

Genome-Wide Identification of Potential Biomarkers in Multiple Myeloma Using Meta-Analysis of mRNA and miRNA Expression Data

Amit Katiyar

Indian Council of Medical Research

Gurvinder Kaur

All India Institute of Medical Sciences

Lata Rani

All India Institute of Medical Sciences

Harpreet Singh

Indian Council of Medical Research

Punit Kaur

All India Institute of Medical Sciences

Ritu Gupta (✉ dritugupta@gmail.com)

All India Institute of Medical Sciences

Research Article

Keywords: gene expression omnibus (GEO), differentially expressed genes (DEGs), differentially expressed miRNAs (DEMs), biomarker, miRNA-mRNA network, gene-TF regulatory network, miRNA-disease association

DOI: <https://doi.org/10.21203/rs.3.rs-132765/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Multiple myeloma (MM) is a plasma cell malignancy with diverse clinical phenotypes and molecular heterogeneity that is not completely understood. Recent studies have identified differentially expressed patterns of genes (DEGs) or miRNAs (DEMs) in MM. But these signatures overlap partially, plausibly due to complexity of myeloma genome, diversity in cell lines studied, molecular technologies and analytical tools utilized in these studies. This warrants further investigations since DEGs/DEMs can impact clinical outcomes and guide personalized therapy. We conducted the genome-wide meta-analysis of expression datasets on DEGs/DEMs in MM and derive net putative signatures and potential biomarkers for MM. A set of 110 DEMs and 3,817 DEGs were identified to be differentially expressed. Among these, 86 DEMs (60 downregulated; 26 upregulated) correlated with 1,970 target DEGs (1373 downregulated; 597 upregulated). Signatures of 23 DEMs ('Union 23') and 196 DEGs ('Union 196') were deduced that shared 10 DEMs and 13 DEGs with published signatures, respectively. The study has identified five topmost nodal genes (*APP*, *KIAA0101*, *CDK2*, *ESR1* and *FN1*) derived from functional modules in PPI networks and has paved the way for further studies to establish their prognostic potential and role in therapeutics for MM. The integrated bioinformatics methods and expression profiling techniques may lead to the identification of putative hub genes and expression signatures that can serve as predictive biomarkers of MM progression.

Introduction

Multiple myeloma (MM) is a neoplasm of plasma cells with heterogeneous clinical symptoms, complex cytogenetic aberrations, multiple copy number variations, mutations, alternate splicing events and epigenetic modifications. Nearly 30% of the MM patients have hyperdiploid karyotypes while most of the remaining is nonhyperdiploid with recurrent translocations between IgH locus and multiple partner genes (such as cyclin D family, cMAF and MMSET). At least 8 to 10 molecular subgroups of MM can be risk stratified based on the genomic and transcriptomic patterns that tend to correlate with different clinical outcomes¹.

Alterations in expression profiles of genes and small non-coding RNAs, especially, the miRNA, are frequently encountered in MM. Expression profiles of differentially expressed genes (DEGs) are of paramount importance and have provided critical prognostic insights in MM. Computational and functional analysis of hub genes, nodes, networks and pathways in MM have led to the development of risk scoring systems, relating to the 8 genetic subgroups², 70 genes UAMS70 risk signatures³, IFM15 risk stratification⁴, 5 gene stemness score⁵, UAMS 17³, CINGLEC 214⁶, EMC 92⁷, HZD 97⁸, Millenium 100⁹, M3CN¹⁰ and others.

Similarly, global miRNA expression studies¹¹⁻¹⁴ have elucidated a multitude of DEMs in MM. Some of the DEMs occur in association with specific cytogenetic subgroups of MM¹⁵. For example, 1q gain has been correlated with overexpression of miR-1231, 205, 215, 488; 19q gain with upregulation of miR-520a-5p, miR-518d-5p, miR-498, miR-520g; del13q with downregulated miR15a/16 cluster, miR-17-92 family

(miR-17, miR-19a, miR-20a); and 17pdel with reduced expression of miR-22. Similarly, deregulation of miR-133b, miR-135b, miR-155, miR-193a, miR-203, miR-146a, miR-215, miR-342, miR-375, miR-650 have been correlated with t(4;14), miR-95, miR-125a, miR-184, miR-199a, miR-215, miR-375, miR-650, miR-99 with t(11;14), and miR-1, miR-99b, miR-125a, miR-133a, miR-135b, miR-196b, miR-214, miR-375, miR-642 with t(14;16)^{16–20}. Recent studies have established the prognostic, predictive and diagnostic potential of not only cellular but also circulating miRNAs in plasma and other body fluids in MM²¹.

Although a series of MM associated DEM/DEG signatures have been identified over the years, these remain mostly heterogeneous and challenging to interpret. There are still lacunae in our knowledge and more studies are needed to understand the signatures that are best valued in clinics for early prognostication of disease progression or relapse, prediction of response to therapy and personalization of treatment for MM. In this study, we have performed a meta-analysis of available multiple datasets on DEGs and DEMs in MM patients obtained through genome wide microarray profiling approach to derive putative signatures and potential biomarkers for MM.

Methods

This study was conducted in compliance with ethical guidelines of the All India Institute of Medical Sciences (AIIMS), New Delhi after obtaining approval from the AIIMS ethics committee. Study individuals were enrolled following their voluntary written informed consent.

Inclusion and *exclusion* criteria for published datasets

The keyword “multiple myeloma” with “homo sapiens” was used to mine the publicly available datasets from the Gene Expression Omnibus (GEO) database^{22,23} of NCBI (<http://www.ncbi.nlm.nih.gov/geo/>) for miRNA and mRNA expression profiles found in Multiple Myeloma (MM) patients and healthy controls. Datasets obtained from Monoclonal Gammopathy of Undetermined Significance (MGUS) and Plasma Cell Leukemia (PCL) patients were excluded since the data size was limited. Data emerging from cells or cell lines that were cultured *in vitro* and/or treated with drugs too were omitted in this study.

miRNA and mRNA expression datasets

Expression profiles from nine publicly available datasets of mRNAs (GEO accession: GSE16558, GSE13591, GSE36474 and GSE5900), and miRNAs (GEO accession: GSE16558, GSE17306, GSE17498, GSE24371, GSE40308 and GSE54156) associated with MM were retrieved from publicly available GEO repository (Table S1 in Supplementary File 1) as per the inclusion and exclusion criteria. In addition, two datasets of miRNA and mRNA expression profiles (44 newly diagnosed MM patients and 4 controls) were generated in-house by Agilent array. These miRNA and mRNA datasets have been submitted to GEO and assigned with accession numbers GSE125363 and GSE125361, respectively. In this study, miRNA expression profiles corresponding to 291 MM and 31 healthy control plasma cell samples in total, whereas mRNA expression profiles representing 262 MM and 39 healthy plasma cell samples in total were collated and analyzed (Table S1 in Supplementary File 1). Datasets obtained from GEO repository

were not subjected to any additional normalization, as all the data obtained had already been processed/normalized and were cross-comparable.

Preprocessing and mining of DEMs/DEGs from GEO repository

GEO2R²³ (<http://www.ncbi.nlm.nih.gov/geo/geo2r/>) web tool was used to identify DEGs and DEMs among MM and control plasma cell samples. GEO2R compared two or more groups of samples in a GEO profile using the GEOquery and Limma (Linear Models for Microarray Analysis) R package²⁴. The t-test and the Benjamini and Hochberg method were used to calculate the p-values and false discovery rate (FDR), respectively²⁵. The adjusted (adj.) $p \leq 0.05$ and $|\logFC| \geq 1.5$ were set as the cut-off criterion for identifying DEGs and DEMs.

Genome-wide miRNA and mRNA expression profiling

Total RNA was isolated from CD138+ plasma cells enriched with MACS beads (Miltenyi Biotech, Germany), collected from 44 newly diagnosed treatment naïve MM patients diagnosed as per IMWG guidelines²⁶ (Table S2 in Supplementary File 1) and 4 controls (pooled from 10 Hodgkin's disease bone marrow samples). Total RNA was extracted using the miRVana miRNA isolation kit (Thermofisher Scientific, MA, USA).

