Research Journal of Oncology

Open access Research article

Investigation of Non-Synonymous Snps in Genes Associated with Oxidative Stress that may be Important in Thyroid Carcinogenesis

Teixeira ES[1], Dal' Bó IF[1], Nascimento M[1], Leão SLS[1], Ferreira Filho AC[1], Torres IOS[1], Rabi LT[1,2], Peres KC[1]; Cunha LL[3], Bufalo NE[1,4,5], Ward LS[1*]

[1]Department of Cancer Molecular Genetics, University of Campinas (UNICAMP), Brazil
[2]Department of Health Science, Paulista University (UNIP), Brazil
[3]Departmet of Endocrinology, Federal University of São Paulo, Brazil
[4]Department of Pathology, São Leopoldo Mandic Institute and Research Center, Brazil
[5]Department of Medicine, Max Planck University Center (UNIMAX), Brazil

ABSTRACT

Thyroid cells have intense circulation of free radicals and oxidizing metabolites such as hydrogen peroxide, from the synthesis of thyroid hormones, and iodide, from the iodination of thyroglobulin. Without an efficient antioxidant system, the generation of reactive oxygen species (ROS) can cause deleterious effects leading to DNA damage. ROS have been associated with many diseases, including cancer. Mitochondrial superoxide dismutase MnSOD (SOD2), glutathione peroxidase (GPX-1), glucose-6-phosphate dehydrogenase (G6PD), and p22phox (one of the subunits of the NOX enzyme complex) are transcribed by the SOD2, GPX-1, G6PD and CYBA genes, respectively. They play an important role in the generation of reactive species and in redox control and are crucial in cellular protection against oxidative stress. Genetic variants can affect protein function and therefore promote disturbances of redox balance, which increases the risk of cell damage by ROS. The connection between oxidative stress and thyroid diseases has been extensively investigated and suggests an important role for SOD2, GPX-1, G6PD and CYBA variants. To better understand the role of variants in the function of the corresponding proteins and their potential effect on thyroid carcinogenesis, we used bioinformatics tools to perform in silico analyzes of non-synonymous SNPs (nsSNPs) of these genes. A total of 1662 nsSNPs were retrieved from the NCBI database dbSNP data and analyzed by a suite of computational platforms: SIFT, PROVEAN, PolyPhen 2.0, PANTHER, SNAP 2, PhD-SNP, SNPs and GO, PMut, Mupro and I-Mutant v3. 23 nsSNPs were predicted by the tool consensus to be harmful. In conclusion, we demonstrate that in silico study can provide a solid foundation and assist researchers in the selection of SNPs, optimizing laboratory experimental analyses.

Keywords: Polymorphism; Antioxidant; Bioinformatics; Thyroid; Carcinogenesis

ABBREVIATIONS

(ROS) Reactive Oxygen Species; (SOD2) Superoxide Dismutase 2 Mitochondrial; (GPX-1) Glutathione Peroxidase); (G6PD) Glucose-6-Phosphate Dehydrogenase; (CYBA) Cytochrome B-245 Alpha Chain; (SNPs) Single Nucleotide Polymorphisms; (nsSNPs) Single Nucleotide Polymorphisms Nonsynonymous; (GSH) Glutathione; (PPP) Pentose Phosphate; (NADPH) Nicotinamide Adenine Dinucleotide Hydrogen Phosphate; (NADP) Nicotinamide Adenine Dinucleotide Phosphate; (PCR) Polymerase Chain Reaction; (SSCP) Single Strand Conformational Polymorphism; (AA) Amino Acid Sequence; (SIFT) Sorting Intolerant From Tolerant; (PolyPhen-20 Polymorphism Phenotyping v2; (PROVEAN) Protein Variation Effect Analyzer; PANTHER Protein

Received:	29-June-2022	**Manuscript No:**	IPRJO-22-13721
Editor assigned:	01-July-2022	**PreQC No:**	IPRJO-22-13721 (PQ)
Reviewed:	15-July-2022	**QC No:**	IPRJO-22-13721
Revised:	20-July-2022	**Manuscript No:**	IPRJO-22-13721 (R)
Published:	27-July-2022	**DOI:**	10.36648/iprjo.6.4.16

Corresponding author Laura Sterian Ward, Department of Cancer Molecular Genetics, University of Campinas (UNICAMP), Brazil, E-mail: ward@fcm.unicamp.br

Citation Teixeira ES, Dal' Bó IF, Nascimento M, Leão SLS, Ferreira Filho AC, et al. (2022) Investigation of Non-Synonymous Snps in Genes Associated with Oxidative Stress that may be Important in Thyroid Carcinogenesis. Res J Onco Vol.6:4

Analysis Through Evolutionary Relationships; (PHD-SNP) Predictor of Human Delerious Single Nucleotide Polymorphisms

INTRODUCTION

Numerous risk factors have been explored and identified as potential triggers or regulators of the pathogenesis of thyroid cancer (TC), including the production of free radicals and reactive oxygen species (ROS), and the genetic variations represented mainly by single nucleotide polymorphisms (SNPs) of genes involved in this process. One of the most important risk factors for TC, especially for the most common type, papillary carcinoma, is ionizing radiation. Ionizing radiation stimulates the expression of ROS generating systems, which cause DNA damage, promoting chromosomal instability, tumorigenesis and dedifferentiation [1-6].

In addition during the synthesis of thyroid hormones, thyroid follicular cells have intense circulation of free radicals and oxidant metabolites such as hydrogen peroxide [7]. Both normal and cancerous thyroid cells have been demonstrated to be particularly sensitive to the action of ROS-induced oxidative damage to DNA. Without an efficient antioxidant system, the generation of reactive oxygen species (ROS) can cause deleterious effects leading to DNA damage that ultimately favors mutagenesis. Some antioxidant enzymes are important for thyroid protection and many studies have investigated genetic variations in genes encoding these enzymes and their relationship to cancer risk, but the results have been inconclusive, and data on the risk of thyroid cancer are still lacking [3,4].

