

DOI:10.16781/j.0258-879x.2017.07.0877

• 论 著 •

## 基于高通量芯片和生物信息学挖掘多发性骨髓瘤的发病差异基因

吴随一<sup>1</sup>, 周潇逸<sup>2</sup>, 王 飞<sup>2</sup>, 杨依林<sup>2</sup>, 魏显招<sup>2\*</sup>

1. 第二军医大学海军医学系学员3队, 上海 200433

2. 第二军医大学长海医院脊柱外科, 上海 200433

**[摘要]** **目的** 筛选多发性骨髓瘤(MM)患者与正常人群之间的差异表达基因(DEGs), 探究MM的发病机制, 为MM的基因诊断和治疗提供导向。**方法** 从GEO数据库中检索获取MM患者的芯片数据, 通过Morpheus在线工具进行芯片数据质量控制和DEGs的筛选, 运用DAVID数据库对筛选获得的DEGs行基因富集和通路分析, 通过STRING数据库构建蛋白相互作用网络, 并采用Cytoscape软件行模块分析。**结果** 共获得16 211个DEGs, 包括7 586个上调基因和8 625个下调基因( $P < 0.05$ )。基因本体(GO)分析结果表明, 生物学过程中上调DEGs主要涉及鞘糖脂代谢等30个功能簇, 下调DEGs涉及细胞分裂等163个功能簇; 分子功能中上调DEGs主要涉及蛋白质结合等29个功能簇, 下调DEGs主要涉及组蛋白结合等59个功能簇; 细胞成分中上调DEGs主要集中在细胞溶质等27个功能簇中, 而下调DEGs主要集中在核质等78个功能簇中。京都基因与基因组百科全书(KEGG)分析结果表明, 上调DEGs主要涉及溶酶体相关通路等26条通路, 下调DEGs主要涉及DNA复制等27条通路。蛋白互作网络分析示CDK1、TOP2A、AURKB、BRCA1、CHEK1、PTEN、RAD51、GMP5、CDC45和CDKN2A 10个基因为富集程度最高的核心DEGs, 模块分析显示得分最高的3个基因模块主要与核分裂、DNA复制和核酸代谢过程相关。**结论** 通过多种生物信息学方法筛选获得了MM患者和健康对照组的DEGs, 并从不同角度阐释了MM发病机制的相关基因及其表达特征, 为MM特异性诊断标志和靶向治疗等提供了依据。

**[关键词]** 多发性骨髓瘤; 差异表达基因; 高通量基因芯片; 生物信息学

**[中图分类号]** R 733.3 **[文献标志码]** A **[文章编号]** 0258-879X(2017)07-0877-08

## Excavation of multiple myeloma associated differential genes based on high-throughput microarray and bioinformatics

WU Sui-yi<sup>1</sup>, ZHOU Xiao-yi<sup>2</sup>, WANG Fei<sup>2</sup>, YANG Yi-lin<sup>2</sup>, WEI Xian-zhao<sup>2\*</sup>1. 3<sup>rd</sup> Students Team, Faculty of Naval Medicine, Second Military Medical University, Shanghai 200433, China

2. Department of Spine Surgery, Changhai Hospital, Second Military Medical University, Shanghai 200433, China

**[Abstract]** **Objective** To screen the differentially expressed genes (DEGs) between multiple myeloma (MM) patients and the healthy controls, to explore the pathogenesis of MM, and to provide a theoretical basis for gene diagnosis and gene therapy. **Methods** Gene expression profiles of MM patients were obtained from GEO database. Morpheus Online (<https://software.broadinstitute.org/morpheus/>) was used to determine the chip-data quality control and DEGs screening; DAVID Online (<https://david.ncifcrf.gov/>) was used to perform the gene ontology and pathway enrichment analysis; STRING Online (<https://string-db.org/>) was used to integrate the protein-protein interaction (PPI) network, and Cytoscape was used to screen the modules of PPI. **Results** A total of 16 211 DEGs (7 586 up-regulated genes and 8 625 down-regulated genes) were identified ( $P < 0.05$ ). The up-regulated DEGs enriched in biological process terms mainly involved 30 functional categories, like glycosphingolipid metabolic process, and the down-regulated mainly involved 163 functional categories, like cell division; the up-regulated DEGs enriched in molecular function terms mainly involved 29 functional categories, like protein binding, and the down-regulated mainly involved 59 functional categories, like histone binding; the up-regulated DEGs enriched in cellular component terms mainly involved 27 functional categories, like cytosol, and the down-regulated mainly involved 78 functional categories, like nucleoplasm. The KEGG analysis showed that the up-regulated DEGs mainly involved 26 pathways, like lysosome related pathway, and the down-regulated were mainly related to 27 pathways, like DNA replication. Ten genes, including CDK1, TOP2A,

**[收稿日期]** 2017-02-25 **[接受日期]** 2017-06-26

**[作者简介]** 吴随一, 第二军医大学海军临床医学专业2013级学员。E-mail: wusuiyi@foxmail.com

\* 通信作者 (Corresponding author). Tel: 021-31161697, E-mail: weixianzhao@126.com

AURKB, BRCA1, CHEK1, PTEN, RAD51, GMP5, CDC45 and CDKN2A, were the hub DEGs with highest enrichment. Module analysis showed that the three most significant DEGs were mainly related to nuclear division, DNA replication and nucleic acid metabolism process. **Conclusion** A series of DEGs between MM patients and the healthy controls have been screened with a variety of bioinformatics methods, and gene and expression features of MM pathogenesis have been explained in various perspectives, which may provide the basis for targeted diagnosis therapy of MM.

