Identification of key genes in colorectal cancer using random walk with restart

XIAOFENG CUI¹, KEXIN SHEN¹, ZHONGSHI XIE¹, TONGJUN LIU² and HAISHAN ZHANG¹

¹Department of Gastrointestinal Colorectal and Anal Surgery, China-Japan Union Hospital, Jilin University, Changchun, Jilin 130033; ²Department of General Surgery, Jilin University, Changchun, Jilin 130022, P.R. China

Received November 18, 2015; Accepted November 15, 2016

DOI: 10.3892/mmr.2016.6058

Abstract. As the most common type of cancer and the second leading cause of cancer-associated mortality, colorectal cancer (CRC) has received increasing attention. The aim of the present study was to investigate the mechanisms of CRC by analyzing the microarray dataset, GSE32323. The GSE32323 dataset was downloaded from the Gene Expression Omnibus, and included 17 pairs of matched cancer and normal colorectal tissue samples. The differentially expressed genes (DEGs) were screened using the Linear Models for Microarray Data package and a search of CRC genes, also denoted as seed genes, was performed using the Online Mendelian Inheritance in Man database. Subsequently, the protein-protein interaction (PPI) network was downloaded from the Search Tool for the Retrieval of Interacting Genes database and the sub-network (CRC.PPI) of the DEGs and seed genes were obtained. In addition, the top 50 nodes with highest affinity scores in the CRC.PPI were identified using random walk with restart analysis. The potential functions of the DEGs included in the top 50 nodes were analyzed using the Database for Annotation, Visualization and Integrated Discovery online tool. Using the Drug Gene Interaction database, drug-gene interaction analysis was performed to identify antineoplastic drug interacts with genes. A total of 1,640 DEGs between the CRC and normal samples were screened. The obtained seed genes included cyclin D1 (CCND1) and aurora kinase A (AURKA). The enriched functions for the 31 DEGs in the PPI network of the top 50 nodes were predominantly associated with cell cycle. The DEGs may function in CRC by interacting with other genes in the PPI network of the top 50 nodes, for example, DEP domain-containing MTOR-interacting protein (DEPTOR)-CCND1, AURKA-breast carcinoma ampli-fied sequence-1 (BCAS1), CCND1-BCAS1, CCND1-neural precursor cell expressed developmentally downregulated 9 (NEDD9) and CCND1-mitogen-activated protein kinase kinase 2 (MAP2K2). Only three DEGs (CCND1, AURKA and DEPTOR) had interactions with their corresponding anti-neoplastic drugs. Taken together, DEPTOR, AURKA, CCND1, BCAS1, NEDD9 and MAP2K2 may act in CRC.

Introduction

As a type cancer occurring in the rectum or colon, colorectal cancer (CRC) is also termed rectal cancer or colon cancer (1). Patients with CRC often have symptoms of persistent tiredness, altered bowel movements, blood in stools and weight loss (2). With a mortality rate of ≥200,000 per year in Europe (3), CRC is the most common type of cancer and ranks as at second leading cause of cancer-associated mortality (4).

Several studies have been performed to investigate the effects of genes on CRC. For example, cyclooxygenase-2, which is mediated by vascular endothelial growth factor, can function in the tumor angiogenesis of CRC (5). In mutations of the catalytic subunit α of phosphatidylinositol-3-kinase (PIK3CA), all mutations are functionally active in colon cancer, therefore, it may be associated with carcinogenesis (6). Mutations in Kirsten-ras, adenomatous polyposis coli and p53 induce the transition from healthy colonic epithelia to CRC (7). The epigenetic loss of function of secreted frizzled-related protein may provide constitutive WNT signaling, which is essential for downstream mutations of complement in the progression of CRC (8). As an epigenetic alteration, loss of imprinting can affect the insulin-like growth factor II gene and may be a useful predictor for the risk of CRC (9).

