**Exploration of Molecular Mechanisms of Diffuse Large B-cell Lymphoma Development Using a Microarray**

Zong-Xin Zhang, Cui-Fen Shen, Wei-Hua Zou, Li-Hong Shou, Hui-Ying Zhang, Wen-Jun Jin*

**Abstract**

**Objective:** We aimed to identify key genes, pathways and function modules in the development of diffuse large B-cell lymphoma (DLBCL) with microarray data and interaction network analysis. **Methods:** Microarray data sets for 7 DLBCL samples and 7 normal controls was downloaded from the Gene Expression Omnibus (GEO) database and differentially expressed genes (DEGs) were identified with Student’s t-test. KEGG functional enrichment analysis was performed to uncover their biological functions. Three global networks were established for immune system, signaling molecules and interactions and cancer genes. The DEGs were compared with the networks to observe their distributions and determine important key genes, pathways and modules. **Results:** A total of 945 DEGs were obtained, 272 up-regulated and 673 down-regulated. KEGG analysis revealed that two groups of pathways were significantly enriched: immune function and signaling molecules and interactions. Following interaction network analysis further confirmed the association of DEGs in immune system, signaling molecules and interactions and cancer genes. **Conclusions:** Our study could systemically characterize gene expression changes in DLBCL with microarray technology. A range of key genes, pathways and function modules were revealed. Utility in diagnosis and treatment may be expected with further focused research.

**Keywords:** Diffuse large B-cell lymphoma (DLBCL) - differentially expressed genes - pathway - module

---

**Introduction**

Diffuse large B-cell lymphoma (DLBCL) is the most common non-Hodgkin’s lymphoma. According to statistics, DLBCL accounts for about 3% of primary brain tumors (Davis et al., 1999) and 30% - 40% of adult non-Hodgkin’s malignant lymphoma (Kawano et al., 2004). It can be originated from lymph nodes or extranodal organs and tissues, and transformed from the inert lymphoma. It has a high recurrence rate about 35% - 60% in 1-2 years and a low survival rate as the 5-year survival rate is only 20% - 40%. In the last few decades, incidence of DLBCL exhibits an increasing trend. Therefore, developing effective treatments is an urgent task (Akbulut, 2012).

Previous studies have reported abnormalities in multiple biological pathways that involved in the pathogenesis of DLBCL, such as chemotactic factor signaling pathways (Smith et al., 2007), NFκB signaling pathway (Montesinos-Rongen et al., 2012), host inflammatory response (Monti et al., 2005) and chronic active B-cell-receptor signaling (Davis et al., 2010). Lenz et al report oncogenic CARD11 mutations in DLBCL and consider that it can be a target for therapy (Lenz et al., 2008). Yao et al find that expression of dendritic cell marker CD21 is a positive prognostic factor for DLBCL (Yao et al., 2012). Expression of B-cell CLL/lymphoma 6 (BCL6) and BCL2 is also thought to be able to predicting survival in patients with DLBCL (Lossos et al., 2001; Iqbal et al., 2006). However, so far, the pathogenesis of DLBCL is not yet entirely clear.

In present study, microarray technology was adopted to globally characterize the changes in gene expression. Then interaction network analysis was conducted to identify key genes, pathways and modules, which would be beneficial to elucidate the mechanisms of DLBCL.

**Materials and Methods**

**Microarray data**

Microarray data set GSE25297 (Sung et al., 2011) was downloaded from GEO database (http://www.ncbi.nlm.nih.gov/geo/), including 7 DLBCL samples and 7 normal controls. Raw data was collected with Agilent-014850 Whole Human Genome Microarray 4x44K G4112F. Probe annotation was also provided by Agilent.

Following steps were done to process the raw data: (1) Agilent Feature Extraction Software (v9) was used to extract DNA hybridization signals; (2) VSN normalization was performed and expression levels were determined; (3) probes linked with more than one gene were removed; (4) multiple probes pointing to the same genes were averaged. Finally, expression profiling data with 14 samples and
18,459 genes were obtained.

