

## Sequence and structure based assessment of non-synonymous SNPs in hypertrichosis universalis

Rabiya Waheed, Mohammad Haroon Khan\*, Raisa Bano & Hamid Rashid

Department of Bioinformatics, Mohammad Ali Jinnah University, Islamabad, Pakistan; Mohammad Haroon Khan - Email: haroon.khan@jinnah.edu.pk; Phone: +92 51 111 87 87 87, Ext. 108; \*Corresponding author

Received March 09, 2012; Accepted March 21, 2012; Published April 13, 2012

### Abstract:

Hairs are complex structures, making a protective layer and serves different biological functions. TRPS1, a transcription factor is one of the candidate genes causing congenital hypertrichosis, an excessive hair growth at inappropriate body parts. SNPs of TRPS1 were retrieved from dbSNP which were screened by SIFT and PolyPhen servers based on their functional impacts. Out of the screened SNPs, rs181507248 and rs146506752 were predicted as intolerant and damaging by both the servers. The predicted tertiary structure of the native TRPS1 after refinement and validation was successfully submitted to the Protein Model Database and was assigned with PMDB ID PM0077843, as it was previously unpredicted. It was observed through the structure based analysis that, the SNPs rs181507248 and rs146506752 caused significant changes in the secondary and tertiary structures as well as the physiochemical properties of TRPS1 protein. It can thus be concluded that the changed properties due to these single nucleotide polymorphisms effect the interactions of TRPS1 which result in congenital hypertrichosis.

**Keywords:** Hypertrichosis, TRPS1, structure prediction, SNP

### Background:

Hairs are one of the defining features of mammals, are a complex structure, epidermal in its origin, grow from the dermal skin layer, making a protective layer and serves for a number of important biological functions. They may be of many types from very fine to the stiff quills, but soft woolly hairs are more common. Human body is covered with very fine vellus hairs except some parts like lips, parts of the genitals, nipples, soles and palms. Hair growth on the human body starts at the age of six months when the fetus is in the uterus. It has been observed that hair grow at different rates among different ethnic races.

Hypertrichosis is an excessive hair growth at inappropriate body parts, which can be localized or generalized and may be acquired or congenital [1]. Congenital hypertrichosis, is an inherited autosomal dominant trait [2], characterized by increased lanugo hair at birth covering the whole body except

some parts like palms and soles etc [3]. The affected individuals display certain abnormal facial characteristics like triangular, coarse face, broad intercanthal distance and long palpebral fissures [4], small anteverted nares, thick nasal wings, a trapezoid mouth and bone abnormalities [5, 6].

Cytogenetics analysis has revealed an important association between rearrangements of chromosome 8, with the break point 8q22-8q24 [7, 5]. Trichorhinophalangeal syndrome 1 (TRPS1), a transcription factor is one of the candidate that maps to 8q23 [8]. Trps1 has also been found expressing in highly proliferative epithelial cells of rat, mouse and human hair follicles [9]. In this paper, we considered the congenital hypertrichosis associated SNPs of structural and UTR regions of TRPS1 with respect to their effect on structure and also to reveal the extent to which these SNPs affect the respective protein. This study will also help to better understand the potential molecular causes of several other genetic disorders,

which will provide guidance for further studies towards developing treatment of the genetic disorders.

## Methodology:

A single nucleotide substitution, addition or deletion can significantly affect structure and function of a gene product or some times of multiple gene products. Associating sequence variations with heritable phenotypic characters is an important area of research in molecular biology. Single nucleotide polymorphisms (SNPs) are the most common sequence variations found nearly in all genomes. For detailed structure based assessment of non-synonymous SNPs, the human TRPS1 protein sequence was retrieved from Ensembl ([www.ensembl.org/](http://www.ensembl.org/)) and it's associated SNPs from dbSNP (<http://www.ncbi.nlm.nih.gov/projects/SNP/>). The deleterious or intolerant SNPs were determined through SIFT (Sorting intolerant from tolerant) [10], sorting intolerant from tolerant amino acid substitutions based on sequence homology and predict whether an amino acid substitution in a protein will have a phenotypic effect. Functionally important residue positions should be conserved, whereas unimportant appear diverse in an alignment of the protein family. The damaging SNPs were further sorted from benign by using PolyPhen server ([genetics.bwh.harvard.edu/pph/](http://genetics.bwh.harvard.edu/pph/)), predicting the possible impact of residue substitution on the structure and function of a human protein. It identifies homologues of the input through BLAST in the nrdb database and computes the absolute difference value between profile scores of both allelic variants in the polymorphic position.