[**Key words**] multiple myeloma; differentially expressed genes; high-throughput microarray; bioinformatics

[Acad J Sec Mil Med Univ, 2017, 38(7): 877-884]

多发性骨髓瘤(multiple myeloma, MM)是骨髓浆细胞克隆性增殖的血液系统恶性肿瘤<sup>[1]</sup>,约占全身恶性肿瘤的1%,占有血液系统肿瘤的10%。MM患者常伴有骨骼病变、高钙血症、严重的免疫缺陷以及细菌感染<sup>[2]</sup>。在我国,MM是第二常见的血液恶性肿瘤<sup>[3]</sup>,主要通过化疗、造血干/祖细胞移植、药物治疗和免疫治疗等方式进行治疗<sup>[4]</sup>,但术后5年生存率仅为44.9%<sup>[5]</sup>。目前,已有许多研究开始探索预防和治疗MM的新方法。B淋巴细胞刺激因子可作为诊断和治疗MM的生物标志物<sup>[6]</sup>;血清中自由轻链比例升高可作为无症状性MM进展的预测指标<sup>[7]</sup>;而在药物治疗方面,选择性蛋白酶抑制剂硼替佐米也表现出良好的治疗效果<sup>[8]</sup>。Lund等<sup>[9]</sup>报道沙利度胺联合美法仑+泼尼松治疗可延长MM患者的生存时间,但该治疗方式可能会引起疲劳、皮疹、多发性神经病和静脉血栓栓塞等并发症<sup>[10]</sup>。由于MM具有异质性较强、老年患者多见、对化疗药物的反应各异等特点,仍需进一步寻找MM早期特异性的诊断标志物及有效的治疗方法,延长患者的生存时间和提高其生活质量。

高通量基因芯片在肿瘤的分子诊断、肿瘤分类、患者预后和预测分析等方面应用广泛,有望为研究MM的发病机制和分子诊断提供新途径。本研究运用生物信息学技术对MM患者和健康对照组人群的基因表达芯片数据进行分析,挖掘MM患者和健康对照组人群的差异表达基因(differentially expressed genes, DEGs),并通过基因富集分析和通路分析、蛋白相互作用网络的构建以及蛋白网络的模块分析,探索MM相关基因通路和功能的变化,旨在为MM的早期诊断和基因治疗提供新的研究思路 and 理论依据。

## 1 材料和方法

1.1 基因表达芯片数据获取 由于MM的发病率相对其他肿瘤较低,肿瘤基因组图谱(the cancer

genome atlas, TCGA)数据库(<https://cancergenome.nih.gov/>)中未找到相关芯片样本集,故本研究的芯片数据集中于GEO数据库(<https://www.ncbi.nlm.nih.gov/gds/>)。从GEO数据库中根据以下3条标准进行芯片筛选:(1)MM患者标本,而非动物模型;(2)多发性;(3)有原始的高通量芯片数据。得到André等<sup>[11]</sup>上传的基于GPL570芯片分析平台(Affymetrix Human Genome U133 Plus 2.0 Array)的GSE36474样本集、López-Corral等<sup>[12]</sup>上传的基于GPL6244芯片分析平台(Affymetrix Human Gene 1.0 ST Array)的GSE47552样本集、Gutiérrez等<sup>[13]</sup>上传的基于GPL96芯片分析平台(Affymetrix Human Genome U133A Array)的GSE6691样本集共3个不同的芯片样本集。GSE36474芯片样本细胞种类为间充质干细胞,样本量为7;GSE47552芯片样本细胞种类为浆细胞,样本量为46;GSE6691芯片样本细胞种类为浆细胞,样本量为21。

1.2 样本均衡性检验和DEGs的筛选 分别对3个样本集进行均衡性检验,主要包括各个样本集MM组和对照组的性别、年龄等。3个样本集的原始文件被用于DEGs的筛选。将芯片样本集分别导入Morpheus在线工具(<https://software.broadinstitute.org/morpheus/>)进行数据质量控制。质量控制时要求:(1)基因中位数值至少发生2倍改变;(2)基因表达量变异P值小于0.05;(3)数据缺失值不超过50%。基因过滤后,根据层次聚类将数据分为两组:有相似表达模式的MM组和健康对照组。对MM组和健康对照组行t检验,满足 $P < 0.05$ 且倍数变化 $> 2$ 或 $< 0.5$ 的为DEGs,其中倍数变化 $< 0.5$ 为表达下调DEGs,倍数变化 $> 2$ 为表达上调DEGs。

1.3 DEGs的功能分析和基因富集分析 通过DAVID数据库(<https://david.ncifcrf.gov/>)在基因功能层面对DEGs进行基因本体(gene ontology, GO)富集分析<sup>[14]</sup>和京都基因与基因组百科全书(Kyoto

encyclopedia of genes and genomes, KEGG) 分析<sup>[15]</sup>。将上调或下调的 DEGs 分别输入 DAVID 数据库中, 并将基因名称转化为 Official-Gene-Symbol, 选取 GO 富集结果中的 GOTERM\_BP\_DIRECT(生物过程)、GOTERM\_MF\_DIRECT(分子功能)和 GOTERM\_CC\_DIRECT(细胞成分)进行 GO 富集分析。此外, 选取通路结果中的 KEGG\_PATHWAY, 对 DEGs 的集中信号通路进行分析。