Manganese superoxide dismutase (SOD2) is the main antioxidant in mitochondria, catalyzing the dismutation of superoxide anions into H_2O_2, which is then reduced to water by catalase (CAT) or glutathione peroxidase. Increased SOD2 expression has been associated with a greater increase in tumor burden accompanied by increased cell proliferation. In contrast, SOD2 overexpression reduced tumor proliferation and mortality in FCT mice. In human cancers, downregulation of SOD2 gene expression was observed in FTC but not in PTC.

Glutathione peroxidases constitute a family of related oxidoreductases distributed in all living domains, involved in the termination reaction of the ROS pathway. Cytosolic glutathione peroxidase (GPX1), an intracellular antioxidant enzyme located in the cytosol, mitochondria and selenium containing peroxisomes, is highly abundant in the thyroid and has been implicated in the development of head and neck, lung, and breast cancer [8-12].

The pentose phosphate (PPP) pathway plays an important role in the biosynthesis of ribonucleotide precursors and nicotinamide adenine dinucleotide hydrogen phosphate (NADPH). The G6PD enzyme is critical for the conversion of nicotinamide adenine dinucleotide phosphate (NADP) to NADPH during cellular metabolism within the PPP pathway. The conversion of NADP to NADPH is critical for the production of glutathione, an important antioxidant that helps protect erythrocytes against oxidative stress. Recent studies suggest that G6PD exerts an additive or synergistic effect in inhibiting cell growth in thyroid cancer cells [13,14].

The family proteins NOXs, unlike other oxidoreductases that generate ROS only as a byproduct along the catalytic pathways, are enzymes specialized in the production of ROS. In recent years, studies have demonstrated the role of NOX in a variety of physiological and pathophysiological processes, including cancer development, and some studies have reported that the inhibition of NOX activity can inhibit tumor growth and promote cancer cell death. The NOX complex is formed by subunits, including the p22phox protein, encoded by the CYBA gene. The function of p22phox has the potential to influence the activity of the protein, being able to alter the generation of ROS in different tissues and under different conditions [15-21].

The study of SNPs plays an important role in the identification of genetic variants and aids in the search for potential biomarkers for the investigation of consequences on protein function and its role in human diseases. However, there are approximately 10 million SNPs in the human genome.

Some SNPs are functional, that is, they can influence the corresponding gene expression by direct or indirect pathways. Synonymous or silent variants are present in the coding regions of the gene and do not result in amino acid changes, but can produce serious splicing defects. In contrast, non-synonymous SNPs (nsSNPs), although also in the coding region of the gene, cause amino acid exchange [22-24].

Missense nsSNPs make nucleotide substitutions and cause an amino acid change that can alter the protein sequence. In protein coding regions, missense like nsSNPs can lead to changes in protein structure and function, hence altering phenotype and causing severe genetic disorders. Some nsSNPs are important for clinical application and may function as predisposition, diagnostic or prognostic markers [22,23,25].

Numerous techniques are available for the identification of SNPs, including amplification and DNA sequencing, allele specific polymerase chain reaction (PCR), and single strand conformational polymorphism analysis (SSCP) (29). These techniques, unfortunately, are only suitable for the analysis of a small number of SNPs in a relatively small number of individuals due to high cost. In recent years, in silico analysis has facilitated the investigation of SNPs, playing an important role in biology. Several bioinformatics tools are currently available for the analysis of structural and functional changes of synonymous or non-synonymous SNPs [26-29].

As technologies advance, there is a continuous influx of new variants in different genes. However, information on the clinical impact of these variants is still scarce. To avoid the labor and cost of investigations of these SNPs with structural and functional consequences on all new SNPs, in silico analysis offers the opportunity to predict their outcome and select candidates for in vivo experiments [30].

SNPs for SOD2, GPX-1, G6PD and CYBA have not been analyzed so far in silico. Therefore, we designed a strategy to analyze the entire coding region of the corresponding genes using different bioinformatics algorithms. Analyses were performed to predict high risk nsSNPs in coding regions that are likely to have an effect on protein function and structure, hence deserving further evaluation and validation in large cohorts of patients and of functional assays.

MATERIALS AND METHODS

Retrieval of Datasets

Based The SNPs associated with the SOD2, G6PD, GpX1, and CYBA genes were retrieved from the single nucleotide polymorphism (dbSNP) database (http://www.ncbi.nlm.nih.gov/snp/) and are commonly referenced. by their reference string IDs (rsIDs). Amino acid sequences (AAs) were retrieved from UniProt (https://www.UniProt.org/UniProt/), ID: P04179; ID: P11413; ID: P07203 and ID: P13498. Information about human genes and proteins was collected from the Online Mendelian Inheritance in Man (OMIM) database (https://www.omim.org), and the ClinVar database was used for amino acid change searches to identify disease associated variants [31].

Validation of Tolerated and Deleterious SNPs

We selected nsSNPs that could potentially influence protein function, subsequently altering the carrier phenotype. To predict and analyze the effect of nsSNPs of the SOD2, G6PD, GPX1, and CYBA genes on the function of each protein, the following in silico tools were used:

SIFT (Sorting intolerant from tolerant) (https://sift.bii.a-star.edu.sg/) is a tool that employs sequence homology to predict the impact of amino acid substitutions on protein function. SIFT can differentiate functionally neutral amino acid changes from functionally deleterious ones [32]. SIFT assumes that important positions in a protein sequence must be conserved throughout evolution; therefore, substitutions at these positions can affect protein function. The SIFT score ranges from 0 to 1, and scores ≤ 0.05 are predicted to be deleterious substitutions, while scores >0.05 are considered tolerated. Reference IDs (rsIDs) for each gene were provided as input values, and the score values along with their interpretations were recorded.

