Pathol. Oncol. Res. (2015) 21:719–725
DOI 10.1007/s12253-014-9880-9

RESEARCH

Identification of Key Genes Associated with Colorectal Cancer Based on the Transcriptional Network

Guoting Chen · Hengping Li · Xianping Niu · Guofeng Li · Ning Han · Xin Li · Guang Li · Yangzhou Liu · Guixin Sun · Yong Wang · Zengchun Li · Qinchuan Li

Received: 1 July 2014 / Accepted: 10 December 2014 / Published online: 23 January 2015

Abstract Colorectal cancer (CRC) is among the most lethal human cancers, but the mechanism of the cancer is still unclear enough. We aimed to explore the key genes in CRC progression. The gene expression profile (GSE4183) of CRC was obtained from Gene Expression Omnibus database which included 8 normal samples, 15 adenoma samples, 15 CRC samples and 15 inflammatory bowel disease (IBD) samples. Thereinto, 8 normal, 15 adenoma, and 15 CRC samples were chosen for our research. The differentially expressed genes (DEGs) in normal vs. adenoma, normal vs. CRC, and adenoma vs. CRC, were identified using the Wilcoxon test method in R respectively. The interactive network of DEGs was constructed to select the significant modules using the Pearson's correlation. Meanwhile, transcriptional network of DEGs was also constructed using the g: Profiler. Totally, 2, 741 DEGs in normal vs. adenoma, 1,484 DEGs in normal vs. CRC, and 396 DEGs in adenoma vs. CRC were identified. Moreover, function analysis of DEGs in each group showed FcR-mediated phagocytosis pathway in module 1, cardiac muscle contraction pathway in module 6, and Jak-STAT signaling pathway in module 19 were also enriched. Furthermore, MZF1 and AP2 were the transcription factor in module 6, with the target SP1, while SP1 was also a transcription in module 20. DEGs like *NCF1*, *AKT*, *SP1*, *AP2*, *MZF1*, and *TPM* might be used as specific biomarkers in CRC development. Therapy targeting on the functions of these key genes might provide novel perspective for CRC treatment.

Keywords Colorectal cancer · Differentially expressed genes · Transcriptional network · Module selection · Functional analysis

Guoting Chen and Hengping Li contributed equally to this work.

Highlights: 1. *NCF1* might be a biomarker for colon adenoma.
2. *AKT* and *TPM* contribute to the CRC development.
3. *MZF1* might act as a promoter in CRC progression.
4. *SP1*, regulated by *AP2*, might prevent the CRC progression.
5. Three significant pathways related to CRC were enriched.

G. Chen · X. Niu · G. Li · N. Han · X. Li · G. Li · Y. Liu · G. Sun · Y. Wang · Z. Li · Q. Li (✉)
Department of Emergency Surgery, East Hospital, Tongji University School of Medicine, No. 150, Jimo Road, Shanghai 200120, China
e-mail: li.qinchuan@163.com

H. Li
Department of General Surgery, Shandong Laigang Hoaspital, Laiwu, Shandong Province 271126, China

Introduction

Colorectal cancer (CRC) is a common rectal malignant tumor with the second mortality rate in the world [1]. Statistics indicate that CRC is the fourth most common malignancy in China while it is the third place in Europe [2]. The 5-year survival rate for early-stage CRC today stands at 90 %, and the overall death rate for CRC in the USA has fallen by 40 % since the 1970s [3]. Hence, in review of these, the early diagnosis and the enhancement of therapies became necessary in the discrimination between genetically and expressionally different tumors [4]. Although great progresses have been made in the prevention, detection and treatment for CRC. These mechanisms, as well as new approaches to prevention and therapy, are necessary for future clinical application.

Most of the CRC happens from normal to adenoma, and then to carcinoma [5]. Genes mutation accumulated in each progress during of cancelations. Gene expression analysis of colorectal biopsies using high density oligonucleotide microarrays may help to detect such gene expression patterns that would establish the basis for new molecule biological diagnostic methods [6]. In contrast to the gene-level analyses, module-based approaches use gene modules as the basic building blocks for analysis. Modules can be defined in a knowledge driven fashion based on existing knowledge on

pathways, biological process, and protein complexes [7]. Some studies have demonstrated that several genes could act as the biomarker for CRC treatment. For instance, NLK (nemo-like kinase) is a tumor suppressor in CRC development [8], and overexpression of SATB1 (SATB homeobox 1) promotes the CRC cell growth and metastasis [9]. Also, mutation of KRAS (Kirsten rat sarcoma viral oncogene homolog) is involved in CRC metastasis [10] and anti-EGFR1 (epidermal growth factor receptor) is allowed to use if the tumor carries wild RAS in CRC [11]. However, mechanism of CRC remains largely known.