1.4 蛋白质相互作用网络的建立和模块分析 为进一步挖掘 DEGs 中的核心基因, 研究通过 STRING 数据库<sup>[16]</sup>(<http://string-db.org/>)在线分析 DEGs 对应编码蛋白的相互作用, 筛选有实验验证的、综合得分为 0.4 以上(满分为 1)的基因构建相互作用网络。使用信息分析学软件 Cytoscape (3.4.0 版本)<sup>[17]</sup>分析 STRING 数据库构建出的蛋白质相互作用网络, 通过 Cytoscape 的 MCODE 功能得出蛋白质相互作用网络中的相关模块, 标准设置如下: (1) MCODE 得分 > 3; (2) 节点数 > 4.0。分析每个相关模块的功能和通路丰度。

1.5 统计学处理 MM 组和健康对照组的均衡性检验和 DEGs 筛选通过 Morpheus 在线工具自带统计学工具行  $\chi^2$  检验和  $t$  检验。GO 分析通过 DAVID 数据库自带统计学工具使用弗朗尼校正法 (Bonferroni), 本杰明假阳性率法 (Benjamini false discovery rate) 和靴带法 (boot straping) 等方法进行检验。KEGG 分析通过 DAVID 数据库自带统计学工具进行 Fisher 精确概率检验及基因集富集分析 (gene set enrichment analysis)。蛋白质相互作用网络通过 Cytoscape 软件进行 Pearson 相关系数分析后, 将基因表达数据进行层次聚类, 连接方法为平均连接聚类法 (average linkage) 和中位数标准化。检验水准 ( $\alpha$ ) 为 0.05。

## 2 结果

2.1 样本均衡性检验和 DEGs 的筛选 3 个样本集中 MM 组与健康对照组的性别、年龄等差异均无统计学意义 ( $P > 0.05$ ), 共对基因表达芯片数据中的 61 例 MM 患者和 13 名健康对照组受试者的数据进行了分析。获得 16 211 个 DEGs, 其中表达上调的基因有 7 586 个, 下调的有 8 625 个。DEGs 热图(前 50 位上调基因和前 50 位下调基因, 以 GSE36474 样本集为例)见图 1。

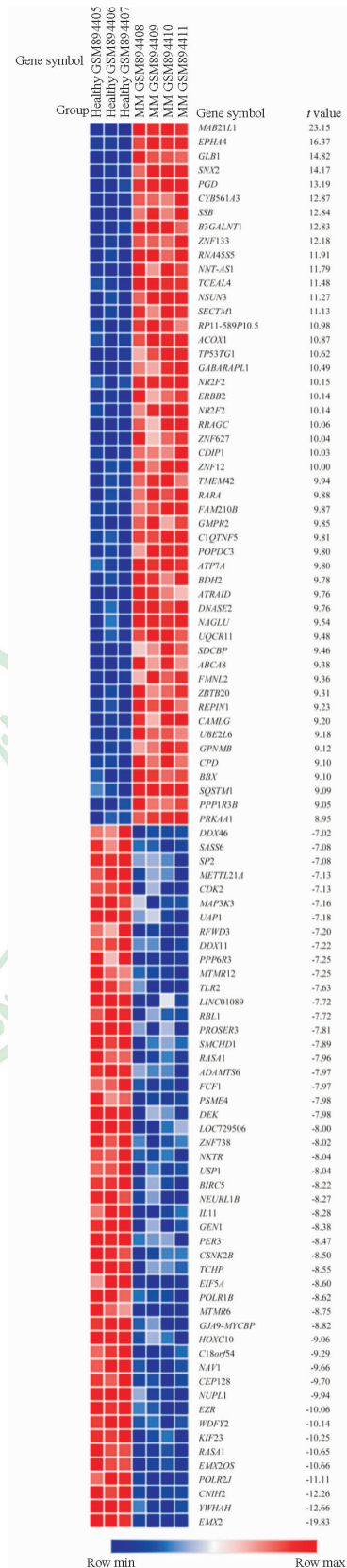


图 1 GSE36474 样本集中 DEGs 的热图  
Fig 1 Heat map of differentially expressed genes (DEGs) in sample set GSE36474

The top 50 up-regulated genes and the top 50 down-regulated genes. Red: Up-regulation; Blue: Down-regulation. MM: Multiple myeloma

2.2 GO 功能分析 将所有筛选所得的 DEGs 通过 DAVID 数据库进行分析,得出有代表性的 GO 类别和 KEGG 通路。GO 分析结果显示,生物学过程中上调的 DEGs 主要涉及鞘糖脂代谢过程 (glycosphingolipid metabolic process)、DNA 转录过程 (transcription, DNA-templated) 和干扰素  $\gamma$  介导的信号转导通路 (interferon-gamma-mediated signaling pathway) 等 30 个功能簇;下调的 DEGs 主要涉及细胞分裂 (cell division)、DNA 复制 (DNA replication) 和有丝分裂核分裂 (mitotic nuclear division) 等 163 个功能簇。关于分子功能,上调的 DEGs 主要涉及蛋白结合 (protein binding)、转录因子与特异序列 DNA 结合的活性 (transcription

factor activity, sequence-specific DNA binding) 和烟酰胺腺嘌呤二核苷酸结合 (NAD binding) 等 29 个功能簇;而下调的 DEGs 主要涉及组蛋白结合 (histone binding)、多聚 A 尾 RNA 结合 [poly(A) RNA binding] 和 DNA 结合 (DNA binding) 等 59 个功能簇。此外,GO 细胞成分分析显示,上调的 DEGs 主要集中在细胞溶质 (cytosol)、溶酶体膜 (lysosomal membrane) 和溶酶体 (lysosome) 等 27 个功能簇中;而下调的 DEGs 主要集中在核质 (nucleoplasm)、细胞核 (nucleus) 和细胞质 (cytoplasm) 等 78 个功能簇中。部分 GO 分析结果见表 1。

