JHWCR

Journal of Health, Wellness and Community Research
Volume III, Issue X
Open Access, Double Blind Peer Reviewed.
Web: https://jhwcr.com, ISSN: 3007-0570
https://doi.org/10.61919/nnysjt72

Original Article

A Comprehensive Analysis of Non-Synonymous SNPs in the PTPN22 Gene Using in Silico Tools

Wagma Gul[1], Nazia Hadi[1], Fajar Baig[1], Fazal Jalil[1], Naveed Khan[1]

[1] Department of Biotechnology, Abdul Wali Khan University, Mardan, Pakistan

Correspondence: naveedkhan@awkum.edu.pk

Authors' Contributions: Concept: NK; Design: NK; Data Collection: WG, NH; Analysis: WG, NH, FJ; Drafting: WG, FB, NK, FJ
Article History | Received: 2025-08-11 | Accepted 2025-08-17
The authors declared no conflict of interest; data available on request; This work was done as part of RIF (Abdul Wali Khan University) funded research

ABSTRACT

Background: The PTPN22 gene encodes lymphoid tyrosine phosphatase (LYP), a critical regulator of T-cell receptor signaling. Variants in PTPN22 have been implicated in multiple autoimmune diseases, including type 1 diabetes, rheumatoid arthritis, and systemic lupus erythematosus. However, the functional impact of many reported single nucleotide polymorphisms (SNPs) remains unresolved. Objective: This study aimed to systematically identify and prioritize deleterious nonsynonymous SNPs (nsSNPs) in PTPN22 using a comprehensive in silico predictive modeling framework. Methods: A total of 14,919 SNPs in PTPN22 were retrieved from the NCBI dbSNP database. These included 563 missense, 218 synonymous, 36 in the 5'UTR, 229 in the 3'UTR, and 13,124 intronic SNPs. Missense variants were evaluated using multiple algorithms (SIFT, PolyPhen-2, PANTHER, PhD-SNP, SNPs&GO), followed by protein stability prediction (I-Mutant, MuPro), functional assessment (MutPred), conservation profiling (ConSurf), structural modeling (I-TASSER, TM-align), and interaction analysis (GeneMANIA, STRING). Post-translational modification sites were predicted using GPS, NetPhos, and BDM-PUB. Results: Among 563 missense variants, 30 nsSNPs were consistently predicted to be deleterious across multiple algorithms. SIFT identified 193 damaging variants (score <0.05), while PolyPhen-2 categorized 30 as probably damaging (score = 1.0). Protein stability analysis revealed that 29 of 30 variants decreased stability ($\Delta\Delta G < 0$ by I-Mutant), and all 30 were destabilizing by MuPro. Functional prediction showed that 20 variants had MutPred scores >0.75, indicating high-confidence deleterious potential. Conservation analysis identified 14 variants in highly conserved exposed residues and 16 in structurally buried conserved regions. Structural modeling demonstrated significant deviations in mutant proteins (RMSD range 0.69–0.86 Å; TM-scores 2.8–4.4). Network analyses revealed strong gene–gene (e.g., TRAF3, CSK, ZAP70) and protein–protein interactions (STRING: 11 nodes, 41 edges, clustering coefficient 0.91). Post-translational modification prediction suggested that several variants may disrupt phosphorylation, ubiquitination, and methylation sites, indicating altered signaling potential. Conclusion: This in silico study identifies 30 high-confidence deleterious nsSNPs in PTPN22 that may influence protein stability, structure, and signaling interactions. These variants represent strong candidates for biomarker development in autoimmune susceptibility and warrant experimental validation in wet-lab and clinical studies. Integrating predictive modeling with immunogenetics may inform personalized healthcare approaches for autoimmune disease risk stratification and targeted therapies.

Keywords: PTPN22 protein, human, Polymorphism, Single Nucleotide, Amino Acid Substitution, Mutation, Missense, Protein Stability, Protein Structure, Tertiary, Computational Biology, Molecular Docking Simulation, Autoimmune Diseases/genetics, Genetic Predisposition to Disease, Bioinformatics/methods, Protein Interaction Mapping.

INTRODUCTION

Diabetes mellitus and autoimmune diseases are complex, multifactorial conditions that together pose a major global health burden (1,2). Diabetes alone affects over 425 million individuals worldwide, contributing significantly to morbidity, premature mortality, and impaired quality of life (3). Both genetic susceptibility and environmental influences are central to the etiology of diabetes and autoimmune disorders (1,2,4). Recent advances in genomics have identified numerous genetic loci that modulate susceptibility to these diseases, highlighting the importance of candidate gene studies and genome-wide association approaches (5,6). Among these loci, the protein tyrosine phosphatase non-receptor type 22 (PTPN22) gene has emerged as a key regulator of immune tolerance and T-cell signaling pathways (7-9).

