Identification of the key genes connected with plasma cells of multiple myeloma using expression profiles

Kefeng Zhang, Zhongyang Xu, Zhaoyun Sun

1Spinal Surgery, Jining No 1 People’s Hospital, Jining; 2Department of Orthopedics, The People’s Hospital of Laiwu City, Laiwu, Shandong Province, People’s Republic of China

Objective: To uncover the potential regulatory mechanisms of the relevant genes that contribute to the prognosis and prevention of multiple myeloma (MM).

Methods: Microarray data (GSE13591) were downloaded, including five plasma cell samples from normal donors and 133 plasma cell samples from MM patients. Differentially expressed genes (DEGs) were identified by Student’s t-test. Functional enrichment analysis was performed for DEGs using the Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases. Transcription factors and tumor-associated genes were also explored by mapping genes in the TRANSFAC, the tumor suppressor gene (TSGene), and tumor-associated gene (TAG) databases. A protein–protein interaction (PPI) network and PPI subnetworks were constructed by Cytoscape software using the Search Tool for the Retrieval of Interacting Genes (STRING) database.

Results: A total of 63 DEGs (42 downregulated, 21 upregulated) were identified. Functional enrichment analysis showed that HLA-DRB1 and VCAM1 might be involved in the positive regulation of immune system processes, and HLA-DRB1 might be related to the intestinal immune network for IgA production pathway. The genes CEBPD, JUND, and ATF3 were identified as transcription factors. The top ten nodal genes in the PPI network were revealed including HLA-DRB1, VCAM1, and TFRC. In addition, genes in the PPI subnetwork, such as HLA-DRB1 and VCAM1, were enriched in the cell adhesion molecules pathway, whereas CD4 and TFRC were both enriched in the hematopoietic cell pathway.

Conclusion: Several crucial genes correlated to MM were identified, including CD4, HLA-DRB1, TFRC, and VCAM1, which might exert their roles in MM progression via immune-mediated pathways. There might be certain regulatory correlations between HLA-DRB1, CD4, and TFRC.

Keywords: multiple myeloma, functional enrichment, transcription factor analysis, PPI network, pathway enrichment

Introduction

Multiple myeloma (MM) is a malignant plasma cell (PC) disorder and is characterized by the infiltration of malignant PCs, which are the endpoint of B-cell differentiation, in the bone marrow.1-4 The mortality of MM is relatively high; approximately 63,000 individuals die from MM annually, accounting for 0.9% of all cancer deaths.5 Therefore, many studies have been conducted to explore the molecular mechanisms of MM etiology.6,7

MM is known to be associated with heterogeneous cytogenetic abnormalities, including hyperdiploid and nonhyperdiploid karyotypes.8 In patients with a nonhy-

hyperdiploid karyotype, many chromosomal translocations are involved, which could...
lead to the altered expression of related proteins (cyclin D1 and cyclin D3) and genes (WHSC1 and MAI). Additionally, the interaction of genes and proteins are also implicated in the pathogenesis of MM. For instance, a previous study demonstrated that IL-6 contributes to the pathogenesis of MM and that the concentrations of insulin-like growth factor binding protein-1 and the soluble IL-6 receptor are positively associated with the risk of MM.9

Despite these significant discoveries, due to the difficulty of obtaining the metaphases of malignant PC clone, there is limited knowledge regarding the genetics of MM due to the difficulty of obtaining malignant PC clones in metaphase. Fortunately, the application of advanced molecular techniques such as microarrays and next-generation sequencing facilitate the improvement of our understanding from a genetic level10 Egan et al11 and Chapman et al12 have analyzed the genomic events that initiate MM using the genomic sequencing.

