Pathway-based network analysis of myeloma tumors: monoclonal gammopathy of unknown significance, smoldering multiple myeloma, and multiple myeloma

L. Dong[1], C.Y. Chen[2], B. Ning[3], D.L. Xu[3], J.H. Gao[3], L.L. Wang[3], S.Y. Yan[3] and S. Cheng[3]

[1]Department of Hematology,
Qian Fo Shan Hospital Affiliated to Shandong University, Jinan, China
[2]Department of Hematology,
Qilu Hospital of Shandong University, Jinan, China
[3]Department of Osteology,
Peoples' Hospital of Dongying City in Shandong Province, Dongying, China

Corresponding author: B. Ning
E-mail: ningbomed@yeah.net

ABSTRACT. Although many studies have been carried out on monoclonal gammopathy of unknown significances (MGUS), smoldering multiple myeloma (SMM), and multiple myeloma (MM), their classification and underlying pathogenesis are far from elucidated. To discover the relationships among MGUS, SMM, and MM at the transcriptome level, differentially expressed genes in MGUS, SMM, and MM were identified by the rank product method, and then co-expression networks were constructed by integrating the data. Finally, a pathway-network was constructed based on Kyoto Encyclopedia of Genes and Genomes pathway enrichment analysis, and the relationships between the pathways were identified. The results indicated that

there were 55, 78, and 138 pathways involved in the myeloma tumor developmental stages of MGUS, SMM, and MM, respectively. The biological processes identified therein were found to have a close relationship with the immune system. Processes and pathways related to the abnormal activity of DNA and RNA were also present in SMM and MM. Six common pathways were found in the whole process of myeloma tumor development. Nine pathways were shown to participate in the progression of MGUS to SMM, and prostate cancer was the sole pathway that was involved only in MGUS and MM. Pathway-network analysis might provide a new indicator for the developmental stage diagnosis of myeloma tumors.

Key words: Myeloma tumor; Biological analysis; Pathway-network

INTRODUCTION

Myeloma tumors can be divided into monoclonal gammopathy of unknown significances (MGUS), smoldering multiple myeloma (SMM), and the clinically active multiple myeloma (MM) according to their clinical features such as the content of blood monoclonal immunoglobulin, the number of plasmocytes, and the damage to peripheral target organs and kidneys (Kyle and Rajkumar, 2009). MGUS and SMM are asymptomatic plasma cell dyscrasias, with a propensity to progress to symptomatic MM (Kyle et al., 2007). In recent years, there have been improvements in the risk stratification models (including molecular markers) of both disorders, which have led to a better understanding of the biology and of the probability of progression of MGUS and SMM; however, there is no accepted secondary prevention strategy available for MGUS and SMM (Bladé and Rosinol, 2006; Hillengass et al., 2011; Korde et al., 2011). The malignant transformation rate of MGUS is 1 to 1.5% per year and it can develop into MM or other malignancies; the underlying genetic transformation might be the main factor determining the outcome (Fonseca et al., 2002). SMM carries more risk of malignant transformation than does MGUS, with the rate of progression to MM being 10 to 20%.

The differentiation between MGUS, SMM, and MM is difficult, as there are no clear markers that are able to distinguish a clonal premalignant plasma cell from a clonal malignant myeloma cell. Though many clinical studies have been performed, the molecular mechanisms and relationship among MGUS, SMM, and MM remain unclear, and the diagnosis of MM still depends on the overt clinical manifestations of serious end-stage organ damage. Malpas et al. (2004) considered that the progression of MGUS to SMM and MM is associated with an expanding bone marrow tumor mass and increasingly severe organ impairment or symptoms. Although advances have been made in research on MM pathogenesis, predicting which patients with MGUS will or will never progress to MM is still impossible. Greenberg et al. (2012) investigated whether there was a comparable risk for light-chain MGUS among 911 relatives of the same heavy-chain MGUS/MM probands versus a reference population, and indicated that the prevalence of light-chain MGUS was significantly higher among first-degree relatives of MM probands than in the reference population. Additional studies have revealed that the secretion of B cell activating factor (BAFF) and the expression level of receptors were elevated in the serum and cell lines from patients with MM. Simultaneously, the expression level of BAFF in bone marrow stromal

cells of patients with MM was significantly higher than that in their peripheral blood and cell lines, and the expression of BAFF was found to be closely related to the development of MM (Tai et al., 2006; Jiang et al., 2009; Ju et al., 2009; Shen et al., 2011). Furthermore, insulin-like growth factor (IGF)-1 and interleukin (IL)-6 were shown to promote the proliferation and growth of MM cells. Birmann et al. (2009) found that the variation of genes related to IGF-1 and IL-6 signaling had a suggestive association with MM.

With the fast-growing knowledge of the human interactome, network-based approaches have become more powerful and informative for the study of disease mechanisms (del Sol et al., 2010). It is likely that MGUS, SMM, and MM are related through shared genes, proteins, and biological processes or pathways. A thorough understanding of these connections could reveal the shared biological mechanisms and provide useful information for the diagnosis and therapy of the disease. Along these lines, computational methods have been proposed to detect disease-related networks (Miller et al., 2008; Ray and Zhang, 2010; Aung et al., 2014; Xing et al., 2014). Therefore, we investigated the differences in the gene expression profiles of MGUS, SMM, and MM, performed function and pathway enrichment analysis, and constructed a pathway network to determine the relationship between these disorders to recognize the underlying pathogenesis, which might provide novel biological markers for the diagnosis, treatment, and gene therapy of this disease.

