International Journal of Chemical and Biochemical Sciences
(ISSN 2226-9614)

Journal Home page: www.iscientific.org/Journal.html

In silico analysis of single nucleotide polymorphisms of NAT-2 gene and its implications in periodontitis

Sangeetha Subramanian[1], Kavitha Sanjeev[2], PSG Prakash[1], Devapriya Appukuttan[1]*

[1]*Department of Periodontics, SRM Dental College, Ramapuram, Chennai, India.*

[2]*Department of Conservative Dentistry and Endodontics, SRM Dental College, Ramapuram, Chennai, India.*

Abstract

Single Nucleotide Polymorphisms **(**SNPs) are the most common source of genetic polymorphism and they may bring about changes in the translated amino acid sequence. The N-Acetyl transferase-2 (NAT-2) gene polymorphism confers changes in the acetylation process and characterizes smokers into different acetylation phenotypes. This gene environment interaction is very significant and plays a major role in the occurence and progression of inflammatory diseases like periodontitis. Bio-computational methods like SIFT, Polyphen-2 and Provean can predict the structural and functional consequences of the target proteins. The study observed that NAT-2 gene polymorphisms at sites 341T>C and 590G>A were found to be deleterious among the six sites analysed using in silico approach.

Keywords: In silico, NAT-2 gene polymorphism, Periodontitis, Polyphen-2, Provean.

Full length article **Corresponding Author*, e-mail: sangeetha_doc@yahoo.com

1. Introduction

SNPs are the most common source of genetic polymorphism in the human genome and accounts for about 90% of all human DNA polymorphism [1]. The SNPs that are non-synonymous stand in need of consideration as they bring about changes in the translated amino acid sequence. It is probable that changes in resultant amino acid may affect the structural and functional variation of coded protein [2]. The alteration in protein function may have an impact on the homeostasis and health and is known to have been associated with many diseases. However it is worth to note that not all non-synonymous SNPs (nsSNPs) causes protein damage leading to disease. Non-synonymous SNPs may affect the protein function in various ways either by reducing its solubility or by destabilizing its structure which may affect gene regulation by altering transcription and translation [3]. Some of the protein changes remain neutral without causing deleterious effects. Hence the identification of such genetic variations and their possible impact on homeostasis needs to be explored. *N*-acetyltransferase, a cytosolic phase II conjugation enzyme primarily deals with the metabolism and acetylation of diverse number of aromatic and heterocyclic amine carcinogens present in the diet, cigarette smoke and drugs [4]. The enzyme is encoded by a polymorphic gene present in the short arm of submetacentric human chromosome 8. The NAT2 acetylation polymorphism is very important in clinical pharmacology and toxicology based on which the individuals have been characterized into different acetylation phenotypes. The NAT2 genotype confers a slow, intermediate or rapid acetylation phenotype, leading to variations in drug metabolic rates and susceptibility to toxicity [5]. About 65 allele variants of NAT2 have been reported among human populations and in the coding region, six frequently occurring SNPs are 282C>T, 341T>C, 481C>T, 590G>A, 803A>G and 857G>A [6]. Most of these variant alleles are implicated in inflammatory diseases and cancer pathology [7,8,9]. Periodontitis, an immuno inflammatory disease is influenced and propagated by risk factors like smoking. As NAT 2 gene is involved in the metabolizing smoke contents, its polymorphism may have an effect in the occurrence of Periodontitis. Therefore, bio-computational methods that predicts the structural and functional consequences of the target proteins is a prerequisite to understand the nature of nsSNPs. The software tools like the Sorting Intolerant from Tolerant (SIFT) algorithm provides users with the predictions whether the given amino acid substitution can possibly affect protein function based on sequence homology and the physical properties of amino acids [10]. Basically SIFT distinguishes the deleterious nsSNPs from the neutral one and has become one of the standard tools for characterizing missense variation. There are other in silico approaches like PolyPhen-2, Provean, MutPred 2, SNP & GO which can identify the effect of the resultant amino acid due to SNPs. The current

study is based on computational analysis of NAT-2 gene to evaluate whether the mutations in the NAT2 gene leads to functional alteration and phenotypic variations and also to assess the importance of the NAT2 genetic polymorphism in inflammatory diseases like periodontitis using an insilico approach.

2.Materials and Method

There are various computational techniques that can be used to identify missense variants which may have an effect on the structure and function of a protein [11,12]. The prediction based on the computational approaches could narrow down the candidate mutations for further validation.

