finding a multiple-position site factorizes into the product of single-position probabilities (Ben-Gal et

al., 2005; Djordjevic et al., 2003). The frequency of each nucleotide (A, T, C and G) at each column of the alignment is calculated in the matrix. After being constructed, the PWM can be used to search for subsequences from a given sequence. The subsequences selected according to PWM are functionally similar to those used to build the PWM, meaning they are putative binding sites for transcription factors (Stormo, 2000). The match between a subsequence and a PWM is calculated by a set score, which measures the similarity of the subsequence to the PWM.

The TRANSFAC database is the largest available collection of eukaryotic transcription factors, their binding sites and the nucleotide distribution matrices as well as their regulated genes (Matys et al., 2006; Wingender et al., 1996). Because it contains the largest amount of accessible information on transcription factors, especially a lot of nucleotide distribution matrices, this database is widely and commonly used to build PWM and consequently used to search for putative transcription factors and their binding sites at a given region. Version 8.3 of TRANSFAC database contains entries for 5711 transcription factors, 3451 of which are vertebrate, including 1357 from *Homo sapiens*, and entries for 14406 binding sites, 4485 of which are vertebrate, including 1907 from *Homo sapiens* (Matys et al. 2006).

PROMO version 3.0 is a search tool for transcription factors and their binding sites in DNA sequences from a species or groups of species of interest, written in C^{++} (Farre et al., 2003; Messeguer et al., 2002). PROMO version 3.0 employs TRANSFAC (either version 6.4 or version 8.3) as transcription factor binding sites database to build the position weight matrices (Farre et al., 2003; Messeguer et al., 2002). An attractive feature of PROMO v3.0 is its taxonomic searching for species-specific transcription factors and their corresponding binding sites in the selected species, e.g. *Homo sapiens* (Farre et al., 2003; Messeguer et al., 2002). Therefore, false positives may be largely reduced by selecting certain species.

To find the CpG islands, computational prediction is one possible approach in addition to footprinting and other experimental methods. CpG island searcher with latest version: 10/29/04 (Takai and Jones, 2002) and EMBOSS CpGPlot (Mullan and Bleasby, 2002) are two easy and convenient tools for computationally predicting CpG sites. The CpG island searcher hunts for CpG islands using the algorithm of Takai and Jones' (Takai and Jones, 2002). The function of the cpgplot is to plot CpG rich areas and cpgreport to report all CpG rich regions, while isochore plots GC content over a sequence. Both tools provide different parameters for CpG island prediction, such as G+C content, length of the sequence (length) and frequency of occurrence of CpG dinucleotides (ObsCpG/ExpCpG).

The definition of CpG island was firstly described in 1987 by Gardiner- Garden and Frommer (Gardiner-Garden and Frommer, 1987). To define this, several parameters and factors were made. They introduced the percentage of G+C content (%GC), frequency of occurrence of CpG dinucleotides (ObsCpG/ExpCpG), and length of the GC-rich region. The ratio observed/expected CpG (ObsCpG/ExpCpG) was calculated as follows:

$$\text{ObsCpG/ExpCpG} = \frac{\text{Number of CpG}}{\text{Number of C} \times \text{Number of G}} \times N ,$$

where N represents the total number of nucleotides in the sequence being analyzed (Gardiner-Garden and Frommer, 1987). The observed value of CpG sites (ObsCpG) was based on the exact number of the CpG sites observed in the selected region, while the expected value of CpG sites (ExpCpG) was based on the G+C content and random occurrence of nucleotides in the same region. In their study, they set a 100-base window (N=100). This window shifted 1 bp downstream after calculating both %GC and ObsCpG/ExpCpG until certain criteria were met. For their study, they defined the CpG-rich region, namely CpG island, must contain percentage of G+C content above 50%, hold the ratio observed/expected CpG over 60%, and have a length larger than 200 bp. From then on, these criteria have been widely accepted when searching the CpG islands. These criteria have been experimentally plausible and are still used. In this study, these three main parameters were set as theirs.

1.5. Aims

As part of an ongoing exploration of the role of Syk in tumor cell migration, and a possible role for the EBV protein LMP2A in this migration, I aimed to address the following three issues: (1) to study the expression pattern of the two isoforms of Syk in epithelia cell lines; (2) by using bioinformatic approach to build foundations for studies of transcriptional regulation of Syk expression, including possible epigenetic mechanisms; and (3) to explore the physical interaction between EBV- LMP2A and Syk protein by using co-immunoprecipitation.

2. Results

2.1 Expression of spleen tyrosine kinase isoforms at the protein level in some different cell lines

Syk expression was studied in different cell lines for subsequent investigation of functions of Syk isoforms. There are two isoforms of Syk in human proteome – Syk (L) around 72kDa and Syk (S) about 68 kDa (Duta et al., 2006). Recent evidence indicated that Syk participates in the suppression of tumorigenesis of epithelial tumor, such as breast carcinoma (Coopman et al., 2000) and gastric carcinoma (Wang et al., 2004). However, it is still unclear which isoform of Syk executes the tumor suppressor function and which mechanism(s) it employs to suppress oncogenesis and metastasis.

Several epithelial cells and one B cell control were chosen for the study. The CNE1 cell line originates from nasopharyngeal carcinoma, Rko cells originate from bladder carcinoma, HeLa cells (used here as a Syk negative control, (Renedo et al., 2001) are from human cervical carcinoma, while the Raji cell line (used here as a Syk (L) positive control) (Wang et al., 2003) was derived from B cells (Burkitt's lymphoma cells). Syk cDNA was used as positive control and transiently transfected into CNE1 and HeLa, lysates from all cells were subjected to Western blot (Figure 5).

Syk from transfected cDNA migrated faster than the endogenous full length Syk in the SDS-PAGE, indicating *Syk* cDNA encoded a protein slightly shorter than Syk (L). After transfection with the same amount of *Syk* cDNA for the same time length, HeLa expressed higher level of Syk than CNE1. In Figure 5, Raji cells expressed quite high level of endogenous Syk, while Rko cells was also shown high Syk expression.

Raji cells only expressed full length Syk (Wang et al., 2003), so it serves as a Syk (L) positive but Syk (S) negative control. In HeLa cells, Syk expression was almost undetectable (Lauvrak et al., 2006; Renedo et al., 2001). Thus, HeLa was used as a negative control for both Syk (L) and Syk (S). Comparing Syk isoforms in HeLa and Raji with the other cell lines in Figure 5, it is hard to detect spliced Syk (S) isoform in any of the cells by WB since there was one band about 68 KDa even in HeLa cells without Syk (L) or Syk (S) expression. This is perhaps because of the interruptions of some unspecific bands below probably overlapping with the Syk (S) (at the place where the red arrow points in Figure 5). In Figure 5, some unspecific bands below endogenous Syk and exogenous Syk illustrated the importance of the quality of commercial antibodies. Several commercial sources of antibodies were tried and this was the best. The non-specific reactions might mask the short isoform of Syk. Therefore, reverse-transcriptase PCR (RT-PCR) was further used to explore possible expression of the two isoforms at the transcriptional level.

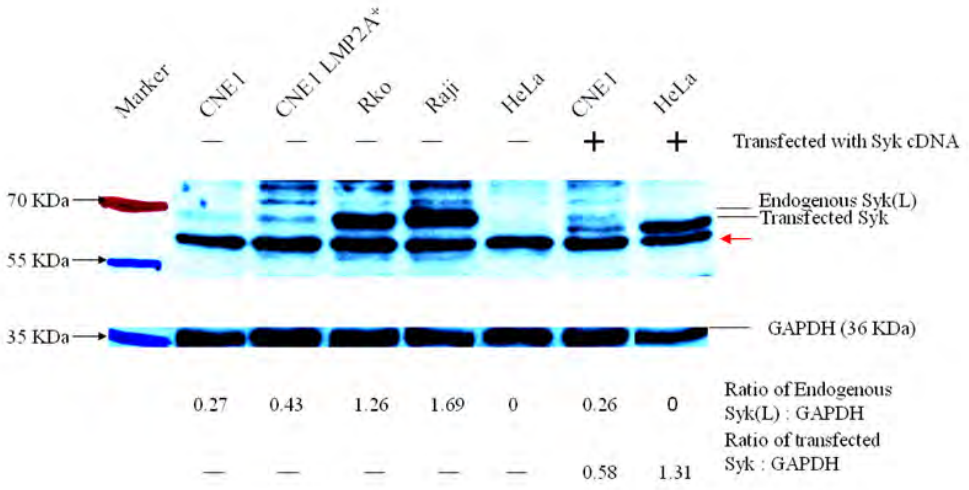


Figure 5. Western blot analyses of Syk expression portrait of Syk in different cell lines. CNE1 and HeLa cells were transiently transfected with *Syk* cDNA and then cultured for 32 hr. 2×10^5 cells of all the chosen cell lines CNE1, CNE1 LMP2A stable transfectant, HeLa, Rko, Raji, CNE1 Syk-transient transfectant as well as HeLa Syk-transient transfectant cells were separated by 9% SDS-polyacrylamide gel electrophoresis (SDS-PAGE), and transferred to nitrocellulose membrane and probed with rabbit anti-Syk polyclonal Ab N19 or mouse anti-GAPDH monoclonal 6C5 Ab. Secondary horse raddish peroxidase-conjugated antibodies allowed detection by enhanced chemiluminescent. Red arrow points to the place where Syk (S) would be on the PAGE.

2.2 Expression of spleen tyrosine kinase isoforms at the transcription level in several different cell lines

The short isoform of Syk RNA is 69 nt shorter than the full length counterpart due to skipping of Exon 7 (Harrison et al. , 2010). To study the Syk expression characteristics in different cell lines, reverse transcription PCR (RT-PCR) was carried out. The concentration of Magnesium ion was optimized using total RNA from Rko cells (Figure 6). The three higher concentrations all resulted in two bands – one at 392 bp representing Syk (L) while the other at 323 bp representing Syk (S). Thus 1.5 mM Mg^{2+} was chosen for subsequent study.

Next, total RNAs obtained from Raji, HeLa, 5637 and Rko cells were reverse transcribed respectively to the corresponding cDNA (Figure 7A and 7B). Both isoforms of Syk were detected but at different expression level in 5637 cells and Rko cells; Syk (L) was detected in CNE1 cells but Syk (S) was almost absent; the Syk (L) was the only isoform in Raji cells, identical to results from Wang L. and his colleagues (Wang et al., 2003); while there was very little or even no Syk expression in HeLa cells.