In 2012, Khamas et al (10) performed oligonucleotide microarray analysis to identify the differentially expressed genes (DEGs) between CRC samples and normal samples. Using the same data as that used by Khamas et al (10), the present study aimed to further identify the DEGs and CRC genes. A sub-network of the DEGs and seed genes were obtained from the downloaded protein-protein interaction...
(PPI) network, termed the CRC.PPI. In addition, random walk with restart (RWR) analysis was performed to identify the 50 nodes with the top affinity scores in the CRC.PPI. The potential functions of the DEGs included in the 50 key nodes were then analyzed using Gene Ontology (GO) and pathway enrichment analyses. In addition, drug-gene interaction analysis was performed to identify interactions between antineoplastic drugs and genes.

Materials and methods

Microarray data. The expression profile of GSE32323, deposited by Khamas et al (10) was downloaded from the Gene Expression Omnibus (www.ncbi.nlm.nih.gov/geo/), which was based on the platform of the GPL570 (HG-U133_Plus_2) Affymetrix Human Genome U133 Plus 2.0 array (Affymetrix, Inc., Santa Clara, CA, USA). The GSE32323 profile included a collection of 17 pairs of matched cancer-normal colorectal tissue samples. The CRC samples were obtained from the American Type Culture Collection (Manassas, VA, USA) and the Cell Resource Center for Biomedical Research of Tohoku University (Sendai, Japan). The cells were cultured in media obtained from Sigma-Aldrich; Merck Millipore (Darmstadt, Germany) or Gibco; Thermo Fisher Scientific, Inc. (Waltham, MA, USA), supplemented with 100 µg/ml streptomycin (Invitrogen); Thermo Fisher Scientific, Inc.), 100 U/ml penicillin and 10% heat-inactivated fetal bovine serum (Nichirei Biosciences, Tokyo, Japan).

Identification of DEGs and CRC genes. Following downloading of the GSE32323 profile, the microarray data were preprocessed using the Affy package (11) in Bioconductor (www.ncbi.nlm.nih.gov/geo/) and the Affy probe annotation file from Brain Array Lab (brainarray.mhri.med.umich.edu/brainarray/default.asp). In brief, the process included background correction, quantile normalization, probe summarization, and log2 conversion. The mean values of probes mapped with the same gene were obtained as ultimate gene expression values. Subsequently, the Linear Models for Microarray data (LIMMA; limma.html) (12) was used to screen the DEGs of the CRC samples, compared with the normal samples. An adjusted P-value of P<0.05 and |logfold-change|>1 were used as the cut-off criteria. The CRC genes were identified from a search using the Online Mendelian Inheritance in Man (OMIM) database (www.ncbi.nlm.nih.gov/omim) (13) and were then annotated as seed genes.

PPI network construction and RWR analysis. The PPI network was downloaded from the Search Tool for the Retrieval of Interacting Genes database (string-db.org/) (14). The sub-network of the DEGs and seed genes was then obtained and termed the CRC.PPI. The RWR algorithm was originally used for image segmentation (15). With a restarting probability (r) of 0.9, the R package dnet (16) was used to perform RWR analysis to identify the top 50 nodes with the highest affinity scores in the CRC.PPI. The affinity score referred to the proximity between two nodes and the parameter ‘r’ is the probability of moving to seed nodes.

Functional and pathway enrichment analysis. The Database for Annotation, Visualization and Integrated Discovery (DAVID; david.abcc.ncifcrf.gov/) is a tool for agglomerating a large quantity of gene/protein symbols into gene clusters (17). Gene Ontology (GO) analysis is used to predict the potential functions of large-scale transcriptomic data or genomic data (18). The Kyoto Encyclopedia of Genes and Genomes database is a base for functional analysis combing with networks of genes and other molecules (19). Using the DAVID online tool, GO functional and pathway enrichment analyses were performed for the DEGs in the PPI network of the top 50 nodes. P<0.05 was used as the cut-off criterion.

Drug-gene interaction analysis. The Drug-Gene Interaction database (dgidb.genome.wustl.edu) (20) was used to identify interactions between antineoplastic drugs and genes. The actions of the drugs inhibit genes expressed at high levels and promote genes expressed at low levels.