PolyPhen-2 (Polymorphism Phenotyping v2) (http://genetics.bwh.harvard.edu/pph2/) classifies and predicts the functional impacts of each AA substitution on the structural and functional properties of the protein. PolyPhen2 classifies the SNPs into 3 different classes: (1) benign [score=0.0], (2) possibly damaging or (3) probably damaging [score=1.0]. The FASTA sequences of the proteins were used as input to the PolyPhen2 web server [33].

Protein Variation Effect Analyzer (PROVEAN) (http://provean.jcvi.org) is a server that predicts the functional impact of amino acid substitutions in a protein, providing high throughput results at the genomic and protein levels for human and mouse variants. The FASTA sequence of the proteins was used as input in the PROVEAN tool. The variant is considered "detrimental" if the final score is less than -2.5 and is considered "neutral" if the score is greater than -2.5 [34].

The SNAP2 neural network based functional tool predicts the effects of nsSNPs on protein function and secondary structure, making predictions across the characteristics of the wild type protein and its variants. The prediction score ranges from -100 (for neutral prediction) to +100 (strong effect), which represents the probability of nsSNPs modifying native protein function. SNAP2 provides a heatmap with possible substitution at each position of the protein, where scores >50 are displayed in dark red, indicating a greater likelihood of pathogenicity [35].

PANTHER (Protein Analysis through Evolutionary Relationships) (http://www.pantherdb.org) estimates the position specific evolutionary conservation of the amino acid sequence, predicting the probability of nsSNPs causing a functional impact on the protein. PANTHER uses it as a measure of the period of time (in millions of years) a position is preserved in the protein. The longer the preservation time is, the greater the likelihood of a functional impact on the protein [36].

Identifying Disease-Associated nsSNPs

The SNPs and GO algorithm is a web server that predicts the impact of protein mutations using information from the three main roots encoded by genetic ontology (GO) terms: molecular function, biological process and cellular component. From the FASTA sequence of the protein, SNPs and GO predicts the probability of disease related mutations with 82% accuracy. The FASTA sequence of each full length protein was used as an input option. The results based on the discrimination of "Disease" and "neutral" variations were recorded.

PMut is a functional tool based on neural network (NN) intelligence that allows one to accurately and quickly display pathological characteristics caused by a single amino acid substitution. The prediction can be considered neutral or disease causing. The input mechanism for PMut is the FASTA protein sequence or SwissProt code. The result of pathogenicity is applied with a variation index from 0 to 1, where index >0.5 indicates pathological mutations [37,38].

The Predictor of Human Delerious Single Nucleotide Polymorphisms (PHD-SNP) (http://snps.biofold.org/phd-snp/phd-snp.html) was also used to determine the effects of amino acid exchange in causing disease [39].

Prediction of Stability Related Mutations

Prediction of the functional impact of mutations on protein stability was verified using the tools I-Mutant v3.0 and MUpro. I-Mutantv3.0 (http://gpcr2.biocomp.unibo.it/cgi/predictors/I-Mutant3.0/I-Mutant3.0.cgi), a support vector machine based algorithm, estimates the variation in protein stability change and this prediction is based on the variation in free energy by the ΔΔG/DDG (kcal/mol) value upon mutation of a single site in the protein structure or sequence. A ΔΔG value less than (<−0.5 kcal/mol) indicates that the variant decreases protein stability. A ΔΔG>0.5 kcal/mol indicates that the variant enhances protein stability [40].

MUpro (http://mupro.proteomics.ics.uci.edu/) is also a web server that predicts changes in protein stability after a single amino acid substitution, and the cutoff value of ΔΔG is the same as that used in the I-mutant [38].

Protein-Protein Interaction

Mutations can alter the structure and function of the protein and therefore the interaction of the mutated protein with other proteins can be affected [41].

STRING is a web-based tool for gene retrieval and protein interaction and was used to investigate the interaction of SOD2, GPX1, G6PD and NOX with other proteins. This usable resource

aggregates available information on protein-protein associations, scores predicted interactions, and points to research findings [40].

RESULTS

SNP Dataset

We evaluated a total of 593 nsSNPs of SOD2, 389 nsSNPs of G6PD, 324 nsSNPs of GPX1 and 356 missense SNPs of CYBA using state of the art bioinformatics tools, all retrieved from dbSNP.

The nsSNPs IDs were sent as input to the SIFT server, and the results are shown in **Table 1**. The lower the tolerance index is, the greater the functional impact the amino acid switch is likely to have Among the 1662 nsSNPs analyzed in SIFT, 143 nsSNPs were identified as deleterious with a tolerance index ≤ 0.05. Among 143 deleterious nsSNPs, 22 nsSNPs were considered highly deleterious.

