

Prioritization of Deleterious Variations in the Human Hypoxanthine-Guanine Phosphoribosyltransferase Gene

Semira Kheiri (MSc)

Department of Biotechnology,
Laboratory Sciences Research Center,
Golestan University of Medical
Sciences, Gorgan, Iran

Mahdieh Safar zad (MSc)

Department of Biochemistry,
Laboratory of Biotechnology, School
of Advanced Technologies in
Medicine, Golestan University of
Medical Sciences, Gorgan, Iran

Mohammad Shariati (BSc)

Department of Microbiology,
Laboratory of Biotechnology, School
of Advanced Technologies in
Medicine, Golestan University of
Medical Sciences, Gorgan, Iran

Hoda Sohrabi (MSc)

Department of Medical Biotechnology,
Department of Medical Biotechnology,
School of Advanced Technologies in
Medicine, Golestan University of
Medical Sciences, Gorgan, Iran

Corresponding author: Hoda Sohrabi

E-mail: hsohrabi65@yahoo.com

Tel: +98-9111780406

Address: Golestan University of
Medical Sciences, Gorgan, Iran

Received : 28 Sep 2016

Revised: 09 Aug 2016

Accepted: 11 Aug 2016

ABSTRACT

Background and Objectives: Non-synonymous single nucleotide polymorphisms are typical genetic variations that may potentially affect the structure or function of expressed proteins, and therefore could be involved in complex disorders. A computational-based analysis has been done to evaluate the phenotypic effect of non-synonymous single nucleotide polymorphisms in the gene encoding the human hypoxanthine-guanine phosphoribosyltransferase (HGPRT-1). HGPRT-1 is an enzyme involved in purine recycling pathway and its deficiency is associated with several human genetic disorders.

Methods: We provide a list of all amino acid replacements in the human HGPRT-1 from the dbSNP, Uniprot and dbEST databases. Sorting intolerant from tolerant (SIFT) and PolyPhen softwares were also used in our study.

Results: Of 94 amino acid substitutions, rs 267606863 was predicted to be the most deleterious. Substitutions of S110L and S104A in flexible loop and D194N, D201Y, H204R, Y195C, F199V and H204D in hood domain were predicted as functionally damaging.

Conclusion: It could be concluded that these intolerant changes may lie within a functional region of the protein and may affect the stability and folding of HGPRT-1. These variants could be used for future functional and molecular epidemiology studies of HGPRT-1-related disorders.

Keywords: Polymorphism, Single Nucleotide, Amino acid substitution, Hypoxanthine Phosphoribosyltransferase.

INTRODUCTION

Hypoxanthine-guaninephosphoribosyl transferase-1 (HGPRT-1) is an enzyme associated with Lesch-Nyhan syndrome (LNS), which leads to hyperuricemia (1, 2). HGPRT-1 deficiency may also cause megaloblastic anemia. This enzyme plays a critical role in purine salvage pathway and catalyzes the synthesis of guanosine monophosphate and inosine monophosphate from guanine and hypoxanthine, respectively (3). Moreover, knockout of *HGPRT-1* in mice inhibited the increase of dopamine in brain, but no neurological alterations have been shown (4). Moreover, uric acid overproduction is present in all HPR1-deficient patients and is associated with lithiasis and gout (5). Single nucleotide polymorphisms (SNPs) are known as the most common changes in the human genome, however only a small group of these variations can be considered as markers of susceptibility to various common human diseases (6). However, understanding the functions of coding substitutions can noticeably help understand the genetics of the human phenotypic variations. Among these alterations, non-synonymous (ns) SNPs, especially missense SNPs that cause amino acid substitutions in the encoded protein sequence are thought to have significant deleterious effects on the structure and/or function of the related protein (7, 8). Experimental analysis of all functionally important replacements is difficult. Predicting the deleterious variants for a key gene, such as *HGPRT-1*, from the majority of benign (no observable phenotypic impact) variants has been receiving much attention. Thus, computational strategies based on the biochemical severity of the amino acid substitution and protein sequence and/or structural information are proposed for the systematic analysis of genetic alterations (6, 9). In this study, we predicted variations of the *HGPRT-1* gene that potentially affect its protein function.

