การหาลำดับเบสและยีนที่มีดีเอ็นเอเมทิเลชั่นที่แตกต่างกัน
ระหว่างเซลล์เม็ดเลือดขาวและสเปิร์ม

นางสาว จิรนันท์ วราชิต

IDENTIFICATION OF DIFFERENTIALLY METHYLATED SEQUENCE
AND GENE BETWEEN WHITE BLOOD CELLS AND SPERM

MISS JIRANAN WARACHIT

A Thesis Submitted in Partial Fulfillment of the Requirements

for the Degree of Master of Science in Medical Science

Program of Medical Science

Faculty of Medicine

Chulalongkorn University

Academic Year 2000

Thesis Title   IDENTIFICATION OF DIFFERENTIALLY METHYLATED SEQUENCE AND GENE BETWEEN WHITE BLOOD CELLS AND SPERM.

By   Miss Jiranan Warachit

Field of Study   Medical Science

Thesis Advisor   Associate Professor Apiwat Mutirangura

       Accepted by the Faculty of Medicine, Chulalongkorn University in Partial Fulfillment of the Requirements for the Master's Degree

---Dean of Faculty of Medicine

(Professor Pirom Kamolratanakul)

Thesis Committee

---Chairman

(Associate Professor Vilai Chintanez)

---Thesis Advisor

(Associate Professor Apiwat Mutirangura)

---Member

(Assistant Professor Montakarn Tansatit)

---Member

(Dr.Vorasuk Shotelersuk)

จิรนันท์ วราชิต : การหาลำดับเบสและยีนที่มีดีเอ็นเอเมททิเลชั่นที่แตกต่างกันระหว่างเซลล์เม็ดเลือดขาว และสเปิร์ม (IDENTIFICATION OF DIFFERENTIALLY METHYLATED SEQUENCE AND GENE BETWEEN WHITE BLOOD CELLS AND SPERM.) อ.ที่ปรึกษา : รศ.อภิวัฒน์ มุทิรางกูร. ISBN 974-346-836-6

ดีเอ็นเอเมททิเลชั่นมักจะเกิดขึ้นบนเบสไซโตซีนที่มีลักษณะเป็น CpG dinucleotide คู่เบสนี้หากอยู่รวม กันเป็นกลุ่มๆ อย่างหนาแน่นเรียกว่า CpG island ซึ่งเชื่อว่าบริเวณนี้เป็นส่วนหนึ่งของดีเอ็นเอบริเวณที่มีการแสดงออก หรือยีนนั่นเอง และจากการที่มีการศึกษาเกี่ยวกับเมททิเลชั่น พบว่าเมททิเลชั่นที่เกิดขึ้นบน CpG island มักจะมีส่วน เกี่ยวข้องกับการควบคุมการทำงานของยีนนั้นด้วยเสมอ และยังเกี่ยวข้องกับการทำงานของดีเอ็นเอในหลายๆขั้นตอน เช่น ลักษณะ X-inactivation, genomic imprinting, ลักษณะจำเพาะต่อเนื้อเยื่อ, การเจริญพัฒนาของตัวอ่อน รวมทั้ง ความผิดปกติของแบบแผนเมททิเลชั่นยังมีความเกี่ยวข้องกับการเกิดโรคมะเร็งและโรคทางพันธุกรรมบางชนิด เช่น Prader-Willi และ Angel man syndrome เป็นต้น จะเห็นได้ว่าเมททิเลชั่นมีส่วนสำคัญในหลายๆ ขบวนการของเซลล์ และใน 30,000 ยีนของ mammalian genome คาดว่ามีประมาณ 2,000 ยีน ที่ถูกควบคุมโดยดีเอ็นเอเมททิเลชั่น แต่ ในปัจจุบันมีประมาณ 40 ยีนเท่านั้นที่ถูกค้นพบและถูกศึกษาเกี่ยวกับเมททิเลชั่น ดังนั้นการศึกษานี้ตั้งขึ้นโดยมีจุด ประสงค์เพื่อจะค้นหายีนที่มีเมททิเลชั่นเกี่ยวข้องกับการทำงานเพิ่มเติมจากความรู้ที่มีอยู่ในปัจจุบัน โดยอาศัย CpG island เป็นจุดสังเกตในการเปรียบเทียบระหว่างจีโนมของเซลล์เม็ดเลือดและสเปิร์ม

ในการศึกษาครั้งนี้จะใช้เทคนิค methylation-sensitive Representational Difference Analysis (ms-RDA) ซึ่งเป็นเทคนิค RDA ที่ใช้คุณสมบัติของ methylation-sensitve restriction endonuclease (HpaII) เข้าช่วยใน การศึกษาหาความแตกต่างระหว่างสองจีโนม ซึ่ง RDA product จะถูกใส่เข้าสู่เวคเตอร์และ transform เข้าสู่ E.coli แล้วใช้ M13 primer ในการทำ PCR เพิ่มจำนวนชิ้นส่วนของดีเอ็นเอที่เราใส่เข้าไปเพื่อไปทำการทดสอบใน Southern blotting hybridization โดยจะใช้เป็น probe ไป hybridized กับดีเอ็นเอของเซลล์เม็ดเลือดและสเปิร์มที่ถูกตัดด้วยเอ็น ไซม์ HpaII และ isozyme ของมัน คือ MspI จากนั้นจะทำการหาลำดับเบสของโคลนที่ถูกคัดเลือก เพื่อนำไปเปรียบ เทียบข้อมูลของลำดับเบสที่มีอยู่กับฐานข้อมูลใน GenBank

ผลการทดลองใน 105 โคลนที่เราทำการศึกษาพบว่ามี 6 โคลนที่มีเมททิเลชั่นแตกต่างกันระหว่างเซลล์ เม็ดเลือดขาวและสเปิร์มโดยมี hypermethylation ในเซลล์เม็ดเลือดขาว จากข้อมูลของลำดับเบสมี 3 โคลนที่มี GC content มากกว่า 50 เปอร์เซ็นต์ และเป็น CpG island ตามกฎเกณฑ์ (มีความยาว 200 bp ปริมาณ GC มากกว่า 50 เปอร์เซ็นต์ สัดส่วน CpG/GpC มากกว่า 0.5) จากการเทียบข้อมูลลำดับเบสกับฐานข้อมูลโดยโปรแกรม BLAST พบว่า มี homology กับยีนของมนุษย์ 3 ยีนคือ ยีน Ribosomal DNA ยีน Niemann-Pick C1 protein (NPC1) และกับ cDNA ของมนุษย์ซึ่งอยู่บนโครโมโซม 19q13.2 และมี 1 clone ที่ไม่พบ homolohy กับ sequence ใดๆ ในฐานข้อมูล

JIRANAN WARACHIT IDENTIFICATION OF DIFFERENTIALLY METHYLATED
SEQUENCE AND GENE BETWEEN BLOOD CELLS AND SPERM THESIS ADVISOR
ASSO.PROF.APIWAT MUTIRANGURA. 110 pp. ISBN 974-346-836-6

DNA methylation of cytosine within 5'CpG islands initiated in the germ line effects control of gene expression required for X-chromosome inactivation, genomics imprinting, and cell differentiation for normal embryonic development in mammals. The aberrant methylation occurs in the process of aging and carcinogenesis. Of the 30,000 genes contained within the mammalian genome, 2,000 are estimated to be regulated by DNA methylation, although only 40 genes have been clearly identified yet. The purpose of this study was to identify methylated DNA sequences by exhibiting different patterns upon CpG island of comparison between white blood cells and sperm.

Employing a cross-sectional analytical study, we applied methylation-sensitive representational difference analysis (ms-RDA) to identify differentially methylated DNA sequences between white blood cells and sperm. The RDA product were inserted into a vector and cloned in E.coli. The DNA clones, obtained by amplified using the M13 primer, were tested as to this authenticity by Southern blot and subsequent hybridization to DNA extracted from white blood cells and sperm after treatment with methylation-sensitive restriction endonuclease, HpaII and its isozyme, MspI. We selected those hybridization products based on the autoradiogram hinted at differential methylation to direct sequencing and compare the data with the GenBank database.

From 105 clones we found 6 clones were hypermethylated in blood compared with sperm. By DNA sequencing, we found that 3 clones had a GC content>50% and satisfied the minimal criteria for CpG islands (200bp, GC content > 50%, CpG/GpC > 0.5). By analyzing the Blast program, there were 3 known human gene sequences were identical to these clones. There were cDNA sequences on 19q13.2, intergenic sequences of ribosomal DNA and 3' to exon 1 of Niemann-Pick C1 protein (NPC1) gene. However, no homology was found from the sixth.

Program Molecular Biology and Genetics   **Student's signature**……………………

Field of study   Medical Science   **Advisor's signature**……………………

Academic year   2000

ACKNOWLEDGEMENT

TABLE OF CONTENES

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES

LIST OF ABBREVIATIONS

CpG	=	Dinucleotide containing cytosine and guanine respectively, p represents the phosphate group
DNA	=	Deoxyribonucleic acid
DNTPs	=	Deoxyribonucleotide containing the base adenine, thymine, cytosine and guanine respectively
dATP	=	deoxyadeninetriphosphate
dGTP	=	deoxyguaninetriphosphate
dTTP	=	deoxythyminetriphosphate
dCTP	=	deoxycytocinetriphosphate
α-^{32}P dCTP	=	α-^{32}P deoxycytosinetriphosphates
A T C G	=	Nucleotide containing the base adenine, thymine, cytosine and guanine respectively
Bp	=	Base pair

RDA	=	Representational difference analysis method
Ms-RDA	=	Methylation sensitive representational difference analysis method
°C	=	Degree celsius
kb	=	Kilobase
mg	=	Milligram
ml	=	Millilitre
µl	=	Microlitre
µM	=	Micromolar
TE	=	Tris-ethylene diamine tetraacetic acid
rpm	=	Round per minute
SDS	=	Sodium dodecyl solphate
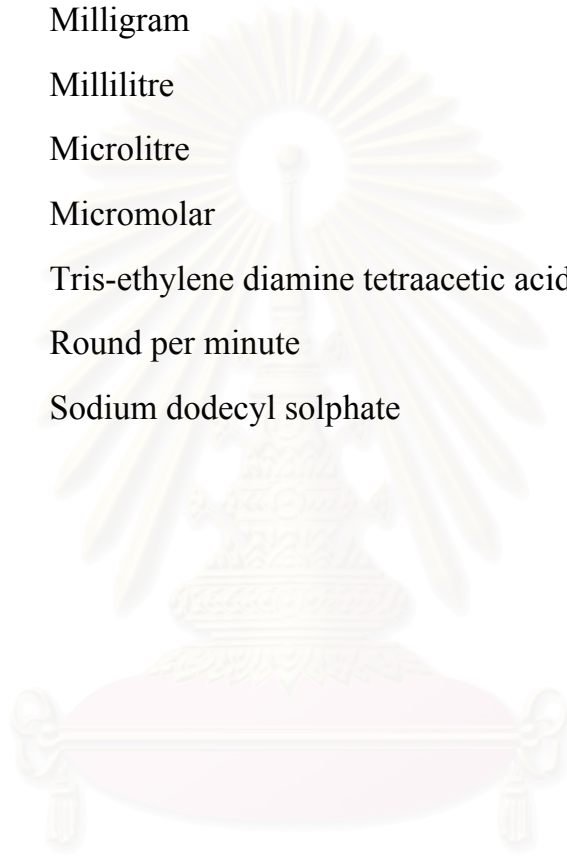
CHAPTER I

Introduction

DNA methylation is catalyzed by enzyme DNA methyltransferase at the 5 position of cytosine. DNA methyltransferase can establish methylation in two processes, *de novo*, and maintain. *De novo* methylation occurs during gametogenesis and will progress during the development of primodium germ cell to be gametes. New established methylation gives the difference in methylation level between male and female gamete, this phenomenon was known as genomic imprinting. The level of DNA methylation changes again by the process of demethylation, which reduce the levels of methylation in fertilized cell. By this reason, the levels of methylation in the early development are lesser than in the gametogenesis state. In blastula state, *de novo* methylation improves the level of methylation that may be different by cell type and its function in developmental process. Lastly, the methylation pattern in somatic cell will be maintained by a maintenance methylation through cell division. By this reason, the methylation is unequal during the lifetime, because of the different methylation pattern in the genome.

Overall, DNA methylation plays the role in many biological processes but our knowledge of methylation site in genome is limited. Only 5-15% of CpG site could be examined for methylation by using restriction endonuclease analysis. Of 30,000 genes in mammalian genome, 2,000 are estimated to be regulated by methylation. Until the present only 40 genes have been clearly identified. By this reason, this study is designed to find the DNA fragments which have the different

methylation pattern between somatic and gametic cells, represented by white blood cell and sperm respectively. The methylation sensitive RDA is a tool that can detect different methylation pattern in genome via the function of methylation sensitive enzyme. This DNA subtraction between white blood cells and sperm is expected to get some new knowledge and recover new methylated control genes.

CHAPTER II

Literature Review

1. Creation of genomic DNA methylation pattern

DNA methylation is a common eukaryotic DNA modification, especially in plants and mammals, and is one of many epigenetic (alteration in gene expression without a change in nucleotide sequence) phenomena. DNA methylation patterns are closely correlated with patterns of gene expression. In addition, heavy methylated DNA is generally associated with chromatin organization that is inhibitory to transcription (Antequera *et al.*, 1989; Tazi and Bird, 1990). Finally, this epigenetic modification is an important process for controlling normal development and tissue specific gene expression.

1.1 Distribution of methylated cytosines and gene regulation control

DNA methylation is created by enzyme DNA methyltransferase at the five position of cytosine (Figure 2-1). Methylation occurred predominantly within CpG dinucleotide (*p* represents the phosphate group), approximately $3x10^7$ sites through out the mammalian genome(Ushijima *et al.*, 1997). New methylation patterns are established by *de novo* methylation during gametogenesis and early embryogenesis (Trasler, 1998). The methylation is maintained during DNA replication by adding methyl group to the newly synthesized strand in the position symmetric to the old methylation position on the template strand.
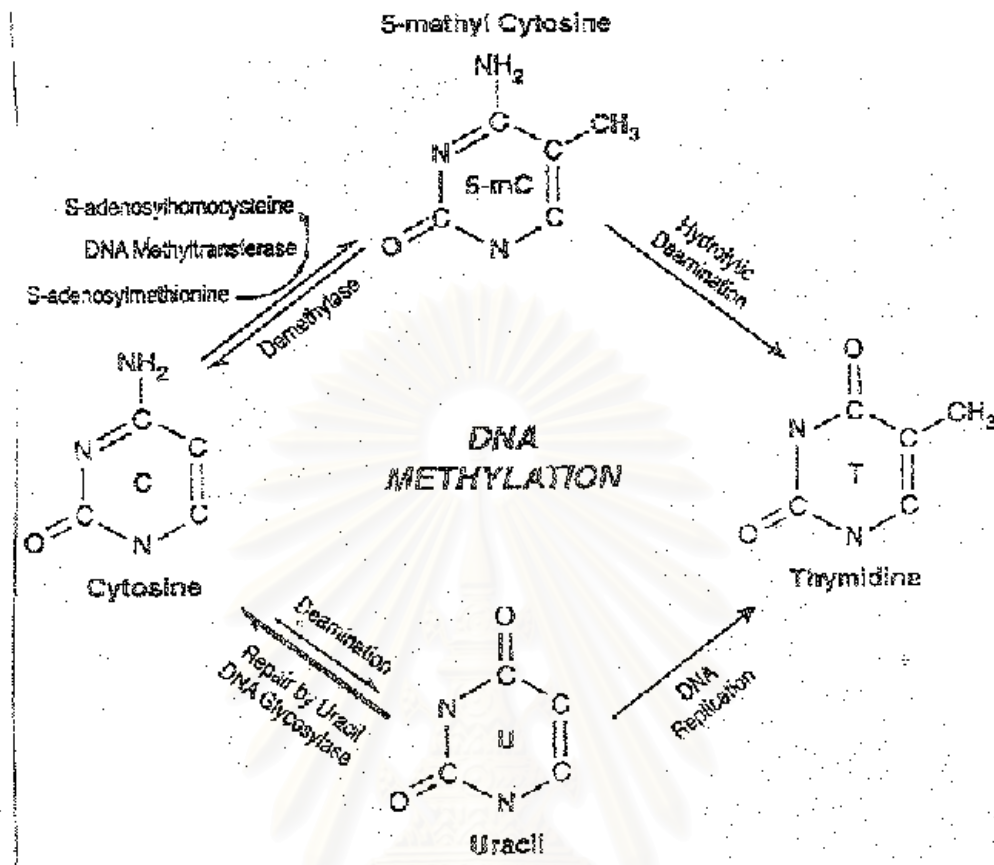
Figure 2-1 The biochemical pathways for cytosine methylation

The methylation could regulate gene expression by two main ways, firstly : the projecting methyl group into the major groove of DNA may block insertion of regulating binding protein into the binding site (major groove). For example, the addition of a methyl group to a cytosine in a *cis* element of the tyrosine aminotransferase gene of rat prevents a regulatory protein from binding to the element. Secondly, the addition of methyl group to carbon atom causes a shift of the conformational equilibrium of the DNA away from the standard B-form toward the other forms (example Z-form). The Z form becomes the lower energy state and DNA may rewind

spontaneously form B to Z conformation and remain stable in the Z-form under cellular condition. Z-form is not appropriate for the bind of regulating binding protein (Figure 2-2).
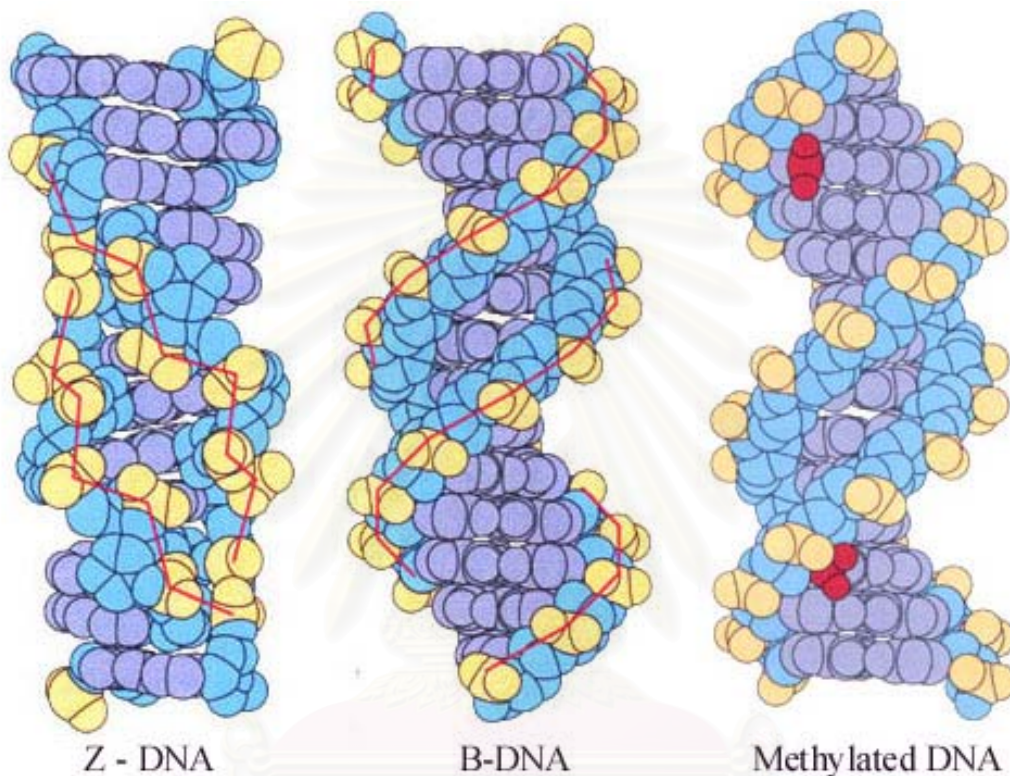
Figure 2-2 Comparison between B and Z conformation of DNA, and the position of methyl groups (red) in the major groove of B-DNA double helix.

Through the life time, DNA methylation could be passed on during DNA replication and maintained by clonal inheritance or lost either by a poorly understood demethylation process. The alter of gene expression could be the interaction of tissue specific or developmentally specific proteins. Even though the mechanism of methylation is uncertain, studies using 5-azacytidine confirm that demethylation induces gene activation. 5-azacytidine can irreversibly inhibit 5-methyltransferase. Exposure dividing

cell to this substance will inhibit the function of methyltransferase enzyme and convert many inactive genes to active forms, including many of those in the inactivate X-chromosome of mammalian female (Karp,1996).

1.2 *Methylated cytosine and CpG islands*

CpG dinucleotide is a connecting base-paired between cytosine and guanine in the same strand. Groups of the CpG dinucleotide are often clustered into CpG islands, the area of high CpG density, 0.5-5kb and found every 100 kb. The presence of CpG island is determinated base on these criterias; (1) 200 bp minimum length, (2) ratio of CpG/GpC more than 0.5, (3) GC content more than 50% (Bird *et al*., 1985; Gardiner-Garden and Frommer, 1987). In vertebrate DNA, the CpG sequence is a signal for methylation by a specific cytosine DNA methyltransferase. In higher animals, only as little as 3% to 4% of the total cytosines were methylated. Nevertheless, if there are cytosines on both sides of double stranded DNA, as much as 80% to 100% of the dinucleotide, there will be two cytosines possible to methylation. On the contrary, if only one is methylated, the doublet is said to be hemimethylated.

CpG islands are frequently associated with location of genes. In the case of genes showing widespread expression, CpG islands are almost always found at the 5' ends of genes. It occurs near transcription start sites of approximately 50% of all mammalian genes (Cross and Bird, 1995). Frequently CpG dinucleotides occur in the center of cis control elements of genes, a position in which their interference with recognition and binding site

is expected to be, and of the extending into the first exon. However, for gene which shows restricted expression patterns, the associated CpG islands are quite often found some distance downstream of the transcription initiation site (Figure 2-3). Many studies suggest that the methylated CpG on CpG islands involve gene expression control. It is estimated that there are 45,000 CpG island in human genome and approximately 50% of 70,000 genes in human are associated with the islands (Larsen et al., 1992; Antequera and Bird 1993).
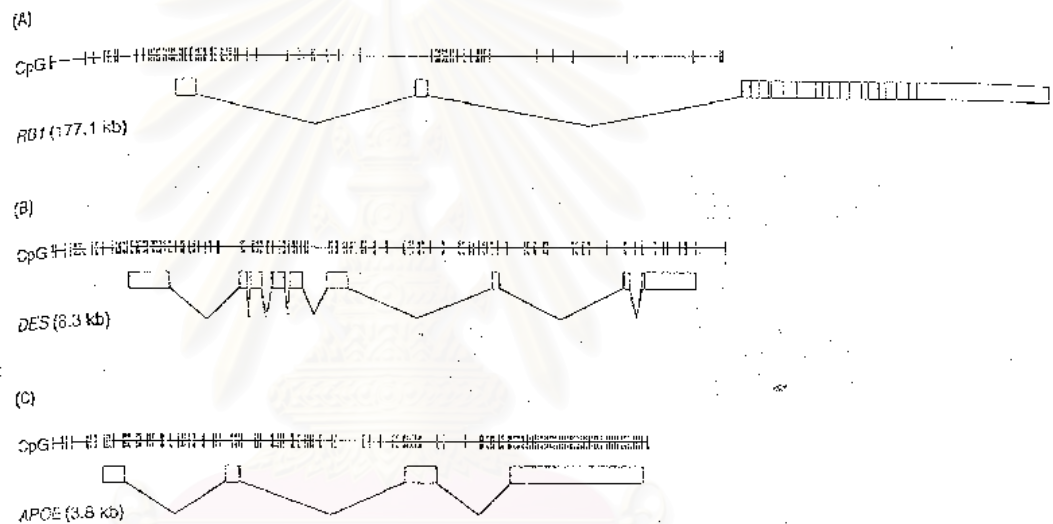
Figure 2-3 CpG island structure in three human gene. Note that CpG islands are often located at the 5'end of a gene (as in *RBI* and *DES*) but occasionally may be found at the other positions (as in the case of *APOE*)

1.3 DNA methyltransferase and level of DNA methylation

The enzyme that transfers methyl group from s-adenosyl methionine to the cytosine ring, cytosine 5-methyltransferases, or DNA methyltransferases have been characterized in the number of eukaryotes. The active mammalian DNA methyltransferase, encoded by *Dnmt1* gene, is capable of methylating both unmethylated DNA (by *de novo* methylation)

and hemimethylated DNA (by maintenance methylation). This gene is highly conserved among eukaryotes and its orthologs have been identified in various species, including human (*DNMT1*).(Singal and Ginder, 1999) The DNA methyltransferase recognized hemi-methylated DNA while replicating and add methyl group to the newly synthesized strand in the position symmetric to the old methylation position on the template strand. This process is the maintenance of methylation, the newly synthesized strand will receive the same methylation pattern as the parental DNA (Figure 2-4).
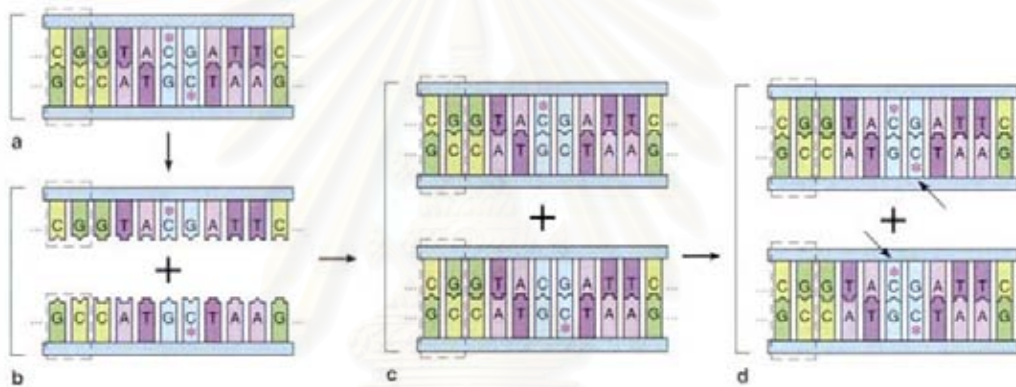
Figure 2-4 The mechanism by which methylation pattern must be passed on during DNA replication. The dashed box encloses and methylation doublet is show in blue

The high level of DNA methyltransferase expresses in male germ cells, mature oocytes and in the early embryo. Five exons of *Dnmt1* are used, but three alternate for exon-1 are employed in different cell types; somatic cells, spermatocytes and oocytes. The oocyte-specific exon is associated with the production of very large amounts of active Dnmt1 protein, which is truncated at the N terminus and sequestered in the cytoplasm during the later stages of growth. The spermatocyte-specific exon interferes with translation

and prevents production of *Dnmt1* during the crossing-over stage of male meiosis(Figure 2-5). However, there are dramatic change during gametogenesis and developing embryo. The genome of the primordial germ cells of the embryo is not methylated in any extent. *De novo* methylation gives rise the newly methylation to developing primordial germ cells, sperm and egg. Anyways, both levels and pattern of methylation are different between oocyte and sperm. In the early embryo, the wave of demethylation occurs at the preimplantation stage, morula and early blastula. But shortly afterwards, the large scale *de novo* methylation begins at the pregastrulation stage. The latter is particularly pronounced in somatic lineages, and to a lesser extent in trophoblast lineages giving rise to placenta and yolk sac, but does not occur in the primordial germ cells (Mertineit *et al*., 1998).
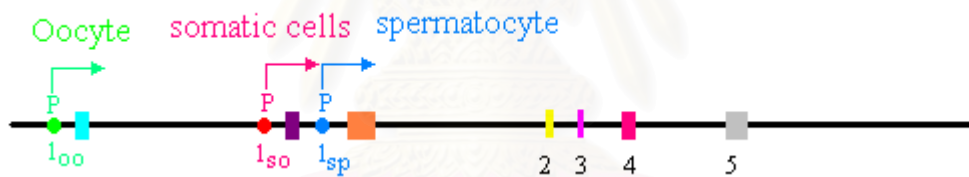
Figure 2-5 The map of *Dnmt1* methyltransferase gene show sex-specific regulation

2. <u>Role of DNA methylation and cell biology</u>

The function of DNA methylation involves in the gene expression control and plays the important role in many biological processes such as genomic imprinting, defense mechanism, X chromosome inactivation, development and cell differentiation. Furthermore much attention in the methylation field has focused on CpG islands primarily because of the propensity of such sequences to become aberrantly hypermethylated in tumors, resulting in the transcriptional silencing of the associated gene.