MATERIAL AND METHODS

Date recruitment and preprocessing

The microarray datasets of MGUS, SMM, and MM were all downloaded from ArrayExpress (http://www.ebi.ac.uk/arrayexpress/). The exact sample information for each disease is shown in Table 1. The probe-level data in ArrayExpress files were converted into expression measures and then read by the Affymetrix GeneChip package (Santa Clara, CA, USA) on the platform HGU133 Plus and HGU133A; 54675 genes were eventually obtained.

Table 1. Information on the samples downloaded from ArrayExpress.

	Diseased subjects	Normal subjects	Data sources
MGUS	27	20	E-MTAB-363
			E-GEOD-6477
SMM	24	15	E-GEOD-6477
MM	239	20	E-MTAB-363
			E-GEOD-6477

MGUS = monoclonal gammopathy of unknown significance; SMM = smoldering multiple myeloma; MM = multiple myeloma.

In order to eliminate the influence of nonspecific hybridization, background correlation and normalization were carried out by the robust multi-array average (RMA) method (Ma et al., 2006) and a quartile-based algorithm (Rifai and Ridker, 2001), respectively. Perfect match and mismatch values were revised using the mas 5 algorithm (Zhang et al., 2003), the value of which was selected via the median method. Then, the data were screened by the feature filter method of the Genefilter package (Gentleman et al., 2014) (http://www.bioconductor. org/packages/release/bioc/html/genefilter.html), and the probe-sets associated with each gene

were analyzed by the maximum-based method to choose those most significantly differentially expressed. Finally, an expression profile dataset including 12,493 common genes was obtained.

Differentially expressed (DE) genes identified by rank products

The rank product is a biologically motivated test for the detection of DE genes in replicated microarray experiments. The gene expression analyses of the diseased and normal subjects in MGUS, SMM, and MM were all conducted by the rank product method (Breitling et al., 2004; Koziol, 2010), which is based on the calculation of rank products (RPs) and has a higher sensitivity and specificity than the *t*-test in detecting the DE genes. For this study, genes with RP changes of two-fold and the percentage of false-positives (pfp) lower than 0.01 were identified as DE genes. This was determined as follows:

Given *n* genes and *k* replicates, let $e_{g,I}$ be the fold change and $r_{g,i}$ the rank of gene *g* in the *i-th* replicate. The rank product was computed via the geometric mean

$$RP\,(g) = \left(\prod_{i=1}^{k} r_{g,i}\right)^{1/k} \qquad \text{(Equation 1)}$$

A simple permutation-based estimation was used to determine how often a given RP value or better was observed in a random experiment. The pfp was calculated as follows:

$$E_{RP}(g) = c/p \qquad \text{(Equation 2)}$$

$$pfp(g) = E_{RP}(g)/rank(g) \qquad \text{(Equation 3)}$$

where $E_{RP(g)}$ is the average expected value for the rank product; *p* is the permutations of *k* rank lists of length *n; pfp (g)* is the percentage of false positives; *c* is the times the rank products of the genes in the permutations are smaller or equal to the observed rank product; and rank (*g*) is the rank of gene *g* in a list of all genes sorted by increasing RP.

Construction of co-expression networks for DE genes

Many functionally related genes are frequently co-expressed and participate in similar biological processes and pathways (Lee et al., 2004); however, the investigation of differentially co-expressed (DC) genes has been hampered by the large cardinality of the space to be interrogated as well as the outlines. In order to identify the DC genes in MGUS, SMM, and MM, an empirical Bayesian approach was utilized, which provided a false discovery rate (FDR) controlled list of interesting pairs along with pair-specific posterior probabilities (Dawson and Kendziorski, 2012). First, the heterogeneity of the independent studies regarding MGUS, SMM, and MM was analyzed by using chi-square tests before the meta-analysis was performed. The heterogeneity was calculated as follows:

$$I^2 = \frac{Q - (k - 1)}{Q} \times 100\% \qquad \text{(Equation 4)}$$

$$Q = \sum W_i(d_i - \bar{d})^2 = \sum w_i\, d_i{}^2 - \frac{(\sum w_i d_i)^2}{\sum w_i} \quad \text{(Equation 5)}$$

$$w_i = \frac{1}{Var(d_i)} \quad \text{(Equation 6)}$$

where k is the number of the genes included in meta-analysis; Q is the heterogeneity card value; d_i is the standardized mean difference; and w_i is the weight of each study.

Finally, a gene pair was identified as being comprised of DC genes according to the calculation steps under a hard thresholding mechanism if the posterior probability of the DC genes exceeded a critical value that controlled the posterior expected FDR at 5%.

Functional and pathway enrichment analyses

Biological pathways represent elaborate series of cascading biochemical reactions that occur within cells and possibly receive external signals (Schilling et al., 1999). Cellular functions such as the cell cycle, cell respiration, or apoptosis are all governed by pathways. Biochemical compounds such as nucleic acids, proteins, and complexes that participate in reactions form a network of biological processes and are grouped into pathways. In order to facilitate the functional analysis and pathway enrichment analysis of the gene products that participate in the development of MM and of their relation to MGUS, SMM, and MM, Gene Ontology (GO) functional enrichment and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses were performed by using the online tool Database for Annotation, Visualization and Integrated Discovery (DAVID) (Huang et al., 2009). GO terms and KEGG pathways with P values less than 0.01 were selected based on the expression analysis systematic explorer (EASE) test (Hosack et al., 2003) implemented in DAVID.

