

## Original Article

# Gene expression profiling based on microarray among monoclonal gammopathy of undetermined significance, smoldering multiple myeloma and multiple myeloma

Baotong Feng<sup>1</sup>, Butong Ding<sup>2</sup>, Yao Sun<sup>3</sup>, Nongjian Guo<sup>4</sup>

Departments of <sup>1</sup>Hematology, <sup>2</sup>Radiotherapy, Weifang Yidu Central Hospital of Qingzhou, Qingzhou 262500, Shandong, P.R. China; <sup>3</sup>Department of Pathology, Ji'nan Central Hospital, Ji'nan 250013, Shandong, P.R. China; <sup>4</sup>Department of Hematology, Central Hospital of Ji'nan, No.105, Jiefang Road, Lixia District, Ji'nan 250013, Shandong, P.R. China

Received July 4, 2016; Accepted October 29, 2016; Epub January 15, 2017; Published January 30, 2017

**Abstract:** Multiple myeloma (MM) and smoldering multiple myeloma (SMM) are different types of plasma cell malignancy which are believed to be developed from Monoclonal Gammopathy of Undetermined Significance (MGUS). In this study, to further identify the molecular pathogenesis of these neoplasms of plasma cells, several mRNA expression libraries of MGUS, SM and MM were pooled from GEO & Array Express database. The mRNA expression profiles among these diseases were analyzed for the total differentially expressed genes (DEGs) and most significantly changed DEGs. Moreover, the protein-protein interaction networks (PPIs) for selected DEGs were constructed as well. Based on our result, although there were DEGs shared by MGUS, SM and MM, it is notable that the Top DEGs or the total number of DEGs are different among them, which suggesting a different gene expression pattern for these plasma neoplasms. Our data suggested there were different molecular mechanism or pathogenesis for each disease and may imply novel therapeutic targets.

**Keywords:** Monoclonal gammopathy of undetermined significance, multiple myeloma, smoldering multiple myeloma, differentially expressed genes, microarray

## Introduction

Multiple myeloma (MM) is one type of plasma cell malignancy which is characterized by so-called CRAB criteria (hypercalcemia, renal failure, anemia and bony lesions) [1-3]. Currently, MM has been considered to be developed from the clinically asymptomatic monoclonal gammopathy (AMG), which is a premalignant disorder of plasma-cell proliferation [4-6]. AMG or the earlier phases of plasma cell dyscrasias contains Monoclonal gammopathy of undetermined significance (MGUS) and smoldering multiple myeloma (SMM), which are characterized by no myeloma-related organ or tissue injury [4, 6]. It appears that almost all cases of MM are preceded from a premalignant MGUS stage and contains a phase of smoldering multiple myeloma (SMM) without end organ damage [3, 5]. Therefore, presenting with MGUS suggests a life-long risk for the patients to

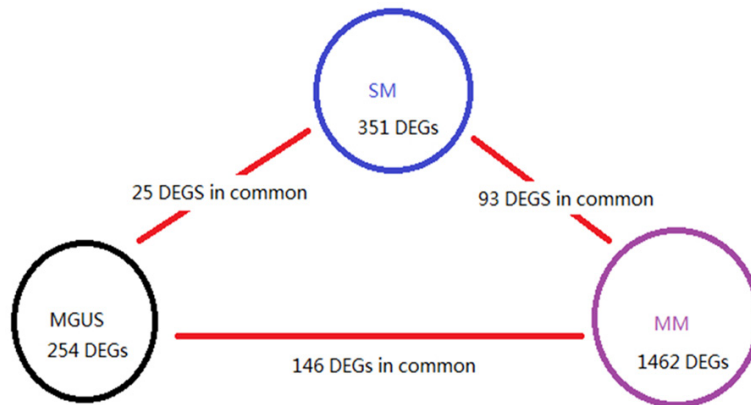
develop MM, or other plasma cell related disorders such as Waldenström's macroglobulinemia, primary amyloidosis, B-cell lymphoma, or chronic lymphocytic leukemia as well [5]. Based on previous reports, one year after the onset of disease, MGUS may develop into MM or similar lymphoproliferative disorder at a rate about 1-2% and the chance to develop MM increased as the extension of the MGUS duration [7]. Moreover, progression of MM is also age-related since elder MGUS patients demonstrated a higher prevalence of MM [8, 9]. Up to date, retrospective analysis based models had been proposed and used for prediction of progression from MGUS to MM [10].

