

Identification of Most Lethal Non-synonymous SNPs in Tollip *gene*- An In-Silico Analysis

Usman Ghani

Center for Applied Molecular Biology and Forensic University of Punjab Lahore

Rahmat Ullah

Center for Applied Molecular Biology University of Punjab Lahore

Tayyaba Anjum

Center for Applied Molecular Biology University of Punjab Lahore

Yasir Ali

Abdul Wali Khan University Mardan

Muhammad Farooq Sabar (✉ farooq.camb@pu.edu.pk)

Center for Applied molecular Biology, University of Punjab Lahoure

Shehryar Khan

Abdul Wali khan University Mardan

Iqra Zulfiqar

Center for Applied Molecular Biology University of Punjab Lahore

Saddam Hussain

Shaheed Benazir Bhutto University Sheringal

Research Article

Keywords: Toll-interacting protein (TOLLIP),TOLLIP lethal variants, structural and functional variants, Interleukin.

Posted Date: September 23rd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-488110/v4>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

The Toll-interacting protein (TOLLIP), first detected by hybrid screening using Interleukin-1 receptor accessory protein in 2000, is ubiquitous and its TLR signaling cascade gets negatively regulated by TOLLIP in particular by impeding the TLR4 and TLR2 pathways. Toll-interacting protein facilitates TLR and TGF- β type 1 receptor intracellular localization and lysosomal degradation and exerts its anti-apoptosis and pro-autophagy effects through interaction with a target of Myb1 membrane trafficking protein 1 (TOM1) in autoimmunity. It also protects intestinal epithelial cells from apoptosis induced by (TNF- α) (IFN- γ) signaling and acts as a cargo adaptor linking (ATG8) autophagy gene 8 and microtubule-associated protein 1 light chain 3. Ubiquitin-modified cell debris along with coated autophagosomes removes harmful protein aggregates and maintains cellular homeostasis. It is therefore structurally and functionally important to detect TOLLIP polymorphisms to indicate the possible malfunctions and therapeutics. We have identified the gap of available data on nsSNPs in the TOLLIP gene in previous studies. Hence, We have used a wide range of bioinformatic techniques in this study to identify the most destructive nsSNPs in the TOLLIP gene. The *in-silico* tools such as PROVEAN, SIFT, SNP&GO, PhD SNP, and PolyPhen2 have been used followed by I Mutant, MutPred, and ConSurf. The 3-D mapping was carried out with I-TASSER and Phyre2. Though, STRING and GeneMANIA proposed the gene to gene interaction of TOLLIP. Our study identified G19D (rs866744102), G32R (rs1308704061), D71N (rs777772934), and E72G (rs1202660177) as the four most lethal non-synonymous SNPs in TOLLIP genes, which may play an essential part in defects of TOLLIP Protein and probably cause a different type of diseases. This is the first study of its kind, and it could pave the way in the field of medical drugs in the future. The influence of these nsSNPs in the normal functioning of TOLLIP can also be evaluated in animal models to identify the underlying biochemical pathway involved in diseases.

Introduction

Various types of mutations may occur in the human DNA chain. The single base pair of variations referred to as SNPs (Single Polymorphisms of Nucleotides) which comprises about 90% of them. The coding area of human genomes is known to be about 500,000 SNPs¹. The nsSNPs (non-synonymous SNPs) lead to the anatomical and physiological variation of human proteins². Such nonsynonymous SNPs can affect amino acids that can be helpful or dangerous to the structure and/or function of the protein³. These alterations are associated with changes to gene regulation, protein structure volatility^{4,5}. These changes are related to protein structure instability⁶, hydrophobicity, structure, and loading of protein⁷, translation, relative strength, protein relations, and dynamics (Inter/Intra)⁸. The SNPs in the Toll interacting protein gene have also been described as downregulating interleukin-6 (TNF - α) pro-inflammatory cytokines and causing interleukin (IL-6) anti-inflammatory cytokines, (TNF- α) tumor necrosis factor-alpha and induces the anti-inflammatory cytokines interleukin-10 (IL-10)⁹. TOLLIP gene polymorphisms are linked to infectious diseases because they control the Toll-like receptors, a family of evolutionarily conserved specific receptors that identify and respond to PAMPs formed by certain microbes such as PGN, LPS, CpG-DNA, and single and double-stranded RNA (dsRNA)¹⁰. Thus, TLRs are an important component of innate immune responses, acting as a buffer by initiating signaling cascades to trigger immune and inflammatory genes against invading pathogens. However, disturbed activation of TLR pathways leads to an unnecessary release of signaling molecules, and pro-inflammatory cytokines, these processes may lead to inflammation, development of infectious and autoimmune diseases¹¹. Therefore, the signaling equilibrium of TLR pathways must be carefully maintained to preserve the immune response. TOLLIP became a new component of the IL-1R pathway in mice and then was published in other

species too¹². TOLLIP protein consists of three conserved domains; an N-terminal Tom1 binding domain, a central conserved C2 domain, and a C-terminal Ubiquitin to ER lysosomal degradation domain. In mammals, TOLLIP proteins work by shutting down MyD88-dependent signaling pathways by inactivating IRAK-1 or by explicitly binding to the TLR2 and TLR4¹³. TOLLIP also contains protein that regulates other inflammatory processes. TOLLIP C2 domain can bind with phosphatidic acid and a CUE domain which can interact with ubiquitinated proteins¹⁴. The C2 domain is involved in phagocytosis¹⁵. The JAK/STAT (Janus Kinase/Signal Transducers and Activators of Transcription) pathway is shown to be built around lysosomes¹⁶. C2 domain helps in liberating and transferring the transmembrane proteins from lysosome and Mitochondrial TOLLIP plays a substantial role in stimulating oxygen-derived free radicals¹⁷. TOLLIP by its upregulated expression in the intestine epithelial cells induces tolerance to the normal enteric flora. Failure in the upregulated expression in the TOLLIP gene can lead to chronic inflammation in a patient with inflammatory bowel disease¹⁸. TOLLIP has been associated with atopic dermatitis but the actual mechanism of the disease is still not known¹⁹. TOLLIP has been shown to act as a negative regulator by attenuating the function of IL1 β ²⁰. Studies are expected to be conducted to understand TOLLIP and its functional and structural variants. We used bioinformatics methods to identify the most harmful nsSNPs found in the TOLLIP protein. We suggested 3D versions of wild-type TOLLIP protein and the protein that could be deleterious due to the non-synonymous SNPs on the gene. For TOLLIP, this is the first study of its kind that covers protein structure prediction and mutation analysis in the context of TOLLIP.

Methodology

The study has been completed in several steps. Below is the figure depicting the tools utilized to carry out the work.

Extracting non-synonymous SNPs

The database of Single Nucleotide polymorphism was accessed via the NCBI. Complete information of TOLLIP's gene including its SNPs and data regarding its Minor allele frequencies (MAF), location, and the changes in residues discussed by Bhagwat²¹, out of which the non-synonymous SNPs of the desired gene were extracted.

Identification of Damaging nsSNPs

The effect of spotted nsSNPs on the TOLLIP gene was recognized by employing the *SIFT*²², *PROVEAN*²³, *PhD-SNP*²⁴, and *SNPs, and GO*²⁵ software. The resultant mutual nsSNPs marked as intolerant by all of these tools were then further get filtered by another tool known as *PolyPhen*²⁶.

