

RESEARCH ARTICLE

Identification and *in silico* analysis of functional SNPs of human TAGAP protein: A comprehensive study

Maria Arshad*, Attya Bhatti*, Peter John

Atta-Ur-Rahman School of Applied Biosciences, National University of Sciences & Technology, Islamabad, Pakistan

* maria.arshad14@gmail.com (MA); attyabhatti@asab.nust.edu.pk (AB)



Abstract

Genetic polymorphisms in *TAGAP* gene have been associated with many diseases including rheumatoid arthritis, multiple sclerosis and other autoimmune disorders. Identifying functional SNPs in such disease associated genes is an uphill task hence before planning larger population study, it is better to scrutinize putative functional SNPs. In this study we used various computational approaches to identify nsSNPs which are deleterious to the structure and/or function of *TAGAP* protein that might be causing these diseases. Computational analysis was performed by five different *in silico* tools including SIFT, PROVEAN, PolyPhen-2, PhD-SNP and SNPs&GO. The study concludes that mutations of Glycine → Glutamic Acid at position 120, Glycine → Tryptophan at position 141 and Valine → Methionine at position 151 are major mutations in native *TAGAP* protein which might contribute to its malfunction and ultimately causing disease. The study also proposed 3D structures of native *TAGAP* protein and its three mutants. Future studies should consider these nsSNPs as main target mutations in various diseases involving *TAGAP* malfunction. This is the first comprehensive study, where *TAGAP* gene variants were analyzed using *in silico* tools hence will be of great help while considering large scale studies and also in developing precision medicines for cure of diseases related to these polymorphisms. Furthermore, animal models of various autoimmune diseases and having these mutations might be of help in exploring their precise roles.

OPEN ACCESS

Citation: Arshad M, Bhatti A, John P (2018) Identification and *in silico* analysis of functional SNPs of human TAGAP protein: A comprehensive study. PLoS ONE 13(1): e0188143. <https://doi.org/10.1371/journal.pone.0188143>

Editor: Yang Zhang, University of Michigan, UNITED STATES

Received: July 31, 2017

Accepted: November 1, 2017

Published: January 12, 2018

Copyright: © 2018 Arshad et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: The author(s) received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

Introduction

Genetic polymorphisms in human genome are mostly (90%) single nucleotide polymorphisms (SNPs) which are single base pair changes in alleles and are considered to be the most common kind of variations in DNA sequence. The SNPs in coding region of human genome are of much importance and around 500,000 SNPs fall in this region [1]. Among these the non-synonymous SNPs (nsSNPs), also named as missense SNPs, are highly significant as they are responsible for amino acid residue substitutions resulting in functional diversity of proteins in humans [2]. Functional variations can have deleterious or neutral effects on protein structure

or function [3]. Damaging effects might include destabilization of protein structure, altering gene regulation [4], affecting protein charge, geometry, hydrophobicity [5], stability, dynamics, translation and inter/intra protein interactions [2,6,7], hence structural integrity of cells comes under risk [8]. Thus it can be avowed that nsSNPs might get linked with many human diseases because of these missense SNPs.

Numerous studies in the past have shown that nsSNPs are responsible for about 50% of mutations which are involved in various genetic disorders [9,10] including many inflammatory and autoimmune disorders [11–15]. A study analyzed genetic variations in ABCA1 gene and predicted their deleterious effects causing familial hypoalphalipoproteinemia and tangier disease [16]. Similar study identified missense SNPs in STEAP2 which cause its upregulation leading to prostate cancer [17]. The nsSNPs in NKX2-5 gene were found associated with congenital heart defects because of their damaging effects on structural features of the protein [18]. A recent study proposed that nsSNPs in MITF gene might cause malignant melanoma [19]. Another latest study on nsSNPs and their effects on patients with non-small cell lung cancer treated with immunotherapy suggested that the combination of deleterious SNPs and known pathogenic lesions might help in getting advantage from immunotherapy [20]. This study is aimed to investigate nsSNPs of T-cell Activation Rho GTPase Activating Protein (TAGAP) and their effects on its structure and function. This protein, located on chromosome 6q25, acts as a molecular switch and is considered important in modulating cytoskeletal changes [21], in activation of T cells [22] and is therefore of particular interest in the context of T cell-driven autoimmune disease processes. For this purpose, we made use of various bioinformatics tools to come up with the most deleterious and damaging nsSNPs of TAGAP protein. 3D models of TAGAP protein and its mutant forms are also proposed in this study. This is the first ever study which covers an extensive *in silico* analysis of nsSNPs of TAGAP protein hence this work might be useful in future in developing precision medicines for the treatment of diseases caused by these genomic variations.

Materials and methods

Retrieving nsSNPs

Information of missense SNPs (SNP ID, protein accession number, position, residue change and global minor allele frequency (MAF) was retrieved from NCBI dbSNP database (<https://www.ncbi.nlm.nih.gov/projects/SNP/>) [23]. All 275 nsSNPs were filtered for investigation.

Identifying the most damaging nsSNPs

We utilized five different bioinformatics tools to predict functional effects of nsSNPs recruited from dbSNP database. These algorithmic programs included: SIFT-Sorting Intolerant From Tolerant [http://sift.jcvi.org/www/SIFT_seq_submit2.html] [24,25], PROVEAN-Protein Variation Effect Analyzer [<http://provean.jcvi.org/index.php>] [26], PolyPhen-2-Polymorphism Phenotyping v2 [<http://genetics.bwh.harvard.edu/pph2/>] [27], Ph^D-SNP -Predictor of human Deleterious Single Nucleotide Polymorphisms [<http://snps.biofold.org/phd-snp/phd-snp.html>] [28] and SNPs&GO [<http://snps.biofold.org/snps-and-go/snps-and-go.html>] [29]. The SNPs predicted deleterious by at least four *in silico* tools were considered high risk nsSNPs and investigated further.

Identifying structural and functional properties

For sorting disease associated or neutral amino acid substitutions in humans, MutPred v1.2 was consulted which is a web based application tool that effectively screens amino acid

substitutions [30]. It also helps predicting molecular cause of the disease. MutPred is based upon gain/loss of 14 different functional and structural properties like loss of a phosphorylation site or gain of helical propensity. Protein sequence (FASTA format) of TAGAP and its amino acid substitutions were submitted. Output provides p-value where; $p < 0.05$ and $p < 0.01$ were considered as confident and very confident hypotheses, respectively.

Analyzing protein stability

To check the stability of target protein, I-Mutant 2.0 was used which is a support vector machine based web server that helps in predicting any change in stability of protein after getting mutated. The tool uses data derived from ProTherm which is currently the most comprehensive database of experimental data on protein mutations. It predicts reliability index (RI) of the results ranging from 0–10, where 10 being the highest reliability [31,32]. We submitted TAGAP protein sequence to predict effects of the most damaging nsSNPs on the protein. Conditions for all submissions were set at temperature 25°C and pH 7.0.

Analyzing protein evolutionary conservation

ConSurf is a bioinformatics tool that we used to estimate evolutionary conservation of amino acid positions using protein sequence [33]. Analysis is based on phylogenetic relations between homologous sequences [34–36]. Degree of conservation of amino acid residues was estimated using 50 homologous sequences. We selected those highly conserved residues for further analysis which were located at the sites of high risk nsSNPs.

3D protein modeling

The 3D models for wild type TAGAP protein and TAGAP mutated with high risk nsSNPs were generated using two homology modeling tools; Phyre2 and I-TASSER [37,38–40]. The resultant structures were viewed by Chimera 1.11 which is an extensive program for interactive visualization and analysis of molecular structures and related data [41] and later verified by ERRAT which is a program for verifying protein structures [42]. Afterwards TM-align was used to compare wild type protein structure with mutant protein structures. This algorithm computes template modeling-score (TM-score) and root mean square deviation (RMSD) along with superposition of the structures. TM-score gives the values in 0 and 1, where 1 indicates perfect match between two structures. While higher RMSD indicates greater variation between wild type and mutant structures [43,44].

Predicting post translational modification (PTM) sites

The putative methylation sites in the TAGAP protein sequence were predicted by PSSMe and BPB-PPMS. The former tool identifies methylation sites based on information gain feature optimization method. The higher support vector machines (SVMs) probability indicates higher probability of lysine (or arginine) to get methylated. False-positive predictions were controlled by focusing on sites with stringency setting higher than 50% [45]. Methodology of BPB-PPMS is based on Bi-profile Bayes combined with SVMs having threshold value of 0.5 [46]. Likely phosphorylation sites in TAGAP protein at serine, threonine and tyrosine residues were predicted using NetPhos 3.1 and GPS 3.0. NetPhos 3.1 uses ensembles of neural networks to complete this task and residues having scores >0.5 threshold are considered phosphorylated [47]. Likewise in GPS 3.0, higher value depicts higher potential of the residue to get phosphorylated [48]. Putative protein ubiquitylation sites were predicted by UbPred and BDM-PUB. In

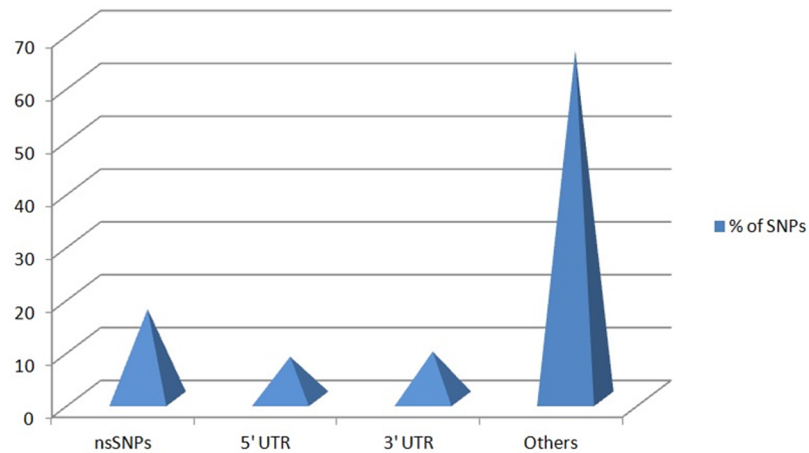


Fig 1. Clustered Pyramid showing the percentages of the SNPs in TAGAP gene. (nsSNPs: 17%; 5'UTR SNPs: 8%; 3'UTR SNPs: 9%; Other SNPs: 66%).

<https://doi.org/10.1371/journal.pone.0188143.g001>

UbPred, lysine residues with a score of ≥ 0.62 were considered ubiquitylated [10] and in BDM-PUB, balanced cut-off option was selected [30].

Results and discussion

nsSNPs retrieved from dbSNP database

We used dbSNP database to retrieve SNPs of interest as it is the most extensive SNP database [23]. There were total 1721 SNPs, of which 275 were nsSNPs, 147 occurred in 5'UTR, 162 in 3'UTR region, and the rest were of other types (Fig 1). We selected only nsSNPs for our investigation, details are provided in S1 Table. The global MAFs of these SNPs are shown in graphical form in Fig 2.

Deleterious nsSNPs identified in TAGAP

We subjected all nsSNPs to five different *in silico* nsSNP prediction algorithms to investigate whether these SNPs have any effect on structure or function of TAGAP protein. *In silico* tools used for this purpose were: SIFT, PROVEAN, PolyPhen-2, PhD-SNP and SNPs&GO. According to SIFT results, nsSNPs scoring tolerance index (TI) of ≤ 0.05 are considered intolerant. In

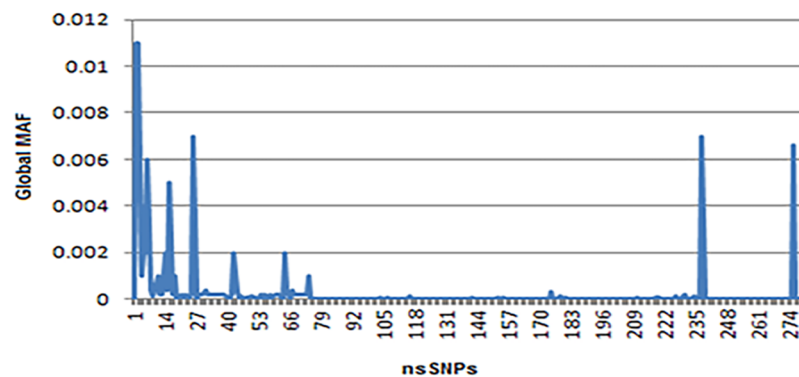


Fig 2. Graphical representation of global MAFs of nsSNPs.

