Open Access Bioinformatics

Dovepress
open access to scientific and medical research

Open Access Full Text Article

ORIGINAL RESEARCH

Influence of SMAD1 gene in osteoporosis: A bioinformatics approach

This article was published in the following Dove Press journal:
Open Access Bioinformatics
21 July 2010
Number of times this article has been viewed

Anand Anabarasu
Arpita Kundu

Bioinformatics Division, School of Biosciences and Technology, VIT University, Tamil Nadu, India.

Abstract: In the present study we have analyzed the role of SMAD family member 1 protein (SMAD1) gene products in relation to bone morphogenesis and osteoporosis. Out of 1045 single nucleotide polymorphisms (SNP) investigated, we find that one nonsynonymous SNP (nsSNP), rs1804647, to have significant damaging effects as predicted by all the tools used in the analysis. This nsSNP resulted in a change of amino acid from a positive charged residue, Lysine, to a strong negatively charged residue, Glutamate, and hence the change of residue with opposite charges might lead to structural defects and result in altered function. The results presented in this report will be a good starting point for genetic analysis of SMAD1 genes in patients with osteoporosis which might lead to more conclusive evidence of the association of this gene with osteoporosis.

Keywords: osteoporosis, nsSNP, SMAD1, rs1804647

Introduction

Osteoporosis is a disease associated with progressive deterioration of bone which results in fragility fractures that occur with very little trauma.[1,2] It is a common disease that affects millions of people worldwide. It is estimated that over 200 million people have osteoporosis.[3]

Genetic factors play a major role in the determination of bone mineral density (BMD) and osteoporosis risk.[4] Multiple chromosomal loci have been mapped by linkage approaches which potentially carry hundreds of genes involved in the determination of bone mass and quality. Association studies based on candidate gene polymorphisms and subsequent meta-analyses, and the more recent genome-wide association studies, have clearly identified a handful of genes associated with BMD and/or fragility fractures.[4]

Molecular genetic studies have implicated the association of single nucleotide polymorphisms (SNP) with osteoporosis.[5] It is anticipated that both the diagnosis and the treatment of this disease will be revolutionized by the integration of genomics and informatics.[2] Also, it is predicted that a genetic algorithm will be developed to identify at-risk patients before they develop osteoporosis, so that preventive measures can be instituted.[2]

The last two decades have seen a tremendous growth of genomic knowledge that is highly relevant to the treatment and management of complex genetic diseases such as osteoporosis.[6,7] In recent years, bioinformaticians have extracted new and meaningful information from literature.[7] The bioinformatics resources and tools for systematic analysis of SNPs are well described in a recent review.[8]

Correspondence: Anand Anabarasu
Bioinformatics Division, School of Biosciences and Technology, VIT University, Vellore 632014, Tamil Nadu, India.
Tel +91 416 220 2556
Email aanand@vit.ac.in

The role of SMA- and MAD-related (SMAD) protein in bone morphogenesis is documented by researchers.[7,9] Even though there is a lot of information on the clinical impact, diagnosis and treatment of osteoporosis, there are not many reports on the systematic analysis of SMAD1 genes in relation to osteoporosis. This report systematically analyzes the SMAD1 gene and its possible implications on osteoporosis using bioinformatics tools. Our results from the present study might be useful for further investigations on the human genetics of osteoporosis.

Materials and methods

Data set

The SNPs associated with SMAD1 gene were obtained from the single nucleotide polymorphism database (dbSNP)[10] which is a public-domain archive established by NCBI, having broad collection of SNPs of any gene of an organism serving as a public repository for genetic variation.

There are 1045 SNPs associated with SMAD1, these are commonly referred to by their reference sequence IDs (rsID), and they are listed in Table 1.

Predicting damaging nsSNPs using SIFT

We used the program Sorting Intolerant from Tolerant (SIFT) designed by Ng and Henikoff to identify the damaging non-synonymous SNP's (nsSNP) in the data set.[11,12] SIFT predicts whether an amino acid substitution affects protein function and it distinguishes between functionally neutral and deleterious amino acid changes in mutagenesis studies and on human polymorphisms. SIFT also searches for similar protein sequences from different species in the database, obtains the multiple alignments of these sequences, and then calculates from the alignment the tolerance index (from 0 to 1) for all possible substitutions at each position. The higher a tolerance index, the less functional impact a particular amino acid substitution is likely to have.[11–13]

Structural alteration by nsSNPs using PolyPhen

The structural alteration as a result of nsSNP was determined using the program Polymorphism Phenotyping (PolyPhen).[14–16] PolyPhen assesses the possible damaging effect of amino acid substitution, based on whether the substitution was

