Identification potential biomarkers and therapeutic agents in multiple myeloma based on bioinformatics analysis

X.-G. WANG, Y. PENG, X.-L. SONG, J.-P. LAN

Hematology Department, Zhejiang Provincial People’s Hospital, Hangzhou City, China

Abstract. – OBJECTIVE: The study aimed to identify potential therapeutic biomarkers and agents in multiple myeloma (MM) based on bioinformatics analysis.

MATERIALS AND METHODS: The microarray data of GSE36474 were downloaded from Gene Expression Omnibus database. A total of 4 MM and 3 normal bone marrow mesenchymal stromal cells (BM-MSCs) samples were used to identify the differentially expressed genes (DEGs). The hierarchical clustering analysis and functional enrichment analysis of DEGs were performed. Furthermore, co-expression network was constructed by Cytoscape software. The potential small molecular agents were identified with Connectivity Map (cMap) database.

RESULTS: A total of 573 DEGs were identified in MM samples comparing with normal samples, including 322 down- and 251 up-regulated genes. The DEGs were separated into two clusters. Down-regulated genes were mainly enriched in cell cycle function, while up-regulated genes were related to immune response. Down-regulated genes such as checkpoint kinase 1 (CHEK1), MAD2 mitotic arrest deficient-like 1 (MAD2L1) and DBF4 zinc finger (DBF4) were identified in cell cycle-related co-expression network. Up-regulated gene of guanylate binding protein 1, interferon-inducible (GBP1) was a hub node in immune response-related co-expression network. Additionally, the small molecular agent vinblastine was identified in this study.

CONCLUSIONS: The genes such as CHEK1, MAD2L1, DBF4 and GBP1 may be potential therapeutic biomarkers in MM. Vinblastine may be a potential therapeutic agent in MM.

Key Words: Multiple myeloma, Bioinformatics analysis, Differentially expressed genes, Biomarker, Therapeutic agent.

Introduction

Multiple myeloma (MM) is a malignancy of plasma cells characterized by unlimited proliferation of abnormal plasma cells in bone marrow and high levels of monoclonal protein in the blood. MM accounts for approximately 10% of hematologic malignancies and 1% of all cancers1,2. Myeloma patients usually have bone lesions, hypercalcemia, severe immunodeficiency and susceptibility to bacterial infections3. MM is the second most common hematologic malignancy in the United States4. The 5-year survival rate of MM after surgery is only 44.9%5. Therefore, uncovering therapeutic biomarkers and agents in MM would supply new insights for the diagnosis and treatment of MM.

Numerous studies have been done to prevent and treat myeloma in MM patients. The B-lymphocyte stimulator is determined to be a biomarker for the diagnosis and treatment of MM5. A high free light chain ratio is a predictor of imminent progression in smoldering MM6. It has been reported7 that gene B-cell CLL/lymphoma 2 (BCL2) is a diagnostic biomarker in MM due to controlling the NF-κB activation-signaling pathway. Currently, some agents have been used in MM therapy. Bortezomib is a selective proteasome inhibitor that has shown encouraging results in patients with MM and other malignant diseases8. Lund et al9 reported that addition of thalidomide to melphalan and prednisone treatment prolonged survival in MM. However, they may cause side effects, such as polyneuropathy, skin rash, fatigue, and venous thromboembolism10. Thus, it is essential to develop a safer drug for the treatment of MM. Given the limitation of these studies, we have taken into consideration all differentially expressed genes (DEGs) to explore new potential therapeutic biomarkers in MM.

The microarray data (GSE36474) were analyzed with Biometric Research Branch Array-Tools (National Cancer Institute, Bethesda, MA, USA) to identify DEGs involved in MM by André et al11. In this study, we downloaded this data and used limma package to identify the DEGs in bone marrow mesenchymal stromal cells (BM-
Biomarkers and agents in MM

MSCs) between MM and normal samples. Additionally, other methods were performed to analyze this data, including hierarchical clustering analysis, functional enrichment analysis, co-expression network construction and identification of small molecular agents. The purpose of this study was to identify new potential therapeutic biomarkers and agents in MM.

