Bulletin of Pure and Applied Sciences
Vol.39A (Zoology), No.1,
January-June 2020: P.172-188
Original Research Article

Print version ISSN 0970 0765
Online version ISSN 2320 3188
DOI 10.5958/2320-3188.2020.00020.0
Available online at www.bpasjournals.com

Evaluation of Nonsynonymous Single Nucleotide Variations in *NOS2* Gene Identified Through Whole Exome Sequencing: A Bioinformatics Approach

[1]Tejaswini Prakash
[2]MS. Basavaraj
[3]Nallur B Ramachandra*

Author's Affiliation:
[1]Genetics and Genomics Lab, Department of Studies in Genetics and Genomics, University of Mysore, Manasagangothri, Mysuru, Karnataka 570006, India
E-mail: prakash.tejaswini@gmail.com
[2]University Health Centre, University of Mysore, Mysuru, Karnataka 570005, India
E-mail: drmsbasavaraj@gmail.com
[3]Genetics and Genomics Lab, Department of Studies in Genetics and Genomics, University of Mysore, Manasagangothri, Mysuru, Karnataka 570006, India
E-mail: nallurbr@gmail.com

***Corresponding author:**
Nallur B Ramachandra
Genetics and Genomics Lab, Department of Studies in Genetics and Genomics, University of Mysore, Manasagangothri, Mysuru, Karnataka 570006, India

E-mail:
nallurbr@gmail.com

Received on 14.03.2020
Accepted on 07.05.2020

Abstract:
Majority of the human diseases are accounted for by nonsynonymous single nucleotide variations (nsSNVs) that occur in the coding region of genes and alter the amino acid residues at specific positions. The future of genomics research is in identification of nsSNVs that contribute to disease pathophysiology by disrupting protein function. A robust pipeline with several integrated computational prediction tools facilitates prioritization of significant disease-associated nsSNVs. In this study, we analysed 30 rare frequency missense variations in *NOS2* identified through Whole Exome Sequencing (WES) of six coronary artery disease (CAD) using multiple bioinformatics softwares and tools. We employed a stringent filtering workflow to identify and assess the pathogenic effect of nsSNVs on NOS2 structure and function. We used a combination of deleterious variation detection tools, protein stability changes prediction, post-translational modification site prediction, protein-protein interaction and enrichment analyses to discern disease-associated variations. Our findings implicate four amino acid substitutions - K730N, P769R, P958S and L1012S as candidates in pathological process of *NOS2*.

Keywords: Coronary artery disease, Whole Exome Sequencing, *In Silico* analyses, Bioinformatics, *NOS2*

INTRODUCTION

Coronary Artery Disease (CAD) is the leading cause of mortality, accounting for approximately 17 million deaths every year worldwide (Benjamin *et al.*, 2019). This can be attributed to the steady rise in the prevalence of various CAD-risk factors such as diabetes, obesity, hypercholesterolemia, hypertension, sedentary lifestyle, age and family history (Hajar, 2017). The most common pathological mechanism for clinical manifestation of CAD is atherosclerosis, a chronic inflammatory process elicited by progressive accumulation of lipids and fibrous elements in arterial walls, leading to platelet activation and thrombus formation, which eventually results in complications like angina pectoris, myocardial infarction (MI) and sudden cardiac death (Aldons, 2012). Much like most common diseases, CAD results from complex interplay between multiple genetic and environmental

factors with a myriad of mediators including growth factors, cell-adhesion molecules, cytokines & chemokines and nitric oxide (Singh *et al.*, 2002).

Recent advances in high-throughput sequencing technologies have led to the identification of numerous genomic variations, especially nonsynonymous single nucleotide variants (nsSNVs) that alter protein structure and function, leading to pathogenic phenotypes. Growth in number of these variants has made the task of determining potential protein disrupting nsSNVs laborious and time-consuming (Elkhattabi *et al.*, 2019). This can be resolved using computational approaches that distinguish neutral variants from pathogenic variants and facilitate prioritization of nsSNVs that are deleterious to the protein activity.

Whole Exome Sequencing (WES) is a powerful and reliable tool to identify population specific, low and rare frequency variations in candidate genes associated with complex, multifactorial diseases (Do *et al.*, 2012). In view of this, we performed WES on six CAD subjects to identify pathogenic rare variants in candidate genes for CAD.

MATERIALS AND METHODS

Whole exome sequencing (WES) at 100X coverage was performed on six subjects with clinically diagnosed coronary artery disease (CAD) who had an incidence of acute myocardial infarction (MI) and had undergone either percutaneous transluminal coronary angioplasty (PTCA) or coronary artery bypass graft surgery (CABG). All subjects had a medical history of type 2 diabetes mellitus (T2DM) with multiple affected members in the family and were between the ages 30 and 80 years. Informed consent and genetic registry was obtained from all subjects selected for WES. The study was approved by the Institutional Human Ethics Committee, University of Mysore.

Exome libraries were prepared using Agilent SureSelectXT Human All Exon V5 + UTRs kit and sequenced in Illumina HiSeq 2500. Data analysis was performed using Strand NGS software version 3.1, build 234566 (Strand Life Sciences, 2016) against hg19 reference genome. Annotation of variants was performed on wANNOVAR (Chang and Wang, 2012) for chromosomal position, altered nucleotide base, gene, transcript ID, amino-acid change, MAF, read depth, zygosity and mutation effects. The following custom-designed pipeline was used to identify pathogenic variants in CAD associated genes.

Gene Prioritization

The annotated files were filtered based on regions that a variant lies in, gene ID associated with variants and variant function. Genes that were annotated to be in the exonic region with functional consequences such as stop gain, frameshift insertions/deletions, nonsynonymous single nucleotide variants (nsSNV) were retained.

Genes with strong association to CAD were identified from existing literature using query terms "genetics of CAD", "candidate genes for CAD", "CAD and role of genetics", "genes for CAD". In addition, curated databases such as CADgene (Coronary Artery Disease Gene Database) (http://www.bioguo.org/CADgene/) (Liu *et al.*, 2011), T2DiACoD (A Gene Atlas of Type 2 Diabetes Mellitus Associated Complex Disorders) (www.http://t2diacod.igib.res.in/) (Rani *et al.*, 2017) and C/VD database (Cardio/Vascular Disease Database) (http://www.padb.org/cvd/) (Fernandes *et al.*, 2018) were also used to select candidate genes. Genes that were found to carry variations in all six subjects were selected for downstream analysis.

