

Identification of Novel Therapeutic Targets in Myelodysplastic Syndrome Using Protein-Protein Interaction Approach and Neural Networks

Ali A¹, Junaid M¹, Khan A¹, Kaushik AC¹, Mehmood A¹, Saleem S², Nangraj AS¹ and Dong-Qing Wei*¹

¹State Key Laboratory of Microbial Metabolism, and College of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, China

²Quaid-e-Azam University Islamabad Pakistan

Abstract

A Myelodysplastic syndrome (MDS) is a disorder characterized by active but ineffective hematopoiesis that leads to pancytopenia. MDS, also termed as myeloid neoplasms, is described by different level of cytopenia that is a different level of blood cells in the body. Various genes mutations have been reported to associate with MDS. To investigate the mechanisms at molecular level underlying MDS patients carrying genetic mutations, the gene expression profiles of MDS the patients were compared to that of healthy individuals and analyzed by bioinformatics tools. In biological networks, genes having important functional roles can be identified by a measure of the node. Networks of genes an in co-expression, candidate hubs also called extremely associated genes have been connected with the key disease-related pathway. Thus, this technique was used to discover the MDS related genes hub. Affymetrix Human Genome U133 plus 2.0 gene expression dataset of microarray GSE58831 was retrieved from GEO (Gene Expression Omnibus) database that contained four 159 diseased samples and 17 samples of control. Based on statistical method and co-expression networking, DEGs gene was detected. DAVID an online tool was employed for Gene ontology (GO) function and KEGG pathway enrichment analysis of DEGs. Besides, PPI (Protein-protein interaction) networks were developed by mapping the DEGs with respect to protein-protein interaction set available in databases for the identification of the pathways involving DEGs. PPI interaction networks were divided into subnetworks via MCODE algorithm and were examined by Cytoscape. Interferon Signaling Pathway, cellular response to zinc ions and negative growth regulation. Immune response, negative regulation of transcription from RNA polymerase II promotor, positive regulation of smooth muscle cell proliferation and cellular response to Dexamethasone stimulus, extracellular matrix, extracellular space, and extracellular region were the main enriched processes and pathways in these DEGs and many of the hub genes' (UBC, TP53, EGFR, GADPH, CREBBP, HDAC1, STAT1, IL6, ESR1, SMAD4) reported in this study were purposed as novel therapeutic targets against MDS disease.

Keywords: Myelodysplastic syndromes; Gene mutations; Protein-protein interaction network; Hub the genes; Sub-networks; DAVID

Introduction

A Myelodysplastic syndrome (MDS), also known as preleukemia, is a state of disease characterized by active but ineffective hematopoiesis leading to pancytopenia [1]. Fatigue, breath shortness, paleness, bleeding, and rashes are the general symptoms observed in patients with MDS and also known as myeloid neoplasms, described by different level of Cytopenia that is a different level of blood cells in the body. Association of Cytopenia with dysplasia usually led to acute myeloid leukemia. Epidemics of MDS are frequently reported in older individuals [2,3]. Many factors have been reported as causes of MDS including gene mutation that is broadly considered as a major factor contributing to MDS. Gene mutations result in genotypic alteration and thus, lead to cytogenetic shifts in gene expression. These cytogenetics shifts are usually characterized by abnormal transcription of the gene, epigenetic, cell signaling and effects of gene dosage. In addition, many frequent cytogenetic aberrations were observed including a long arm of chromosome 7, 20 and 5 that lead to a complex karyotype. In MDS, most commonly mutated genes are of RNA splicing regulators and epigenetic modifiers, along with pathways of signal transduction and transcription factors have been frequently targeted towards this syndrome [4-6]. MDS was reported to occur because of mutations in different genes primarily includes *SF3B1*, *SRSF2*, *ZRSR2*, *U2AF1*, *DNMT3A*, *EZH2*, *TP53*, *RUNX1*, and *TET2*. *SF3B1* and demonstrated independent expression with low mutation frequency, reduced expression of *TET2* in the stem and progenitor cells, and increased DNA methylation activity in MDS [7]. Mutations in *TET2* occurred with same frequency in MDSs [8]. Different mutations were identified

