Biomedicine & Pharmacotherapy 108 (2018) 288–296

Contents lists available at ScienceDirect

Biomedicine & Pharmacotherapy

journal homepage: www.elsevier.com/locate/biopha

Genome-wide methylation analysis identifies novel prognostic methylation markers in colon adenocarcinoma

Weinan Xue[a,1], Xiangxin Wu[a,1], Fan Wang[b], Peng Han[a], Binbin Cui[a,*]

[a] Department of Colorectal Surgery, The Affiliated Tumor Hospital of Harbin Medical University, Harbin, 150081, China
[b] Department of Epidemiology, School of Public Health, Harbin Medical University, Harbin, 150081, China

ARTICLE INFO

Keywords:
Colon adenocarcinoma
DNA methylation profiles
Methylation markers

ABSTRACT

Previous studies have indicated that abnormal methylation is a critical and early event in the pathogenesis of most types of human cancer, which contributes to tumorigenesis. However, there has been little focus on the potential of DNA methylation patterns as predictive markers for the prognosis of colon adenocarcinoma (COAD). In the present study, a genome-wide comparative analysis of DNA methylation profiles was performed between 315 COAD samples and 38 matched tumor-adjacent normal tissue samples. A total of 675 differentially methylated regions (DMRs) associated with 630 genes were identified, including 654 hypermethylated regions (UMRs) and 21 hypomethylated regions, which were capable of distinguishing COAD samples from non-malignant tissue samples. Although most of the DMRs appeared to be located within the gene body or promoter regions, UMRs were mostly located within CpG islands. Functional analysis suggested that genes associated with DMRs were enriched in many of the core cancer-signaling pathways known to be important in COAD biology. A survival analysis was also performed, which identified 7 DMRs as potential candidate markers with the ability to classify patients into high and low-risk groups with significantly different overall survival. The present study provides a better understanding of the molecular mechanisms underlying COAD, and demonstrates the utility of aberrant DNA methylation in the prognosis of COAD.

1. Introduction

Colon adenocarcinoma (COAD) is one of the most common types of gastrointestinal cancer, and a leading cause of cancer mortality and morbidity worldwide. However, there has been a significant increase in its incidence and mortality rates over the past decade [1]. Surgical resection is the standard treatment for localized colon cancer (stage I-III), while other treatment methods, including cryotherapy, radiofrequency ablation and hepatic arterial infusion of chemotherapeutic agents are utilized to treat patients who are not surgical candidates [2]. Although the American Joint Committee on Cancer tumor, lymph node, metastases (TNM) staging system has contributed to the management of COAD patients, it remains inadequate for the prediction of prognosis, due to the molecular heterogeneity of COAD [3]. Therefore, the identification of sensitive and specific molecular markers for the recognition of high-risk patients is urgently required to enable personalized therapies and improve clinical outcomes.

DNA methylation is a major epigenetic mechanism, and an important regulator of gene expression, which can inhibit the binding of transcription factors or the recruitment of repression proteins [4]. Previous studies have demonstrated that DNA methylation in epigenetic gene regulation plays crucial roles in normal development and cell functions, including imprinting, X-inactivation and tissue-specific gene expression [5]. Previous studies have also shown that DNA methylation profiles differ significantly between cancer and normal tissues [6–8]. Increasing reports of extensive deregulation of DNA methylation in cancer cells has indicated that abnormal methylation changes, which lead to inappropriate gene expression, are a critical and early event in the pathogenesis of most types of human cancers, and contribute to tumorigenesis [9–11]. In addition, altered DNA methylation patterns have been shown to be the first detectable neoplastic changes associated with tumorigenesis. Furthermore, they may be easily detected in the plasma or serum of cancer patients, thereby highlighting the potential of DNA methylation as a novel molecular marker for cancer diagnosis and prognosis [12,13].

The aim of the present study was to investigate the altered DNA methylation patterns between COAD samples and adjacent tissue samples using high-throughput methylation profiles from a large number of

* Corresponding author.
E-mail address: binbincuihrb@126.com (B. Cui).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.biopha.2018.09.043
Received 4 July 2018; Received in revised form 26 August 2018; Accepted 8 September 2018

