

Original Research Article

DOI: 10.26479/2020.0603.05

IDENTIFICATION OF DRIVER GENE IMPORTANT FOR PROGRESSION FROM COLONIC INFLAMMATION TO COLORECTAL CANCER

A Chauhan, S Kalra*

Department of Biochemistry, School of Sciences,
Indira Gandhi National Open University, New Delhi, India.

ABSTRACT: The inflammation and tumorigenesis have been connected evidently from many studies in the past. Inflammation is understood to be an important driver of tumorigenesis in colorectal cancer (CRC). Our study is directed towards finding the driver genes which are differentially expressed in inflammation but can act as a driver for the transformation of colonic inflammation to CRC. We have identified a hub gene from the network of driver genes which links inflammatory and colorectal cancer pathway. NFATc1 is identified as an important gene which may interact with other gene leading to CRC. Survival analysis and copy number variation (CNV) analysis of identified hub gene has been done.

Keywords: colonic inflammation, CRC, differential expression, driver gene, hub gene, NFATc1, CNV.

Article History: Received: April 30, 2020; Revised: May 13, 2020; Accepted: May 29, 2020.

Corresponding Author: Dr. S Kalra* Ph.D.

Department of Biochemistry, School of Sciences,
Indira Gandhi National Open University, New Delhi, India.

1. INTRODUCTION

The connection between inflammation and tumorigenesis in the last decade has received a great deal of supporting evidence from genetic, pharmacological, and epidemiological data. Inflammation as a hallmark of cancer may arise from a variety of factors, including infection, environmental carcinogens (tobacco, alcohol, radiation), cellular senescence, and obesity [1,2]. Molecular pathobiology of colorectal cancer (CRC) implicates pro-inflammatory conditions to promote the tumor progression, invasion, and metastasis. It is well known that patients with inflammatory

bowel disease are at higher risk of CRC. Many studies exist reiterating the link between inflammation and CRC. Inflammation involves interaction between various immune cells, inflammatory cells, chemokines, cytokines, and pro-inflammatory mediators [3,4]. Mounting evidence suggests that treatment with anti-inflammatory medications may prevent or delay the development of CRC in hereditary and sporadic cases [5], underscoring a potentially ubiquitous role of inflammation in the development and progression of all types of CRC. The present study is expected to support the idea of inflammation as a promoter for CRC progression. Colonic inflammation genes have been identified which may act as driver genes for CRC progression with NFATc1 as hub gene. In this study, we selected the microarray datasets GSE11223 [6] from the GEO database to identify differentially expressed genes (DEGs). Kyoto encyclopedia of genes and genomes (KEGG) analysis was used to investigate DEGs. Common genes from DEGs and driver genes retrieved from the DriverDb3 database [7] were selected. A hub gene was identified by constructing a protein-protein interaction (PPI) network. Furthermore, the survival analysis was used on the gene expression profiling interactive analysis (GEPIA) [8]. Copy number variation analysis was done using the DriverDb3 database.

2. MATERIALS AND METHODS

Data acquisition and pre-processing

The differential gene expression between colonic inflammation and normal samples, GDS3268 microarray dataset was downloaded from the GEO website (<http://www.ncbi.nlm.nih.gov/geo/>). These RNA profiles were provided on platform GPL1708 (Agilent-012391 Whole Human Genome Oligo Microarray G4112A). Data normalization and variance filtration were done using the Multi Experiment Viewer, MeV 4.9 [9]. The variance filter which allows removal of genes with low variation of expression over the loaded samples was used. This filter is used to remove 'flat genes' which don't vary much in expression over the conditions of the experiment. Expression values were log₂ transformed. Flowchart of methodology used in the study is given in Figure 1.

Identification of DEGs

To identify the differentially expressed genes between colonic inflammation and normal tissue samples, the normalized and log₂ transformed values of genes were given to k-means clustering using the Pearson correlation coefficient method, followed by statistical t-test analysis between inflammation and normal tissue samples. Settings $p < 0.1$ and fold change > 2 were used to identify the differentially expressed genes (DEGs) for colonic inflammation.