2.1 Data collection

The NCBI database of SNPs was used to obtain the SNPs information including SNP IDs, gene IDs of the human arylamine N-acetyltransferase-2 gene. The protein sequence (FASTA format) and protein ID were retrieved from the unitprot.org.

2.1.1 Sorting Intolerant from Tolerant (SIFT)

SIFT is a biotool (https://sift.bii.a-star.edu.sg) used to predict if an amino acid substitution in a protein will have an effect on function. This is based on sequence homology and the physical properties of amino acids and ascertain whether the mutation is tolerated or not. The reference SNP cluster ID of the polymorphic sites of NAT-2 gene were retrieved from dbSNP NCBI database. It was then entered into SIFT dbSNP rsIDs column and submitted for prediction of its function. The SIFT score ranges from 0 to 1 and rsID that scores less than 0.05 are expected to be deleterious and that which scores above 0.05 appears to be tolerated.

2.1.2 Polymorphism Phenotyping-2 (Polyphen-2)

PolyPhen-2 is available as a Web server (http://genetics.bwh.harvard.edu/pph2/) which predicts the possible impact of amino acid substitutions based on the stability and function of human proteins using structural and comparative evolutionary considerations. In the polyphen-2 web interface, UniProtKB accession number for NAT-2 gene is provided in the Protein or SNP identifier text box. Then the Position of the substitution in the protein sequence along with wild-type (query sequence) amino acid residue AA_1 and the substitution residue AA_2 entered and submitted. Prediction outcome can be probably damaging, possibly damaging, *or* benign depending on the scores that ranges from 0.0 (tolerated) to 1.0 (deleterious). Variants with scores of 0.0 are predicted to be benign; 0.15 to 1.0 are possibly damaging and 0.85 to 1.0 are more confidently predicted to be damaging. The predicted damaging effect is also indicated by a vertical black marker inside a color gradient bar, where green is benign and red is damaging.

2.1.3 Protein Variation Effect Analyzer (PROVEAN)

PROVEAN (https://provean.jcvi.org) can generate predictions not only for single amino acid substitutions but also for multiple amino acid substitutions, insertions, and deletions using the same underlying scoring scheme. A protein sequence in FASTA format or amino acid variations are entered as input which is subjected to a BLAST search to identify homologous sequences and generates PROVEAN scores. The cut off value of provean score is -2.5. The amino acid variants with a score equal to or below -2.5 are considered "deleterious and above -2.5 are considered "neutral."

2.1.4 MutPred 2

The query input requires FASTA format of the protein along with the sequence ID. The list of substitutions of aminoacid and its position is also entered. The output consists of a general score which is the probability to ascertain whether the amino acid substitution is pathogenic. In MutPred, a score of 0.50 is suggestive of pathogenicity if interpreted as a probability.

2.1.5SNP & GO

The SNP & GO (GO-Gene Ontology) is accessed through (http://snps-and-go.biocomp.unibo.it/snps-and-go/) which uses evolutionary information and function as encoded in the GO sequence-associated terms. This biotool predicts whether the variation in amino acid has any effect on the gene function. The prediction output of SNPs&GO is a score ranging between 0 and 1 that represents the probability of a SAV (Single amino acid variation) to be pathogenic. Depending on the score, a reliability index (RI) ranging from 0 to 10 is defined to estimate the level of confidence of the prediction.

3. Results and Discussion

In the present study, about five biotools have been used to predict the effects of amino acid replacements in the coding region of NAT-2 gene. SIFT, Polyphen-2 and Provean are sequence based predictions which gives an indication as to whether the mutation is tolerant or deleterious. SNPs&GO and mutpred-2 represents single amino acid variation to be pathogenic not. Table 1 lists the effect of six polymorphic sites of NAT-2 gene which are analysed using SIFT approach. The SIFT ranged from score 0.043 to 1 with 4 sites being tolerated and two sites deleterious. 282C>T and 481C>T are not considered for analysis with other biotools as they are silent mutations. Table 2 shows Polyphen-2 scores two sites with benign condition and two sites with damaging positions. Provean predictions for NAT-2 gene sites resulted with two neutral and two deleterious sites. Mutpred2 analysis carried out represented only one site being pathogenic out of the four sites. In SNPs & GO evaluation all the six sites are found to be neutral Table 3. Cigarette smoking is one of the well established risk factors in periodontitis [13]. There is now remarkable evidence that genes that metabolise smoke contents also play a role in the predisposition and progression of periodontal disease [14]. The gene environment interaction studies may possibly provide valuable insights into the pathogenesis of complex periodontal diseases [15]. The NAT2 gene has significant implications in categorizing the individuals as fast or slow phenotype. As NAT2 gene is involved in smoke metabolism, the assessment of the polymorphic sites is very much relevant to understand the effect of its SNP in the pathogenicity of inflammatory diseases like periodontitis [16]. As far as NAT-2 gene is concerned, the frequently occurring six polymorphic sites in Indian population were analysed to determine its effect for periodontitis patients [17].