Using the gene expression profile of GSE4183 [12], Nam et al. proved the significant pathways in early-onset CRC in terms of differentiation and dedifferentiation [13]. Heijink et al. identified several novel therapeutic targets for CRC chemoprevention [14]. In this paper, we used the microarray bioinformatics to screen the differentially expressed genes (DEGs) in adenoma and CRC samples compared to the normal samples using the same gene expression profile. Comprehensive bioinformatics was used to build the interactive network and to analyze the significant modules information to provide deeper insight into the biological mechanisms of CRC. This approach was benefit for predicting the key genes that are most likely associated with CRC and identify the molecular mechanisms of CRC.

Materials and Methods

Microarray Data Resource

The microarray and other forms of high flux data produced by the scientific community were archived and freely released from the Gene Expression Omnibus (GEO) database of National Center for Biotechnology Information (NCBI) [15], which is the biggest completely public storage. The gene expression profile (GSE4183) [12] was downloaded from Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/geo) based on the platform of GLP570 Affymetrix Human Genome U133 plus 2.0 Array (HG-U133_Plus_2). Total 53 chips were available for further analysis, including 8 healthy normal samples, 15 patients with adenoma, 15 with CRC and 15 with inflammatory bowel disease (IBD). Thereinto, we chose the first three kinds samples, including 8 normal samples, 15 adenoma samples, and 15 CRC samples for our research. The total 38 samples were separated into three groups, normal vs. adenoma, normal vs. CRC, and CRC vs. adenoma.

Data Preprocessing and DEGs Screening

The original data were performed normalization by the robust multiarray average (RMA) algorithm [16] in R. Wilcoxon test method [17] in R language was used to select the differentially expressed genes (DEGs) of the three groups respectively. The normalized p value and FDR (false discovery rate)<0.01 were chosen as the cut-off criterion.

Interactive Network Construction of DEGs

The Pearson correlation coefficient [18] was used to calculate the correlation coefficients of all the selected DEGs, the formula was given as follows:

$$r = \frac{\mathrm{cov}(x, y)}{\sqrt{D(x)^*}\sqrt{D(y)}}$$

Whereas, x, y stand for the expression value vector, each value in the expression value vector stands for the gene expression value of one DEG. The node in the interactive network stands for the DEG, edge represents the communication degree. DEGs with the correlation coefficient r≥0.75 were used to construct the interactive network.

Module Selection and Function Annotation

Based on the interactive network, Markov's algorithms [19] was used to select the significant modules via an online software (http://biit.cs.ut.ee/graphweb/). All parameters used were the default parameters. g: Profiler (http://biit.cs.ut.ee/gprofiler/gorth.cgi), a web-based toolset for functional profiling of gene lists from large-scale experiments [20], is a functional annotation software that similar to the Database for Annotation, Visualization and Integrated Discovery (DAVID) [21], including the gene ontology (GO) [22] function, kyoto encyclopedia of genes and genomes (KEGG) pathway [23], and enrichment analysis of TRANSFAC transcription [24] combine region. The integrated resources of g: Profiler was more than that in DAVID, thus, more significant enrichment categories would be collected by g: Profiler. The selected DEGs in each group were put into the online biological tool of g: Profile to detect the biological pathways and TRANSFAC combine region to predict TFs binding site.

Results

DEG Screening

Totally, 2,741 DEGs in normal vs. adenoma, 1,484 DEGs in normal vs. CRC, and 396 DEGs in CRC vs. adenoma were screened (Fig. 1).