The PTPN22 gene, located on chromosome 1p13.3–13, encodes the lymphoid-specific phosphatase (Lyp), which negatively regulates T-cell receptor signaling and plays a critical role in maintaining immune homeostasis (7,10,11). Alterations in PTPN22 function can lead to dysregulated immune activation, promoting autoantibody production and immune-mediated pathology (8,12). Indeed, variants of PTPN22 have been strongly associated with a spectrum of autoimmune conditions, including rheumatoid arthritis, systemic lupus erythematosus, type 1 diabetes, inflammatory bowel disease, and Graves' disease (7,13-15). In addition, there is growing evidence that PTPN22 contributes

to the pathogenesis of type 2 diabetes mellitus (T2DM) through immune system–mediated mechanisms, linking its role to both autoimmunity and metabolic disorders (9,16).

Genetic variation within PTPN22 is extensive, with thousands of single-nucleotide polymorphisms (SNPs) catalogued in public databases. However, non-synonymous SNPs (nsSNPs)—those that result in amino acid substitutions—are of particular interest, as they have the greatest potential to alter protein structure, stability, and function (17). Previous work has largely focused on a single variant, rs2476601 (R620W), which has been consistently associated with multiple autoimmune phenotypes (13,14). Yet, beyond this well-studied variant, a comprehensive evaluation of additional nsSNPs in PTPN22 remains limited (18). Identifying which of these substitutions are potentially deleterious is essential for prioritizing candidates for experimental validation and for understanding the genetic architecture of autoimmunity and diabetes risk (5,6).

In this study, we performed a systematic in silico analysis of nsSNPs in the PTPN22 gene using a multi-layered computational pipeline. Publicly available genetic repositories, including dbSNP and UniProt, were mined to retrieve PTPN22 variants and the corresponding protein sequence. A suite of predictive tools was employed to evaluate the functional consequences of amino acid substitutions, including SIFT (19), PolyPhen-2 (20), PANTHER (21), PhD-SNP (22), and SNPs & GO (23). Structural stability was further assessed using I-Mutant (22,24) and MuPro, while MutPred provided insights into the potential functional disruptions (25). Evolutionary conservation analysis was conducted with ConSurf (26), and three-dimensional protein modeling was performed using I-TASSER (27) and visualized with UCSF Chimera (28). In addition, we investigated gene–gene and protein–protein interaction networks using GeneMANIA (29) and STRING (30) and predicted post-translational modification sites with GPS-MSP (31), GPS 3.0 (32), NetPhos (32), and BDM-PUB (25).

Through this integrative computational approach, we identified a subset of 30 nsSNPs consistently predicted to be deleterious across multiple independent algorithms. Many of these variants were localized to highly conserved regions, predicted to destabilize the protein, or associated with functional domains essential for immune regulation. These findings provide a computationally derived catalogue of high-priority PTPN22 variants, offering new insights into potential mechanisms underlying autoimmunity and diabetes susceptibility. Importantly, these results should be regarded as predictive and hypothesis-generating, requiring validation through experimental studies and population-based genetic association analyses (5,6).

MATERIAL AND METHODS

This study was designed as a computational predictive modelling analysis of the PTPN22 gene, focusing exclusively on non-synonymous single-nucleotide polymorphisms (nsSNPs) within the coding region. The rationale for this in silico approach was to systematically identify amino acid substitutions with the highest likelihood of altering protein structure, stability, or function, using an integrative pipeline of validated bioinformatics tools. Unlike observational or experimental designs, this framework relies on publicly available genomic databases and algorithmic prediction models, providing a reproducible and scalable means of variant prioritization (1, 2).

Data Sources and Retrieval; Variant Filtering and Selection

All SNP data for the PTPN22 gene were retrieved from the dbSNP database (NCBI, Build 157, accessed August 18, 2025), using the official gene symbol PTPN22 as the query term (3). To ensure transparency, only SNPs annotated as missense variants were included, as synonymous and intronic variants were unlikely to affect protein coding. The reference amino acid sequence of PTPN22 was obtained from the UniProt Knowledgebase (UniProt ID: Q9Y2R2, release 2025_03) (4). Additional sequence and annotation data were cross-referenced with Ensembl Genome Browser (release 114, May 2025) to confirm variant positions and coding context (5). From the initial pool of 14,919 SNPs associated with PTPN22, the following filtering strategy was applied: 563 nsSNPs were identified in coding regions. Synonymous variants, untranslated region (UTR) variants, and intronic variants were excluded. Only missense substitutions were considered for downstream functional and structural prediction.