Collection of SNV information

The annotated files were screened for single nucleotide variants (SNVs) in the exonic region of *NOS2* gene with a read depth of >30 to ensure confidence of the called SNV. The resulting SNVs were classified based on minor allele frequency (MAF) into common SNPs (>5% MAF), low frequency variations (1-5% MAF) and rare variants (<5% MAF) (Bomba *et al.*, 2017). The list of low and rare frequency SNVs were screened for nsSNV and were verified in dbSNP database

(http://www.ncbi.nlm.nih.gov/snp/) (Sherry *et al.*, 2001) and Genome Aggregation Database (genomAD) (Karczewski *et al.*, 2019)
(https://gnomad.broadinstitute.org/gene/ENSG00000007171?dataset=gnomad_r2_1) using NOS2 as the query input. Duplicate nsSNVs among the six subjects were removed and only unique variations were retained.

Prediction of functional effect of missense variations on protein function

The pathogenicity of identified missense variations was assessed using ten *in silico* prediction tools: SIFT (Sim *et al.*, 2012), PolyPhen2 (Adzhubei *et al.*, 2013), MutationAssessor (Reva *et al.*, 2011), MutationTaster (Schwarz *et al.*, 2014), PhD-SNP (Capriotti *et al.*, 2006), SNAP2 (Hecht *et al.*, 2015), FATHHM (Shihab *et al.*, 2013), M-Cap (Jagadeesh *et al.*, 2016), PMut (López-Ferrando *et al.*, 2017) and PROVEAN (Choi and Chan, 2015). The tools employ different strategies to predict the effect of mutations on protein function and hence classify nsSNVs based on different features. Input information such as chromosome location, amino acid sequence of NOS2, reference and altered allele/amino acid, position of variation in the protein etc., were used to suit tool requirement.

Sequence Conservation Profiling

ConSurf web server (http://consurf.tau.ac.il/) was used to investigate evolutionary conservation of identified NOS2 amino acid variants based on phylogenetic relationships between sequence homologs using HMMER homolog search algorithm (Ashkenazy *et al.*, 2016). This web server gives the residues a score from 1 to 9, with 1 being the least conserved region and 9 being the highly conserved region. ConSurf was also used to predict the nature (buried/exposed; functional/structural) of the amino acid residue of interest.

Prediction of Protein Stability Changes

The change in protein stability brought about by deleterious nsSNVs was assessed using two prediction tools: I-Mutant v3.0 (http://gpcr2.biocomp.unibo.it/cgi/predictors/I-Mutant3.0/I-Mutant3.0.cgi) (Capriotti *et al.*, 2005) and MUpro (http://mupro.proteomics.ics.uci.edu/) (Cheng *et al.*, 2006) which function on machine learning algorithms to accurately predict Gibb's free energy change (DDG) upon single point mutations and direction of protein stability using protein sequence data. For both prediction tools, the amino acid sequence for NOS2 (NCBI Reference Sequence: NP_000616.3) obtained from NCBI protein database https://www.ncbi.nlm.nih.gov/protein/NP_000616.3) was used in FASTA format as input along with position of variation and altered amino acid. Rest were set to default parameters.

Post-translational Modification Sites Prediction

Post-translational modification (PTM) sites in NOS2 were predicted using ModPred (http://montana.informatics.indiana.edu/ModPred/index.html) (Pejaver *et al.*, 2014) web server. NOS2 protein sequence was submitted in FASTA format and queried for all 23 types of modifications.

Protein-Protein Interaction Prediction and Enrichment Analysis

Protein-protein interaction (PPI) network was predicted using STRING v11.0 (Search Tool for the Retrieval of Interacting Genes/Proteins) (https://string-db.org/) (Szklarczyk *et al.*, 2019) using NOS2 as the input protein name and *Homo sapiens* as the organism of interest. Text mining, experiments and databases were used as the active interaction sources, with other default parameters. Enrichment analysis was carried out on EnrichR (Kuleshov *et al.*, 2016) using primary interacting partners of NOS2 obtained from PPI network 'Pathways' category of enrichment was assessed for functional association of the queried gene set with a biological pathway annotated by KEGG and WikiPathways, Reactome and Panther databases. Pathways with significant p-value (<0.05) were selected.

RESULTS

WES data of six subjects who presented with CAD were analysed in the current study. The base quality score (Phred score) was >30 in all six samples, ascertaining high quality of base calls. We obtained an average of 2, 36, 207 variants called and annotated from all six subjects combined.

Gene Prioritization

A candidate gene list approach was employed to identify potential pathogenic variants in CAD associated genes. Filtering the annotated files for variations in the exonic region produced a set of approximately 11,000 genes. Review of literature and curated databases resulted in identification of 673 candidate genes related to CAD, which were manually screened against the genes in exonic regions and *NOS2* gene was identified to carry variations in all six subjects.

Collection of SNV information

Thirty distinctive missense rare variants in exonic region of *NOS2* NM_000625 transcript were identified, including 21 novel variations. All variations were found at a read depth of >30. Nine of the 30 nsSNVs (c.G3365A, c.C3265T, c.C2117T, c.T2226G, c.G3247T, c.C3139T, c.G3247A, c.G2249A and c.C2872T corresponding to p.R1122H, p.R1089W, p.P706L, p.N742K, p.G1083W, p.G1083R, p.R1047C, p.R750H and p.P958S) have been previously reported (rs748134272, rs201947529, rs201789784, rs773983928, rs775310737, rs772991341, rs150704221 and rs1312528971) while 21 of them were not described in dbSNP and genomAD and are hence considered novel variations. Additional information including chromosomal position, reference and altered base pair, cDNA position, consequent amino acid change, exon and associated dbSNP IDs are presented for all 30 rare variations in Table 1.

Prediction of functional effect of missense variations on protein function

All 30 missense variations of *NOS2* gene were analysed on a set of 10 prediction tools - SIFT, PolyPhen2, MutationAssessor, MutationTaster, PhD-SNP, SNAP2, FATHHM, M-Cap, PMut and PROVEAN to assess their functional impact on protein function. Out of 30, 10 variants - R1122H, H1042P, R1122C, G1083W, R1047C, G1015E, G1083R, L1012S, Y1055D and R990L were predicted to be damaging by all 10 prediction tools while 9 tools predicted 7 variations - R1089W, W704R, G659E, G774R, P769R, S945G and E920A to be deleterious. 8 of the 10 tools predicted 5 variations - E1125D, K730N, R750H, A680V and P958S as pathogenic whereas 4 variants - Y1123F, P706L, L662V and N742K were predicted to be deleterious by 7 tools. 6 tools predicted 3 variations - P706S, V1085M and Q899L as damaging to protein function. V882L was predicted neutral/benign by all prediction tools except for SIFT, MutationTaster and PROVEAN. We selected missense variations that were predicted to be damaging and deleterious by a minimum of 7 prediction tools for further analysis (Table 2).