2.1 DNA methylation and genomic imprinting

The genomic imprinting is an unusual yet important mechanism of gene regulation by which only one of the parental copies of a gene express. The allele inherited from one parent behaves differentially than the allele derived from the other parent. The difference between the parental genomes is believed to be due to gamete specific differential modification(Figure 2-6). DNA methylation is one of the main candidates to mark either the maternal or paternal allele of certain genes for preferential expression in various tissue in the offspring. Genomic imprinting is initiated in the germ line and leads to the transcription silencing of one allele, which is usually referred as the imprinted allele. The global demethylation occuring at the early stage of primordial germ cell development also includes imprinted gene. *De novo* methylation in gametogenesis marked difference between maternal and paternal genomes. After fertilization, zygote develops to be the individual, there are functional differences between alleles and the imprint persists throughout life. Recently, there are only 26 loci in human genome found parent of origin effect (Sanford *et al*., 1987; Hoffman and Vu, 1996).

Figure 2-6 Concept of genomic imprinting.
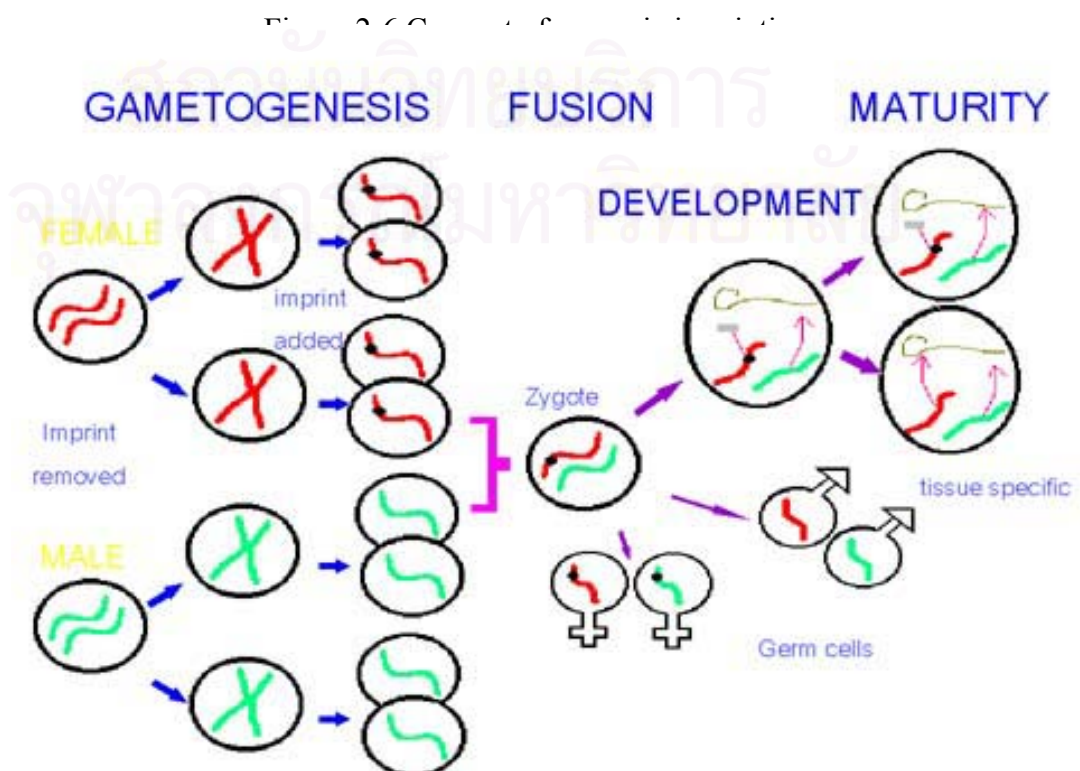
There are many mechanisms that DNA methylation controls imprinted gene, such as the antisense transcription occludes the *Igf2r* promoter (Figure 2-7), or DNA methylation promotes the binding of an imprinting factor to prevent the binding of transcriptional factor (Wutz *et al*., 1997). Imprinting is normally required for normal development. The aberrant in imprinting pattern , such as loss of imprinting, involves in the variety of diseases and cancers. Oncogenesis may also occur when the normal imprinting process is disrupted and a normally silenced imprinted gene express. The effect of imprinting may be detected in many gene deletion syndromes. For example, both Prader-Willi and Angelman syndromes are distinct disorders that are associated with deletion of the same region of chromosome 15. If the deletion occurred on the paternal allele, phenotype is Prader-Willi syndrome, whereas if maternal allele is deleted, the result is Angelman syndrome. In some cases of Prader-Willi or Angelman syndrome, there is uniparental disomy for the maternal or paternal chromosome (Yang *et al*., 1998). Presumably, there are imprinted genes in this region, and some of which are expressed on the paternal and some on the maternal. Finally, number of genes on chromosome 15q11-q13 such as *SNRPN, UBE3A* were proved to have paternal and maternal allele specific expression respectively (Hoffman and Vu, 1996).
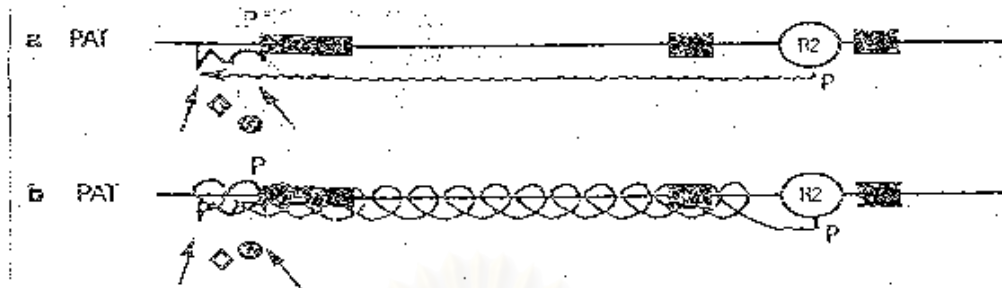
Figure 2-7 The model of paternal silencing of *Igf2r* gene

2.2 DNA methylation and X-chromosome inactivation

X chromosome inactivation is a process that occurs in all mammals, resulting in selective inactivation of allele on one of the two X chromosomes in female. It provides the mechanism of dosage compensation. Not all genes on the X chromosome are subjected to inactivation. Genes which escape X inactivation include ones that have a functional homology on the Y chromosome, and some of genes dosage do not seem to be important. In mammals, X chromosome inactivation appears to be initially controlled by a single gene, *Xist* gene, at the early stage in development (blastula stage). (Goto and Monk, 1998) The inactivation occurred randomly, either maternal or paternal X chromosome may be inactivated. X chromosome inactivation is permanent and clonally propagated. *Xist* gene, or *XIST* gene in human, encodes a mature 15kb RNA product which located on Xic (X chromosome inactivation center which is located at Xq13 in human) and uniquely encoded only by the inactive X. The accumulation of *Xist* RNA along the length of the inactive chromosome is thought to be required for the nucleation and spread of heterochromatin. Genes on the inactived chromosome have been found to be extensively methylated, suggesting that DNA methylation is

involved in the maintainance of X chromosome inactivation (Driscall and Migeon, 1990; Duthie *et al*, 1999).

2.3 DNA methylation and mismatch repair

Moreover, methylation has some important roles in mismatch repair machanism. While the replication process occurred, a mispairing that occurs between an incoming deoxyribonucleoside triphosphate and DNA template. The wrong nucleotide can be incorporated into the new DNA chain, producing mutation. The high fidelity of DNA replication depends on several "proofreading" mechanisms that act sequentially to remove errors brought about in these ways. The mismatch repair system (mismatch proofreading system) detects the distortion on the outside of their helix that results from the misfit between noncomplementary base pairs. The recognition system depends on the methylation. Bacterial DNA chains are highly methylated at all times in the cell cycle except for a brief period just after their initial assembling during replication. During this brief period, when the template chain is highly methylated and copy chain is still unmethylated, an excision repair will take place. Since the enzyme prefers unmethylated DNA, excision removing mispaired nucleotide are made in the copy rather than the template chain. Equivalent mechanisms repairing mismatched base have also been demonstrated in eukaryotes. At the moment of the basic for chain selection in eukaryotes, the mismatch repair remains unknown. However, although relatively few bases are methylated in eukaryotic DNA, the degree of methylation might also provide one sign post distinguishing between the template and copy chain. Detection of single-chain nicks may also contribute for recognition of newly synthesized chain in eukaryotes.

2.4 DNA methylation and development

DNA methylation patterns are established during gametogenesis and are very different in the genome of male and female gametes. During the gametogenesis and in developing embryo, there are dramatically change in methylation. As the germ cells begin to develop, *de novo* methylation occurred leading the high level of methylation in sperm DNA, but much lower in oogonia (Razin and Kafri, 1997). Sex specific difference in methylation patterns are found, notably at imprinted loci. Many genes were studied about the methylation status in mice gamete such as, intercisternal A particle (IAP) retroviral sequence, L_1 element and major urinary protein (MUP) sequences were unmethylated in oocyte. In contrast these sequences were methylated in spermatozoa. On the other hand, *Alu* elements, which are repetitive sequence in DNA, are methylated in oocyte and unmethylated in sperm. (Figure 2-8)
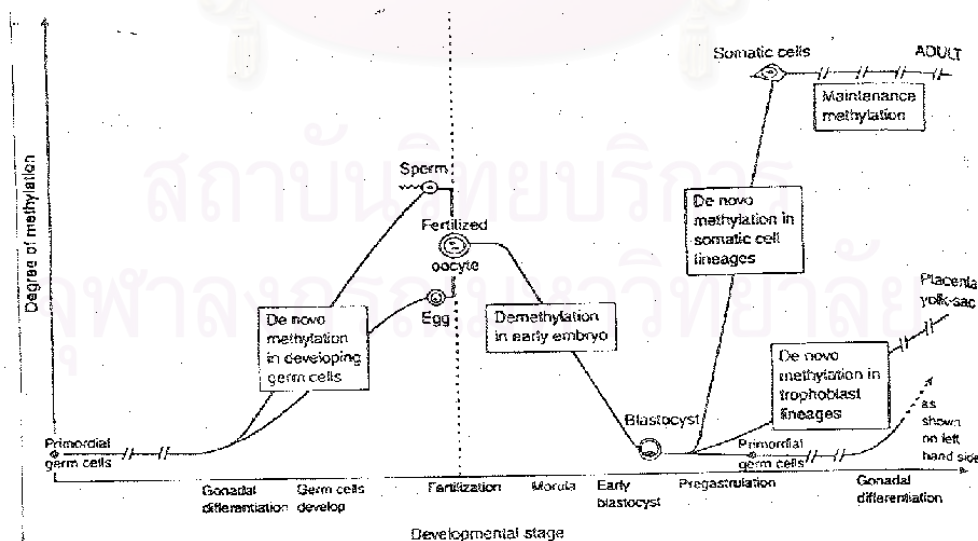
Figure 2-8 Changes in DNA methylation during mammalian development

After fertilization, in the early morula stages of development, the degree of modification represents an average of the maternal and paternal genomes. A further demethylation event may occur at the blastocyte stage and is followed by a wave of *de novo* methylation which takes place around the time of implantation. The study showed that all of the sequences described above, including L, IAP, MUP, any unmethylated sequences become methylated in the embryo by *de novo* methylation, shortly after implantation. Suggested that the dramatically change in methylation involve in the gene regulation control in each step of development, confirmed by the study of mutation in DNA methyltransferase gene. The target mutation of DNA methyltransferase gene causes homozygous to die at mid-gestation prior to germ cell development. This study prefered that the *de novo* methylation in early development stage was important for development of embryo (Mertineit *et al*.,1998).

2.5 Methylation and defense mechanism

Such parasitic DNA elements, transposible like, account for almost 40% of the human genome (exons account for about 5%) and it has been proposed that DNA methylation may have arisen as a genome-defense system to silence expression of these elements and limit their spread through the genome. The alternative theory proposed is that cytosine methylation in mammal is a nuclear host defense system that evolved primarily to counter the treads posed by endogenous parasite mobile element. Cytosine methylation inactivates the promoter of most viruses and transposon,

including the retrovirus and Alu element, and such sequences are methylated in the DNA of differentiate cell.

In both cultured cells translated with foreign DNA and transgenic organism, the newly integrated foreign DNA frequently become *de novo* methylated. It has been proposed that *de novo* constitutes a cellular defense mechanism to silence integrated foreign DNA or genes. Orend et al. have shown that, upon integration, *de novo* methylation spreads from the center of the integrated collinear viral DNA. Methylation of specific site in the adenoviral promoter results in promoter inactivation. Herpes virus also undergoes *de novo* methylation in mammalian cells. Epstein-Barr virus DNA has been found to be methylated in normal lymphocytes of healthy volunteers. Methylation-mediated inactivation of foreign gene expression in specific cell types has important therapeutic and pharmacological implication in that inhibition of methylation of therapeutically introduced genes might enhance gene therapy significantly by preventing transcriptional silencing.

2.6 Aberration of DNA Methylation and Cancer

A role of DNA methylation in oncogenesis has been hypothesized for many years. Numerous studied have suggested aberration in DNA methyltransferase activity in tumor tissue. The potential contribution of DNA methylation to oncogenesis appears to be mediated by one or more of the following mechanism; (a) DNA hypomethylation are common finding in tumorigenesis. Hypomethylation may cause the activation of proto-oncogene. A good inverse correlation between methylation and gene expression was observed in the antiapoptic *bcl-2* gene in B-cell chronic

lymphocytic leukemia and for k-*ras* proto-oncogene in lung and colon carcinomas. (b) hypermethylation of promotor sequence suggests an alternative means for the inactivation of tumor-suppressor genes in cancer. This may result from the increased DNA methyltransferase levels that have been demonstrated in various cancers such as retinoblastoma gene (*Rb*), *p16* and *p15* (Figure 2-9).
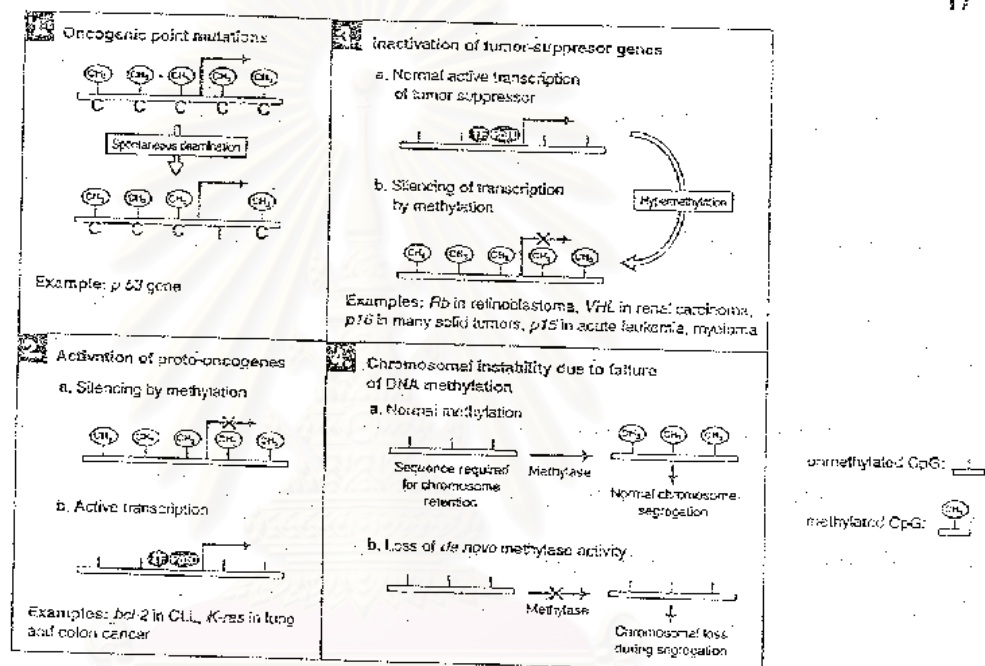
Figure 2-9   The model for the mechanism of cytosine methylation can promote oncogenesis.

3. **Background of the experiment approach**

This study aimed to find loci which have different methylation status between white blood cells and sperm. In addition, the DNAs will be further characterized their chromosome location, association with gene and methylation status. To differentiate the distinct methylation DNA fragment between leukocytes and sperms, methylation-sensitive representational difference analysis (ms-RDA) was used to subtract out DNA with the same

methylation status and select only different fragments to be cloned. Each clone would be further proved the methylation status by Southern blot using methylation sensitive restriction enzyme. Then each clones was sequenced and analyzed its homology to report sequences by BLAST family of program. Finally, tissue specific methylation pattern would be analyzed by Southern blotting hybridization of HpaII digested DNA and methylation specific PCR (MS-PCR).

3.1   Methylation-sensitive representational difference analysis (ms-RDA)

This study using methylation-sensitive difference analysis (ms-RDA) to detect different methylation status between two complex genomes, white blood cells and sperm. With this method, series of subtractive hybridization are performed using two representation of the two genomes to be compared. The representation of each genome was prepared using the methylation sensitive restriction enzyme, HpaII which recognized the four bases without methylation. We expect, by its function, HpaII to give the different product between white blood cells and sperm which performed as the tester and driver respectively. The digestion product had been ligated with a universal adapter and amplified. Restriction fragment whose size and sequence suitable for PCR amplification are enriched in the amplicon, and the other fragment remain unamplified. This step reduce the complexity of the genome and is essential for the efficient subtractive hybridization. The excessive reduction of the complexity will result in the loss of target.

The mixture ratio of tester and driver DNAs was optimized to detect the differences in the methylation status of the single copy per diploid

genome. Before each round, all the tester DNA fragments are ligated to nonphosphorylated oligonucleotide adapters so that they have a long oligonucleotide at their 5'ends. Each round starts by mixing the tester DNA sample with the large excess of driver DNA samples followed by denaturation by heat or alkali to form separate strands and reformation of hybrid double helices by reannealing the complementary strands. After filling in the oligonucleotide cohesive ends with *Taq* DNA polymerase by using the long nucleotide as the primer. The target fragments are predominantly self-reannealling and form homoduplexes with oligonucleotide on the both ends. Thus the target DNA fragments are selectively amplified. At the same time, the single strand of non-target tester fragment predominantly form heteroduplexes (hybrids) with driver DNA fragments. The tester-driver hybrids have the primer sequences on one end only, so they fail to participate in the exponential amplification and are therefore subtracted. After all, we subclone the difference product into bacterial cells for the further study in characteristic of DNA methylation(Figure 2-10). (Lisitsyn *et al*., 1993; Lisitsyn 1995; Ushijima et al., 1997; Toyota et al., 1999)

Tester

Driver (in excess)

Denaturation and reassociation

Target

Hybrids

Driver
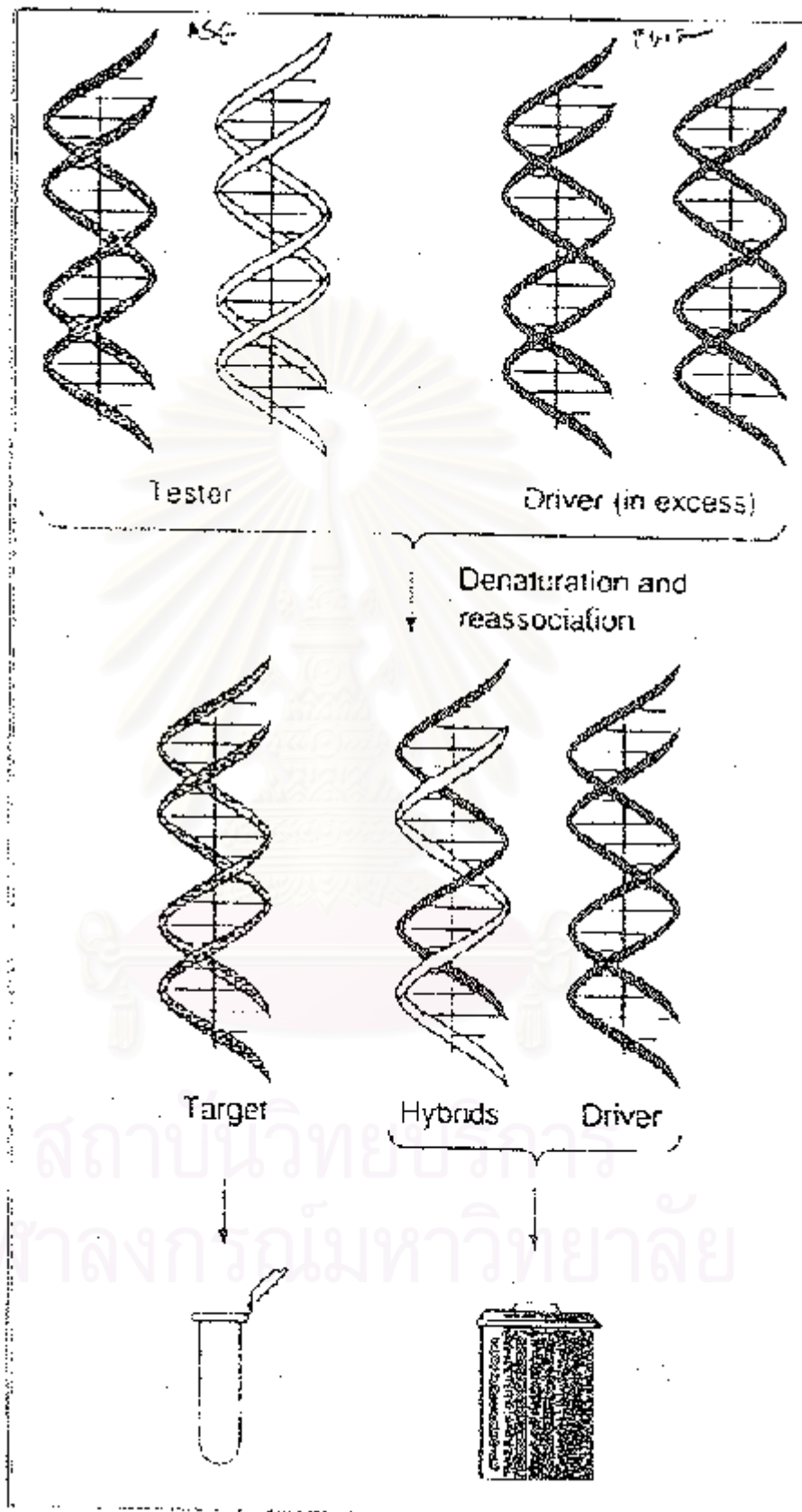
Figure 2-10 The outline of methylation-sensitive representational different analysis in brief.