but a mutation in *SF3B1* gene was concluded as a most important factor. MDS epidemics due to a mutation in *SF3B1* gene contributed a total of 19.9% of all the reported cases. Patients with *SF3B1* gene were usually reported with more complications during the lifespan in comparison to the patient with MDS due to other risk factors. *SF3B1* gene was also documented with mutations in a variety of other tumor types [9]. According to IPSS (International Prognostic Scoring System), MDS was categorized in different risk groups, such as primary and secondary MDS [10]. Abnormalities were observed like clonal karyotype that formed nearly 40-50% of primary MDS and about 90% of secondary MDS [11]. To date, no inclusive treatment other than Azacitidine, Decitabine, and Lenalidomide is available in the markets and approved by Food and Drug Administration of United States (US) for MDS. However, allogenic therapy was reported as conclusive but currently, less than 10% patients undergo such stem cell transplant [12]. Pellagatti [13] detected several downregulated genes and gene pathways in MDS using gene expression profiling [13-18]. Recently Gerstung et al. [19]

***Corresponding author:** Dong-Qing Wei, State Key Laboratory of Microbial Metabolism, College of Life Sciences, Biotechnology, Shanghai Jiao Tong University, Shanghai, China, Tel: 862154740000; E-mail: dqwei@sjtu.edu.cn

Received March 10, 2018; **Accepted** March 26, 2018; **Published** March 30, 2018

Citation: Ali A, Junaid M, Khan A, Kaushik AC, Mehmood A, et al. (2018) Identification of Novel Therapeutic Targets in Myelodysplastic Syndrome Using Protein-Protein Interaction Approach and Neural Networks. J Comput Sci Syst Biol 11: 184-189. doi:10.4172/jcsb.1000270

Copyright: © 2018 Ali A, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

described a huge mutation in 738 patients with MDS and presented a comprehensive map for the mutational landscape of myelodysplasia screen in 111 cancer genes. Expression changes were typically lower than 10% as a reoccurrence of mutation and could not be reliably mapped in unknown and small subgroups. The genomics data reported by Gerstung et al. [19] characterized the gene expression profiles of 159 MDS patients by comparing with 17 normal individuals to explore the expression pattern of the genome in MDS affected individuals [19].

To date, the only agent with high efficacy such as Hypo ethylating agent (HMA) has been employed for the treatment that improved clinical outcomes in 40-60% of patient, however, no universal and inclusive drugs are till date available against MDS. Therefore, we used microarray gene expression data reported by Gerstung et al. [19] to discover the potential drug targets. This study follows the statistical test models along with genes enrichment and protein-protein interaction analysis to identify the novel drug targets as a treatment for MDS. This examination may facilitate and give a better understanding of the detailed molecular mechanisms underlying MDS and thus, assist in the selection of suitable and effective treatment strategies for patients with MDS.

Methods

Microarray data

Gene expression profiling data of MDS was download from GEO (Gene Expression Omnibus) [20] reported by Gerstung et al. [19], under accession number GSE58831 [19]. This dataset was based on Affymetrix Human Genome U133 plus 2.0 Array GPL570 platforms. A total of 176 samples were reported in this dataset, that includes 17 normal candidates and remaining 159 candidates carried different mutations for MDS. Quantile normalization [20] was carried out to normalize the dataset via integrated GCRMA package of R v3.0.2 [21,22]. On the basis of Gerstung et al. [19] reported information, data were divided into two groups as normal and diseased (Table 1).

Identification of DEGs

For the identification of differentially expressed genes, statistical tools were applying to compare the normal samples with the diseased samples to understand the molecular markers disturbed in MDS. Student's t-test, Pearson correlation test, and Benjamin-Hochberg method were also applied for multiple testing via R v 3.0.2 as a standardized method to identify the database of essential genes (DEGs). The parameters were fixed for identification of these essential genes contained Benjamin-Hochberg [22] multiple testing methods (FDR <0.05) with fold-change ≥ 1 and the adjusted p value <0.05.

Gene ontology and pathway enrichment analysis of DEGs

Gene ontology (GO) analysis such as GO Biological Processes, Molecular Function and KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways analysis are the most common and useful annotation of different Genes and its products. Attributes of high throughput genomics and transcriptomic data could be obtained through GO [23,24]. For the functional analysis, DAVID is an essential and most frequent functioning online server which can functionally annotate genes with high success [25]. Here, DAVID was used with a P value <0.05 to annotate the functional role, KEGG pathways and GO enrichment analysis of the identified DEGs [26].

