Identification of hub genes with prognostic values in multiple myeloma by bioinformatics analysis

Jie Yang¹, Fei Wang¹, Shanliang Zhong² and Baoan Chen³

¹Department of Hematology and Oncology, Zhongda Hospital, School of Medicine, Southeast University, Nanjing, People’s Republic of China; ²Center of Clinical Laboratory Science, Jiangsu Cancer Hospital Affiliated to Nanjing Medical University, Nanjing, People’s Republic of China

ABSTRACT
Purpose: Multiple myeloma (MM) is a malignant disease with abnormal proliferation of clonal plasma cells. Hypoxia is an important factor in the pathogenesis and development of MM. However, the underlying mechanisms are not fully understood.

Material & Methods: To determine hub genes related to hypoxia in MM, this study took integrated bioinformatics analysis with two expression datasets (GSE80140 and GSE80545) downloaded from Gene Expression Omnibus (GEO) database. Differentially expressed genes (DEGs) were filtered under the condition of both p-value < 0.05 and \( \log_{2}\text{FoldChange} > 1 \). Then, gene ontology (GO) and Kyoto encyclopedia of genes and genomes enrichment (KEGG) analysis, and protein–protein interaction (PPI) network construction were utilized to further explore these DEGs. PrognoScan evaluated all the candidate hub genes for survival analysis.

Results: In total, three hub genes, including FH, TSTA3, and POLR3G, were screened out to be related to hypoxia in MM. Patients with the lower expression level of FH, TSTA3, and POLR3G have statistically significantly longer disease-specific survival (Cox \( p < 0.05 \)).

Conclusion: We identified FH, TSTA3, and POLR3G as hub genes which can affect MM patients’ outcome and new biomarkers for diagnosis and prognosis of MM. Further functional and mechanistic studies are need to develop in order to make them as potential target for clinical treatment.

1. Introduction

Multiple myeloma (MM) is a hematological malignancy with about 1% morbidity and mortality of all tumors. In many countries, it has been the second most common malignant disease of the blood system [1]. Recently, there are data showing that the incidence of MM in Asian countries is increasing rapidly [2]. The characteristics of MM mainly include anemia, bone pain, and renal failure, leading to death eventually [3]. With the great progress of treatments including immunoregulatory drugs, proteasome inhibitors, histone deacetylase inhibitors, and monoclonal antibodies in the past few decades, the survival period of MM has been significantly extended, but it still cannot be completely cured [4,5]. As a result, identifying new biomarkers is particularly essential for early diagnosis, effective treatment, and better prognosis of MM.

The hypoxic bone marrow (BM) microenvironment is an important factor for the occurrence and progression of MM. It is mainly manifested in the following aspects: (1) hypoxia reduces the adhesion of tumor cells and accelerates tumor development process [6,7]; (2) hypoxia promotes the expression of vascular endothelial growth factor A, and then neovascularization in the BM is accelerated [8]; (3) hypoxia upregulates the expression of glycolysis-related genes such as lactate dehydrogenase A in MM cells, and the process of cellular glycolysis is significantly enhanced [9]; (4) MM cells interact with BMstromal cells in the hypoxic microenvironment, and some clones enter a dormant state and then have drug resistance [10].

In recent years, high-throughput platforms in gene expression (GE) is more and more important as a method to explore pathogenesis, prognosis prediction, and potential new target drugs of various diseases, including tumor [11]. To date, several chip analyses have revealed the pathogenesis of MM [6,12,13]. However, no chip analysis on hypoxia in MM is conducted, and related critical genes and signaling pathways remain to be clarified. So integrated bioinformatics analysis, including various methods, is utilized to explore hub genes related to hypoxia in MM.

In this study, two microarray datasets (GSE80140 and GSE80545) were downloaded from National Center for Biotechnology Information-Gene Expression Omnibus database (NCBI-GEO) (https://www.ncbi.nlm.nih.gov/)
We picked out differentially expressed genes (DEGs) and further analyzed these DEGs’ functions and pathway enrichment. Survival analysis was conducted with the online clinical data, and the results showed that expression levels of FH, TSTA3, and POLR3G were statistically significantly related to MM patients’ outcomes. These analysis results provided important information for further research in MM.

