



A Qualitative Transcriptional Signature for the Risk Assessment of Precancerous Colorectal Lesions

Qingzhou Guan^{1,2†}, Qihong Zeng^{2†}, Weizhong Jiang^{3†}, Jiajing Xie², Jun Cheng², Haidan Yan², Jun He², Yang Xu², Guoxian Guan^{3*}, Zheng Guo^{2*} and Lu Ao^{2*}

¹ Collaborative Innovation Center for Chinese Medicine and Respiratory Diseases Co-Constructed by Henan Province & Education Ministry of P.R. China, Academy of Chinese Medical Sciences, Henan University of Chinese Medicine, Zhengzhou, China, ² Key Laboratory of Medical Bioinformatics, Key Laboratory of Ministry of Education for Gastrointestinal Cancer, School of Basic Medical Sciences, Fujian Medical University, Fuzhou, China, ³ Department of Colorectal Surgery, The Affiliated Union Hospital of Fujian Medical University, Fuzhou, China

OPEN ACCESS

Edited by:

Hongwei Wang,
Sun Yat-sen University, China

Reviewed by:

Dechao Bu,
Chinese Academy of Sciences (CAS),
China

Juntao Li,
Henan Normal University, China

*Correspondence:

Lu Ao
lukey@fjmu.edu.cn
Zheng Guo
guoz@ems.hrbmu.edu.cn
Guoxian Guan
gxguan1108@163.com

† These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Computational Genomics,
a section of the journal
Frontiers in Genetics

Received: 18 June 2020

Accepted: 01 December 2020

Published: 15 January 2021

Citation:

Guan Q, Zeng Q, Jiang W, Xie J,
Cheng J, Yan H, He J, Xu Y, Guan G,
Guo Z and Ao L (2021) A Qualitative
Transcriptional Signature for the Risk
Assessment of Precancerous
Colorectal Lesions.
Front. Genet. 11:573787.
doi: 10.3389/fgene.2020.573787

It is meaningful to assess the risk of cancer incidence among patients with precancerous colorectal lesions. Comparing the within-sample relative expression orderings (REOs) of colorectal cancer patients measured by multiple platforms with that of normal colorectal tissues, a qualitative transcriptional signature consisting of 1,840 gene pairs was identified in the training data. Within an evaluation dataset of 16 active and 18 inactive (remissive) ulcerative colitis subjects, the median incidence risk score of colorectal carcinoma was 0.6402 in active ulcerative colitis subjects, significantly higher than that in remissive subjects (0.3114). Evaluation of two other independent datasets yielded similar results. Moreover, we found that the score significantly positively correlated with the degree of dysplasia in the case of colorectal adenomas. In the merged dataset, the median incidence risk score was 0.9027 among high-grade adenoma samples, significantly higher than that among low-grade adenomas (0.8565). In summary, the developed incidence risk score could well predict the incidence risk of precancerous colorectal lesions and has value in clinical application.

Keywords: qualitative transcriptional signature, incidence risk score, colorectal cancer, ulcerative colitis, colorectal adenoma

INTRODUCTION

Colorectal cancer (CRC) is one of the commonest malignancies worldwide, with high morbidity and mortality rates (Arnold et al., 2017). The condition mainly develops from malignant transformation of acquired precancerous lesions (Conteduca et al., 2013; Brenner et al., 2014), such as inflammatory bowel disease (IBD) and colorectal adenomas. Chronic IBD is a major type of precancerous colorectal lesions, which has two forms: ulcerative colitis (UC) and Crohn's disease (CD). Long-term exposure to chronic inflammation is the primary risk factor for CRC pathogenesis (Axelrad et al., 2016; Fujita et al., 2018). The progression of CRC ranges from the physiological state to quiescent chronic inflammation that then progresses to active chronic inflammation

Abbreviations: CRC, Colorectal Cancer; UC, Ulcerative Colitis; REO, Relative Expression Ordering; IBD, Inflammatory Bowel Disease; CD, Crohn's Disease; RMA, the Robust Multi-array Average; FPKM, Fragments Per Kilobase of transcript per Million fragments.

without dysplasia. Dysplasia eventually develops and progresses to outright malignancy (Bjerrum et al., 2014). Moreover, the proportion of bowel affected by IBD and the severity of inflammation likewise affect CRC risk of patients with IBD (Feagins et al., 2009). Another major type of precancerous colorectal lesions is colorectal adenomas (Conteduca et al., 2013; Brenner et al., 2014). The development sequence is from normal colonic mucosa to small tubular adenoma to large adenoma and finally to cancer (Levine and Ahnen, 2006). The risk of developing colorectal cancer for patients with adenomas is two to four times higher than those patients without adenomas (Lotfi et al., 1986; Avidan et al., 2002; Rex et al., 2006). Individuals suffering signs and symptoms suggestive of CRC, including IBD, changes in bowel habits, and bloody stools are advised to seek medical help, including endoscopic or radiologic imaging examination and non-invasive tests (Quaife et al., 2014; Mhaidat et al., 2018). Colonoscopy is an essential step in the diagnosis of CRC with 4–21% inaccurate rate (Stanciu et al., 2007; Huang et al., 2016), which is mainly influenced by the image quality and the endoscopists' experience (Costamagna and Marchese, 2010; Kaminski et al., 2010; Tumino et al., 2017). Currently, established non-invasive tests, such as the fecal occult blood test (FOBT), have a low sensitivity (Schottinger et al., 2016) and positive predictive value (Azimafousse Assogba et al., 2015; Brown et al., 2018). Methylated Septin9 (^mSEPT9) has superior sensitivity compared to FOBT, which range from 36.6 to 95.6% (Warren et al., 2011; Ahlquist et al., 2012; Toth et al., 2012; Lee et al., 2013; Church et al., 2014). Moreover, some serum protein biomarkers, including carcinoembryonic antigen, CA19.9, and CA125, are used for monitoring the prognosis of CRC patients (Duffy et al., 2007; Soler et al., 2016). However, patients can be identified as either cancerous or non-cancerous based on colonoscopy or these diagnostic signatures (Guan et al., 2018). None could accurately assess the risk of cancer incidence among non-cancerous patients with precancerous colorectal lesions. And, to the best of our knowledge, there is currently no such molecule-based incidence risk score. Thus, it is of great clinical value to construct a molecular signature to assess the incidence of precancerous colorectal lesions converting to CRC, which could eventually aid in the prevention of CRC occurrence.

