Original Research Article

Progress in Microbes and Molecular Biology

HH PUBLISHER

Landscape Of *HOXA* Genes Methylation in Colorectal Cancer

Muhiddin Ishak[1], Rashidah Baharudin[1], Loh Teng-Hern Tan[2], Learn-Han Lee[2*], Nurul-Syakima Ab Mutalib[1*]

[1]UKM Medical Molecular Biology Institute (UMBI), Universiti Kebangsaan Malaysia, Jalan Yaacob Latif, 56000 Cheras, Kuala Lumpur, Malaysia
[2]Novel Bacteria and Drug Discovery Research Group (NDBB), Microbiome and Bioresource Research Strength (MBRS), Jeffrey Cheah School of Medicine and Health Sciences, Monash University Malaysia, 47500 Bandar Sunway, Selangor Darul Ehsan, Malaysia

Abstract: Colorectal cancer (CRC) is among the most common cancers worldwide and the second leading cause of cancer-related death in Malaysia. The *HOXA* gene cluster is a family of Homeobox A genes encoding transcriptional regulators that play vital roles in cancer susceptibility and progression. Dysregulated *HOXA* expression influences various aspects of carcinogenesis processes. Therefore, this study aims to elucidate the methylation landscape of *HOXA* genes in CRC. Twelve pairs of CRC — adjacent normal tissues were subjected to Infinium DNA MethyEPIC array. Differentially methylated regions were identified using the ChAMP Bioconductor and methylation levels of *HOXA* genes were manually curated. We identified 100 significantly differentially methylated probes annotated to *HOXA* genes. *HOXA3* has the highest number of differentially methylated probes (n=27), followed by *HOXA2* (n=20) and *HOXA4* (n=14). The majority (43%) of the probes were located at the transcription start site (TSS) 200, which is one of the gene promoters. In respect to CpG islands (CGI), the probes were equally located in the island and shore regions (47% each) while a minor percentage was in the shelf (6%). Our work gave a comprehensive assessment of the DNA methylation pattern of *HOXA* genes and provide the first evidence of *HOXA2, HOXA3* and *HOXA4* differential methylation in Malaysian CRC. The new knowledge from this study can be utilized to further increase our understanding of CRC methylomics, particularly on the homeobox A genes. The prognostic and diagnostic roles of the differentially methylated *HOXA* genes warrant future investigations.

Keywords: Homeobox A genes; colorectal cancer; DNA methylation; *HOXA2*; *HOXA-AS3*

***Correspondence:** Nurul-Syakima Ab Mutalib, UKM Medical Molecular Biology Institute (UMBI), Universiti Kebangsaan Malaysia, Jalan Yaacob Latif, 56000 Cheras, Kuala Lumpur, Malaysia; syakima@ppukm.ukm.edu.my. Learn-Han Lee, Novel Bacteria and Drug Discovery Research Group (NDBB), Microbiome and Bioresource Research Strength (MBRS), Jeffrey Cheah School of Medicine and Health Sciences, Monash University Malaysia, 47500 Bandar Sunway, Selangor Darul Ehsan, Malaysia; lee.learn.han@monash.edu.

Received: 10th April 2020
Accepted: 12th May 2020
Published Online: 18th May 2020

Citation: Ishak M, Baharudin R, Tan LT-H *et al*. Landscape Of *HOXA* Genes Methylation in Colorectal Cancer. Prog Microbes Mol Biol 2020; 3(1): a0000085. https://doi.org/10.3687/pmmb.a0000085

Introduction

Cancer is a continuous global burden and colorectal cancer (CRC) placed as the fourth most frequently diagnosed cancer worldwide[1] and second in Malaysia[2,3]. Cancer occurs through the accumulation of multiple genetics and epigenetics changes[4]. Somatic mutation in *APC*, *BRAF*, *KRAS*, *PIK3CA* and *TP53*[5–8] are identified in CRC at varying frequencies and are perceived as the drivers of CRC formation. Although there are many efforts on investigating the molecular alterations involved in CRC pathogenesis, the existing knowledge remains inadequate for an early diagnosis and prognosis assessment. Therefore, further understanding of epigenetic components involved in CRC carcinogenesis is highly sought after and will unravel new genes which can be utilized as the diagnostic, prognostic and predictive biomarkers to prevent CRC-related mortality.

Epigenetics mechanism can be generally categorized into histone modification and DNA methylation[9], with the latter being the most widely investigated. The clinical application of DNA methylation markers to determine at-risk patient populations, improve diagnostic criteria, and provide prognostic factors to guide treatment decisions are becoming increasingly relevant. This

is because DNA methylation can be reversed, thereby providing alternative treatment options for patients with methylation phenotype. Clinically, DNA methylation has also been demonstrated to have significant utility owing to its stability and relative ease of testing[9]. For instance, *SEPT9* gene is the first US Food and Drug Administration (FDA)–approved diagnostic assay for CRC screening (https://www.epiprocolon.com). However, the clinical utility of the methylated *SEPT9* assay is still limited owing to patients' heterogeneity factor which includes various demographic characteristics and pathological features[10]. The search for a good diagnostic, prognostic and predictive DNA methylation markers in CRC is an active area of research.

Homeobox genes play a role as the master regulators of morphogenesis and are aberrantly expressed in cancer[11]. These genes possess a highly conserved DNA sequence that code for the homeodomain proteins, which act as transcription factors that bind specifically to the DNA motifs and regulate the genes involved in cellular processes including adhesion, proliferation and differentiation[11]. Over 200 homeobox genes have been discovered in the human genome and are divided into four *HOX* gene clusters namely *HOXA*, *HOXB*, *HOXC*, and *HOXD*, positioned at the chromosome 7p15, 17q21.2, 12q13, and 2q31 loci, respectively. Of these, there are 11 *HOXA* genes[12] (Table 1).

Table 1. The list of 11 HOXA genes under HOXL subclass homeoboxes [11].