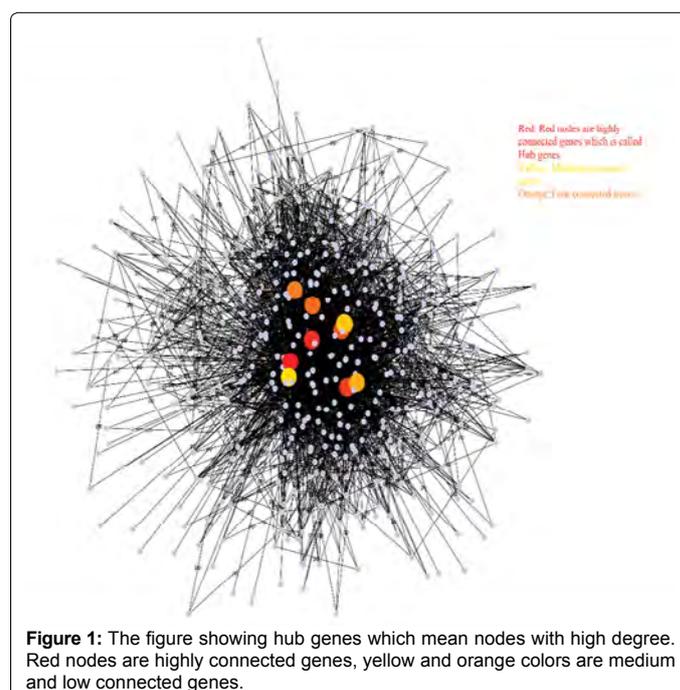
PPI Network Generation

Cytoscape [13] is reliable software for the construction, mapping,

visualization, and analysis of protein-protein interaction (PPI) networks. It works parallel with large databases which provides information regarding protein-protein, protein-DNA, and genetic interactions. STRING [27], BioGrid [28], GeneMANIA [29] for the retrieval of protein interactions. In the present study, we used GeneMANIA, STRING, and BioGrid to retrieve the interactions and construction of Protein-protein interaction network for the identified DEGs. Many topological parameters are available to analyze and compare the network. Cytoscape is freely available software which also provides an integrated function "Network Analyzer" to analyze the gene/protein network. We also used "Network Analyzer" to calculate the parameters for all the constructed networks. The primary parameters which were analyzed include power law of node distribution, distribution of node degree, clustering coefficient, network centralization and density to distinguish the three constructed networks [30].

Hub Genes Identification and Molecular Complex Detection Analysis of DEGs

A number of plugins are available for Cytoscape to perform different analyses. Identifying hub genes which can be employed as probable drug targets were identified using a well-known integrated plugin Cytohubba [31]. Cytohubba provides the eleven methods of topological analysis comprising degree, Edge Percolated Component, Maximum Neighborhood Component, Density of Maximum Neighborhood Component, Maximal Clique Centrality and six centralities (Bottleneck, Eccentricity, Closeness, Radiality, Betweenness, and Stress) based on shortest paths. It uses ranking features to rank different nodes in a network and based on their values hub genes are reported. "Molecular Complex Detection" (MCODE) is a novel clustering algorithm which identifies sub-modules, as shown in Figure 1, in large PPI networks. It allows fine-tuning of clusters of interest for protein networks. We used MCODE [32] along with Cytohubba to identify the interconnected dense sub-modules in the network. The hub genes and sub-modules were subjected to enrichment analysis again for subsequent verification using BinGO [33], an integrated app in Cytoscape.



Develop Algorithm for PPI using Neural Network Algorithms

Step 1: Normalization of dataset: Normalized dataset computed by $V_{new} = (V_{old} - \text{Min}V) / (\text{Max}V - \text{Min}V) * (D_{max} - D_{min}) + D_{min}$.

Step 2: Input the data for training, the interrelated values of input and output execute for training using neural network algorithms.

Step 3: Set network constraint.

Start

[1] Define the sample input and define buffer to store all the samples.

[2] Let the learning count=0;

[3] Let the learning count increase by 1;

[4] Training stage iteration begins.

[5] Input one sample and if the desired job output for the sample is a new process; there are no knowledge datasets under this process. Assign a new process and put new knowledge datasets under this process, Go to step 6.