Examining structural and functional effects

For the purpose of examining the conformational and functional changes of targeted protein due to the presence of nsSNPs, a *Mutpred*²⁷ tool was used. This web application detects the molecular disorder by investigating any amino acid's replacement in the protein structure. The variation in conformation and configuration of protein such as an increase in helical tendency along with any disruption caused by the loss of phosphorylation site in protein's normal function get screened by this server. In this study, we inserted the FASTA sequence of TOLLIP protein and marked the regions where the changes in amino acids were observed (deleterious nsSNPs). The resultant values can be declared as confident and very confident when the probability value ranges from less than 0.05 to less than 0.01, respectively.

Analysis of TOLLIP's stability

To carry out the analysis of targeted TOLLIP's protein's stability, I-Mutant 2.0²⁸ was utilized in the study. Any alterations in the stability of protein due to mutation are estimated by this web server. The sequence of TOLLIP protein was subjected to this tool by setting the conditions of pH as 7 and temperature at 25 °C. The reliability index (RI) exhibited by this server in the presence of any mutant protein normally falls in the range of 0–10.

Identification of Evolutionary Conserved Amino Acids

Phylogenetic association among the homologous sequences of conserved amino acids in TOLLIP protein was identified by the ConSurf^{29,30,31,32}. The degree of evolutionary conservation of estimated amino acids was achieved by using 50 dissimilar homologous sequences. The detectable damaging nsSNPs were further studied which was spotted in affiliation with the highly conserved amino acids.

Predicting 3D TOLLIP's Structure

3-dimensional homology modeling of TOLLIP protein was achieved through Phyre2³³. It developed the four mutants' models because of the toxic non-synonymous SNPs, in addition to the TOLLIP's wild-type. Both wild-type and mutants' models were compared by Template modeling score. The accurate measurement of similarity score was attained by the calculated values of RMSD and TM. Studies have shown that an elevated level of root mean square deviation value is directly related to the enhanced level of alteration between wild-type and mutant^{34,35}. By following their values, the deleterious mutants were then given as input to I-TASSER^{36,37,38}. The molecular features and interactive visualization of the resulting protein structure were examined using Chimera 1.11³⁹.

Estimation of PTM Sites

Post-transcriptional modification in TOLLIP protein such as methylation, phosphorylation, and ubiquitylation was determined by employing (GPS-MPS 3.0)⁴⁰, (MODEPRED and NetPhos 3.1)⁴¹, (BDM-PUB and UbPred)⁴² respectively. The sites of phosphorylation were assessed in the Tyrosine, serine, and threonine amino acid of TOLLIP protein by setting the threshold value of NetPhos to (0.5)⁴³. The elevated values than the standard indicated the presence of a phosphorylation site. Similarly, the standard score for determining ubiquitylation set at 0.62⁴⁴, lysine possess this site or not depends on the score equal/ highest or lower than standard, separately. All the biological events are triggered by the PTMs because these are directly involved in the normal functionality and structure maintenance of protein to carry out processes like interactions among proteins as well as cell signaling cascade^{45,46}. Likewise, whenever a lysine residue of targeted protein gets methylated it ultimately causes the disruption in the binding capacity of DNA which results in the alteration of expression of gene⁴⁷. On the other hand, repairing DNA damage is aided by the ubiquitylation⁴⁸. Similarly, signal transduction pathways also get activated or deactivated when certain amino acid residues get phosphorylated to achieve the desired structural confirmation by phosphorylation^{49,50,51,52}.

Gene-Gene Interaction of TOLLIP

The influence of non-synonymous SNPs on other genes was confirmed by the two famous tools known as STRING and GeneMANIA.^{53,54} The interaction among the genes based on biochemical or signaling pathways, having resemblance in protein's domain analyzed by the GeneMANIA. The work of STRING is to exhibit the combined score of interactive genes. In this step, we provided TOLLIP as our targeted input gene and then carried out the analysis.

Results

Recruited nsSNPs

The SNPs of the TOLLIP gene extracted from the database of NCBI were **11535**, which further comprised of **196** non-synonymous SNPs, **320** positioned in the 5' Untranslated region whereas **836** located in the 3' untranslated region. The remaining 6104 were other forms of SNPs including Intronic SNPs, X synonymous SNPs, intronic SNPs, **198** synonymous SNPs, splice site SNPs, and uncategorized SNPs, their diagrammatical depiction has been shown in Fig. 2. Our target was 196 nsSNPs which were further examined out of this data. The findings of this analysis have shown that 4 nonsynonymous SNPs can shorten the protein by causing variations in it which can stimulate the termination of mRNA's translation known as truncated protein. The S1 table contains information regarding nsSNPs.

Identification of Damaging nsSNPs

To further detect the effect of 196 nsSNPs on the structure as well as on the function of the TOLLIP gene the subsequent bioinformatics tools have presented the given results. The protein variation effect analyzer (PROVEAN) has shown 74 nsSNPs for possessing the damaging impact because the final score of variants came out to be lower than the established value of threshold (-2.5). The findings of another program known as Sorting intolerant from tolerant (SIFT) have shown 118 nsSNPs to be intolerant because the resultant values came out to be lower than the Tolerance Index (0.05). On the other hand, 39 nsSNPs have been characterized as diseased by the derived results of SNPs and GO. We picked 4 those mutual nsSNPs from these tools which exhibited the most deleterious impact and submitted them to the Polymorphism Phenotyping v2 (Polyphen2). This tool is used to predict the impact as probably, possibly damaging as well as benign. The findings from this software have further confirmed the impact of this 4 nsSNPS as the most probable damaging and they were then brought for further analysis.

Table 1
Identification of damaging nsSNPs by PROVEAN, SNP&GO, SIFT, PHD SNP, and Polyphen2

rs ID	Amino Acid change	provean	SNP & Go	Sift		PHD SNP	Polyphen-2
		score	probability	RI	score		
rs866744102	G19D	-6.001	0.730	5	0	5	0.999
rs1308704061	G32R	-7.000	0.780	6	0	2	1.000
rs777772934	D71N	-4.017	0.757	5	0.05	6	0.999
rs1202660177	E72G	-6.617	0.744	5	0	4	0.995

Structural and Functional Effect of nsSNPs Prediction

MutPred server was used for selected 4 nsSNPs. This web application has characterized the nsSNPs that can cause the disruption in the function as well as in the structure of the protein. The p-values derived from this server has shown in Table 2.

Table 2
p-value of the most damaging nsSNPs of TOLLIP gene

Mutation	P-values	Mutation	P-values
G19D	0.607	D71N	0.412
G32R	0.599	E72G	0.571

TOLLIP's Stability Prediction

The neural network-based web-server was used for all the selected nsSNP of TOLLIP protein to predict its stability known as I-Mutant. The stability of protein increased/decreased depends on amino acid substitution and their result was obtained with RI range from 0–10, which are given in the table. In the four selected (rs866744102) G19D (rs1308704061) G32R (rs777772934 D71N) (rs1202660177) E72G all of them have shown decreased stability which may cause greater damaging of TOLLIP protein stability.