HGNC ID (gene)	**Approved symbol**	**Approved name**
HGNC:5099	*HOXA1*	Homeobox A1
HGNC:5103	*HOXA2*	Homeobox A2
HGNC:5104	*HOXA3*	Homeobox A3
HGNC:5105	*HOXA4*	Homeobox A4
HGNC:5106	*HOXA5*	Homeobox A5
HGNC:5107	*HOXA6*	Homeobox A6
HGNC:5108	*HOXA7*	Homeobox A7
HGNC:5109	*HOXA9*	Homeobox A9
HGNC:5100	*HOXA10*	Homeobox A10
HGNC:5101	*HOXA11*	Homeobox A11
HGNC:5102	*HOXA13*	Homeobox A13

Dysregulated homeobox gene expression is a frequent in cancer and one of the mechanisms causing such dysregulation is DNA methylation. The *HOXA* cluster is often methylated in the non–small cell lung cancer (NSCLC)[13,14], while *HOXA5* and *HOXA11* promoter methylation diminishes their tumour-suppressive function through transcriptional silencing[15,16]. *HOXA11* is hypermethylated in gastric cancer tissues and is significantly associated with TNM III and IV patients[17]. Hypomethylation *HOXA10* and *HOXA11* able to discriminate ovarian cancer tissue from the normal tissue[18]. *HOXA9* was reported to be hypermethylated in half of the ovarian cancer patients and is significantly associated with endometrioid histological subtype[19]. In bladder cancer, *HOXA9* promoter methylation has been linked with cisplatin chemotherapy-resistant and metastatic bladder cancer, and reversing the DNA methylation using decitabine sensitized the cancer cells to cisplatin[20]. Yet, little is known about *HOXA* methylome in CRC. Only recently, Li and colleagues reported the DNA methylation status of three HOXA genes, which are *HOXA2*, *HOXA5* and *HOXA6* in CRC using targeted bisulfite sequencing assay[21]. Nevertheless, a comprehensive, unbiased methylome profile of *HOXA* genes has not been described. Therefore, this study aims to investigate the methylation landscape of *HOXA* genes in CRC context.

Material and Methods

Clinical specimens

The archived 12 pairs of tumour-adjacent normal fresh frozen colon tissues (n=24) from CRC patients were retrieved from the UMBI-HCTM Biobank. The tissues were collected according to the procedures approved by UKM Research Ethics Committee. As a quality control procedure, all tissues were cryosectioned, followed by haematoxylin and eosin staining for the pathologist to determine the percentage of tumour cells and normal cells contents. Only tumour samples with ≥ 80% cancerous cells and normal adjacent colon tissues with ≤ 20% necrosis were selected for DNA extraction using Allprep DNA/RNA/miRNA Universal Kit (Qiagen, USA) according to the manufacturer's instructions. Then, the integrity of DNA was assessed using agarose gel electrophoresis while the quantity and purity were evaluated using Nanodrop 2000c Spectrometer (Thermo Fisher Scientific, USA).

Bisulfite Conversion and Methylation Microarray

Five hundred nanogram (500 ng) of DNA was subjected to bisulfite conversion to change all unmethylated cytosine to uracil using the EZ DNA methylation — Gold kit (Zymo Research, USA) according to the manufacturer's protocol. The effectiveness of bisulfite conversion was determined using Universal Methylated DNA Standard & Control Primers (Zymo Research, USA) according to the manufacturer's protocol. The Infinium DNA MethylationEPIC assay, covering 850,000 CpG dinucleotides spread over the whole genome, was performed according to the manufacturer's specifications (Illumina, Inc.).

Methylation Microarray Data Analysis

The raw idat files obtained from methylation microarray were subsequently analyzed using GenomeStudio V1.9.0 and CHAMP Bioconductor packages[22]. Filters were applied to all datasets where CpG sites with detection p-values ≥ 0.01 in one or more samples were omitted from further analysis. To reduce the technical biases intrinsic to the probe design, the raw intensities were SWAN-normalized prior to statistical analysis[23]. β-values were then extracted and subjected to further statistical analysis.

Expression of *HOXA* genes from The Cancer Genome Atlas (TCGA) study

The mRNA expression data of selected *HOXA* genes in CRC were retrieved using web-based FireBrowse Gene Expression Viewer tool (https://gdac.broadinstitute.org/)

from Broad Institute. This tool provides access to results of various omics analyses involving more than 14,000 cancer cases, from 38 types of cancers based on TCGA data version 2016_01_28.

Statistical Analysis

A T statistic from the limma Bioconductor package was used to determine the differentially methylated CpG sites[24,25]. The CpG sites were further filtered at an adjusted p-value < 0.05 to identify significant differentially methylated *HOXA* genes. To substantiate the specificity and accuracy of the differentially methylated probes, the discriminative performance of the probes was evaluated by receiver operating characteristic (ROC) curves, and the area under the ROC curve (AUC), specificity, and sensitivity at the optimal cut-offs were determined using GraphPad Prism V8 (GraphPad Software, Inc., USA).

Results

Locations of differentially methylated loci in *HOXA* genes

We analysed the differential methylation status of 12 CRC tissue samples with the 12 adjacent cancer-free colonic tissue samples and only differentially methylated regions with adjusted p-value < 0.05 were reported. From the list of differentially methylated probes, we further filtered for *HOXA* genes. Here, we found that there are 100 CRC-associated differentially methylated probes in 11 *HOXA* genes, noncoding *HOXA-AS3* and *HOXA10-HOXA9* readthrough. *HOXA3* has the highest number of differentially methylated probes (n=27), followed by *HOXA2* (n=20) and *HOXA4* (n=14) (Figure 1A). The majority (43%) of the probes were located at the transcription start site (TSS) 200 (Figure 1B), which is one of the gene promoters. In respect to CpG islands (CGI), the probes were equally located in the island region and shore regions (47% each) while a minor percentage was in the shelf (6%) (Figure 1C).

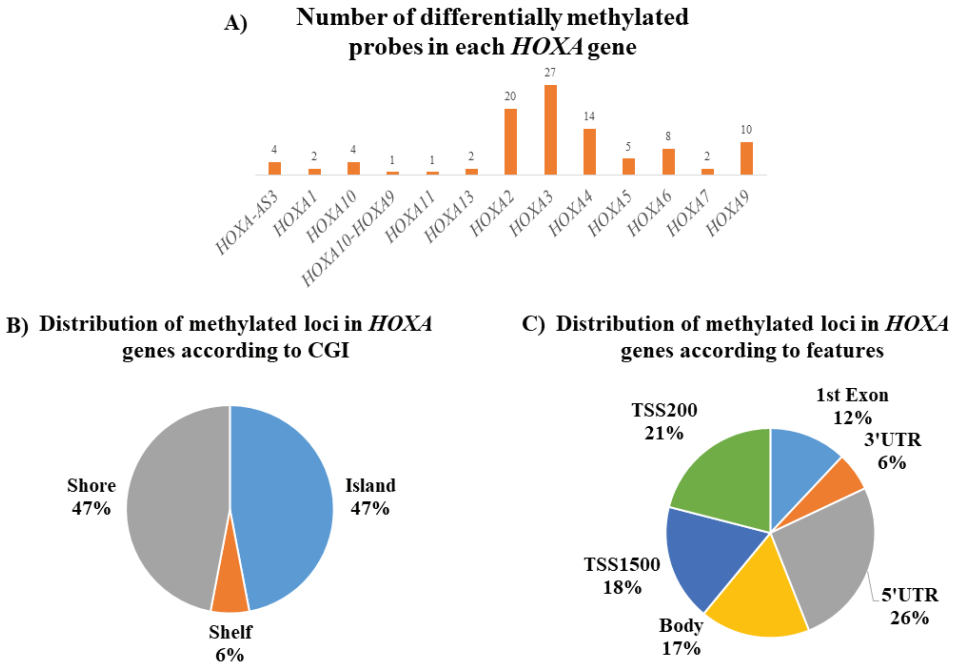

Figure 1. Differentially methylated HOXA genes in CRC. (A) The number of differentially methylated probes in each HOXA genes. (B) Distribution of methylated loci in HOXA genes with respect to features. (C) Distribution of methylated loci in HOXA genes with respect to CGI.