Microarray data (GSE13591) were established by Agnelli et al13 who analyzed these data using a combined FISH and microarray approach and identified near-tetraploidy as a hallmark of the tumor. They also highlighted that loss of heterozygosity is a prominent mechanism in the regulation of mRNA and gene expressions. In their further studies based on these microarray data, the authors provide an elaborate elucidation of the Hedgehog pathway in MM, laying an elaborate foundation for the use of Hedgehog inhibitor detection in clinical trials.14 Other investigations based on these microarray data (GSE13591) have also been carried out. Lionetti et al15 defined the microRNA/mRNA regulatory network of MM. HOXB7 expression in MM is identified by both microarray data (GSE13591) analysis and further verification of experiments.16,17 Moreover, another study reconstructed gene regulatory networks by combining gene expression profiles from seven publicly available datasets; however, the emphasis of this study was to identify critical genes that predict overall survival in the prognosis of MM.18

Nevertheless, none of the aforementioned studies further analyzed the functions, pathways, or the potential correlations of the identified differentially expressed genes (DEGs) between normal and MM PCs. For this study, we downloaded the microarray data (GSE13591) and reanalyzed them using bioinformatics methods including the identification of DEGs in MM, as well as performed the Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), protein–protein interaction (PPI), and PPI subnetwork analyses of these DEGs. These approaches are based on multiple bioinformatic packages, which employ powerful statistical methods,19,20 thus facilitating the identification of critical DEGs and their pathways that are involved in the development of normal PCs into MM. In addition, along with predicting the interactions of the DEGs at the protein level, we aimed to uncover the potential regulatory mechanisms of the relevant genes that contribute to the prognosis and prevention of MM.

Methods

Microarray data and data preprocessing

The gene expression profile GSE13591, deposited by Agnelli et al13 was downloaded from the Gene Expression Omnibus (GEO) database.21 The microarray platform was the GPL96 (HG-U133A) Affymetrix Human Genome U133A Array (Affymetrix, Inc., Santa Clara, CA, USA). The expression profile included 138 samples, comprising five PC samples obtained from normal donors (control group) and 133 PC samples derived from MM patients (MM group).

The raw expression profile data were preprocessed using the Affy package in Bioconductor22 and the Affymetrix annotation files from the Brain Array Lab (Microarray Lab, University of Michigan, Ann Arbor, MI, USA). The background correction, quantile data normalization, and probe summarization were performed by the Robust Multi-array Average algorithm to obtain a gene expression matrix.23

Identification of DEGs

The DEGs between the MM and control groups were identified based on Student’s t-test in the linear models for the microarray data (Limma) package.24 Raw P-values were adjusted into false discovery rates (FDRs) by the Benjamini and Hochberg method.25 The thresholds for DEG selection were FDR < 0.05 and |log2 FC (fold change)| > 1.

Functional enrichment analysis of DEGs

The GO26,27 and KEGG28 pathway enrichment analyses were performed to identify the biological functions and pathways of the identified DEGs, based on the hypergeometric distribution algorithm. P<0.05 was chosen as the cut-off value for enriched functions and pathways.

Gene function annotation of DEGs

By mapping the DEGs using the TRANSFAC database,30 the transcription factors among the DEGs were identified. Combined with information from the tumor suppressor gene (TSGene)31 and tumor-associated gene (TAG) databases,32 all known oncogenes and tumor suppressor genes among the DEGs were extracted.
PPI network construction and module analyses

The DEGs were mapped using the Search Tool for the Retrieval of Interacting Genes (STRING) database to reveal potential correlations between the DEGs at the protein level. Only the interaction pairs that contained at least one DEG were selected to construct the PPI network, with the cut-off value of a confidence score >0.85. Cytoscape software was used to visualize the PPI network. Furthermore, the PPI subnetwork was identified using BioNet, and FDR <0.01 was chosen as the cut-off value. The KEGG pathway enrichment analysis of the genes in the PPI subnetwork was performed.

Results

DEGs between the MM and control groups

A total of 166 transcripts, corresponding to 63 DEGs, including 42 downregulated DEGs and 21 upregulated DEGs, were identified between the MM and control groups, as shown in the heat map of the cluster analysis of DEGs (Figure 1).