Construction of a pathway-network based on KEGG enrichment analysis

Network-based approaches for the study of a number of diseases have provided insights into disease mechanisms, and network-based biomarkers have been shown to be superior to single gene-based and pathway-based biomarkers with respect to both accuracy and robustness (Ideker and Sharan, 2008). To reveal the relationships among MGUS, SMM, and MM and to identify the genes that play vital roles in the pathways of the disease, a pathway-network based on KEGG enrichment analysis was constructed. If, for example, there were three genes that were each present in the same pathway of two different diseases, e.g., MGUS and MM, then the count value for this pathway in MGUS and MM was calculated as 3. If the count value in one pathway was not more than 3, then the pathway interaction was regarded as a small interaction and discarded. Through this process, eventually the KEGG intersection map of MGUS, SMM, and MM was obtained.

RESULTS

Identification of DE genes

After normalization and processing of the transcription profile data for MGUS, SMM, and MM samples with the rank product method, 46 up- and 126 downregulated genes in MGUS, 69 up- and 210 downregulated genes in SMM, and 711 up- and 206 downregulated genes in MM were identified, respectively.

Construction of co-expression networks of DE genes in MGUS, SMM, and MM

Genes with similar functions usually have similar expression patterns in the co-expression network. The expression values of all screened DE genes were compared and analyzed to construct the co-expression networks. Finally, we obtained separate co-expression networks for MGUS, SMM, and MM. The transcription factors (TFs) *MAFB*, *NAF*, *TCF7L2*, *SMMAD1*, *HIF1A*, and *KLF4* were identified in the MGUS network. *MAFB*, a basic leucine zipper TF that plays an important role in the regulation of lineage-specific hematopoiesis, had the most edges with other genes (Figure 1). Only two TFs (*MAFB* and *HIF1A*) were found to be involved in the SMM network, but *MAFB* has a direct connection with *RNASE1*, which exhibits many interactions with other genes involved in the development of SMM (Figure 2). For disease that had progressed to MM, 559 genes, including 26 TFs were involved in the regulation process (Figure 3). The specific characteristics of each network are shown in Table 2.
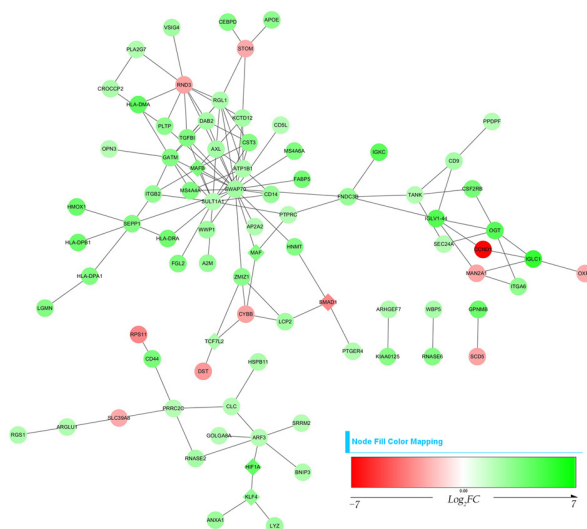
Figure 1. Co-expression network constructed for MGUS. The co-expression network shows the co-expressed genes in MGUS, and includes 84 nodes and 124 edges. The green nodes represent the upregulated genes and the red nodes represent the downregulated genes. Diamonds represent the transcription factors. From the network diagram, *MAFB*, which is a lineage-specific hematopoiesis related gene, appears to play an important role in MGUS. The color intensity corresponds to the absolute values of the log2-transformed fold-change of co-expressed genes (up- or downregulated). Most of the genes were upregulated. The black edges indicate the correlation of gene pairs that were identified as differentially co-expressed (DC) genes under a hard thresholding mechanism (FDR at 5%). MGUS = monoclonal gammopathy of unknown significance; FDR = false discovery rate.
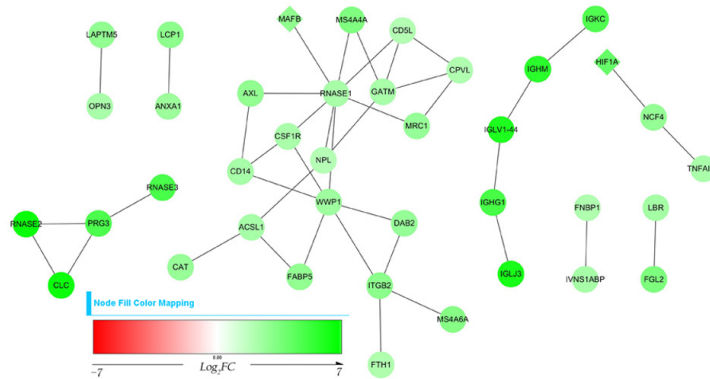
Figure 2. Co-expression network constructed for SMM. The co-expression network shows the co-expressed genes in SMM, and includes 39 nodes and 41 edges. The green nodes represent the upregulated genes and the diamonds represent the transcription factors. The black edges indicate the correlation of gene pairs that were identified as differentially co-expressed (DC) genes under a hard thresholding mechanism (FDR at 5%). The color intensity corresponds to the absolute values of the log2-transformed fold change of DC genes (up- or downregulated); all the genes in the co-expression network of SMM were upregulated. SMM = smoldering multiple myeloma; FDR = false discovery rate.
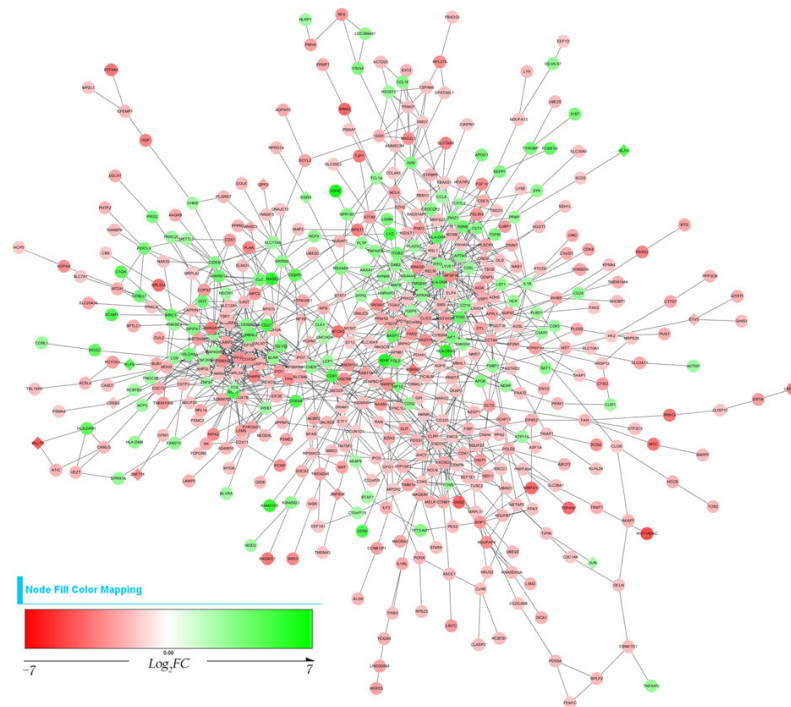
Figure 3. Co-expression network constructed for MM. The co-expression network shows the co-expressed genes in MM, and includes 599 nodes and 1088 edges. The green nodes represent the upregulated genes and the red nodes represent the downregulated genes. The color intensity corresponds to the absolute values of the log2-transformed fold-change of the differentially regulated genes (up- or downregulated). The diamonds represent the transcription factors. The black edges indicate the correlation of gene pairs that were identified as differentially co-expressed (DC) genes under a hard thresholding mechanism (FDR at 5%). When the disease progressed to MM, more transcription factors were involved and most of the genes were downregulated.