For the genome wide miRNA expression profiling, RNA was labeled and hybridized to an unrestricted human microRNA v19 Microarray slide (Agilent 046064, GPL18044) (Agilent Technologies, Santa Clara, CA, USA) according to the manufacturer's protocol. Briefly, 100 ng of total RNA was labeled with Cyanine3 (Cy3) using miRNA Complete Labeling and Hybridization Kit (Agilent Technologies, Santa Clara, CA, USA). The Cy3-labeled samples were resuspended in hybridization buffer and hybridized onto Human miRNA 8X60K format microarrays (Agilent Technologies, Santa Clara, CA, USA) at 55°C for 20h. After hybridization, microarrays were washed with gene expression wash buffer and the fluorescent signals were scanned using SureScan microarray scanner D (Agilent Technologies, Santa Clara, CA, USA) using one colour scan settings (Scan resolution 3 μm , Dye channel set to Green, Green PMT = 100%). The data generated on miRNA expression in MM using microarrays has been submitted to GEO database with accession no GSE125363.

To correlate whether the miRNA alteration affects gene expression, mRNA expression array analysis was also performed on 44 MM patient samples and 4 controls (pooled from 10 Hodgkin's disease bone marrow samples). Double-stranded cDNA was generated from 200ng total RNA (isolated with miRVana kit) using the low input quick amp labelling kit (Agilent Technologies, Agilent Technologies, Santa Clara, CA, USA) using T7 primer, dNTPs and affinity script RNase block. Next, cDNA was transcribed to cRNA using T7 RNA polymerase and NTP mix and labeled with Cyanine3 using Cy3-CTP. The labeled cRNA was purified according to manufacturer's protocol using RNeasy extraction kit (Qiagen, Hilden, Germany). The concentration of Cyanine3 and cRNA was measured using NanoDrop ND1000 spectrophotometer. Samples with specific activity $\geq 6\text{pmol Cy3}/\mu\text{g cRNA}$ were hybridized onto a SurePrint G3 human GE v3 8x60K microarray slide (Agilent 072363, GPL20844) (Agilent technologies, Santa Clara, CA, USA), and

incubated for 17 hours at 65°C in a hybridization oven. The slides were washed and scanned in SureScan microarray scanner D (Agilent technologies, Santa Clara, CA, USA) with scan settings (Scan resolution 3 µm, Dye channel set to Green, Green PMT = 100%). The data generated on mRNA expression in MM has been submitted to GEO database with accession no. GSE125361.

Preprocessing and mining of DEMs/DEGs from Agilent array

Microarray images (*.tiff) obtained from SureScan scanner were quantified using Agilent Feature Extraction Software (version 11.5.1.1) (Agilent Technologies, Santa Clara, CA, USA). The raw data obtained (tab-delimited text file per hybridisation) was subsequently processed *with* the *Limma R package available in the Bioconductor repository* (<http://www.bioconductor.org>). The raw intensity data were background corrected using normexp method and subsequently normalized using quantile method for one-color. Expression level variations between replicates were analyzed by pairwise comparisons using the lmFit function. The fitted model object was further processed by the eBayes function to produce empirical Bayes test statistics for each gene, including moderated t-statistics, p-values and log-odds of differential expression. The t test and Benjamini and Hochberg method were used to calculate the p-values and false discovery rate (FDR)²⁵. The adjusted $p \leq 0.05$ and $|\logFC| \geq 1.5$ were set as the cut-off criterion for identifying DEGs and DEMs.

Meta-analysis of DEGs/DEMs datasets

A widely used meta-analysis approach^{27–29} was applied to integrate the gene/miRNA expression profiles obtained independently from GEO repository as well as datasets generated at our centre following microarray hybridization (Table S1 in Supplementary File 1). The gene and miRNA probes were assigned as per HGNC (HUGO Gene Nomenclature Committee) and miRBase-22 identifiers, respectively using the g:Profiler³⁰ (<https://biit.cs.ut.ee/gprofiler/>). The differentially expressed genes/miRNAs obtained through R/Bioconductor limma package²⁴ from these individual studies were merged by taking the union across them. When multiple probes referred to the same gene/miRNA, the expression values obtained from these probes were minimized to a single value by averaging the expression value (when in the same direction of expression) or were discarded (when had diverse directions of expression). The probes with unknown gene or unknown miRNA identifiers or annotated as antisense RNA, chromosomes, hypothetical loci, non-coding RNAs, non-functional proteins, non-protein coding genes, pseudo-genes and uncharacterized genes were discarded. The DEGs identified were mapped in DisGeNET³¹ to determine their known disease associations.

Core analysis using IPA

Ingenuity Pathway Analysis (IPA, Ingenuity Systems, USA; www.qiagen.com/ingenuity) was used to identify the biological functions, diseases, canonical pathways, and regulatory networks of the functional miRNA-mRNA target interactions. Tab-delimited text files containing gene/miRNA IDs, expression data (fold change), and p-values were uploaded into IPA for their core analysis. The statistical significance of

the enrichment was calculated using hypergeometric test and adjusted by FDR method (adj. p-value \leq 0.05). The putative targets of DEMs were identified by microRNA Target Filter, according to at least one out of four different databases (TargetScan, miRecords, Tarbase, or Ingenuity expert findings). The miRNA-mRNA pairs with inverse correlation expression trends were filtered and score ranks for the top functions (molecular, cellular and biological), diseases, toxicology, and gene signaling networks were calculated using IPA-generated negative logarithm p-values i.e., $-\log_{10}(\text{p-value})$ and associated Z- and network scores.

Construction of protein-protein interaction (PPI) network

To examine the interactive associations among the DEGs at the protein level, MM related genes were mapped on protein-protein interaction (PPI) data using NetworkAnalyst³² (version 3.0; <http://www.networkanalyst.ca>). The network was built based on the original seed proteins by executing the zero order interactions to evade “hairball effect.” Literature-curated comprehensive PPI data was used to predict interaction network³³. Network modules containing densely connected group of proteins were predicted using the random walk approach. The significant p-value of a given module was calculated with Wilcoxon rank-sum test³⁴. The enriched pathways of DEGs in significant modules (≥ 10 DEGs) were analysed with a threshold of $p \leq 0.05$ using DAVID (database for annotation, visualization and integrated discovery) functional annotation tool.

Biomarker candidates and TF regulatory network

The candidate gene biomarkers were predicted using Ingenuity Pathway Analysis (IPA-biomarkers analysis; <http://www.ingenuity.com>). The adjusted (adj.) $p \leq 0.05$ and $|\log_{2}FC| \geq 1.5$ were set as the cut-off criterion. Upstream regulators (TFs) of biomarker candidates were predicted using NetworkAnalyst (<http://www.networkanalyst.ca>). TF-gene interaction analysis was performed using the ENCODE database. The miRNA-disease association was predicted by HMDD v3.2 (the Human microRNA Disease Database; <http://www.cuilab.cn/hmdd>) and miRNet 2.0 (<https://www.mirnet.ca/>). The regulatory associations between TFs and miRNAs were predicted using TransmiR v2.0 (<http://www.cuilab.cn/transmir>). Functional and pathway enrichment analyses of upstream regulators were investigated using DAVID (<https://david.ncifcrf.gov/>) with adjusted (adj.) $p \leq 0.05$ and $|\log_{2}FC| \geq 1.5$ cut-off criterions. The enrichment results were visualized with bubble plots by the R package ggplot2 (<http://ggplot2.org>).

Results

Identification of DEGs/DEMs

Following consolidation of data and its meta-analysis, a total of 3817 DEGs and 110 DEMs with FDR \leq 0.05 and $\log_{2}FC \geq 1.5$ threshold values were identified in MM patients as compared to control plasma cells (*Supplementary Tables S3-S4*). Among these DEMs, 34 miRNAs were upregulated, and 76 miRNAs were downregulated, whereas 859 DEGs were upregulated and 2958 genes were downregulated in MM

(Table S3-S4 in Supplementary File 1). The DEGs identified from this study were mapped across the available data of MM present in DisGeNET which established their known associations with MM (*Table S5 in Supplementary File 1*).