Enrichment analysis of DEGs

GO enrichment analysis revealed that 262 functional terms for downregulated DEGs and 78 for upregulated DEGs were enriched; the top ten functional terms ranked by statistical significance were listed in Table 1 (for down- and

![Color Key](https://example.com/color-key.png)

**Figure 1** The heat map of the cluster analysis of differentially expressed genes.

**Note:** The X-axis represents samples, while the Y-axis represents genes.
upregulated DEGs, respectively). As presented in Table 1, the top three significantly enriched GO terms for downregulated DEGs were regulation of immune system processes, immune response, and positive regulation of immune system processes (in which HLA-DRB1, VCAM1, A2M, CXCL12, and HMOX1 might be involved; \( P=9.99\times10^{-16} \)), whereas the top three for upregulated DEGs were primary metabolic processes (in which ATF3, CCNB1IP1, DBF4, DUSP5, and EEF1E1 may participate; \( P=0.0008866 \)), organic substance metabolic processes (\( P=0.0014942 \)), and response to axon injury (\( P=0.0016905 \)).

The KEGG pathway enrichment analysis revealed that 22 pathways for downregulated DEGs and two pathways for upregulated DEGs were enriched. The top ten significant pathways for downregulated DEGs and the two pathways for upregulated DEGs were shown in Table 2. Among them, the top three KEGG pathways for downregulated DEGs were intestinal immune network for immunoglobulin (Ig) A production (\( P=5.55\times10^{-14} \)), antigen processing and presentation (\( P=7.19\times10^{-14} \)), and asthma (\( P=7.84\times10^{-14} \)). The enrichment result implied that the genes HLA-DMA, HLA-DPA1, HLA-DBB1, HLA-DQA1, HLA-DQA2, HLA-DRB1, HLA-DRB3, and HLA-DRB4 might be involved in the top three pathways. The two enriched pathways for upregulated genes were the mitogen-activated protein kinase signaling pathway (\( P=0.0088912 \)) involving DUSP5, GADD45A, and JUND, and the cell cycle (\( P=0.0178289 \)), involving DBF4 and GADD45A.

### Analysis of gene function annotation

By integrating the information in the TRANSFAC database, there were three transcription factors identified, including one downregulated gene (CEBP) and two upregulated genes (JUND and ATF3). Additionally, three TAGs, including one downregulated gene (CEBP) and two upregulated genes (JUND and EEF1E1), were selected.
Table 2 The top ten enriched KEGG pathways of downregulated DEGs and the two enriched KEGG pathways of upregulated DEGs

