CHAPTER IV

PART II: GENOME-WIDE DNA METHYLATION SIGNATURE IN CHOLANGIOCARCINOMA

1. Introduction

Epigenetic change is a heritable mechanism which regulates the pattern of gene expression and does not involve an alteration in the primary nucleotide sequence of a gene [44-45]. As a major epigenetic mechanism, DNA methylation has been described to play the important role in normal cells for X-chromosome inactivation, control of imprinting genes, suppression of testis specific genes, and regulation of cell type specific expression patterns [48]. At present, DNA methylation study has been increasing in term of the relation to several human diseases including cancer. It is obvious that aberration of DNA methylation is pivotal in the development of most cancers, with genome-wide and gene-specific hypermethylation and hypomethylation of genes and genome regions [14, 165]. Interestingly, the reversibility of epigenetic mechanism provides an opportunity for novel anti-cancer agents such as DNA methyltransferase inhibitors to re-express epigenetically silenced regulatory genes [64].

CCA is a fatal malignancy originating from biliary epithelium which classified into intrahepatic (ICC) and extrahepatic cholangiocarcinoma (ECC). CCA accounts for around 10-25% of the primary liver cancers in most parts of the world [103]. CCA has been considered as a rare disease. However, CCA is difficult to diagnose, aggressive and has late clinical stage presentation, low responsive rate by existing chemotherapy and non-surgical therapeutic regimes in vast majority of patients. The potentially curative of surgical resection has been reported, but only few patients are suitable for this treatment option [5-6, 166]. Globally, the incidence rates of CCA markedly vary among different areas which related to the distribution of risk factors. The highest incidence has been reported in the endemic area of liver fluke infection-the northeastern Thailand with age-standardized incidence rate (ASIR) up to 317.6 per 100,000 people [4]. It exceeds hundred times higher than in western countries where

the ASIR is less than 2 per 100,000 population [167-170]. Several reports have indicated the rise of incidence and mortality rates of CCA, particularly ICC worldwide [167, 169, 171-174]. Aberrant methylation profiles in CCA have been described by several studies [73-74, 175], although, only a few of the DNA methylation changes have been uncovered for this fatal cancer. This might be due to the fact that there are limitation of techniques used in those studies detecting DNA methylation in only candidate loci using, for example, methylation specific PCR (MSP). To overcome this challenge, we applied high-throughput technological advances of epigenetic assessment to quantify DNA methylation level in genome-wide scale using Illumina's Infinium® HumanMethylation27 BeadChip. The Methylation BeadChip represents the methylation levels of 27,578 CpG sites (within and out of CpG-islands; CGI) spanning 14,485 genes in human genome [84]. This technology can provide an appropriate system to investigate DNA methylation signature of cholangiocarcinoma as biomarkers for early detection, diagnosis, and prognosis.

2. Materials and methods

2.1 Samples

Genomic DNA samples of 32 primary intrahepatic CCA (Thailand; n=27, West; n=5), 6 matched pairs of adjacent normal samples (Thailand), 5 CCA cell lines (Thailand; n=3, West; n=2) and a normal biliary cell line (MMNK1) were used in our study. All Thai samples and CCA cell lines were kindly supplied by the Liver Fluke and Cholangiocarcinoma Research Center, Khon Kaen University (Khon Kaen, Thailand), and 27 samples of Thai primary CCA were divided into short (n=13) and long survival time (n=14) groups. The Western samples were provided by Department of Histopathology, Hammersmith Hospital Campus (Imperial College, London, UK). A biliary normal cell line (MMNK1) was provided by Dr. Naoya Kobayashi (Department of Gastroenterological Surgery, Transplant and Surgical Oncology, Okayama University Graduate School of Medicine, Dentistry and Pharmaceutical Sciences, Okayama, Japan). The list of samples was detailed in Appendix J, Table J-1.

2.2 Bisulfite conversion and Infinium Methylation Assay

DNA samples were modified with sodium bisulfite using the EZ DNA Methylation Gold kit (Zymo Research, Cambridge, UK) according to the manufacturer's instruction. For the Infinium Methylation Assay, 4 μl of bisulfite converted DNA (~200 ng) was used in the whole genome amplification (WGA) reaction. The procedure of WGA reaction and Methylation BeadArray were performed following the standard Infinium Methylation Assay protocols as previously described [84]. The ratio of fluorescent signals from the BeadChip was computed as β-values between 0 and 1 which reflects the methylation level of each CpG site [176]. The Methylation BeadChip represents the methylation levels of 27,578 CpG sites (within CGI; n=15,369 and out of CGI; n=12,209) in human genome.

2.3 Pyrosequencing

To evaluate the reliability of Methylation BeadChip data, we validated methylation status of selected genes showing significantly differential methylation using Pyrosequencing. Pyrosequencing assays were designed to detect methylation of 2 homeobox genes including *HOXA9* and *HOXD9*. In brief, bisulfite converted DNA which prior modified using EpiTect® Bisulfite Kit (Qiagen, West Sussex, UK) was amplified with specific PCR primers for each gene and PCR product was then sequenced with specific sequencing primer of each assay as described in Tables 4-1 and 4-2 for the sequences and conditions. For pyrosequencing, we used the PyroGold Reagent kit on a Pyrosequencing 96HS plate according to the manufacturer's protocol. The sequencing procedure was performed following to the manufacturer's instruction (Biotage, Uppsala, Sweden). The methylation percentage at an individual CpG site was generated and analysed as quantitative information using the Pyro Q-CpG^{TM} software (Biotage).

Table 4-1 Primer sequences and PCR product sizes of *HOXA9* and *HOXD9* for pyrosequencing.

