Int J Pharm Bio Sci 2013 Oct; 4(4): (B) 504 - 512

Research Article **Bio informatics**

International Journal of Pharma and Bio Sciences **ISSN 0975-6299**

COMPUTATIONAL ANALYSIS OF DISEASE ASSOCIATED nsSNPs in MMP2

H. JEMMY CHRISTY *

Department of Bioinformatics,Sathyabama University, Chennai-119, Tamilnadu, India.

ABSTRACT

MMPs have significant role in pathological processes and are associated with cancer invasion, and metastasis, cartilage destruction in arthritis. Non-synonymous single nucleotide polymorphisms (nsSNPs) occurring in protein coding regions of MMP potentially affect their function, and results in diseases. In this study we performed an insilco analysis of SNPs associated with MMP2 genes. Total of 683 SNPs are reported for MMP2 gene, coding region. In that only 39 are nsSNPs and 46 SNPs are in UTRs, 277 SNPs are in intronic region of mRNA. Our postulation is based on nsSNPs in the coding region as well as SNPs in the UTR region. Only 8SNPs out of 39nsSNP were identified as deleterious and had significant impact in structural stability which was analyzed using modeled and mutant proteins and native protein structures. Hence, these SNPs can be used for the larger population based studies based on MMP2 associated diseases.

KEYWORDS: **FastSNP, I-Mutant, Matrixmettalloproteinase2, PolyPhen, SIFT, SNP, F-SNP.**

H.JEMMY CHRISTY
Department of Bioinformatics,Sathyabama University, Chennai-119, Tamilnadu, India.

INTRODUCTION

MMP2s play a vital role physiological and pathological processes in human like angiogenesis, cell proliferation, apoptosis, alteration of cell motility, effects on the immune system and host defense, modulation of the bioactivity of chemokines [1,2] connective tissue remodeling associated with cancer invasion and metastasis, cartilage destruction in arthritis, atherosclerotic plaque rupture, and the development of aneurysms [3] . Clinical case studies revealed the association between circulating levels of MMP-2 and cardiac dysfunction in patients with hypertrophic cardiomyopathy[4] ischemic cardiomyopathy[5], and with congestive heart failure[6, 7]. Increased MMP-2 levels also associated with left ventricular remodeling after myocardial infarction [8]and predict poor outcome in patients with congestive heart failure [9]. SNP in the Matrix metalloproteinase 2 gene involved in multi centric osteolysis and arthritis syndrome.[10] and colorectal cancer[11]. TypeIV collagenase (72 KD) is officially designated asMatrixmetallopeptidase 2 (MMP2), and also known as gelatinase [12]. AccordingtoNCBI (http://www.ncbi.nlm.nih.gov/sites/entrez),
MMP2 is located on chromosome 16 and the protein has 660 amino acid residues. Loss of MMP activity through mutations and single nucleotide polymorphisms leads to various diseasessothatMMP2sgenes polymorphisms studies will explore their implementation in common complex diseases and traits such as arthritis, cancer and cardiovascular diseases. In this study we performed computational analysis of the SNP's in the MMP2 gene to identify the deleterious nsSNP that are likely to affect the function. Those SNP's can be used as biomarkers diseases and thus facilitate early diagnosis in high risk patients.

MATERIALS AND METHODS

Datasource and selection of nsSNP

A total of 683 found to be associated with MMP2 gene were retrieved from the dbSNP[13] and "PDB" file for MMP2 Protein were retrieved from the RCSB Protein Data Base[14] for computational analysis in this study.