<table>
<thead>
<tr>
<th>DEGs</th>
<th>KEGG pathway</th>
<th>Gene counts</th>
<th>P-value</th>
<th>Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Downregulated</td>
<td>Intestinal immune network for IgA production</td>
<td>9</td>
<td>5.55E-14</td>
<td>CXCL12, HLA-DMA, HLA-DPA1, HLA-DBP1, HLA-DQA1, HLA-DQA2, HLA-DRB1, HLA-DRB3, HLA-DRB4</td>
</tr>
<tr>
<td></td>
<td>Antigen processing and presentation</td>
<td>10</td>
<td>7.19E-14</td>
<td>CTSB, HLA-DMA, HLA-DPA1, HLA-DBP1, HLA-DQA1, HLA-DQA2, HLA-DRB1, HLA-DRB3, HLA-DRB4, HSPA1A</td>
</tr>
<tr>
<td></td>
<td>Asthma</td>
<td>8</td>
<td>7.84E-14</td>
<td>HLA-DMA, HLA-DPA1, HLA-DBP1, HLA-DQA1, HLA-DQA2, HLA-DRB1, HLA-DRB3, HLA-DRB4</td>
</tr>
<tr>
<td></td>
<td>Staphylococcus aureus infection</td>
<td>9</td>
<td>2.08E-13</td>
<td>C1QA, HLA-DMA, HLA-DPA1, HLA-DBP1, HLA-DQA1, HLA-DQA2, HLA-DRB1, HLA-DRB3, HLA-DRB4</td>
</tr>
<tr>
<td></td>
<td>Allograft rejection</td>
<td>8</td>
<td>5.09E-13</td>
<td>HLA-DMA, HLA-DPA1, HLA-DBP1, HLA-DQA1, HLA-DQA2, HLA-DRB1, HLA-DRB3, HLA-DRB4</td>
</tr>
<tr>
<td></td>
<td>Graft-versus-host disease</td>
<td>8</td>
<td>1.25E-12</td>
<td>HLA-DMA, HLA-DPA1, HLA-DBP1, HLA-DQA1, HLA-DQA2, HLA-DRB1, HLA-DRB3, HLA-DRB4</td>
</tr>
<tr>
<td></td>
<td>Type 1 diabetes mellitus</td>
<td>8</td>
<td>1.89E-12</td>
<td>HLA-DMA, HLA-DPA1, HLA-DBP1, HLA-DQA1, HLA-DQA2, HLA-DRB1, HLA-DRB3, HLA-DRB4</td>
</tr>
<tr>
<td></td>
<td>Autoimmune thyroid disease</td>
<td>8</td>
<td>9.59E-12</td>
<td>HLA-DMA, HLA-DPA1, HLA-DBP1, HLA-DQA1, HLA-DQA2, HLA-DRB1, HLA-DRB3, HLA-DRB4</td>
</tr>
<tr>
<td></td>
<td>Rheumatoid arthritis</td>
<td>9</td>
<td>2.37E-11</td>
<td>CXCL12, HLA-DMA, HLA-DPA1, HLA-DBP1, HLA-DQA1, HLA-DQA2, HLA-DRB1, HLA-DRB3, HLA-DRB4</td>
</tr>
<tr>
<td></td>
<td>Viral myocarditis</td>
<td>8</td>
<td>1.15E-10</td>
<td>HLA-DMA, HLA-DPA1, HLA-DBP1, HLA-DQA1, HLA-DQA2, HLA-DRB1, HLA-DRB3, HLA-DRB4</td>
</tr>
<tr>
<td>Upregulated</td>
<td>MAPK signaling pathway</td>
<td>3</td>
<td>0.0088912</td>
<td>DUSP5, GADD45A, JUND</td>
</tr>
<tr>
<td></td>
<td>Cell cycle</td>
<td>2</td>
<td>0.0178289</td>
<td>DBF4, GADD45A</td>
</tr>
</tbody>
</table>

**Abbreviations:** KEGG, Kyoto Encyclopedia of Genes and Genomes; DEGs, differentially expressed genes; IgA, immunoglobulin A; MAPK, mitogen-activated protein kinase.

PPI network and subnetwork analysis

The PPI network of genes was constructed, as shown in Figure 2. The top ten genes/proteins with the highest degrees (the numbers of PPI pairs that the gene/protein possessed) in the PPI network were HLA-DRB1, VCAM1, A2M, CXCL12, CEBPD, HMOX1, CD79A, TFRC, CD19, and GADD45A, with the degrees of 113, 88, 82, 70, 67, 46, 44, 42, 37, and 35, respectively. The PPI subnetwork of genes included 15 nodal genes, of which the top two were HLA-DRB1 and VCAM1, both with a degree of 4, while the degree of TFRC was 1 (Figure 3).

For the genes identified in the PPI subnetwork, a total of 20 KEGG pathways were enriched (Table 3). The top three pathways were cell adhesion molecules (P=3.86E-06; in which HLA-DRB1, CD4, HLA-DMA, VCAM1, and HLA-DMB might be involved), the B-cell receptor signaling pathway (P=1.13E-05; in which CD19, CD81, CD79A, and INPP5D might participate), and antigen processing and presentation (P=1.19E-05, in which HLA-DRB1, CD4, HLA-DMA, and HLA-DMB might be involved).

Discussion

In this study, a total of 63 DEGs (42 downregulated, 21 upregulated) between the MM and control groups were identified. Based on functional and pathway enrichment analyses by GO and KEGG, genes such as HLA-DRB1 and VCAM1 might be involved in the positive regulation of immune system processes; meanwhile, HLA-DRB1 might also participate in the intestinal immune network for the IgA production pathway.