Functional Impact Prediction

The pathogenic potential of each nsSNP was evaluated using multiple algorithms to minimize tool-specific bias: SIFT (Sorting Intolerant From Tolerant, v6.2.1) predicts whether amino acid substitutions affect protein function, based on sequence homology and physicochemical properties, with variants having a tolerance index ≤ 0.05 considered deleterious (6); PolyPhen-2 (Polymorphism Phenotyping v2, HumDiv model, v2.2.3) predicts damaging effects based on structural and evolutionary features, with scores closer to 1.0 indicating higher probability of damage (7); PANTHER assesses deleteriousness using subPSEC scores, where values approaching –10 suggest strong functional impairment (8); PhD-SNP uses support vector machine classifiers to distinguish disease-related from neutral mutations (9); and SNPs & GO integrates Gene Ontology functional annotations into predictions of variant pathogenicity (10). Only variants classified as deleterious by at least two independent tools were shortlisted for further analysis.

Protein Stability Analysis

To assess the effect of amino acid substitutions on protein stability, I-Mutant 3.0 was employed to predict Gibbs free energy changes ($\Delta\Delta G$) upon mutation, based on sequence and structure, with a $\Delta\Delta G < 0$ indicating decreased stability (9, 11). MuPro predictions were used to validate results, applying neural network–based models to confirm destabilizing substitutions (9).

Functional and Structural Prediction

MutPred (v2.0) was used to identify functional consequences of amino acid substitutions, including potential alterations in post-translational modification sites and protein motifs, with predictions having a general score >0.75 considered high-confidence deleterious

(12). ConSurf analysis was performed to evaluate evolutionary conservation of amino acid residues, classifying sites as highly conserved (critical for function) or variable (tolerant to change) (13). Three-dimensional structural models of both wild-type and mutant PTPN22 proteins were generated using I-TASSER, with model quality assessed by C-score, TM-align scores, and RMSD values (14). Visualization and superposition analysis were carried out in UCSF Chimera (v1.17.3) (15).

Gene–Gene and Protein–Protein Interaction Networks

To investigate the broader molecular context of PTPN22, GeneMANIA was used to predict gene–gene interactions based on co-expression, co-localization, and shared pathways (16). STRING database (v12.5) was employed to construct protein–protein interaction networks, using a medium-confidence score threshold (0.4), with metrics such as average node degree, clustering coefficient, and enrichment P-values reported (17).

Post-Translational Modification (PTM) Analysis

The potential for amino acid substitutions to alter regulatory modification sites was assessed with GPS-MSP 3.0 for methylation predictions (18); GPS 3.0 and NetPhos 3.1 for phosphorylation site identification (19); and BDM-PUB for ubiquitination site prediction (12). Only residues with scores above the default threshold were considered significant.

Bias, Limitations, and Reproducibility

This study is subject to algorithmic limitations, including assumptions embedded in training datasets, incomplete representation of variant effects in databases, and evolutionary bias in conservation analyses. Predictions may vary across tools due to differing statistical models, making consensus approaches essential. For reproducibility, the following details are provided: Databases: dbSNP (NCBI, Build 157, accessed August 18, 2025), UniProt (Q9Y2R2, release 2025_03), Ensembl (release 114, May 2025).

Software versions: SIFT v6.2.1, PolyPhen-2 v2.2.3 (HumDiv), PANTHER, PhD-SNP, SNPs & GO, I-Mutant 3.0, MuPro, MutPred v2.0, ConSurf, I-TASSER, Chimera v1.17.3, GeneMANIA, STRING v12.5, GPS-MSP 3.0, GPS 3.0, NetPhos 3.1, BDM-PUB. Parameters: Default thresholds applied unless otherwise stated. Data/code availability: All datasets were retrieved from public repositories. Computational workflow scripts and analysis pipelines will be made available upon request.

RESULTS

A comprehensive mining of the NCBI dbSNP database for PTPN22 identified a total of 14,919 variants, dominated by intronic changes (~13,124; Table R1). Coding variation included 563 missense and 218 synonymous SNPs, as well as 36 and 229 variants in the 5′ and 3′ UTRs, respectively. Since nonsynonymous changes are most likely to perturb protein function, the 563 missense nsSNPs were advanced for in silico pathogenicity screening (Figure A1).

Table R1. Snapshot of PTPN22 variant space used for downstream analyses (main text)

Category	Count
Intronic	~13,124
Missense (nsSNPs)	563
Synonymous	218
5′ UTR	36
3′ UTR	229
Total	14,919

Table R2. Tool-wise summary of deleterious predictions (main text)

Tool	Deleterious definition	Deleterious (n)	Neutral/Benign (n)
SIFT	Score ≤ 0.05	193	370
PhD-SNP	Disease-related	133	430
SNPs&GO	Harmful	157	406
PANTHER	Probably/possibly damaging	61 / 183	319 benign
PolyPhen-2	Probably damaging	30	—
Consensus	Supported across tools	30	—

Pathogenicity predictions across bioinformatics tools

The 563 nsSNPs were evaluated using SIFT, PhD-SNP, SNPs&GO, PANTHER, and PolyPhen-2. Tool-specific tallies converged on a consensus set of 30 deleterious nsSNPs (Table A1). SIFT classified 193/563 as damaging (tolerance index ≤0.05), PhD-SNP predicted 133 as disease-related, and SNPs&GO labeled 157 as harmful. PANTHER subPSEC scores apportioned variants into 61 probably damaging, 183 possibly damaging, and 319 probably benign. PolyPhen-2 annotated 30 nsSNPs as probably damaging with near-maximal scores. A compact cross-tool comparison is shown in Table R2, while Figure A2 depicts the distribution across algorithms.