Results and Discussion

DEG analysis. Compared with the normal samples, there were 1,640 DEGs screened in the CRC samples, including 850 upregulated genes and 790 downregulated genes. The CRC genes or seed genes from the OMIM database are listed in Table I, including cyclin D1 (CCND1) and aurora kinase A (AURKA).

PPI network analysis. Following downloading of the PPI network and obtaining the CRC.PPI of the DEGs and seed genes, RWR analysis was performed and nodes with the top 50 affinity scores, which included DEP domain-containing MTOR-interacting protein (DEPTOR), breast carcinoma amplified sequence-1 (BCAS1), neural precursor cell expressed developmentally downregulated 9 (NEDD9), mitogen-activated protein kinase kinase 2, (MAP2K2) in the CRC.PPI were identified. The PPI network of the top 50 nodes with the highest affinity scores, which comprised 14 upregulated genes, 17 downregulated genes and 19 non-DEGs, had 224 interactions. These interactions included DEPTOR-CCND1, AURKA-BCAS1, CCND1-BCAS1, CCND1-NEDD9 and CCND1-MAP2K2 (Fig. 1). The DEGs included in the top 50 nodes are listed in Table II. The enriched functions of the 31 DEGs in the PPI network of the top 50 nodes were primarily associated with cell cycle, including cell cycle phase (P=0.005651), M phase (P=0.019763) and intracellular signaling cascade (P=0.019784), as shown in Table III. The enriched pathway for the 31 DEGs in the PPI network of the top 50 nodes included cytokine-cytokine receptor interaction (P=0.04602; Table III).

Drug-gene interaction analysis. The interactions between antineoplastic drugs and the 31 DEGs in the CRC.PPI network were screened. Only three DEGs (CCND1, AURKA and DEPTOR), had interactions with their corresponding antineoplastic drugs.

For the analysis of drug-gene interactions in the present study, a total of 1,640 DEGs were screened in the CRC samples, compared with normal samples, which included 850 upregulated genes and 790 downregulated genes. The
Table I. Seed genes from an Online Mendelian Inheritance in Man database search.

<table>
<thead>
<tr>
<th>Gene/locus</th>
<th>ID</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>PLA2G2A</td>
<td>5320</td>
<td>Phospholipase A2, group IIA (platelets, synovial fluid)</td>
</tr>
<tr>
<td>NRAS</td>
<td>4893</td>
<td>Neuroblastoma RAS viral (v-ras) oncogene homolog</td>
</tr>
<tr>
<td>ODC1</td>
<td>4953</td>
<td>Ornithine decarboxylase 1</td>
</tr>
<tr>
<td>CTNNB1</td>
<td>1499</td>
<td>Catenin (cadherin-associated protein), β1, 88 kDa</td>
</tr>
<tr>
<td>PIK3CA</td>
<td>5290</td>
<td>Phosphatidylinositol-4,5-bisphosphate 3-kinase, catalytic subunit α</td>
</tr>
<tr>
<td>FGFR3</td>
<td>2261</td>
<td>Fibroblast growth factor receptor 3</td>
</tr>
<tr>
<td>TLR2</td>
<td>7097</td>
<td>Roll-like receptor 2</td>
</tr>
<tr>
<td>APC</td>
<td>324</td>
<td>Adenomatous polyposis coli</td>
</tr>
<tr>
<td>MCC</td>
<td>4163</td>
<td>Mutated in colorectal cancers</td>
</tr>
<tr>
<td>PTEN12</td>
<td>5782</td>
<td>Protein tyrosine phosphatase, non-receptor type 12</td>
</tr>
<tr>
<td>PDGFRL</td>
<td>5157</td>
<td>Platelet-derived growth factor receptor-like</td>
</tr>
<tr>
<td>RAD54B</td>
<td>25788</td>
<td>RAD54 homolog B (S. cerevisiae)</td>
</tr>
<tr>
<td>TLR4</td>
<td>7099</td>
<td>Toll-like receptor 4</td>
</tr>
<tr>
<td>PTENJ</td>
<td>5795</td>
<td>Protein tyrosine phosphatase, receptor type, J</td>
</tr>
<tr>
<td>CCND1</td>
<td>595</td>
<td>Cyclin D1</td>
</tr>
<tr>
<td>MLH3</td>
<td>27030</td>
<td>MutL homolog 3</td>
</tr>
<tr>
<td>AKT1</td>
<td>207</td>
<td>V-akt murine thymoma viral oncogene homolog 1</td>
</tr>
<tr>
<td>BUB1B</td>
<td>701</td>
<td>BUB1 mitotic checkpoint serine/threonine kinase B</td>
</tr>
<tr>
<td>TP53</td>
<td>7157</td>
<td>Tumor protein p53</td>
</tr>
<tr>
<td>FLCN</td>
<td>201163</td>
<td>Folliculin</td>
</tr>
<tr>
<td>AXIN2</td>
<td>8313</td>
<td>Axin 2</td>
</tr>
<tr>
<td>DCC</td>
<td>1630</td>
<td>Deleted in colorectal carcinoma</td>
</tr>
<tr>
<td>BAX</td>
<td>581</td>
<td>BCL2-associated X protein</td>
</tr>
<tr>
<td>AURKA</td>
<td>6790</td>
<td>Aurora kinase A</td>
</tr>
<tr>
<td>EP300</td>
<td>2033</td>
<td>E1A binding protein p300</td>
</tr>
</tbody>
</table>