	Edges	Nodes	TFs
			Table 2. Characteristics of the MGUS, SMM, and MM networks.
MGUS	124	84	*TCF7L2, SMAD1, MAFB, MAF, KLF4, HIF1A*
SMM	41	39	*MAFB, HIF1A*
MM	1088	559	*ZNF91, ZNF148, WHSC1, TCF7L2, STAT1, SMAD1, SATB1, NR4A2, NFYB, MAFB, KLF6, KLF4, JUN, HIF1A, FOSB, FOS, EZH2, ETV5, ESRRG, EGR3, DPF2, CREM, CREB3, ZNF711, ZNF693, MLLT3, FOXN3*

MGUS = monoclonal gammopathy of unknown significance; SMM = smoldering multiple myeloma; MM = multiple myeloma; TF = transcription factors.

Functional annotation of DE genes

To functionally annotate the differentially co-expressed genes in each network, the online biological classification tool DAVID was used and the significant enrichment of the genes observed in multiple GO categories is shown in Table 3. The most significant enrichments in MGUS were in the GO categories of protein complex binding (P value = 4.91E-004) and MHC class II receptor (P value = 6.86 E-004). SMM had more GO categories and included all the categories in MGUS. The most significant enrichment in SMM was the GO category MHC class II receptor (P value = 1.87E-004). The other significant GO categories included protein complex binding (P value = 2.59E-004), cytokine binding (P value = 3.06E-004), kinase binding (P value = 5.11E-004), and protein dimerization activity (P value = 8.56E-004). The GO categories in MM were very different from those in MGUS and MM. The most significant enrichment in MM was the GO category RNA binding, with P value = 1.84E-005. The other significant GO categories included enzyme binding (P value = 2.86E-004) and structure-specific DNA binding (P value = 9.63E-004). There were also many other GO categories correlated with DNA structure, protein kinase, and translation factor activity.

Pathway enrichment analysis of DE genes

To further investigate the function of the DE genes in MGUS, SMM and MM, we mapped all the up- and downregulated genes in each network to the KEGG database and identified all the pathways in which they occurred (Table 4). There were six significant pathways in MGUS, 11 in SMM, and four in MM, separately. The most significant pathway in MGUS was cell adhesion molecules (CAMs) with a P value = 2.77E-005. The pathway of asthma was also significant with a P value = 9.09E-004. The most significant pathway in SMM and MM was asthma with P values = 2.30E-007 and 3.95E-004, respectively. The other significant pathways in SMM were CAMs with a P value = 1.21E-005 and systemic lupus erythematosus with a P value = 5.19E-004. MM also had other pathways that differed from those identified in MGUS and SMM, including DNA replication. The significant pathways shared between MGUS, SMM, and MM included CAMs and asthma.

Construction of the pathway-network for MGUS, SMM, and MM

A network provides a means of intuitively visualizing and characterizing complex

systems, and has proven to be particularly valuable in modeling disease systems. According to the KEGG analysis of MGUS, SMM, and MM, the pathways that received a count of 3 or less (i.e., three or fewer genes involved in each) were discarded, and a network based on the resulting pathway enrichment analysis was constructed. The network and the relationships between pathways are shown in Figure 4. The number of common pathways between MGUS and SMM, SMM and MM, and MGUS and MM was 18, 10, and 10, respectively. The pathways of cytokine-cytokine receptor interaction, cancer, Wnt signaling, hematopoietic cell lineage, lysosome, and CAMs were the common pathways among MGUS, SMM, and MM. The pathways of viral myocarditis, allograft rejection, asthma, type I diabetes mellitus, antigen processing and presentation, intestinal immune network for IgA production, leukocyte transendothelial migration, graft-versus-host disease, and autoimmune thyroid disease were involved in MGUS and SMM, but not in MM. The prostate cancer pathway was only seen to be involved in MGUS and MM, but not in SMM. The pathway-network might provide useful information for disease diagnosis.