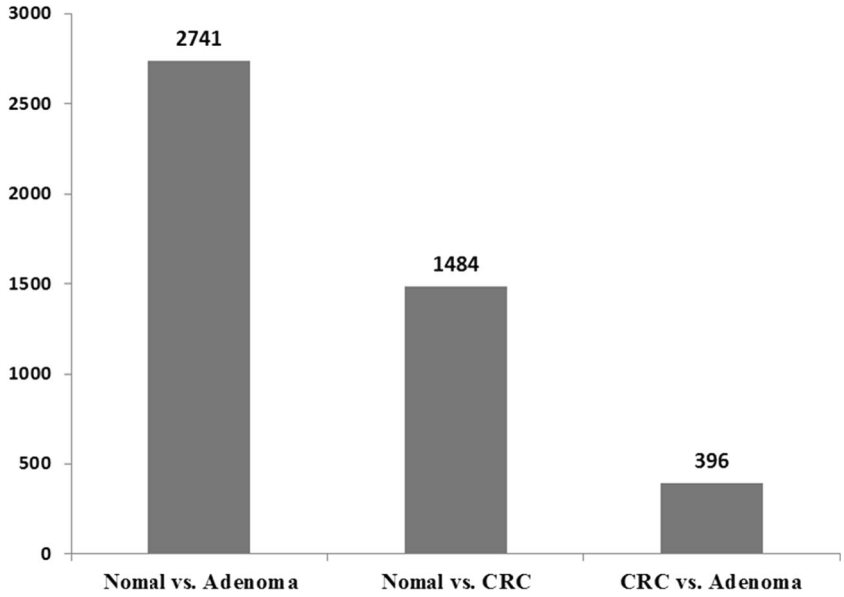

Fig. 1 DEGs in normal vs. adenoma, normal vs. CRC, and CRC vs. adenoma

Interactive Network of DEGs

The Pearson correlation coefficient was used to calculate all the 3,445 DEGs. Consequently, 3,183 DEGs were selected to construct the interactive network with the $r \geq 0.75$. Gene products in same module often have some or similar function, they work together to carry out one bio-function. Therefore, we used software Cytoscape to visualized network and its plugin of Cytoscape-Mcode to make module division, another plugin of Cytoscape-Bingo to annotate module function based on Markov algorithm, and got the significant function of each module. Total 20 modules were identified, and the genes of each module were in the file module_members.xls.

Significant Modules and Functional Enrichment Analysis

We selected 20 significant modules using the Wilcoxon test method in R. After being annotated in the online g: Profiler, 3 significant pathways associated with 3 modules were enriched. For instance, FcR-mediated phagocytosis related to module 1 (Fig. 2), cardiac muscle contraction related to module 6, and Jak-STAT signaling pathway related to module 19 (Fig. 3).

We could see from the Fig. 2 that LAT (linker for activation of T cells) was the common DEG in three groups, *WASP* (Wiskott-Aldrich syndrome protein) was the common DEG between normal vs. adenoma group and normal vs. CRC group, while *CD45* and *DOCK180* (dedicator of cytokinesis 180) were the common genes bwteen normal vs. adenoma and CRC vs. adenoma group. DEGs like *PKC* (protein kinase C), *PAP*, *PLC*γ (phospholipase C), and *NCF1* (neutrophil cytosolic factor 1) were the genes in normal vs. adenoma group.

Besides, NCX, TPM (tropomyosin) and Cyto were the common DEGs in normal vs. adenoma and adenoma vs. CRC, suggested that this pathway might play an important role in the CRC biogenesis).

In addition, Fig. 3 showed that IL2/3 and STAT were the common DEGs in the normal vs. adenoma and normal vs. CRC, while AKT was the common gene in normal vs. adenoma and adenoma vs. CRC group. However, IL6 was the common gene in the three groups.

Transcriptional Regulatory Analysis of DEGs

In order to investigate the relationships between the DEGs and their transcription factor and targets, the dendrogram was constructed to show the relationships between transcription factor and targets (Fig. 4). For example, most of the DEGs in module 6 were regulated by the transcription factor AP-2 (adaptor-related protein complex 2). Thereinto, SP1 (specificity protein 1) was the target gene of AP-2 while it was also a transcription factor, regulating the most DEGs in module 20. Besides, MZF1 (myeloid zinc finger 1) was also a transcription factor in module 6, with the target AP-2. These regulatory relationships were mostly associated with bioprocess, such as the regulation of cell growth, tumorigenesis, cell proliferation, cell differentiation and cell cycle regulation.

Discussion

CRC is one of the second malignant tumors in the world, and it brings about huge economic burden on both society and the patients' family [1]. However, the mechanism of CRC has not been fully discussed. In our work, 2,741 DEGs in normal vs. adenoma, 1,484 DEGs in normal vs. CRC, and 396 DEGs in adenoma vs. CRC were identified.