Materials and Methods

Affymetrix Microarray Data

The gene expression profile data of GSE36474 based on the platform of GPL570 (Affymetrix Human Genome U133 Plus 2.0 Array) were obtained from Gene Expression Omnibus (GEO) database in National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/geo/), which was deposited by André et al. GSE36474 datasets detected the transcript of BM-MSCs from 4 untreated MM patients (47, 52, 56, 65 years) and 3 healthy donors (48, 50, 58 years).

Data Preprocessing and Differential Expression Analysis

The probe-level data in CEL files were converted into expression measures. The missing data were imputed. Then, the data were quantile normalized with the Affy package in R. The limma package in R was used to identify DEGs between MM and normal samples. Log2-fold change (log2FC) was calculated to identify genes with expression-level differences. The p-value < 0.01 and log2FC > 1 were used as the cutoff criteria. For hierarchical clustering of samples and identified DEGs, clustering analysis was performed to create a clustering graph of samples and genes with heatmap package in R using Euclidean distance measure.

Gene Ontology Enrichment Analysis

GO analysis has become a commonly used method for functional studies of large-scale transcriptomic or genomic data. Database for Annotation, Visualization and Integrated Discovery (DAVID) is a gene functional enrichment analysis tool to understand the biological meaning for investigators. GO categories are divided into three systems: molecular function (MF), biological process (BP), and cellular component (CC). In this study, DAVID was used to identify GO categories for BP with false discovery rate (FDR) < 0.05.

Construction of Co-Expression Network

The genes with similar expression profiles are considered to have the common regulatory function and a high degree of correlation. Search Tool for the Retrieval of Interacting Genes (STRING) is an online database which collects comprehensive information of proteins. Genes in the most significant GO terms were used to construct the co-expression network. The STRING online tool was applied to analyze the co-expressed gene pairs. In the network, genes represent nodes and the interactions between the nodes represent edges. Connectivity degree represents the number of edges linked to a given node. The important nodes with high degree in the network were obtained, namely hub nodes. The co-expression value > 0.4 was used as the cutoff criterion. If the co-expressed pairs over 2000, the co-expresses value > 0.8 was used as the cutoff criterion. The co-expression network was constructed with Cytoscape software.

Network-Based Pathway Analysis

Kyoto Encyclopedia of Genes and Genomes (KEGG) knowledge database is applied to identify the functional and metabolic pathway. KEGG Orthology-Based Annotation System (KOBAS) is a software tool for automated annotation and pathway identification. DEGs in co-expression networks were analyzed in this step. KOBAS was used to identify pathways based on the hypergeometric distribution with p-value < 0.05.

Identification of Small Molecular Agents

Connectivity Map (cMap) is a public database which contains 453 Affymetrix gene expression signatures of 164 bioactive molecules. DEGs in co-expression network were compared with the data from the cMap database. The query small molecules were output with a connectivity score from +1 to −1. The high positive connectivity score indicated that the query small molecular induced the disease, while a high negative connectivity score indicated corresponding molecule suppressed disease. The |connectivity score| > 0.8 was used as the cutoff criterion.

Results

Identification of DEGs

As shown in Figure 1, the raw expression data were normalized after preprocessed. Compared
with normal samples, a total of 573 genes were differentially expressed in MM samples, including 322 down- and 251 up-regulated genes.

The heat map for DEGs was shown in Figure 2. The clustering analysis results showed that DEGs were classified into 2 clusters. Genes in cluster 1 (red bar in Figure 2) were up-regulated in MM samples, while genes in cluster 2 (blue bar in Figure 2) were down-regulated in MM samples.

**Gene Ontology Enrichment Analysis**

The GO BP terms of down- and up-regulated genes were shown in Table I. The down-regulated genes were significantly related to cell cycle, M phase and cell cycle phase. A total of 125 DEGs were enriched in the GO term of cell cycle. On the other hand, the up-regulated genes were mainly enriched in immune response and response to the virus. A total of 23 up-regulated genes were identified in immune response function.

**Construction of Co-Expression Network**

As shown in Figure 3, the cell cycle-related network (A) was constructed with 35 nodes and 213 edges (co-expresses value > 0.8). The immune response-related network (B) was constructed with 13 nodes and 23 edges and up-regulated gene guanylate binding protein 1, interferon-inducible (GBP1, degree = 8) was a hub node in this network.