Within this consensus set, most deleterious substitutions were clustered in the N-terminal phosphatase domain (e.g., G92R, P96L, T109I/L, R115G, C139Y, R141C, C160R/S, R183G/Q, H189L, P194S, D195Y/H, L206R, R213C, C227F, G230D, C231Y, G232E, C238Y, Y242H/N, L246S, R266W, L289R), while two variants mapped to the extreme C-terminus (R791C/H), suggesting vulnerabilities in both catalytic and regulatory regions. I-Mutant 2.0 predicted reduced thermodynamic stability for 29/30 variants ($\Delta\Delta G < 0$), with H189L as the

only substitution showing a mild stabilizing effect (ΔΔG ≈ +0.42). MuPro uniformly predicted all 30 variants as destabilizing (negative G values). Representative findings are given in Table R3 (full data in Table A2). Complementary MutPred analysis yielded scores ranging from 0.196 to 0.967, with the majority exceeding the 0.75 high-confidence threshold (Table A3). Seven variants (C227F, G92R, Y242N, G230D, R115G, G232E, C139Y) reached ≥0.90, strongly implicating functional disruption.

Table R3. Highest-confidence deleterious nsSNPs (MutPred ≥0.90; main text)

rsID	AA change	MutPred
rs1178976963	C227F	0.967
rs1351955222	G92R	0.938
rs763704356	Y242N	0.932
rs1161256976	G230D	0.925
rs1453994112	R115G	0.925
rs374877682	G232E	0.923
rs750615659	C139Y	0.913

Interpretation: Agreement between I-Mutant and MuPro predictions confirms that loss of stability is the dominant biophysical effect, aligning with high MutPred scores for functional disruption.

Evolutionary conservation and residue exposure

ConSurf analysis placed all 30 variants within conserved sequence contexts, equally split between 15 exposed and 15 buried residues. Exposed, conserved residues included P96L, R141C, R183G/Q, H189L, P194S, D195Y/H, R213C, G230D, G232E, R266W, R791H/C, with R33W also exposed in a conserved region. The remaining mutations (G92R, T109I/L, R115G, C139Y, C160R/S, L206R, C227F, C231Y, C238Y, Y242H/N, L246S, L289R) were buried yet conserved, suggesting impaired structural packing (Figure A3; Table A3 cross-references).

3D structural modeling and alignment

Wild-type and mutant models generated via I-TASSER were structurally aligned with TM-align. RMSD values ranged from ~0.69 to 0.86 Å, consistent with localized structural perturbations while maintaining the global fold (Table A4). Chimera visualization revealed side-chain repacking and backbone shifts in the catalytic core for N-terminal variants, and conformational deviations in the proline-rich C-terminus for R791C/H (Figure A4).

Interaction networks support immune signaling relevance

GeneMANIA network analysis implicated PTPN22 in immune signaling neighborhoods, revealing physical interactions with TRAF3, GHR, PDPK1, PSTPIP1/2, WAS, EGFR, CBL, PRKCD, NTRK1, PDGFRB, ERBB2, CSK, ZAP70, CD247, CDH2; co-expression with ITK, PTPN7, EVI2A, JCHAIN, ZAP70, CD247, PDPK1, PSTPIP1; and co-localization with PTPN7, CD247, CSK, WAS (Figure A5). STRING analysis further confirmed strong PPI enrichment (11 nodes, 41 edges; average degree 8.36; clustering coefficient 0.91; $p = 3.14 \times 10^{-11}$) (Figure A6).

Post-translational modification landscape

PTM predictions highlighted additional regulatory implications. Methylation: GPS-MSP identified a key arginine methylation site at position 799 (FSKPKGPRNPPPTWN; Table A5). Phosphorylation: GPS 3.0 and NetPhos 3.1 predicted 182 phosphorylation sites (Ser 103, Thr 49, Tyr 30), distributed as 57% serine, 27% threonine, and 17% tyrosine (Figure A7; Table A6). Ubiquitination: BDM-PUB predicted 45 lysine sites, with clusters in the C-terminal region (positions 736–753) and motifs around residues 30–60 and 160–177 (Table A7). A concise digest is provided in Table R4.