Table 1: List of single base missense variations in *NOS2* identified from WES

Chromosome Position (bp)	Ref	Alt	Variation Type	Transcript ID	cDNA Position	Exon	AA Change	dbSNP ID
chr17: 26084359 - 26084359	T	G	Single base Missense	NM_000625	c.A3375C	27	p.E1125D	NA
chr17: 26084366 - 26084366	T	A			c.A3368T	27	p.Y1123F	NA
chr17: 26084369 - 26084369	C	T			c.G3365A	27	p.R1122H	rs748134272
chr17: 26085996 - 26085996	G	A			c.C3265T	26	p.R1089W	rs201947529
chr17: 26087090 - 26087090	T	G			c.A3125C	25	p.H1042P	NA
chr17: 26093592 - 26093592	C	A			c.G2190T	19	p.K730N	NA
chr17: 26094781 - 26094781	G	A			c.C2117T	18	p.P706L	rs201789784
chr17: 26094788 - 26094788	A	T			c.T2110A	18	p.W704R	NA
chr17: 26096053 - 26096053	G	C			c.C1984G	17	p.L662V	NA
chr17: 26096061 - 26096061	C	T			c.G1976A	17	p.G659E	NA
chr17: 26084370 - 26084370	G	A			c.C3364T	27	p.R1122C	NA
chr17: 26093556 - 26093556	A	C			c.T2226G	19	p.N742K	rs773983928
chr17: 26094782 - 26094782	G	A			c.C2116T	18	p.P706S	NA
chr17: 26086008 - 26086008	C	T			c.G3253A	26	p.V1085M	NA
chr17: 26086014 - 26086014	C	A			c.G3247T	26	p.G1083W	rs775310737
chr17: 26087076 - 26087076	G	A			c.C3139T	25	p.R1047C	rs772991341
chr17: 26087171 - 26087171	C	T			c.G3044A	25	p.G1015E	NA
chr17: 26089980 - 26089980	C	A			c.G2644T	22	p.V882L	NA
chr17: 26086014 - 26086014	C	T			c.G3247A	26	p.G1083R	rs775310737
chr17: 26087180 - 26087180	A	G			c.T3035C	25	p.L1012S	NA
chr17: 26089928 - 26089928	T	A			c.A2696T	22	p.Q899L	NA
chr17: 26092669 - 26092669	C	T			c.G2320A	20	p.G774R	NA
chr17: 26092683 - 26092683	G	C			c.C2306G	20	p.P769R	NA
chr17: 26086098 - 26086098	A	C			c.T3163G	26	p.Y1055D	NA
chr17: 26092740 - 26092740	C	T			c.G2249A	20	p.R750H	rs150704221
chr17: 26094859 - 26094859	G	A			c.C2039T	18	p.A680V	NA

chr17: 26087690 - 26087690	C	A			c.G2969T	24	p.R990L	NA
chr17: 26088186 - 26088186	G	A			c.C2872T	23	p.P958S	rs1312528971
chr17: 26088225 - 26088225	T	C			c.A2833G	23	p.S945G	NA
chr17: 26089865 - 26089865	T	G			c.A2759C	22	p.E920A	NA

Table 2: Prediction of damaging missense variants in NOS2 using *in silico* tools

AA Change	SNV Effect Prediction										Total
	SIFT	PolyPhen2 (HumDiv)	Mutation Assessor	MutationTaster	PhD-SNP	FATHMM	M-CAP	SNAP2	Pmut	PROVEAN	
E1125D	Damaging	Probably Damaging	Medium	Disease Causing	Disease	Tolerated	Possibly Pathogenic	Effect	TRUE	Neutral	8
Y1123F	Tolerated	Probably Damaging	Medium	Disease Causing	Disease	Tolerated	Possibly Pathogenic	Neutral	TRUE	Deleterious	7
R1122H	Damaging	Probably Damaging	Medium	Disease Causing	Disease	Damaging	Possibly Pathogenic	Effect	TRUE	Deleterious	10
R1089W	Damaging	Benign	High	Disease Causing	Disease	Damaging	Possibly Pathogenic	Effect	TRUE	Deleterious	9
H1042P	Damaging	Possibly Damaging	Medium	Disease Causing	Disease	Damaging	Possibly Pathogenic	Effect	TRUE	Deleterious	10
K730N	Damaging	Probably Damaging	Medium	Disease Causing	Neutral	Tolerated	Possibly Pathogenic	Effect	TRUE	Deleterious	8
P706L	Damaging	Probably Damaging	Medium	Disease Causing	Disease	Tolerated	Possibly Pathogenic	Neutral	FALSE	Deleterious	7
W704R	Damaging	Probably Damaging	Medium	Disease Causing	Disease	Tolerated	Possibly Pathogenic	Effect	TRUE	Deleterious	9
L662V	Damaging	Possibly Damaging	Medium	Disease Causing	Neutral	Tolerated	Possibly Pathogenic	Neutral	TRUE	Deleterious	7
G659E	Damaging	Probably Damaging	High	Disease Causing	Disease	Tolerated	Possibly Pathogenic	Effect	TRUE	Deleterious	9
R1122C	Damaging	Probably Damaging	High	Disease Causing	Disease	Damaging	Possibly Pathogenic	Effect	TRUE	Deleterious	10
N742K	Damaging	Possibly Damaging	Low	Disease Causing	Neutral	Tolerated	Possibly Pathogenic	Effect	TRUE	Deleterious	7
P706S	Damaging	Possibly	Medium	Disease	Neutral	Tolerated	Possibly	Neutral	FALSE	Deleterious	6