The genomic and gene-related regions of the significant differentially methylated *HOXA* genes were distributed differently. Generally, 53 probes (in seven genes) were hypermethylated compared to 47 loci (eight genes) that were hypomethylated. The largest portion of hypomethylated sites (55.3%) were in the shore and subsequently decreased in other categories (island 42.6% and shelf 2.1%). In contrast, more than half (50.9%) of the significantly hypermethylated loci of *HOXA* genes were on the island, followed by the shore (39.6%), and shelf (9.4%). None of the loci was identified in opensea region. Meanwhile, most of the significantly hypomethylated loci were in the 5'UTR (46.8%), followed by TSS200 (14.9%), gene body (12.8%), and 8.5% in each 1st exon, 3'UTR and TSS1500. Meanwhile, around a quarter (26.4%) of the significant hypermethylated loci were located in TSS200 and TSS1500, while the rest were mainly found in the gene body (20.8%), 1st exon (15.1%), and, to a lesser extent, in the 5' and 3' UTR (7.5% and 3.8%, respectively).

Differentially methylated HOXA genes

All of the 11 *HOXA* genes are significantly differentially methylated. Due to the power of comprehensive contents in the microarray platform, we also identified significant hypomethylation of the noncoding *HOXA-AS3* and *HOXA10-HOXA9* readthrough. High resolution, probe-level analyses revealed hypomethylation of 20 loci in *HOXA3*, with the remaining seven loci were hypermethylated. In *HOXA2* and *HOXA9*, all the probes were hypermethylated, while probes in *HOXA4* were hypomethylated. *HOXA6* exhibited eight hypermethylated loci and only one was hypomethylated. The summary of hypo- and hypermethylated probes in each gene were summarized in Table 2.

Table 2. The number of hypermethylated and hypomethylated probes in the *HOXA* genes.

Genes	Number of hypermethylated probes	Genes	Number of hypomethylated probes
HOXA1	2	*HOXA-AS3*	4
HOXA2	20	*HOXA10*	4
HOXA3	7	*HOXA10-HOXA9*	1
HOXA5	5	*HOXA11*	1
HOXA6	7	*HOXA13*	2
HOXA7	2	*HOXA3*	20
HOXA9	10	*HOXA4*	14
		HOXA6	1

The 100 significant probes with the methylation changes (Δβ) are illustrated in Table 3. It is worth mentioning that the loci in *HOXA2* were the locations with the highest methylation changes; these probes were hypermethylated in CRC as compared to the normal colon. On the other hand, the loci in *HOXA3* were the most hypomethylated probes in CRC as compared to the normal colon.

Table 3. The 100 significant differentially methylated probes in *HOXA* genes.

Genes	Probes	Adjusted p-value	Δβ	Feature	CGI
HOXA2	cg06055873	2.66E-05	0.368	1st Exon	shore
HOXA2	cg24058604	1.35E-04	0.362	TSS200	shore
HOXA2	cg05921905	2.49E-04	0.362	TSS200	shore
HOXA2	cg04737131	3.75E-04	0.357	TSS1500	shore
HOXA2	cg17353412	6.30E-06	0.357	1st Exon	shore
HOXA2	cg20747380	2.23E-05	0.356	1st Exon	shore
HOXA2	cg02979457	3.64E-04	0.326	TSS200	shore
HOXA2	cg22943986	1.19E-03	0.322	TSS1500	shore
HOXA2	cg06786372	5.16E-03	0.319	Body	shore
HOXA2	cg26069745	1.55E-04	0.317	1st Exon	shore
HOXA2	cg06769202	8.92E-04	0.312	TSS200	shore
HOXA2	cg09871315	1.95E-03	0.293	TSS1500	shore
HOXA2	cg23979631	2.88E-04	0.291	TSS200	shore
HOXA5	cg03744763	7.31E-03	0.286	TSS1500	island
HOXA1	cg07450037	1.02E-03	0.282	Body	shore
HOXA2	cg20087093	2.13E-03	0.276	TSS1500	shore
HOXA2	cg02803819	7.12E-03	0.258	Body	shelf
HOXA9	cg12600174	1.19E-02	0.249	TSS200	island
HOXA3	cg27539480	4.52E-03	0.244	3'UTR	shore
HOXA2	cg23206851	6.94E-03	0.243	TSS1500	shore
HOXA3	cg02627455	5.21E-03	0.240	5'UTR	shelf
HOXA2	cg13661519	4.09E-03	0.240	Body	shelf
HOXA3	cg07153966	2.74E-02	0.234	Body	island
HOXA9	cg21001184	1.42E-02	0.228	TSS200	island
HOXA9	cg03217995	3.19E-02	0.226	Body	shore
HOXA2	cg00188704	3.46E-02	0.222	Body	shelf
HOXA1	cg03116258	4.38E-04	0.220	1st Exon	shore
HOXA5	cg14882265	4.94E-02	0.211	TSS1500	island
HOXA2	cg02225599	1.51E-02	0.204	TSS1500	island
HOXA3	cg14216068	9.16E-03	0.194	3'UTR	island
HOXA3	cg09591524	3.06E-02	0.186	5'UTR	island
HOXA6	cg14044640	3.34E-02	0.184	TSS200	island
HOXA5	cg03368099	2.16E-02	0.183	TSS1500	island
HOXA7	cg20725013	2.91E-02	0.182	Body	shore
HOXA3	cg02439266	1.46E-02	0.182	5'UTR	island
HOXA5	cg01748892	3.58E-03	0.180	TSS1500	island