Identification of DEGs as driver genes for CRC progression

DriverDb3, a database that incorporates 6079 cases of exome-seq data, annotation databases and published bioinformatics algorithms dedicated to driver gene/mutation identification, was introduced to recruit driver genes for colon adenocarcinoma. These genes and the genes which are differentially expressed in colonic inflammation were compared. Those common in both the lists

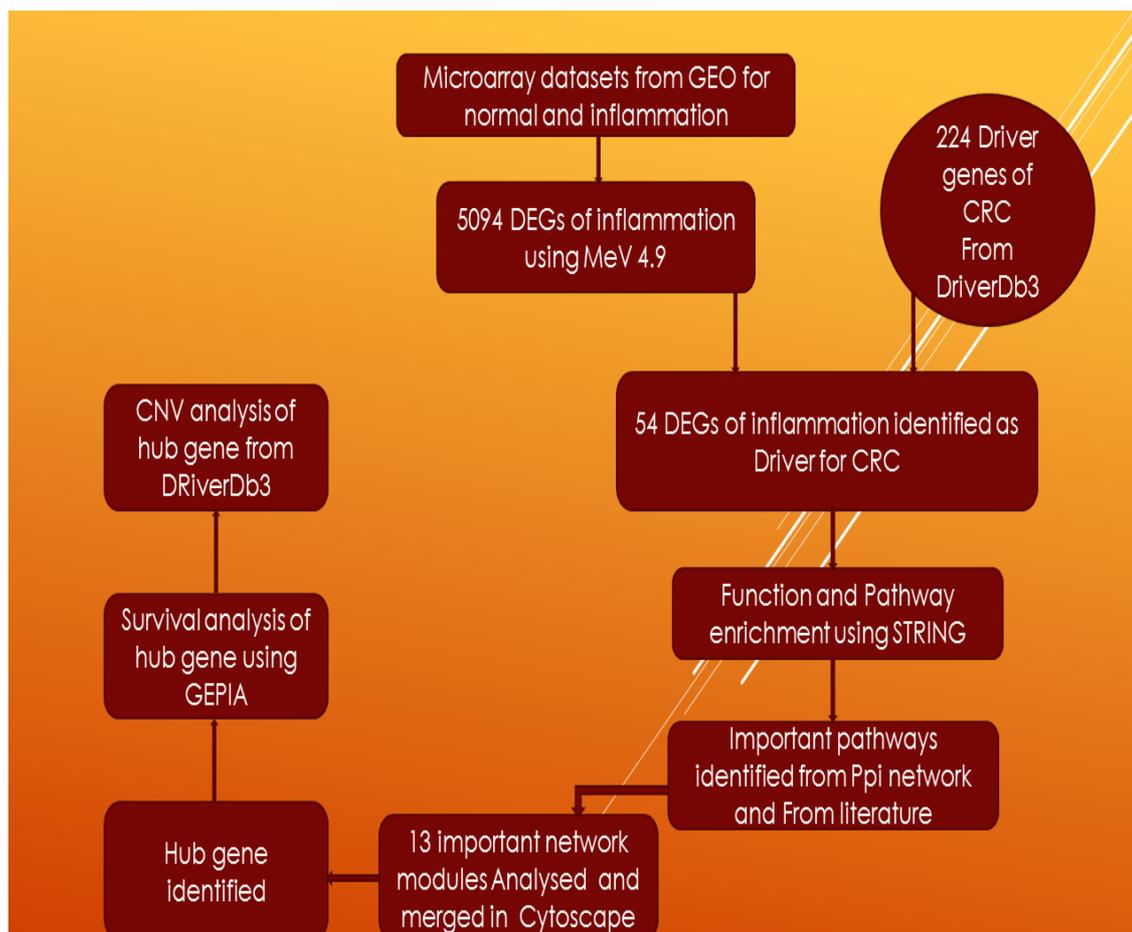


Figure 1. Flowchart of methodology

were identified as potential driver genes from colonic inflammation to CRC progression.

Function and pathway enrichment analyses of DEGs

Identified driver genes were uploaded to Search Tool for the Retrieval of Interacting Genes (STRING) database to construct protein-protein interaction (PPI) [10]. Pathway enrichment analysis using Kyoto Encyclopedia of Genes and Genomes (KEGG) was done. KEGG is a database resource for understanding high-level functions and biological systems from large-scale molecular datasets generated by high-throughput experimental technologies. p value < 0.05 is considered to be statistically significant. All the modules were analyzed in Cytoscape; an open-source bioinformatics software platform for visualizing molecular interaction networks [11]. The interactions with a combined score > 0.4 were considered significant and extracted for constructing PPI networks.