3.2 Southern blotting hybridization

The Southern blotting hybridization couple with the function of methylation sensitive restriction enzyme were used for the examination of the different methylation status between white blood cells compared with sperm and the methylation status in various tissues. The Southern blotting hybridization was designed in 1975 by E.M.Southern for the propose of the study in base sequence homology between DNA and DNA.

In the Southern blotting hybridization process, the DNA samples, white blood cell and sperm DNA, were digested with the restriction endonuclease. In this study we used HpaII and MspI. Both of them have the same four bases recognition sites, but HpaII did not cut if there is methylation within the position. On the other hand, MspI cut every recognition site. By this reason, we used MspI digested DNA as the control of unmethylated DNA. Digested DNA will be separated by electrophoresis. Then the DNA were transferred to nitrocellulose membrane. The hybridization using radiation labeled DNA clone was performed. After all, the position of annealing was detected by exposure with X-ray film.

By the hypothesis that there are methylation status difference between white blood cell and sperm, we expect the methylation sensitive enzyme will give the different digested DNA fragment between two genome. Digested DNA fragment had been sized fractionated by gel electrophoresis. During the electrophorase, DNA, which are negatively

charge, are repelled from the negative electrode towards the positive electrode. The migration speed depends on fragment length. After hybridization with labeled DNA clone, the fragment which had complementary sequence can be detected by applying on X-ray film. As the result, the different methylation status fragment will perform the different band position in fractionated white blood cell DNA when compare with sperm DNA. The unmethylated DNA fragment will perform the same band position when compared with MspI digested DNA.

3.3 BLAST family of program

Sequence of each clone was analyzed its homology and reported the sequences by BLAST (Basic Local Alignment Search Tool) family of program. They are designed to explore all of the available sequence databases regardless of whether the query is protein or DNA. In short, users input either a nucleotide or amino acid query sequence, and search a nucleotide or amino acid sequence database. The score assigned in a BLAST search have a well-defined statistical interpretation, marking real matches easier to distinguish from random background hits. BLAST uses a heuristic algorithm which seeks local as opposed to global alignments and is therefore able to detect relationships among sequences which share only isolated regions of similarity.

3.4 The methylation specific PCR (MS-PCR)

The tissue-specific methylation pattern was analyzed by using Southern blotting hybridization and methylation specific PCR (MS-PCR). The MS-PCR was used in the DNA sample because it is the simple, rapid

and inexpensive method for determination of methylation patterns from very small samples of DNA. This method is combined with the function of bisulfite modification and PCR.

Chemical modification of cytosine to uracil by bisulfite treatment is the basis of this method. Under the appropriate conditions, cytosines in the DNA sample are converted to uracil. However, cytosines that are methylated (5-methylcytosine) are resistant to this modification and remain cytosines. The primer had been designed for the modified DNA. After amplification they are digested with the restriction endonuclease, which is chosen by mentioning in the remaining cytosine. The amplified DNA with remain cytosine on the restriction site have been digested, that mean on that site is methylated. On the other hand, the undigested sample means that there are unmethylated cytosine on the restriction site. Amplifying and sequencing of this modified DNA provides detailed information of the methylation status of all CpG sites within the amplification region.

CHAPTER III

Material and Equipment

1. Specimens

The specimen used in this study were consisted of both peripheral blood leukocytes and fresh sperm from five donors.

2. Material

2.1 Pipette tip : 10 μl, 100 μl, 1,000 μl (Elkay, USA)

2.2 Microcentrifuge tube : 0.2 ml, 0.5 ml, 1.5 ml (Bio-rad, Elkay, USA)

2.3 Polypropylene conical centrifuge tube : 15 ml (Elkay, USA)

2.4 Falcon 2059 polypopylene tube : 15 ml (Elkay, USA)

2.5 Screw cap microcentrifuge tube : 2 ml (Elkay, USA)

2.6 Petridise : 15x100 mm (Nunc, Denmark)

2.7 Beaker : 50 ml, 100 ml, 200 ml, 500 ml, 1,000 ml (Pyrex)

2.8 Flask : 250 ml, 500 ml, 1,000 ml (Pyrex)

2.9 Reagent bottle : 100 ml, 250 ml, 500 ml, 1,000 ml (Duran, USA)

2.10   Cylinder : 25 ml, 50 ml, 100 ml, 250 ml, 500 ml, 1,000 ml (Witeg, Germany)

2.11   Glass hybridization tube : 100 ml (Witeg, Germany)

2.12   Glass pipette : 5 ml, 10 ml (Witeg, Germany)

2.13   Pipette rack (Autopack, USA)

2.14   Microcentrifuge Tube rack (USA/ Scientific plastics)

2.15   Thermometer (Precision, Germany)

2.16   Parafilm (American National Can, USA)

2.17   Polaroid film (Polaroid, UK)

2.18   3 MM Chromatography paper (Whatman, UK)

2.19   Plastic wrap

2.20   Stirring-magnetic bar

2.21   PhosphoImager (Molecular Dynamics, USA)

2.22   Surgical bladder (Surgeon)

2.23   Nylon tranfer membrane (Amersham, England)

2.24   Alcohol lamp

3. Equipment

3.1 Automatic adjustable micropipette : P2 (0.1-2μl), P10 (0.5-10μl), P20 (5-20 μl), P100 (20-100 μl), P1000 (0.1-1 ml) (Gilson, France)

3.2 Pipette boy (Tecnomara, Switzerland)

3.3 X-ray film cassette (Kodak, USA)

3.4 Vortax (Scientific Industry, USA)

3.5 pH meter (EutechCybernetics)

3.6 Stirring hot plate (Bamstead/Thermolyne, USA)

3.7 Balance (Precisa, Switzerland)

3.8 Microcentrifuge (Fotodyne, USA)

3.9 DNA Thermal cycler 480 (Perkin Elmer, Cetus USA)

3.10   Thermal cycler (Omnigene, Hybraid USA)

3.11   Hybridization oven/ shaker (Stuart Scientiic, USA)

3.12   Power supply model 250 (Gibco BRL, Scothland)

3.13   Horizon 11-14,(Gibco BRL, Scothland)

3.14   Heat block (Bockel, Germany)

3.15   Incubator (Memmert)

3.16   Thermostat shaking-water bath (Heto, Denmark)

3.17   Liquid nitrogen tank (Minnesota Valley engineering, USA)

3.18   Spectronic spectrophotometers (Genesys5, Milton Roy USA)

3.19   UV Transilluminator (Foto/prep, Fotodyne USA)

3.20   UV-absorbing face shield (Spectronics, USA)

3.21   Polaroid camera (Fotodyne, USA)

3.22   Radiation safety shieling-screen (C.B.S. Scientific, USA)

3.23   Beta microcentrifuge tube racks (C.B.S. Scientific, USA)

3.24   Beta waste safes (C.B.S. Scientific, USA)

3.25   Refrigerator 4 ^{o}C (Misubishi, Japan)

3.26   Deep freeze -20 ^{o}C, -80 ^{o}C (Revco)

3.27   Water Purification equipment (Water pro Ps, Labconco USA)

3.28   Molecular Dynamics Storm 840 (Molecular Dynamics, USA)

3.29   Image Erase (Molecular Dynamics, USA)

3.30   Water bath

3.31   AB Prism 377 DNA sequencer (Perkin Elmer)

4. Reagent

4.1   General reagent

4.1.1   Absolute ethanol (Merck)

4.1.2   Absolute methanol (Merck)

4.1.3   Acetic acid (Merck)

4.1.4   Agar (Gibco BRL)

4.1.5   Agarose, molecular glade (Promega)

4.1.6   Ammonium persulfate (Merck)

4.1.7   Ammonium acetate (Merck)

4.1.8   Boric acid (Merck)

4.1.9   Bromphenol blue (Pharmacia)

4.1.10   Chloramphenical (Gibco BRL)

4.1.11   Dimethyl formamide

4.1.12 Disodium   ethylenediamine   tetracetic   acid : EDTA

   Merck)

4.1.13 Denhardt's solution (Amresco)

4.1.14 Ethidium bromide (Gibco BRL)

4.1.15 Ficoll (Pharmacia)

4.1.16 Formamide (Pharmacia)

4.1.17 Glucose (Merck)

4.1.18 Glyceral (Merck)

4.1.19 Hydrochloric acid (Merck)

4.1.20 IPTG

4.1.21 Magnesium chloride (Merck)

4.1.22 Mineral oil (Sigma)

4.1.23 Phenol-Chloroform-Isoamyl alcohol (Sigma)

4.1.24 Potassium chloride (Merck)

4.1.25 Selmon sperm

4.1.26 Sodium acetate (Merck)

4.1.27 Sodium chloride (Merck)

4.1.28 Sodium dodecyl sulfate (Sigma)

4.1.29 Sodium hydroxide (Merck)

4.1.30 Sodium phosphate (Merck)

4.1.31 Sucrose (BDH)

4.1.32 Tris base (USB)

4.1.33 Tryptone (Gibco BRL)

4.1.34 X-gal

4.1.35 Yeast extract (Gibco BRL)

4.2 Reagent kits

4.2.1 PCR kit (Perkin Elmer Cetus)

4.2.2 StrataPrep PCR Purification kit. (Stratagene)

4.2.3 PCR-Script^{TM} Cam cloning kit (Stratagene)

4.2.4   Nick translation kit (Roache)

4.2.5   Big Dye terminator cycle sequencing ready reaction kit

4.3 Enzymes

4.3.1 *Taq* polymerase (Perkin Elmer)

4.3.2 Proteinase K (Amresco)

4.3.3 HpaII (Biolab)

4.3.4 MspI (Biolab)

4.3.5   DpnII (Biolab)

5. **Radioactive radivue**

α-^{32}P dCTP (Amersham phamasia biotech)

CHAPTER IV

Methods

1. Sampling of specimens

The specimen used in this study were consisted of both peripheral blood leukocytes and fresh semen from five donors.

2. DNA extraction

2.1   *Peripheral blood leukocyte* : The extraction of DNA from peripheral blood leukocyte was performed as followed :

2.1.1 5-10 ml of whole blood is centrifuged for 10 minutes at 3,000 rpm.

2.1.2 Remove supernatant and collect buffy coat to a new polypropylene tube. Then add 10 volumes of cold lysis buffer1 (or 10 ml), mix thoroughly and incubate at -20°C for 5 minutes.

2.1.3 Centrifuge for 8 minutes at 1,000 g, then remove supernatant.

2.1.4 Add 3 ml cold lysis buffer1, mix thoroughly and centrifuge for 8 minute at 1,000g.

2.1.5 Discard supernatant afterward add 900μl lysis buffer2, 10 μl Proteinase K solution (20 mg Proteinase K in 1.0 ml of 1% SDS-2mM EDTA, should be prepare 30 min before use.), and 10%SDS 50μl. Mix vigorously for 15 seconds.

2.1.6 Incubate the tube(s) in 55 $^{\circ}$C shaking waterbath overnight for complete digestion.

2.1.7 Add 1 ml phenol-chloroform-isoamyl alcohol shake vigorously for 15 seconds and centrifuge at 6,000 rpm for 5 minutes.

2.1.8 Transfer the supernatant from each tube (contains DNA) to a new microcentrifuge tube.

2.1.9 Add 0.5 volumes of 7.5M CH_3COONH_4 and 1 volume of 100% ethanol, mix by inversion. The DNA should immediately form a stringy precipitate. Recover the DNA by centrifugation at 14,000 rpm for 15 minutes. Then remove supernatant.

2.1.10 Rinse the pellet with 70% ethanol. Decant the ethanol and air dried the pellet. (It is important to rinse well to remove any residual salt and phenol.)

2.1.11 Resuspend the digested DNA in 20-300 µl of the double distilled water at $37\,^{\circ}C$ until dissolved.

2.2 *Sperm* : The extraction of DNA from sperm was performed as followed :

2.2.1 5 ml of semen is centrifuge 8 minutes at 3,300 rpm.

2.2.2 Remove supernatant and add 2 volumes of PBS buffer. Then centrifuge 8 minutes at 3,300 rpm. Repeat this step at least twice.

2.2.3 Discard supernatant afterward add 1 volume of lysis buffer2, and 1/10 volume Proteinase K solution (20 mg Proteinase K in 1.0 ml of 1% SDS-2mM EDTA, should be prepare 30 min before use.), 10%SDS 50µl. Mix vigorously for 15 seconds

2.2.4 Incubate the tube(s) in 55 °C shaking waterbath overnight for complete digestion.

2.2.5 Add 1 ml phenol-chloroform-isoamyl alcohol shake vigorously for 15 seconds and centrifuge at 6,000 rpm for 5 minutes.

2.2.6 Transfer the supernatant from each tube (contains DNA) to a new microcentrifuge tube.

2.2.7 Add 0.5 volumes of 7.5M CH_3COONH_4 and 1 volume of 100% ethanol, mix by inversion. The DNA should immediately form a stringy precipitate. Recover the DNA by centrifugation at 14,000 rpm for 15 minutes. Then remove supernatant.

2.2.8 Rinse the pellet with 70% ethanol. Decant the ethanol and air dried the pellet. (It is important to rinse well to remove any residual salt and phenol.)

2.2.9 Resuspend the digested DNA in 20-300 μl of the double distilled water at 37 °C until dissolved.

2.3 *Various tissues* : The extraction of DNA from various tissues (heart, brain, lung, liver, stomach, spleen, kidney and bone marrow) was performed as followed :

2.3.1 Used 1.2 ml of digestion buffer (Lysis buffer2 and 1/10 volume of Proteinase K solution; 20 mg Proteinase K in 1.0 ml of 1% SDS-2mM EDTA, should be prepare 30 min before use) per 100 mg of tissue (wash tissue with PBS buffer to remove residual blood).

2.3.2 Incubate the tube(s) in 55 ^{o}C shaking waterbath overnight for complete digestion.

2.3.3 Add 1 volume of phenol-chloroform-isoamyl alcohol shake vigorously for 15 seconds and centrifuge at 6,000 rpm for 5 minutes.

2.3.4 Transfer the supernatant from each tube (contains DNA) to a new microcentrifuge tube.

2.3.5 Add 0.5 volumes of 7.5M CH_3COONH_4 and 2 volumes of 100% ethanol, mix by inversion. The DNA should immediately form a stringy precipitate. Recover the DNA by centrifugation at 14,000 rpm for 15 minutes. Then remove supernatant.

2.3.6 Rinse the pellet with 70% ethanol. Decant the ethanol and air dried the pellet. (It is important to rinse well to remove any residual salt and phenol.)

2.3.7 Resuspend the digested DNA in 20-300 μl of the double distilled water at 65 ^{o}C until dissolved.

3. Calculation of DNA concentration

The reading at 260 nm is used for calculating the concentration of nucleic acid of the samples. An OD of 1 corresponds to approximately 50 μg/ml for double-strand DNA. Therefore DNA concentration is calculated from the following :

DNA concentration = OD x 50 x dilution ratio (μg/ml)

4. Methylation-sensitive Representational Different Analysis (ms-RDA)

4.1 *Preparation of tester and driver amplicom*

4.1.1 Digestion of genomic DNA of the tester and the driver was performed as followed :

    a. Add 1x of digest buffer2 and 10U of HpaII (used 5U per 1μg DNA) to 2μg of white blood cell or sperm genomic DNA. Then incubate at 37°C for 3-4 hours.

    b. Extract with phenol-chlorophrome-isoamyl alcohol and precipitate with 70% ethanol.

    c. Dry the pellet and dissolve it in 30μl of TE

4.1.2 Ligation of the R-Hpa adapter and digested DNA was performed as followed :

    a. Ligase R-Hpa adapter to digested DNA by mix the following : 1.2μg digested DNA, 500ρmole of desalted R-Hpa24 oligonucleotide and desalted R-Hpa12 oligonucleotide, 1x ligase buffer then bring the volume to 60μl.

    b. Anneal oligo in PCR machine at 50°C for 1 minute, and cooling to 10°C at 1°C /minute.

    c. Add 1200U T$_4$DNA ligase, and incubate overnight at 16°C.

4.1.3 Amplification of the tester and the driver amplicon was performed as the followed :

    a. Dilute ligations by adding 200μl TE

    b. The PCR reaction was performed in a total volume of 200μl using 4μl diluted ligation in 1x PCR buffer (335mM Tris HCl, pH8.8 at 25 °C; 20mM MgCl$_2$; 80mM (NH$_4$)$_2$SO$_4$; 166μg/μl BSA), 340μM each of

deoxynucleotide triphosphates (dNTPs), 100ρmole R-Hpa24 primer.

    c. In the multiplex PCR reaction, the initial denaturation step was 72°C for 3 minutes then add 5U *Taq* DNA polymerase. Afterwards, incubate 5 minutes at 72°C the followed by 30 cycles of denaturation at 95°C for 1 minute, annealing and extension at 72°C for 3 minutes and a final extension at 72°C for 10 minutes. At this point, check 5μl on a 1.3% agarose gel (a smear ranging in size from ~1.5-0.2kb should be seen)

4.2 First round RDA

    4.2.1 Generation of representation was performed as followed:

        a. Combine 4 reactions into 1.5ml microcentrifuge tube and extract twice with a volume of phenol-chloroform-isoamyl alcohol and once with a volume of chloroform-isoamyl alcohol.

        b. Add 2/3 volume of 5M NH_4Oac, vortex. Then add 2 volumes of 100% ethanol and store at -20°C for 20 minutes.

        c. Centrifuge at 14,000 rpm for 15 minutes, and wash the pellet with 70% ethanol. Dry the pellet and resuspend amplicon to 0.5μg/μl with TE.

    4.2.2 Restriction of representation was performed as followed :

a. Digest 300µg and 20µg of driver amplicon and tester amplicon respectively with HpaII (used 5U per 1µg DNA) at 37°C for 4 hours.

b. Preparation of the driver performed as followed :

    i. Extract digested driver amplicon twice with a volume of phenol-chloroform and once with a volume of chloroform.

    ii. Afterwards, add 1/10 volume of NaOAc (pH5.3), 700µl isopropanol, and precipitate on ice for 20 minutes.

    iii. Centrifuge at 14,000 rpm for 14 minutes at 4°C. Wash the pellet with 70% ethanol and resuspend at 0.5µg/µl.

c. Preparation of tester performed as the followed :

    i. Separate digested tester amplicon with 1.2% agarose gel electrophoresis.

    ii. Excise amplicon-containing gel slice (leaving behind the digested linkers), place in the falcon tube and weigh.

    iii. Add 3 volumes of QX1 buffer and 30µl of QIAEXII, mix thoroughly. Then incubate at 50°C for 10 minutes by mix thoroughly every 2 minutes.

    iv. Centrifuge at high speed for 30 seconds and remove supernatant.

    v. Wash the pellet with 500µl of QX1 buffer then wash twice with 500µl of PE buffer.

Air dry the pellet and resuspend in 120μl ddH$_2$O.

vi. Ligase J-Hpa adapter to purified tester by mixed the following: 1μg purified tester DNA, 500 ρmole each of desalted J-Hpa24 oligonucleotide and desalted J-Hpa12 oligonucleotide, 1x ligase buffer then bring the volume to 60μl.

vii. Anneal oligo in PCR machine at 50°C for a minute then cooling to 10°C at 1°c per minute.

viii. Add 1200U T$_4$ ligase, and incubate overnight at 16°C.

ix. Dilute ligated J-oligo tester to approximately 10ng/μl by adding 120μl TE.

4.2.3 Subtractive hybridization was performed as followed :

a. Mixed 40μg of driver with 200ng diluted J-ligated tester.

b. Extract once with phenol-chloroform then precipitate with 30μl of 10M NH$_4$Oac and 380μl 100% ethanol.

c. Incubate at -70°C for 10 minutes. Afterwards, incubate at 37°C for a minute and centrifuge at 14,000 rpm for 15 minutes.

d. Wash the pellet twice with 70% ethanol, air dry and resuspend thoroughly in EEX3 buffer (30mM EPPS, pH8.0 at 20°C; 3mM EDTA) by pipetting

for at least 2 minutes then incubate at 37°C for 5 minutes.

e. Vortexing and spinning to bottom of the tube and overlay with 35µl of mineral oil.

f. Denature at 98°C for 5 minutes in PCR machine and cool to 67°C then immediately add 1µl of 5M NaCl directly into DNA.

g. Incubate at 67°C for 20 hours to allow complete hybridization.

4.2.4 Generation of first difference product was performed as the followed :

a. Remove as much mineral oil as possible. Then dilute with 8µl of TE containing 5µg/µl yeast RNA, pipette vigorously.

b. Add 25µl TE, pipette vigorously. Follow with 362µl TE and vortexing.

c. The PCR reaction was performed in a total volume of 200µl using 20µl diluted hybridization mix in 1x PCR buffer, 4mM each of deoxynucleotide triphosphate (dNTPs). Set up 4 reactions for each subtraction.

d. In the multiplex PCR reaction, the initial step was 72°C for 3 minutes to melt away 12 mer. Then pause and add 5U of *Taq* DNA polymerase. Afterwards, incubate 5 minutes at 72°C, then add 250ρmole of J-Hpa24 and followed by 12 cycles of denaturation at 95°C for a minute, annealing

and extension at 70°C for 3 minutes (when generating second difference product, perform anneal and extension at 72°C) and a final extension at 72°C for 10 minutes. Then cool to room temperature.

e. Combine the 4 reactions into 1.5ml microcentrifuge tube. Extract twice with phenol-chloroform-isoamyl alcohol and once with chloroform-isoamyl alcohol.

f. Add 2μg glycogen, 75μl of 3M NaOAc (pH5.3), 800μl isopropanol and precipitate on ice for 20 minutes.

g. Centrifuge 14,000 rpm for 14 minutes at 4°C then wash the pellet with 70% ethanol and resuspend the pellet in 40μl of 0.2x TE.

h. Digest PCR product with Mung Bean Nuclease by using 20μl DNA in 1x Mung Bean Nuclease Buffer (NEB), 20U Mung Bean Nuclease (MBN). Then incubate 30°C for 35 minutes.

i. Stop reaction by adding 160μl of 50mM Tris.HCl (pH8.9) and incubating at 98°C for 5 minutes then chill on ice.

j. During MBN incubation, set up final PCR reaction on ice (4 reactions for each subtraction) by mixing 120μl ddH$_2$O, 1x PCR buffer, 4mM each of deoxynucleotide triphosphate (dNTPs), 250ρmole

J-Hpa24. Afterwards, add 20μl MBN treated DNA.

k. In the multiplex PCR reaction, the initial denaturation step was 95°C for 1 minute then cooling to 80°C and add 5U *Taq* DNA polymerase. Then followed by 30 cycles of denaturation at 95°C for 1 minute, annealing and extension at 70°C for 3 minutes and a final extension at 72°C for 10 minute and cool to 4°C.

l. Combine 4 reactions and extract twice with phenol-chloroform-isoamyl alcohol and once with chloroform-isoamyl alcohol.

m. Add 75μl of 3M NaOAc (pH5.3), 800μl isopropanol and precipitate on ice for 20 minutes.

n. Centrifuge at 14,000 rpm for 14 minute then wash the pellet with 70% ethanol and resuspend the pellet in 100μl TE (at approximate 0.5 μg/μl) This is the First Difference Product (DP1).

4.3 Second and third round RDA

Digest 20μg DP1 with HpaII and purified. Then ligase digested Dp1 (diluted 1:10 with TE) with J-Hpa adapter and repeat subtractive hybridization step and generation of second difference product.

Repeat previous step again but change adapter to N-Hpa adapter. The third difference product will be clone.

5. **Cloning**

5.1 *Purifying the PCR products with the StrataPrep PCR Purification Kit*

> 5.1.1 Add a volume of DNA binding solution equal to the volume of the aqueous of the PCR product to the microcentrifuge tube and mix thoroughly.
>
> 5.1.2 Transfer the solution mixture to the microspin cup that is seated in a 2 ml receptacle tube. Snap the cap of the receptacle tube onto the top of the microspin cup.
>
> 5.1.3 Spin the tube at maximum speed for 30 second.. Open the cap of the receptacle tube, remove and retain the microspin cup and discard the DNA-binding solution.
>
> 5.1.4 Add 750 μl of 1xPCR wash buffer (5x PCR wash buffer, 20ml of 100% ethanol) to the microspin cup.
>
> 5.1.5 Spin the tube at maximum speed for 30 seconds, remove and retain the microspin cup, and discard the wash buffer.
>
> 5.1.6 Place the microspin cap back in the receptacle tube. Spin the tube at maximum speed for 30 seconds, make sure that all of wash buffer is remove from the microspin cap.
>
> 5.1.7 Transfer the micro spin cap to a fresh 1.5ml microcentrifuge tube. Add 50 μl of elution buffer directly onto the top of the fiber matrix at the bottom of the microspin cup.
>
> 5.1.8 Incubate the tube at room temperature for 5 minutes. Spin the tube in a microcentrifuge at maximum speed for 30 seconds, discard the microspin cup

5.2 *Polishing the purified products*

5.2.1 The polishing reaction was performed by using 10μl of the purified PCR product, 1μl of 10mM dNTP mix (2.5 mM each), 1.3μl of 10x polishing buffer and 0.5U of cloned *Pfu* DNA polymerase.

