Published in final edited form as:

Cancer Prev Res (Phila). 2013 December ; 6(12): . doi:10.1158/1940-6207.CAPR-13-0104.

Aberrant DNA methylation at genes associated with a stem cell-like phenotype in cholangiocarcinoma tumours

Ruethairat Sriraksa^{#1,2}, Constanze Zeller^{#3}, Wei Dai³, Afshan Siddiq⁴, Andrew J Walley⁴, Temduang Limpaboon^{1,2}, and Robert Brown³

¹Centre for Research and Development of Medical Diagnostic Laboratories, Faculty of Associated Medical Sciences, Khon Kaen University, Khon Kaen, 40002, Thailand ²Liver Fluke and Cholangiocarcinoma Research Center, Faculty of Medicine, Khon Kaen University, Khon Kaen, 40002, Thailand ³Epigenetics Unit, Department of Surgery and Cancer, Imperial College London, IRDB Building, Hammersmith Hospital Campus, London, W12 0NN, UK ⁴Department of Genomics of Common Disease, School of Public Health, Burlington-Danes Building, Hammersmith Hospital Campus, London W12 0NN, UK

These authors contributed equally to this work.

Abstract

Genetic abnormalities of cholangiocarcinoma have been widely studied; however, epigenomic changes related to cholangiocarcinogenesis have been less well characterised. We have profiled the DNA methylomes of 28 primary cholangiocarcinoma and six matched adjacent normal tissues using Infinium's HumanMethylation27 BeadChips with the aim of identifying gene sets aberrantly epigenetically regulated in this tumour type.

Using a linear model for microarray data we identified 1610 differentially methylated autosomal CpG sites with 809 CpG sites (representing 603 genes) being hypermethylated and 801 CpG sites (representing 712 genes) being hypomethylated in cholangiocarcinoma versus adjacent normal tissues (false discovery rate = 0.05). Gene ontology and gene set enrichment analyses identified gene sets significantly associated with hypermethylation at linked CpG sites in cholangiocarcinoma including *homeobox* genes and target genes of PRC2, EED, SUZ12 and histone H3 trimethylation at lysine 27. We confirmed frequent hypermethylation at the *homeobox* genes *HoxA9* and *HoxD9* by bisulfite pyrosequencing in a larger cohort of cholangiocarcinoma (n = 102).

Our findings indicate a key role for hypermethylation of multiple CpG sites at genes associated with a stem cell-like phenotype as a common molecular aberration in cholangiocarcinoma. These data have implications for cholangiocarcinogenesis, as well as possible novel treatment options using histone methyltransferase inhibitors.

Corresponding author: Prof Robert Brown, Epigenetics Unit, Department of Surgery and Cancer, Imperial College London, IRDB Building, Hammersmith Hospital Campus, London, W12 0NN, United Kingdom. Phone: +44 (0) 2075 941804; Fax: +44 (0) 2075 942129; b.brown@imperial.ac.uk; Correspondence may also be addressed to Temduang Limpaboon, Centre for Research and Development of Medical Diagnostic Laboratories, Faculty of Associated Medical Sciences, Khon Kaen University, Khon Kaen, 40002, Thailand. Phone: +66 (0) 4336 2028; Fax: +66 (0) 4320 2088; temduang@kku.ac.th.

Disclosure of potential conflicts of interest: There are no relationships that the authors believe could be construed as resulting in an actual, potential, or perceived conflict of interest with regards to this manuscript.

Keywords

DNA methylation; cholangiocarcinoma; cancer; stem cells; HOX

Introduction

Cholangiocarcinoma (CCA) is a hepatic malignancy originating from the biliary tract which accounts for 10%-20% of primary liver cancers (1, 2). Depending on the anatomical site of the tumour, CCA are classified into being either of extrahepatic or intrahepatic origin (3). The global incidence rates of intrahepatic CCA, in particular, have markedly increased worldwide (3, 4). The geographic distribution of risk factors account for geographic differences in incidence rates with the highest incidence reported in the endemic area of liver fluke infection in northeastern Thailand. Here, the age-standardized incidence rate (ASIR) is up to 317.6 per 100,000 people which exceeds the ASIR of less than 2 per 100,000 people in Western countries by a hundred-fold (5-9). However, CCA incidences have also been steadily rising in Western countries such as the UK or US over the last 40 years (3). CCA are particularly difficult to diagnose and often present at a late clinical stage which drastically reduces the chances for successful treatment. Surgical resection has been reported to be curative, however, only very few patients are eligible for this treatment option (10). The vast majority of patients will receive non-surgical therapeutic regimes such as chemotherapy with low responsive rates observed overall.

Epigenetic changes in tumours are associated with cancer development and progression (11). Aberrant DNA methylation occurs very early during carcinogenesis before the onset of malignancy and is thought to interact with genetic aberrations in driving the tumourigenic phenotype of cancer cells (12). Epigenetic silencing at CpG islands, in particular, is observed in many solid tumour types. Aberrant methylation of specific candidate genes in CCA has been described by several studies (13-16), however there have been no studies looking at methylation profiles throughout the genome. We have quantified DNA methylation levels at a genome-wide scale using Illumina's Infinium® HumanMethylation27 BeadChip. The Methylation BeadChip covers 27,578 CpG sites across 14,485 consensus coding sequences (CCDS) and well-known cancer genes in the human genome (17).