2. Material and methods

2.1. Microarray data information and DEG identification

The two GE datasets GSE80140 and GSE80545 were downloaded from NCBI-GEO [14]. Both GSE80140 and GSE80545 were generated with Agilent-039494 SurePrint G3 Human GE v2 8 × 60K Microarray 039381. GSE80140 compared four myeloma cell lines (RPMI8226, KMS-11, KMS-12BM, and MM.1S) cultured under hypoxia (1% O2) with those cultured under normoxia (20% O2), respectively. GSE80545 profiled the GE in four primary myeloma patient samples exposed to normoxia versus those exposed to normoxia for 48 h. R package ‘limma’ was utilized to screen out DEGs between normoxic and hypoxic groups [15]. Genes satisfied with p-value < 0.05 and |log2FoldChange| > 1 were picked out as DEGs. Then, all DEGs were visualized by volcano maps, and the top 50 dramatically altered genes were selected to draw heatmaps by R package ‘ggplot2’ [16]. Wayne diagram produced by the Venny 2.1.0 (http://bioinfogp.cnb.csic.es/tools/venny/index.html) online tool was used to show the common genes.

2.2. Gene ontology (GO) and Kyoto encyclopedia of genes and genomes (KEGG) analysis

DAVID (https://david-d.ncifcrf.gov/) was used to conduct GO enrichment analysis and KEGG pathway analysis. GO enrichment includes a biological process (BP), molecular function (MF), and cellular component (CC) as three subontologies. Analysis results were displayed with the ‘GOplot’ package of R [17]. Pathway analysis results were further analyzed with ClueGO plug-ins of Cytoscape software 3.7.2 [18]. A p-value less than 0.05 was considered the cut-off criterion.

2.3. Protein–protein interaction (PPI) analysis

A PPI network was established using the STRING database (https://string-db.org) in order to evaluate the proteins’ interrelation encoded by selected genes. Cytoscape software 3.7.2 was used to visualize the genes with a minimum interaction score of more than 0.4.

2.4. Survival analysis

The correlation between DEG expression and survival rate in MM was analyzed with the PrognoScan database (http://dna00.bio.kyutech.ac.jp/PrognoScan/) [19]. We researched the relationship between GE and MM disease-specific survival (DSS) based on the GSE2658 dataset (n = 559) provided by Zhan [20]. Cox p-value < 0.05 was treated as statistically significant.

3. Results

3.1. Identification of DEGs in MM

According to the standard of p-value < 0.05 and |log2FC| > 1, 760 and 1499 DEGs were identified using ‘limma’ R package from GSE80140 and GSE80545 (Figure 1(a)), respectively. The top 50 significant DEGs were used to depicted the heatmaps (Figure 1(b)). We next utilized Wayne diagram and identified 119 up-regulated genes and 67 down-regulated genes as common to the two datasets, respectively (Figure 1(c)).

3.2. GO and KEGG enrichment analysis

We performed GO and KEGG analysis with candidate DEGs using a functional annotation tools in DAVID in order to further explore their functions. As shown in Figure 2(a), the DEGs were examined according to three subontologies, including BP, MF, and CC, in GO enrichment analysis. Histone H3-K9 dimethylation pathway (p = 1.78E-04, FDR = 0.071474), catalytic activity pathway (p = 6.95E-04, FDR = 0.09272), and nucleoplasm pathway (p = 5.47E-06, FDR = 9.74E-04) were selected as the most significant pathway in each subontology, respectively (Table 1). Forty-eight candidate DEGs were enriched in these three terms for further analysis. KEGG pathway enrichment analysis result showed that metabolic pathways (p = 1.73E-04, FDR = 0.016285) were the top enriched pathway according to their p-values, and there were 25 DEGs involved in this pathway (Figure 2(b)).

As a result, 48 and 25 DEGs were selected in the GO and KEGG enrichment analysis, respectively (Table 2). Then, we screened out eight common genes with overlapping DEGs selected in the two analyses above. They were TSTA3, FH, FLAD1, PNPO, CAD, POLR3G, SCLY, and POLR3K.

3.3. PPI network construction

In order to further understand the correlation of candidate DEGs and pick out the hub genes, we constructed the PPI network involving all the DEGs from the GO and KEGG enrichment analysis based on the STRING database (https://string-db.org). We used the STRING data-base to evaluate the proteins’ interaction encoded by selected genes. Cytoscape software 3.7.2 was used to visualize the genes with a minimum interaction score of more than 0.4.

3.4. Survival analysis

The correlation between DEG expression and survival rate in MM was analyzed with the PrognoScan database (http://dna00.bio.kyutech.ac.jp/PrognoScan/) [19]. We researched the relationship between GE and MM disease-specific survival (DSS) based on the GSE2658 dataset (n = 559) provided by Zhan [20]. Cox p-value < 0.05 was treated as statistically significant.