5.2.2 Mix the polishing reaction gently and add a 20μl mineral oil overlay. Incubate the polishing reaction for 30 minutes at 72°C in a water bath.

5.2.3 Store the polished PCR product at 4°C until use.

5.3 *Inserting of the PCR products*

Insert the PCR product into the pPCR-Script Cam SK(+) Cloning Vector by add 200ng of polished PCR product directly to the cloning reaction.

5.4 *Ligation the insert:*

5.4.1 The ligation reaction was perform in the total volume of 10μl using 1μl of the pPCR-Script Cam SK(+) cloning vector (10ng/μl), 1μl of PCR-Script 10x reaction buffer, 0.5 μl of 10mM rATP, 4μl of the blunt-ended PCR product, 1μl of *SrfI* restriction enzyme (5U/μl), 1μl of T4 DNA ligase (4U/μl).

5.4.2 Mix the ligation reaction gently and incubate 1 hour at room temperature. Then heat the ligation reaction for 10 minutes at 65°C. Store the ligation reaction on ice until use for transformation.

5.5 *Transformation into the Epicurian Coli XL10-Gold Kan ultracompletent cells:*

5.5.1 Thaw the XL10-Gold Kan ultracompletent cell on ice.

5.5.2 Gently mix the cells by hand, and aliquot 40μl of into a 15ml Falcon 2059 polypropylene tube for each of the following reactions: the experimental ligation reaction, the ligation reaction containing the PCR test insert, and the pUC18 control plasmid.

5.5.3 Add 1.6 15ml Falcon 2059 polypropylene tube of the XL10-Gold β-mercaptoethanol mix provided with the kit to the 4015ml Falcon 2059 polypropylene tube of bacteria. Swirl the contents of the tube gently. Incubate the cells on ice for 10 minutes, swirling gently every 12 minutes.

5.5.4 Add 215ml Falcon 2059 polypropylene tube of the cloning reaction (the ligation reaction) to the transformation reaction and swirl the reaction gently.

5.5.5 Incubate the tubes on ice for 30 minutes. Heat pulse the tubes in a 42°C water bath for 30 seconds. (The duration of the heat pulse is *criteria* for obtaining the highest efficiencies. Do not exceed 42°C)

5.5.6 Incubate the tubes on ice for 2 minutes. Add 0.45ml of preheated (42°C) NZY$^+$ broth to each tube and incubate the tubes at 37°C for 1 hour with shaking at 225-250 rpm.

6. Selected colony

Plate LB-chloramphenical agar plate (LB agar with 50μg/ml Chloramphenical which had spread with 100μl of 10mM IPTG and 100μl of 2% X-gal 30 minutes prior to plating transformations) with 200 μl

transformations solution (1:10,0000 dilution) using a sterile spreader. Incubate at 37°C overnight, then pick up a single white colony on agar plate to LB-chloramphenical broth in 15ml Falcon 2059 polypropylene tube. Incubate at 37°C overnight. Centrifuge at 3000 rpm and remove supernatant. Then add 40μl PBS buffer to resuspend the pellet. Pipette thoroughly and remove the solution to a microcentrifuge tube with 40μl PBS buffer. Keep bacterial clone as the bacterial stock by aliquot 20μl of bacterial solution to sterile screw cap tube with 500μl LB-chloramphenical broth medium and 450μl of 20% glycerol. Otherwise solution that remain to use as template in PCR analysis.

7. PCR analysis

The PCR reaction was performed in a total volume of 50 μl using 50 ng of bacterial clone DNA in 200 μM each of deoxynucleotide triphosphates (dNTPs), 10mM tris HCI pH8.4, 50mM KCI, 1.5mM MgCl$_2$. Each of the M13 primer pair was performed in optimal concentration 0.5 μM. In the multiplex PCR reaction, the initial denaturation step was 95°C for 5 minutes then followed by 35 cycles of denaturation at 95 ^{o}C for 1 minute, annealing at 50 ^{o}C for 1 minute, extension at 72 ^{o}C for 2 minutes and a final extension at 72 ^{o}C for 7 minutes. Afterward, separate PCR product by submarine agarose gel electrophoresis; a 10 μl of PCR product with 2 μl of 10x Ficoll loading buffer, vortex and spin briefly, then load sample onto 0.8% agarose gel, electrophorase agarose gel in 1x TBE at 0.95v/cm until dye front reach the end of gel. Estimate size of PCR product by compare with 100bp marker.

8. Southern blotting

The southern blot was performed as the follows :

8.1 Digest genomic DNA by restriction enzyme (HpaII and MspI)

    8.1.1 Add 15 μg of genomic DNA (extracted from peripheral blood leukocytes, sperm and various tissues) to microcentrifuge tube

    8.1.2 Add 20 μl enzyme buffer (1/10 volume of total volume) and 75 U enzyme. Mix vigorously for 2-3 seconds then incubate reaction at $37\,^{\circ}C$ overnight.

8.2 Purification of digested product

    8.2.1 The digested product with equal volume of phenol-chloroform-isoamyl alcohol, vortex and mix thoroughly.

    8.2.2 Centrifuge 5 minutes at 14,000 rpm.

    8.2.3 Transfer the supernatant to the new microcentrifuge tube, add 20μg/μl glycogen carrier, 1/10 volume of 3M CH_3COONa and 2 volume of 100% ethanol and gently mix. The DNA should immediately form a stringy precipitate. Recover the DNA by centrifugation at 14,000 rpm for 15 minutes.

    8.2.4 Rinse the pellet with 70% ethanol. Decant the ethanol and air dried the pellet. (It is important to rinse well to remove any residual salt and phenol.)

    8.2.5 Resuspend the digested DNA in 20 μl of the double distilled water at $37\,^{\circ}C$ until dissolved.

    8.2.6 To calculate the DNA concentration, measure the density of the digested DNA at 260 nm and 280 nm. If OD

ratio of A_{260} to A_{280} is >1.6, the digested DNA is pure enough for further study.

8.3 Separate digested product

8.3.1 Purified digested product with 3 µl of 10x Ficoll loading buffer, vortex and spin briefly.

8.3.2 Load sample onto 0.8% agarose gel. Electrophorase agarose gel in 1x TBE at 0.35v/cm until dye front reach the end of gel.

8.4 Transfer DNA to nitrocellulose membrane

8.4.1 Denature DNA by transfer gel to denaturing solution for 30 minutes. Then wash gel by neutralizing solution for 15 minutes, twice.

8.4.2 Transfer DNA to Nylon transfer membrane with 20x SSC by using standard protocol for southern blotting.

8.4.3 Bake membrane 1-2 hours at 80 ^{o}C before used.

9. Hybridization

9.1 Place the membrane in the glass hybridization tube. Add 10 ml hybridization cocktail. Prehybridize 1-3 hours on a rotation in 42 ^{o}C hybridization oven.

9.2 Label DNA probe with P^{32} by Nick translation method as followed:

9.2.1 The Nick translation reaction was performed in a total volume of 20 µl using 1 µg of PCR product in 1µl each of dATP, dGTP, dTTP and α-^{32}P dCTP and 2µl each of buffer and mixture enzyme (from Nick Translation kit).

9.2.2 Incubate the reaction at 15 ^{o}C for 35 minutes. Stop the reaction with heating to 95 ^{o}C for 10 minutes then place on ice.

9.2.3 Add 50 μg COT1 DNA, mix thoroughly then incubate at 68 ^{o}C for 1-2 hours.

9.3 Add probe to hybridization tube. Incubate overnight at 42 ^{o}C with rotation.

9.4 Wash membrane with wash buffer1 for 1 hours, follow with wash buffer2 for 30 minutes twice.

9.5 Wrap membrane with clean plastic wrap and exposed to Phospho imager for 3-6 hours.

10. Analysis the DNA band pattern

Scan Phospho imager with Molecular dynamic storm. Analyze the DNA band pattern using the following pattern as the standard. Select the DNA band pattern from Southern blotting hybridization which have the difference band pattern between white blood cells and sperm infer the following pattern.
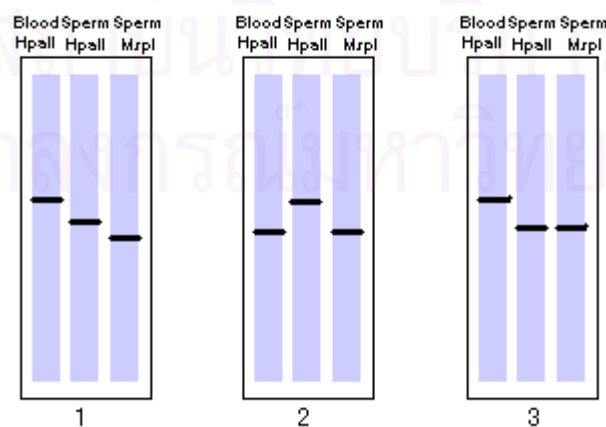
Figure 4-1 The expected DNA band pattern of Southern blotting hybridization
Picture 1 : Methylation in white blood cells with lesser expansion in sperm
Picture 2 : Methylation in sperm, demethylation in white blood cells
Picture 3 : Methylation in white blood cells, demethylation in sperm

11. Sequencing

11.1 Preparing the Plasmid DNA by alkaline lysis miniprep method

11.1.1 Smear LB-chloramphenical agar plate with selected bacteria clone solution from bacterial stock. Incubate at 37°C overnight,

11.1.2 pick up the single white colony on the agar plate to inoculate the LB-chloramphenical broth in 15ml Falcon 2059 polypropylene tube and incubate at 37°C overnight.

11.1.3 Aliquot 1.5ml to the microcentrifuge tube and centrifuge for 30 seconds at maximum speed then discard supernatant.

11.1.4 Resuspend the pellet with 100μl GTE solution, incubate for 5 minutes at room temperature.

11.1.5 Add 200μl NaOH/SDS solution, mix well and place on ice for 5 minutes.

11.1.6 Add 150μl potassium acetate solution, vortex and place on ice for 5 minutes. Then centrifuge for 3 minutes at maximum speed.

11.1.7 Transfer the supernatant to the new microcentrifuge tube and add 0.8 ml of 95%ethanol, incubate at room temperature for 2 minutes. Centrifuge 1 minute at maximum speed and discard supernatant.

11.1.8 Wash the pellet with 1ml 70%ethanol and air dry the pellet. Resuspend the pellet with 30µl TE buffer

11.2   Preparing sequencing reaction

The PCR reaction was performed in the total volume of 20 µl using 200ng of purified plasmid DNA in 8 µl Terminator Ready Reaction mix and either T7or T3 primer was performed in optimum concentration 3.2 ρmol. In the multiplex PCR reaction performed 25 cycles of denaturation at 96°C for 30 seconds, annealing at 50°C for 15 seconds and extension at 60°C for 4 minutes. Then rapid thermal ramp to 4°C and hold until ready to purify.

11.3   Purifying extension products by Gel column.

11.3.1   Gently tap the column to cause the gel material to settle to the bottom of the column.

11.3.2   Remove the upper end cap and add 0.8ml of deionized water.

11.3.3   Replace the upper end cap and invert the column a few times to mix the water and gel material.

11.3.4   Allow the gel to hydrate at room temperature for at least 2 hours.

11.3.5   Remove any air bubbles by inverting or tapping the column and allowing the gel to settle.

11.3.6   Remove the upper end cap first then remove the bottom cap. Allow the column to drain completely by gravity.

11.3.7   Insert the column into the wash tube provided.

11.3.8   Spin the column in the microcentrifuge at 730 g for 2 minutes to remove the interstitial fluid.

11.3.9   Remove the column from the wash tube and insert it into a sample collection tube

11.3.10   Remove the extension reaction mixture and load it carefully on top of gel material.

11.3.11   Spin the column in a microcentrifuge at 730g for 2 minutes

11.3.12   Discard the column. And dry the sample in microcentrifuge tube in at 90°C until dry do not over-dry.

11.4   *Electrophoresis on the ABI Prism 377 DNA sequencer*

11.4.1   Prepare a loading buffer by combining the following in a 5:1 ratio of deionizied formamide:25 mM EDTA (pH 8.0) with blue dextran (50mg/mL).

11.4.2   Resuspend each sample pellet in 6μl loading buffer. Vortex and spin.

11.4.3   Heat the samples at 95°C for 2 minutes to denature. Place on ice until ready to load.

11.4.4   Load 1 μl of each sample into a separate lane of the gel.

12. Analysis of homology

CpG island character of each clone had been predicted based on the following criterias; (1) minimum leangth 200 bp, GC content >50%, CpG/GpC>0.5. Putative promoter sequence and exon prediction were

predicted using the computer program NNPP and TSSG available through the Baylor College of Medicine launcher at *http://dot.imgen.bcm.tmc.edu:9331.* Sequence homologies were identified using the BLAST program of the National Center for Biotechnology Information available at *http://www.ncbi.nlm.nih.gov/BLAST/.* Search for the nucleotide sequence against the DNA sequence in the EST, HTGs, and nr. (see also in the Figure 4-2 to 4-8 and appendix B)
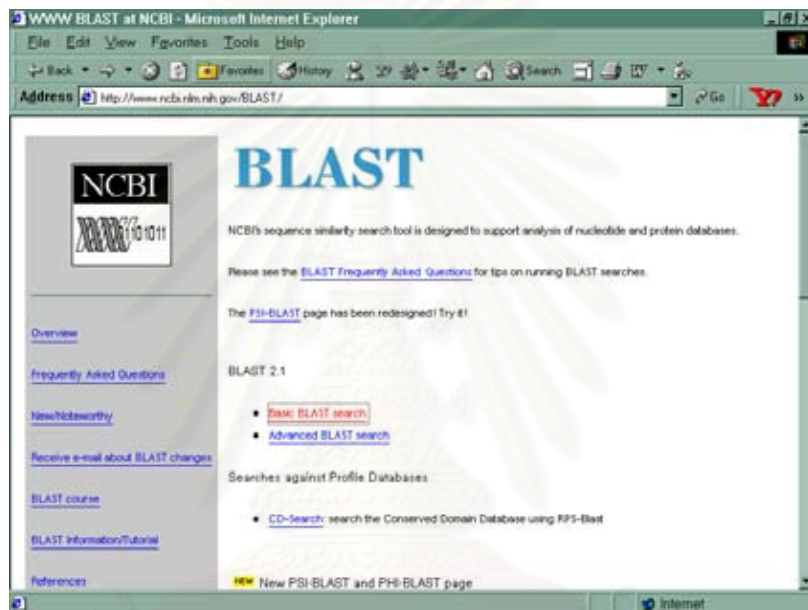
Figure 4-2   Web page of BLAST program which provide by National Center for Biotechnology Information. Chose basic BLAST search for this study
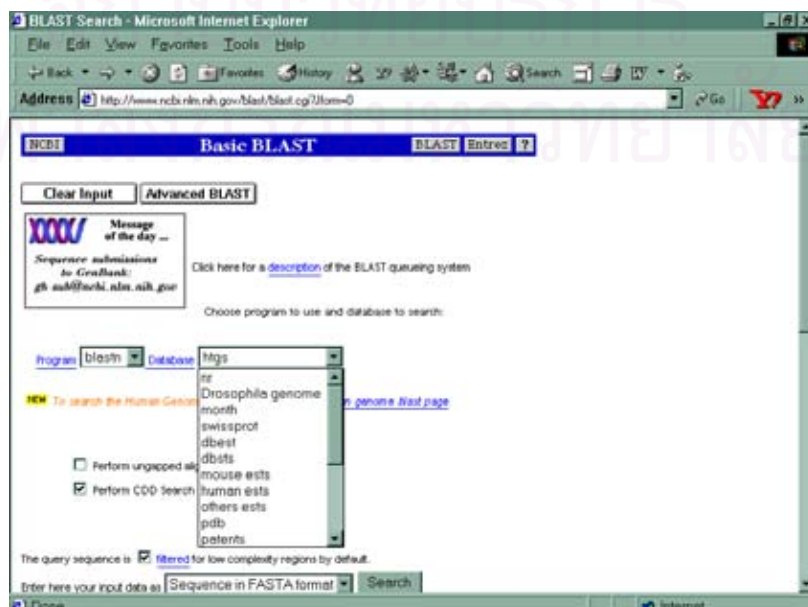
Figure 4-3 Web page of basic BLAST program. Choosing *blastn* and then
chose the organism which we focus on study (in this study we chose *nr, human ets*
and *htgs* , see also in appendix B)
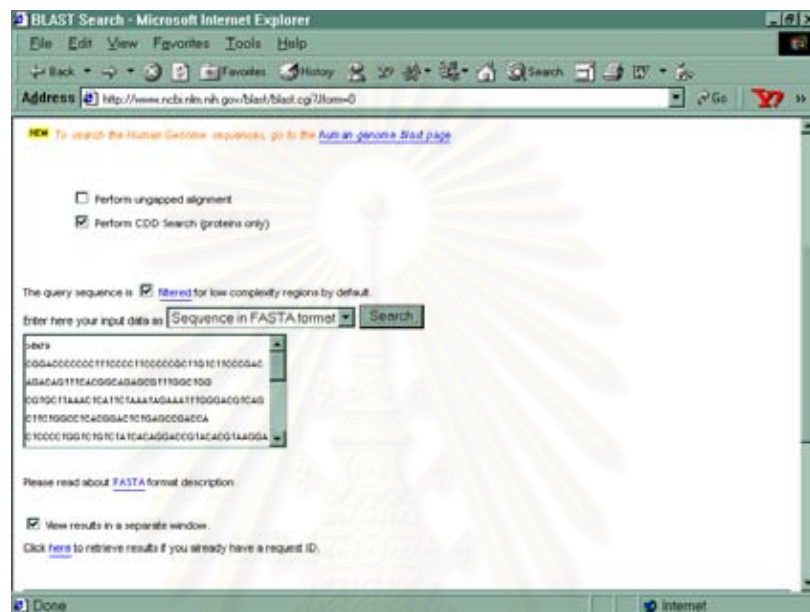
Figure 4-4 Web page of basic BLAST program (cont.) Put our sequence in
FASTA format form on the box. Then push *search* for homology search.

Figure 4-5 Web page of basic BLAST program (cont.) Then the result page will appears then press *Format result* for reviewing the homology search result.
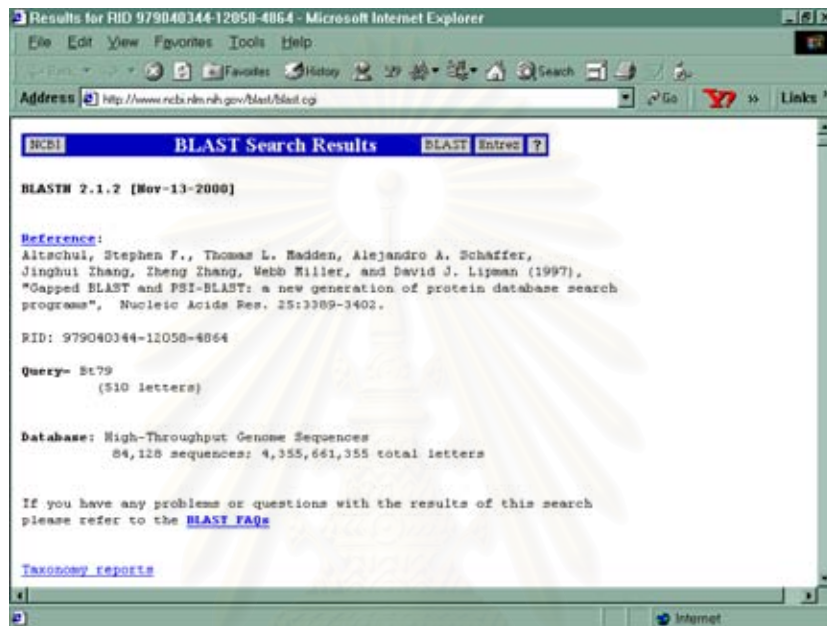
Figure 4-6 Web page of homology search result report of BLAST program. The report show name, the length of our query and database which had been searched of homology.
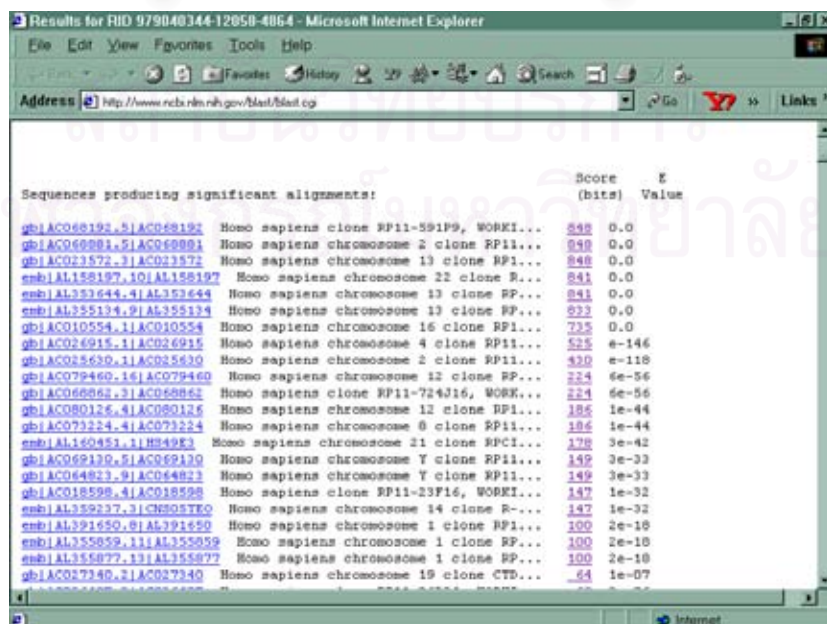
Figure 4-7 Web page of homology search result report of BLAST program. (cont.) The report show list of homology sequence, its GenBank number, E value and computing score.

Figure 4-8 Web page of homology search result report of BLAST program. (cont.) The report show the alignment between our query and each homology sequence. Its GenBank accession number, computing score , E-value, the percent of identity and the percent of gap had been reported.

13.   Methylation Analysis in Various Tissue by Methylation-specific PCR (MS-PCR)

13.1 *Primer design for MS-PCR*

Primer were designed to amplify the methylated and unmethylated allele equally (see primer sequence in appendix). The primer design mention about the difference between unmethylated allele and methylated allele after amplified and digest with chosen restriction endonuclease because of base conversion from cytosine to uracil after treated with bisulfite.

13.2 MS-PCR

I. Preparing of DNA template

    a. Dilute 1-2 µg DNA of each various tissue in 50 µl ddH$_2$O

    b. Add 5.5 µl 2M NaOH (from fresh stock) and mix well

    c. Incubate at 37°C for 10 minutes.

    d. Add 30 µl of the diluted hydroquinone (dilution 1:10 of 55mg hydroquinone in 5 ml ddH$_2$O), then vortex.

    e. Add 50 µl bisulfite (bisulfite 1.88g in 5 ml ddH$_2$O, bring pH to 5.0 with 5 drops of 19.5M NaOH), then vortex.

    f. Take off the oil and add 1 ml Wizard^{TM} resin to each tube and mix.

    g. Add to syringe attached to column anchored on the vacuum manifold and apply vacuum.

    h. Once drained, wash with 2ml 80% isopropanol and apply vacuum.

    i. Once drained, elute DNA from column by adding 50 µl heated (50-70°C) ddH$_2$O and centrifuge 1 minute at maximum speed.

    j. Add 5.5 µl 3M NaOH to the elute and incubate at room temp for 5 minutes.

    k. Add 66 µl 5M NH$_4$OAc and 2.3 volume of 95% ethanol.

    l. Precipitate overnight at -20°C, centrifuge 25 minutes at maximum speed, wash with 70% ethanol and dry pellet.

    m. Elute DNA with 20µl TE buffer, then ready for PCR.

II. Reaction and condition

The PCR reaction was performed in a total volume of 25μl using 3μl bisulfite treated DNA in 1x PCR buffer, 1.5mM MgCl$_2$, 0.2 mM each of deoxynucleotide triphosphates (dNTPs). Each of primer pair was performed in optimal concentration 1.0 μM. In the multiplex PCR reaction, the initial denaturation step was 95°C for 10 minutes then followed by 35 cycles of denaturation at 94°C for 1 minute, annealing at 48°C for 1 minute, extension at 72°C for 1 minutes and a final extension at 72°C for 7 minutes. Then, digested 20 μl of PCR product with restriction endonuclease (in this experiment is DpnII). Afterward, separate PCR product by submarine agarose gel electrophoresis; a 10 μl of digesting product with 2 μl of 6x Ficoll loading buffer, vortex and spin briefly, then load sample onto 2% agarose gel, electrophorase agarose gel in 1x TBE at 0.95v/cm until dye front reach the end of gel. Estimate size of digesting product by compare with 100bp marker.

