RESEARCH ARTICLE

Analysis of Key Genes and Pathways Associated with Colorectal Cancer with Microarray Technology

Yan-Jun Liu, Shu Zhang*, Kang Hou, Yun-Tao Li, Zhan Liu, Hai-Liang Ren, Dan Luo, Shi-Hong Li

Abstract

Objective: Microarray data were analyzed to explore key genes and their functions in progression of colorectal cancer (CRC). Methods: Two microarray data sets were downloaded from Gene Expression Omnibus (GEO) database and differentially expressed genes (DEGs) were identified using corresponding packages of R. Functional enrichment analysis was performed with DAVID tools to uncover their biological functions. Results: 631 and 590 DEGs were obtained from the two data sets, respectively. A total of 32 common DEGs were then screened out with the rank product method. The significantly enriched GO terms included inflammatory response, response to wounding and response to drugs. Two interleukin-related domains were revealed in the domain analysis. KEGG pathway enrichment analysis showed that the PPAR signaling pathway and the renin-angiotensin system were enriched in the DEGs. Conclusions: Our study to systematically characterize gene expression changes in CRC with microarray technology revealed changes in a range of key genes, pathways and function modules. Their utility in diagnosis and treatment now require exploration.

Keywords: Colorectal cancer - differentially expressed genes - microarray - pathway - functional enrichment analysis

Introduction

Colorectal cancer (CRC) is the third most commonly diagnosed cancer in the world, but it is more common in developed countries. Around 60% of cases were diagnosed in the developed world. It is estimated that worldwide, in 2008, 1.23 million new cases of colorectal cancer were clinically diagnosed, and that it killed 608,000 people (Ferlay et al., 2010).

Bad lifestyle and aging are the main risk factors while genetic disorders also contribute to the incidence of CRC. It typically starts in the lining of the bowel and if left untreated, can grow into the muscle layers underneath, and then through the bowel wall. Therefore, early diagnosis is critical to decrease the mortality.

Gene mutations (Johnson et al., 2005; Talseth-Palmer et al., 2010) and SNPs (Menin et al., 2006; Aizat et al., 2011) have been found to be linked with CRC and some of them are suggested to be biomarkers. Discovering and validating protein biomarkers are the research hotspots (Zhai et al., 2012). Liu et al suggest that CD73 is a novel prognostic biomarker for human colorectal cancer (Liu et al., 2012). Of course circulating proteins exhibit more usefulness in clinical applications and some have been validated, like cytokeratin 18 (Greystoke et al., 2012) and M2-pyruvate kinase (Meng et al., 2012). Besides, Takahashi et al report that MiR-148a can act as a predictive biomarker in patients with advanced colorectal cancer (Takahashi et al., 2012). Various technologies are utilized to carry out the researches, among which proteomic tool (de Wit et al., 2012) and microarray technology (Amid et al., 2012) are two powerful methods enabling global identification of potential biomarkers.

Though many achievements have been obtained, our understandings about the disease are far from enough to control it clinically. In present study, two microarray data sets were analyzed to identify differentially expressed genes (DEGs). These DEGs can be potential biomarkers. Functional enrichment analysis was then carried out for the DEGs to elucidate their biological functions and thus help to uncover the pathogenesis of CRC.

Materials and Methods

Microarray data and DEGs

Microarray data (GSE4107 (Hong et al., 2007) and GSE8671 (Sabates-Bellver et al., 2007)) were downloaded from GEO database (Edgar et al., 2002). Both raw data were collected in the following platform: GPL570 [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array. Package limma (Smyth, 2004) of R was used to identify DEGs from each data set. The original data
were processed by Bioconductor with RMA method and default settings, and then linear model was adopted. Fold change of >2 and \( p \)-value of < 0.05 were set as the cut-offs to screen DEGs.

**Screening of common DEGs**

The rank product package (Hong et al., 2006) was used to identify the common DEGs between control group and disease group. Briefly, genes were ranked based on up- or down-regulation by the disease group in each experiment. Then a combined probability was calculated for each gene as a rank product (RP). The RP values were used to rank the genes based on how likely it was to observe them by chance at that particular position on the list of DEGs. The RP can be interpreted as a \( p \)-value. To determine significance levels, the RP method uses a permutation-based estimation procedure to transform the \( p \)-value into an e-value that addresses the multiple testing problems derived from testing many genes simultaneously. Genes with a percentage of false-positives (PFP) ≤ 0.05 were considered differentially expressed between treatments and control in each experiment.