CpG-island	Primer sequence	Product size (bp)
HOXA9	F: 5′-biotin-GGTGATGGTTATTATTGGGGTTT -3′ R: 5′-AAAAACTAACCCAAAATCCC-3′ S: 5′--CTAACCCAAAATCCCC-3′	120
HOXD9	F: 5′-GGGTATATGGATTTGGGTTTG-3′ R: 5′-biotin-TCCCACCCAACATTACACTATC-3′ S: 5'-TAATTATAGTAGTTTTTAGTTTA-3'	228

F: forward, R: reverse, S: sequencing

Table 4-2 PCR conditions for pyrosequencing of *HOXA9* and *HOXD9*.

Reagent	Volume (µl)	
	HOXA9	***HOXD9***
Sterile water	16.2	16.2
10X PCR buffer with $MgCl_2$	4.5	4.5
Primer mix (F+R; 5 µM each)	1	1
10mM dNTPs	0.5	0.5
Fast start *Taq* DNA polymerase	0.8	0.8
Bisulfite modified DNA	2	2
Total	25 µl/reaction	
Thermal cycles:		
Step 1: Denaturation; 95 °C, 6 min		
Step 2: Denaturation; 95 °C, 30 sec, Annealing; 63 °C for *HOXA9*, or 60 °C for *HOXD9*, 30 sec		
Step 3: Extension; 72 °C, 30 sec (repeat steps 2 and 3; 35 cycles)		
Step 4: Final extension; 72 °C, 5 min		

2.4 Statistical and bioinformatics analysis

For pre-processing data, quartile normalization was applied to β-value across all the samples. Assay reproducibility was determined from the correlation of technical replicates for each run of the BeadChip analysed using linear regression by comparing β values of each replicate data set. Hierarchical clustering (Euclidean distance and average linkage methods) was analyzed to visualize and indentify the separation of DNA methylation patterns between primary CCA, cell lines and adjacent normal samples. Hierarchical clustering analysis was performed using MultiExperiment Viewer (MeV version 4.5, http://mev.tm4.org). To avoid the effect of gender-related DNA methylation, we excluded CpG sites on the X and Y chromosomes from further analysis. Therefore, 26,486 autosomal CpG sites of both within and out of CpG-islands (n=14,862 and 11,624 CpG sites, respectively) were used for hierarchical clustering. Differential methylation was analyzed using the statistic test in limma to select CpG sites with false discovery rate (FDR) <0.05 and which have greater average difference between normal and tumors compared to the difference between replicates (FDR <0.05). Specific functional term analysis of CpG sites showing differential DNA methylation was determined using the Database for Annotation, Visualization and Integrated Discovery (http://david.abcc.ncifcrf.gov). Gene Set Enrichment Analysis (GSEA, R software using GSA package, http://www.r-project.org) was used to determine if predefined sets of genes are differentially methylated in CCA compared to adjacent normal samples [191]. For the validation of methylation results, the association of methylation β-values from the BeadArray and methylation levels detected by Pyrosequencing was analysed using Spearman correlation for each selected gene.

3. Results

3.1 Methylation assay reproducibility

Technical replicates were performed for each Methylation BeadChip using the same converted DNA sample. Bisulfite-modified DNA from PEO1 ovarian cancer cell line was used for this purpose. In this study, 5 data sets of replicate from each BeadChip were analysed to examine assay reproducibility. We obtained highly reproducible methylation profiles between these technical replicates with an average R^2=0.989 (range 0.988 to 0.990, P<0.000001) by β-value comparisons (Figures 4-1, A and B).

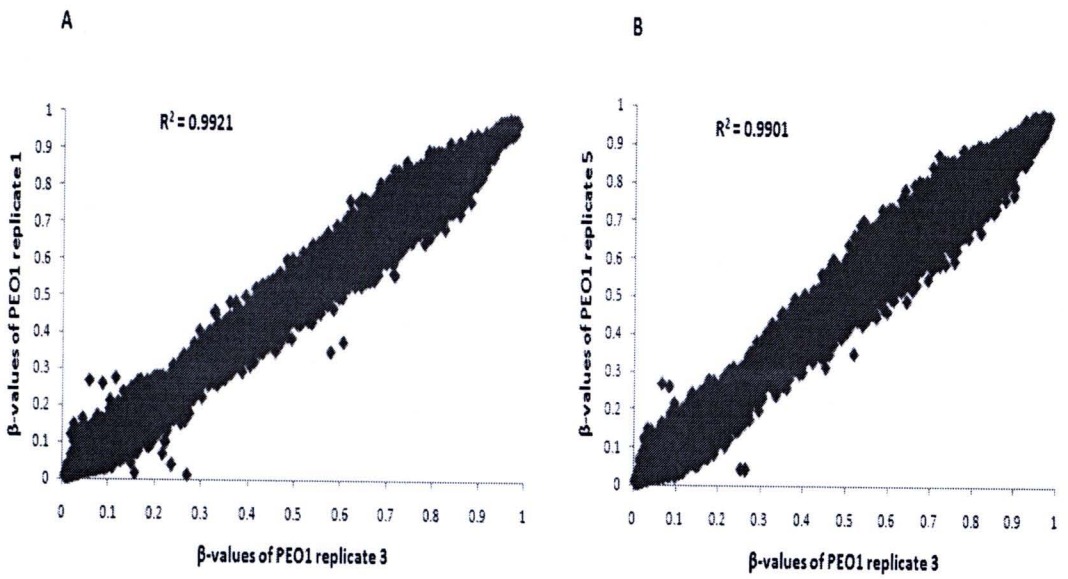


Figure 4-1 Methylation assay reproducibility. The β-values (for all 27,578 CpG sites) obtained from one replicate experiment were plotted against those from another technical replicate. (A and B) Representative plots showed the reproducibility of different technical replicates on bisulfite-converted DNA derived from PEO1 ovarian cancer cell line.