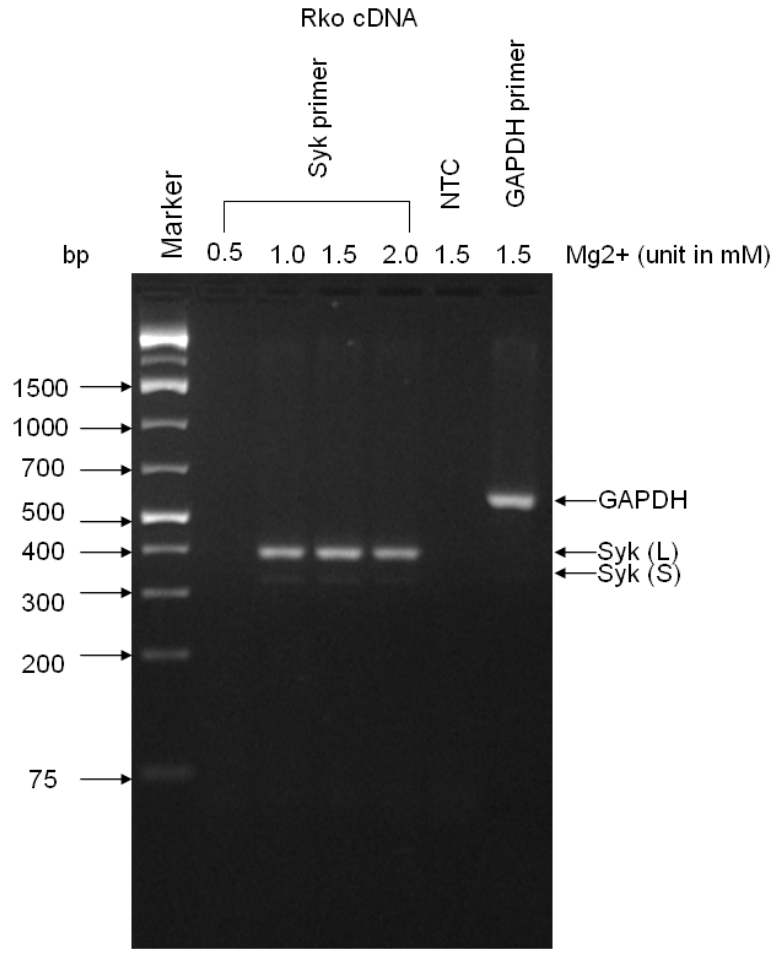


Figure 6. Condition optimization for concentration of Magnesium ion for RT-PCR. 3 μg of total RNA from Rko cells was used for reverse transcription to cDNA using primers for *Syk* at different magnesium ion concentrations was selected from 0.5 mM, to 1.0 mM, 1.5 mM and 2.0 mM. PCR product using GAPDH primers at 1.5 mM of magnesium ion was used as internal control. The no template control (NTC) contained ddH_2O instead of template. Samples were then analyzed by electrophoresis on a 2% agarose gel.

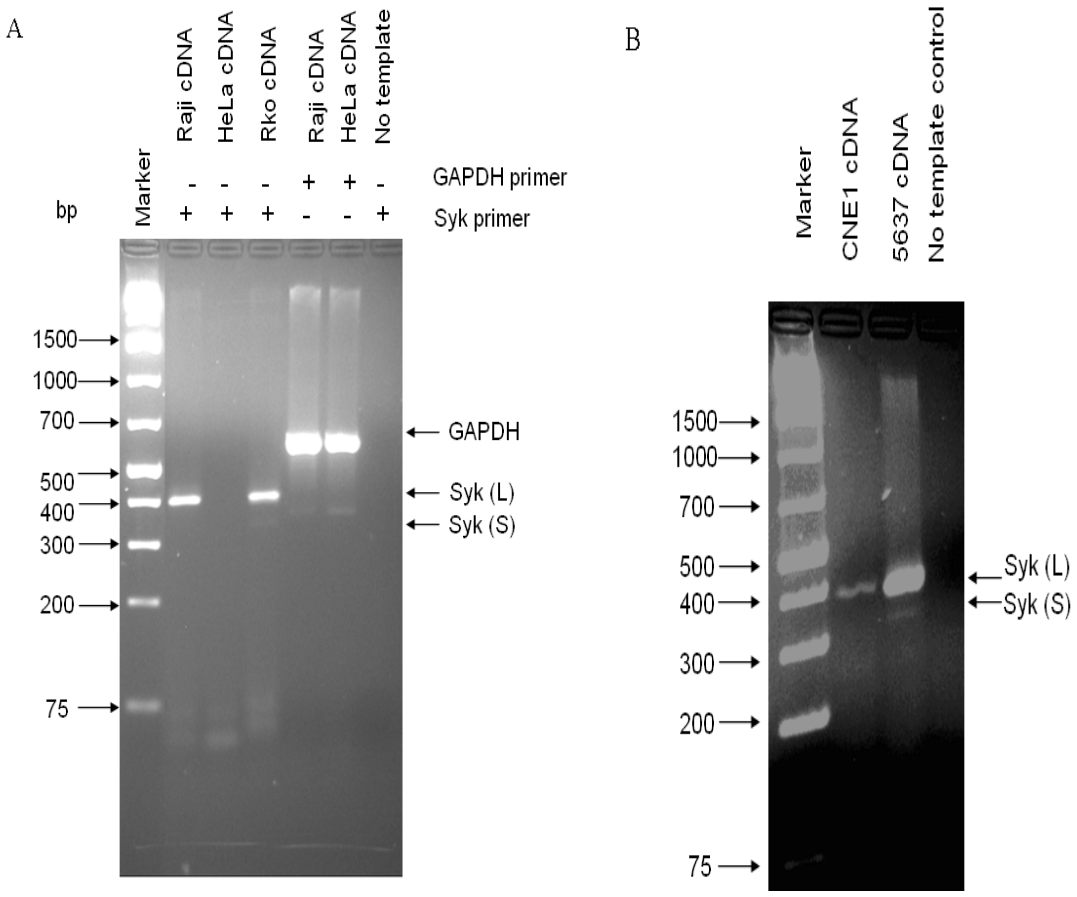


Figure 7. Analyses of Syk expression at the transcription level in different cell lines. 3 μg of each total RNA sample from Raji, HeLa, 5637, CNE1 and Rko cells was used for reverse transcription to cDNA using *Syk* primers and GAPDH primers. Products were analyzed by electrophoresis on a 2% agarose gel. (A) Raji cDNA, HeLa cDNA and Rko total RNA transcripts were used for Syk isoform analysis. PCR product using GAPDH primer and Raji cDNA or HeLa total RNA transcripts were used as internal control. The no template control (NTC) contained ddH_2O instead of template. (B) Products from CNE1 and 5637 total RNA transcripts were used for Syk isoform analysis. No template control was performed by using ddH_2O instead of templates.

2.3 Computational prediction for putative transcription factors as well as CpG islands in *spleen tyrosine kinase* promoter

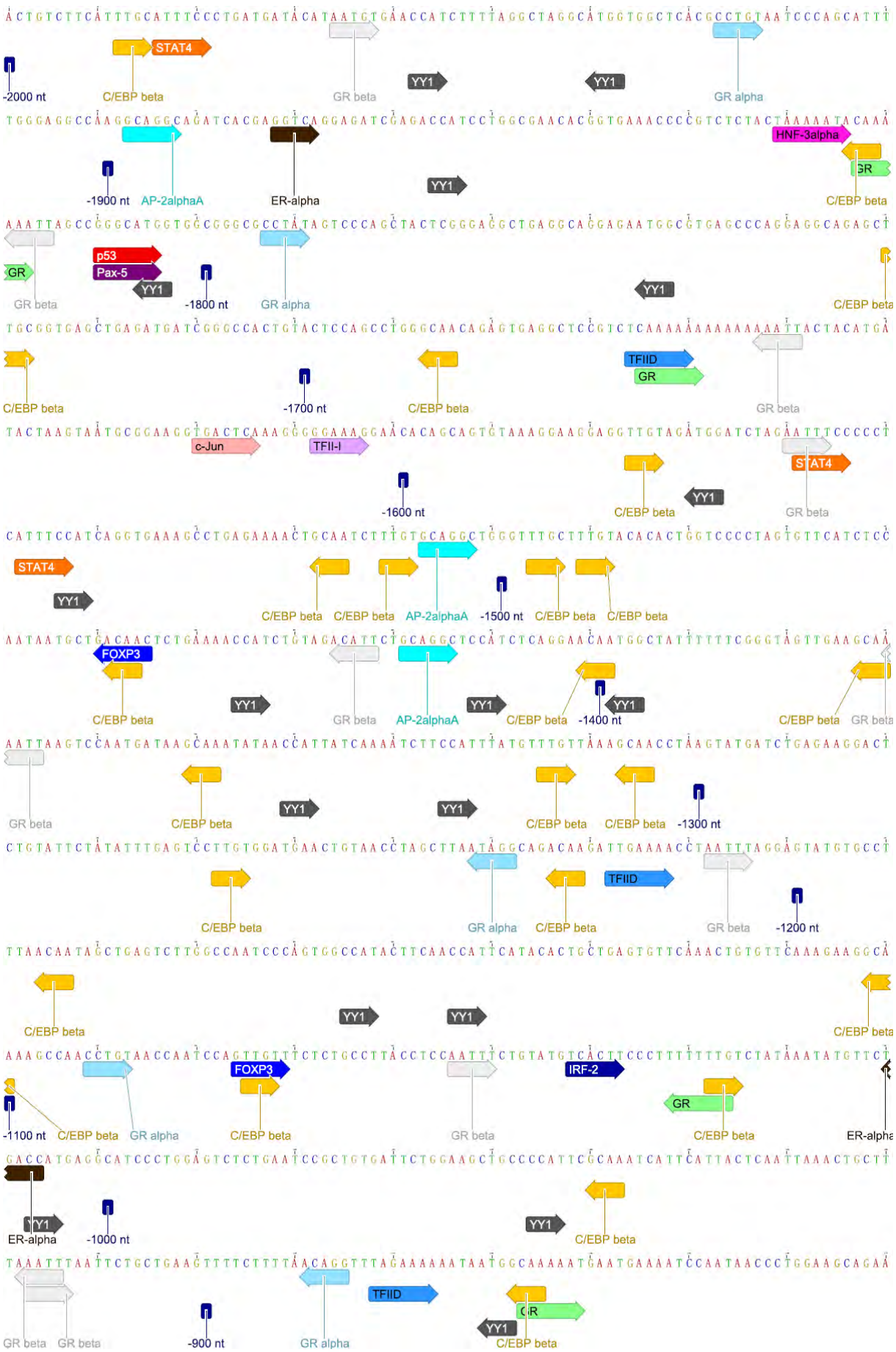
The promoter region of Syk was selected from the NCBI database. The Syk gene occupies from 93564012 nt to 93660832 nt (Maglott et al., 2007); http://www.ncbi.nlm.nih.gov/sites/entrez?Db=gene&Cmd=retrieve&dopt=full_report&list_uids=6850&log$=databasead&logdbfrom=nuccore) (Figure 3). A region from 93562012 nt to 93564211 nt was selected (corresponding from -2000 to +200 in the Syk promoter region) for bioinformatic analysis. This would most likely include the proximal promoter control region. With perfect matching (a dissimilarity rate of 0) to the positional weight matrix (PWM), PROMO v3.0 predicted 21 different human transcription factors for altogether 130 binding sites in such region (Figure 8). These transcription factors included C/EBPbeta, STAT4, GR-beta, YY1, AP-2alphaA, HNF-3alpha, GR, TFIID, FOXP3, GR-alpha, ER-alpha, TCF-4E, Pax-5, GCF, p53, ENKTF-1, Elk-1, TFII-I, XBP-1, IRF-2 and c-Jun (details in Figure 8).

A high threshold level of similarity could reduce the number of false positive predictions (Zheng et al., 2003). However, transcription factor binding sites rarely show perfect similarity to the consensus sequence. Therefore, 5% and 15% dissimilarities were allowed. As expected, the kinds and number of transcription factor putative binding sites increased considerably (Figure 9 and details in Appendix Table 7 and Table 8). At 5% dissimilarity, the number of different transcription factors was increased to 55 with 458 total binding sites. When 15% dissimilarity was allowed (a default value in PROMO v3.0), 83 different transcription factors with 1012 putative binding sites were predicted (Figure 9 and details in Appendix Table 7).

DNA methylation at CpG sites in a CpG island in the proximal promoter region of certain gene is critical for regulation of gene expression (Jones and Baylin, 2002). When methylated, the CpG sites may block gene transcription by either blocking the transcription factor binding sites or recruiting other proteins that compete with transcription factors for the same sites (Jones and Baylin, 2002). Therefore, CpG island prediction can be used to analyze whether such epigenetic modification on gene transcription is likely.