Table 3
I-Mutant Outcome for the chosen nsSNPs

SNP id	Amino acid Change	Stability	RI
rs866744102	G19D	D	7
rs1308704061	G32R	D	6
rs777772934	D71N	D	9
rs1202660177	E72G	D	9

Note: The pH was established at 7.0 and Temperature at 25°C.

Evolutionary Conservation of TOLLIP Protein

In order to carry out the determination of functional regions in TOLLIP protein, ConSurf recognized the four nsSNPs. Its findings have estimated (G19D, G32R) and (D71N, E72G) as highly conserved buried and exposed respectively. Hence, its analysis has clarified that for the proper functioning and conformation of TOLLIP protein the non-synonymous SNPs traced at the extremely conserved regions tend to be very harmful.

Table 4
Conservation profiles of the selected nsSNPs

SNP ID	Amino Acid Chang	Conservation Score	Prediction
rs866744102	G19D	9	highly conserved and buried
rs1308704061	G32R	9	highly conserved and buried
rs777772934	D71N	9	highly conserved and exposed(f)
rs1202660177	E72G	9	highly conserved and exposed(f)

3d-modeling of TOLLIP'S Mutants

I-Mutant predicted that four nsSNPs played a role in decreasing the TOLLIP protein's stability and then were selected for the ultimate comparative modeling of protein. The sequences of TOLLIP protein were submitted to I-TASSER with a single amino acid of wild type and mutants to generate TOLLIP protein structure. It is the most reliable and advanced tool for predicting protein structure. 5 models for each TOLLIP mutant and protein were then generated by this method. 2co9 (83% identity) and 2nbiA (85% coverage of the threading alignment) were the templates employed in this method. I-TASSER used ten threading programs to develop 3-D protein structures such as wdPPAS, FFAS-3D, SPARKS-X, pGenTHREADER, HHSEARCH (1 and 2 as well) MUSTER, and Neff-PPAS. The findings derived from I-TASSER were then deferred to the algorithm (TM) to calculate its score. RMSD value of every mutant was also investigated (see Table 6). The visualization of protein's structure and molecular depiction of TOLLIP protein was studied via Chimera 1.11

Table 5: TM-score and RMSD values of 4 particular nsSNPs

SNP ID	Residual Change	TM Score	RMSD Values
rs866744102	G19D	0.82278	3.13
rs1308704061	G32R	0.79103	3.13
rs777772934	D71N	0.63305	3.42
rs1202660177	E72G	0.77515	3.16

0.0 < TM-score < 0.30, random structural similarity 0 ± 0.3 and $0.5 < \text{TM-score} < 1.00$, in about the same fold

0.5 ± 1

Investigated PTMs (Post Transcriptional Modifications)

We investigated the following modifications in this study and their results are as follows:

Methylation:

In the human TOLLIP gene, no site has recognized as methylated by employing the GPS-MPS 3.0 for this objective.

Phosphorylation:

TOLLIP protein phosphorylation sites predicted by **MODEPRED** and **NetPhos 3.1** are as follows:

21 residues of (S:35 %, T:35%, and Y:29%) and 16 residues of (Ser:10%, Thr:9%, Tyr:4%) to carry out the phosphorylation were marked by the MODEPRED and Netphos 3.1 respectively. The outcome from both of these servers has been compared in the following table.

Table 6
TOLLIP protein phosphorylation prediction using MODEPRED and NetPhos 3.1

	NetPhos 3.1		MODEPRED					
	Position	Score	Kinase	Residue	Modification	Score	Confidence	Remarks
Serine (S)	42	0.558	DNAPK	T3	Phosphorylation	0.56	Low	Novel prediction
	48	0.523	unsp	T4	Phosphorylation	0.54	Low	Novel prediction
	118	0.997	unsp	T7	Phosphorylation	0.65	Low	Novel prediction
	118	0.513	cdc2	Y18	Phosphorylation	0.72	Medium	Novel prediction
	133	0.984	unsp	Y26	Phosphorylation	0.53	Low	Novel prediction
	133	0.716	PKC	Y33	Phosphorylation	0.54	Low	Novel prediction
	145	0.517	DNAPK	Y36	Phosphorylation	0.81	High	Novel prediction
	147	0.664	PKC	T38	Phosphorylation	0.70	Medium	Novel prediction
	147	0.553	unsp	Y66	Phosphorylation	0.51	Low	Novel prediction
	163	0.500	DNAPK	S76	Phosphorylation	0.53	Low	Novel prediction
	202	0.561	cdk5	Y102	Phosphorylation	0.61	Low	Novel prediction
	202	0.538	p38MAPK	S103	Phosphorylation	0.52	Low	Novel prediction
	222	0.998	unsp	S160	Phosphorylation	0.53	Low	Novel prediction
	222	0.528	RSK					
	243	0.546	CKII					
Threonine (T)	14	0.632	unsp					
	45	608	unsp					
	45	0.508	PKC					
	49	0.887	unsp					
	49	0.573	PKC					
	49	0.526	DNAPK					
	80	0.545	p38MAPK					

	NetPhos 3.1		MODEPRED
	99	0.509	cdc2
	126	0.661	PKC
	129	0.651	unsp
	183	0.532	PKC
	195	0.514	cdc2
Tyrosine (Y)	68	0.931	unsp
	78	0.505	EGFR
	108	0.525	EGFR
	191	0.556	unsp

Ubiquitylation:

Ubpred as well as BDM-PUB determined the TOLLIP protein's ubiquitylation. 7 sites have been recognized for Ubiquitylation in the TOLLIP's protein's lysine residue.

Table 7
Ubiquitylation sites in TOLLIP protein Prediction using BDM-PUB and UbPred

Peptide	Position	BDM-PUB Score (Threshold 0.3)	UbPred Score (Threshold 0.62)
STQRGPAKLAKNYGM	13	1.37	Not Ubiquitylated
RGPAKLAKNYGMTRM	16	1.33	Not Ubiquitylated
PTAHNGAKNPRWNKV	46	1.34	Not Ubiquitylated
PESLRQGKVEDKWYS	96	0.87	Not Ubiquitylated
RQGKVEDKWYLSGR	100	0.95	yes
SGRQGDDKEGMINLV	112	1.07	Not Ubiquitylated
LEAQRGNKDAAINSL	210	1.98	Not Ubiquitylated

From these findings, it is apparent that phosphorylation is the only modification that can have major effects on the structure and function of the TOLLIP protein.

Gene- Gene Interaction of TOLLIP

The Gene's interactions have been elucidated by the geneMANIA to exhibit the TOLLIP's physical interactions with TOMIL2, IL1RL2, CSTF1, ITCH, IRAK2, IL1RAP, TOM1, IL1R1. TRAF6, SNX8, TOM1L1, IRAK1, OXR1, are co-expressed with TOM1L2 and MYD88. It has not shown co-localization with any gene. In pathways, it has a relation with IRAK2, IL1RAP, IRAK2, IL1RN, EHHADH, MYD88, IL1B, IRAK3, TRAF6, IL1A, IL1R1, IRAK1, and IRAK4. TOLLIP has no genetic interactions with other genes. Predictions resulted from STRING showed that IRAK1 is the most interactive gene with TOLLIP.

