

Evaluation and identification of damaged single nucleotide polymorphisms in COL1A1 gene involved in osteoporosis

Tariq Ahmad Masoodi, Mohammed A. Alsaif, Sulaiman A. Al Shammari, Adel A. Alhamdan

Health Care Development for Elderly Research Chair, College of Applied Medical Sciences, King Saud University, Riyadh, Kingdom of Saudi Arabia

Submitted: 27 April 2011

Accepted: 16 December 2011

Arch Med Sci 2013; 9, 5: 899–905

DOI: 10.5114/aoms.2012.28598

Copyright © 2013 Termedia & Banach

Corresponding author:

Tariq Ahmad Masoodi
Community Health
Sciences Department
Health Care Development
for Elderly Research Chair
College of Applied
Medical Sciences
King Saud University
PO Box 10219
Riyadh 11433
Kingdom of Saudi Arabia
Phone: +966 (1)4693705
Fax: +0966 (1) 4693698
E-mail: tahamasoodi@gmail.com

Abstract

Introduction: Single-nucleotide polymorphisms (SNPs) are biomarkers for exploring the genetic basis of many complex human diseases. The prediction of SNPs is promising in modern genetic analysis but it is still a great challenge to identify the functional SNPs in a disease-related gene. The computational approach has overcome this challenge and an increase in the successful rate of genetic association studies and reduced cost of genotyping have been achieved. The objective of this study is to identify deleterious non-synonymous SNPs (nsSNPs) associated with the COL1A1 gene.

Material and methods: The SNPs were retrieved from the Single Nucleotide Polymorphism Database (dbSNP). Using I-Mutant, protein stability change was calculated. The potentially functional nsSNPs and their effect on proteins were predicted by PolyPhen and SIFT respectively. FASTSNP was used for estimation of risk score.

Results: Our analysis revealed 247 SNPs as non-synonymous, out of which 5 nsSNPs were found to be least stable by I-Mutant 2.0 with a DDG value of > -1.0 . Four nsSNPs, namely rs17853657, rs17857117, rs57377812 and rs1059454, showed a highly deleterious tolerance index score of 0.00 with a change in their physicochemical properties by the SIFT server. Seven nsSNPs, namely rs1059454, rs8179178, rs17853657, rs17857117, rs72656340, rs72656344 and rs72656351, were found to be probably damaging with a PSIC score difference between 2.0 and 3.5 by the PolyPhen server. Three nsSNPs, namely rs1059454, rs17853657 and rs17857117, were found to be highly polymorphic with a risk score of 3-4 with a possible effect of non-conservative change and splicing regulation by FASTSNP.

Conclusions: Three nsSNPs, namely rs1059454, rs17853657 and rs17857117, are potential functional polymorphisms that are likely to have a functional impact on the COL1A1 gene.

Key words: in silico analysis, dbSNP, SIFT, PolyPhen.

Introduction

Osteoporosis, a serious skeletal disease commonly observed among the elderly, is associated with substantial morbidity and socio-economic burden [1, 2]. In the United States alone, more than 40 million people either already have osteoporosis or are at high risk due to low bone mass [http://www.niams.nih.gov/Health_Info/Bone/Osteoporosis/osteoporosis_ff.asp]. In Saudi Arabia, a high prevalence of osteoporosis in the elderly has

been observed [3]. It is diagnosed when the bone mineral density (BMD) is greater than 2.5 standard deviations below peak bone mass according to the criteria of the World Health Organization [4]. Osteoporosis can occur in both men and women and at any age, but it is most common in older women. It has been reported that one in two women and one in 5 men over the age of 50 sustain fractures due to osteoporosis [3]. Collagen type I α 1 (COL1A1) encodes the primary subunit of α -1-chain type I collagen, the main structural and most abundant protein in bone. Within this gene, > 400 human disease-associated mutations have been identified, the majority of which are linked to osteoporosis. The COL1A1 gene is a strong functional candidate for the genetic regulation of bone mass and susceptibility to fragility fractures [5].