MicroRNA–mRNA regulatory network analysis

To verify the targets of differentially expressed miRNAs in MM datasets, miRNA–mRNA regulatory network was constructed using the 110 DEMs and 3817 DEGs. Network analysis showed an association of 86 DEMs, including 26 upregulated and 60 downregulated miRNAs, with 1970 target genes, including 1373 downregulated and 597 upregulated DEGs in MM (Figure 1). The observed targets were statistically significant with $p\text{-value} \leq 0.05$ and fold change ≥ 1.5 . The most significantly upregulated and downregulated miRNAs in MM were observed to be hsa-miR-16-5p (5.39 logFC) and hsa-miR-575 (-6.85 logFC), respectively, whereas the most significantly upregulated and downregulated DEGs were KIT (5.02 logFC) and IGLL1/IGLL5 (-4.64 logFC), respectively (Table 1; Table S6 and S7 in Supplementary File 1). Hsa-miR-650 (3.10 logFC) was predicted to target maximum number of genes (242 downregulated DEGs), whereas hsa-miR-29c-5p (-1.80 logFC) was predicted to target the least number of genes (6 upregulated DEGs) (Table S8 in Supplementary File 1). Conversely, several genes were predicted to be the common targets of different miRNAs (Table S9 in Supplementary File 1).

Correlation between canonical pathways, diseases and functions

To gain further insights into the pathogenesis of MM, all significant MM-correlated genes/miRNAs were investigated by IPA core analysis that revealed 534 human canonical pathways significantly enriched for 549 overlapping genes associated with MM (*Table S10 in Supplementary File 1*). The top five significant enriched pathways based on their significance (*lowest BH-adjusted p-value ≤ 0.05*) were hepatic fibrosis / hepatic stellate cell activation ($p=2.20E-09$), antigen presentation pathway ($p=1.48E-05$), cardiac-adrenergic signaling ($p=2.49E-05$), atherosclerosis signaling ($p=5.77E-05$) and osteoarthritis pathway ($p=8.66E-05$) (Figure 2).

DEGs and DEMs were further determined to outline the most enriched diseases and functions in multiple myeloma. On annotation, most of the DEGs were found to be involved in cancer, organismal injury and abnormalities, gastrointestinal disease, endocrine system disorders and dermatological diseases and conditions (*Figure S1a in Supplementary File 2*), whereas DEMs were found to be enriched in cancer, organismal injury and abnormalities, reproductive system disease, inflammatory disease, and inflammatory response (*Figure S1b in Supplementary File 2*). The topmost significant diseases and biofunctions identified for DEMs and DEGs are shown in Table 2. Besides the leading pathways and cellular functions, gene networks were constructed to connect key genes and enriched categories of diseases and functions based on the correlation between DEGs. Core analysis-based network revealed 25 significant networks and each individual *network had a maximum of 35 focus genes*. *Top ranked network (network 1) with an IPA score of 42 contained 35 focus molecules*. The downregulated amyloid precursor protein (APP) in network 1 plays a central role (hub gene) in the formation of this complex network (Figure S2 in Supplementary File 2). Top functions of these genes were mainly connected to cancer,

dermatological diseases and conditions, organismal injury and abnormalities. Likewise, miR-network consisted of 16 major networks with *a maximum of 26 focus genes*. Most of the genes in miRNA network 1 were again mainly connected to cancer, organismal injury and abnormalities and reproductive system disease. Gene/miRNA networks and their related top diseases and functions are listed in Table 3 and Table S11 in Supplementary File 1.

Identification of functional modules in PPI network

Protein–protein interactions (PPI) network was constructed using aberrantly expressed genes identified in MM to predict biologically significant modules containing a group of proteins that execute similar functions. The zero-order interaction network scattered in 1–3 sub-networks including one big network with highest nodes and edges. The network analysis disclosed 933 nodes (47.36% of DEGs) associated with 2004 edges in the network. The PPI degree of the genes revealed 37.73% (n=352) of the nodes with a degree of one, whereas 62.27% (n=581) of nodes with a degree greater than one. Out of 581 nodes, a total of 76 nodes were observed with ≥ 10 degrees or connections with other nodes (Table S12 in Supplementary File 1). Among, these 76 nodes, 52 nodes were observed to be targeted by more than one MM-associated DEMs (S12 in Supplementary File 1). The betweenness centrality of nodes were observed for 55.84% (n=5214) of nodes with a range of 1.47 to 161738.69 in the constructed network.

The modules containing a group of proteins with identical functions were detected using the random walk approach. A total of 33 independent functional modules were observed, whereas five modules (module no 1, 2, 5, 6 and 8) were highly connected with more than 10 nodes and $p \leq 0.05$ (Table 4; Table S13 in Supplementary File 1). The predicted modules were connected to neighbouring modules and ranged in size from 5 to 182 genes. The distribution of highly connected hub nodes (degree ≥ 10) in module 1 encompassed 15 genes namely KIAA0101 (degree 48; betweenness centrality 3438.67), CDK2 (degree 44; betweenness centrality 4142.06), ESR1 (degree 35; betweenness centrality 2851.19), FN1 (degree 34; betweenness centrality 2151.94), SUMO1 (degree 32; betweenness centrality 2794.83), MCC (degree 20; betweenness centrality 1187.16), ARRB2 (degree 19; betweenness centrality 846.69), NCL (degree 15; betweenness centrality 785.87), GNB2L1 (degree 12; betweenness centrality 843.36), SF3B3 (degree 12; betweenness centrality 728.39), LMNA (degree 11; betweenness centrality 468.3), RPL12 (degree 11; betweenness centrality 348.79), MCM7 (degree 10; betweenness centrality 890.36), MYL12A (degree 10; betweenness centrality 325.59), and RPL24 (degree 10; betweenness centrality 334.01) (Figure 3a). The network of 10 hub nodes including APP (degree 103; betweenness centrality 10586.07), SRC (degree 16; betweenness centrality 1307.26), STAT1 (degree 14; betweenness centrality 882.44), JUN (degree 13; betweenness centrality 949.37), PPARG (degree 13; betweenness centrality 760.39), RB1 (degree 12; betweenness centrality 522.03), KAT5 (degree 11; betweenness centrality 1265.96), CDKN1A (degree 11; betweenness centrality 911.75), LRIF1 (degree 10; betweenness centrality 531.37), and CCND1 (degree 10; betweenness centrality 179.47) was observed in module 2 (Figure 3b). Likewise, modules 5, 6 and 8 contained only 1 significant hub node i.e. EGR1 (degree 29; betweenness centrality 518), ONECUT1 (degree 14; betweenness centrality 103), and HNF1A (degree 9; betweenness centrality 36), respectively (Figure 3c-e).

Identification of biomarker candidates for multiple myeloma

The common molecular biomarker candidates among DEGs and DEMs for diagnosis, disease progression, efficacy, prognosis, response to therapy and safety were identified using the IPA software and HMDD/miRNet database, respectively. The analysis revealed 196 (9.95%) potential biomarkers out of 1970 observed DEGs that could bear clinical value for MM and were designated as 'Union 196' signature (Figure 4a). These included common biomarker candidates predominantly with diagnosis (n=134; 68.37%), efficacy (n=93; 47.45%), prognosis (n=55; 28.06%), disease progression (n=34; 17.35%), response to therapy (n=23; 11.73%), and safety (n=14; 7.14%) (Table S14 and S15 in Supplementary File 1). Among the target gene candidate biomarkers, 44.38% (n=87) of targets qualified for more than one role (Supplementary Tables S14 and S15). For example, gene CDKN2A and IL10 were observed to be implicated in diagnosis, disease progression, efficacy, prognosis, response to therapy and safety. The lists of common gene biomarkers are given in Table 5.

In addition, miRNA disease databases such as HMDD and miRNet revealed 23 aberrantly expressed miRNAs as potential biomarkers with clinical utility for MM (Table S16 in Supplementary File 1) and were designated as 'Union 23' signature (Figure 4b). The systematic literature review of 'Union 23' signatures disclosed that 30.43% (n=7) miRNAs were known circulating biomarkers as diagnostics and prognostic in multiple myeloma. Other miRNAs were identified as epigenetically regulated miRNAs (n=4), miRNAs as therapeutic targets (n=3) and dysregulated miRNAs that resulted in disease phenotypes (n=6) and are given in Table 6.