Our recent studies (Ao et al., 2018; Guan et al., 2018) demonstrated that compared with the quantitative transcriptional signatures, the qualitative transcriptional signatures—namely the relative expression orderings (REOs) of gene pairs within individual samples—are robust against experimental batch effects and could be directly applied to samples at the individualized level (Guan et al., 2016; Ao et al., 2018). In addition, we have reported that the qualitative transcriptional signatures are also highly robust against specimens with different proportions of tumor epithelial cells from different tumor locations from the same patients (Cheng et al., 2017), partial RNA degradation during specimen preparation and storage (Chen et al., 2017) and amplification bias for minimum specimens (Liu et al., 2017). The qualitative transcriptional signature is thus much more applicable to clinical application. Furthermore, the qualitative

transcriptional signature is also suitable for application to inaccurately sampled specimens in clinical settings (Ao et al., 2018; Guan et al., 2019).

Based on the unique advantages of the qualitative transcriptional signature, we selected gene pairs with stable but reversal REO patterns among CRC and normal colorectal tissues as the signature for calculating incidence risk scores of precancerous lesions. These scores, in turn, were used to predict the incidence risk of malignant transformation among non-cancer patients suffering such precancerous colorectal lesions. Score performance was evaluated using multiple independent datasets via the comparison of CRC incidence risk scores among non-cancer patients with precancerous lesions (i.e., UC and adenomas) at different disease stages. Results revealed that the incidence risk scores of high-grade precancerous lesions were significantly higher than those of low-grade lesions, suggesting that the signature could well predict the incidence risk of CRC in patients suffering precancerous colorectal lesions.

MATERIALS AND METHODS

Data and Preprocessing

The gene expression profiles used in this study were shown in **Tables 1, 2**, including CRC samples and normal samples. All normal colorectal tissue samples were obtained from individuals that were demonstrated to lack polyps and a known family history of previous CRC. Notably, all UC and adenoma tissue samples analyzed in this study were obtained from biopsy. All data were downloaded from Gene Expression Omnibus (Barrett et al., 2013) (GEO¹) and The Cancer Genome Atlas (International Cancer Genome Consortium et al., 2010) (TCGA²).

For array-based data measured using the Affymetrix platform, raw mRNA expression data (.CEL files) were downloaded and the Robust Multi-array Average (RMA) algorithm was used for

¹<http://www.ncbi.nlm.nih.gov/geo/>

²<http://cancergenome.nih.gov/>

TABLE 1 | Datasets used for building the CRC incidence-risk score.

Platform	GEO Acc	Normal	GEO Acc	Cancer	
Affymetrix	GSE4107	10	GSE27854	115	
	GSE9348	12	GSE21510	123	
	GSE47908	15	GSE17536	177	
	GSE36807	7	GSE10714	7	
	GSE10714	3	GSE4183	15	
	GSE4183	8	GSE9348	70	
			GSE32323	17	
			GSE33113	90	
	Illumina RNA_seq	GSE48634	69	GSE37178	84
		GSE56789	40	GSE33126	9
GSE53306		12	GSE31279	44	
			GSE50760	36	
			TCGA	556	

Empty spaces indicate that there is no sample in the corresponding category.

TABLE 2 | Datasets used for evaluating the performance of the score.

Platform	GEO Acc	Active UC	Inactive UC	UC_inflammation	UC_without_inflammation
Affymetrix	GSE13367	16	18		
	GSE9452			8	13
RNA_seq	GSE53306	16	12		
		High-grade adenoma	Low-grade adenoma		
Affymetrix	GSE37364	13	16		
	GSE8671	10	14		
		Normal	CRC		
Affymetrix	GSE22619	10			
Affymetrix PrimeView™ Array	Our_Data1		33		
Illumina HiSeqXten	Our_Data2		13		

Empty spaces indicate that there was no sample in the corresponding category; Active UC and Inactive UC denote samples with active and inactive (remissive) ulcerative colitis, respectively; UC_inflammation and UC_without_inflammation denote samples with and without macroscopic signs of inflammation, respectively; High-grade adenoma and Low-grade adenoma denote colorectal adenoma samples with high-grade and low-grade dysplasia, respectively. Normal denotes the normal colorectal tissue samples were obtained from individuals that were demonstrated to lack polyps and a known family history of previous CRC; Our_Data1 and Our_Data2 respectively denotes 33 CRC biopsies and 13 CRC surgical resection specimens measured in our previous study.

background adjustment without quantile normalization (Irizarry et al., 2003). For array- or sequence-based data measured using the Illumina platform, processed data were directly downloaded. For sequence-based data from TCGA, the FPKM (fragments per kilobase of transcript per million fragments mapped) (Trapnell et al., 2010) value was downloaded.

For array-based data, the probe ID was mapped to the Entrez gene ID with the corresponding platform file. If a probe was mapped to zero or multiple genes, then data of this probe were discarded. If multiple probes were mapped to the same gene, the expression value of this gene was defined as the arithmetic mean of the values of those probes. For sequence-based data from ArrayExpress, gene symbols were mapped to Entrez gene ID with the biological database network (Mudunuri et al., 2009) (bioDBnet³). For sequence-based data from TCGA, Ensembl gene IDs were mapped to unique Entrez gene IDs of protein coding genes.

Feature Selection

Based on gene expression profiles for CRC and normal colorectal tissues in training data (shown in **Table 1**), we selected gene pairs with reversal REOs by comparing highly stable gene pairs of CRC with normal tissue samples (the threshold was 90% in this study).