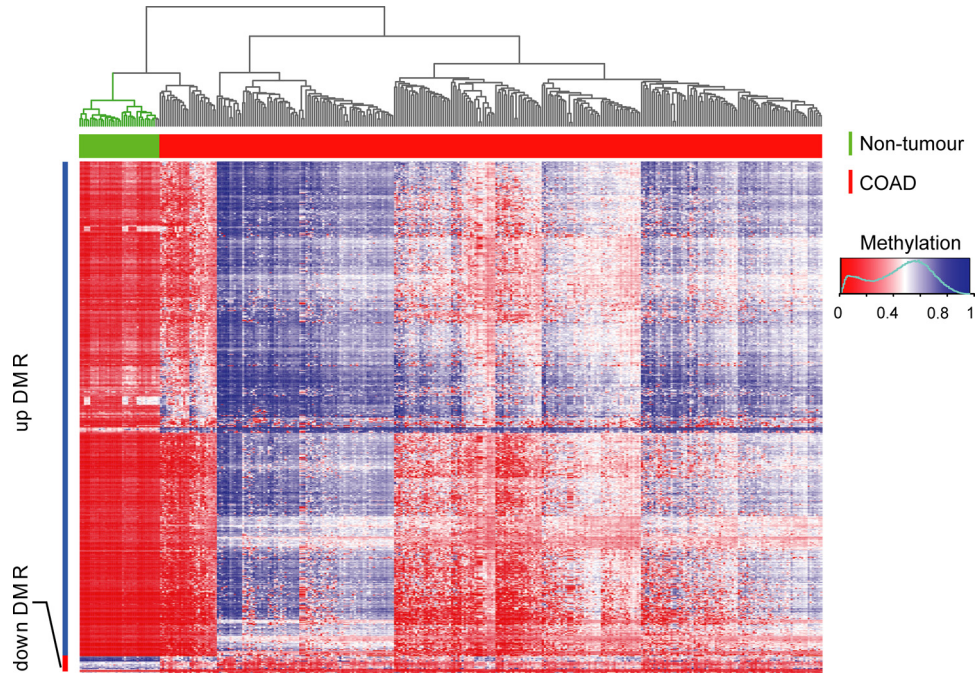

Fig. 1. Unsupervised hierarchical clustering and heat map of 353 samples based on the methylation levels of 675 differentially methylated regions.

patients. The present study aimed to identify specific DNA methylation sites as potential biomarkers, with the ability to predict the overall survival of patients with COAD.

2. Materials and methods

2.1. Data sources

DNA methylation data from 353 samples (315 COAD samples and 38 matched tumor-adjacent normal tissue samples) generated using the Illumina Human Methylation 450k Array were obtained from The Cancer Genome Atlas (TCGA, https://cancergenome.nih.gov/). RNA-Seq gene expression data of 329 COAD samples generated using Illumina HiSeq were obtained from UCSC Xena (https://xenabrowser.net/datapages/?dataset=TCGA.COAD.sampleMap/HiSeqV2&host=https://tcga.xenahubs.net) [14]. Clinical information for the COAD samples, including tumor stage, survival status and time was obtained from the GDC Data Portal (https://portal.gdc.cancer.gov/).

2.2. Preprocessing and analysis of DNA methylation data

The DNA methylation data were preprocessed and normalized using the minfi package as previously described [15], and subset-quantile within array normalization [16] was performed for probe filtering, color bias correction and background subtraction, as well as subset quantile normalization [17]. Differentially methylated regions (DMRs) between COAD samples and the adjacent tissue samples were identified using the minfi package [15]. The methylation level of the DMRs was estimated using the average methylation levels of multiple CpG probes mapped to the DMR.

2.3. Functional enrichment analyses

Gene Ontology (GO) enrichment analysis was performed using the clusterProfiler package for genes associated with DMRs, to identify over-represented GO terms in three categories (biological processes, molecular function and cellular component) [18]. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis for genes associated with DMRs was performed to identify statistically significant enriched pathways using KOBAS 3.0 software [19].

2.4. Statistical analyses

Hierarchical clustering of the methylation level of DMRs was performed using the distance matrix of Pearson correlation. COAD patients were divided into two patient groups defined as high- or low-methylation, using the median of the methylation levels as the cutoff. The Kaplan-Meier survival curve and log-rank tests were used to assess the survival differences between the two patient groups with high- or low-methylation using a “survival” package. If a CpG was within 100 kb of the transcription start site (TSS) of a gene according to a previous study [17], the correlation between DNA methylation level and gene expression was measured using the Pearson correlation coefficient as follows:

$$\mathrm{PCC}(dmr,\ gene) = \frac{1}{n-1}\sum_{i-1}^{n}\left(\frac{e(dmr,\ i) - \overline{e(dmr)}}{\sigma(dmr)} \times \frac{e(gene) - \overline{e(gene)}}{\sigma(gene)}\right)$$

Where e (gene, i) represents gene expression of sample i; $\sigma(gene)$ is the standardization of gene expression of n samples; e(dmr,i) represents the methylation level of DMR in the sample i; $\sigma(dmr)$ represents the standardization of methylation level of DMR of n samples. An association was considered significant if the P-value was < 0.05. |The correlation coefficient matrix in methylation levels among 7 prognostic methylation markers was visualized using the Performance Analytics package (https://rdrr.io/rforge/PerformanceAnalytics/.)

3. Results

3.1. Identification of DMRs between COAD samples and adjacent tissue samples

DNA methylation data of 485,577 loci were obtained from the TGCA. After preprocessing of the data and quality control, 467,971 probes were retained for further analysis. The DMRs between the COAD samples and the adjacent tissue samples were then determined using a