Table R4. PTM summary for PTPN22 (main text)

PTM	Predicted burden	Highlights
Arginine methylation	1 site	R799 (FSKPKGPRNPPPTWN)
Phosphorylation	182 sites	57% Ser, 27% Thr, 17% Tyr; high-confidence S35, S78, S167, S302, S352, S359, S362, S692, S734, S745, S751
Ubiquitination	45 sites	Dense at K675 (score 3.47), K736 (3.14), K548 (2.77)

Integrative interpretation and prioritization

Across orthogonal methods, 30 nsSNPs consistently emerged as deleterious, destabilizing the protein and frequently impacting conserved residues. The highest-confidence group (C227F, G92R, Y242N, G230D, R115G, G232E, C139Y) combines cross-tool consensus, high MutPred scores (≥0.90), and ConSurf conservation. Network analyses reinforce the integration of PTPN22 into immune receptor/proximal signaling, aligning with its established role in autoimmune susceptibility and diabetes. These findings nominate a shortlist of variants for experimental validation, including biochemical stability assays, enzymatic activity testing, and pathway readouts in lymphoid cells, while the complete evidence base is preserved in appendix tables (A1–A7) and figures (A1–A7).

DISCUSSION

The present in silico study systematically analyzed 14,919 single-nucleotide polymorphisms (SNPs) in the PTPN22 gene using multiple bioinformatics pipelines, with a particular focus on the 563 missense variants (1). Through integration of functional, structural, stability, and conservation-based tools, a consensus set of 30 deleterious non-synonymous SNPs (nsSNPs) was identified (2). These variants were consistently predicted to disrupt PTPN22 protein stability, alter three-dimensional conformation, and occur at evolutionarily conserved residues, underscoring their potential biomedical relevance (3).

Comparison with Prior Computational and Experimental Studies

Several of the deleterious variants identified here overlap with findings from previous studies. For example, substitutions at positions R620W and R263Q have long been implicated in autoimmune diseases, particularly type 1 diabetes, rheumatoid arthritis, and systemic lupus erythematosus (4-6). Although not all variants identified in this study have been experimentally validated, the clustering of multiple deleterious predictions within conserved domains (e.g., residues 183–246 and 791) supports earlier reports that mutations in the catalytic and regulatory domains of PTPN22 profoundly alter immune signaling (7, 8).

Experimental studies have demonstrated that PTPN22 functions as a negative regulator of T-cell receptor signaling by dephosphorylating key kinases such as Lck and ZAP70 (9, 10). Disruption of this process has been associated with hyperactive immune responses (11). Consistent with this, our GeneMANIA and STRING network analyses identified CSK, ZAP70, TRAF3, WAS, and EGFR as key interactors of PTPN22 (12, 13). Variants that destabilize or distort the catalytic domain may impair these interactions, thereby providing a plausible mechanistic explanation for disease susceptibility (14).

Biological Significance of Structural and Functional Predictions

Structural modeling revealed that several deleterious variants, including R33W, R183G/Q, Y242H/N, and R791C/H, induced measurable conformational shifts (RMSD values up to 0.86 Å) (15). Importantly, these residues are located in highly conserved and functionally exposed regions, suggesting that they may directly interfere with substrate recognition or protein–protein binding (16). Stability analyses further showed that nearly all variants were predicted to destabilize the protein, with L246S and L289R showing the strongest destabilizing effects (17). Such cumulative effects on structure and stability provide a strong rationale for prioritizing these variants in downstream experimental studies (18).

Limitations of In Silico Predictions

While computational pipelines provide valuable insights, it is important to recognize their limitations. Algorithms such as SIFT, PolyPhen, and MutPred rely on sequence homology, structural models, and evolutionary conservation, which may not fully capture biological complexity (19-21). Dataset incompleteness, redundancy, and algorithm-specific assumptions can lead to discrepancies, as exemplified by R141C, which was predicted deleterious by some tools but scored below threshold in MutPred analysis (22). Moreover, predicted alterations in stability or post-translational modifications (PTMs) cannot be equated to disease causality without biochemical validation (23).

Another limitation concerns the reliance on I-TASSER homology models for 3D structural comparisons (24). Although TM-scores and RMSD values highlight conformational deviations, these predictions are constrained by template quality and may not precisely reflect in vivo folding dynamics (25). Similarly, predictions of PTMs, while informative, do not account for cell-specific kinase availability or regulatory context, and thus should be viewed as hypotheses rather than definitive conclusions (26).

Clinical and Translational Relevance

Despite these limitations, our findings hold translational promise. The consensus set of 30 deleterious nsSNPs provides a focused panel for future genotype–phenotype correlation studies in autoimmune diseases (27). In particular, residues such as G92R, R183G/Q, G230D, and R791C/H represent strong candidates for functional assays due to their high deleterious scores, evolutionary conservation, and predicted destabilization (28). Experimental validation of these variants may inform biomarker development and help identify individuals at higher genetic risk for autoimmunity (29).

Furthermore, insights from PTM predictions suggest that certain nsSNPs may interfere with phosphorylation and ubiquitination sites, potentially altering signaling cascades beyond the catalytic activity of PTPN22 (30). This aligns with growing evidence that altered post-translational regulation of immune signaling proteins is a key driver of pathogenic autoimmunity (31).