Analysis of coding nsSNPs using a Sequence Homology Tool (SIFT)

The Sorting Intolerant from Tolerant (SIFT) server available at (http://sift.jcvi.org) [15] was used to predict the deleterious coding non-synonymous SNPs. The algorithm is based on a modified version of PSI BLAST and Dirichlet mixture regularization to construct a multiple sequence alignment of proteins that can be globally aligned to the query sequence and belong to the same clade[16,17]. The basic principle of this program is that it generates alignments with a large number of homologous sequences and assigns scores to each residue, ranging from zero to one. SIFT scores were classified as intolerant (0.00–0.05), potentially intolerant (0.051–0.10), borderline (0.101–0.20), or tolerant (0.201–1.00)[18]. Lower the tolerance index, more functional impact on a particular amino acid substitution likely to have. The analysis was performed by using MMP2 amino acid sequence (NP_001121363).

Analysis of coding nsSNPs using a Structural Homology-Based Method (PolyPhen)

PolyPhen[19] is a computational tool available athttp://coot.embl.de/PolyPhenfor identification of potentially functional nsSNPs. Predictions are based on a combination of phylogenetic, structural, and sequence annotation information characterizing a substitution and its position in the protein[20].The online input form was filled with the MMP2 amino acid sequence in FASTA format, and the position and substitution of each of the 39 nsSNPs analyzed by SIFT were also submitted for PolyPhen analysis. PolyPhen then searched for 3D protein structures, multiple alignments of homologous sequences and amino acid contact information in several protein structure databases, calculated position-specific independent counts (PSIC) scores for each of the two amino acid residues entered (the original residue and the nsSNP, and then computed the PSIC scores difference of the two

residues. PolyPhen scores were classified as "benign" or "probably damaging" [21]. The higher a PSIC score difference, the higher functional impact a particular amino acid substitution is likely to have. PolyPhen scores of >2.0 are expected to be "probably damaging" to protein structure and function, and PolyPhen scores of 1.99–1.50 are expected to be "possibly damaging" to protein function [20]. The query options were left with default values.

Analysis of deleterious nsSNP based on integrated method by F-SNP

F-SNP database provides integrated information about the functional effects of SNPs[21]. Each SNP's meant for MMP2 was examined for deleterious effects with respect to each functional category (i.e., protein coding, splicing regulation, transcriptional regulation, and post-translation.) and listed in the Table 1.

Functional significance of SNP's in untranslated regions

The polymorphisms in the 3′ UTR affect gene expression during translation of mRNA while the polymorphisms in the 5′ UTR influence RNA half-life by altering polyadenylation[22,23]. Hence, the UTRs were also analyzed for their functional SNPs. FastSNP prioritizes SNPs according to twelve phenotypic risks and putative functional effects, such as changes to the transcriptional levels and pre-mRNA splicing and protein structure. Input of the candidate gene symbol (MMP2) was used for analysis. The SNP prioritization result was a list of SNPs with its risk ranking and possible function types. Risk level is ranked as 0, 1, 2, 3, 4 or 5, which signify the levels of "no risk", "very low risk," "low risk," "medium risk," "high risk," and "very high risk," respectively[24]. The FastSNP search was performed by querying by gene symbol (MMP2) to lists the SNPs in the UTRs region.

Modeling of protein structure amino acid substitutions caused by nsSNPs, energy minimization and calculating the RMSD

(A) Finding the closest related protein

The EMBL-EBI Web-based tool PDBsum was used to find the proteins related to the MMP2 gene[25]. The closest matched structure for MMP2 protein 1CK7 was selected for further structural analysis of mutant and wild type SNP'S in MMP2 Protein.

(B) Modeling amino acid substitution, energy minimization and RMSD calculation

Pymol (v4.04) was used to generate the mutated models of each of the selected PDB entry 1CK7 for the corresponding amino acid substitutions [26]. Pymol allows browsing through a rotamer library to change amino acids. A "mutagenesis tool" was used to replace the native amino acid with a new one. The mutation tool facilitates the replacement of the native amino acid by the "best" rotamer of the new amino acid. The ".pdb" files were saved for all the models. Discovery studio (DS) server was used to perform energy minimization for all the native and mutated models of 1CK7. DS-energy minimization makes use of quantum mechanics force fields for energy minimization according to conjugate gradient method .RMSD's between the native structure and each mutant were calculated using DS Server.