		Damaging		Causing			Pathogenic				
V1085M	Damaging	Benign	Low	Disease Causing	Disease	Damaging	Possibly Pathogenic	Neutral	TRUE	Neutral	6
G1083W	Damaging	Probably Damaging	High	Disease Causing	Disease	Damaging	Possibly Pathogenic	Effect	TRUE	Deleterious	10
R1047C	Damaging	Probably Damaging	High	Disease Causing	Disease	Damaging	Possibly Pathogenic	Effect	TRUE	Deleterious	10
G1015E	Damaging	Probably Damaging	High	Disease Causing	Disease	Damaging	Possibly Pathogenic	Effect	TRUE	Deleterious	10
V882L	Damaging	Benign	Low	Disease Causing	Neutral	Tolerated	Likely Benign	Neutral	FALSE	Deleterious	3
G1083R	Damaging	Probably Damaging	High	Disease Causing	Disease	Damaging	Possibly Pathogenic	Effect	TRUE	Deleterious	10
L1012S	Damaging	Probably Damaging	High	Disease Causing	Disease	Damaging	Possibly Pathogenic	Effect	TRUE	Deleterious	10
Q899L	Damaging	Probably Damaging	Low	Disease Causing	Neutral	Tolerated	Possibly Pathogenic	Neutral	TRUE	Deleterious	6
G774R	Damaging	Probably Damaging	High	Disease Causing	Disease	Tolerated	Possibly Pathogenic	Effect	TRUE	Deleterious	9
P769R	Damaging	Probably Damaging	High	Disease Causing	Disease	Tolerated	Possibly Pathogenic	Effect	TRUE	Deleterious	9
Y1055D	Damaging	Probably Damaging	High	Disease Causing	Disease	Damaging	Possibly Pathogenic	Effect	TRUE	Deleterious	10
R750H	Damaging	Possibly Damaging	Medium	Disease Causing	Neutral	Tolerated	Possibly Pathogenic	Effect	TRUE	Deleterious	8
A680V	Damaging	Probably Damaging	Medium	Disease Causing	Disease	Tolerated	Possibly Pathogenic	Effect	FALSE	Deleterious	8
R990L	Damaging	Probably Damaging	Medium	Disease Causing	Disease	Damaging	Possibly Pathogenic	Effect	TRUE	Deleterious	10
P958S	Damaging	Probably Damaging	Medium	Disease Causing	Disease	Tolerated	Possibly Pathogenic	Neutral	TRUE	Deleterious	8
S945G	Damaging	Probably Damaging	High	Disease Causing	Disease	Tolerated	Possibly Pathogenic	Effect	TRUE	Deleterious	9
E920A	Damaging	Probably Damaging	Medium	Disease Causing	Disease	Tolerated	Possibly Pathogenic	Effect	TRUE	Deleterious	9

Sequence Conservation Profiling

As presented in Table 3, results of conservation analysis on ConSurf revealed 23 of 26 missense variations to be located in highly conserved regions, with conservation scores 7, 8 and 9 (Fig., 1). Two variants, R1089W and H1042P, are the least conserved residues with scores 1 and 3 respectively, while one variant P706L is moderately conserved with a conservation score of 6. Of the 23 highly conserved variants, 16 were found to be exposed and functional and the rest 7 (G1083W, G1015E, G1083R, L1012S, G774R, A680V and S945G) were structurally significant (highly conserved and buried). Residues with least and moderately conserved scores were filtered out from next step of analysis.

Table 3: Conservation profile of deleterious amino acid variants in NOS2

AA Variant	Conservation Score	ConSurf Prediction
E1125D	9	Highly conserved and Exposed (f)
Y1123F	9	Highly conserved and Exposed (f)
R1122H	8	Highly conserved and Exposed (f)
R1089W	1	Variable and Exposed
H1042P	3	Variable and Buried
K730N	8	Highly conserved and Exposed (f)
P706L	6	Moderately conserved and Exposed
W704R	8	Highly conserved and Exposed (f)
L662V	9	Highly conserved and Exposed (f)
G659E	9	Highly conserved and Exposed (f)
R1122C	8	Highly conserved and Exposed (f)
N742K	7	Highly conserved and Exposed
G1083W	9	Highly conserved and Buried (s)
R1047C	8	Highly conserved and Exposed (f)
G1015E	9	Highly conserved and Buried (s)
G1083R	9	Highly conserved and Buried (s)
L1012S	9	Highly conserved and Buried (s)
G774R	8	Highly conserved and Buried
P769R	9	Highly conserved and Exposed (f)
Y1055D	9	Highly conserved and Exposed (f)
R750H	8	Highly conserved and Exposed (f)
A680V	8	Highly conserved and Buried
R990L	9	Highly conserved and Exposed (f)
P958S	8	Highly conserved and Exposed (f)
S945G	9	Highly conserved and Buried (s)
E920A	8	Highly conserved and Exposed (f)

Conservation scores: 7,8,9 - highly conserved; 5,6 - moderately conserved; 1-4 – variable; f: functional, s: structural

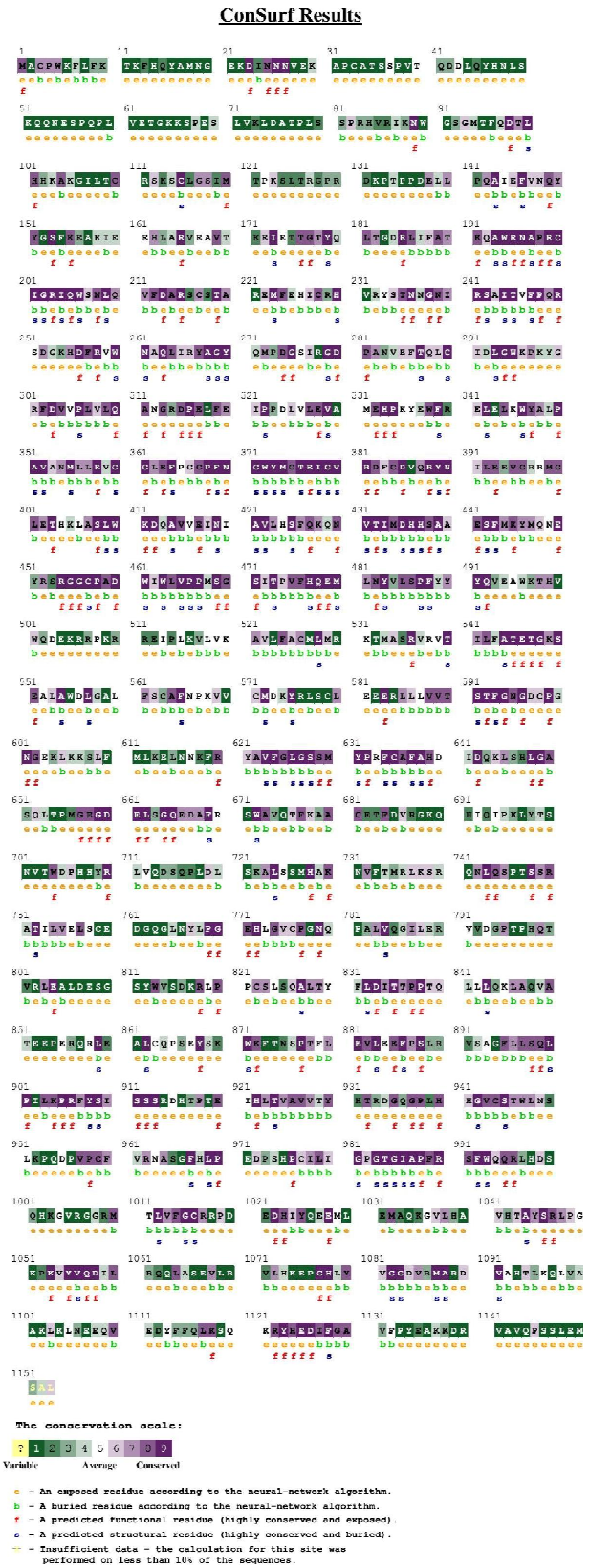

Figure 1: Sequence conservation profiling of NOS2 shows amino acid residues on a scale of 1 to 9 indicating least conserved to highly conserved residue. Structurally exposed and buried residues as well as functionally significant amino acids are depicted.