It has been proposed that there are subpopulations of cells present in tumours that are responsible for sustaining their growth and are phenotypically distinct from the majority of tumour cells (18). These populations of tumour sustaining cells have characteristics of stem cells, which include prolonged proliferative capacity *in vitro* and *in vivo*, expression of specific subsets of genes and resistance to DNA damage (19, 20). Tumours 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 (21-23). Furthermore, it has been suggested that genes involved in regulation of a stem cell state may be more vulnerable to aberrant DNA methylation (24). We have therefore addressed whether there is enrichment for aberrant DNA methylation of genes associated with a stem cell-like state in CCA.

Materials and Methods

Tumour samples and DNA extraction

Twenty-eight primary intrahepatic CCA and six matched adjacent normal samples were supplied by the Liver Fluke and Cholangiocarcinoma Research Center, Khon Kaen University (Khon Kaen, Thailand). Ethical approval was obtained for all samples (Ethics

Committee of Khon Kaen University - HE510651). Histopathological features of the tissues are summarised in Supplementary Table S1. Genomic DNA was extracted using the DNeasy Tissue Kit (Qiagen, Hercules, CA) following the manufacturer's recommendations.

DNA methylation profiling

Array-based DNA methylation profiling was performed using Infinium HumanMethylation27 BeadChips (Illumina, San Diego, CA, USA) which simultaneously interrogate 27,578 CpG sites within the promoter regions of more than 14,000 genes (15,369 CpG sites located within CpG islands, 12,209 CpG sites located outside of CpG islands as defined by Gardiner-Garden and Frommer (25)). 1 µg genomic DNA was bisulfite modified using the EZ DNA Methylation Kit (Zymo Research, Cambridge, UK). 200 ng of converted DNA were further processed to run BeadArrays according to the manufacturer's instructions. The arrays were imaged using a BeadArray™ Reader. Each locus is represented by fluorescent signals from two bead types corresponding to the methylated (M) and unmethylated (U) alleles, respectively. The raw signals of unmethylated and methylated bead types (average 15-30 beads of each type per locus) were background corrected and computed into a beta value (β) using the BeadStudio software Methylation Module (version 1.0.5) (Illumina, San Diego, CA, USA). The beta value represents the ratio of the intensity of the methylated bead type to the combined locus intensity and reflects the methylation status of a specific CpG site. Beta values are continuous variables ranging from 0 (unmethylated) to 1 (fully methylated). Beta values were quartile normalized and subsequent analysis was carried out in R (version 2.10.1).

The robustness and reproducibility of the Infinium HumanMethylation27 BeadChips were verified by using five technical replicates of the ovarian cancer cell line PEO1 for correlation experiments. Beta values of these replicates were highly correlated with an average $R^2 = 0.989$ (range: 0.988 - 0.990, $P < 0.00001$).

Bisulfite modification and pyrosequencing analysis

Validation of candidate loci at the two *HOX* genes, *HOXA9* and *HOXD9*, was performed in a bigger panel of Thai CCA (n = 102) and adjacent normal samples (n = 24) using bisulfite pyrosequencing. 1 µg of genomic DNA was bisulfite modified using the EpiTect Bisulfite Kit (Qiagen, West Sussex, UK) according to the manufacturer's instructions. Up to 100 ng of modified DNA were used as a template in subsequent pyrosequencing reactions. Pyrosequencing primer sets covering differentially methylated CpG sites in the *HOXA9* and *HOXD9* gene promoters were HOXA9_PYRO_F: 5'-Biotin-GGTGATGGTTATTATTGGGGTTT-3', HOXA9_PYRO_R: 5'-AAAACTAACCCAAAATCCC-3', HOXA9_PYRO_S: 5'-CTAACCCAAAATCCCC-3', HOXD9_PYRO_F: 5'-GGGTATATGGATTGTTGTTG-3', HOXD9_PYRO_R: 5'-Biotin-TCCACCCAACATTACACTATC-3', HOXD9_PYRO_S: 5'-TAATTATAGTAGTTTTAGTTTA-3'. Pyrosequencing-PCRs were performed in duplicate for each sample in a 25 µl volume containing 0.2 µl FastStart Taq DNA polymerase (Roche Applied Science, Mannheim, Germany), 2.5 µl FastStart Buffer (Roche Applied Science, Mannheim, Germany), 0.2 mM dNTPs (Applied Biosystems, Foster City, CA), 0.2 µM primers (each) and 2 µl of modified DNA template using the following conditions: 95°C for 6 min, 35 cycles of 95°C for 30 sec, 63°C (*HOXA9*) or 60°C (*HOXD9*) for 30 sec and 72°C for 30 sec followed by final extension at 72°C for 5 min. Methylation frequencies were determined by pyrosequencing of PCR products using the PSQ HS96 Pyrosequencing System (Biotage, Uppsala, Sweden) and converting the resulting pyrograms into numerical values for peak heights. The methylation status of the two genes in each sample was calculated by using the average percentages of methylation across all targeted CpG sites, respectively. For methylation analysis comparing normal adjacent and CCA, a

cut-off was calculated from tested adjacent normal samples which was based on the mean methylation value across samples + $2.069 \times$ standard deviation (cut-off percentage of methylation in *HOXA9* = 21.4% and *HOXD9* = 18.0%). Increased methylation was defined as being above this cut-off.