Table 1: SIFT prediction for SNPs of NAT-2 gene

Polymorphic sites of NAT-2 gene	ID	Amino acid substitution	SIFT	
			Score	Prediction
282C>T	rs1041983	Y94Y	1	Tolerated
341T>C	rs#1801280	I114T	0.043	Deleterious
481C>T	rs1799929	L161L	1	Tolerated
590G>A	rs1799930	R197Q	0.091	Deleterious
803A>G	rs#1208	K268R	1	Tolerated
857G>A	rs#1799931	G286E	0.967	Tolerated

Table 2: Polyphen-2 and Provean prediction for SNPs of NAT-2 gene

Polymorphic sites of NAT-2 gene	Polyphen-2 score	Sensitvity	Specificity	Prediction	Provean	Prediction
341T>C	0.859	0.83	0.93	Possibly damaging	-4.178	Deleterious
590G>A	1.000	0	1.00	Probably damaging	-2.842	Deleterious
803A>G	0.011	0.96	0.78	Benign	1.425	Neutral
857G>A	0.077	0.93	0.85	Benign	1.37	Neutral

Table 3: MutPred-2 and SNPs & GO prediction for SNPs of NAT-2 gen

Polymorphic sites of NAT-2 gene	MutPred-2	SNPs&GO Reliability score	Prediction
341T>C	0.539*	6	Neutral
590G>A	0.193	8	Neutral
803A>G	0.030	10	Neutral
857G>A	0.257	9	Neutral

*Pathogenicity; Altered Ordered interface (p Value=0.05); Altered Metal binding (p Value=0.03)

In silico tools like SIFT, Polyphen-2 and a variety of other techniques help us to predict whether a specific polymorphic site is going to have a functional effect [18,19]. This is very useful in a sense that if case control studies are performed, in silico approach enlighten which polymorphic sites are to be considered for a potential pathogenic effect. Among the NAT-2 gene sites only 341T>C and 590G>A are found to be deleterious with SIFT analysis and all the other sites were tolerated. The 341T>C variant is nonsynonymous substituting the amino acid isoleucine to threonine in the amino acid sequence at position 114. Similarly 590G>A which is also nonsynonymous has been substituted with glutamine in the amino acid sequence instead of arginine. The sites were again confirmed with Polyphen-2 where 341T>C is found to be Possibly damaging and 590G>A was observed to be Probably damaging. The Provean tool assessed the protein sequence and found that these two sites with SAV are potentially deleterious. Mutpred2 analysis of NAT-2 gene was presented in table 3 along with SNPs & GO tool. According to Mutpred2 analysis, only 341T>C resulted to be pathogenic whereas in SNPs &GO tool none of the sites seem to be pathogenic. Variations in the protein have influence not only in the protein structure but also its stability and function. The deleterious sites of NAT-2 gene found in different insilico approaches are to be subjected for experimental assays where smokers with periodontitis patients were to be included as cohorts. The insilico tools like SIFT, Polyphen-2, Provean were consistent in the prediction of 341T>C , 590G>A sites as intolerant which leads to the assumption that it can alter the function of resultant NAT-2 function. Although in silico tools predict the probability of a harmful effect in genetic variations in protein function and structure, it is reported that in silico software may sometimes show inaccurate prediction and further experimental analysis is warranted [20].

4. Conclusion

The NAT-2 enzyme polymorphic sites predicted with insilico approaches namely SIFT, Polyphen-2, Provean reported that 341T>C and 590G>A are deleterious. This may affect the function of NAT-2 enzyme and alter the acetylator phenotypes of the individuals harbouring these sites.

References

[1] S. Nares. (2002). The genetic relationship to periodontal disease. Periodontol .2000.32:36-49.

