Thirty two novel nsSNPs May effect on *HEXA* protein Leading to Tay-Sachs disease (TSD) Using a Computational Approach

Tebyan A. Abdelhameed¹, Mohamed Mustafa Osman Fadul¹, Dina Nasereldin Abdelrahman Mohamed^{1,2}, Amal Mohamed Mudawi^{1,3}, Sayaf Kamal Khalifa Fadul Allah¹, Ola Ahmed Elnour Ahmed¹, Sogoud Mohammednour Idrees Mohammeddeen¹, Aya Abdelwahab Taha khairi¹, Soada Ahmed Osman¹, Ebrahim Mohammed Al-Hajj¹, Mustafa Elhag¹, Mohamed Ahmed Hassan Salih¹.

1. Department of biotechnology, Africa City of Technology-Khartoum, Sudan.

2. Central Laboratory, Ministry of Higher Education and Scientific Research- Khartoum, Sudan.

3. Faculty of Dentistry, University of Khartoum- Khartoum, Sudan.

Corresponding author: Tebyan A. Abdelhameed: tebyanameer2511@gmail.com.

ABSTRACT

Background: Genetic polymorphisms in the *HEXA* gene are associated with a neurodegenerative disorder called Tay-Sachs disease (TSD) (GM2 gangliosidosis type 1). This study aimed to predict the possible pathogenic SNPs of this gene and their impact on the protein using different bioinformatics tools.

Methods: SNPs retrieved from the NCBI database were analyzed using several bioinformatics tools. The different algorithms collectively predicted the effect of single nucleotide substitution on both structure and function of the hexosaminidase A protein.

Results: Fifty nine mutations were found to be highly damaging to the structure and function of the *HEXA* gene protein.

Conclusion: According to this study, thirty two novel nsSNP in *HEXA* are predicted to have possible role in Tay-Sachs Disease using different bioinformatics tools. Our findings could help in genetic study and diagnosis of Tay-Sachs Disease.

Keywords: *Tay-Sachs disease (TSD), GM2 gangliosidosis, hexosaminidase A, HEXA, a neurodegenerative disorder, bioinformatics, single nucleotide polymorphisms (SNPs), computational, insilico.*

INTRODUCTION

Tay-Sachs disease (TSD -[OMIM* 272800] also called GM2 gangliosidosis type 1 is an autosomal recessive lysosomal storage disorder which mainly affects Ashkenazi Jewish (AJ) population with low occurrences in non-AJ populations [1-5]. It's incidence is one in 100,000 live births (carrier frequency of about one in 250) [6]. Deficiency of β -hexosaminidase-A (Hex-A) isozyme which cleaves the terminal N-acetyl hexosamine residues of GM2 ganglioside is the cause of the TSD. This deficiency leads to accumulation of GM2

ganglioside which is caused by mutations in *HEXA* gene the highest concentration of GM2 ganglioside found in neuronal cells and they are the main glycolipids of neuronal cell's plasma membranes responsible of the normal cellular activities. Patients with HexA deficiency suffer from GM2 ganglioside accumulates inside neural's lysosomes. Therefore, HexA deficiency primarily affects the nervous system [1, 2, 4, 6-14].

According to the disease onset age, TSD is differentiated into three types which are: acute infantile, juvenile, and late-onset. Acute infantile TSD is the most common type. It causes progressive weakness and motor skills lost in the infected infants. This progression occurs between the ages of six months to three years. In addition, the infected infant progressively suffers from diminished attentiveness and an exaggerated startle response. As TSD continues to destroy the brain infant suffers from seizures, blindness, and death which usually occurs before four years of age [8, 9, 11, 13-15].

The human *HEXA* gene [OMIM *606869] encodes the alpha subunit of the lysosomal hexosaminidase A isozyme (HexA) which is located on chromosome 15 q23-q24 and contains 14 exons. More than two hundred mutations have been reported in the *HEXA* gene to cause TSD and its variants. These mutations include single base substitutions, small deletion, duplications, insertions splicing alterations, complex gene rearrangement, partial large duplications, partial deletion, splicing mutations, nonsense mutations and missense mutations that lead to the disruption of transcription, translation, folding, dimerization of monomers and catalytic dysfunction of HexA protein [4, 8, 11, 15, 16].

Many studies through different population demonstrate that the molecular base of *HEXA* gene still is unidentified [3, 8, 17-20]. Therefore, this study aims to recognize the possible pathogenic SNPs in *HEXA* gene using bioinformatics prediction tools to define its role of the SNPs on the structure, function, and regulation of their corresponding protein.

Recently, bioinformatics tools enabled drawing out very valuable results from large amounts of raw data like analysis of gene and protein expression, comparison of genetic data (evolution), modeling of 3D structure from sequence of gene or protein based on homologous. It also assists in prediction of damaging SNPs and its relationship with diseases [21]. Our study aimed to define the impact of *HEXA* SNPs on the structure and function of *HEXA* protein using bioinformatics softwares which may have a significant role in disease susceptibility. The usage of the bioinformatics approach has a potent impact on the identification of provider SNPs since it is easy and less costly, less time consuming and can facilitate future genetic studies. This study regarded as the first study that determine the possible damaging *HEXA* gene SNPs using bioinformatics tools which could be useful for further genetic studies.

MATERIALS AND METHODS

Data on *HEXA* gene was obtained from the national center for biological information (NCBI) website (<https://www.ncbi.nlm.nih.gov/>) and the SNPs (single nucleotide polymorphisms) information was retrieved from NCBI SNPs database dbSNP (<https://www.ncbi.nlm.nih.gov/snp/>). The gene ID and sequence was obtained from Uniprot (<https://www.uniprot.org/>). Analysis of the SNPs was done according to figure 1.

Figure (1): Diagram demonstrates *HEXA* gene work flow using different bioinformatics tools

SIFT

(Sorting Intolerant from Tolerant) 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 outcome, considering the position at which the mutation occurred and the type of amino acid change. When a protein sequence submitted, SIFT chooses related proteins and obtains an alignment of these proteins with the query. Based on the change on the type of amino acids appearing at each position in the alignment, SIFT calculates the probability that amino acid at a position is tolerated conditional on the most frequent amino acid being tolerated. If this normalized value is less than a cutoff, the substitution is predicted to be deleterious. SIFT scores <0.05 are predicted by the algorithm to be intolerant or deleterious amino acid substitutions, whereas scores >0.05 are considered tolerant [22]. (<http://sift.bii.a-star.edu.sg/>)

PolyPhen-2

Polymorphism Phenotyping v.2 available at (<http://genetics.bwh.harvard.edu/pph2/>) is a tool that predicts the possible effect of an amino acid substitution on the structure and function of a human protein by using simple physical and comparative considerations. The submission of the sequence allows querying for a single individual amino acid substitution or a coding, non-synonymous SNP annotated in the SNP database. It calculates position-specific independent count (PSIC) scores for each of the two variants and computes the difference of the PSIC scores of the two variants. The higher a PSIC score difference, the higher the functional effect of a particular amino acid substitution is likely to have. PolyPhen scores were designated as probably damaging (0.95–1), possibly damaging (0.7–0.95), and benign (0.00–0.31) [23].

Provean

Protein Variation Effect Analyzer software available at (<http://provean.jevl.org/index.php>) it predicts whether an amino acid substitution has an impact on the biological function of the protein. Provean is useful in filtering sequence variants to identify non-synonymous variants that are predicted to be functionally important [24].

SNAP2

Available at (<https://roslab.org/owiki/index.php/Snap2>), SNAP2 is a trained classifier that is based on a machine learning device called "neural network". It distinguishes between effect and neutral variants/non-synonymous SNPs by taking a variety of sequence and variant features in considerations. The most important input signal for the prediction is the evolutionary information taken from an automatically generated multiple sequence alignment. Also, structural features such as predicted secondary structure and solvent accessibility are considered. If available also annotation (i.e. known functional residues, pattern, regions) of the sequence or close homologs are pulled in. SNAP2 has persistent two-state accuracy (effect/neutral) of 82 % [25].