Table 3. Classification and comparisons of DEGs among MGUS, SMM, and MM according to GO terms with P values < 0.01.

	Term and description	Count	P value
MGUS	GO:0032403~Protein complex binding	9	4.91E-004
	GO:0032395~MHC class II receptor	4	6.86E-004
	GO:0005539~Glycosaminoglycan binding	7	1.97E-003
	GO:0005178~Integrin binding	5	2.20E-003
	GO:0001871~Pattern binding	7	3.17E-003
	GO:0030247~Polysaccharide binding	7	3.17E-003
	GO:0030246~Carbohydrate binding	10	5.84E-003
	GO:0008009~Chemokine activity	4	8.96E-003
SMM	GO:0032395~MHC class II receptor	5	1.87E-004
	GO:0032403~Protein complex binding	12	2.59E-004
	GO:0019955~Cytokine binding	9	3.06E-004
	GO:0019900~Kinase binding	11	5.11E-004
	GO:0046983~Protein dimerization activity	20	8.56E-004
	GO:0005539~Glycosaminoglycan binding	9	1.60E-003
	GO:0004522~Pancreatic ribonuclease	4	1.81E-003
	GO:0019901~Protein kinase binding	9	2.17E-003
	GO:0001871~Pattern binding	9	2.90E-003
	GO:0030247~Polysaccharide binding	9	2.90E-003
	GO:0016892~Endoribonuclease activity, producing 3'-phosphomonoesters	4	3.03E-003
	GO:0019899~Enzyme binding	18	3.51E-003
	GO:0030246~Carbohydrate binding	14	3.86E-003
	GO:0046982~Protein heterodimerization activity	10	5.44E-003
	GO:0016894~Endonuclease activity, active with either ribo- or deoxyribonucleic acids and producing 3'-phosphomonoesters	4	6.72E-003
MM	GO:0003723~RNA binding	63	1.84E-005
	GO:0019899~Enzyme binding	46	2.86E-004
	GO:0043566~Structure-specific DNA binding	18	9.63E-004
	GO:0019901~Protein kinase binding	18	1.12E-003
	GO:0019900~Kinase binding	19	3.90E-003
	GO:0003697~Single-stranded DNA binding	9	5.84E-003
	GO:0004721~Phosphoprotein phosphatase activity	17	8.85E-003
	GO:0008135~Translation factor activity, nucleic acid binding	12	9.96E-003

DEG = differentially expressed genes; MGUS = monoclonal gammopathy of unknown significance; SMM = smoldering multiple myeloma; MM = multiple myeloma; GO = gene ontogeny.

Table 4. Pathway enrichment analysis of MGUS, SMM, and MM with P values < 0.01.

	Term and description	Count	P value	Genes
MGUS	hsa04514: Cell adhesion molecules (CAMs)	11	2.77E-005	VCAM1, NCAM1, PTPRC, ITGA6, NLGN4X, VCAN, ITGB2, HLA-DPA1, HLA-DPB1, HLA-DMA, HLA-DRA
	hsa05310: Asthma	5	9.09E-004	FCER1G, HLA-DPA1, HLA-DPB1, HLA-DMA, HLA-DRA
	hsa04640: Hematopoietic cell lineage	7	1.90E-003	CD9, CD19, ITGA6, CD44, KIT, CD14, HLA-DRA
	hsa05416: Viral myocarditis	6	4.41E-003	CCND1, ITGB2, HLA-DPA1, HLA-DPB1, HLA-DMA, HLA-DRA
	hsa04672: Intestinal immune network for IgA production	5	6.44E-003	HLA-DPA1, HLA-DPB1, HLA-DMA, CXCL12, HLA-DRA
	hsa04612: Antigen processing and presentation	6	8.53E-003	LGMN, HLA-DPA1, HLA-DPB1, CTSB, HLA-DMA, HLA-DRA
SMM	hsa05310: Asthma	9	2.30E-007	HLA-DQB1, RNASE3, EPX, PRG2, FCER1G, HLA-DPA1, HLA-DPB1, HLA-DMA, HLA-DRA
	hsa04514: Cell adhesion molecules (CAMs)	14	1.21E-005	HLA-DQB1, PTPRC, CADM1, CD99, ITGB2, HLA-DMA, ITGB1, VCAM1, NCAM1, ITGA6, VCAN, HLA-DPA1, HLA-DPB1, HLA-DRA
	hsa05322: Systemic lupus erythematosus	10	5.19E-004	C1QA, HLA-DQB1, C1QB, HIST1H2AC, FCGR2C, ELANE, HLA-DPA1, HLA-DPB1, HLA-DMA, HLA-DRA
	hsa04612: Antigen processing and presentation	8	3.38E-003	HLA-DQB1, LGMN, TAP1, HLA-DPA1, HLA-DPB1, CTSB, HLA-DMA, HLA-DRA
	hsa04640: Hematopoietic cell lineage	8	4.12E-003	CD9, ITGA6, CD44, IL6R, KIT, CD14, HLA-DR4, CSF1R
	hsa05200: Pathways in cancer	17	4.44E-003	WNT5A, IL8, MMP9, TGFBR2, KIT, NFKB2, BIRC3, ITGB1, FZD6, CCND1, CDKN1A, HIF1A, CDKN1B, ITGA6, TCEB2, LAMC1, CSF1R
	hsa04672: Intestinal immune network for IgA production	6	5.80E-003	HLA-DQB1, HLA-DPA1, HLA-DPB1, HLA-DMA, CXCL12, HLA-DR4
	hsa05416: Viral myocarditis	7	6.51E-003	HLA-DQB1, CCND1, ITGB2, HLA-DPA1, HLA-DPB1, HLA-DMA, HLA-DRA
	hsa04670: Leukocyte transendothelial migration	9	6.72E-003	VCAM1, GNA13, MMP9, NCF4, RAP1A, CD99, ITGB2, ITGB1, CXCL12
	hsa05020: Prion diseases	5	9.03E-003	C1QA, EGR1, NCAM1, C1QB, LAMC1
	hsa05330: Allograft rejection	5	9.98E-003	HLA-DQB1, HLA-DPA1, HLA-DPB1, HLA-DMA, HLA-DRA
MM	hsa05310: Asthma	9	3.95E-004	RNASE3, EPX, PRG2, FCER1G, HLA-DPA1, HLA-DPB1, HLA-DMA, HLA-DRA
	hsa05322: Systemic lupus erythematosus	16	1.76E-003	HIST1H2AC, HIST1H2BH, SNRPD1, HLA-DMB, HLA-DMA, C1QA, C1QB, TROVE2, H2AFV, FCGR2C, HIST2H2BE, HLA-DPA1, HLA-DPB1, HIST1H3H, HLA-DRA, CD28
	hsa03030: DNA replication	9	1.86E-003	RPA1, PRIM1, MCM7, POLE3, POLD2, MCM3, RNASEH2A, MCM6, RP43
	hsa04514: Cell adhesion molecules (CAMs)	18	5.23E-003	PTPRC, MPZL1, OCLN, CADM1, CD99, ITGB2, HLA-DMB, HLA-DMA, VCAM1, NCAM1, ITGAV, NLGN4X, CNTN1, VCAN, HLA-DPA1, HLA-DPB1, HLA-DRA, CD28