[6] Stop and output training.

Step 4: Calculate the neurons of output, every neurons output signals calculated using $net_j = \sum_{i=1}^m w_{ji} x_i + b_j$ and sigmoidal function is making use of change net_j for every neuron of hidden layers.

Step 5: Signal of output layers calculation using $net_k = TV_k + \delta_k^L$. Where TV_k is target value of output neurons and δ_k^L is the error of neuron.

Step 6: Compute the error of neuron k until network is congregate and the error is computed using $SSE = \sum_{i=1}^n (T_i - Y_i)^2$. Where T_i is actual assessment and Y_i is estimated assessment.

Results

DEGs analysis and co-expression network

The complex biological system is composed of thousands of genes and its products. Genes and respective products interacted randomly and formed a more complicated network. The expression of these genes and respective proteins performed and conveyed much information such as signaling, transport, immunity, and defense and disease susceptibility. Microarray expression analysis may pose many regular variations. Herein, we used "Array Quality Metrics" and "GCRMA" to perform the normalization. Comparing two types of datasets, control and diseased, using statistical analysis; Student's t-test, Pearson correlation test and Benjamin-Hochberg methods, we identified a total of 585 DEGs, wherein the 224 and 361 genes were recorded as Up-regulator and down-regulator genes, respectively.

GO function and KEGG pathway enrichment analysis

GO analysis approach is the easy and accessible approach for the functional annotation of genomics data. Using DAVID database, we explored the functional changes in a patient with MDS. GO enrichment analysis was performed using the identified DEGs results into a diverse array of processes and functions as shown in Table 2. Biological processes of up regulated genes were mainly found in Interferon Signaling Pathway, Cellular Response to Zinc Ions and negative regulation of growth. Immune Response, negative regulation of transcription from

RNA polymerase II promotor, positive regulation of smooth muscle cell proliferation and cellular response to Dexamethasone stimulus were the enriched biological processes in down regulated DEGs. While only molecular function in down regulated DEGs described the transcriptional activator activity, RNA polymerase II core promotor proximal region sequence-specific binding with FDR 0.04. Furthermore, up regulated DEGs were found in extracellular matrix, extracellular space, and extracellular region part. The analysis of KEGG pathway enrichment revealed that down regulated DEGs were mainly involved in only pathway primary for immunodeficiency while up regulated genes were not enriched in any of KEGG pathways (Table 3).

Protein-Protein Interaction Network

A powerful tool to identify and understand the mechanism of cellular networking is protein-protein interaction network. It provides the basic understanding of both healthy and diseased conditions by understanding the distortion in protein cellular network. Nodes are proteins/genes while their interactions are edges in the protein-protein interaction networks. Herein, DEGs obtained from the analysis of microarray gene expression dataset of MDS were mapped and a protein interaction network was constructed.

A total of 585 DEGs were mapped onto protein-protein interaction network in Cytoscape. These DEGs were retrieved from available public databases such as GeneMANIA, STRING, BioGrid and hub genes were identified using a Cytoscape app, Cytohubba. Based on the degree of connectivity between nodes and hub genes were identified as well as ranked. The red nodes were highly connected nodes while the others were low and medium. Cytohubba was used to calculate the degree of each node in the network, that detected the hubs as nodes with degree value >10 (Table 4). The complete DEGs network was divided into a highly dense interconnected module using MCODE algorithm. A total of 9 modules (Sub-networks were identified) using the k-core value of 2.0, node score cut-off of 0.2, maximum depth from the seed node of 100 and graphics processing-unit-based parallelization was employed to find modules efficiently. Only sub-networks with a number of nodes greater than 15 were selected (Figure 2). Finally, functional enrichment of sub-networks was carried out. First, hub genes were checked in sub-networks and presence of hub genes in the sub-networks revealed that our results were reliable.