On the other hand, although MGUS is a major risk factor for patients, majority of MGUS patients still never develop MM in their whole life [5]. According to the criteria of International Myeloma Working Group (IMWS), SMM distin-

## Gene expression profiling among MGUS, SMM and MM

**Table 1.** Data source information

Source	Disease type	Cases	System
GSE6477	SM	9	U133a
E-MTAB-363	MM	9	U133 plus2
E-MTAB-363	MGUS	5	U133 plus2
E-MTAB-363 & GSE6477	Normal	13	U133a & U133 plus2



**Figure 1.** DEGs identified from MGUS, SMM and MM.

guished from MGUS by the presentation of >10% marrow plasmacytosis and/or  $\geq 3$  g/dL serum M-protein but lacking the existence of CRAB [3]. Compared with MGUS, 10% of the SMM patients will develop MM in the first 5 years after diagnosis, which indicated a higher risk than MGUS [11]. However, the molecular mechanism behind disease progression from MGUS to MM is still illusive.

In this study, several mRNA expression libraries of MGUS, SM and MM as well as healthy controls were pooled from GEO & Array Express database (access number: GSE5900, GSE6477, E-MTAB-363). The mRNA expression profiles among these disease were analyzed by Affymetrix Human Genome U133a and Array Affymetrix Human Genome U133 Plus 2.0 Array for identification of different expressed genes (DEGs). The most significantly DEGs were analyzed and the PPI networks were drawn for them.

### Materials and methods

#### Data collection

The mRNA expression profiles for Monoclonal Gammopathy of Undetermined Significance (MGUS), smoldering multiple myeloma (SMM)

and multiple myeloma (MM) were downloaded from GEO & Array Express database (Access No.: GSE5900, GSE6477, E-MTAB-363). Moreover, the microarray data from healthy cases were included as control (Access No. E-MTAB-363 & GSE6477). The mRNA expression level among these packages was analyzed by Affymetrix Human Genome U133a and Array Affymetrix Human Genome U133 Plus 2.0 Array. The detail of these data packs, such as data source, disease types and case numbers were listed as **Table 1**.

#### Data analysis

The raw Affymetrix data (.CEL files) recording the original signal intensity for each probe was obtained from microarray and normalized by Robust Multi-array Average (RMA) approach to reduce the mismatch spots in the R-Software (downloaded from <https://www.r-project.org/>). For different probes which are targeting the same gene, the probe with maximum value was selected as the final expression folds to generate the microarray data.

#### Identification of differentially expressed genes (DEGs)

To identify the differentially expressed genes among the normal controls, MGUS, SMM and MM patients, the Limmaprogram from R-Software was used for data analysis. Briefly, after setting the false discovery rate (FDR) to 0.05, the value of expression folds for all samples was compared via Bayesian linear regression to obtain the corresponding *P*-value. The genes expressed in MGUS, SMM and MM with  $P < 0.05$  when comparing with normal control was considered as differentially expressed genes (DEGs).

#### The construction of protein-protein interaction (PPI) network

The DEGs were first inputted into STRING data based for Protein-Protein Interaction network