PHD-SNP

Prediction of human Deleterious Single Nucleotide Polymorphisms is available at (<http://snps.biofold.org/phd-snp/phd-snp.html>). The working principle is Support Vector Machines (SVMs) based method trained to predict disease-associated nsSNPs using sequence information. The related mutation is predicted as disease-related (Disease) or as neutral polymorphism (Neutral) [26].

SNP and GO

It is a server for the prediction of single point protein mutations likely to be involved in the causing of diseases in humans.[26] (<https://snps-and-go.biocomp.unibo.it/snps-and-go/>)

Pmut

Available at (<http://mmb.irbbarcelona.org/PMut/>), Pmut is based on the use of different kinds of sequence information to label mutations, and neural networks to process this information. It provides a very simple output: a yes/no answer and a reliability index [27].

IMUTANT

Imutant is a suit of support vector machine based predictors which integrated into a unique web server. It offers the opportunity to predict automatically protein stability changes upon single site mutations starting from protein sequence alone or protein structure if available. Also, it gives the possibility to predict human deleterious SNPs from the protein sequence alone. It is available at (http://gpcr.biocomp.unibo.it/~emidio/I-Mutant3.0/old/IntroI-Mutant3.0_help.html) [28].

Project Hope

Online software is available at: (<http://www.cmbi.ru.nl/hope/method/>). It is a web service where the user can submit a sequence and mutation. The software collects structural information from a series of sources, including calculations on the 3D protein structure,

sequence annotations in UniProt and prediction from other software. It combines this information to give analysis for the effect of a certain mutation on the protein structure. HOPE will show the effect of that mutation in such a way that even those without a bioinformatics background can understand it. It allows the user to submit a protein sequence (can be FASTA or not) or an accession code of the protein of interest. In the next step, the user can indicate the mutated residue with a simple mouse click. In the final step, the user can simply click on one of the other 19 amino acid types that will become the mutant residue, and then full report well is generated [29].

UCSF Chimera (University of California at San Francisco)

UCSF Chimera (<https://www.cgl.ucsf.edu/chimera/>) is a highly extensible program for interactive visualization and analysis of molecular structures and related data, including density maps, supramolecular assemblies, sequence alignments, docking results, trajectories, and conformational ensembles. High-quality images and animations can be generated. Chimera includes complete documentation and several tutorials. Chimera is developed by the Resource for Biocomputing, Visualization, and Informatics (RBVI), supported by the National Institutes of Health (P41-GM103311).[30]

PolymiRTS

PolymiRTS is a software used to predict 3UTR (un-translated region) polymorphism in microRNAs and their target sites available at (<http://compbio.uthsc.edu/miRSNP/>). It is a database of naturally occurring DNA variations in microRNAs (miRNA) seed region and miRNA target sites. MicroRNAs pair to the transcript of protein coding genes and cause translational repression or mRNA destabilization. SNPs in microRNA and their target sites may affect miRNA-mRNA interaction, causing an effect on miRNA-mediated gene repression, PolymiRTS database was created by scanning 3UTRs of mRNAs in human and mouse for SNPs in miRNA target sites. Then, the effect of polymorphism on gene expression and phenotypes are identified and then linked in the database. The PolymiRTS data base also includes polymorphism in target sites that have been supported by a variety of experimental methods and polymorphism in miRNA seed regions. [31]

GeneMANIA

It is gene interaction software that finds other genes which is related to a set of input genes using a very large set of functional association data. Association data include protein and genetic interactions, pathways, co-expression, co-localization and protein domain similarity. GeneMANIA also used to find new members of a pathway or complex, find additional genes you may have missed in your screen or find new genes with a specific function, such as protein kinases. available at (<https://genemania.org/>) [32].

RESULTS

The functional effect of Deleterious and damaging nsSNPs of *HEXA* by SIFT, PolyPhen-2, Provean, and SNAP2

Table (1): Prediction of functional effect of Deleterious and damaging nsSNPs by SIFT, Polyphen-2, PROVEAN and SNAP2

Mutation	SIFT PRE	SIFT SCO	POLYPHEN PRE	POLYPHEN SCO	PROVEAN SCO	PROVEAN PRE	SNAP2 SCO	SNAP2 PRE
G511D	D	0	PROB DAMAGE	1	-6.394	Deleterious	76	effect
G511V	D	0	PROB DAMAGE	1	-8.221	Deleterious	61	effect
R499H	D	0	PROB DAMAGE	1	-4.642	Deleterious	97	effect
R499C	D	0	PROB DAMAGE	1	-7.427	Deleterious	97	effect
A496P	D	0	PROB DAMAGE	1	-4.242	Deleterious	11	effect
W485R	D	0	PROB DAMAGE	1	-12.998	Deleterious	97	effect
E482K	D	0	PROB DAMAGE	1	-3.714	Deleterious	98	effect
P475R	D	0	PROB DAMAGE	1	-8.356	Deleterious	76	effect
W474C	D	0	PROB DAMAGE	1	-12.209	Deleterious	83	effect
R472S	D	0	PROB DAMAGE	1	-5.648	Deleterious	71	effect
E462V	D	0	PROB DAMAGE	1	-6.621	Deleterious	85	effect
G454A	D	0	PROB DAMAGE	1	-5.675	Deleterious	79	effect
G454S	D	0	PROB DAMAGE	1	-5.675	Deleterious	98	effect
G454R	D	0	PROB DAMAGE	1	-7.566	Deleterious	93	effect
P439L	D	0	PROB DAMAGE	1	-9.569	Deleterious	50	effect
W420C	D	0	PROB DAMAGE	1	-12.44	Deleterious	94	effect
S417Y	D	0	PROB DAMAGE	1	-4.879	Deleterious	76	effect
D378Y	D	0	PROB DAMAGE	1	-8.787	Deleterious	9	effect
I335N	D	0	PROB DAMAGE	1	-6.341	Deleterious	62	effect
C328Y	D	0	PROB DAMAGE	1	-10.723	Deleterious	49	effect
D322N	D	0	PROB DAMAGE	1	-4.877	Deleterious	78	effect
D322Y	D	0	PROB DAMAGE	1	-8.731	Deleterious	88	effect
G321E	D	0	PROB DAMAGE	1	-7.803	Deleterious	86	effect
G321R	D	0	PROB DAMAGE	1	-7.803	Deleterious	89	effect
L319H	D	0	PROB DAMAGE	1	-6.661	Deleterious	57	effect
H318L	D	0	PROB DAMAGE	1	-10.742	Deleterious	88	effect
Y316C	D	0	PROB DAMAGE	1	-7.45	Deleterious	48	effect
Y298C	D	0	PROB DAMAGE	1	-8.772	Deleterious	52	effect
D258H	D	0	PROB DAMAGE	1	-6.827	Deleterious	96	effect
R252H	D	0	PROB DAMAGE	1	-4.877	Deleterious	53	effect
R252C	D	0	PROB DAMAGE	1	-7.803	Deleterious	68	effect
G250V	D	0	PROB DAMAGE	1	-8.778	Deleterious	67	effect
G250S	D	0	PROB DAMAGE	1	-5.852	Deleterious	41	effect
R247W	D	0	PROB DAMAGE	1	-7.803	Deleterious	89	effect
A246T	D	0	PROB DAMAGE	1	-3.768	Deleterious	31	effect
A246S	D	0	PROB DAMAGE	1	-2.726	Deleterious	13	effect
F211S	D	0	PROB DAMAGE	1	-7.73	Deleterious	88	effect
S210F	D	0	PROB DAMAGE	1	-5.823	Deleterious	80	effect
H204R	D	0	PROB DAMAGE	1	-7.797	Deleterious	89	effect
W203G	D	0	PROB DAMAGE	1	-12.67	Deleterious	77	effect
H202R	D	0	PROB DAMAGE	1	-7.797	Deleterious	86	effect
N199S	D	0	PROB DAMAGE	1	-4.873	Deleterious	27	effect
N199I	D	0	PROB DAMAGE	1	-8.772	Deleterious	71	effect
K197T	D	0	PROB DAMAGE	1	-5.848	Deleterious	95	effect
Y180H	D	0	PROB DAMAGE	1	-4.448	Deleterious	95	effect
H179L	D	0	PROB DAMAGE	1	-10.6	Deleterious	76	effect
R178H	D	0	PROB DAMAGE	1	-4.815	Deleterious	94	effect
R178C	D	0	PROB DAMAGE	1	-7.708	Deleterious	96	effect
D175A	D	0	PROB DAMAGE	1	-7.708	Deleterious	88	effect
R170Q	D	0	PROB DAMAGE	1	-3.854	Deleterious	96	effect
R170W	D	0	PROB DAMAGE	1	-7.708	Deleterious	98	effect
R166H	D	0	PROB DAMAGE	1	-4.818	Deleterious	18	effect
R166C	D	0	PROB DAMAGE	1	-7.708	Deleterious	37	effect
I161T	D	0	PROB DAMAGE	1	-4.534	Deleterious	43	effect
R137Q	D	0	PROB DAMAGE	1	-3.654	Deleterious	72	effect
S129F	D	0	PROB DAMAGE	1	-3.307	Deleterious	48	effect
L127R	D	0	PROB DAMAGE	1	-5.481	Deleterious	95	effect
Y116F	D	0	PROB DAMAGE	1	-3.682	Deleterious	60	effect
Y68D	D	0	PROB DAMAGE	1	-8.377	Deleterious	65	effect