Discussion

Numerous studies have been carried out previously to elucidate the association of the TOLLIP gene's polymorphism with many disorders such as Leprosy or Hansen's disease⁵⁵, tuberculosis⁵⁶, Visceral Leishmaniasis⁵⁷, idiopathic pulmonary fibrosis⁵⁸. In this study, we have examined the most damaging non-synonymous SNPs in the TOLLIP gene that may have a critical role in certain diseases.

The results have demonstrated the 196 nsSNPs, the data of other kinds of single nucleotide polymorphism extracted from the employed servers were excluded in this study. The present results have shown that the TOLLIP protein possesses the four most damaging nsSNPs. PROVEAN shows the highest score for D71N and G19D as -4.017 and -6.001 respectively, while for E72G score is -6.617, and the lower score for G32R came out to be -7.000. PolyPhen2, with a scale of 0 to 1 including G19D, G32R, D71N, and E72G scores turn out to be 0.999, 1.000, 0.999 and 0.995 respectively, and all had termed as probable damage by Polyphen2. We accessed Ensemble genome browser 96 to cross-check these nsSNPs via many servers, such as MetaIR, REVEL, Mutation Assessors, and CADD. Mutation Assessor predicted that all four mutations are the most harmful. Likewise, the rest of the tools predicted that all of these 4 nsSNPs are toxic. CADD demonstrated the score of E72G as 34 which is the highest as compared to score of G19D, G32R, and D71N as 26, 26 and 27 respectively (here the score of CADD 30 refers to the 0.1% and 20 to the 1% of the most damaging SNPs in the genome of human). The impact of characteristics like variation in the organized interface, methylation loss, and intensification in acetylation predicted by the MutPred. Out of four mutants, G19D represented 0.607 as the peak *p-value*, whereas three (G32R, D71N, and E72G) mutations demonstrated 0.599, 0.412 and 0.571 as the lowest values. The function and structure of TOLLIP protein get affected by the presence of highlighted nsSNPs in the above results. Though, the particular stability data of TOLLIP protein obtained via I-mutant elucidated that G19D, G32R, D71N, and E72G nsSNPs reduced the stability of TOLLIP protein, See Table 3. ConSurf predicted protein conservation profile. According to Berezin⁵⁹, the highly preserved residues are those which are expected to become essential in their conformational and functional aspect depending on the core and location on the surface of the protein. Likewise, Miller and Kumar⁶⁰ described that amino acids involved in important biological processes are located in the most conserved region. Hence, relating to this information it is now obvious that non-synonymous SNP found on the very conserved regions can cause the maximum impairment to TOLLIP. Out of 4 nsSNPs, G19D and G32R marked on conserved and buried regions which are considered as structurally important while D71N, E72G spotted on highly conserved and exposed sections which are functionally very significant. The adverse effect of these nsSNPs on the TOLLIP protein is further confirmed. I-TASSER was employed to get the demonstration of TOLLIP protein's structure. The FASTA sequence of the protein was utilized as the only source of input. This server is well-developed because it collects the prototypes and initiates the simulation of protein. In order to measure the RAMPAGE value of TOLLIP protein, we made the Ramachandran plot. According to (Morris) If RAMPAGE values exceed 80%, the protein configuration will be known as stronger structures. Hence, for TOLLIP the allowed wild-type RAMPAGE value came out to be 86.8 percent and for outlier residues, it turned out to be 17.4 percent. The mutant nsSNPs such as G19D, G32R, D71N, and E72G depicted the permissible percentage (82.9%, 85.2%, 82.6%, and 88.1%) and outlier values such as (13.1%, 12.6%, 17.4% and 12.9%) correspondingly. Another factor that verifies these four nsSNPs in TOLLIP as the most toxic alterations is the maximum deviation score of RMS (root mean square). Both wild and mutant structures get remarkably dissimilar when the resultant Root mean square deviation value surpasses the 2 angstroms. In the sites of TOLLIP protein PTMs, the likelihood of PTMs at these nsSNP positions was investigated. It is noteworthy that not a single site of phosphorylation and ubiquitylation was spotted in the most dangerous nsSNPs. We searched for another site on nsSNPs that identified by Modepred and NetPhos3.1. Modepred shows at twelve sites expected to be phosphorylated and NetPhos3.1 shows 31 sites which are can be phosphorylated (Table 4). ubiquitylation sites projected by BDM-PUB and UbPred. BDM-PUB both detected 7 sites.

Predictions from STRING and GeneMANIA indicate that IRK1 is the most interactive gene with the TOLLIP gene, which is associated with many diseases and also indicates its importance in many diseases.

Thus, it can be inferred that all 4 of the most destructive nsSNPs in the gene TOLLIP eventually affect and interrupt the regular work of other expressive genes based on their interaction patterns and their co-expression profiled with many diseases such as in pathways, it has relation IRAK2, IL1RAP, IRAK2, IL1RN, EHHADH, MYD88, IL1B, IRAK3, TRAF6, IL1A, IL1R1, IRAK1 and IRAK4 genes which indicates its importance.

Conclusion

In our study, we conclude from the derived findings that, these four selected non-synonymous SNPs can cause damage to the TOLLIP protein. The malfunctioning in the conformation and function of this protein might lead to a variety of diseases. The four major damaging nsSNPs found in this study are rs866744102 (glycine → aspartic acid at position 19), rs1308704061 (glycine → arginine at position 32), rs777772934 (aspartic acid → asparagine at position 71), and rs1202660177 (glutamic acid → glycine at position 72). The input SNP Id of these four non-synonymous SNPs has exhibited decreased stability as well. According to the ConSurf predictions, D71N and E72G have located in highly conserved and buried regions as functional residue, though all these four nsSNPs are present in highly conserved regions indicates their threatening impact on the TOLLIP protein. In this study, we have also analyzed the interaction of the TOLLIP gene with other genes. The results of STRING have revealed that TOLLIP is most interactive with the IRAK1 (interleukin-1 receptor-associated kinase-1) gene. On the other hand, the TOLLIP gene is not co-localized with any gene. Hence, in this study, we have carried out an in-silico analysis of these four most damaging non-synonymous SNPs by using various tools of bioinformatics so that, this derived data may contribute to carrying out the careful investigation that should be done to identify the comprehensive effect of these non-synonymous SNPs on the structure as well as on function of TOLLIP protein.