Syk gene expression could be regulated by DNA methylation (Yuan et al., 2001; Yuan et al., 2006). For experimental methylation study, Yuan and his colleges chose a region from -350 nt to +257 nt of Syk promoter, spanning from 350 nt upstream of the transcription start site (TSS), through 107 nt-long Exon 1, to the first 150 nt of Intron 1. I chose a region from -400 nt to +200 nt that spans part of the region upstream of TSS, through the Exon 1, to the first 83 nt of Intron 1. In this region, there are 57 CpG sites (Appendix Table 9, Figure 10). Both CpG island searcher and EMBOSS CpGPlot predicted that there was one CpG island in this region (Figure 11 and Figure 12). EMBOSS CpGPlot predicted that a region spanning -221 to +144 nt from the selected region is most likely one CpG island (Figure 12). From Figure 10, it can be seen that some CpG sites overlapped with some predicted transcription factor binding sites, e.g. for p53, when perfect matching.

p53 has been found recently to play a role in Syk regulation (Vrba et al., 2008; Xu and el-Gewely, 2001). Nevertheless, the exact binding sites for p53 and the mechanism behind its up-regulation of *Syk* expression are not known. With perfect matching, there was one putative p53 binding site from +24 nt to +30, which also overlapped with the cytosine residue in the CpG dinucleotides at site +30 (Figure 10). At 5% dissimilarity, 19 putative p53 binding sites were found, 13 out of which were located in the region from -400 nt to +200 nt (Figure 13 and Appendix Table 10). These 13 binding sites included 10 sites (about 76.9%) that overlapped with at least one CpG site (Figure 13 and Appendix Table 10). When the dissimilarity was set to 15%, the number of putative p53 binding sites increased to 38 (Figure 13 and Appendix Table 11), among which 27 sites (Figure 13 and Appendix Table 11) were in the selected region (from -400 nt to +200 nt). Out of these, 21 putative binding sites (Figure 13 and Appendix Table 11) overlapped with CpG sites. Taken together, CpG methylation in the proximal promoter region of the *Syk* gene might play a critical role in modulating p53-mediated *Syk* expression.


STAT4
C/EBP beta
GR beta
YY1
YY1
GR alpha
-2000 nt
HNF-3alpha
-1900 nt
AP-2alphaA
ER-alpha
YY1
C/EBP beta
p53
Pax-5
YY1
GR
GR beta
-1800 nt
GR alpha
YY1
C/EBP beta
C/EBP beta
-1700 nt
C/EBP beta
TFIID
GR
GR beta
c-Jun
TFII-I
STAT4
-1600 nt
C/EBP beta
YY1
GR beta
STAT4
YY1
C/EBP beta
C/EBP beta
AP-2alphaA
-1500 nt
C/EBP beta
C/EBP beta
FOXP3
C/EBP beta
YY1
GR beta
AP-2alphaA
YY1
C/EBP beta
-1400 nt
YY1
C/EBP beta
GR beta
GR beta
C/EBP beta
YY1
YY1
C/EBP beta
C/EBP beta
-1300 nt
C/EBP beta
GR alpha
C/EBP beta
TFIID
GR beta
-1200 nt
C/EBP beta
YY1
YY1
C/EBP beta
FOXP3
IRF-2
-1100 nt
C/EBP beta
GR alpha
C/EBP beta
GR beta
GR
C/EBP beta
ER-alpha
YY1
ER-alpha
-1000 nt
YY1
C/EBP beta
GR beta
GR beta
-900 nt
GR alpha
TFIID
GR
YY1
C/EBP beta

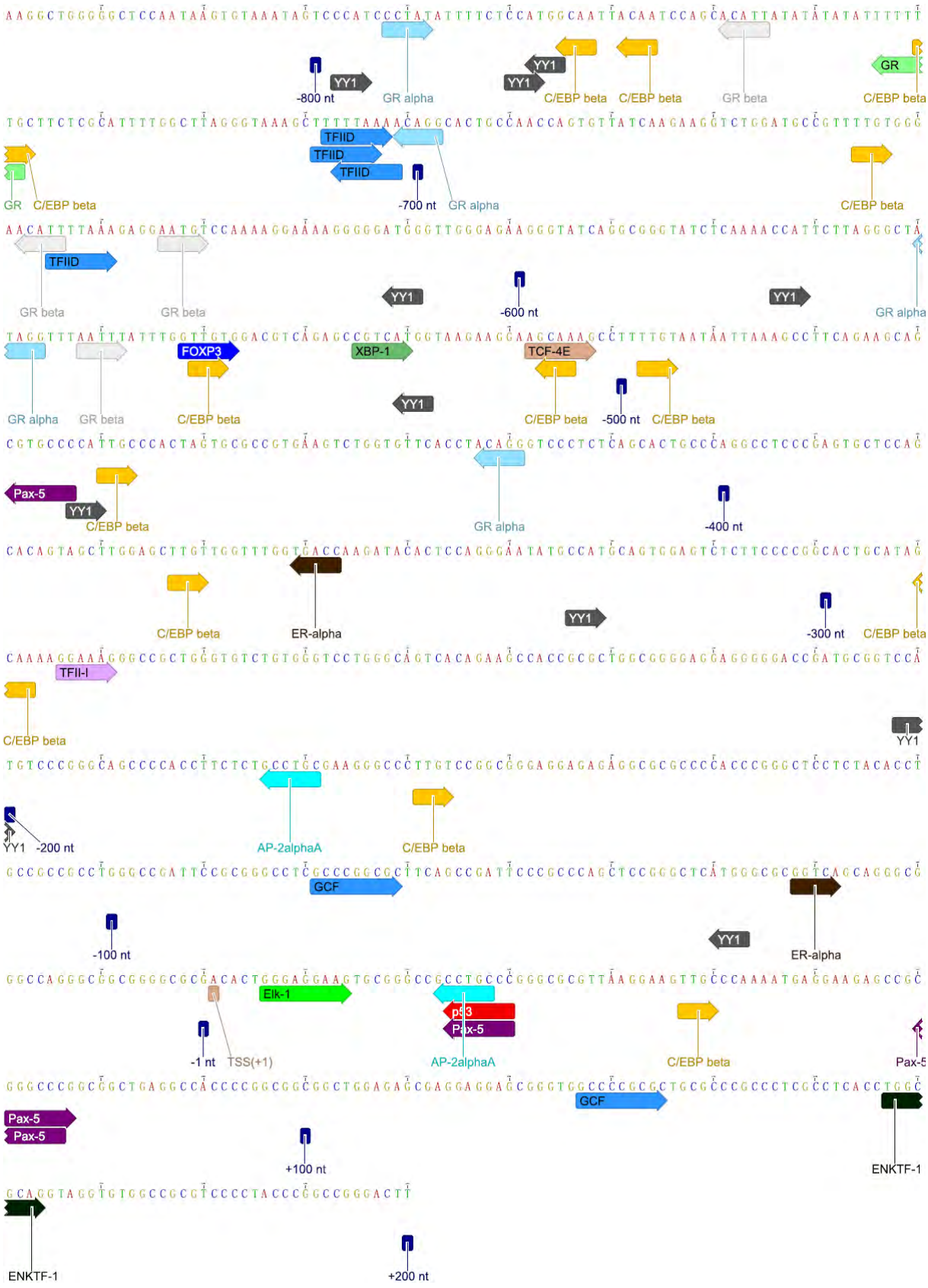


Figure 8. All the putative transcription factor binding sites in the *Syk* promoter from -2000 nt to +200 nt at 0 dissimilarity (perfect matching). TSS (+1) represents the transcription start site (represented by the grayish pink rectangle below A). The arrows with distinct colors represent different transcription factors.

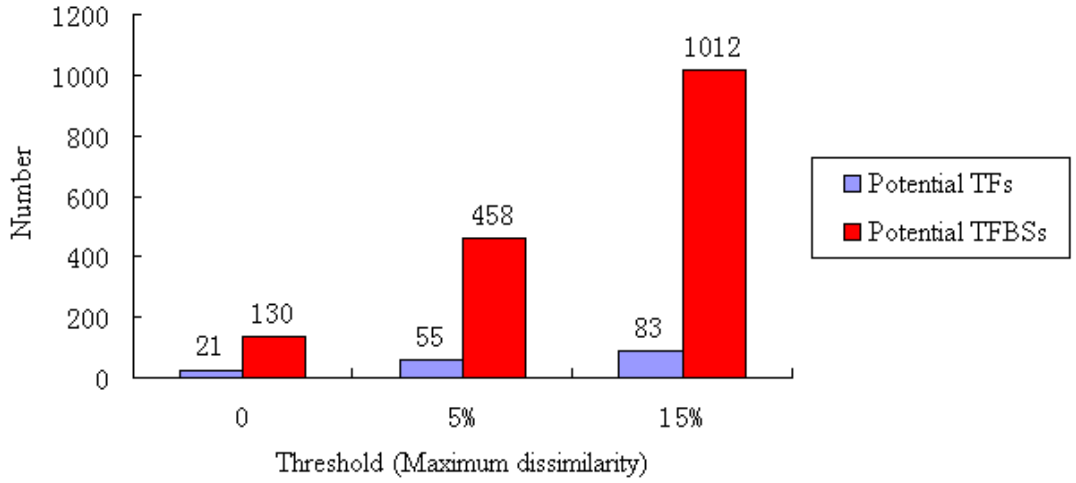
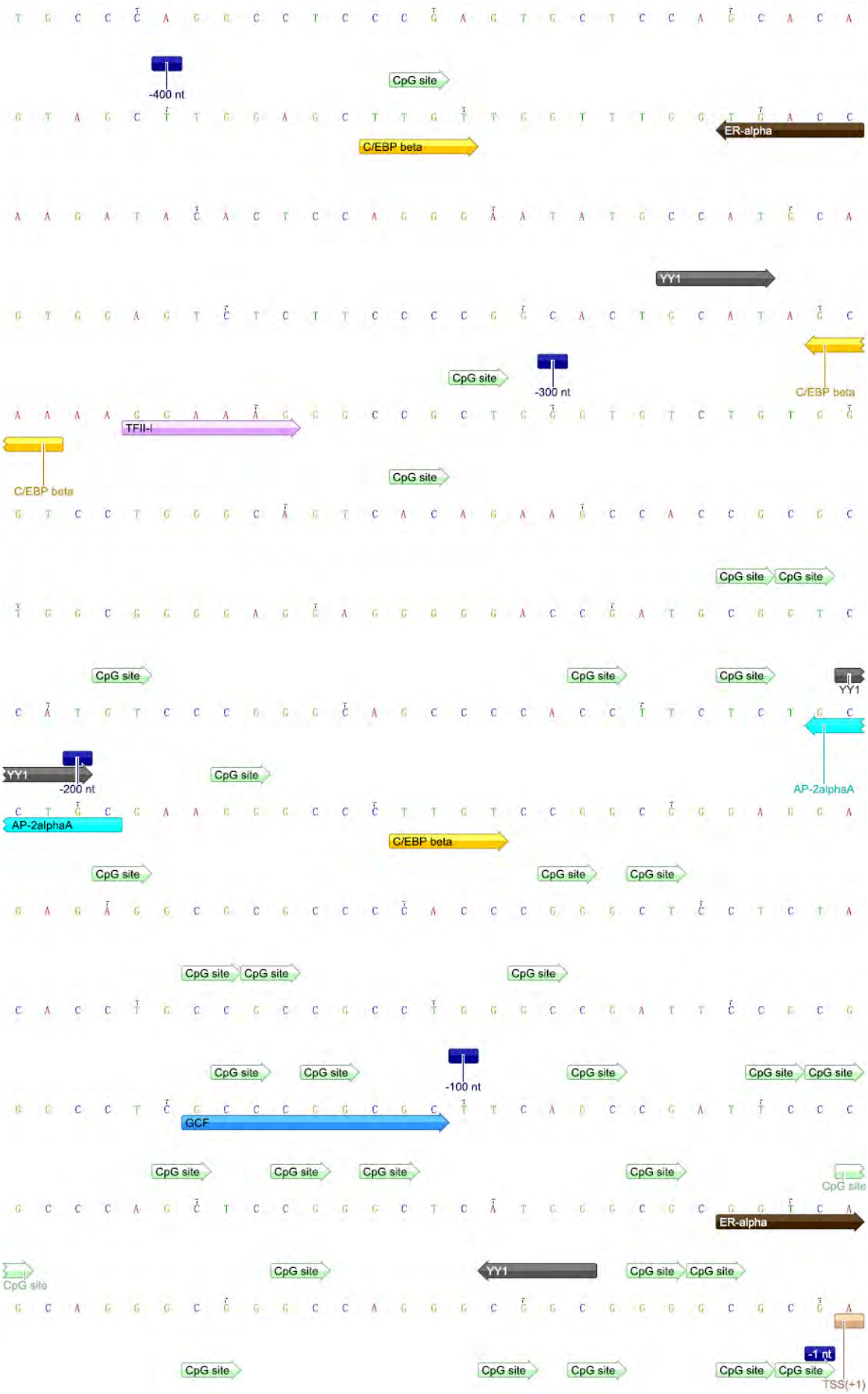


Figure 9. Number of potential transcription factors that may bind *Syk* promoter region under different thresholds. TF represents transcription factor. TFBS represents transcription factor binding site. The x-axis represents maximum dissimilarity used for prediction of transcription factors and binding sites. The y-axis represents the number of the potential transcription factors or putative transcription factor binding sites. The blue bars represent the "number of kinds of potential TFs". The red rectangles represent the "number of potential TF binding sites".