In the training datasets, the gene measurement of each gene was converted to the corresponding rank within each sample (i.e., the smallest measurement was converted to the minimum rank and the largest measurement was converted to the maximum rank). Then, pair wise comparisons were performed for all within-sample genes to identify stable gene pairs for a specific tissue type. For one sample, the REO pattern of two genes, i and j , was denoted as $G_i > G_j$ (or $G_i < G_j$) if the rank of gene i was higher (or lower) than that of gene j . If gene pair (i, j) had the same REO pattern as did a majority of samples (e.g., 90%), it was considered a stable REO pattern and the gene pair was defined as a stable gene pair. For two groups of samples, a gene pair with a stable but reversal REO pattern in the two groups was defined as

a reversal gene pair. For the reversal gene pairs identified in the above process, REOs of gene pairs in CRC tissues were defined as the CRC signature that was applied for defining the incidence risk score of precancerous colorectal lesions.

Risk Scoring Model Construction

In this study, we defined the REO patterns $G_i > G_j$ and $G_i < G_j$ to characterize CRC and normal tissues, respectively. For each non-cancer patient with a precancerous lesion, the risk score was simply calculated as the percentage of REOs characterizing CRC. The incidence risk score for a particular sample was calculated as:

$$\text{Score} = n/m,$$

where m was the number of gene pairs included in the signature, n was the number of gene pairs with the same REO patterns characterizing CRC. The higher the incidence risk score was, the greater the cancer incidence risk that the patient had.

The performance of the risk score was then evaluated using samples from patients with precancerous colorectal lesions (including UC and adenomas) at different disease stages from multiple datasets (**Table 2**).

Functional Analyses

A total of 244 pathways covering 6,934 unique genes were obtained from Kyoto Encyclopedia of Genes and Genomes (KEGG) database⁴ (Kanehisa and Goto, 2000). The hypergeometric distribution model was used to calculate the significance of enriched pathways with interested genes (Fury et al., 2006). And p values were adjusted with Benjamini-Hochberg method.

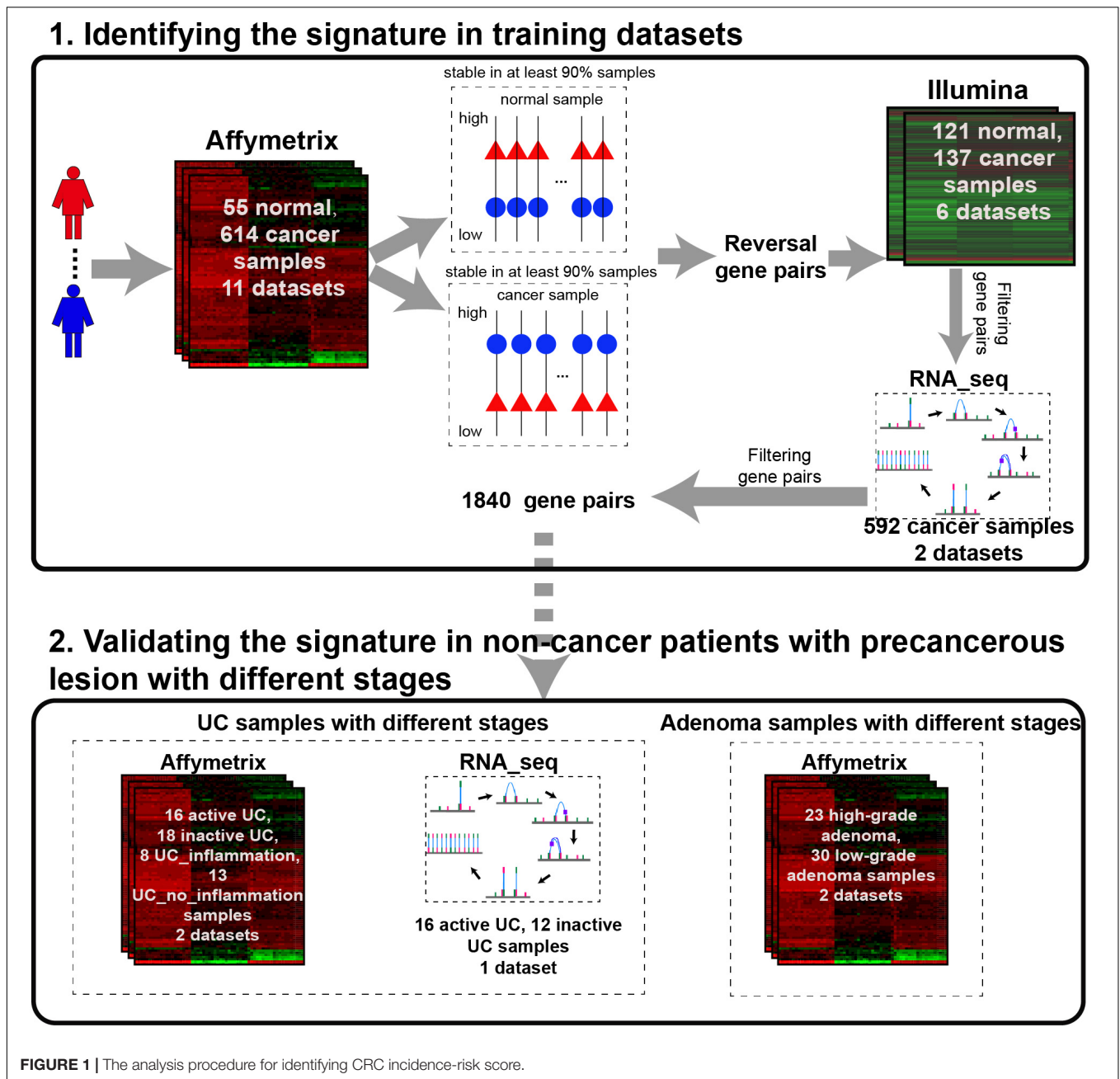
RESULTS

Acquisition of REO Features

The analysis procedure of this study is described in **Figure 1**. Considering that carcinogenesis of CRC is a continuous,