Predicting the change in stability due to mutation

To predict the change in the stability of the protein upon mutation, a support vector machine (SVM)-based tool server, I-Mutant 2.0 used [27]. This tool automatically predicts protein stability changes upon single point mutations. Prediction can be performed using either protein structure or sequence. I-Mutant 2.0 can be used both as a classifier for predicting the sign of the protein stability change upon mutation and as a regression estimator for predicting the related change in Gibbs-freeenergy ($\Delta\Delta G$)[28].nsSNPs(Deleterious) position in functional domains of MMP2. To find the nsSNPs and the amino acid changes they may cause in different domains of the protein structures, the Prosit-ExPaSy tool[30]was used. The UniProtKB accession number P08253 was provided for the query column and the UniProt database was searched for motifs and domains of MMP2. The results were obtained as the categorized sequence of amino acids with their respective positions in the protein subsequences and domains.

RESULTS AND DISCUSSION

SNP Dataset

At dbSNP, MMP2 gene contains data for 683 SNPs. out of 683 SNPs, 76 SNP's are in the coding region.In that only 39 are nsSNPs and46 SNP's are in UTRs. There are 7 SNPs in the 5′ UTR, 24 SNPs in the 3′ UTR and 277 SNPs intronic region of mRNA. Our postulation is based on nsSNPs in the coding region as well as SNPs in the 5′ and 3′ UTR and region.

Functional Analysis of the Coding nsSNPs and their impact on MMP2gene using a Sequence Homology Tool (SIFT)

Protein sequence of MMP2 and their corresponding mutational positions as well as amino acid residue variants for the 39 nsSNP were used for analysis. According to Ng and Henikoff [18] only 8 nsSNPs were identified to be deleterious with a tolerance index score $\leq$ 0.05and the results were listed in (Table.1). In that only three nsSNPs like rs11542001, rs111609606, and rs121908741 were showed a higher score for deleterious SNP prediction. Theother 5 nsSNPs like rs17859943, rs28730814,rs59727333,rs112710941, rs16955280 showed a tolerance index of(0.01-0.04).

Analysis of the Functional Impact of coding nsSNPs using a Structural Homology-Based Method (PolyPhen)

The structural level of alterations were analysed using Polyphen program.39 nsSNPs were used as Polyphen and their PSIC Score ranging from 0.114-1.775.These 8 nsSNPs were also predicted as deleterious in SIFT analysis.Out of these 8 deleterious nsSNPs one changed from aromatic polar uncharged to aliphatic nonpolar in mutant type (F->L), two were unchanged,then one from aliphatic nonpolar to positive charge(G->D) ,two polar charged sulphur containing amino acid to aliphatic nonpolar(M->I) and aromatic polar uncharged(C->F) among the mutant type. In high scored nsSNP rs11542001 wild type aromatic polar uncharged amino acid changed in to aliphatic nonpolar in mutant type. Polymorphism in sulphur containing amino acid also significant impact in structural stability

Analysis of nsSNP using F-SNP

F-SNP assesses the deleterious effect of SNPs by calculating a specific functional significance (FS) score for each nsSNP. The deleterious SNP has a FS score value between 0.5 and 1. Among all five nsSNPs, five of them (rs11542001, rs17859943, rs59727333, rs111609606 and rs16955280) are found to have good significant FS scores of 0.518, 0.749, 0.617 and 0.524 and 0.719 respectively and can be considered to be damaging. They are found to be deleterious by having changes in the protein coding region. Putative ESEs (Exonic splicing enhancers) are also predicted for these nsSNPs by the change in splicing regulation region.These nsSNPs affect the activity of ESE's and there by stimulating the splicing from the exon's location.

Table 1

List of ns SNP's analysed by SIFT, POLYPHEN & F-SNP Methods.