(i) in an active or binding site;
(ii) affecting interactions with ligands present in the experimentally resolved structure;
(iii) leading to hydrophobicity or electrostatic charge change in a buried site;
(iv) destroying a S-S bond;
(v) affecting protein's solubility;
(vi) inserting proline (P) in an α-helix or
(vii) incompatible with the profile of amino acid substitutions observed at that site in the set of homologous proteins.[15]

Functional analysis and selection for nsSNPs using FASTSNP

We used the functional analysis and selection tool for single nucleotide polymorphisms (FASTSNP) web server that allows users to efficiently identify the SNPs that will most likely have functional effects. It prioritizes SNPs according to 13 phenotypic risks and putative functional effects, such as changes to the transcriptional level, pre-mRNA splicing, and protein structure.[17]

The functional report on a SNP contains seven sections on the SNP's functional effects, namely:

(i) genomic information, presents the nearby sequence, the alleles and the allele frequency among different ethnic groups;
(ii) functional effects summary, presents the risk assessment;
(iii) transcription regulatory, shows the predicted transcription factor binding sites generated or disrupted by the different SNP alleles;
(iv) alternative splicing regulatory, reports exonic splicing enhancer/silencer motifs changed by the SNP alleles leading to exon skipping or inclusion;
(v) mRNA/protein domain effects, presents all spliced forms of mRNAs and protein variants extracted from GenBank.[18] The protein domains that the SNP locates in are highlighted;
(vi) protein structure effects, reports whether the SNP may cause a significant structural change in a protein; and
(vii) SwissProt[19] feature table, provides information regarding other known mutations or variations of the translated protein of mRNAs related to the SNP.[17]

Some of these sections were specific to coding or non-coding SNPs and they will appear or not appear in the functional report accordingly.[17]

Gene information for nsSNPs from Ensembl

We found the gene related information for the nsSNPs from the Ensembl database.[20–22] The Ensembl database consists of comprehensive genome information system that features