HOXA3	cg12305431	1.79E-02	0.175	5'UTR	shelf
HOXA5	cg13694927	1.20E-02	0.168	TSS1500	island
HOXA9	cg26476852	4.68E-02	0.167	1st Exon	island
HOXA9	cg20399871	2.10E-02	0.164	1st Exon	island
HOXA6	cg03529432	4.98E-02	0.161	TSS200	island
HOXA9	cg16104915	1.81E-02	0.154	TSS200	island
HOXA6	cg22469274	4.68E-02	0.154	TSS200	island
HOXA6	cg09936824	3.98E-02	0.153	TSS1500	island
HOXA6	cg19183743	4.34E-02	0.150	TSS1500	shore
HOXA7	cg21778348	3.36E-02	0.149	Body	island
HOXA9	cg16913789	4.62E-02	0.146	Body	island
HOXA2	cg01217984	1.10E-02	0.134	TSS1500	island
HOXA9	cg05065989	1.07E-02	0.129	TSS200	island
HOXA9	cg07778029	3.18E-02	0.125	1st Exon	island
HOXA6	cg04265576	2.56E-02	0.123	TSS200	island
HOXA9	cg03698009	4.16E-02	0.108	Body	island
HOXA6	cg12810523	4.84E-02	0.103	TSS200	island
HOXA10	cg08938793	4.30E-02	-0.041	3'UTR	shore
HOXA6	cg23590202	8.97E-04	-0.052	TSS1500	shore
HOXA13	cg01363170	4.33E-02	-0.056	3'UTR	shelf
HOXA-AS3	cg10374314	4.89E-02	-0.070	Body	shore
HOXA10	cg05092861	5.66E-03	-0.078	TSS200	shore
HOXA4	cg23884241	2.58E-02	-0.078	1st Exon	island
HOXA4	cg04317399	1.43E-02	-0.084	1st Exon	island
HOXA13	cg02366798	4.96E-02	-0.089	3'UTR	shore
HOXA4	cg03724423	2.08E-02	-0.096	TSS1500	shore
HOXA4	cg11410718	1.61E-02	-0.108	TSS200	island
HOXA4	cg07317062	2.15E-02	-0.114	5'UTR	island
HOXA10	cg01078824	2.94E-02	-0.118	TSS200	shore
HOXA4	cg19142026	1.84E-02	-0.137	5'UTR	island
HOXA4	cg17591595	4.32E-02	-0.151	TSS1500	shore
HOXA4	cg22997113	1.80E-02	-0.170	1st Exon	island
HOXA3	cg16406967	1.99E-02	-0.196	5'UTR	island
HOXA3	cg22798849	3.57E-02	-0.198	5'UTR	island
HOXA3	cg18680977	1.76E-02	-0.209	5'UTR	island
HOXA3	cg16748008	4.02E-02	-0.220	5'UTR	island
HOXA4	cg11532431	2.10E-02	-0.231	Body	island
HOXA11	cg05516617	8.55E-03	-0.231	3'UTR	shore
HOXA3	cg23403004	2.42E-04	-0.244	5'UTR	shore
HOXA3	cg04778178	2.48E-04	-0.254	5'UTR	island
HOXA3	cg16644023	4.88E-02	-0.255	5'UTR	shore
HOXA3	cg15982700	4.41E-02	-0.259	5'UTR	shore
HOXA3	cg23806243	2.21E-03	-0.260	5'UTR	shore
HOXA3	cg00318947	1.81E-02	-0.270	5'UTR	shore
HOXA4	cg00562553	1.01E-02	-0.272	1st Exon	island
HOXA4	cg20171892	5.24E-04	-0.273	Body	island
HOXA3	cg04351734	2.10E-02	-0.274	5'UTR	island
HOXA10	cg05517976	1.19E-02	-0.275	TSS200	shore
HOXA4	cg11227540	1.46E-02	-0.279	Body	shore
HOXA4	cg09799676	3.54E-03	-0.286	Body	island
HOXA3	cg21556281	1.04E-02	-0.291	5'UTR	shore
HOXA-AS3	cg14429861	7.94E-03	-0.299	TSS200	shore
HOXA3	cg18430152	7.29E-04	-0.305	5'UTR	island

HOXA10-HOXA9	cg22274074	1.72E-03	-0.307	TSS1500	shore
HOXA3	cg14072564	6.32E-04	-0.311	5'UTR	island
HOXA4	cg17132446	4.36E-03	-0.323	Body	shore
HOXA3	cg01820751	2.39E-03	-0.328	5'UTR	shore
HOXA-AS3	cg06188746	3.12E-03	-0.331	TSS200	shore
HOXA-AS3	cg18091117	3.05E-03	-0.331	TSS200	shore
HOXA3	cg03483713	4.12E-04	-0.346	5'UTR	shore
HOXA3	cg26297005	1.81E-03	-0.361	5'UTR	island
HOXA3	cg15725372	6.95E-04	-0.367	5'UTR	island
HOXA3	cg00431187	7.17E-04	-0.381	5'UTR	shore
HOXA3	cg09798023	2.84E-04	-0.390	5'UTR	shore

TSS: transcription start site

UTR: untranslated regions

Expression of *HOXA* genes and their correlation with methylation level

Using FireBrowse, gene expression of the *HOXA* genes were retrieved from COAD[5] and COADREAD[26] studies (Figure 2). The data were presented as log2 fold change. The expression of seven *HOXA* genes was downregulated (*HOXA1*, *HOXA2*, *HOXA4*, *HOXA5*, *HOXA6*, *HOXA7*, and *HOXA13*) while four of genes (*HOXA3*, *HOXA9*, *HOXA10*, *HOXA11*) were upregulated. The expression profiles of *HOXA1*, *HOXA2, HOXA5, HOXA6, HOXA7, HOXA10*, and *HOXA11* are inversely related to the methylation level as predicted, but not the *HOXA4*, *HOXA9* and *HOXA13*.