Table 1: Deleterious nsSNPs using SIFT

S. No	GENE	rsID	AA change	Position	Prediction	Score
1	SOD2	rs5746129	R/W	156	D	0.01
2	SOD2	rs11575993	L/F	84	D	0.014
3	SOD2	rs11575993	L/F	38	D	0.045
4	SOD2	rs185564053	G/W	18	D	0
5	SOD2	rs370671213	H/Q	55	D	0
6	SOD2	rs372074075	T/M	136	D	0.015
7	SOD2	rs373540824	E/K	67	D	0.022
8	SOD2	rs375177938	H/Y	51	D	0.049
9	SOD2	rs375884951	L/Q	38	D	0
10	SOD2	rs376398472	V/F	142	D	0.008
11	GPX1	rs11552757	A/V	161	D	0
12	GPX1	rs112304179	F/L	171	D	0.028
13	GPX1	rs183107871	E/D	165	D	0.005
14	GPX1	rs200311870	Q/R	84	D	0
15	GPX1	rs201944086	P/R	77	D	0.031
16	GPX1	rs370228556	D/V	191	D	0.001
17	GPX1	rs370714711	G/V	170	D	0.02
18	GPX1	rs373838463	L/Q	168	D	0.042
19	GPX1	rs377594183	P/L	154	D	0.011
20	G6PD	rs1050828	V/M	98	D	0.01
21	G6PD	rs1050828	V/M	68	D	0.011
22	G6PD	rs5030868	S/F	218	D	0.008
23	G6PD	rs5030868	S/F	188	D	0.008
24	G6PD	rs5030869	A/T	381	D	0.006
25	G6PD	rs5030869	A/T	365	D	0.007
26	G6PD	rs34193178	D/H	350	D	0.019
27	G6PD	rs34193178	D/H	396	D	0.02
28	G6PD	rs34193178	D/H	380	D	0.022
29	G6PD	rs72554664	R/H	509	D	0.022
30	G6PD	rs72554664	R/H	493	D	0.03
31	G6PD	rs72554664	R/H	463	D	0.035
32	G6PD	rs74575103	R/H	331	D	0.011
33	G6PD	rs74575103	R/H	315	D	0.016
34	G6PD	rs74575103	R/H	285	D	0.016
35	G6PD	rs78365220	L/P	128	D	0.049
36	G6PD	rs78478128	A/G	44	D	0.002
37	G6PD	rs78478128	A/G	74	D	0.003
38	G6PD	rs137852314	G/S	163	D	0.032
39	G6PD	rs137852314	G/S	193	D	0.036
40	G6PD	rs137852316	R/H	439	D	0.001
41	G6PD	rs137852316	R/H	423	D	0.001

42	G6PD	rs137852316	R/H	393	D	0.001
43	G6PD	rs137852317	G/R	493	D	0
44	G6PD	rs137852317	G/R	477	D	0.001
45	G6PD	rs137852318	D/H	312	D	0.002
46	G6PD	rs137852318	D/H	328	D	0.003
47	G6PD	rs137852318	D/H	282	D	0.003
48	G6PD	rs137852319	F/L	216	D	0.001
49	G6PD	rs137852319	F/L	246	D	0.002
50	G6PD	rs137852321	R/H	417	D	0.044
51	G6PD	rs137852323	G/C	456	D	0
52	G6PD	rs137852323	G/C	440	D	0
53	G6PD	rs137852323	G/C	410	D	0
54	G6PD	rs137852324	R/H	500	D	0
55	G6PD	rs137852324	R/H	484	D	0
56	G6PD	rs137852324	R/H	454	D	0
57	G6PD	rs137852325	R/K	444	D	0.001
58	G6PD	rs137852325	R/K	428	D	0.002
59	G6PD	rs137852326	V/L	213	D	0.044
60	G6PD	rs137852327	V/M	337	D	0.001
61	G6PD	rs137852327	V/M	321	D	0.001
62	G6PD	rs137852327	V/M	291	D	0.001
63	G6PD	rs137852328	R/L	257	D	0.001
64	G6PD	rs137852328	R/Q	257	D	0.003
65	G6PD	rs137852329	N/K	393	D	0.012
66	G6PD	rs137852330	R/C	198	D	0
67	G6PD	rs137852330	R/C	228	D	0
68	G6PD	rs137852332	R/P	228	D	0
69	G6PD	rs137852332	R/P	198	D	0
70	G6PD	rs137852332	R/H	228	D	0
71	G6PD	rs137852332	R/H	198	D	0
72	G6PD	rs137852333	P/S	399	D	0.005
73	G6PD	rs137852333	P/S	383	D	0.007
74	G6PD	rs137852333	P/S	353	D	0.007
75	G6PD	rs137852334	R/C	433	D	0.018
76	G6PD	rs137852334	R/C	417	D	0.02
77	G6PD	rs137852336	G/D	440	D	0
78	G6PD	rs137852336	G/D	410	D	0
79	G6PD	rs137852336	G/D	456	D	0.012
80	G6PD	rs137852337	R/P	485	D	0.004
81	G6PD	rs137852337	R/P	439	D	0.008
82	G6PD	rs137852337	R/P	469	D	0.009
83	G6PD	rs137852341	G/V	131	D	0.035
84	G6PD	rs137852341	G/V	161	D	0.047
85	G6PD	rs137852343	F/L	203	D	0
86	G6PD	rs137852343	F/L	173	D	0
87	G6PD	rs137852344	P/R	513	D	0
88	G6PD	rs137852344	P/R	467	D	0
89	G6PD	rs137852344	P/R	497	D	0.001
90	G6PD	rs137852345	A/V	391	D	0.001
91	G6PD	rs137852345	A/V	361	D	0.001
92	G6PD	rs137852345	A/V	407	D	0.005
93	G6PD	rs137852346	C/Y	269	D	0.003