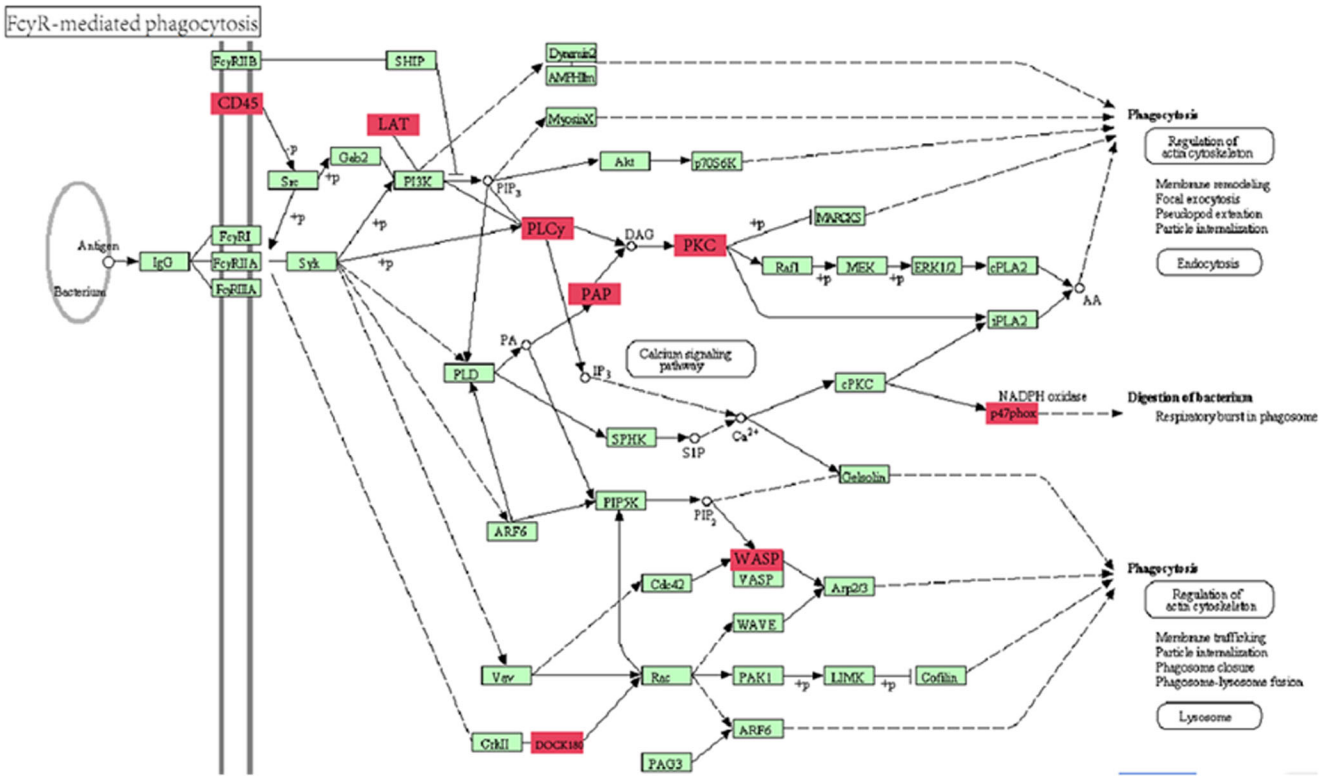

Fig. 2 FcR-mediated phagocytosis pathway in module 1. Red color stands for the selected DEGs

Besides, FcR-mediated phagocytosis pathway in module 1, cardiac muscle contraction pathway in module 6, and Jak-STAT signaling pathway in module 19 were also enriched. Also, MZF1 and AP2 were the transcription factor in module 6, with the target SP1, while SP1 was also a transcription factor in module 20.

Our findings showed that *NCF1* that involved in FcR-mediated phagocytosis pathway, was the DEG in normal vs. adenoma group instead of in normal vs. CRC group. CD11b is important to affect the FcR-mediated immunity to melanoma [25]. NCF1 is a cytosolic subunit of neutrophil NADPH oxidase that has the ability to produce superoxide anion