Prediction of Protein Stability Changes

The effect of 23 missense variants, predicted to be deleterious and damaging in previous steps of analysis, on the NOS2 protein stability was analysed using MuPro and I-mutant 3.0 servers. MuPro predicted all 23 variants to decrease the stability of protein with DDG scores falling under 0. However, I-mutant 3.0 results indicated that G659E, G774R and Y1055D increase the stability of NOS2 with free energy change values -0.10, -0.30 and -0.75 kcal/mol respectively. Rest of the variants were predicted to destabilize the protein stability, concurrent with MuPro prediction as presented in Table 4. Variants that were predicted to decrease protein stability by both tools were selected for post-translational modification (PTM) prediction.

Table 4: Prediction of protein stability changes induced by NOS2 damaging amino acid variations

AA Variant	Mupro (SVM)	I-Mutant 3.0
	DDG Value (kcal/mol)	DDG Value (kcal/mol)
E1125D	-1.25 (Decrease)	-0.49 (Decrease)
Y1123F	-1.06 (Decrease)	-0.12 (Decrease)
R1122H	-1.31 (Decrease)	-1.36 (Decrease)
K730N	-0.69 (Decrease)	-0.46 (Decrease)
W704R	-1.43 (Decrease)	-1.18 (Decrease)
L662V	-0.68 (Decrease)	-1.47 (Decrease)
G659E	-0.30 (Decrease)	-0.10 (Increase)
R1122C	-1.03 (Decrease)	-0.97 (Decrease)
N742K	-1.48 (Decrease)	-0.45 (Decrease)
G1083W	-0.75 (Decrease)	-0.37 (Decrease)
R1047C	-0.89 (Decrease)	-1.01 (Decrease)
G1015E	-0.50 (Decrease)	-0.69 (Decrease)
G1083R	-0.72 (Decrease)	-0.54 (Decrease)
L1012S	-2.20 (Decrease)	-2.27 (Decrease)
G774R	-0.56 (Decrease)	-0.30 (Increase)
P769R	-0.82 (Decrease)	-0.71 (Decrease)
Y1055D	-0.84 (Decrease)	-0.75 (Increase)
R750H	-1.32 (Decrease)	-1.26 (Decrease)
A680V	-0.20 (Decrease)	-0.22 (Decrease)
R990L	-0.10 (Decrease)	-0.18 (Decrease)
P958S	-0.68 (Decrease)	-1.83 (Decrease)
S945G	-1.06 (Decrease)	-0.79 (Decrease)
E920A	-0.93 (Decrease)	-0.88 (Decrease)

Post-translational Modification Sites Prediction

NOS2 protein sequence was used to predict PTMs present within the protein. Of the 20 amino acid residues (wild type and variant) we investigated, 4 conserved residues showed PTM sites with medium confidence and are discussed here. In the wild type, residue K730 was predicted as a proteolytic cleavage site which changed into a site of GPI anchor amidation with change of lysine to asparagine (K730N). At P769, change from proline to arginine (P769R) was predicted to confer proteolytic cleavage with medium confidence and ADP-ribosylation with low confidence to an unmodified native residue. P958 was predicted to be a hydroxylation site in its native state with medium confidence which was seen to change into proteolytic cleavage site, on varying from proline to serine (P958S). However, prediction of change for this modification was at a low confidence. Similarly, L1012, a proteolytic cleavage site predicted with medium confidence, changed into a

phosphorylation site with low confidence as the residue changed from lysine to serine (L1012S). ModPred also predicted with medium confidence that the native residue changed to proteolytic cleavage site upon altering to serine. Rest of the amino acid variants were predicted to either carry PTM sites with low confidence scores or remained unmodified (Table 5).

Table 5: Prediction of post-translational modification sites by ModPred

Amino Acid Residue (Wild Type)	PTM	Confidence	Amino Acid Variant	PTM	Confidence
E1125	Proteolytic cleavage	Low	1125D	Proteolytic cleavage	Low
Y1123	Unmodified	Low	1123F	Unmodified	Low
R1122	Unmodified	Low	1122H	Proteolytic cleavage	Low
K730	Proteolytic cleavage	Medium	730N	Proteolytic cleavage /GPI anchor amidation	Medium
W704	Unmodified	NA	704R	Proteolytic cleavage	Low
L662	Proteolytic cleavage	Low	662V	Unmodified	NA
R1122	Unmodified	NA	1122C	Proteolytic cleavage	Low
N742	Proteolytic cleavage	Low	742K	Proteolytic cleavage	Low
G1083	Unmodified	NA	1083W	Unmodified	NA
R1047	Unmodified	NA	1047C	Unmodified	NA
G1015	Proteolytic cleavage	Low	1015E	Proteolytic cleavage	Medium
G1083	Unmodified	NA	1083R	Proteolytic cleavage	Low
L1012	Proteolytic cleavage	Medium	1012S	Phosphorylation /Proteolytic cleavage	Low/Medium
P769	Unmodified	NA	769R	Proteolytic cleavage /ADP-ribosylation	Medium/Low
R750	ADP-ribosylation /Proteolytic cleavage	Medium/Low	750H	Unmodified	NA
A680	Unmodified	NA	680V	Unmodified	NA
R990	Proteolytic cleavage/ADP-ribosylation	Low	990L	Unmodified	NA
P958	Hydroxylation	Medium	958S	Proteolytic cleavage	Low
S945	Unmodified	NA	945G	Unmodified	NA
E920	Proteolytic cleavage	Low	920A	Unmodified	NA

Protein-Protein Interaction (PPI) Prediction and Enrichment Analysis

The predicted PPI network showed that NOS2 physically interacts with 10 key proteins involved in CAD development and associated processes, namely, Calmodulin 1, 2 and 3 (CALM1, CALM2, CALM3), tumor necrosis factor (TNF), SRC proto-oncogene, non-receptor tyrosine kinase (SRC), Jun Proto-Oncogene, AP-1 Transcription Factor Subunit (JUN), Signal transducer and activator of transcription 3 (STAT3), Catalase (CAT), Arginase-1 & 2 (ARG1, ARG2) (Fig., 2).