94	G6PD	rs137852346	C/Y	315	D	0.004
95	G6PD	rs137852346	C/Y	299	D	0.015
96	G6PD	rs137852347	Y/H	368	D	0
97	G6PD	rs137852347	Y/H	352	D	0
98	G6PD	rs137852347	Y/H	322	D	0
99	G6PD	rs137852349	Y/H	100	D	0
100	G6PD	rs137852349	Y/H	70	D	0
101	G6PD	rs267606836	R/W	212	D	0.001
102	G6PD	rs267606836	R/W	182	D	0.002
103	G6PD	rs387906468	E/K	368	D	0.001
104	G6PD	rs387906468	E/K	398	D	0.002
105	G6PD	rs387906468	E/K	414	D	0.007
106	G6PD	rs387906471	E/K	333	D	0.001
107	G6PD	rs387906471	E/K	287	D	0.001
108	G6PD	rs387906471	E/K	317	D	0.002
109	G6PD	rs1050827	Q/H	11	D	0.032
110	G6PD	rs1050827	Q/H	41	D	0.048
111	G6PD	rs138687036	R/C	81	D	0.003
112	G6PD	rs138687036	R/C	111	D	0.006
113	G6PD	rs141830127	S/N	84	D	0.013
114	G6PD	rs141830127	S/N	114	D	0.029
115	G6PD	rs281860640	S/N	209	D	0.001
116	G6PD	rs281860640	S/N	179	D	0.001
117	G6PD	rs370451233	D/G	143	D	0.029
118	G6PD	rs387906467	R/H	403	D	0.02
119	G6PD	rs387906467	R/H	387	D	0.029
120	G6PD	rs387906467	R/H	357	D	0.03
121	G6PD	rs387906470	R/C	403	D	0
122	G6PD	rs387906470	R/C	387	D	0
123	G6PD	rs387906470	R/C	357	D	0
124	CYBA	rs8053867	E/D	12	D	0.017
125	CYBA	rs28941476	G/R	24	D	0.005
126	CYBA	rs104894510	H/R	94	D	0.012
127	CYBA	rs104894513	R/Q	90	D	0
128	CYBA	rs104894514	Q/R	118	D	0.002
129	CYBA	rs104894515	P/Q	156	D	0
130	CYBA	rs119103269	A/T	125	D	0.006
131	CYBA	rs149344911	V/M	76	D	0.009
132	CYBA	rs179363890	L/P	52	D	0.001
133	CYBA	rs179363891	G/V	25	D	0
134	CYBA	rs179363892	R/W	90	D	0
135	CYBA	rs179363893	E/V	53	D	0
136	CYBA	rs179363894	A/V	124	D	0.023
137	CYBA	rs201755210	S/L	98	D	0.045
138	CYBA	rs9940427	R/S	130	D	0.015
139	CYBA	rs11547384	Y/C	41	D	0.001
140	CYBA	rs13306297	R/Q	158	D	0.01
141	CYBA	rs145267803	P/S	379	D	0.007
142	CYBA	rs367729578	C/Y	386	D	0.037
143	CYBA	rs374698190	G/S	394	D	0

D=Damaging, N=Neutral

Validation of Tolerated and Deleterious SNPs

Of the 38 nsSNPs considered deleterious by SIFT, 22 were evaluated as deleterious in the consensus between the Poly-Phen-2, PROVEAN, SNAP2 and PANTHER tools, and only 2 variants (G25V and G394S) of the CYBA gene were considered neutral in more than one of the four tools, according to **Table 2**.

Table 2: Prediction of functional effects of nsSNPs using PolyPhen-2, PROVEAN, SNAP2 and PANTHER.

S. No	GENE	AA change	MAF*	PolyPhen-2	PROVEAN	SNAP2	PANTHER
1	SOD2	G18W	<0.01	D	D	D	N
2	SOD2	H55Q	<0.01	D	D	D	D
3	SOD2	L38Q	<0.01	D	D	D	D
4	GPX1	A161V	<0.01	D	D	D	D
5	GPX1	Q84R	<0.01	D	D	D	D
6	G6PD	G493R	<0.01	D	D	D	D
7	G6PD	G456C	<0.01	D	D	D	D
8	G6PD	R500H	<0.01	D	D	D	D
9	G6PD	R198C	<0.01	D	D	D	D
10	G6PD	R198P	ND	D	D	D	D
11	G6PD	G440D	<0.01	D	D	D	D
12	G6PD	F203L	<0.01	D	D	D	D
13	G6PD	P467R	<0.01	D	D	D	D
14	G6PD	Y368H	ND	D	D	D	D
15	G6PD	Y100H	<0.01	D	D	D	D
16	G6PD	R357C	ND	D	D	D	D
17	CYBA	R90Q	<0.01	D	D	D	D
18	CYBA	P156Q	<0.01	D	D	D	D
19	CYBA	G25V	<0.01	D	N	N	D
20	CYBA	R90W	<0.01	N	D	D	D
21	CYBA	E53V	<0.01	N	D	D	D
22	CYBA	G394S	<0.01	N	N	D	D

D=Damaging, N=Neutral, MAF=de minor allele frequence

Disease-Associated nsSNPs

All 22 identified nsSNPs were further analyzed by SNPs and GO, PMut and PHD-SNPs. The 8 nsSNPs were predicted to be associated with the disease by three methods (**Table 3**).

Table 3: Prediction of disease-related mutations using SNPs&GO, PMUT and phD-SNPs.

S. No	GENE	AA change	SN-Ps&GO	PMUT	phD-SNP
1	SOD2	G18W	D	D	N
2	SOD2	H55Q	N	N	D
3	SOD2	L38Q	N	N	D
4	GPX1	A161V	D	D	D
5	GPX1	Q84R	D	D	D
6	G6PD	G493R	D	D	N
7	G6PD	G456C	D	D	N
8	G6PD	R500H	D	D	N
9	G6PD	R198C	D	D	N
10	G6PD	R198P	D	D	D
11	G6PD	G440D	D	D	N
12	G6PD	F203L	D	D	N
13	G6PD	P467R	D	D	N
14	G6PD	Y368H	D	D	N
15	G6PD	Y100H	D	D	N
16	G6PD	R357C	D	D	N
17	CYBA	R90Q	D	D	D
18	CYBA	P156Q	N	D	D
19	CYBA	G25V	D	D	D
20	CYBA	R90W	D	D	D
21	CYBA	E53V	D	D	D
22	CYBA	G394S	D	D	D

D=Damaging N=Neutral

Validation of Stability-Related Mutations

I-Mutant v3.0 and MUpro estimate the effect of substitution on protein stability by calculating the reliability index. Of the 22 missense SNPs analyzed, 16 were predicted to cause a decrease in stability (**Table 4**).