Single-nucleotide polymorphisms (SNPs) are the most common mutations of DNA sequence variation for mapping complex genetic traits. About 500,000 SNPs fall within the coding regions of the human genome. Among these, the non-synonymous SNPs cause changes in the amino acid residues. These are likely to be an important factor contributing to the functional diversity of the encoded proteins in the human population [6]. It has been worked out that non-synonymous SNPs (nsSNPs) affect the functional roles of proteins in the signal transduction of visual, hormonal, and other stimulants [7, 8]. These nsSNPs affect gene expression by modifying DNA and transcription factor binding [9, 10] and deactivate active sites of enzymes or change splice sites, thereby producing defective gene products [11, 12].

Epidemiological association studies focus a great amount of effort on identifying SNPs in genes that may have an association with disease risk, and often the SNPs that have an association with disease are non-synonymous. Many molecular epidemiological studies focus on studying SNPs found in coding regions in the hope of finding significant association between SNPs and disease susceptibility, but often find little or no association [13]. With the availability of high-throughput SNP detection techniques, the population of nsSNPs is increasing rapidly, providing a platform for studying the relationship between genotypes and phenotypes of human diseases. Our ability to better select an nsSNP for an association study can be enhanced by first examining the potential impact an amino acid variant may have on the function of the encoded protein using different SNP detection programs such as I-Mutant, Sort Intolerant from Tolerant (SIFT) and Polymorphism Phenotype (PolyPhen) [13]. Discovering the deleterious nsSNPs out of a pool of all the SNPs will be very useful for epidemiological population-based studies.

So the main aim of this study is to identify deleterious nsSNPs associated with the COL1A1 gene.

Material and methods

Methodology

Methodology used was the same as described earlier [6, 13, 14].

SNP dataset from dbSNP

This computational analysis used Single Nucleotide Polymorphism Database (dbSNP) (<http://www.ncbi.nlm.nih.gov/SNP/>) to identify SNPs and their related protein sequence for the COL1A1 gene [15].

Analysis of protein stability change by I-Mutant 2.0

We predicted nsSNP causing protein stability change using the I-Mutant 2.0 tool [16] available from the University of Bologna (<http://gpcr.biocomp.unibo.it/>). I-Mutant 2.0 is a support vector machine (SVM) based tool for the automatic prediction of protein stability change upon single amino acid substitution. The protein stability change was predicted from the COL1A1 protein sequence (NP_000079). The software computed the predicted free energy change value or sign (DDG) which is calculated from the unfolding Gibbs free energy value of the mutated protein minus the unfolding Gibbs free energy value of the native protein (kcal/mol). A positive DDG value indicates that the mutated protein possesses high stability and vice versa.

Evaluation of coding single nucleotide polymorphisms

There are many web-based resources available that allow one to predict whether non-synonymous coding SNPs may have functional effects on proteins. We chose SIFT [17] available from <http://sift.jcvi.org/> to perform protein conservation analysis and predict the phenotypic effect of amino acid substitutions. The SIFT is based on the premise that protein evolution is correlated with protein function. Variants that occur at conserved alignment positions are expected to be tolerated less than those that occur at diverse positions. The algorithm uses a modified version of PSIBLAST [18] and Dirichlet mixture regularization [19] to construct a multiple sequence alignment of proteins that can be globally aligned to the query sequence and belong to the same clade. The underlying 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. The SIFT scores ≤ 0.05 are predicted by the algorithm to be intolerant or deleterious amino acid substitutions, whereas scores > 0.05 are considered tolerant [20]. The higher the tolerance index of a particular amino acid substitution, the smaller is its likely impact.

Simulation of functional change in nsSNP by PolyPhen server

PolyPhen [21] available from Harvard School of Medicine (<http://genetics.bwh.harvard.edu/pph/>) is a computational tool for 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. For a given amino acid variation, PolyPhen performs several steps: (a) extraction of sequence-based features of the substitution site from the UniProt database, (b) calculation of profile scores for two amino acid variants, (c) calculation of structural parameters and contacts of a substituted residue. PolyPhen scores were classified as 'benign' or 'probably damaging' [22]. Input options for the PolyPhen server are protein sequence or accession number together with sequence position with two amino acid variants. We submitted the query in the form of a protein sequence with mutational position and two amino acid variants. PolyPhen searches for three-dimensional protein structures, multiple alignments of homologous sequences and amino acid contact information in several protein structure databases. Then it calculates position-specific independent count (PSIC) scores for each of two variants, and computes the difference of the PSIC scores of the two variants. The higher a PSIC score difference, the higher the functional impact a particular amino acid substitution is likely to have. A PSIC score difference of 1.5 or above is considered to be damaging.