MGUS = monoclonal gammopathy of unknown significance; SMM = smoldering multiple myeloma; MM = multiple myeloma.
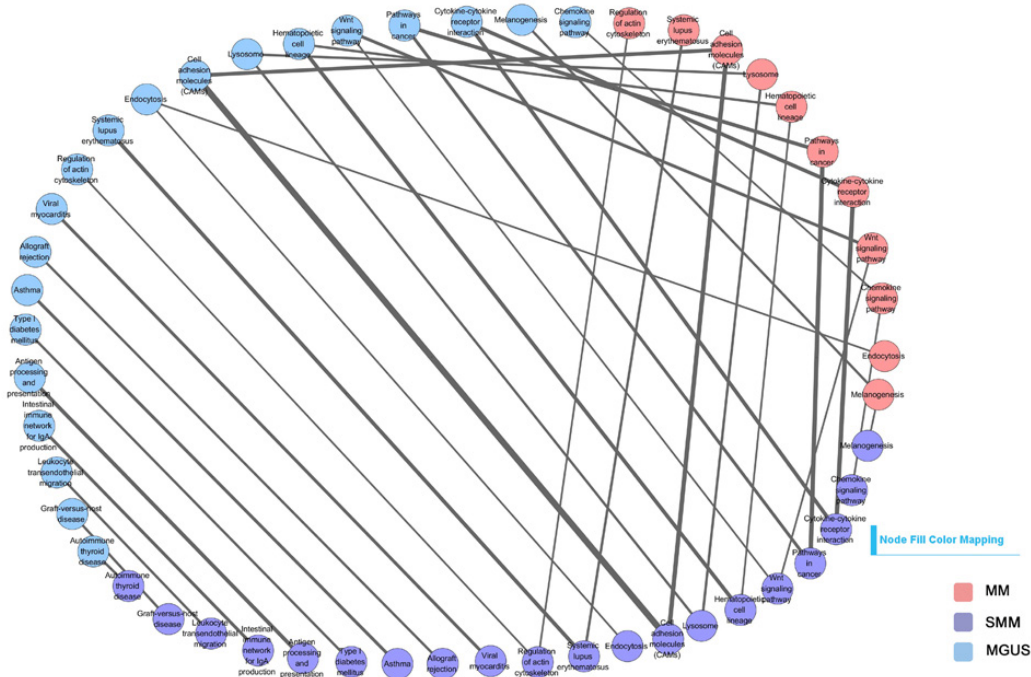
Figure 4. Pathway-network constructed based on KEGG analysis. In the pathway-network, the blue nodes represent the pathways in MGUS, the purple nodes represent the pathways in SMM, and the red nodes represent the pathways in MM. The edges that connect the same pathway in different diseases represent that the pathway involved has more than three common genes. The thickness of the edges indicates the number of common genes included in the diseases.

DISCUSSION

MM is an incurable plasma cell malignancy and the progression of MGUS or SMM to MM is often associated with expanding bone marrow and increasingly severe organ impairment or other symptoms. In this study, the gene expression patterns and pathways involved in each disease were analyzed, and then a pathway-network was constructed to better understand the relationships among them.