MATERIAL AND METHODS

The nsSNPs in the *HGPRT-1* gene and related protein sequences were retrieved from the National Center for Biotechnology Information (NCBI) database of SNPs, dbSNP (<http://www.ncbi.nlm.nih.gov/projects/SNP>) and Uniprot (<http://www.uniprot.org>) for the

computational analysis. To assess amino acid substitutions in HGPRT-1, we compared the amino acid sequence of HGPRT-1 (NP_000185.1) against the most recent EST database (build 130) using the tBlastn algorithm

(<http://www.ncbi.nlm.nih.gov/BLAST/>) (10-12). Overall, 110 HGPRT-1 or HGPRT-1-related sequences were screened. Only variants that were observed more than once were scored as possible amino acid substitutions and the rest were not included in the study. In addition, nonsense (*) or ambiguous (B, Z, X, etc.) amino acid substitutions were not considered in the analysis.

Sorting intolerant from tolerant (SIFT) is based on a premise that protein evolution is correlated with protein function. SIFT is a sequence homology-based tool that sorts intolerant from tolerant amino acid substitutions and predicts whether an amino acid substitution in a protein will have a phenotypic effect (13). We used SIFT to detect the deleterious coding variants and submitted the query in the form of either SNP IDs or as protein sequences. SIFT analysis was performed by allowing the algorithm to search for homologous sequences using the default settings (SWISS-PROT 45 and TrEMBL 28 databases, median conservation score 3.00, remove sequences >90% identical to query sequence). The tool generates alignments with a large number of homologous sequences and assigns scores to each residue, ranging from 0 to 1. Scores close to 0 indicate evolutionary conservation and intolerance to substitution, while scores close to 1 indicate tolerance to substitution. SIFT scores <0.05 are predicted by the algorithm to be intolerant or deleterious amino acid substitutions, whereas scores >0.05 are considered tolerant. The higher the tolerance index of a particular amino acid substitution, the lesser is its probable impact. Polymorphism phenotyping (PolyPhen) is an automatic tool for prediction of the possible impact of an amino acid substitution on the structure and function of a human protein (14). This prediction is based on straightforward empirical rules that are applied to the sequence and phylogenetic and structural information SWALL database ID or accession number together with sequence position with two amino acid alterations. We submitted the

identifying the substitution. Input options for the PolyPhen server are protein sequence, query in the form of protein sequence with mutational position and two amino acid variants.

PolyPhen searches for 3D protein structures, multiple alignments of homologous sequences and amino acid information in several protein structure databases. Then, 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.

Higher PSIC score differences indicate greater functional impact of a particular amino acid substitution. PolyPhen scores were designated as “probably damaging” (0.95–1), “possibly damaging” (0.7–0.95), and “benign” (0.00–0.31).

RESULTS

The human *HGPRT-1* gene retrieved from the dbSNP, dbEST and Uniprot databases contained 94 coding variants. Corresponding information for the variants including allele changes, amino acid substitutions and substitution positions are listed in table 1. In the PolyPhen analysis, protein sequences with mutational position and amino acid variants were submitted as inputs to the PolyPhen. This algorithm predicted 36 (38.2%) of these variants to be probably damaging, 19 (20.2%) to be possibly damaging and 39 (41.4%) to be benign substitutions. Moreover, all studied variants were predicted to be absolutely tolerated by the SIFT software (SIFT score= 1). The software showed low confidence that means the protein alignment does not have enough sequence diversity.