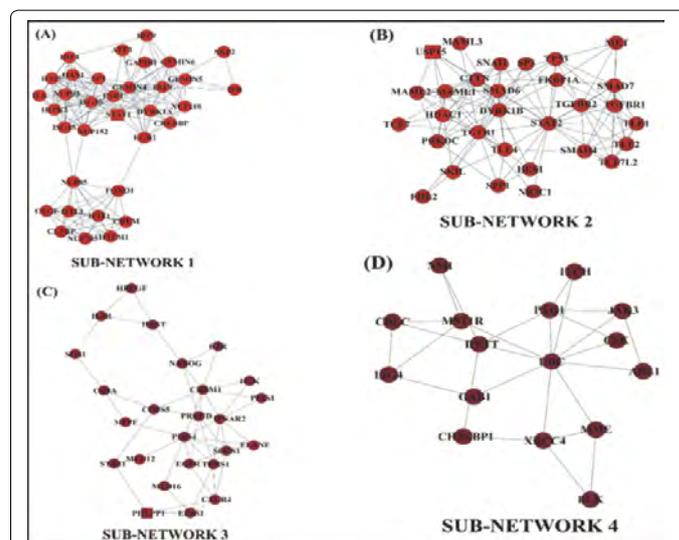


Figure 2: The depiction is illustrating the subnetworks identified in this study by using MCODE algorithm.

No.	Groups	Samples No	Condition
1	Group I	159	MDS Mutated
2	Group II	17	Normal

Table 1: Samples classification into groups.

Term	Category	P-Value	FDR	Description
GO-0006955	BP	1.50E-07	2.47E-04	Immune response.
GO-0000122	BP	8.78E-07	0.001442789	Negative regulation of transcription from RNA Polymerase II promoter.
GO-0048661	BP	3.59E-06	0.005902204	Positive regulation of smooth muscles cell proliferation.
GO-0071549	BP	1.33E-05	0.021784047	Cellular response to dexamethasone stimulus.
GO-0060337	BP	1.33E-06	0.002119021	Type I interferon Signaling pathway.
GO-0071294	BP	1.92E-05	0.030481573	Cellular response to Zinc ion.
GO-0045926	BP	1.92E-05	0.030481573	Negative regulation of growth.
GO-0005667	CC	2.73E-05	0.034948878	Transcription factor complex.
GO-0001077	MF	2.93E-05	0.040182076	Transcription activator activity, RNA polymerase II core promoter proximal region sequence-specific binding.

Table 2: The table is categorizing and describing the functional annotation of the DEGs identified in this analysis.

KEGG	P-Value	FDR	Description
KEGG-Pathway	4.97E-05	0.05222364	Primary Immunodeficiency

Table 3: Enriched KEGG pathways of DEGs with low P-values in samples from MDS.

Gene	Gene name	Degree	Go Term
UBC	Ubiquitin C	234	
TP53	Tumor protein 53	130	Negative regulation of transcription from RNA polymerase II promoter.
EGFR	Epidermal growth receptor factor	108	Positive regulation of smooth muscles cells Proliferation.
GADPH	Glyceraldehyde 3-phosphate dehydrogenase	86	Transcription activation, initiation of apoptosis.
CREBBP	Histone Deacetylase 1	82	Negative regulation of transcription from RNA polymerase II promoter.
HDAC1	Histone Deacetylase 1	76	Negative regulation of transcription from RNA polymerase II promoter.
STAT1	signal transducer and activator of transcription 1	66	Positive regulation of smooth muscles cells Proliferation
IL6	Interleukin 6	64	Positive regulation of smooth muscles cells Proliferation.
ESR1	Estrogen receptor 1	61	Cell signaling.
SMAD4	Mothers against decapentaplegic homolog 4	57	Cell signaling.

Table 4: Top 10 hub nodes identified in PPI network for DEGs from MDS patients and normal.

Figure 2 showing hub genes which mean nodes with high degree. Red nodes are highly connected genes, yellow and orange color are medium and low connected genes. Subnetworks 1,2,3 and 4 were enriched in GO terms related to chemical component extracellular region part, extracellular matrix. Subnetworks 1 and 2 were enriched in Immune response, Negative regulation of transcription from RNA Polymerase II promoter. Subnetwork 3 and 4 were enriched in the Positive regulation of smooth muscles cell proliferation, Cellular response to Zinc ion. P-values of all the enriched GO terms were in the range of 1.50E-07 to 8.78E-07 (Table 5 and Figure 1).

Interaction of Hub Nodes with Interaction

Calculate the neurons of output, every neurons output signals calculated using neural network algorithms where Table 6 indicate the best optimized proteins interactions.

