EUROPEAN JOURNAL OF BIOMEDICAL AND PHARMACEUTICAL SCIENCES
http://www.ejbps.com

ISSN 2349-8870
Volume: 7
Issue: 1
514-521
Year: 2020

INSILICO ANALYSIS OF A SINGLE NUCLEOTIDE POLYMORPHISM (SNPS) IN HUMAN *GSTM1* GENE ASSOCIATED WITH CANCER DEVELOPMENT

Dalia Mursi Abdelhalim[1,2], Afra M. Albkrye[1], Mohammed Ahmed Salih[3], Amna Elsadig Elsafi Abodlaa[11,4], Hind Abdelaziz Elnasri[1] and Mona A. M. Khaier[1]*

[1]Department of Molecular Biology and Bioinformatics, Collage of Veterinary Medicine, University of Bahri. Khartoum, Sudan.
[2]Central Laboratory. Ministry of Higher Education and Scientific Research, Khartoum, Sudan.
[3]Department of Bioinformatics, Africa city of Technology, Khartoum. Sudan.
[4]Research Diagnosis Unit/Khartoum Hospital. Khartoum, Sudan.

***Corresponding Author: Mona A. M. Khaier**
Department of Molecular Biology and Bioinformatics, Collage of Veterinary Medicine, University of Bahri. Khartoum, Sudan.

Article Received on 21/11/2019 **Article Revised on 11/12/2019** **Article Accepted on 31/12/2019**

ABSTRACT

Background: The gene for Glutathione-S-Transferase (*GSTM1*), is a member of the GST-superfamily, it encodes enzymes responsible for detoxification of free radicals and other carcinogens. The activity of these enzymes may differ due to polymorphisms which ultimately results in inter-individual susceptibility to cancer development. **Aim:** This study was carried out to analyse the nsSNPs in the *GSTM1* gene, to identify the possible mutations and propose a modelled structure for the mutant protein. **Methods:** Through the in silico approach, using a combination of SIFT, Poly Phen-2, I-Mutant Suite, SNPs & Go and Project Hope software. **Results:** Among the 62 nsSNPs, only 21 were found to be deleterious. From these seven SNPs (rs371083091, rs376564748, rs553341658, rs758844606, rs142484086, rs533860247, and rs777299993) were reported to be highly damaging. A modelled structure of the seven highly deleterious and highly damaging mutant proteins were proposed. **Conclusions:** These results provide useful information in selecting target SNPs that are likely to have an impact on *GSTM1* gene activity and contribute to an individual's susceptibility to the disease.

KEYWORDS: *GSTM1* gene, In Silico, SIF, Polyphen-2, I-Mutant Suite, SNPs & Go, Project Hope.

1. INTRODUCTION

The Glutathione S-Transferases (GST) are a family of enzymes responsible for the metabolism of a broad range of xenobiotics and carcinogens.[1] Glutathione-S-Transferase (*GSTM1*), a member of the GST-superfamily, is widely studied as a risk for devolving cancer.[2] Human GSTs are divided into three main families: cytosolic, mitochondrial and membrane-bound microsomal.[3] At present, eight distinct classes of the soluble cytoplasmic mammalian glutathione S-transferases have been identified: alpha, kappa, mu, omega, pi, sigma, theta and zeta. This gene encodes a glutathione S-transferase that belongs to the mu class. The mu class of enzymes functions in the detoxification of electrophilic compounds, including carcinogens, therapeutic drugs, environmental toxins and products of oxidative stress, by conjugation with glutathione. The genes encoding the mu class of enzymes are organized in a gene cluster on chromosome 1p13.3 and are known to be highly polymorphic. These genetic variations can change an individual's susceptibility to carcinogens and toxins as well as affecting the toxicity and efficacy of certain drugs. Null mutations of this mu gene have been linked with an increase in a number of cancers, likely due to an increased susceptibility to environmental toxins and carcinogens. Multiple protein isoforms are encoded by transcript variants of this gene.[4] An individual's response to a drug depends on both environmental factors and genetic factors. Individual variations in response to cancer therapy, and to toxic chemicals are related to the genetic differences in the drug metabolizing enzymes. A primary cause of cancer treatment failure and patient relapse is an acquired or intrinsic resistance to anticancer therapies.[5] Genetic variation in the human genome is an emerging resource for studying cancer and other diseases. Single-nucleotide polymorphisms (SNPs) are the most common type of DNA sequence variation, accounting for approximately 90% of the DNA polymorphism in humans.[6] A non-synonymous single nucleotide polymorphism (nsSNP), which is present within the exon of region, is responsible for the incorporation of an alternative amino acid and known to be one of the main causes for major genetic disorders. However, tolerant nsSNPs are not deleterious and are not involved in any genetic disorders, whereas deleterious nsSNPs have a profound influence on protein