-400 nt
CpG site
ER-alpha
C/EBP beta
YY1
CpG site
-300 nt
C/EBP beta
TFII-I
C/EBP beta
CpG site
CpG site
CpG site
CpG site
CpG site
YY1
YY1
-200 nt
CpG site
AP-2alphaA
AP-2alphaA
C/EBP beta
CpG site
CpG site
CpG site
CpG site
CpG site
CpG site
-100 nt
CpG site
CpG site
GCF
CpG site
CpG site
CpG site
CpG site
CpG site
ER-alpha
CpG site
CpG site
YY1
CpG site
CpG site
-1 nt
CpG site
CpG site
CpG site
CpG site
CpG site
TSS(+1)

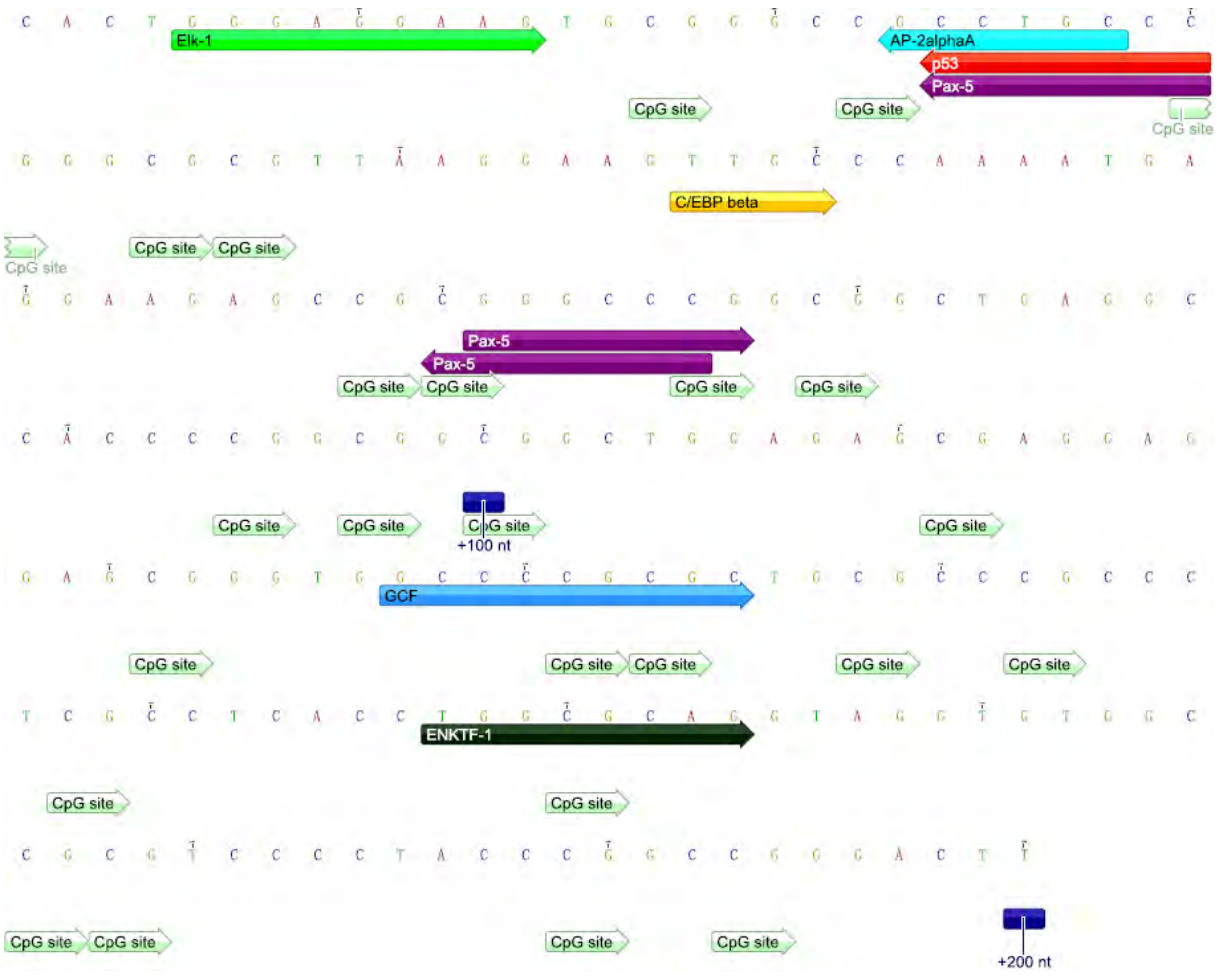


Figure 10. GpG sites and the putative transcription factor binding sites in the *Syk* promoter from -400 nt to +200 nt at 0 dissimilarity. TSS (+1) represents the transcription start site (represented by the grayish pink rectangle below A). The arrows with distinct colors represent different transcription factors.

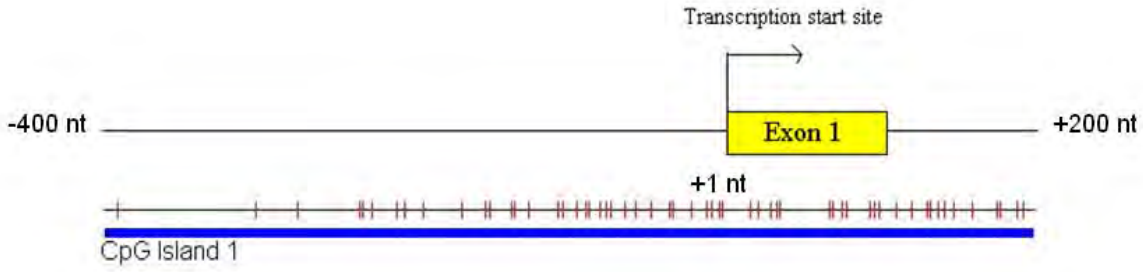


Figure 11. CpG island prediction in the *Syk* promoter region (-400 nt to +200 nt) using CpG island searcher. CpG island prediction was made by CpG island searcher. After inputting the fasta sequence of *Syk* promoter region and setting the threshold of each parameter (such as percentage of G+C content, frequency of occurrence of CpG dinucleotides, and length, mentioned in the introduction). To search the CpG islands, criteria was set as Gardiner- Garden and Frommer's that has been widely used for searching CpG islands. Percentage of G+C content (%GC) was set more than 50%, frequency of occurrence of CpG dinucleotides (ObsCpG/ExpCpG) was above 0.6. The bars represent CpG dinucleotides.

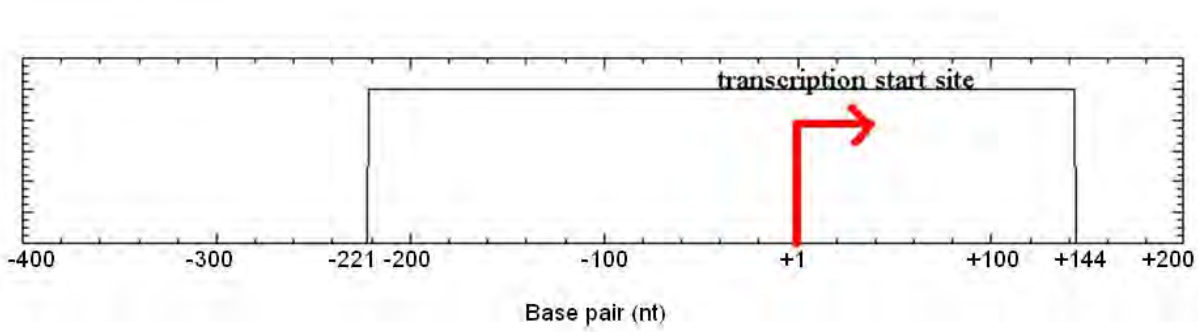


Figure 12. CpG island prediction in the *Syk* promoter region (-400 nt to +200 nt) using EMBOSS CpGPlot. The criterion of each parameter (such as percentage of G+C content, frequency of occurrence of CpG dinucleotides, and length, mentioned in the introduction) was set as Gardiner- Garden and Frommer's that has been widely used for searching CpG islands. After inputting the fasta sequence of *Syk* promoter region, CpG island prediction was made by EMBOSS CpGplot. Percentage of G+C content (GC%) was set more than 50%, frequency of occurrence of CpG dinucleotides (ObsCpG/ExpCpG) was above 0.6, and length of CpG island should be more than 200 nt. After inputting the fasta sequence of *Syk* promoter region, CpG island prediction was made by EMBOSS CpGplot. The flat line starting from -221 to +144 nt represents the putative CpG island in this region.

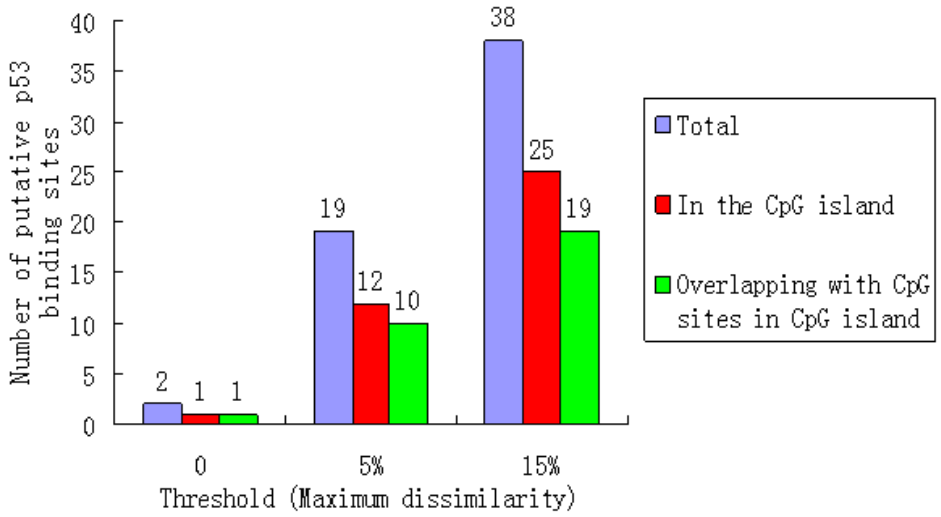


Figure 13. Prediction of putative p53 binding sites in the *Syk* promoter regions from -2000nt to +200nt and from -400 nt to +200 nt under different thresholds. The x-axis represents maximum dissimilarity used for prediction of p53 putative binding sites. The y-axis represents the number of the potential p53 binding sites. "Total" (in blue) represented total number of all the putative p53 binding sites in Syk promoter region from -2000nt to +200nt. "In the CpG island" (in red) indicated the number of potential p53 binding sites in a region of Syk promoter from -400 nt to +200 nt; while "Overlapping with CpG sites in CpG island" (in green) described the number of potential p53 binding sites overlapping with CpG sites in -400 nt to +200 nt of the Syk promoter region.

