

Identifying functional evolution processes according to the pathological stages of colorectal cancer

Bolin Chen^{†§}, Manting Yang[†], Li Gao[†] and Xuequn Shang^{†§*}

[†] School of Computer Science, Northwestern Polytechnical University, Xi'an, China

[§]Key Laboratory of Big Data Storage and Management, Ministry of Industry and Information Technology, Northwestern Polytechnical University, Xi'an, China

*Corresponding email: npu_bioinf@hotmail.com

Abstract—Colorectal cancer (CRC) is one of the malignant tumors with high morbidity and mortality. A prevalent method for studying colorectal cancer is to identify differentially expressed genes (DEGs) between control and patient samples, followed by the pathway enrichment analyses. However, many of those studies ignore the fact that different pathological stages of the cancer are often highly different from each other. The mixture of those heterogeneous samples may lack the efficiency of identifying the real DEGs, and loss the opportunity to analyze the dynamic evolution process of cancer. In this study, we develop a feasible framework to identify function evolution processes of cancers according to their pathological stages. Firstly, the limma package was used to identify DEG sets between control and CRC stage I, II, III, and IV samples, separately. Secondly, a pathway interaction network was constructed by taking a comprehensive analysis of pathways at individual stages, and a functional module interaction network was also generated independently by clustering genes into modules in a PPI network. The relationship between pathways and modules in adjacent stages was analyzed for all stages of CRC. A total of 479, 313, 349, and 383 DEGs were identified and they were enriched in 17, 16, 20, and 24 pathways, respectively. A functional evolution network was constructed by using those pathways, and two significant evolution processes a2-b1-c2-d1 (Mod1) and a1-b2-c1-d2 (Mod2) were identified which may play critical roles in the development of CRC. The framework proposed in this study can be used to explore molecular mechanisms and evolution processes of CRC.

Index Terms—colorectal cancer, differentially expressed gene, module network, functional evolution process

I. INTRODUCTION

Colorectal cancer (CRC), with high morbidity and high mortality, is one of the malignant tumors. The incidence and mortality of CRC are among the top four cancers. It is predicted that 145,600 patients will be diagnosed with CRC and 51,020 patients will die of this cancer in 2019 [1]. Therefore, exploring evolution processes of CRC is conducive to study the molecular mechanism of CRC, and also promote the development of prognostic markers and therapeutic methods.

The occurrence of CRC is closely related to the mutation of the cell signaling pathway. In recent years, several cell signal transduction pathways related to CRC carcinogenesis have been discovered, including Wnt/-catenin [2], Hedgehog [3], p53 [4] signaling pathways and so on. The signaling pathways are of great importance for studying the molecular mechanisms of CRC development. However, the functional

evolution processes underlying the pathological stages of CRC remain unclear.

In this study, we propose a new approach to study the functional evolution processes of the CRC development. We constructed a pathway network and a module network through pathway enrichment analysis and PPI network modules analysis to explore the critical signal pathways related to CRC development, which was contributed to analyse functional evolution processes underlying the pathological stages of CRC.

II. MATERIALS AND METHODS

A. Data Sources and Data Processing

The gene expression dataset was obtained from Gene Expression Omnibus (ID: GSE62932). It includes 4 healthy controls and 64 CRC samples. The number of samples of CRC stage I, II, III, and IV is 12, 17, 20, and 15. The raw microarray dataset was preprocessed through the R language affy package (<https://www.bioconductor.org/>). Then it was normalized with the Robust Multichip Average (RMA) method [5]. The probe data was then mapped to corresponding gene symbols using the annotation information package in GPL570, hgu133plus2.db [6].

B. Differential Expression Analysis

Differentially expressed genes (DEGs) between control samples and CRC stage I, II, III, and IV samples were screened respectively by using the limma [7] algorithm. The Benjamini-Hochberg [8] method was employed to correct the p -value to acquire FDR. Calculating \log_2FC of each gene between two groups. Only the genes that meets $FDR < 0.01$ and $|\log_2FC| \geq 1.5$ were remained as DEGs.