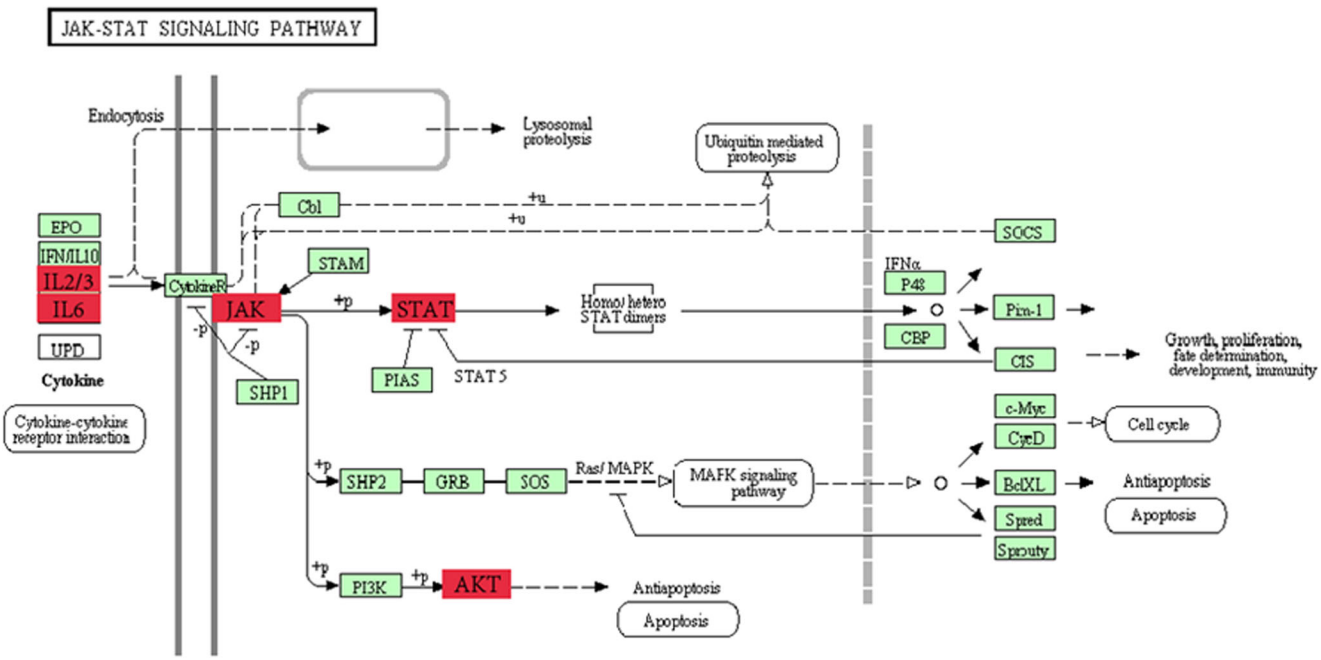

Fig. 3 JAK-STAT signaling pathway in module 19. Red color stands for the selected DEGs