SNP	Aminoacid change	SIFT &PolyPhen prediction	SNP Effect	ESEfinder
rs11542001	F239L	Damaging	Secondary structure changed	Changed
rs17859943	A447V	Damaging	No Change	Changed
rs59727333	K359I	Damaging	Amyloid propensity changed	Changed
rs111609606	M282F	Damaging	Amyloid propensity changed, Increases the aggregation tendency of protein	Changed
rs112710941	C65H	Benign	No Change	None
rs28730814	D450H	Benign	No Change	None
rs16955280	V621L	Damaging	Amyloid propensity changed	Changed
rs121908741	G357D	Damaging	No Change	Changed

Analysis of functional SNP in untranslated region (UTR) by FastSNP

According to FastSNP there were 7 SNPs in 5'UTR region,but only three SNPs namely predicted to be damaging with a risk ranking of 1-3,and their possible functional effect on promoter region were listed in (Table.2).There is no functional significant SNPs in 3'UTR.256 SNPs located in intronic region of mRNA ,2SNPs namely rs17859942, rs243834 show significant role in splicesite variation whereas119 SNPs with score(1-2)which will less impact on protein structure and function.

Table 2
Deleterious SNP's in the UTR region of MMP2.

SNP ID	Nucleotide change	UTR/Intronic region	Possible functional effect	Level of risk
rs17859829	G/T	5'UTR	Promoter/regulatory region	Very Low-Medium(1-3)
rs17859831	G/T	5'UTR	Promoter/regulatory region	Very Low-Medium(1-3)
rs17859832	C/T	5'UTR	Promoter/regulatory region	Very Low-Medium(1-3)
rs17859942	T/C	Intron	Splice site	Medium-High(3-4)
rs243834	A/G	Intron	Splice site	Medium-High(3-4)

Structural analysis of Mutant proteins

Predicted deleterious nsSNP in the coding region were mapped to the PDB id 1CK7 native structure. Amino acid substitution analysis for the mutant types were performed by Pymol mutagenesis for the selected 5nsSNPs (rs11542001, rs111609606, rs121908741).Then energy minimization for all the mutant models and their native structure was performed by DS-server's minimization. The total energy for all the mutant and native models after energy minimization were tabulated in table3. The total energy for native structure of 1CK7 was -45290.997. Mutation implementation in 1CK7 is noticeable in the 1CK7 mutant structures for the nsSNP's with the RMSD value ranging from of 0.80-0.71.RMSD value deviation between the mutant and native strongly influence the functional activity. RMSD values for the mutant structures were listed in table3

Table 3
Energy minimization values for the wild and mutant type MMP2 protein structures with respective RMSD values.

SNP id's	RMSD	Energy minimization(kj/mol)
1CK7 native-type structure		**-45290.997**
1CK7 Mutant F239L(rs11542001)	0.74	-41442.653.
1CK7 Mutant A447V(rs17859943)	0.80	-41497.934
1CK7 Mutant K359I(rs59727333)	0.79	-41458.655
1CK7 Mutant M282F(rs111609606)	0.72	-41526.371
1CK7 Mutant C65H(rs112710941)	0.76	-41475.209
1CK7 Mutant G357D(rs121908741)	0.77	-41381.454
1CK7 Mutant D450H(rs28730814)	0.71	-41118.454
1CK7 Mutant V621L(rs16955280)	0.78	-41534.454