Functional analysis of *HEXA* gene using diseased related softwares (PhD-SNP, SNPs & GO and PMut)

Table (2): Prediction of Disease Related and pathological effect of nsSNPs by PhD-SNP, SNPs & GO and PMut:

Mutation	SNP&GPre	RI	Probability	PhD Pre	RI	Probability	P MUT Mutation	Position	Prediction
G511D	Disease	5	0.742	Disease	8	0.915	G → D (Gly → Asp)	511	0.66 (85%) Disease
G511V	Disease	6	0.8	Disease	8	0.912	G → V (Gly → Val)	511	0.87 (91%) Disease
R499H	Disease	6	0.798	Disease	8	0.883	R → H (Arg → His)	499	0.86 (91%) Disease
R499C	Disease	6	0.807	Disease	8	0.911	R → C (Arg → Cys)	499	0.92 (94%) Disease
A496P	Disease	5	0.74	Disease	8	0.882	A → P (Ala → Pro)	496	0.82 (90%) Disease
W485R	Disease	7	0.834	Disease	8	0.914	W → R (Trp → Arg)	485	0.93 (94%) Disease
E482K	Disease	6	0.817	Disease	7	0.864	E → K (Glu → Lys)	482	0.88 (92%) Disease
P475R	Disease	6	0.8	Disease	8	0.889	P → R (Pro → Arg)	475	0.87 (91%) Disease
W474C	Disease	5	0.753	Disease	8	0.878	W → C (Trp → Cys)	474	0.92 (94%) Disease
R472S	Disease	5	0.763	Disease	6	0.825	R → S (Arg → Ser)	472	0.82 (90%) Disease
E462V	Disease	8	0.885	Disease	9	0.943	E → V (Glu → Val)	462	0.93 (94%) Disease
G454A	Disease	7	0.834	Disease	7	0.843	G → A (Gly → Ala)	454	0.88 (92%) Disease
G454S	Disease	7	0.842	Disease	8	0.893	G → S (Gly → Ser)	454	0.92 (94%) Disease
G454R	Disease	7	0.869	Disease	8	0.917	G → R (Gly → Arg)	454	0.92 (94%) Disease
P439L	Disease	3	0.656	Disease	7	0.858	P → L (Pro → Leu)	439	0.78 (88%) Disease
W420C	Disease	3	0.669	Disease	6	0.814	W → C (Trp → Cys)	420	0.85 (91%) Disease
S417Y	Disease	4	0.718	Disease	7	0.873	S → Y (Ser → Tyr)	417	0.68 (85%) Disease
D378Y	Disease	5	0.759	Disease	6	0.813	D → Y (Asp → Tyr)	378	0.75 (87%) Disease
I335N	Disease	7	0.845	Disease	8	0.877	I → N (Ile → Asn)	335	0.90 (93%) Disease
C328Y	Disease	2	0.615	Disease	6	0.817	C → Y (Cys → Tyr)	328	0.68 (85%) Disease
							D → N (Asp → Asn)	322	0.89 (92%) Disease
D322N	Disease	7	0.874	Disease	9	0.934		322	0.93 (94%) Disease
D322Y	Disease	8	0.877	Disease	9	0.971	D → Y (Asp → Tyr)	322	0.93 (94%) Disease
G321E	Disease	8	0.878	Disease	9	0.932	G → E (Gly → Glu)	321	0.93 (94%) Disease
G321R	Disease	8	0.888	Disease	9	0.954	G → R (Gly → Arg)	321	0.92 (94%) Disease
L319H	Disease	5	0.735	Disease	6	0.788	L → H (Leu → His)	319	0.93 (94%) Disease
H318L	Disease	7	0.85	Disease	9	0.929	H → L (His → Leu)	318	0.92 (94%) Disease
Y316C	Disease	4	0.708	Disease	7	0.858	Y → C (Tyr → Cys)	316	0.92 (94%) Disease
Y298C	Disease	3	0.629	Disease	7	0.856	Y → C (Tyr → Cys)	298	0.93 (94%) Disease
D258H	Disease	6	0.812	Disease	8	0.894	D → H (Asp → His)	258	0.92 (94%) Disease
R252H	Disease	1	0.57	Disease	4	0.718	R → H (Arg → His)	252	0.68 (85%) Disease
R252C	Disease	6	0.776	Disease	8	0.878	R → C (Arg → Cys)	252	0.88 (92%) Disease
G250V	Disease	6	0.808	Disease	8	0.877	G → V (Gly → Val)	250	0.93 (94%) Disease
G250S	Disease	3	0.672	Disease	6	0.794	G → S (Gly → Ser)	250	0.68 (85%) Disease
R247W	Disease	5	0.737	Disease	8	0.884	R → W (Arg → Trp)	247	0.80 (89%) Disease
A246T	Disease	5	0.775	Disease	7	0.863	A → T (Ala → Thr)	246	0.92 (94%) Disease
A246S	Disease	5	0.76	Disease	8	0.885	A → S (Ala → Ser)	246	0.87 (92%) Disease
F211S	Disease	7	0.828	Disease	7	0.832	F → S (Phe → Ser)	211	0.92 (94%) Disease
S210F	Disease	6	0.811	Disease	8	0.878	S → F (Ser → Phe)	210	0.88 (92%) Disease
H204R	Disease	8	0.876	Disease	9	0.941	H → R (His → Arg)	204	0.91 (93%) Disease
W203G	Disease	7	0.826	Disease	8	0.893	W → G (Trp → Gly)	203	0.93 (94%) Disease
H202R	Disease	8	0.888	Disease	9	0.946	H → R (His → Arg)	202	0.93 (94%) Disease
N199S	Disease	7	0.832	Disease	8	0.919	N → S (Asn → Ser)	199	0.89 (92%) Disease
N199I	Disease	8	0.91	Disease	9	0.971	N → I (Asn → Ile)	199	0.93 (94%) Disease
K197T	Disease	8	0.883	Disease	9	0.942	K → T (Lys → Thr)	197	0.93 (94%) Disease
Y180H	Disease	2	0.619	Disease	4	0.698	Y → H (Tyr → His)	180	0.81 (89%) Disease
H179L	Disease	2	0.602	Disease	6	0.818	H → L (His → Leu)	179	0.93 (94%) Disease
R178H	Disease	7	0.855	Disease	7	0.866	R → H (Arg → His)	178	0.92 (94%) Disease
R178C	Disease	6	0.815	Disease	7	0.86	R → C (Arg → Cys)	178	0.93 (94%) Disease
D175A	Disease	7	0.826	Disease	6	0.822	D → A (Asp → Ala)	175	0.92 (94%) Disease
R170Q	Disease	7	0.851	Disease	8	0.891	R → Q (Arg → Gln)	170	0.92 (94%) Disease
R170W	Disease	7	0.855	Disease	8	0.918	R → W (Arg → Trp)	170	0.93 (94%) Disease
R166H	Disease	4	0.682	Disease	6	0.809	R → H (Arg → His)	166	0.79 (89%) Disease
R166C	Disease	5	0.73	Disease	8	0.879	R → C (Arg → Cys)	166	0.81 (89%) Disease
I161T	Disease	7	0.87	Disease	8	0.889	I → T (Ile → Thr)	161	0.89 (92%) Disease
R137Q	Disease	4	0.688	Disease	2	0.604	R → Q (Arg → Gln)	137	0.80 (89%) Disease
S129F	Disease	5	0.75	Disease	6	0.801	S → F (Ser → Phe)	129	0.52 (79%) Disease
L127R	Disease	2	0.583	Disease	8	0.904	L → R (Leu → Arg)	127	0.92 (94%) Disease
Y116F	Disease	5	0.736	Disease	7	0.827	Y → F (Tyr → Phe)	116	0.88 (92%) Disease
Y68D	Disease	0	0.508	Disease	7	0.848	Y → D (Tyr → Asp)	68	0.91 (93%) Disease