## Gene expression profiling among MGUS, SMM and MM

**Table 2.** Top15 DEGs in MM

Gen Symbol	Gene Name	logFC (evaluation marked as "-") Log (Ctr/MM)	T-value	P-value
IGLC1	Immunoglobulin lambda constant 1 (Mcg marker)	11.02798	25.82008	9.38E-14
CD19	CD19 molecule	6.466287	15.27624	1.73E-10
LOC96610	BMS1 homolog, ribosome assembly protein (yeast) pseudogene	9.521166	14.36084	4.10E-10
HLA-DRB6	Major histocompatibility complex, class II, DR beta 6 (pseudogene)	5.29852	13.80206	7.10E-10
IGLV1-44	Immunoglobulin lambda variable 1-44	9.2309	13.42191	1.04E-09
BEST1	Bestrophin 1	4.193983	13.11152	1.44E-09
SLC39A8	Solute carrier family 39 (zinc transporter), member 8	-5.75563	-13.1078	1.45E-09
XAGE-4	XAGE-4 protein	5.722528	12.91144	1.78E-09
MCAM	Melanoma cell adhesion molecule	4.852414	12.91111	1.78E-09
C5orf30	Chromosome 5 open reading frame 30	-6.8773	-12.7924	2.02E-09
HLA-DPB1	Major histocompatibility complex, class II, DP beta 1	6.958092	12.4308	2.99E-09
PLAUR	Plasminogen activator, urokinase receptor	5.430296	12.28823	3.50E-09
HIST2H2BE	Histone cluster 2, H2be	-6.83745	-12.2655	3.59E-09
MYO1D	Myosin ID	5.887433	12.25336	3.64E-09
SERBP1	SERPINE1 mRNA binding protein 1	-6.56065	-11.742	6.48E-09

**Table 3.** Top15 DEGs of MGUS

Gene Symbol	Gene Name	logFC (evaluation marked as "-") Log (Ctr/MGUS)	T-value	P-value
NDNF	Neuron-derived neurotrophic factor	-8.784605773	-24.9818	6.98E-11
IGLC1	Immunoglobulin lambda constant 1 (Mcg marker)	9.1781025	17.30071	3.34E-09
CD19	CD19 molecule	6.112124909	16.11877	6.96E-09
CCND1	Cyclin D1	-9.582893043	-14.4741	2.12E-08
FAM49A	Family with sequence similarity 49, member A	-5.831240288	-13.159	5.62E-08
KCNMB2	Potassium large conductance calcium-activated channel, subfamily M, beta member 2	-8.598320019	-12.876	7.01E-08
LSAMP	Limbic system-associated membrane protein	-7.083451807	-12.6058	8.69E-08
CXCL12	Chemokine (C-X-C motif) ligand 12	8.521326135	12.42521	1.01E-07
C12orf75	Chromosome 12 open reading frame 75	6.752004091	12.13187	1.28E-07
SLC39A8	Solute carrier family 39 (zinc transporter), member 8	-4.203070561	-12.0466	1.38E-07
ACN9	ACN9 homolog (S. cerevisiae)	5.047266434	11.29034	2.64E-07
HSH2D	Hematopoietic SH2 domain containing	5.999476449	10.66738	4.64E-07
KIT	V-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog	-7.081336132	-10.4588	5.64E-07
ACOXL	AcyI-CoA oxidase-like	-5.497865819	-10.4191	5.86E-07
OGT	O-linked N-acetylglucosamine (GlcNAc) transferase	7.502858743	10.3726	6.12E-07

(PPI) analysis, then PPI network were drawn by Cytoscape (Version 3.20).

### Results

#### Identification of differentially expressed genes (DEGs)

We first set the data from normal healthy people as control group (Access No. E-MTAB-363 & GSE6477, 13 cases) to evaluate the baseline of genes expression. Then data from SM patients groups (Access No. GSE6477, 9 cases), MM patients (Access No. E-MTAB-363, 9 cases) and MGUS patients (Access No. E-MTAB-363, 5

cases) was pooled and compared to the control groups. By comparing the microarray data from these cases, there were totally 254, 351 and 1462 DEGs identified for MGUS, SMM and MM, respectively (**Figure 1**). The top DEGs (the most differentially expressed genes) for MGUS, SMM and MM were listed as **Tables 2-4**, respectively. Moreover, there were 25 DEGs shared by MGUS and SMM (MGUS/SMM), while 146 and 93 DEGs were shared by MGUS/MM and SM/MM, respectively (**Figure 1**).