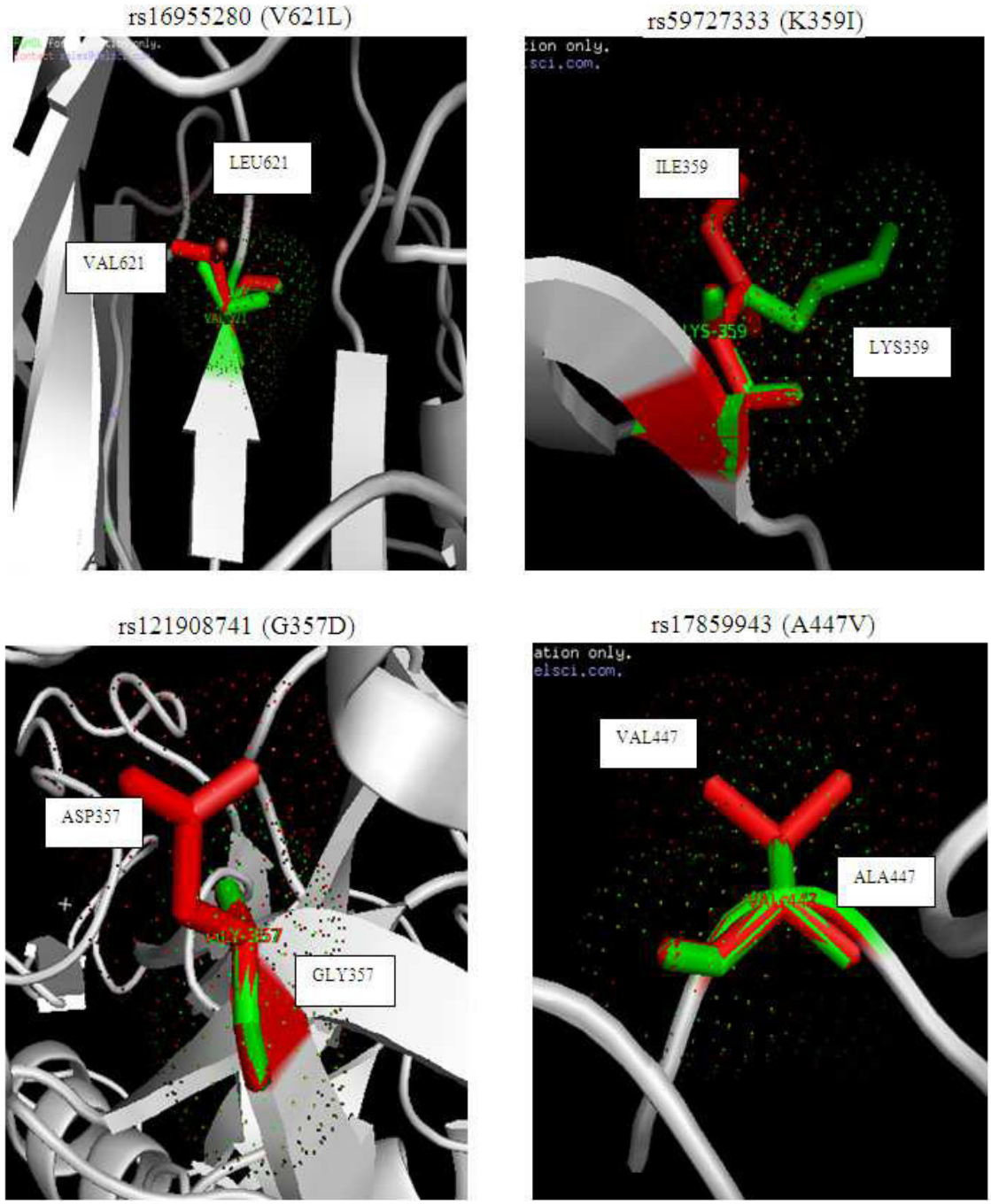
Changes in MMP2 protein due to ns deleterious SNP's