S. No	GENE	AA change	I-Mutant v3.0	MUpro
1	SOD2	G18W	Increase	Increase
2	SOD2	H55Q	Decrease	Decrease
3	SOD2	L38Q	Decrease	Decrease
4	GPX1	A161V	Increase	Increase
5	GPX1	Q84R	Decrease	Decrease

6	G6PD	G493R	Decrease	Decrease
7	G6PD	G456C	Decrease	Decrease
8	G6PD	R500H	Decrease	Decrease
9	G6PD	R198C	Decrease	Decrease
10	G6PD	R198P	Decrease	Decrease
11	G6PD	G440D	Decrease	Decrease
12	G6PD	F203L	Decrease	Decrease
13	G6PD	P467R	Decrease	Decrease
14	G6PD	Y368H	Decrease	Decrease
15	G6PD	Y100H	Decrease	Decrease
16	G6PD	R357C	Decrease	Decrease
17	CYBA	R90Q	Decrease	Decrease
18	CYBA	P156Q	Decrease	Decrease
19	CYBA	G25V	Increase	Increase
20	CYBA	R90W	Increase	Increase
21	CYBA	E53V	Increase	Increase
22	CYBA	G394S	Increase	Increase

↓=Decrease stability, ↑=Increase stability

Protein-Protein Interaction

The STRING protein-protein interaction network queried with SOD2, GPX1, G6PD and CYBA. SOD2 and GPX1 proteins interact in the standard human protein-protein association network in STRING. We did not observe interactions between SOD2 or GPX1 and the interaction network of G6PD and CYBA (**Figure 1**).

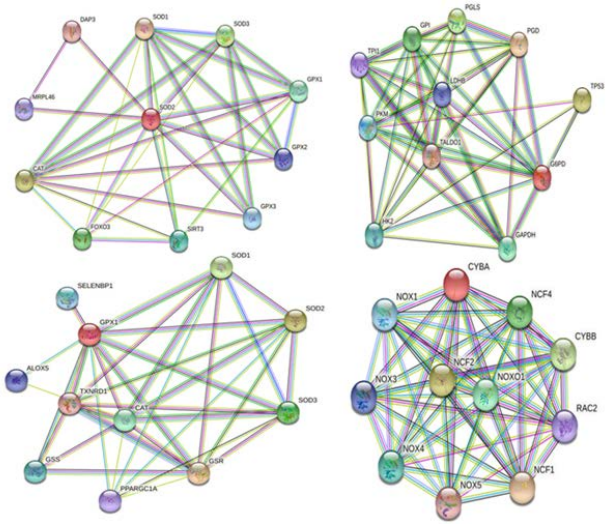

Figure 1: STRING protein–protein interaction network. Network of functions between bonds with the 10 most significant proteins.

We also evaluated the characteristics of the wild type and the mutated residues using the HOPE tool. The three variants of SOD2 (G18W, H55Q and L38Q) had changes in amino acid size; hydrophobicity; and binding to neighboring molecules. Glycine is the most flexible residue of all amino acids. This flexibility may be necessary for protein function and stability. In G18W, switching to tryptophan at this position can abolish function and modulate protein structure. The amino acid glutamine (Q) at position 55 (H55Q) and at position 38 (L38Q) are located in a domain that is important for binding with other molecules. It is possible that this change disturbs these contacts, affecting the interaction and thus interfering with the signal transfer from the binding domain to the activity domain.

Both GPX1 mutant residues (A161V and Q84R) are located in domains that are important for binding to other molecules and other domains and, in addition to disturbing this contact; these residues can alter the function of the protein. Wild type and mutant residues from all CYBA variants (R90Q, P156Q, G25V, R90W, E53V, and G394S) differ in size, hydrophobicity, and charge. The hydrophobicity reported in the variants (R90W and E53V) can result in the loss of hydrogen bonds and/or disturb the correct folding of the protein. In addition, modifications caused by the change in residue charge can cause loss of interactions with other molecules or residues and impair protein function. The conformation of the protein can also be disturbed by an amino acid change. The proline present in P156Q is an amino acid known to be very rigid, that induces a special conformation of the backbone. A modification of this amino acid in this position can change the flexibility and induce modifications in the protein conformation.

The G6PD variants (G493R, G456C, G456C, G440D, F203L, Y368H, Y100H and R500H) were not found in the HOPE platform and their characteristics could not be evaluated. All variants of G6PD (R198C, R198P, P467R and R357C) show differences in terms of size, hydrophobicity and load of the mutant and wild type. These features can disrupt the ionic interaction of molecules and contact with other residues and other domains. The loss of proline at position 467 (P467R) is likely to be detrimental to protein structure. This substitution modulates the twist angles necessary to maintain protein stability.

In **Figure 2** the schematic structures of the original and mutant amino acid are presented faithfully to the output data provided by the HOPE tool.

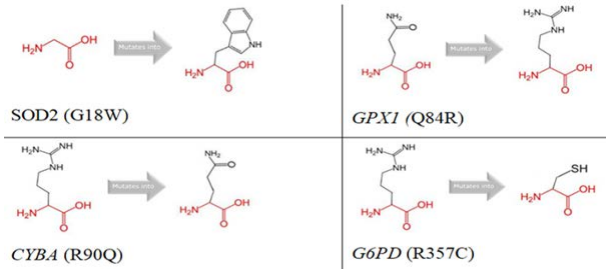

Figure 2: Schematic structures of the original (left) and mutant (right) amino acid. The backbone is represented in red and the side chain in black.

DISCUSSION

nsSNPs cause a substitution in the amino acid sequence of the polypeptide chain and can result in structural and functional abnormalities. Their investigation has important clinical applications. Previous genetic studies have examined the association of some polymorphisms in genes related to the oxidative stress pathway with thyroid cancer development. However, the large number of variants described in these genes impairs validation. In addition, there is insufficient evidence of any association with human diseases for most of these SNPs, making it difficult to sort out the polymorphisms worth further bench investigation or validation in patient cohorts [42-45].

Approximately 500,000 SNPs have been reported in coding re-

gions of the human genome, and many studies focus on nsSNPs that, by altering the amino acid residues of protein sequences, can cause harmful effects on protein functions or structures. With the large amount of human genome data available and the increasingly common use of in silico analysis, it has been possible to reduce the search for nsSNPs and thus save time and cost before proceeding with laboratory experiments.