Table 1 SNP rs ID's associated with SMAD1

rs43206128	rs41837184	rs41837183	rs41837182	rs41837181	rs41837180	rs41837179
rs41837178	rs41837177	rs41837176	rs41837175	rs41837174	rs41837173	rs41837172
rs41837171	rs41837170	rs41837169	rs41837168	rs41837167	rs41837166	rs41837165
rs41837164	rs41837163	rs41837162	rs41837161	rs41837160	rs41837159	rs41837158
rs41837157	rs41837156	rs41837155	rs41837154	rs41837153	rs41837152	rs41837151
rs41837150	rs41837149	rs41837148	rs41837147	rs41837146	rs41837145	rs41837144
rs41837143	rs41837142	rs41837141	rs41837140	rs41837139	rs41837138	rs41837137
rs16383403	rs16383402	rs16383401	rs16383400	rs16383399	rs16383398	rs16383397
rs15529039	rs15529035	rs15529034	rs15529032	rs15529030	rs15529027	rs15529025
rs15529023	rs15529021	rs15529018	rs15529016	rs15529014	rs15529011	rs15529009
rs15529006	rs15529004	rs15529002	rs15528997	rs15528993	rs15528990	rs15528988
rs15528985	rs15528983	rs15528980	rs15528978	rs15528976	rs15528973	rs15528971
rs15528968	rs15528966	rs14445871	rs14445870	rs14445869	rs14445868	rs14445867
rs14445866	rs14445865	rs14445864	rs14445863	rs14445862	rs14445861	rs14445860
rs14445859	rs14445858	rs14445857	rs14445856	rs14445855	rs14445854	rs14445853
rs14445852	rs14445851	rs14445850	rs14445849	rs14445848	rs14445847	rs14445846
rs14445845	rs14445844	rs14445843	rs14445842	rs14445841	rs14445840	rs14445839
rs14445838	rs14445837	rs14445836	rs14445835	rs14445834	rs14445833	rs14445832
rs14445831	rs14445830	rs14445829	rs14445828	rs14445827	rs14445826	rs14445825
rs14445824	rs14445823	rs14445822	rs14445821	rs14445820	rs14445819	rs14445818
rs13513493	rs13513492	rs13513491	rs13513490	rs13513489	rs13513488	rs13513487
rs13513486	rs10722400	rs13455744	rs13455743	rs13455742	rs52353258	rs49393035
rs48744858	rs40324585	rs39689443	rs39378399	rs38904637	rs38890813	rs38861705
rs38837210	rs38736131	rs38728861	rs38673698	rs38672482	rs38598229	rs38594114
rs38586608	rs38571957	rs38478504	rs38408710	rs38407259	rs38381145	rs38379589
rs38379316	rs38357614	rs38343211	rs38319122	rs38304454	rs38286771	rs38216312
rs38198301	rs38151706	rs38122743	rs38104923	rs38084479	rs38069777	rs38046037
rs38028566	rs38028025	rs38009371	rs37999190	rs37996620	rs37970866	rs37890809
rs37857319	rs37844585	rs37829955	rs37829805	rs37820241	rs37798728	rs37776122
rs37769262	rs37751971	rs37723161	rs37678138	rs37670970	rs37662143	rs37622281
rs37606105	rs37601352	rs37595569	rs37592912	rs37570279	rs37565989	rs37557932
rs37547978	rs37529033	rs37512483	rs37502918	rs37462850	rs37462363	rs37447335
rs37424711	rs37395093	rs37380944	rs37372985	rs37358262	rs37346703	rs37344855
rs37328013	rs37316393	rs37311870	rs37278533	rs37227073	rs37197406	rs37185197
rs37183904	rs37176820	rs37173505	rs37141214	rs37138939	rs37133086	rs37095121
rs37078372	rs37072603	rs37055222	rs37049288	rs37041092	rs37036754	rs37035385
rs37026804	rs37021065	rs37014646	rs36984116	rs36981578	rs36974734	rs36961475
rs36935369	rs36934698	rs36926658	rs36915208	rs36913122	rs36912217	rs36910109
rs36904384	rs36854139	rs36836173	rs36826819	rs36807995	rs36785399	rs36781909
rs36781325	rs36780156	rs36773895	rs36770138	rs36763326	rs36756065	rs36734144
rs36720417	rs36693953	rs36682785	rs36678280	rs36676848	rs36652809	rs36628123
rs36623863	rs36620169	rs36616718	rs36604460	rs36604173	rs36598919	rs36592125
rs36590740	rs36589398	rs36589170	rs36587668	rs36577742	rs36574562	rs36574154
rs36574562	rs36574154	rs36550928	rs36548387	rs36543793	rs36550928	rs36548387
rs36543793	rs36518106	rs36516399	rs36513012	rs36500074	rs36499024	rs36485545
rs36471433	rs36463229	rs36433155	rs36417697	rs36409573	rs36392860	rs36385065
rs36380669	rs36378657	rs36374178	rs36372803	rs36355585	rs36355433	rs36355396
rs36343803	rs36336450	rs36334006	rs36308116	rs36307375	rs36304409	rs36273831
rs36262476	rs36258235	rs13471524	rs6320822	rs6320755	rs6320308	rs6320203
rs6157027	rs41182187	rs41162462	rs41127100	rs41097188	rs41052054	rs41042179
rs41001288	rs40914529	rs40829302	rs40801804	rs40785650	rs40718880	rs40714934
rs40707256	rs40699558	rs40696624	rs40624834	rs40600779	rs22438341	rs22438340
rs22438338	rs22435243	rs22435242	rs22430308	rs22429717	rs22429143	rs22423498
rs22423496	rs22422179	rs22422178	rs22420143	rs22416419	rs22413764	rs22410926
rs22408159	rs22407129	rs22407128	rs22407127	rs22407126	rs22407125	rs22407124
rs22407110	rs22407109	rs22407108	rs22407107	rs22407106	rs22407105	rs22407104
rs22407103	rs22407101	rs22407094	rs22407093	rs22407092	rs22407091	rs22407090
rs22407089	rs22407088	rs22407087	rs22407086	rs22407079	rs22407078	rs22407077

(*Continued*)

Table 1 (*Continued*)