3.2 DNA methylation profiles in cholangiocarcinoma and normal

Unsupervised hierarchical clustering was applied to visualize and indentify the quantitative β-values of each CpG dinucleotides obtained from the set of samples examined. The clustering showed that DNA methylation profiles of all autosomal CpG sites (n=26,486) within and out of CpG-islands of cell lines, primary CCA and adjacent normal samples had similar pattern of clusters. Two major clusters were found to separate between cell lines and primary tissue samples. The close clustering was found between primary CCA samples and CCA cell lines whereas normal biliary cell line (MMNK1) was individually separated. Moreover, primary CCA samples from Thailand and the West were also grouped together therefore we combined them as the same tumor group for further analysis. Two distinguished subgroups were found to separate between primary CCA and adjacent normal samples with few errors of co-clustering (Figures 4-2 A, B and C).

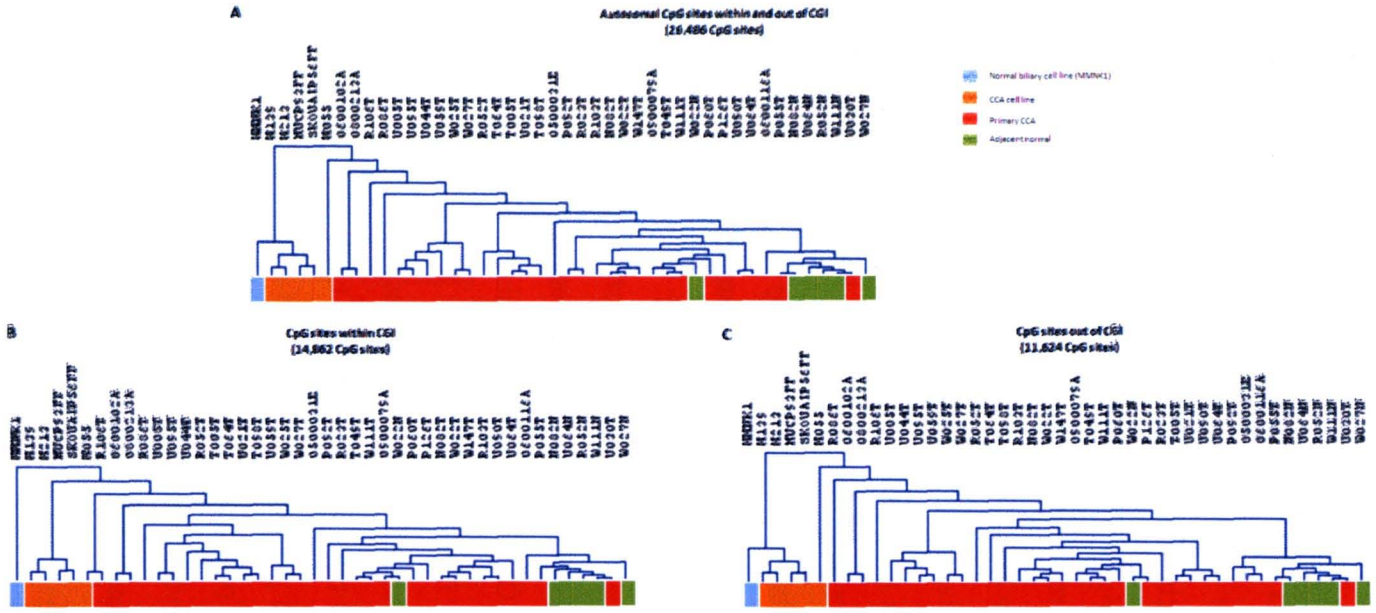


Figure 4-2 Hierarchical clustering of DNA methylation profiles. Clustering analysis of all autosomal CpG sites showed 2 clusters separating between cell lines and primary tissue samples. In the subgroups, primary CCA were closely clustered together as well as the cluster of normal adjacent samples (A). By clustering CpG sites within CGIs (n=14,862), primary CCA and CCA cell lines were closely clustered together whereas normal biliary cell line (MMNK1) was individually separated. All the adjacent normal and primary CCA were separated into two distinguished clusters except three tumors (0600116A, P055T and U030T) were co-clustered with normal samples as well as one normal was found in the cluster of primary CCA (B). The similar patterns of clustering were obtained for CpG sites out of CGIs (n=11,624) (C).

3.3 DNA methylation changes between primary CCA and adjacent normal samples

To identify which CpG sites contribute the most to DNA methylation signature of cholangiocarcinoma, we used the statistic test in package limma to select CpG sites with false discovery rate (FDR) <0.05 and have more dramatic average difference between normal and tumors compared to the difference between replicates (FDR <0.05). Hundreds of CpG sites within and out of CGI exhibiting differential methylation in 2 comparisons (6 matched pairs; normal vs. Thai CCA, and 6 normal vs. 32 CCA) were detailed as Table 4-3 and listed in Appendix J (Table J-2 and 4). The differential hypermethylation were found in 513 CpG sites within CGI and 258 CpG sites out of CGI. For CpG sites within CGI, two major clusters were found from the heat map of differential methylation of 513 CpG dinucleotides within CGI separating between high methylation (β-values) group of tumors and medium-low methylated group of samples (tumors and normal tissues). Primary CCA samples (Thailand and West) were closely clustered together displaying medium and high level of methylation whereas adjacent normal samples grouped in their own sub-cluster showing low methylation. There was co-clustering of 4 tumors (U030T, P055T, 0900079A and 0600116A) within the cluster of normal as represented in Figure 4-3A. Similar pattern of clustering was found among 258 differential hypermethylated CpG sites out of CGI. Again the heirarchial clustering separated adjacent normal samples from primary tumors with only a few tumors of co-clustering with the normals (Figure 4-3B). Similar to the hypermethylated CpG sites' clustering, differential hypomethylated CpG sites both within and out of CGI were also found to saparate adjacent normal samples and primary CCA (Figures 4-4, A and B) and the list of differentially hypomethylated CpG sites detailed in Appendix J (Tables J-3 and 5). In addition, the difference of methylation changes was not found to separate short and long survival group of CCA patients based on heirarchial clustering.