C. Pathway Enrichment Analysis and Pathway Interaction Network of all Stages

The various signal pathways in cells are not isolated, but rather interact and correlate to form a complex signal network system [9]. The ClueGO [10] was used to conduct pathway enrichment analysis by using the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways. The relationship of pathways were analyzed by Kappa Score (criteria: p -value ≤ 0.05 , Kappa Score threshold=0.4), and a pathway interaction network was constructed and visualized by using Cytoscape [11].

D. Module Relation Network between Adjacent Stages

STRING (Search Tool for the Retrieval of Interacting Genes) [12] database was utilized to obtain the PPIs between DEGs identified above. Four PPI sub-networks were constructed by selecting PPIs scores larger than 0.4, and those sub-networks were then visualized in Cytoscape [11].

Identifying functional modules in a PPI network is important for understanding the functional evolution processes and complex molecular mechanisms of cancer. ClusterONE [13] was employed to identify functional modules (criteria: Min size=3, overlap=0.8, similarity=coefficient) in four PPI networks, respectively. Modules of different stages of the PPI sub-networks may share the same DEGs, which makes it possible to calculate overlapping score between modules of adjacent stages. A module relationship network was then generated and visualized in Cytoscape [11].

III. RESULTS

A. Differentially Expressed Genes

Based on the gene expression profile dataset GSE62932, a total of 479, 313, 349, and 383 DEGs with $FDR < 0.01$ and $|\log_2 FC| \geq 1.5$ were identified between control samples and CRC stage I, II, III, and IV, respectively. The corresponding DEG sets were defined as DEG-stage1, DEG-stage2, DEG-stage3, and DEG-stage4.

B. Pathway Enrichment Analysis and Pathway Interaction Network of all Stages

Four DEG sets were significantly enriched in 17, 16, 16, and 24 pathways. The total number of non-overlapped pathways was 35 pathways. In order to explore the development process of various stages of CRC, a pathway interaction network was constructed by taking the those pathways as vertices and their functional relationship as edges. The details of this network was illustrated in Fig 1. The edge between pathways was evaluated according to Kappa Scores, where the value of those scores were indicated by the width of lines.

C. Module Relation Network between Adjacent Stages

Four PPI networks was established based on PPIs of four DEG sets respectively. A total of 75, 53, 57, and 67 modules were identified among four PPI networks obtained above, respectively. The number of overlapped genes between modules of adjacent stages serves as the weights of the edges.

A functional evolution network was constructed to find evolution process which are closely related to CRC stage. There were two significant evolution paths in the module network (see Fig 2 for details): a2-b1-c2-d1(Mod1) and a1-b2-c1-d2(Mod2). Mod1 and Mod2 were enriched in 10 and 7 pathways (criteria: $p - value \leq 0.05$), respectively (details shown in Table I and Table II). A total 17 pathways were identified, which were dynamically changed with pathological stages of CRC.

TABLE I
KEGG PATHWAYS OF A2-B1-C2-D1 (MOD1)

ID	Term	<i>p-value</i>	Count
KEGG:04080	Neuroactive ligand-receptor interaction	4.19E-16	18
KEGG:04672	Intestinal immune network for IgA production	1.35E-03	3
KEGG:04060	Cytokine-cytokine receptor interaction	2.62E-09	12
KEGG:04062	Chemokine signaling pathway	1.63E-11	12
KEGG:04923	Regulation of lipolysis in adipocytes	1.89E-03	3
KEGG:04924	Renin secretion	2.46E-04	4
KEGG:04657	IL-17 signaling pathway	7.68E-04	4
KEGG:05132	Salmonella infection	5.71E-04	4
KEGG:05134	Legionellosis	1.01E-04	4
KEGG:05146	Amoebiasis	8.66E-04	4

TABLE II
KEGG PATHWAYS OF A1-B2-C1-D2 (MOD2)