References

1. Collins FS, Brooks LD, Chakravarti A. A DNA polymorphism discovery resource for research on human genetic variation. *Genome research*. 1998 Dec 1;8(12):1229-31.
2. Lander ES. The new genomics: global views of biology. *Science*. 1996 Oct 25;274(5287):536–9.
3. Capriotti E, Altman RB. Improving the prediction of disease-related variants using protein three-dimensional structure. *BMC bioinformatics*. 2011 Dec;12(4):1–1.
4. 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 Mar 23;307(2):683–706.
5. Kucukkal TG, Petukh M, Li L, Alexov E. Structural and physico-chemical effects of disease and non-disease nsSNPs on proteins. *Current opinion in structural biology*. 2015 Jun 1;32:18–24.
6. Barroso I, Gurnell M, Crowley VE, Agostini M, Schwabe JW, Soos MA, Maslen GL, Williams TD, Lewis H, Schafer AJ, Chatterjee VK. Dominant negative mutations in human PPAR γ associated with severe insulin resistance, diabetes mellitus and hypertension. *Nature*. 1999 Dec;402(6764):880–3.
7. Petukh M, Kucukkal TG, Alexov E. On human disease-causing amino acid variants: Statistical study of sequence and structural patterns. *Human mutation*. 2015 May;36(5):524–34.
8. Thomas R, McConnell R, Whittacker 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. *The*

- American Journal of Human Genetics. 1999 Jul 1;65(1):39–49.
9. Mukherjee S, Huda S, Sinha Babu SP. Toll-like receptor polymorphism in host immune response to infectious diseases: A review. *Scandinavian journal of immunology*. 2019 Jul;90(1):e12771.
 10. Arancibia SA, Beltrán CJ, Aguirre IM, Silva P, Peralta AL, Malinarich F, Hermoso MA. Toll-like receptors are key participants in innate immune responses. *Biological research*. 2007;40(2):97–112.
 11. Gao W, Xiong Y, Li Q, Yang H. Inhibition of toll-like receptor signaling as a promising therapy for inflammatory diseases: a journey from molecular to nano therapeutics. *Frontiers in physiology*. 2017 Jul 19;8:508.
 12. Luiz DP, Júnior CD, Bonetti AM, Brandeburgo MA. Tollip or not Tollip: what are the evolving questions behind it?. *PLoS One*. 2014 May 14;9(5):e97219.
 13. Brissoni B. *Characterization of Tollip in the Interleukin-1 Receptor/Toll Like Receptors signaling pathways* (Doctoral dissertation, Université de Lausanne, Faculté de biologie et médecine).
 14. Kowalski EA. *Toll-Interacting Protein Regulation of Low-grade Non-resolving Inflammation* (Doctoral dissertation, Virginia Tech).
 15. Blasius AL, Beutler B. Intracellular toll-like receptors. *Immunity*. 2010 Mar 26;32(3):305–15.
 16. Casanova JL, Abel L, Quintana-Murci L. Human TLRs and IL-1Rs in host defense: natural insights from evolutionary, epidemiological, and clinical genetics. *Annual review of immunology*. 2011 Apr 23;29:447 – 91.
 17. Li C, Zienkiewicz J, Hawiger J. Interactive sites in the MyD88 Toll/interleukin (IL) 1 receptor domain responsible for coupling to the IL1 β signaling pathway. *Journal of Biological Chemistry*. 2005 Jul 15;280(28):26152–9.
 18. Abreu MT. Toll-like receptor signalling in the intestinal epithelium: how bacterial recognition shapes intestinal function. *Nature Reviews Immunology*. 2010 Feb;10(2):131–44.
 19. Schimming TT, Parwez Q, Petrasch-Parwez E, Nothnagel M, Epplen JT, Hoffjan S. Association of toll-interacting protein gene polymorphisms with atopic dermatitis. *BMC dermatology*. 2007 Dec;7(1):1–8.
 20. Didierlaurent A, Brissoni B, Velin D, Aebi N, Tardivel A, Käslin E, Sirard JC, Angelov G, Tschopp J, Burns K. Tollip regulates proinflammatory responses to interleukin-1 and lipopolysaccharide. *Molecular and cellular biology*. 2006 Feb 1;26(3):735 – 42.
 21. Bhagwat M. Searching NCBI's dbSNP database. *Current protocols in bioinformatics*. 2010 Dec;32(1):1–9.
 22. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature protocols*. 2009 Jul;4(7):1073.
 23. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the functional effect of amino acid substitutions and indels. *PloS one*. 2012 Oct 8;7(10):e46688.
 24. 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 Nov 15;22(22):2729–34.
 25. 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 May;14(3):1–7.
 26. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. A method and server for predicting damaging missense mutations. *Nature methods*. 2010 Apr;7(4):248–9.
 27. Li B, Krishnan VG, Mort ME, Xin F, Kamati KK, Cooper DN, Mooney SD, Radivojac P. Automated inference of molecular mechanisms of disease from amino acid substitutions. *Bioinformatics*. 2009 Nov 1;25(21):2744-50.

28. Capriotti E, Fariselli P, Casadio R. I-Mutant2. 0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic acids research*. 2005 Jul 1;33(suppl_2):W306-10.
29. Berezin C, Glaser F, Rosenberg J, Paz I, Pupko T, Fariselli P, Casadio R, Ben-Tal N. ConSeq: the identification of functionally and structurally important residues in protein sequences. *Bioinformatics*. 2004 May 22;20(8):1322-4.
30. 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 research*. 2010 Jul 1;38(suppl_2):W529-33.
31. Celniker G, Nimrod G, Ashkenazy H, Glaser F, Martz E, Mayrose I, Pupko T, Ben-Tal N. ConSurf: using evolutionary data to raise testable hypotheses about protein function. *Israel Journal of Chemistry*. 2013 Apr;53(3-4):199–206.
32. Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, Pupko T, Ben-Tal N. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic acids research*. 2016 Jul 8;44(W1):W344-50.
33. 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.
34. Carugo O, Pongor S. A normalized root-mean-square distance for comparing protein three-dimensional structures. *Protein science*. 2001 Jul;10(7):1470–3.
35. Zhang Y, Skolnick J. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic acids research*. 2005 Jan 1;33(7):2302-9.
36. Zhang Y. I-TASSER server for protein 3D structure prediction. *BMC bioinformatics*. 2008 Dec;9(1):1–8.
37. Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nature protocols*. 2010 Apr;5(4):725–38.
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 Jan;12(1):7–8.
39. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of computational chemistry*. 2004 Oct;25(13):1605–12.
40. Deng W, Wang Y, Ma L, Zhang Y, Ullah S, Xue Y. Computational prediction of methylation types of covalently modified lysine and arginine residues in proteins. *Briefings in Bioinformatics*. 2017 Jul 1;18(4):647 – 58.
41. 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. *Molecular & cellular proteomics*. 2008 Sep 1;7(9):1598 – 608.
42. Li B, Krishnan VG, Mort ME, Xin F, Kamati KK, Cooper DN, Mooney SD, Radivojac P. Automated inference of molecular mechanisms of disease from amino acid substitutions. *Bioinformatics*. 2009 Nov 1;25(21):2744-50.
43. Blom N, Gammeltoft S, Brunak S. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *Journal of molecular biology*. 1999 Dec 17;294(5):1351–62.
44. Radivojac P, Vacic V, Haynes C, Cocklin RR, Mohan A, Heyen JW, Goebel MG, Iakoucheva LM. Identification, analysis, and prediction of protein ubiquitination sites. *Proteins: Structure, Function, and Bioinformatics*. 2010 Feb 1;78(2):365 – 80.
45. Dai C, Gu W. p53 post-translational modification: deregulated in tumorigenesis. *Trends in molecular medicine*. 2010 Nov 1;16(11):528 – 36.