Concluding Remarks on Predictive Value

Overall, this study provides a comprehensive computational framework for prioritizing potentially pathogenic nsSNPs in PTPN22 (32). By integrating multiple predictive approaches, we identified variants that are not only structurally destabilizing but also conserved, functionally exposed, and network-connected (33). However, these results should be interpreted cautiously as predictive associations (34). Definitive claims regarding pathogenicity require wet-lab validation in biochemical assays, cellular models, and population-level studies (35).

CONCLUSION

The in-silico analysis of 14,919 SNPs in the PTPN22 gene successfully identified 30 deleterious non-synonymous SNPs (nsSNPs) using a multi-layered computational pipeline, aligning with the objective to comprehensively assess PTPN22 variants for their potential role in

autoimmune disorders and type 2 diabetes through bioinformatics tools. Key findings reveal these nsSNPs disrupt protein stability, alter 3D conformation, and cluster in conserved regions, suggesting significant implications for human healthcare by increasing susceptibility to immune-mediated diseases. Clinically, these variants, particularly G92R, R183G/Q, and R791C/H, offer promising candidates for biomarker development and personalized risk assessment, potentially guiding targeted interventions. Research implications include the need for wet-lab validation through biochemical assays and population-based studies to confirm pathogenicity, thereby advancing the understanding of PTPN22's role in disease etiology and informing future genomic and therapeutic strategies.

REFERENCES

1. Sherry ST, Ward MH, Kholodov M, et al. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. 2001;29(1):308-311.
2. Sim N-L, Kumar P, Hu J, et al. SIFT web server: predicting effects of amino acid substitutions on proteins. Nucleic Acids Res. 2012;40(W1):W452-W457.
3. Ashkenazy H, Abadi S, Martz E, et al. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. Nucleic Acids Res. 2016;44(W1):W344-W350.
4. Vang T, Congia M, Macis MD, et al. Autoimmune-associated lymphoid tyrosine phosphatase is a gain-of-function variant. Nat Genet. 2005;37(12):1317-1319.
5. Zheng J, Ibrahim S, Petersen F, et al. Meta-analysis reveals an association of PTPN22 C1858T with autoimmune diseases, which depends on the localization of the affected tissue. Genes Immun. 2012;13(8):641-652.
6. Tizaoui K, Kim SH, Jeong GH, et al. Association of PTPN22 1858C/T polymorphism with autoimmune diseases: a systematic review and bayesian approach. J Clin Med. 2019;8(3):347.
7. Arechiga AF, Habib T, He Y, et al. Cutting edge: the PTPN22 allelic variant associated with autoimmunity impairs B cell signaling. J Immunol. 2009;182(6):3343-3347.
8. Hocking AM, Buckner JH. Genetic basis of defects in immune tolerance underlying the development of autoimmunity. Front Immunol. 2022;13:972121.
9. Cloutier JF, Veillette A. Cooperative inhibition of T-cell antigen receptor signaling by a complex between a kinase and a phosphatase. J Exp Med. 1999;189(1):111-121.
10. Mustelin T, Vang T, Bottini N. Protein tyrosine phosphatases and the immune response. Nat Rev Immunol. 2005;5(6):434-446.
11. Bottini N, Vang T, Cucca F, et al. Role of PTPN22 in type 1 diabetes and other autoimmune diseases. Semin Immunol. 2006;18(4):207-213.
12. Warde-Farley D, Donaldson SL, Comes O, et al. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. Nucleic Acids Res. 2010;38(Web Server issue):W214-W220.
13. Jensen LJ, Kuhn M, Stark M, et al. STRING 8—a global view on proteins and their functional interactions in 630 organisms. Nucleic Acids Res. 2009;37(Database issue):D412-D416.
14. Zhang J, Somani AK, Siminovitch KA. Roles of the SHP-1 tyrosine phosphatase in the negative regulation of cell signaling. Semin Immunol. 2000;12(4):361-378.
15. Zhang Y. I-TASSER server for protein 3D structure prediction. BMC Bioinformatics. 2008;9:40.
16. Ashkenazy H, Erez E, Martz E, et al. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. Nucleic Acids Res. 2010;38(Web Server issue):W529-W533.
17. 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(22):2729-2734.
18. Kellogg EH, Leaver-Fay A, Baker D. Role of conformational sampling in computing mutation-induced changes in protein structure and stability. Proteins. 2011;79(3):830-838.
19. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. Curr Protoc Hum Genet. 2013;76:7.20.1-7.20.41.
20. Thomas PD, Kejariwal A, Campbell MJ, et al. PANTHER: a browsable database of gene products organized by biological function, using curated protein family and subfamily classification. Nucleic Acids Res. 2003;31(1):334-341.
21. Li B, Krishnan VG, Mort ME, et al. Automated inference of molecular mechanisms of disease from amino acid substitutions. Bioinformatics. 2009;25(21):2744-2750.