Analysis and validation of hub genes

A network of genes from STRING pathway analysis was used to find out the hub genes. Cytoscape was used to find out the hub genes by merging all modules. To confirm the reliability of hub genes from our detection, we analyzed their prognostic and expression in CRC using Gene Expression Profiling Interactive Analysis (GEPIA), an interactive web application tool for gene expression analysis, containing 8,587 normal samples and 9,736 tumors samples from the Genotype- Tissue

Expression databases and TCGA databases [12]. The box plot was performed to visualize the relationships.

Copy number variation analysis

Copy number variation (CNV) analysis of the selected hub gene was done to see the loss or gain in CRC using DriverDb3. It uses two CNV tools: iGC and DIGGIT [13] to analyze the gene. Color codes are used to identify the CNV in the specific cancer types and whether it is significant.

3. RESULTS AND DISCUSSION

Identification of DEGs

Analysis with the MeV 4.9 showed that a total of 5,059 DEGs were expressed in colonic inflammation samples. Adjusted p value <0.01 and $|\log_2FC| >2$ were set as the cut-off criteria to select the significant DEGs. The volcano plot of DEGs of CRC in each dataset was presented in Figure 2.

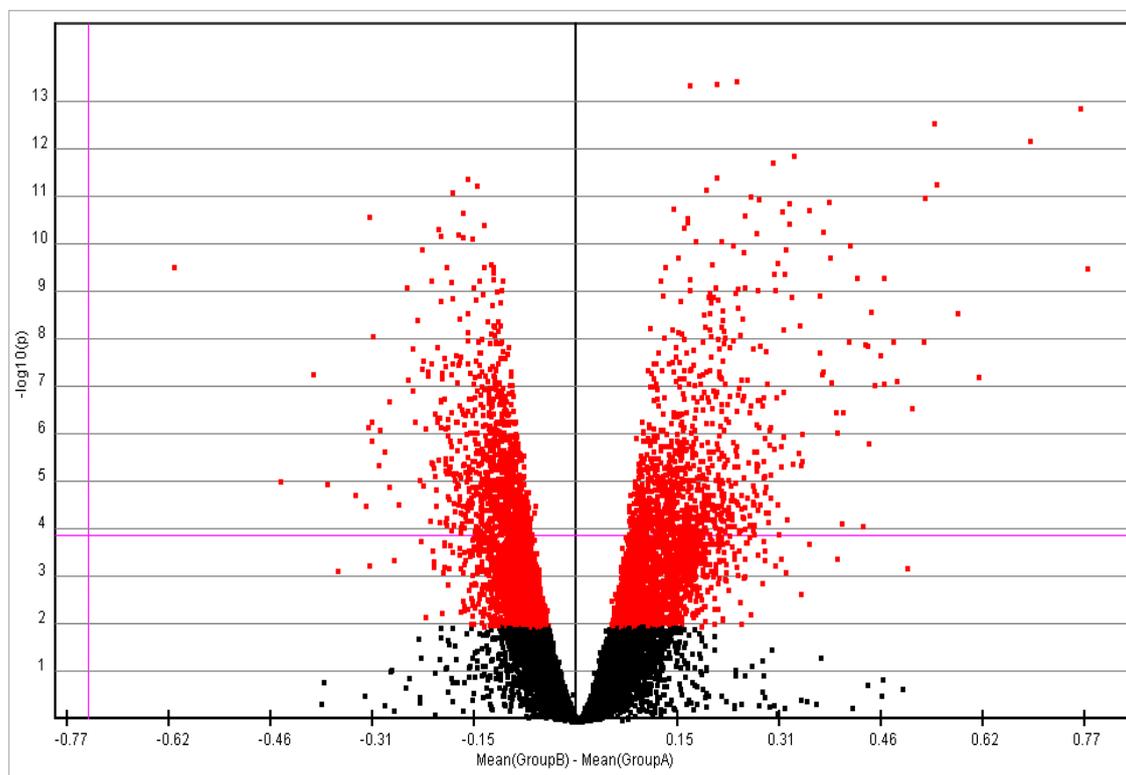


Figure 2. Volcano plot of gene expression profile data of inflamed and normal tissues in dataset. Red dots toward right: significantly upregulated genes in inflammation; red dots towards left: significantly downregulated genes in inflammation; Black dots: no differentially expressed genes. Adj. $p < 0.01$ and $|\log_2FC| > 2$ were considered as significant.