structure and its interaction. Therefore, it is important to differentiate deleterious nsSNPs from tolerant nsSNPs to characterize the genetic basis of human diseases.[7] Discovering of such deleterious nsSNPs is the main task of Pharmacogenomics.[8] A possible way to overcome this problem would be to prioritize SNPs according to their functional significance by using Bioinformatics prediction tools, which may help to discriminate neutral SNPs from SNPs of likely functional importance and could also be useful to reveal the structural basis of disease mutations.

This study was carried out to analyse the nsSNPs in the *GSTM1* gene, to identify the possible mutations and propose a modelled structure for the mutant protein. The coding nsSNPs were selected for our investigation.

2. MATERIALS AND METHODS

The SNPs information (Protein accession number and SNP ID) of the *GSTM1* gene was retrieved from the NCBI dbSNP (http://www.ncbi.nlm.nih.gov/snp/), and UniprotKB Kb at ExPASY databases (http://www.uniprot.org/).

Different freely available computational algorithms based on sequence homology, physicochemical properties of the amino acid residue and an in silico site directed mutagenesis tool is used to identify the possible deleterious mutations.

2.1. Sorting Intolerant From Tolerant (SIFT)

(http://siftdna.org/www/SIFT_dbSNP.html) This is a sequence homology-based tool that sorts intolerant from tolerant amino acid substitutions and predicts whether an amino acid substitution in a protein will have a phenotypic effect. SIFT takes a query sequence and uses multiple alignment information to predict tolerated and deleterious substitutions for every position of the query sequence, it is a multistep procedure that searches for similar sequences, chooses closely related sequences that may share similar function to the query sequence, obtains the alignment of these chosen sequences, and calculates normalized probabilities for all possible substitutions from the alignment. Positions with normalized probabilities less than 0.05 are predicted to be deleterious, those greater than or equal to 0.05 are predicted to be tolerated.[9]

2.2. Polymorphism Phenotyping v2 (PolyPhen-2)

(http://genetics.bwh.harvard.edu/pph2/) PolyPhen-2 is a structural-homology-based tool that predicts the impact of an amino acid substitution on the structure and function of a human protein. Predictions are based on a combination of phylogenetic, structural and sequence annotation information characterizing a substitution and its position in the protein.

The output of the PolyPhen-2 prediction pipeline is a prediction of probably damaging, possibly damaging, or benign, along with a numerical score ranging from 0.0 (benign) to 1.0 (damaging). A prediction of probably damaging means that the query substitution is predicted to be damaging with high confidence, while a prediction of benign means that the query substitution is predicted to be benign with high confidence. A prediction of possibly damaging means that the query substitution is predicted to be damaging, but with low confidence. Though the possibly damaging score is often interpreted as an indication of a mild effect or low penetrance.[10]