Discussion

Protein-protein interaction network has become a powerful tool for identification of targets and analysis of different diseases. In the current

era, PPI network analysis has been widely utilized to understand the mechanism of different diseases, identifying drug targets and metabolic process. Analysis of gene expression dataset and identification of differentially expressed genes in a disease condition compared to the normal run a way of targeting different nodes for the discovery of novel drug candidates. Here, we used microarray gene expression dataset submitted to GEO under accession number GSE58831. Different statistical tools were used such as, Student's t-test, Pearson correlation, and Benjamin Hochberg multiple testing method (FDR<0.05 with a fold change>_1) and adjusted P-value 0.05) for the identification of DEG, that result into a total of 585 differentially expressed genes, in which 361 were downregulated and 224 were upregulated. Among the downregulated genes, RAG1 (recombination activating gene 1) was found to be the most downregulated one with a -4.69-fold change followed by MME (Membrane metallic-endopeptidase) and ARPP21 (cAMP-regulated phosphoprotein 21) with 444 Fold change and -4.43, respectively. Of the identified upregulated, DEGs HBG2///HBG1 (hemoglobin subunit gamma 2///hemoglobin subunit gamma 1), HBG2///HBG1 (hemoglobin subunit gamma 2///hemoglobin subunit gamma 1) and HBG2///HBG1 (hemoglobin subunit gamma

Sub Network	Score	Protein	Interaction
1	10.545	34	174
2	8.552	30	124
3	4.167	25	50
4	3.6	16	27

Table 5: Statistics for top five subnetworks identified by MCODE method in PPI network.

Hub node	Interaction	SSE Score
UBC	185	3.015
EGFR	167	3.004
TP53	145	2.054
STAT1	143	2.045
ESR1	127	1.055

Table 6: Interaction of Hub nodes with interaction.

2///hemoglobin subunit gamma 1) were the top with 265 Fold change followed by 2.37 and 2.33 respectively. Upon subjection to enrichment analysis, these DEGs revealed that they were involved in a diverse array of processes. Among them, most enriched GO biological processes were positive and negative of transcription from Immune response, Negative regulation of transcription from RNA Polymerase II promotor, Positive regulation of smooth muscles cell proliferation. Mapping of DEGs on Cytoscape, top 10 hub genes were identified using Cytohubba. For hub genes degree in Cytohubba was set as a parameter. Among the top hub genes UBC with highest degree 234, TP53, EGFR, GADPH, CREBBP, HDAC1, STAT1, IL6, ESR1 and SMAD4 with a degree 130,108,86,82,76,66,64,61 and 57 respectively. These hub genes were also found in sub-networks as shown in the table as a way of validating out the result.

The enrichment of hub genes given in Table 4 reports the same category of processes and thus validated our results. The Ubiquitin C (UBC) and Tumor protein (TP53) were found as “super hubs” genes. Zhang et al. [33] reported the Tumor protein 53(TP53) genes mutations occurring in patients with myelodysplastic syndrome (MDS) which was associated with abnormalities like a high risk of karyotype including 17p and complex cytogenetics. The existence of TP53 mutations in 17% of MDS was isolated with del(5q). TP53 was reported as a most frequently mutated gene with a complex karyotype [34,35]. Signal transducer and activator of transcription 1 play a vital role in body’s immune response. Herold et al. [36] reported that interleukin-6(IL-6) exerts positive effect and played a great role in infections, inflammations and possible involvement in leukemogenesis. Mothers against decapentaplegic homolog 4(SMAD4) gene was an important element of tumor suppressor as reported by Dolatshad et al. [36] that also showed differential Exon usage overlapping with other gene. Li et al. [37] documented the epidermal growth factor receptor (EGFR) gene represent somewhat the malignant proliferation and suppress apoptosis through an unknown mechanism in MDs because it overexpressed in most cases [37]. Furthermore, our finding revealed that a number of Hub genes were associated with proteins because they disturbed the progression and development of MDS, and provided the important route towards therapeutics development against MDS [38].

Conclusion

To examine the relationships between the importance of genes and several topological characteristics in human PPI network, system biology approach was used in the present study. There are several previously reported works with mutations and pathways associated with different gene mutation and pathways involved in its pathogenesis. Because of this, gene expression data of MDS patients and Healthy individual’s samples were used to identify the DEGs relation to MDS.