TF-gene/miRNA coregulatory networks

We further investigated the TF-miRNA-target gene regulatory network for meta-signature gene/miRNAs identified in this study. The gene-TF regulatory network of 8 gene biomarkers (≥ 5 applications) revealed 104 interaction pairs among 7 genes (SRC, IL10, CDKN2A, MKI67, MET, CCL2, and ESR1) and 99 transcription factors (TFs) (Table S17 in Supplementary File 1). Among them, top three upregulated genes including SRC (SRC Proto-Oncogene, Non-Receptor Tyrosine Kinase) was found to be regulated by the highest 53 TFs, CDKN2A (Cyclin Dependent Kinase Inhibitor 2A) was regulated by 11 TFs and MET (MET Proto-Oncogene, Receptor Tyrosine Kinase) was regulated by 10 TFs (Figure 5a-c). Similarly, top three downregulated genes such as IL10 (Interleukin 10) interacts with 15 TFs, MKI67 (Marker of Proliferation Ki-67) interacts with 11 TFs, and CCL2 (C-C Motif Chemokine Ligand 2) interacts with 3 TFs (Figure 5d-f). TF-gene interactions with connectivity degree ≥ 2 are shown in Table 7.

TF-miRNA regulations analysis of top 3 up- and downregulated miRNA biomarkers based on the number of targets showed the association between 6 miRNAs and 202 TFs (Table S18 in Supplementary File 1). From the data of TransmiR, upregulated DEMs such as hsa-mir-135b, hsa-mir-29b, and hsa-mir-16 were found to be regulated by 58, 16 and 9 TFs, respectively (Figure 6a-c). Likewise downregulated DEMs including hsa-mir-106b, hsa-mir-148a and hsa-let-7a were found to be regulated by 148, 82 and 14 TFs, respectively (Figure 6d-f). TF-miRNA interactions are shown in Table 7.

The top 5 enriched biological functions of TFs were investigated and subsequently compared for up- and downregulated meta-signature gene/ miRNAs (Supplementary Tables S19). TFs of upregulated genes were enriched in pathway namely “transcriptional misregulation in cancer”, whereas TFs of downregulated genes were not observed to be significantly enriched in any pathways. Moreover, we found that the TFs of up- and downregulated miRNAs were commonly enriched in four pathways, including transcriptional misregulation in cancer, pathways in cancer, HTLV-1 infection pathway and prostate cancer (Figure S3 in Supplementary File 2). Similarly, TFs were commonly enriched in three biological processes, including negative regulation of transcription from RNA polymerase II promoter, transcription, DNA-templated, and positive regulation of transcription from RNA polymerase II promoter. The top five commonly enriched biological processes (for TFs of up- and downregulated miRNAs) were negative regulation of transcription from RNA polymerase II promoter, transcription, DNA-templated, transcription from RNA polymerase II promoter, positive regulation of transcription, DNA-templated, and positive regulation of transcription from RNA polymerase II promoter. The biological functions of TFs are given in Supplementary Tables S19.

Discussion

In this study, a meta-analysis of mRNA and miRNA expression profiles has been carried out on more than 300 MM patients including 44 Indian myeloma patients (represented in 11 GSE-GEO datasets) in order to compute altered mRNA and miRNA patterns and potential biomarkers of clinical relevance in multiple myeloma. Overall, this study has imputed two net signatures, ‘Union 196’ for DEGs and ‘Union 23’ for DEMs in MM that appear to have a unified representation of several other analogous signatures reported in the literature (Figure 4a and b).

The present study has revealed that 78.18% (86/110) of DEMs and 51.61% (1970/3817) of DEGs were significantly altered, inversely correlated and involved in regulatory networking in multiple myeloma. Some of the miRNAs such as miR-29b/c found to be downregulated in MM in this study and previous studies are known to reduce cell proliferation and possess tumor suppressive effects³⁵. Similarly, miRNA-135b is primarily oncogenic and has been consistently found to be upregulated not only in this study but also others³⁵.

Computational studies in the past have shown that 63 miRNAs undergo temporal changes during plasma cell differentiation and that 47 of these form network relationships with 141 targets¹⁴. Among these, 24 miRNAs were reported to be aberrantly expressed in MM patients (n=62) and 13 in myeloma cell lines (n=20). Out of these, 4 and 3 DEMs observed in MM patients and myeloma cell lines, respectively are in common with the list of 23 DEMs (Union 23 signature) observed in the present study. These overlapping miRs include miR-16, miR-148a, miR-29b, miR-30b, miR-155, miR-106b and miR-18a. Another similar study¹³ mined two miRNA and two mRNA microarray GEO datasets and identified 39 DEMs and 32 hub genes. Among these DEMs, miR-155, miR-135, miR-148 and miR-203 were found to be deregulated in their study¹³ as well as in Union 23 profile in the present work. Likewise, another meta-analysis of 7 datasets

including MM patients and cell lines¹² highlighted 13 DEMs, of which hsa-miR-106b is also represented in Union 23 profile. A recent meta-analysis by Xu et al¹¹ reported 7 DEMs of prognostic significance among which deregulated miR-92a and miR-16 are common to the Union 23 signature.

Recent studies have eluded the predictive potential of circulating exosomal miRNAs found in plasma and other body fluids in MM²¹. Low circulating levels of let-7b and miRNA-18a have been significantly associated with poor overall survival as well as progression free survival³⁶ and were found to be downregulated in this composite analysis too. Similarly, higher circulating miR-193a-5p has been shown to be an important prognostic biomarker for reduced probability of early relapse and better PFS following autologous SCT³⁷. On the contrary, down regulation of circulating miRNA-193 has been associated with favoured progression of cancer in osteosarcoma and colorectal cancer^{38,39}. This report further supports our study, where a related miRNA-193b was found to be downregulated in MM and it may be worthwhile to examine the DEMs identified in this study in circulation among myeloma patients.

In addition, top 5 up- and down-regulated miRNAs identified from this study were detected to be involved in MM. For example, overexpression of miR-26a inhibited MM cell growth by suppressing cyclin-dependent kinase 6 expression (CDK6)⁴⁰. Upregulated expression of miR-103a-3p was observed to be associated with renal impairment risk in MM patients and may use as novel biomarkers for RI risk in MM patients⁴¹. Higher expression of miR-155 has been associated with longer survival in MM primary samples⁴². The overexpression of let-7 family member such as let-7b-5p and miR-16 was identified to be acts as a tumor suppressor in MM⁴³⁻⁴⁴. Likewise, low expression of hsa-miR-575 was significantly associated with poor survival of MM patients⁴⁵. Low expression of miR-198 in MM was significantly associated with repressed MM cell proliferation, colony formation and migration⁴⁶. Downregulation of miR-198 was significantly associated with its tumor suppressor activity in MM⁴⁷. Similarly deregulation of miR-572 and upregulation of miR-638 was exclusively associated with MM^{16, 48}. In contrast, down-regulated expression of miR-638 was observed in our datasets as reported in other cancers including breast cancers⁴⁹ and oral squamous cell carcinoma⁵⁰.

The IPA core analysis disclosed that 9.95% (n=196) of DEGs are putative biomarkers that could be useful in diagnosis, disease progression, efficacy, prognosis, response to therapy and safety. Further investigation revealed that 44.38% (n=87) of targets were involved in more than one functional role. It is known that proteins with the highest degree have the highest betweenness in the network. As hub proteins are accountable for holding networks together⁵¹⁻⁵², they are more likely to be master regulators of signaling and transcription and can be used as therapeutic targets or biomarkers⁵³. The target genes identified in this study were subjected to PPI network which disclosed a total of 76 nodes with ≥ 10 connections with other nodes and were designated as 'hub' genes. All hub genes were observed to be targeted by MM associated DEMs and could act as possible biomarkers for this disease. It is noteworthy that five top hub genes lying in the centre of functional networks delineated include *APP*, *KIAA0101*, *EGR1*,

ONECUT1, and *HNF1A*, out of which the *APP* biomarker⁵⁴⁻⁵⁵ supersedes the rest in terms of its degrees of interaction.

Conclusions

The regulatory crosstalk between DEGs and DEMs in MM is highly complex. Their subnetworks operate differentially in different patients and capturing their interactions may be prudent to understand and perceive biomarkers of personified clinical significance. The top five hub genes and their functional networks deduced in this study may need to be validated in further studies or prospective trials for their clinical significance.

Abbreviations

BP, biological process; **DAVID**, database for annotation, visualization and integrated discovery; **DEG**, differentially expressed genes; **DEM**, differentially expressed miRNAs; **FDR**, false discovery rate; **GEO**, gene expression omnibus; **GO**, gene ontology; **IPA**, ingenuity pathway analysis; **KEGG**, Kyoto encyclopedia of genes and genomes; **MM**, multiple myeloma; **miRNA**, microRNA; **NCBI**, National Center for Biotechnology Information; **PPI**, protein-protein interaction.

