A Comprehensive Analysis Of Multiple Myeloma Genes Using Advanced Bioinformatics Tools

S Bulusu, A Kumar, S Parija, A Sinha

Abstract

Multiple myeloma is a monoclonal B-cell malignancy which originates in lymph node germinal centers but locates and expands in bone marrow. Multiple Myeloma (MM) occurs when plasma cells grow excessively and interfere with the production of red blood cells, white blood cells, and the platelets. Genetic abnormalities in multiple myeloma have been a subject of interest to researchers, as an understanding of the complex karyotypes can be of great value in treatment of MM.The aim of our study is to provide a comprehensive analysis of MM genes for understanding the cytogenetics of the disease. The present investigation is based upon the identification of over 35 important multiple myeloma genes located on different chromosomes.Various aspects of these genes were studied using different in-silico tools. An attempt was made to study the domains and patterns exhibited by the proteins and correlate them to the landmark events in the formation of multiple myelomas. A close scrutiny of the nucleotide sequences of these genes was conducted for identifying any internal repeats within the sequence and their possible significance as biomarkers.

INTRODUCTION

Multiple myeloma is a malignancy of the immunoglobulin-producing plasma cells. Excessive plasma cells form tumors called myelomas. Though it is a hematological malignancy like leukemia, the myeloma cells rarely enter the blood stream and instead accumulate in bone marrow. As the tumors grow in the bone marrow, they cause a disruption of normal bone marrow function, giving rise to anemia. Other complications of multiple myeloma include increased susceptibility to infection, renal failure and bone fractures. High levels of calcium in the blood are seen in about a quarter of myeloma patients as a consequence of the increased bone destruction releasing calcium into the bloodstream.

Cytogenetic aspects of Multiple myeloma have attracted the attention of researchers, as the karyotypes in MM are typically complex, in contrast to other hematologic malignancies. The genetic imbalances in multiple myeloma have been defined at molecular level and have shown that genomic changes may affect almost all chromosomes, as shown by extensive fluorescence in situ hybridization (FISH) and comparative genomic hybridization (CGH) 9.

Several recurring chromosomal abnormalities have been identified including structural rearrangements (e.g., translocations of 14q32), trisomies (e.g., chromosome 15), and monosomies (e.g., chromosome 13)8. The frequency and extent of karyotypic abnormalities correlates with the disease stage, duration, and response to treatment.13.

An analysis of the genes and the encoded proteins associated with multiple myeloma would be a necessary prerequisite towards a better understanding of the disease for an early diagnosis and more effective methods of treatment. The present study was taken up as a preliminary step and an attempt was made to analyze some important genes of MM.

Besides collecting protein and nucleotide sequence data from large public Databases, an attempt was made to use most advanced bioinformatics tools to perform sequence retrieval, phylogenetic analysis, domain and pattern recognition, protein-to-protein interactions, micro satellite detection and repeat analysis etc.

METHODOLOGY

After a survey of the available literature, over 35 important myeloma genes were identified for investigation.

A brief list of the major Databases and Bioinformatics tools employed during the study is given below.

Databases: GenBank, GENECARDS Swiss-Prot, HUGO.

Figure 1

In-Silico Tools	Purpose
BLAST	Protein Sequences
Entrez Gene	Nucleotide Sequences
SMART	Functional Domains
Biology work Bench, CLUSTALW	Phylogeny and Clustering of Proteins
SSRIT	Analysis of Tandem Repeats.
IBM Pattern Discovery	Patterns in Protein Sequences.

The expressed proteins were analyzed phylogenetically for their clustering. Each cluster was then examined with respect to the functional domains, which would help to authenticate the clustering obtained in the preliminary phylogenetic analysis. Special tools were employed for detecting protein-to-protein Interactions.

RESULTS & DISCUSSION

The following is the list of the identified genes along with their Ids and location on chromosome. The nucleotide and corresponding protein sequences were downloaded from NCBI and used for further analysis.