CHAPTER V

Results

1. RDA result and PCR analysis of clone

Both white blood cell and sperm DNA were digested with HpaII, methylation sensitive restriction enzyme. HpaII recognized four base sequences (CCGG) without methylated cytosine. Average expected fragment of human DNA genome which digested with HpaII is approximately 0.6 kb (assuming 40% G+C and a CpG frequency 20% of this expected). By its function, methylation sensitive enzyme will give the different DNA fragment between white blood cell and sperm DNA, if there is different methylation pattern between the two genomes. Both digestion products have been ligated with RHpa adapter by the function of ligase enzyme. The multiplex PCR reaction promotes the higher amounts of the digestion products, and the tester and driver amplicons are performed in blunt end. Then the PCR amplification, approximately 200-1000 are enriched in the amplicon.
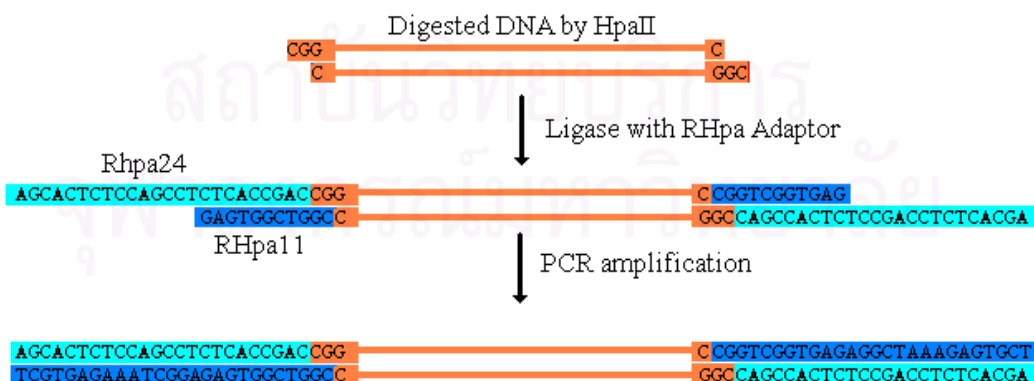
Figure 5-1 The preparation of tester and driver amplicon for ms-RDA
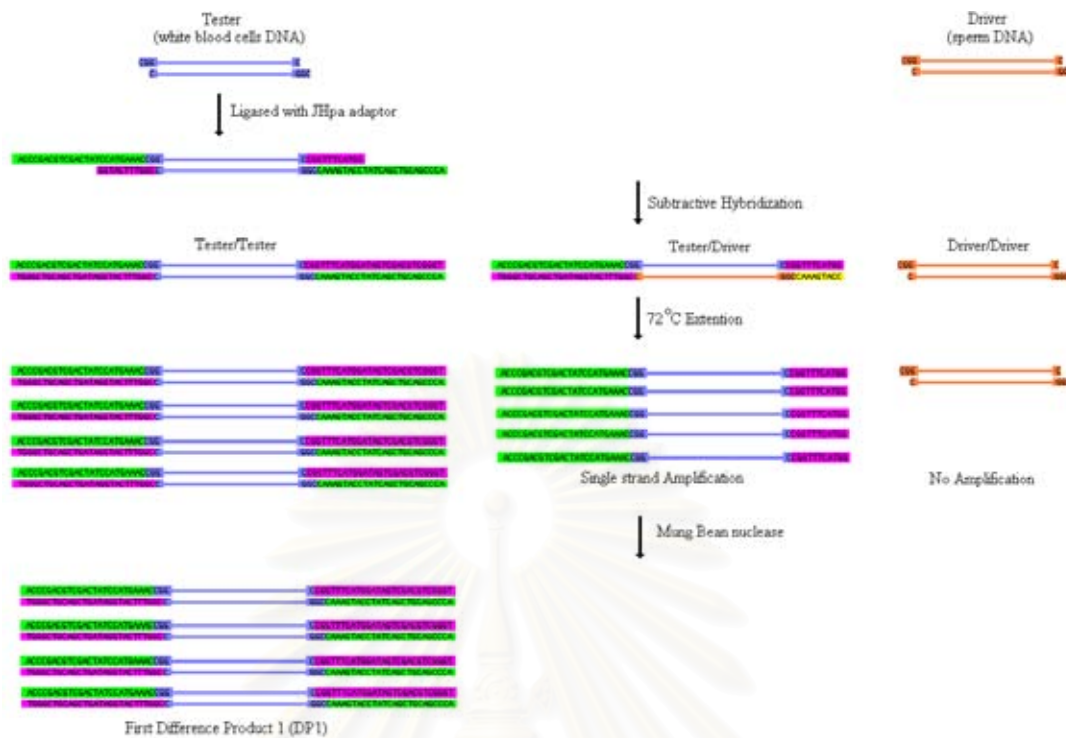
Figure 5-2 The description of ms-RDA method.

After the removal of adapter, only the tester amplicon (white blood cell genome) will be ligated with adapter, JHpa, and performed in the sticky end. Denaturation and reassociation in subtractive hybridization between tester and driver allowed the formation of three types of reannealed duplexed: (a) tester/tester, the target strands self reanneal in homoduplexes and have adaptor oligonuclotide sequences present on both 5'ends; (b) hybrids between tester and driver, that have adaptor oligo nucleotide on one end only and (c) driver/driver, the self-reannealed driver duplexes. In the multiplex PCR reaction, the initial step at 72°C will extend to the both ends of tester/tester by use Jhpa24 adaptor as the tamplate. On the other hand, hybrids between tester and driver, the extention will occur on one end only. The target fragments are selectively amplified after filling in the

oligonucleotide cohesive ends with *Taq* DNA polymerase. The Mung bean nuclease will digest the single strand DNA, therefore only tester/tester is the difference product.
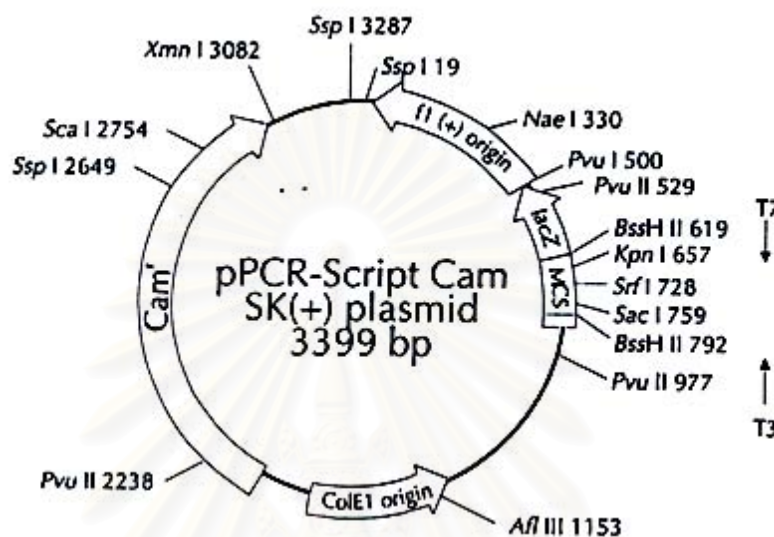
Figure 5-3 Circular map and polylinker sequence of the pPCR-Script Cam SK(+) cloning vector. The polylinker of the pPCR-Script Cam SK(+) cloning vector with the alterd nucleotides underlinedis shown above.

The difference product were subcloned in the pPCR-Script Cam SK (+) plasmid. (Figure 5-3) The plasmid contains the multiple cloning sites (MCS), but in this study the DNA fragment were inserted at SrfI site because this site had been digested in the *lacZ* gene. *lac Z* gene encodes the enzyme β-galactosidase, which hydrolyzes lactose into galactose and glucose. The plasmids were induced into competent *E.coli* cells, bacteria that carry the regulation elements and the Z gene of the lac operon (*lacZ* gene) can be induced by isopropyl-indolyl-β-D-galactoside (X-gal) and give rise to a blue insoluble derivative. The insertion at the *SrfI* site must interrupt the function of *lacZ* gene, and the bacterial colony with insertion will perform the white

colony. The white colonies were picked up and used as the template for PCR reaction to identify the insertion. In the MCS, there are many primer sequences such as M13, T3 and T7 (Figure 5-3). Plasmid DNA of each clone was performed in PCR amplification by using M13 oligonucleotides as the primers. There are variations in amplification fragment size in range between 200-700 bp (Figure 5-4 and Table 5-1). The amplified product whose size larger than 400 bp had been chosen to prove the different in methylation status by Southern blotting and hybridization. In the table, we calculated the length of inserted DNA fragment by erasing the length of multiple cloning size to M13 forward and reverse primers (227 bp).
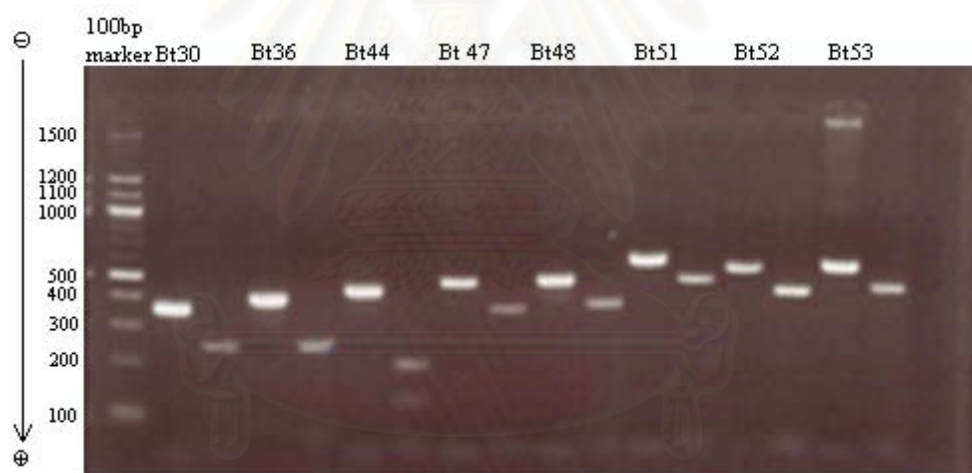
Figure 5-4 The example of the estimate PCR product fragment by using agarose gel electrophoresis.

Table 5-1 The result of PCR analysis and Southern blotting hybridization

CLONE	PCR Fragment using M13 as the primer (bp)	Calculated inserted DNA fragment (bp)	Southern blot hybridization band pattern
Bt1	400	173	NI
Bt2	300	73	NI
Bt3	450	223	NI
Bt4*	470	243	DP
Bt5	600	373	NI
Bt6	550	323	NI
Bt7	420	193	NI
Bt8	400	173	NI
Bt9	300	73	NI
Bt10	800	573	NI
Bt11	600	373	NI
Bt12	1000	773	NI
Bt13	320	93	NI
Bt14	470	243	NI
Bt15	270	53	NI
Bt16	350	123	NP
Bt17	340	113	NI
Bt18	300	73	NI
Bt19	500	273	NI
Bt20*	550	323	DP
Bt21	300	73	NI
Bt22	510	283	NI
Bt23	400	173	NI
Bt24	300	73	NI
Bt25	650	423	NI
Bt26	450	223	NI
Bt27	340	113	NI
Bt28	650	423	NI
Bt29	360	133	NI
Bt30	350	123	NP
Bt31	520	293	NI
Bt32	550	323	NI
Bt33	350	123	NI

NI : No information

NP : Non different band pattern

DP : Different band pattern

Table 5-1(cont.)The results of PCR analysis and Southern blotting hybridization

CLONE	PCR Fragment using M13 as the primer (bp)	Calculated inserted DNA fragment (bp)	Southern blot hybridization band pattern
Bt34	800	573	NI
Bt35	600	373	NI
Bt36	400	173	NP
Bt37	300	73	NI
Bt38*	400	173	DP
Bt39	600	373	NI
Bt40	700	473	NI
Bt41	400	173	NI
Bt42	420	193	NI
Bt43	700	473	NI
Bt44*	450	223	DP
Bt45	300	73	NI
Bt46	250	23	NI
Bt47	480	253	NI
Bt48	500	273	NI
Bt49	600	373	NI
Bt50	600	373	NI
Bt51	600	373	NI
Bt52	550	323	NI
Bt53	560	333	NI
Bt54	370	143	NP
Bt55	500	273	NI
Bt56	550	323	NI
Bt57	300	73	NI
Bt58	600	373	NI
Bt59	500	273	NI
Bt60	280	53	NI
Bt61	500	273	NI
Bt62	500	273	NI
Bt63	280	53	NI
Bt64	300	73	NI
Bt65	320	93	NI
Bt66	320	93	NI
Bt67	400	173	NI

NI : No information

NP : Non different band pattern

DP : Different band pattern

Table 5-1 (cont.) The results of PCR analysis and Southern blotting hybridization

CLONE	PCR Fragment using M13 as the primer (bp)	Calculated inserted DNA fragment (bp)	Southern blot hybridization band pattern
Bt68	280	53	NI
Bt69	320	73	NI
Bt70	500	273	NI
Bt71	450	223	NI
Bt72	600	373	NI
Bt73	350	123	NI
Bt74	320	193	NI
Bt75*	350	123	DP
Bt76	750	523	NI
Bt77	300	73	NI
Bt78	450	223	NI
Bt79*	700	473	DP
Bt80	440	243	NI
Bt81	300	73	NI
Bt82	240	63	NI
Bt83	400	173	NI
Bt84	400	173	NI
Bt85	400	173	NI
Bt86	500	273	NP
Bt87	650	423	NI
Bt88	470	243	NI
Bt89	300	73	NI
Bt90	400	173	NI
Bt91	350	123	NI
Bt92	400	173	NI
Bt93	350	123	NI
Bt95	350	123	NI
Bt96	350	123	NI
Bt99	400	173	NI
Bt100	380	153	NI
Bt101	400	173	NI
Bt102	350	123	NI
Bt103	350	123	NI
Bt104	350	123	NI
Bt105	380	153	NI

NI : No information

NP : Non different band pattern

DP : Different band pattern

2. Southern blotting hybridization pattern analysis of clone

The colony whose insertion fragment larger than 300 bp had been chosen for further study by Southern blotting and hybridization. The Southern blot for detection of the difference methylation status between white blood cells and sperm was prepared by using ms-restriction enzyme, HpaII and its isochizomer which do not recognize methylation, MspI. The HpaII digested product of white blood cells, placenta and sperm DNA had been loaded to the first, second and third lane respectively (Figure 5-5). Then the MspI digested product of sperm which performed as control had been loaded to the fourth lane (Figure 5-5). The electrophorase was performed in the optimum condition then followed by the transference to nitrocellulose membrane. The restriction DNA fragments migrate along the gel toward the positive electrode by their own electrical charge (DNA is a negative charge molecule). The rate of migration with respect to size reflects the ability of different molecules to thread through the gel network. Larger molecules move more slowly than smaller molecules because they meet more resistance in passing through the gel. Because the relative rates at which the molecules move depend on size, shape and charge density, the initial mixture of the molecules gradually separates into series of distinct bands moving at different rates through the gel. The fastest moving band contain the smallest, most compact molecules with the highest charge density.

DNA molecules separated by gel electrophoresis are removed from slab gel toward nitrocellulose membrane. A gel contain separate DNA into band had been soaked with denaturation solution to unpair the DNA molecule. A piece of filter paper soaked in salt solution is placed on one side

of the gel. Then a piece of nitrocellulose membrane, on which the blot will forms, is placed on the other side, backed by the several sheets of dry filter paper. As the result solution is drawn through the gel to the dry filter paper, it carries the molecules in the bands into nitrocellulose membrane, which can directly bind unpaired DNA nucleotide chain, where the bands are deposited and tightly bound as blots.

The hybridizations were performed by using inserted DNA fragments whose size larger than 400 bp as probes. The probes were labeled with radioactive α-^{32}P by nick translation method. DNAseI in the enzyme mixture induce nick on the nucleotide. By the function of DNA polymeraseI, polimerization at a nick is coupled with the 5'-> 3' exonuclease. Then the α-^{32}P dCTP will replace the old dCTP. Hybridization depends on the fact that DNA nucleotide chain unwind from the double helix at elevated temperature and will rewind with complementary sequence when the DNA is cooled. In this study, repetitive sequence was blocked by unwind salmon sperm and human placenta DNA. The labeled probes also unwinded before mix with hybridization solution. A hybridization solution containing a radioactive probe was poured over single-chain DNA bound in Southern blot. The labeled probe hybridized only with a blot containing a complementary DNA chain. The membrane was washed to remove unbound probe molecules, leaving the DNA molecule of interest as an individual blot marked with radioactivity.

We hybridized prepared Southern blot with the probes which created from selected colonies. There were four patterns of hybridization signals.

First, no signal present (Figure 5-5a). Second, there are no specific band on the blot but the signal was a smear hybridization signal (as shown in Figure 5-5 b). Thirdly, insignificance specific hybridization: The hybridization result performed as the band which indifferent in methylation status between white blood cell and sperm DNA. Finally, significance specific hybridization: The hybridization performed as the band which have different methylation statuses between white blood cells and sperm DNA. (as shown in Figure 5-5c, 5-6 to 5-11)

(a)   (b)   (c)
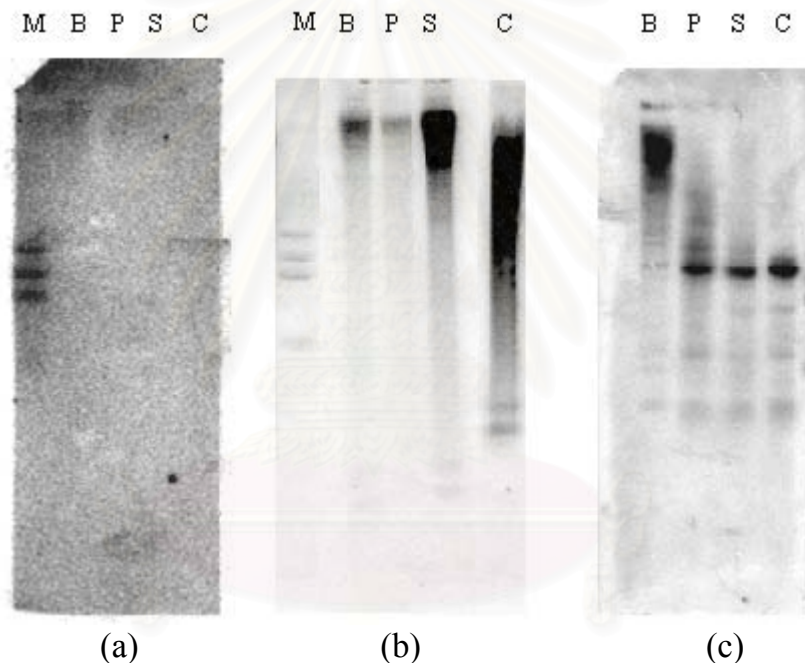
Figure 5-5 The example of Southern blotting hybridization pattern (a)Undetectable: The result of Southern blot hybridization preformed as the blank, only marker can be detected. (b) Non-specific hybridization: There are too many DNA fragments on the blot that can hybridize with the probe, so they preformed as the range of hybridization not a band. (c) Significance specific hybridization: The hybridization result performed as the band which different in methylation status between white blood cell and sperm DNA. M is 100 bp marker. B, P and S represent digested blood, placenta and sperm DNA by HpaII, respectively. M is 100 bp marker. C is degested sperm DNA by MspI.

The following is the Southern blotting hybridization result of 6 clones from 52 clones which present significant specific hybridization result after

hybridization. *B, P* and *S* lane in the hybridization results represent white blood cell sperm and placenta DNA which had been digested with HpaII respectively. Control represent by *C* is sperm DNA, digested with MspI, the isoshisozyme of HpaII but unrecognized methylation. Thus control showed the unmethylated status of each clone.

Hybridization Result of Bt4

The control lane showed two bands at approximately 250 and 500 bp. Blood, placenta and sperm DNA showed bands at approximately 200, 500, 700, 800 and 900 bp. There are differences in density of each band between each lanes (see Figure 5-6). This means that there may be two homologous loci. Both had methylated and unmethylated sequences in each tissues. Additionally, the extension of methylation in each cell is heterogenous.

Hybridization Result of Bt20

The control lane showed band at approximately 1100 bp which was different from amplified fragment size by M13. *S* and *P* lane revealed band at 1100 bp which is the same size as in control. Interstingly white blood cell showed intense high molecular weigh DNA (see Figure 5-7). From this result, large spreading of methylation on Bt20 occurs in most of white blood cells. In other word there is methylation difference between white blood cell and sperm DNA on their DNA sequences, homology to Bt20.

Hybridization Result of Bt38

The control lane demonstrated a major band at approximately 600 bp. This was different from its cloning size. *P* and *S* lane showed the same DNA size as control. Interstingly, blood DNA revealed not only 600 bp fragment

but also high molecular weigh methylated DNA. Thus there is partial methylation, extending several kb, occurs in white blood cells genome (see Figure 5-8).

Hybridization Result of Bt44

In Bt44 case, all bands presented at the higher molecular weight, larger than 1.5 kb, hence we couldn't estimate the exactly size of each band. Blood, placenta and sperm demonstrated distinction in their expression of methylation along the length of their genome in which blood cells was the largest.

Hybridization Result of Bt75

The control lane showed a band at approximately 1050 bp and its methylation pattern is similar to Bt20. There are unmethylated DNA in placenta and sperm. However, a long methylated DNA could be discovered from blood and placenta.