Moreover, total of 21 DEGs were shared by all three types of diseases. The official gene names for them are UNC13B, CD19, ST3GAL6,

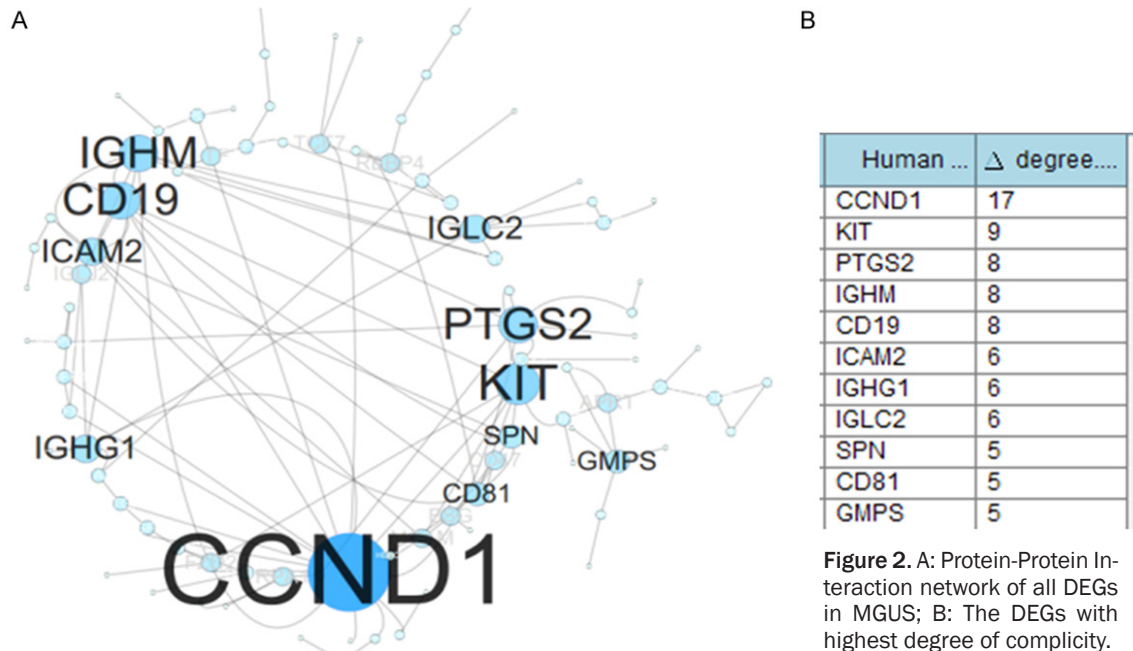
## Gene expression profiling among MGUS, SMM and MM

**Table 4.** Top15 DEGs of SMM

Gene Symbol	Gene Name	logFC (evaluation marked as "-") Log (Ctr/SM)	T-value	P-value
ANKRD12	Ankyrin repeat domain 12	-4.0798	-19.7017	4.91E-13
RNASE2	Ribonuclease, RNase A family, 2 (liver, eosinophil-derived neurotoxin)	-8.07298	-16.9039	5.71E-12
CLC	Charcot-Leyden crystal galectin	-8.0734	-16.8206	6.17E-12
TIAL1	TIA1 cytotoxic granule-associated RNA binding protein-like 1	-3.6075	-15.9966	1.37E-11
TJP1	Tight junction protein 1	5.436485	14.64327	5.49E-11
SMCHD1	Structural maintenance of chromosomes flexible hinge domain containing 1	-3.67077	-11.277	2.99E-09
PRG3	Proteoglycan 3	-6.12255	-10.275	1.18E-08
PRG2	Proteoglycan 2, bone marrow (natural killer cell activator, eosinophil granule major basic protein)	-7.4738	-10.154	1.40E-08
AKAP17A	A kinase (PRKA) anchor protein 17A	-3.00035	-9.37787	4.41E-08
COBLL1	Cordon-bleu WH2 repeat protein-like 1	-4.29677	-9.30799	4.91E-08
CLK1	CDC-like kinase 1	-4.28546	-9.21761	5.63E-08
EXOC1	Exocyst complex component 1	-3.98055	-8.91387	9.03E-08
ANKRD10	Ankyrin repeat domain 10	-2.71892	-8.75124	1.17E-07
RB1	Retinoblastoma 1	-3.47373	-8.75064	1.17E-07
METTL3	Methyltransferase like 3	-2.77435	-8.64896	1.37E-07