表 1 MM 患者中 DEGs 的 GO 分析

Tab 1 Gene ontology analysis of differentially expressed genes (DEGs) associated with MM

Expression	Category	Term	Count	Percent (%)	P value	
Up-regulated	GOTERM_BP_DIRECT	GO:0006687-glycosphingolipid metabolic process	12	1.44	1.48E-06	
	GOTERM_BP_DIRECT	GO:0006351-transcription, DNA-templated	116	13.96	3.76E-05	
	GOTERM_BP_DIRECT	GO:0060333-interferon-gamma-mediated signaling pathway	12	1.44	1.41E-04	
	GOTERM_BP_DIRECT	GO:0051607-defense response to virus	19	2.29	1.55E-04	
	GOTERM_BP_DIRECT	GO:0060337-type I interferon signaling pathway	11	1.32	2.62E-04	
	GOTERM_MF_DIRECT	GO:0005515-protein binding	295	35.50	3.49E-04	
	GOTERM_MF_DIRECT	GO:0003700-transcription factor activity, sequence-specific DNA binding	62	7.46	3.97E-04	
	GOTERM_MF_DIRECT	GO:0051287-NAD binding	8	0.96	0.0012	
	GOTERM_MF_DIRECT	GO:0008307-structural constituent of muscle	8	0.96	0.0012	
	GOTERM_MF_DIRECT	GO:0035091-phosphatidylinositol binding	11	1.32	0.0024	
	GOTERM_CC_DIRECT	GO:0005829-cytosol	197	23.71	7.99E-08	
	GOTERM_CC_DIRECT	GO:0005765-lysosomal membrane	33	3.97	9.11E-08	
	GOTERM_CC_DIRECT	GO:0005764-lysosome	29	3.49	2.43E-07	
	GOTERM_CC_DIRECT	GO:0043202-lysosomal lumen	16	1.93	1.59E-06	
	GOTERM_CC_DIRECT	GO:0005739-mitochondrion	82	9.87	1.88E-04	
	Down-regulated	GOTERM_BP_DIRECT	GO:0051301-cell division	77	9.66	3.28E-35
		GOTERM_BP_DIRECT	GO:0006260-DNA replication	50	6.27	2.10E-30
		GOTERM_BP_DIRECT	GO:0007067-mitotic nuclear division	60	7.52	2.93E-29
		GOTERM_BP_DIRECT	GO:0034080-CENP-A containing nucleosome assembly	26	3.26	2.31E-24
GOTERM_BP_DIRECT		GO:0007062-sister chromatid cohesion	36	4.51	1.01E-23	
GOTERM_MF_DIRECT		GO:0042393-histone binding	27	3.38	2.45E-12	
GOTERM_MF_DIRECT		GO:0044822-poly(A) RNA binding	95	11.91	1.27E-11	
GOTERM_MF_DIRECT		GO:0003677-DNA binding	122	15.30	3.35E-10	
GOTERM_MF_DIRECT		GO:0005524-ATP binding	109	13.67	1.58E-09	
GOTERM_MF_DIRECT		GO:0019904-protein domain specific binding	30	3.76	4.76E-09	
GOTERM_CC_DIRECT		GO:0005654-nucleoplasm	270	33.87	4.33E-53	
GOTERM_CC_DIRECT		GO:0005634-nucleus	373	46.81	3.93E-40	
GOTERM_CC_DIRECT		GO:0005737-cytoplasm	308	38.64	8.86E-19	
GOTERM_CC_DIRECT		GO:0000775-chromosome, centromeric region	22	2.76	7.15E-16	
GOTERM_CC_DIRECT		GO:0000228-nuclear chromosome	21	2.63	1.43E-15	

The top 5 significant GO terms of each category of the up-regulated DEGs or down-regulated DEGs, respectively. MM; Multiple myeloma; BP; Biological process; MF; Molecular function; CC; Cellular component

2.3 KEGG 信号通路分析 通过 KEGG 分析富集得到差异最显著的上调和下调 DEGs 所在的信号通路。上调的 DEGs 主要涉及溶酶体相关通路 (lysosome)、病毒致癌作用 (viral carcinogenesis) 和丙型肝炎相关通路 (hepatitis C) 等 26 条信号通路;

而下调的 DEGs 主要涉及 DNA 复制 (DNA replication)、Fanconi 贫血通路 (Fanconi anemia pathway) 和卵母细胞减数分裂 (oocyte meiosis) 等 27 条信号通路 (表 2)。

表 2 MM 患者中 DEGs 的 KEGG 分析

Tab 2 KEGG pathway analysis of differentially expressed genes (DEGs) associated with MM