Table 4-3 Differentially methylated CpG sites normal and CCA samples.

Comparison	Limma	Number of CpG sites by average difference of β value	Number of differential CpG
CpGs within CGI			
6 matched pairs (Normal vs. CCA)	709	1452 (cutoff.uper = 0.08, FDR=0.050)	hyper:407
		499 (cutoff.lower = - 0.10, FDR=0.040)	hypo: 199
6 normal vs. 32 CCA	902	1665 (cutoff.uper = 0.08, FDR=0.052)	hyper: 513
		639 (cutoff.lower) = -0.09, FDR=0.044)	hypo: 203
CpGs out of CGI			
6 matched pairs (Normal vs. CCA)	1665	856 (cutoff.uper = 0.11, FDR=0.030)	hyper:458
		2120 (cutoff.lower = -0.09, FDR=0.050)	hypo:958
6 normal vs. 32 CCA	1057	625 (cutoff.uper = 0.11, FDR=0.042)	hyper: 258
		2029 (cutoff.lower = -0.09, FDR=0.052)	hypo: 651

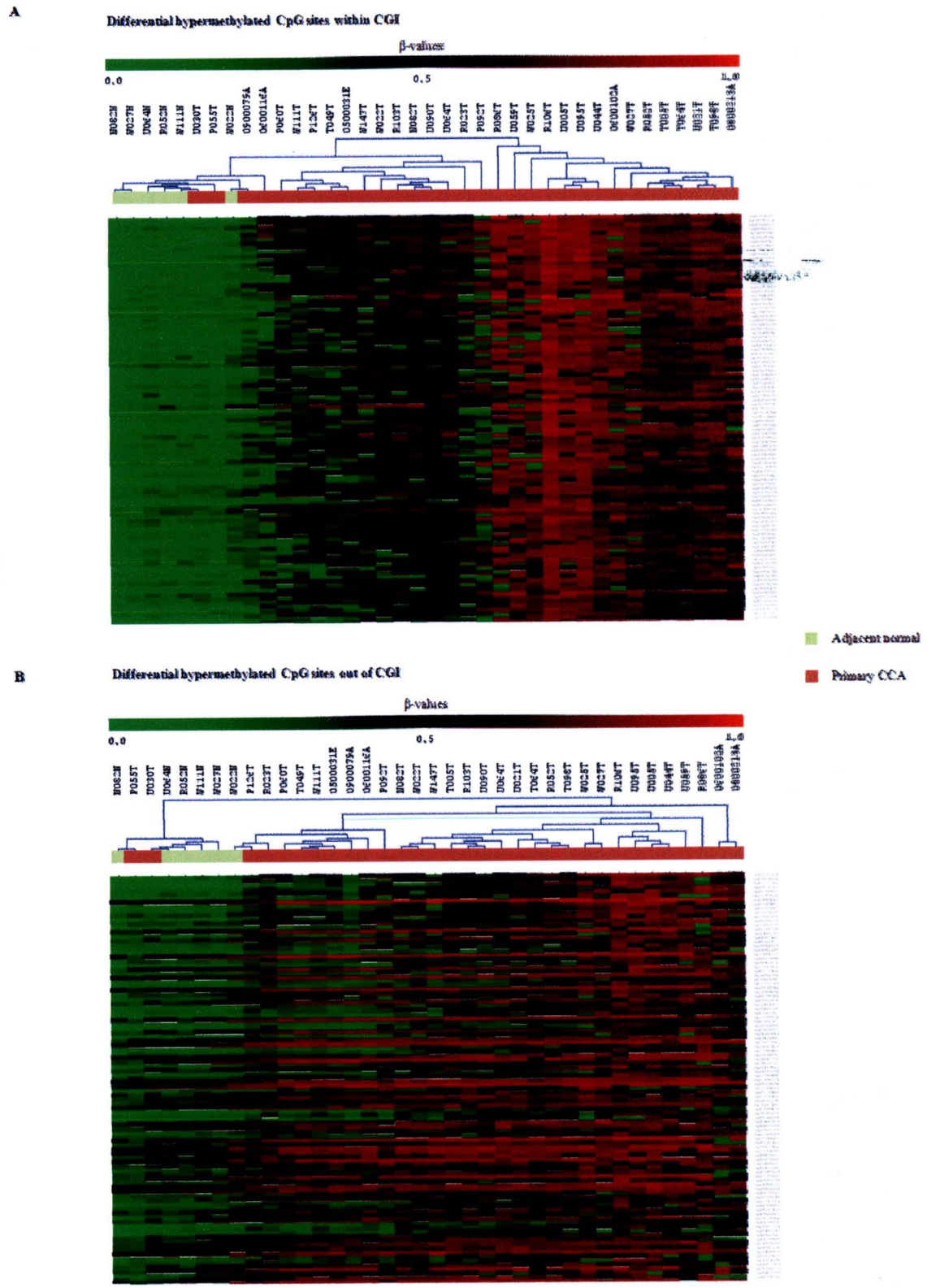


Figure 4-3 Differential DNA hypermethylated CpG sites within (A) and out of CGIs (B) between tumors and adjacent normal samples.