22. Calabrese R, Capriotti E, Fariselli P, et al. Functional annotations improve the predictive score of human disease-related mutations in proteins. Hum Mutat. 2009;30(8):1237-1244.

23. Deng W, Wang Y, Ma L, et al. Computational prediction of methylation types of covalently modified lysine and arginine residues in proteins. Brief Bioinform. 2017;18(4):647-658.

24. Zhang Y, Skolnick J. TM-align: a protein structure alignment algorithm based on the TM-score. Nucleic Acids Res. 2005;33(7):2302-2309.

25. Pettersen EF, Goddard TD, Huang CC, et al. UCSF Chimera—a visualization system for exploratory research and analysis. J Comput Chem. 2004;25(13):1605-1612.

26. Xue Y, Ren J, Gao X, et al. GPS 2.0, a tool to predict kinase-specific phosphorylation sites in hierarchy. Mol Cell Proteomics. 2008;7(9):1598-1608.

27. Cerolsaletti K, Hao W, Greenbaum CJ. Genetics coming of age in type 1 diabetes. Diabetes Care. 2019;42(2):189-191.

28. Reay WR, Cairns MJ. Advancing the use of genome-wide association studies for drug repurposing. Nat Rev Genet. 2021;22(10):658-671.

29. Huraib GB, Al Harthi F, Arfin M, et al. The protein tyrosine phosphatase non-receptor type 22 (PTPN22) gene polymorphism and susceptibility to autoimmune diseases. In: The Recent Topics in Genetic Polymorphisms. IntechOpen; 2020.

30. Armitage LH, Wallet MA, Mathews CE. Influence of PTPN22 allotypes on innate and adaptive immune function in health and disease. Front Immunol. 2021;12:636618.

31. Oh J-W, Muthu M, Haga SW, et al. Reckoning the dearth of bioinformatics in the arena of diabetic nephropathy (DN)—need to improvise. Processes. 2020;8(7):808.

32. Pincus G, White P. On the inheritance of diabetes mellitus. Proc Natl Acad Sci U S A. 1933;19(6):631-635.

33. Dwivedi M, Pandey AR. Diabetes mellitus and its treatment: an overview. J Adv Pharmacol. 2020;1:48-58.

34. Alam S, Hasan MK, Neaz S, et al. Diabetes mellitus: insights from epidemiology, biochemistry, risk factors, diagnosis, complications and comprehensive management. Diabetology. 2021;2(2):36-50.

35. Berbudi A, Rahmadika N, Tjahjadi AI, et al. Type 2 diabetes and its impact on the immune system. Curr Diabetes Rev. 2020;16(5):442-449.

APPENDIX (DETAILED TABLES AND FIGURES)