**GO enrichment and IntroPro domain analysis**

Gene Ontology (GO) Biological Process (BP) data and functional domain data were extracted using the DAVID (Huang da et al., 2009). GO terms and domains with less than 2 genes were discarded. Over-represented groups of GO BP terms and IntroPro functional domains (Hunter et al., 2009) were identified using a hypergeometric test, with a threshold of \( p \)-value <0.05.

**Pathway analysis**

We adopted an impact analysis that includes the statistical significance of the set of pathway genes but also considers other crucial factors such as the magnitude of each gene’s expression change, the topology of the signaling pathway, their interactions, etc (Draghici et al., 2007).

In this model, the Impact Factor (IF) of a pathway \( Pi \) is calculated as the sum of two terms:

\[
IF(Pi) = \log\left(\frac{1}{Pi}\right) + \sum_{g \in g} \left| \frac{PF(g)}{\alpha \cdot E \cdot |N_e(Pi)|} \right|
\]

**Results**

**Differentially expressed genes**

A total of 631 genes from GSE4107 and 590 genes from GSE8671 were selected out as their fold change larger than 2 and \( p \)-value less than 0.05. The rank product package was used to further determine the differential expression and finally 32 genes with a percentage of false-positives (PFP) < 0.05 were obtained as the common DEGs.

**Functional enrichment analysis results**

Biological processes enrichment analysis was performed for the DEGs obtained from GSE4107, GSE8671 and meta-analysis with DAVID tool to gain insights into their functions. \( P \) value <0.05 was set as the threshold and 278, 155 and 13 GO terms were obtained, respectively. Package VennDiagram of R was chosen to generate Venn diagram (Figure 1). A total of 10 terms were shared by the three groups. They were related with inflammatory response, response to wounding, etc. (Figure 2).

**DOMAIN analysis results**

To add meaningful information to the results from the GO term enrichment, we extended our investigation to IntrPro protein domains. Common and significant
functional domains (p-value<0.05, hypergeometric test) were shown. Most of significantly overrepresented groups included domains were related with interleukin (Table 1).

**KEGG pathway enrichment analysis results**

We carried out an impact analysis integrating many factors including the statistical significance of differentially expressed genes in the pathway, the expression level change, the topology of the signaling pathway, their interactions and so on (p-value<0.05, hypergeometric test). The impact analysis method revealed many significant pathways contained PPAR signaling pathway (Takahashi et al., 2005) and renin-angiotensin system (Burrell et al., 2004) (Table 2).

**Discussion**

Microarray technology is an effective tool to globally uncover changes in gene expression and thus elucidate the molecular mechanisms of complex diseases like CRC. In present study, two microarray data sets were obtained to identify DEGs. A total of 32 DEGs were observed in both data sets, suggesting a high confidence. 5 out of the 32 DEGs were associated with inflammatory response and they were CR2 (CD21), IL8, chemokine (C-C motif) ligand 21 (CCL21), chemokine (C-X-C motif) ligand 13 (CXCL13) and EPH receptor A3 (EPHA3). Since inflammation is closely related with cancer (Marx, 2004), it is not strange that inflammation-related DEGs account for a high percentage. CCL21 is a member of chemokines, which are involved in immunoregulatory and inflammatory processes. They stimulate chemotaxis for different types of immunocytes. Shields et al. suggest that CCL21 is involved in altering the microenvironment and thus facilitates tumor progression (Shields et al., 2010). Koizumi et al report that CCL21 promote the metastasis of human non-small cell lung cancer (Koizumi et al., 2007). The study by Li and others further indicate that CCL21 plays a key role in colon cancer metastasis through regulation of matrix metalloproteinase-9 (Dong et al., 2011). CXCL13 is a member of CXC chemokines that promotes the migration of B lymphocytes (Ansel et al., 2002). It has been found to be related with breast cancer (Panse et al., 2008) and prostate cancer (Singh et al., 2009). El-Hai bi et al further point out that CXCL13 mediates prostate cancer cell proliferation through JNK signaling and invasion through ERK activation (El-Hai bi et al., 2011). EPHA3 belongs to the ephrin receptor subfamily of the protein-tyrosine kinase family and its implication in cancer has been indicated (Surawski et al., 2004; Pasquale, 2010). While several studies investigate the somatic mutations in cancer (Davies et al., 2005; Wood et al., 2006), some look into the underlying mechanisms (Lisabeth et al., 2012) and potential clinical applications as targets (Garber, 2010). The study by Xi et al. confirms the clinicopathological significance and prognostic value of EPHA3 expression in colorectal carcinoma (Xi and Zhao, 2011). IL8 is also closely related with cancer (Lokshin et al., 2006). The study by Rubie et al suggest an association between IL-8 expression, induction and progression of colorectal carcinoma and the development of colorectal liver metastases (Rubie et al., 2007). In accordance with previous findings, IL8-related domain was significantly enriched in present study.