Prediction of Change in *HEXA* Stability due to Mutation Using I-Mutant 3.0 Server

Table (3): Prediction of nsSNPs Impact on Protein structure Stability by I-Mutant

dbSNPs#	Amino Acid Change	I-Mutant PRE	I Mutant SCO
rs1448744870	G511D	Decrease	-0.77
rs1448744870	G511V	Decrease	-0.35
rs121907956	R499H	Decrease	-1.27
rs121907966	R499C	Decrease	-0.82
rs891448877	A496P	Increase	-0.24
rs121907968	W485R	Decrease	-1.22
rs121907952	E482K	Decrease	-0.89
rs758829206	P475R	Decrease	-1.02
rs121907981	W474C	Decrease	-1.21
rs369117208	R472S	Decrease	-1.36
rs863225434	E462V	Increase	0.2
rs1229811721	G454A	Decrease	-0.26
rs121907978	G454S	Decrease	-0.97
rs121907978	G454R	Decrease	-0.14
rs1438534723	P439L	Decrease	-0.11
rs121907958	W420C	Decrease	-1.72
rs1321597225	S417Y	Increase	-0.05
rs772807186	D378Y	Increase	-0.03
rs1431312050	I335N	Decrease	-1.68
rs761861041	C328Y	Increase	-0.23
rs772180415	D322N	Decrease	-0.53
rs772180415	D322Y	Increase	0.03
rs1316178162	G321E	Decrease	-0.53
rs556877212	G321R	Decrease	-0.41
rs1265331930	L319H	Decrease	-2.32
rs766622566	H318L	Increase	0.38
rs1222599742	Y316C	Decrease	-1.15
rs1396944040	Y298C	Decrease	-1.1
rs121907971	D258H	Decrease	-0.55
rs762255098	R252H	Decrease	-1.42
rs566580738	R252C	Decrease	-1.07
rs121907959	G250V	Decrease	-0.55
rs1057521137	G250S	Decrease	-1.3
rs121907970	R247W	Decrease	-0.49
rs758166013	A246T	Decrease	-0.69
rs758166013	A246S	Decrease	-0.87
rs121907974	F211S	Decrease	-1.64
rs121907961	S210F	Increase	0.15
rs121907976	H204R	Decrease	-0.22
rs1002712424	W203G	Decrease	-2.34
rs761536872	H202R	Increase	-0.14
rs1345279330	N199S	Decrease	-0.55
rs1345279330	N199I	Increase	0.77
rs121907973	K197T	Decrease	-0.38
rs28941771	Y180H	Decrease	-1.1
rs747372270	H179L	Increase	-0.01
rs28941770	R178H	Decrease	-1.1
rs121907953	R178C	Decrease	-0.86
rs1057519460	D175A	Decrease	-0.82
rs121907957	R170Q	Decrease	-1.07
rs121907972	R170W	Decrease	-0.44
rs398123442	R166H	Decrease	-1.43
rs759487463	R166C	Decrease	-1.08
rs765136263	I161T	Decrease	-2.07
rs769724246	R137Q	Decrease	-0.64
rs1295567201	S129F	Increase	0.32
rs121907975	L127R	Decrease	-1.46
rs1282031681	Y116F	Decrease	-0.37
rs760202948	Y68D	Decrease	-1.11

Modeling of amino acid substitution effects on *HEXA* protein structure using Chimera and Project Hope

Figure 2: rs1335478824 (E462V): Effect of change in the amino acid position 462 from to Glutamic Acid to Valine on the *HEXA* protein 3D structure using Project Hope and Chimera Softwares.

Figure 3: rs121907970 (R247W): Effect of change in the amino acid position 247 from Arginine to Tryptophan on the *HEXA* protein 3D structure using Project Hope and Chimera Softwares.

Figure 4: rs121907972 (R170W): Effect of change in the amino acid position 170 from Arginine to Tryptophan on the *HEXA* protein 3D structure using Project Hope and Chimera Software.

Figure 4: rs28941770 (R178W): Effect of change in the amino acid position 178 from Arginine to Tryptophan on the *HEXA* protein 3D structure using Project Hope and Chimera Software.

Interactions of *HEXA* gene with other Functional Genes illustrated by GeneMANIA

Figure 5: *HEXA* Gene Interactions and network predicted by GeneMania.

Table (4): The *HEXA* gene functions and its appearance in network and genome

Function	FDR	Genes in network	Genes in genome
vacuolar part	3.17E-06	8	262
glycosphingolipid metabolic process	2.79E-05	5	55
lysosomal lumen	4.47E-05	5	67
vacuolar lumen	4.47E-05	5	69
glycolipid metabolic process	0.000146	5	91
sphingolipid metabolic process	0.000216	5	102
keratan sulfate catabolic process	0.00065	3	12
membrane lipid metabolic process	0.00079	5	140
keratan sulfate metabolic process	0.011247	3	32
sulfur compound catabolic process	0.014531	3	36
glycosaminoglycan catabolic process	0.042837	3	53
aminoglycan catabolic process	0.043915	3	55

Table (5): The gene co-expressed, share domain and Interaction with *HEXA* gene network