2.3. I-Mutant

I-Mutant version 3 (http://gpcr2.biocomp.unibo.it/cgi/predictors/I-Mutant3.0/I-Mutant3.0.cgi)
Predictor of stability change upon single point protein mutation, it offers the opportunity to predict automatically protein stability changes upon single-site mutations starting from protein sequence alone or protein structure when available. Moreover it gives the possibility to predict human deleterious Single Nucleotide Polymorphism starting from the protein sequence alone.[11]

2.4. Single Nucleotide Polymorphism (SNPs) and Gene Ontology (SNPs &GO)

http://snps.biofold.org/snps-and-go//snps-and-go.html
These are support vector machine (SVM) based accurate methods used to predict the disease related mutations from protein sequences with a scoring accuracy of 82% and Matthews correlation coefficient of 0.63. For SNPs & GO, FASTA sequence of whole protein is considered to be an input option and output will be the prediction results based on the discrimination among disease related and neutral variations of protein sequence. The probability score higher than 0.5 reveals the disease related effect of mutation on the parent protein function.[12,13]

2.5. GeneMANIA

(http://www.genemania.org).GeneMANIA is an online server that gives the function, interaction and the network of the gene based on genomics and proteomics data through searching many large, publicly available biological datasets to find related genes.[14]

2.6 Project HOPE

(HOPE; http://www.cmbi.ru.nl/hope/home) is an automatic mutant analysis server to study the insight structural features of native protein and the variant models. HOPE provides the 3D structural visualization of mutated proteins, and gives the results by using UniProt and DAS prediction servers. Input method of Project HOPE carries the protein sequence and selection of Mutant variants. HOPE server predicts the output in the form of structural variation between mutant and wild type residues.[15]

3. RESULTS

3.1. SNP dataset from dbSNP

The human *GSTM1* gene contains a total of 3429 SNPs, of which 62 are nsSNPs (missense) and 91 are coding synonymous SNPs. The noncoding SNPs consisted of 323 SNPs. The coding nsSNPs were selected for investigation.

3.2. Identification of deleterious and damaging nsSNPs

Through SIFT, 21 nsSNPs(34%) were predicted to be deleterious with a tolerance score of less than or equal to 0.0, eight(G12V, R18H, D42G, Q72P, S73N, D162V I194Tand W215C) were reported to be highly deleterious with a tolerance score of 0.00. Table 1.

3.3. Prediction of protein structural and functional modifications

Further analyzed the nsSNPs using PolyPhen-2 .of the 21 deleterious nsSNPs predicted by sift analysis, 14 nsSNPs were predicted to be ‘‘probably damaging’’, and four nsSNPs were found to be ‘‘possibly damaging’’. , three nsSNPs were characterized as benign. Table: 1.

Table 1: nsSNPs predicted with SIFT and polyphen softwere.

SNP	Substitution	SIFT Score	SIFT Prediction	Polyphen Prediction	Polyphen Score
rs371083091	G12V	0	Deleterious	Probably Damaging	1
rs536289169	A16V	0.01	Deleterious	Benign	0.076
rs376564748	R18H	0	Deleterious	Probably Damaging	1
rs553341658	Y23C	0.02	Deleterious	Probably Damaging	1
rs11546855	D42G	0	Deleterious	Probably Damaging	0.997
rs760927586	N59S	0.01	Deleterious	Possibly Damaging	0.893
rs758844606	Q72P	0	Deleterious	Probably Damaging	1
rs778256041	S73N	0	Deleterious	Probably Damaging	0.997
rs746113233	R82H	0.04	Deleterious	Probably Damaging	0.996
rs199816990	R96L	0.02	Deleterious	Benign	0.007
rs202002774	M105T	0.04	Deleterious	Possibly Damaging	0.773
rs748101558	Y116C	0.03	Deleterious	Probably Damaging	0.982
rs142484086	R145W	0.02	Deleterious	Probably damaging	0.999
rs544766466	D162V	0	Deleterious	Probably Damaging	0.998
rs72549312	P179L	0.03	Deleterious	Benign	0.023
rs753904217	R187H	0.03	Deleterious	Probably damaging	0.996
rs878938889	G190R	0.02	Deleterious	Possibly Damaging	0.908
rs199721250	I194T	0	Deleterious	Probably Damaging	0.996
rs371247780	R202H	0.01	Deleterious	Possibly Damaging	0.818
rs533860247	A213T	0.01	Deleterious	Probably Damaging	0.999
rs777299993	W215C	0	Deleterious	Probably Damaging	1