³<https://biodbnet-abcc.ncicrf.gov/db/db2db.php>

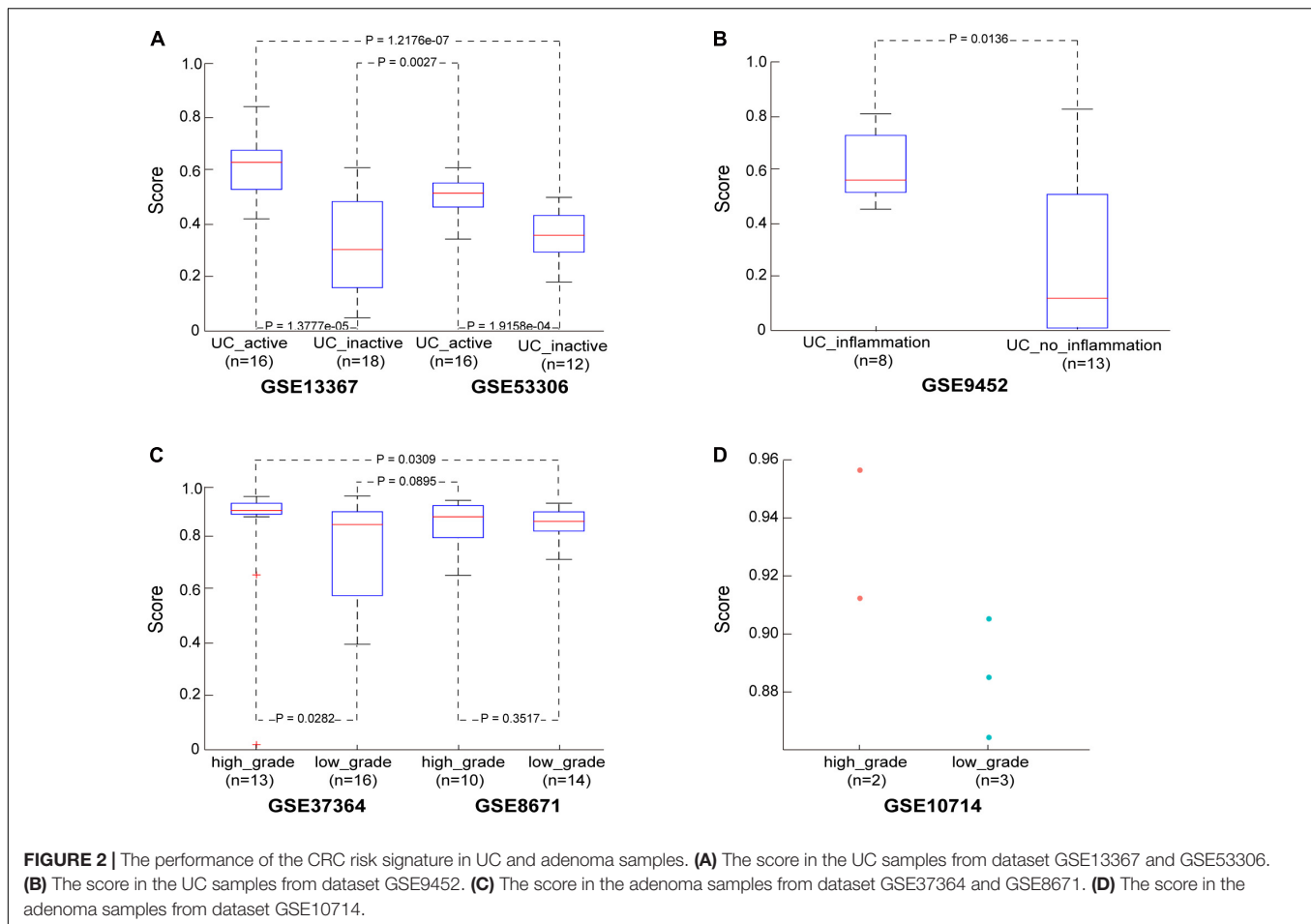
⁴<https://www.kegg.jp/kegg/pathway.html>



multistep malignant transformation of normal colorectal tissues, we initially identified gene pairs with stable but reversal REOs between CRC and normal colorectal tissue samples with a threshold of 90% (see section “Material and Methods”).

For the 614 CRC and 55 normal colorectal tissue samples from the 11 datasets measured using the Affymetrix microarray platform (see **Table 1**), 356,573 gene pairs with stable (threshold of 90%) but reversal REO patterns between CRC and normal tissues were identified; these gene pairs were defined as reversal gene pairs. Similarly, for the 137 CRC and 121 normal colorectal tissue samples from the six datasets measured using the Illumina microarray platform (see **Table 1**), 406,957 reversal

gene pairs were identified. We found 18,135 gene pairs that were consistently detected in the above two lists of reversal gene pairs. Among those 18,135 gene pairs, we further selected 1,840 gene pairs that had identical REO patterns in at least 90% of the 556 cancer samples in TCGA and 36 cancer samples in the GSE50760 dataset which were measured using the RNA_seq platform. Finally, the 1,840 gene pairs (see **Supplementary Table S1**) were selected and the REOs of the selected gene pairs of CRC tissues were defined as the CRC signature that was applied for defining the incidence risk score of precancerous colorectal lesions (see section “Material and Methods”).



Validation and Functional Analysis of the CRC Incidence Risk Signature

For each sample, the risk score was simply calculated as the percentage of REOs characterizing CRC, which was close to 0 and 1 in normal and CRC tissues, respectively. Among training data with all merged samples, score medians were 0.0174 and 0.9891 for normal and CRC tissues, respectively. In the GSE22619 dataset of 10 normal colorectal tissues, we found the median score to be 0.1139. With RNA-seq, we studied 13 CRC tissue samples obtained via surgical resection and 33 CRC tissues obtained from biopsy (Guo et al., 2018), finding score medians of 0.9919 and 0.9271, respectively. This data indicated that the risk score was applicable for CRC tissues obtained for biopsy, although there were variations in data.