**Network-Based Pathway Analysis**

A total of 2 pathways enriched by down-regulated genes were obtained in this analysis (Table II), including cell cycle and oocyte meiosis pathways. Genes such as checkpoint kinase 1 (CHEK1), MAD2 mitotic arrest deficient-like 1 (MAD2L1), DBF4 zinc finger (DBF4), TTK protein kinase (TTK, alias MSP1), extra spindle pole bodies homolog 1 (ESPL1) and cyclin A2 (CCNA2) were identified in cell cycle pathway.
Biomarkers and agents in MM

Figure 2. Heat map and clustering analysis of the differentially expressed genes (DEGs) between multiple myeloma and normal samples. The dendrogram to the left of the heat map shows clustering of the DEGs. The dendrogram above the heat map shows clustering of the samples. Red color represents high expression level, and blue color represents low expression level. The DEGs were mainly classified into 2 clusters. The red bar represents the cluster 1 and the blue bar represents the cluster 2.

Table I. The Gene Ontology (GO) biological process terms enriched by down- and up-regulated differentially expressed genes (DEGs).

<table>
<thead>
<tr>
<th>GO category</th>
<th>GO term</th>
<th>Count</th>
<th>p-value</th>
<th>FDR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Down-regulated DEGs</td>
<td>Cell cycle</td>
<td>125</td>
<td>2.65E-85</td>
<td>4.43E-82</td>
</tr>
<tr>
<td>GO:0007049</td>
<td>M phase</td>
<td>89</td>
<td>9.32E-79</td>
<td>1.56E-75</td>
</tr>
<tr>
<td>GO:00022403</td>
<td>Cell cycle phase</td>
<td>95</td>
<td>3.06E-77</td>
<td>5.12E-74</td>
</tr>
<tr>
<td>GO:00022402</td>
<td>Cell cycle process</td>
<td>103</td>
<td>9.49E-74</td>
<td>1.59E-70</td>
</tr>
<tr>
<td>GO:0007067</td>
<td>Mitosis</td>
<td>68</td>
<td>2.70E-63</td>
<td>4.51E-60</td>
</tr>
<tr>
<td>GO:0000280</td>
<td>Nuclear division</td>
<td>68</td>
<td>2.70E-63</td>
<td>4.51E-60</td>
</tr>
<tr>
<td>GO:0000087</td>
<td>M phase of mitotic cell cycle</td>
<td>68</td>
<td>1.07E-62</td>
<td>1.78E-59</td>
</tr>
<tr>
<td>GO:00048285</td>
<td>Organelle fission</td>
<td>68</td>
<td>5.70E-62</td>
<td>9.53E-59</td>
</tr>
<tr>
<td>GO:00051301</td>
<td>Cell division</td>
<td>74</td>
<td>7.07E-62</td>
<td>1.18E-58</td>
</tr>
<tr>
<td>GO:0000278</td>
<td>Mitotic cell cycle</td>
<td>80</td>
<td>8.10E-62</td>
<td>1.35E-58</td>
</tr>
<tr>
<td>GO:0006259</td>
<td>DNA metabolic process</td>
<td>71</td>
<td>4.65E-41</td>
<td>7.77E-38</td>
</tr>
<tr>
<td>GO:0007049</td>
<td>Cell cycle</td>
<td>125</td>
<td>2.65E-85</td>
<td>4.43E-82</td>
</tr>
<tr>
<td>GO:0000279</td>
<td>M phase</td>
<td>89</td>
<td>9.32E-79</td>
<td>1.56E-75</td>
</tr>
<tr>
<td>GO:00022403</td>
<td>Cell cycle phase</td>
<td>95</td>
<td>3.06E-77</td>
<td>5.12E-74</td>
</tr>
<tr>
<td>GO:00022402</td>
<td>Cell cycle process</td>
<td>103</td>
<td>9.49E-74</td>
<td>1.59E-70</td>
</tr>
<tr>
<td>GO:0007067</td>
<td>Mitosis</td>
<td>68</td>
<td>2.70E-63</td>
<td>4.51E-60</td>
</tr>
<tr>
<td>GO:0000280</td>
<td>Nuclear division</td>
<td>68</td>
<td>2.70E-63</td>
<td>4.51E-60</td>
</tr>
<tr>
<td>Up-regulated DEGs</td>
<td>Immune response</td>
<td>23</td>
<td>6.95E-05</td>
<td>0.011488</td>
</tr>
<tr>
<td>GO:0006955</td>
<td>Response to virus</td>
<td>10</td>
<td>1.06E-05</td>
<td>0.017506</td>
</tr>
<tr>
<td>GO:0001568</td>
<td>Blood vessel development</td>
<td>11</td>
<td>0.001215</td>
<td>0.019902</td>
</tr>
<tr>
<td>GO:0001944</td>
<td>Vasculature development</td>
<td>11</td>
<td>0.004165</td>
<td>0.023955</td>
</tr>
<tr>
<td>GO:0007389</td>
<td>Pattern specification process</td>
<td>11</td>
<td>0.002301</td>
<td>0.037382</td>
</tr>
</tbody>
</table>