SIFT Tolerance Index: Ranges from 0 to 1. The amino acid substitution is predicted deleterious if the score is <= 0.05, and tolerated if the score is > 0.05.

POROBABLY DAMAGING (more confident prediction) / POSSIBLY DAMAGING (less confident prediction), PSIC SD: Position-Specific Independent Counts software if the score is => 0.5.

3.4. Prediction of disease related mutations by SNPs&GO

Out of the total nsSNPs screened by SIFT and PolyPhen, 18 nsSNPs were submitted to the SNPs&GO software. Elevnwere predicted to be disease related while 7 SNPs were neutral. While when useing PHD16 SNPs were disease and two were neutral. Table 2.

3.5. Prediction of protein structural stability

To improve the prediction accuracy of structure-based tools, I-Mutant Suite was used for 18 possibly and probably damaging’’ nsSNPs,16 mutations in *GSTM1* gene decrease effective stability of the protein, while two were increase effective stability According to I-Mutant software, we found the proprieties of amino acids as of the protein. Table3.

Table 2: Shows the SNPs that were predicted using SNPS&GO.

SNP	SNPS&GO			PHD		
	Prediction	RI	Probability	Prediction	RI	Probability
rs371083091	Disease	8	0.885	Disease	9	0.943
rs536289169	Disease	5	0.747	Disease	7	0.849
rs376564748	Disease	2	0.583	Disease	4	0.722
rs553341658	Disease	4	0.714	Disease	6	0.816
rs11546855	Disease	4	0.714	Disease	6	0.816
rs760927586	Neutral	4	0.312	Disease	4	0.679
rs758844606	Disease	7	0.872	Disease	9	0.932
rs778256041	Disease	3	0.635	Disease	7	0.851
rs746113233	Disease	1	0.557	Disease	4	0.714
rs202002774	Neutral	1	0.467	Disease	0	0.51
rs748101558	Neutral	1	0.462	Disease	4	0.702
rs142484086	Neutral	1	0.462	Disease	4	0.702
rs544766466	Disease	7	0.87	Disease	8	0.911
rs753904217	Neutral	5	0.264	Disease	3	0.646
rs878938889	Neutral	5	0.257	Neutral	1	0.47
rs199721250	Disease	1	0.575	Disease	5	0.737
rs371247780	Neutral	6	0.203	Neutral	1	0.454
rs533860247	Neutral	2	0.41	Disease	1	0.561
rs777299993	Disease	3	0.64	Disease	2	0.621

Neutral: Neutral variation, Disease: Disease associated variation RI: Reliability Index Probability: Disease probability (if >0.5 mutation is predicted Disease).

Table 3 Shows the SNPs that were predicted using I-Mutant.

SNP	SVM2	RI	DDG
rs536289169	Increase	5	0.29
rs376564748	Decrease	0	-0.64
rs553341658	Decrease	2	-0.5
rs11546855	Decrease	7	-1.33
rs760927586	Decrease	7	-1.5
rs758844606	Decrease	1	-0.17
rs778256041	Decrease	6	-0.81
rs371083091	Decrease	6	-0.75
rs746113233	Decrease	9	-1.49
rs202002774	Decrease	6	-0.64
rs748101558	Decrease	3	-0.39
rs142484086	Increase	0	0.3
rs544766466	Increase	0	-0.21
rs753904217	Decrease	10	-2.42
rs878938889	Decrease	6	-1.78
rs199721250	Decrease	9	-1.66
rs371247780	Decrease	9	-2.18
rs533860247	Decrease	1	-0.29
rs777299993	Decrease	9	-2.34
rs199816990	Increase	3	0.18
rs72549312	Decrease	9	-1.56

RI: Reliability Index; DDG: ΔΔG sign; SVM: Support Vector Machine DDG value: DG (New Protein)-DG (Wild Type) in Kcal/mole; SVM2 value: DDG < 0: decrease stability, DDG>0:increase stability.