Declarations

Author contributions

Conceived and designed the experiments: RG, and PK. Performed the experiments: LR. Analysed the data: AK, GK, HS, PK and RG. Wrote the manuscript: AK, GK, LR, RG, HS and PK. All authors have read and approved the final manuscript.

Conflict of interest statement

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The financial support was provided by Department of Biotechnology (BT/ PR7136 /MED /30/901/2012 and BT/MED/30/SP11006/2015), Ministry of Science & Technology, Government of India, and All India Institute of Medical Sciences, New Delhi (A411 and AC002/RS) to RG for carrying out the research work.

Data availability

Gene expression (GSE125361) and miRNA expression (GSE125363) signatures in multiple myeloma have been submitted to the National Center for Biotechnology Information (NCBI; <https://www.ncbi.nlm.nih.gov/geo>) under BioProject accession number PRJNA515992.

References

1. Szalat R, Avet-Loiseau H, Munshi NC. Gene Expression Profiles in Myeloma: Ready for the Real World? *Clin Cancer Res* 2016;22:5434–42.
2. Zhan F, Barlogie B, Arzoumanian V, Huang Y, Williams DR, Hollmig K, Pineda-Roman M, Tricot G, van Rhee F, Zangari M, Dhodapkar M, Shaughnessy JD. Gene-expression signature of benign monoclonal gammopathy evident in multiple myeloma is linked to good prognosis. *Blood* 2007;109:1692–700.
3. Shaughnessy JD, Zhan F, Burington BE, Huang Y, Colla S, Hanamura I, Stewart JP, Kordsmeier B, Randolph C, Williams DR, Xiao Y, Xu H, et al. A validated gene expression model of high-risk multiple myeloma is defined by deregulated expression of genes mapping to chromosome 1. *Blood* 2007;109:2276–84.
4. Decaux O, Lodé L, Magrangeas F, Charbonnel C, Gouraud W, Jézéquel P, Attal M, Harousseau J-L, Moreau P, Bataille R, Campion L, Avet-Loiseau H, et al. Prediction of survival in multiple myeloma based on gene expression profiles reveals cell cycle and chromosomal instability signatures in high-risk patients and hyperdiploid signatures in low-risk patients: a study of the Intergroupe Francophone du Myélome. *J Clin Oncol* 2008;26:4798–805.
5. Bai H, Chen B. A 5-Gene Stemness Score for Rapid Determination of Risk in Multiple Myeloma. *Oncotargets Ther* 2020;13:4339–48.
6. Chung T-H, Mulligan G, Fonseca R, Chng WJ. A novel measure of chromosome instability can account for prognostic difference in multiple myeloma. *PLoS ONE* 2013;8:e66361.
7. Kuiper R, Broyl A, de Knecht Y, van Vliet MH, van Beers EH, van der Holt B, el Jarari L, Mulligan G, Gregory W, Morgan G, Goldschmidt H, Lokhorst HM, et al. A gene expression signature for high-risk multiple myeloma. *Leukemia* 2012;26:2406–13.
8. Dickens NJ, Walker BA, Leone PE, Johnson DC, Brito JL, Zeisig A, Jenner MW, Boyd KD, Gonzalez D, Gregory WM, Ross FM, Davies FE, et al. Homozygous Deletion Mapping in Myeloma Samples Identifies Genes and an Expression Signature Relevant to Pathogenesis and Outcome. *Clin Cancer Res* 2010;16:1856–64.
9. Mulligan G, Mitsiades C, Bryant B, Zhan F, Chng WJ, Roels S, Koenig E, Fergus A, Huang Y, Richardson P, Trepicchio WL, Broyl A, et al. Gene expression profiling and correlation with outcome in clinical trials of the proteasome inhibitor bortezomib. *Blood* 2007;109:3177–88.
10. Liu Y, Yu H, Yoo S, Lee E, Laganà A, Parekh S, Schadt EE, Wang L, Zhu J. A Network Analysis of Multiple Myeloma Related Gene Signatures. *Cancers (Basel)* 2019;11.
11. Xu P, Xia T, Ling Y, Chen B. MiRNAs with prognostic significance in multiple myeloma: A systemic review and meta-analysis. *Medicine (Baltimore)* 2019;98:e16711.
12. Xiang T, Hu A-X, Sun P, Liu G, Liu G, Xiao Y. Identification of four potential predicting miRNA biomarkers for multiple myeloma from published datasets. *PeerJ* 2017;5:e2831.
13. Gao H, Wang H, Yang W. Identification of key genes and construction of microRNA-mRNA regulatory networks in multiple myeloma by integrated multiple GEO datasets using bioinformatics analysis. *Int*

J Hematol 2017;106:99–107.

14. Kassambara A, Jourdan M, Bruyer A, Robert N, Pantesco V, Elemento O, Klein B, Moreaux J. Global miRNA expression analysis identifies novel key regulators of plasma cell differentiation and malignant plasma cell. *Nucleic Acids Res* 2017;45:5639–52.
15. Misiewicz-Krzeminska I, Krzeminski P, Corchete LA, Quwaider D, Rojas EA, Herrero AB, Gutiérrez NC. Factors Regulating microRNA Expression and Function in Multiple Myeloma. *Noncoding RNA [Internet]* 2019 [cited 2020 Jul 15];5. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6468559/>
16. Gutiérrez NC, Sarasquete ME, Misiewicz-Krzeminska I, Delgado M, De Las Rivas J, Ticona FV, Ferriñán E, Martín-Jiménez P, Chillón C, Risueño A, Hernández JM, García-Sanz R, et al. Deregulation of microRNA expression in the different genetic subtypes of multiple myeloma and correlation with gene expression profiling. *Leukemia* 2010;24:629–37.
17. Lionetti M, Biasiolo M, Agnelli L, Todoerti K, Mosca L, Fabris S, Sales G, Deliliers GL, Bicciato S, Lombardi L, Bortoluzzi S, Neri A. Identification of microRNA expression patterns and definition of a microRNA/mRNA regulatory network in distinct molecular groups of multiple myeloma. *Blood* 2009;114:e20–6.
18. Corthals SL, Jongen-Lavrencic M, de Knecht Y, Peeters JK, Beverloo HB, Lokhorst HM, Sonneveld P. Micro-RNA-15a and micro-RNA-16 expression and chromosome 13 deletions in multiple myeloma. *Leuk Res* 2010;34:677–81.
19. Pichiorri F, De Luca L, Aqeilan RI. MicroRNAs: New Players in Multiple Myeloma. *Front Genet* 2011;2:22.
20. Handa H, Murakami Y, Ishihara R, Kimura-Masuda K, Masuda Y. The Role and Function of microRNA in the Pathogenesis of Multiple Myeloma. *Cancers (Basel) [Internet]* 2019 [cited 2020 Jul 15];11. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6896016/>
21. Federico C, Sacco A, Belotti A, Ribolla R, Cancelli V, Giacomini A, Ronca R, Chiarini M, Imberti L, Marini M, Rossi G, Presta M, et al. Circulating microRNAs and Their Role in Multiple Myeloma. *Noncoding RNA [Internet]* 2019 [cited 2020 Jul 15];5. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6631121/>
22. Barrett T, Edgar R. Mining Microarray Data at NCBI's Gene Expression Omnibus (GEO). *Methods Mol Biol* 2006;338:175–90.
23. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, Yefanov A, Lee H, et al. NCBI GEO: archive for functional genomics data sets –update. *Nucleic Acids Res* 2013;41:D991–5.
24. Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* 2004;3:Article3.
25. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B (Methodological)* 1995;57:289–300.