To elucidate the functions and pathways that were associated with the CRC incidence risk signature, KEGG pathway analysis were performed. Functional enrichment analysis of 1,580 genes included in the signature showed that 10 pathways were significantly enriched ($p < 0.05$, **Supplementary Figure S1**), including p53 signaling pathway and ECM-receptor interaction pathway. The p53 signaling pathway has been reported to be involved in cell cycle regulation and suppression of tumor expression (Hu et al., 2007; Kruiswijk et al., 2015; Tanikawa et al., 2017). ECM-receptor interaction pathway plays an important role in the process of CRC (such as tumor shedding, adhesion,

degradation, movement and hyperplasia) and could promote the development of epithelial mesenchymal transition (EMT) in cancer cells (Rahbari et al., 2016). Moreover, ECM also plays a key role in the process of other cancer types, such as prostate cancer and gastric cancer (Andersen et al., 2018; Yan et al., 2018).

The Performance of the CRC Incidence Risk Score in UC Samples

The typical pathogenesis of CRC is the transformation from normal cells to quiescent chronic inflammation. Dysplasia eventually arises from persistent inflammation and ultimately progresses to outright malignant transformation (Bjerrum et al., 2014). Thus, we evaluated the performance of our score in UC samples at different stages of the disease course.

Then, in the GSE13367 dataset consisting of 16 active and 18 inactive (remissive) UC samples, the median CRC incidence risk score of active UC samples was 0.6402, significantly higher than that in remissive UC samples (Wilcoxon rank sum test; $p = 1.3777e-05$). Similar findings were also obtained in the GSE53306 dataset consisting of 16 active and 12 remissive UC samples (Wilcoxon rank sum test; $p = 1.9158e-04$). Detailed results are shown in **Figure 2A** and **Supplementary Table S2**. Importantly, the incidence risk scores of the 16 active UC samples from the GSE13367 dataset were also significantly higher than those in the 12 remissive UC samples from the GSE53306 dataset

(Wilcoxon rank sum test; $p = 1.2176e-07$). Similar results were also obtained when analyzing active and remissive UC samples from datasets GSE53306 and GSE13367 (Wilcoxon rank sum test; $p = 0.0027$), respectively.

We also evaluated the applicability of our score to the GSE9452 dataset, which included 8 UC samples with and 13 UC samples without (UC_inflammation; UC_without_inflammation) macroscopic signs of inflammation. We found that CRC incidence risk scores in UC_inflammation samples (median = 0.5682) were significantly higher than those in UC_without_inflammation samples (median = 0.1228) (Wilcoxon rank sum test; $p = 0.0136$), as shown in **Figure 2B** and **Supplementary Table S3**. These findings further confirmed that our score could predict UC sample cancer incidence risk.

The Performance of the CRC Incidence Risk Score in Adenoma Samples

The transformation of normal colorectal tissue to adenomatous tissue and finally to outright malignancy is the typical pathogenic process of CRC (Conteduca et al., 2013; Brenner et al., 2014). We thus also evaluated our score in colorectal adenoma samples at different disease stages.

In the GSE37364 dataset consisting of 13 high-grade and 16 low-grade dysplasia colorectal adenoma samples, the median incidence risk score in the high-grade dysplasia samples was 0.9076, significantly higher than that in the low-grade dysplasia samples (median = 0.8543) (Wilcoxon rank sum test; $p = 0.0282$). In the GSE8671 dataset, the CRC incidence risk scores in the 10 high-grade dysplasia samples (median = 0.8837) were only slightly higher than those in the 14 low-grade dysplasia samples (median = 0.8663, Wilcoxon rank sum test, $p = 0.3517$), as shown in **Figure 2C** and **Supplementary Table S4**, which may be ascribed to insufficient sample size. Then, we merged samples from the GSE37364 and GSE8671 datasets to further evaluate the performance of our score. For merged data with 23 high-grade and 30 low-grade dysplasia samples, the median CRC incidence risk score of high-grade dysplasia samples was 0.9027, significantly higher than that of low-grade dysplasia samples (median = 0.8565) (Wilcoxon rank sum test; $p = 0.0191$). The CRC incidence risk scores of the 13 high-grade dysplasia samples from the GSE37364 dataset were also significantly higher than those of the 14 low-grade dysplasia samples from the GSE8671 dataset (Wilcoxon rank sum test; $p = 0.0309$). Similar results were also obtained for high-grade dysplasia samples from the GSE8671 dataset and low-grade dysplasia samples from the GSE37364 dataset (Wilcoxon rank sum test; $p = 0.0895$). We also evaluated our score using the GSE10714 dataset consisting of 2 high-grade and 3 low-grade dysplasia colorectal adenoma samples. The median incidence risk score in the high-grade dysplasia samples was 0.9345, also higher than that in the low-grade dysplasia samples (median = 0.8853), as shown in **Figure 2D**.

Our aforementioned results further demonstrated that our score could well predict the incidence risk of CRC in non-cancer patients with precancerous lesions and that it was applicable to samples from different sources.

DISCUSSION

CRC mainly develops from malignant transformation of acquired precancerous lesions, such as IBD and colorectal adenomas. Based on the qualitative transcriptional characteristics, we developed a signature to assess the incidence risk of precancerous colorectal lesions to CRC by calculating the percentage of gene pairs in our signature that characterized the CRC tissue. For non-cancer patients with precancerous colorectal lesions, such as UC and adenomas, at different stages in the disease course and using data from multiple datasets, our score was well verified. Moreover, we also found that the CRC incidence risk scores of pan-colitis samples were higher than those of left-sided colitis samples, as was previously reported (Fujita et al., 2018).

The top five highest-frequency genes in the CRC incidence risk signature were *ABCG2*, *SLC51B*, *CLDN1*, *TEX11*, and *SLC25A34*, which have been reported to be related with the pathogenesis of CRC. *ABCG2* plays an important role in the progression and metastasis of CRC (Liu et al., 2010). The increased expression of *SLC51B* (also called *OSTβ*) in feces, one of the key membrane transporters of bile acids, is positively correlated with the incidence of CRC (Ballatori et al., 2009). Studies have suggested that expression of *CLDN1* is a prognostic factor in CRC patients (Nakagawa et al., 2011), while *TEX11* likely serves as a biomarker of early onset CRC (Luo et al., 2013). All of these genes are potential targets for future therapeutic interventions.