Expression	Term	Name	Count	Percent (%)	P value	Genes
Up-regulated	hsa04142	Lysosome	24	2.89	2.33E-10	NAGLU, LIPA, GM2A, AP1G1, PSAP, ATP6AP1, HEXA, ATP6V1H, CTSA, MANBA, ASAH1, GLB1, DNASE2, GNS, LAMP1, LAMP2, NPC2, IDS, TPP1, IGF2R, SMPD1, GALC, NEU1, GBA
	hsa05203	Viral carcinogenesis	22	2.65	5.62E-05	HIST1H2BC, LTBR, RBL2, HIST1H2BE, HIST1H2BF, IL6ST, HIST1H2BG, BAD, HLA-E, GTF2B, STAT3, GTF2H1, TRADD, IRF9, EP300, CDKN1B, CDKN2A, HIST1H2BI, BAX, GTF2A2, MAPK3, PIK3R2
	hsa05160	Hepatitis C	14	1.68	0.0023	IRF9, TNFRSF1A, RNASEL, IFIT1, SOS1, MAPK3, TLR3, SCARB1, BAD, STAT1, STAT3, PIK3R2, TRADD, STAT2
	hsa00511	Other glycan degradation	5	0.60	0.0048	HEXA, NEU1, MANBA, GLB1, GBA
	hsa04146	Peroxisome	10	1.20	0.0053	ACOX2, ACOX1, ECH1, GSTK1, IDH1, MPV17, HSD17B4, DDO, PHYH, ACAA1
Down-regulated	hsa03030	DNA replication	18	0.02	1.52E-14	POLA1, MCM2, RNASEH2A, MCM3, MCM4, RNASEH2B, MCM5, RFC5, RPA1, PRIM1, DNA2, RFC3, MCM7, POLE2, RFC2, POLE3, PRIM2, FEN1
	hsa03460	Fanconi anemia pathway	16	0.01	3.17E-09	USP1, EME1, BRCA2, BRIP1, RMI2, RMI1, BRCA1, RAD51, RPA1, FANCM, FANCD2, FANCI, FANCG, FANCA, UBE2T, FANCB
	hsa04114	Oocyte meiosis	19	0.02	7.18E-07	CDK1, YWHAZ, ANAPC5, CPEB3, PPP2R5C, PPP3R1, PKMYT1, AURKA, ESPL1, CDC25C, CDC27, SMC3, CDK2, CCNE2, YWHAH, MAD2L1, YWHAQ, BUB1, FBXO5
	hsa04914	Progesterone-mediated oocyte maturation	14	0.01	7.65E-05	CDK1, ANAPC5, CPEB3, PKMYT1, RAF1, CDC25C, CDC27, CDK2, CDC25A, MAD2L1, BUB1, PIK3R5, CCNA2, AKT3
	hsa05130	Pathogenic <i>Escherichia coli</i> infection	11	0.01	4.92E-05	ACTB, ACTG1, CDC42, EZR, ARPC2, ARPC5L, TUBA3C, TUBA3D, TUBB6, TUBA1B, TUBA1C

The top 5 significantly enriched pathways of the up-regulated DEGs or down-regulated DEGs, respectively. MM: Multiple myeloma

2.4 蛋白质相互作用网络中的模块分析 基于 STRING 数据库筛选出得分最高的前 10 位核心基因, 分别是细胞周期蛋白依赖性激酶 1 (cyclin-dependent kinase 1, CDK1)、拓扑异构酶 II  $\alpha$  亚基 [topoisomerase (DNA) II  $\alpha$ , TOP2A]、光激酶 B (aurora kinase B, AURKB)、乳腺癌易感基因 1 (breast cancer susceptibility gene 1, BRCA1)、细胞周期检测点激酶 1 (checkpoint kinase 1, CHEK1)、磷酸酶张力蛋白同源物 (phosphatase and tensin homolog, PTEN)、RAD51、单磷酸鸟嘌呤合成酶

(guanine monophosphate synthetase, GMPS)、细胞分裂周期 45 同源物 (cell division cycle 45 homolog, CDC45) 和细胞周期蛋白依赖性激酶抑制剂 2A (cyclin-dependent kinase inhibitor 2A, CDKN2A)。在这些核心基因中, CDK1 的相关节点数最高, 为 157 个。此外, 使用 Cytoscape 的 MCODE 功能对 STRING 数据库分析得出的蛋白网络中的 1180 个节点蛋白的 8250 条关系进行了分析, 筛选出 MCODE 得分最高的 3 个模块 (图 2, 表 3)。对 3 个模块中的基因均进行了功能分析, 结

果显示,模块1~3中丰度最高的基因主要与核分裂(nuclear division)、DNA复制(DNA replication)和

核酸代谢过程(nucleic acid metabolic process)相关。

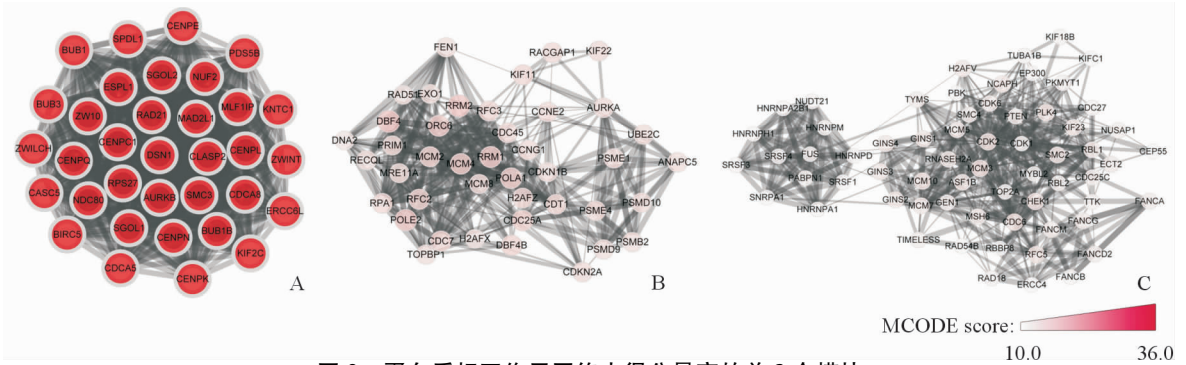


图2 蛋白质相互作用网络中得分最高的前3个模块

Fig 2 Top 3 modules from the protein-protein interaction network

A: Enriched pathways of module 1; B: Enriched pathways of module 2; C: Enriched pathways of module 3

表3 蛋白质相互作用网络中得分最高的前3个模块及其功能分析

Tab 3 Top 3 modules from the protein-protein interaction network and the related functions