26. Rajkumar SV, Dimopoulos MA, Palumbo A, Blade J, Merlini G, Mateos M-V, Kumar S, Hillengass J, Kastritis E, Richardson P, Landgren O, Paiva B, et al. International Myeloma Working Group updated criteria for the diagnosis of multiple myeloma. *Lancet Oncol* 2014;15:e538-548.
27. Katiyar A, Sharma S, Singh TP, Kaur P. Identification of Shared Molecular Signatures Indicate the Susceptibility of Endometriosis to Multiple Sclerosis. *Front Genet [Internet]* 2018 [cited 2020 Jul 21];9. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5820528/>
28. Arasappan D, Tong W, Mummaneni P, Fang H, Amur S. Meta-analysis of microarray data using a pathway-based approach identifies a 37-gene expression signature for systemic lupus erythematosus in human peripheral blood mononuclear cells. *BMC Med* 2011;9:65.
29. Taminau J, Lazar C, Meganck S, Nowé A. Comparison of Merging and Meta-Analysis as Alternative Approaches for Integrative Gene Expression Analysis. *ISRN Bioinform [Internet]* 2014 [cited 2020 Jul 21];2014. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4393058/>
30. Reimand J, Kull M, Peterson H, Hansen J, Vilo J. g:Profiler—a web-based toolset for functional profiling of gene lists from large-scale experiments. *Nucleic Acids Res* 2007;35:W193-200.
31. Piñero J, Queralt-Rosinach N, Bravo À, Deu-Pons J, Bauer-Mehren A, Baron M, Sanz F, Furlong LI. DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes. *Database (Oxford) [Internet]* 2015 [cited 2020 Jul 15];2015. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4397996/>
32. Zhou G, Soufan O, Ewald J, Hancock REW, Basu N, Xia J. NetworkAnalyst 3.0: a visual analytics platform for comprehensive gene expression profiling and meta-analysis. *Nucleic Acids Res* 2019;47:W234–41.
33. Breuer K, Foroushani AK, Laird MR, Chen C, Sribnaia A, Lo R, Winsor GL, Hancock REW, Brinkman FSL, Lynn DJ. InnateDB: systems biology of innate immunity and beyond—recent updates and continuing curation. *Nucleic Acids Res* 2013;41:D1228-1233.
34. Haynes W. Wilcoxon Rank Sum Test [Internet]. In: Dubitzky W, Wolkenhauer O, Cho K-H, Yokota H, eds. *Encyclopedia of Systems Biology*. New York, NY: Springer, 2013 [cited 2020 Jul 15]. 2354–5. Available from: https://doi.org/10.1007/978-1-4419-9863-7_1185
35. Zhu B, Ju S, Chu H, Shen X, Zhang Y, Luo X, Cong H. The potential function of microRNAs as biomarkers and therapeutic targets in multiple myeloma. *Oncol Lett* 2018;15:6094–106.
36. Manier S, Liu C-J, Avet-Loiseau H, Park J, Shi J, Campigotto F, Salem KZ, Huynh D, Glavey SV, Rivotto B, Sacco A, Roccaro AM, et al. Prognostic role of circulating exosomal miRNAs in multiple myeloma. *Blood* 2017;129:2429–36.
37. Park S-S, Lim J-Y, Kim TW, Ko YH, Jeon W-J, Lee S-Y, Lee J-H, Min C-K. Predictive impact of circulating microRNA-193a-5p on early relapse after autologous stem cell transplantation in patients with multiple myeloma. *British Journal of Haematology* 2020;189:518–23.
38. Gougelet A, Pissaloux D, Besse A, Perez J, Duc A, Dutour A, Blay J-Y, Alberti L. Micro-RNA profiles in osteosarcoma as a predictive tool for ifosfamide response. *International Journal of Cancer* 2011;129:680–90.

39. Zhang P, Ji D-B, Han H-B, Shi Y-F, Du C-Z, Gu J. Downregulation of miR-193a-5p correlates with lymph node metastasis and poor prognosis in colorectal cancer. *World J Gastroenterol* 2014;20:12241–8.
40. Xu YY, Song YQ, Huang ZM, Zhang HB, Chen M. MicroRNA-26a inhibits multiple myeloma cell growth by suppressing cyclin-dependent kinase 6 expression. *Kaohsiung J Med Sci.* 2019;35(5):277-283.
41. Ren D, Cai Y, Xu G. Potential of microRNA expression profile in predicting renal impairment risk in multiple myeloma patients. *Transitional cancer Research.* 2020; 9(3). doi: 10.21037/tcr.2020.01.41
42. Krzeminski P, Sarasquete ME, Misiewicz-Krzeminska I, Corral R, Corchete LA, Martín AA, García-Sanz R, San Miguel JF, Gutiérrez NC. Insights into epigenetic regulation of microRNA-155 expression in multiple myeloma. *Biochim Biophys Acta.* 2015;1849(3):353-66.
43. Xu H, Liu C, Zhang Y, Guo X, Liu Z, Luo Z, Chang Y, Liu S, Sun Z, Wang X. Let-7b-5p regulates proliferation and apoptosis in multiple myeloma by targeting IGF1R. *Acta Biochim Biophys Sin (Shanghai).* 2014; 46(11):965-72.
44. Zhang L, Zhou L, Shi M, Kuang Y, Fang L. Downregulation of miRNA-15a and miRNA-16 promote tumor proliferation in multiple myeloma by increasing CABIN1 expression. *Oncol Lett.* 2018;15(1):1287-1296.
45. Chi J, Ballabio E, Chen XH, Kušec R, Taylor S, Hay D, Tramonti D, Saunders NJ, Littlewood T, Pezzella F, Boultonwood J, Wainscoat JS, Hatton CS, Lawrie CH. MicroRNA expression in multiple myeloma is associated with genetic subtype, isotype and survival. *Biol Direct.* 2011;18:6:23.
46. Chonglei B., Zhou J., Chung T.H., Huang G., Yan J., Chng W.J. Identification of tumor suppressive microRNAs in multiple myeloma by pharmacologic unmasking. *Clin. Lymphoma Myeloma Leuk.*2013;13:S37–S38.
47. Bi C, Chung TH, Huang G, Zhou J1, Yan J, Ahmann GJ, Fonseca R and Chng WJ: Genome-wide pharmacologic unmasking identifies tumor suppressive microRNAs in multiple myeloma. *Oncotarget.*2015; 6:26508–26518.
48. Yyusnita, Norsiah, Zakiah I, Chang KM, Purushotaman VS, Zubaidah Z and Jamal R: MicroRNA (miRNA) expression profiling of peripheral blood samples in multiple myeloma patients using microarray. *Malays J Pathol.* 2012; 34:133–143.
49. Tan X, Peng J, Fu Y, An S, Rezaei K, Tabbara S, Teal CB, Man YG, Brem RF, Fu SW. miR-638 mediated regulation of BRCA1 affects DNA repair and sensitivity to UV and cisplatin in triple-negative breast cancer. *Breast Cancer Res.* 2014; 16(5):435.
50. Tang KL, Tang HY, Du Y, Tian T, Xiong SJ. MiR-638 suppresses the progression of oral squamous cell carcinoma through wnt/ β -catenin pathway by targeting phospholipase D1. *Artif Cells Nanomed Biotechnol.* 2019; 47(1):3278-3285.
51. Raman K. Construction and analysis of protein–protein interaction networks. *Automated Experimentation* 2010;2:2.
52. Yu H, Kim PM, Sprecher E, Trifonov V, Gerstein M. The Importance of Bottlenecks in Protein Networks: Correlation with Gene Essentiality and Expression Dynamics. *PLOS Computational Biology*

2007;3:e59.

53. Jeong H, Mason SP, Barabási AL, Oltvai ZN. Lethality and centrality in protein networks. *Nature* 2001;411:41–2.
54. Pandey P, Sliker B, Peters HL, Tuli A, Herskovitz J, Smits K, Purohit A, Singh RK, Dong J, Batra SK, Coulter DW, Solheim JC. Amyloid precursor protein and amyloid precursor-like protein 2 in cancer. *Oncotarget* 2016;7:19430–44.
55. Galvão F, Grokoski KC, da Silva BB, Lamers ML, Siqueira IR. The amyloid precursor protein (APP) processing as a biological link between Alzheimer’s disease and cancer. *Ageing Res Rev* 2019;49:83–91.

Tables

Table 1: List of top 5 up- and down-regulated miRNAs and genes in multiple myeloma.