The prevalence of MGUS, a premalignant plasma-cell disorder, is markedly age dependent and significantly higher in males (Cohen et al., 1998). Though the genes related to MGUS might cluster within families, the roles of genomic background and gene environment are poorly defined (Lynch et al., 2001). Lopez-Corral et al. (2012) found that the incidence of copy number abnormalities and the genomic complexity progressively increased from MGUS to MM. However, the current diagnostic criteria for MGUS consist only of the concentration of monoclonal protein in serum, bone marrow plasmacytosis, and whether the patient has organ damage. In our study, many genes related to the immune system (e.g., *HLA-DMA*, *HLA-DRA*, and *HLA-DPB1*) were found to be dysregulated; specifically, CCND1, which is a G1-specific cyclin that has been linked to lymphoid, parathyroid, and breast tumors (Rosenwald et al., 1993), was significantly downregulated in MGUS. The disordered expression of these

genes in MGUS might lead to premalignant plasma-cell proliferative disorder and thus drive the progression to SMM or MM.

According to the pathway enrichment analysis, the process of transformation from MGUS to SMM might not represent a linear acquisition of changes, but rather a branched heterogeneous process. Many pathways were found to have participated in the development of the three diseases simultaneously. Due to the considerable molecular heterogeneity, reliable markers to predict the prognosis for individual patients with SMM have not been identified. For example, Gkotzamanjdou et al. (2013) studied the role of angiogenesis in the pathogenesis of SMM and their results revealed that the ratios of circulating angiopoietin-1 to angiopoietin-2 in SMM were not different from those in MGUS and normal samples. In contrast, our functional enrichment analysis revealed that many biological processes occurred in SMM that were the same as those in MGUs, such as protein complex binding, MHC class II receptors, integrin binding, and cytokine binding. The genes that belong to the ribonuclease RNase A family (*RNASE2* and *RNASE3*), fatty acid binding protein gene (*FABP5*), and membrane-spanning genes (*MS4A4A*) were all upregulated. This was consistent with a change in enzymes that are correlated with DNA or RNA.

Despite advances in the understanding of MM pathogenesis, the molecular pathways underlying the development of MM are still unknown. Our functional and pathway enrichment analyses showed that translation factor activity, nucleic acid binding, phosphoprotein phosphatase activity, and DNA replication were all active; of these, RNA binding was the most significant process with a P value = 1.84E-005 and a count value of 63. This was consistent with a previous study wherein the progression of MGUS to SMM and finally to MM were found to be accompanied by a progressive, significant increase in the looseness of local chromatin structure, gene expression levels, and DNA repair efficiency (Gkotzamanidou et al., 2014). Yoshizawa et al. (2012) found that the expression level of miR-92a (which is stably expressed in plasma) in symptomatic patients with MM was significantly downregulated compared with that in normal subjects, but that there was no obvious difference in patients with SMM compared with normal subjects. As secreted proteins with functions in growth, differentiation, and activation, cytokines and chemokines play an important role in regulating the nature of immune responses. The contents of cytokines and chemokines in the peripheral blood and bone marrow supernatant of patients with MM were greatly increased compared to those in patients with SMM (Zingone et al., 2014). This finding might shed light on the decision to initiate chemotherapy and on disease status monitoring.

Currently, no specific biological markers have been identified that could distinguish the developmental stage of the disease. The progression of MGUS to SMM and MM is very complex. Our pathway-network proved that there were some exclusive pathways in each developmental stage, which might be considered as indicators for disease diagnosis and for distinguishing the developmental stage, the identification of which represents the first step for disease therapy. Given that some patients with MGUS will never progress to MM in their lifetime, and that there is currently no therapy to prevent the progression of MGUS to MM, thus continuous monitoring is important for patients with MGUS. For active SMM, such as for patients with extreme bone marrow plasmacytosis or extremely abnormal serum immunoglobulin free light chain ratio, the acceptance of MM-appropriate therapy should be considered.

Conflicts of interest

ACKNOWLEDGMENTS

REFERENCES

Aung HH, Tsoukalas A, Rutledge JC and Tagkopoulos I (2014). A systems biology analysis of brain microvascular endothelial cell lipotoxicity. *BMC Syst. Biol.* 8: 80.

Birmann BM, Tamimi RM, Giovannucci E, Rosner B, et al. (2009). Insulin-like growth factor-1 and interleukin-6-related gene variation and risk of multiple myeloma. *Cancer Epidemiol. Biomarkers Prev.* 18: 282-288.

Bladé J and Rosinol L (2006). Smoldering multiple myeloma and monoclonal gammopathy of undetermined significance. *Curr. Treat. Options Oncol.* 7: 237-245.

Breitling B, Armengaud P, Amtmann A and Herzyk P (2004). Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. *FEBS Lett.* 573: 83-92.

Cohen HJ, Crawford J, Rao MK, Pieper CF, et al. (1998). Racial differences in the prevalence of monoclonal gammopathy in a community-based sample of the elderly. *Am. J. Med.* 104: 439-444.

Dawson JA and Kendziorski C (2012). An empirical Bayesian approach for identifying differential coexpression in high-throughput experiments. *Biometrics* 68: 455-465.

del Sol A, Balling R, Hood L and Galas D (2010). Disease as network perturbations. *Curr. Opin. Biotechnol.* 21: 566-571.

Fonseca R, Bailey RJ, Ahmann GJ, Rajkumar SV, et al. (2002). Genomic abnormalities in monoclonal gammopathy of undetermined significance. *Blood* 100: 1417-1424.

Gentleman R, Carey V, Huber W and Hahne F (2014). Genefilter: methods for filtering genes from microarray experiments. R package version 1.50.0.

Gkotzamanidou M, Christoulas D, Souliotis VL, Papatheodorou A, et al. (2013). Angiogenic cytokines profile in smoldering multiple myeloma: no difference compared to MGUS but altered compared to symptomatic myeloma. *Med. Sci. Monit.* 19: 1188-1194.