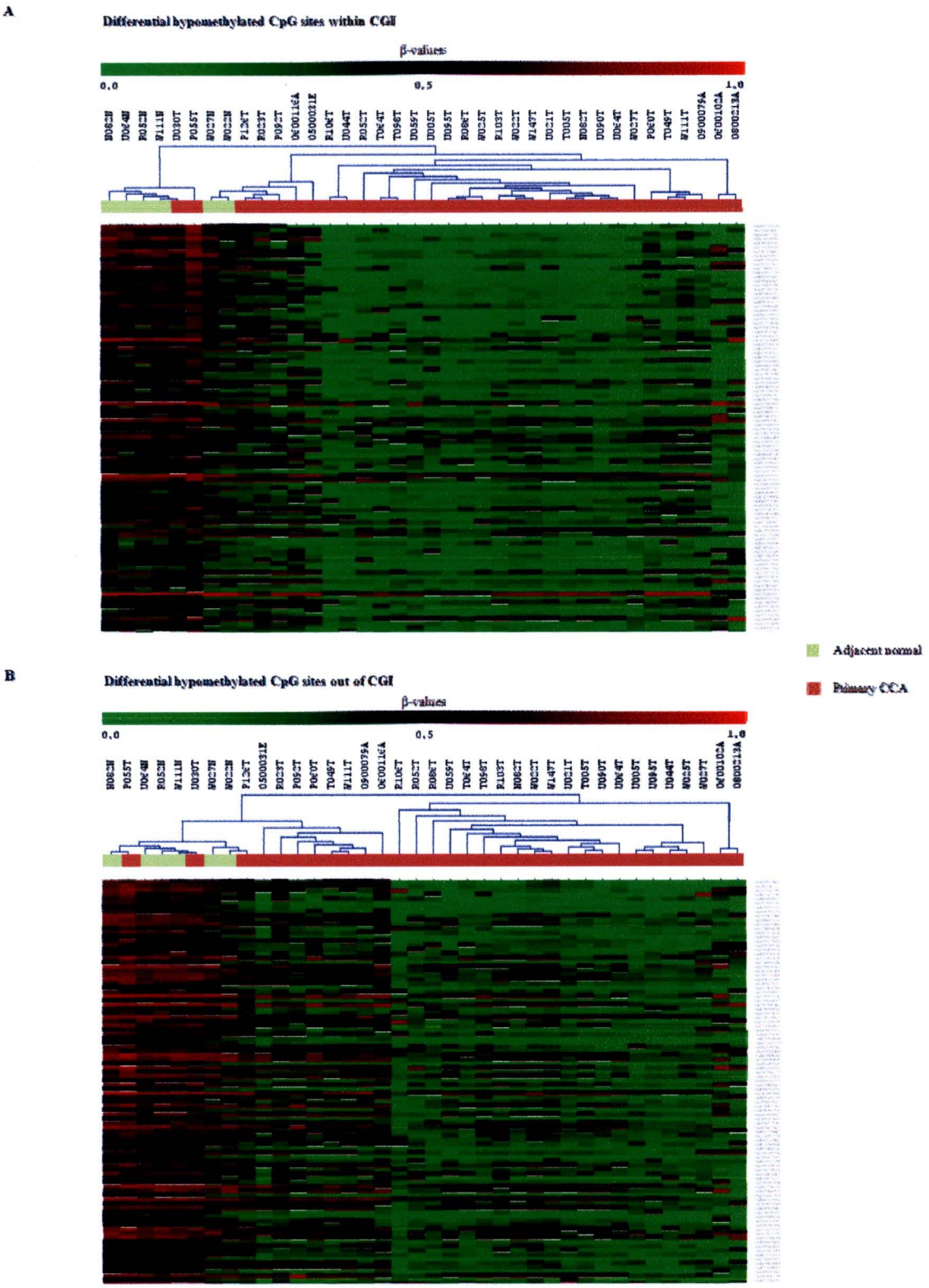


Figure 4-4 Differential DNA hypomethylated CpG sites within (A) and out of CGIs (B) between tumors and adjacent normal samples.

3.4 Specific functional terms of DNA methylation profiles

The specific functional analysis was performed in CpG sites showing hypermethylation and hypomethylation within or out of CGI obtained by the analysis using the Database for Annotation, Visualization and Integrated Discovery. There were a number of significant functional terms enriched in the genes associated with hypermethylated CGIs and CpG sites out of CGI (FDR <0.05) in comparison between adjacent normal and CCA (Thailand and West). Among hypermethylated CpG sites within CGI obtained, *homeobox* genes were noticeably found significant hypermethylation in CCA (FDR $=3.65 \times 10^{-11}$) as listed in Table 4-4 and hierarchical clustering shown in Figure J-1 (Appendix J). Moreover, we also found that the CpG sites within CGI hypermethylated in tumor samples from West were significantly associated with EGF like genes (FDR=0.036), but this is not observed in the tumor samples from Thailand. No specific functional terms were indentified in the hypomethylated CpGs within and out of CGI except immune system process identified in the hypomethylated CpG sites out of CGI in comparison between matched pair samples (adjacent normal vs. CCA, n=6) from Thailand (FDR=0.020) as differentially hypomethylated CpG sites listed in Table J-6 and hierarchical clustering shown in Figure J-2 (Appendix J).

Table 4-4 Homeobox genes associated with differentially hypermethylated CpG sites within CpG-islands.