Bioinformatic and statistical analysis

Unsupervised clustering—Unsupervised hierarchical clustering was performed using DNA methylation profiles from primary CCA and adjacent normal samples in TIGR MultiExperiment Viewer (version 4.5, <http://mev.tm4.org>) using Euclidean distance as the distance metric and average linkage clustering to merge clusters. CpG sites on the X and Y chromosomes were excluded from the analysis to avoid gender-related effects leaving 26,486 autosomal CpG sites (14,862 CpG sites within CpG islands, 11,624 CpG sites outside of CpG islands) for hierarchical clustering.

Gene annotation—Gene annotation was carried out using the Mar. 2006 (NCBI/36/hg18) assembly at the UCSC database (genome.ucsc.edu/). Individual CpG sites located within a CpG island as defined by Gardiner-Garden and Frommer (25) are linked to an associated gene if the CpG island was within 2 kb distance from the respective transcription start site (26). Non-CpG island-related CpG sites are linked to a gene if they were located within 2 kb from the transcription start site.

Differential methylation analysis—Differential methylation at CpG sites was analyzed by a linear model for microarray data (LIMMA) comparing between the group of CCA and the group of adjacent normal. A false discovery rate (FDR) = 0.05 was used as a cut-off for differential methylation analysis. The difference of five technical replicates of the ovarian cancer cell line PEO1 was calculated as the expected difference (null distribution). The observed difference was calculated from the grouped samples. Using bootstrap resampling method, we extracted the same number of CpG sites ($n = 26,486$) from the null distribution 200 times. Each time, a FDR was calculated using the formula below. The median of FDRs was used as an estimation of the probability that CpG sites with differential methylation above the selected cut-off were identified by chance. This median FDR was set at 0.05 corresponding to $|\Delta\beta| = 0.15$.

$$FDR = \frac{\# \text{CpG_sites_over_cutoff_in_null_data}}{\# \text{CpG_sites_over_cutoff_in_real_data}}$$

Gene ontology analysis—A list of genes associated with hypermethylated or hypomethylated CpG sites in CCA compared to adjacent normal samples determined by LIMMA analysis FDR = 0.05 and $|\Delta\beta| = 0.15$ was used for Interpro and gene ontology (GO) analyses using the functional annotation tool in the database DAVID (<http://david.abcc.ncifcrf.gov/summary.jsp>). The functional terms enriched in this set of genes were examined by a modified Fisher Exact test corrected by FDR compared to the functional terms relevant to a full list of genes presented on HumanMethylation27 BeadChip. The significant level was determined at FDR = 0.05.

Gene set enrichment analysis (GSEA)—GSEA (27) was performed to determine whether 13 predefined sets of genes associated with embryonic stem cells (Supplementary Table S2) showed statistically significant, concordant methylation differences between 28 CCA as compared to six adjacent normal samples. For this, beta values of multiple CpG sites associated with the same gene were averaged. A full set of averaged methylation data was used instead of a subset of data determined by a cut-off in Gene ontology analysis. The

significant level was set at $FDR = 0.05$ which was estimated from a 1000 times permutation test. GSA package was used in this analysis.

Correlation between methods—Correlation of methylation levels of *HOXA9* and *HOXD9* as detected by HumanMethylation27 BeadChips and pyrosequencing were performed using Spearman correlation.

Data deposition—Methylation profiling data are available in NCBI's Gene Expression Omnibus (GEO accession number: GSE38860) and can be accessed under: <http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=zzqplqssqyeoizs&acc=GSE38860>.

Results

DNA methylation profiles of primary CCA and adjacent normal tissues

We analyzed the DNA methylation status of 27,578 CpG sites in 28 primary intrahepatic CCA and six matched adjacent normal samples using Infinium HumanMethylation27 BeadChips. Unsupervised hierarchical clustering was performed to determine the similarity of DNA methylation profiles across primary tumours and adjacent normal tissues using CpG methylation levels (β values) of 26,486 CpG dinucleotides obtained from each sample, excluding CpG sites on the X and Y chromosomes to avoid gender-related DNA methylation effects. Hierarchical clustering analysis returned two major clusters broadly separating between primary CCA tissues and adjacent normal tissues (Figure 1 and Supplementary Figure S1). CCA formed several subclusters with one adjacent normal tissue (W022N) being embedded in the cluster of CCA. Adjacent normal tissues, on the other hand, clustered tightly together with only two tumours (P055T and U030T) co-clustering with this subgroup of tissues. Re-evaluation of histopathology of an adjacent biopsy to the one used to extract DNA of these exceptions confirmed normal pathology and tumour pathology, respectively. Re-evaluation of tumours showed >80% tumour by histological cellularity.