2.4 Co-immunoprecipitation analysis of the interaction between latent membrane protein 2A and spleen tyrosine kinase

Syk was found to interact with LMP2A in epithelial cells to mediate cell migration (Lu et al., 2006). Fu Chen et al found that phosphorylated Syk could interact with phosphorylated LMP2A in epithelial cells (Fu Chen, Maria Werner, and Ingemar Ernberg, unpublished data). However, whether one or both isoforms are involved in such interaction is still unclear. To explore this, co-immunoprecipitation was used. 5637 cells without or with LMP2A stable transfection were transiently transfected with Syk-coding cDNA. Raji cells that expressed high level of endogenous Syk (L) were used as control. After lysing cells, Syk was precipitated using anti-Syk antibodies, precipitated proteins were analyzed by western blot (Figure 14). However, I could not detect LMP2A. Many non-specific bands were found (in the bottom left rectangle, Figure 14). This might be due to the quality of N19 antibody used for immunoprecipitation. Another possible reason is that some unknown processes during experiments, e.g. dephosphorylation of Syk or LMP2A might explain the failure since only phosphorylated Syk could interact with phosphorylated LMP2A (Fu Chen, Maria Werner, and Ingemar Ernberg, unpublished data). In the future, reversed co-immunoprecipitation should be tried, precipitating with antibodies against LMP2A, e.g. 14B7, to see if it pulls down Syk. Unfortunately, there was no time to optimize this protocol.

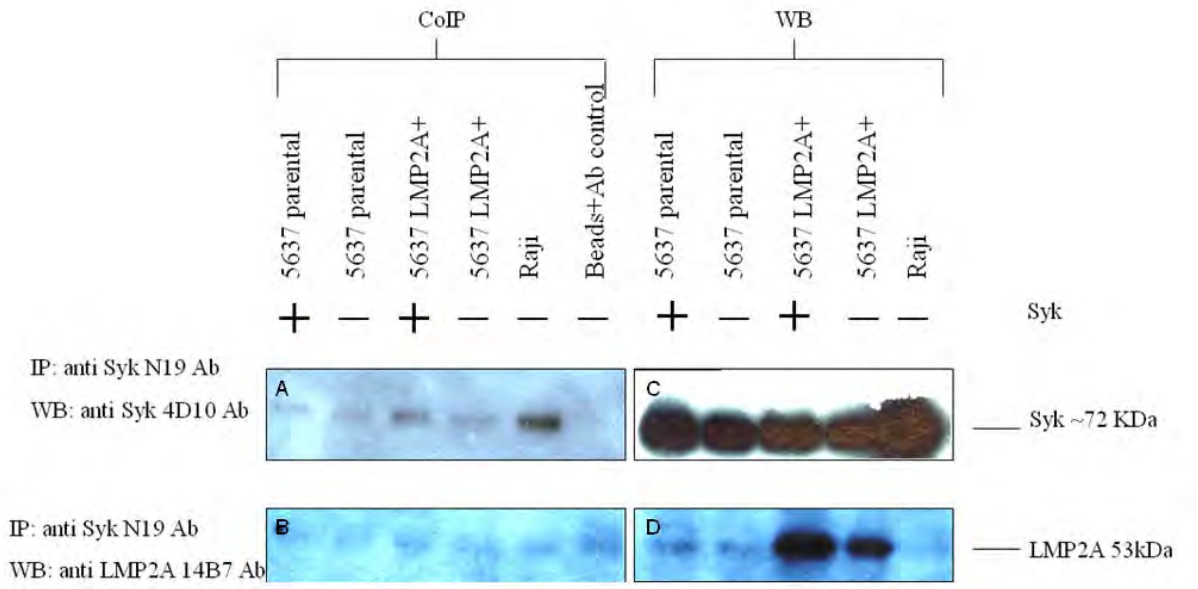


Figure 14. Immunoprecipitation and western blot detection of interaction of Syk with LMP2A. 5637 cells stably expressing LMP2A (clone 4, afterwards referred as 5637/2A4) and 5637 wildtype cells, and Raji cells (positive Syk expression) were used. 5637/2A4 and 5637 cells were transiently transfected with *Syk* cDNA and then cultured for 48 hr. "+" represents cells transfected with *Syk* cDNA; and "-" represents cells without *Syk* cDNA transfection. All cells were lysed. Total cell lysates were subjected to immunoprecipitation (rabbit anti-Syk polyclonal N19 antibody) and then analyzed by western blot for Syk (panel A) using mouse anti-Syk monoclonal 4D10 antibody or for LMP2A (panel B) using rat anti-LMP2A monoclonal 14B7 antibody. The total Syk or LMP2A were detected by western blot using mouse anti-Syk monoclonal 4D10 antibody (panel C) or using rat anti-LMP2A monoclonal 14B7 antibody (panel D). Then corresponding peroxidase-conjugated secondary antibodies were used and the blots were visualized using enhanced chemiluminescent detection. IP: immunoprecipitation. WB: western blot.

3. Discussion

There are two isoforms of Syk protein in *Homo sapiens* (Figure 3). To find out which isoform of Syk is involved in suppression is of great interest. After transfecting cDNA that encodes functional Syk protein (inhibiting invasion and migration of epithelial tumor, Fu Chen, et al. unpublished data) into CNE1 and HeLa cells, western blot analysis showed that Syk from the transfected cDNA was in between of the place of endogenous Syk (L) and Syk (S) (Figure 5; Fu Chen, Li-Sophie Zhao Rathje, Maria Werner, Zuobai Wang, Ingemar Ernberg, unpublished data). Sequencing would be needed to find out which part of Syk is missing in Syk (L). Transfected Syk was expressed much higher in HeLa cells compared with CNE1 cells (Figure 5). One possible reason is that CNE1 cells might have a different time point for peak Syk expression compared with HeLa cells. However, although a series of time points for Syk expression at protein level in HeLa cells and CNE1 cells (0h, 24h, 36h, and 48h) were taken to study the peak expression time point, both cells show highest expression 24h after transfection (Fu Chen, Li-Sophie Zhao Rathje, Maria Werner, Zuobai Wang, Ingemar Ernberg, unpublished data). Thus, perhaps longer time and more time points would be needed to check if this is the reason for higher expression of *Syk* cDNA in HeLa cells. Another interpretation is that *Syk* cDNA could be more transcribed and further translated in HeLa cells than CNE1 cells due to different intracellular background and signal pathway. To check this, *Syk* mRNA at a series of time points after *Syk* cDNA transfection would be extracted and investigated via RT-PCR.

At the transcription level of the parental Syk (not the transfected Syk), both isoforms of Syk were detected in 5637 and Rko cells, although the short isoform was much less abundant than the long one in each cell lines; Syk (L) was detected in CNE1 cells but Syk (S) was almost absent; only Syk (L) was found in Raji cell, which is correlated to the findings of Wang and colleagues (Wang et al., 2003); while very low level of or even no Syk was expressed in HeLa cell (Figure 7) . Thus, for the epithelial cell types (5637, Rko, CNE1 and HeLa), Syk expression pattern differs. The reason why distinct epithelial cell types differ in Syk expression pattern is still unclear. Perhaps, the progress of aggressiveness results in changes of the Syk expression pattern from Syk (L) to Syk (L) and Syk (S) to nonexpression. The different expression pattern may also result from different cell background (tissue specific expression). If the former is the truth, new early diagnosis of epithelial tumor may be developed via detecting Syk expression pattern. Therefore, it is very meaningful. Analyses of biopsies from patients suffering from epithelial tumor may help understand the reason.

In terms of both the translational level and the transcriptional level, HeLa cells could serve as an ideal host for *Syk* cDNA transfection in terms of extremely low production of Syk (Lauvrak et al., 2006; Renedo et al., 2001). However, after *Syk* cDNA transfection into HeLa cells, the number of host cells seems to decrease sharply, resulting in great occurrence of cellular fragments (Fu Chen, Li-Sophie Zhao Rathje, Maria Werner, Zuobai Wang, Ingemar Ernberg, unpublished data). This is perhaps because of overexpression of *Syk* resulting in a great burden to the HeLa cells, exerting a bad or even lethal effect on the host cells. Another interpretation is that Syk protein serves as a tumor suppressor, inducing death of malignant

cells. More studies, such as investigating the apoptotic factors in the HeLa cells with *Syk* transfection, are needed in order to find out the reason affecting viability of these transfectants. In a word, in terms of host cell viability after transfection, HeLa cells do not seem to be an acceptable control and host cells transfected with *Syk* cDNA for study.

In several studies (Coopman and Mueller, 2006; Wang et al., 2003), MDA-MB-231, a malignant breast tumor cell line in lack of Syk expression, was used as host cells for *Syk* transient transfection and further served as a model system for studying Syk's role in suppression of epithelial cells. After *Syk* transient transfection into MDA-MB-231 cells, the cell motility decreased sharply but the cell viability did not change much. Thus, in order to find out which isoform of Syk carries out the tumor suppressor role, MDA-MB-231 could be separately transfected with Syk (L) and Syk (S) cDNA. Then, the aggressiveness, expression microarray and signal pathways of these two different transfectants could be compared. Meanwhile, which exact part(s) of the Syk (L) and Syk (S) is/are involved in the signal transduction leading to tumor suppression could be studied.

Regulation of *Syk* expression at the transcription level is also under intensive study. Although DNA methylation at the promoter region of *Syk* already has been found to be critical for regulation of *Syk* expression (Yuan et al., 2001; Yuan et al., 2006), little evidence suggests transcription factor-mediated *Syk* expression. Hence, it is very attractive to study regulation of *Syk* expression by transcription factors. Under perfect matching condition, the number of transcription factors and their putative binding sites were 21 and 130 respectively. After increasing the maximum dissimilarity to 5% and 15%, both the kinds of transcription factors and the number of potential binding sites increased sharply (Table 7, Table 8 and Figure 9). Among these transcription factors, p53 has already been found to positively regulate *Syk* expression, although no exact site *in vivo* has been found out (Vrba et al., 2008; Xu and el-Gewely, 2001). Since methylation at CpG sites in CpG island could regulate gene expression, the overlap of p53 putative binding sites with CpG sites in the CpG island region of *Syk* (-400 to +200 nt) was analyzed. It was found that putative p53 binding sites were more condense in such region both at 5% and at 15% maximum dissimilarity and that the putative p53 binding sites overlapped with CpG sites. Thus, after CpG methylation, p53 binding to their binding sites in the *Syk* promoter region may be interfered. However, experiments using this information should be taken in the future to analyze if CpG methylation really hinders p53 binding to *Syk* promoter region. If binding of p53 could be interfered by DNA methylation, new questions come. Is this interference due to blockage of p53 binding sites by methyl-CpG recruited proteins or the methyl group in the cytosine residues? Possible ways to study this are mutation and RNAi. After mutation of the methyl cytosine binding proteins or screening the methyl cytosine binding proteins, the binding of p53 to methyl CpG sites would be studied. If p53 could still bind to the methyl CpG sites, the blockage of p53 binding to CpG sites would be due to the binding of the methyl cytosine binding proteins; otherwise, methyl groups play the role in hindering p53 binding.