**Table 5.** Total number of disease specific DEGs and the top 10 Gene symbol in MGUS, MM and SMM

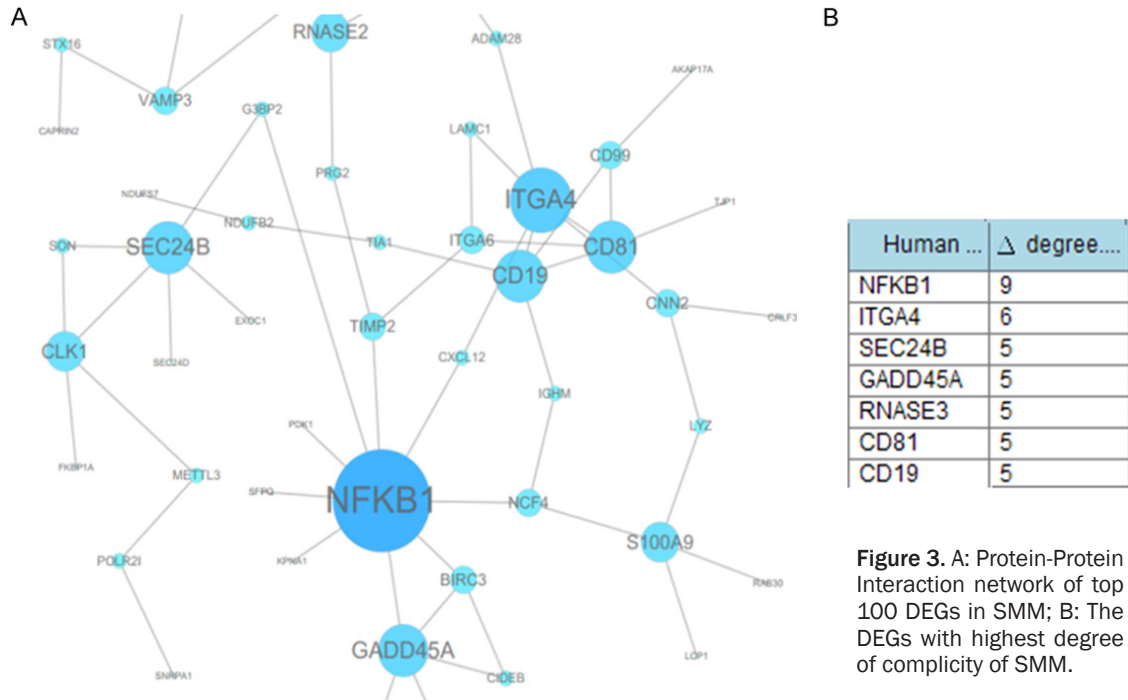
Disease	Numbers of disease specific DEGS	Top10 Gene Symbol
MGUS	104	KCNMB2 LSAMP C12orf75 MRPS15 C14orf142 RGS14 CEP97 AKT3 ZNRD1 DHRS9
MM	1245	C5orf30 PLAUR MYO1D SERBP1 ADNP2 PSAPL1 GALNT3 ZNF665 TRIM52 STAT3
SMM	5	ANKRD12 RNASE2 CLC TJP1 SMCHD1 PRG3 PRG2 AKAP17A COBLL1 CLK1