Gene 1	Gene 2	Weight	Network group	Network
CTSA	GAA	0.012222	Co-expression	Ramaswamy-Golub-2001
P2RX4	HEXA	0.019946	Co-expression	Ramaswamy-Golub-2001
SUSD1	HEXA	0.025539	Co-expression	Ramaswamy-Golub-2001
CTSA	HEXB	0.006801	Co-expression	Wang-Maris-2006
P2RX4	HEXA	0.023653	Co-expression	Wang-Maris-2006
CLCN7	GAA	0.017076	Co-expression	Wang-Maris-2006
CLCN7	CTSA	0.01052	Co-expression	Wang-Maris-2006
HEXB	HEXA	0.02151	Co-expression	Mallon-McKay-2013
GALNS	HEXA	0.010783	Co-expression	Mallon-McKay-2013
SUSD1	HEXA	0.018532	Co-expression	Mallon-McKay-2013
SUSD1	P2RX4	0.016401	Co-expression	Mallon-McKay-2013
CLCN7	HEXA	0.011458	Co-expression	Mallon-McKay-2013
CLCN7	HEXDC	0.007927	Co-expression	Mallon-McKay-2013
CLCN7	GAA	0.010732	Co-expression	Bild-Nevins-2006 B
CLCN7	CTSA	0.008945	Co-expression	Bild-Nevins-2006 B
PPIB	HEXA	0.015699	Co-expression	Bild-Nevins-2006 B
FAN1	HEXA	0.026124	Co-expression	Burington-Shaughnessy-2008
CTSA	GAA	0.012194	Co-expression	Dobbin-Giordano-2005
AGA	HEXB	0.0211	Co-expression	Dobbin-Giordano-2005
GALNS	GM2A	0.027189	Co-expression	Bahr-Bowler-2013
CTSA	HEXA	0.008463	Co-expression	Bahr-Bowler-2013
P2RX4	HEXA	0.008105	Co-expression	Bahr-Bowler-2013
P2RX4	CTSA	0.011332	Co-expression	Bahr-Bowler-2013
NAGA	GAA	0.006624	Co-expression	Bahr-Bowler-2013
NAGA	SUSD1	0.005711	Co-expression	Bahr-Bowler-2013
TOR1A	HEXA	0.009849	Co-expression	Bahr-Bowler-2013
TOR1A	P2RX4	0.012126	Co-expression	Bahr-Bowler-2013
AGA	STIM2	0.031346	Co-expression	Alizadeh-Staudt-2000
HIST1H2BO	GALNS	0.006821	Co-expression	Innocenti-Brown-2011
CTSA	HEXA	0.025624	Co-expression	Innocenti-Brown-2011
CTSA	GALNS	0.010198	Co-expression	Innocenti-Brown-2011
CLCN7	HEXDC	0.004031	Co-expression	Innocenti-Brown-2011
PPIB	P2RX4	0.005575	Co-expression	Innocenti-Brown-2011
PPIB	AGA	0.007841	Co-expression	Innocenti-Brown-2011
AGA	P2RX4	0.02524	Co-expression	Rieger-Chu-2004
TOR1A	HEXA	0.012116	Co-expression	Rieger-Chu-2004
GAA	HEXDC	0.021471	Co-expression	Noble-Diehl-2008
P2RX4	CTSA	0.008663	Co-expression	Noble-Diehl-2008
NAGA	HEXB	0.009936	Co-expression	Noble-Diehl-2008
CLCN7	HEXDC	0.015451	Co-expression	Noble-Diehl-2008
PPIB	HEXA	0.011401	Co-expression	Noble-Diehl-2008
PPIB	HEXB	0.004546	Co-expression	Roth-Zlotnik-2006
AGA	HEXA	0.02231	Co-expression	Boldrick-Relman-2002
GAA	GM2A	0.010571	Co-expression	Perou-Botstein-2000
TOR1A	GM2A	0.023196	Co-expression	Smirnov-Cheung-2009
TOR1A	HEXB	0.016741	Co-expression	Smirnov-Cheung-2009
HEXB	GM2A	0.012785	Co-expression	Wang-Cheung-2015
CTSA	HEXA	0.012801	Co-expression	Wang-Cheung-2015
P2RX4	HEXB	0.007097	Co-expression	Wang-Cheung-2015
NAGA	HEXB	0.011548	Co-expression	Wang-Cheung-2015
NAGA	P2RX4	0.013862	Co-expression	Wang-Cheung-2015
STX2	P2RX4	0.012284	Co-expression	Wang-Cheung-2015
GAA	HEXA	0.013943	Co-expression	Chen-Brown-2002
CLCN7	GAA	0.010428	Co-expression	Chen-Brown-2002
PPIB	HEXA	0.016275	Co-expression	Chen-Brown-2002
PPIB	TOR1A	0.014532	Co-expression	Wu-Garvey-2007
AGA	STIM2	0.044308	Co-expression	Rosenwald-Staudt-2001
GALNS	HEXA	0.024879	Co-localization	Schadt-Shoemaker-2004
GAA	HEXA	0.015767	Co-localization	Schadt-Shoemaker-2004
GAA	GM2A	0.021491	Co-localization	Schadt-Shoemaker-2004
CTSA	HEXA	0.01313	Co-localization	Schadt-Shoemaker-2004
CTSA	GALNS	0.01848	Co-localization	Schadt-Shoemaker-2004

CSNK1D	HEXA	0.020417	Co-localization	Schadt-Shoemaker-2004
AGA	HEXA	0.019586	Co-localization	Schadt-Shoemaker-2004
AGA	GM2A	0.014327	Co-localization	Schadt-Shoemaker-2004
AGA	GALNS	0.017438	Co-localization	Schadt-Shoemaker-2004
AGA	GAA	0.0216	Co-localization	Schadt-Shoemaker-2004
AGA	CTSA	0.009104	Co-localization	Schadt-Shoemaker-2004
NAGA	HEXA	0.022976	Co-localization	Schadt-Shoemaker-2004
NAGA	AGA	0.018217	Co-localization	Schadt-Shoemaker-2004
CLCN7	HEXA	0.02073	Co-localization	Schadt-Shoemaker-2004
CLCN7	CTSA	0.011321	Co-localization	Schadt-Shoemaker-2004
CLCN7	AGA	0.023337	Co-localization	Schadt-Shoemaker-2004
TOR1A	HEXA	0.018053	Co-localization	Schadt-Shoemaker-2004
TOR1A	CLCN7	0.018929	Co-localization	Schadt-Shoemaker-2004
STX2	HEXA	0.025706	Co-localization	Schadt-Shoemaker-2004
STX2	TOR1A	0.017325	Co-localization	Schadt-Shoemaker-2004
PPIB	HEXA	0.009866	Co-localization	Schadt-Shoemaker-2004
PPIB	GALNS	0.010641	Co-localization	Schadt-Shoemaker-2004
PPIB	CTSA	0.006791	Co-localization	Schadt-Shoemaker-2004
PPIB	NAGA	0.010904	Co-localization	Schadt-Shoemaker-2004
HEXB	HEXA	0.00598	Co-localization	Johnson-Shoemaker-2003
GALNS	HEXA	0.015838	Co-localization	Johnson-Shoemaker-2003
FAN1	KTN1	0.011546	Co-localization	Johnson-Shoemaker-2003
GAA	HEXA	0.014948	Co-localization	Johnson-Shoemaker-2003
GAA	HEXB	0.012258	Co-localization	Johnson-Shoemaker-2003
GAA	GALNS	0.029843	Co-localization	Johnson-Shoemaker-2003
P2RX4	HEXB	0.009507	Co-localization	Johnson-Shoemaker-2003
CSNK1D	HEXA	0.01438	Co-localization	Johnson-Shoemaker-2003
NAGA	HEXB	0.009874	Co-localization	Johnson-Shoemaker-2003
PPIB	HEXB	0.005274	Co-localization	Johnson-Shoemaker-2003
PPIB	P2RX4	0.010852	Co-localization	Johnson-Shoemaker-2003
KTN1	HEXA	0.138413	Physical Interactions	Kristensen-Foster-2012
HIST1H2BO	HEXA	0.123828	Physical Interactions	Kristensen-Foster-2012
STIM2	HEXA	0.796225	Physical Interactions	Wan-Emili-2015
STIM2	HEXB	0.35232	Physical Interactions	Wan-Emili-2015
FAN1	HEXA	0.11305	Physical Interactions	Hein-Mann-2015
CLCN7	KTN1	0.020924	Physical Interactions	Hein-Mann-2015
BIOGRID-SMALL-SCALE-STUDIES				
GM2A	HEXA	1	Physical Interactions	IREF-HPRD
GM2A	HEXA	0.777151	Physical Interactions	Wu-Stein-2010
GM2A	HEXA	0.414214	Predicted	Wu-Stein-2010
HEXB	HEXA	0.765367	Predicted	INTERPRO
HEXB	HEXA	0.142195	Shared protein domains	INTERPRO
RP11-106M3.2	HEXA	0.110406	Shared protein domains	INTERPRO
RP11-106M3.2	HEXB	0.110406	Shared protein domains	INTERPRO
HEXDC	HEXA	0.050149	Shared protein domains	INTERPRO
HEXDC	HEXB	0.050149	Shared protein domains	INTERPRO
HEXDC	RP11-106M3.2	0.051545	Shared protein domains	INTERPRO
GAA	HEXA	0.009137	Shared protein domains	INTERPRO
GAA	HEXB	0.009137	Shared protein domains	INTERPRO
GAA	RP11-106M3.2	0.009392	Shared protein domains	INTERPRO
GAA	HEXDC	0.01046	Shared protein domains	INTERPRO
NAGA	HEXA	0.016292	Shared protein domains	INTERPRO
NAGA	HEXB	0.016292	Shared protein domains	INTERPRO
NAGA	RP11-106M3.2	0.016746	Shared protein domains	INTERPRO
NAGA	HEXDC	0.018651	Shared protein domains	INTERPRO
NAGA	GAA	0.010607	Shared protein domains	INTERPRO
HEXB	HEXA	0.379855	Shared protein domains	PFAM
RP11-106M3.2	HEX	0.379855	Shared protein domains	PFAM
RP11-106M3.2	HEXB	0.379855	Shared protein domains	PFAM
HEXDC	HEXA	0.283013	Shared protein domains	PFAM
HEXDC	HEXB	0.283013	Shared protein domains	PFAM
HEXDC	RP11-106M3.2	0.283013	Shared protein domains	PFAM