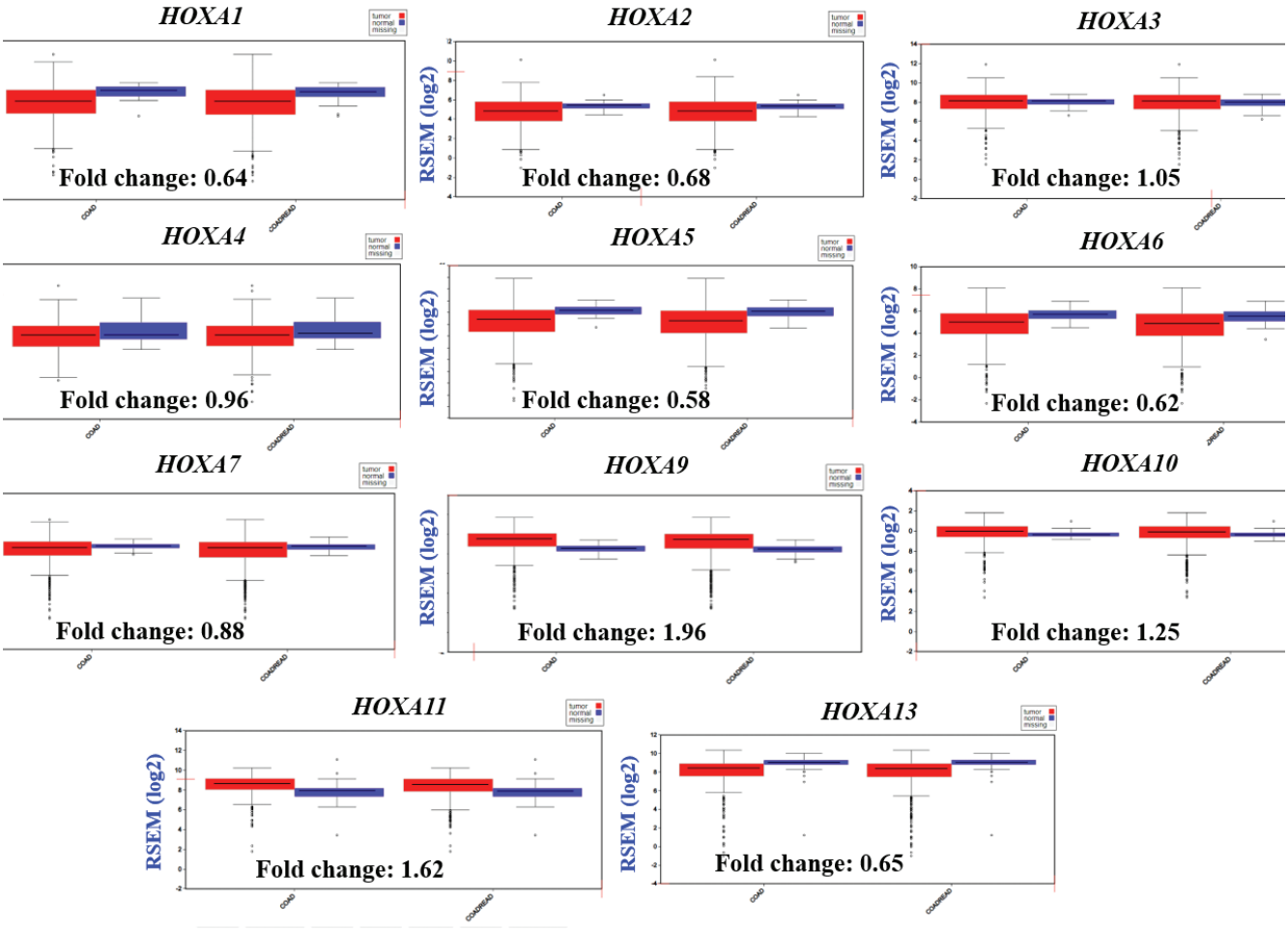

Figure 2. Gene expression of *HOXA* genes from TCGA COAD and COAREAD studies.

Receiver operating characteristics (ROC) curve analysis

Lastly, the sensitivity and specificity of the methylation levels were further assessed using receiver-operator curve (ROC) analysis. The methylation levels of 10 topmost hypermethylated CpG sites significantly differentiated the CRCs from the normal colonic tissues (*p*-value 0.0032 to 0.0002) (Table 4). The highest discriminative accuracy was demonstrated by *HOXA2* cg06055873 (AUC = 0.9514, confident interval = 0.8547 to 1.000, *p*-value = 0.0002). Other candidate probes also reached high diagnostic accuracy (Table 4; Figure 3).

Table 4. Receiver operating characteristics (ROC) curve analysis of the top 10 differentially methylated probes in *HOXA2* gene.

Gene_probe	Area	Std. Error	95% confidence interval	*P* value
HOXA2 cg06055873	0.9514	0.04935	0.8547 to 1.000	0.0002
HOXA2 cg24058604	0.9444	0.04489	0.8565 to 1.032	0.0002
HOXA2 cg05921905	0.875	0.08441	0.7096 to 1.04	0.0018
HOXA2 cg04737131	0.8958	0.07498	0.7489 to 1.043	0.001
HOXA2 cg17353412	0.9444	0.05546	0.8357 to 1.053	0.0002
HOXA2 cg20747380	0.9375	0.06155	0.8169 to 1.058	0.0003
HOXA2 cg02979457	0.8681	0.07918	0.7129 to 1.023	0.0022
HOXA2 cg22943986	0.8611	0.09216	0.6805 to 1.042	0.0027
HOXA2 cg06786372	0.8542	0.08306	0.6914 to 1.017	0.0032
HOXA2 cg26069745	0.9167	0.0691	0.7812 to 1.052	0.0005