In addition to the interaction between Syk and LMP2A, these days, it is found that Syk could also interact with the integrin, eg. integrin α II β3 in platelet (Miranti et al., 1998) and integrin

α6β4 in epithelial cells (Fu Chen, Li-Sophie Zhao Rathje, Maria Werner, Zuobai Wang, Ingemar Ernberg, unpublished data). Integrin is involved in cell adhesion and migration, and a lot of signaling pathways contributing to development and oncogenesis (Guo and Giancotti, 2004; Nikolopoulos et al., 2004). Platelets have a unique integrin signaling pathway including Src kinases, Syk, Vav1, and Cbl, which may exert influences on actin polymerization and cytoskeletal rearrangements (Miranti et al., 1998). However, in epithelial cells, the physiological consequences of interaction between integrin α6β4 and Syk are still under study. Based on the current findings from Fu Chen, Ingemar Ernberg, et al., binding of Syk to integrin and its downstream signaling pathways seem to inhibit invasion of carcinoma cells (Fu Chen, Ingemar Ernberg et al., unpublished data). Thus, new questions come as in the epithelial cells which hold both the Syk-interacting proteins integrin and LMP2A, which Syk would prefer to act with and what this preference would result become interesting. The next work after this research studying the Syk-LMP2A interaction and role of different isoforms of Syk would further go to details of Syk-Integrinα6β4 interaction and the effect of this interaction on the Syk-LMP2A interaction and oncogenesis.

4. Materials and methods

4.1 Cell lines and cell culture

Cell lines are described in Table 2. Cells were cultured at 37℃ with 5% CO_2 concentration.

Table 2. Cell lines used in this study.

Cell lines	Properties
5637[(1)]	Human bladder carcinoma cell line
5637 LMP2A transfectant subclone 4[(1)]	Human bladder carcinoma cell line with stable LMP2A transfection
CNE1[(1)]	Human nasopharyngeal carcinoma cell line
CNE1 LMP2A transfectant subclone 14[(1)]	Human nasopharyngeal carcinoma cell line with stable LMP2A transfection
HeLa[(1)]	Human cervical carcinoma cell line, used in this research as a Syk negative control
Rko[(2)]	Human colorectal carcinoma cell line
Raji[(3)]	Human Burkitt's lymphoma cell line

Footnotes:

(1) Cells were cultured in IMDM medium (HyClone, #SH30228.01) containing 10% FBS (GIBCO, #10500-064), 100 units/ml penicillin and 100 μg/ml streptomycin (HyClone, #SV30010).

(2) Cells were cultured in RPMI-1640 medium (HyClone, #SH30027.01) containing 10% FBS (GIBCO, #10500-064), 100 units/ml penicillin and 100 μg/ml streptomycin (HyClone, #SV30010) and 25mM HEPES (HyClone, #SH30237.01)

(3) Cells were cultured in RPMI-1640 medium (HyClone, #SH30027.01) containing 10% FBS (GIBCO, #10500-064), 100 units/ml penicillin and 100 μg/ml streptomycin (HyClone, #SV30010)

4.2 Transient transfection with *spleen tyrosine kinase* DNA

Transfection was carried out with Lipofectamine™ 2000 reagent (Invitrogen) following the manufacturer's protocol. Lipofectamine™ 2000 and DNA were firstly individually diluted into Opti-MEM with ratios according to the manufacturer's protocol and mixed sufficiently (http://tools.invitrogen.com/Content/SFS/ProductNotes/F_Lipofectamine%202000b-040923-RD-MKT-TL-HL050602.pdf). Dilutions were incubated five minutes at room temperature. 1 μl lipofectamine™ 2000 reagent per 2 μg *Syk* cDNA (gift from Fu Chen) were combined for 20 minutes at room temperature. The mixture was added to 90-95% confluence of cells that were cultured at 37℃ and 5% CO_2 for 5 h. Then, the mixture was replaced by normal medium without penicillin or streptomycin for culturing cells for 48 h.

4.3 RNA extraction and reverse transcription PCR

Total RNA from different cell lines were extracted using TRIzol Reagent (Invitrogen, #15596-026) according to the manufacturer's instruction. RNA concentrations were measured using a Nanodrop spectrophotometer (Thermo Fisher Scientific Inc.). RNA samples were electrophoresized and normalized on an agarose gel (1% agarose in 0.5×Tris-acetate-EDTA buffer, TAE; diluted 100 times in ddH_2O from 50×TAE; 50×TAE: 2M Tris-acetate, 50 mM Na_2EDTA). 3 μg RNA was reverse-transcribed using 4 μM random hexamer primer (Fermentas, #R0192), 10 mM DTT, 1 mM dNTP mix, 10 U RNase inhibitor (Fermentas, #EO038), 0.6 U AMV reverse transcriptase (Promega, #M510F) and 5 μl of 5 AMV reverse transcriptase buffer (Promega, #M515A) in a total volume of 25 μl at 42℃ for 1 hour and 95℃ for 10 min. The obtained cDNA was used for PCR amplification (for conditions see Table 3). The primer sequences are shown in Table 4. PCR program for all primers was set as follows: 94℃ for 3 min, 35 cycles of 94℃ for 30 sec, 54℃ for 1 min, and 72℃ for 1 min, and 72℃ for 10 min. The PCR products were electrophoresized on a agarose gel (2% agarose in 0.5×Tris-borate-EDTA buffer, TBE; diluted 20 times in ddH_2O from 10×TBE; 10×TBE: 890mM Tris-borate, 890mM boric acid, 20mM EDTA).

Table 3. PCR coditions

Reagents	Volume (μl)
cDNA	4
ddH_2O	14.875
10×PCR buffer(1)	2.5
10μM dNTP(1)	0.5
25mM magnesium ion(1)	1.5
Primer (Syk primer or GAPDH primer): Forward strand(1)	0.5
Primer (Syk primer or GAPDH primer): Reverse strand(1)	0.5
Taq DNA polymerase(2)	0.625
Total volume	25

Footnote:

(1): From Fermentas

(2): From Invitrogen

Table 4. PCR primers used in this study.

Primer	Target	Sequence (5' → 3')
Syk796F	Forward primer for *Syk*	AGACAACAACGGCTCCTAC
Syk1187R	Reverse primer for *Syk*	CAAGTTCTGGCTCATACGG
GAPDH300F	Forward primer for *GAPDH*	GCTTGTGATCAATGGAAATC
GAPDH869R	Reverse primer for *GAPDH*	TCATATTTGGCAGGTTTTTC

4.4 Western blot

For Western blot, 3×10^5 cells of chosen cell lines cultured in separated wells of six-well plate were lysed for 15 min at 4 ℃ with 80 μl 1% NP40 lysis buffer (see Table 6). Then, extracts were centrifuged at 10,600×g at 4 ℃ for 15 min to get rid of the debris. The total protein concentrations of different samples were measured according to instruction of Bio-Rad Protein Assay (BIO-RAD) using 16 μg/ml bovine serum albumin to make the standard curve.

Equal amounts of protein (maximum volume 25 μl) was mixed with corresponding equal volume of 2×loading buffer (see Table 6) and heated together for 5 min and then loaded for SDS-polyacrylamide gel electrophoresis (PAGE) in the running buffer (see Table 6). Nitrocellulose membrane (GE Healthcare Amersham, Hybond™-C Extra) was cut into the size similar to gel and then rinsed in transfer buffer (see Table 6). The proteins were transferred from the gel to nitrocellulose membrane at 0.2 mA for 1 hour in transfer buffer. 5% nonfat dry milk in Tris buffered saline (see Table 6) containing 0.1% Tween20 (TBST, see Table 6) was used for blocking for 1 hr at room temperature. Then, 3 ml 1:1000 dilution of N19 antibody or 1:1000 dilution of 14B7 antibody in the freshly-made blocking buffer (Table 5) were used for probing Syk and LMP2A respectively. The membrane was then rinsed 3 times in TBST totally for 30 min, followed by probing by 20 ml 1:1000 dilution of different peroxidase-conjugated secondary antibodies in the freshly-made blocking buffer and another round of 3 times of TBST rinsing totally for 30 min. Signals were detected by enhanced chemiluminescence (GE Healthcare) was chosen for developing purpose.

Table 5. Antibodies used in this study.

Antibody	Property	Manufacturer	Catalog number
4D10	Mouse anti-Syk monoclonal IgG2a molecules	Santa Cruz Biotechnology, CA, US	J0102
14B7	Rat anti-LMP2A monoclonal IgG2a molecules	ITN GmbH, Neuherberg, Germany	
N19	rabbit anti-Syk polyclonal IgG molecules	Santa Cruz Biotechnology, CA, US	D2809
6C5	Mouse anti-GAPDH monoclonal IgG1 molecules	Santa Cruz Biotechnology, CA, US	E1407
Rabbit anti-Rat immunoglobulins/HRP	Peroxidase-conjugated rabbit secondary antibody against rat IgG molecules	Dako Cytomation, Glostrup, Denmark	P0450
Goat anti-mouse IgG HRP-conjugate	Peroxidase-conjugated goat secondary antibody against mouse IgG molecules	BIO-RAD, CA, US	172-1011

Table 6. Reagents for polyacrylamide gel electrophoresis

Reagent	Recipe
1% NP40 lysis buffer	1% Nonidet-P40, 150 mM NaCl, 50 mM Tris–HCl (pH 7.4), 2 mM EDTA, 10 μg (each) pepstatin and leupeptin per mL, 0.5 mM phenylmethylsulfonyl fluoride, 1 mM sodium orthovanadate
2×Loading buffer	2 ml glycerol (100%), 2.5 ml 0.5 M Tris-HCl (pH 6.8), 0.5 ml 0.05% bromphenol blue, 4 ml 10% SDS (laryl sulfate), 1 ml β-mercaptoethanol
Running buffer	1% SDS, 24.76 mM Tris base, and 191.82 mM glycine
Transfer buffer	25 mM Tris base, 192 mM glycine, and 10% methanol
Stacking solution	0.5 M Tris-HCl (pH 6.8), 0.4% SDS
Separation solution	1.5 M Tris-HCl (pH 8.8), 0.4% SDS

4% stacking gel	0.5 ml 40% acrylamide/1.5% bisacrylamide (Bio-Rad), 1.25 ml stacking solution, 3.25 ml ddH_2O, 5 µl 1,2-bis(dimethylamino)ethane (TEMED), 25 µl 10% ammonium persulfate (APS)
9% separation gel	2.25 ml 40% acrylamide/1.5% bisacrylamide (Bio-Rad), 2.5 ml stacking solution, 5.25 ml ddH_2O, 10 µl 1,2-bis(dimethylamino)ethane (TEMED), 50 µl 10% ammonium persulfate (APS)
Tris buffered saline	25 mM Tris-HCl (pH 7.4), 150 mM NaCl, 2 mM KCl
Tris buffered saline with 0.1% Tween	0.1% Tween 20, 25 mM Tris-HCl (pH 7.4), 150 mM NaCl, 2 mM KCl,
Blocking buffer	5% nonfat dry milk powder (w/v), 0.1% Tween 20, 25 mM Tris-HCl (pH 7.4), 150 mM NaCl, 2 mM KCl

4.5 Co-Immunoprecipitation

1×10^6 cells of chosen cell lines cultured in 10 cm petri dish were lysed for 15 min at 4 ℃ with 1 ml of 1% NP40 lysis buffer (see Table 6). Then, extracts were centrifuged at 10,600× g at 4 ℃ for 15 min to get rid of the debris for preclarifying. The supernatant was added 100 µl protein A agarose resins (Thermo Fisher Scientific, US; binding to the Fc region of IgG molecules) and well mixed by rotating for 30 min at 4 ℃. 1.5 µg of N19 primary antibody against Syk (details in Table 5) were added to the supernatants, rotating at 4 ℃ for 16 hours. Afterwards, the beads were washed in NP-40 lysis buffer five times, 1 ml each sample every time. Then, pellets were obtained following centrifugation at 1,300×g for 45 sec at 4 ℃, and were resuspended in 45 µl of 2× loading buffer followed by being boiled for 5 min. The beads were removed by centrifugation at 17,900×g at room temperature for 1 min and the supernatants collected. The samples dissolved in 2× loading buffer were loaded and separated by 9% gel. Nitrocellulose membrane (GE Healthcare Amersham, Hybond™-C Extra) was cut into the size similar to gel and then rinsed in transfer buffer (see Table 6). The proteins were transferred from the gel to nitrocellulose membrane at 0.2 mA for 1 hour in transfer buffer. 5% of nonfat dry milk in Tris buffered saline (TBS) containing 5% milk and 0.1% Tween20 (TBST) was used for blocking for 1 hr at room temperature. Then, 3 ml 1:1000 dilution of N19 antibody or 1:1000 dilution of 14B7 antibody (Table 5) were used for probing Syk and LMP2A respectively, The membrane was then rinsed 3 times in TBST totally for 30 min, followed by probing by 20 ml 1:1000 dilution of different corresponding peroxidase-conjugated secondary antibodies (see Table 5) and another round of 3 times of TBST rinsing totally for 30 min. Signals were detected by enhanced chemiluminescence (GE Healthcare) was chosen for developing purpose.