Identification of DEGs as driver genes for CRC progression

204 driver genes for colon adenocarcinoma were retrieved from the DriverDb3 database and compared with 5059 DEGs in colonic inflammation. 54 inflammation genes were identified which may act as driver genes for CRC progression (Figure 3).

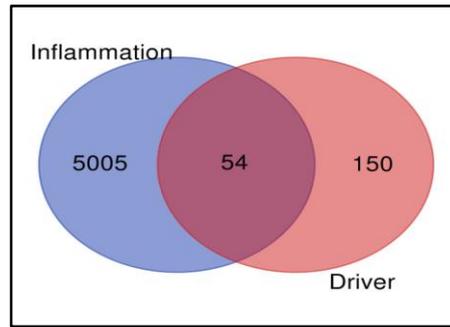


Figure 3. Venn diagram of overlapping 54 inflammation DEGs and driver genes.

Enrichment analysis

To explore the biological functions of identified 54 DEGs, we performed functional and pathway enrichment analyses using STRING. The results of KEGG pathway analysis revealed that these DEGs were mainly enriched in Erb signaling, Wnt signaling, apoptosis, platelet activation, mapk signaling, p13k-Akt signaling and p53 pathways (Figure 4).

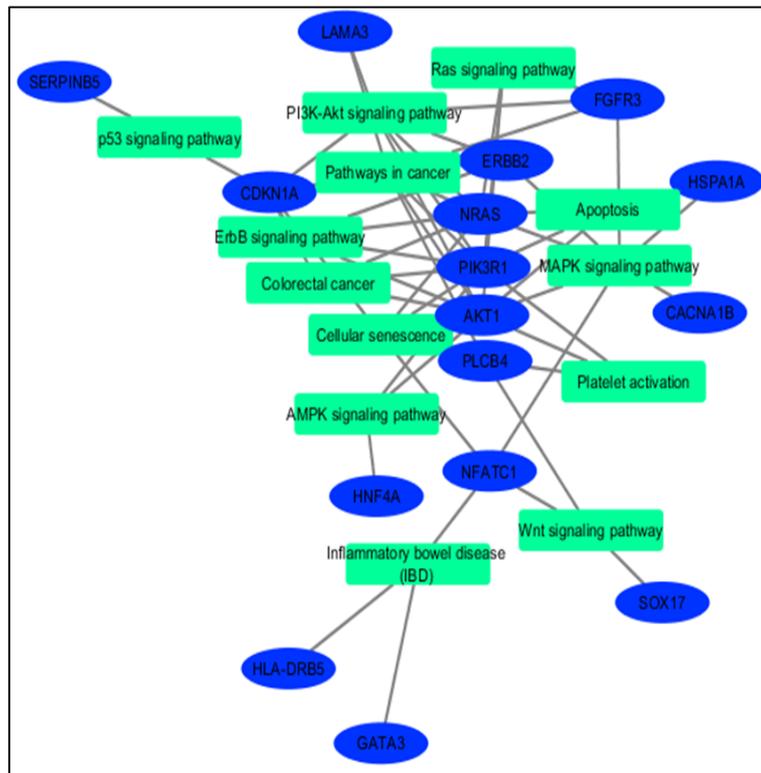


Figure 4. Thirteen most significant modules extracted from PPI network and KEGG pathway analysis of module genes

Role of these pathways in colonic inflammation and colorectal is corroborated by many studies. The study on platelet activation indicates that many features of the altered platelet function detected in human IBD can be reproduced in animal models of colonic inflammation [14]. P13k-Akt signaling and Wnt signaling have role in colonic inflammation. Studies have shown that KYNA reduces PI3K/Akt and p38, ERK1/2 cellular kinases in the HT-29 cell line, and inhibits the Wnt pathway [15]. It has been observed that chronic inflammation precedes colitis-associated tumor development

which is a subtype of colorectal cancer. Chronic inflammation causes oxidative damage to DNA, leading to the p53 mutations observed in tumor cells and the inflamed, but non dysplastic epithelium[16,17] stating the role of the P53 signaling pathway in inflammation.