SDC1 (2p24.1) NP_002988; FGFR3 (4p16.3) AAA52450; MMSET (4p16.3) O96028; P18 (5q12.3) NP_776190; LTA (6p21.31) BAF31278; TNF (6p21.31) BAE78639; IRF4 (6p25-p23)NP_002451; IL6 (7p21) NP_000591; HGF (7q21) AAA52649; C-myc (8q24.21) NP_002458; TNFSF15 (9q32) NP_005109; FAS (10q23-q24.1) AAC16237; TNFRSF6 (10q24.1) NP_000034; PTEN (10q23) NP_000305; MYEOV (11q13) NP_620123;

CCND1 (11q13) NP_444284; BCL1 (11q13) NP_444284; TNFRSF1A (12p13) EAW88805; BCL2L14 (12p13-12p12) EAW88805; CDK4 (12q14) NP_000066; BCL7A (12q24.1) CAA62011; FLT 1 (13q12) NP_002010; RB1 (13q14.2) NP_000312; MYETS1 (13q14) AAN16377; FAM10A4 (13q14) Q8IZP2; LIG4 (13q22-q34) NP_001091738; TPP2 (13q32-33) NP_003282; TNFAIP (14q32) NP_006282; B2M (15q21-q22) NP_004039; P15 (16p13.3) AAK61226; CD19 (16p11.2) AAD02340; C-maf (16q23) NP_085132; TP53 (17p13.1) NP_000537; FASN (17q25) NP_004095; BCL5 (17q22) NP_114432; TNFSF12 (17p13) NP_742086.

PHYLOGENETIC ANALYSISS

Figure 2
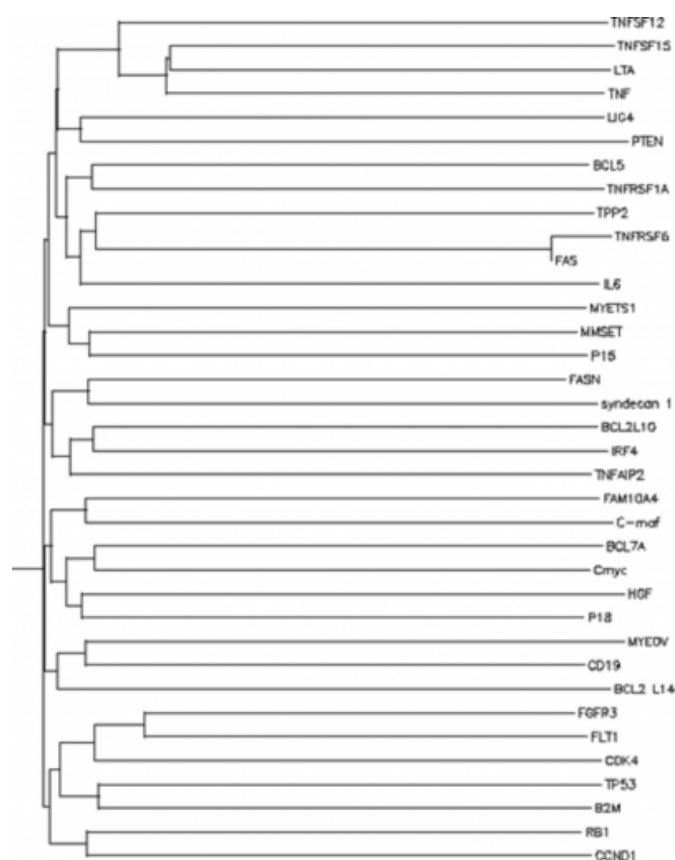
It is evident from the dendrogram that there is a grouping of the genes in accordance with the proteins encoded by them. From the ClustalW dendrogram, it is clear that the genes are grouped into 3 categories. Further for convenience of analysis, we grouped them into 7 clusters. As the different domains present in the sequence determine the functions of the proteins, a thorough analysis of the functional domains was performed to authenticate the dendrogram

DOMAIN ANALYSIS

Maximum number of genes falls into the first category which comprises of 4 clusters, organized into 2 branches. Clusters 1, 2 & 3, which are located on one branch, predominantly show the presence of Tumor Necrosis Factor (TNF) and related domains. TNF-alpha, a family of TNF can stimulate cell proliferation and induce cell differentiation under certain conditions. Death domains present along with TNFRSF have a role in apoptosis and various other domains present in genes MYETS1, MMSET, FAS etc. play important roles in gene transcription, translation, mRNA trafficking, cytoskeleton organization, epithelial development, cell adhesion, protein folding, chromatin remodeling and zinc sensing, to mention but a few.