Different bioinformatics tools were utilized to discover the hub genes, enriched GO terms and KEGG pathways. GO terms related to adhesion, signaling were the main terms enriched by DEGs. UBC, TP53, EGFR, GADPH, CREBBP, HDAC1, STAT1, IL6, ESR1, and SMAD4 were identified as possible potential genes targets for MDS disease. However, further studies are required to determine the clinical utility of these observations in the therapeutic management of MDS related neurological disease.

Acknowledgements

Part of our computations was carried out at the High-Performance Computing Center of Shanghai Jiao Tong University.

Funding

This work is supported by the Key Research Area Grant 2016YFA0501703 from the Ministry of Science and Technology of China also grants from the state key Lab on Microbial Metabolism, and joint research fund for Medical and Engineering and Scientific research at Shanghai Jiao Tong University.

Authors’ Contributions

AA has made a contribution to the acquisition, analysis, and interpretation of the data and drafted the manuscript. MJ and AK are involved in the design of the study. DW is the academic supervisor and involved in the supervision of the study. We are also thankful to ACK for the critical evaluation of the manuscript. Aamir Mehmood, Shoaib Saleem and Asma Sindhoo Nangraj is involved in the Proof reading of the paper. All authors read and approved the final manuscript.

Competing Interests

The authors declare that there is no conflict of interest.

References

- Albitar M, Manshouri T, Shen Y, Liu D, Beran M, et al. (2002) Myelodysplastic syndrome is not merely “preleukemia”. *Blood* 100: 791-798.
- Cogle CR (2015) Incidence and burden of the myelodysplastic syndromes. *Curr Hematol Malig Rep* 10: 272-281.
- Fenaux P, Mufti GJ, Hellstrom-Lindberg E, Santini V, Finelli C, et al. (2009) Efficacy of azacitidine compared with that of conventional care regimens in the treatment of higher-risk myelodysplastic syndromes: a randomised, open-label, phase III study. *Lancet Oncol* 10: 223-232.
- Lindsley RC, Ebert BL (2013) Molecular pathophysiology of myelodysplastic syndromes. *Annu Rev Pathol* 8: 21-47.
- Bejar R, Levine R, Ebert BL (2011) Unraveling the molecular pathophysiology of myelodysplastic syndromes. *J Clin Oncol* 29: 504-515.
- Kulasekararaj AG, Mohamedali AM, Mufti GJ (2013) Recent advances in understanding the molecular pathogenesis of myelodysplastic syndromes. *Br J Haematol* 162: 587-605.
- Ganguly BB, Kadam NN (2016) Mutations of myelodysplastic syndromes (MDS): an update. *Mutat Res Rev Mutat Res* 769: 47-62.
- Rocquain J, Carbuccia N, Trouplin V, Raynaud S, Murati A, et al. (2010) Combined mutations of *asx1*, *cbl*, *flt3*, *idh1*, *idh2*, *jak2*, *kras*, *npm1*, *nras*, *runx1*, *tet2* and *wt1* genes in myelodysplastic syndromes and acute myeloid leukemias. *BMC Cancer* 10: 401-407.
- Papaemmanuil E, Cazzola M, Boultonwood J, Malcovati L, Vyas P, et al. (2011) Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts. *New England Journal of Medicine* 365: 1384-1395.
- Giagounidis AA, Germing U, Aul C (2007) Current treatment strategies in low-risk myelodysplastic syndromes. *Cancer Treatment Reviews* 33: 19-24.