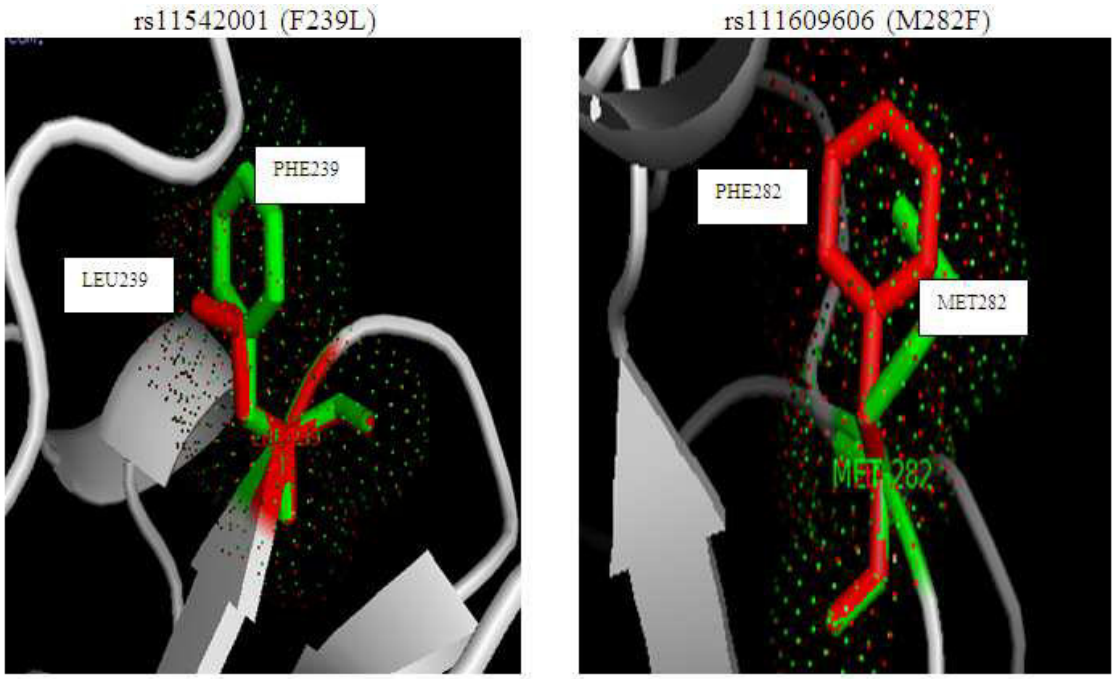
More negative is the $\Delta\Delta G$ value,indicating the less stability in the specified mutation according to I Mutant 2.0server. We obtained 8 nsSNPs that were found to be less stable by this server as shown in Table 4 .Out of 8 nsSNPs six nsSNPs, rs11542001,rs1785994 rs59727333,rs111609606,rs16955280,rs12190 8741 showed a $\Delta\Delta G$ value of > -1.0. The remaining 2 nsSNPs showed a $\Delta\Delta G$ value of < -1.0, as depicted in Table 4. Since these six nsSNPs showed higher $\Delta\Delta G$ value, we considered these nsSNPs to be least stable and deleterious.

Table 4
Protein stability analysis for the MMP2 mutant proteins.

WT	Position	MT	SNP's ID	ΔΔG Value
F	239	L	rs11542001	-2.69
A	447	V	rs17859943	-1.77
K	359	I	rs59727333	-1.11
M	282	F	rs111609606	-2.28
C	65	H	rs112710941	-0.21
G	357	D	rs121908741	-1.66
V	621	L	rs16955280	-2.06
D	450	H	rs28730814	0.21

Figure 1
Structural deviations in the mutant MMP2 structure due to deleterious ns SNP's





Deleterious ns SNP's influence in the functional domain of MMP2.

The X-ray crystallographic-derived structure of the MMP-2 (Protein Data BankID-1CK7) molecule contains a prodomain, catalytic domain, and fibronectin and hemopexin domains.Other functional region's and identified ns deleterious SNP's were listed in the table. Altered Fibronectin expression has a vital role in carcinoma development. There were 3 SNP's located in the fibronectin II domain region.Altered Hemopexin like domain function may have detrimental impact in their function.