In the PPI network, genes such as HLA-DRB1 and TFRC were ranked in the top ten nodal genes with high degrees. In the PPI subnetwork, critical nodal genes, including HLA-DRB1, VCAM1, CD4, and TFRC, were identified.

**CD4** encodes a T-lymphocyte membrane glycoprotein, which is a receptor for human immunodeficiency virus and functions to initiate or augment the early phase of T-cell activation. In patients with MM, although impaired T-cell function and viral-specific cytotoxic T-lymphocytes have been reported, the presence of expanded T-cell populations such as clonal CD4+ T-cells has also been detected. CD4 plays significant roles in immune responses mediated by cells and antibodies. T-helper 17 cells (T+17) are a subset of CD4+ T-cells) have been suggested as a therapeutic target in MM to improve immune function, as the number of T+17 cells is remarkably increased in MM, accompanied by elevated levels of several proinflammatory cytokines. The antigen specificity of T-cells makes great sense for the immune system. In the present study, CD4 was predicted as a crucial nodal gene in the PPI subnetwork and it was implicated in the antigen processing and presentation pathway, indicating...
Figure 2 Protein–protein interaction network of genes.

Note: The red nodes represent upregulated differentially expressed genes and the green nodes represent downregulated differentially expressed genes, while the yellow nodes represent no differentially expressed genes.
Table 3 The 20 enriched KEGG pathways of the protein–protein interaction subnetwork

<table>
<thead>
<tr>
<th>KEGG pathway</th>
<th>Gene counts</th>
<th>P-value</th>
<th>Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell adhesion molecules</td>
<td>5</td>
<td>3.86E-06</td>
<td>HLA-DRB1, CD4, HLA-DMA, VCAM1, HLA-DMB</td>
</tr>
<tr>
<td>B-cell receptor signaling pathway</td>
<td>4</td>
<td>1.13E-05</td>
<td>CD19, CD81, CD79A, INPP5D</td>
</tr>
<tr>
<td>Antigen processing and presentation</td>
<td>4</td>
<td>1.19E-05</td>
<td>HLA-DRB1, CD4, HLA-DMA, HLA-DMB</td>
</tr>
<tr>
<td>Hematopoietic cell lineage</td>
<td>4</td>
<td>2.13E-05</td>
<td>CD19, HLA-DRB1, CD4, TFRC</td>
</tr>
<tr>
<td>Rheumatoid arthritis</td>
<td>4</td>
<td>2.43E-05</td>
<td>HLA-DRB1, IL18, HLA-DMA, HLA-DMB</td>
</tr>
<tr>
<td>Asthma</td>
<td>3</td>
<td>2.57E-05</td>
<td>HLA-DRB1, HLA-DMA, HLA-DMB</td>
</tr>
<tr>
<td>Primary immunodeficiency</td>
<td>3</td>
<td>4.12E-05</td>
<td>CD19, CD79A, CD4</td>
</tr>
<tr>
<td>Allograft rejection</td>
<td>3</td>
<td>4.88E-05</td>
<td>HLA-DRB1, HLA-DMA, HLA-DMB</td>
</tr>
<tr>
<td>Graft-versus-host disease</td>
<td>3</td>
<td>6.66E-05</td>
<td>HLA-DRB1, HLA-DMA, HLA-DMB</td>
</tr>
<tr>
<td>Type 1 diabetes mellitus</td>
<td>3</td>
<td>7.70E-05</td>
<td>HLA-DRB1, HLA-DMA, HLA-DMB</td>
</tr>
<tr>
<td>Intestinal immune network for IgA production</td>
<td>3</td>
<td>0.000107226</td>
<td>HLA-DRB1, HLA-DMA, HLA-DMB</td>
</tr>
<tr>
<td>Malaria</td>
<td>3</td>
<td>0.000128658</td>
<td>CD81, IL18, VCAM1</td>
</tr>
<tr>
<td>Autoimmune thyroid disease</td>
<td>3</td>
<td>0.000136377</td>
<td>HLA-DRB1, HLA-DMA, HLA-DMB</td>
</tr>
<tr>
<td>Staphylococcus aureus infection</td>
<td>3</td>
<td>0.000161334</td>
<td>HLA-DRB1, HLA-DMA, HLA-DMB</td>
</tr>
<tr>
<td>Phagosome</td>
<td>4</td>
<td>0.000186619</td>
<td>HLA-DRB1, HLA-DMA, TRC, HLA-DMB</td>
</tr>
<tr>
<td>Viral myocarditis</td>
<td>3</td>
<td>0.000330845</td>
<td>HLA-DRB1, HLA-DMA, HLA-DMB</td>
</tr>
<tr>
<td>Leishmaniasis</td>
<td>3</td>
<td>0.000359628</td>
<td>HLA-DRB1, HLA-DMA, HLA-DMB</td>
</tr>
<tr>
<td>Toxoplasmosis</td>
<td>3</td>
<td>0.002107591</td>
<td>HLA-DRB1, HLA-DMA, HLA-DMB</td>
</tr>
<tr>
<td>African trypanosomiastis</td>
<td>2</td>
<td>0.002195842</td>
<td>IL18, VCAM1</td>
</tr>
<tr>
<td>Systemic lupus erythematosus</td>
<td>3</td>
<td>0.002295939</td>
<td>HLA-DRB1, HLA-DMA, HLA-DMB</td>
</tr>
</tbody>
</table>