Distinctive hyper- and hypomethylation patterns in CCA as compared to adjacent normal tissues

We further examined differences in methylation ($|\Delta\beta| \geq 15\%$) between the 28 CCA and six adjacent normal tissues at all 26,486 autosomal CpG sites using a linear model for microarray data (LIMMA) (28). We identified 1610 methylation changes with 809 CpG sites (associated with 603 genes) being hypermethylated and 801 CpG sites (associated with 712 genes) being hypomethylated in 28 CCA compared to six adjacent normal samples at a false discovery rate (FDR) of less than 5% (Figure 2, Supplementary Tables S3 and S4). Within the set of 28 CCA, six tumours were matched to the adjacent normal samples used in this study. Among the methylation changes identified in the group comparison, 781/809 hypermethylated CpG sites (associated with 585 genes) and 748/801 hypomethylated CpG sites (associated with 665 genes) remained significant ($FDR = 0.05$ and $|\Delta\beta| \geq 15\%$) when comparing the six pairs of matched tumours and adjacent normal samples (Supplementary Tables S5 and S6).

In order to explore any patterns within the set of genes showing differential methylation at linked CpG sites in CCA versus adjacent normal, we interrogated gene ontology (GO) (29) and InterPro categories (30) to identify enrichment of gene functions. Table 1 shows the top 20 highly significant results for genes being hypermethylated in CCA compared to adjacent normal with $P < 3.62 \times 10^{-5}$ after controlling for a $FDR < 0.05$. The *homeobox* class was the most significantly enriched category within the set of genes associated with hypermethylated CpG sites in CCA ($FDR = 2.18 \times 10^{-20}$). In addition, molecular functions and biological

processes related to *homeobox* such as transcription factor activity ($\text{FDR} = 1.89 \times 10^{-11}$) or embryonic morphogenesis ($\text{FDR} = 3.66 \times 10^{-9}$) were found among the top ten of terms (Table 1). Among the set of genes which were hypomethylated in CCA versus adjacent normal tissue we observed significant enrichment of gene groups associated with the plasma membrane ($\text{FDR} = 3 \times 10^{-4}$), immune response ($\text{FDR} = 0.005212$), and the regulation and induction of apoptosis ($\text{FDR} = 0.016337$ and $\text{FDR} = 0.032143$) (Table 2).

We validated our array-based findings in a bigger panel of CCA ($n = 102$) and adjacent normal samples ($n = 24$) using bisulfite pyrosequencing to analyze differentially methylated loci at the two *HOX* genes, *HOXA9* and *HOXD9* identified as differentially methylated. Hypermethylation of *HOXA9* and *HOXD9* was found in 86.3% (88/102) and 89.2% (91/102) of primary CCA, respectively. Primary CCA showed significantly increased methylation ($P < 0.001$, unpaired t-test) as compared to adjacent normal samples at both loci (Figure 3A and B). We found a strong correlation between pyrosequencing-based detection of CpG methylation and beta values obtained from HumanMethylation27 BeadChips (Spearman correlation coefficient > 0.9 , $P < 0.001$) for both genes supporting the observed hypermethylation of linked CpG sites in CCA.

Hypermethylation of PRC2 targets is enriched in CCA compared to adjacent normal tissue

Homeobox genes have a vital role in embryonic development and differentiation. Regulation of *HOX* involves the epigenetic key mark trimethylation at histone H3 lysine 27 (H3K27me3) which is established and maintained by polycomb group proteins (PcG), in particular the polycomb repressive complex 2 (PRC2), which maintains early patterns of *HOX* gene repression (31). It has been previously suggested that PRC2 target genes in embryonic stem cells may be more vulnerable to epigenetic silencing by DNA methylation. PcG protein repression targets many more genes besides the *homeobox* genes (32, 33) in undifferentiated embryonic cells, so we addressed whether additional targets of PcG proteins would be enriched within the set of aberrantly methylated genes in CCA. To this end, we performed gene set enrichment analysis (GSEA) interrogating 13 published embryonic gene signature sets associated with human embryonic stem cells (hES) for their methylation pattern at loci associated with CpG islands in CCA and adjacent normal tissues (Supplementary Table S2). Five signatures were significantly hypermethylated in CCA as compared to adjacent normal at $P < 0.05$ and a $\text{FDR} < 0.05$ encompassing target genes of the PRC2 components EED ($P = 0.009$ and $\text{FDR} = 0.04$) and SUZ12 ($P = 0.017$ and $\text{FDR} = 0.04$), targets of H3K27me3 ($P = 0.017$ and $\text{FDR} = 0.04$), PRC2 target genes (shared targets of EED, SUZ12 and H3K27me3; $P = 0.011$ and $\text{FDR} = 0.04$) and *homeobox* genes ($P = 0.014$ and $\text{FDR} = 0.04$) (Table 3). Overall, the observed enrichment of differential methylation in CCA at embryonic gene sets, particularly those associated with the PRC2 and its components indicates aberrant DNA methylation of genes associated with a more stem cell-like phenotype in CCA tumours as compared to adjacent normal tissue.