Gkotzamanidou M, Terpos E, Bamia C, Kyrtopoulos SA, et al. (2014). Progressive changes in chromatin structure and DNA damage response signals in bone marrow and peripheral blood during myelomagenesis. *Leukemia* 28: 1113-1121.

Greenberg AJ, Rajkumar SV, Larson DR, Dispenzieri A, et al. (2012). Increased prevalence of light chain monoclonal gammopathy of undetermined significance (LC-MGUS) in first-degree relatives of individuals with multiple myeloma. *Br. J. Haematol.* 157: 472-475.

Hillengass J, Moehler T and Hundemer M (2011). Monoclonal gammopathy and smoldering multiple myeloma: diagnosis, staging, prognosis, management. *Recent Results Cancer Res.* 183: 113-131.

Hosack DA, Dennis G, Sherman BT, Lane HC, et al. (2003). Identifying biological themes within lists of genes with EASE. *Genome Biol.* 4: R70.

Huang da W, Sherman BT and Lempicki RA (2009). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 37: 1-13.

Ideker T and Sharan R (2008). Protein networks in disease. *Genome Res.* 18: 644-652.

Jiang P, Yueguo W, Huiming H, Hongxiang Y, et al. (2009). B-Lymphocyte stimulator: a new biomarker for multiple myeloma. *Eur. J. Haematol.* 82: 267-276.

Ju S, Wang Y, Ni H, Wang X, et al. (2009). Correlation of expression levels of BLyS and its receptors with multiple myeloma. *Clin. Biochem.* 42: 387-399.

Korde N, Kristinsson SY and Landgren O (2011). Monoclonal gammopathy of undetermined significance (MGUS) and smoldering multiple myeloma (SMM): novel biological insights and development of early treatment strategies. *Blood* 117: 5573-5581.

Koziol JA (2010). The rank product method with two samples. *FEBS Lett.* 584: 4481-4484.

Kyle RA and Rajkumar SV (2009). Treatment of multiple myeloma: a comprehensive review. *Clin. Lymphoma Myeloma* 9: 278-288.

Kyle RA, Remstein ED, Therneau TM, Dispenzieri A, et al. (2007). Clinical course and prognosis of smoldering (asymptomatic) multiple myeloma. *N. Engl. J. Med.* 356: 2582-2590.

Lee HK, Hsu AK, Sajdak J, Qin J, et al. (2004). Coexpression analysis of human genes across many microarray data sets. *Genome Res.* 14: 1085-1094.

López-Corral L, Sarasquete ME, Beà S, García-Sanz R, et al. (2012). SNP-based mapping arrays reveal high genomic complexity in monoclonal gammopathies, from MGUS to myeloma status. *Leukemia* 26: 2521-2529.

Lynch HT, Sanger WG, Pirruccello S, Quinn-Laquer B, et al. (2001). Familial multiple myeloma: a family study and review of the literature. *J. Natl. Cancer Inst.* 93:1479-1483.

Ma L, Robinson LN and Towle HC (2006). ChREBP. Mlx is the principal mediator of glucose-induced gene expression in the liver. *J. Biol. Chem.* 281: 28721-28730.

Malpas JS, Bergsagel DE, Kyle R and Anderson K (2004). Multiple myeloma: biology and management. Oxford University Press, Oxford.

Miller JA, Oldham MC and Geschwind DH (2008). A systems level analysis of transcriptional changes in Alzheimer's disease brain. *J. Neurosci.* 28: 1410-1420.

Ray M and Zhang W (2010). Analysis of Alzheimer's disease severity across brain regions by topological analysis of gene co-expression networks. *BMC Syst. Biol.* 4: 136.

Rifai N and Ridker PM (2001). Proposed cardiovascular risk assessment algorithm using high-sensitivity C-reactive protein and lipid screening. *Clin. Chem.* 47: 28-30.

Rosenwald IB, Lazaris-Karatzas A, Sonenberg N and Schmidt EV (1993). Elevated levels of cyclin D1 protein in response to increased expression of eukaryotic initiation factor 4E. *Mol. Cell Biol.* 13: 7358-7363.

Schilling CH, Schuster S, Palsson BO and Heinrich R (1999). Metabolic pathway analysis: basic concepts and scientific applications in the post-genomic era. *Biotechnol. Prog.* 15: 296-303.

Shen X, Zhu W, Zhang X, Xu G, et al. (2011). A role of both NF-κB pathways in expression and transcription regulation of *BAFF-R* gene in multiple myeloma cells. *Mol. Cell Biochem.* 357: 21-30.

Tai YT, Li XF, Breitkreutz I, Song W, et al. (2006). Role of B-cell-activating factor in adhesion and growth of human multiple myeloma cells in the bone marrow microenvironment. *Cancer Res.* 66: 6675-6682.

Xing SM, Wang J, He X, Lai J, et al. (2014). Identification of disease related miRNAs based on co-expression network in spinal cord injury. *Int. J. Neurosci.* [Epub Ahead of Print] Doi: 10.3109/00207454.2014.930741.

Yoshizawa S, Ohyashiki JH, Ohyashiki M, Umezu T, et al. (2012). Downregulated plasma miR-92a levels have clinical impact on multiple myeloma and related disorders. *Blood Cancer J.* 2: e53.

Zhang L, Miles MF and Aldape KD (2003). A model of molecular interactions on short oligonucleotide microarrays. *Nat. Biotechnol.* 21: 818-821.

Zingone A, Wang W, Corrigan-Cummins M, Wu SP, et al. (2014). Altered cytokine and chemokine profiles in multiple myeloma and its precursor disease. *Cytokine* 69: 294-297.