In the present study, we performed in silico analyses to identify potential harmful nsSNPs in the SOD2, G6PD, GPX1 and CYBA genes [42,46].

We studied the functional, structural and stability consequences of 1662 nsSNPs from SOD2, G6PD, GPX1 and CYBA. Using a series of easily available bioinformatics tools, we were able to select 22 variants that have a high probability of being deleterious and affecting thyroid cancer risk and/or prognosis.

We also analyzed the properties of the amino acids generated by these 22 variants, providing more information about the role of each change and allowing the formulation of new hypotheses about their effects on protein function. Each amino acid has its own size, charge, and specific hydrophobicity values, and the evaluation of these characteristics can help select pathogenic variants. G18W, A161V, Q84R, R90W, E53V, P156Q, and P467R may modify protein function and structure, and their roledeserve further investigation in thyroid cancer.

This research has an obvious limitation, as we only performed in silico analyses. A large-scale study associated with nsSNPs with different populations and laboratory experiments may provide a more robust validation of our results. However, this research can provide a solid foundation for in vivo experiments, assist in the selection of SNPs of interest and thus help laboratory experimental analyses.

CONCLUSION

This research has an obvious limitation, as we only performed in silico analyses. A large-scale study associated with nsSNPs with different populations and laboratory experiments may provide a more robust validation of our results. However, this research can provide a solid foundation for in vivo experiments, assist in the selection of SNPs of interest and thus help laboratory experimental analyses.

ACKNOWLEDGEMENTS

The authors also thank American Journal Experts for the language services provided and the Coordination for the Improvement of Higher Education Personnel (CAPES) for financial support to the postgraduate students involved in this project. LSW is a recipient of the Brazilian National Council for Scientific and Technological Development (CNPq) researcher category 1 grant.

AUTHOR CONTRIBUTIONS

All authors contributed to the concept and design of this study or to data acquisition and interpretation. All authors contributed to the review of the manuscript and read and approved the submitted version.

DECLARATION OF CONFLICT OF INTEREST

The authors declare that the research was carried out in the absence of any commercial or financial relationship that could be interpreted as a potential conflict of interest.

REFERENCES

1. Versteyhe S, Driessens N, Ghaddhab C, Tarabichi M, Hoste C, et al. (2013) Comparative analysis of the thyrocytes and T cells: Responses to H_2O_2 and radiation reveals an H_2O_2-induced antioxidant transcriptional program in thyrocytes. J Clin Endocrinol Metab 98(10):E1645-54.
2. Panicker V (2011) Genetics of thyroid function and disease. Clin Biochem Rev 32(4):165-75.
3. Erdamar H, Cimen B, Gulcemal H, Saraymen R, Yerer B, et al. (2010) Increased lipid peroxidation and impaired enzymatic antioxidant defense mechanism in thyroid tissue with multinodular goiter and papillary carcinoma. Clin Biochem 43(7-8):650-4.
4. Xing M (2012) Oxidative stress: A new risk factor for thyroid cancer. Endocr Relat Cancer 19(1):C7-11.
5. Ameziane R, Buffet C, Leboulleux S, Dupuy C (2019) Oxidative stress in thyroid carcinomas: Biological and clinical significance. Endocr Relat Cancer 26(3):R131-R43.
6. Young O, Crotty T, O'Connell R, O'Sullivan J, Curran AJ (2010) Levels of oxidative damage and lipid peroxidation in thyroid neoplasia. Head Neck 32(6):750-6.
7. Carvalho DP, Dupuy C (2017) Thyroid hormone biosynthesis and release. Mol Cell Endocrinol 458:6-15.
8. Banescu C, Trifa AP, Voidazan S, Moldovan VG, Macarie I, et al. (2014) CAT, GPX1, MnSOD, GSTM1, GSTT1, and GSTP1 genetic polymorphisms in chronic myeloid leukemia: A case-control study. Oxid Med Cell Longev.;2014:875861.
9. Ashtekar A, Huk D, Magner A, La Perle KMD, Boucai L, et al. (2018) alterations in sod2-induced oxidative stress affect endocrine cancer progression. J Clin Endocrinol Metab 103(11):4135-45.
10. Howie AF, Walker SW, Akesson B, Arthur JR, Beckett GJ (1995) Thyroidal extracellular glutathione peroxidase: A potential regulator of thyroid-hormone synthesis. Biochem J 308 (Pt 3):713-7.
11. Schmutzler C, Mentrup B, Schomburg L, Hoang-Vu C, Herzog V, et al. (2007) Selenoproteins of the thyroid gland: Expression, localization and possible function of glutathione peroxidase 3. Biol Chem 388(10):1053-9.
12. Ubos E, Loscalzo J, Handy DE (2011) Glutathione peroxidase-1 in health and disease: From molecular mechanisms to therapeutic opportunities. Antioxid Redox Signal 15(7):1957-97.
13. Liu CL, Hsu YC, Lee JJ, Chen MJ, Lin CH, et al. (2020) Targeting the pentose phosphate pathway increases reactive ox-

ygen species and induces apoptosis in thyroid cancer cells. Mol Cell Endocrinol 499:110595.