Identification of hub genes

Thirteen gene modules having a role in important pathways were selected based upon the literature and merged using Cytoscape to find out the hub genes in the network. Nuclear Factor of Activated TCells 1 (NFATc1) gene was identified as a hub gene (Figure 5).

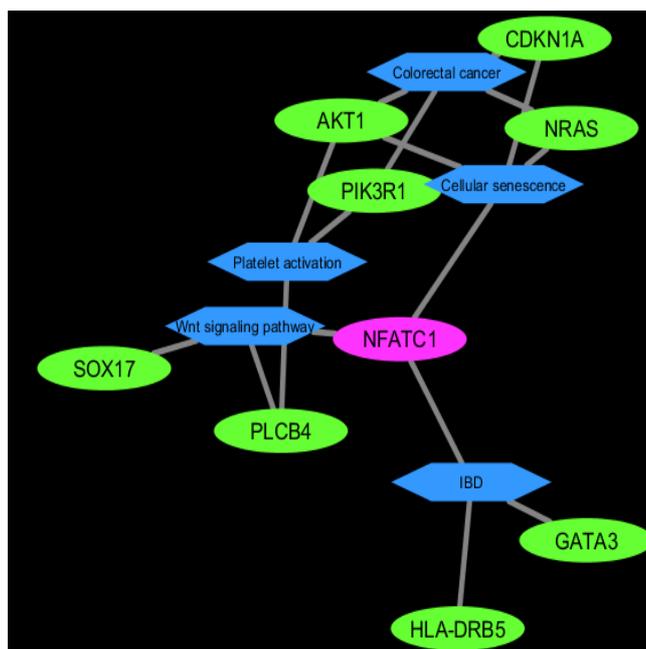


Figure 5. Interactions of hub gene;NFATc1 in inflammation and cancer.

Nuclear Factor of Activated T Cells 1 (NFATc1) gene was identified as a hub gene as it was connecting inflammatory bowel disease network module with Wnt signaling, cellular senescence module. It was also connecting to CRC through NRAS, and to the platelet activation pathway through Wnt pathway. There are studies having pieces of evidence that mesalamine, a known drug approved by the FDA used to treat inflammatory bowel disease directly acts on the Wnt pathway [18,19]. As a hub gene if the NFATc1 gene is removed from the network it could collapse the whole network. NFATc1 has been identified as a promoter for invasion of cancer [20].

NFATc1 has shown a role in CRC. There have been studies showing evidence that NFATc1 regulates colon cancer cell behavior and its transcriptional targets constitute a novel, biologically-anchored gene expression signature for the identification of colon cancers with a high risk of metastatic recurrence [21]. Singh et al [22] demonstrated in their study on pancreatic and colon cancer cell lines that NFAT transcription factors are important mediators of TGF β switch from suppressor to promoter of cancer. These studies support our inference that NFATc1 is an important gene for the transition of inflammation to CRC.

Analysis and validation of hub gene

The mining of the GEPIA database demonstrated that the identified hub gene exhibited significant differences in expression between inflammation and normal tissues. This result further confirmed that the expression level of this hub gene is closely correlated with the onset of CRC. It was found that the hub gene NFATc1 was down regulated (Figure 6).

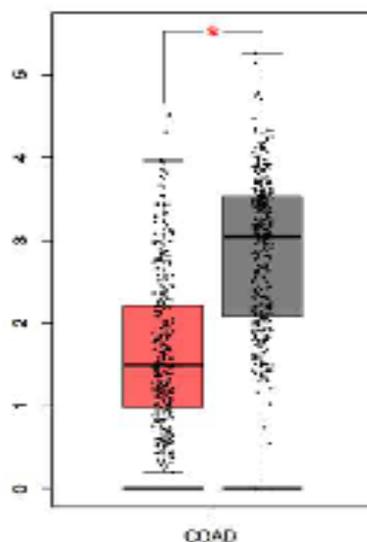


Figure 6. The expression level and prognostic value of hub genes according to the GEPIA database. Expression of NFATc1 is decreased significantly in colorectal adenocarcinoma (COAD) (red) as compared to normal (black).