rs22407076	rs22407075	rs22407074	rs22407073	rs22407072	rs22407071	rs22407070
rs22407064	rs22407062	rs22407060	rs22407059	rs22407057	rs22407056	rs22407055
rs22407054	rs22407052	rs22407049	rs22407047	rs22407045	rs22407044	rs22407043
rs22407042	rs22407041	rs22405061	rs22403115	rs22401316	rs22399442	rs22398238
rs22398211	rs22398210	rs22379962	rs22379948	rs22379160	rs22364266	rs22364212
rs22363681	rs22353531	rs22353529	rs22349453	rs22349452	rs22349451	rs22348997
rs22348996	rs9197660	rs9197659	rs9190264	rs9190263	rs9159692	rs9159691
rs9159690	rs9159689	rs9159688	rs9127901	rs9127900	rs9113580	rs9082059
rs9082058	rs9015277	rs9015276	rs9015275	rs8967641	rs8967640	rs8958253
rs8958252	rs8948231	rs8948230	rs8948229	rs8948228	rs8916864	rs8904715
rs8878619	rs8878618	rs8878617	rs8878616	rs8878615	rs8871149	rs8871148
rs8871147	rs8821428	rs8616390	rs8616389	rs8548136	rs8548135	rs8548134
rs8548133	rs8434954	rs8434953	rs8434952	rs8434951	rs8412032	rs26109380
rs25383970	rs25350003	rs25327873	rs25326101	rs25321520	rs25306350	rs25297286
rs25294121	rs25292078	rs25285454	rs25281219	rs25279061	rs25263171	rs25241614
rs25239333	rs25238019	rs25231997	rs24891022	rs62446864	rs62446863	rs62446862
rs61613428	rs61128823	rs61098313	rs60956809	rs60584748	rs60532382	rs60323633
rs60103805	rs59784627	rs59525016	rs59034773	rs58891283	rs58572713	rs58441521
rs57305857	rs57232114	rs56022562	rs55811804	rs55731773	rs41275991	rs41275988
rs36125163	rs35821025	rs35711502	rs35482863	rs35425165	rs35302357	rs35241091
rs35151755	rs35045679	rs34957085	rs34885661	rs34725384	rs34668775	rs34495080
rs34457367	rs34230821	rs34149226	rs28937323	rs28937322	rs28936972	rs17159296
rs17159291	rs17159289	rs17159287	rs17159280	rs17159279	rs17159275	rs17159273
rs17159272	rs17159267	rs17159262	rs60939160	rs16875290	rs17159297	rs13228758
rs13228097	rs13228094	rs13228006	rs13226718	rs12532057	rs11982890	rs11981106
rs11980069	rs11979352	rs11976576	rs56650146	rs11761080	rs11553503	rs11553502
rs11553498	rs11553497	rs11465024	rs35809100	rs11454228	rs17778009	rs60345184
rs10951271	rs10487700	rs10487699	rs10280491	rs61212867	rs10265589	rs10264469
rs10249885	rs10239660	rs57200522	rs10239537	rs10233132	rs10229836	rs7808770
rs6970807	rs6950015	rs10347965	rs6947982	rs6945586	rs6945546	rs6945475
rs6945223	rs6945035	rs6462214	rs4000210	rs3886641	rs3840589	rs11403655
rs3840588	rs3807633	rs10386194	rs9792086	rs56638865	rs3807632	rs2970504
rs2952805	rs2952804	rs2952795	rs4461798	rs2893390	rs2893389	rs2893388
rs10378014	rs3735431	rs2709809	rs2709806	rs3807631	rs17700042	rs57637303
rs2709778	rs60761314	rs2709772	rs2709771	rs2529442	rs60428326	rs2529441
rs3779252	rs2529439	rs11553501	rs2529438	rs3764826	rs57476542	rs2527880
rs60677610	rs2527879	rs2527878	rs10378720	rs2270025	rs2240501	rs58255920
rs10378822	rs2240401	rs60101002	rs2240400	rs34886142	rs2230310	rs2214837
rs2190241	rs2074779	rs2072236	rs1986757	rs56754355	rs10338017	rs1986756
rs59471849	rs1558064	rs1468402	rs17413053	rs3189753	rs1130653	rs3189564
rs11553500	rs17856227	rs17856223	rs1049402	rs3189560	rs1049399	rs2709804
rs758000	rs61005429	rs728539	rs722423	rs17159306	rs1130593	rs3189633
rs2230311	rs2070205	rs1801345	rs1049584	rs3189770	rs4593	rs62343509
rs62343505	rs62343494	rs62343493	rs62343492	rs61303468	rs61102345	rs60788260
rs60537353	rs60369536	rs60139040	rs60073922	rs60072759	rs60014433	rs59984540
rs59971057	rs59933736	rs59909582	rs59846899	rs59268619	rs59193620	rs59063315
rs59057604	rs59054276	rs59029464	rs58968834	rs58913698	rs58688997	rs58688046
rs58468364	rs58339165	rs57963489	rs57797623	rs57612157	rs57581646	rs57545372
rs57257834	rs57044560	rs56986943	rs56985736	rs56837794	rs56756555	rs56695028
rs60478333	rs56250124	rs55948666	rs55835219	rs55700163	rs36007951	rs35999258
rs35986161	rs35956618	rs35923324	rs35882080	rs35842798	rs35813714	rs35812565
rs35774186	rs35756866	rs35747692	rs35710925	rs35668194	rs35612203	rs35584992
rs35528395	rs35483247	rs35476542	rs35386763	rs35281972	rs35241245	rs35184283
rs35058397	rs35046379	rs62343508	rs34978105	rs34968029	rs34959603	rs34880052
rs34832344	rs34791274	rs34791030	rs34629544	rs34610916	rs34504494	rs34443486
rs34405116	rs34373983	rs34333288	rs34318288	rs34303116	rs34169615	rs34142651
rs56664494	rs34107624	rs34082880	rs34079792	rs34077417	rs34013277	rs34011978

(*Continued*)