Table 5

Domain and Functional region associated with the deleterious ns SNP's

Domain/Functional Region	Start	End	ns SNP's ID
Fibronectin type-II 1	228	276	rs11542001
Fibronectin type-II 3	344	392	rs59727333, rs121908741
Required for inhibitor TIMP2 binding	414	660	rs28730814, rs17859943, rs16955280
Activation peptide	30	109	rs112710941
Hemopexin-like 4	617	660	rs16955280

CONCLUSION

In our analysis, we found out that 6nsSNP (rs11542001, rs17859943, rs59727333, rs111609606, rs16955280, rs121908741) as less stable, deleterious, possibly damaging and have high risk score in SIFT & PolyPhen prediction. The mutant protein structures of these nsSNP also showed energy and RMSD values compared to the native type structure. We therefore concluded that this nsSNP as potentially functional polymorphic. SNP's rs59727333(K359I) create missense mutation i.e.; Positively charged basic amino acid changed into hydrophobic amino acid ,so that prefers buried in hydrophobic core, Similarly rs121908741(G357D) i.e aliphatic to positively charged acid, create an impact in Fibronectin II domain which in turn connected with carcinoma development. rs11542001 (F239L) lies in the domain region of fIbronectin II also reported for their increased expression in lung cancer patients. It is necessary to screen the nsSNP located in this region as Fibronectin is considered as a novel target for anticancer drug development. rs16955280 (V621L) lies in the

Hemopexin like domain region and missense mutation create an amyloid propensity change .The other two SNP's are not connected to any change in domain structure and not involved in structural changes. The results presented from this in silico study will open up new prospect for genetic analysis of MMP gene and their correlation with clinical data will be very useful in understanding their disease association.

CONFLICT OF INTEREST : NONE.

REFERENCES

1. Cauwe B, Van den Steen PE, Opdenakker G The biochemical, biological, and pathological kaleidoscope of cell surface substrates processed by matrix metalloproteinases. Crit Rev Biochem Mol Biol 42: 113-185, (2007).
2. Arakaki PA, Marques MR, Santos MCLG MMP-1 polymorphism and its relationship to pathological processes. J Biosci 34: 313-320, (2009).
3. Ye S Polymorphism in matrix metalloproteinase gene promoters: implication in regulation of gene expression and susceptibility of various diseases. Matrix Biol 19: 623-629, (2000)
4. Noji Y**,** Shimizu M, Ino H, Higashikata T, Yamaguchi M, Nohafa A, Horita T, Shimizu K, Ito Y, Matsuda T, Namura M, Mabuchi H. Increased circulating matrix metaloproteinase-2 in patients with hypertrophic cardiomyopathy with systolic dysfunction. Circ J 68: 355–360, (2004).
5. Tziakas DN, Chalikias GK, Papaioakeim M, Hatzinikolaou EI, Stakos DA, Tentes IK, Papanas N, Kortsaris A, Maltezos E, Hatseras DI. Comparison of levels of matrix metalloproteinase-2 and -3 in patients with ischemic cardiomyopathy versus nonischemic cardiomyopathy. Am J Cardiol 96: 1449–1451, (2005).
6. Altieri P, Brunelli C, Garibaldi S, Nicolino A, Ubaldi S, Spallarossa P, Olivotti L, Rossettin P, Barsotti A, Ghigliotti G. Metalloproteinases 2 and 9 are increased in plasma of patients with heart failure. Eur J Clin Invest 33: 648–656, (2003).
7. Yamazaki T**,** Lee JD, Shimizu H, Uzui H, Ueda T**.** Circulating matrix metalloproteinase-2 is elevated in patients with congestive heart failure. Eur J Heart Fail 6*:* 41–*45, (*2004*).*
8. Matsunaga T**,** Abe N, Kameda K, Hagii J, Fujita N, Onodera H, Kamata T, Ishizaka H, Hanada H, Osanai T, Okumura K. Circulating level of gelatinase activity predicts ventricular remodeling in patients with acute myocardial infarction. Int J Cardiol 105*:* 203–*208, (*2005*).*
9. George J**,** Patal S, Wexler D, Roth A, Sheps D, Keren G. Circulating matrix metalloproteinase-2 but not matrix metalloproteinase-3, matrix metalloproteinase-9, or tissue inhibitor of metalloproteinase-1 predicts outcome in patients with congestive heart failure.. Am Heart J 150*: 484–*487, (2005*).*
10. Martignetti J.A., Aqeel A.A., Sewairi W.A., Boumah C.E., Kambouris M., Mayouf S.A., Sheth K.V., Eid W.A., Dowling O., Harris J. Mutation of the matrix metalloproteinase 2 gene (MMP2) causes a multicentric osteolysis and arthritis syndrome. Nat. Genet.; 28:261–265, (2001).
11. Xu E, Lai M, Lü B, Xing X, Huang Q, Xia X. A single nucleotide polymorphism in the matrix metalloproteinase-2 promoter is associated with colorectal cancer. Biochem Biophys Res Commun.; 324:999–1003, (2004).
12. Nagase H, Barrett AJ, Woessner JF, Jr. Nomenclature and glossary of the matrix metalloproteinases. Matrix (Stuttgart, Germany) 1: 421−424, (1992).
13. Bhagwat M: Searching NCBI's dbSNP database,Curr Protoc Bioinformatics Chapter 1: Unit 1.19. (2010)
14. Dutta S, M Berman H, F Bluhm W Using the tools and resources of the RCSB protein data bank. Curr Protoc