**Screening of differentially expressed genes**

Microarray data for DLBCL were compared with that for control to identify DEGs using Student’s t-test. Multiple-correction was conducted for the p value with Benjamin & Hochberg (BH) method to reduce false positive. FDR ≤ 0.1 and fold change ≥ 2 were set as the cut-off.

**Functional enrichment analysis**

KEGG annotation is useful in exploring the biological functions of the DEGs. The EASE that developed by DAVID (Huang da et al., 2009) was chosen to complete this task, and it is similar to Fisher’s exact test. EASE of 0.1 was selected as the threshold.

**Establishment of global network for immune system**

Global network for immune system-related pathways was established to observe the distribution of the DEGs implicated in immune. 15 relevant pathways were acquired from the KEGG pathway database: hematopoietic cell lineage, complement and coagulation cascades, toll-like receptor signaling pathway, NOD-like receptor signaling pathway, RIG-I-like receptor signaling pathway, cytosolic DNA-sensing pathway, natural killer cell mediated cytotoxicity, antigen processing and presentation, T cell receptor signaling pathway, B cell receptor signaling pathway, chemokine receptor and metastasis of several cancers (Kakinuma and Hwang, 2006; Koizumi et al., 2007). Besides, chemokine CXCL12-CXCR4 is found to be implicated in the cross-talk between chemokine as well as cytokine-cytokine receptor interaction. These pathways might play important parts in the incidence and progression of DLBCL. Previous studies have reported the close relationships between chemokine as well as chemokine receptor and metastasis of several cancers (Kakinuma and Hwang, 2006; Koizumi et al., 2007). Besides, chemokine CXCL12-CXCR4 is found to be implicated in the cross-talk between central nervous system and immune system (Klein and Rubin, 2004).

**Results**

**Differentially expressed genes**

DEGs were identified with Student’s t-test and the p values were corrected by BH method. The corrected FDR<0.05 and fold change ≥2 were set as the cut-offs. A total of 945 DEGs were obtained, including 272 up-regulated genes and 673 down-regulated genes, and the ratio was 1:2.47. More DEGs were down-regulated. These DEGs might play key roles in the development of DLBCL and thus further analysis was performed.

**KEGG functional enrichment analysis results**

KEGG functional enrichment analysis was conducted and p values were corrected with BH method. A total of 18 pathway were revealed as p value < 0.05, and 5 remained significant after multiple correction as FDR ≤ 0.05 (Table 1).

The five pathways could be divided into two groups. The first group was associated with immune system: hematopoietic cell lineage and T cell receptor signaling pathway. The second group was about signaling molecules & interactions: cell adhesion molecules (CAMs) and cytokine-cytokine receptor interaction. These pathways might play important parts in the incidence and progression of DLBCL.

**DEGs and immune system**

As shown in KEGG pathway enrichment analysis, some DEGs were significantly enriched in immune system-related pathways, thus a global network about...
immune system was built up to systemically screen key genes, such as hub genes or functional modules, which play important parts in the pathogenesis of DLBCL. A network including 657 genes and 3611 interactions were obtained (Figure 1), among which 55 were found to be DEGs with annotations, 51 down-regulated and 3 up-regulated. Most of the DEGs were down-regulated, suggesting decline in immune function.

Figure 1 A showed the distribution of the 55 DEGs in the global network. The DEGs were almost located in the center of the network, suggesting their important regulatory roles and influence on the immune function. Then the degree for each DEG was calculated and the top 11 genes with most degree were PHGDH (25), PDGFRB (46), PIK3CD (31), PIK3R5 (29), FZD10 (22), CCND1 (18), ITGA3 (18), FLT3 (14), IGF1 (9), SHC1 (7), COL4A1 (6), COL4A2 (6) and COL4A4 (6). PDGFRB (46), PIK3CD (31), PIK3R5 (29), FZD10 (22), CCND1 (18), ITGA3 (18), FLT3 (14), IGF1 (9), SHC1 (7), COL4A1 (6), COL4A2 (6) and COL4A4 (6). PDGFRB had the maximum degree and it was a platelet-derived growth factor (Bornfeldt et al., 1995). Previous study indicates that it is related with glioma and lower level was observed in glioma with high grade malignancy (Lokker et al., 2002). It interacts with PIK3CD and PIK3R5. The DEGs and signaling molecules & interactions