Table 1 (*Continued*)

rs60015014	rs33928689	rs57951197	rs28607764	rs28579035	rs28541642	rs56683612
rs28534855	rs28522520	rs28495533	rs28464746	rs28408896	rs28397904	rs17797966
rs17797805	rs17741593	rs17426247	rs59063334	rs17020321	rs17020317	rs17020315
rs17020313	rs17020310	rs17020304	rs17020302	rs17020297	rs17020284	rs17020281
rs17020269	rs17020255	rs17020253	rs17020251	rs59444418	rs17020248	rs17020237
rs17020236	rs58221759	rs17020235	rs17020202	rs59633267	rs17020201	rs56640118
rs58653480	rs17020200	rs16998662	rs16998659	rs13435697	rs13149786	rs13149771
rs13149754	rs13145825	rs13144151	rs13138805	rs58040113	rs13137598	rs13120843
rs13118865	rs13113936	rs13110369	rs13104775	rs17426240	rs12646702	rs12504239
rs58910572	rs11946962	rs11946830	rs11945076	rs58166623	rs11944685	rs11944363
rs11939520	rs11939179	rs11938489	rs11937875	rs11937328	rs11936636	rs11935522
rs57735613	rs11935237	rs13129371	rs11934722	rs11931456	rs13107924	rs11930852
rs13104519	rs11930324	rs11736932	rs11725101	rs56452565	rs58073224	rs11724813
rs11547180	rs11547178	rs11537832	rs58395199	rs11392205	rs11355954	rs11302430
rs33926573	rs11293293	rs33995597	rs11285972	rs11100885	rs11100884	rs17798152
rs60525943	rs56482779	rs11100883	rs56920353	rs11100882	rs10715854	rs35752574
rs10713518	rs10694111	rs10677232	rs36210455	rs10599198	rs10557196	rs33996344
rs10548153	rs10519733	rs10033759	rs9992345	rs9968492	rs9685596	rs9683725
rs7699931	rs7699352	rs59616816	rs7698944	rs7695203	rs61314879	rs7694604
rs7694086	rs56553319	rs7689005	rs7686994	rs60923728	rs7685592	rs7685387
rs7684336	rs7683590	rs58089363	rs7682401	rs7681713	rs7677879	rs7677863
rs7677160	rs7675811	rs59990034	rs7673821	rs7672606	rs59927196	rs7672306
rs60192542	rs7670486	rs7670403	rs17798495	rs7662543	rs56661211	rs7662541
rs56554331	rs7661162	rs7658436	rs7658004	rs57170185	rs7655568	rs7435311
rs17435952	rs6854034	rs61107117	rs6852200	rs61409198	rs6848098	rs6845829
rs57611059	rs6833260	rs6829932	rs6829091	rs6823513	rs6822700	rs6816306
rs6813833	rs6537359	rs13121278	rs17797649	rs6537356	rs17426233	rs57727279
rs6537355	rs60611965	rs35037580	rs5862739	rs5862738	rs4395527	rs3948257
rs3943707	rs3816967	rs3775325	rs3775323	rs3775322	rs60112569	rs3756021
rs57309032	rs17741322	rs2289737	rs17227189	rs2118438	rs57644485	rs2068991
rs61103877	rs2043779	rs57261547	rs2013367	rs59728727	rs17741821	rs1992165
rs1804647	rs1497126	rs1437827	rs1437826	rs1370571	rs57628021	rs1055440
rs3816966	rs58152281	rs1016792	rs60988923	rs1016791	rs60581465	rs959641
rs3775324	rs920261	rs58547099	rs768139	rs763560	rs17798104	rs714195
rs15392	rs11416.	rs14270	rs62343506			

an integrated set of genome annotation, databases and other information for chordate and selected model organism and disease vector genomes.[20–22] The Ensembl genome browser provides visualization for genome annotations, alignments, variation and functional genomics data and supporting additional data integration.

Ensembl gene sets are created using an automated analysis pipeline that has been significantly optimized based on the completeness of the genome sequence.[20–22]

Functional significance of nsSNPs by UTR scan

The untranslated regions (UTR) functional significance of each SNP was determined using UTResource, an internet based UTR analyzing tool, consisting of a resource of sequence analysis of 5' and 3' UTR of eukaryotic mRNAs (http://www.ba.itb.cnr.it/BIG/UTRScan). Briefly, two or three sequences of each UTR SNP which have different nucleotide at SNP position were analyzed by UTRscan which looks for UTR functional elements by searching through user submitted sequence data for the patterns defined in the UTRsite collection.

The collection of functional sequence patterns located in 5' or 3' UTR sequences were stored in UTRsite. If different sequences of each UTR SNP were found to have different functional pattern(s), that UTR SNP is predicted to have functional significance.[13]

Modeling wild and mutant protein structures

We utilized the SWISS-MODEL[23–25] workspace, a web-based integrated service dedicated to protein structure homology

modeling for modeling the native and mutant protein structures. Swiss-Model assists and guides the user in building protein homology models at different levels of complexity. A personal working environment is provided for each user where several modeling projects can be carried out in parallel.

The program builds a homology model based on four main steps:

(1) identification of structural template(s),
(2) alignment of target sequence and template structure(s),
(3) model building and
(4) model quality evaluation.