Copy number variation analysis of NFATc1

CNV analysis of NFATc1 was done to see its role in the development and progression of CRC through the alterations of their gene expression levels on colorectal adenocarcinoma (Figure 7). It showed the loss of CNV.

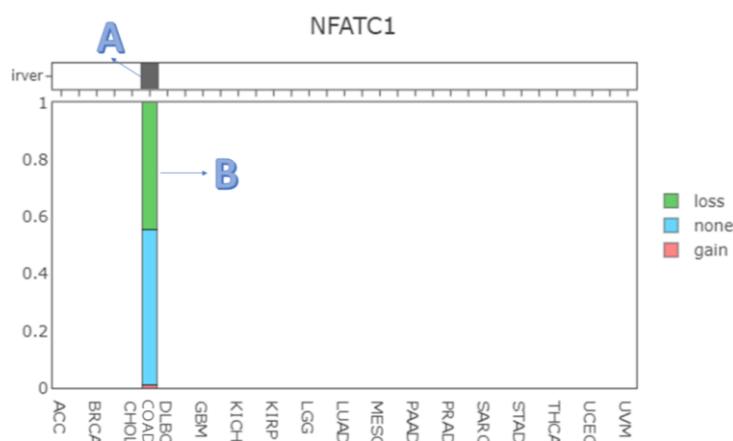


Figure 7. This graph shows the copy number variation (CNV) of a user-selected gene in colorectal adenocarcinoma (COAD). Gene is analyzed by two CNV tools: iGC and DIGGIT. On top of the bar chart, CNV Driver (A) presents the number of tools that identified the CNV in the specific cancer types and whether it is significant. CNV driver in dark grey implied that it is identified by both

tools. (B) The green color represents loss of copy number; red is gain copy number; blue represents no copy number change.

These results further confirmed that the expression level of the hub genes was closely correlated with the onset of CRC. Cancer is a heterogeneous disease with many genetic variations. Lines of evidence have shown CNVs of certain genes are involved in the development and progression of many cancers through the alterations of their gene expression levels on the individual or several cancer types [23]. Our study indicated that NFATc1 was down-regulated in CRC due to loss of CNV. From our study, the role of NFATc1 in the progression and development of CRC looks quite evident.

4. CONCLUSION

Our study found key genes involved in inflammation and potential genes to drive cancer from inflammation to CRC stages, which may provide novel insights and potential biomarkers for prognosis, as well as providing new CRC treatments. Further studies targeting this hub gene to see its role in the transformation of inflammation into CRC are needed.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

HUMAN AND ANIMAL RIGHTS

No Animals/Humans were used for studies that are base of this research.

CONSENT FOR PUBLICATION

Not applicable.

AVAILABILITY OF DATA AND MATERIALS

The authors confirm that the data supporting the findings of this research are available within the article.

FUNDING

None.

ACKNOWLEDGEMENT

None

CONFLICT OF INTEREST

Authors have no conflict of interest.

REFERENCES

1. Clevers H. At the crossroads of inflammation and cancer. *Cell* 2004;118:671–674.
2. Askling J, Dickman PW, Karlen P, et al. Family history as a riskfactor for colorectal cancer in inflammatory bowel disease. *Gastroenterology* 2001;120:1356–1362.
3. Atreya I, Neurath MF. Immune cells in colorectal cancer: prognostic relevance and therapeutic strategies. *Expert Rev Anticancer Ther* 2008;8:561–572.
4. Waldner MJ, Neurath MF. Cytokines in colitis associated cancer: potential drug targets? *Inflamm Allergy Drug Targets* 2008;7:187–194.