Accession no	miRNA name (miRBase-22)	adj.P.Val	logFC	Regulation
MIMAT0000069	hsa-miR-16-5p	5.40E-02	5.39	Up
MIMAT0000414	hsa-let-7g-5p	1.10E-03	4.77	Up
MIMAT0000646	hsa-miR-155-5p	2.97E-03	4.69	Up
MIMAT0000101	hsa-miR-103a-3p	7.85E-06	4.61	Up
MIMAT0000082	hsa-miR-26a-5p	1.28E-02	4.05	Up
MIMAT0003240	hsa-miR-575	1.12E-02	-6.85	Down
MIMAT0003308	hsa-miR-638	1.43E-03	-6.43	Down
MIMAT0003237	hsa-miR-572	7.10E-03	-6.31	Down
MIMAT0015070	hsa-miR-3188	4.10E-04	-6.22	Down
MIMAT0000228	hsa-miR-198	1.43E-03	-6.17	Down
Accession no	Gene name	adj.P.Val	logFC	Regulation
HGNC:6342	KIT	1.81E-02	5.02	Up
HGNC:3539	F2RL2	2.27E-03	4.43	Up
HGNC:1073	BMP6	1.41E-07	4.37	Up
HGNC:7656	NCAM1	2.69E-02	4.33	Up
HGNC:6757	MAB21L1	2.06E-02	4.09	Up
HGNC:5870/38476	IGLL1/IGLL5	1.46E-02	-4.64	Down
HGNC:11763	TFRC	2.65E-06	-3.95	Down
HGNC:6720	LTF	2.03E-02	-3.94	Down
HGNC:13371	MS4A4A	8.61E-06	-3.78	Down
HGNC:7545	MYB	2.61E-02	-3.78	Down

Table 2: Top five diseases and bio functions identified by IPA analysis of miRNA and mRNA in MM

Name (miRNA)	p-value range	Focus molecules
Diseases and Disorders		
Cancer	4.82E-02 - 7.77E-38	63
Organismal Injury and Abnormalities	4.98E-02 - 7.77E-38	68
Reproductive System Disease	1.81E-02 - 7.77E-38	48
Inflammatory Disease	4.98E-02 - 3.26E-34	48
Inflammatory Response	4.98E-02 - 3.26E-34	42
Molecular and Cellular Functions		
Cellular Development	4.76E-02 - 2.62E-13	41
Cellular Growth and Proliferation	4.76E-02 - 2.62E-13	37
Cellular Movement	3.59E-02 - 5.41E-07	24
Cell Cycle	4.79E-02 - 5.49E-07	14
Cell Death and Survival	3.46E-02 - 1.88E-05	24
Physiological System Development and Function		
Organismal Development	4.76E-02 - 2.26E-15	21
Digestive System Development and Function	4.11E-10 - 4.11E-10	7
Hepatic System Development and Function	4.11E-10 - 4.11E-10	7
Organ Development	4.76E-02 - 4.11E-10	9
Cardiovascular System Development and Function	4.76E-02 - 6.38E-09	16
Name (genes)		
Diseases and Disorders		
Cancer	9.49E-09 - 1.09E-92	1882
Organismal Injury and Abnormalities	9.57E-09 - 1.09E-92	1901
Gastrointestinal Disease	6.79E-09 - 2.55E-56	1695
Endocrine System Disorders	3.66E-10 - 9.36E-53	1616
Dermatological Diseases and Conditions	6.43E-09 - 8.48E-43	1193
Molecular and Cellular Functions		
Cellular Movement	6.25E-09 - 1.47E-27	534
Cell Death and Survival	2.64E-09 - 2.73E-24	679

Cellular Growth and Proliferation	8.28E-09 - 4.26E-22	717
Cellular Development	8.28E-09 - 5.30E-21	752
Cell-To-Cell Signaling and Interaction	9.78E-09 - 1.98E-20	346
Physiological System Development and Function		
Tissue Morphology	6.44E-09 - 3.27E-24	538
Organismal Survival	3.74E-09 - 2.21E-22	545
Lymphoid Tissue Structure and Development	2.36E-10 - 4.26E-22	349
Connective Tissue Development and Function	3.48E-09 - 5.30E-21	387
Tissue Development	8.28E-09 - 5.30E-21	586

Table 3: Top five *associated network functions* predicted by an IPA analysis of miRNA and mRNA in MM

ID	Associated Network Functions (miRNAs)	Score	Focus Molecules
1	Cancer, Organismal Injury and Abnormalities, Reproductive System Disease	61	26
2	Glomerular Injury, Inflammatory Disease, Inflammatory Response	27	14
3	Cancer, Organismal Injury and Abnormalities, Reproductive System Disease	22	12
4	Glomerular Injury, Inflammatory Disease, Inflammatory Response	22	12
5	Cancer, Gastrointestinal Disease, Organismal Injury and Abnormalities	20	11
ID	Associated Network Functions (genes)		
1	Cancer, Dermatological Diseases and Conditions, Organismal Injury and Abnormalities	42	35
2	RNA Post-Transcriptional Modification, Protein Synthesis, Auditory Disease	37	33
3	Cell Cycle, Cell-To-Cell Signaling and Interaction, Nervous System Development and Function	37	33
4	RNA Post-Transcriptional Modification, Gene Expression, Protein Synthesis	35	32
5	Connective Tissue Disorders, Organismal Injury and Abnormalities, Skeletal and Muscular Disorders	35	32

Table 4: Top five significant functional modules and associated hub genes

Module	Size	p-value	ID	Hub gene	Degree (≥ 10)	Betweenness
1	182	3.35E-23	9768	KIAA0101	48	3438.672
			1017	CDK2	44	4142.063
			2099	ESR1	35	2851.185
			2335	FN1	34	2151.942
			7341	SUMO1	32	2794.829
			4163	MCC	20	1187.158
			409	ARRB2	19	846.688
			4691	NCL	15	785.867
			10399	GNB2L1	12	843.3601
			23450	SF3B3	12	728.3915
			4000	LMNA	11	468.303
			6136	RPL12	11	348.7949
			4176	MCM7	10	890.3589
			10627	MYL12A	10	325.5931
			6152	RPL24	10	334.0069
2	163	4.54E-20	351	APP	103	10586.07
			6714	SRC	16	1307.26
			6772	STAT1	14	882.44
			3725	JUN	13	949.37
			5468	PPARG	13	760.39
			5925	RB1	12	522.03
			10524	KAT5	11	1265.96
			1026	CDKN1A	11	911.75
			55791	LRIF1	10	531.37
595	CCND1	10	179.47			
5	34	6.94E-05	1958	EGR1	29	518
6	16	2.57E-03	3175	ONECUT1	14	103
8	10	9.60E-03	6927	HNF1A	10	36

Table 5: Common molecular biomarker candidates for diagnosis, disease progression, efficacy, prognosis, response to therapy and safety in multiple myeloma

Gene biomarker	Description	Regulation	Biomarker applications
CDKN2A	Cyclin Dependent Kinase Inhibitor 2A	Up	diagnosis, disease progression, efficacy, prognosis, response to therapy, safety
SRC	SRC Proto-Oncogene, Non-Receptor Tyrosine Kinase	Up	diagnosis, disease progression, efficacy, prognosis, response to therapy
MET	MET Proto-Oncogene, Receptor Tyrosine Kinase	Up	diagnosis, disease progression, efficacy, prognosis, response to therapy
KIT	KIT Proto-Oncogene, Receptor Tyrosine Kinase	Up	diagnosis, efficacy, prognosis, safety
CDKN1A	Cyclin Dependent Kinase Inhibitor 1A	Up	diagnosis, efficacy, prognosis, response to therapy
STAT1	Signal Transducer And Activator Of Transcription 1	Up	diagnosis, efficacy, prognosis, response to therapy
CCND1	Cyclin D1	Up	diagnosis, efficacy, prognosis, response to therapy
HGF	Hepatocyte Growth Factor	Up	diagnosis, disease progression, efficacy, prognosis
PTGS2	Prostaglandin-Endoperoxide Synthase 2	Up	diagnosis, disease progression, efficacy, prognosis
IL10	Interleukin 10	Down	diagnosis, disease progression, efficacy, prognosis, response to therapy, safety
CCL2	C-C Motif Chemokine Ligand 2	Down	diagnosis, efficacy, prognosis, response to therapy, safety
IGFBP3	Insulin Like Growth Factor Binding Protein 3	Down	diagnosis, disease progression, efficacy, prognosis, safety
ESR1	Estrogen Receptor 1	Down	diagnosis, disease progression, efficacy, prognosis, response to therapy
MKI67	Marker Of Proliferation Ki-67	Down	diagnosis, disease progression, efficacy, prognosis, response to therapy
MTHFR	Methylenetetrahydrofolate Reductase	Down	diagnosis, efficacy, response to therapy, safety
CALCA	Calcitonin Related Polypeptide Alpha	Down	diagnosis, efficacy, prognosis, safety
LEP	Leptin	Down	diagnosis, efficacy, prognosis, safety
HIF1A	Hypoxia Inducible Factor 1 Subunit Alpha	Down	diagnosis, efficacy, prognosis, response to therapy
CDH1	Cadherin 1	Down	diagnosis, disease progression, efficacy,

			prognosis
PTK2	Protein Tyrosine Kinase 2	Down	diagnosis, disease progression, efficacy, prognosis
CSF1	Colony Stimulating Factor 1	Down	diagnosis, disease progression, efficacy, prognosis