3-D viewing and RMSD calculations

The Swiss PDB viewer (Deep View) version 4.01 was used for viewing the modeled structures and for calculation of the root mean square deviation (RMSD) between the native and mutant structures.[26] DeepView is an application that provides a user friendly interface to analyze several proteins at the same time. The proteins can be superimposed in order to deduce structural alignments and compare their active sites or any other relevant parts. Amino acid mutations, hydrogen bonds, angles and distances between atoms can be obtained using Deep View.[26]

Results and Discussion

We used the SIFT tool to predict whether an amino acid substitution at a particular position in a protein molecule will have a phenotypic effect. We find that of 1045 rs IDs associated with SMAD1 gene, only two, rs1804647 and rs17159287, are non-synonymous.

SIFT gives the results of these nsSNPs with their amino acid change and the tolerance index score which predicts the functional impact of the protein of these SNPs. The lesser the score, the more functional impact, ie, the corresponding SNP is predicted to be damaging. Of these two SNPs, rs1804647 is predicted to be damaging with a tolerance index score of 0.01. (Score of $\leq$0.05 is the cut-off to identify damaging SNPs)[11–13] as shown in Table 2. The nucleotide change is in the 510 position where **A**AA is replaced with **G**AA ie, Adenine to Guanine resulting in the substitution of amino acid Lysine to Glutamate. (The base represented in bold caption is the nsSNP).

We employed the PolyPhen based methodology[14–16] to compute the position-specific independent counts (PSIC) score difference of the two identified nsSNPs (rs1804647 and rs17159287) to assess the functional impact of the amino acid substitution predicted by SIFT. To evaluate the ability of the PolyPhen method to distinguish between damaging and neutral amino acid types; we applied it to the data on known deleterious mutations and to the data on species divergence.[15] The deleterious data set consists of natural replacements known to cause disease phenotypes, variants observed in individuals affected by genetic disorders and artificial replacements known to damage structure, function and the stability of a protein.[15] Score of $\geq$1.5 is the cut-off to identify damaging SNPs by Polyphen. SNP with PSIC score difference of higher value is predicted to be damaging. We found that of the two nsSNPs, rs1804647 is predicted to be damaging with a PSIC score difference of 1.852, as depicted in Table 3.

FASTSNP was employed to efficiently identify and prioritize high risk nsSNPs according to their phenotypic risks and putative functional effects.[17] FASTSNP rates the phenotypic deleterious risks from 0 to 5 point scale; 0 representing no known effect and 5 indicating very high damaging effect, while a score of 3–4 indicates moderate to high damaging effect.[17] We found rs17159287 to have low to medium level of risk, whereas rs1804647 was identified to have significantly damaging effect with a score in the range of 3–4; as shown in Table 4. This may be due to splicing regulation functional effect by possibly breaking the exonic splicing enhancer/silencer binding site in the coding sequence leading to abolished protein domain.[17]

Gene related information the non-synonymous SNPs associated with SMAD1, with their corresponding change in alleles and amino acids were obtained from Ensembl database and we found three of the 1045 associated with SAMD1 listed as nsSNPs. These are: rs1804647, rs61748163; and rs62343507, depicted in Table 5.

Table 2 nsSNPs predicted by SIFT

SNP ID	Amino acid change[a]	Protein	Damaging amino acid[a]	Damage	Score	Median sequence IC
rs1804647	E32K	NP_005891	E	YES	0.01	3.29
rs17159287	Q334R	NP_002038	Q	NO	0.59	2.44

[a]**Abbreviations:** E, Glutamate; K, Lysine; Q, Glutamine; R, Arginine.

Table 3 nsSNPs predicted by PolyPhen

SNP ID	Protein	Position	Amino acid 1	Amino acid 2	PSIC[a]	Prediction
rs17159287	P41250 NP_002038.1	388	R	Q	0.298	Benign
rs1804647	1. Q15797 NP_001003688.1	32	K	E	1.852	Possibly damaging
	2. Q15797 NP_005891.1	32	K	E	1.852	Possibly damaging

[a]Position Specific Independent Count.

There is a significant correlation between all the SNP tools with respect to rs1804647 and hence this SNP might be expected to have high deleterious impact on the function of the bone morphogenesis process. The other two SNPs (rs61748163 and rs62343507) did not show any result with other tools used and hence, we did not consider them for our study.

To determine the functional significance of the nsSNP rs1804647, we used UTRScan and found a change in a pattern in 3' UTR region. Polymorphism in the 3' UTR region may affect gene expression by affecting RNA half-life or influencing ribosomal translation of mRNA.[27]

The functional element change related to this nsSNP was found to be 15-lipoxygenase differentiation control element (15-LOX-DICE). 15-LOX-DICE controls 15-LOX synthesis which catalyses the degradation of lipids and is an important factor responsible for the degradation of mitochondria during reticulocyte maturation.[13] The mutation Lysine→Glutamate is observed in the protein due to nucleotide change in the coding sequence of the SNP in 510 position where **A**AA is replaced with **G**AA. From this we infer that the nsSNP, rs1804647 is significantly damaging which might result in osteoporosis.