5. Friis S, Riis AH, Erichsen R, Baron JA, Sorensen HT. Low-dose aspirin or nonsteroidal anti-inflammatory drug use and colorectal cancer risk: A population-based, case-control study. *Ann. Intern. Med.* 2015; 163: 347–55.
6. Noble CL, Abbas AR, Cornelius J, Lees CW et al. Regional variation in gene expression in the healthy colon is dysregulated in ulcerative colitis. *Gut* 2008;57(10):1398-405.
7. Cheng W, Chung I, Chen C, Sun H, Fen J, Tang W et al. DriverDB: an exome sequencing database for cancer driver gene identification. *Nucleic Acids Research*.2013; 42: D1048-D1054.
8. Tang, Z., Li, C., Kang, B., Gao, G., Li, C., & Zhang, Z. GEPIA: A web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Research* 2017, 45(W1), W98–102.
9. Saeed AI, Sharov V, White J, Li J, Liang W, BhagabatiN, et al. TM4: a free, open-source system for microarray data management and analysis. *Biotechniques*. 2003; 34(2):374-378.
10. Snel B., Lehmann G., Bork P., Huynen M.A.STRING: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene. *Nucleic Acids Res.* 2000; 28:3442–3444.
11. Morris J, Lotia S, Wu A, Doncheva N, Albrecht M, Pico A et al. sets App for Cytoscape: Set operations for Cytoscape Nodes and Edges. *F1000Research*. 2015;3:149-158.
12. Cerami, E., Gao, J., Dogrusoz, U., Gross, B. E., Sumer, S. O., Aksoy, B.A., & Schultz, N. (2012). The cBio cancer genomics portal: An openplatform for exploring multidimensional cancer genomics data.*Cancer Discovery*, 2(5), 401–404.
13. Driver-gene inference by genetic-genomics and information theory (DIGGIT) algorithm to identify disease driver mutations. *Science-Business eXchange*. 2014;7(43):1271-1271.
14. Yan S, Russell J, Harris N, Senchenkova E, Yildirim A, Granger D. Platelet Abnormalities during Colonic Inflammation. *Inflammatory Bowel Diseases*. 2013;19(6):1245-1253.
15. M. Freewan, M. D. Rees, T. S. Plaza, E. Glaros, Y. J. Lim,X. S. Wang, et al., Human Indole amine 2,3-dioxygenase isa catalyst of physiological hemeperoxidase reactions-implications for the inhibition of dioxygenase activity by hydrogen peroxide, *J. Biol. Chem.*, 2013, 288, 1548–1567.
16. Kraus S, Arber N. Inflammation and colorectal cancer. *Curr Opin Pharmacol*2009;9:405–410.
17. Choi J, Yoon SH, Kim JE, Rhee KH, Youn HS, Chung MH. Gene-specific oxidative DNA damage in *Helicobacter pylori*-infected human gastric mucosa. *Int J Cancer* 2002;99:485–490.
18. Bajpai M, Seril D, Van Gurp J, Geng X, Alvarez J, Minacapelli C et al. Effect of Long-Term Mesalamine Therapy on Cancer-Associated Gene Expression in Colonic Mucosa of Patients with Ulcerative Colitis. *Digestive Diseases and Sciences*. 2018;64(3):740-750.
19. Stolfi C, Pellegrini R, Franzè E, Pallone F, Monteleone G. Molecular basis of the potential of mesalazine to prevent colorectal cancer. *World Journal of Gastroenterology*. 2008;14(28):4434.
20. Oikawa T, Nakamura A, Onishi N, Yamada T, Matsuo K, Saya H. Acquired Expression of

- NFATc1 Downregulates E-Cadherin and Promotes Cancer Cell Invasion. *Cancer Research*. 2013;73(16):5100-5109.
21. Tripathi M, Deane N, Zhu J, An H, Mima S, Wang X et al. Nuclear Factor of Activated T-cell Activity Is Associated with Metastatic Capacity in Colon Cancer. *Cancer Research*. 2014; 74(23):6947-6957.
22. Singh G, Singh SK, Konig A, Reutlinger K, Nye MD, Adhikary T, Eilers M, Gress TM, Fernandez-Zapico ME, Ellenrieder V. Sequential activation of NFAT and c-Myc transcription factors mediates the TGF-beta switch from a suppressor to a promoter of cancer cell proliferation. *J Biol Chem*. 2010; 285:27241–27250.
23. Shao X, Lv N, Liao J, Long J, Xue R, Ai N et al. Copy number variation is highly correlated with differential gene expression: a pan-cancer study. *BMC Medical Genetics*. 2019 ;20:175.