Module	Gene set	P value	FDR	Nodes
1	Nuclear division	1.49E-42	2.55E-40	ERCC6L, ZWILCH, CDCA5, BUB1B, CASC5, CDCA8, PDS5B, SMC3, AURKB, DSN1, SGOL1, RAD21, NUF2, KNTC1, BUB3, BUB1, CLASP2, NDC80, ZWINT, CENPE, ESPL1, ZW10, BIRC5, KIF2C, MAD2L1
	Mitosis	1.49E-42	2.55E-40	ERCC6L, ZWILCH, CDCA5, BUB1B, CASC5, CDCA8, PDS5B, SMC3, AURKB, DSN1, SGOL1, RAD21, NUF2, KNTC1, BUB3, BUB1, CLASP2, NDC80, ZWINT, CENPE, ESPL1, ZW10, BIRC5, KIF2C, MAD2L1
	M phase of mitotic cell cycle	3.26E-42	3.47E-40	ERCC6L, ZWILCH, CDCA5, BUB1B, CASC5, CDCA8, PDS5B, SMC3, AURKB, DSN1, SGOL1, RAD21, NUF2, KNTC1, BUB3, BUB1, CLASP2, NDC80, ZWINT, CENPE, ESPL1, ZW10, BIRC5, KIF2C, MAD2L1
2	DNA replication	3.29E-33	2.12E-30	CDT1, FEN1, RRM1, RFC3, RRM2, MCM8, RFC2, PRIM1, RPA1, RECQL, MRE11A, CDC7, CDC25A, POLA1, CDC45, ORC6, RAD51, DBF4, CCNE2, POLE2, MCM4, DNA2, MCM2
	DNA metabolic process	6.43E-31	2.07E-28	FEN1, MCM8, PRIM1, CDC45, ORC6, EXO1, TOPBP1, CDT1, RRM1, RFC3, RRM2, CDKN2A, RFC2, H2AFX, RPA1, RECQL, MRE11A, CDC7, KIF22, CDC25A, POLA1, RAD51, DBF4, CCNE2, POLE2, MCM4, DNA2, MCM2
	Cell cycle	3.53E-24	7.58E-22	PSMD10, CDKN1B, MCM8, KIF11, DBF4B, AURKA, PSMD9, CDC45, RACGAP1, PSMB2, EXO1, CDT1, CDKN2A, UBE2C, H2AFX, RPA1, MRE11A, CDC7, KIF22, CDC25A, POLA1, RAD51, DBF4, CCNG1, PSME1, ANAPC5, MCM2
3	Nucleic acid metabolic process	4.26E-23	3.79E-20	TOP2A, RNASEH2A, MCM7, SRSF1, MCM10, TYMS, RAD54B, PABPN1, CHEK1, RBBP8, HNRNPA1, GEN1, RFC5, GINS1, GINS2, FANCM, FUS, GINS3, GINS4, FANCA, FANCB, CDC6, CDC25C, FANCG, MSH6, HNRNPM, NUDT21, FANCD2, HNRNPH1, ERCC4, HNRNPA2B1, CDK2, MCM3, HNRNPD, SRSF3, SRSF4, MCM5, SNRPA1, RAD18
	DNA metabolic process	1.38E-22	6.17E-20	TOP2A, RNASEH2A, MCM7, MCM10, TYMS, RAD54B, CHEK1, RBBP8, GEN1, RFC5, GINS1, GINS2, FANCM, GINS3, GINS4, FANCA, FANCB, CDC6, CDC25C, FANCG, MSH6, FANCD2, ERCC4, CDK2, MCM3, MCM5, RAD18
	M phase	6.16E-20	1.45E-17	FANCA, KIF23, TTK, FANCB, CDC6, CDC25C, PKMYT1, SMC4, NCAPH, SMC2, RAD54B, MSH6, FANCD2, KIFC1, CHEK1, CDC27, CDK2, PBK, NUSAP1, TIMELESS, CDK1, CEP55

FDR: False discovery rate

### 3 讨论

本研究对 GEO 数据库中的基因表达芯片数据

文件 GSE36474、GSE47552 和 GSE6691 中 MM 患者和健康人的样本信息进行了生物信息学分析,最终筛选出 16 211 个 DEGs,其中 7 586 个表达上调,

8 625 个表达下调。与健康对照人群相比,MM 患者的样本中上调表达的基因主要有 *CDK1*、*TOP2A*、*AURKB*、*BRCA1*、*CHEK1*、*PTEN*、*RAD51*、*GMPS*、*CDC45* 和 *CDKN2A* 等。

通过 GO 分析,本研究发现 MM 患者的上调 DEGs 主要涉及鞘糖脂代谢过程、DNA 转录过程和干扰素  $\gamma$  介导的信号转导通路等生物学过程。鞘糖脂是一类广泛分布在细胞膜上的糖脂类物质,在调控细胞的识别、黏附、增殖和凋亡等生物学过程中有重要作用。Ersek 等<sup>[18]</sup> 研究发现,MM 中激活破骨细胞需要鞘糖脂的合成,并且激活的破骨细胞能够进一步促进鞘糖脂的合成,形成正反馈作用。在骨髓平衡紊乱疾病的药物治疗中,限制鞘糖脂的合成能够有效预防 MM 的发生<sup>[19]</sup>。干扰素  $\gamma$  有着强大的抗肿瘤效果,应用于干扰素  $\gamma$  治疗 MM 疗效明显<sup>[20]</sup>。Haydaroglu 等<sup>[21]</sup> 比较了包括干扰素  $\gamma$  在内的多种不同药物治疗 MM 的效果,结果发现干扰素  $\gamma$  相关基因及其等位基因可能在 MM 中发挥作用;本研究在 GO 分析中也发现 MM 患者中表达显著上调的基因集中在干扰素  $\gamma$  介导的信号转导通路中。