Proteins are complex organic compounds, essential to the structure and function of all living cells and viruses. Usually they do not exist as linear polypeptides but rather as compact and folded structures. Protein functions are determined by their overall three-dimensional conformation. They can be better understood in terms of the simple secondary structure elements put together in high order arrangements which ultimately define their biological functions. Prediction of protein structure and its features is therefore an important area of computational biology. The screened SNPs were substituted in the native TRPS1 protein sequence to get the mutants, secondary structures were predicted using GOR IV method [11] and were aligned through PRALINE (<http://www.ibi.vu.nl/programs/pralinewww/>), a multiple sequence alignment program based upon progressive alignment to compare the proteins for their variations in the secondary structure elements. The three-dimensional structure of TRPS1 protein and its mutants were predicted through I-Tasser server [12] which builds the 3D models based on multiple-threading alignments. The predicted models were refined through FG-MD (<http://zhanglab.ccmb.med.umich.edu/FG-MD/>) to enhance the quality of predicted structures and validated for quality assurance through multiple servers like ProSA-web [13], WHATIF [14] and RAMPAGE [15]. Ramachandran plot is a two-dimensional plot of  $\phi$ - $\psi$  angles for the assessment of protein backbone, depicts information about the protein structure and conformation and also provides information about the residues lying in favored, allowed or outlier region [16].

The predicted models were compared in 3D against the native protein through PDBeFold for the assessment of structural impacts enforced by the SNPs. Physicochemical properties for the native and mutants were predicted through ProtParam for further analysis of the impact of SNPs on molecular weight, theoretical PI, amino acid composition, instability index, grand average of hydrophobicity etc which can thus affect the interactions of TRPS1 protein in the network.

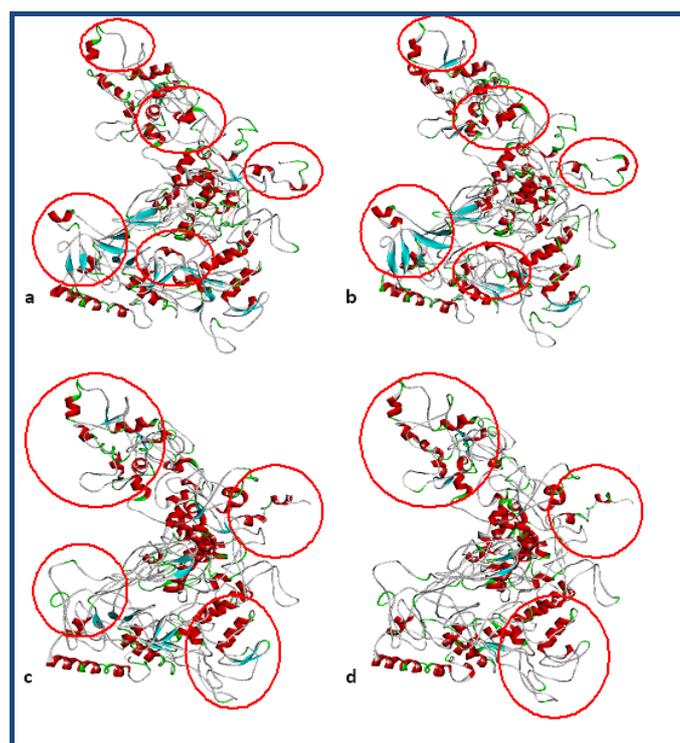
## Discussion:

The TRPS1 gene, located on chromosome 8 from 116,420,723 to 116,681,227 provides instructions for making a protein that regulates the activity of many other genes. A total of 38 SNPs of TRPS1 protein were retrieved from dbSNP which were further screened by SIFT and PolyPhen based on their functional impacts. It was observed that, 17 SNPs were categorized as intolerant that can affect protein function on the basis of their tolerance index  $<0.05$  and three were categorized as damaging by Polyphen server on the basis of PSIC value. Out of the screened SNPs, rs181507248 (tolerance index 0.01) and rs146506752 (tolerance index 0.00) were categorized both as intolerant by SIFT as well as damaging by PolyPhen (PSIC value 1.773 and 1.692 respectively), were modeled and compared to the native TRPS1 protein for investigating their structural variations. rs181507248 cause substitution of R→L at position 814 while rs146506752 causes V→I substitution at residue position 639. There were observed 278 Alpha Helices in the native protein which remained conserved after substitutions V639I (rs181507248) and R814L (rs146506752), but changes were observed in the extended strands and random coils. Substitution V639I (rs181507248) caused a decrease in extended strands from 215 to 213 while increase in random coils from 801 to 803. Similarly, R814L (rs146506752) caused an increase in extended strands from 215 to 216 while decrease in random coils from 801 to 800.

The alternations of key residues in a protein cause loss of its normal biological functions. Theoretical pI of mutant TRPS1 protein with rs181507248 was lower than normal (7.53) by 0.26 while that of the mutant with rs146506752 by 0.37, which means increase in net negative charge and thus promoting hydrophobicity. Both the SNPs caused changes in the secondary structure elements and physicochemical properties thus it was inferred that these changes may be translated into the tertiary structure. Protein tertiary structures are vital to narrate the function of protein and thus were predicted for the native TRPS1 Protein as it was previously unpredicted. The predicted native model was properly refined and validated for quality enhancement. It was ensured that there lie no errors in valine, threonine, leucine, isoleucine, arginine, phenylalanine, tyrosine, aspartic acid, glutamic acid and torsion angle conventions. Bond lengths were in agreement with typical bond lengths, side chain of residues was planar and lie within the expected RMS deviations, there were no atoms with wrong handedness and z-score was observed -2.89. After fulfilling the necessary formalities, the model was successfully submitted to the Protein Model Database and was assigned with PMDB ID PM0077843.