<https://doi.org/10.1371/journal.pone.0188143.g002>

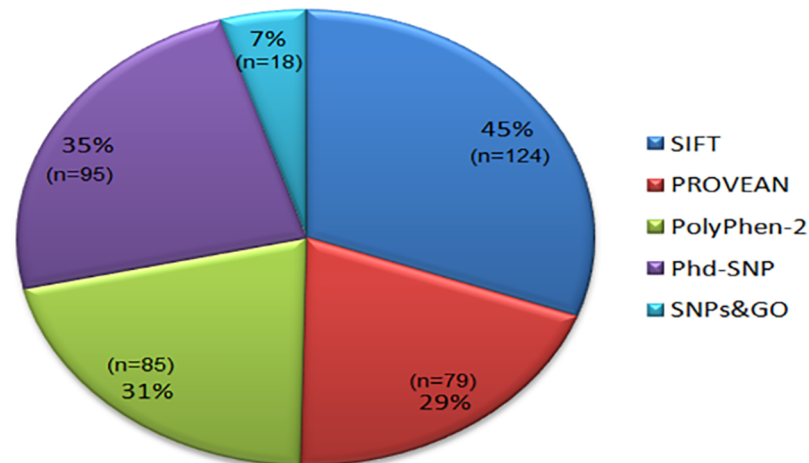


Fig 3. Pie chart showing percentage of damaging nsSNPs identified. The distribution of damaging nsSNPs by percentage (%) and number (n) identified by five *in silico* tools; SIFT, PROVEAN, PolyPhen-2, Phd-SNP and SNPs&GO.

<https://doi.org/10.1371/journal.pone.0188143.g003>

PROVEAN, the variants are predicted as deleterious when final score is below threshold value of -2.5 and neutral when it is above this value. PolyPhen-2 results predicted probably damaging, possibly damaging and benign nsSNPs, with probably damaging as being the most confident prediction as compared to other two. These predictions are based on position specific independent count score difference, where score 1 is considered the most damaging. The Phd-SNP predicted 95 nsSNPs as diseased while SNPs&GO revealed the most unique results showing only 18 nsSNPs as diseased (Fig 3, S2 Table).

We shortlisted those nsSNPs which are common in at least 4 of these algorithmic tools, and also which scored 0 in SIFT and 1 in PolyPhen-2 so that only the highly deleterious SNPs would be analyzed. Total 14 nsSNPs out of 275 met the criteria and we classified them as high risk. Interestingly, 9 of these nsSNPs lie in the only RhoGAP domain of TAGAP protein (Table 1) and all additional investigations were held for only these 9 nsSNPs.

Functional and structural modifications of TAGAP predicted by MutPred

The shortlisted 9 nsSNPs were submitted to this server and the resultant probability scores are given in Table 2. The structural and functional alterations predicted include loss of disorder, catalytic residue, glycosylation and gain of phosphorylation, solvent accessibility, ubiquitination and molecular recognition features (MoRF) binding. Their P values are provided in S3 Table. According to these predictions, it can be stated that several nsSNPs might be the reason behind any possible structural and functional modifications of TAGAP protein.

Stability modification prediction

We predicted any stability alterations in the TAGAP protein with the help of I-Mutant which completes this task by considering the single-site mutations [21,32]. The 9 nsSNPs that have been found in the RhoGAP domain were submitted to I-Mutant 2.0 server to predict their RI and free energy change values. Results revealed that all these nsSNPs decrease stability of TAGAP protein (Table 3). Hence these polymorphisms in the RhoGAP domain might cause maximum damage to the protein by affecting its stability. According to some studies, decreased protein stability causes increase in degradation, misfolding and aggregation of proteins [49–51].

Table 1. High risk nsSNPs identified by five *in silico* programs.

Amino acid Change	SIFT		PROVEAN		PolyPhen-2 (HumDiv)		PhD-SNP		SNPs&GO	
	Pred	TI	Sc	(cutoff = -2.5)	Effect	Sc	Pred	RI	Pred	RI
C94R	Intolerant	0	-11.11	Deleterious	Pro-damg	1	Diseased	7	Neutral	1
L100F*	Intolerant	0	-3.769	Deleterious	Pro-damg	1	Diseased	2	Neutral	7
T118M*	Intolerant	0	-5.67	Deleterious	Pro-damg	1	Diseased	4	Neutral	4
G120E*	Intolerant	0	-7.671	Deleterious	Pro-damg	1	Diseased	8	Diseased	4
F122L*	Intolerant	0	-5.753	Deleterious	Pro-damg	1	Diseased	7	Diseased	1
A126T*	Intolerant	0	-3.269	Deleterious	Pro-damg	1	Diseased	2	Neutral	8
E136K*	Intolerant	0	-3.702	Deleterious	Pro-damg	1	Diseased	0	Neutral	7
G141W*	Intolerant	0	-7.671	Deleterious	Pro-damg	1	Diseased	6	Diseased	0
V151M*	Intolerant	0	-2.596	Deleterious	Pro-damg	1	Diseased	2	Neutral	3
N205S*	Intolerant	0	-4.878	Deleterious	Pro-damg	1	Diseased	5	Diseased	3
S476F	Intolerant	0	-4.967	Deleterious	Pro-damg	1	Diseased	1	Neutral	8
S490P	Intolerant	0	-2.6	Deleterious	Pro-damg	1	Diseased	1	Diseased	2
F511S	Intolerant	0	-3.352	Deleterious	Pro-damg	1	Diseased	7	Neutral	2
F718S	Intolerant	0	-5.07	Deleterious	Pro-damg	1	Diseased	8	Diseased	3

* shows the positions in the RhoGAP domain region of the TAGAP protein,

Pred = prediction, Sc = score, Pro-damg = probably damaging.