Count: enriched gene number in the GO category; FDR: false discovery rate.
X.-G. Wang, Y. Peng, X.-L. Song, J.-P. Lan

Identification of Small Molecular Agents

A total of 11 small molecular agents were obtained, such as vinblastine, MS-275, menadione and quinostatin. As shown in Table III, the small molecular agent vinblastine (connectivity score = -0.922) had the highest negative score.

Discussion

In this study, the gene expression profile data of GSE36474 were downloaded from GEO database to identify DEGs between MM and normal samples using bioinformatics analysis. In total, 573 DEGs including 322 down- and 251 up-regulated genes were selected in MM samples. Down-regulated genes were related to cell cycle, while up-regulated genes were mainly enriched in the immune response. In order to better understand the interactions of DEGs, we established the co-expression network with significant gene pairs. The KEGG pathway analysis indicated that cell cycle was the most enriched pathway, which was responding to the results of GO analysis. Down-regulated genes such as CHEK1, MAD2L1 and DBF4 were identified in this function. Moreover, up-regulated gene GBP1 was a hub node in

Table II. The Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis of differentially expressed genes in the co-expression networks.

<table>
<thead>
<tr>
<th>KEGG pathway term</th>
<th>Genes</th>
<th>p-value</th>
<th>FDR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell cycle</td>
<td>MAD2L1, DBF4, TTK, CHEK1, ESPL1, CCNA2</td>
<td>1.89E-06</td>
<td>0.001089</td>
</tr>
<tr>
<td>Oocyte meiosis</td>
<td>MAD2L1, SGOL1, FBXO5, ESPL1, AURKA</td>
<td>3.94E-05</td>
<td>0.022668</td>
</tr>
</tbody>
</table>

FDR: false discovery rate.

Table III. The list of small molecular agents.

<table>
<thead>
<tr>
<th>cMap name</th>
<th>Score</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vinblastine</td>
<td>-0.922</td>
<td>0.00004</td>
</tr>
<tr>
<td>Irinotecan</td>
<td>0.904</td>
<td>0.00176</td>
</tr>
<tr>
<td>Withaferin A</td>
<td>0.905</td>
<td>0.00008</td>
</tr>
<tr>
<td>Puromycin</td>
<td>0.906</td>
<td>0.00008</td>
</tr>
<tr>
<td>Rottlerin</td>
<td>0.921</td>
<td>0.00112</td>
</tr>
<tr>
<td>I,4-chrysenquinone</td>
<td>0.935</td>
<td>0.00817</td>
</tr>
<tr>
<td>Piperlongumine</td>
<td>0.955</td>
<td>0.00374</td>
</tr>
<tr>
<td>Quinostatin</td>
<td>0.962</td>
<td>0.00256</td>
</tr>
<tr>
<td>0297417-0002B</td>
<td>0.969</td>
<td>0.00004</td>
</tr>
<tr>
<td>Menadione</td>
<td>0.972</td>
<td>0.00133</td>
</tr>
<tr>
<td>MS-275</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

Figure 3. The co-expression networks. A, Cell cycle-related network. Blue node represents down-regulated genes. B, Immune response-related network. Red node represents up-regulated genes.
the immune response-related network. In addition, small molecular agent vinblastine was identified as one important potential therapeutic agent in MM.