46. Shiloh Y, Ziv Y. The ATM protein kinase: regulating the cellular response to genotoxic stress, and more. *Nature reviews Molecular cell biology*. 2013 Apr;14(4):197–210.
47. Lanouette S, Mongeon V, Figeys D, Couture JF. The functional diversity of protein lysine methylation. *Molecular systems biology*. 2014 Apr;10(4):724.
48. Gallo LH, Ko J, Donoghue DJ. The importance of regulatory ubiquitination in cancer and metastasis. *Cell Cycle*. 2017 Apr 3;16(7):634–48.
49. Deutscher J, Saier Jr MH. Ser/Thr/Tyr protein phosphorylation in bacteria—for long time neglected, now well established. *Journal of molecular microbiology and biotechnology*. 2005;9(3–4):125–31.
50. Puttick J, Baker EN, Delbaere LT. Histidine phosphorylation in biological systems. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*. 2008 Jan 1;1784(1):100-5.
51. Cieśla J, Frączyk T, Rode W. Phosphorylation of basic amino acid residues in proteins: important but easily missed. *Acta Biochimica Polonica*. 2011 May 27;58(2).
52. Sawicka A, Seiser C. Sensing core histone phosphorylation—A matter of perfect timing. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*. 2014 Aug 1;1839(8):711-8.
53. Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, Franz M, Grouios C, Kazi F, Lopes CT, Maitland A. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic acids research*. 2010 Jul 1;38(suppl_2):W214-20.
54. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic acids research*. 2003 Jul 1;31(13):3784-8.
55. Montoya-Buelna M, Fafutis-Morris M, Tovar-Cuevas AJ, Alvarado-Navarro A, Valle Y, Padilla-Gutierrez JR, Muñoz-Valle JF, Figuera-Villanueva LE. Role of toll-interacting protein gene polymorphisms in leprosy Mexican patients. *BioMed research international*. 2013 Oct;2013.
56. Wu S, Liu X, Chen L, Wang Y, Zhang M, Wang M, He JQ. Polymorphisms of TLR2, TLR4 and TOLLIP and tuberculosis in two independent studies. *Bioscience Reports*. 2020 Aug 28;40(8).
57. Sarmah P, Bharali R, Khatonier R, Khan AM. Polymorphism in Toll interacting protein (TOLLIP) gene and its association with Visceral Leishmaniasis. *Gene Reports*. 2020 Sep 1;20:100705.
58. Bonella F, Campo I, Zorzetto M, Börner E, Ohshimo S, Theegarten D, Taube C, Costabel U. Potential Clinical Utility of MUC5B und TOLLIP Single Nucleotide Polymorphisms (SNPs) in the Management of Patients with IPF.
59. Berezin C, Glaser F, Rosenberg J, Paz I, Pupko T, Fariselli P, Casadio R, Ben-Tal N. ConSeq: the identification of functionally and structurally important residues in protein sequences. *Bioinformatics*. 2004 May 22;20(8):1322-4.
60. Miller MP, Kumar S. Understanding human disease mutations through the use of interspecific genetic variation. *Human molecular genetics*. 2001 Oct 2;10(21):2319-28.