Table 1- SIFT and Polyphen predictions for nsSNPs in the human *HGPRT1* gene

#rs ID	Allele change	Amino acid substitution	substitution position	PolyPhen prediction	PolyPhen score
rs137852477	ATT ⇒ ATG	I [Ile] ⇒ M [Met]	132	Possibly damaging	0.930
rs137852478	GAT ⇒ GTT	D [Asp] ⇒ V [Val]	80	Possibly damaging	0.732
rs137852479	GAT ⇒ GGT	D [Asp] ⇒ G [Gly]	201	Possibly damaging	0.951
rs137852480	CTA ⇒ CCA	L [Leu] ⇒ P [Pro]	41	probably damaging	1.000
rs137852481	TTC ⇒ TTA	F [Phe] ⇒ L [Leu]	74	probably damaging	0.999
rs137852482	TCA ⇒ TTA	S [Ser] ⇒ L [Leu]	110	probably damaging	0.987
rs137852483	GTC ⇒ GAC	V [Val] ⇒ D [Asp]	130	probably damaging	0.978
rs137852484	GCA ⇒ TCA	A [Ala] ⇒ S [Ser]	161	Possibly damaging	0.905
rs137852485	AGC ⇒ AGA	S [Ser] ⇒ R [Arg]	104	probably damaging	1.000
rs137852486	TTC ⇒ GTC	F [Phe] ⇒ V [Val]	199	probably damaging	0.985
rs137852487	GGG ⇒ GAG	G [Gly] ⇒ E [Glu]	70	probably damaging	1.000
rs137852488	GGC ⇒ CGC	G [Gly] ⇒ R [Arg]	71	probably damaging	0.971
rs137852489	CAG ⇒ AAG	Q [Gln] ⇒ K [Lys]	109	Benign	0.001
rs137852490	CAT ⇒ GAT	H [His] ⇒ D [Asp]	204	Possibly damaging	0.747
rs137852491	AGG ⇒ AAG	R [Arg] ⇒ K [Lys]	45	probably damaging	1.000
rs137852492	GAC ⇒ TAC	D [Asp] ⇒ Y [Tyr]	177	probably damaging	1.000
rs137852493	CCA ⇒ CTA	P [Pro] ⇒ L [Leu]	176	probably damaging	0.994
rs137852494	CGA ⇒ GGA	R [Arg] ⇒ G [Gly]	51	Benign	0.081
rs137852495	ATG ⇒ ACG	M [Met] ⇒ T [Thr]	57	Benign	0.009
rs137852496	ATG ⇒ AAG	M [Met] ⇒ K [Lys]	143	probably damaging	1.000
rs137852498	ACC ⇒ ATC	T [Thr] ⇒ I [Ile]	168	Benign	0.258
rs137852499	GGT ⇒ AGT	G [Gly] ⇒ S [Ser]	16	Possibly damaging	0.792
rs137852500	GGA ⇒ AGA	G [Gly] ⇒ R [Arg]	58	Benign	0.029
rs137852501	CTG ⇒ GTG	L [Leu] ⇒ V [Val]	78	probably damaging	0.961
rs137852502	GAT ⇒ GGT	D [Asp] ⇒ G [Gly]	52	Possibly damaging	0.869
rs137852503	GGC ⇒ GAC	G [Gly] ⇒ D [Asp]	140	probably damaging	1.000
rs137852504	GAC ⇒ GAG	D [Asp] ⇒ E [Glu]	194	Possibly damaging	0.520
rs137852506	CTC ⇒ TTC	L [Leu] ⇒ F [Phe]	65	probably damaging	0.999
rs267606863	GAC ⇒ AAC	D [Asp] ⇒ N [Asn]	194	probably damaging	0.973
rs387906725	CGT ⇒ CAT	R [Arg] ⇒ H [His]	48	probably damaging	1.000
rs398123242	AAA ⇒ GAA	K [Lys] ⇒ E [Glu]	166	Possibly damaging	0.790
rs369065223	TCA ⇒ TTA	S [Ser] ⇒ L [Leu]	123	Benign	0.047
rs371800609	CGA ⇒ CAA	R [Arg] ⇒ Q [Gln]	51	Benign	0.059
rs374646638	GAC ⇒ GAA	D [Asp] ⇒ E [Glu]	18	Benign	0.000
rs534390401	CCA ⇒ GCA	P [Pro] ⇒ A [Ala]	169	Benign	0.006