3.6. Networks and interaction by GeneMania

Functions and interaction of *GSTM1* with functional similar genes illustrated by GenMANIA software. The genes that are co expressed or have physical properties with *GSTM1* are shown in Figure 2. The description of genes co expressed and shared domain with *GSTM1* are listed in Table 4. The *GSTM1*network genes functions and its appearance in network and genome are listed in Table 5.

3.7 The 3D analysis of wild type and mutant protein structures by project HOPE

The 7 most deleterious and damaging ns SNPs predicted damaging by Polyphen-2 with score (1-0.999) were submitted to Project HOPE software. rs371083091, rs376564748, rs553341658, rs758844606, rs142484086, rs533860247, rs777299993. Polymorphisms change in the amino acids were showed in the Figure2.

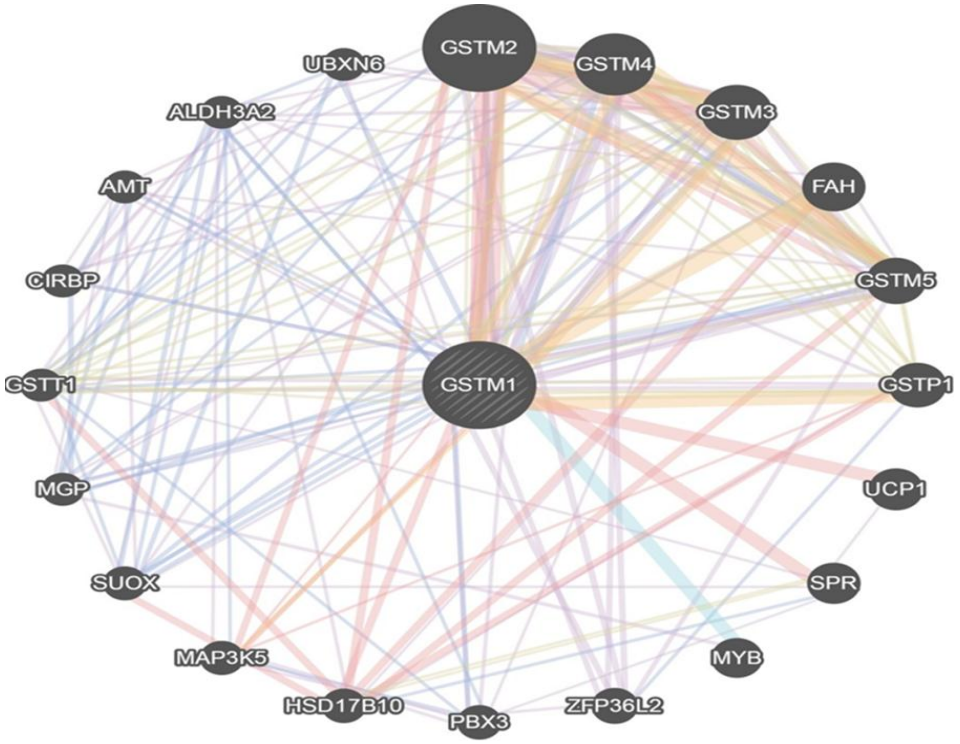


Figure 1: Genes interact with *GSTM1* gene.

Table 4: Genes Co-expressed and Share a domain with *GSTM1*.