The nsSNP, rs17159287 has tolerance index score of 0.59 with SIFT, a PSIC score difference of 0.298 with PolyPhen, a relatively low risk score of 2–3 with FASTSNP and was not listed in Ensembl database. Hence we did not consider this nsSNP to be significantly damaging.

SNP rs61748163 has amino acid change from Serine to Asparagine, both are of same properties and showed no change in structure when native and mutant proteins were modeled. SNP, rs62343507 has amino acid change from Tryptophan to Arginine, also showed no change in structure when native and mutant proteins were modeled. Hence, we did not consider these nsSNPs (rs61748163 and rs62343507) listed in Ensembl to be significantly damaging and they did not show any results with other bioinformatics tools.

To assess the 3-D structural change brought about by the nsSNP rs1804647, we performed structural analysis comparing the native and mutant protein structures. The native structure for protein Q15797 and mutant structure for nsSNP rs1804647 were modeled using SWISS-MODEL. To analyze their structural differences, superimposition of both the structures was done using the SWISS PDB Viewer tool and we found an RMSD of 2.95 Å between the native and mutant protein structures as shown in Figure1. This nsSNP also showed a high PSIC score difference of 1.852. Hence, there might be a considerable change in the structure because of this nsSNP. Even though, there is only one amino acid change, the change in amino acid is of the different type. Here the amino acid Lysine is a cationic residue in the native protein and the substituted residue in the mutant protein molecule is Glutamate which is an anionic residue. Due to the complete change in charges brought about by this mutation, there seems to be a considerable change in structure.

The function of a protein is dependent on its 3-D structure and shape which is determined by its primary structure or sequence of amino acids. Since there is a change in amino acid sequence due to the single nucleotide polymorphism,

Table 4 nsSNPs predicted by FAST SNP

SNP ID	Position of change in amino acid	Level of risk	Possible functional effect
rs1804647	32	3–4 (moderate to high)	Splicing regulation
rs17159287	388	2–3 (low to medium)	Splicing regulation

Table 5 nsSNPs predicted by ENSEMBL database

AA position	SNP ID	SNP type	Alleles[a]	Amino acid change[a]
32	rs1804647	Non-synonymous	A/G	K, E
281	rs61748163	Non-synonymous	A/G	S, N
358	rs62343507	Non-synonymous	C/T	W, R

[a]**Abbreviations:** A, Adenine; G, Guanine; C, Cytosine; T, Thymidine; K, Lysine; E, Glutamate; S, Serine; N, Asparagine; W, Tryptophan; R, Arginine.

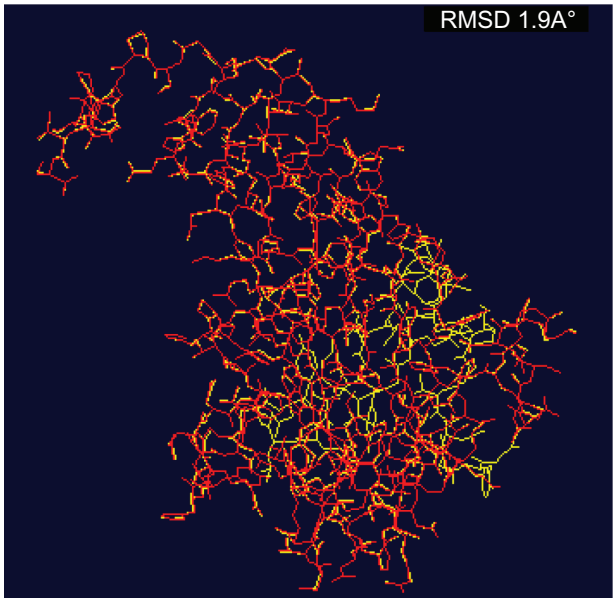

Figure 1 Superimposed structures of native (yellow) and mutant (red) proteins.

non functional protein may be produced which will lead to altered overall function and hence this particular nsSNP may be responsible for a defective bone morphogenic protein and result in osteoporosis.

Results obtained from this in-silico study open new prospects for understanding osteoporosis and its consequences. These results, if correlated with clinical data, will be very useful in understanding the genetics of osteoporosis.

The methodology used to examine the effects of SNPs using a succession of tools could be structured and proposed as a standard screening process for SNPs' effects assessment on genes.

Conclusion

The nsSNP rs1804647 is predicted to be damaging by SIFT, PolyPhen, FASTSNP and listed as deleterious in Ensembl database. The change of nucleotide in 510th position from A to G (**A**AA→**G**AA) results in the amino acid change from Lysine to Glutamate. Since the change of amino acid residues is of opposite charge it may lead to conformational change of the three dimensional structure of the protein and hence result in altered function.