KEGG 通路分析结果表明,DEGs 中丰度最高的信号通路是溶酶体相关通路。李礼轩和李佳<sup>[22]</sup> 研究发现,通过慢病毒介导的 shRNA 沉默溶酶体相关膜蛋白 2A 的表达可抑制 MM 细胞增殖。细胞自噬是依赖溶酶体途径对细胞内物质进行处理的重要过程,自噬活性与 MM 的发生和发展密切相关<sup>[23]</sup>。KEGG 通路分析显示下调的 DEGs 所在的信号通路主要集中在细胞周期。孙婉玲等<sup>[24]</sup> 分析了 MM 细胞的 DNA 含量和细胞周期,发现 MM 患者处于 S+G<sub>2</sub>/M 期的骨髓瘤细胞(CD138<sup>+</sup>)比例明显上升,这和 KEGG 的通路富集结果相对应。细胞增殖失调是肿瘤细胞的特点之一,肿瘤细胞通常对调节细胞周期的基因(如 *CDK1*、*CHEK1* 和 *CDC45* 等)直接调控,使细胞增殖紊乱。

通过 Cytoscape 对 STRING 数据库产生的蛋白相互作用网络进行分析,本研究获得 DEGs 中得分最高的前 10 位核心基因,*CDK1* 在真核细胞中通过调节中心体周期以及有丝分裂的发生,对细胞周期的调控起着关键作用,*CDK1* 可以促进 G<sub>2</sub>-M 过渡、加速 G<sub>1</sub> 过程和促进 G<sub>1</sub>-S 过渡<sup>[25]</sup>。*CDK1* 是细胞进入 S 期和进行有丝分裂所必需的。*BRCA1* 在乳腺癌发病早期表达上升,并且特异性介导 E3 泛素蛋白连接酶多聚泛素链中六聚赖氨酸的形成,进而促进细胞对受损 DNA 的修复。*BRCA1-BARD1* 二聚体协调调控 DNA 修复、蛋白质泛素化和转录调控

等细胞通路,从而维持基因组的稳定性<sup>[26]</sup>。已有研究报道 *CHEK1* 在人类 MM 中调节细胞 G<sub>0</sub>/G<sub>1</sub> 期<sup>[27]</sup>,Tu 等<sup>[28]</sup> 研究发现在 MM 中抑制 *CHEK1* 基因的表达可以缩短 S 期,促进骨髓瘤细胞凋亡。本研究发现 *CHEK1* 表达上调,这和细胞周期密切相关,表明 *CHEK1* 可能是治疗 MM 的一个潜在靶点。

本研究挖掘到了一系列关键基因,但也存在部分局限:(1)前期缺少相关临床标本,未进行基础实验验证。目前课题组已收集标本,相关实验正在进行中。(2)数据进行了质量控制,基本信息的均衡性也满足要求,但患者个体间的状态差异未全部提及(GSE36474 患者均为 MM III a 期,而 GSE47552 和 GSE6691 分期未知),提示我们在进行验证实验时需保持患者状态的均一性,减少组内个体差异。(3)由于 TCGA 数据库中未收录 MM 的芯片信息,故研究仅分析了 GEO 数据库的芯片信息。

综上,本研究通过 Morpheus 在线工具、DAVID 数据库、STRING 数据库和 Cytoscape 等多种生物信息学工具对基因芯片数据进行了筛选、统合、挖掘和分析,探索 MM 相关分析标签、DEGs、信号通路和功能模块的变化,从多个角度分析 MM 的表达特征。研究表明,鞘糖脂代谢过程和细胞周期可能与 MM 的发展密切相关,干扰素  $\gamma$  有望成为 MM 的潜在治疗剂,*CDK1*、*BRCA1-BARD1* 和 *CHEK1* 基因可作为 MM 治疗的潜在靶点。这些结果为进一步的生物学验证和基础研究提供了依据,也为今后 MM 的诊断和治疗研究提供了新方向。

## [参考文献]

- [1] LONIAL S, ANDERSON K C. Association of response endpoints with survival outcomes in multiple myeloma[J]. *Leukemia*, 2014, 28: 258-268.
- [2] RAAB M S, PODAR K, BREITKREUTZ I, RICHARDSON P G, ANDERSON K C. Multiple myeloma[J]. *Lancet*, 2009, 374: 324-339.
- [3] 陈协群. 中国多发性骨髓瘤诊治指南(2015 年修订)[J]. *中华内科杂志*, 2016, 54: 1066-1070.
- [4] 吴玉玲, 黄东平. 多发性骨髓瘤的治疗进展[J]. *现代医药卫生*, 2016, 32: 2828-2830.
- [5] WANG X G, PENG Y, SONG X L, LAN J P. Identification potential biomarkers and therapeutic agents in multiple myeloma based on bioinformatics analysis[J]. *Eur Rev Med Pharmacol Sci*, 2016, 20: 810-817.
- [6] JIANG P, YUEGUO W, HUIMING H, HONGXIANG Y, MEI W, JU S. B-lymphocyte stimulator: a new biomarker for multiple myeloma[J].