Illumina probeID	Chromosome	Gene symbol	Average difference of β-value	Adjusted *P* value
cg08089301	17	HOXB4||MIR10A	0.587580174	0.000444431
cg14991487	2	HOXD9	0.565482409	0.000419343
cg26521404	7	HOXA9	0.556921432	0.00013706
cg15520279	2	HOXD8	0.52973297	0.000238346
cg14458834	17	HOXB4||MIR10A	0.529447356	0.004100231
cg06760035	17	HOXB4||MIR10A	0.513227818	0.003488409
cg08097882	13	POU4F1	0.495017636	0.004309362
cg21546671	17	HOXB4||MIR10A	0.491646159	0.007894958
cg01354473	7	HOXA9	0.476575288	0.000444431
cg21243096	1	POU3F1	0.467396053	0.00044688
cg19456540	14	SIX6	0.459741205	0.002465508
cg01381846	7	HOXA9	0.449383811	0.000238346
cg07109287	9	LHX2	0.442729689	0.01211636
cg00243313	5	IRX4	0.440490182	0.001914296
cg17241310	1	BARHL2	0.412076765	0.003150287
cg03963198	5	IRX4	0.397730023	0.002015882
cg23432345	7	HOXA7	0.396606992	0.00657172
cg27009703	7	HOXA9	0.396038394	0.000238346
cg12686016	7	HOXA1	0.395030924	0.010389371

Table 4-4 Homeobox genes associated with differentially hypermethylated CpG sites within CpG-islands (cont.).

Illumina probeID	Chromosome	Gene symbol	Average difference of β-value	Adjusted *P* value
cg26609631	13	GSX1	0.394847015	0.015120308
cg07778029	7	HOXA9	0.392103227	0.00932533
cg26504021	5	C5orf38||IRX2	0.391198295	0.001440328
cg18722841	11	PHOX2A	0.384689826	0.003231667
cg24199834	4	POU4F2	0.382765773	0.001597804
cg03700462	7	HOXA1	0.375279886	0.009118704
cg20291049	2	POU3F3	0.372026098	0.002961335
cg04317399	7	HOXA4	0.369970674	0.002578253
cg22660578	17	LHX1	0.369746106	0.014521661
cg15433631	5	C5orf38||IRX2	0.366626182	0.000422876
cg13262687	4	POU4F2	0.342572333	0.007078103
cg18750960	2	HOXD4||MIR10B	0.337913402	0.011343676
cg03874199	2	HOXD12	0.329747523	0.02077764
cg19576304	18	RAX	0.327835326	0.015174878
cg11525285	14	VSX2	0.321716318	0.017463069
cg23130254	2	HOXD12	0.321153174	0.002288042
cg02245378	2	CCDC140||PAX3	0.318902386	0.006742063
cg08876932	11	PHOX2A	0.309764318	0.015003336
cg27626299	7	EVX1	0.306819152	0.045749382
cg16632715	2	HOXD11	0.304487727	0.019335644
cg07175883	2	HOXD13	0.304038758	0.020695512

Table 4-4 Homeobox genes associated with differentially hypermethylated CpG sites within CpG-islands (cont.).

Illumina probeID	Chromosome	Gene symbol	Average difference of β-value	Adjusted *P* value
cg20073553	4	NKX3-2	0.30191703	0.028137524
cg15760840	7	HOXA11AS||HOXA11	0.300180667	0.01274376
cg21591742	2	HOXD10	0.296696242	0.005899971
cg02248486	7	HOXA5||HOXA6	0.293674	0.004878959
cg03285457	10	FLJ41350||LBX1	0.278315561	0.020007102
cg25047280	7	HOXA9	0.261830932	0.033055951
cg25596297	2	CCDC140||PAX3	0.257000894	0.043026675
cg14611174	14	SIX6	0.250822311	0.011729868
cg12128839	7	HOXA5||HOXA6	0.249603924	0.017792787
cg17233506	17	HOXB1	0.244479447	0.000926895
cg21815667	2	HOXD8	0.229934432	0.020299368
cg24650501	1	LMX1A	0.224619318	0.036132796
cg00767581	2	HOXD4||MIR10B	0.222300985	0.01465648
cg16977035	11	ALX4	0.183036091	0.003512148
cg07823492	17	HOXB1	0.154912	0.005376829

To demonstrate the utility of differential methylation between CCA and normal samples, we examined methylation states of 16 gene sets (ES1, ES2, NANOG, OCT4, SOX2, NOS, NOSTF, SUZ12, EED, H3K27, PRC1, E2F6, PRC2, PRC2_targets1, PRC2_targets2, homebox) using Gene Set Enrichment Analysis by comparison between adjacent normal and tumor samples from Thailand, as well as between adjacent normal and tumor samples from West. We observed the homeobox genes (P=0.014 and FDR=0.1387), EED targets (P=0.026 and FDR=0.1387) and PRC2_targets1 (P=0.026 and FDR=0.1387) hypermethylated in tumor samples from Thailand. Seventy seven out of 131 homeobox genes are also PRC2_targets1 and 94 out of 131 are EED targets. The overlap genes might be the ones mainly contributing to hypermethylation in tumor in all three gene sets. These three gene sets were also hypermethylated in tumors from the West (homeobox $P<10^{-6}$ and FDR$<10^{-4}$, PRC2_targets1 (P=0.03 and FDR=0.12), EED targets (P=0.04 and FDR=0.128). SUZ12 targets and H3K27 targets were also hypermethylated in tumors from West, which were the same trend as tumors from Thailand. For the gene sets with hypomethylation, OCT4 targets (P=0.026 and FDR=0.2514), PRC2 (P=0.046 and FDR=0.2514), NOS targets (P=0.048 and FDR=0.2514) were significantly different in tumors from West as well as NOS targets were also hypomethylated in CCA from Thailand (P=0.04 and FDR=0.2987) as shown in Tables 4-5 and 4-6 for the list of gene sets.