Another group of pathways enriched in DEGs was signaling molecules & interactions. Therefore, a global network was generated to identify key genes closely associated with DLBCL. The global network was comprised of 332 genes and 1071 interactions, and a total of 53 DEGs were included, 44 down-regulated and 9 up-regulated (Figure 3). Most of the genes were down-regulated, implying the decreased activity in cell communication.

Figure 3 A shows the distribution of the 53 DEGs in the global network. The top 11 genes with most degree were ITGA3 (36), ITGA9 (35), SDC1 (27), CLDN14 (19), CLDN15 (19), CLDN5 (19), COL4A1 (15), COL4A2 (15), COL4A3 (15) and COL5A3 (15). Two modules were observed in Figure 3 B: one with CLDN15 (19), CLDN5 (19) and COL4A1 (15), and another one with CD4.

DEGs and global network for cancer genes

A global network for cancer genes was established to investigate the involvement of the DEGs in DLBCL. The global network was comprised of 332 genes and 1481 interactions (Figure 4). 29 had annotations in this network, 23 down-regulated and 6 up-regulated. Most of the cancer-related DEGs were down-regulated.

Figure 4 A shows the distribution of the 29 DEGs in the global network. The top 11 genes with most degree were PDGFRB (46), PIK3CD (31), PIK3R5 (29), FZD10 (22), CCND1 (18), ITGA3 (18), FLT3 (14), IGF1 (9), SHC1 (7), COL4A1 (6), COL4A2 (6) and COL4A4 (6). PDGFRB had the maximum degree and it was a platelet-derived growth factor (Bornfeldt et al., 1995). Previous study indicates that it is related with glioma and lower level was observed in glioma with high grade malignancy (Lokker et al., 2002). It interacts with PIK3CD and PIK3R5.
results of the present study indicate that PDGF-D promotes malignant mesothelioma cell chemotaxis through PDGF-ββ receptor signaling pathways along a PI3 kinase/PDK1/Akt/Rac1/ROCK axis and relevant to ERK activation.

Then DEGs observed in the above two networks were entered into this network and distribution was shown in Figure 5.

As shown in Figure 5, some of the DEGs related with immune and signaling molecules & interactions interacted with cancer genes. These genes were located in part of the network, rather than the whole, in accordance with the knowledge that cancer is caused by multiple factors. The DEGs shared by both networks of signaling molecules & interactions and cancer genes were WT1, CCL21, COL4A1, LAMC2, COL4A2, COL4A3, LAMC3, FLT3, VEGFC and TGFBR2, most of which were laminin and matrix membrane collagen. And those for networks of immune system and cancer genes were PTPN11, SOS1, PIK3R5, PIK3CD, SHC1 and PAK6, most of which were phosphatidylinositol and tyrosine kinase.

Discussion

Gene expression data for DLBCL was compared with that for normal control to identify DEGs. Then their biological functions were investigated by KEGG functional enrichment analysis and two groups of pathways were revealed. The first one was immune function-related pathways, such as hematopoietic cell lineage and cell receptor signaling pathway, and the second one was signaling molecules & interactions, such as cell adhesion molecules and cytokine-cytokine receptor interaction. Thus we considered that abnormalities in immune function and signaling molecules & interactions were major causes of DLBCL.

Some key genes were identified with network analysis. These genes had the highest degree and thus imposed significant impact on the whole network. Perturbations in these genes might contribute most to the incidence of DLBCL.