Due to the lack of corresponding clinical follow-up data, we could not verify whether individuals without cancer and with high CRC incidence risk score, as identified by our signature eventually develop cancer. In future research, we plan to collaborate with a tertiary healthcare facility to better evaluate UC or adenoma, to perform sampling at more diverse sites pre-malignant sites and samples at different stages of illness progression. Patients should be closely followed in the future to further evaluate the utility of our score and to compare the CRC incidence-risk score with the time from diagnosis to cancer incidence.

In summary, our score, based on qualitative transcriptional parameters, is robust against batch effects as well as amplification bias for minimum specimens. Thus, our calculated score is applicable to be used in the individualized diagnosis and is generally suitable for analyzing inaccurately sampled tissues in the clinical setting.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

QG, QZ, and WJ conceived the study, analyzed the data, made figures, performed the statistical analysis, and drafted

the manuscript. JX, JC, HY, JH, and LA searched the data and participated in the statistical analysis. YX participated in discussing and revising the manuscript. ZG and GG conceived of the study, participated in its design and coordination, helped to draft the manuscript, and supervised the work. QG, QZ, and LA revised the manuscript. All authors read and approved the final manuscript.

FUNDING

This work was supported by the National Natural Science Foundation of China (Grant Nos: 81872396 and 61801118), the Joint Scientific and Technology Innovation Fund of Fujian Province (Grant Nos: 2016Y9044 and 2018Y9065), young and middle-aged backbone training project in the health system of Fujian province (Grant No: 2016-ZQN-26), Fujian Provincial Finance Department Special Fund (Grant No: 2015-1297), China National Postdoctoral Program for Innovative Talents (Grant No: BX20200115), China Postdoctoral Science Foundation (Grant No: 2020M682314), the Joint Research Program of Health and

Education in Fujian Province (Grant No: 2019-WJ-32), and the Natural Science Foundation of Fujian Province (Grant No: 2020J01600).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2020.573787/full#supplementary-material>

Supplementary Figure 1 | Bar plot of KEGG pathways enriched by genes involved in the CRC risk signature.

Supplementary Table 1 | The gene pair of the CRC incidence-risk score signature.

Supplementary Table 2 | CRC incidence risk scores in active and inactive (remissive) UC samples.

Supplementary Table 3 | CRC incidence risk scores in UC_inflammation and UC_without_inflammation samples.

Supplementary Table 4 | CRC incidence risk scores in high-grade and low-grade dysplasia adenoma samples.

REFERENCES

- Ahlquist, D. A., Taylor, W. R., Mahoney, D. W., Zou, H., Domanico, M., Thibodeau, S. N., et al. (2012). The stool DNA test is more accurate than the plasma septin 9 test in detecting colorectal neoplasia. *Clin. Gastroenterol. Hepatol.* 10, 272–277.e1. doi: 10.1016/j.cgh.2011.10.008
- Andersen, M. K., Rise, K., Giskeodegard, G. F., Richardsen, E., Bertilsson, H., Storkersen, O., et al. (2018). Integrative metabolic and transcriptomic profiling of prostate cancer tissue containing reactive stroma. *Sci. Rep.* 8:14269. doi: 10.1038/s41598-018-32549-32541
- Ao, L., Zhang, Z., Guan, Q., Guo, Y., Zhang, J., Lv, X., et al. (2018). A qualitative signature for early diagnosis of hepatocellular carcinoma based on relative expression orderings. *Liver Int.* 38, 1812–1819. doi: 10.1111/liv.13864
- Arnold, M., Sierra, M. S., Laversanne, M., Soerjomataram, I., Jemal, A., and Bray, F. (2017). Global patterns and trends in colorectal cancer incidence and mortality. *Gut* 66, 683–691. doi: 10.1136/gutjnl-2015-310912
- Avidan, B., Sonnenberg, A., Schnell, T. G., Leya, J., Metz, A., and Sontag, S. J. (2002). New occurrence and recurrence of neoplasms within 5 years of a screening colonoscopy. *Am. J. Gastroenterol.* 97, 1524–1529. doi: 10.1111/j.1572-0241.2002.05801.x