4.6 Bioinformatic analysis of *spleen tyrosine kinase* promoter

The NCBI database (http://www.ncbi.nlm.nih.gov/sites/entrez?Db=gene&Cmd=retrieve&dopt=full_report&list_uids=6850&log$=databasead&logdbfrom=nuccore) was searched for the *Syk* gene sequence (Maglott et al., 2007). Then, a region spanning from 2000 nt upstream of the first transcribed site (transcription start site, TSS) to 200 nt downstream was analyzed for promoter information. PROMO 3.0 (Farre et al., 2003; Messeguer et al., 2002) was used to identify the

putative transcription factors and their binding sites in this Syk promoter region. Then, the resulted transcription factors and their binding sites were analyzed using Geneious Pro version 4.8.5. Meanwhile, a region ranging from 400 nt upstream of the TSS to 200 nt down stream was analyzed for CpG island prediction using website based analyzing software - EMBOSS CpGPlot (http://www.ebi.ac.uk/Tools/emboss/cpgplot/index.html). This region was also analyzed further using Geneious Pro version 4.8.5.

5. Acknowledgement

Helps from other nice people were strongly indispensable to those results obtained in this thesis. Herein, at the end of this thesis, I would like firstly to thank Professor Ingemar Ernberg, Post-doc Fu Chen. Both Professor Ingemar Ernberg and Post-doc Fu Chen gave me a lot of ideas and taught me many things from principles to experiments. Their helps contributed hugely to my work. I will never forget those nice teachers.

And I would also present my gratefulness and memories to those passed away. You will forever live in my heart.

I would like to send my sincere thankfulness to Professor Karin Carlson as well. Her valuable suggestions helped me a lot when I wrote my thesis and defended my thesis.

I am also very grateful to all the other members in our group. PhD student Aymeric Fouquier d'Hérouël taught me a lot of information about bioinformatics. Post-doc Jiezhi Zou, Post-doc Li-Sophie Zhao Rathje, PhD student Ziming Du and PhD student Qin Li taught me a lot of experimental details. Associated Professor Lifu Hu gave me a lot of backgrounds in the field of CpG islands. I also appreciate to PhD Maria Werner, master student Ilquar Abdullayev, Imran Nawaz Bugti and all the rest members of our group.

Professor Domènec Farré, who told me a lot of useful information about PROMO v3.0, is another person to whom I want to send my regards.

I would also like to send my gratefulness to National Library of Medicine (NLM) Web for providing me with the *Syk* genome picture in the public domain information.

Here, my dear classmates and friends in Uppsala and Stockholm also deserve my warmest regards. They helped me a lot and relieved me from certain stress when I stay in Sweden. I am really touched by them.

In the end, I would like to send my warmest regards to my dearest parents, grandparents and my friends in China (PRC). They supported me a lot during these years. I miss them all; and extremely grateful to them.

6. References

Adams, A., and Lindahl, T. (1975). Epstein-Barr virus genomes with properties of circular DNA molecules in carrier cells. Proc Natl Acad Sci U S A *72*, 1477-1481.

Allen, M.D., Young, L.S., and Dawson, C.W. (2005). The Epstein-Barr virus-encoded LMP2A and LMP2B proteins promote epithelial cell spreading and motility. J Virol *79*, 1789-1802.

Ben-Gal, I., Shani, A., Gohr, A., Grau, J., Arviv, S., Shmilovici, A., Posch, S., and Grosse, I. (2005). Identification of transcription factor binding sites with variable-order Bayesian networks. Bioinformatics *21*, 2657-2666.

Bornkamm, G.W. (2009). Epstein-Barr virus and the pathogenesis of Burkitt's lymphoma: more questions than answers. Int J Cancer *124*, 1745-1755.

Bornkamm, G.W., and Hammerschmidt, W. (2001). Molecular virology of Epstein-Barr virus. Philos Trans R Soc Lond B Biol Sci *356*, 437-459.

Coopman, P.J., Do, M.T., Barth, M., Bowden, E.T., Hayes, A.J., Basyuk, E., Blancato, J.K., Vezza, P.R., McLeskey, S.W., Mangeat, P.H., and Mueller S.C. (2000). The Syk tyrosine kinase suppresses malignant growth of human breast cancer cells. Nature *406*, 742-747.

Coopman, P.J., and Mueller, S.C. (2006). The Syk tyrosine kinase: a new negative regulator in tumor growth and progression. Cancer Lett *241*, 159-173.

Djordjevic, M., Sengupta, A.M., and Shraiman, B.I. (2003). A biophysical approach to transcription factor binding site discovery. Genome Res *13*, 2381-2390.

Duta, F., Ulanova, M., Seidel, D., Puttagunta, L., Musat-Marcu, S., Harrod, K.S., Schreiber, A.D., Steinhoff, U., and Befus, A.D. (2006). Differential expression of spleen tyrosine kinase Syk isoforms in tissues: Effects of the microbial flora. Histochem Cell Biol *126*, 495-505.

Epstein, M.A., Achong, B.G., and Barr, Y.M. (1964). Virus Particles in Cultured Lymphoblasts from Burkitt's Lymphoma. Lancet *1*, 702-703.

Farre, D., Roset, R., Huerta, M., Adsuara, J.E., Rosello, L., Alba, M.M., and Messeguer, X. (2003). Identification of patterns in biological sequences at the ALGGEN server: PROMO and MALGEN. Nucleic Acids Res *31*, 3651-3653.

Feldman, A.L., Sun, D.X., Law, M.E., Novak, A.J., Attygalle, A.D., Thorland, E.C., Fink, S.R., Vrana, J.A., Caron, B.L., Morice, W.G., Remstein E.D., Grogg K.L., Kurtin P.J., Macon W.R., and Dogan A. (2008). Overexpression of Syk tyrosine kinase in peripheral T-cell lymphomas. Leukemia *22*, 1139-1143.

Gardiner-Garden, M., and Frommer, M. (1987). CpG islands in vertebrate genomes. J Mol Biol *196*, 261-282.

Gershenzon, N.I., Stormo, G.D., and Ioshikhes, I.P. (2005). Computational technique for improvement of the position-weight matrices for the DNA/protein binding sites. Nucleic Acids Res *33*, 2290-2301.

Guo, W., and Giancotti, F.G. (2004). Integrin signalling during tumour progression. Nat Rev Mol Cell Biol *5*, 816-826.

Harrison, M.J., Tang, Y.H., and Dowhan, D.H. (2010). Protein arginine methyltransferase 6 regulates multiple aspects of gene expression. Nucleic Acids Res *38*, 2201-2216.

Inatome, R., Yanagi, S., Takano, T., and Yamamura, H. (2001). A critical role for Syk in endothelial cell proliferation and migration. Biochem Biophys Res Commun *286*, 195-199.

Jones, P.A., and Baylin, S.B. (2002). The fundamental role of epigenetic events in cancer. Nat Rev Genet *3*, 415-428.

Kel, A.E., Gossling, E., Reuter, I., Cheremushkin, E., Kel-Margoulis, O.V., and Wingender, E. (2003). MATCH: A tool for searching transcription factor binding sites in DNA sequences. Nucleic Acids Res *31*, 3576-3579.

Latchman, D.S. (1997). Transcription factors: an overview. Int J Biochem Cell Biol *29*, 1305-1312.

Lauvrak, S.U., Walchli, S., Iversen, T.G., Slagsvold, H.H., Torgersen, M.L., Spilsberg, B., and Sandvig, K. (2006). Shiga toxin regulates its entry in a Syk-dependent manner. Mol Biol Cell *17*, 1096-1109.

Longnecker, R., and Miller, C.L. (1996). Regulation of Epstein-Barr virus latency by latent membrane protein 2. Trends Microbiol *4*, 38-42.

Lu, J., Lin, W.H., Chen, S.Y., Longnecker, R., Tsai, S.C., Chen, C.L., and Tsai, C.H. (2006). Syk tyrosine kinase mediates Epstein-Barr virus latent membrane protein 2A-induced cell migration in epithelial cells. J Biol Chem *281*, 8806-8814.

Luangdilok, S., Box, C., Patterson, L., Court, W., Harrington, K., Pitkin, L., Rhys-Evans, P., P, O.c., and Eccles, S. (2007). Syk tyrosine kinase is linked to cell motility and progression in squamous cell carcinomas of the head and neck. Cancer Res *67*, 7907-7916.

Maglott, D., Ostell, J., Pruitt, K.D., and Tatusova, T. (2007). Entrez Gene: gene-centered information at NCBI. Nucleic Acids Res *35*, D26-31.

Matys, V., Kel-Margoulis, O.V., Fricke, E., Liebich, I., Land, S., Barre-Dirrie, A., Reuter, I., Chekmenev, D., Krull, M., Hornischer, K., Voss N., Stegmaier P., Lewicki-Potapov B., Saxel H., Kel A.E., and Wingender E. (2006). TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. Nucleic Acids Res *34*, D108-110.

Messeguer, X., Escudero, R., Farre, D., Nunez, O., Martinez, J., and Alba, M.M. (2002). PROMO: detection of known transcription regulatory elements using species-tailored searches. Bioinformatics *18*, 333-334.

Miller, C.L., Lee, J.H., Kieff, E., and Longnecker, R. (1994). An integral membrane protein (LMP2) blocks reactivation of Epstein-Barr virus from latency following surface immunoglobulin crosslinking. Proc Natl Acad Sci U S A *91*, 772-776.

Miranti, C.K., Leng, L., Maschberger, P., Brugge, J.S., and Shattil, S.J. (1998). Identification of a novel integrin signaling pathway involving the kinase Syk and the guanine nucleotide exchange factor Vav1. Curr Biol *8*, 1289-1299.

Mullan, L.J., and Bleasby, A.J. (2002). Short EMBOSS User Guide. European Molecular Biology Open Software Suite. Brief Bioinform *3*, 92-94.

Nikolopoulos, S.N., Blaikie, P., Yoshioka, T., Guo, W., and Giancotti, F.G. (2004). Integrin beta4 signaling promotes tumor angiogenesis. Cancer Cell *6*, 471-483.

Pang, M.F., Lin, K.W., and Peh, S.C. (2009). The signaling pathways of Epstein-Barr virus-encoded latent membrane protein 2A (LMP2A) in latency and cancer. Cell Mol Biol Lett *14*, 222-247.

Pegtel, D.M., Subramanian, A., Sheen, T.S., Tsai, C.H., Golub, T.R., and Thorley-Lawson, D.A. (2005). Epstein-Barr-virus-encoded LMP2A induces primary epithelial cell migration and invasion: possible role in nasopharyngeal carcinoma metastasis. J Virol *79*, 15430-15442.

Reik, W. (2007). Stability and flexibility of epigenetic gene regulation in mammalian development. Nature *447*, 425-432.