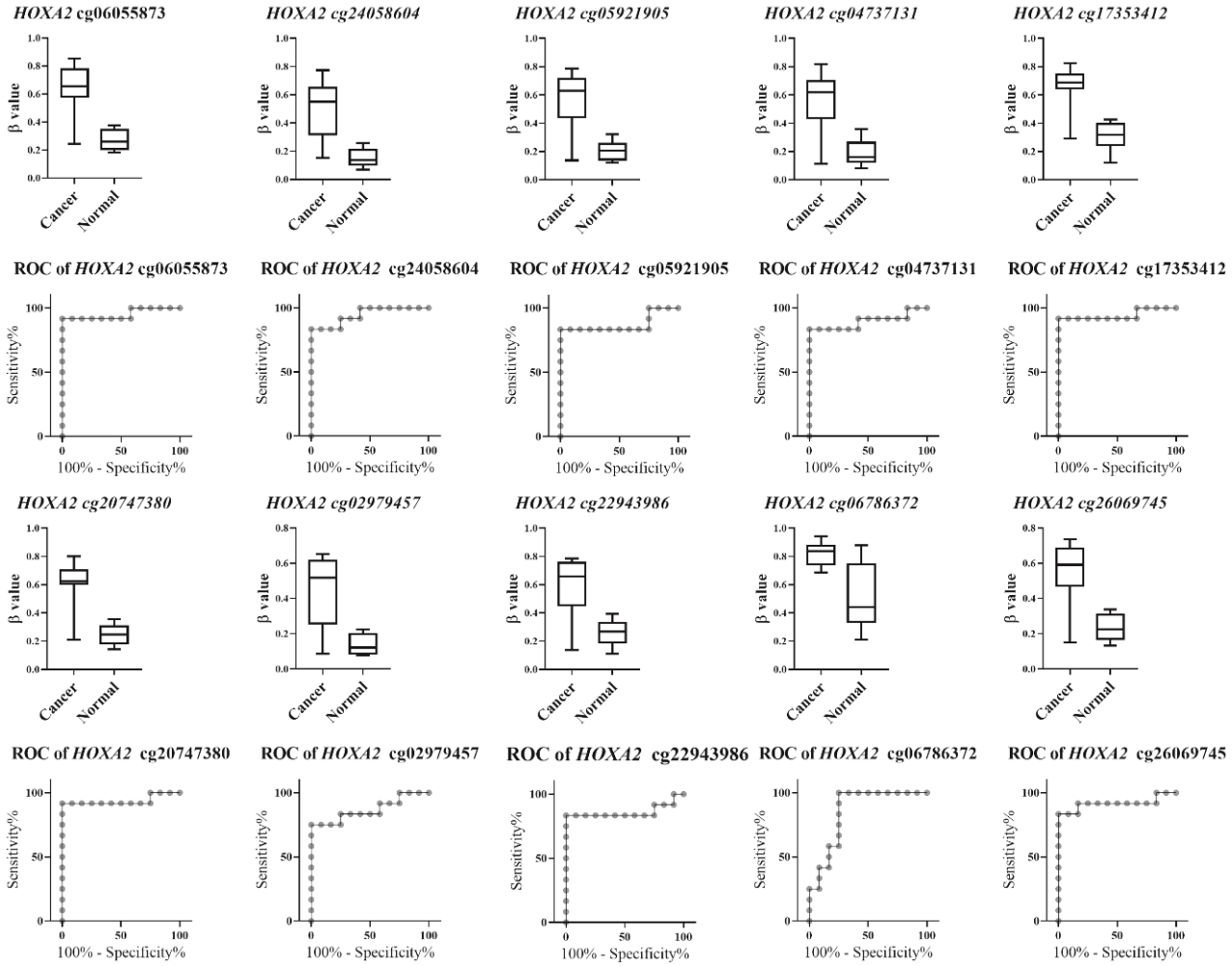

Figure 3. Box plot illustrating the comparison of β values and ROC curve-based evaluation of the diagnostic accuracy for the top 10 hypermethylated *HOXA2* probes in cancerous and normal tissues.

Discussion

In this study, we analyzed in greater detail the genome-wide methylation patterns of *HOXA* genes from 12 CRCs compared with their adjacent normal tissues. Interestingly, due to the high throughput nature of microarray platform, our analysis revealed that all of the 11 *HOXA* genes are significantly differentially methylated in CRCs. We are the first to report this new finding and shed new light on the possible role of these genes in colorectal carcinogenesis. Epigenetic alteration of *HOXA* genes has been widely studied in many cancers, especially non-small cell lung cancer (NSCLC)[27–29], yet, the investigation about this gene cluster in CRC is lacking. To date, there is only a handful of study which investigates the methylome of *HOXA* genes[21,30–32].

HOXA2 is the most significantly hypermethylated *HOX* gene in our study, and the methylated loci were mostly located in the promoter regions (TSS200 and TSS1500). Hypermethylation of this gene in CRC has also been recently reported in concordance with our finding[21]. In addition, the significant association between *HOXA2* methylation with age, node status (N), stage, metastasis (M), lymphovascular invasion, perineural invasion, as well as the number of lymph node was also demonstrated[21]. Another study has shown that *HOXA2* is hypermethylated in the rectal cancer mucosa compared to the nonmalignant rectal mucosa[30]. DNA methylation is known to be inversely correlated with mRNA expression; however, the published data on *HOXA2* expression is lacking and the aforementioned two studies did not investigate the gene expression levels. Therefore we attempted to investigate the mRNA expression using the TCGA CRC dataset and the finding is in agreement with our hypothesis. The expression of *HOXA2* is indeed downregulated in CRC compared to the normal tissues by 0.68 fold, and thus, could be explained by its hypermethylation status. Nevertheless, *HOXA2* promoter is also found to be hypermethylated in nasopharyngeal cancer, whereby it associates with low mRNA expression in the biopsies and cell lines[33]. Moreover, in a study involving 101 patients from stage I–III NSCLC, methylation of *HOXA2* was proposed to have prognostic significance in squamous cell carcinoma (SCC) subtypes patients[34]. Due to the limited number of patients and the lack of clinical information in our study, the association with clinical features were not established and warrant further investigation. In addition, ROC curves for *HOXA2* gene show exceptional diagnostic ability in differentiating CRC from the normal healthy tissue, especially in Stage I patients with the AUC = 0.9979 [21]. We also observed a similar trend, whereby the loci in *HOXA2* are the most significantly hypermethylated and exhibited high discriminative accuracy.

HOXA5 and *HOXA6* promoters were both hypermethylated in our study and this is supported by recent data by Li and colleagues[21]. The authors went on to demonstrate a significant association between *HOXA5* hypermethylation and age, tumour (T), metastasis (M), stage, and patients' tumour status, while *HOXA6* hypermethylation is correlated with age and presence of *KRAS* mutation. Similarly, with *HOXA2* genes, *HOXA5* and *HOXA6* have not been studied in detail in CRC. In other malignancies, such as NSCLC, *HOXA5* is hypermethylated[35] and a separate study showed that low *HOXA5* expression indicates unfavourable prognosis and reduces cell proliferation by via p21 expression[36]. Our TCGA analyses revealed downregulation of *HOXA5* and *HOXA6* by 0.58 and 0.62 fold, respectively. The relationship between downregulation of *HOXA5* and *HOXA6* with CRCs patients prognosis is the subject for further research. *HOXA5* also plays a role in haematopoietic differentiation, whereby *HOXA5* is hypermethylated in the development of acute myeloid leukaemia (AML)[37]. Additionally, *HOXA6* hypermethylation was reported in oral cancer [38] and more recently in meningiomas[39].

We observed the hypomethylation of *HOXA3* and *HOXA4* among our patients, which is in disagreement with other CRC studies [31, 32]. It is unclear what causes the discrepancies and it will be worthwhile to reconfirm this finding in a larger cohort. By looking at the gene expression of *HOXA3* and *HOXA4* in CRC, several studies partly support our findings. For instance, Zhang and colleagues[40] demonstrated that *HOXA3* expression is increased in both CRC tissues and cell lines. Their analysis of the relationship between *HOXA3* and tumour progression has revealed that elevated *HOXA3* expression is linked with poor survival rates in CRC. On the other hand, our finding on *HOXA4* hypomethylation is also partially supported by Bhatlekar and colleagues whereby *HOXA4* is found to be overexpressed in CRC[41]. They further demonstrated that overexpression of *HOXA4* encourages self-renewal, leading to the overabundance of colon cancer stem cell[42], which play an essential role in the metastasis and relapse of this disease. Taken together, *HOXA3* and *HOXA4* hypomethylation, as identified from our study, may play an important role in CRC.