- Eur J Haematol, 2009, 82: 267-276.
- [7] LARSEN J, KUMAR S K, DISPENZIERI A, KYLE R, KATZMANN J, RAJKUMAR S V. Serum free light chain ratio as a biomarker for high-risk smoldering multiple myeloma [J]. *Leukemia*, 2013, 27: 941-946.
- [8] RICHARDSON P G, SONNEVELD P, SCHUSTER M W, IRWIN D, STADTMAUER E A, FACON T, et al. Bortezomib or high-dose dexamethasone for relapsed multiple myeloma [J]. *N Engl J Med*, 2005, 352: 2487-2498.
- [9] LUND J, UTTERVALL K, LIWING J, GAHRTON G, ALICI E, ASCHAN J, et al. Addition of thalidomide to melphalan and prednisone treatment prolongs survival in multiple myeloma—a retrospective population based study of 1 162 patients [J]. *Eur J Haematol*, 2014, 92: 19-25.
- [10] HAAS P, DENZ U, IHORST G, ENGELHARDT M. Thalidomide in consecutive multiple myeloma patients: single-center analysis on practical aspects, efficacy, side effects and prognostic factors with lower thalidomide doses [J]. *Eur J Haematol*, 2008, 80: 303-309.
- [11] ANDRÉ T, MEULEMAN N, STAMATOPOULOS B, DE BRUYN C, PIETERS K, BRON D, et al. Evidences of early senescence in multiple myeloma bone marrow mesenchymal stromal cells [J/OL]. *PLoS One*, 2013, 8: e59756. doi: 10.1371/journal.pone.0059756.
- [12] LÓPEZ-CORRAL L, CORCHETE L A, SARASQUETE M E, MATEOS M V, GARCÍA-SANZ R, FERMIÑÁN E, et al. Transcriptome analysis reveals molecular profiles associated with evolving steps of monoclonal gammopathies [J]. *Haematologica*, 2014, 99: 1365-1372.
- [13] GUTIÉRREZ N, OCIO E, DE LAS RIVAS J, MAISO P, DELGADO M, FERMINAN E, et al. Gene expression profiling of B lymphocytes and plasma cells from Waldenström's macroglobulinemia: comparison with expression patterns of the same cell counterparts from chronic lymphocytic leukemia, multiple myeloma and normal individuals [J]. *Leukemia*, 2007, 21: 541-549.
- [14] Gene Ontology Consortium. Gene ontology consortium: going forward [J]. *Nucleic Acids Res*, 2015, 43: D1049-D1056.
- [15] KANEHISA M, GOTO S. KEGG: Kyoto encyclopedia of genes and genomes [J]. *Nucleic Acids Res*, 2000, 28: 27-30.
- [16] SZKLARCZYK D, FRANCESCHINI A, WYDER S, FORSLUND K, HELLER D, HUERTA-CEPAS J, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life [J]. *Nucleic Acids Res*, 2015, 43: D447-D452.
- [17] SHANNON P, MARKIEL A, OZIER O, BALIGA N S, WANG J T, RAMAGE D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks [J]. *Genome Res*, 2003, 13: 2498-2504.
- [18] ERSEK A, XU K, ANTONOPOULOS A, BUTTERS T D, SANTO A E, VATTAKUZZI Y, et al. Glycosphingolipid synthesis inhibition limits osteoclast activation and myeloma bone disease [J]. *J Clin Invest*, 2015, 125: 2279-2292.
- [19] ERSEK A, KARADIMITRIS A, HORWOOD N. Effect of glycosphingolipids on osteoclastogenesis and osteolytic bone diseases [J]. *Front Endocrinol (Lausanne)*, 2012, 3: 106.
- [20] 周亚兰, 米瑞华, 艾昊, 陈琳, 袁芳芳, 尹青松, 等. 沙利度胺联合干扰素在复发难治及体能、经济状况欠佳多发性骨髓瘤患者中的应用 [J]. *白血病·淋巴瘤*, 2016, 25: 479-482.
- [21] HAYDAROGLU H, OGUZKAN BALCI S, PEHLI VAN S, OZDILLI K, GUNDOGAN E, OKAN V, et al. Effect of cytokine genes in the pathogenesis and on the clinical parameters for the treatment of multiple myeloma [J]. *Immunol Invest*, 2017, 46: 10-21.
- [22] 李礼轩, 李佳. 慢病毒介导的 shRNA 沉默溶酶体相关膜蛋白 2A 表达可抑制多发性骨髓瘤细胞增殖 [J]. *细胞与分子免疫学杂志*, 2015, 31: 605-608.
- [23] MIZUSHIMA N, KOMATSU M. Autophagy: renovation of cells and tissues [J]. *Cell*, 2011, 147: 728-741.
- [24] 孙婉玲, 武永吉, 汪玄, 李辉, 庄俊玲. 多发性骨髓瘤细胞 DNA 含量和细胞周期分析 [J]. *中国实验血液学杂志*, 2008, 16: 824-828.
- [25] BENANTI J A. Create, activate, destroy, repeat: Cdk1 controls proliferation by limiting transcription factor activity [J]. *Curr Genet*, 2016, 62: 271-276.
- [26] MEHRGOU A, AKOUCHEKIAN M. The importance of *BRCA1* and *BRCA2* genes mutations in breast cancer development [J]. *Med J Islam Repub Iran*, 2016, 30: 369.
- [27] PEI X Y, DAI Y, YOUSSEFIAN L E, CHEN S, BODIE W W, TAKABATAKE Y, et al. Cytokinetically quiescent ( $G_0/G_1$ ) human multiple myeloma cells are susceptible to simultaneous inhibition of Chk1 and MEK1/2 [J]. *Blood*, 2011, 118: 5189-5200.
- [28] TU Y S, KANG X L, ZHOU J G, LV X F, TANG Y B, GUAN Y Y. Involvement of Chk1-Cdc25A-cyclin A/CDk2 pathway in simvastatin induced S-phase cell cycle arrest and apoptosis in multiple myeloma cells [J]. *Eur J Pharmacol*, 2011, 670: 356-364.