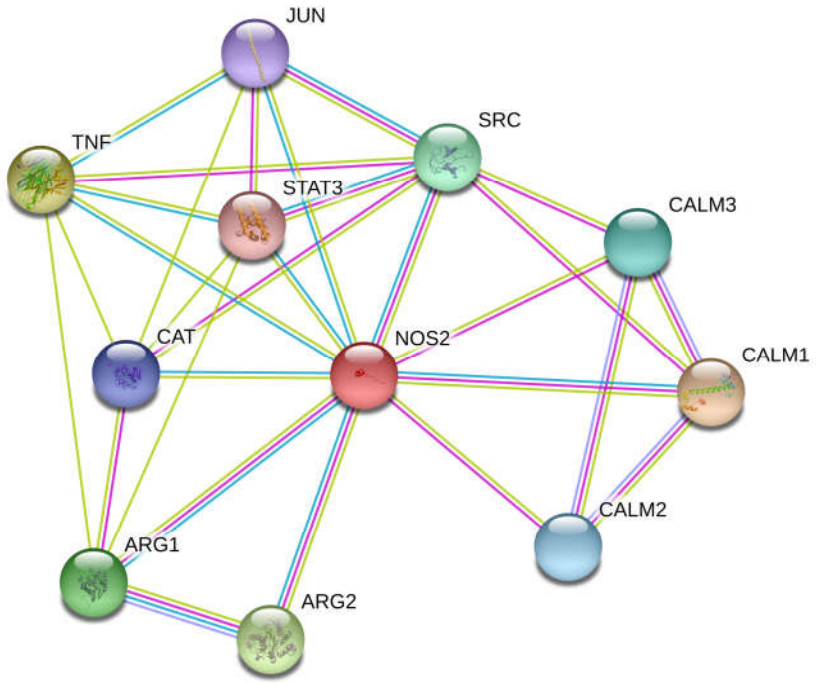

Figure 2: Functional protein-protein interaction network of NOS2 as predicted by STRING server. NOS2 is seen to interact with Calmodulin 1, 2 and 3 (CALM1, CALM2, CALM3), tumor necrosis factor (TNF), SRC proto-oncogene, non-receptor tyrosine kinase (SRC), Jun Proto-Oncogene, AP-1 Transcription Factor Subunit (JUN), Signal transducer and activator of transcription 3 (STAT3), Catalase (CAT), Arginase-1 & 2 (ARG1, ARG2).

Enrichment analysis revealed overrepresentation of these genes in pathways and processes related to CAD such as physiological and pathological hypertrophy of the heart, fluid shear stress and atherosclerosis, vascular smooth muscle contraction and platelet activation, signalling and aggregation. Results are presented in Table 6.

Table 6: Overrepresentation of genes in pathways related to CAD and their corresponding p-value

Term	P-value	Adjusted P-value	Genes
Physiological and Pathological Hypertrophy of the Heart	9.33E-10	4.40E-07	JUN;STAT3;CALM1;CALM2
Fluid shear stress and atherosclerosis	5.17E-13	1.59E-10	JUN;SRC;CALM3;CALM1;CALM2;TNF
Vascular smooth muscle contraction	5.92E-05	4.14E-04	CALM3;CALM1;CALM2
Platelet activation, signaling and aggregation	4.04E-04	0.088397269	SRC;CALM1

DISCUSSION

Nitric oxide (NO), an important messenger molecule which has various actions throughout the human body, is synthesized by nitric oxide synthases (NOS) family of enzymes. Many biological processes like neurotransmission, anti-tumoral and anti-microbial activities have NO as a key mediator (Huang *et al.*, 2018). Nitric oxide synthase enzyme 2 which is responsible for synthesis of inducible nitric oxide (iNOS) is encoded by *NOS2*. The gene *NOS2* is about 43.76 kb in length, spanning from 26083792 bp to 26127555 bp (GRCh37) on chromosome 17q11.2 on the reverse strand. *NOS2* is linked to several inflammatory diseases such as multiple sclerosis, rheumatoid arthritis, Sjögren's syndrome, cerebral and cardiac ischemia. iNOS protein has also been detected in infiltrating macrophages, vasculature and tumor cells of human lung, brain, colon and breast (Kröncke *et al.*, 1998).

NO functions as a significant vasodilator in the cardiovascular system. The bioavailability of NO controls development of coronary endothelial dysfunction and arterial hypertension (Levy *et al.*, 2009). Although NO produced by endothelial synthases (NOS3) is believed to be the main player in development and progression of these processes, recent evidences have shown that inducible NO (NOS2) also influences regulation of NO levels in blood in hypertensive patients (Topchieva *et al.*, 2019). In cardiovascular phenotypes, as a result of elevated levels of pro-inflammatory proteins, *NOS2* activity is shown to be upregulated in macrophages, determining increased concentration of NO in blood, possibly leading to cytotoxicity (Pacher *et al.*, 2007). While it is established that NOS2 activity confers CAD risk due to its pro-inflammatory properties, there is growing evidence for its role in cardioprotection. *NOS2* carries a shear-stress responsive element which induces NO production that mediates CAD protection through proliferation of vascular smooth muscle cells and inhibition of platelet aggregation and leucocyte adhesion (Levy *et al.*, 2009). Further, several studies have shown that upregulation of *NOS2* mediates anti-stunning and anti-infarct actions during late phase myocardial ischemia (Xi *et al.*, 1999; Bolli, 2001). Thus, *NOS2* functions as an important mediator of cardiovascular outcomes. Identification of deleterious gene variants in the coding region of *NOS2* provides key insights into their effect on protein activity.

In the current study, we identified 30 rare frequency nsSNVs in *NOS2* from WES of six CAD subjects (Table 1) using a stringent variant calling and filtering pipeline. The quality of called bases was assessed using Phred-scaled quality scores which is an estimation of error and accuracy, conventionally ranging between 2 and 40, with higher scores indicating a higher probability of called bases being accurate and vice versa (Ewing and Green, 1998). The variants called in this study were set to Phred score of >30 that indicates 99.9% accuracy (error rate = 1 in 1000), suggesting that the variants identified are true calls.

NOS2 comprises of 27 exons and harbours the identified missense variants from exons 17 to 27 (Fig., 3). The protein interaction network predicted complex interaction between *NOS2* and 10 candidate genes (Fig., 2) with essential roles in platelet activation, signalling and aggregation, fluid shear stress and atherosclerosis, vascular smooth muscle contraction, physiological and pathological hypertrophy of the heart (Table 6). Prediction of pathogenicity of nsSNVs through single bioinformatics tool may not be reliable (Kalia *et al.*, 2016). Therefore, we used a combination of several complementary prediction tools to identify pathogenic variants which employ different strategies to predict the effect of mutations on protein function. Our analysis on the functional effect of the 30 nsSNVs revealed a total of 26 variations (86.6%) predicted to be damaging and deleterious by at least seven prediction tools (Table 2). The amino acid substitutions of these variants were subjected to sequence conservation profiling using ConSurf. Amino acid residues that are highly conserved through evolution are crucial parts of protein secondary structures and domains. Therefore, variations in these regions are an indication of altered protein structure and function (Wang *et al.*, 2020). The advantage of ConSurf is that it not only estimates the degree of conservation of amino acids but also gives insight into significance of said residue in terms of structure and function based on its location either inside the protein core or on its surface (Ashkenazy *et al.*, 2016). R1089W, H1042P and P706L were predicted to be least conserved (Fig., 1).