Cluster 4 consists of 5 genes whose domains are in varied pattern without much uniformity. FASN has a number of fatty acid synthetase domains and there are records that cancer cells frequently exhibit a significant increase in over-expression and activity of fatty acid synthase (FASN). The transcription factor IRF4 is required for the generation of plasma cells. Available data suggest that IRF-4, by modulating the efficiency of the Fas-mediated death signal, is a novel participant in the regulation of lymphoid cell apoptosis.

II. In the second category, there are two clusters (Clusters 5 & 6), distinctly on two branches.

Genes of Cluster 5 mainly consist of domains like PAN_AP, KR domains, ZnF_C2HC domain and Ring finger and B-Box-type zinc finger with different binding functions. These domains are believed to play a role in binding mediators (e.g., membranes, other proteins or phospholipids). Zinc-binding motifs are stable structures, and they rarely undergo conformational changes upon binding their target. Znf-containing proteins function in gene transcription, translation, mRNA trafficking, cytoskeleton organization, cell adhesion, protein folding, chromatin remodeling etc. Besides, there are myc-n, Ig and BCL domains that influence cell proliferation and apoptosis.

Cluster 6 is a small cluster with 3 genes namely MYEOV, CD9, BCL2L14 all of which, when dys-regulated promote mitotic progression and prevent apoptosis.

III. In the third category (cluster 7), there is a predominance of IG domains in combination with Cyclin domains and tumor suppressor related domains.

IG domains: The Ig monomer is a "Y"-shaped molecule that consists of two identical heavy chains and two identical light chains connected by disulfide bonds. Each chain is composed of structural domains called Ig domains.

Under normal conditions the immune response usually activates many B cell clones resulting in heterogeneous antibodies. On the contrary B cell tumors provide source of homogenous antibodies. As such, in Multiple Myeloma, cancer of B cells outgrows other cells of the marrow and results in homogenous antibodies (monoclonal antibodies).

Cyclin domains: Cyclins are a family of proteins involved in the progression of cells through the cell cycle and regulation of cyclin dependent kinases (CDKs). When dysregulated expression of cyclin D1, D2, or D3 occurs in tumors, it may render the cells more susceptible to proliferative stimuli.

In addition to the Ig and cyclin domains, there are other domains in this gene cluster like TyrKcTyrosine kinase, catalytic domain and P53 domains which play a role in a multitude of cellular processes, including division, proliferation, apoptosis and differentiation.

Patterns exhibited by the protein sequences and their significance:

In the Myeloma genes analyzed in the present study for patterns, we could obtain interesting correlations. The patterns detected were present in many other diseases, especially cancers indicating that the presence of such patterns can serve as biomarkers for detecting cancers.

The following is the list of pattern exhibited by genes and corresponding diseases

LTA- multiple aneoplasia type IIA (MEN2A) LIG4-immunodeficiency diseases PTEN- Fanconi anemia BCL5-breast cancer, Fanconi anemia IL6- systemic juvenile rheumatoid arthritis MMSET- Other tumor suppressor genes. FASN- Huntington disease, myelodysplastic syndrome (MDS), squamous cell carcinoma antigen TNFAIP2- childhood pre-B acute lymphoblastic leukemia, small cell lung carcinomas BCL7A- B-cell non-Hodgkin lymphoma, ovarian cancer, breast, lung, colon, pancreatic and ovarian cancers C-myc- acute leukemias, oncogenesis, a variety of hematopoietic tumors HGF- Overexpressed in tumor cells lines FGFR3- Bladder cancer, multiple Myeloma skeletal dysplasia, colorectal tumors. TP53 - mutated or inactivated in about 60% of cancers.