Hybridization Result of Bt79

The Bt79 unmethylated DNA fragment was 500 bp which was the same as its insert. Fully and large extension of methylation could be discovered from WBC DNA. However, the methylation status from placenta and sperm DNA are heterogenous. Both of them revealed both unmethylated and heterogenous extension of their methylated DNA.
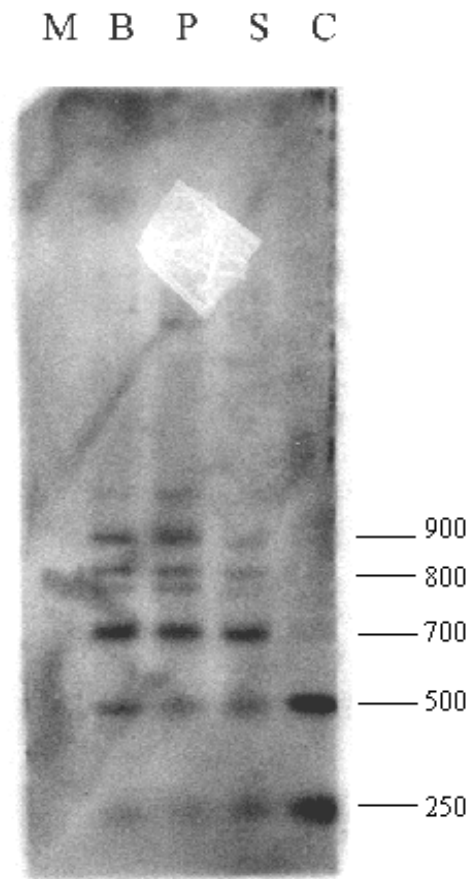
Figure 5-6 The hybridization result of Bt 4. M is 100 bp marker. B, P and S represent digested blood, placenta and sperm DNA by HpaII, respectively. C is degested sperm DNA by MspI.
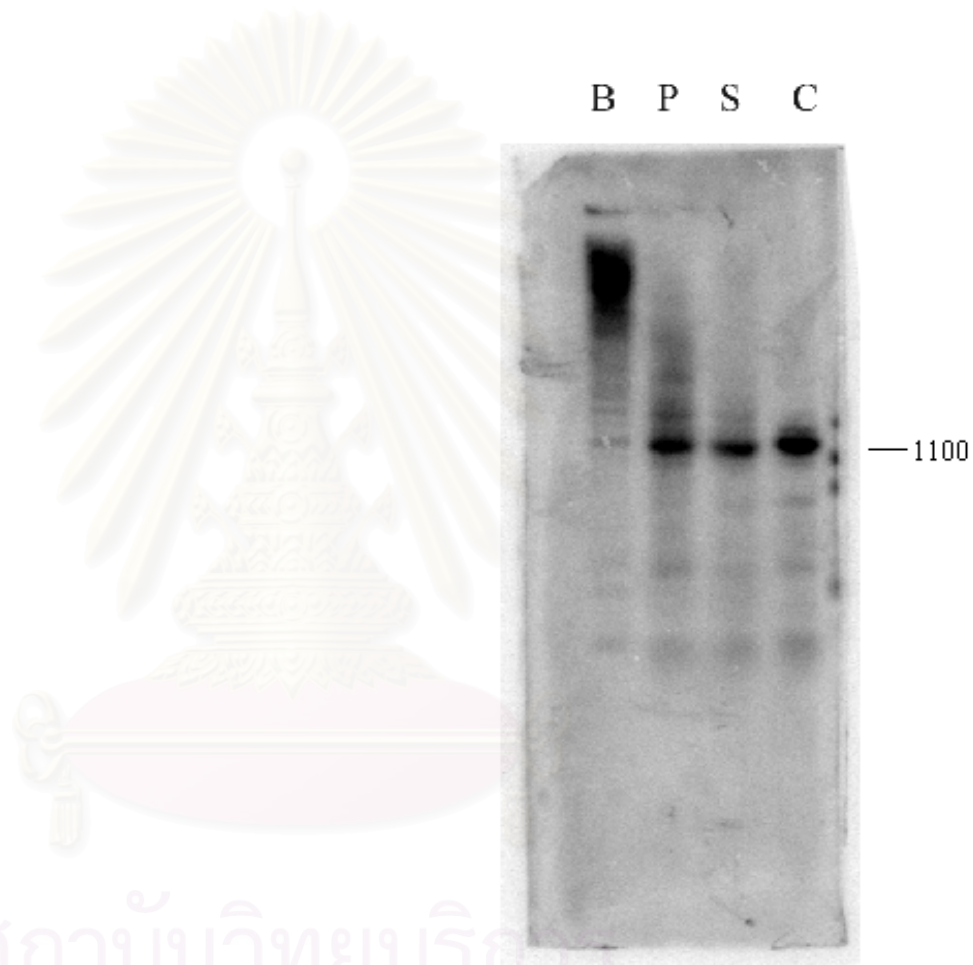
Figure 5-7 The hybridization result of Bt 20. B, P and S represent digested blood, placenta and sperm DNA by HpaII, respectively. C is degested sperm DNA by MspI.
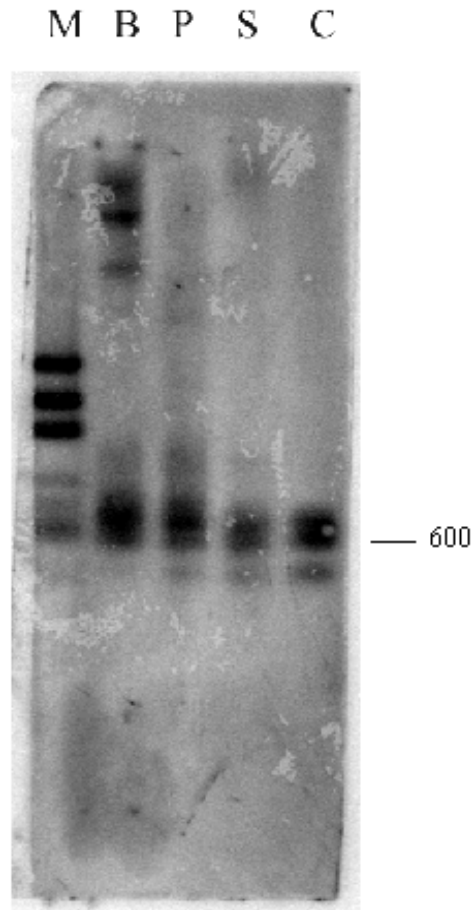
Figure 5-8 The hybridization result of Bt 38. M is 100 bp marker. B, P and S represent digested blood, placenta and sperm DNA by HpaII, respectively. C is degested sperm DNA by MspI.
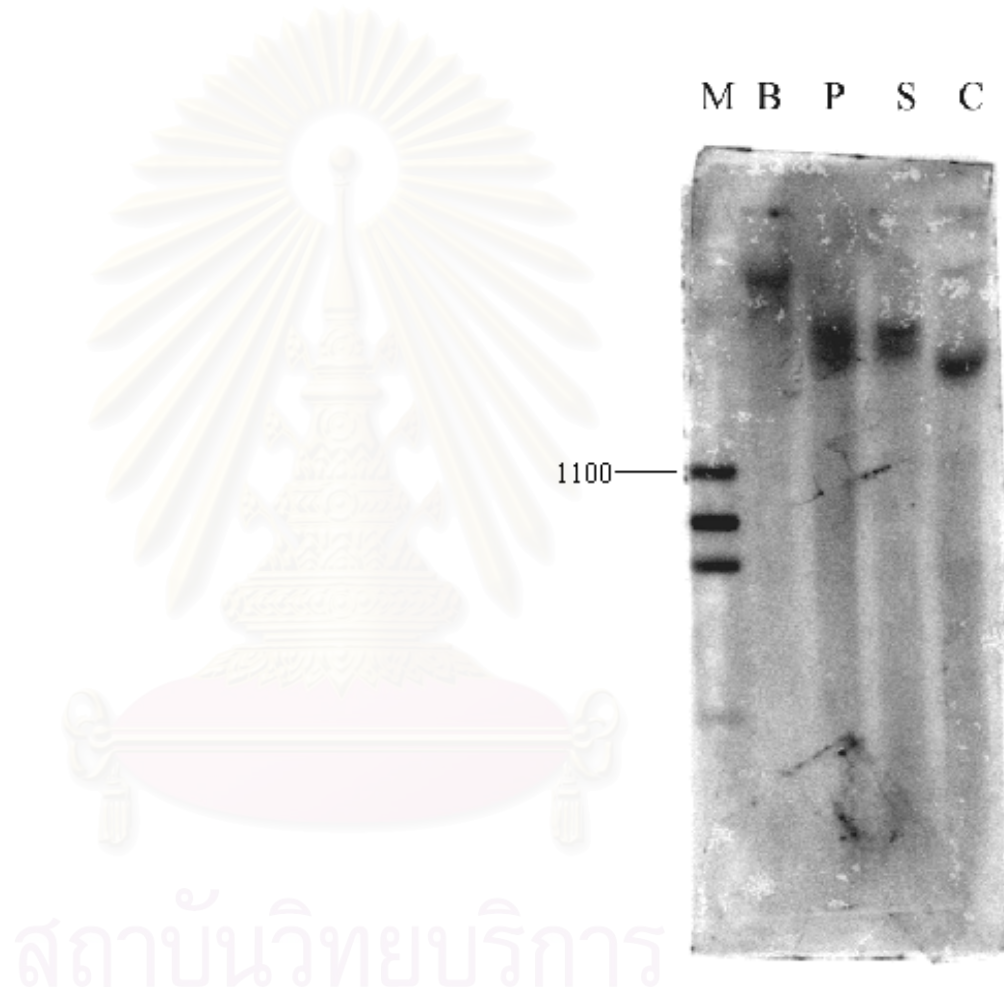
Figure 5-9 The hybridization result of Bt 44. M is 100 bp marker. B, P and S represent digested blood, placenta and sperm DNA by HpaII, respectively. C is degested sperm DNA by MspI.
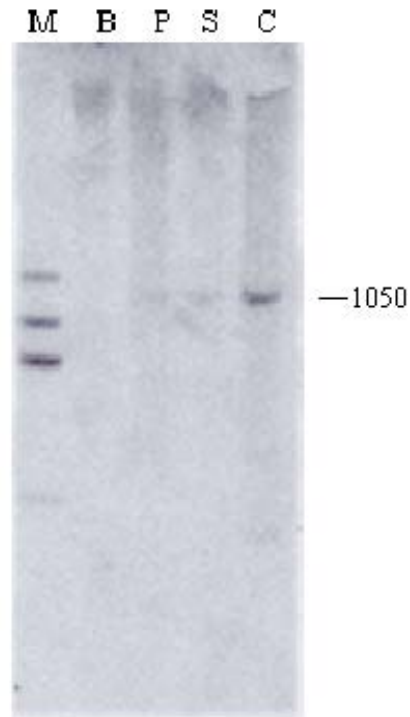
Figure 5-10 The hybridization result of Bt 75. M is 100 bp marker. B, P and S represent digested blood, placenta and sperm DNA by HpaII, respectively. C is degested sperm DNA by MspI.
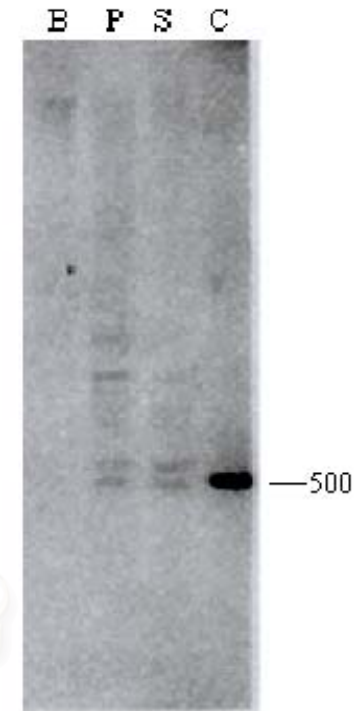
Figure 5-11 The hybridization result of Bt 79. M is 100 bp marker. B, P and S represent digested blood, placenta and sperm DNA by HpaII, respectively. C is degested sperm DNA by MspI.

3. <u>Analysis of DNA sequence by BLAST program</u>

Plasmid DNA from all of the clones, Bt4, 20, 38, 44, 75, 79, had been sequenced using automated DNA sequencer. Obtained sequences of pPCR-Script^{TM} Cam SK(+) and adaptors of ms-RDA method (RHpa) were manually removed. By DNA sequencing, we found that 3 clones had a GC content>50% and satisfied the minimal criteria for CpG islands (200bp, GC content > 50%, CpG/GpC > 0.5; Gardiner-Garden M, Frommer M; 1987). The edited sequences were submitted for GenBank accession number using dbGSS submission. The size of each clone, percentage of GC nucleotide, CpG/GpC, sequence homology, chromosomal location, and Genbank accession number are summarized in Table 5-2.

The result from sequencing of each clone

The sequence of each clones are as the followings:

Bt4

Edited sequence size: 135 bp

Base count (Plus strand)　　　49A　　24T　　30C　　32G

```
5'GAGCAGCAGC AGCATTCGCC TACGGATTTC TAGAAAAATA AGATGTCATG ATGAAGGATA
3'CTCGTCGTCG TCGTAAGCGG ATGCCTAAAG ATCTTTTTAT TCTACAGTAC TACTTCCTAT

GTAAACATCA ACCGGCTCTC ACTGCACGTT GAGAGAGTCA CAAAGCGCTA GTTCACAACA
CATTTGTAGT TGGCCGAGAG TGACGTGCAA CTCTCTCAGT GTTTCGCGAT CAAGTGTTGT

GGAAAAAACG GCAGC 3'
CCTTTTTTGC CGTCG 5'
```

Bt20

Edited sequence size　206　bp

Base count (Plus strand)　　　45A　　45T　　85C　　31G

```
5'TCTGACCTCC CAGGCATTCC CGACCCCCAC TTCTGCACTC TCTCTGACCG CTGGTAGAGT
3'AGACTGGAGG GTCCGTAAGG GCTGGGGGTG AAGACGTGAG AGAGACTGGC GACCATCTCA

GACTCCAGAC ACTCAAACAT CCACAAAGAC CCCACACCCC ATGTCTGAGG GCGACTTTAG
CTGAGGTCTG TGAGTTTGTA GGTGTTTCTG GGGTGTGGGG TACAGACTCC CGCTGAAATC
```

ACCCGCATCA ACCTCTCATC GACCCTGTCT GTCCTCTCAG TACCCACCGA CCCCCACTCG
TGGGCGTAGT TGGAGAGTAG CTGGGACAGA CAGGAGAGTC ATGGGTGGCT GGGGGTGAGC

ACGACCTACT ACCGTCATTC TTGTAT 3'
TGCTGGATGA TGGCAGTAAG AACATA 5'

Bt38

Edited sequence size: 293 bp

Base count (Plus strand)   93A   61T   72C   68G

  5'CGAGTCACAG CGTGACACTG CCCCACTCCA GTTTTAAAAG AGGAAACTAC AGACACACAA
  3'GCTCAGTGTC GCACTGTGAC GGGGTGAGGT CAAAATTTTC TCCTTTGATG TCTGTGTGTT

ATACTGAGGT GCTCTTTCCC AAAGCTATTA AACCAAGCAA AAGTCCCCAG AAGGAGTCAG
TATGACTCCA CGAGAAAGGG TTTCGATAAT TTGGTTCGTT TTCAGGGGTC TTCCTCAGTC

CAGGACACCG AGTTCCAGGC CCCATCCTGT CGCTGGCTAA TAGTGACAAG GGGAAATTGA
GTCCTGTGGC TCAAGGTCCG GGGTAGGACA GCGACCGATT ATCACTGTTC CCCTTTAACT

ATACCGCCCT GTGAGTCTGA GTGTACTGAG GTGTAAAAAG GAGTTGATGA AAAGTAACCT
TATGGCGGGA CACTCAGACT CACATGACTC CACATTTTTC CTCAACTACT TTTCATTGGA

GCCCTCTTTA GAAGACCAAG CTGACTTATG TCAAGTACTC AAGGCAGTTT CACG 3'
CGGGAGAAAT CTTCTGGTTC GACTGAATAC AGTTCATGAG TTCCGTCAAA GTGC 5'

Bt44

Edited sequence size: 178 bp

Base count (Plus strand)   12A   75T   61C   30G

  5'TCTATTTGTC TTTCTCCCTC CCTGTCTGTT TCTCTCTCTC TCTCTCTCTC TCTCTCTCTC
  3'AGATAAACAG AAAGAGGGAG GGACAGACAA AGAGAGAGAG AGAGAGAGAG AGAGAGAGAG

  CCTCTGTCTG TCTGTTTCTC TCTATCTCTC GCTGTCCATC TCTGTCTTTC TATGTCTGTC
  GGAGACAGAC AGACAAAGAG AGATAGAGAG CGACAGGTAG AGACAGAAAG ATACAGACAG

  TCTTTCTCTG TCAGTCTGTC AGACACCCCC GTGCCGGTCG GTGAGAGGCT GGAGAGTG 3'
  AGAAAGAGAC AGTCAGACAG TCTGTGGGGG CACGGCCAGC CACTCTCCGA CCTCTCAC 5'

Bt75

Edited sequence size: 155 bp

Base count (Plus strand)   28A   57T   32C   38G

  5'TTTTGTTTCG TTTCGTTCCT AAGTTCTGGG GTATATGTGC AGGATGTGCA GATTTGTTAC
    3'AAAACAAAGC AAAGCAAGGA TTCAAGACCC CATATACACG TCCTACACGT CTAAACAATG

  TAAGGTTAAC GTGTGCCATG GTGGTTTGCT GCTACCTGTC AACCCATCAC CTAGGTATTA
  ATTCCAATTG CACACGGTAC CACCAAACGA CGATGGACAG TTGGGTAGTG GATCCATAAT

GGCCCAGCAT GCAGTAGCTG TTTTTCTTAA CGCTC 3'
CCGGGTCGTA CGTCATCGAC AAAAAGAATT GCGAG 5'

Bt₇₉

Edited Sequence size: 510 bp

Base count (Plus strand)   105A   123T   151C   131G

5'CGGACCCCCC CTTTCCCCTT CCCCCGCTTG TCTTCCCGAC AGACAGTTTC ACGGCAGAGC
3'GCCTGGGGGG GAAAGGGGAA GGGGGCGAAC AGAAGGGCTG TCTGTCAAAG TGCCGTCTCG

GTTTGGCTGG CGTGCTTAAA CTCATTCTAA ATAGAAATTT GGGACGTCAG CTTCTGGCCT
CAAACCGACC GCACGAATTT GAGTAAGATT TATCTTTAAA CCCTGCAGTC GAAGACCGGA

CACGGACTCT GAGCCGACCA CTCCCCTGGT CTGTCTATCA CAGGACCGTA CACGTAAGGA
GTGCCTGAGA CTCGGCTGGT GAGGGGACCA GACAGATAGT GTCCTGGCAT GTGCATTCCT

GGAGAAAAAT CGTAACGTTC AAAGTCAGTC ATTTTGTGAT ACAGAAATAC ACGGATTCAC
CCTCTTTTTA GCATTGCAAG TTTCAGTCAG TAAAACACTA TGTCTTTATG TGCCTAAGTG

CCAAAACACA GAAAGCAAGT CTTTTAGAAA TGGCCTTAGC CCTGGTGTCC GTGCCAGCGA
GGTTTTGTGT CTTTCGTTCA GAAAATCTTT ACCGGAATCG GGACCACAGG CACGGTCGCT

TTCTTTTCGG TTTGGACCTT GACTGAGAGG ATTCCCAGTC GGTCTCTCGT CTCTGGACGG
AAGAAAAGCC AAACCTGGAA CTGACTCTCC TAAGGGTCAG CCAGAGAGCA GAGACCTGCC

AAGTTCCAGA TGATCCGATG GTGGGGGACT TAGGCTGCGT CCCCCCAGGA GCCCTGGTCG
TTCAAGGTCT ACTAGGCTAC CACCCCCTGA ATCCGACGCA GGGGGGTCCT CGGGACCAGC

ATTAGTTGTG GGGATCGCCT TGGAGGGCGC GGTGACCCAC TGTGCTGTGG GAGCCTCCAT
TAATCAACAC CCCTAGCGGA ACCTCCCGCG CCACTGGGTG ACACGACACC CTCGGAGGTA

CCTTCCCCCC ACCCCCTCCA CAGGGGATCC 3'
GGAAGGGGGG TGGGGGAGGT GTCCCCTAGG 5'

Each sequence was further explored by the BLAST program of the National Center for Biotechnology Information available at http://www.ncbi.nlm.nih.gov/BLAST/. 3 clones were identical to human gene sequences. One was an upstream sequence of Niemann-Pick C1 gene (*NPC1*) and the other two were ribosomal DNA intragenic spacer sequence. 2 clones were identical to human cDNA clone randomly sequenced deposit in GenBank, and one had no significance match in database.

Both Bt4 and Bt79 identical to intergenic spacer of ribosomal RNA gene (rRNA gene, GenBank accession number: U13369). Approximately 400 copies of rRNA gene in human genome located on the short arm of five acrocentric chromosomes (chromosome13, 14, 15, 21 and 22). Bt4 and Bt75 located at the position 28868 bp to 29003 bp with 97% identity and 37407 bp to 37912 bp with 98% identity of rRNA gene (GenBank accession number: U13369) respectively. This position located previously the initiation of rRNA gene promoter on the intergenic spacer (GenBank accession number: X68195). In addition Bt4 was identical to Human DNA insert showing sperm-specific hypomethylation (GenBank accession number: X06588) with 98% identity. From a previous study about ribosomal RNA sequence by Maden, BE et al; 1978, suggesting that the position, which Bt4 located on, is hypomethylated in sperm and may be the position which sensitive for variation. Thus the Bt4 and Bt79 locations may play an important role in its expression regulation (Gonzalez, IL and Sylvester, JE:1995). In our study, by the Southern blotting hybridization results of these two clones confirmed that there was hypomethylation in sperm when compared with the methylation status in white blood cell and placenta. The estimated 400 copies per diploid genome may cause the variation of band performance in the Southern blotting hybridization result of Bt4. The following shows BLAST result alignment of Bt4 and Bt79 The relation pattern of Bt4, Bt79 and ribosomal RNA gene are showed in Figure 5-12a.

BLAST result of Bt4
Database : nr

>gi|555853|gb|U13369.1|HSU13369 Human ribosomal DNA complete repeating unit
 Length = 42999

 Score = 218 bits (110), Expect = 5e-55
 Identities = 133/137 (97%), Gaps = 3/137 (2%)
 Strand = Plus / Minus

Query: 1   gagcagcagcagcattcgcctacggatttctagaaaaataagatgtcatgatgaaggata 60
             | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct: 29003 gagcagcagcagcattcgcctacggatttctagaaaa-taagatgtcatgatgaaggatn 28945

Query: 61   gtaaacatcaaccggctctcactgca-cgttgagagagtcacaaa-gcgctagttcacaa 118
             | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct: 28944 gtaaacatcaaccggctctcactgcagcgttgagagagtcacaaaagcgctagttcacaa 28885

Query: 119   caggaaaaaacggcagc 135
             | | | | | | | | | | | | | | | | |
Sbjct: 28884 caggaaaaaacggcagc 28868

>gi|36566|emb|X06588.1|HSSPM4 Human DNA insert showing sperm-specific
hypomethylation (Sp-0.3-10)Length = 335

 Score = 111 bits (56), Expect = 9e-23
 Identities = 63/64 (98%), Gaps = 1/64 (1%)
 Strand = Plus / Plus

Query: 73   cggctctcactgcacgttgagagagtcac-aaagcgctagttcacaacaggaaaaaacgg 131
             | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct: 1   cggctctcactgcacgttgagagagtcacaaaagcgctagttcacaacaggaaaaaacgg 60

Query: 132 cagc 135
           | | | |
Sbjct: 61 cagc 64

BLAST result of Bt79
Database : nr

>gi|36165|emb|X68195.1|HSRSPAC H.sapiens genomic DNA of ribosomal RNA intergenic
spacer sequence Length = 7431

 Score = 934 bits (471), Expect = 0.0
 Identities = 503/511 (98%), Gaps = 2/511 (0%)
 Strand = Plus / Plus

Query: 1   cggaccccccctttcccccttcccccgcttgtcttcccgacagacagtttcacggcagagc 60
             | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct: 6922 cggacccccccttttcccccttcccccgcttgtcttcccgacagacagtttcacggcagagc 6981

Query: 61   gtttggctggcgtgcttaaactcattctaaatagaaatttgggacgtcagcttctggcct 120

```
            | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct: 6982 gtttggctggcgtgcttaaactcattctaaatagaaatttgggacgtcagcttctggcct 7041


Query: 121  cacggactctgagccgaccactccctggtctgtctatcacaggaccgtacacgtaagga 180
            | | | | | | | | | | | | | | | | |   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct: 7042 cacggactctgagccgaggagtccctggtctgtctatcacaggaccgtacacgtaagga 7101


Query: 181  ggagaaaaatcgtaacgttcaaagtcagtcattttgtgatacagaaatacacggattcac 240
            | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct: 7102 ggagaaaaatcgtaacgttcaaagtcagtcattttgtgatacagaaatacacggattcac 7161


Query: 241  ccaaaacacagaaagcaagtcttttagaaatggccttagccctggtgtccgtgccagcga 300
            | | | | | | | | | | | | | | |   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |   | |
Sbjct: 7162 ccaaaacacagaaa-ccagtcttttagaaatggccttagccctggtgtccgtgccagtga 7220


Query: 301  ttcttttcggtttggaccttgactgagaggattcccagtcggtctctcgtctctggacgg 360
            | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct: 7221 ttcttttcggtttggaccttgactgagaggattcccagtcggtctctcgtctctggacgg 7280


Query: 361  aagttccagatgatccgat-ggtgggggacttaggctgcgtccccccaggagccctggtc 419
            | | | | | | | | | | | | | | | | | | |   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct: 7281 aagttccagatgatccgatgggtgggggacttaggctgcgtcccccaggagccctggtc 7340


Query: 420  gattagttgtggggatcgccttggagggcgcggtgacccactgtgctgtgggagcctcca 479
            | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct: 7341 gattagttgtggggatcgccttggagggcgcggtgacccactgtgctgtgggagcctcca 7400


Query: 480  tccttcccccaccccctccacaggggatcc 510
            | | | | | | | | | | | | | | | | | | | | |   | | | | | | | | | | | |
Sbjct: 7401 tccttcccccacccccctccccaggggatcc 7431
```

>gi|555853|gb|U13369.1|HSU13369 Human ribosomal DNA complete repeating unit
 Length = 42999

 Score = 926 bits (467), Expect = 0.0
 Identities = 499/507 (98%), Gaps = 2/507 (0%)
 Strand = Plus / Plus

```
Query: 1     cggacccccctttcccttcccccgcttgtcttcccgacagacagtttcacggcagagc 60
             | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct: 37407 cggacccccctttcccttcccccgcttgtcttcccgacagacagtttcacggcagagc 37466


Query: 61    gtttggctggcgtgcttaaactcattctaaatagaaatttgggacgtcagcttctggcct 120
             | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct: 37467 gtttggctggcgtgcttaaactcattctaaatagaaatttgggacgtcagcttctggcct 37526
```

```
Query: 121   cacggactctgagccgaccactcccctggtctgtctatcacaggaccgtacacgtaagga 180
              | | | | | | | | | | | | | | | |  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct: 37527  cacggactctgagccgaggagtcccctggtctgtctatcacaggaccgtacacgtaagga 37586
```

```
Query: 181   ggagaaaaatcgtaacgttcaaagtcagtcattttgtgatacagaaatacacggattcac 240
              | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct: 37587  ggagaaaaatcgtaacgttcaaagtcagtcattttgtgatacagaaatacacggattcac 37646
```

```
Query: 241   ccaaaacacagaaagcaagtcttttagaaatggccttagccctggtgtccgtgccagcga 300
              | | | | | | | | | | | | | | | |  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  | |
Sbjct: 37647  ccaaaacacagaaa-ccagtcttttagaaatggccttagccctggtgtccgtgccagtga 37705
```

```
Query: 301   ttcttttcggtttggaccttgactgagaggattcccagtcggtctctcgtctctggacgg 360
              | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct: 37706  ttcttttcggtttggaccttgactgagaggattcccagtcggtctctcgtctctggacgg 37765
```

```
Query: 361   aagttccagatgatccgat-ggtgggggacttaggctgcgtccccccaggagccctggtc 419
              | | | | | | | | | | | | | | | | | | |  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct: 37766  aagttccagatgatccgatgggtgggggacttaggctgcgtccccccaggagccctggtc 37825
```

```
Query: 420   gattagttgtggggatcgccttggagggcgcggtgacccactgtgctgtgggagcctcca 479
              | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct: 37826  gattagttgtggggatcgccttggagggcgcggtgacccactgtgctgtgggagcctcca 37885
```

```
Query: 480   tccttcccccaccccctccacagggg 506
              | | | | | | | | | | | | | | | | | | | | |  | | | | | |
Sbjct: 37886  tccttcccccaccccctcccagggg 37912
```

Both Bt20 and Bt75 sequences are identical to *Homo sapiens* chomosome19 cosmid genomic DNA (GenBank accession number: AF025422). Bt20 and Bt75 located on the position 1538bp to 1743bp and the position 4993bp to 5099bp with 95% identity of this chromosome19 cosmid, respectively. The chromosome 19 cosmid is a genomic directly submission sequence maped on 19q13.2. In addition, Bt20 sequence was also identical to human retina cDNA clone (GenBank accession number: W27289) with 100% identity and Bt75 was also identical to cDNA from neuroendocrine lung carcinoid clone (GenBank accession number: AA808534). This indicates that both clones are parts of genes. Therefore, their methylations

maybe important in their tissue specific regulation of expression. The following show BLAST result alignment of Bt20 and Bt75. The relation pattern of Bt20, Bt75 and *Homo sapiens* chomosome19 cosmid had shown in Figure 5-12b.