The DEGs like PIK3CD, PIK3R5, HLA-G, CLDN14, CLDN15, CLDN5, VCAM1, CCL14, CCL19 and CCL21 were found to be linked with immune function. Phosphatidylinositol-4,5-bisphosphate 3-kinase, catalytic subunit delta (PIK3CD) and phosphoinoside-3-kinase, regulatory subunit 5 (PIK3R5) belong to PI3 family, which is found to be involved in tumor growth (Katsos et al., 2001; Koutros et al., 2010) and metastasis (Wee et al., 2008). Therefore, we considered that these hub genes might play critical roles in the development of DLBCL. Courtenay et al think that the PI3K pathway can be drug target for cancer (Courtenay et al., 2010). Major histocompatibility complex class I G (HLA-G) belongs to the HLA class I heavy chain paralogues. It is suspected to provide an escape mechanism for cancer cells (Davidson et al., 2005; Rouas-Freiss et al., 2005). In addition, two modules were extracted from the immune system-related network. Most of the genes in the first module (Figure 2 A) were claudins, a class of membrane proteins participating in tight junctions. They were implicated in signal transduction and maintenance of cell polarity. Previous studies have indicated that claudins were over-expressed in cancers (Rangel et al., 2003), and even involved in cancer cell metastasis (Dhawan et al., 2005) and invasion processes (Agarwal et al., 2005; Oku et al., 2006; Takehara et al., 2009). Some studies consider that some members of claudins can be targets for treatment (Michi et al., 2001; Morin, 2005). Most of the genes in the second module (Figure 2 B) were chemokines and their receptors. Chemokine (C-C motif) ligand 21 (CCL21) is reported to guide aggregation of B cells in the thyroid gland of transgenic mice (Muniz et al., 2011). Chemokines were also closely linked with cancer cell survival and proliferation (Xu et al., 2011).

DEGs such as ITGA3, ITGA9, SDC1, CLDN14, CLDN15, CLDN5, COL4A1, COL4A2, COL4A3, COL5A3 and COL6A2 were associated with signaling molecules & interactions. Integrin alpha 3 (ITGA3) and integrin alpha 9 (ITGA9) are members of integrin family. Chen et al. report that ITGA3 is implicated in the metastasis and invasion of meningiomas, and its expression level is negatively correlated with proliferative activity and degree of malignancy (Chen et al., 2009). The integrins and corresponding signaling pathways can be drug targets for treatment of cancers (Hannigan et al., 2005; Cai and Chen, 2006; Mitra and Schlaepfer, 2006). Collagen type IV alpha 1 (COL4A1), collagen type IV alpha 2 (COL4A2), collagen type IV alpha (COL4A), collagen type V alpha 3 (COL5A3) and collagen type VI alpha 2 (COL6A2) are various collagen chains, components of basement membrane, which plays an important role in cancer metastasis and invasion (Okawa et al., 2006). Havenith et al report the prognostic value of type IV collagen in colorectal cancer (Havenith et al., 1988). Two modules were also identified. The first module (Figure 3 B upper) contains CLDN15, CLDN5 and COL4A1 and was associated with signal transduction and maintenance of cell polarity. The second module (Figure 3 B lower) includes CD4 and was linked with inflammatory response. Genes in this module were components of second class of histocompatibility complex and participated in antigen presentation. CD4 was located at the center of the module and found to be down-regulated, implying the decreased capacity of immune system which provides favorable conditions for the formation and growth of cancer cells.

Moreover, several DEGs usually formed a function module, which worked together to finish certain biological process. Alterations in these modules might provide favorable environment for the development of DLBCL.

Some DEGs participating in the networks for immune function and signaling molecules & interactions were also observed in the network for cancer genes, which further confirm their implications in the cancer genesis. The analysis suggested that PIK3CD, PIK3R5 and ITGA3 were in very important positions.

Overall, a range of critical genes, pathways and function modules were revealed in present study, which were really helpful in improving our understandings about the pathogenesis of DLBCL. In addition, more investigations are needed to fulfill their clinical applications.
References


Davis FG, McCarthy BJ, Berger MS (1999). Centralized databases available for describing primary brain tumor incidence, survival, and treatment: Central Brain Tumor Registry of the United States; Surveillance, Epidemiology, and End Results; and National Cancer Data Base. Neuro Oncol, 1, 205-11.