Uncontrolled cell proliferation is the hallmark of MM, and tumor cell has typically acquired damage to genes that directly regulate their cell cycles. CHEK1 has been reported to regulate the G1/S phase in human MM cells. CHEK1 (aka CHK1) encodes the serine/threonine protein kinase. Inhibition of CHEK1 diminishes the S phase arrest and enhances the apoptosis in MM. In this study, we found that CHEK1 was decreased in MM samples and it was related to cell cycle function. The above studies demonstrated that CHEK1, as a negative regulator of apoptosis, might be a molecular target for MM therapeutics.

MAD2L1 plays an important role in maintaining the mitotic spindle checkpoint function, which is corresponding to the GO term of the cell cycle in our study. Genetic variant in MAD2L1 confers susceptibility in lung cancer, which results from reduced spindle checkpoint function due to attenuated function of MAD2L1. Decreased expression of MAD2L1 is reported in breast cancer cell lines exhibiting chromosome instability and aneuploidy. The work of Shaughnessy et al. found that MAD2L1 was a significant gene related to chromosome segregation (a phase of the cell cycle) in MM. In this study, MAD2L1 was a down-regulated gene in MM and enriched in the most significant function of cell cycle. Thus, we inferred that MAD2L1 might be associated with MM development via mediating cell cycle. It may be a potential therapeutic biomarker in MM.

DBF4 encoding a serine-threonine kinase has been reported to interact with cell division cycle 7-related protein kinase (CDC7) and CDK7. CDC7-DBF4 complex is an essential kinase for regulating initiation of DNA replication and plays a central role in cell proliferation. Bonte et al. reported that CDC7-DBF4 overexpression in multiple cancers and tumor cell lines correlated with p53 inactivation. p53, a tumor suppressor, plays a key role in inhibiting the tumor growth. Inactivation of p53 has been reported to be related to MM development or progression. In the present study, DBF4 was a down-regulated gene, suggesting that reduced expression of CDC7-DBF4 might lead to p53 activation. Therefore, we inferred that DBF4 might be effective in controlling tumor growth and cell proliferation in MM.

In addition to cell cycle-related genes known to be down-regulated in MM, up-regulated gene GBP1 was a hub node in the immune response-related network. GBP1 encodes an interferon-inducible GTPase that belongs to the dynamin family of guanine nucleotide binding proteins. The immune response gene GBP1 is the key mediator of the inhibitory effects of inflammatory cytokines (e.g., interleukin-6, transforming growth factor-beta and interleukin-17) on endothelial cells proliferation and invasiveness. Elevated interleukin-17 promotes myeloma cell growth and inhibits immune function in MM. It also has been reported that GBP1 expression is inhibited by vascular endothelial growth factor (VEGF). VEGF stimulates proliferation and migration of MM cells in both autocrine and paracrine mechanisms. GBP1 was found to be up-regulated in MM and closely associated with immune response. Overexpressed GBP1 appeared to decrease MM cell proliferation and growth to some extent.

Furthermore, vinblastine was identified to be the significant small molecular agent in the development of MM. Vinblastine, an antimicrotubule drug, belongs to vinca alkaloids. It has been reported that vinca alkaloids inhibit cell proliferation by altering the dynamics of tubulin addition and loss at the ends of mitotic microtubules. Vinblastine has been widely used as an antineoplastic drug to treat certain kinds of cancer, including breast cancer, Hodgkin’s lymphoma, non-small cell lung cancer, head and neck cancer, and testicular cancer. However, the evidence concerning the drug application of vinblastine in MM is rare. In our study, the small molecular drug vinblastine had the highest negative score, implying the capable of reversing MM. Therefore, it may be a potential novel agent for the treatment of MM. Further animal experiments and clinical trials are needed to investigate the usage and dosage of vinblastine in MM.

**Conclusions**

Our study shows that cell cycle may be closely associated with MM development. The genes such as CHEK1, MAD2L1, DBF4 and GBP1 may be the potential therapeutic biomarkers in MM. Vinblastine may be a potential therapeutic agent in MM. These findings would supply new insights for the diagnosis and treatment of MM. However, further experiments with larger sample size are still needed to confirm our results.
Conflict of Interest
The Authors declare that there are no conflicts of interest.

References
4) Institute NC. See stat fact sheets: Myeloma 2014.