BLAST result of Bt20
Database : nr

>gb|AF025422|AF025422 Homo sapiens chromosome 19 cosmid F15386, genomic sequence,
complete sequence [Homo sapiens] Length = 38000
Score = 329 bits (166), Expect = 2e-88
Identities = 196/206 (95%)
Strand = Plus / Plus

```
Query:   1  tatgttcttactgccatcatccagcagctcacccccagccacccatgactctcctgtctg 60
            |||||||||||||||||||||||||||||| ||||||||| |||||||||||||||||||
Sbjct: 1538 tatgttcttactgccatcatccagcagctcactcccagccatccatgactctcctgtctg 1597
```

```
Query:  61  tcccagctactctccaactacgcccagatttcagcgggagtctgtaccccacaccccaga 120
            ||||||||||||||||||||||| |||||||||||||||||||||| || ||| |||||||||
Sbjct: 1598 tcccagctactctccaactacgtccagatttcagcgggagtcggtgtcccgcaccccaga 1657
```

```
Query: 121  aacacctacaaactcacagacctcagtgagatcctcgccagtctctctcacgtcttcacc 180
            |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Sbjct: 1658 aacacctacaaactcacagacctcagtgagatcctcgccagtctctctcatgtcttcacc 1717
```

```
Query: 181  cccagcccttacggaccctccagtct 206
            ||||||||||||||||| |||| ||||
Sbjct: 1718 cccagcccttacggactctcctgtct 1743
```

Database : Human EST

>gi|1307012|gb|W27289.1|W27289 27h10 Human retina cDNA randomly primed sublibrary
Homo sapiens cDNA. Length = 626
Score = 393 bits (198), Expect = e-107
Identities = 204/206 (99%)
Strand = Plus / Plus

```
Query:   1  agactggagggtccgtaagggctgggggtgaagacgtgagagagactggcgaccatctca 60
            |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Sbjct: 109  agactggagggtccgtaagggctgggggtgaagacgtgagagagactggcgaggatctca 168
```

```
Query:  61  ctgaggtctgtgagtttgtaggtgtttctggggtgtggggtacagactcccgctgaaatc 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

Sbjct: 169 ctgaggtctgtgagtttgtaggtgtttctggggtgtggggtacagactcccgctgaaatc 228

Query: 121 tgggcgtagttggagagtagctgggacagacaggagagtcatgggtggctgggggtgagc 180
            | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct: 229 tgggcgtagttggagagtagctgggacagacaggagagtcatgggtggctgggggtgagc 288

Query: 181 tgctggatgatggcagtaagaacata 206
            | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct: 289 tgctggatgatggcagtaagaacata 314

BLAST result of Bt75
Database : nr

>gb|AF025422.1|AF025422 Homo sapiens chromosome 19 cosmid F15386, genomic sequence,
complete sequence Length = 38000
 Score = 172 bits (87), Expect = 6e-41
 Identities = 102/107 (95%)
 Strand = Plus / Minus

Query: 13   aagttctggggtatatgtgcaggatgtgcagatttgttactaaggtaaacgtgtgccatg 72
            | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct: 5099 aagttctggggtatatgtgcaggatgcgcagatttgttacaaaggtaaacgtgtgccgtg 5040

Query: 73   gtggtttgctgcacctgtcaacccgtcacctaggtattaagcccagc 119
            | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct: 5039 gtggtttgctgcacctgtcgacacgtcacctaggtattaagcccagc 4993

Database : Human est

> AGI2877940|AA808534 oe55g03.s1 NCI_CGAP_Lu5 Homo sapiens cDNA clone IMAGE:14155723′ similar to
contains L1.t3 L1 repetitive element ; mRNA sequence

 Length 408 bp
 Score = 244 bits (127), Expect = 8e-63
 Identities = 142/147 (96%), Gaps = 1/147 (0%)
 Strand = Plus / Plus

Query: 9   cgtttcgttcctaagttctggggtatatgtgcaggatgtgcagatttgttactaaggtta 68
            | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct: 1   cgtttcgtttctaagttctggggtatatgtgcaggatgtgcagatttgttactaaggtaa 60

Query: 69   acgtgtgccatggtggtttgctgctacctgtcaacccatcacctaggtattaggcccagc 128
            | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct: 61   acgtgtgccatggtggtttgctgc-acctgtcaacccgtcacctaggtattaagcccagc 119

Query: 129 atgcagtagctgtttttcttaacgctc 155

```
          | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct: 120 atgcagtagctgtttttcttaacgctc 146
```

Bt38 is identical to the position 3' to the first exon of Niemann-Pick C1 gene (NPC1 gene). It locates on the position 162644 bp to162936 bp with 98% identity of Homo sapiens chromosome18 clone (GenBank accession number: AC010853). In the further exploration of this clone found that the first exon of NPC1 gene (GenBank accession number: AF157365) also identical to this Homo sapiens chromosome18 clone at the position 164122 bp to 166489 bp with 100% identity. So the position which Bt38 located is an intron between exon1 and exon2 of NPC1 gene (position 189 bp to 482 bp downstream from the last base of exon1). NPC1 gene was mapped on 18q13.2 locus and contained 25 exons, varying in size from 74 to 788 bp, spread over 47 kb (Morris et al; 1999). NPC1 is an integral membrane protein with multiple transmembrane domains that appears to be localized to a late endosomal compartment. Its function plays a central role in modulating intracellular sorting of cholesterol and glycosphingolipids (Neufeld et al; 1999). By the result of Southern blotting hybridization in our study showed the methylation status different between white blood cells and sperm, we could assume that there is DNA methylation involve in the control of cholesterol transportation in some white blood cells by controlling NPC1 expression.

BLAST result of Bt38
Database : htgs

>gi|7637768|dbj|AC010853.2|AC010853 Homo sapiens chromosome 18 clone RP11-349 D12 map 18, WORKING DRAFT SEQUENCE, 27 unordered pieces. Length = 180370

 Score = 546 bits (284), Expect = e-153
 Identities = 290/293 (98%)
 Strand = Plus / Plus

Query: 1   cgagtcacagcgtgacactgccccactccagttttaaaagaggaaactacagacacacaa 60

```
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 162644 cgagtcacagcgtgacactgccccactccagttttaaaagaggaaactacagacacacaa 162703


Query: 61     atactgaggtgctctttcccaaagctattaaaccaagcaaaagtccccagaaggagtcag 120
            ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Sbjct: 162704 atactgaggtgctctttcccaaagctattaaaccaagcaaaagccccagaaggagtcag 162763


Query: 121    caggacaccgagttccaggccccatcctgtcgctggctaatagtgacaaggggaaattga 180
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Sbjct: 162764 caggacaccgagttccaggccccatcctgtcgctggctaatagtgacaaggggaaattca 162823


Query: 181    ataccgccctgtgagtctcagtgtactgaggtgtaaaaaggagttgatgaaaagtaacct 240
            ||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Sbjct: 162824 ataccgccctgtgagtctcagtgtcctgaggtgtaaaaaggagttgatgaaaagtaacct 162883