Another important group of DEGs are associated with response to drugs and they are UDP glucuronosyltransferase 1 family, polypeptide A6 (UGT1A6), butyrylcholinesterase (BCHE), fatty acid binding protein 4, adipocyte (FABP4) and ATP-binding cassette, sub-family G (WHITE), member 2 (ABCG2). UGT1A6 is a UDP-glucuronosyltransferase participating in transforming small lipophilic molecules like drugs and hormones. It has been found that genetic variants of UGT1A6 influence risk of colorectal adenoma recurrence (Hubner et al., 2006). Bigler et al report that UGT1A6 genotypes influence the protective effect of aspirin on colon adenoma risk (Bigler et al., 2001), implying its regulatory role in the effectiveness of chemopreventive drugs (Samowitz et al., 2006). BCHE also encodes an enzyme related with drug metabolism, like suxamethonium. Brass et al first report the amplification of the genes BCHE in 40% of squamous cell carcinoma of the lung (Brass et al., 1997). Bernardi et al find similar phenomenon in breast cancer (Bernardi et al., 2010). Montenegro et al find that butyrylcholinesterase activity decreases in human colon adenocarcinoma (Montenegro et al., 2006).

Several DEGs implicated in enzyme linked receptor protein signaling pathway were also uncovered: chordin-like 1 (CHRDL1), angiopoietin-like 1 (ANGPTL1), gremlin 2 (GREM2) and EPHA3. CHRDL1 and GREM2 are involved in regulating the intestinal stem cell niche (Ricci-Vitiani et al., 2009; Todaro et al., 2010), and further studies may reveal the whole regulatory mechanisms and thus support the targeted therapy.

Besides, we found that PPAR signaling pathway was significantly enriched, which was in accordance with

<table>
<thead>
<tr>
<th>Table 1. DEGs Contained in Each Enriched Domain in CRC</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Term</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>IPR001811:Small chemokine, interleukin-8-like</td>
</tr>
<tr>
<td>IPR002473:Small chemokine, C-X-C/Interleukin 8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 2. Significant Pathways Involved in CRC</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pathway Name</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>PPAR signaling pathway</td>
</tr>
<tr>
<td>Renin-angiotensin system</td>
</tr>
</tbody>
</table>

DOI:http://dx.doi.org/10.7314/APJCP.2013.14.3.1819
previous study (Yang and Frucht, 2001). The relationship between the renin-angiotensin system and malignancy is also determined (Ager et al., 2008; George et al., 2010). These results further confirm the usefulness of our findings.

Overall, our study provides a range of DEGs, some of which have been confirmed to be related with CRC. Subsequent functional enrichment analysis indicates their biological roles and the results are beneficial to promote relevant studies. Like EphA3, which have been validated to be a good biomarker, more targets would be identified if further studies were carried out, which will improve the clinical outcomes for patients with CRC.

References


Analysis of Key Genes and Pathways Associated with Colorectal Cancer with Microarray Technology

DOI:http://dx.doi.org/10.7314/APJCP.2013.14.3.1819