[2] M.S. Hassan, A.A. Shaalan, M.I. Dessouky, A.E. Abdelnaiem & M. El Hefnawi. (2019). Evaluation of computational techniques for predicting non-synonymous single nucleotide variants pathogenicity. Genomics.111: 869-882

[3] A. Ajith & U. Subbiah. (2023). In silico screening of non-synonymous SNPs in human TUFT1 gene. Journal of Genetic Engineering and Biotechnology. 21:95.

[4] C.B. Ambrosone, S. Kropp, J. Yang, S. Yao, P.G. Shields & J. Chang-Claude. (2008). Cigarette smoking, N-acetyltransferase 2 genotypes, and breast cancer risk: pooled analysis and meta-analysis. Cancer Epidemiology, Biomarkers & Prevention. 17:15-26.

[5] D.W. Hein. (2009) N-acetyltransferase SNPs: emerging concepts serve as a paradigm for understanding complexities of personalized medicine. Expert Opinion on Drug Metabolism & Toxicology. 5:353-366.

[6] N. Khan, V. Pande & A. Das. (2013). NAT2 sequence polymorphisms and acetylation profiles in Indians. Pharmacogenomics.14:289-303.

[7] D.W. Hein. (2006). N-acetyltransferase 2 genetic polymorphism: effects of carcinogen and haplotype on urinary bladder cancer risk. Oncogene. 25:1649-1658.

[8] Z. Huang, L. Yuan, Z. Jiang & D. Wang. (2015). Associations of polymorphisms in NAT2 gene with risk and metastasis of osteosarcoma in young Chinese population. OncoTargets and Therapy. 8:2675-2680.

[9] X.L. Zhuo, J.J. Ling, Y. Zhou, H.Y. Zhao, Y.F. Song & Y.H. Tan. (2012). NAT2 polymorphisms with oral carcinoma susceptibility: a meta-analysis. Molecular Biology Reports. 39:8813-8819.

[10] P.C. Ng & S. Henikoff. (2003). SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Research. 31:3812-3814.

[11] H. Tang & P.D. Thomas. (2016). Tools for Predicting the Functional Impact of Nonsynonymous Genetic Variation. Genetics. 203:635-647.

[12] T. Gong, L. Yang, F. Shen, H. Chen, Z. Pan, Q. Zhang, Y. Jiang, F. Zhong, P. Yang & Y. Zhang. (2021). Computational and Mass Spectrometry-Based Approach Identify Deleterious Non-Synonymous Single Nucleotide Polymorphisms (nsSNPs) in JMJD6. Molecules. 26:4653.

[13] E.T. Knight, J. Liu, G.J. Seymour, C.M. Jr Faggion & M.P. Cullinan. (2016). Risk factors that may modify the innate and adaptive immune responses in periodontal diseases. Periodontology. 71:22-51.

[14] T. Kocher, H. Sawaf, J. Fanghänel, R. Timm & P. Meisel. (2002). Association between bone loss in periodontal disease and polymorphism of N-acetyltransferase (NAT2). Journal of Clinical Periodontology. 29:21-27.

[15] C. Ober & D. Vercelli. (2011). Gene-environment interactions in human disease: nuisance or opportunity? Trends in Genetics. 27:107-115.

[16] P. Meisel, R. Timm, H. Sawaf, J. Fanghänel, W. Siegmund & T. Kocher. (2000). Polymorphism of the N-acetyltransferase (NAT2), smoking and the potential risk of periodontal disease. Archives of Toxicology.74:343-348.

[17] A. Anitha & M. Banerjee. (2003). Arylamine N-acetyltransferase 2 polymorphism in the ethnic populations of South India. International Journal of Molecular Medicine. 11:125-131

[18] S.E. Flanagan, A.M. Patch & Ellard. (2010). Using SIFT and PolyPhen to predict loss-of-function and gain-of-function mutations. Genetic Testing and Molecular Biomarkers. 14:533-537.

[19] Y.B. Jarrar, A.A. Balasmeh & W. Jarrar. (2018). Sequence analysis of the N-acetyltransferase 2 gene (NAT2) among Jordanian volunteers. Libyan Journal of Medicine.13:1408381.

[20] S. Ali, U. Ali, A. Qamar, I. Zafar, M. Yaqoob, Q.U. Ain, S. Rashid, R. Sharma, H.A. Nafidi, Y.A.B. Jardan & M. Bourhia. (2023) Predicting the effects of rare genetic variants on oncogenic signaling pathways: A computational analysis of HRAS protein function. Frontiers in Chemistry. 11:1173624.