Knowledge of protein structure provides an insight into its interactions [17], which define the protein's biological role and functions [18]. 3D structure are more conserved than sequence

[19], thus native and mutant structures were superimposed in three dimensions which actually produces a measure to assess the level of similarity of the aligned structures. RMSD value of 2.316Å was observed for R814L (rs181507248) and 2.11Å for V639I (rs146506752). Generally RMSD value between 0 and 1.5Å represent very similar structures while increase in RMSD means increase structural dissimilarity. Moreover, small RMSD computed over large structures are also very significant as compared to larger RMSD values computed over structures with small number of residues. TRPS1 protein is a large protein so the observed RMSD values of 2.11Å and 2.316Å means significant structural variations (**Figure 1**), reflecting significant functional variations.



**Figure 1:** Tertiary structures of the Native and mutated TRPS1 proteins predicted through I-Tasser and refined through FG-MD, circles shows the areas of structural variations. a) Unrefined tertiary structure of Native TRPS1 (b) Refined tertiary structure of Native TRPS1; (c) Tertiary structure of V639I (rs181507248); (d) Tertiary structure of R814L (rs146506752)

### Conclusion:

SNPs represent the most common and frequent type of sequence variations in human DNA. Non-synonymous coding SNPs are considered to have the highest impact on phenotype of an organism. Thirty eight non-synonymous coding SNPs in the structural and UTR regions of TRPS1, associated with hypertrichosis universalis were considered in this study with

respect to their effect on structure and the extent to which they affect the respective protein. Sequence and Structure based computations were systematically evaluated and applied to the single nucleotide polymorphisms and it was observed that two of them (rs181507248 and rs146506752) are damaging as they affect the protein structure both at secondary and tertiary levels and also affecting other physiochemical properties of TRPS1 Protein, so damaging its interactions in the network also. It is hoped that this study will help to better explain the consequences posed by these SNPs. Our comprehensive investigation will also provide a valuable insight into some of the features that have not been previously studied and computational confirmation for some of the previous results by other researchers. It is obvious from the results that these computational approaches can be used against other genetic disorders as well.

### References:

- [1] Wendelin DS *et al.* *J Am Acad Dermatol.* 2003 **48**: 161 [PMID: 12582385]
- [2] Lee IJ *et al.* *Pediatr Dermatol.* 1993 **10**: 263 [PMID: 8415305]
- [3] DeRaeve L & Keymolen K, *Arch Dermatol.* 2011 **147**: 746 [PMID: 21690547].
- [4] Baumeister FA *et al.* *Clin Genet.* 1993 **44**: 121 [PMID: 8275569]
- [5] Wuyts W *et al.* *Am J Med Genet.* 2002 **113**: 326 [PMID: 12457403]
- [6] Belengeanu V *et al.* *Clin Dysmorphol.* 2004 **13**: 265 [PMID: 15365467]
- [7] Tadin-Strapps M *et al.* *Cytogenet Genome Res.* 2004 **107**: 68 [PMID: 15305058]
- [8] Momeni P *et al.* *Nat Genet.* 2000 **24**: 71 [PMID: 10615131]
- [9] Fantauzzo KA *et al.* *Gene Expr Patterns.* 2008 **8**: 51 [PMID: 18054290]
- [10] Ng PC & Henikoff S, *Nucleic Acids Res.* 2003 **31**: 3812 [PMID: 12824425]
- [11] Garnier J *et al.* *Methods Enzymol.* 1996 **266**: 540 [PMID: 8743705]
- [12] Roy A *et al.* *Nature Protocols.* 2010 **5**: 725 [PMID: 20360767]
- [13] Wiederstein M & Sippl MJ, *Nucleic Acids Research.* 2007 **35**: W407 [PMID: 17517781]
- [14] Vriend G, *J Mol Graph.* 1990 **8**: 52 [PMID: 2268628].
- [15] Lovell SC *et al.* *Proteins.* 2002 **50**: 437 [PMID: 12557186].
- [16] Ho BK & Brasseur R, *BMC Structural Biology.* 2005 **5**: 14 [PMID: 16105172].
- [17] Aydin Z *et al.* *BMC Bioinformatics.* 2011 **12**: 154 [PMID: 21569525].
- [18] Cheng J *et al.* *Nucleic Acids Res.* 2005 **33**: W72 [PMID: 15980571].
- [19] Capriotti E & Marti-Renom MA, *BMC Bioinformatics.* 2010 **11**: 322 [PMID: 20550657]

Edited by P Kanguane

Citation: Waheed *et al.* 8(7): 316-318 (2012)

**License statement:** This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited.