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 systemically 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
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:

$$IF(P_i) = \log \left( \frac{1}{p_i} \right) + \sum_{g \in P_i} \left| \frac{PF(g)}{\alpha \cdot E \cdot N_g(P_i)} \right|$$

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 IntroPro protein domains. Common and significant
et al., 2011). EPHA3 belongs to the ephrin receptor signaling and invasion through ERK activation (El-Haibi et al., 2009). El-Haibi et al. further point out that CXCL13 that promotes the migration of B lymphocytes (Ansel et al., 2004) and prostate cancer (Singh et al., 2009). CXCL13 is an important mediator of prostate cancer cell proliferation through JNK signaling and invasion through ERK activation (El-Haibi et al., 2011). EPHA3 belongs to the ephrin receptor subfamily of the protein-tyrosine kinase family and its implication in cancer has been indicated (Suraw ska 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.

**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-Haibi et al further point out that CXCL13 mediates prostate cancer cell proliferation through JNK signaling and invasion through ERK activation (El-Haibi et al., 2011). EPHA3 belongs to the ephrin receptor domain of differentially expressed genes involved in colorectal cancer.

### Table 1. DEGs Contained in Each Enriched Domain in CRC

<table>
<thead>
<tr>
<th>Term</th>
<th>META Count</th>
<th>P value</th>
<th>GSE4107 Count</th>
<th>P value</th>
<th>GSE8671 Count</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPR001811: Small chemokine, interleukin-8-like</td>
<td>3</td>
<td>2.30E-03</td>
<td>6</td>
<td>9.75E-03</td>
<td>16</td>
<td>6.23E-13</td>
</tr>
<tr>
<td>IPR002473: Small chemokine, C-X-C/Interleukin-8</td>
<td>2</td>
<td>2.41E-02</td>
<td>3</td>
<td>4.98E-02</td>
<td>8</td>
<td>7.77E-08</td>
</tr>
</tbody>
</table>

### Table 2. Significant Pathways Involved in CRC

<table>
<thead>
<tr>
<th>Pathway Name</th>
<th>META Impact Factor</th>
<th>p-value</th>
<th>GSE4107 Impact Factor</th>
<th>p-value</th>
<th>GSE8671 Impact Factor</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>PPAR signaling pathway</td>
<td>5.228</td>
<td>5.36E-03</td>
<td>5.154</td>
<td>5.78E-03</td>
<td>7.013</td>
<td>9.00E-04</td>
</tr>
<tr>
<td>Renin-angiotensin system</td>
<td>3.595</td>
<td>2.75E-02</td>
<td>4.118</td>
<td>1.63E-02</td>
<td>4.297</td>
<td>0.013605</td>
</tr>
</tbody>
</table>

In summary, the study provides insights into the molecular mechanisms underlying colorectal cancer through microarray analysis. The top pathways including PPAR signaling and renin-angiotensin system were significantly enriched, highlighting key genes and pathways that are potentially targeted for therapeutic intervention. Further studies are needed to validate these findings and explore potential clinical applications.
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