<https://doi.org/10.1371/journal.pone.0188143.t001>

Conservation profile of deleterious nsSNPs in TAGAP

Evolutionary information is essential to detect mutations which might affect human health [52]. Using ConSurf web server, we calculated the evolutionary conservation of amino acid residues of TAGAP protein to further explore the possible effects of 9 most deleterious nsSNPs. Results were obtained in the form of structural representation of the protein (S1 Fig). ConSurf identifies structural and functional residues by combining evolutionary conservation data with solvent accessibility predictions. Highly conserved residues are predicted as either functional or structural based on their location either on protein surface or inside its core [33]. Amino acids which are involved in vital biological processes for example in interactions among different proteins, are more conserved than others. Taking this into consideration, those nsSNPs which are located at these conserved regions are considered immensely damaging to protein as compared to those at non-conserved sites [9,53].

Results obtained via ConSurf represented all residues of TAGAP showing their structural and functional conservation levels. But we focused only on those residues which matched their positions with 9 high risk nsSNPs which we have identified. The results predicted L100, T118, E136, G141, V151 and N205 as functional residues making them highly conserved and exposed. While G120, F122 and A126 are predicted as structural residues which mean that they are highly conserved and buried (Table 4). The results further confirmed these 9 high risk nsSNPs as being really deleterious to the structure and/or function of TAGAP protein.

Table 2. Probability scores of deleterious mutations.

Mutation	P-value	Mutation	P-value
T118M	0.618	G141W	0.663
L100F	0.573	V151M	0.676
F122L	0.846	A126T	0.804
G120E	0.902	E136K	0.498
N205S	0.896		

<https://doi.org/10.1371/journal.pone.0188143.t002>

Table 3. I-MUTANT 2.0 and TM-align predictions for nsSNPs in RhoGAP domain of TAGAP.

nsSNP ID	Amino Acid Change	Stability	RI	DDG	TM-Score	RMSD
rs748659041	100, L → F	Decrease	9	-0.62	1	0
rs368265576	118, T → M	Decrease	6	-0.29	0.98778	0.83
rs764717611	120, G → E	Decrease	2	-0.84	0.78894	1.98
rs763380333	122, F → L	Decrease	3	-0.6	0.98778	0.83
rs780953963	126, A → T	Decrease	7	-1.04	0.7912	1.87
rs866898464	136, E → K	Decrease	7	-1.12	0.98778	0.83
rs765146154	141, G → W	Decrease	7	-0.58	0.78894	1.98
rs777042268	151, V → M	Decrease	9	-1.53	0.78851	1.99
rs778438807	205, N → S	Decrease	2	-2.45	0.7928	1.84

DDG: free energy change value. 0.0 < TM-score < 0.30, random structural similarity 0–0.3 and 0.5 < TM-score < 1.00, in about the same fold 0.5–1.

<https://doi.org/10.1371/journal.pone.0188143.t003>

Comparative modeling of wild type TAGAP and its mutants

To determine whether the 9 high risk nsSNPs alter the wild type structure of TAGAP protein, we first used Phyre2 to generate 3D structures of wild type protein and 9 mutants. Each nsSNP was individually substituted into the wild type sequence of TAGAP and the sequences were submitted to Phyre2 homology modeling tool [37]. Phyre2 used c1xa6A as the template for predicting 3D models and the structures were then visualized by Chimera 1.11 [41]. We extended our analysis by calculating the TM-scores and RMSD values for each mutant model. The TM-score is used to evaluate the topological similarity between wild type and mutant models, while RMSD helps in measuring average distance between α -carbon backbones of wild type and mutant models [43–44]. The greater the RMSD value the greater is the deviation of mutant structure from that of the wild type (Table 3). The mutant model for V151M showed the maximum RMSD value of 1.99 followed by those of G120E and G141W having RMSD value 1.98. Interestingly, the mutant model L100F showed no variation from wild type structure as depicted by its RMSD value 0. The nsSNP models of T118M, F122L and E136K showed very slight variation from wild type TAGAP protein model (RMSD = 0.83).

Based on higher RMSD values, we finally selected only three mutants; G120E, G141W and V151M to remodel them using I-TASSER (being the most advanced modeling tool) [38–41] so that we could come up with the most reliable protein structures. The templates used by this server were: 5c5sA, 3cxlA, 3flc2A, 3msxB and 5ircA. I-TASSER produced 5 models each for

Table 4. ConSurf predictions showing conservation profile of amino acids in TAGAP.

SNP ID	Residue & Position	Conservation Score	Prediction
rs748659041	L100	8	Highly conserved and exposed (f)
rs368265576	T118	9	Highly conserved and exposed (f)
rs764717611	G120	9	Highly conserved and buried (s)
rs763380333	F122	9	Highly conserved and buried (s)
rs780953963	A126	9	Highly conserved and buried (s)
rs866898464	E136	9	Highly conserved and exposed (f)
rs777042268	G141	8	Highly conserved and exposed (f)
rs778438807	V151	9	Highly conserved and exposed (f)
rs765146154	N205	9	Highly conserved and exposed (f)

(f): predicted functional residue, (s): predicted structural residue.