SNPs effect on 3'UTR Region (miRNA binding sites) using PolymiRTS Database

Table (6): prediction of SNPs at the 3'UTR Region using PolymiRTS

Location	dbSNP ID	Variant type	Wobble base pair	Ancestral Allele	Allele	miR ID	Conservation	miR Site	Function Class	Exp Support	context+ score change
72635788	rs35949555	SNP	Y	A	A	hsa-miR-338-5p	3	acacAT A TTGTtc	D	N	-0.099
						hsa-miR-511-3p	4	ACACAT A ttgttc	D	N	-0.158
72635829	rs11629508	SNP	N	T	T	hsa-miR-127-5p	2	gctGCT T CAAttt	D	N	-0.153
					G	hsa-miR-7843-3p	2	gctGCT T CAAttt	D	N	-0.155
						hsa-miR-103a-3p	2	gcTGCT G CAattt	C	N	-0.198
						hsa-miR-107	2	gcTGCT G CAattt	C	N	-0.198
						hsa-miR-1301-3p	2	gctGCT G CAAttt	C	N	-0.202
						hsa-miR-5047	2	gctGCT G CAAttt	C	N	-0.195
72635843	rs76075374	SNP	Y	G	G	hsa-miR-186-3p	4	cTTT G Gtttct	D	N	-0.177
					A	hsa-miR-888-5p	4	ccTT TG AGTttct	C	N	-0.08
72635869	rs60288568	SNP	N	C	C	hsa-miR-149-3p	6	ccacct C CCTCCC	D	N	-0.225
						hsa-miR-4419a	2	ccac CT CCCTCcc	D	N	-0.146
						hsa-miR-4510	2	ccac CT CCCTCcc	D	N	-0.184
						hsa-miR-4728-5p	6	ccacct C CCTCCC	D	N	-0.225
						hsa-miR-6127	2	ccac CT CCCTCcc	D	N	-0.165
						hsa-miR-6129	2	ccac CT CCCTCcc	D	N	-0.165
						hsa-miR-6130	2	ccac CT CCCTCcc	D	N	-0.174
						hsa-miR-6133	2	ccac CT CCCTCcc	D	N	-0.146
						hsa-miR-6785-5p	7	ccacct C CCTCCC	D	N	-0.3
						hsa-miR-6797-5p	5	ccacc T CCTCCc	D	N	-0.172
					G	hsa-miR-6883-5p	7	ccacct C CCTCCC	D	N	-0.235
						hsa-miR-1827	2	ccac CT GCCTCcc	C	N	-0.149
						hsa-miR-3612	4	ccacc TG CCTCCc	C	N	-0.176
72635903	rs3087652	SNP	Y	G	G	hsa-miR-650	5	ccacc TG CCTCCc	C	N	-0.139
						hsa-miR-4677-3p	3	gTCACAG A agagc	D	N	-0.066
					A	hsa-miR-4679	3	gTCACAG A agagc	D	N	-0.072
						hsa-miR-2113	4	gtCACAA A agagc	C	N	0.008
						hsa-miR-5010-3p	4	gtCACAA A agagc	C	N	0.015
						hsa-miR-6808-3p	3	GTCACAA A agagc	C	N	-0.199
72635982	rs112626309	SNP	N	C	C	hsa-miR-758-3p	2	GTCACAA A agagc	C	N	-0.225
					T	hsa-miR-614	2	ccaAGG C GTTggt	D	N	-0.269
72636051	rs184059814	SNP	N	T	T	hsa-miR-6813-3p	5	CCAAGG T gttggt	C	N	-0.225
						hsa-miR-4433-3p	2	cACTC T GAccat	D	N	-0.219
						hsa-miR-4433b-3p	2	CACTC T gaccat	D	N	-0.124
						hsa-miR-4459	2	caCTC T GAccat	D	N	-0.065
					C	hsa-miR-4768-3p	2	caCTC T GAccat	D	N	-0.061
						hsa-miR-6746-5p	2	caCTC CC GAccat	C	N	-0.17
72636063	rs188399769	SNP	N	C	C	hsa-miR-6771-5p	2	caCTC CC GAccat	C	N	-0.243
						hsa-miR-149-3p	2	tcaact C CCTCCC	D	N	-0.213
						hsa-miR-4728-5p	2	tcaact C CCTCCC	D	N	-0.22
						hsa-miR-6785-5p	2	tcaact C CCTCCC	D	N	-0.22
72636100	rs139968433	SNP	N	T	G	hsa-miR-6883-5p	2	tcaact C CCTCCC	D	N	-0.22
						hsa-miR-4438	4	tAGC C TGTgtgct	C	N	-0.167
72636235	rs1054432	SNP	N	C	T	hsa-miR-6768-5p	4	tagC C TGTGTgct	C	N	-0.107
						hsa-miR-3163	5	attt T TATAAat	C	N	0.054
72636339	rs181233715	SNP	N	T	T	hsa-miR-340-5p	5	atttt T TATAAat	C	N	0.015
						hsa-miR-6729-3p	4	aGGGG A TGgagc	D	N	-0.181

DISCUSSION

In this study, we revealed 32 novel nsSNP in *HEXA* gene (mentioned in Table (7) out of 59 most damaging missense mutations (in coding region of gene) by using different softwares. We will discuss some of the SNPs that revealed to be most damaging SNPs by this computational analysis study (E462V, R247W, R178H, and R170W).

E462V (1385A>T) mutation was identified in different Indian populations and mentioned to be one of the most commonly reported missense mutations in Indian TSD patient [15, 33, 34]. It was identified as a founder mutation in about 21-40% of TSD patients from Gujarat in India [35-37]. Depending on previous study findings, patients originating from Gujarat state could be screened for E462V mutation [8, 38]. E462V firstly described in Mistri et al. (2012) study as a novel mutation found in six unrelated families from Gujarat indicating a founder effect. They found that it is occurring at the functionally active site of the alpha subunit of hexosaminidase A. In their analysis they found that E462 is a highly conserved region. This mutation was present in the homozygous state in all six patients exhibiting infantile TSD. This is in accordance with previous observations that missense mutations responsible for infantile TSD were generally located in a functional importance region, such as the active site. Also, they recommend screening of E462V SNP in TSD patients from Gujarat [8] and that is in agreement with this study finding in that this SNP is predicted to be a damaging mutation.