To the best of our knowledge, hypomethylation of the long noncoding RNA *HOXA-AS3* and *HOXA10-HOXA9* readthrough has never been reported before. Therefore we are the first to notice such observation in CRC. *HOXA10-HOXA9* readthrough represents a naturally occurring read-through transcription between the *HOXA10* and neighbouring *HOXA9* and is a candidate for nonsense-mediated mRNA decay (NMD), which does not produce any protein product. Published literature regarding this readthrough is severely lacking. On the other hand, *HOXA-AS3* has been gaining more attention from cancer researchers. In lung cancer, *HOXA-AS3* expression was significantly increased and inhibition of *HOXA-AS3* impairs cancer cell proliferation, migration, and invasion[43]. *In vitro* experiment further supported its oncogenic role, where the A549-derived xenografts with silenced *HOXA-AS3* had significantly reduced tumour weights and volumes. These findings suggest the potential application of *HOXA-AS3* inhibition as an effective targeted therapy for lung cancer patients[43]. The authors also concluded that the upregulated *HOXA-AS3* expression was shown to be caused by histone acetylation, and the link between histone deacetylation and DNA methylation has been established[44]. Furthermore, *HOXA-AS3* confers resistance towards cisplatin treatment via interaction with *HOXA3*

in NSCLC[45]. In glioma, *HOXA-AS3* upregulation promotes tumour progression and predicts poor prognosis[46]. It is probable that the hypomethylation of *HOXA-AS3* in our CRC patients could lead to its increased expression. It will be interesting to validate this observation and investigate its function in CRC.

An interrogation *HOXA*-associated oncogenes or tumour suppressors as prospective mechanisms as predictive biomarkers may offer novel therapeutic strategies for treating cancers[47], including CRC. Yet, the obstacle lays in the fact that our knowledge and understanding of *HOXA* genes in the context of CRC are still insufficient. In this study, we further extended the understanding of CRC pathology by investigating the methylome landscape of *HOXA* genes. Nevertheless, our study is not without limitation. While our sample size is rather small, the hypo- and hypermethylation of the *HOXA* genes reported in this study are relevant to carcinogenesis as reported in several studies. For future study, validation of *HOXA* methylation changes in cancer tissues from a larger cohort is necessary, and the association with survival and other clinicopathological data is warranted. Furthermore, an integrated analysis with gene expression data will be of importance to further establish the correlation between *HOXA* methylation and gene regulation.

Conclusion

Using the latest methylation microarray platform, we report a detailed, unbiased landscape of *HOXA* genes methylome and discovered epigenetically regulated candidate genes in CRC carcinogenesis. Specifically, our results provide the primary evidence that aberrant methylation of *HOXA2, HOXA3* and *HOXA4* in Malaysian CRC. The new knowledge from this study can be utilized to further increase our understanding of CRC methylomics, particularly on the homeobox A genes. The prognostic and diagnostic roles of the differentially methylated *HOXA* genes warrant future investigations.

Author Contributions

MI and RB performed the lab experiments, data analysis and manuscript writing. LT-HT, L-HL and NS-AM provided vital guidance for the project and improvement of the writing. The project was conceptualised by NS-AM.

Conflict of interest

The authors declare that there is no conflict of interest in this work.

Acknowledgement

This work was fully supported by Universiti Kebangsaan Malaysia under Research University Grant Scheme (GUP-2018-070 Geran Universiti Penyelidikan).

References

1. Bray F, Ferlay J, Soerjomataram I, *et al.*, Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018; 68(6): 394–424.
2. Azizah AM, Nor Saleha IT, Noor Hashimah A, *et al.*, *Malaysian national cancer registry report 2007-2011*. 2016, Putrajaya, Malaysia: National Cancer Institute.
3. Ishak M, Baharudin R, Rose IM, *et al.*, Genome-wide open chromatin methylome profiles in colorectal cancer. Biomol 2020; 10(5): 719–719.
4. Chow YP, Yunos RIM, Rose IM, *et al.*, Characterization of somatic mutations in Malaysian luminal breast cancer. Prog Microbes Mol Biol 2018; 1(1).
5. Atlas TCG, Comprehensive molecular characterization of human colon and rectal cancer. Nature 2012; 487(7407): 330–337.
6. Abdul S-N, Ab Mutalib N-S, Sean KS, *et al.*, Molecular characterization of somatic alterations in Dukes' B and C colorectal cancers by targeted sequencing. Front Pharmacol 2017; 8: 465–465.
7. Yusof AM, Tieng FYF, Muhammad R, *et al.*, In-depth characterization of miRNome in papillary thyroid cancer with BRAF V600E mutation. Prog Microbes Mol Biol 2020; 3(1).
8. Yunos R-IM, Ab Mutalib N-S, Khor SS, *et al.*, Whole exome sequencing identifies genomic alterations in proximal and distal colorectal cancer. Prog Microbes Mol Biol 2019; 2(1).
9. Werner RJ, Kelly AD, and Issa J-PJ, Epigenetics and Precision Oncology. Cancer J 2017; 23(5): 262–269.
10. Wang Y, Chen P-M, and Liu R-B, Advance in plasma SEPT9 gene methylation assay for colorectal cancer early detection. World J Gastrointest Oncol 2018; 10(1): 15–22.
11. Rodrigues MFSD, Esteves CM, Xavier FCA, *et al.*, Methylation status of homeobox genes in common human cancers. Genomics 2016; 108(5-6): 185–193.
12. Holland PW, Booth HA, and Bruford EA, Classification and nomenclature of all human homeobox genes. BMC Biol 2007; 5: 47–47.
13. Pradhan MP, Desai A, and Palakal MJ, Systems biology approach to stage-wise characterization of epigenetic genes in lung adenocarcinoma. BMC Syst Biol 2013; 7(1): 141–141.
14. Jaffe AE, Murakami P, Lee H, *et al.*, Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies. Int J Epidemiol 2012; 41(1): 200–209.
15. Wang C-C, Su K-Y, Chen H-Y, *et al.*, HOXA5 inhibits metastasis via regulating cytoskeletal remodelling and associates with prolonged survival in non-small-cell lung carcinoma. PLoS One 2015; 10(4): e0124191–e0124191.
16. Hwang J-A, Lee BB, Kim Y, *et al.*, HOXA11 hypermethylation is associated with progression of non-small cell lung cancer. Oncotarget 2013; 4(12): 2317–2325.
17. Bai Y, Fang N, Gu T, *et al.*, HOXA11 gene is hypermethylation and aberrant expression in gastric cancer. Cancer Cell Int 2014; 14(1): 79–79.
18. Cheng W, Jiang Y, Liu C, *et al.*, Identification of aberrant promoter hypomethylation of HOXA10 in ovarian cancer. J Cancer Res Clin Oncol 2010; 136(8): 1221–1227.
19. Wu Q, Lothe RA, Ahlquist T, *et al.*, DNA methylation profiling of ovarian carcinomas and their in vitro models identifies HOXA9, HOXB5, SCGB3A1, and CRABP1 as novel targets. Mol Cancer 2007; 6: 45–45.
20. Xylinas E, Hassler MR, Zhuang D, *et al.*, An epigenomic approach to improving response to neoadjuvant cisplatin chemotherapy in bladder cancer. Biomol 2016; 6(3).
21. Li D, Bai Y, Feng Z, *et al.*, Study of promoter methylation patterns of HOXA2, HOXA5, and HOXA6 and its clinicopathological characteristics in colorectal cancer. Front Oncol 2019; 9(394).
22. Tian Y, Morris TJ, Webster AP, *et al.*, ChAMP: updated methylation analysis pipeline for Illumina BeadChips. Bioinf 2017; 33(24): 3982–3984.
23. Maksimovic J, Gordon L, and Oshlack A, SWAN: Subset-quantile within array normalization for illumina infinium HumanMethylation450 BeadChips. Genome Biol 2012; 13(6): R44–R44.
24. Ritchie ME, Phipson B, Wu D, *et al.*, limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 2015; 43(7): e47–e47.
25. Wettenhall JM and Smyth GK, limmaGUI: a graphical user interface for linear modeling of microarray data. Bioinformatics 2004; 20(18): 3705–3706.
26. Giannakis M, Mu XJ, Shukla SA, *et al.*, Genomic correlates of immune-cell infiltrates in colorectal carcinoma. Cell Rep 2016; 15(4):