Analysis of functional nsSNPs and estimation of risk score by FASTSNP

The Functional Analysis and Selection Tool for Single Nucleotide Polymorphism (FASTSNP) is a web server (<http://fastsnp.ibms.sinica.edu.tw/>) which connects many programs and databases for processing analysis [23]. We used FASTSNP for the prediction of the functional effect of nsSNPs and estimation of their risk score. FASTSNP uses a decision tree for prioritizing the functional effect and estimating risk score. The nsSNPs were submitted for FASTSNP analysis and output files were displayed as a decision tree.

Results

SNP dataset from dbSNP

The COL1A1 gene investigated in this work was retrieved from the dbSNP database. It contained a total of 716 SNPs, of which 247 were nsSNPs, 25 were synonymous SNPs, and 32 were in non-coding regions, which comprise 1 SNP in the 5' UTR and 31 SNPs in the 3' UTR. The rest were in the intron region. We selected non-synonymous coding SNPs for our investigation.

Identification of functional nsSNP by I-Mutant 2.0

The more negative the free energy value (DDG value), the more likely a given point mutation is to be less stable and deleterious. We obtained 23 nsSNPs that were found to be less stable by this server, as shown in Table I. Out of 23 nsSNPs, 5 nsSNPs, namely rs1059454, rs17853657, rs17857117, rs41316719 and rs72656344, showed a DDG value of > -1.0 . The remaining nsSNPs showed a DDG value of < -1.0 , as depicted in Table I. Out of 23 nsSNPs that showed negative DDG, three nsSNPs, namely rs17853657, rs17857117 and rs57377812, changed their amino acid from non-polar to polar amino acid, and two nsSNPs, namely rs1059454 and rs72656307, changed their amino acids from polar to non-polar. Four nsSNPs, namely rs1135345, rs1800211, rs72656344 and rs72656351, changed their amino acid from polar to polar mutation and the remaining ones changed from non-polar to non-polar mutation. Since the amino acid mutations in the first five nsSNPs changed their physicochemical properties, we considered these nsSNPs to be less stable and deleterious by this analysis.

Predictions of deleterious and damaging coding nsSNPs

Protein conservation analysis was performed using a sequence-homology based tool, SIFT. Two hundred and forty-seven nsSNPs retrieved from the COL1A1 gene were submitted independently to the SIFT program to check its tolerance index. Our results showed that 19 nsSNPs were deleterious, having a tolerance index score of ≤ 0.05 . The results are shown in Table I. We observed that, out of 19 deleterious nsSNPs, 12 nsSNPs showed a highly deleterious tolerance index score of 0.00. Among these deleterious 19 nsSNPs, two nsSNPs showed a nucleotide change from A→G, one from A→C, one from C→T, two from C→G and the other 13 from G→T (Table I). Also, according to the SIFT results, three nsSNPs, namely rs17853657, rs17857117 and rs57377812, changed their amino acid from non-polar to polar amino acid, and one nsSNP, namely rs1059454, changed its amino acid from polar to non-polar amino acid in the mutant protein. We found that these four nsSNPs that are seen to be deleterious according to SIFT were also found less stable by the I-Mutant 2.0 server. Therefore, these four nsSNPs were found deleterious by this investigation.

Identification of damaged COL1A1 nsSNPs by PolyPhen server

To identify the COL1A1 nsSNPs that affected protein structure, the COL1A1 nsSNPs were analyzed for predicting a possible impact of amino acids on the structure and function of the protein using

the PolyPhen server. The COL1A1 protein sequence (NP_000079) with each nsSNP position and their 2 amino acid variants was submitted as input for analyzing the protein structural change due to amino acids. Our result showed 7 nsSNPs, namely rs1059454, rs8179178, rs17853657, rs17857117, rs72656340, rs72656344 and rs72656351, to be probably damaging, with a PSIC score difference between 2.0 and 3.5. The rs1059454, rs17853657 and rs17857117 which were observed to be the cause of protein lower stability by the I-Mutant 2.0

server and SIFT were also predicted to be probably damaging by the PolyPhen server. In addition, the other four nsSNPs are highly confidently predicted as probably damaging nsSNPs and the remainder as benign by PolyPhen (Table I).