Gene Symbol	Description	CO- EXPRESSION	Shared domain
GSTM2	glutathione S-transferase mu 2 (muscle)	Yes	Yes
GSTM4	glutathione S-transferase mu 4	Yes	Yes
GSTM3	glutathione S-transferase mu 3	Yes	Yes
FAH	fumarylacetoacetate hydrolase (fumarylacetoacetase)	yes	No
GSTM5	glutathione S-transferase mu 5	Yes	Yes
GSTP1	glutathione S-transferase pi 1	Yes	Yes
UCP1	uncoupling protein 1 sepiapterin reductase	Yes	No
SPR	(7,8-dihydrobiopterin:NADP+ oxidoreductase	Yes	No
MYB	MYB proto-oncogene, transcription factor	Yes	No
ZFP3BL2	ZFP36 ring finger protein-like 2	Yes	No
PBX3	PBX homeobox 3	Yes	No
HSD17B10	hydroxysteroid 17-beta dehydrogenase 10	Yes	Yes
MAP3K5	mitogen-activated protein kinase kinase kinase 5	Yes	No
SUOX	sulfite oxidase	Yes	No
MGP	matrix Gla protein	Yes	No
GSTT1	glutathione S-transferase theta 1	Yes	Yes
CIRBP	cold inducible RNA binding protein	Yes	No
AMT	aminomethyltransferase	Yes	No
ALDH3A2	aldehyde dehydrogenase 3 family member A2	Ye	No
UBXN6	UBX domain protein 6	Yes	No

Table 5: *GSTM1* network genes functions and its appearance in network and genome.

Function	FDR	Genes in network	Genes in genome
glutathione transferase activity	8.01E-13	7	20
glutathione derivative metabolic process	3.04E-12	7	27
glutathione derivative biosynthetic process	3.04E-12	7	27
glutathione metabolic process	2.63E-11	7	37
transferase activity, transferring alkyl or aryl (other than methyl) gr	3.14E-11	7	39
peptide metabolic process	2.22E-09	7	71
sulfur compound metabolic process	7.17E-08	8	208
cellular amide metabolic process	8.11E-08	7	122
cellular response to xenobiotic stimulus	2.45E-07	7	151
response to xenobiotic stimulus	2.45E-07	7	151
xenobiotic metabolic process	2.45E-07	7	150
cellular modified amino acid metabolic process	2.45E-07	7	145
modified amino acid binding	5.1935E-05	4	32
benzene-containing compound metabolic process	1.78E-04	3	10
amino acid binding	5.59E-04	4	59
sulfur compound binding	2.63E-03	4	88
peptide binding	4.47E-03	4	102
amide binding	4.47E-03	4	105
organic acid binding	4.77E-03	4	108
carboxylic acid binding	4.77E-03	4	108
response to toxic substance	1.66E-02	3	48
oxidoreductase activity, acting on CH-OH group of donors	7.94E-02	3	82

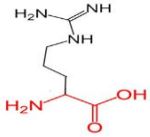
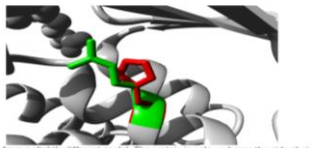
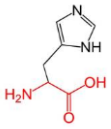
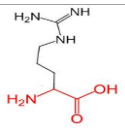
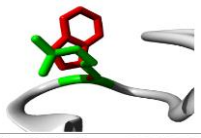
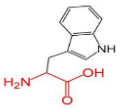
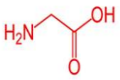
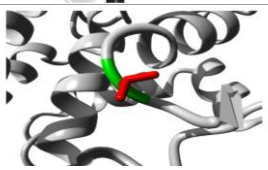
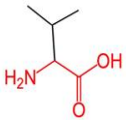
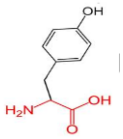
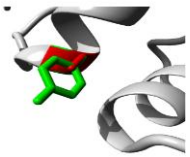
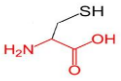
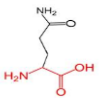
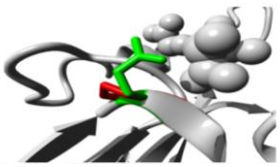
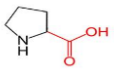
SNP	Wild	3D structure	Mutant type	Amino acid change
Rs376564748				R 18 H
Rs142484086				R 145W
Rs371083091				G12V
Rs553341658				Y 23 C
Rs758844606				Q72 P

Figure 2: 3D model by project Hope, wild type and mutant were coloured green and red respectively.