DISCUSSION

HGPRT-1 is an important enzyme that has a key role in purine recycling pathway. Numerous investigations showed that alterations in the *HGPRT-1* gene may affect its phenotype in several human disorders (15, 16). LNS is a rare X-linked recessive inherited neurogenetic disorder caused by deletion of exons 2 and 3 of the *HGPRT* gene, which results in deficiency of the purine salvage enzyme (HGPRT) (17, 18). The human *HGPRT-1* gene is located in the q26-q27 region of the X chromosome and has eight introns and nine exons (19, 20). The coding region of this gene contains 654 nucleotides encoding a core domain containing the phosphoribosylpyrophosphate (PRPP) binding site and a hood domain (residues 188-204) that binds to the purine base substrate (21, 22). Results of the PolyPhen analysis suggested that amino acid substitutions S110L and S104A located at the flexible loop (residues 101-111), which is involved in interactions with the bound Mg^{2+} PRPP molecule and F199V, H204D, D194N, D201Y, H204R and Y195C in the hood domain contributing to the purine base binding may be highly deleterious and affect the protein function. According to SIFT, amino acid changes at surface of this protein probably have less structural and functional effects because all the studied alterations have been absolutely tolerated (score= 1.00) and the role of evolutionary players may interpret these observations. Furthermore, among the damaging replacements, only D194N (rs267606863) occurs at the Mg^{2+} binding site of the enzyme that has been determined to be a highly effective substitution on the structure of this protein. Considering the diversity of rules for predicting the effect of alterations on protein function, the outcomes could be somewhat dissimilar. However, the most reliable result is predictions of an SNP that overlap in various algorithms. There is no direct approach to

evaluate the accuracy of these predictions made by SIFT and PolyPhen, as it is possible that the algorithms may have used different data sets (23). Hence, even if the predictions made by these softwares were not completely consistent for this subset of mutations, these HGPRT-1 variants should still be regarded as candidates for SNP screening. Furthermore, molecular modeling and experimental investigations can be used to confirm our results and clarify possible regulatory functions of the studied deleterious replacements.

CONCLUSION

Screening of the *HGPRT-1* gene mutations revealed the existence of many polymorphic sites, some of which are associated with phenotypic traits and diseases risk. However, given the high number of SNPs in this gene, association studies should be carried on the genetic variants with functional significance. Use of computational tools may be an alternative approach to select functional SNPs. Furthermore, molecular modeling and experimental investigations can be used to confirm our results and clarify possible regulatory functions of the studied deleterious replacements.

ACKNOWLEDGMENTS

This study received approval from the Golestan University of Medical Sciences, Iran. We would like to thank our colleagues, especially Mr. Esmail Samadian for useful comments.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