Investigation of functional effect and estimated risk of COL1A1 nsSNPs

In order to identify nsSNPs with a high possibility of having a functional effect, FASTSNP was

Table I. I-Mutant, SIFT and PolyPhen results of COL1A1

SNP ids	Alleles	Amino acid change	DDG	Tolerance index	Predicted impact	PolyPhen prediction
rs1059454	A/C	T1431P	-2.19	0.00	Intolerant	Probably damaging
rs1135345	A/G	E591K	-0.87	0.89	Tolerant	Benign
rs1135348	C/G	G1019A	-0.23	0.00	Intolerant	Benign
rs1800211	A/G	R564H	-0.36	0.36	Tolerant	Benign
rs8179178	G/T	G197C	0.35	0.00	Intolerant	Probably damaging
rs17853657	G/T	P1460H	-1.71	0.03	Intolerant	Probably damaging
rs17857117	C/G	P1438R	-1.12	0.00	Intolerant	Probably damaging
rs41316713	A/G	R1141Q	0.15	0.04	Intolerant	Benign
rs41316719	A/G	V1177I	-1.21	0.00	Intolerant	Benign
rs57377812	C/T	G476R	-0.37	0.01	Intolerant	Benign
rs66548636	G/T	G389C	-0.22	0.32	Tolerant	Benign
rs66761141	G/T	G407C	0.05	0.00	Intolerant	Benign
rs66893386	G/T	G404C	-0.18	0.06	Tolerant	Benign
rs66929517	G/T	G815V	-0.09	0.09	Tolerant	Benign
rs66948146	G/T	G1187V	0.18	0.04	Intolerant	Benign
rs67182491	G/T	G383C	0.33	0.06	Tolerant	Benign
rs67445413	G/T	G866C	0.00	0.04	Intolerant	Benign
rs67682641	G/T	G530C	-0.21	0.00	Intolerant	Benign
rs72656307	C/T	R1093C	-0.35	0.38	Tolerant	Benign
rs72656312	G/T	G1124C	0.73	0.06	Tolerant	Benign
rs72656318	G/T	G1145C	-0.12	0.00	Intolerant	Benign
rs72656321	G/T	G1151V	-0.61	0.02	Intolerant	Benign
rs72656324	G/T	G1166C	-0.44	0.00	Intolerant	Benign
rs72656329	G/T	G1178V	-0.51	0.01	Intolerant	Benign
rs72656331	G/T	G1184V	0.13	0.00	Intolerant	Benign
rs72656340	A/G	M1264V	-0.70	0.78	Tolerant	Probably damaging
rs72656343	G/T	W1312C	-0.05	0.11	Tolerant	Benign
rs72656344	C/T	H1323Y	-2.15	1.00	Tolerant	Probably damaging
rs72656351	G/T	D1441Y	-0.99	0.00	Intolerant	Probably damaging
rs72667029	G/T	G200V	-0.29	0.29	Tolerant	Benign
rs72667031	G/T	G203V	0.37	0.06	Tolerant	Benign
rs72667037	G/T	G221C	0.55	0.03	Intolerant	Benign
rs72667038	G/T	G224C	-0.23	0.06	Tolerant	Benign

SNP IDs in bold are predicted to be highly polymorphic

applied for the detection of nsSNP influence on cellular and molecular biological function, e.g. transcriptional and splicing regulation. In addition the estimation of risk score was also calculated by FASTSNP. The functional effect and estimated risk of COL1A1 nsSNPs are shown in Table II. Eight COL1A1 nsSNPs exhibited a medium-high risk score (risk score = 3-4). The functional nsSNPs were rs1059454, rs8179178, rs17853657, rs17857117, rs41316713, rs41316719, rs72656312 and rs72656329. The remain-

ing nsSNPs showed low-medium risk (risk score = 2-3). The two functional nsSNPs (rs72656329 and rs41316719) detected by FASTSNP were also predicted to be polymorphic by I-Mutant 2.0 and SIFT. The nsSNPs rs72656312 and rs41316713 were also predicted to be deleterious by SIFT. The nsSNP rs8179178 was also predicted to be functionally damaging by SIFT and PolyPhen software. But the most important finding detected by FASTSNP was the three nsSNPs, namely rs1059454, rs17853657