3. Results

3.1. Identification of DEGs in MM

According to the standard of p-value < 0.05 and |log2FC| > 1, 760 and 1499 DEGs were identified using ‘limma’ R package from GSE80140 and GSE80545 (Figure 1(a)), respectively. The top 50 significant DEGs were used to depicted the heatmaps (Figure 1(b)). We next utilized Wayne diagram and identified 119 up-regulated genes and 67 down-regulated genes as common to the two datasets, respectively (Figure 1(c)).

3.2. GO and KEGG enrichment analysis

We performed GO and KEGG analysis with candidate DEGs using a functional annotation tools in DAVID in order to further explore their functions. As shown in Figure 2(a), the DEGs were examined according to three subontologies, including BP, MF, and CC, in GO enrichment analysis. Histone H3-K9 dimethylation pathway (p = 1.78E-04, FDR = 0.071474), catalytic activity pathway (p = 6.95E-04, FDR = 0.09272), and nucleoplasm pathway (p = 5.47E-06, FDR = 9.74E-04) were selected as the most significant pathway in each subontology, respectively (Table 1). Forty-eight candidate DEGs were enriched in these three terms for further analysis. KEGG pathway enrichment analysis result showed that metabolic pathways (p = 1.73E-04, FDR = 0.016285) were the top enriched pathway according to their p-values, and there were 25 DEGs involved in this pathway (Figure 2(b)).

As a result, 48 and 25 DEGs were selected in the GO and KEGG enrichment analysis, respectively (Table 2). Then, we screened out eight common genes with overlapping DEGs selected in the two analyses above. They were TSTA3, FH, FLAD1, PNPO, CAD, POLR3G, SCLY, and POLR3K.

3.3. PPI network construction

In order to further understand the correlation of candidate DEGs and pick out the hub genes, we constructed the PPI network involving all the DEGs from the GO and KEGG enrichment analysis based on the STRING database (https://string-db.org) to evaluate the proteins’ interaction encoded by selected genes. Cytoscape software 3.7.2 was used to visualize the genes with a minimum interaction score of more than 0.4.
Figure 1. Identification of differentially expressed genes in two GE profiles (GSE80140 and GSE80545). (a) Respective volcanic maps of the two datasets. Up (red plots), the expression is up-regulated; down (green plots), the expression is down-regulated; no (gray plots), no significantly dysregulated. (b) Heatmaps of top 50 dramatically altered genes of DEGs in GSE80140 and GSE80545. (C) Commonly changed DEGs in the two datasets (119 up – and 67 down-regulated genes). DEGs, differentially expressed genes.
Then, we used Cytoscape plug-ins cytoHubba to screen all the candidate hub genes according to nodes rank (Figure 3(c,d)). All the candidate genes were listed in Table 3. Then, we identified FH, POLR3K, TSTA3, POLR3G, and PNPO as common candidate hub genes in the two sets.

Finally, we assessed the potential prognostic value of candidate hub genes for MM utilizing PrognoScan [19]. Based on the GSE2658 dataset ($n=559$), three candidate hub DEGs were acquired to be statistically significant with DSS in MM, including FH (Cox $p=0.000127$), TSTA3 (Cox $p=0.035654$), and POLR3G (Cox $p=0.018991$) (Figure 4).

### Table 1. Top 5 enriched terms of GO analysis.

| Term          | Description                                      | Count | p-value     |
|---------------|--------------------------------------------------|-------|-------------|
| BP            | GO:0036123 Histone H3-K9 dimethylation           | 3     | 1.78E-04    |
|               | GO:0006418 tRNA aminoacylation for protein translation | 5     | 2.35E-04    |
|               | GO:0043276 Anoikis                                | 3     | 3.53E-04    |
|               | GO:0006260 DNA replication                        | 7     | 0.001308    |
|               | GO:0006520 Cellular amino acid metabolic process  | 4     | 0.003634    |
| MF            | GO:0003824 Catalytic activity                     | 21    | 6.95E-04    |
|               | GO:0044822 Poly(A) RNA binding                   | 8     | 4.81E-04    |
|               | GO:0005224 ATP binding                            | 20    | 0.023579    |
|               | GO:0005315 Protein binding                        | 81    | 0.028966    |
|               | GO:0003873 Phosphofructo-2-kinase activity        | 2     | 0.030916    |
| CC            | GO:0005654 Nucleoplasm                           | 42    | 5.47E-06    |
|               | GO:0005739 Mitochondrion                         | 26    | 1.20E-05    |
|               | GO:0005743 Mitochondrial inner membrane           | 12    | 4.19E-04    |
|               | GO:0005829 Cytosol                               | 40    | 0.001316    |
|               | GO:0005730 Nucleolus                             | 16    | 0.001598    |