Figures

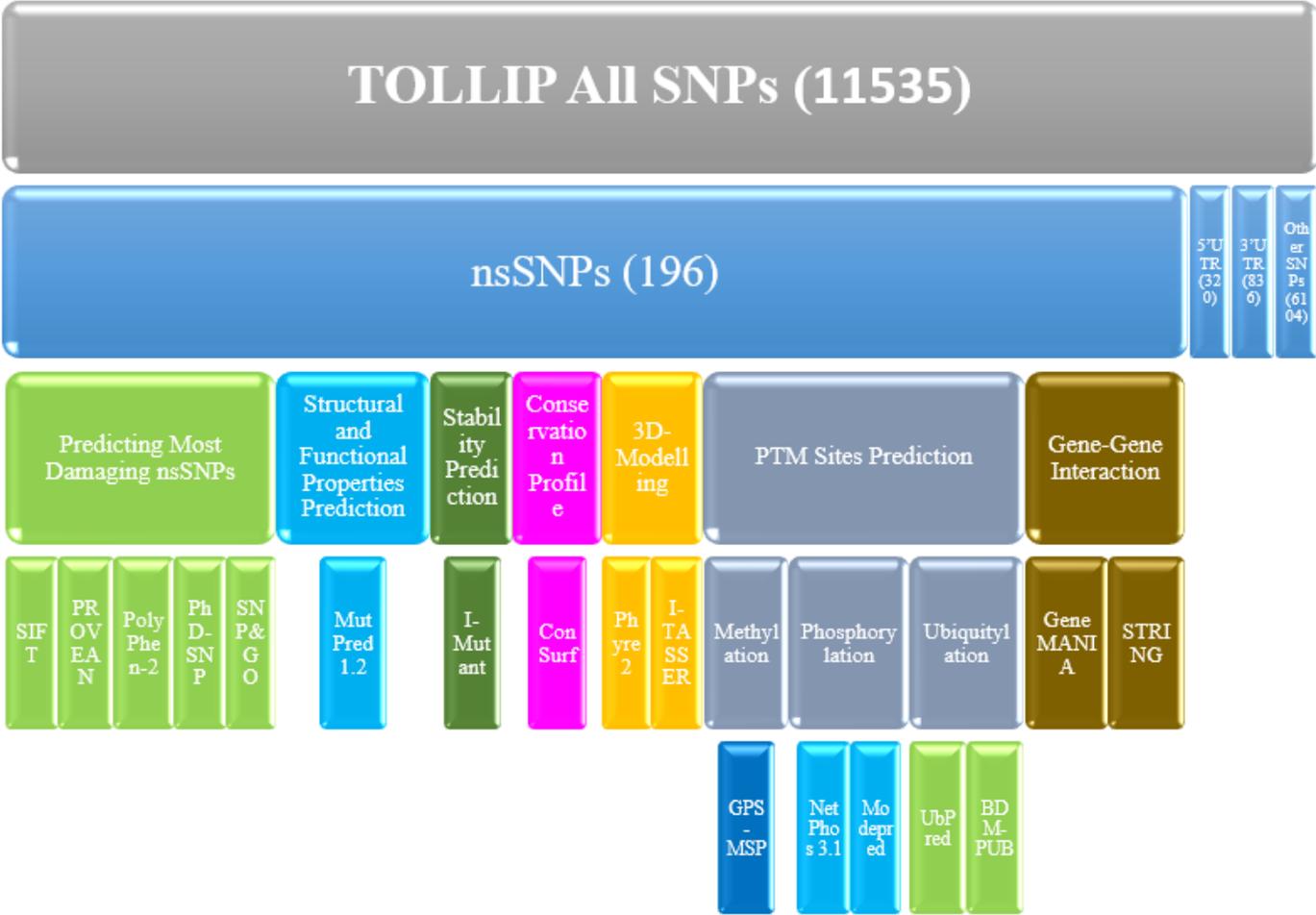


Figure 1

A Diagrammatical representation of Utilized Tools

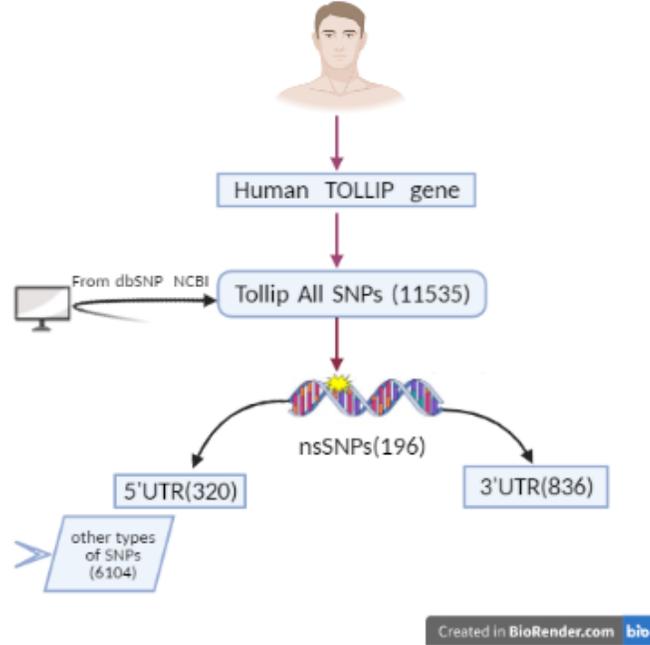


Figure 2

A General Depiction of collected Data of SNPs and nsSNPs

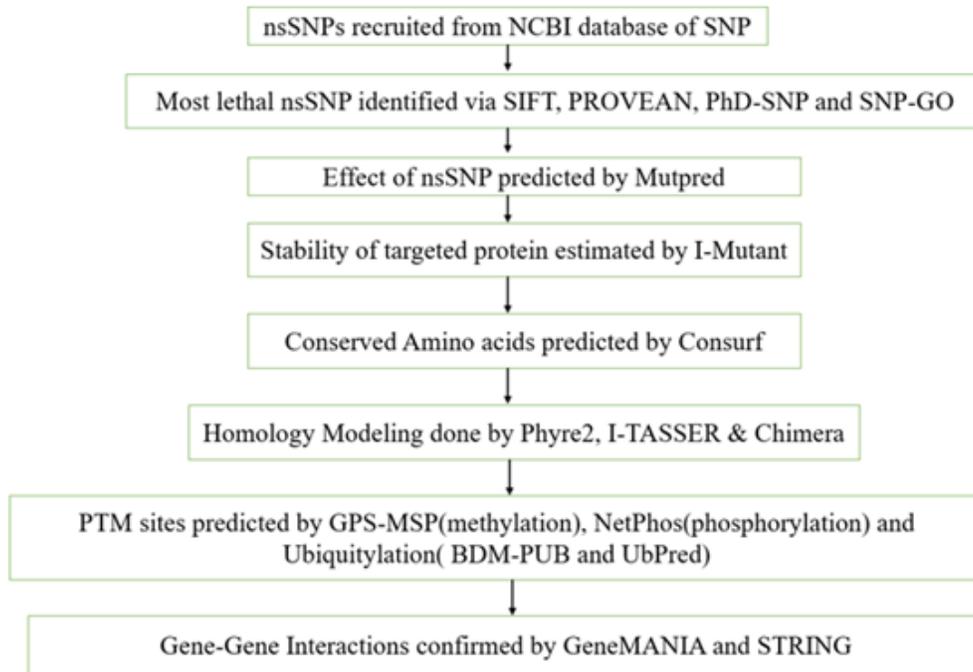


Figure 3

Summary of complete methodology in flow chart format

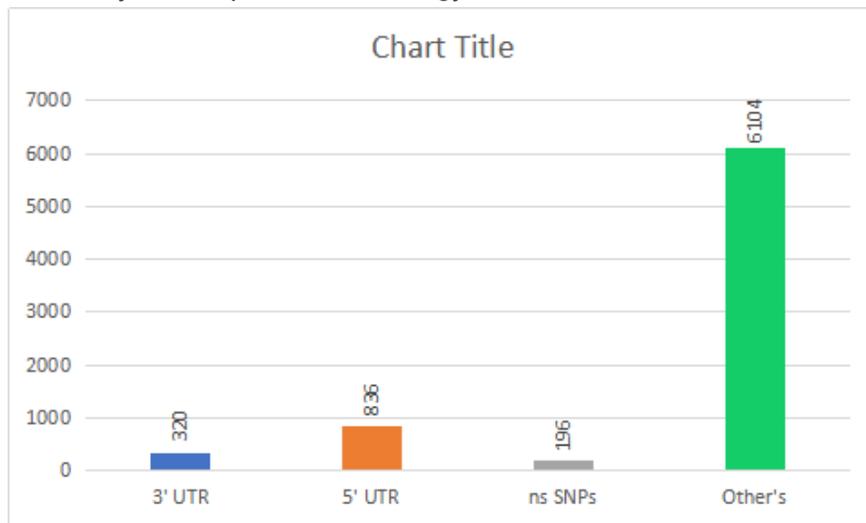


Figure 4

All SNPs in TOLLIP gene.

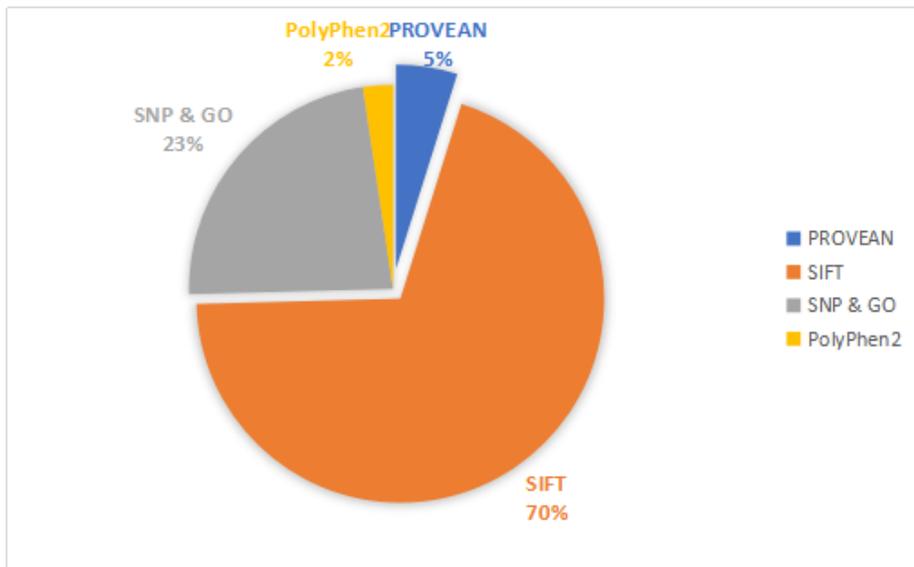


Figure 5

Numbers (n) and percentage (%) of damaging nsSNPs examined by PROVEAN, SIFT, SNP&GO, and Polyphen2