Table 4-5 Gene Set Enrichment Analysis of hypermethylation compared between CCA and adjacent normal samples.

Hypermethylation	Gene_set	Gene_set_name	Score	*P*-value	FDR
Thailand	16	homeobox	0.5533	0.014	0.1387
	9	eed	0.23	0.026	0.1387
	14	prc2_targets1	0.3953	0.026	0.1387
	10	h3k27	0.2487	0.05	0.1856
	8	suz12	0.2722	0.058	0.1856
West	16	homeobox	0.6951	0	0
	8	suz12	0.3858	0.018	0.12
	10	h3k27	0.386	0.028	0.12
	14	prc2_targets1	0.5022	0.03	0.12
	9	eed	0.306	0.04	0.128

Table 4-6 Gene Set Enrichment Analysis of hypomethylation compared between CCA and adjacent normal samples.

Hypomethylation	Gene_set	Gene_set_name	Score	*P*-value	FDR
Thailand					
	6	nos	-0.3273	0.04	0.2987
	4	OCT4	-0.2905	0.05	0.2987
	2	ES2	-0.4901	0.056	0.2987
	3	nanog	-0.2897	0.118	0.336
	1	ES1	-0.2808	0.12	0.336
	12	E2F6	-0.3232	0.126	0.336
	5	sox2	-0.2345	0.152	0.336
	15	prc2_targets2	-0.2078	0.168	0.336
	11	prc1	-0.2687	0.192	0.3413
	13	prc2	-0.2634	0.238	0.3808

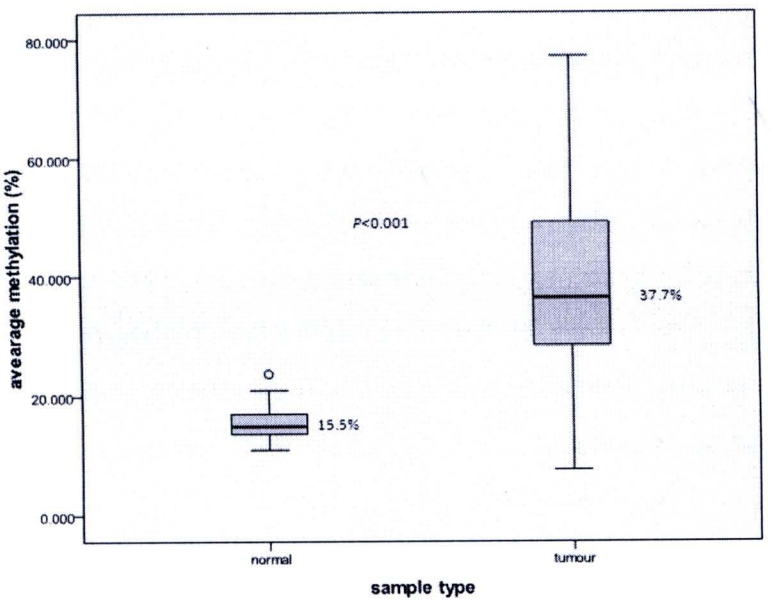
Table 4-6 Gene Set Enrichment Analysis of hypomethylation compared between CCA and adjacent normal samples (cont.).

Hypomethylation	Gene_set	Gene_set_name	Score	*P*-value	FDR
West					
	4	OCT4	-0.376	0.026	0.2514
	13	prc2	-1.2808	0.046	0.2514
	6	nos	-0.354	0.048	0.2514
	1	ES1	-0.575	0.066	0.2514
	3	nanog	-0.5016	0.084	0.2514
	5	sox2	-0.4782	0.1	0.2514
	2	ES2	-0.4434	0.11	0.2514
	12	E2F6	-0.499	0.128	0.256
	15	prc2_targets2	-0.1529	0.244	0.4338

3.5 Validation of Infinium HumanMethylation27 BeadChip by pyrosequencing

The methylation of selected genes from Infinium Methylation assay was validated using pyrosequencing. Assays were designed and optimized to detect methylation covered an identical CpG site on the Methylation BeadChip for selected genes. These included 2 homeobox genes (*HOXA9* and *HOXD9*) showing significantly differential methylation between CCA and adjacent normal tissues. We evaluated the correlation of our Infinium Methylation assay by comparing methylation β values with methylation levels detected by pyrosequencing of 2 homeobox genes. The quantitative pyrosequencing data of both CpG-islands showed strong correlation with data obtained from the Methylation BeadChip (Spearman correlation coefficient >0.9, $P<0.001$) (Appendix K, Figure K-1). By pyrosequencing, the results showed that *HOXA9* and *HOXD9* were significantly increased hypermethylated in primary CCA (> 85% of tumors) compared to adjacent normal samples (Figures 4-5, A and B) and the overall pyrosequencing results of *HOXA9* and *HOXD9* are presented in Figures K-2 and K-3 (Appendix K).