ID	Term	<i>p-value</i>	Count
KEGG:03030	DNA replication	8.21E-04	3
KEGG:03460	Fanconi anemia pathway	2.67E-03	3
KEGG:04110	Cell cycle	1.40E-17	15
KEGG:04115	p53 signaling pathway	3.14E-05	5
KEGG:04218	Cellular senescence	1.25E-06	8
KEGG:04114	Oocyte meiosis	1.87E-07	8
KEGG:04914	Progesterone-mediated oocyte maturation	3.00E-08	8

IV. DISCUSSIONS

In this study, four DEG sets were screened between gene expression profiles of normal and four stages of CRC samples, respectively. A variety of signal pathways related to CRC were identified based on the pathway network and the module network through pathway enrichment analysis and PPI network modules analysis.

There are 8 pathways connected closely each other in Fig 1. Several studies have shown that dietary fats may potentially lead to the development of cancer by regulating the expression of genes in cells to alter metabolism, proliferation and differentiation of cells [14]. Therefore, fatty acid metabolism plays a promoting role in cancer development [15]. Numerous highly polymorphic genes in Cytochrome P450 may affect metabolic ability, which may lead to a different susceptibility to CRC [16]. Some researches shows that Cytochrome P450 was significantly related to CRC development [17]. We considered that closely related metabolic pathways in the network may play an important role in the development of CRC, which contributed to further research on CRC related pathways.

KEGG pathways of Mod1 and Mod2 have great significance to the analysis of the functional evolution process of cancer at different stages. Chronic intestinal inflammation is one of the important risk factors of CRC. Tumors may generate once the regulation of IgA-secreting cells is imbalanced [18]. Lack of IgA in tumor sites may cause increased inflammation and promote tumor progression, because IgA may