R247W (739C>T) accompanied by R249W (745C>T) regarded in Cao, Z., et al.(1997) study as benign mutations that reduce the alpha subunit of protein through affecting its stability in vivo without affecting its processing i.e. phosphorylation, targeting or secretion. Studies also demonstrated that these benign mutations could be readily differentiated from disease-causing mutations using a transient expression system[39].

R247W(739C>T) together with R249W (745C>T) occur in 2% of Ashkenazi Jewish and in 35% in non-Jews and considered to be pseudodeficient mutations, because of that the two common variations of the *HEXA* gene lead to reduced activity against the artificial MUG (4-methylumbelliferyl-2-acetamido-2-deoxy-D-glucopyranoside) substrate but do not impair the ability to degrade the physiological substrate GM2,sothey are associated with no clinical phenotypes [8, 15, 40-44]. These results disagree with this study which record that these SNPs are causing disease and effect on protein structure, function and stability by using different tools to prove that.

R170W (508C>T) mutation identified in previous studies to be associated with an infantile form of TSD [19, 34, 45-47]. It shows complete inactivation of Hex A by the consistent with the infantile TSD phenotype. In areported case, the R170W substitution is associated with the expression of a subunit precursor while no association with its maturation and targeting to the lysosome [46]. R170W was one of the amino acid substitutions which have been reported with a study in the alpha subunit of Hex A. The dysfunctional and destabilizing defects in Hex α - reflects biochemical and phenotypic abnormalities in TSD [47]. In another study, R170W identified as a known pathogenic mutation which disrupts the beta sheet. This mutation has been previously reported in different ethnic groups [Japanese, French-Canadian (Estreeregion, Quebec) and Italian patients]. The side chain of R170 in domain II forms a hydrogen bond with the E141 in domain I. The substitution of R with W with a bulkier side chain has a large effect on the interface of domains I and II. This would destabilize the

domain interface and cause degradation of the α -subunit[8].So they agree with this study result in that R170W SNP consider as harmful mutation.

R178H(533G>A) is the most reported mutation responsible for the TSD B1 variant[45, 48-53]. It is mainly found in a higher frequency in Portugal [45, 48-50, 52, 54, 55]. It found in *HEXA* gene and affect the active site residue in the α subunit of Hex A enzyme, affecting its structure and subsequently its function by inactivating it leading to Tay-Sachs disease [45, 47, 49, 51, 54-56]. And that is confirming this study outcome to be a damaging SNP. Using I-Mutant software, 12SNPs were predicted to increase the stability of protein while 47 SNPs were predicted to decrease it.

Out of the 154 SNPs in 3'UTR, 11SNPs were found to have an effect on it. 48 functional classes were predicted among 11 SNPs; 28 alleles disrupted a conserved miRNA site and 20 derived alleles creating a new site of miRNA and this might result in deregulation of the gene function. GeneMANIA revealed *HEXA* gene functions and activities such which they are following: aminoglycan catabolic process, glycolipid metabolic process, glycosaminoglycan catabolic process, glycosphingolipid metabolic process, keratan sulfate catabolic process, keratan sulfate metabolic process, lysosomal lumen, membrane lipid metabolic process, sphingolipid metabolic process, sulfur compound catabolic process, vacuolar lumen, and vacuolar part. As bioinformatics softwares play an important role, invivo/invitro analysis remains highly recommended confirming these present study findings. These findings can be used in a helpful way to improve the diagnosis of TaySaches Disease.

CONCLUSION

This study revealed 32 novel nsSNPout of 59 damaging SNPs in the *HEXA* gene that leads to TaySaches disease, by using different bioinformatics tools. Also, 48 functional classes were predicted in11 SNPs in the 3'UTR, among the 21 SNPs, 28 alleles disrupted a conserved miRNA site and 20 derived alleles created a new site of miRNA. This might result in the de-regulation of the gene function. Hopefully, these results will help in genetic studying and diagnosis of TaySaches disease improvement.

ACKNOWLEDGMENT

The authors desire to acknowledge the exciting cooperation of Africa City of Technology – Sudan.

DATA AVAILABILITY

All relevant data used to support the findings of this study are included within the manuscript and supplementary information files.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest regarding the publication of this paper.

REFERENCES

1. Mehta, N., et al., *Tay-Sachs Carrier Screening by Enzyme and Molecular Analyses in the New York City Minority Population*. Genet Test Mol Biomarkers, 2016. **20**(9): p. 504-9.
2. Stepien, K.M., et al., *Haematopoietic Stem Cell Transplantation Arrests the Progression of Neurodegenerative Disease in Late-Onset Tay-Sachs Disease*. JIMD Rep, 2018. **41**: p. 17-23.
3. Hoffman, J.D., et al., *Next-generation DNA sequencing of HEXA: a step in the right direction for carrier screening*. Mol Genet Genomic Med, 2013. **1**(4): p. 260-8.
4. Lew, R., L. Burnett, and A. Proos, *Tay-Sachs disease preconception screening in Australia: self-knowledge of being an Ashkenazi Jew predicts carrier state better than does ancestral origin, although there is an increased risk for c.1421 + 1G > C mutation in individuals with South African heritage*. J Community Genet, 2011. **2**(4): p. 201-9.
5. Kelly, T.E., et al., *Tay-Sachs disease: high gene frequency in a non-Jewish population*. Am J Hum Genet, 1975. **27**(3): p. 287-91.
6. Solovyeva, V.V., et al., *New Approaches to Tay-Sachs Disease Therapy*. Front Physiol, 2018. **9**: p. 1663.
7. Zokaeem, G., et al., *A shortened beta-hexosaminidase alpha-chain in an Italian patient with infantile Tay-Sachs disease*. Am J Hum Genet, 1987. **40**(6): p. 537-47.
8. Mistri, M., et al., *Identification of novel mutations in HEXA gene in children affected with Tay Sachs disease from India*. PLoS One, 2012. **7**(6): p. e39122.
9. Triggs-Raine, B.L., et al., *Sequence of DNA flanking the exons of the HEXA gene, and identification of mutations in Tay-Sachs disease*. Am J Hum Genet, 1991. **49**(5): p. 1041-54.
10. Karumuthil-Meethil, S., et al., *Novel Vector Design and Hexosaminidase Variant Enabling Self-Complementary Adeno-Associated Virus for the Treatment of Tay-Sachs Disease*. Hum Gene Ther, 2016. **27**(7): p. 509-21.
11. Sheth, J., et al., *Identification of deletion-duplication in HEXA gene in five children with Tay-Sachs disease from India*. BMC Med Genet, 2018. **19**(1): p. 109.
12. Karimzadeh, P., et al., *GM2-Gangliosidosis (Sandhoff and Tay Sachs disease): Diagnosis and Neuroimaging Findings (An Iranian Pediatric Case Series)*. Iran J Child Neurol, 2014. **8**(3): p. 55-60.
13. Vu, M., et al., *Neural stem cells for disease modeling and evaluation of therapeutics for Tay-Sachs disease*. Orphanet J Rare Dis, 2018. **13**(1): p. 152.
14. Colaianne, A., S. Chandrasekharan, and R. Cook-Deegan, *Impact of gene patents and licensing practices on access to genetic testing and carrier screening for Tay-Sachs and Canavan disease*. Genet Med, 2010. **12**(4 Suppl): p. S5-S14.
15. Sheth, J., et al., *Expanding the spectrum of HEXA mutations in Indian patients with Tay-Sachs disease*. Mol Genet Metab Rep, 2014. **1**: p. 425-430.
16. Korneluk, R.G., et al., *Isolation of cDNA clones coding for the alpha-subunit of human beta-hexosaminidase. Extensive homology between the alpha- and beta-subunits and studies on Tay-Sachs disease*. J Biol Chem, 1986. **261**(18): p. 8407-13.
17. Jamali, S., et al., *Three novel mutations in Iranian patients with Tay-Sachs disease*. Iran Biomed J, 2014. **18**(2): p. 114-9.
18. Lew, R.M., et al., *Tay-Sachs disease: current perspectives from Australia*. Appl Clin Genet, 2015. **8**: p. 19-25.
19. Martin, D.C., et al., *Evaluation of the risk for Tay-Sachs disease in individuals of French Canadian ancestry living in new England*. Clin Chem, 2007. **53**(3): p. 392-8.
20. Zhang, J., et al., *Prenatal Diagnosis of Tay-Sachs Disease*. Methods Mol Biol, 2019. **1885**: p. 233-250.
21. Elbager, S., et al., *Computational Analysis of Deleterious Single Nucleotide Polymorphisms (SNPs) in Human CALR Gene*. 2018.
22. Sim, N.L., et al., *SIFT web server: predicting effects of amino acid substitutions on proteins*. Nucleic Acids Res, 2012. **40**(Web Server issue): p. W452-7.