Discussion

Aberrant DNA methylomes are one of the key features of cancer (11). Recent advances in DNA methylation technology have allowed for genome-wide profiling of cancer tissues providing a vast amount of previously unknown methylation changes which further our understanding of the underlying pathways associated with tumorigenesis. To our knowledge, this is the first report on a genome-wide DNA methylome analysis in CCA, a cancer associated with liver-fluke infection in endemic countries. We show that DNA methylation is a common molecular aberration in CCA with DNA methylation profiles of primary CCA being different from those of adjacent normal tissues. Differential methylation between CCA and adjacent normal tissues comprised hypermethylation and

hypomethylation events which were equally present at CpG sites within and outside of CpG islands. As previously discussed, our findings of distinct methylation changes in CCA as compared to normal tissues relate to liver-fluke associated CCA whereas the sporadic type found in Western countries has not been included in this study. While CCA from different geographical sources may share common biological mechanisms occurring during tumourigenesis additional studies are required to analyse differential methylation in Western CCA versus normal and to filter out differences between Western and Thai CCA. To demonstrate the utility in relevant biological processes of differential methylation between CCA and adjacent normal samples, we further showed that hypermethylation in CCA is significantly over-represented at genes involved in developmental processes, most importantly *homeobox* and genes involved in transcription factor activity and embryonic morphogenesis. More specifically, we examined methylation states of gene sets associated with a stem-cell like phenotype and identified hypermethylation at specific target genes of the PcG complexes including PRC2 targets, *homeobox* genes, EED targets, SUZ12 targets and H3K27me3 target genes. More than 55% of *homeobox* genes are also the target genes of PRC2, EED, SUZ12 and H3K27me3 indicating that the overlap of these gene sets might be the main target of hypermethylation in CCA. PcG complexes are involved in transcriptional repression of genes. They play a role to maintain heritable transcription patterns of the *homeobox* genes during development and differentiation, in conjunction with their positively acting counterparts known as Trithorax group (TrxG) proteins (34, 35). A 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 (36). Several reports have described that tumours 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 (21-23). Taken together, our findings support the theory of a stem cell origin for CCA.

Consistent with the observation of a stem-cell associated phenotype and in support of our observation of hypermethylation of PRC2 target genes in CCA, a recent study has shown overexpression of EZH2, the catalytic subunit of PRC2, in CCA (37). Aberrant levels of EZH2 have previously been linked to tumour development and have been observed in a variety of tumours including prostate, breast and melanoma (38-40). Moreover, high EZH2 levels have been associated with poor prognosis in these tumours types. In CCA, EZH2 expression levels were significantly and positively associated with step-wise progression of CCA from low-grade dysplasia, high-grade dysplasia, carcinoma *in situ* to invasive carcinoma suggesting that chromatin remodelling complexes are involved in development and progression of this cancer type (37). High EZH2 was linked to increased methylation of the *p16^{Ink4a}* promoter (37) supporting the notion that alternative complementary silencing mechanisms such as DNA methylation may induce or permanently replace initial silencing of PRC2 target genes. In our study, low methylation levels of the *homeobox* genes *HOXA9* and *HOXD9*, both PRC2 target genes, were verified in adjacent normal tissues. Although the number of adjacent normal tissues initially used for DNA methylome analysis was comparatively small, the inclusion of more normal adjacent samples in our pyrosequencing analyses indicates that low methylation levels can be observed across additional normal samples at *homeobox* genes. Methylation of *homeobox* in normal tissue could reflect a mark of rare cancer stem cells which then spread in tumours. In this respect, our observation of DNA hypermethylation at PRC2 target genes including *homeobox* in CCA not only shows potential to serve as a biomarker for early detection of the disease. PRC2 and EZH2, in particular, might further be attractive therapeutic and, even more important, chemopreventative targets in CCA. In glioblastoma multiform, an aggressive type of brain tumour, pharmacologic EZH2 inhibitors such as DZNep have been shown to strongly impair

cancer stem cell self-renewal *in vitro* and tumour-initiating capacity *in vivo* (41). Therefore, DZNep as well as novel, more specific pharmacologic agents (42) targeting histone lysine methyltransferases (HKMT) may have potential to serve as a novel approach for early intervention of the often terminal CCA.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We would like to greatly thank the Liver Fluke and Cholangiocarcinoma Research Center, Khon Kaen University for samples and clinical data.

Grant support: This work was supported by the Higher Education Research Promotion and National Research University Project of Thailand, the Office of the Higher Education Commission, through the Center of Excellence in Specific Health Problems in Greater Mekong Sub-region cluster (SHeP-GMS), Khon Kaen University, Thailand (PD54201 to R. Sriraksa and NRU542015 to T. Limpaboon); and the Imperial Experimental Cancer Medicine Centre and Cancer Research UK programme grant (C536/A13086 to R. Brown).