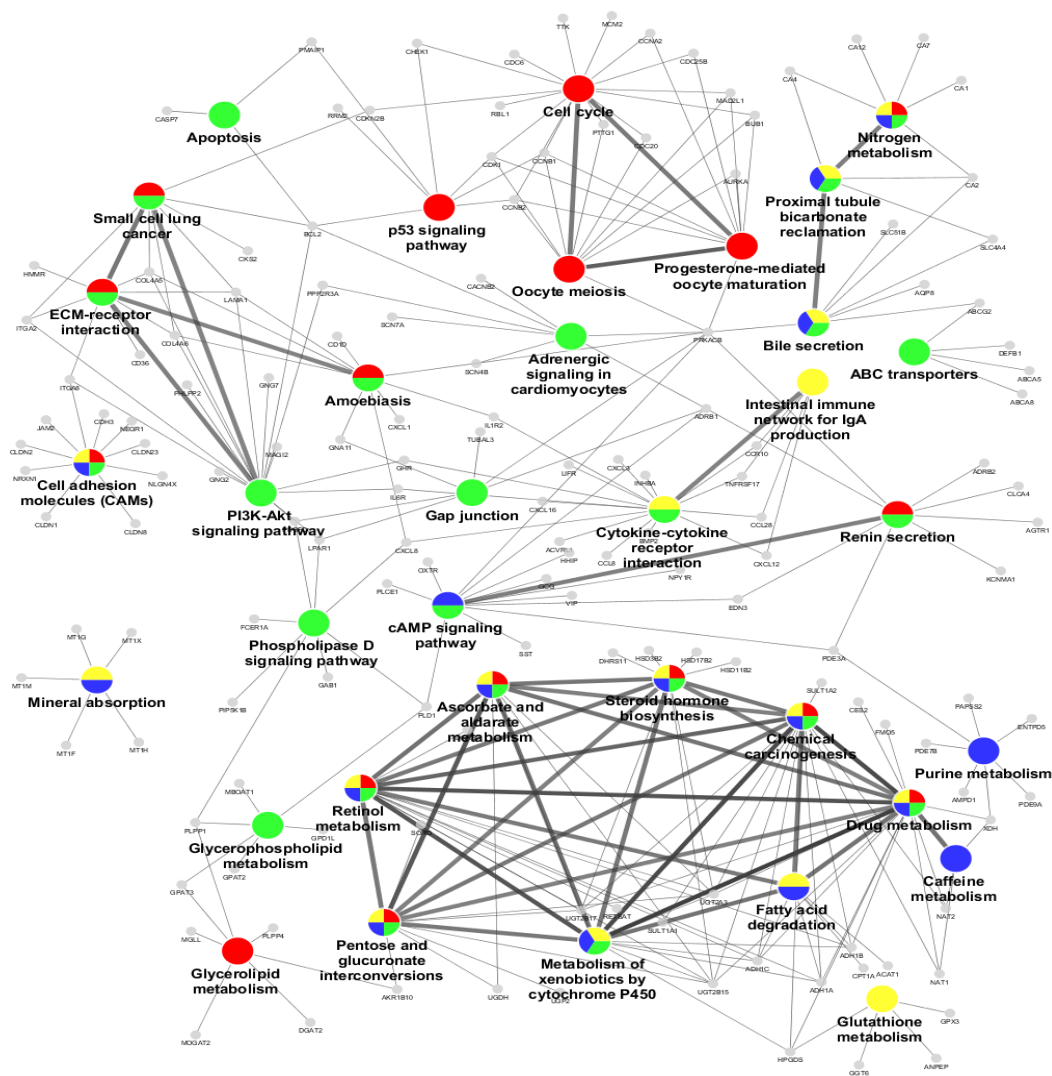


Fig. 1. Interaction network of pathways of all stages of CRC. Each vertex was represented in a color pie chat, where red, yellow, blue, and green represent pathways part of DEG-stage1, DEG-stage2, DEG-stage3, and DEG-stage4, respectively.

decrease bacteria-induced inflammation in tumor sites [19], [20]. Moreover, Chemokine signaling affects CRC invasion and/or metastasis by altering tumor microenvironment [21]. Numerous investigations have indicated that p53 is a tumor suppressor gene [22] and plays a critical role in development of tumors. It contributes to not only control cell cycle [22], but also keep chromosome stability [23] and mitochondrial genetic stability [24], [25]. Above all, we guess that other pathways are also associated with the development of CRC. More details and evidence are investigated in further studies.

Overall, some pathways we obtained have been proven to be involved in the development of CRC, and we predict that others might also be related to the pathogenesis of CRC, which indicates that the method of obtaining related pathways of CRC is effective. Besides, this method also can be performed to identify pathways in other cancer, such as liver cancer, lung cancer, which also contribute to analyze the dynamic evolution process of cancer.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China under [Grant No. 61602386, 61972320, 61702161 and 61772426]; the Natural Science Foundation of Shaanxi Province under [Grant No. 2017JQ6008]; the Fundamental Research Funds for the Central Universities under [Grant No. 3102019DX1003]; and the Top International University Visiting Program for Outstanding Young scholars of Northwestern Polytechnical University.

REFERENCES

- [1] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2019," *CA: a cancer journal for clinicians*, vol. 69, no. 1, pp. 7–34, 2019.
- [2] A. Klaus and W. Birchmeier, "Wnt signalling and its impact on development and cancer," *Nature Reviews Cancer*, vol. 8, no. 5, p. 387, 2008.
- [3] S. You, J. Zhou, S. Chen, P. Zhou, J. Lv, X. Han, and Y. Sun, "Ptch1, a receptor of hedgehog signaling pathway, is correlated with metastatic potential of colorectal cancer," *Upsala journal of medical sciences*, vol. 115, no. 3, pp. 169–175, 2010.

