Experimental and Molecular Pathology 96 (2014) 361–366

Contents lists available at ScienceDirect

Experimental and Molecular Pathology

journal homepage: www.elsevier.com/locate/yexmp

Gene expression profiling and bioinformatics analysis of gastric carcinoma

Ningning Liu [a,b], Xuan Liu [a], Ning Zhou [b], Qiong Wu [b], Lihong Zhou [a], Qi Li [a,*]

[a] Department of Medical Oncology, Shuguang Hospital, Shanghai University of Traditional Chinese Medicine, Shanghai 201203, PR China
[b] Interventional Cancer Institute of Integrative Medicine & Putuo Hospital, Shanghai University of Traditional Chinese Medicine, Shanghai 200062, PR China

ARTICLE INFO

Article history:
Received 18 September 2013
Available online 28 February 2014

Keywords:
Gastric carcinoma
Expression profiles
Co-expression module
Transcription factor binding sites
MicroRNA
Bioinformatics analysis

ABSTRACT

Gastric cancer remains one of the major health problems worldwide, and it is one of the most common cancers and the leading cause of cancer-related deaths in China. This study was to analyze the expression profiles of genes in gastric carcinoma, and predict potential regulating factors. The gene expression profile data GSE13911 was downloaded from Gene Expression Omnibus and the differentially expressed genes (DEGs) were identified by t-test. Gene modules were constructed using hierarchical clustering in R based on average linkage and Pearson's correlation coefficient and functional analysis for these genes were performed with DAVID. Genes in each module with Pearson's correlation coefficient >0.3 were obtained to construct co-expression network. Protein–protein interactions (PPIs) were identified by comparing protein–protein interaction (PPI) network with co-expression networks. In addition, the potential regulatory microRNAs and the transcription factors for each module were screened out. In this study, six modules associated with protein degradation, cell cycle, protein trafficking and immunoreaction were identified. COPS5 (COP9 Subunit 5) was the core protein in the largest PPI network of module 1. The transcription factors MYC and MAZ (Myc-associated zinc-finger protein) were enriched in module 1. A total of 9 microRNA–target bi-clusters were identified and module 1 enriched 20 genes targeting to miR-17–92 gene cluster(miR-17/20ab)and miR-106b–25 gene cluster (miR-106b/93). In conclusion, we constructed 6 gene modules and screened out some genes, transcriptional factors and microRNAs that may be used as potential molecular biomarkers for gastric carcinoma.

Introduction

Gastric cancer, a highly invasive and aggressive malignancy that is characterized by resistance to apoptosis and radio resistance, is among the most common cancers and is the leading cause of cancer-related death in China (Aggarwal et al., 2006). So far, gastric carcinoma is still one of the common malignant tumors that threaten human health severely, while China is among the countries where there is high risk of gastric carcinoma throughout the world (Lv et al., 2011). Studies have shown that individuals infected with *Helicobacter pylori* have an increased risk of gastric carcinoma (Islami and Kamangar, 2008; Kamangar et al., 2007). Some *H. pylori* bacteria could inject a toxin produced by cagA into the junctions of the stomach lining meet thus increase the risk of gastric cancer (Bagnoli et al., 2005; Wen and Moss, 2009). Recently, along with the improvement of living and health conditions, the optimization of dietetic structure and food storage methods, the popularity of Gastroscopy Examination, and the timely detection and therapy of *H. pylori*, the incidence of gastric carcinoma over the world has decreased. Even though, of malignant diseases, it is still a threat to the middle and old, especially the ones in lower living conditions (Thun et al., 2010). As there is little or even no peculiar symptom in early stage of gastric carcinoma, the diagnosis is often delayed. When the typical symptoms of tumor appeared, like anemia, weight loss, and so on, gastric carcinoma had developed into late-stage. Besides, when it is diagnosed, the tumor cells have intruded into muscular is propriety (Okines et al., 2010), which brings great trouble in therapy.

With the development of modern medicine and technology, a series of intensive studies of gastric carcinoma were promoted. The application of endoscopic technology drove postoperative 5-year survival rate of early gastric carcinoma up to more than 90%, but the therapeutic effect to progression was still poor (Dai et al., 2005). Lots of researches indicated that variation of gene function was the key in the occurrence and development of gastric carcinoma. And numerous gene variations related to gastric carcinoma were discovered as well. For example, the polymorphism of the prostate stem cell antigen (PSCA) gene, which is possibly involved in regulating gastric epithelial-cell proliferation, influences susceptibility to diffuse-type gastric cancer (Sakamoto et al., 2008). Polymorphisms in interleukin (IL)-1β and its endogenous receptor antagonist are associated with risk of *H. pylori*-related gastric cancer

* Corresponding author at: Department of Medical Oncology, Shuguang Hospital, Shanghai University of Traditional Chinese Medicine, No. 528 Zhangheng Road, Shanghai 201203, PR China. Fax: +86 2120256533.
E-mail address: Lzwf@hotmail.com (Q. Li).

http://dx.doi.org/10.1016/j.yexmp.2014.02.007
0014-4800/

(El-Omar et al., 2003). The most widely studied is the GSTM1 null polymorphism. It was reported that the GSTM1 null allele may predispose endometriotic lesions to malignant transformation to endometrioid and clear cell ovarian cancer (Baxter et al., 2001). The most consistent results are the increased gastric cancer risk associated with IL1B and NAT1 variants, which may account for up to 48% of attributable risk of gastric cancer. Only polymorphisms at HLA-DQ, TNF and CYP2E genes may confer some protective effect against gastric cancer (González et al., 2002). However, the molecular mechanism of occurrence, infiltration and metastasis in gastric carcinoma has not been clarified distinctly till now. Fortunately, cDNA microarray technology combining bioinformatics analysis made it able to analyze the expression changes of mRNA in the development and progression of gastric carcinoma, comprehensively. In this way, we overcome the shortages like insufficient information and ambiguous inner link in the research of individual genes (Irizarry et al., 2003b).