Figure 1. Protein-protein interaction network of the top 50 nodes with the highest affinity scores.
obtained seed genes included \textit{CCND1} and \textit{AURKA}. The enriched functions for the 31 DEGs in the PPI network of the top 50 nodes were primarily associated with cell cycle. According to the drug-gene interaction analysis, only three DEGs (\textit{CCND1}, \textit{AURKA} and \textit{DEPTOR}) had interactions with their corresponding antineoplastic drugs. \textit{DEPTOR}, which is negatively correlated with tumor progression and the activity of mammalian target of rapamycin complex 1 (mTORC1) in CRC, may function as a marker for prognosis and treatment of CRC (21). The PI3K/Akt/mTOR signaling pathway is one of the primary mechanisms involved in maintaining tumor progression and metastasis, and mTOR activation/inhibition functions in CRC (22). It has been reported that \textit{DEPTOR} may be closely correlated with CRC. The variants in \textit{CCND1} may have potential associations with familial CRC (23), and the \textit{CCND1} A870G polymorphism may increase the risk of CRC (24). The epidermal growth factor A61G and \textit{CCND1} A870G polymorphisms may be valuable molecular biomarkers for the prognosis of patients with CRC treated with the single-agent, cetuximab (25). Thus, the expression levels of \textit{CCND1} may be associated with CRC. The PPI network of the top 50 nodes in the present study also showed \textit{CCND1} had an interaction with \textit{DEPTOR}, indicating that \textit{CCND1} may be involved in CRC by mediating \textit{DEPTOR}.

\textit{AURKA}, which is located on chromosome 20q, is involved in adenoma-to-carcinoma progression and may function as an indicator of poor prognosis (27-33). \textit{NABC1}, also known as \textit{BCAS1}, and its variant, \textit{NABC15B}, are downregulated in CRC (34). These findings suggest that \textit{AURKA} and \textit{BCAS1} may be associated with CRC. In the PPI network of the top 50 nodes in the present study, it was also found that \textit{BCAS1} interacted with \textit{AURKA}, suggesting that \textit{BCAS1} may also be involved in CRC through regulating \textit{AURKA}.