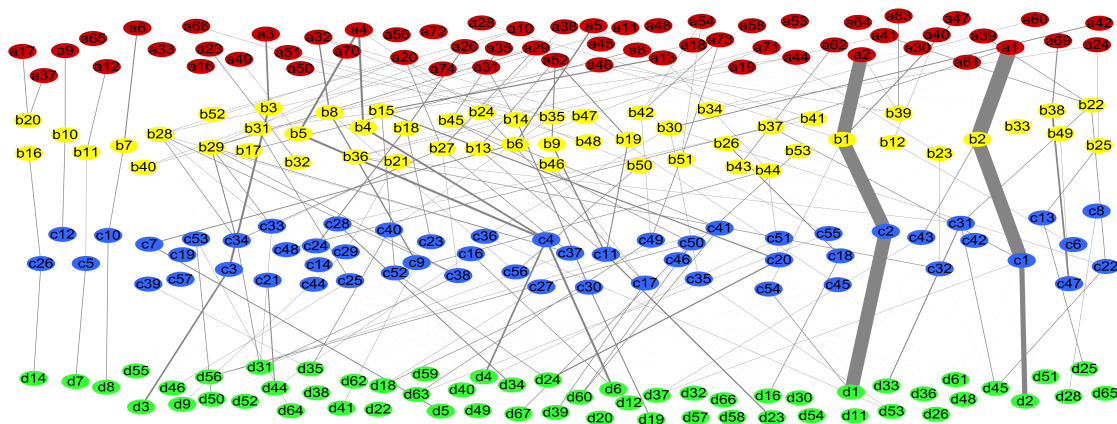


Fig. 2. Modules relation network between adjacent stages of CRC. Red, yellow, blue, and green circles represent modules belonging to DEG-stage1, DEG-stage2, DEG-stage3, and DEG-stage4, respectively.

