Integration Analysis of Differential Expressed Genes Regulated by Differential Methylated Regions for Predicting Prognosis in Human Colon Cancer

Xiaowen Peng  
Guangzhou First People's Hospital

Dong Dong  
Guangzhou First People's Hospital

Wentao Ou  
Guangzhou First People's Hospital

Yongke Xie  
Guangzhou First People's Hospital

Yuqi Luo (✉ eyluoyuqi@scut.edu.cn)  
Hospital, School of Medicine  https://orcid.org/0000-0002-4236-4449

Primary research

Keywords: colon cancer, DNA methylation, gene expression regulation, prognosis prediction

DOI: https://doi.org/10.21203/rs.3.rs-72157/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

Background Colon cancer is a leading cause of cancer-associated death globally, and numerous evidences show that different expressed gens (DEGs) regulated by differential methylated regions (DMRs) act an important role in tumor biology. However, the specific regulatory mechanism of DEGs related to DERs in colonic carcinogenesis is still unclear.

Materials and methods

RNA sequencing data and DNA methylation data of 455 colon adenocarcinoma (COAD) cases and 41 normal controls were downloaded from The Cancer Genomic Atlas (TCGA) to investigate the significant DEGs and DMRs. Gene ontology (GO) functional enrichment and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses were performed by DAVID database. To identify the hub genes regulated by methylation, univariate cox and multivariate cox regression analyses were concluded. Furthermore, Riskscore and nomogram were built to identify the prognosis prediction power of the hub genes in colon cancer patients.

Results

A total of 133 DEGs regulated by DMRs were identified through analyzing RNA-seq data and DNA methylation data from TCGA; GO functional enrichment and KEGG pathway enrichment analysis showed that the genes involved in the initiation and progression of colon cancer. Univariate cox regression analysis and multivariate cox regression analysis focused on the 7 hub genes associated with overall survival, whose expression negatively correlated with their methylated level; Riskscore and nomogram model showed that the hub gens served as potential biomarker for the prognosis prediction of colon cancer patient.

Conclusion

Our funding suggests that the DEGs regulated by DMRs involve in the carcinogenesis and development of colon cancer, and the aberrant methylated DEGs associated with overall survival of patients may be potential diagnosis and therapeutic targets for colon cancer.

Background

In recent years, the morbidity and mortality of colon cancer increase rapidly, both of them have ranged fourth worldwide. Though surgical-based comprehensive treatments improve the prognosis of colon cancer, because of lacking available means for early diagnosis, the mortality still maintains a high level for the patients who suffer advanced stage cancer. The carcinogenesis and development of colon cancer is very complicated, its origin is the aberrant gene expression, and various factors which can change gene expressed level involve in procedure of the cancer. Hence, study the specific biomarkers and therapeutic targets is of great value in improving the prognosis of colon cancer.
Accumulating evidences have validated that epigenetic modification might promote the carcinogenesis and development of colon cancer via regulating various gene expression. DNA methylation is an important modification in the region of epigenetics, the oncogenes or antioncogenes exhibit irregulated expression if the gene regulatory regions exit abnormal DNA methylation. For example, the oncogene CCAAT/enhancer-binding protein-beta (C/EBP-β), long interspersed nuclear element-1 (LINE-1), F2RL3 and AHRR and undergoes over expression due to the hypomethylation in the promoter sites(1-4). Therefore, study the methylation regulated different expressed genes (DEGs) is necessary for understanding the mechanism of cancer initiation and development. However, studies on correlation between DNA methylation and the regulations of gene expression remain inadequate, and the potential value of different methylation regions (DMRs) correlate with DEGs on predicting prognosis in human colon cancer still requires to be studied in depth.

In last decades, with the development of next generation sequence technology and microarray platform, accumulating DEGs and epigenetic alterations such as DMRs have been revealed by bioinformatic analysis. For instance, Liu(5) et al validated that the DNA Methyltransferase Inhibitor Guadecitabine (SGI-110) altered the expression of oncogenes or antioncogenes by regulating DNA methylation; Qu et al(6) analyzed 57 AML patients with normal karyotype by using Illumina's methylation 450k BeadChip platform and showed that abnormal DNA methylation was altered significantly at enhancer regions and that the methylation levels at specific enhancers predict overall survival of AML patients. However, there is still a lack of integrated analysis of the gene expression regulated by DNA methylation in human colon cancer. Similarly, studies on DNA methylation in predicting the prognosis of patients in large cohorts are deficient.