References

1. Blechacz B, Gores GJ. Cholangiocarcinoma: advances in pathogenesis, diagnosis, and treatment. *Hepatology*. 2008; 48:308–21. [PubMed: 18536057]
2. Huang L, Frampton G, Liang LJ, Demorrow S. Aberrant DNA methylation profile in cholangiocarcinoma. *World J Gastrointest Pathophysiol*. 2010; 1:23–9. [PubMed: 21607139]
3. Patel T. Cholangiocarcinoma. *Nat Clin Pract Gastroenterol Hepatol*. 2006; 3:33–42. [PubMed: 16397610]
4. Khan SA, Miras A, Pelling M, Taylor-Robinson SD. Cholangiocarcinoma and its management. *Gut*. 2007; 56:1755–6. [PubMed: 17998326]
5. Jepsen P, Vilstrup H, Tarone RE, Friis S, Sorensen HT. Incidence rates of intra- and extrahepatic cholangiocarcinomas in Denmark from 1978 through 2002. *J Natl Cancer Inst*. 2007; 99:895–7. [PubMed: 17551150]
6. Landis SH, Murray T, Bolden S, Wingo PA. Cancer statistics, 1998. *CA Cancer J Clin*. 1998; 48:6–29. [PubMed: 9449931]
7. Shaib Y, El-Serag HB. The epidemiology of cholangiocarcinoma. *Semin Liver Dis*. 2004; 24:115–25. [PubMed: 15192785]
8. Sriamporn S, Pisani P, Pipitgool V, Suwanrungruang K, Kamsa-ard S, Parkin DM. Prevalence of *Opisthorchis viverrini* infection and incidence of cholangiocarcinoma in Khon Kaen, Northeast Thailand. *Trop Med Int Health*. 2004; 9:588–94. [PubMed: 15117303]
9. Taylor-Robinson SD, Toledano MB, Arora S, Keegan TJ, Hargreaves S, Beck A, et al. Increase in mortality rates from intrahepatic cholangiocarcinoma in England and Wales 1968–1998. *Gut*. 2001; 48:816–20. [PubMed: 11358902]
10. Lazaridis KN, Gores GJ. Cholangiocarcinoma. *Gastroenterology*. 2005; 128:1655–67. [PubMed: 15887157]
11. Jones PA, Baylin SB. The epigenomics of cancer. *Cell*. 2007; 128:683–92. [PubMed: 17320506]
12. Wolff EM, Chihara Y, Pan F, Weisenberger DJ, Siegmund KD, Sugano K, et al. Unique DNA methylation patterns distinguish noninvasive and invasive urothelial cancers and establish an epigenetic field defect in premalignant tissue. *Cancer Res*. 2010; 70:8169–78. [PubMed: 20841482]
13. Kim BH, Cho NY, Choi M, Lee S, Jang JJ, Kang GH. Methylation profiles of multiple CpG island loci in extrahepatic cholangiocarcinoma versus those of intrahepatic cholangiocarcinomas. *Arch Pathol Lab Med*. 2007; 131:923–30. [PubMed: 17550320]
14. Lee S, Kim WH, Jung HY, Yang MH, Kang GH. Aberrant CpG island methylation of multiple genes in intrahepatic cholangiocarcinoma. *Am J Pathol*. 2002; 161:1015–22. [PubMed: 12213730]