Bioinformatics Dec; Chapter 1: Unit1.9, (2007).
15. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc; 4(7):1073-81,(2009).
16. Altschul SF, Madden TL, Schäffer AA, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research. 25(17):3389–3402,(1997).
17. Sjölander K, Karplus K, Brown M, et al. Dirichlet mixtures: a method for improved detection of weak but significant protein sequence homology. Computer Applications in the Biosciences. 12(4):327–345, (1996).
18. Ng PC, Henikoff S SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res 31: 3812–3814,((2003).
19. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR .Poly Phen. Nat Methods **7**(4):248-249, (2010).
20. Xi T, Jones IM, Mohrenweiser HW. Many amino acid substitution variants identified in DNA repair genes during human population screenings are predicted to impact protein function. Genomics.; 83(6):970–979,(2004).
21. Lee PH, Shatkay H, An integrative scoring system for ranking SNPs by their potential deleterious effects. Bioinformatics 25: 1048–1055, (2009).
22. Sandberg R, Neilson JR, Sarma A, Sharp PA, Burge CB (2008) Proliferating cells express mrnas with shortened 3′ untranslated regions and fewer microRNA target sites. Science 320: 1643–1647.
23. Wang G, Guo X, Floros J Differences in the translation efficiency and mRNA stability mediated by 5′-UTR splice variants of human SP-A1 and SP-A2 genes. AJP - Lung Physiol 289: L497–L508, (2005).
24. Hsiang-Yu Yuan, Jen-Jie Chiou, Wen-Hsien Tseng, Chia-Hung Liu, Chuan-Kun Liu, Yi-Jung Lin, Hui-Hung Wang, Adam Yao, Yuan-Tsong Chen, and Chun-Nan Hsu.FASTSNP: an always up-to-date and extendable service for SNP function analysis and prioritization.Nucleic Acids Res., 1 July; 34: W635 - W641,(2006).
25. Laskowski R A PDBsum new things. Nucleic Acids Res., 37, D355-D359, (2009).
26. The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC.
27. Capriotti E, Fariselli P, Casadio R (2005) I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. Nucleic Acids Res 33: W306–W310.
28. Sigrist CJA, Cerutti L, Hulo N, Gattiker A, Falquet L, Pagni M, Bairoch A, Bucher P.PROSITE: a documented database using patterns and profiles as motif descriptors.Brief Bioinform. 3:265-274(2002).