### Table 2. DEGs identified from GO and KEGG analysis.

| DEGs Gene names          | GO enrichment analysis |
|--------------------------|-----------------------|
| KDM3A, SUV39H2, SUV39H1, PFKFB3, NTHL1, MRT04, CSTF2, DDX39A, PAK11P1, A5F1B, FKBP4, UBE2T, HNRNPB, POP1, KDM48, ORC6, UTP20, FEN1, CDG6, MRPS26, GAR1, TIPIN, MAFF, PNPO, RNASEH2A, RTEL1, BANF1, CAD, PSMG1, RRP9, E2F1, MPHOSPH6, KPN2AP2, GMNN, METTL1, POLR3G, SNRP25, SNRPD1, WDR77, POLR3 K, TIMM50, CDC25A, FH, FLAD1, PFKFB4, ISOC2, TSTA3, SCLY |  
| TSTA3, FH, GAL3ST1, PYCRL, ATP8, ATP5G1, HSD17B10, HK2, SRF, GC5H, IL4I1, FLAD1, PNPO, SEP55, GG76, CAD, G0T2, CO03, P4HA1, MTHFD1, NDUF56, POLR3G, SCLY, POLR3K, DCTP19 |  

**Figure 2.** GO and KEGG enrichment analysis of DEGs. (a) GO analysis including BP, MF, and CC three subontologies. (b) KEGG pathway enrichment analysis. GO, gene ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; DEGs, differentially expressed genes; BP, biological process; MF, molecular function; CC, cellular component.
Figure 3. PPI network analysis. (a, b) The PPI analysis at STRING. (c, d) Cytoscape plug-ins cytoHubba analysis of candidate genes after PPI analysis. (a, c) Genes identified from GO analysis. (c, d) Genes identified from KEGG analysis. PPI, protein–protein interaction.

Table 3. The genes with the score of each analysis through the Cytoscape ‘cytoHubba’ module analysis

| Rank | Name  | GO analysis Score | Rank | Name  | KEGG analysis Score |
|------|-------|------------------|------|-------|---------------------|
| 1    | FEN1  | 112              | 23   | KDM4B | 4                   |
| 2    | CDC6  | 108              | 24   | UTP20 | 4                   |
| 3    | GMNN  | 72               | 25   | SUV3H1 | 4               |
| 4    | TIPIN | 54               | 26   | NTHL1 | 4                   |
| 5    | CDC25A| 36               | 27   | SNRP25 | 3                |
| 6    | ASF1B | 32               | 28   | FH    | 3                   |
| 7    | UBE2T | 30               | 29   | PFKFB3 | 2                |
| 8    | ORC6  | 30               | 30   | RTEL1 | 2                   |
| 9    | KPNA2 | 27               | 31   | POLR3K | 2                |
| 10   | RRP9  | 22               | 32   | POP1  | 2                   |
| 11   | SNRPD1| 17               | 33   | SUV3H2 | 2                |
| 12   | RNA5EH2A | 17 | 34 | MHSOPH3H6 | 2 |
| 13   | HNRNPA8 | 14 | 35 | TSTA3 | 1 |
| 14   | DDX39A | 14 | 36 | PFKFB4 | 1 |
| 15   | MRT04 | 14               | 37   | POLR3G | 1                |
| 16   | CSTF2 | 13               | 38   | ISOC2 | 1                   |
| 17   | GAR1  | 10               | 39   | PNPO  | 1                   |
| 18   | E2F1  | 8                | 18   | SEPHS1 | 1               |
| 19   | METTL1 | 7                | 19   | SLCY | 1                   |
| 20   | WDR77T | 6                | 20   | COQ3  | 1                   |
| 21   | PAK11P1 | 6             | 21   | MT-ATP8 | 1          |
| 22   | KDM3A | 5                |      |       |                     |
4. Discussion

Due to the complex unknown pathogenesis of MM, it remains an incurable hematologic malignancy which leads to high rates of mortality. The hypoxic microenvironment in BM is closely related to the occurrence, development, treatment, and prognosis of MM. Therefore, it is essential to explore hypoxia-related hub genes with prognostic values to provide potential biomarkers for diagnosis and follow-up treatment of MM.