In this study, basing on the construction of gene co-expression network, we recognized the modules whose expression profiles highly resembled in both gastric carcinoma and normal gastric tissues. By way of analyzing their biological functions and predicting the regulatory factors (transcription factors or microRNA), we may light the further insight of gastric carcinoma development at molecular level, and at the same time provide methods for advanced diagnostics and prognosis.

Materials and methods

Data of gene chip

The gene expression profile of GSE13911 (D'Errico et al., 2009) was downloaded from a public functional genomics data repository GEO (Gene Expression Omnibus, http://www.ncbi.nlm.nih.gov/geo/) database. A total of 62 specimens, including 19 pairs of MSS(micro-satellite stability)gastric carcinoma and the matched normal tissues, and 12 pairs of MSI(micro-satellite instability)ones, were available based on the HG-U133 plus 2.0 Platform. Nonspecific probes were removed, and the rest were combined into one probe set with the Genome-wide expression profile provided by scholars in the Michigan University (Dai et al., 2005). Using the robust multiparty average (RMA) algorithm (Irizarry et al., 2003b), the expression index was recalculated and expression profiling data was obtained with Entrez Gene as the identifier. From "Homo_sapiens.gene_info", an annotation file in NCBI, we got the official names of human genes. "Present," "Marginal," and "Absent" calls were made in R using the MAS5 algorithm in the affy package (Irizarry et al., 2003a). Genes registered "Present" calls totally by the MAS5 algorithm for all the 62 samples were selected to analyze.

Screening of differentially expressed genes (DEGs)

t-Test was used to identify genes that were significantly differentially expressed among the 31 couples of gastric carcinoma and the matched normal tissues. The p-values were adjusted to q-values for multiple testing correction based on qvalue package in R. Only the genes with $q < 0.05$ were considered to have significant differences in expression.

Construction of module and co-expression network

The genes with "Present" calls were clustered with respect to expression pattern using hierarchical clustering based on average linkage and Pearson's correlation coefficient as a distance measure and finally a dendrogram was created. Clusters in dendrogram were detected using Dynamic Hybrid tree cut algorithm (Langfelder et al., 2008) depending on clusters' shapes. The cluster joining height was set to 0.25. While constructing the co-expression network of genes in one module, Pearson's correlation coefficient above 0.3 was used as the cut-off criterion. With the gene expression profiles of each module in gastric carcinoma, figures were established in R. Then we uploaded genes in each module to DAVID (Da Wei Huang and Lempicki, 2008) to identify overrepresented GO categories in biological process and KEGG signal pathway (Kanehisa and Goto, 2000). The FDR < 0.001 was used as the cut-off criterion.

Integration of protein–protein interaction (PPI) and co-expression network

The PPI data was downloaded from BioGRID (Chatr-aryamontri et al., 2013). The overlapping interaction relationships were selected by comparing the co-expression network of each module with PPI network in Cytoscape (Shannon et al., 2003). After the identification of sub-networks by MCL algorithm (Enright et al., 2002), BiNGO (Maere et al., 2005) was used to analyze their over-represented GO categories.

Gene set enrichment analysis of transcription factor binding sequence in co-expression network

Genes of each module were uploaded to MSigDB (Liberzon et al., 2011). Afterwards, motifs were identified within the region from upstream 2 kb to the downstream 2 kb of promoter. Then we evaluated whether each motif was significantly enriched in the module by Fisher's exact test. FDR less than 0.05 was chosen as the cut-off criterion.

Enrichment analysis of microRNA targets

The predicted target genes of microRNA were downloaded from TargetScan (Friedman et al., 2009). Of all the 62 samples, the genes whose detection call results totally registered "Present" were selected. Then microRNAs with at least one target gene were retained. We defined microRNA as the row and target gene as the column to establish a binary matrix. If there is target relationship between microRNA and a gene, the corresponding element value was set to 1, otherwise 0. Bimax algorithm (Prelić et al., 2006) was applied to detect the bi-cluster, whose element values were all 1 in the binary matrix, on condition that there were at least 100 target genes and 10 microRNAs in the bi-clusters and that the algorithm recognized less than 30 bi-clusters. The target genes in the obtained bi-clusters were matched with the genes in each module. And the overlap was analyzed by Fisher's exact test to evaluate the significance. After Bonferroni's correction for multiple testing, p-value < 0.05 was set as the cut-off criterion.

Results

Construction gene module

There were 3156 genes registered "Present" totally in the 62 samples, 1439 genes of which were grouped into 6 modules. Module 1 was the largest one and module 6 was the smallest one, which contained 895 and 39 genes, respectively. The expression profiles of genes in 6 modules were shown in Fig. 1. The remaining 1717 genes were assigned to module 0 (Supplementary Fig. 1) for analyzing as well. Comparison of each module and the differentially expressed genes revealed that 857 of the 895 genes in module 1 were significantly up-regulated in gastric carcinoma (q-value < 0.05), while there were also different numbers of DEGs in the other modules. There were 157, 64, 12, 8 and 577 genes that were significantly up-regulated gastric carcinoma in modules 2, 3, 4, 5, 6 and 0, respectively.

Biological function analysis

We uploaded genes of each module to the online software DAVID to identify overrepresented GO categories and KEGG signal pathways (Table 1). Except module 4 and module 6, the other modules were all enriched in different biological processes or signal pathways. In detail, module 1 was associated with protein degradation and cell cycle;