Table II. Functional effect and estimated risk (FASTSNP)

SNP ids	Alleles	Amino acid change	Possible effect	Risk score
rs1059454	A/C	T1431P	Non-conservative change, splicing regulation	3-4
rs1135345	A/G	E591K	Conservative change, splicing regulation	2-3
rs1135348	C/G	G1019A	Conservative change, splicing regulation	2-3
rs1800211	A/G	R564H	Conservative change, splicing regulation	2-3
rs8179178	G/T	G197C	Splicing site	3-4
rs17853657	G/T	P1460H	Non-conservative change, splicing regulation	3-4
rs17857117	C/G	P1438R	Non-conservative change, splicing regulation	3-4
rs41316713	A/G	R1141Q	Non-conservative change, splicing regulation	3-4
rs41316719	A/G	V1177I	Splicing site	3-4
rs57377812	C/T	G476R	Conservative change, splicing regulation	2-3
rs66548636	G/T	G389C	Conservative change	2-3
rs66761141	G/T	G407C	Conservative change	2-3
rs66893386	G/T	G404C	Conservative change, splicing regulation	2-3
rs66929517	G/T	G815V	Conservative change, splicing regulation	2-3
rs66948146	G/T	G1187V	Conservative change, splicing regulation	2-3
rs67182491	G/T	G383C	Conservative change, splicing regulation	2-3
rs67445413	G/T	G866C	Conservative change	2-3
rs67682641	G/T	G530C	Conservative change	2-3
rs72656307	C/T	R1093C	Conservative change	2-3
rs72656312	G/T	G1124C	Splicing site	3-4
rs72656318	G/T	G1145C	Conservative change, splicing regulation	2-3
rs72656321	G/T	G1151V	Conservative change, splicing regulation	2-3
rs72656324	G/T	G1166C	Conservative change, splicing regulation	2-3
rs72656329	G/T	G1178V	Splicing site	3-4
rs72656331	G/T	G1184V	Conservative change, splicing regulation	2-3
rs72656340	A/G	M1264V	Conservative change, splicing regulation	2-3
rs72656343	G/T	W1312C	Conservative change, splicing regulation	2-3
rs72656344	C/T	H1323Y	Conservative change, splicing regulation	2-3
rs72656351	G/T	D1441Y	Conservative change	2-3
rs72667029	G/T	G200V	Conservative change, splicing regulation	2-3
rs72667031	G/T	G203V	Conservative change, splicing regulation	2-3
rs72667037	G/T	G221C	Conservative change, splicing regulation	2-3
rs72667038	G/T	G224C	Conservative change, splicing regulation	2-3

Note: SNP IDs in bold are predicted to be highly polymorphic

and rs17857117, that were also found polymorphic by I-Mutant 2.0, SIFT as well as by PolyPhen.

Discussion

Our analysis revealed 247 SNPs as non-synonymous out of which 5 nsSNPs, namely rs1059454, rs17853657, rs17857117, rs41316719 and rs72656344, were found to be least stable by I-Mutant 2.0 with a DDG value of > -1.0 . Four nsSNPs, namely rs17853657, rs17857117, rs57377812 and rs1059454, showed a highly deleterious tolerance index score of 0.00 with a change in their physicochemical properties by the SIFT server. Seven nsSNPs, namely rs1059454, rs8179178, rs17853657, rs17857117, rs72656340, rs72656344 and rs72656351, were found to be probably damaging, with a PSIC score difference between 2.0 and 3.5 by the PolyPhen server. Three nsSNPs, namely rs1059454, rs17853657 and rs17857117, were found to be highly polymorphic with a risk score of 3-4 with a possible effect of non-conservative change and splicing regulation by FASTSNP.