In the current study, we downloaded RNA-seq data, DNA methylation data and clinical data of colon cancer from The Cancer Genome Atlas (TCGA, http://cancergenome.nih.gov) project(7). DEGs and DMRs were identified, meanwhile, DEGs regulated by DMRs were screened by analyzing their correlations. For further study on the function and value in predicting prognosis of the genes, GO functional enrichment and KEGG pathway enrichment analysis were performed. Moreover, the correlations among methylation status, gene expression level and over survival of colon cancer patients were analyzed, and risks analysis of prognosis-related DMRs was performed, to discuss the potential biomarkers of diagnosis and predicting prognosis for colon cancer patients.

Methods

Data and sources

The raw data and clinical information were downloaded from the TCGA-COAD project (https://cancergenome.nih.gov/) (7). This dataset includes 455 COAD (Colon adenocarcinoma) and 41 normal (non-tumor) samples. The COAD samples were randomly separated into two subsets with equal size, training dataset (228 tumor / 41 normal sample) and testing dataset (227 tumor / 0 normal sample).
Identification of differential expressed genes with altered methylation status

The training dataset included colon cancer and normal samples were utilized to identify the differential expressed genes (DEGs) with altered methylation status. Briefly, Deseq2(8) was applied to identify DEGs by comparing colon cancer and normal samples(8). The Benjamini-Hochberg method was used to adjust the p-value. A gene with FDR < 0.05 and |log2FC|>1 is identified as DEG. Then, differential methylated sites were identified in the colon cancer group compared with normal group by wilcoxon test. We retained differential methylation regions (DMRs) with p-value < 0.05(9). Furthermore, we identified hypermethylated genes and hypo-methylated genes based on the loci relative to the DMRs in the colon and normal samples.

GO functional enrichment and KEGG pathway enrichment analysis

To explore the function of DEGs regulated by methylation in the carcinogenesis and development of colon cancer, GO functional enrichment analysis was performed using DAVID database (https://david.ncifcrf.gov/) and three categories: cellular component (CC), molecular function (MF), and biological process (BP) were analyzed(10). In addition, KEGG pathways were also analyzed using DAVID database.

Screening for MDGE signatures and establishment of prognostic model

The associations between the expression level of each gene and the overall survival (OS) are evaluating by univariate cox regression analysis in our training dataset We retained the genes with p-value < 0.2. The remaining genes were further screened and confirmed by the multivariate cox regression analysis. Genes with p-value < 0.05 were selected as potential markes. Then, the survival analysis of the remaining genes was performed by Kaplan-Meier method with the Log-rank test; genes with log-rank p-value < 0.05 were retained. The prognosis risk score was defined as follows(11):

Risk score =

β is the regression coefficient of gene, which represents the contribution of gene to the prognostic risk score. Based on the risk score, patients can be assigned to a high-risk or low-risk group according to the median cutoff of the prognosis risk score. Then the Kaplan-Meier survival curves were calculated to compare survival and recurrence risk between the high and low-risk groups. The time-dependent receiver operating characteristic (ROC) curve analysis within 1 year, 3 years and 5 years were performed to evaluate the predictive accuracy and sensitivity of our prognostic model.

Association analysis of risk score and clinical features

The prognostic effect of various clinicopathological features including age, gender, tumor stage, were evaluated by univariate cox regression analysis and multivariate cox regression analysis. Then, the nomogram was constructed based on the results of the multivariate Cox regression analyses of risk score and clinicopathological features using rms package(12).
Statistics analysis

Data were analyzed using R package. Data were represented as mean ± standard deviation (S.D.). All tests were two sided, and P < 0.05 was considered statistically significant.

Results

Selection of DEGs and DMGs in colon cancer simples

We downloaded RNA-seq data from 455 colon cancer and 41 normal tissues, and screened 4,180 up-regulated and 7,821 down-regulated DEGs, the DEGs were shown as heatmap in Figure 1B. Meanwhile, we analyzed DMRs data from TCGA database, and identified 373 hypermethylated gens and 571 hypomethylated genes, the DMRs were shown as heatmap in Figure 1C. Next, we analyzed the intersection of down-regulated genes and hypermethylated genes, there were 95 genes met the condition (Figure 2A); similarly, there were 38 genes exited in the intersection of up-regulated genes and hypomethylated genes (Figure 2B). From Figure 2C, we found that there existed negative correlation between the expression levels of DEGs and the methylated levels. Therefore, a total of 133 genes met our requirements and were the candidate genes for further analysis.