In this study, we analyzed the GE in two microarray datasets based on myeloma cell lines and primary MM patient samples. About 119 up-regulated and 67 down-regulated common DEGs were identified with \( p < 0.05 \) and \( \log_{2}\text{FC} > 1 \) standard. Then, we utilized GO term enrichment, signaling pathway enrichment, and PPI network establishment to further explore these DEGs, and FH, POLR3K, TSTA3, POLR3G, and PNPO were identified as candidate hub genes. Subsequently, survival analysis with PrognoScan based on clinical data of GSE2658 dataset revealed FH, TSTA3, and POLR3G were the hub genes associated with DSS of MM patients.

Fumarate hydratase (FH) is an integral enzyme component of the Krebs cycle and tricarboxylic acid cycle, and it can also act in different kinds of cancer [21]. Ha YS et al. [22] evaluated FH mRNA levels in 140 tumor specimens from the patients with primary renal cell cancer and in 62 specimens of corresponding normal-appearing kidney tissue. The results showed that FH was down-regulated in tumor specimens. Sudarshan S [23], with his team, demonstrated that reduced expression of FH was associated with HIF-2α accumulation and tumor progression. However, another study found that FH inhibition, together with elevated intracellular fumarate, coincides with HIF upregulation. Besides, fumarate could increase the expression level of HIF-regulated genes [24]. TSTA3, also known as GFUS (GDP-L-fucose synthase), is the substrate of several fucosyltransferases involved in the expression of many glycoconjugates [25]. Changes in glycoprotein structure were common in cancer cells that can affect the adhesion, invasion, and metastasis of cells. Yang J et al. [25] found high expression of TSTA3 was a potential biomarker for poor prognosis of esophageal squamous cell carcinoma patients. The same result was found in breast cancer, and the study further revealed that miR-125a-5p and miR-125b could regulate the expression of TSTA3 [26]. POLR3G (RNA polymerase III subunit G), a version of RNA polymerase III, which is enriched in stem and cancer cells [27]. Recently, Petrie JL et al. [28] demonstrated the inhibition effect of POLR3G on cell differentiation and down-regulated POLR3G significantly reduced in the proliferation, invasiveness, tumor-initiating activity, and viability of prostate cancer cells.

5. Conclusion

In summary, we identified DEGs by integrated bioinformatics analysis to find hypoxia-related hub genes with prognostic value in MM. The study results provide several crucial DEGs for further research on the molecular pathogenic mechanisms and biomarkers of MM so that patients can get target therapy and a longer survival period in the future. It’s worth noting that different research results mentioned above showed that the same gene might express at different levels in different tumors, especially in hematologic malignancies. More molecular biological researches are still required to verify the DEGs’ function in MM.

Disclosure statement

The authors declare that there is no conflict of interests.
Funding
This study was supported by Natural Science Foundation of Jiangsu Province [Grant Number BK20180372], Jiangsu Provincial Medical Youth Talent [Grant Number QNRC2016812], and Key Medical Subjects of Jiangsu Province [Grant Number ZDXKB2016020].

References
[1] Fitzmaurice C, Dicker D, Pain A, et al. The global burden of cancer 2013. JAMA Oncol. 2015;14(14):505–527.
[2] Lu J, Lee JH, Huang SY, et al. Continuous treatment with lenalidomide and low-dose dexamethasone in transplant-ineligible patients with newly diagnosed multiple myeloma in Asia: subanalysis of the FIRST trial. Br J Haematol. 2017;176(5):743–749.
[3] Pawlyn C, Morgan GJ. Evolutionary biology of high-risk multiple myeloma. Nat Rev Cancer. 2017;17(9):543–556.
[4] Manier S, Salem KZ, Park J, et al. Genomic complexity of multiple myeloma and its clinical implications. Nat Rev Clin Oncol 2017;14(2):100-113.
[5] Sonneveld P, Avet-Loiseau H, Lonial S, et al. Treatment of multiple myeloma with high-risk cytogenetics: a consensus of the international myeloma working group. Blood. 2016;127(24):2955–2962.
[6] Zhu FX, He YC, Zhang JY, et al. Using prognosis-related gene expression signature and connectivity Map for personalized drug repositioning in multiple myeloma. Med Sci Monit. 2019;25:3247–3255.
[7] Viziteu E, Grandmougin C, Goldschmidt H, et al. Chetomin, targeting HIF-1alpha directly targeting SOX6. Nucleic Acids Res. 2014;24(1):37–51.
[8] Jin Z, Zhou S, Ye H, et al. The mechanism of SP1/p300 complex promotes proliferation of multiple myeloma cells through regulating IQGAP1 transcription. Biomed Pharmacother. 2019;119:109434.