23. Adzhubei, I., D.M. Jordan, and S.R. Sunyaev, *Predicting functional effect of human missense mutations using PolyPhen-2*. Curr Protoc Hum Genet, 2013. **Chapter 7**: p. Unit7 20.
24. Choi, Y. and A.P. Chan, *PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels*. Bioinformatics, 2015. **31**(16): p. 2745-7.
25. Bromberg, Y. and B. Rost, *SNAP: predict effect of non-synonymous polymorphisms on function*. Nucleic Acids Res, 2007. **35**(11): p. 3823-35.
26. Capriotti, E., et al., *WS-SNPs&GO: a web server for predicting the deleterious effect of human protein variants using functional annotation*. BMC Genomics, 2013. **14 Suppl 3**: p. S6.
27. Lopez-Ferrando, V., et al., *PMut: a web-based tool for the annotation of pathological variants on proteins, 2017 update*. Nucleic Acids Res, 2017. **45**(W1): p. W222-W228.
28. Capriotti, E., P. Fariselli, and R. Casadio, *I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure*. Nucleic Acids Res, 2005. **33**(Web Server issue): p. W306-10.
29. Venselaar, H., et al., *Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces*. BMC Bioinformatics, 2010. **11**: p. 548.
30. Pettersen, E.F., et al., *UCSF Chimera--a visualization system for exploratory research and analysis*. J Comput Chem, 2004. **25**(13): p. 1605-12.
31. Bhattacharya, A., J.D. Ziebarth, and Y. Cui, *PolymiRTS Database 3.0: linking polymorphisms in microRNAs and their target sites with human diseases and biological pathways*. Nucleic Acids Res, 2014. **42**(Database issue): p. D86-91.
32. Warde-Farley, D., et al., *The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function*. Nucleic Acids Res, 2010. **38**(Web Server issue): p. W214-20.
33. Khera, D., et al., *Tay-Sachs disease: a novel mutation from India*. BMJ Case Rep, 2018. **11**(1).
34. Sheth, J., *Molecular study of lysosomal storage disorders in India*. Mol Cytogenet, 2014. **7**(Suppl 1 Proceedings of the International Conference on Human): p. I30.
35. Tamhankar, P.M., et al., *Clinical, biochemical and mutation profile in Indian patients with Sandhoff disease*. J Hum Genet, 2016. **61**(2): p. 163-6.
36. Aggarwal, S. and S.R. Phadke, *Medical genetics and genomic medicine in India: current status and opportunities ahead*. Mol Genet Genomic Med, 2015. **3**(3): p. 160-71.
37. Ankala, A., et al., *Clinical applications and implications of common and founder mutations in Indian subpopulations*. Hum Mutat, 2015. **36**(1): p. 1-10.
38. Udwadia-Hegde, A. and O. Hajirnis, *Temporary Efficacy of Pyrimethamine in Juvenile-Onset Tay-Sachs Disease Caused by 2 Unreported HEXA Mutations in the Indian Population*. Child Neurol Open, 2017. **4**: p. 2329048X16687887.
39. Cao, Z., et al., *Benign HEXA mutations, C739T(R247W) and C745T(R249W), cause beta-hexosaminidase A pseudodeficiency by reducing the alpha-subunit protein levels*. J Biol Chem, 1997. **272**(23): p. 14975-82.
40. Wendeler, M. and K. Sandhoff, *Hexosaminidase assays*. Glycoconj J, 2009. **26**(8): p. 945-52.
41. Park, N.J., et al., *Improving accuracy of Tay Sachs carrier screening of the non-Jewish population: analysis of 34 carriers and six late-onset patients with HEXA enzyme and DNA sequence analysis*. Pediatr Res, 2010. **67**(2): p. 217-20.
42. Strom, C.M., et al., *Molecular screening for diseases frequent in Ashkenazi Jews: lessons learned from more than 100,000 tests performed in a commercial laboratory*. Genet Med, 2004. **6**(3): p. 145-52.
43. McGinniss, M.J., et al., *Eight novel mutations in the HEXA gene*. Genet Med, 2002. **4**(3): p. 158-61.
44. Monaghan, K.G., et al., *Technical standards and guidelines for reproductive screening in the Ashkenazi Jewish population*. Genet Med, 2008. **10**(1): p. 57-72.

45. Ohno, K., et al., *Structural consequences of amino acid substitutions causing Tay-Sachs disease*. Mol Genet Metab, 2008. **94**(4): p. 462-8.
46. Fernandes, M.J., et al., *Identification of candidate active site residues in lysosomal beta-hexosaminidase A*. J Biol Chem, 1997. **272**(2): p. 814-20.
47. Tsuji, D., *[Molecular pathogenesis and therapeutic approach of GM2 gangliosidosis]*. Yakugaku Zasshi, 2013. **133**(2): p. 269-74.
48. Tutor, J.C., *Biochemical characterization of the GM2 gangliosidosis B1 variant*. Braz J Med Biol Res, 2004. **37**(6): p. 777-83.
49. Maegawa, G.H., et al., *The natural history of juvenile or subacute GM2 gangliosidosis: 21 new cases and literature review of 134 previously reported*. Pediatrics, 2006. **118**(5): p. e1550-62.
50. Gort, L., et al., *GM2 gangliosidoses in Spain: analysis of the HEXA and HEXB genes in 34 Tay-Sachs and 14 Sandhoff patients*. Gene, 2012. **506**(1): p. 25-30.
51. Mark, B.L., et al., *Crystallographic evidence for substrate-assisted catalysis in a bacterial beta-hexosaminidase*. J Biol Chem, 2001. **276**(13): p. 10330-7.
52. Pastores, G.M. and G.H.B. Maegawa, *GM2-Gangliosidoses*, in *Rosenberg's Molecular and Genetic Basis of Neurologic and Psychiatric Disease*, R.N. Rosenberg and J.M. Pascual, Editors. 2015, Elsevier. p. 321-329.
53. Zampieri, S., et al., *Molecular analysis of HEXA gene in Argentinean patients affected with Tay-Sachs disease: possible common origin of the prevalent c.459+5A>G mutation*. Gene, 2012. **499**(2): p. 262-5.
54. Perez, L.F., et al., *Thermodynamic characterisation of the mutated isoenzyme A of beta-N-acetylhexosaminidase in GM2-gangliosidosis B1 variant*. Clin Chim Acta, 1999. **285**(1-2): p. 45-51.
55. Chiricozzi, E., et al., *Chaperone therapy for GM2 gangliosidosis: effects of pyrimethamine on beta-hexosaminidase activity in Sandhoff fibroblasts*. Mol Neurobiol, 2014. **50**(1): p. 159-67.
56. Maegawa, G.H., et al., *Pyrimethamine as a potential pharmacological chaperone for late-onset forms of GM2 gangliosidosis*. J Biol Chem, 2007. **282**(12): p. 9150-61.