REFERENCES

1. Hladnik U, Nyhan WL, Bertelli M. *Variable expression of HPRT deficiency in 5 members of a family with the same mutation*. Archives of neurology. 2008; 65(9): 1240-3. doi: 10.1001/archneur.65.9.1240.
2. Chandekar P, Madke B, Kar S, Yadav N. *Lesch-nyhan syndrome in an Indian child*. Indian journal of dermatology. 2015; 60(3): 298-300. doi: 10.4103/0019-5154.156392.
3. Torres RJ, Puig JG. *Hypoxanthine-guanine phosphoribosyltransferase (HPRT) deficiency: Lesch-Nyhan syndrome*. Orphanet journal of rare diseases. 2007; 2 (1): 48. doi.org/10.1186/1750-1172-2-48.
4. Engle SJ, Womer DE, Davies PM, Boivin G, Sahota A, Simmonds HA, et al. *HPRT-APRT-deficient mice are not a model for Lesch-Nyhan syndrome*. Human molecular genetics. 1996; 5(10): 1607-10.
5. Torres RJ, Prior C, Puig JG. *Efficacy and safety of allopurinol in patients with hypoxanthine-guanine phosphoribosyltransferase deficiency*. Metabolism. 2007; 56(9): 1179-86.
6. Wang L-L, Li Y, Zhou S-F. *Prediction of deleterious non-synonymous single nucleotide polymorphisms of genes related to ethanol-induced toxicity*. Toxicology letters. 2009; 187(2): 99-114. doi: 10.1016/j.toxlet.2009.02.007.
7. Bao L, Cui Y. *Functional impacts of non-synonymous single nucleotide polymorphisms: Selective constraint and structural environments*. FEBS letters. 2006; 580(5): 1231-4. DOI:10.1016/j.febslet.2006.01.035.
8. Chasman D, Adams RM. *Predicting the functional consequences of non-synonymous single nucleotide polymorphisms: structure-based assessment of amino acid variation*. Journal of molecular biology. 2001;307(2):683-706. DOI:10.1006/jmbi.2001.4510.
9. Jordan DM, Ramensky VE, Sunyaev SR. *Human allelic variation: perspective from protein function, structure, and evolution*. Current opinion in structural biology. 2010; 20(3): 342-50. doi: 10.1016/j.sbi.2010.03.006.
10. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. *Basic local alignment search tool*. Journal of molecular biology. 1990; 215(3): 403-10. DOI:10.1016/S0022-2836(05)80360-2.
11. Gish W, States DJ. *Identification of protein coding regions by database similarity search*. Nature genetics. 1993; 3(3): 266-72.
12. Boguski MS, Lowe TM, Tolstoshev CM. *dbEST—database for “expressed sequence tags”*. Nature genetics. 1993; 4(4): 332-3.
13. Ng PC, Henikoff S. *SIFT: Predicting amino acid changes that affect protein function*. Nucleic acids research. 2003; 31(13): 3812-4.
14. Ramensky V, Bork P, Sunyaev S. *Human non-synonymous SNPs: server and survey*. Nucleic acids research. 2002; 30(17): 3894-900.
15. Fu R, Ceballos-Picot I, Torres RJ, Larovere LE, Yamada Y, Nguyen KV, et al. *Genotype–phenotype correlations in neurogenetics: Lesch-Nyhan disease as a model disorder*. Brain. 2014;137(5):1282-303. doi: 10.1093/brain/awt202.
16. Monzani PS, Trapani S, Thiemann OH, Oliva G. *Crystal structure of Leishmania tarentolae hypoxanthine-guanine phosphoribosyltransferase*. BMC structural biology. 2007;7(1): 59.
17. Baba S, Saito T, Yamada Y, Takeshita E, Nomura N, Yamada K, et al. *Novel mutation in HPRT1 causing a splicing error with multiple variations*. Nucleosides, Nucleotides and Nucleic Acids. 2017; 36(1): 1-6. doi: 10.1080/15257770.2016.1163381.
18. Thumfart J, Weschke B, Ringe H, Weinhold N, Müller D. *Acute renal failure unmasking Lesch-Nyhan disease in a patient with tuberous sclerosis complex*. European Journal of Paediatric Neurology. 2016; 20(4): 649-51. doi: 10.1016/j.ejpn.2016.04.014.
19. Keebaugh AC, Sullivan RT, Thomas JW, Program NCS. *Gene duplication and inactivation in the HPRT gene family*. Genomics. 2007; 89(1): 134-42. DOI:10.1016/j.ygeno.2006.07.003.
20. Flier JS, Moore MJ, Wilson JM, Young AB, Kelley WN. *Hypoxanthine-guanine phosphoribosyltransferase deficiency: The molecular basis of the clinical syndromes*. New England Journal of Medicine. 1983; 309(15): 900-10.
21. Duan J, Nilsson L, Lambert B. *Structural and functional analysis of mutations at the human hypoxanthine phosphoribosyl transferase (HPRT1) locus*. Human mutation. 2004; 23(6): 599-611.
22. Balendiran GK, Molina J, Xu Y, Torres-Martinez J, Stevens R, Focia PJ, et al. *Ternary complex structure of human HGPRTase, PRPP, Mg²⁺, and the inhibitor HPP reveals the involvement of the flexible loop in substrate binding*. Protein Science. 1999; 8(5): 1023-31. DOI:10.1110/ps.8.5.1023.
23. Naderi M, Gharaei R, Soleymani-Nejadian E, Samadian E. *In Silico survey of functional coding variants in human AEG-1 gene*. Egyptian Journal of Medical Human Genetics. 2013;14(4):419-22.