15. Yang B, House MG, Guo M, Herman JG, Clark DP. Promoter methylation profiles of tumor suppressor genes in intrahepatic and extrahepatic cholangiocarcinoma. *Mod Pathol*. 2005; 18:412–20. [PubMed: 15467712]
16. Sriraksa R, Zeller C, El-Bahrawy MA, Dai W, Daduang J, Jearanaikoon P, et al. CpG-island methylation study of liver fluke-related cholangiocarcinoma. *Br J Cancer*. 2011; 104:1313–8. [PubMed: 21448164]
17. Bibikova M, Le J, Barnes B, Saedinia-Melnyk S, Zhou L, Shen R, et al. Genome-wide DNA methylation profiling using Infinium(R) assay. *Epigenomics*. 2009; 1:177–200. [PubMed: 22122642]
18. Mackillop WJ, Ciampi A, Till JE, Buick RN. A stem cell model of human tumor growth: implications for tumor cell clonogenic assays. *J Natl Cancer Inst*. 1983; 70:9–16. [PubMed: 6571928]
19. Dick JE. Stem cell concepts renew cancer research. *Blood*. 2008; 112:4793–807. [PubMed: 19064739]
20. Wicha MS, Liu S, Dontu G. Cancer stem cells: an old idea--a paradigm shift. *Cancer Res*. 2006; 66:1883–90. discussion 95–6. [PubMed: 16488983]
21. Beachy PA, Karhadkar SS, Berman DM. Tissue repair and stem cell renewal in carcinogenesis. *Nature*. 2004; 432:324–31. [PubMed: 15549094]
22. Reya T, Morrison SJ, Clarke MF, Weissman IL. Stem cells, cancer, and cancer stem cells. *Nature*. 2001; 414:105–11. [PubMed: 11689955]
23. Stingl J, Caldas C. Molecular heterogeneity of breast carcinomas and the cancer stem cell hypothesis. *Nat Rev Cancer*. 2007; 7:791–9. [PubMed: 17851544]
24. Ohm JE, Baylin SB. Stem cell chromatin patterns: an instructive mechanism for DNA hypermethylation? *Cell Cycle*. 2007; 6:1040–3. [PubMed: 17457052]
25. Gardiner-Garden M, Frommer M. CpG islands in vertebrate genomes. *J Mol Biol*. 1987; 196:261–82. [PubMed: 3656447]
26. Saxonov S, Berg P, Brutlag DL. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc Natl Acad Sci U S A*. 2006; 103:1412–7. [PubMed: 16432200]
27. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005; 102:15545–50. [PubMed: 16199517]
28. Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol*. 2004; 3 Article3.
29. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*. 2000; 25:25–9. [PubMed: 10802651]
30. Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, et al. InterPro: the integrative protein signature database. *Nucleic Acids Res*. 2009; 37:D211–5. [PubMed: 18940856]
31. Margueron R, Reinberg D. The Polycomb complex PRC2 and its mark in life. *Nature*. 2011; 469:343–9. [PubMed: 21248841]
32. Boyer LA, Lee TI, Cole MF, Johnstone SE, Levine SS, Zucker JP, et al. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell*. 2005; 122:947–56. [PubMed: 16153702]
33. Lee TI, Jenner RG, Boyer LA, Guenther MG, Levine SS, Kumar RM, et al. Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell*. 2006; 125:301–13. [PubMed: 16630818]
34. Jacobs JJ, van Lohuizen M. Cellular memory of transcriptional states by Polycomb-group proteins. *Semin Cell Dev Biol*. 1999; 10:227–35. [PubMed: 10441076]
35. Orlando V. Polycomb, epigenomes, and control of cell identity. *Cell*. 2003; 112:599–606. [PubMed: 12628181]
36. Widschwendter M, Fiegl H, Egle D, Mueller-Holzner E, Spizzo G, Marth C, et al. Epigenetic stem cell signature in cancer. *Nat Genet*. 2007; 39:157–8. [PubMed: 17200673]

37. Sasaki M, Yamaguchi J, Itatsu K, Ikeda H, Nakanuma Y. Over-expression of polycomb group protein EZH2 relates to decreased expression of p16 INK4a in cholangiocarcinogenesis in hepatolithiasis. *J Pathol.* 2008; 215:175–83. [PubMed: 18393368]
38. Bachmann IM, Halvorsen OJ, Collett K, Stefansson IM, Straume O, Haukaas SA, et al. EZH2 expression is associated with high proliferation rate and aggressive tumor subgroups in cutaneous melanoma and cancers of the endometrium, prostate, and breast. *J Clin Oncol.* 2006; 24:268–73. [PubMed: 16330673]
39. Varambally S, Dhanasekaran SM, Zhou M, Barrette TR, Kumar-Sinha C, Sanda MG, et al. The polycomb group protein EZH2 is involved in progression of prostate cancer. *Nature.* 2002; 419:624–9. [PubMed: 12374981]
40. Yu J, Rhodes DR, Tomlins SA, Cao X, Chen G, Mehra R, et al. A polycomb repression signature in metastatic prostate cancer predicts cancer outcome. *Cancer Res.* 2007; 67:10657–63. [PubMed: 18006806]
41. Suva ML, Riggi N, Janiszewska M, Radovanovic I, Provero P, Stehle JC, et al. EZH2 is essential for glioblastoma cancer stem cell maintenance. *Cancer Res.* 2009; 69:9211–8. [PubMed: 19934320]
42. Cherblanc F, Chapman-Rothe N, Brown R, Fuchter MJ. Current limitations and future opportunities for epigenetic therapies. *Future medicinal chemistry.* 2012; 4:425–46. [PubMed: 22416773]

Figure 1. Unsupervised hierarchical clustering of DNA methylation profiles of primary CCA and matched adjacent normal tissues.

Clustering analysis of all 26,486 autosomal CpG sites returned two major clusters separating between primary CCA tissues and adjacent normal tissues with three exceptions: adjacent normal sample W022N which clustered with CCA and primary CCA P055T and U030T which clustered with normal.

Figure 2. Differential methylation in tumours (n = 28) as compared to adjacent normal samples (n = 6).

The difference of methylation (X axis) versus $-\log_{10}(\text{FDR})$ (Y axis) obtained from LIMMA analysis is shown. The broken red line shows the significance cut-off corresponding to a $\text{FDR} < 0.05$ and difference $|\Delta\beta| \geq 0.15$ used in this study. Red spots indicate hypermethylated CpG sites, blue spots show hypomethylated CpG sites. The difference cut-off was estimated from a null distribution of differential methylation between five technical replicates on Infinium's HumanMethylation27 beadchips.