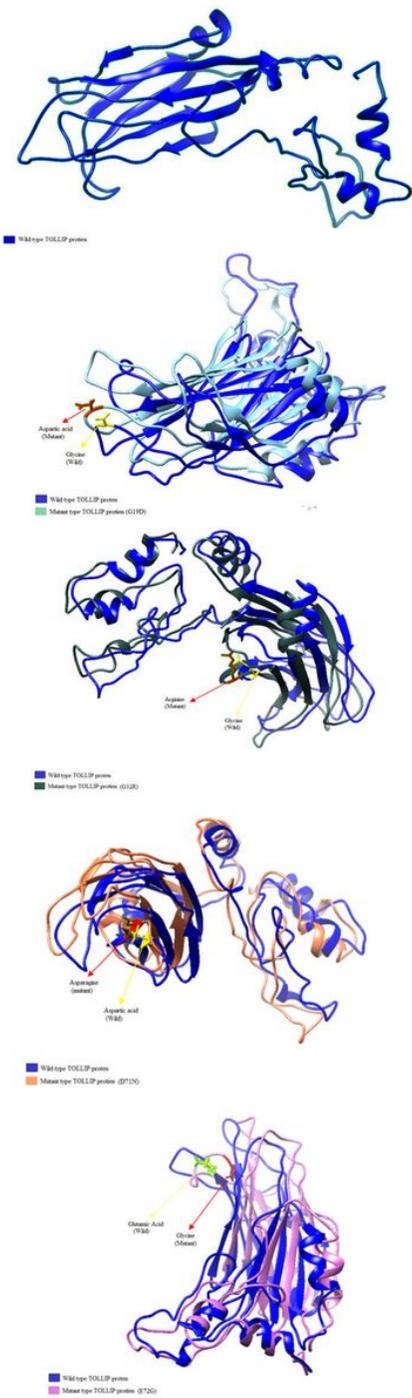


Figure 6

Top pane: A) Wild-type TOLLIP protein Structure. Middle panel: (B & C) TOLLIP'S protein superimposed structures and its mutants G19D and G32R, respectively. Bottom panel: (D & E) Superimposed Structure of TOLLIP and its mutants D71N and E72G, respectively.

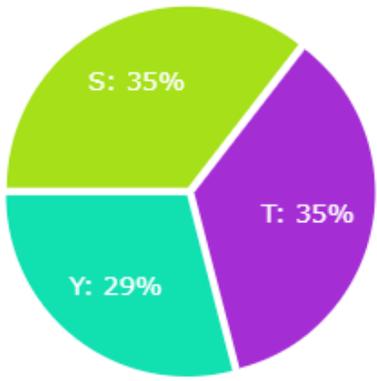


Figure 7

(A) 21 residues containing phosphorylation predicted by MODEPRED.

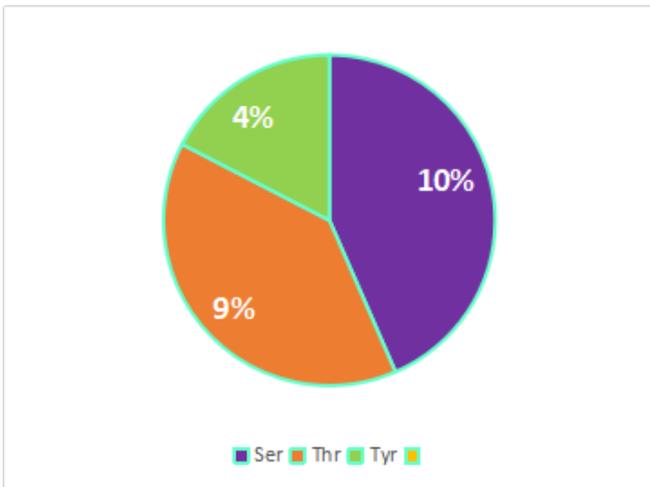


Figure 8

(B) 16 residues containing phosphorylation predicted by NetPhos 3.1.

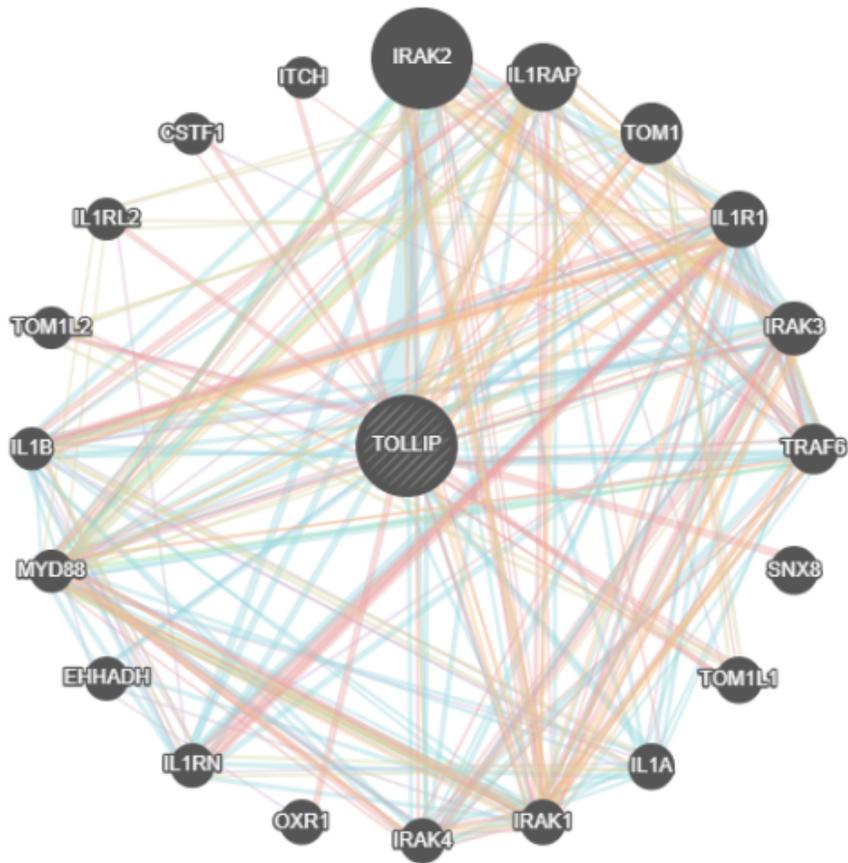


Figure 9

TOLLIP's gene-gene interaction exhibited by GeneMANIA (Second Figure 8 in the manuscript.)

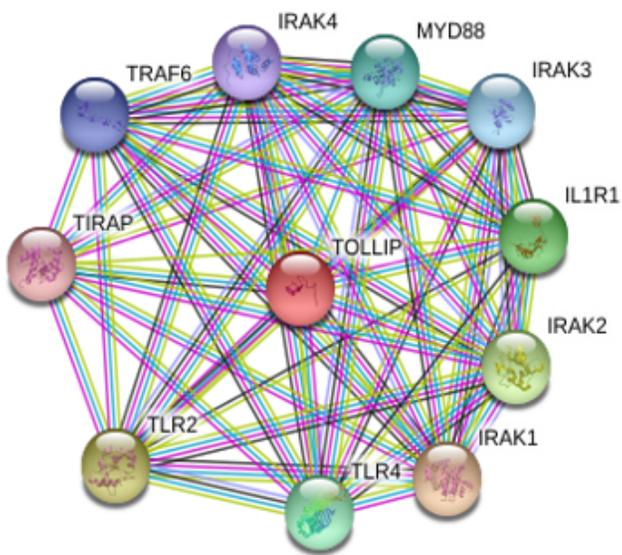


Figure 10

TOLLIP's gene-gene interaction exhibited by STRING.