<https://doi.org/10.1371/journal.pone.0188143.t004>

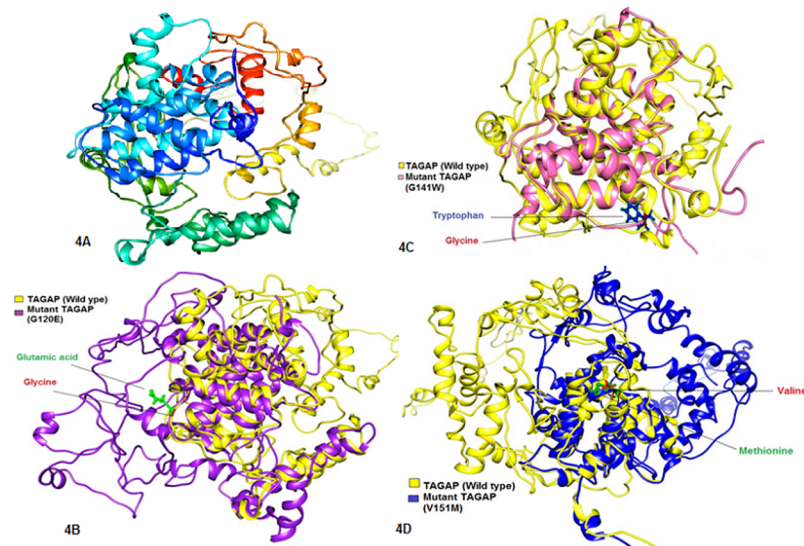


Fig 4. Comparison of wild type TAGAP protein structure with its mutant forms. (A) 3D model of wild type TAGAP protein. (B) Superimposed structures of wild type TAGAP protein and its mutant having mutation from Glycine to Glutamic Acid at position 120. (C) Superimposed structures of wild type TAGAP protein and its mutant having mutation from Glycine to Tryptophan at position 141. (D) Superimposed structures of wild type TAGAP protein and its mutant having mutation from Valine to Methionine at position 151.

<https://doi.org/10.1371/journal.pone.0188143.g004>

TAGAP and its mutants. We selected only that model having minimum C-score and also had higher ERRAT values when submitted to ERRAT program (G120E; C-score = -2.95, ERRAT = 84, G141W; C-score = -3.01, ERRAT = 91.6, V151M; C-score = -2.95, ERRAT = 89.4). Hence these three mutant models were finally superimposed over the wild type protein model (Fig 4). Similar studies have been carried out on various genes and proteins like GDH protein, MBL2 gene etc using different bioinformatics tools [54,55]. Such studies can offer novel therapeutic markers for a range of diseases.

Predicted post translational modifications

PTMs are important in regulating structures and functions of proteins hence are involved in many biological events, for example, protein-protein interactions and cell signaling etc [56,57]. In this study, we sought to investigate whether the high risk nsSNPs have any effect on PTMs in TAGAP. For this, we used various *in silico* tools to predict probable PTM sites in the TAGAP protein.

Methylation. Methylation of lysine residues in certain histones effect their binding with the neighboring DNA and this alters the expression of genes on that DNA. The PSSMe tool predicted a total of 21 lysine residue sites that can get methylated, whereas BPB-PPMS predicted only 6 residues that undergo methylation. Only 3 lysine residues at positions 195, 256 and 517 were common findings of both PSSMe and BPB-PPMS tools (Fig 5).

Phosphorylation. Phosphorylation of proteins is an important regulatory mechanism as it acts as their molecular switch to perform various functions like conformational changes in protein structure, during signal transduction pathways, activating some proteins and deactivating others [58–61]. The putative phosphorylation sites in the TAGAP were predicted by using NetPhos 3.1 and GPS 3.0 servers which made predictions for 17 and 3 kinases respectively (S4 Table). NetPhos 3.1 predicted total 97 residues (Ser:76, Thr:18, Tyr:3) having potential of getting phosphorylated. On the contrary, GPS 3.0 predicted only 10 residues, of which 9

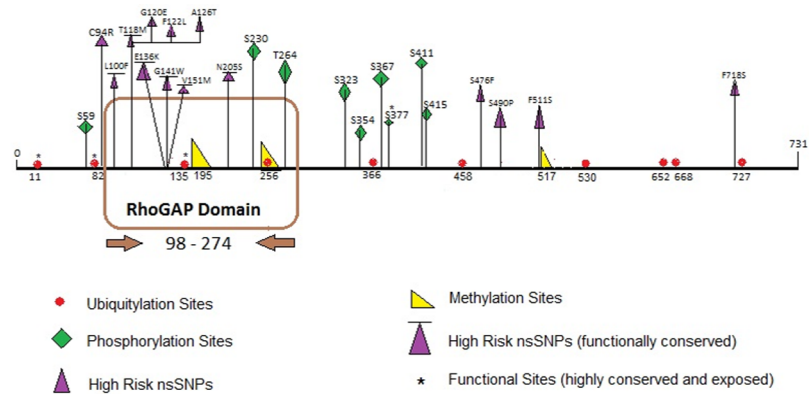


Fig 5. Putative PTM sites and high risk nsSNPs in TAGAP protein. Schematic illustration of locations of high risk nsSNPs and PTMs (ubiquitylation, phosphorylation and methylation) in TAGAP protein. The nsSNPs that are also predicted as functionally conserved by ConSurf are all present in the RhoGAP domain of the protein starting from L100F to N205S. The site K256 is predicted to undergo both methylation and ubiquitylation.

<https://doi.org/10.1371/journal.pone.0188143.g005>

were serine specific sites and 1 was threonine specific phosphorylation site. No phosphorylation at tyrosine residues was predicted by GPS 3.0. The common sites predicted by both servers are shown in Fig 5. According to ConSurf results, only S377 of these sites is a highly conserved and exposed residue making it a functional residue hence depicting its significance.

Ubiquitylation. Ubiquitylation is a PTM which serves as a degradation mechanism for proteins and helps in DNA damage repair [62]. These modification sites in the TAGAP were predicted by UbPred and BDM-PUB tools. UbPred predicted 12 lysine residues in the TAGAP protein that experience ubiquitylation while BDM-PUB predicted that 33 lysine residues undergo ubiquitylation. Total 10 residues were common predictions by both UbPred and BDM-PUB (Fig 5). Only 3 of these putative ubiquitylation sites; K11, K82 and K135 were predicted to be important functional residues (highly conserved and exposed) according to the ConSurf results (S1 Fig). All putative PTM sites in TAGAP protein along with the high risk nsSNPs identified in this study are illustrated in Fig 5 while the results of all the PTMs are provided in S4 Table.

A few of these PTMs also coincided in position with nsSNPs (S1 Table) in the TAGAP, i.e. K82, S367 and S411 (Table 5) and of these, K82 is highly conserved among TAGAP homologues according to the ConSurf results.

Though the upshots of TAGAP methylation, phosphorylation and ubiquitylation have not been reported yet, various studies have shown that these modifications can significantly alter the protein function by varying its location, stability or inter-protein interactions etc. It is possible that ubiquitylation of lysine residues in the TAGAP at sites K52, K82, K114 and K518 and phosphorylation of serine residues at sites S367 and S411 are vital for some of the protein's

Table 5. Low risk nsSNPs identified considering PTMs and ConSurf predictions.