Protein-to-protein associations and functional partners:

Information about interactions that take place among the

important disease genes and their so called functional partners improves the overall understanding of diseases and can provide a basis for new therapeutic approaches. These functional partners could very well be the targets of new therapies too, where therapies face short comings in targeting the actual disease proteins.

In the present investigation the functional partners of each gene were detected by special tools and the interactions that exist between them are studied in order to detect their roles in MM. For e.g. CCND1 gene encodes for cyclin proteins important for a temporal coordination of each mitotic event. These cyclins forms a complex with CDK4 or CDK6 whose activity is required for cell cycle G1/S transition. CCND1 protein is also shown to interact with tumor suppressor protein RB and is regulated positively by it. Mutations, amplification and overexpression of CCND1 which alters cell cycle progression may contribute to tumorigenesis.

Figure 3
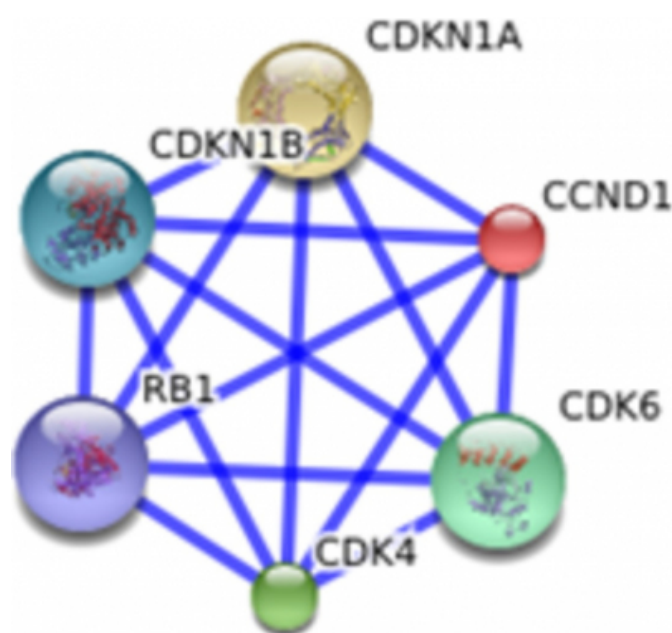
Further, it is interesting to note that not only most of the Myeloma genes had their own unique functional partners but also they were interacting with other myeloma genes. For e.g., MYEOV gene, RB1, P15, MMSET and CDK4 have CCND1, as their functional partner. Similarly, MMSET also acts in association with FGFR3. However, for 6 genes in the present study, there were no partners recorded in the set parameters.

Detection of Tandem repeats

At the early stage of development of some cancers, microsatellites are believed to change their length. Such microsatellites are useful as markers for early cancer detection. Mutations in DNA-repair genes cause microsatellites to get longer or shorter, a phenomenon called microsatellite instability (MSI)12. MSI is associated with defective DNA mismatch repair in various human malignancies13. In view of their role in detection of diseases, a cluster-wise analysis of Simple Sequence Repeats of the myeloma genes was taken up.

Results reveal that all the genes except CD19 show the presence of tandem repeats as di-, tri-, tetra- and/or pentamers. Frequencies are represented in the form of graphs. It is evident from the graphs that dimers were more frequent in occurrence.