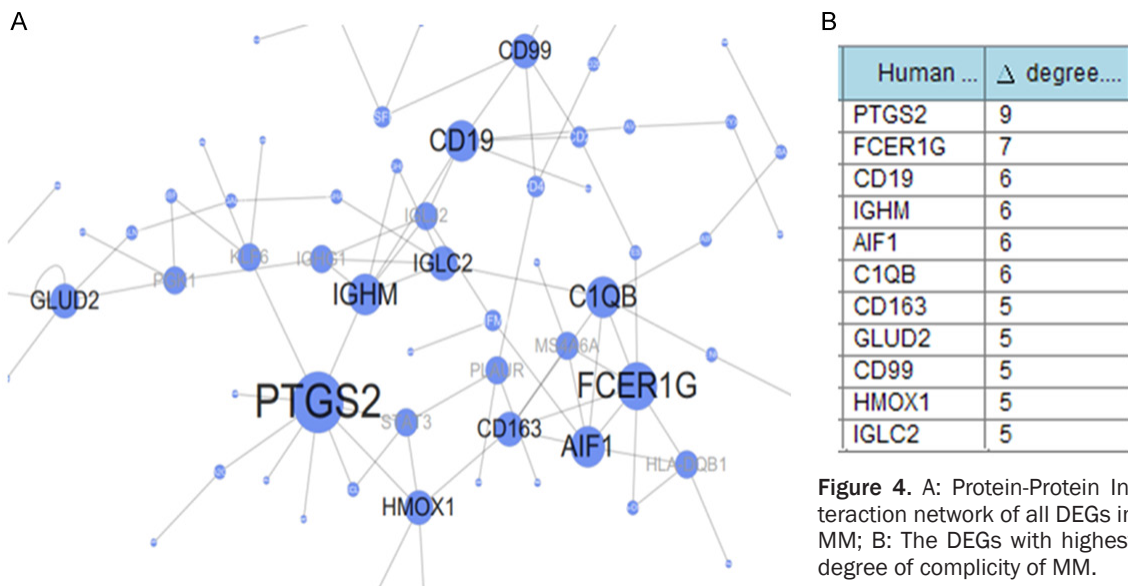
TMEM156, IGKC, CD81, IGHM, CXCL12, HLA-DMA, IGHG1, IGLC1, IGLJ3, IGHD, C1QB, NDNF,

C1QA, MLLT3, IGK, HMOX1, OGT and HLA-DPB1. Moreover, based on our analysis, there

## Gene expression profiling among MGUS, SMM and MM



**Figure 3.** A: Protein-Protein Interaction network of top 100 DEGs in SMM; B: The DEGs with highest degree of complicity of SMM.



**Figure 4.** A: Protein-Protein Interaction network of all DEGs in MM; B: The DEGs with highest degree of complicity of MM.

are DEGs presented in only one type of diseases as well, which suggested a disease specific gene expression pattern in MGUS, SMM and MM (Table 5).

### *The protein-protein interaction (PPI) network for DEGs*

Since our analysis identified the shared DEGs among MGUS, SMM and MM as well as disease

specific DEGs, it is interesting to understand the protein-protein interaction network for these DEGs which may underline the novel regulation mechanism for disease progression or identify the key regulator for plasma cell malignancy. By utilizing the String (Protein interaction data base), we first analyzed the protein interaction network for all DEGs in MGUS (Figure 2). The size of the key nodes represents the complicity of the protein interaction net-

work for indicated genes. As demonstrated in **Figure 2A**, the CCND1, KIT, PTGS2, IGHM, CD19, ICAM2 and IGHG1 had been proposed to hold the most complicated interaction network in MGUS.

On the other hand, the PPI network was analyzed and drawn for the top 100 and 200 DEGs of SMM and MM (**Figures 3A** and **4A**), respectively. Based on the degree of complicity, NFKB1, ITGA, SEC24B, GADD45A, RNASE3, CD81 and CD19 demonstrated the highest complicity for PPI among all DEGs for SMM (**Figure 3B**). Moreover, PTGS2, FCER1G, CD19, IGHM, AIF1, C1QB, CD163, GLUD2, CD99, HMOX1 and IGLC2 holds the highest complicity for PPI among all DEGs for MM (**Figure 4B**).

### Discussion

Gene-expression profiling (GEP) based on microarray analysis for cancer cell provide a novel methods to analyze the changes of gene expression for cancer cell in a genomic-wide manner [12]. Previous report suggested that GEP for MM also provide a novel approach for disease prognosis or guide the selection of therapeutic targets as well [13].