SNP ID	Mutation	Deleterious Predictions	PTM	ConSurf Prediction
rs375785212	K82N	3	Ubiquitylation	F
rs776525307	S367I	4	Phosphorylation	E
rs182059529	S411C	1	Phosphorylation	E

e: exposed residue (highly conserved and buried), f: functional residue (highly conserved and exposed).

<https://doi.org/10.1371/journal.pone.0188143.t005>

essential functions, and that the missense SNPs: K52T, K82N, K114E, K518T, S367I and S411C somehow impair those functions. Conversely, these nsSNPs might also destabilize the protein that might eventually enhance the harms of PTM impairment.

Conclusions

This study suggests that structure and/or function of *TAGAP* protein can be disturbed by various nsSNPs. In native protein of *TAGAP* gene, three major mutations found were: Glycine → Glutamic Acid at position 120 ([rs764717611](#)), Glycine → Tryptophan at position 141 ([rs777042268](#)) and Valine → Methionine at position 151 ([rs778438807](#)). These mutations occur in RhoGAP domain of *TAGAP* protein hence are of particular concern as this is the only functional domain of the protein. Therefore, these nsSNPs can be strongly considered as key candidates in causing diseases related to *TAGAP* malfunction and hence will help in effective drug discovery and developing precision medicines. Thorough investigations and wet lab experimentation are needed to explore the effects of these polymorphisms on structure and function of the protein. Also, various diseased animal models comprising these major mutations in the *TAGAP* protein might be very supportive in exploring their job in the disease.

Supporting information

S1 Fig. ConSurf prediction showing conservation profile of amino acids in *TAGAP* protein.

(PDF)

S1 Table. All 275 nsSNPs including their ID, allele change, protein accession number, position, amino acid change and global maf.

(DOCX)

S2 Table. The results for all 275 nsSNPs by five *in silico* tools.

(DOCX)

S3 Table. Effects of nsSNPs on structural & functional properties of *TAGAP* by MutPred server.

(DOCX)

S4 Table. Post translational modifications.

(DOCX)

Acknowledgments

The authors are grateful to Dr. Amjad Ali from Atta-Ur-Rahman School of Applied Biosciences, National University of Sciences and Technology, for his guidance in this work.

Author Contributions

Conceptualization: Maria Arshad.

Data curation: Maria Arshad.

Methodology: Maria Arshad.

Supervision: Attya Bhatti.

Validation: Attya Bhatti.

Writing – original draft: Maria Arshad.

Writing – review & editing: Attya Bhatti, Peter John.

References

1. Collins FS, Brooks LD, Chakravarti A. A DNA polymorphism discovery resource for research on human genetic variation. *Genome Res.* 1998; 8: 1229–1231. PMID: [9872978](#)
2. Lander ES. The new genomics: global views of biology. *Science.* 1996; 274: 536–539. PMID: [8928008](#)
3. Capriotti E, Altman RB. Improving the prediction of disease-related variants using protein three-dimensional structure. *BMC Bioinformatics.* 2011; 12: S3. <https://doi.org/10.1186/1471-2105-12-S4-S3> PMID: [21992054](#)
4. Barroso I, Gurnell M, Crowley VE, Agostini M, Schwabe JW, Soos MA, et al. Dominant negative mutations in human PPARgamma associated with severe insulin resistance, diabetes mellitus and hypertension. *Nature.* 1999; 402: 880–883. <https://doi.org/10.1038/47254> PMID: [10622252](#)
5. Petukh M, Kucukkal TG, Alexov E. On human disease-causing amino acid variants: statistical study of sequence and structural patterns. *Hum Mutat.* 2015. 36; 524–534. <https://doi.org/10.1002/humu.22770> PMID: [25689729](#)
6. Chasman D, Adams RM. Predicting the functional consequences of nonsynonymous single nucleotide polymorphisms: structure-based assessment of amino acid variation. *J Mol Biol.* 2001; 307: 683–706. <https://doi.org/10.1006/jmbi.2001.4510> PMID: [11254390](#)
7. Kucukkal TG, Petukh M, Li L, Alexov E. Structural and physico-chemical effects of disease and non-disease nsSNPs on proteins. *Curr Opin Struct Biol.* 2015; 32: 18–24. <https://doi.org/10.1016/j.sbi.2015.01.003> PMID: [25658850](#)
8. Thomas R, McConnell R, Whittaker J, Kirkpatrick P, Bradley J, Sandford R. Identification of mutations in the repeated part of the autosomal dominant polycystic kidney disease type 1 gene, PKD1, by long-range PCR. *Am J Hum Genet.* 1999; 65: 39–49. <https://doi.org/10.1086/302460> PMID: [10364515](#)
9. Doniger SW, Kim HS, Swain D, Corcuera D, Williams M, Yang SP, Fay JC. A catalog of neutral and deleterious polymorphism in yeast. *PLoS Genet.* 2008; 4: e1000183. <https://doi.org/10.1371/journal.pgen.1000183> PMID: [18769710](#)
10. Radivojac P, Vacic V, Haynes C, Cocklin RR, Mohan A, Heyen JW, et al. Identification, Analysis and Prediction of Protein Ubiquitination Sites. *Proteins.* 2010; 78: 365–380. <https://doi.org/10.1002/prot.22555> PMID: [19722269](#)
11. Begovich AB, Carlton VE, Honigberg LA, Schrodi SJ, Chokkalingam AP, Alexander HC, et al. A missense single-nucleotide polymorphism in a gene encoding a protein tyrosine phosphatase (PTPN22) is associated with rheumatoid arthritis. *Am J Hum Genet.* 2004; 75: 330–337. <https://doi.org/10.1086/422827> PMID: [15208781](#)
12. Sobieszczyk ME, Lingappa JR, McElrath MJ. Host genetic polymorphisms associated with innate immune factors and HIV-1. *Curr Opin HIV AIDS.* 2011; 6: 427–434. <https://doi.org/10.1097/COH.0b013e3283497155> PMID: [21734565](#)
13. Azad AK, Sadee W, Schlesinger LS. Innate immune gene polymorphisms in tuberculosis. *Infect Immun.* 2012; 80: 3343–3359. <https://doi.org/10.1128/IAI.00443-12> PMID: [22825450](#)
14. Santana-de Anda K, Gómez-Martín D, Díaz-Zamudio M, Alcocer-Varela J. Interferon regulatory factors: beyond the antiviral response and their link to the development of autoimmune pathology. *Autoimmun Rev.* 2011; 11: 98–103. <https://doi.org/10.1016/j.autrev.2011.08.006> PMID: [21872684](#)
15. Heim MH. Innate immunity and HCV. *J Hepatol.* 2013; 58: 564–574. <https://doi.org/10.1016/j.jhep.2012.10.005> PMID: [23063572](#)
16. Marín-Martín FR, Soler-Rivas C, Martín-Hernández R, Rodríguez-Casado A. A Comprehensive In Silico Analysis of the Functional and Structural Impact of Nonsynonymous SNPs in the ABCA1 Transporter Gene. *Cholesterol.* 2014; 639751. <https://doi.org/10.1155/2014/639751> PMID: [25215231](#)
17. Naveed M, Tehreem S, Mubeen S, Nadeem F, Zafar F, Irshad M. In-silico analysis of non-synonymous SNPs of STEAP2: To provoke the progression of prostate cancer. *Open Life Sciences.* 2016; 11: 402–416.
18. Samad FA, Suliman BA, Basha SH, Manivasagam T, Essa MM. A Comprehensive In Silico Analysis on the Structural and Functional Impact of SNPs in the Congenital Heart Defects Associated with NKX2-5 Gene—A Molecular Dynamic Simulation Approach. *PLoS One.* 2016; 11: e0153999. <https://doi.org/10.1371/journal.pone.0153999> PMID: [27152669](#)
19. Naveed M, Anwar F, Kazmi SK, Tariq F, Tehreem S, Abbas G, et al. In Silico Screening and Pathway Analysis of Disease-Associated nsSNPs of MITF Gene: A study on Melanoma. *IJCSIS.* 2017; 15: 31–54.