Renedo, M.A., Fernandez, N., and Crespo, M.S. (2001). FcgammaRIIA exogenously expressed in HeLa cells activates the mitogen-activated protein kinase cascade by a mechanism dependent on the endogenous expression of the protein tyrosine kinase Syk. Eur J Immunol *31*, 1361-1369.

Sada, K., Takano, T., Yanagi, S., and Yamamura, H. (2001). Structure and function of Syk protein-tyrosine kinase. J Biochem *130*, 177-186.

Scholle, F., Bendt, K.M., and Raab-Traub, N. (2000). Epstein-Barr virus LMP2A transforms epithelial cells, inhibits cell differentiation, and activates Akt. J Virol *74*, 10681-10689.

Stewart, Z.A., and Pietenpol, J.A. (2001). Syk: a new player in the field of breast cancer. Breast Cancer Res *3*, 5-7.

Stormo, G.D. (2000). DNA binding sites: representation and discovery. Bioinformatics *16*, 16-23.

Takai, D., and Jones, P.A. (2002). Comprehensive analysis of CpG islands in human chromosomes 21 and 22. Proc Natl Acad Sci U S A *99*, 3740-3745.

Thompson, M.P., and Kurzrock, R. (2004). Epstein-Barr virus and cancer. Clin Cancer Res *10*, 803-821.

Turner, M., Schweighoffer, E., Colucci, F., Di Santo, J.P., and Tybulewicz, V.L. (2000). Tyrosine kinase SYK: essential functions for immunoreceptor signalling. Immunol Today *21*, 148-154.

Vrba, L., Junk, D.J., Novak, P., and Futscher, B.W. (2008). p53 induces distinct epigenetic states at its direct target promoters. BMC Genomics *9*, 486.

Wang, L., Devarajan, E., He, J., Reddy, S.P., and Dai, J.L. (2005). Transcription repressor activity of spleen tyrosine kinase mediates breast tumor suppression. Cancer Res *65*, 10289-10297.

Wang, L., Duke, L., Zhang, P.S., Arlinghaus, R.B., Symmans, W.F., Sahin, A., Mendez, R., and Dai, J.L. (2003). Alternative splicing disrupts a nuclear localization signal in spleen tyrosine kinase that is required for invasion suppression in breast cancer. Cancer Res *63*, 4724-4730.

Wang, S., Ding, Y.B., Chen, G.Y., Xia, J.G., and Wu, Z.Y. (2004). Hypermethylation of Syk gene in promoter region associated with oncogenesis and metastasis of gastric carcinoma. World J Gastroenterol *10*, 1815-1818.

Wilhelmsen, K., Litjens, S.H., and Sonnenberg, A. (2006). Multiple functions of the integrin alpha6beta4 in epidermal homeostasis and tumorigenesis. Mol Cell Biol *26*, 2877-2886.

Wingender, E., Dietze, P., Karas, H., and Knuppel, R. (1996). TRANSFAC: a database on transcription factors and their DNA binding sites. Nucleic Acids Res *24*, 238-241.

Xu, H., and el-Gewely, M.R. (2001). P53-responsive genes and the potential for cancer diagnostics and therapeutics development. Biotechnol Annu Rev *7*, 131-164.

Yanagi, S., Inatome, R., Ding, J., Kitaguchi, H., Tybulewicz, V.L., and Yamamura, H. (2001). Syk expression in endothelial cells and their morphologic defects in embryonic Syk-deficient mice. Blood *98*, 2869-2871.

Yuan, Y., Mendez, R., Sahin, A., and Dai, J.L. (2001). Hypermethylation leads to silencing of the SYK gene in human breast cancer. Cancer Res *61*, 5558-5561.

Yuan, Y., Wang, J., Li, J., Wang, L., Li, M., Yang, Z., Zhang, C., and Dai, J.L. (2006). Frequent epigenetic inactivation of spleen tyrosine kinase gene in human hepatocellular carcinoma. Clin Cancer Res *12*, 6687-6695.

Zheng, J., Wu, J., and Sun, Z. (2003). An approach to identify over-represented cis-elements in related sequences. Nucleic Acids Res *31*, 1995-2005.

7. Appendix

Table 7. Numbers of putative transcription factor binding sites in the *Syk* promoter region* at maximum dissimilarity 5%.

Transcription factor	Number of binding sites in the promoter region of *Syk*
C/EBP beta	68
POU2F1	1
TFⅡ-Ⅰ	13
STAT4	22
STAT1beta	2
GR-beta	74
PXR-1:RXR-alpha	2
YY1	27
AP-2alphaA	29
PITX2	1
RXR-alpha	7
HNF-3alpha	7
TFⅡD	10
c-ETS-2	8
c-Jun	2
TBP	1
ER-alpha	4
P53	19
XBP-1	4
AR	1
CREB	1
TCF-4E	1
Pax-5	18
T3R-beta1	4
ENKTF-1	2
GCF	7
E2F-1	1
Sp1	4
FOXP3	10
PR B	6
PR A	6
GATA-2	1
GATA-1	8
c-Ets-1	14
VDR	3
HNF-1A	4
RelA	1
C/EBP alpha	4
IRF-2	1

TCF-4	2
SRY	1
RAR-beta	1
GR	8
Ik-1	1
GR-alpha	24
IRF-1	5
Elk-1	5
c-Myb	1
PU.1	1
LEF-1	1
NF-AT1	1
AP-1	1
AhR	1
NF-Y	6
NF-kappaB1	1
Total number of binding sites: 458	

Footnote:

Syk promoter region * represents a region from -2000 nt to +200 nt.

Table 8. Numbers of putative transcription factor binding sites in the *Syk* promoter region* at maximum dissimilarity 15%.

Transcription factor	Number of binding sites in the promoter region of *Syk*
C/EBP beta	68
POU2F1	4
NF-AT2	3
TFⅡ-Ⅰ	47
STAT4	24
c-Ets-1	24
RelA	2
STAT1beta	5
XBP-1	41
GR-beta	82
VDR	6
FOXP3	36
PXR-1:RXR-alpha	6
YY1	27
GR-alpha	99
AP-2alphaA	34
ENKTF-1	24
AhR	2
AhR-Arnt	2
Ik-1	3
NF-1	14

IRF-1	6
c-Ets-2	10
NF-AT1	6
NF-kappa B	2
NF-kappa B1	2
NF-AT1	1
c-Myb	9
LEF-1	5
HNF-1C	4
GATA-2	1
GATA-1	8
Elk-1	11
C/EBPalpha	15
HNF-1A	4
TBP	1
ER-alpha	4
E2F-1	30
POU2F2(Oct-2.1)	2
AR	3
CREB	1
ATF-2	1
PEA3	6
PR B	18
PR A	18
RXR-alpha	15
RAR-beta	4
TF Ⅱ D	28
MEF-2A	5
HNF-3alpha	25
HOXD9	2
HOXD10	2
GR	28
PPAR-alpha:RXR-alpha	3
RAR-beta:RXR-alpha	2
Pax-5	38
P53	38
ATF3	3
SRY	5
TCF-4E	5
EBF	4
AP-1	2
GCF	8
NFI/CTF	20
c-Jun	10

COUP-TF1	1
ELF-1	2
PU.1	2
HNF-1B	2
MAZ	1
ATF	1
T3R-beta1	7
ETF	6
NF-Y	7
STAT5A	1
IRF-2	1
TCF-4	2
USF2	3
Sp1	4
WT1	1
RBP-J kappa	1
GTF	1
c-Fos	1
Total number of binding sites: 1012	

Footnote:

Syk promoter region * represents a region from -2000 nt to +200 nt.

Table 9. GpG sites in the *Syk* promoter region *.

Start sites	End sites
-392	-391
-303	-302
-276	-275
-236	-235
-234	-233
-228	-227
-212	-211
-207	-206
-195	-194
-170	-169
-155	-154
-152	-151
-138	-137
-136	-135
-127	-126
-108	-107
-105	-104
-96	-95
-90	-89
-88	-87

-81	-80
-77	-76
-74	-73
-65	-64
-58	-57
-48	-47
-36	-35
-34	-33
-22	-21
-12	-11
-9	-8
-4	-3
-2	-1
17	18
22	23
30	31
34	35
36	37
68	69
70	71
76	77
79	80
94	95
97	98
100	101
111	112
121	122
131	132
133	134
138	139
142	143
148	149
160	161
176	177
178	179
189	190
193	194

Footnote:

Syk promoter region * represents a region from -2000 nt to +200 nt.

Table 10. p53 and its putative binding sites in the Syk promoter region * at maximum dissimilarity 5%.

No.	Start site	End site	Sequence
1	-1811	-1805	GGGCATG
2	-1690	-1684	GGGCAAC

3	-963	-957	GCTGCCC
4	-470	-464	CGTGCCC
5	-462	-456	ATTGCCC
6	-407	-401	ACTGCCC
7(1)	-255	-249	GGGCAGT
8(2)	-194	-188	GGGCAGC
9(2)	-138	-132	CGCGCCC
10(2)	-83	-77	CTCGCCC
11(2)	-60	-54	CCCGCCC
12(2)	-39	-33	GGGCGCG
13(2)	-25	-19	GGGCGGG
14(1)	-21	-15	GGGCCAG
15(2)	-7	-1	GGGCGCG
16(2)	24	30	CCTGCCC
17(2)	31	37	GGGCGCG
18(1)	46	52	GTTGCCC
19(2)	140	146	CCCGCCC

Footnote:

(1): putative p53 binding sites in the *Syk* promoter region from -400 nt to +200 nt of Syk promoter region.

(2): putative p53 binding sites overlapping with whole CpG site(s) or part of CpG site(s) in the *Syk* promoter region from -400 nt to +200 nt of *Syk* promoter region.

Syk promoter region * represents a region from -2000 nt to +200 nt.

Table 11. p53 and its putative binding sites in the Syk promoter region * at maximum dissimilarity 15%.

No.	Start site	End site	Sequence
1	-1811	-1805	GGGCATG
2	-1799	-1793	GGGCGCC
3	-1750	-1744	TGAGCCC
4	-1710	-1704	GGGCCAC
5	-1690	-1684	GGGCAAC
6	-963	-957	GCTGCCC
7	-822	-816	GGGCTCC
8	-566	-560	GGGCTAT
9	-470	-464	CGTGCCC
10	-462	-456	ATTGCCC
11	-407	-401	ACTGCCC
12(1)	-280	-274	GGGCCGC
13(2)	-255	-249	GGGCAGT
14(1)	-194	-188	GGGCAGC
15(2)	-192	-186	GCAGCCC
16(2)	-167	-161	AGGGCCC
17(2)	-166	-160	GGGCCCT
18(1)	-138	-132	CGCGCCC
19(1)	-126	-120	GGGCTCC

20[2]	-100	-94	GGGCCGA
21[1]	-87	-81	GGGCCTC
22[1]	-83	-77	CTCGCCC
23[1]	-60	-54	CCCGCCC
24[1]	-47	-41	GGGCTCA
25[1]	-39	-33	GGGCGCG
26[1]	-25	-19	GGGCGGG
27[2]	-21	-15	GGGCCAG
28[1]	-15	-9	GGGCGGC
29[1]	-7	-1	GGGCGCG
30[1]	18	24	GGGCCGC
31[1]	24	30	CCTGCCC
32[1]	31	37	GGGCGCG
33[2]	46	52	GTTGCCC
34[1]	70	76	CGGGCCC
35[1]	71	77	GGGCCCG
36[1]	124	130	GTGGCCC
37[1]	136	142	TGCGCCC
38[1]	140	146	CCCGCCC

Footnote:

(1): putative p53 binding sites overlapping with whole CpG site(s) or part of CpG site(s) in the *Syk* promoter region from -400 nt to +200 nt of *Syk* promoter region.

(2): putative p53 binding sites in the *Syk* promoter region from -400 nt to +200 nt of Syk promoter region.

Syk promoter region * represents a region from -2000 nt to +200 nt.