Multiple myeloma is a specially neoplasm form which is characterized by the abnormal proliferation of terminally differentiated B cells or plasma cells [14]. It is believed that oncogenes expression promoted by immunoglobulin enhancers as a result of chromosome translocation is a major genetic cause for MGUS and MM [15]. Since the discovery of chromosome 13 deletions and abnormalities of chromosome 1q21 were associated with poor prognosis of MM in patients [15, 16], there were evidences suggested that all these major cytogenetic abnormalities in MM could be observed in MGUS and increasing chromosomal instability (CIN) is indispensable for normal plasma cell transformation to MM [13, 14]. Moreover, despite variety genetic abnormalities were observed, gene expression patterns within tumor cells are remarkably stable and reproducible for different MM patients [12].

Although genes expression signature has been proposed for prognosis of MM [17], the function for different DEGs identified among MGUS, SMM and MM is still largely unknown. In our study, we systematically analyzed the GEDs

among three forms of plasma neoplasm. Based on our results, there were totally 21 genes were identified as the common DEGs which is shared by all three disease types. However, it is notable that the Top DEGs or total number of DEGs is different among them, suggesting a different genes expression pattern for these plasma neoplasms.

In our study, the PPI network for top DEGs of each disease had been constructed. In addition to the variable DEGs among three diseases, the PPI network and the DEGs located in the center of each network is also highly variable. In MGUS, CCND1 (cyclin D1) was identified as the gene shown highest complexity in PPI network, while NFKB1 and PTGS2 was identified as highest in SMM and MM, respectively. It appears that PPI networks are highly diversified among these diseases, which may implied different molecular mechanism of these DES and PPI played during the disease progression. Since NFKB1 and CCND1 were involved in cell proliferation, cell cycles and apoptosis, it is notable that PTGS2 which is coding for Prostaglandin-endoperoxide synthase 2 was demonstrated highest complexity in the PPI network of MM, which may suggest a unique therapeutic target for MM.

Taken together, these date suggested there may be different therapeutic target for each form of disease but need further validation. In sum, our study will provided new information for future development of therapy for MGUS, SMM and MM.

### Disclosure of conflict of interest

None.

**Address correspondence to:** Nongjian Guo, Department of Hematology, Central Hospital of Ji'nan, No.105, Jiefang Road, Lixia District, Ji'nan 250013, Shandong, P.R. China. E-mail: guonongjian2016@sina.com

### References

- [1] Spitzer TR, Sachs DH and Cosimi B. Multiple myeloma. *N Engl J Med* 2011; 364: 2364; author reply 2364.
- [2] Palumbo A and Anderson K. Multiple myeloma. *N Engl J Med* 2011; 364: 1046-1060.
- [3] Papanikolaou X, Rosenthal A, Dhodapkar M, Epstein J, Khan R, van Rhee F, Jethava Y,