Abbreviations: KEGG, Kyoto Encyclopedia of Genes and Genomes; IgA, immunoglobulin A.

Figure 3 Protein–protein interaction in the subnetwork of genes.

Notes: The red nodes represent upregulated differentially expressed genes and the green represent downregulated differentially expressed genes, while the white indicate no differentially expressed genes. The fold change of gene expression is shown in color (deeper color indicates higher fold change of gene expression). The square nodes denote genes with lower importance in the subnetwork, and round nodes denote genes with higher importance.
that this gene might play a central role in the immune system through the involvement of this pathway in the development of MM cells from normal PCs.

**HLA-DRB1** (major histocompatibility complex, class II, DR Beta 1), which encodes more than 1,196 alleles, is a polymorphic locus of the HLA system.\textsuperscript{41} It is a member of the HLA class II alpha chain paralogs, which comprise an alpha (DPA) and a beta (DBP) chain and are expressed in antigen-presenting cells.\textsuperscript{44,45} Via presenting peptides from extracellular proteins, HLA-DRB1 plays a pivotal role in the immune system.\textsuperscript{46} For instance, it could bind peptides from antigens and present them on the cell surface to be recognized by CD4 T-cells.\textsuperscript{57} Additionally, HLA-DPB1 alleles are found to be positively associated with IgA deficiency in German individuals.\textsuperscript{48} Although few studies exist regarding the immune function of HLA-DPB1 related to MM, our pathway analysis suggests that HLA-DRB1 might be involved in the intestinal immune network for the IgA production pathway, implying that HLA-DPB1 might exert a regulatory role in MM development via participation in this pathway. Notably, combined with the potential role of CD4 and the predicted correlations in the PPI subnetwork, a regulatory relationship between HLA-DPB1 and CD4 might exist.

Transferrin receptor (TFRC) is closely related to cellular iron intake.\textsuperscript{49} TFRC is also related to the immune response.\textsuperscript{50} The blocking of CD4 results in an increasing peptide dose of CD22 and consequently has an intermediate effect on the cell adhesion molecule pathway.\textsuperscript{51} CD4 might exert a regulatory role in MM development via participation in this pathway. Notably, combined with the potential role of CD4 and the predicted correlations in the PPI subnetwork, a regulatory relationship between HLA-DPB1 and CD4 might exist.

Notably, certain regulatory correlations might exist between HLA-DRB1, CD4, and TFRC. However, because the results are based on microarray data with a small sample size, more experimental validations are warranted.

**Disclosure**

The authors report no conflicts of interest in this work.

**References**