Functional and pathway enrichment analysis for candidate genes

To understand potential biological function of the 133 candidate genes, GO functional and KEGG pathway enrichment analysis was performed for them. A total of 17 enriched GO terms in biological process (BP) and 6 terms in molecular function (MF) were identified (Figure 2D). From the results of GO analysis, we found that the candidate genes mainly enriched in cancer-associated functions, such as xenobiotic glucuronidation, cellular hormone metabolic and second-massager-mediate signaling. In addition, there were 10 pathways significantly enriched from the results of KEGG pathway enrichment analysis (Figure 2E). “Drug metabolism cytochrome P450”, “Complement and coagulation cascades” and “Chemical carcinogenesis” involved in the carcinogenesis and development of clone cancer based on the prevenient reports.

Identification of key methylated DEGs associated with poor prognosis

The univariate cox regression analysis confirmed 32 genes that significantly related with prognosis. Subsequently, the multivariate cox regression analysis focused on 7 genes, which were CDH4, CR2, KRT85, LGI4, NPAS4, RUVBL1 and SP140(Table 1). Next, we selected the 7 genes for further study. The expression levels of the 7 genes in cancer and normal tissues were shown in Figure 3A, the methylation levels of the 7 genes in cancer and normal tissues were shown in Figure 3B. However, from the results of Kaplan-Meier survival analysis, we found that the expression level of RUVBL1 significantly correlated with the overall survival rate of colon patients (Figure 3C).

Table 1 Cox regression analysis of clinical factors and hub genes
### Building a risk score to predict prognosis

To estimate the prediction power of the key methylated DEGs, a risk score was built. Using the median of the risk score as the cutoff point, in training cohort, the patients in high risk score group had a poorer overall survival than those patients in low risk group (Figure 4 A and B). Meanwhile, a ROC model was built, and we found that the AUC>0.5 and it met power of prognosis prediction (Figure 4C). Similarly, in validating cohort, the risk score model also achieved the prognosis prediction power (Figure 4D-F).

### Nomogram analysis for prognosis prediction

From the multivariate Cox regression analysis, we found that the pathology stage and RiskScore were the independent predicting factors for overall survival (Figure 5A). Finally, we construed a simple-to-use nomogram based on RiskScore and clinical characterization, such as, gender, age at diagnosis and pathology stage of colon cancer patients (Figure 5B). The nomogram provided some useful information in prediction of survival for the patients based on multivariate cox regression, and it suggested a good prediction.

### Discussion

Many reports validate that epigenetic regulation involves in the carcinogenesis and development of cancers, and DNA methylation which is the most common form of epigenetic modification plays an
important role in the regulation of gene expression. The alteration of DNA methylation in gene promoter region changes the expression level of gene. In general, the hypermethylation inhibits the gene expression, by contrary, hypomethylation promotes the gene expression (13-15). For colon cancer, DNA methylation alterations are found in many patients, for example, SST1 pericentromeric repeats existed hypomethylation, which resulted in the mutation of TP53, and the mutated TP53 associated with genome damage, which related to the tumorigenesis and development of colon cancer (16). Similarly, hypermethylation appeared in the promoter of antioncogenes SFPR1, SFPR2 and WIF1, leaded to downregulation expression of genes, inhibition of gene function, action of Wnt/β-catenin signal pathway and promotion of colon cancer (17); ADHFEI is also an antioncogene, the hypermethylation of ADHFEI promotes proliferation of the colon cancer cell via regulating cell cycle progression (18). The abnormal methylation of the genes above predicted poor prognosis of colon cancer. However, the systematic analysis of the correlation between DEGs and DMRs was still deficient. Therefore, to provide potential prognosis and target therapeutic targets, it is meaningful to study the gene expression regulated by DNA methylation.

Our study also validated that there existed a lot of DEGs and DMRs in colon cancer tissues (Figure 1B&C). Based on methylated regulation pattern, we focused on 133 genes, 95 of them were down regulated but hypermethylated, and 38 of them were up regulated but hypomethylated (Figure 2A). Similarly, negative regulated correction was found between DEGs and DMRs (Figure 2B). DAVID gene enrichment analysis and KEGG pathway enrichment analysis are useful for predicting the function and pathway of gene set. From the results of GO functional analysis, we found that the 133 DEGs regulated by DMRs played an important role in carcinogenesis and development of colon cancer. In biological process (BP) of GO enrichment analysis, as many as 9 terms related to metabiotic regulation. A lot of reports show that retinoic acid metabolic process associates with colonic tumorigenesis and metastasis, such as retinoic acid metabolizing enzymes CYP26B1, LRAT and CYP26A1 over expressed in colorectal cancer tissues and that LRAT and CYP26B1 significantly associated with the prognosis of the colorectal cancers (19). Xenobiotic metabolic process also involves in the carcinogenesis and development of cancer, because colonic epithelium is exposed to various compounds from diet, and the compounds can be metabolized to the procarcinogens, which are the risks of colonic cancer (20, 21). Similarly, in cellular companion (CC) module, there were 4 terms relative to molecular binding. Some reports found that retinoid acid binding receptor inhibited cancer cell apoptosis by regulating miR-22/NUR77 axis (21, 22). Interestingly, from the results of KEGG pathway enrichment analysis, we found that the genes mainly enriched in molecular metabolism, being consistent with BP of GO enrichment, and the drug, retinoid and porphyrin metabolism are proved involving in irregulated tumor cell metabolism, and resulted in chemotherapy resistant (23).