11. Nimer SD (2006) Clinical management of myelodysplastic syndromes with interstitial deletion of chromosome 5q. *J Clin Oncol* 24: 2576-2582.
12. Steensma DP (2015) Myelodysplastic syndromes: diagnosis and treatment. *Mayo Clinic Proceedings* 90: 969-973.
13. Pellagatti A, Cazzola M, Giagounidis A, Perry J, Malcovati L (2010) Deregulated gene expression pathways in myelodysplastic syndrome hematopoietic stem cells. *Leukemia* 24: 756-764.
14. Mills KI, Kohlmann A, Williams PM, Wieczorek L, Liu WM (2009) Microarray-based classifiers and prognosis models identify subgroups with distinct clinical outcomes and high risk of AML transformation of myelodysplastic syndrome. *Blood* 114: 1063-1072.
15. Chen G, Zeng W, Miyazato A, Billings E, Maciejewski JP, et al. (2004) Distinctive gene expression profiles of CD34 cells from patients with myelodysplastic syndrome characterized by specific chromosomal abnormalities. *Blood* 104: 4210-4218.
16. Theilgaard-Mönch K, Boultonwood J, Ferrari S, Giannopoulos K, Hernandez-Rivas JM (2011) Gene expression profiling in MDS and AML: potential and future avenues. *Leukemia* 25: 909-920.
17. Pellagatti A, Cazzola M, Giagounidis AA, Malcovati L, Della Porta MG, et al. (2006) Gene expression profiles of CD34+ cells in myelodysplastic syndromes: involvement of interferon-stimulated genes and correlation to FAB subtype and karyotype. *Blood* 108: 337-345.
18. Sridhar K, Ross DT, Tibshirani R, Butte AJ, Greenberg PL (2009) Relationship of differential gene expression profiles in CD34+ myelodysplastic syndrome marrow cells to disease subtype and progression. *Blood* 114: 4847-4858.
19. Gerstung M, Pellagatti A, Malcovati L, Giagounidis A, Della Porta MG, et al. (2015) Combining gene mutation with gene expression data improves outcome prediction in myelodysplastic syndromes. *Nat Commun* 6: 5901-5910.
20. Edgar R, Domrachev M, Lash AE (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* 30: 207-210.
21. Hochberg Y, Benjamini Y (1990) More powerful procedures for multiple significance testing. *Stat Med* 9: 811-818.
22. Gene Ontology Consortium (2006) The gene ontology (GO) project in 2006. *Nucleic Acids Res* 34: 322-326.
23. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, et al. (2000) Gene Ontology: tool for the unification of biology. *Nat Genet* 25: 25-29.
24. Dennis G, Sherman BT, Hosack DA, Yang J, Gao W, et al. (2003) DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol* 4: R60.
25. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, et al. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 13: 2498-2504.
26. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, et al. (2014) STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res* 43: 447-452.
27. Stark C, Breitkreutz BJ, Reguly T, Boucher L, Breitkreutz A, et al. (2006) BioGRID: a general repository for interaction datasets. *Nucleic Acids Res* 34: 535-539.
28. Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, et al. (2010) The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res* 38: 214-220.
29. Assenov Y, Ramirez F, Schelhorn SE, Lengauer T, Albrecht M (2007) Computing topological parameters of biological networks. *Bioinformatics* 24: 282-284.
30. Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, et al. (2014) CytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Syst Biol* 8: 11-15.
31. Bader GD, Hogue CW (2003) An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics* 4: 2-8.
32. Maere S, Heymans K, Kuiper M (2005) BiNGO: A Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* 21: 3448-3449.
33. Kulasekararaj AG, Smith AE, Mian SA, Mohamedali AM, Krishnamurthy P, et al. (2013) TP53 mutations in myelodysplastic syndrome are strongly correlated with aberrations of chromosome 5, and correlate with adverse prognosis. *Br J Haematol* 160: 660-672.
34. Ok CY, Patel KP, Garcia-Manero G, Routbort MJ, Fu B, et al. (2015) Mutational profiling of therapy-related myelodysplastic syndromes and acute myeloid leukemia by next generation sequencing, a comparison with de novo diseases. *Leu Res* 39: 348-354.
35. Pellagatti A, Cazzola M, Giagounidis A, Perry J, Malcovati L, et al. (2010) Deregulated gene expression pathways in myelodysplastic syndrome hematopoietic stem cells. *Leukemia* 24: 756-764.
36. Dolatshad H, Pellagatti A, Fernandez-Mercado M, Yip BH, Malcovati L, et al. (2015) Disruption of SF3B1 results in deregulated expression and splicing of key genes and pathways in myelodysplastic syndrome hematopoietic stem and progenitor cells. *Leukemia* 29: 1092-1103.
37. Li X, Pu Q, Beckham C, Deeg JH (2003) Expression of epidermal growth factor receptor and its relationship to apoptosis in myelodysplastic syndromes. *Zhonghua Xue Ye Xue Za Zhi* 24: 22-24.
38. Shityakov S, Dandekar T, Förster C (2015) Gene expression profiles and protein-protein interaction network analysis in AIDS patients with HIV-associated encephalitis and dementia. *HIV/AIDS* 7: 265-276.