Figure 4
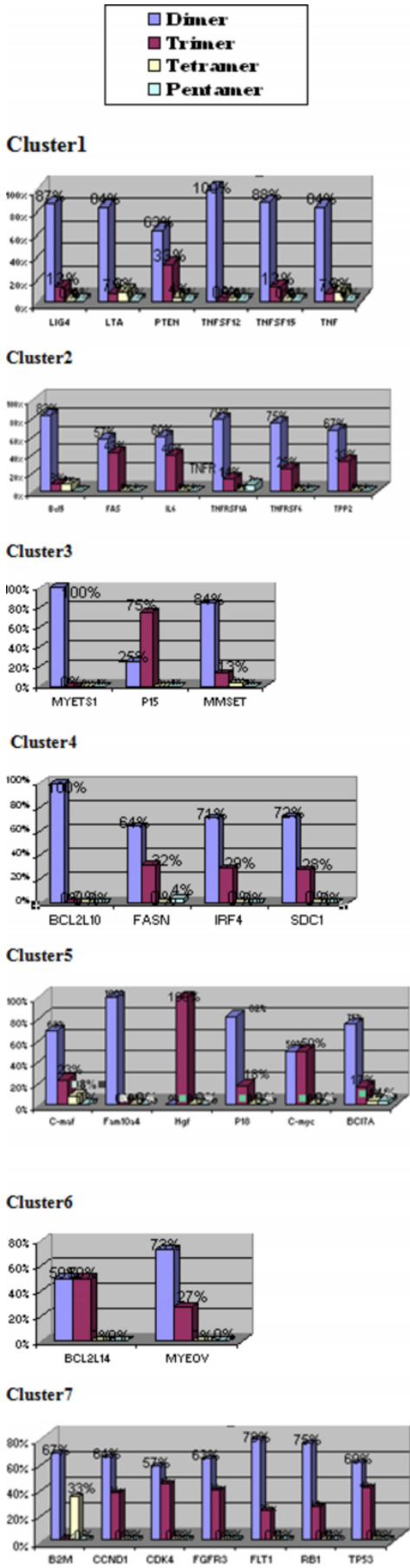
SSR effects in coding regions on phenotypes have been extensively studied only in human diseases, revealing abundant evidence in cancers and neuronal disorders 16.

Analysis of available DNA sequences of human, mouse, worm (Caenorhabditis elegans), and yeast genomes shows that the distribution functions of all possible dimeric SSRs are exponential in coding DNA6. The potential size expansion of di- or tetranucleotide SSRs at the 3' and 5' regions and at introns could lead to disruption of the original protein and/or formation of new genes by frame shift2, 11.

Dinucleotide repeated sequences are preferential sites for recombination because of their high affinity for recombination enzymes4. Some SSR sequences may influence recombination directly by their effects on DNA structure.

SSR can affect enzymes controlling cell cycles. For instance, the human CHK1 gene has a role in controlling cell cycle progression, and its coding region contains an (A)9 tract5 that is a potential site of mutations in tumors with SSR instability3. Alterations in the CHK1 gene in human colon and endometrial cancers were associated with the presence of a high degree of poly (A) tract instability.

About 15% of sporadic colorectal cancers, as well as cancers at several other sites, show SSR instability1. The progressive accumulation of SSR instability may also contribute to gastric cancer development10. It has been shown that 14 neurological disorders result from the expansion of unstable trinucleotide repeats.

The demonstrated widespread non randomness, selectivity, and other various patterns displayed by repeated elements, call for special attention at all stages of using this class of markers in all the fields, from designing of the experiments, to data analysis and interpretation15.

During the further analysis of the tandem repeats we have come up with interesting observations like the excessive presence of dimers in connection with the occurrence of recombination events that take place during the progress of multiple myeloma.

CONCLUSION

The present study on Multiple myeloma was taken up to provide a comprehensive understanding of the cytogenetics of the disease with respect to the various aspects of the genes involved.

Besides collecting protein and nucleotide sequence data from large public Databases, an attempt was made to use most advanced bioinformatics tools to perform sequence retrieval, phylogenetic analysis, domain and pattern recognition, protein-to-protein interactions, microsatellite detection and repeat analysis etc.

The genes in phylogenetic analysis showed clustering in accordance with their domains. A good correlation was observed between the different myeloma genes and their functional partners with respect to the pathogenecity of the disease. The tandem repeats detected in the nucleotide sequences showed relevance to the chromosomal recombinations which are the important events of cytogenetics of MM.

The possibility of comparing our data with other better-known cancers is also being looked into. It is proposed to create a comprehensive Database ultimately covering various aspects of multiple myeloma in near future.

Though Multiple Myeloma is a serious malignancy, it is treatable. With new treatments the average survival of 5 years for patients diagnosed with Multiple Myeloma may be further extended 7.