Query: 241    gccctctttagaagaccaagctgacttatgtcaagtactcaaggcagtttcac 293
            |||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 162884 gccctctttagaagaccaagctgacttatgtcaagtactcaaggcagtttcac 162936
```

There is no information about the identity of Bt44 in GenBank database. Hopefully, when the human genome project is complete there will be information regarding Bt44.

Table 5-2: Summary of 6 differentially methylated clones isolated by MS-RDA

Clone	Bt4	Bt20	Bt38	Bt44	Bt75	Bt79
Size (bp)	135	206	293	178	155	510
%CG	47	56	47	52	45	55
CpG/GpC	0.5	1.8	0.4	5.0	0.4	1.1
CpG island[a]	No	Yes	No	Yes	No	No
Blast Homology	Ribosomal DNA gene	Retina cDNA	NPCI gene	NI	Lung cDNA	Ribosomal DNA gene
Blast accession No.	U13369 X06588	AF025422[b] W27289[c]	AC010853 AF157365	[b]NI	AF025422[b] AA808534[c]	U13369 X68195[c]
Chromosome map	NI[d]	19q13.2	18q11	NI	19q13.2	NI
GenBank accession No.	AZ301005	AZ301006	AZ301007	AZ301008	AZ301008	AZ301009

[a] The presence of CpG islaNI was determined based on criteria described previously(R): minimum length 200 bp; GC content >50% ; CpG/GpC >0.5.

[b] Regions sequenced as part of the human genome project.

[c] Regions sequenced as part of randomly primer subcloned.

[d] NI : No Information data.
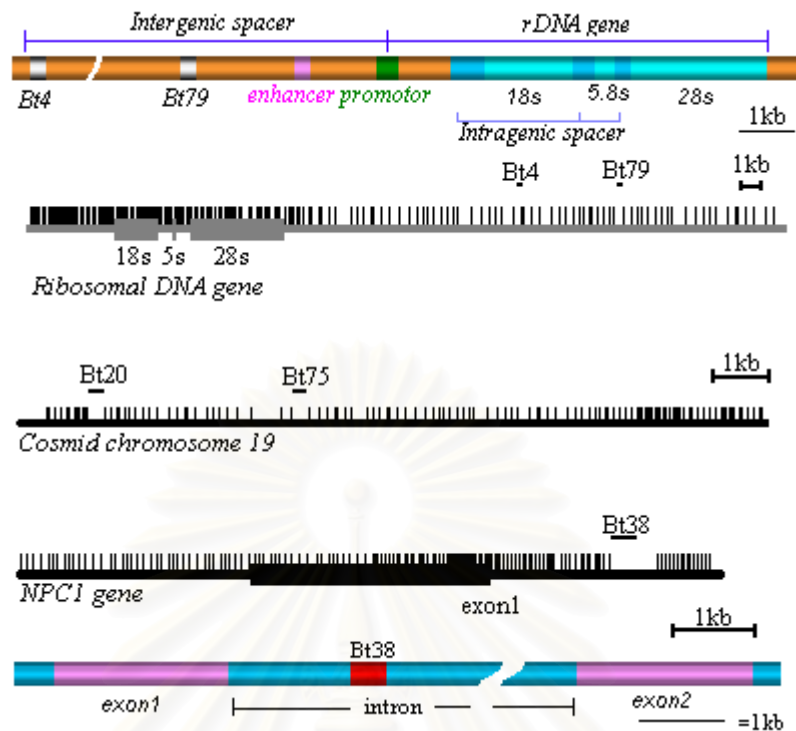
Figure 5-12 Relation between isolated clones and known genes or sequences.Two of the genes and one cosmid sequence is available from GenBank are shown, *Filled boxes*; exons, *Opened boxes*; the position of cDNA, *Short vertical lines*; the position of CpG dinucleotide, *Bars at the top*; position of Bt clones.

4. **Methylation status analysis in various tissue**

NPC1 upstream sequence, Bt38, was analyzed methylation status by MS-PCR. Chemical modification of cytosine to uracil by bisulfite is the basis of the MS-PCR. Under appropriate conditions, cytosines in bisulfite treated DNA are converted to uracil. However, cytosine that are methylated (5-methylcytosine) resists to this modification and remain cytosines. Amplifying this modified DNA provides detailed information of the methylation status of all CpG sites within the amplified region (Frommer et al., 1992). By sequence analysis of Bt38 found that Bt38 identity to *NPC1* gene, so primer for MS-PCR had been designed from the upstream sequence of *NPC1*.

After the modification of DNA by bisulfite, the two daughter strands of any given gene are no longer complementary after treatment. Either strand can serve as the template for subsequent PCR amplification, and the methylation pattern of each strand could then be determined. CpG dinucleotides are almost always symmetrically methylated, meaning that detection on one strand implies similar methylation pattern on the other strand. The primers was designed for amplification 197 bp in length (at position 155978 bp to156175 bp of *NPC1* gene). The amplified product has restriction site of DpnII at the position 149 bp to provide digesting products 149 bp and 48 bp if the location is methylated. Both methylated and unmethylated sequences were discovered from all white blood cell lines and all tissue types.
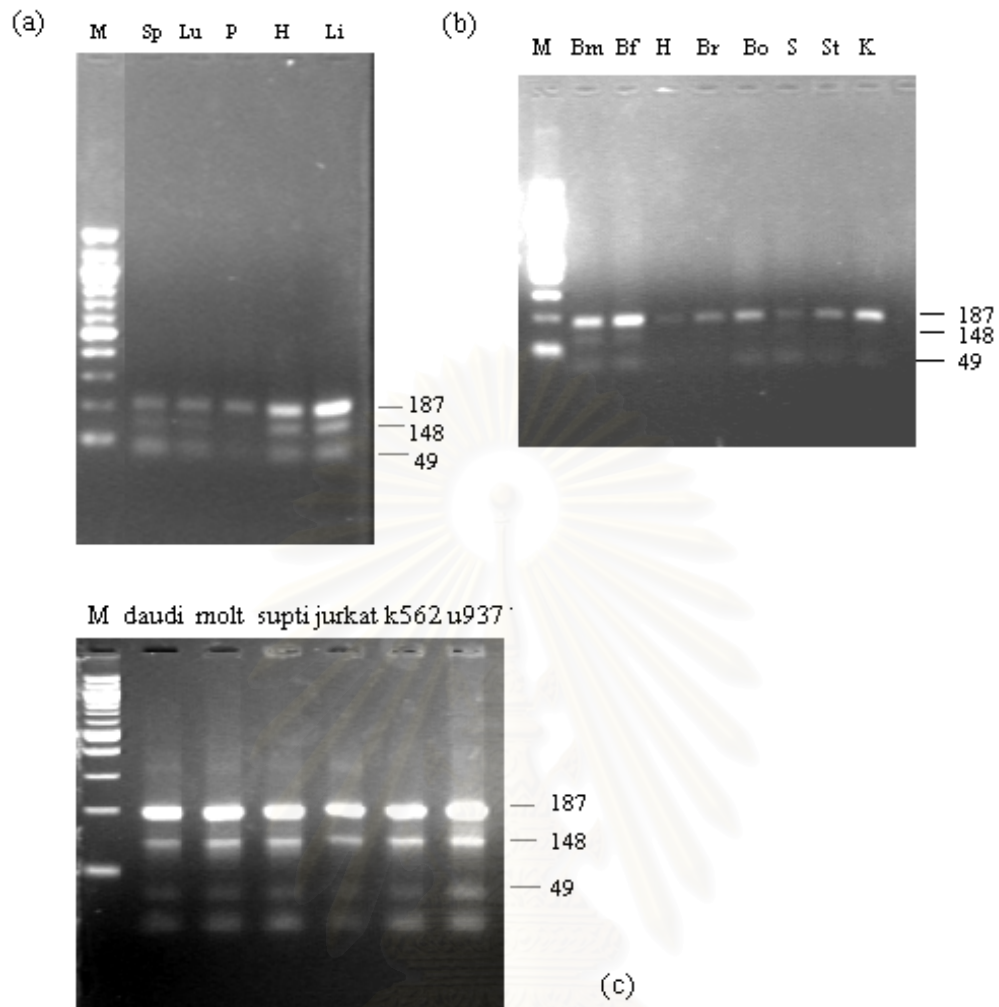
Figure 5-14 The MS-PCR result of Bt38 in various tissues. (a) and (b) show the result in various tissues. *M* is 100bp marker. *Sp, Lu, P, H, Li, Bm, Bf, Br, Bo, S, St* and *K* represent DNA of spleen, lung, placenta, heart, liver, male blood, female blood, brain, bone marrow, sperm, stomach and kidney, respectively. (c) shows the results in cell line of various types of white blood cells. *duadi* is B-cell line. *molt*, *supti* and *jurkat* are T-cell lines. *k562* is erythroid myeloid cell line. *u937* is monocytic cell line.

CHAPTER VI

Discussion

Using MS-RDA teachnique, we succeeded in subtracting methylated DNA fragment between white blood cells and sperm. All showed methylated DNA in white blood cells. Comparing the methylation pattern of placenta, white blood cells and sperm suggested tissue specific methylation. There are two known genes, *NPC1* and ribosomal DNA gene, proved to be control by DNA methylation from this study.

Homozygous or double heterozygous mutation of NPC1 can cause Niehmann-Pick disease type C (NPC). This is a lipid storage disease that can present in infants, children, or adults. Neonates can present with ascites and severe liver disease and/or respiratory failure from infiltration of the lungs (Rutledge; 1989). Other infants, without liver or pulmonary disease, can present hypotonia and development delay. The classic presentation is in middle to late childhood with the insidious onset of ataxia, vertical supranuclear gaze palsy (VSGP), and dementia. Dystonia and seizures are common. Dysarthria and dysphagia become disabling, making oral feeding impossible (Brady et al; 1989). About 95% of patients with NPC have a defective *NPC1* gene which located on 18q11-q12 (Carstea et al; 1997), and about 5% are presumed to have mutations in an as yet unmapped and unidentified gene, NPC2. DNA analysis of the NPC1 gene in research studies reveals disease-causing mutations in about 70-90% of patients.

The diagnosis of NPC can be confirmed by biochemical testing that demonstrates impaired cholesterol esterification and positive filipin staining in cultured fibroblasts (Pentchev et al; 1985). The diagnosis often delayed substantially because many clinicians do not realize that the absence of organomegaly dose not rule out lipid storage disease and because routine tests for metabolic disease, such as urine screens and lysosomal enzyme panels, are normal in NPC. Biochemical testing for carrier status is unreliable. Molecular testing for carrier detection and prenatal testing which available on a research basis only, is primarily used to identity mutation of the *NPC1* gene in a proband for genetics counseling of family members. Other tests, including tissue biopsies and tissue lipid analysis, which were essential for diagnosis before recognition of biochemical defect in NPC, are now rarely needed. These tests include examination of bone marrow, spleen and liver, or brain may show polymorphous cytoplasmic bodies (Boustany et al; 1990).

NPC is inherited in an autosomal recessive manner. A couple who have had a child with NPC are at 25% risk in each pregnancy of having a affected child. The phenotype usually runs through in families. Unaffected sibs who are older than the proband have a 2/3 risk of carrying one abnormal *NPC* allele. Prenatal testing is possible for pregnancies at 25% risk using a) biochemical testing when the proband has a classical biochemical phenotype but not when the proband has a variant biochemical phenotype, b) molecular diagnosis (available on research basis) when the proband has identified mutations in the *NPC1* gene (Vanier et al; 1996).

The *NPC1* gene contains 25 exons, varying in size from 74 to 788 bp spread over 47 kb (Morris et al; 1999). The NPC1 gene product is an integral membrane protein with multiple transmembrane domains that appears to be localized to a late endosomal compartment. Its function is imperfectly understood, but it clearly plays a central role in modulating intracellular sorting of cholesterol and glycosphingolipids (Neufeld et al; 1999). Most patients are compound heterozygotes for point mutations producing missense and nonsense mutations; deletions and splice site mutations have also been reported. Deficiency of *NPC1* gene products leads to a complex pattern of intracellular lipid storage, including excess unesterified cholesterol, GM2 and GM3 gangliosides, lactosylceramide, and lysobiosphosphatidic acid. (Watari et al; 1999) Approximately 100 mutations have been identified. These are distributed throughout the gene with an apparent clustering in the region encoding the carboxy-terminal half of NPC1 protein (Greer et al; 1999). Insufficient data are available to attempt genotype-phenotype correlations. Our study showed that there are methylation in subpopulation of cells from all tissues or cell types. The role and consequence of this methylation on cell function remain to be elucidate.

The human ribosomal RNA genes (rDNA) are tandemly repeated on the p-arms of five pairs of acrocentric chromosomes. Each unit of rDNA contains 13kb transcribed region and 30 kb intergenic spacer (IGS). Expecting that among 400 copies of rDNA per diploid genome, variation would be arise. Earlier studies have shown a variability at two levels: (a) in the numbers of large repeated sequence blocks (LaVolpe et al., 1984; Erickson and Schmickel, 1985; Sylvester et al., 1986, 1989). This length variation arise from unequal homologous exchanges. (b) in the form of

microsatellite variation (gonzalez et al., 1985, 1990; Maden et al., 1987; Sasaki et al., 1987). This variation arises from slipped-strand mispairing during DNA replication.

The IGS contains internally repeated sequences, which can be classified as tandemly repeated blocks and as nontandem blocks. Varying numbers of the tandemly repeated blocks are seen as length variation among 400 copies of rDNA in a diploid genome (detectable on Southern blots). There also are small length variations among the repeat blocks due to variation in microsetellite-type repeats (detactable by sequencing).

Two short sequences that were reported in a publication that were reported in a publication that characterized them as hypermethylated in embryonal carcinoma cell DNA and hypermethylated in sperm DNA (Zhang *et al*., 1987) matched the IGS almost perfectly. These clones are clearly rDNA-derieved. They are: (1) Clone sp0.3-10 (Accession No. X06588), estimated at 400 copies per genome by Zhang et al.; 1987. The clone matches rDNA 28,593-28927. (2) Clone sp0.3-23 (Accession No. X07493), estimated at 1000 copies per genome. This second clone matches rDNA 21,378-21,719 and 23,871-24,193, in LR1 and LR2, respectively.

Sequence analysis indicates that rDNA base composition varies along its length, that sequence motifs associated with specific functions are present, that many retroposons are present, and that mutation rates very along its length. The 90bp repeat blocks found 2.3 kb upsteam of the main rDNA promoter fit this description, and preliminary evidence from our laboratory

suggests that they might be a part of a larger region that increases *in vitro* transcription activity.

Recombination and gene conversion are driving force behind the concerted evolution of tandem genes including rDNA (Arnheim et al, 1980). Direct evidence for recombination in rDNA includes: (1) The spread of retroposons at identical and defined positions in all copies of rDNA (Gonzalez et al., 1989, 1993). (2) The presence of variable numbers of the tandem R repeats (La Volpe, et al., 1984, 1985) and of variable numbers of 90 bp blocks (Sylvester et al., 1989). (3) A chromosomal translocation that joined a 28S gene region consisting of [GGC] repeats with a similar region of the dystrophin gene (Bodrug et al., 1987). (4) The instability of multiple rDNA units in YACs: no YACs containing more than one rDNA unit could be stably cloned (labella and Schlessinger, 1989).

Our data was surprising since ribosome is house keeping gene but methylation usually causes gene inactivation. It would be very interesting to further explore whether there is distinct expression of each rDNA locus among each cell type according to DNA methylation. The nature of the methylation sites is not yet clear, although exocriments with primate cells indicate that 5meCpG is among the patterns capable of directing strand selection (Hare and Taylor, 1985). In fact, the presence of methylated CpGs in the transcribed region has been demonstrated by Kawasaki et al. (1992)

CHAPTER VII

Conclusion

From ms-RDA method, 6 clones revealed hypermathylated sequences in white blood cells, 3 clones had a GC content > 50% and satisfied the minimal criteria for CpG islands (200 bp, GC content >50%, CpG/GpC >0.5). Furthermore, 5 of these 6 clones located on 3 human single copy gene sequences. 2 clones (Bt20 and Bt75) were identical to human cDNA sequences located on chromosome 19q13.2. One (Bt38) was upstream sequence of *NPC1*. The other two clones were intergenic sequence of ribosomal DNA. The sixth clone, Bt44, had no information of sequence identity. Interstingly, further exploration of *NPC1* methylation status by MS-PCR showed methylated and unmethylated DNA sequences in all cell types, including sperm. This suggested that all cells are mosaic regarding methylation of this gene. This information is crucial to further explore the physiologic role of tissue specific DNA methylation and cholesterol transportation.

REFERENCE

Antequera F, Bird A. Number of CpG island and genes in human and mouse. Proc. Natl Acad Sci USA. 1993; 90: 11995.

Anteque F, Bird A. CpG islands as genomic footprinting of promotors that are associate with replication origins. Current biology 1999; 9: R661-R667.

Bestor TH. Cytosine methylation and unequal developmental potentials of the oocyte and sperm genomes. Am J Hum Genet. 1998; 62(4): 1269-1273.

Bestor TH, Tycko B. Creation of genomic methylation pattern. Nat Gene. 1996; 12: 363-366.

Bird A, Taggart M, Frommer M, Miller OJ, Macleod D. A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. CELL. 1985; 40(1): 91-95

Bird AP. Trends Genet. 1987; 3:343.

Bohinski RC. (1987) Modern Concepts in Biochemistry. Fifth Edition. Allyn and Bacon, Inc.

Boustany RN, Kaye E, Alroy J. Ultrastructural finding in skin from patients with Niehmann-Pick disease, type C. Pediatr Neurol 1990; 6:177-183.

Crady RO, Filling-Katz MR, Barton NW, Pentchev PG. Niehmann-Pick types C and D. Neurol Clin 1989; 7:75-88.

Carstea ED, et al. Niehmann-Pick C1 disease gene: homology to mediators of cholesterol homeostasis. Science 1997; 277: 228-231.

Cavillé J, *et al*. Identification of brain-specific and imprinted small nucleolar RNA gene exhibiting an usual genomic organization. <u>PNAS</u>. 2000; 97(26): 14311-14316.

Cedar H. DNA methylation and gene activity. <u>CELL.</u> 1988; 53: 3-4.

Constĩncia M, Pickard B, Kelsey G, Reik W. Imprinting mehanism. <u>Genome Research.</u> 1998; 8: 881-900.

Cross SH, Bird AP. CpG island and genes. <u>Curr Opin Genet Dev</u> 1995;5:309.

Cui H, Horon IL, Ohlsson R, Hamilton SR, Feinberg AP. Loss of imprinting in normal tissue of colorectal cancer pateints with microsatellite instability. <u>Nature Medicine</u>. 1998;4(11):1276-1280.

Delgado S, Gomaz M, Bird A, Antequera F. Initiation of DNA replication at CpG islands in mammalian chromosomes. <u>The EMBO Journal.</u> 1998; 17(8): 2426-2435.

Driscoll DJ, Migeon BR. Sex difference in methylation of the single copy genes in human meiotic germ cells: implication for X chromosome inactivation, parental imprinting, and origin of CpG mutation. <u>Somatic Cell Mol Genet</u> 1990; 16: 261-282.

Duthie SM, *et al*. *Xist* RNA exhibits a banded localization on the inactive X chromosome and is excluded from autosomal material in /cis/. <u>Human Molec Gent.</u> 1999; 8: 195-204.

Esteller M, Sanchez-Cespedes M, Rosell R, Sidransky D, Baylin SB, Herman JG. Detection of aberrant promotor hypermethylation of tumor supressor genes in serum DNA from non-small cell lung cancer patients. <u>Cancer Research</u>. 1999; 1(59): 67-70.

Farrell WE, Simpson DJ, Frost ST, Clayton RN. Methylation mechanism in pituitary tumorigenesis. <u>Endrocr Relat Cancer</u> 1999; 6(4):437-447.

Feinberg AP, Vogelstein B. Hypomethylation distinguishes genes of some human cancers for their normal counterparts. Nature. 1983; 301: 89.

Forrester WC, Fernandez LA, Grosschedl R. Nuclear matrix attechment regions antagonize methylation-dependent repression of long range enhancer promotor interactions. Genes & Development. 1999; 13: 3003-3014.

Gardiner-Garden M, Frommer M. CpG islands in vertebrate genomes.J Mol Biol. 1987; 196: 261-282.

Gonzalez IL, Sylvester JE. Complete sequence of 43-kb human ribosomal DNA repeat analysis of the intergenic spacer. GENOMICS. 1995; 27: 320-323.

Goto T, Monk M. Rehulation of X-chromosome inactivation in development in mice and humans. Microbiol Mol Biol Dev 1998; 62(2): 362-378.

Graff JR, Gabrielson E, Fujii H, Baylin SB, Herman JG. Methylation pattern of the E-cadherin 5' CpG island are unstable and reflect the dynamic, heterogeneous loss of E-cadherin expression during metastatic progression. The J of Biological Chemistry. 2000; 275(4): 2727-2732.

Graff JR, Herman JG, Lapidus RG, Chopra H, Xu R, Jarrard DF, Isaacs WB, Pitha PM, Davidson NE, Baylin SB. E-cadherin expression is silenced by DNA hypermethylation in human breast and prostate carcinomas.Cancer Res. 1995 Nov 15;55(22):5195-5199.

Greer WL, Dobson MJ, Girouard GS, Byers DM, Riddell DC, Neumann PE. Mutations in NPC1 highlight a conserved NPC1-specific cysteine-rich domain. Am J Hum Genet 1999; 65: 1252-1260.

Greger V, Debus N, Lohmann D, Hopping W, Passarge E, Horrsthemke B. Frequency and parental origin of hypermethylated *RB1* allele in retinoblastoma. <u>Hum Genet</u> 1994; 94: 491.

Hanada M, Delia D, Aiello A, Stadtmuer E, Reed JC. bcl-2 gene hypomethylation and high level expression in B-cell chronic lymphocytic leukemia. <u>Blood.</u> 1993; 82: 1820.

Herman JG, Latif F, Weng Y, Lerman MI, Zbar B, Liu S, Samid D, Duan DS, Gnarra JR, Linehan WM, et al. Silencing of the VHL tumor-suppressor gene by DNA methylation in renal carcinoma. <u>Proc Natl Acad Sci USA.</u> 1994 Oct 11;91(21):9700-9704.

Herman JG, Merlo A, Mao L, Lapidus RG, Issa JP, Davidson NE, Sidransky D, Baylin SB. Inactivation of the CDKN2/p16/MTS1 gene is frequently associated with aberrant DNA methylation in all common human cancers. <u>Cancer Res.</u> 1995 Oct 15;55(20):4525-4530.

Hoffman AR, Vu TH. Genomic Imprinting. <u>Scientific American SCIENCE & MEDICINE.</u> 1996 ; January/Febuary: 52-61.

Karp G. (1996) Cell and Molecular Biology, Concept and Experiments. John Wiley & Sons, Inc.

Keshet I, Lieman-Hurwitz J, Cedar H. DNA methylation affects the formation of active chromatin. <u>CELL.</u> 1986; 44: 535.

Korf BR. (1996) Human Genetics ; A Problem-based Approach. Blackwell Science, Inc.

Larsen F, Gundersen G, Lopez R, Prydz H. CpG island as gene markers in human genome. <u>Genomics.</u> 1992; 13: 1095-1107.

Lee JT, Davidow LS, Warshawsky D. *Tsix*, a gene antisen to *Xist* at the X-inactivation centre. <u>Nature Genet.</u> 1999; 21: 400-404.

Lee S, Shroyer KR, Markham NE, Cool CD, Voelkel NF, Tuder KM. Monoclonal endothelial cell proliferation is present in primary but not secondary pulmonary hypertension. J Clin Fnvest. 1998; 101(5): 927-934.

Lisitsyn N, Lisitsina N, Wigler M. Cloning the difference between two complex genomes. Science. 1993; 259: 946-951.

Lisitsyn NA. Representational difference analysis : finding the difference between genomes. TIG. 1995; 11(8) : 304-307.

Maden BE, Dent CL, Farrell TE, Garde J, McCallum FS, Wakeman JA. Clone of human ribosomal DNA containing the complete 18s rRNA and 28s rRNA genes charecterization, a detailed map of the human ribosomal transcription unit and diversity among clones. Biochem J. 1987; 246(22): 519-527.

Meehan R, Lewis J, Cross S, Nan X, Jeppesen P, Bird A. Transcriptional repression by methylation of CpG. J Cell Sci Suppl 1992; 16: 9-14.

Mertineit C, Yoker JA, Taketo T, Laird DW, Trasier JM, Bestor TH. Sex-specific exons control DNA methyltransferase in mammalian germ cell. Development. 1998; 125: 889-897.

Monk M, Boubelik M, Lehnert S. Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development. Development. 1987; 99: 371-382.

Momparler RL, Bovenzi V. DNA methylation and cancer. J Cell Physiol 2000; 183(2): 145-154.

Morison IM, Reeve AE. A catalogue of imprinted genes and parent-of-origin effects in humans and animals. Human Molecular Genetics. 1998; 7(10) :1599-1609.

Morris JA, Zhang D, Coleman KG, Nagle J, Pentchev PG, Carstea ED. The genomic organization and polymorphism analysis of the human Niehmann-Pick C1 gene. Biochem Biophys Res Commun 1999; 261: 493-498.

Moore T, *et al*. Multiple imprinted sense and antisense transcripts, differential methylation and tandem repeats in a putative imprinting control region upstream of mouse *Igf2*. Proc Natl Acad Sci. 1997; 94: 12509-12514.

Nan X, Tate P, Li E, Bird A. DNA methylation specifics chromosomal localization of *MeCP2*. Molecular and cellular biology. 1996; 16(1): 414-421.

Neufeld EB, et al. The Niehmann-Pick C1 protein resides in a vascular compartment linked to retrograde transport of multiple lysosome cargo. J Biol Chem 1999; 274:9627-9635.

Numata M., Ono T, Iseki S. Expression and localization of the mRNA for DNA(cytosine-5)-methyltransferase in mouse seminiferous tubule. J Histochem Cytochem. 1994; 42: 1271-1276.

Oswald J. Active demethylation of the paternal genome in the mouse zygote. Current Biology. 2000; 10: 475-478.

Paulson EJ, Speck SH. Differential methylation of Epstein-Barr Virus latency promotors facilitates viral persistance in healthy seropositive individuals. J of Virology. 1999; 73(12): 9959-9968.

Pentchev PG, Comly ME, Kruth HS, Vanier MT, Wenger DA, Patel S, Brady RO. Adefect in cholesterol esterification in Niemann-Pich disease (type C) patients. Proc Natl Acad Sci USA 1985; 82: 8247-8251.

Razin A, Cedar H. DNA methylation and gene expression. Microbiol Rev. 1991; 55: 451.

Razin A, Kafri T. DNA methylation from embryo to adult. Proc Nucl Acid Res. 1994; 48: 53-81.

Rutledge JC. Prgressive neonatal liver failure due to type C Niehmann-Pick disease. Pediatr Pathol 1989; 9: 779-784.

Sanford JP, Clark HJ, Chapman VM, Rossant J. Difference in DNA methylation during oogenesis and spermatogenesis and thair persistence during early embryogenesis in the mouse. Genes Dev. 1987; 1: 1039-1046.

Santoso B, Ortiz BD, Winoto A. Control of organ specific demethylation by an element of the T-cell receptor-alpha locus control region. J Biol Chem 2000; 275(3): 1952-1958.

Schutle M, et al. Identification by RDA of the homozygous deletion in pancreatic carcinoma that lies within the BRCA2 region. Proc Natl Acad Sci. 1995; 92: 5950-5954.

Siegfried Z, Eden S, Mendelsohn M, Feng X, Tsuberi B, Cedar H. DNA methylation represses transcription in vivo. Nature Genetics. 1999; 22: 203-206.

Singal R, Ginder G. DNA Methylation. Blood. 1999; 93(12): 4053-4070.

Strachan T, Read AP. Human Molecular genetics. Second edition. John Wiley&Sons Inc, New York, USA. 1999.

Tate PH, Bird AP. Effect of DNA methylation on DNA-binding protein and gene expression. Curr Opin Genet Dev. 1993; 3: 226.

Tazi J, Bird A. Alternative chromatin structure of CpG islands. CELL. 1990; 60: 990.

Tilghman SM. The sins of the fathers and mothers: genomic imprinting in mammalian development. CELL. 1999; 96: 185-193.

Tilghman SM, Caspary T, Ingram RS. Competitionn edge at the imprinted Prader-Willi/Angelman region. Nature Genetics. 1998; 18: 206-208.

Toyota M, et al. Identification of differentiallly methylated sequence in colorectal cancer by methylated CpG island amplification. Cancer Research. 1999; 59: 2307-2312.

Trasler JM. Origin and roles of genomic methylation patterns in male germ cells. CELL & DEVELOPMENTAL BIOLOGY 1998; 9: 467-474. Tulchinsky E, *et al*. Transcription analysis of *mls1* gene with specific reference to 5' flanking sequences. Proc Natl Acad Sci. 1992; 89: 9141-9150.

Vanier MT, Duthel S, Rodriguez-Lafrasse C, Pentchev P, Carstea ED. Genetic heterogeniety in Niehmann-Pick C disease: a study using somatic cell hybridization and linkage analysis. AM J Hum Genet 1996; 58: 118-125.

Ushijima T, Morimura K, Hosoya Y, Okonogi H, Tatematsu M, Sukimura T, Nakao M. Establishment of methylation-sensitive-representational difference analysis and isolation of hypo- and hypermethylated genomic fragment in mouse liver tumors. Proc. Natl. Acad. Sci. USA. 1997; 94: 2284-2289.

Watari H, et al. Niehmann-Pick C1 protein: obligatory roles for N-terminal domains and lysosomal targeting in cholesterol mobilization. Proc Natl Acad Sci USA 1999; 96: 805-810.

Wutz A, Smrzka OW, Schweifer N, Schellander K, Wagner EF, Barlow DP. Imprinted expression of the *Igf2r* gene depends on and intronic CpG island. NATURE. 1997; 389: 745-749.

Yang T. et. al. A mouse model for Prader-Willi syndrome imprinting-center
mutations. Nature Genetics. 1998; 19: 25-31.

APPENDIX A

BUFFERS AND REAGENTS

1. **1M Tris (pH 7.0)**

Tris base	121.1 g
dH$_2$O	700 ml

   *adjust the pH to 7.0 by adding conc.HCl

   Adjust volume to 1.0 litre with dH$_2$O, and sterlize by autoclaving.

2. **0.5M EDTA (pH 8.0)**

Disodium ethylenediamine tetraacetate.2H$_2$O	186.6 g
dH$_2$O	700 ml

   *adjust the pH to 8.0 by adding conc.NaOH

   Adjust volume to 1.0 litre with dH$_2$O, and sterlize by autoclaving

3. **7.5M Ammonium acetate (CH$_3$COONH$_4$)**

Ammonium acetate	57.81 g
dH$_2$O	80 ml

   Adjust volume to 100 ml with dH$_2$O, and sterlize by autoclaving

4. **3M Sodium acetate (CH$_3$COONa) (pH 5.3)**

Sodium acetate	40.82 g
dH$_2$O	80 ml

   *adjust the pH to 5.3 by adding conc.HCl

   Adjust volume to 100 ml with dH$_2$O, and sterlize by autoclaving

5. **20 mg/ml Proteinase K (stock solution)**

    Dissolve Proteinase K   20   g

    dH$_2$O   1   ml

    *store at -20 ^{o}C

6. **10% Sodium dodecyl sulfate (SDS)**

    SDS (electrophoresis grade)   100   g

    dH$_2$O   870   ml

    *adjust the pH to 7.2 by adding conc.HCl (a few drop)

    Adjust volume to 1.0 litre with dH$_2$O.

7. **Digestion buffer**

    100 mM NaCl

    10   mM Tris-Cl (pH 8.0)

    11   mM EDTA (pH 8.0)

    0.5%   SDS

    0.1 mg/ml PK*

    *PK is labile and must be added fresh with each use.

8. **10x Ficoll loading buffer**

    Ficoll   25   g

    Bromphenol blue   0.025   g

    0.5M EDTA (pH 8.0)   0.2   Ml

    Adjust volume to 10 ml with dH$_2$O.

    Store at -20 ^{o}C

9. **Neutrolizing solution**

Tris base	60.57	g
dH$_2$O	700	ml

*adjust the pH to 7.2 by adding conc.HCI

Adjust volume to 500 ml with dH$_2$O.

0.5M EDTA (pH 8.0)	2	ml
NaCl	87.6	g

Adjust volume to 1 litre with dH$_2$O.

10. **Denaturing solution**

NaCl	87.6	g
NaOH	20	g
dH$_2$O	800	ml

Adjust volume to 1.0 litre with dH$_2$O.

11. **0.8% Agarose gel (w/v)**

Agarose	0.8	g
1x TBE	100	ml

Dissolve by heating and occational ixing untill no granules of agarose are visible

Add ethidium bromide 50 μg (0.5μg/ml)

12. **TE buffer**

Tris base	1.21	g
5M EDTA	200	μl

Adjust pH to 7.5 with conc. HCl

Adjust volume to 1.0 litre with dH$_2$O

13. Hybridization cocktail

100% Formamide	5.0	ml
20x SSPE	2.5	ml
100X Danhardt's solution	0.5	ml
10% SDS	0.1	ml
Salmon sperm	100	μg/ml
dH$_2$O	1.9	ml

14. 10x TBE buffer (pH 8.3)

Tris base	121.1	g
EDTA.2H$_2$O	3.7	g
dH$_2$O	800	ml
Slowly add the boric acid, anhydrous	55.6	g

*adjust the pH to 8.3 by adding conc.HCI

Adjust volume to 1.0 litre with dH$_2$O.

15. LB broth

Tryptone	10	g
Yeast extract	5	g
NaCl	10	g

Adjust pH to 7.0 with 5N NaOH

Adjust volume to 1.0 litre with sterile water, and sterlize by autoclaving 25 minute,

Cool to 50 ^{o}C or below.

Add antibiotic (Chloramphenical)	50	μg/ml

16. **LB agar**

Tryptone	10	g
Yeast extract	5	g
NaCl	10	g
Agar	20	g

Adjust pH to 7.0 with 5N NaOH

Adjust volume to 1.0 litre with sterile water, and sterlize by autoclaving 25 minutes, Cool to 50 °C or below.

Add antibiotic (Chloramphenical)	50	μg/ml

Pour into petri dishes (~25 ml/100 mm plate)

17. **NZY$^+$ broth**

NZ amine (casien hydrosylate)	10	g
Yeast extract	5	g
NaCl	5	g

Adjust volume to 1.0 litre with sterile water, and sterlize by autoclaving 25 minutes.

Add the following supplement before use

MgCl$_2$	1.25	ml
MgSO$_4$	12.5	ml
20% Glucose (w/v)	20	ml

Filter sterilze.

18. **Lysis bufferII**

NaCl	292.20	g
0.5 M EDTA	48	ml

Adjust volume to 1.0 litre with dH$_2$O

19. **Lysis bufferI**

Sucrose	109.54	g
Tris base	1.21	g
MgCl$_2$	1.02	g
Triton X-100	10	ml

Adjust volume to 1.0 litre with dH$_2$O

20. **Phosphate buffered saline (PBS)**

NaCl	8.00	g
KCl	0.20	g
Na$_2$HPO$_4$	1.40	g
KH$_2$PO$_4$	0.24	g
dH$_2$O	800	ml

Adjust pH to 7.4 with conc. HCl.

Adjust volume to 1.0 litre with dH$_2$O

21. **Elution buffer**

Tris base	1.21	g

Adjust pH to 8.5 with conc. HCl

Adjust volume to 1.0 litre with dH$_2$O

22. **2X Wash buffer**

Tris base	1.21	g

Adjust pH to 7.5 with conc. HCl

NaCl	5.84	g
5M EDTA	0.5	ml

Adjust volume to 1.0 litre with dH$_2$O.

23. 2% X-gal

X-gal

Dimethylformamide (DMF)

Adjust volume to 100 ml with dH_2O.

24. 5x PCR buffer

335mM Tris base

Adjust pH to 8.8 at 25°C with conc. HCl

20mM $MgCl_2$

80mM $(NH_4)_2SO_4$

166µl/ml Borineserum albumin (BSA)

Adjust volume to 10 ml with dH_2O.

25. GTE solution

50mM Glucose

25mM Tris-Cl

10mM EDTA

Adjust pH to 8.0.

Autoclave and store at 4°C

26. NaOH/SDS

2M NaOH	100 µl
10% SDS	100 µl

Adjust volume to 1 ml with dH_2O.

APPENDIX B

Nucleotide sequence database for BLASTN

nr. The nr (nonredundant) database contains all nucleotide sequences present in GenBank, EMBL, and DDBJ. Italso contains nucleotide sequences obtained from PDB (sequences associated with 3-dimensional structures in the Brookhaven Protein Data Bank). Nr conprises only sequences that are normally well annotated, so it does not contain expr, essed sequence tag (EST), sequence-tagged site (STS), genome survey sequence (GSS), oe high-throughput genome (HTG) sequences. Although nr may contain multiple copies of similar sequence, identical sequences are merged into one entry. Tobe merged, two sequences must have identical lenghts and every nucleotide at every position must be same.

month. The month database contains all nucleotide sequences present in GenBank, EMBL, DDBJ, and PDB that were release within the last 30 days. Unlike nr, it also contains EST, STS, GSS and HTG sequences released within the last month.

EST. EST accesses a nonredundant copy of all ESTs present in GenBank, EMBL and DDBJ (Boguski et al., 1993). ESTs are short sequences, a few hundred nucleotide in length, which are derivedby partial, single-pass sequencing of insert of randomly selected cDNA clones (Adams et al., 1991). Since the number of ESTs is increasing rapidly, it is an important database to search for noval cDNAs. As of August, 1998, ~70% of

the sequence in GenBank were ESTs; of these, 61% were from human, 20% from mouse.

STS. STS contain a nonredundant copy of all STSs present in GenBank, EMBL, and DDBJ. An STS is a short unique genomic sequence that is used as a sequence landmark for genomic mapping efforts (Olson et al., 1989) As of August, 1998, 83% of the sequences in the STS database were from human.

HTGS. HTGS contains "unfinished" DNA sequences generated by the high-throughput sequencing centers (Ouellette and Boguski, 1997). A typical HTG record might consist of all the first-pass sequence data generated from a single cosmid, BAC YAC, or P1 clone. The record is composed of two or more sequence fragments that have a total lenght of ≥2 kb and contain one or more gaps. The sequences are normally updated by the sequencing centers as more data become available. A single accession number is assigned to this collection of sequences. The accession number dose not change as the record is updated, and only the most recent version of the record remain in GenBank. Phase 1 HTG sequences are unordered, unoriented contigs with gaps. Phase 2 HTG sequences are ordered, oriented contig with or without gaps. All HTG records contain a prominent warning that the sequence data is unfinished and may contain errors. When a record is considered finished, it becomes a Phase 3 HTG and is moved to the nr database with the same accession number. HTGS is a valuable source of new genomic sequences not yet in nr.

GSS. GSS includes short, single-pass genomic data identified by various means (Smith et al., 1994). Many of the sequence have been mapped.

As of August, 1998, 80% of the sequences in GSS were from human, 14% were from *Arabidopsis thalina.*

Alu. The Alu database contains representative *Alu* repeats from all *Alu* subfamilies (Claverie and Makalowski, 1994). If a query sequence containing an *Alu* repeat is used in a BLAST search of the above nucleotide databases, many of the resulting high-scoring hits will also contain *Alu* sequences. It may be useful, especially with a genomic sequence query, to perform a search of the Alu database to identity the location of any *Alu* repeats that might produce high-scoring and potentially misleading hits in queries of other database.

Vector. The vector database contains nucleotide sequences of a nuber of standard cloning vectors. New sequences should be screened against the Vector database to assure that they do not contain any Vector contamination.

Mito. The Mito database contains representative mitochondrial sequences from many families. Nuclear-derived sequences may be screened against the Mito database to assure that they do not contain any mitochondrial contamination.

APPENDIX C

Sequence Identifier Syntax

The syntax of sequences header line used by the NCBI BLAST server depends on the database from which each sequence was obtained. Table A lists the identifiers for the database from which the sequences were derived. For example, and identifiers for the database from which the sequences were derived. For example , and identifier might be *gb|M73307|AGMA13GT,* where the gb tag indicates that the identifier refers to a GenBank sequence, M73307 is its GenBank accession number, and AGMA13GT is its GenBank locus.

NCBI assigns gi identifiers for all sequences contained within NCBI's sequence databases (Ostell and Kans, 1998). The gi identifier provides a uniform and stable naming convention whereby a specific sequence is assigned its unique gi identifier. If the nucleotide or protein sequence changes, a new gi identifier is assigned, evenif the accession number of the record remains unchanged. Thus, gi identifiers provide a mechanism for identifying the exact sequence that was used or retrieved in the given search.

For search of the nr protein database where the sequence are derived from conceptual translations of sequences from the nucleotide database the gi syntax is *gi|gi_identifier.* An example would be *gi|451623 (U04987) env|Simian immunodeficiency...,* where 451623 is the gi identifier and U04987 is the accession number of the nucleotide sequence from which it was derived.

Users may select the *-gi* option for BLAST output, which will produce a header line with the gi identifier concatenated with the database identifier of the database from which it was derived. For example, *gi|176485|gb|M73307|AGMA* would be useful a match from a nucleotide database, and *gi|129295|sp|P01013|OVAX_Chick* for a protein database.

The gnl (general) identifier allows database no listed in table 11.3.3 to be identified with same syntax.An example here is the PID identifier *gnl|PID|e1632*. PID stands for Protein-ID, and the e in e1632 indicates that this ID was issued by EMBL. As mentioned above, use of the *-gi* option produce the NCBI gi (in addition to the PID), which user can also use to retrive sequences of interest.

Table A   Identifier Syntax for Sequence Database

Database	Identifier syntax
GenBank	*gb|accession|locus*
EMBL Data Libraly	*emb|accession|locus*
DDBJ (DNA data bank of Japan)	*dbj|accession|locus*
NBRF PIR	*pir|entry*
Protein Research Foundation	*prf|name*
Swiss-Prot	*sp|accession|entry name*
Brookhaven Protein Data Bank	*pdb|entry|chain*
Patents	*pat|country|number*
GenInfo Backbone Id	*bbs|number*
General database identifier	*gnl|database|identifier*

APPENDIX D

Sequence of Adaptor and Primer

Adaptor Sequences

RHpa adaptor

    RHpa24 5'-AGC ACT CTC CAG CCT CTC ACC GAC-3'

    RHpa11 5'-CGCTCGGTGAG-3'

NHpa adaptor

    NHpa24 5'-AGG CAA CTG TGC TAT CCG AGG GAC-3'

    NHpa11 5'-CGG TCC CTC GG-3'

JHpa adaptor

    JHpa24 5'-ACC GAC GTC CAC TAT CCA TGA AAC-3'

    JHpa11 5'-CGG TTT CAT GGT-3'

Primer sequences

M13 primer

    M13-forward 5'-GTA AAA CGA CGG CCA GT-3'

    M13-reverse   5'-AAT TAA CCC TCA CTA AAG GG-3'

Primer for sequencing

    T7 primer 5'-GTA ATA CGA CTC ACT ATA GGG C-3'

AUTHOR BIOGRAPHY

Miss Jiranan Warachit was born on April 4, 1978 in Chumporn, Thailand. She recieved her Bachlor degree of Science in 1998 from the Department of Botany, Faculty of Science, Chulalongkorn University, Bankok, Thailand. She has enrolled Chulalongkorn University in graduate programme for Master degree of Medical Science since 1998.