A major interest in human genetics is to distinguish mutations that are functionally neutral from those that contribute to disease. Amino acid substitutions currently account for approximately half of the known gene lesions responsible for human inherited disease. Therefore, the identification of nsSNPs that affect protein functions and relate to disease is an important task. The effect of many nsSNPs will probably be neutral as natural selection will have removed mutations at essential positions. Assessment of non-neutral SNPs is mainly based on phylogenetic information (i.e. correlation with residue conservation) extended to a certain degree with structural approaches. However, there is increasing evidence that many human disease genes are the result of exonic or non-coding mutations affecting regulatory regions [14]. Much attention has been focused on modeling by different methods the possible phenotypic effect of SNPs that cause amino acid changes, and only recently has interest focused on functional SNPs affecting regulatory regions or the splicing process. Moreover, because of their widespread distribution on the species genome, SNPs are particularly important and valuable as genetic makers in research on diseases and the corresponding drugs. To date, millions of human SNPs have been reported by high-throughput methods. The vast number of SNPs causes a challenge for biologists and bioinformaticians although they provide a lot of information about the relationships between individuals. Besides numerous ongoing efforts to identify millions of these SNPs, there is now also a focus on studying associations between disease risk and these genetic variations using a molecular epidemiological approach. This plethora of SNPs points out a major difficulty faced by scientists in planning costly population-based

genotyping, which is to choose target SNPs that are most likely to affect phenotypic functions and ultimately contribute to disease development [14].

Currently, most molecular studies focus on SNPs located in coding and regulatory regions, yet many of these studies have been unable to detect significant associations between SNPs and disease susceptibility. To develop a coherent approach for prioritizing SNP selection for genotyping in molecular studies, an evolutionary perspective to SNP screening is applied. The hypothesis is that amino acids conserved across species are more likely to be functionally significant. Therefore, SNPs that change these amino acids might be more likely to be associated with disease susceptibility. It is becoming clear that application of the molecular evolutionary approach may be a powerful tool for prioritizing SNPs to be genotyped in future molecular epidemiological studies [14]. Therefore, our analysis will provide useful information in selecting SNPs of the COL1A1 gene that are likely to have a potential functional impact.

Although computational tools show their potential in reducing the number of nsSNPs for disease association studies by filtering nsSNPs that are most likely to be disease related, error predictions do occur. Various computational tools used in this analysis determine the functional effects of SNPs only with respect to a single biological function. Therefore, much time and effort is required from researchers to identify the appropriate tools and interpret the predictions. There are also some aspects affecting the prediction correctness for prediction programs like SIFT and PolyPhen. SIFT and PolyPhen depend on diverse databases for SNP information. Polluted databases with incorrect SNP reports and bias of the data towards disease-associated allelic variants are likely to lead to over-prediction of the number of deleterious nsSNPs [24]. Furthermore, tools finding SNPs may identify base alterations between the functional gene and a pseudogene and mistakenly report these alterations as SNPs in the functional protein. Including nsSNPs mistakenly mapped from pseudogenes in the SNP database will affect the prediction accuracy of predictive tools using SNP information from these databases [25].

In conclusion, in our analysis, three nsSNPs (rs1059454, rs17853657 and rs17857117) were found to be less stable, deleterious, probably damaging and to have a high risk score by I-Mutant 2.0, SIFT, PolyPhen and FASTSNP, respectively. We therefore conclude that these three nsSNPs are potentially functionally polymorphic. To those conducting large-scale population-based epidemiological studies, the idea of prioritizing nsSNPs in the investigation of association of SNPs with disease risk is of great interest. The use of these servers to select potentially polymorphic nsSNPs for epidemiological stu-

dies can be an efficient way to explore the role of genetic variation in disease risk and to curtail cost. Furthermore, the predicted impact of these nsSNPs can be tested using animal models or cell lines to determine whether functionality of the protein has indeed been altered.