- [4] X.-L. Li, J. Zhou, Z.-R. Chen, and W.-J. Chng, "P53 mutations in colorectal cancer-molecular pathogenesis and pharmacological reactivation," *World journal of gastroenterology: WJG*, vol. 21, no. 1, p. 84, 2015.
- [5] L. Gautier, L. Cope, B. M. Bolstad, and R. A. Irizarry, "affy: analysis of affymetrix genechip data at the probe level," *Bioinformatics*, vol. 20, no. 3, pp. 307–315, 2004.
- [6] M. Carlson, S. Falcon, H. Pages, and N. Li, "hgu133plus2. db: Affymetrix human genome u133 plus 2.0 array annotation data (chip hgu133plus2)," *R package version*, vol. 3, no. 3, 2016.
- [7] G. K. Smyth, M. Ritchie, N. Thorne, and J. Wettenhall, "Limma: linear models for microarray data. in bioinformatics and computational biology solutions using r and bioconductor. statistics for biology and health," 2005.
- [8] J. A. Ferreira, "The benjamini-hochberg method in the case of discrete test statistics," *The international journal of biostatistics*, vol. 3, no. 1, 2007.
- [9] R. Cheong, A. Rhee, C. J. Wang, I. Nemenman, and A. Levchenko, "Information transduction capacity of noisy biochemical signaling networks," *science*, vol. 334, no. 6054, pp. 354–358, 2011.
- [10] G. Bindea, B. Mlecnik, H. Hackl, P. Charoentong, M. Tosolini, A. Kirilovsky, W.-H. Fridman, F. Pagès, Z. Trajanoski, and J. Galon, "Cluego: a cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks," *Bioinformatics*, vol. 25, no. 8, pp. 1091–1093, 2009.
- [11] P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker, "Cytoscape: a software environment for integrated models of biomolecular interaction networks," *Genome research*, vol. 13, no. 11, pp. 2498–2504, 2003.
- [12] D. Szklarczyk, A. Franceschini, S. Wyder, K. Forslund, D. Heller, J. Huerta-Cepas, M. Simonovic, A. Roth, A. Santos, K. P. Tsafou *et al.*, "String v10: protein–protein interaction networks, integrated over the tree of life," *Nucleic acids research*, vol. 43, no. D1, pp. D447–D452, 2014.
- [13] T. Nepusz, H. Yu, and A. Paccanaro, "Detecting overlapping protein complexes in protein-protein interaction networks," *Nature methods*, vol. 9, no. 5, p. 471, 2012.
- [14] D. B. Jump and S. D. Clarke, "Regulation of gene expression by dietary fat," *Annual review of nutrition*, vol. 19, no. 1, pp. 63–90, 1999.
- [15] B. Hoeft, J. Linseisen, L. Beckmann, K. Müller-Decker, F. Canzian, A. Hüsing, R. Kaaks, U. Vogel, M. U. Jakobsen, K. Overvad *et al.*, "Polymorphisms in fatty acid metabolism-related genes are associated with colorectal cancer risk," *Carcinogenesis*, vol. 31, no. 3, pp. 466–472, 2009.
- [16] S. Kürty, B. Buecher, S. Robiou-du Pont, C. Scoul, V. Sébille, H. Colman, C. Le Houérou, T. Le Neel, J. Bourdon, R. Faroux *et al.*, "Combinations of cytochrome p450 gene polymorphisms enhancing the risk for sporadic colorectal cancer related to red meat consumption," *Cancer Epidemiology and Prevention Biomarkers*, vol. 16, no. 7, pp. 1460–1467, 2007.
- [17] G. Özhan, M. Muter, G. Ercan, and B. Alpertunga, "Genetic variations in the xenobiotic-metabolizing enzymes cyp1a1, cyp1a2, cyp2c9, cyp2c19 and susceptibility to colorectal cancer among turkish people," *Genetic testing and molecular biomarkers*, vol. 18, no. 4, pp. 223–228, 2014.
- [18] X. Wang, G.-l. Hao, B.-y. Wang, C.-c. Gao, Y.-x. Wang, L.-s. Li, and J.-d. Xu, "Function and dysfunction of plasma cells in intestine," *Cell & bioscience*, vol. 9, no. 1, p. 26, 2019.
- [19] S. Le Gouvello, S. Bastuji-Garin, N. Aloulou, H. Mansour, M.-T. Chaumette, F. Berrehar, A. Seikour, A. Charachon, M. Karoui, K. Leroy *et al.*, "High prevalence of foxp3 and il17 in mmr-proficient colorectal carcinomas," *Gut*, vol. 57, no. 6, pp. 772–779, 2008.
- [20] M. Tosolini, A. Kirilovsky, B. Mlecnik, T. Fredriksen, S. Mauger, G. Bindea, A. Berger, P. Bruneval, W.-H. Fridman, F. Pagès *et al.*, "Clinical impact of different classes of infiltrating t cytotoxic and helper cells (th1, th2, treg, th17) in patients with colorectal cancer," *Cancer research*, vol. 71, no. 4, pp. 1263–1271, 2011.
- [21] Y. Itatani, K. Kawada, S. Inamoto, T. Yamamoto, R. Ogawa, M. Taketo, and Y. Sakai, "The role of chemokines in promoting colorectal cancer invasion/metastasis," *International journal of molecular sciences*, vol. 17, no. 5, p. 643, 2016.
- [22] K. Suzuki and H. Matsubara, "Recent advances in p53 research and cancer treatment," *BioMed Research International*, vol. 2011, 2011.
- [23] G. Achanta, R. Sasaki, L. Feng, J. S. Carew, W. Lu, H. Pelicano, M. J. Keating, and P. Huang, "Novel role of p53 in maintaining mitochondrial genetic stability through interaction with dna pol γ ," *The EMBO journal*, vol. 24, no. 19, pp. 3482–3492, 2005.
- [24] G. Liu, J. M. Parant, G. Lang, P. Chau, A. Chavez-Reyes, A. K. El-Naggar, A. Multani, S. Chang, and G. Lozano, "Chromosome stability, in the absence of apoptosis, is critical for suppression of tumorigenesis in trp53 mutant mice," *Nature genetics*, vol. 36, no. 1, p. 63, 2004.
- [25] P. Iglesias, A. Salas, and J. A. Costoya, "The maintenance of mitochondrial genetic stability is crucial during the oncogenic process," *Communicative & integrative biology*, vol. 5, no. 1, pp. 34–38, 2012.