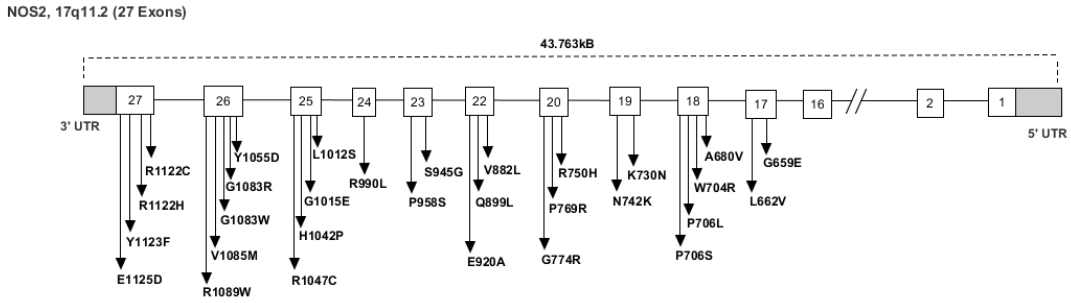

Figure 3: Single base missense variations identified across 27 exons of *NOS2*.

An important aspect of protein functionality is its stability state (Seifi and Walter, 2018). Our results from protein stability predictions indicated 20 variants decrease the stability of NOS2 (Table 4). Of these, 4 substitutions K730N, P769R, P958S and L1012S might interfere with post-translational modifications of NOS2 as predicted by ModPred (Table 5). These amino acid changes differ from the wild-type in polarity and structural properties. K730N, P958S and L1012S variations change to smaller amino acids while P769R change to bigger amino acid than the wild-type residue. The wild-type residues in all four substitutions were more hydrophobic as compared to the variants. The substitutions also influenced polarity of the amino residues, changing the positively charged lysine to neutrally charged asparagine in K730N and neutrally charged proline to positively charged arginine in P769R. Interestingly, K730N, P769R and P958S were found to be in FAD-binding FR-type domain; L1012S is located in the NAD-binding-1 domain (Venselaar *et al.*, 2010). Since the variations introduce amino acids with different properties than wild-type, function of these domains may be significantly disturbed. Hence, K730N, P769R, P958S and L1012S can be considered important candidates in the *NOS2* pathological process.

CONCLUSIONS

Bioinformatics studies are very useful in predicting effect of genomic variations on protein structure and function. Present study of *NOS2* nsSNVs implicates K730N, P769R, P958S and L1012S as the most pathogenic variations among the 30 nsSNVs identified through WES of six CAD subjects. To our best knowledge, this is the first report on *in silico* analysis of *NOS2* nsSNVs using WES. These findings may further be validated through appropriate wet-lab experiments to establish and confirm their effect on protein activity.

ACKNOWLEDGEMENT

We thank the participants of our study and technical staff at the University Health Centre. We thank Institution of Excellence, University of Mysore for providing the funds to carry out this study. We also thank Department of Studies in Genetics and Genomics, University of Mysore for providing facility to conduct this work.

CONFLICT OF INTEREST

The authors declare no competing financial interests.

REFERENCES

1. Adzhubei, I., Jordan, D.M. and Sunyaev, S,R. (2013). Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet.* 7: 7.20. doi:10.1002/0471142905.hg0720s76.

2. Aldons, J.L. (2012). Genetics of Atherosclerosis. *Trends Genet*. 28(6): 267-275.
3. Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T. and Ben-Tal, N. (2016). ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res*. 44(W1): W344-350.
4. Benjamin, E.J., Muntner, P., Alonso, A., Bittencourt, M.S., Callaway, C.W., Carson, A.P., Chamberlain, A.M., Chang, A.R., Cheng, S., Das, S.R., Delling, F.N., Djousse, L., Elkind, M.S.V., Ferguson, J.F., Fornage, M., Jordan, L.C., Khan, S.S., Kissela, B.M., Knutson, K.L., Kwan, T.W., Lackland, D.T., Lewis, T.T., Lichtman, J.H., Longenecker, C.T., Loop, M.S., Lutsey, P.L., Martin, S.S., Matsushita, K., Moran, A.E., Mussolino, M.E., O'Flaherty, M., Pandey, A., Perak, A.M., Rosamond, W.D., Roth, G.A., Sampson, U.K.A., Satou, G.M., Schroeder, E.B., Shah, S.H., Spartano, N.L., Stokes, A., Tirschwell, D.L., Tsao, C.W., Turakhia, M.P., VanWagner, L.B., Wilkins, J.T., Wong, S.S. and Virani, S.S. (2019). American Heart Association Council on Epidemiology & Prevention Statistics Committee and Stroke Statistics Subcommittee, Heart Disease and Stroke Statistics-2019 Update: A Report From the American Heart Association. *Circulation*. 139(10): e56-e528.
5. Bolli, R. (2001). Cardioprotective function of inducible nitric oxide synthase and role of nitric oxide in myocardial ischemia and preconditioning: an overview of a decade of research. *J Mol Cell Cardiol*. 33(11): 1897-1918.
6. Bomba, L., Walter, K. and Soranzo, N. (2017). The impact of rare and low-frequency genetic variants in common disease. *Genome Biol*. 18(1): 77.
7. Capriotti, E., Calabrese, R. and Casadio, R. (2006). Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. *Bioinformatics*. 22(22): 2729-2734.
8. Capriotti, E., Fariselli, P. and Casadio, R. (2005). I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Res*. 33: W306-10.
9. Chang, X. and Wang, K. (2012). wANNOVAR: annotating genetic variants for personal genomes via the web. *J Med Genet*. 49(7): 433-436.
10. Cheng, J., Randall, A. and Baldi, P. (2006). Prediction of protein stability changes for single-site mutations using support vector machines. *Proteins*. 62(4): 1125-1132.
11. Choi, Y. and Chan, A.P. (2015). PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics*. 31(16): 2745-2747.
12. Do, R., Kathiresan, S. and Abecasis, G.R. (2012). Exome sequencing and complex disease: practical aspects of rare variant association studies. *Hum Mol Genet*. 21(R1): R1-9.
13. Elkhattabi, L., Morjane, I., Charoute, H., Amghar, S., Bouafi, H., Elkarhat, Z., Saile, R., Rouba, H. and Barakat, A. (2019). In Silico Analysis of Coding/Noncoding SNPs of Human RETN Gene and Characterization of Their Impact on Resistin Stability and Structure. *J Diabetes Res*. 2019. https://doi.org/10.1155/2019/4951627.
14. Ewing, B. and Green, P. (1998). Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res*. 8(3): 186-194.
15. Fernandes, M., Patel, A. and Husi, H. (2018). C/VDdb: A multi-omics expression profiling database for a knowledge-driven approach in cardiovascular disease (CVD). *PLoS One*. 13(11): e0207371.
16. Hajar, R. (2017). Risk Factors for Coronary Artery Disease: Historical Perspectives. *Heart Views*. 18(3): 109-114.
17. Hecht, M., Bromberg, Y. and Rost, B. (2015). Better prediction of functional effects for sequence variants. *BMC Genomics*. 16: S1.
18. Huang, W., Zhang, K., Zhu, Y., Wang, Z., Li, Z. and Zhang, J. (2018). Genetic polymorphisms of NOS2 and predisposition to fracture non-union: A case control study based on Han Chinese population. *PLoS One*. 13(3): e0193673.
19. Jagadeesh, K.A., Wenger, A.M., Berger, M.J., Guturu, H., Stenson, P.D., Cooper, D.N., Bernstein, J.A. and Bejerano, G. (2016). M-CAP eliminates a majority of variants with uncertain significance in clinical exomes at high sensitivity. *Nat Genet*. 48(12): 1581-1586.
20. Kalia, N., Sharma, A., Kaur, M., Kamboj, S.S. and Singh, J. (2016). A comprehensive in silico analysis of non-synonymous and regulatory SNPs of human MBL2 gene. *Springerplus*. 5(1): 811.

21. Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., Gauthier, L.D., Brand, H., Solomonson, M., Watts, N.A., Rhodes, D., Singer-Berk, M., England, E.M., Seaby, E.G., Kosmicki, J.A., …MacArthur, D.G. (2019). The mutational constraint spectrum quantified from variation in 141,456 humans. In bioRxiv (p. 531210). https://doi.org/10.1101/531210.
22. Kröncke, K.D., Fehsel, K. and Kolb-Bachofen, V. (1998). Inducible nitric oxide synthase in human diseases. *Clin Exp Immunol*. 113(2): 147-156.
23. Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A., McDermott, M.G., Monteiro, C.D., Gundersen, G.W. and Ma'ayan, A. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res*. 44(W1): W90-97.
24. Levy, A.S., Chung, J.C., Kroetsch, J.T. and Rush, J.W. (2009). Nitric oxide and coronary vascular endothelium adaptations in hypertension. *Vasc Health Risk Manag*. 5: 1075-1087.
25. Liu, H., Liu, W., Liao, Y., Cheng, L., Liu, Q., Ren, X., Shi, L., Tu, X., Wang, Q.K. and Guo, A.Y. (2011). CADgene: a comprehensive database for coronary artery disease genes. *Nucleic Acids Res*. 39: D991-996.
26. López-Ferrando, V., Gazzo, A., de la Cruz, X., Orozco, M. and Gelpí, J.L. (2017). PMut: a web-based tool for the annotation of pathological variants on proteins, 2017 update. *Nucleic Acids Res*. 45(W1): W222-228.
27. Pacher, P., Beckman, J.S. and Liaudet, L. (2007). Nitric oxide and peroxynitrite in health and disease. *Physiol Rev*. 87(1): 315-424.
28. Pejaver, V., Hsu, W.L., Xin, F., Dunker, A.K., Uversky, V.N. and Radivojac, P. (2014). The structural and functional signatures of proteins that undergo multiple events of post-translational modification. *Protein Sci*. 23(8): 1077-1093.
29. Rani, J., Mittal, I., Pramanik, A., Singh, N., Dube, N., Sharma, S., Puniya, B.L., Raghunandanan, M.V., Mobeen, A. and Ramachandran, S. (2017). T2DiACoD: A Gene Atlas of Type 2 Diabetes Mellitus Associated Complex Disorders. *Sci Rep*. 7(1): 6892.
30. Reva, B., Antipin, Y. and Sander, C. (2011). Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Res*. 39(17): e118.
31. Schwarz, J.M., Cooper, D.N., Schuelke, M. and Seelow, D. (2014). MutationTaster2: mutation prediction for the deep-sequencing age. *Nat Methods*. 11(4): 361-362.
32. Seifi, M. and Walter, M.A. (2018). Accurate prediction of functional, structural, and stability changes in PITX2 mutations using in silico bioinformatics algorithms. *PLoS One*. 13(4): e0195971.
33. Sherry, S.T., Ward, M.H., Kholodov, M., Baker, J., Phan, L., Smigielski, E.M. and Sirotkin, K. (2001). dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res*. 29(1): 308-311.
34. Shihab, H.A., Gough, J., Cooper, D.N., Stenson, P.D., Barker, G.L., Edwards, K.J., Day, I.N., and Gaunt, T.R. (2013). Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. *Hum Mutat*. 34(1): 57-65.
35. Sim, N.L., Kumar, P., Hu, J., Henikoff, S., Schneider, G. and Ng, P.C. (2012). SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res*. 40: W452-457.
36. Singh, R.B., Mengi, S.A., Xu, Y-J., Arneja, A.S. and Dhalla, N.S. (2002). Pathogenesis of atherosclerosis: A multifactorial process. *Exp Clin Cardiol*. 7(1): 40-53.
37. Strand Life Sciences Pvt. Ltd., Strand NGS. ©Strand Life Sciences, Bangalore, India, 2016.
38. Szklarczyk, D., Gable, A.L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N.T., Morris, J.H., Bork, P., Jensen, L.J. and Mering, C.V. (2019). STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*. 47(D1): D607-613.
39. Topchieva, L.V., Balan, O.V., Korneeva, V.A. and Malysheva, I.E. (2019). The Role of Inducible NOS2 Gene Polymorphism in the Development of Essential Arterial Hypertension. *Bull Exp Biol Med*. 168(1): 79-83.
40. Venselaar, H., Te Beek, T.A., Kuipers, R.K., Hekkelman, M.L. and Vriend, G. (2010). Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. *BMC Bioinformatics*. 11: 548.

41. Wang, Z., Huang, C., Lv, H., Zhang, M. and Li, X. (2020). In silico analysis and high-risk pathogenic phenotype predictions of non-synonymous single nucleotide polymorphisms in human Crystallin beta A4 gene associated with congenital cataract. *PLoS One*. 15(1): e0227859.
42. Xi, L., Jarrett, N.C., Hess, M.L. and Kukreja, R.C. (1999). Essential role of inducible nitric oxide synthase in monophosphoryl lipid A–induced late cardioprotection: evidence from pharmacological inhibition and gene knockout mice. *Circulation*. 99(16): 2157-2163.