□ Gene candidates with ≥ 4 biomarker applications

Table 6: List of miRNAs that could result in disease phenotypes (multiple myeloma) when permuted.

miRNA biomarker	Regulation	Evidence	Description	Causality
hsa-miR-16-5p	Up	target gene	MicroRNAs 15a and 16 regulate tumor proliferation in multiple myeloma.	YES
hsa-miR-16-5p	Up	target gene	miR-15a and miR-16 affect the angiogenesis of multiple myeloma by targeting VEGF.	YES
hsa-miR-199a-3p	Up	target gene	Targeting of multiple myeloma-related angiogenesis by miR-199a-5p mimics: in vitro and in vivo anti-tumor activity.	YES
hsa-miR-29b-3p	Up	therapeutic target	miR-29b-based epi-therapeutic approaches in the treatment of this malignancy	YES
hsa-miR-148a-3p	Down	target gene	miR-148a participates in the growth of RPMI8226 multiple myeloma cells by regulating CDKN1B.	YES
hsa-miR-197-5p	Down	target gene	miR-137 and miR-197 Induce Apoptosis and Suppress Tumorigenicity by Targeting MCL-1 in Multiple Myeloma.	YES
hsa-miR-451a	Down	target gene	MicroRNA-451 regulates stemness of side population cells via PI3K/Akt/mTOR signaling pathway in multiple myeloma.	YES

Table 7: Upstream transcription factors (TFs) of dominant gene/miRNA biomarkers in multiple myeloma

Candidates	Regulation	Transcription factors (TFs)
*Gene biomarker		
SRC	Up	TEAD1, MBD1, ETV1, KDM1A, TFDP1, MXD3, ELF3, FOXA3, EGR2, ETV4, GMEB2, KLF16, HHEX, ZHX2, PBX2, ARID4B, RCOR2, ZFP64, RARA, CBFB, ADNP, NR2F1, SMARCA5, EED, ZKSCAN1, GATAD2A, KLF9, MXD4, GLIS2, TEAD3, KLF11, THRB, HBP1, RAD21, RFXANK, HMG20B, DRAP1, MLX, TFAP4, GATAD1, TGIF2, ZNF580, SOX13, HDAC2, WRNIP1, E2F5, ZBTB26, SSRP1, EBF1, DMAP1, SIN3A, BCL6, HDAC6
CDKN2A	Up	ZFP37, L3MBTL2, ZNF2, CTBP2, ZNF71, CTCF, EZH2, ZNF423, ZNF76, PRDM1, SUZ12
MET	Up	BCL11B, HBP1, RFXANK, ELF3, ZNF366, ZNF101, STAT3, TGIF2, ZNF24, MBD1
IL10	Down	CBFB, SUPT5H, GTF2E2, TBX21, JUNB, RUNX3, MLLT1, ZNF143, POU2F2, TCF7, ID3, SPI1, TRIM22, HES1, NFIC
MKI67	Down	TBX21, ZFP2, ZFP37, IRF4, GLIS2, FOXM1, EGR1, NRF1, KLF8, ZNF2, ZNF501
CCL2	Down	CEBPG, YY1, KLF16
#miRNA biomarker		
hsa-mir-135b	Up	AR, BRD4, CBFA2T2, CDK8, CTNNB1, E2F6, EGR1, ELF1, EP300, ERG, ESR1, EZH2, FOS, FOSL1, FOXM1, GABPA, GATA3, GATA4, GRHL2, HEY1, HIF1A, HNF4A, HSF1, JUN, JUNB, KDM5B, KMT2D, LARP7, MAFK, MAX, MBD3, NANOG, NKX2-1, OTX2, PAX6, PGR, PHF8, PRDM14, RAD21, REST, RUNX1, SFPQ, SIN3A, SMAD1, SMC1A, SOX2, SP1, STAT3, SUPT5H, TAF1, TCF12, TCF7L2, TEAD4, TFAP2C, TRIM25, TRIM28, USF1, ZBTB7A
hsa-mir-29b	Up	CEBPA, DNMT3A, DNMT3B, GATA3, IRF1, MYC, NFE2L2, NFKB1, NR1H3, PPARG, SMAD3, SOX2, SP1, SREBF1, SRF, YY1
hsa-mir-16	Up	E2F1, E2F3, E2F7, ETS1, MAX, MYC, NFKB1, STAT5, ZEB1
hsa-mir-106b	Down	AHR, AML1-ETO, AR, ARNT, ARNTL, ASXL1, ATF2, ATF3, ATF4, BCL6, BCOR, BHLHE40, BRD4, BRPF3, CBFB, CBX3, CCNT2, CDK8, CDK9, CEBPB, CHD2, CHD8, CREB1, CREBBP, CTCFL, CXXC1, E2F1, E2F3, E2F4, E2F6, E2F7, EBF1, EGR1, ELF1, ELK3, EP300, EP400, ERG, ESR1, ETS1, ETV1, EZH2, FGFR1, FLI1, FOS, FOSL2, FOXA1, FOXA2, FOXK1, FOXO1, FOXP1, FOXP2, GABPA, GATA1, GATA3, GTF2B, GTF2I, HCFC1, HDAC1, HEY1, HIF1A, HNF4G, HOXA4, HOXA6, HSF1, ICE1, ING2, IRF1, JUN, JUND, KDM2B, KDM5B, KLF1, KLF11, KLF5, KMT2A, LARP7, LEO1, MAFF, MAFK, MAX, MAZ, MBD3, MED1, MITE, MXI1, MYB, MYC, MYCN, MYH11, NELFE, NFE2, NFE2L2, NFKB1, NFYA, NFYB, NIPBL, NOTCH1, PBX1, PBX3, PHF8, RARA, RB1, RBBP5, RBL2, RELA, RFX5, RUNX1, RUNX1T1, RUNX3, SAP30, SIN3A, SMAD3, SNAI2, SOX2, SP1, SP2, SPI1, STAT1, STAT3, SUMO2, SUPT5H, TAF1, TAF7, TAL1, TAZ, TBP, TCF12, TCF3, TCF7L2, TEAD1, TEAD4, TFAP2A, TFAP4, TRIM24, TRIM25, TRIM28, UBTF, USF1, USF2, VDR, WDR5, YAP1, YY1, ZBTB7A, ZNF143, ZNF384, ZNF83
hsa-mir-	Down	AR, ARID3A, ARNTL, ASXL1, ATF3, BHLHE40, BRD4, CDK9, CEBPB,

148a		CNOT3, CREB1, CTCF, CTCFL, DNMT1, E2F1, E2F4, E2F6, EGR1, EP300, ERG, FGFR1, FLI1, FOXA1, FOXA2, GATA1, GATA2, GTF2I, HCFC1, HDAC2, HEY1, HIF1A, HNF4A, HNF4G, IL6, IRF1, JMJD1C, JUND, KDM2B, KDM5B, KMT2A, LYL1, MAFK, MAX, MAZ, MBD3, MED1, MITF, MXI1, MYB, MYBL2, MYC, MYCN, NFE2, NFIC, NFKB1, NFYB, NR2C2, NR2F1, NR2F2, OTX2, PHF8, RAD21, RUNX3, SETD1A, SP1, TAF1, TBP, TCF3, TCF4, TCF7L2, TEAD4, TFAP2C, TFAP4, TP53, TRIM24, TRIM28, USF1, USF2, WDR5, XBP1, YY1, ZEB1
hsa-let-7a	Down	AR, E2F1, E2F3, EIF2C2, FLI1, FSH, HIF1A, HOXA10, IL6, LIN28, LIN28B, MYC, STAT1, TRIM32

*Up- and down-regulated gene biomarker with ≥ 5 applications

#Up- and down-regulated miRNA biomarker based on the number of targets (inversely correlated)