- Axelrad, J. E., Lichtiger, S., and Yajnik, V. (2016). Inflammatory bowel disease and cancer: the role of inflammation, immunosuppression, and cancer treatment. *World J. Gastroenterol.* 22, 4794–4801. doi: 10.3748/wjg.v22.i20.4794
- Azimafousse Assogba, G. F., Jezewski-Serra, D., Lastier, D., Quintin, C., Denis, B., Beltzer, N., et al. (2015). Impact of subsequent screening episodes on the positive predictive value for advanced neoplasia and on the distribution of anatomic subsites of colorectal cancer: a population-based study on behalf of the French colorectal cancer screening program. *Cancer Epidemiol.* 39, 964–971. doi: 10.1016/j.canep.2015.09.008
- Ballatori, N., Li, N., Fang, F., Boyer, J. L., Christian, W. V., and Hammond, C. L. (2009). OST alpha-OST beta: a key membrane transporter of bile acids and conjugated steroids. *Front. Biosci. (Landmark Ed)* 14:2829–2844. doi: 10.2741/3416
- Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., et al. (2013). NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res.* 41, D991–D995. doi: 10.1093/nar/gks1193
- Bjerrum, J. T., Nielsen, O. H., Riis, L. B., Pittet, V., Mueller, C., Rogler, G., et al. (2014). Transcriptional analysis of left-sided colitis, pancolitis, and ulcerative colitis-associated dysplasia. *Inflamm. Bowel Dis.* 20, 2340–2352. doi: 10.1097/MIB.0000000000000235
- Brenner, H., Kloor, M., and Pox, C. P. (2014). Colorectal cancer. *Lancet* 383, 1490–1502. doi: 10.1016/S0140-6736(13)61649-61649
- Brown, J. P., Wooldrage, K., Wright, S., Nickerson, C., Cross, A. J., and Atkin, W. S. (2018). High test positivity and low positive predictive value for colorectal cancer of continued faecal occult blood test screening after negative colonoscopy. *J. Med. Screen* 25, 70–75. doi: 10.1177/0969141317698501
- Chen, R., Guan, Q., Cheng, J., He, J., Liu, H., Cai, H., et al. (2017). Robust transcriptional tumor signatures applicable to both formalin-fixed paraffin-embedded and fresh-frozen samples. *Oncotarget* 8, 6652–6662. doi: 10.18632/oncotarget.14257
- Cheng, J., Guo, Y., Gao, Q., Li, H., Yan, H., Li, M., et al. (2017). Circumvent the uncertainty in the applications of transcriptional signatures to tumor tissues sampled from different tumor sites. *Oncotarget* 8, 30265–30275. doi: 10.18632/oncotarget.15754
- Church, T. R., Wandell, M., Lofton-Day, C., Mongin, S. J., Burger, M., Payne, S. R., et al. (2014). Prospective evaluation of methylated SEPT9 in plasma for detection of asymptomatic colorectal cancer. *Gut* 63, 317–325. doi: 10.1136/gutjnl-2012-304149
- Conteduca, V., Sansonno, D., Russi, S., and Dammacco, F. (2013). Precancerous colorectal lesions (Review). *Int. J. Oncol.* 43, 973–984. doi: 10.3892/ijo.2013.2041
- Costamagna, G., and Marchese, M. (2010). Progress in endoscopic imaging of gastrointestinal tumors. *Eur. Rev. Med. Pharmacol. Sci.* 14, 272–276.
- Duffy, M. J., van Dalen, A., Haglund, C., Hansson, L., Holinski-Feder, E., Klapdor, R., et al. (2007). Tumour markers in colorectal cancer: European Group on Tumour Markers (EGTM) guidelines for clinical use. *Eur. J. Cancer* 43, 1348–1360. doi: 10.1016/j.ejca.2007.03.021
- Feagins, L. A., Souza, R. F., and Spechler, S. J. (2009). Carcinogenesis in IBD: potential targets for the prevention of colorectal cancer. *Nat. Rev. Gastroenterol. Hepatol.* 6, 297–305. doi: 10.1038/nrgastro.2009.44
- Fujita, M., Matsubara, N., Matsuda, I., Maejima, K., Oosawa, A., Yamano, T., et al. (2018). Genomic landscape of colitis-associated cancer indicates the impact of chronic inflammation and its stratification by mutations in the Wnt signaling. *Oncotarget* 9, 969–981. doi: 10.18632/oncotarget.22867
- Fury, W., Batliwalla, F., Gregersen, P. K., and Li, W. (2006). Overlapping probabilities of top ranking gene lists, hypergeometric distribution, and stringency of gene selection criterion. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 2006, 5531–5534. doi: 10.1109/IEMBS.2006.260828
- Guan, Q., Chen, R., Yan, H., Cai, H., Guo, Y., Li, M., et al. (2016). Differential expression analysis for individual cancer samples based on robust