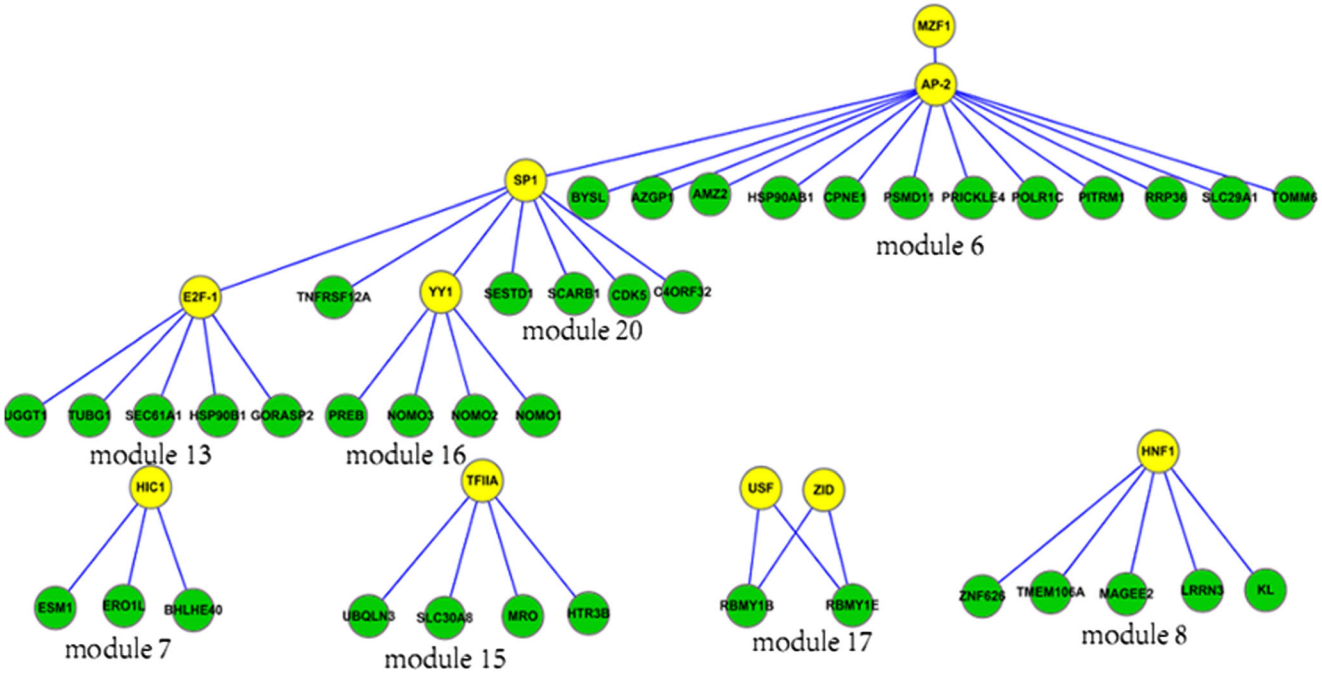

Fig. 4 Transcriptional network of modules. Yellow represents the transcription factor, while green represents the target genes

[26]. Role of NCF1 in CRC has not been fully discussed. However, Suraweera et al. proved that there was no relationship between NCF1 and inflammatory bowel disease [27]. On the basis of our finding, we speculated that *NCF1* gene may be crucial via the FcR-mediated phagocytosis pathway in colon adenoma.

TPM (encoded by *TPM* gene) is a member of the tropomyosin family of highly conserved, widely distributed actin-binding proteins involved in the contractile system of striated and smooth muscles [28]. Downregulation of TPMβ was associated with the early CRC development and might act as a potential biomarker for CRC detection [29]. Also, TPM1 might be an alternative path of CRC development [30]. Our work showed that *TPM* appeared in the cardiac muscle contraction pathway in module 6, indicating that *TPM* might be crucial for the CRC development via the cardiac muscle contraction pathway.

AKT is catalytically inactive in serum-starved primary and immortalized fibroblasts [31]. P-AKT is the biomarker for CRC possible therapeutic target identification [32]. Also, AKT promotes the CRC tumor cell survival and metastasis, then enhances the CRC progression [33]. Meanwhile, Jak-STAT signaling pathway is associated with the transmitting information from extracellular polypeptide signals to target gene promoters in the nucleus [34]. Inhibition of Jak-STAT signaling pathway may prevent the cell proliferation and induce the cell apoptosis in hodgkin lymphoma [35]. Based on our work, we speculated that *AKT* may be a promoter in CRC progression via involving in Jak-STAT signaling pathway.

MZF1 (encoded by *MZF1* gene) is a cancer related transcription factor that belongs to SCAN domain family transcription factors family [36, 37]. MZF1 could induce the cell migration, invasion and metastasis in CRC [38]. Deng et al. reported that MZF1 could promote the CRC cell proliferation via activating the transcription factor p55PIK [39]. Our data showed that MZF1 in module 6 could regulate AP2 in CRC, suggesting that it might important in CRC progression.

Meanwhile, our work showed that the transcription factor SP1 in module 20 was the target for transcription factor AP2 in module 6, suggesting that these two factors might be important in CRC. AP2 (encoded by *AP2* gene) is a heterotetramer consisting of two large adaptins (α and β), facilitates clathrin-mediated endocytosis [40], while SP1 (encoded by *SP1* gene) is a zinc finger transcription factor that involves in many cellular processes, including cell differentiation, cell growth, apoptosis, and response to DNA damage [41]. It has been proved that p53 mutation is one of the microsatellite instability status of CRC [42]. Stabach et al. proved that overexpression of AP2α could alter the transcriptional activity and stability of p53 in tumor tissue [43]. Besides, the nuclear localization of SP1 and its DNA binding activity could be suppressed by the inhibition of O-GlcNAc transferase, and then functioned in the cell death of pancreatic cancer [44]. Also, SP1 was a key modulator of p53-mediated apoptosis in cell apoptosis [45]. Thus, SP1 may be a key factor for CRC progression. Besides, Zhao et al. proved that SP1 suppressed the colon cancer stem cell growth and induced the cell apoptosis in the nude mouse, and SP1 might serve as a novel therapeutic strategy for colon cancer treatment [46].

Based on our data, we speculated that *SP1* might play a key role via regulated by *AP2* in CRC treatment.

In conclusion, our study explored the potential roles of several pathways (FcR-mediated phagocytosis pathway, cardiac muscle contraction pathway, and Jak-STAT signaling pathway) and key genes in CRC development. DEGs like *NCF1*, *AKT*, *TPM*, *AP2*, *SP1*, and *MZF1*, were important in PC development, but their roles are different. Thus, the key genes may serve as potential biomarkers for CRC detection and might provide novel perspective for CRC treatment. Further research is still needed to verify the results in clinical trial.