DISCUSSION

The SNPs are hypothesized to play an important role in several human diseases. Approximately, more than 4 million unique human SNPs have now been reported by a number of SNP databases such as National Centre for Biotechnology Information database, dbSNP.[16] Approximately, 2% of the all known single nucleotide variants associated with genetic diseases are nsSNPs occurring in the coding regions and contributes to the functional diversity of the encoded proteins in the human population.[17] In silico analysis using powerful software tools can facilitate predicting the phenotypic effect of non-synonymous coding SNPs on the physio-chemical properties of the concerned proteins. Such information is critical for genotype– phenotype correlations and also to understand disease biology. Rs371083091 the mutation of a Alanine into a Valine at position 12 is located within a domain, which can disturb this domain and abolish its function, is buried in a domain that is important for binding of other molecules. For rs376564748, the mutation of Arginine into a Histidine at position 18, the wild-type residue charge was positive, the mutant residue charge is neutral, the wild-type residue forms a hydrogen bond with Isoleucine at position 10 Glutamic Acid at position 30 the mutant residue is smaller than the wild-type residue. For the rs553341658 the mutation of a Tyrosine into a Cysteine at position 23 the mutant residue is smaller than the wild-type residue. This will cause a possible loss of external interactions. In case of rs758844606 the mutation of a Glutamine into a Proline at position 72, only this residue type was found at this position. Mutation of a 100% conserved residue is usually damaging for the protein The difference in properties between wild-type and mutant-type can easily cause loss of interactions with the ligand. Because ligand binding is often important for the protein's function, this function might be disturbed by this mutation. rs142484086, the mutation of a Arginine into a Tryptophan at position 145, was found in other study(ref) The wild-type residue forms a hydrogen bond with Leucine at position 142, the wild-type residue forms a salt bridge with: Glutamic Acid at position 92The mutant residue is bigger than the wild-type residue. The residue is located on the surface of the protein, mutation of this residue can disturb interactions with other molecules or other parts of the protein.rs533860247 mutation of a Alanine into a Threonine at position 213, The wild-type residue is predicted (using the Reprof software) to be located in an α-helix. The mutation converts the wild-type residue in a residue that does not prefer α-helices as secondary structure, mutant residue was not among the other residue types observed at this position in other, homologous proteins. The mutant residue is bigger than the wild-type residue. For rs777299993 the mutation of a Tryptophan into a Cysteine at position 215. The wild-type residue is annotated in UniProt to be located in its preferred secondary structure, a β-strand. The mutant residue prefers to be in another secondary structure; therefore the local conformation will be slightly destabilized. Mutation is probably damaging to the protein, the mutant residue is smaller than the wild-type residue. This will cause a possible loss of external interactions. Other Insilco study by *Yadav, et al* were found rs72549312 and rs142484086 SNPs deleterious as our result, while rs184653774 was predicted to be tolerated in our study, the rs142484086 and rs72549313 were not found in previously mentioned study.[18]

CONCLUSIONS

These results provide useful information in selecting target SNPs that are likely to have an impact on *GSTM1* gene activity and contribute to an individual's susceptibility to the disease.

DATA AVAILABILITY

The SNPs information (Protein accession number and SNP ID) of the *GSTM1* gene was retrieved from the NCBI dbSNP (http://www.ncbi.nlm.nih.gov/snp/), and UniprotKB Kb at ExPASY databases (http://www.uniprot.org/).