A. *HOXA9*



B. *HOXD9*

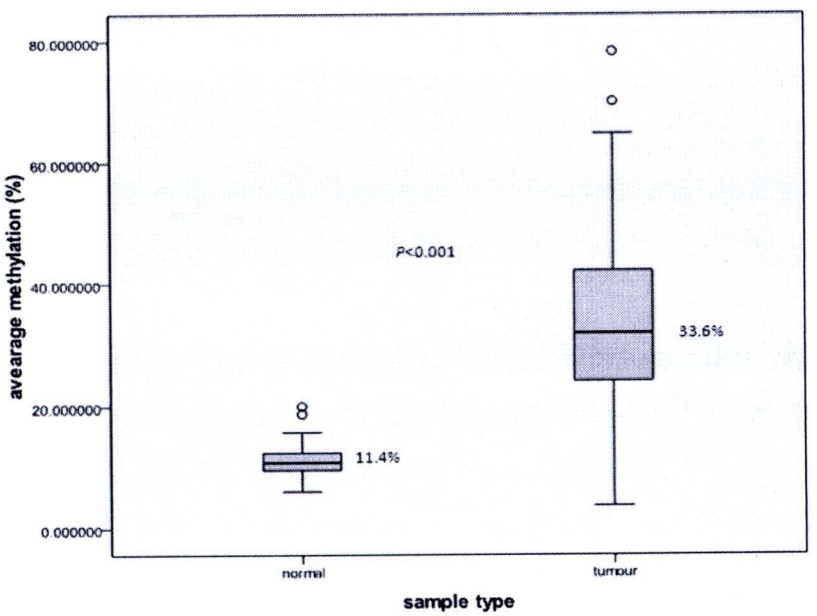


Figure 4-5 Boxplots represent differential methylation of *HOXA9* (A) and *HOXD9* (B) between adjacent normal and primary CCA detected by pyrosequencing.

4. Discussion

To our knowledge, this is the first report of genome-wide DNA methylation study in cholangiocarcinoma showing DNA methylation signatures those distinguish tumors from normal biliary samples. Regarding to the results from Infinium Methylation BeadChips, high reproducibility has been obtained from technical replicate of each BeadChip performed. We have clearly demonstrated that DNA methylation is a common molecular aberration in CCA. Both hypermethylation and hypomethylation of hundreds of CpG sites (within and out of CpG-islands) have been found being changes in CCA compared to adjacent normal samples. Although the difference of methylation changes is not found, it might be due to the effect of large scale data. Therefore, it is necessary to further focus on some candidate set of genes which are related to survival outcome of CCA patients. In the genome-wide methylation data, we have identified a number of specific functional terms including homeobox genes enriched in the genes associated with hypermethylated CpG sites located within and out of CpG-islands.

Our recent study in CCA, a number of differentially hypermethylated CpG sites have been found at genes that are associated with specific functional term of target genes of the Polycomb group (PcG) complexs including *homeobox* genes, EED target, PRC2 targets, SUZ12 targets and H3K27 target genes. Basically, PcG complexs are involved in transcriptional repression of genes. They play a role to maintain heritable transcription patterns of the *HOX* genes during development and differentiation, in conjunction with their positively acting counterparts known as Trithorax Group (TrxG) proteins [192-193].The previous report has described that stem cell PcG targets are up to 12-fold more likely to have cancer-specific promoter DNA hypermethylation than non-targets. These genes might be permanently repressed leading to loss of differentiation properties of the cells [187]. In addition we found the hypomethylation is significantly associated to OCT4, PRC2 and NOS targets which might lead to these genes overexpression. These genes are proposed as transcription factors to maintaining the pluripotent embryonic stem cell (ES) phenotype (self-renewal), especially OCT4 and NOS, and usually overexpressed in ES cell [188-189]. Several reports have described that tumors often lack the terminal differentiation traits possessed by their normal cell states. These parallels have supported the hypothesis

that cancer cells arise from undifferentiated stem or progenitor cells or, alternatively, that cancer cells can undergo progressive de-differentiation during their development [194-196]. Taken together, our finding should support the theory of stem cell origin in cancers, including CCA.

The strong correlation has been found in methylation level of selected *homeobox* genes, *HOXA9* and *HOXD9*, obtained from Illumina and pyrosequencing study indicating the reliability of our high-throughput methylation data. Although there was no any statistic significance between clinicopathological data and methylation of *HOXA9* or *HOXD9* based on pyrosequencing, primary CCA still had significantly high methylation of both loci (>85% of tumors) than those in adjacent normal samples. Our study indicates that *HOXA9* and *HOXD9* as well as other homeobox genes might serve as diagnostic methylation biomarkers in CCA. Previous studies have reported hypermethylation of *homeobox genes* in several types of cancer but limit publication in intrahepatic cholangiocarcinoma. Shiraishi and colleagues [177] have reported that many CpG islands at the *HOXA* and *HOXD* loci were methylated in human lung adenocarcinoma. In ovarian cancer, 51% of samples have been reported hypermethylation of *HOXA9* [178] and methylation of *HOXA11* was independently associated with poor outcome of patients [179]. Moreover, 60% of acute myeloid leukaemia exhibited very high methylation levels of *HOXA5*, far greater than that seen in normal haematopoietic cells [180]. For CCA, a study has demonstrated the methylation of *HOXA1* in 95.2% of extrahepatic type [175]. In addition, the oncogenic potential of *homeobox* genes has been widely described in leukemia as well as in other neoplasias [197-199]. Although several reports have evaluated aberrant expression of *homeobox* genes between normal and tumor tissues, but their functional relationship with the malignant transformation has remained elusive [209]. Abate-Sen [200] has defined that *homeobox* genes can be down-regulated in cancer cells derived from tissues in which a particular gene is normally expressed in adult cells. In the case of CCA, hypermethylation could be a mechanism of down regulation of *homeobox* genes.

Several studies have reported the large scale of DNA methylation profiles in other cancers using high-throughput technologies by Illumina Goldengate Methylation Array or Infinium® HumanMethylation27 BeadChip e.g. ovarian

epithelial carcinoma [181], lymphoma [182], renal cell carcinoma [183] and lung adenocarcinoma and cancer cell lines (prostate, lung, breast and colon cancer) [176].

In conclusion, DNA methyltion signatures with hypermethylation and hypomethylation of multiple CpG sites have been found being differentially changed as a common molecular mechanism in CCA. Our finding provides a useful data resource of epigenetic signature which could serve as the methylation biomarkers of this fatal cancer.