Acknowledgments This study was supported by Shanghai medical key subject construction project (NO.ZK2012A28), Key National Clinical Discipline construction project (2013)

Conflict of Interest The Authors declare that they have no conflicts of interest to disclose.

References

1. Rebecca SM, Deepa Naishadham MA, Ahmedin Jemal DVM (2013) Cancer statistics, 2013. CA Cancer J Clin 63:11–30
2. Chen W-Q, Zeng H-M, Zheng R-S et al (2012) Cancer incidence and mortality in China, 2007. Chin J Cancer Res 24:1–8
3. Goel A, Boland CR (2012) Epigenetics of colorectal cancer. Gastroenterology 143:1442–1460
4. Garagnani P, Pirazzini C, Franceschi C (2013) Colorectal cancer microenvironment: among nutrition, gut microbiota, inflammation and epigenetics. Curr Pharm Des 19:765–778
5. Vinci S, Gelmini S, Mancini I et al (2013) Genetic and epigenetic factors in regulation of microRNA in colorectal cancers. Methods 59: 138–146
6. Wilkinson J (2012) Epigenome enhancer profiling reveals signature for colon cancer. Epigenomics 4(3):248
7. Schweiger MR, Hussong M, Rohr C et al (2013) Genomics and epigenomics of colorectal cancer. Wiley Interdiscip Rev Syst Biol Med 5:205–219
8. Han Y, Kuang Y, Xue X et al (2014) NLK, a novel target of miR-199a-3p, functions as a tumor suppressor in colorectal cancer. Biomed Pharmacother 9:003
9. Zhang Y, Tian X, Ji H et al. (2014) Expression of SATB1 Promotes the Growth and Metastasis of Colorectal Cancer. PloS one 9
10. Van Cutsem E, Köhne C-H, Láng I et al (2011) Cetuximab plus irinotecan, fluorouracil, and leucovorin as first-line treatment for metastatic colorectal cancer: updated analysis of overall survival according to tumor KRAS and BRAF mutation status. JCO 2010(2033):5091
11. Misale S, Yaeger R, Hobor S et al. (2012) Emergence of KRAS mutations and acquired resistance to anti-EGFR therapy in colorectal cancer. Nature
12. Galamb O, Sipos F, Solymosi N et al (2008) Diagnostic mRNA expression patterns of inflamed, benign, and malignant colorectal biopsy specimen and their correlation with peripheral blood results. Cancer Epidemiol Biomarkers Prev 17:2835–2845
13. Nam S, Park T (2012) Pathway-based evaluation in early onset colorectal cancer suggests focal adhesion and immunosuppression along with Epithelial-Mesenchymal transition. PLoS ONE 7:e31685
14. Heijink D, Fehrmann R, de Vries E et al (2011) A bioinformatical and functional approach to identify novel strategies for chemoprevention of colorectal cancer. Oncogene 30:2026–2036
15. Barrett T, Troup DB, Wilhite SE et al (2007) NCBI GEO: mining tens of millions of expression profiles—database and tools update. Nucleic Acids Res 35:D760–D765
16. Irizarry RA, Hobbs B, Collin F et al (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics 4:249–264
17. Gehan EA (1965) A generalized Wilcoxon test for comparing arbitrarily singly-censored samples. Biometrika 52:203–223
18. Adler J, Parmryd I (2010) Quantifying colocalization by correlation: the pearson correlation coefficient is superior to the Mander's overlap coefficient. J Int Soc Anal Cytol 77:733–742. doi:10.1002/cyto.a.20896
19. Collobert R, Bengio S, Mariéthoz J (2002) Torch: a modular machine learning software library. Technical Report IDIAP-RR 02-46, IDIAP
20. Reimand J, Kull M, Peterson H et al (2007) g: Profiler—a web-based toolset for functional profiling of gene lists from large-scale experiments. Nucleic Acids Res 35:W193–W200
21. Alvord G, Roayaei J, Stephens R et al (2007) The DAVID gene functional classification tool: a novel biological module-centric algorithm to functionally analyze large gene lists. Genome Biol 8:R183
22. Ashburner M, Ball CA, Blake JA et al (2000) Gene ontology: tool for the unification of biology. Nat Genet 25:25
23. Ogata H, Goto S, Sato K et al (1999) KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 27:29–34
24. Matys V, Fricke E, Geffers R et al (2003) TRANSFAC: transcriptional regulation, from patterns to profiles. Nucleic Acids Res 31: 374–378
25. van Spriel AB, van Ojik HH, Bakker A et al (2003) Mac-1 (CD11b/CD18) is crucial for effective Fc receptor–mediated immunity to melanoma. Blood 101:253–258
26. Roos D, de Boer M, Koker MY et al (2006) Chronic granulomatous disease caused by mutations other than the common GT deletion in NCF1, the gene encoding the p47phox component of the phagocyte NADPH oxidase. Hum Mutat 27:1218–1229
27. Suraweera N, Zampeli E, Rogers P et al (2004) NCF1 (p47phox) and ncf1 pseudogenes are not associated with inflammatory bowel disease. Inflamm Bowel Dis 10:758–762
28. Parry D, Squire J (1973) Structural role of tropomyosin in muscle regulation: analysis of the x-ray diffraction patterns from relaxed and contracting muscles. J Mol Biol 75:33–55
29. Luo YX, Cui J, Wang L et al (2009) Identification of cancer-associated proteins by proteomics and downregulation of beta-tropomyosin expression in colorectal adenoma and cancer. Proteomics Clin Appl 3:1397–1406
30. Mlakar V, Berginc G, Volavsek M et al (2009) Presence of activating KRAS mutations correlates significantly with expression of tumour suppressor genes DCN and TPM1 in colorectal cancer. BMC Cancer 9:1471–2407
31. Franke TF, Yang S-I, Chan TO et al (1995) The protein kinase encoded by theAktproto-oncogene is a target of the PDGF-activated phosphatidylinositol 3-kinase. Cell 81:727–736
32. Baba Y, Nosho K, Shima K et al (2011) Phosphorylated AKT expression is associated with PIK3CA mutation, low stage, and favorable outcome in 717 colorectal cancers. Cancer 117:1399–1408
33. Agarwal E, Brattain MG, Chowdhury S (2013) Cell survival and metastasis regulation by Akt signaling in colorectal cancer. Cell Signal 25:1711–1719
34. PIM JS JAK-STAT signaling pathway.
35. Diaz T, Navarro A, Ferrer G et al (2011) Lestaurtinib inhibition of the Jak/STAT signaling pathway in hodgkin lymphoma inhibits proliferation and induces apoptosis. PLoS ONE 6:0018856