Figure 3. Differential methylation at loci at *HOXA9* (A) and *HOXD9* (B) in primary CCA compared to adjacent normal samples as detected by bisulfite pyrosequencing. Scatter dot plots represent percentage of methylation and black horizontal lines are average methylation levels across adjacent normal and primary CCA samples, respectively. Significant differences of methylation levels are observed between primary CCA, compared to adjacent normal samples ($P < 0.001$, unpaired t-test, GraphPad Prism software).

Table 1
Significantly enriched top 20 functional terms in genes associated with hypermethylated CpG sites in CCA

Category	Term	n ^a of genes	%	P value ^b	FDR
INTERPRO	IPR001356:Homeobox	46	8.57	1.18E-20	2.18E-20
INTERPRO	IPR017970:Homeobox, conserved site	46	8.57	1.52E-20	2.80E-20
INTERPRO	IPR012287:Homeodomain-related	46	8.57	2.16E-19	3.99E-19
GOTERM_MF	GO:0003700~transcription factor activity	88	16.39	8.33E-12	1.89E-11
GOTERM_CC	GO:0005576~extracellular region	126	23.46	3.07E-11	1.36E-10
GOTERM_BP	GO:0030182~neuron differentiation	54	10.06	4.77E-10	3.22E-10
GOTERM_MF	GO:0043565~sequence-specific DNA binding	63	11.73	2.20E-10	5.00E-10
GOTERM_BP	GO:0048598~embryonic morphogenesis	43	8.01	5.42E-09	3.66E-09
GOTERM_BP	GO:0003002~regionalization	34	6.33	1.25E-08	8.45E-09
GOTERM_BP	GO:0009952~anterior/posterior pattern formation	28	5.21	3.42E-08	2.31E-08
GOTERM_BP	GO:0007389~pattern specification process	38	7.08	9.84E-08	6.64E-08
INTERPRO	IPR001827:Homeobox protein, antenpedia type, conserved site	13	2.42	3.27E-08	6.03E-08
GOTERM_BPa	GO:0001501~skeletal system development	41	7.64	5.28E-07	3.56E-07
GOTERM_BP	GO:0048562~embryonic organ morphogenesis	25	4.66	8.86E-07	5.98E-07
GOTERM_BP	GO:0048568~embryonic organ development	28	5.21	1.81E-06	1.22E-06
GOTERM_BP	GO:0048705~skeletal system morphogenesis	23	4.28	3.86E-06	2.61E-06
GOTERM_BP	GO:0007610~behavior	48	8.94	1.67E-05	1.13E-05
GOTERM_BP	GO:0048704~embryonic skeletal system morphogenesis	16	2.98	2.45E-05	1.66E-05
GOTERM_BP	GO:0050877~neurological system process	68	12.66	3.34E-05	2.26E-05
GOTERM_BP	GO:0006355~regulation of transcription, DNA-dependent	105	19.55	3.62E-05	2.45E-05

^a number,

^b Bonferroni corrected

Table 2
Significantly enriched functional terms in genes associated with hypomethylated CpG sites in CCA

Category	Term	n ^a of genes	%	P value ^b	FDR
GOTERM_CC	GO:0044459~plasma membrane part	123	20.16	8.27E-05	3.00E-04
GOTERM_BP	GO:0006955~immune response	49	8.03	0.007458	0.005212
GOTERM_BP	GO:0042981~regulation of apoptosis	53	8.69	0.023192	0.016337
GOTERM_CC	GO:0005887~integral to plasma membrane	72	11.80	0.004246	0.01543
GOTERM_BP	GO:0043067~regulation of programmed cell death	53	8.69	0.030145	0.02131
GOTERM_BP	GO:0010941~regulation of cell death	53	8.69	0.033297	0.023575
GOTERM_BP	GO:0006917~induction of apoptosis	28	4.59	0.045123	0.032143
GOTERM_BP	GO:0012502~induction of programmed cell death	28	4.59	0.047725	0.034042
GOTERM_CC	GO:0031226~intrinsic to plasma membrane	72	11.80	0.009136	0.033281

^a number,

^b Bonferroni corrected

Table 3
Gene Set Enrichment Analysis between CCA (n=28) and adjacent normal samples (n=6)

Methylation status	Gene set ^a	GSA Score	P value	FDR
Hypermethylation	EED targets	0.2396	0.009	0.04 [*]
	PRC2 targets (1)	0.3557	0.011	0.04 [*]
	H3K27me3 bound	0.211	0.017	0.04 [*]
	Homeobox genes	0.6098	0.014	0.04 [*]
	SUZ12 targets	0.2559	0.017	0.04 [*]
	NOSTF targets	0.1956	0.158	0.34
Hypomethylation	PRC2 targets (2)	-0.402	0.045	0.33
	OCT4 targets	-0.1652	0.093	0.33
	NOS targets	-0.1803	0.106	0.33
	NANOG targets	-0.1923	0.125	0.33
	ES2	-0.2615	0.125	0.33

^a only gene sets with FDR<50% are shown in the table;

^{*} FDR 0.05