\begin{table}[h]
\centering
\caption{31 differentially expressed genes in the protein-protein interaction network of the top 50 nodes.}
\begin{tabular}{lcccr}
\hline
\textbf{Gene} & \textbf{Seed} & \textbf{Affinity score} & \textbf{logFC} & \textbf{P-value} \\
\hline
\textit{PLA2G2A} & Y & 0.033061 & -1.790530 & 2.46E-02 \\
\textit{CCND1} & Y & 0.032774 & 1.195653 & 5.43E-06 \\
\textit{BUB1B} & Y & 0.032755 & 1.290377 & 3.10E-03 \\
\textit{AURKA} & Y & 0.032712 & 1.748420 & 3.98E-05 \\
\textit{RAD54B} & Y & 0.032643 & 1.757562 & 1.64E-08 \\
\textit{LTK} & N & 0.001651 & -1.023600 & 7.66E-04 \\
\textit{RDH5} & N & 0.001267 & -1.146630 & 1.34E-08 \\
\textit{NOSTRIN} & N & 0.001110 & -1.379460 & 1.42E-04 \\
\textit{DHR59} & N & 0.001109 & -3.407330 & 9.77E-05 \\
\textit{RMDN2} & N & 0.001097 & -1.010210 & 7.66E-04 \\
\textit{PAG1} & N & 0.000960 & -1.535700 & 2.55E-05 \\
\textit{SEMA3B} & N & 0.000917 & -1.108180 & 7.87E-07 \\
\textit{BCAS1} & N & 0.000903 & -1.881610 & 1.78E-06 \\
\textit{MAP2K2} & N & 0.000881 & -1.020110 & 3.50E-07 \\
\textit{CSF2RB} & N & 0.000838 & -1.414110 & 8.49E-05 \\
\textit{TUBAL3} & N & 0.000837 & -1.631620 & 2.32E-03 \\
\textit{STAP2} & N & 0.000829 & -1.175130 & 5.35E-03 \\
\textit{CCNYL1} & N & 0.000828 & -1.516030 & 3.86E-06 \\
\textit{NEDD9} & N & 0.000808 & -1.110480 & 6.60E-05 \\
\textit{BCAR3} & N & 0.000777 & -1.380050 & 8.69E-09 \\
\textit{FAS} & N & 0.000768 & -1.110550 & 1.40E-04 \\
\textit{KIAA1804} & N & 0.001365 & 1.058162 & 2.43E-03 \\
\textit{EDAR} & N & 0.001124 & 1.319916 & 2.78E-04 \\
\textit{DEPTOR} & N & 0.001112 & 1.325986 & 4.94E-05 \\
\textit{LIMS1} & N & 0.000946 & 1.088253 & 7.41E-09 \\
\textit{ZNF185} & N & 0.000912 & 1.31012 & 2.84E-06 \\
\textit{NKD2} & N & 0.000910 & 1.495638 & 4.17E-08 \\
\textit{FZD3} & N & 0.000898 & 1.25656 & 2.96E-05 \\
\textit{GLCE} & N & 0.000843 & 1.419823 & 7.12E-06 \\
\textit{TNFRSF6B} & N & 0.000829 & 1.121187 & 3.31E-03 \\
\textit{PM20D2} & N & 0.000822 & 1.602442 & 1.00E-06 \\
\hline
\end{tabular}
\end{table}

\begin{flushright}
FC, fold-change; Y, yes; N, no.
\end{flushright}
As a scaffolding protein, NEDD9, also termed human enhancer of filamentation 1 (HEF1) is involved in mediating several cellular processes, including cell cycle progression, cellular attachment, apoptosis, motility and inflammation (35,36). Prostaglandin E2 (PGE2) promotes the expression of HEF1 and then induces cell cycle progression, and HEF1 is an important downstream mediator of the activity of PGE2 in the progression of CRC (37). The activation and overexpression of extracellular signal-regulated kinase/MAPK is important in the progression of CRC and can be a molecular target for the treatment of the cancer (38). It has been reported that MAP2K inhibitors can be of use in CRC therapy (39). Therefore, the dysregulation of NEDD9 and MAP2K2 may affect CRC. In the PPI network of the top 50 nodes in the present study, CCND1 was also found to have interactions with BCAS1, NEDD9 and MAP2K2, indicating that CCND1 may also function in CRC through BCAS1, NEDD9 and MAP2K2.

In conclusion, the present study performed a comprehensive bioinformatics analysis of genes, which may be involved in CRC. A total of 1,640 DEGs were screened in the CRC samples, compared with normal samples. It was found that DEPTOR, AURKA, CCND1, BCAS1, NEDD9 and MAP2K2 may be correlated with CRC. However, further confirmation is required to elucidate their mechanisms of action in CRC.

References