- within-sample relative gene expression orderings across multiple profiling platforms. *Oncotarget* 7, 68909–68920. doi: 10.18632/oncotarget.11996
- Guan, Q., Yan, H., Chen, Y., Zheng, B., Cai, H., He, J., et al. (2018). Quantitative or qualitative transcriptional diagnostic signatures? a case study for colorectal cancer. *BMC Genom.* 19:99. doi: 10.1186/s12864-018-4446-y
- Guan, Q., Zeng, Q., Yan, H., Xie, J., Cheng, J., Ao, L., et al. (2019). A qualitative transcriptional signature for the early diagnosis of colorectal cancer. *Cancer Sci.* 110, 3225–3234. doi: 10.1111/cas.14137
- Guo, Y., Jiang, W., Ao, L., Song, K., Chen, H., Guan, Q., et al. (2018). A qualitative signature for predicting pathological response to neoadjuvant chemoradiation in locally advanced rectal cancers. *Radiother Oncol.* 129, 149–153. doi: 10.1016/j.radonc.2018.01.010
- Hu, W., Feng, Z., Teresky, A. K., and Levine, A. J. (2007). p53 regulates maternal reproduction through LIF. *Nature* 450, 721–724. doi: 10.1038/nature05993
- Huang, R. X., Xiao, Z. L., Li, F., Ji, D. N., Zhou, J., Xiang, P., et al. (2016). Black hood assisted colonoscopy for detection of colorectal polyps: a prospective randomized controlled study. *Eur. Rev. Med. Pharmacol. Sci.* 20, 3266–3272.
- International Cancer Genome Consortium, Hudson, T. J., Anderson, W., Artez, A., Barker, A. D., et al. (2010). International network of cancer genome projects. *Nature* 464, 993–998. doi: 10.1038/nature08987
- Irizarry, R. A., Hobbs, B., Collin, F., Beazer-Barclay, Y. D., Antonellis, K. J., Scherf, U., et al. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 4, 249–264. doi: 10.1093/biostatistics/4.2.249
- Kaminski, M. F., Regula, J., Kraszewska, E., Polkowski, M., Wojciechowska, U., Didkowska, J., et al. (2010). Quality indicators for colonoscopy and the risk of interval cancer. *N. Engl. J. Med.* 362, 1795–1803. doi: 10.1056/NEJMoa0907667
- Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30. doi: 10.1093/nar/28.1.27
- Kruiswijk, F., Labuschagne, C. F., and Vousden, K. H. (2015). p53 in survival, death and metabolic health: a lifeguard with a licence to kill. *Nat. Rev. Mol. Cell Biol.* 16, 393–405. doi: 10.1038/nrm4007
- Lee, H. S., Hwang, S. M., Kim, T. S., Kim, D. W., Park, D. J., Kang, S. B., et al. (2013). Circulating methylated septin 9 nucleic acid in the plasma of patients with gastrointestinal cancer in the stomach and colon. *Transl. Oncol.* 6, 290–296. doi: 10.1593/tlo.13118
- Levine, J. S., and Ahnen, D. J. (2006). Clinical practice. Adenomatous polyps of the colon. *N. Engl. J. Med.* 355, 2551–2557. doi: 10.1056/NEJMcp063038
- Liu, H., Li, Y., He, J., Guan, Q., Chen, R., Yan, H., et al. (2017). Robust transcriptional signatures for low-input RNA samples based on relative expression orderings. *BMC Genom.* 18:913. doi: 10.1186/s12864-017-4280-4287
- Liu, H. G., Pan, Y. F., You, J., Wang, O. C., Huang, K. T., and Zhang, X. H. (2010). Expression of ABCG2 and its significance in colorectal cancer. *Asian Pac. J. Cancer Prev.* 11, 845–848.
- Lotfi, A. M., Spencer, R. J., Ilstrup, D. M., and Melton, L. J. III (1986). Colorectal polyps and the risk of subsequent carcinoma. *Mayo Clin. Proc.* 61, 337–343. doi: 10.1016/s0025-6196(12)61950-61958
- Luo, T., Wu, S., Shen, X., and Li, L. (2013). Network cluster analysis of protein-protein interaction network identified biomarker for early onset colorectal cancer. *Mol. Biol. Rep.* 40, 6561–6568. doi: 10.1007/s11033-013-2694-2690
- Mhaidat, N. M., Al-Husein, B. A., Alzoubi, K. H., Hatamleh, D. I., Khader, Y., Matalqah, S., et al. (2018). Knowledge and awareness of colorectal cancer early warning signs and risk factors among university students in Jordan. *J. Cancer Educ.* 33, 448–456. doi: 10.1007/s13187-016-1142-y
- Mudunuri, U., Che, A., Yi, M., and Stephens, R. M. (2009). bioDBnet: the biological database network. *Bioinformatics* 25, 555–556. doi: 10.1093/bioinformatics/btn654
- Nakagawa, S., Miyoshi, N., Ishii, H., Mimori, K., Tanaka, F., Sekimoto, M., et al. (2011). Expression of CLDN1 in colorectal cancer: a novel marker for prognosis. *Int. J. Oncol.* 39, 791–796. doi: 10.3892/ijo.2011.1102
- Quaife, S. L., Forbes, L. J., Ramirez, A. J., Brain, K. E., Donnelly, C., Simon, A. E., et al. (2014). Recognition of cancer warning signs and anticipated delay in help-seeking in a population sample of adults in the UK. *Br. J. Cancer* 110, 12–18. doi: 10.1038/bjc.2013.684
- Rahbari, N. N., Kedrin, D., Incio, J., Liu, H., Ho, W. W., Nia, H. T., et al. (2016). Anti-VEGF therapy induces ECM remodeling and mechanical barriers to therapy in colorectal cancer liver metastases. *Sci. Transl. Med.* 8:360ra135. doi: 10.1126/scitranslmed.aaf5219
- Rex, D. K., Kahi, C. J., Levin, B., Smith, R. A., Bond, J. H., Brooks, D., et al. (2006). Guidelines for colonoscopy surveillance after cancer resection: a consensus update by the American Cancer Society and the US multi-society task force on colorectal Cancer. *Gastroenterology* 130, 1865–1871. doi: 10.1053/j.gastro.2006.03.013
- Schottinger, J. E., Kanter, M. H., Litman, K. C., Lau, H., Schwartz, G. E., Brasfield, F. M., et al. (2016). Using literature review and structured hybrid electronic/manual mortality review to identify system-level improvement opportunities to reduce colorectal cancer mortality. *Jt Comm. J. Qual. Patient Saf.* 42, 303–310. doi: 10.1016/s1553-7250(16)42041-42046
- Soler, M., Estevez, M. C., Villar-Vazquez, R., Casal, J. I., and Lechuga, L. M. (2016). Label-free nanoplasmonic sensing of tumor-associate autoantibodies for early diagnosis of colorectal cancer. *Anal. Chim. Acta* 930, 31–38. doi: 10.1016/j.aca.2016.04.059
- Stanciu, C., Trifan, A., and Khder, S. A. (2007). Accuracy of colonoscopy in localizing colonic cancer. *Rev. Med. Chir. Soc. Med. Nat. Iasi.* 111, 39–43.
- Tanikawa, C., Zhang, Y. Z., Yamamoto, R., Tsuda, Y., Tanaka, M., Funauchi, Y., et al. (2017). The transcriptional landscape of p53 signalling pathway. *EBioMedicine* 20, 109–119. doi: 10.1016/j.ebiom.2017.05.017
- Toth, K., Sipos, F., Kalmar, A., Patai, A. V., Wichmann, B., Stoehr, R., et al. (2012). Detection of methylated SEPT9 in plasma is a reliable screening method for both left- and right-sided colon cancers. *PLoS One* 7:e46000. doi: 10.1371/journal.pone.0046000
- Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., et al. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* 28, 511–515. doi: 10.1038/nbt.1621
- Tumino, E., Parisi, G., Bertoni, M., Bertini, M., Metrangolo, S., Ierardi, E., et al. (2017). Use of robotic colonoscopy in patients with previous incomplete colonoscopy. *Eur. Rev. Med. Pharmacol. Sci.* 21, 819–826.
- Warren, J. D., Xiong, W., Bunker, A. M., Vaughn, C. P., Furtado, L. V., Roberts, W. L., et al. (2011). Septin 9 methylated DNA is a sensitive and specific blood test for colorectal cancer. *BMC Med.* 9:133. doi: 10.1186/1741-7015-9-133
- Yan, P., He, Y., Xie, K., Kong, S., and Zhao, W. (2018). In silico analyses for potential key genes associated with gastric cancer. *PeerJ* 6:e6092. doi: 10.7717/peerj.6092

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Guan, Zeng, Jiang, Xie, Cheng, Yan, He, Xu, Guan, Guo and Ao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.