To screen the genes which associate with the prognosis of colon cancer patients, univariate and multivariate cox regression analysis was performed to test the independent significance of different factors. We found that 7 genes from the candidate gens significantly related to the overall survival of colon cancer patients. Meanwhile, the expression levels of the 7 genes were negatively related to their methylated levels, which proved the expression of the genes were regulated by methylation (Figure 4A). From previous reports, we find the 7 genes involve in carcinogenesis and development of various cancers.
For example, in glioblastoma, CDH4 plays a role of oncogene through proliferating and infiltrating the brain parenchyma, which resulted in highly impaired(24); similarly, CDH4 acts as an oncogene role through initiating and maintaining epigenetic suppression of multiple tumor suppressor genes in colorectal cancer(25); in human basal carcinoma, the abnormal expression of KRT85 resulted in cancer stem cell exhaustion(26); LGI4 interacts with ADAM promotes proliferation and differentiation of neuronal precursors and adipocytes; in breast cancer, the interactions between LGI4 and ADAM also relate to carcinogenesis(27). However, though the abnormal expression of the 7 gens could be seemed as the independent risk factors of over survival for colon cancer patients, the expression levels of the gens didn't relate to over survival rate of patients, excepted for RUVBL1 (Figure 4B). RUVBL1 is an oncogene that relate to the prognosis of patients in various cancers, and it promotes carcinogenesis through regulating Wnt/β-catein signal pathway(28-30). Therefore, RUVBL1 has the potential value of prognosis and therapeutic target for cancer patients.

Irregulated DNA methylation is an important factor for carcinogenesis, and it can be used for diagnosis and predicting prognosis of cancer patients, because DNA methylation profiles are tissue or cell specific. In our study, we built a risk score to estimate the power of the 7 hub genes in predicting the prognosis of colon patients. We found that the 7 hub genes had a good performance in prognosis prediction, which illustrated the aberrant methylation in cancer tissue could be used to diagnosis and therapeutic targets. Next, our nomogram showed that Riskscore was a good predictor for 1-, 3- or 5-year overall survival in colon cancer patients, it visualized the correction between Riskscore and clinical features, including age and tumor stage. While, these results should be validated by further studies.

Finally, there are some limitations in this study should be noted. First, our RNA sequencing, DNA methylation and clinical data were obtained from TCGA database, but no clinical samples were used for validating the results. Therefore, it is necessary to select large clinical samples for testing the effectiveness of the biomarkers associated with the hub genes above. Second, we screened the biomarkers through using the method of statistical and bioinformatic analysis, but not biological experiment. So, the mechanisms of the biomarkers are still unknown, and it need some biological experiments to understand the roles of the candidate markers in carcinogenesis and development of colon cancer. Third, there are potential bias in this study, because we didn't analyze some important clinical information, especially for treatment factors (such as operation, chemotherapy and radiotherapy). Therefore, prospective studies and multicenter clinical trials are necessary for further validation.

Conclusion

In conclusion, the DEGs regulated by methylation play an important role in carcinogenesis of colon cancer, and the hub gens regulated by DNA methylation are potential tool for predicting prognosis of colon cancer patients.

Declarations
Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

Not applicable.

Competing interests

All of the authors declare that there is no conflict of interest.

Funding

This study is supported by grants from the Fundamental Research Funds for the Central Universities, South China University of Technology (No. 2018MS21).

Authors' contributions

XWP conceived and designed the analyses; DD analyzed the data; WTOu and YKX analyzed the data; YQL revised the paper.

Acknowledgements

Not applicable.

References


Figures

Figure 1
Figure 2
Figure 3
Figure 4
Figure 5

A

Risk Score

log(HR) (95% CI)

Gender (male vs. female)
Age (≥ 60 vs. < 60)
Tumor stage (III+VI vs. I+II)

B

Points

Risk Score

Tumor stage (III+VI vs. I+II)
Age (≥ 60 vs. < 60)

Total Points

Linear Predictor

1 year survival

3 year survival

5 year survival