14. Sparrow RL (2017) Red blood cell components: Time to revisit the sources of variability. Blood Transfus 15(2):116-25.
15. Montezano AC, Touyz RM (2014) Reactive oxygen species, vascular Noxs, and hypertension: Focus on translational and clinical research. Antioxid Redox Signal 20(1):164-82.
16. Weyemi U, Caillou B, Talbot M, Ameziane-El-Hassani R, Lacroix L, et al. (2010) Intracellular expression of reactive oxygen species-generating NADPH oxidase NOX4 in normal and cancer thyroid tissues. Endocr Relat Cancer 17(1):27-37.
17. Maranchie JK, Zhan Y (2005) Nox4 is critical for hypoxia-inducible factor 2-alpha transcriptional activity in von Hippel-Lindau-deficient renal cell carcinoma. Cancer Res 65(20):9190-3.
18. Thomasz L, Oglio R, Salvarredi L, Perona M, Rossich L, et al. (2018) Regulation of NADPH oxidase NOX4 by delta iodolactone (IL-delta) in thyroid cancer cells. Mol Cell Endocrinol.;470:115-26.
19. San Jose G, Fortuno A, Beloqui O, Diez J, Zalba G (2008) NADPH oxidase CYBA polymorphisms, oxidative stress and cardiovascular diseases. Clin Sci (Lond) 114(3):173-82.
20. Bedard K, Krause KH (2007) The NOX family of ROS-generating NADPH oxidases: Physiology and pathophysiology. Physiol Rev 87(1):245-313.
21. Lambeth JD. (2004) NOX enzymes and the biology of reactive oxygen. Nat Rev Immunol 4(3):181-9.
22. Pirooz HJ, Jafari N, Rastegari M, Fathi-Roudsari M, Tasharrofi N, et al. (2018) Functional SNP in microRNA-491-5p binding site of MMP9 3'-UTR affects cancer susceptibility. J Cell Biochem 119(7):5126-34.
23. Rashid MU, Gull S, Asghar K, Muhammad N, Amin A, et al. (2012) Prevalence of TP53 germ line mutations in young Pakistani breast cancer patients. Fam Cancer 11(2):307-11.
24. Griffiths AJF, Gelbart WM, Lewontin RC, Miller JH (2002) Modern genetic analysis: Integrating genes and genomes
25. Newfeld S, Newfeld SJ. (2002) Modern genetic analysis: Integrating genes and genomes. JSTOR
26. Fredriksson H, Ikonen T, Autio V, Matikainen MP, Helin HJ, et al. (2006) Identification of germline MLH1 alterations in familial prostate cancer. Eur J Cancer 42(16):2802-6.
27. Gillanders EM, Pearson JV, Sorant AJ, Trent JM, O'Connell JR, et al. (2006) The value of molecular haplotypes in a family-based linkage study. Am J Hum Genet.;79(3):458-68.
28. Tanaka Y, Hirata H, Chen Z, Kikuno N, Kawamoto K, et al. (2007) Polymorphisms of catechol-O-methyltransferase in men with renal cell cancer. Cancer Epidemiol Biomarkers Prev 16(1):92-7.
29. Kamio K, Matsushita I, Tanaka G, Ohashi J, Hijikata M, et al. (2004) Direct determination of MUC5B promoter haplotypes based on the method of single-strand conformation polymorphism and their statistical estimation. Genomics 84(3):613-22.
30. Kirk BW, Feinsod M, Favis R, Kliman RM, Barany F (2002) Single nucleotide polymorphism seeking long term association with complex disease. Nucleic Acids Res 30(15):3295-311.
31. Landrum MJ, Chitipiralla S, Brown GR, Chen C, Gu B, et al. (2020) ClinVar: Improvements to accessing data. Nucleic Acids Res 48(D1):D835-D44.
32. Sim NL, Kumar P, Hu J, Henikoff S, Schneider G, et al. (2012) SIFT web server: Predicting effects of amino acid substitutions on proteins. Nucleic Acids Res 40(Web Server issue):W452-7.
33. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, et al. (2010) A method and server for predicting damaging missense mutations. Nat Methods.;7(4):248-9.
34. Choi Y, Chan AP (2015) PROVEAN web server: A tool to predict the functional effect of amino acid substitutions and indels. Bioinformatics 31(16):2745-7.
35. Hecht M, Bromberg Y, Rost B (2015) Better prediction of functional effects for sequence variants. BMC Genomics 16 Suppl 8:S1.
36. Mi H, Ebert D, Muruganujan A, Mills C, Albou LP, et al. (2021) PANTHER version 16: A revised family classification, tree-based classification tool, enhancer regions and extensive API. Nucleic Acids Res 49(D1):D394-D403.
37. Calabrese R, Capriotti E, Fariselli P, Martelli PL, Casadio R (2009) Functional annotations improve the predictive score of human disease-related mutations in proteins. Hum Mutat 30(8):1237-44.
38. Cheng J, Randall A, Baldi P (2006) Prediction of protein stability changes for single-site mutations using support vector machines. Proteins 62(4):1125-32.
39. Capriotti E, Calabrese R, 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):2729-34.
40. Bava KA, Gromiha MM, Uedaira H, Kitajima K, Sarai A (2004) ProTherm, version 4.0: Thermodynamic database for proteins and mutants. Nucleic Acids Res 32(Database issue):D120-1.
41. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, et al. (2015) STRING v10: Protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res 43(Database issue):D447-52.
42. Collins FS, Brooks LD, Chakravarti A (1998) A DNA polymorphism discovery resource for research on human genetic variation. Genome Res 8(12):1229-31.
43. Salimi S, Harati-Sadegh M, Eskandari M, Heidari Z (2020) The effects of the genetic polymorphisms of antioxidant enzymes on susceptibility to papillary thyroid carcinoma. IUBMB Life 72(5):1045-53.
44. Vodusek AL, Goricar K, Gazic B, Dolzan V, Jazbec J (2016)

Antioxidant defence-related genetic variants are not associated with higher risk of secondary thyroid cancer after treatment of malignancy in childhood or adolescence. Radiol Oncol 50(1):80-6.

45. Krhin B, Goricar K, Gazic B, Dolzan V, Besic N (2016) Functional polymorphisms in antioxidant genes in Hurthle cell thyroid neoplasm: An association of GPX1 polymorphism and recurrent Hurthle cell thyroid carcinoma. Radiol Oncol 50(3):289-96.

46. Capriotti E, Altman RB (2011) Improving the prediction of disease-related variants using protein three-dimensional structure. BMC Bioinformatics 12 (Suppl 4):S3.