20. Patel K, Stein MK, Morris LK, Byrd KP, Weksler B, Schwartzberg LS, et al. In silico analysis of non-synonymous SNPs (nsSNPs) and outcomes in non-small cell lung cancer (NSCLC) patients (pts) treated with immunotherapy (IT). *J Clin Oncol*. 2017; 35 (suppl; abstr e14563).
21. Connelly TM, Berg AS, Harris LR, Hegarty JP, Ruggiero FM, Deiling SM, et al. T cell activation Rho GTPase-activating protein expression varies with inflammation location and severity in Crohn's disease. *J Surg Res*. 2014; 190: 457–464. <https://doi.org/10.1016/j.jss.2014.01.019> PMID: 24582067
22. Mao M, Biery MC, Kobayashi SV, Ward T, Schimmack G, Burchard J, et al. T lymphocyte activation gene identification by coregulated expression on DNA microarrays. *Genomics*. 2004; 83: 989–999. <https://doi.org/10.1016/j.ygeno.2003.12.019> PMID: 15177553
23. Bhagwat M. Searching NCBI's dbSNP database. *Curr Protoc Bioinformatics*, Chapter 1: Unit 1. 2010; 19. <https://doi.org/10.1002/0471250953.bi0119s32> PMID: 21154707
24. Ng PC, Henikoff S. Predicting the Effects of Amino Acid Substitutions on Protein Function. *Annu Rev Genomics Hum Genet*. 2006; 7: 61–80. <https://doi.org/10.1146/annurev.genom.7.080505.115630> PMID: 16824020
25. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc*. 2009; 4: 1073–1081. <https://doi.org/10.1038/nprot.2009.86> PMID: 19561590
26. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the Functional Effect of Amino Acid Substitutions and Indels. *PLoS One*. 2012; 7: e46688. <https://doi.org/10.1371/journal.pone.0046688> PMID: 23056405
27. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. *Nat Methods*. 2010; 7: 248–249. <https://doi.org/10.1038/nmeth0410-248> PMID: 20354512
28. Capriotti E, Calabrese R, Casadio R. Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. *Bioinformatics*. 2006; 22: 2729–2734. <https://doi.org/10.1093/bioinformatics/btl423> PMID: 16895930
29. Capriotti E, Calabrese R, Fariselli P, Martelli PL, Altman RB, Casadio R. WS-SNPs&GO: a web server for predicting the deleterious effect of human protein variants using functional annotation. *BMC Genomics*. 2013; 3: S6. <https://doi.org/10.1186/1471-2164-14-S3-S6> PMID: 23819482
30. Li B, Krishnan VG, Mort ME, Xin F, Kamati KK, Cooper DN, et al. Automated inference of molecular mechanisms of disease from amino acid substitutions. *Bioinformatics*. 2009; 25: 2744–2750. <https://doi.org/10.1093/bioinformatics/btp528> PMID: 19734154
31. Bava KA, Gromiha MM, Uedaira H, Kitajima K, Sarai A. ProTherm, version 4.0: thermodynamic database for proteins and mutants. *Nucleic Acids Res*. 2004; 32: D120–121. <https://doi.org/10.1093/nar/gkh082> PMID: 14681373
32. Capriotti E, Fariselli P, Casadio R. I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Res*. 2005; 33: W306–310. <https://doi.org/10.1093/nar/gki375> PMID: 15980478
33. Berezin C, Glaser F, Rosenberg J, Paz I, Pupko T, Fariselli P, et al. ConSeq: The Identification of Functionally and Structurally Important Residues in Protein Sequences. *Bioinformatics*. 2004; 20: 1322–1324. <https://doi.org/10.1093/bioinformatics/bth070> PMID: 14871869
34. Ashkenazy H, Erez E, Martz E, Pupko T, Ben-Tal N. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res*. 2010; 38: W529–533. <https://doi.org/10.1093/nar/gkq399> PMID: 20478830
35. Celniker G, Nimrod G, Ashkenazy H, Glaser F, Martz E, Mayrose I, et al. ConSurf: Using Evolutionary Data to Raise Testable Hypotheses about Protein Function. *Israel Journal Of Chemistry*. 2013; 53: 199–206. <https://doi.org/10.1002/ijch.201200096>
36. Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, Pupko T, et al. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res*. 2016; 8: W344–350. <https://doi.org/10.1093/nar/gkw408> PMID: 27166375
37. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc*. 2015; 10: 845–858. <https://doi.org/10.1038/nprot.2015.053> PMID: 25950237
38. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. The I-TASSER Suite: Protein structure and function prediction. *Nature Methods*. 2015; 12: 7–8. <https://doi.org/10.1038/nmeth.3213> PMID: 25549265
39. Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nature Protocols*. 2010; 5: 725–738. <https://doi.org/10.1038/nprot.2010.5> PMID: 20360767