36. Yue CH, Chiu YW, Tung JN et al (2012) Expression of protein kinase C alpha and the MZF-1 and Elk-1 transcription factors in human breast cancer cells. Chin J Physiol 55:31–36
37. Hsieh YH, Wu TT, Huang CY et al (2007) Suppression of tumorigenicity of human hepatocellular carcinoma cells by antisense oligonucleotide MZF-1. Chin J Physiol 50:9–15
38. Mudduluru G, Vajkoczy P, Allgayer H (2010) Myeloid zinc finger 1 induces migration, invasion, and in vivo metastasis through Axl gene expression in solid cancer. Mol Cancer Res 8:159–169
39. Deng Y, Wang J, Wang G et al (2013) p55PIK transcriptionally activated by MZF1 promotes colorectal cancer cell proliferation. Biomed Res Int 868131:23
40. Raman D, Sai J, Hawkins O et al (2014) Adaptor protein2 (AP2) orchestrates CXCR2-mediated cell migration. Traffic 15:451–469
41. Kadonaga JT, Carner KR, Masiarz FR et al (1987) Isolation of cDNA encoding transcription factor Sp1 and functional analysis of the DNA binding domain. Cell 51:1079–1090
42. Lam AK, Chan SS (2014) Leung M Synchronous colorectal cancer: clinical, pathological and molecular implications. World J Gastroenterol 20(22):6815–6820
43. Stabach PR, Thiyagarajan MM, Woodfield GW et al (2006) AP2alpha alters the transcriptional activity and stability of p53. Oncogene 25:2148–2159
44. Banerjee S, Sangwan V, McGinn O et al (2013) Triptolide-induced cell death in pancreatic cancer is mediated by O-GlcNAc modification of transcription factor Sp1. J Biol Chem 288:33927–33938
45. Li H, Zhang Y, Strose A et al (2014) Integrated high-throughput analysis identifies Sp1 as a crucial determinant of p53-mediated apoptosis. Cell Death Differ 27:69
46. Zhao Y, Zhang W, Guo Z et al (2013) Inhibition of the transcription factor Sp1 suppresses colon cancer stem cell growth and induces apoptosis in vitro and in nude mouse xenografts. Oncol Rep 30: 1782–1792