ACKNOWLEDGEMENT

References

1. Atkin NB (2001) Microsatellite instability. Cytogenetics and Cell Genetics, 92, 177– 181. CrossRef, Medline, ISI, Chemport
2. Bachtrog et al. 1999; --Bachtrog D, Weiss S, Zangerl B, Brem G, Schlötterer C (1999) Distribution of dinucleotide microsatellites in the Drosophila melanogaster genome. Molecular Biology and Evolutionary, 16, 602– 610. Medline, ISI
3. Bertoni F, Codegoni AM, Furlan D, Tibiletti MG, Capella C, Broggini M (1999) CHK1 frameshift mutations in genetically unstable colorectal and endometrial cancers. Genes Chromosomes Cancer, 26, 176– 180. CrossRef, Medline, ISI, Chemport.
4. Biet et al. 1999--Biet E, SunJ,Dutreix M (1999) Conserved sequence preference in DNA binding among recombination proteins: an effect of ssDNA secondary structure. Nucleic Acids Research, 27, 596– 600. CrossRef, Medline, ISI
5. Codegoni et al. 1999--Codegoni AM, Bertoni F, Collela G et al. (1999) Microsatellite instability and frameshift mutations in genes involved in cell cycle progression or apoptosis in ovarian cancer. Oncology Research, 11, 297– 301. Medline, ISI, Chemport
6. Dokholyan et al. 2000--Dokholyan NV, Buldyrev SV,

Havlin S and Stanley HE (2000) Distributions of dimeric tandem repeats in noncoding and coding DNA sequences. Journal of Theoretical Biology, 202, 273– 282. CrossRef, Medline, ISI, CSA

7. International Myeloma Foundation Booklet, USA, 2007

8. John Shaughnessy Jr, Erming Tian, Jeffrey Sawyer, Klaus Bumm, Reid Landes, Ashraf Badros, Christopher Morris, Guido Tricot, Joshua Epstein, and Bart Barlogie High incidence of chromosome 13 deletion in multiple myeloma detected by multiprobe interphase FISH Blood, 15 August 2000, Vol. 96, No. 4, pp. 1505-1511

9. Largo C, Saéz B, Alvarez S, Suela J, Ferreira B, Blesa D, Prosper F, Calasanz MJ, Cigudosa JC. Molecular Cytogenetics Group, Centro Nacional de Investigaciones Oncológicas, Madrid, Spain. Haematologica. 2007 Jun;92(6):795-802

10. Leung et al. 2000-- Leung WK, Kim JJ, Kim JG, Graham DY, Sepulveda AR (2000) Microsatellite instability in gastric intestinal metaplasia in patients with and without gastric cancer. American Journal of Pathology, 156, 537– 543. Medline, ISI

11. Liu Z, Tan G, Li P, Dunham RA (1999) Transcribed dinucleotide microsatellites and their associated genes from channel catfish Ictalurus punctatus. Biochemical and Biophysical Research Communication, 259, 190– 194. CrossRef, Medline, ISI, CSA

12. Miriam Komaromy -Genetic Health 2000

13. Pratt, G 2002 Molecular aspects of multiple myeloma. Molecular Pathology 2002;55:273-283

14. Shinya Oda etal 2005- Two modes of microsatellite instability in human cancer: differential connection of defective DNA mismatch repair to dinucleotide repeat instability. Nucleic Acids Res. 2005; 33(5): 1628–1636

15. YOU-Chun Li, Abraham B. Korol, Tzion Fahima, Avigdor Beiles and Eviatar Nevo 2002- Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review-- Molecular Ecology Volume 11 Issue 12 Page 2453-2465, December 2002

16. You-Chun Li,*, Abraham B. Korol , Tzion Fahima and Eviatar Nevo 2004 Microsatellites Within Genes: Structure, Function, and EvolutioMol. Biol. Evol. 21(6):991-1007

Author Information

Shakuntala Bulusu

Dept of Biotechnology and Molecular Biology, BioAxis DNA Research Centre

Amit Kumar

Department of Bioinformatics and Structural Biology, BioAxis DNA Research Centre

Satyajeet Parija

Department of Bioinformatics, BioAxis DNA Research Centre

Amita Sinha

Dept of Biotechnology and Molecular Biology, BioAxis DNA Research Centre