40. Zhang Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*. 2008; 9:40. <https://doi.org/10.1186/1471-2105-9-40> PMID: 18215316
41. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem*. 2004; 13:1605–1612.
42. Colovos C, Yeates T. ERRAT: an empirical atom-based method for validating protein structures. *Protein Sci*. 1993; 2: 1511–1519.
43. Carugo O, Pongor S. A normalized root-mean-square distance for comparing protein three-dimensional structures. *Protein Sci*. 2001; 10: 1470–1473. <https://doi.org/10.1110/ps.690101> PMID: 11420449
44. Zhang Y, Skolnick J. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res*. 2005; 33: 2302–2309. <https://doi.org/10.1093/nar/gki524> PMID: 15849316
45. Wen PP, Shi SP, Xu HD, Wang LN, Qiu JD. Accurate in silico prediction of species-specific methylation sites based on information gain feature optimization. *Bioinformatics*. 2016; 32: 3107–3115. <https://doi.org/10.1093/bioinformatics/btw377> PMID: 27354692
46. Shao J, Xu D, Tsai SN, Wang Y, Ngai SM. Computational identification of protein methylation sites through bi-profile Bayes feature extraction. *PLoS One*. 2009; 4: e4920. <https://doi.org/10.1371/journal.pone.0004920> PMID: 19290060
47. Blom N, Gammeltoft S, Brunak S. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J Mol Biol*. 1999; 294: 1351–1362. <https://doi.org/10.1006/jmbi.1999.3310> PMID: 10600390
48. Xue Y, Ren J, Gao X, Jin C, Wen L, Yao X. GPS 2.0, a tool to predict kinase-specific phosphorylation sites in hierarchy. *Mol Cell Proteomics*. 2008; 7: 1598–1608. <https://doi.org/10.1074/mcp.M700574-MCP200> PMID: 18463090
49. Du K, Sharma M, Lukacs GL. The DeltaF508 cystic fibrosis mutation impairs domain-domain interactions and arrests post-translational folding of CFTR. *Nat Struct Mol Biol*. 2005; 12: 17–25. <https://doi.org/10.1038/nsmb882> PMID: 15619635
50. Mayer S, Rüdiger S, Ang HC, Joerger AC, Fersht AR. Correlation of levels of folded recombinant p53 in *Escherichia coli* with thermodynamic stability in vitro. *J Mol Biol*. 2007; 372: 268–276. <https://doi.org/10.1016/j.jmb.2007.06.044> PMID: 17631895
51. Singh SM, Kongari N, Cabello-Villegas J, Mallela KM. Missense mutations in dystrophin that trigger muscular dystrophy decrease protein stability and lead to cross-beta aggregates. *Proc Natl Acad Sci*. 2010; 107: 15069–15074. <https://doi.org/10.1073/pnas.1008818107> PMID: 20696926
52. Ramensky V, Bork P, Sunyaev S. Human non-synonymous SNPs: Server and survey. *Nucleic Acids Res*. 2002; 30: 3894–3900. PMID: 12202775
53. Miller MP, Kumar S. Understanding human disease mutations through the use of interspecific genetic variation. *Hum Mol Genet*. 2001; 10: 2319–2328. PMID: 11689479
54. Naveed M, Ahmed I, Khalid N, Mumtaz AS. Bioinformatics based structural characterization of glucose dehydrogenase (gdh) gene and growth promoting activity of *Leclercia* sp. QAU-66. *Braz J Microbiol*. 2014; 45: 603–611. PMID: 25242947
55. Kalia N, Sharma A, Kaur M, Kamboj SS, Singh J. A comprehensive in silico analysis of non-synonymous and regulatory SNPs of human *MBL2* gene. *Springerplus*. 2016; 5: 811. <https://doi.org/10.1186/s40064-016-2543-4> PMID: 27390651
56. Dai C, Gu W. p53 post-translational modification: deregulated in tumorigenesis. *Trends Mol Med*. 2010; 16: 528–536. <https://doi.org/10.1016/j.molmed.2010.09.002> PMID: 20932800
57. Shiloh Y, Ziv Y. The ATM protein kinase: regulating the cellular response to genotoxic stress, and more. *Nat Rev Mol Cell Biol*. 2013; 14: 197–210.
58. Deutscher J, Saier MH. Ser/Thr/Tyr protein phosphorylation in bacteria—for long time neglected, now well established. *J Mol Microbiol Biotechnol*. 2005; 9: 125–131. <https://doi.org/10.1159/000089641> PMID: 16415586
59. Puttick J, Baker EN, Delbaere LT. Histidine phosphorylation in biological systems. *Biochim Biophys Acta*. 2008; 1784: 100–105. <https://doi.org/10.1016/j.bbapap.2007.07.008> PMID: 17728195
60. Ciesla J, Fraczyk T, Rode W. Phosphorylation of basic amino acid residues in proteins: important but easily missed. *Acta Biochim Pol*. 2011; 58: 137–147. PMID: 21623415
61. Sawicka A, Seiser C. Sensing core histone phosphorylation—a matter of perfect timing. *Biochim Biophys Acta*. 2014; 1839: 711–718. <https://doi.org/10.1016/j.bbagrm.2014.04.013> PMID: 24747175
62. Gallo LH, Ko J, Donoghue DJ. The importance of regulatory ubiquitination in cancer and metastasis. *Cell Cycle*. 2017; <https://doi.org/10.1080/15384101.2017.1288326> PMID: 28166483