## Gene expression profiling among MGUS, SMM and MM

- Waheed S, Zangari M, Hoering A, Crowley J, Alapat D, Davies F, Morgan G and Barlogie B. Flow cytometry defined cytoplasmic immunoglobulin index is a major prognostic factor for progression of asymptomatic monoclonal gammopathies to multiple myeloma (subset analysis of SWOG S0120). *Blood Cancer J* 2016; 6: e410.
- [4] Dhodapkar MV, Sexton R, Waheed S, Usmani S, Papanikolaou X, Nair B, Petty N, Shaughnessy JD Jr, Hoering A, Crowley J, Orlowski RZ and Barlogie B. Clinical, genomic, and imaging predictors of myeloma progression from asymptomatic monoclonal gammopathies (SWOG S0120). *Blood* 2014; 123: 78-85.
- [5] Landgren O, Kyle RA, Pfeiffer RM, Katzmann JA, Caporaso NE, Hayes RB, Dispenzieri A, Kumar S, Clark RJ, Baris D, Hoover R and Rajkumar SV. Monoclonal gammopathy of undetermined significance (MGUS) consistently precedes multiple myeloma: a prospective study. *Blood* 2009; 113: 5412-5417.
- [6] Bories C and Jagannath S. Asymptomatic monoclonal gammopathies. *Clin Lymphoma Myeloma Leuk* 2014; 14 Suppl: S78-86.
- [7] Blade J. Clinical practice. Monoclonal gammopathy of undetermined significance. *N Engl J Med* 2006; 355: 2765-2770.
- [8] Iwanaga M, Tagawa M, Tsukasaki K, Kamihira S and Tomonaga M. Prevalence of monoclonal gammopathy of undetermined significance: study of 52,802 persons in Nagasaki City, Japan. *Mayo Clin Proc* 2007; 82: 1474-1479.
- [9] Landgren O, Katzmann JA, Hsing AW, Pfeiffer RM, Kyle RA, Yeboah ED, Biritwum RB, Tettey Y, Adjei AA, Larson DR, Dispenzieri A, Melton LJ 3rd, Goldin LR, McMaster ML, Caporaso NE and Rajkumar SV. Prevalence of monoclonal gammopathy of undetermined significance among men in Ghana. *Mayo Clin Proc* 2007; 82: 1468-1473.
- [10] Perez-Persona E, Vidriales MB, Mateo G, Garcia-Sanz R, Mateos MV, de Coca AG, Galende J, Martin-Nunez G, Alonso JM, de Las Heras N, Hernandez JM, Martin A, Lopez-Berges C, Orfao A and San Miguel JF. New criteria to identify risk of progression in monoclonal gammopathy of uncertain significance and smoldering multiple myeloma based on multiparameter flow cytometry analysis of bone marrow plasma cells. *Blood* 2007; 110: 2586-2592.
- [11] Kyle RA, Remstein ED, Therneau TM, Dispenzieri A, Kurtin PJ, Hodnefield JM, Larson DR, Plevak MF, Jelinek DF, Fonseca R, Melton LJ 3rd and Rajkumar SV. Clinical course and prognosis of smoldering (asymptomatic) multiple myeloma. *N Engl J Med* 2007; 356: 2582-2590.
- [12] Zhou Y, Barlogie B and Shaughnessy JD Jr. The molecular characterization and clinical management of multiple myeloma in the post-genome era. *Leukemia* 2009; 23: 1941-1956.
- [13] Anguiano A, Tuchman SA, Acharya C, Salter K, Gasparetto C, Zhan F, Dhodapkar M, Nevins J, Barlogie B, Shaughnessy JD Jr and Potti A. Gene expression profiles of tumor biology provide a novel approach to prognosis and may guide the selection of therapeutic targets in multiple myeloma. *J Clin Oncol* 2009; 27: 4197-4203.
- [14] Kuehl WM and Bergsagel PL. Multiple myeloma: evolving genetic events and host interactions. *Nat Rev Cancer* 2002; 2: 175-187.
- [15] Hanamura I, Stewart JP, Huang Y, Zhan F, Santra M, Sawyer JR, Hollmig K, Zangari M, Pineda-Roman M, van Rhee F, Cavallo F, Burington B, Crowley J, Tricot G, Barlogie B and Shaughnessy JD Jr. Frequent gain of chromosome band 1q21 in plasma-cell dyscrasias detected by fluorescence in situ hybridization: incidence increases from MGUS to relapsed myeloma and is related to prognosis and disease progression following tandem stem-cell transplantation. *Blood* 2006; 108: 1724-1732.
- [16] Tricot G, Barlogie B, Jagannath S, Bracy D, Mattox S, Vesole DH, Naucke S and Sawyer JR. Poor prognosis in multiple myeloma is associated only with partial or complete deletions of chromosome 13 or abnormalities involving 11q and not with other karyotype abnormalities. *Blood* 1995; 86: 4250-4256.
- [17] Shaughnessy JD Jr, Zhan F, Burington BE, Huang Y, Colla S, Hanamura I, Stewart JP, Kordsmeier B, Randolph C, Williams DR, Xiao Y, Xu H, Epstein J, Anaissie E, Krishna SG, Cottler-Fox M, Hollmig K, Mohiuddin A, Pineda-Roman M, Tricot G, van Rhee F, Sawyer J, Alsayed Y, Walker R, Zangari M, Crowley J and Barlogie B. A validated gene expression model of high-risk multiple myeloma is defined by deregulated expression of genes mapping to chromosome 1. *Blood* 2007; 109: 2276-2284.