GRANT INFORMATION

The authors declare that no grants were involved in supporting this work.

REFERENCES

1. Mannervik, B., *The isoenzymes of glutathione transferase.* Adv Enzymol Relat Areas Mol Biol, 1985; 57: 357-417.
2. Bhattacharjee P, Paul S, Banerjee M, Patra D, Banerjee P, Ghoshal N, et al. Functional compensation of glutathione S-transferase M1 (GSTM1) null by another GST superfamily member, GSTM2. Scientific reports, 2013; 3: 2704.
3. McIlwain, C., D. Townsend, and K. Tew, *Glutathione S-transferase polymorphisms: cancer incidence and therapy.* Oncogene, 2006; 25(11): 1639-1648.
4. Pruitt, K.D., et al., *NCBI Reference Sequences: current status, policy and new initiatives.* Nucleic acids research, 2008; 37(suppl_1): D32-D36.
5. Pruitt KD, Tatusova T, Klimke W, Maglott DR. NCBI Reference Sequences: current status, policy and new initiatives. Nucleic acids research, 2008; 37(suppl_1): D32-D6.
6. Collins, F.S., L.D. Brooks, and A. Chakravarti, *A DNA polymorphism discovery resource for research on human genetic variation.* Genome research, 1998; 8(12): 1229-1231.
7. Schork, N.J., D. Fallin, and J.S. Lanchbury, *Single nucleotide polymorphisms and the future of genetic epidemiology.* Clinical genetics, 2000; 58(4): 250-264.
8. Emahazion T, Feuk L, Jobs M, Sawyer SL, Fredman D, St Clair D, et al. SNP association studies in Alzheimer's disease highlight problems for complex disease analysis. TRENDS in Genetics, 2001; 17(7): 407-13.

9. Ng, P.C. and S. Henikoff, *SIFT: Predicting amino acid changes that affect protein function.* Nucleic acids research, 2003; 31(13): 3812-3814.
10. Adzhubei, I., D.M. Jordan, and S.R. Sunyaev, *Predicting functional effect of human missense mutations using PolyPhen-2.* Current protocols in human genetics, 2013; 7.20: 1-7.20. 41.
11. Capriotti E, Fariselli P, Calabrese R, Casadio R. Predicting protein stability changes from sequences using support vector machines. Bioinformatics, 2005; 21(suppl_2): ii54-ii8.
12. Calabrese R, Capriotti E, Fariselli P, Martelli PL, Casadio R. Functional annotations improve the predictive score of human disease-related mutations in proteins. Human mutation, 2009; 30(8): 1237-44.
13. Thomas PD, Kejariwal A, Campbell MJ, Mi H, Diemer K, Guo N, et al. PANTHER: a browsable database of gene products organized by biological function, using curated protein family and subfamily classification. Nucleic acids research, 2003; 31(1): 334-41.
14. Venselaar H, te Beek TA, Kuipers RK, Hekkelman ML, Vriend G. Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. BMC bioinformatics, 2010; 11(1): 548.
15. Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, et al. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. Nucleic acids research, 2010; 38(suppl_2): W214-W220.
16. Rajasekaran R, Sudandiradoss C, Doss CGP, Sethumadhavan R. Identification and in silico analysis of functional SNPs of the BRCA1 gene. Genomics, 2007; 90(4): 447-52.
17. Fredman D, Siegfried M, Yuan YP, Bork P, Lehväslaiho H, Brookes AJ. HGVbase: a human sequence variation database emphasizing data quality and a broad spectrum of data sources. Nucleic Acids Research, 2002; 30(1): 387-91.
18. Yadav, P., A. Chatterjee, and A. Bhattacharjee, *Identification of deleterious nsSNPs in α, μ, π and θ class of GST family and their influence on protein structure.* Genomics data, 2014; 2: 66-72.