This observation can be considered as a good beginning for further investigations to study the genetic mechanisms behind osteoporosis, and when considered along with different clinical patterns of osteoporosis will result in more definitive evidence for the genetic origin of osteoporosis.

Disclosure

The authors report no conflicts of interest in this work.

References

1. Cline RR, Farley JF, Hansen RA, et al. Osteoporosis beliefs and antiresorptive medication use. *Maturitas*. 2005;50(3):196–208.
2. Baylink DJ, Strong DD, Mohan S. The diagnosis and treatment of osteoporosis: future prospects. *Molecular Medicine Today*. 1999;5(3):133–140.
3. Reginster JY, Burlet N. Osteoporosis: A still increasing prevalence. *Bone*. 2006;38(2 Suppl 1): S4–S9.
4. Ferrari S. Human genetics of osteoporosis. *Best Practice and Research Clinical Endocrinology & Metabolism*. 2008;22(5):723–735.
5. Liu YJ, Shen H, Xiao P, et al. Molecular genetic studies of gene identification for osteoporosis: a 2004 update. *Journal of Bone and Mineral Research*. 2006;21(10):1511–1535.
6. Bajdik CD, Kuo B, Rusaw S, et al. CGMIM: automated text-mining of Online Mendelian Inheritance in Man (OMIM) to identify genetically-associated cancers and candidate genes. *BMC Bioinformatics*. 2005;78(6):1–7.
7. Gajendran VK, Lin JR, Fyhrie DP. An application of bioinformatics and text mining to the discovery of novel genes related to bone biology. *Bone*. 2007;40(5):1378–1388.
8. Mah JTL, Chia KS. A gentle introduction to SNP analysis: resources and tools. *J Bioinfo Comp Biol*. 2007;5(5):1123–1138.
9. Ross S, Hill CS. How the Smads regulate transcription. *The Int J of Biochem. Cell Biol*. 2008;40(3):383–408.
10. Sherry ST, Ward MH, Kholodov M, et al. dbSNP: The NCBI database of genetic variation. *Nucleic Acids Res*. 2001;29(1): 308–311.
11. Ng CP, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res*. 2003;31(13):3812–3814.
12. Ng CP, Henikoff S. Predicting deleterious amino acid substitutions. *Genome Res*. 2001;11:863–874.
13. Shen J, Deininger PL, Zhao H. Applications of computational algorithm tools to identify functional SNPs in cytokine genes. *Cytokine*. 2006;35(1–2):62–66.
14. Ramensky V, Bork P, Sunyaev S. Human non-synonymous SNPs: Server and survey. *Nucleic Acids Res*. 2002;30(17):3894–3900.
15. Sunyaev S, Ramensky V, Koch I, et al. Prediction of deleterious human alleles. *Hum Mol Genet*. 2001;10(6):591–597.
16. Sunyaev S, Ramensky V, Bork P. Towards a structural basis of human non-synonymous single nucleotide polymorphisms. *Trends Genet*. 2000;16(5):198–200.
17. Yuan HY, Chiou JJ, Tseng WH, et al. FASTSNP: an always up-to-date and extendable service for SNP function analysis and prioritization. *Nucleic Acids Res*. 2006;34(Web Server):W635–W641.
18. Benson DA, Karsch-Mizrachi I, Lipman DJ, et al. GenBank: update. *Nucleic Acids Res*. 2004;32(Database):D23–D26.
19. Boeckmann B, Bairoch A, Apweiler R, et al. The Swiss-Prot protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res*. 2003;31(1):365–370.
20. Hubbard TJP, Aken BL, Ayling S, et al. Ensembl 2009. *Nucleic Acids Res*. 2009;37(Database):D690–D697.
21. Flicek P, Aken BL, Beal K, et al. Ensembl 2008. *Nucleic Acids Res*. 2008;36(Database):D707–D714.
22. Hubbard T, Barker D, Birney E, et al. The Ensembl genome database project. *Nucleic Acids Research*. 2002;30(1):38–41.
23. Arnold K, Bordoli L, Kopp J, et al. The SWISS-MODEL Workspace: A web-based environment for protein structure homology modelling. *Bioinformatics*. 2006;22(2):195–201.
24. Kopp J, Schwede T. The SWISS-MODEL Repository of annotated three-dimensional protein structure homology models. *Nucleic Acids Res*. 2004;32(Database):D230–D234.

25. Schwede T, Kopp J, Guex N, et al. SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Res*. 2003;31(13):3381–3385.
26. Guex N, Peitsch MC. SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis*. 1997;18(15):2714–2723.
27. van Deventer SJH. Cytokine and cytokine receptor polymorphisms in infectious disease. *Intensive Care Med*. 2000;26(13):S98–S102.