857–865.
27. Rauch T, Wang Z, Zhang X, *et al.*, Homeobox gene methylation in lung cancer studied by genome-wide analysis with a microarray-based methylated CpG island recovery assay. Proc Natl Acad Sci U S A 2007; 104(13): 5527–5532.
28. Ooki A, Maleki Z, Tsay J-CJ, *et al.*, A panel of novel detection and prognostic methylated DNA markers in primary non-small cell lung cancer and serum DNA. Clin Cancer Res 2017; 23(22): 7141–7152.
29. Nelson HH, Marsit CJ, Christensen BC, *et al.*, Key epigenetic changes associated with lung cancer development: Results from dense methylation array profiling. Epigenetics 2012; 7(6): 559–566.
30. Vymetalkova V, Vodicka P, Pardini B, *et al.*, Epigenome-wide analysis of DNA methylation reveals a rectal cancer-specific epigenomic signature. Epigenomics 2016; 8(9): 1193–1207.
31. Shiao SPK, Xiao H, Dong L, *et al.*, Genome wide DNA differential methylation regions in colorectal cancer patients in relation to blood related family members, obese and non-obese controls — A preliminary report. Oncotarget 2018; 9(39): 25557–25571.
32. Hongwei Wu QZ, Aberrantly Methylated CpG Island Detection in Colon Cancer. J Proteom Bioinf 2014; s9.
33. Li H-P, Peng C-C, Chung IC, *et al.*, Aberrantly hypermethylated Homeobox A2 derepresses metalloproteinase-9 through TBP and promotes invasion in Nasopharyngeal carcinoma. Oncotarget 2013; 4(11): 2154–2165.
34. Heller G, Babinsky VN, Ziegler B, *et al.*, Genome-wide CpG island methylation analyses in non-small cell lung cancer patients. Carcinogenesis 2013; 34(3): 513–521.
35. Kim D-S, Kim M-J, Lee J-Y, *et al.*, Epigenetic inactivation of Homeobox A5 gene in nonsmall cell lung cancer and its relationship with clinicopathological features. Mol Carcinog 2009; 48(12): 1109–1115.
36. Zhang M-l, Nie F-q, Sun M, *et al.*, HOXA5 indicates poor prognosis and suppresses cell proliferation by regulating p21 expression in non small cell lung cancer. Tumour Biol 2015; 36(5): 3521–3531.
37. Strathdee G, Sim A, Soutar R, *et al.*, HOXA5 is targeted by cell-type-specific CpG island methylation in normal cells and during the development of acute myeloid leukaemia. Carcinogenesis 2007; 28(2): 299–309.
38. Xavier FCA, Destro MFdSS, Duarte CME, *et al.*, Epigenetic repression of HOXB cluster in oral cancer cell lines. Arch Oral Biol 2014; 59(8): 783–789.
39. Galani V, Lampri E, Varouktsi A, *et al.*, Genetic and epigenetic alterations in meningiomas. Clin Neurol Neurosurg 2017; 158: 119–125.
40. Zhang X, Liu G, Ding L, *et al.*, HOXA3 promotes tumor growth of human colon cancer through activating EGFR/Ras/Raf/MEK/ERK signaling pathway. J Cell Biochem 2018; 119(3): 2864–2874.
41. Bhatlekar S, Addya S, Salunek M, *et al.*, Identification of a developmental gene expression signature, including HOX genes, for the normal human colonic crypt stem cell niche: overexpression of the signature parallels stem cell overpopulation during colon tumorigenesis. Stem Cells Dev 2014; 23(2): 167–179.
42. Bhatlekar S, Viswanathan V, Fields JZ, *et al.*, Overexpression of HOXA4 and HOXA9 genes promotes self-renewal and contributes to colon cancer stem cell overpopulation. J Cell Physiol 2018; 233(2): 727–735.
43. Zhang H, Liu Y, Yan L, *et al.*, Increased levels of the long noncoding RNA, HOXA-AS3, promote proliferation of A549 cells. Cell Death Dis 2018; 9(6): 1–14.
44. Irvine RA, Lin IG, and Hsieh C-L, DNA methylation has a local effect on transcription and histone acetylation. Mol Cell Biol 2002; 22(19): 6689–6696.
45. Lin S, Zhang R, An X, *et al.*, LncRNA HOXA-AS3 confers cisplatin resistance by interacting with HOXA3 in non-small-cell lung carcinoma cells. Oncogenesis 2019; 8(11).
46. Wu F, Zhang C, Cai J, *et al.*, Upregulation of long noncoding RNA HOXA-AS3 promotes tumor progression and predicts poor prognosis in glioma. Oncotarget 2017; 8(32): 53110–53123.
47. Chen S-L, Qin Z-Y, Hu F, *et al.*, The role of the HOXA gene family in acute myeloid leukemia. Genes 2019; 10(8): 621–621.