References

1. Yazici S, Korkmaz U, Erkan M, et al. The effect of breast-feeding duration on bone mineral density in postmenopausal Turkish women: a population-based study. *Arch Med Sci* 2011; 7: 486-92.
2. Sadat-Ali M, AlElq A. Osteoporosis among male Saudi Arabs: a pilot study. *Ann Saudi Med* 2006; 26: 450-4.
3. Dolan P, Torgerson DJ. The cost of treating osteoporotic fractures in the United Kingdom female population. *Osteoporos Int* 2000; 11: 551-2.
4. Aronow WS. Osteoporosis, osteopenia, and atherosclerotic vascular disease. *Arch Med Sci* 2011; 7: 21-6.
5. Stover DA, Verrelli BC. Comparative vertebrate evolutionary analyses of type I collagen: potential of COL1a1 gene structure and intron variation for common bone-related diseases. *Mol Biol Evol* 2011; 28: 533-42.
6. Rajasekaran R, Sudandiradoss C, Doss CG, Sethumadhavan R. Identification and in silico analysis of functional SNPs of the BRCA1 gene. *Genomics* 2007; 90: 447-52.
7. Dryja TP, McGee TL, Hahn LB, et al. Mutations within the rhodopsin gene in patients with autosomal dominant retinitis pigmentosa. *N Engl J Med* 1990; 323: 1302-7.
8. Smith EP, Boyd J, Frank GR, et al. Estrogen resistance caused by a mutation in the estrogen-receptor gene in a man. *N Engl J Med* 1994; 331: 1056-61.
9. Barroso I, Gurnell M, Crowley VE, et al. Dominant negative mutations in human PPAR γ associated with severe insulin resistance, diabetes mellitus and hypertension. *Nature* 1999; 402: 880-3.
10. Thomas R, McConnell R, Whittacker J, Kirkpatrick P, Bradley J, Sandford R. Identification of mutations in the repeated part of the autosomal dominant polycystic kidney disease type 1 gene, PKD1, by long-range PCR. *Am J Hum Genet* 1999; 65: 39-49.
11. Yoshida A, Huang IY, Ikawa M. Molecular abnormality of an inactive aldehyde dehydrogenase variant commonly found in Orientals. *Proc Natl Acad Sci U S A* 1984; 81: 258-61.
12. Jaruzelska J, Abadie V, d'Aubenton-Carafa Y, Brody E, Munnich A, Marie J. In vitro splicing deficiency induced by a C to T mutation at position -3 in the intron 10 acceptor site of the phenylalanine hydroxylase gene in a patient with phenylketonuria. *J Biol Chem* 1995; 270: 20370-5.
13. Johnson MM, Houck J, Chen C. Screening for deleterious nonsynonymous single-nucleotide polymorphisms in genes involved in steroid hormone metabolism and response. *Cancer Epidemiol Biomarkers Prev* 2005; 14: 1326-9.
14. Doss CG, Sethumadhavan R. Investigation on the role of nsSNPs in HNPCC genes – a bioinformatics approach. *J Biomed Sci* 2009; 16: 42.
15. Sherry ST, Ward MH, Kholodov M, et al. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* 2001; 29: 308-11.
16. Capriotti E, Fariselli P, Casadio R. I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Res* 2005; 33: w306-10.
17. Ng PC, Henikoff S. Predicting deleterious amino acid substitutions. *Genome Res* 2001; 11: 863-74.
18. Altschul SF, Madden TL, Schäffer AA, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997; 25: 3389-402.
19. Sjölander K, Karplus K, Brown M, et al. Dirichlet mixtures: a method for improved detection of weak but significant protein sequence homology. *Comput Appl Biosci* 1996; 12: 327-45.
20. Ng PC, Henikoff S. SIFT: predicting amino acid changes that affect protein function. *Nucl Acids Res* 2003; 31: 3812-4.
21. Sunyaev S, Ramensky V, Bork P. Towards a structural basis of human non-synonymous single nucleotide polymorphisms. *Trends Genet* 2000; 16: 198-200.
22. 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* 2004; 83: 970-9.
23. 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: W635-41.
24. Ramensky V, Bork P, Sunyaev S. Human non-synonymous SNPs: server and survey. *Nucleic Acids Res* 2002; 30: 3894-900.
25. Ng PC, Henikoff S. Accounting for human polymorphisms predicted to affect protein function. *Genome Res* 2002; 12: 436-46.