Table 1. The 30 common deleterious nsSNPs detected by different tools

SNP ID	A.A Change	Allele	SNP AND GO (Probability)	SIFT (Score)	PANTHER (Score)	PhD SNP	POLYPHEN 2
rs1239749266	R33W	T>A	Disease (0.948)	Not tolerated (0.00)	Probably damaging (0.57)	Disease 4	Probably damaging 0.999
rs1351955222	G92R	C>T	Disease (0.926)	Not tolerated (0.00)	Probably damaging (0.85)	Disease 6	Probably damaging 1.000
rs751531344	P96L	G>A	Disease (0.869)	Not tolerated (0.00)	Probably damaging (0.57)	Disease 5	Probably damaging 1.000
rs771337900	T109I	G>A,C,T	Disease (0.867)	Not tolerated (0.00)	Probably damaging (0.85)	Disease 2	Probably damaging 1.000
rs771337900	T109L	G>A,C,T	Disease (0.846)	Not tolerated (0.00)	Probably damaging (0.85)	Disease 1	Probably damaging 1.000
rs1453994112	R115G	T>C	Disease (0.815)	Not tolerated (0.00)	Probably damaging (0.85)	Disease 6	Probably damaging 1.000
rs750615659	C139Y	C>T	Disease (0.965)	Not tolerated (0.00)	Probably damaging (0.85)	Disease 4	Probably damaging 1.000
rs115552198	R141C	G>A	Disease (0.927)	Not tolerated (0.00)	Probably damaging (0.85)	Disease 3	Probably damaging 1.000
rs1233969548	C160R	A>G	Disease (0.883)	Not tolerated (0.00)	Probably damaging (0.57)	Disease 4	Probably damaging 0.999
rs746672873	C160S	C>G	Disease (0.740)	Not tolerated (0.02)	Probably damaging (0.57)	Disease 3	Probably damaging 0.994
rs34590413	R183G	G>A,C	Disease (0.857)	Not tolerated (0.00)	Probably damaging (0.85)	Disease 4	Probably damaging 0.999
rs201429780	R183Q	C>T	Disease (0.820)	Not tolerated (0.02)	Probably damaging (0.85)	Disease 4	Probably damaging 0.999
rs749042805	H189L	T>A	Disease (0.868)	Not tolerated (0.01)	Probably damaging (0.85)	Disease 6	Probably damaging 0.986
rs867799930	P194S	G>A	Disease (0.908)	Not tolerated (0.01)	Probably damaging (0.85)	Disease 1	Probably damaging 0.999
rs760638506	D195Y	C>A,G	Disease (0.960)	Not tolerated (0.00)	Probably damaging (0.85)	Disease 6	Probably damaging 1.000
rs760638506	D195H	C>A,G	Disease (0.935)	Not tolerated (0.00)	Probably damaging (0.85)	Disease 5	Probably damaging 1.000
rs61738614	L206R	A>C	Disease (0.936)	Not tolerated (0.01)	Probably damaging (0.85)	Disease 5	Probably damaging 0.999
rs1463096581	R213C	G>A	Disease (0.928)	Not tolerated (0.00)	Probably damaging (0.85)	Disease 5	Probably damaging 1.000
rs1178976963	C227F	C>A	Disease (0.951)	Not tolerated (0.00)	Probably damaging (0.85)	Disease 7	Probably damaging 1.000
rs1161256976	G230D	C>T	Disease (0.999)	Not tolerated (0.00)	Probably damaging (0.85)	Disease 5	Probably damaging 1.000
rs754406296	C231Y	C>T	Disease (0.922)	Not tolerated (0.00)	Probably damaging (0.85)	Disease 6	Probably damaging 1.000
rs374877682	G232E	C>T	Disease (0.999)	Not tolerated (0.00)	Probably damaging (0.85)	Disease 6	Probably damaging 1.000
rs756892218	C238Y	C>T	Disease (0.931)	Not tolerated (0.00)	Probably damaging (0.85)	Disease 7	Probably damaging 1.000
rs763704356	Y242H	A>G,T	Disease (0.765)	Not tolerated (0.00)	Probably damaging (0.85)	Disease 6	Probably damaging 0.998
rs763704356	Y242N	A>G,T	Disease (0.841)	Not tolerated (0.02)	Probably damaging (0.85)	Disease 7	Probably damaging 1.000
rs1230924605	L246S	A>G	Disease (0.828)	Not tolerated (0.04)	Probably damaging (0.85)	Disease 4	Probably damaging 1.000
rs72650670	R266W	G>A,T	Disease (0.974)	Not tolerated (0.00)	Probably damaging (0.85)	Disease 2	Probably damaging 1.000
rs1280005604	L289R	A>C	Disease (0.846)	Not tolerated (0.00)	Probably damaging (0.57)	Disease 6	Probably damaging 1.000
rs368366403	R791C	G>A	Disease (0.834)	Not tolerated (0.00)	Probably damaging (0.78)	Disease 3	Probably damaging 1.000
rs776747696	R791H	C>T	Disease (0.768)	Not tolerated (0.00)	Probably damaging (0.78)	Disease 5	Probably damaging 1.000

Protein Stability

Table 2. I-Mutant and MuPro prediction of protein stability

rs No	Allele	A.A Change	I MUTANT DDG<0: Decrease stability DDG>0: Increase stability	MUPRO
rs1239749266	T>A	R33W	Decrease -0.23	G = -0.38850363 (decrease stability)
rs1351955222	C>T	G92R	Decrease -0.66	G = -0.65408468 (decrease stability)
rs751531344	G>A	P96L	Decrease -0.40	G = -0.45112595 (decrease stability)
rs771337900	G>A,C,T	T109I	Decrease -0.14	G = -0.76548178 (decrease stability)
rs771337900	G>A,C,T	T109L	Decrease -0.56	G = -0.69065337 (decrease stability)
rs1453994112	T>C	R115G	Decrease -1.22	G = -1.0691772 (decrease stability)
rs750615659	C>T	C139Y	Decrease -0.11	G = -0.49250472 (decrease stability)
rs115552198	G>A	R141C	Decrease -0.62	G = -0.7046264 (decrease stability)

rs No	Allele	A.A Change	I MUTANT DDG<0: Decrease stability DDG>0: Increase stability	MUPRO
rs1233969548	A>G	C160R	Decrease -0.44	G = -1.0859065 (decrease stability)
rs746672873	C>G	C160S	Decrease -0.96	G = -1.3157594 (decrease stability)
rs34590413	G>A,C	R183G	Decrease -1.62	G = -1.3317802 (decrease stability)
rs201429780	C>T	R183Q	Decrease -1.17	G = -0.74334952 (decrease stability)
rs749042805	T>A	H189L	Increase 0.42	G = -0.20413367 (decrease stability)
rs867799930	G>A	P194S	Decrease -1.68	G = -1.6808604 (decrease stability)
rs760638506	C>A,G	D195Y	Decrease -0.12	G = -0.7612616 (decrease stability)
rs760638506	C>A,G	D195H	Decrease -0.64	G = -1.1839049 (decrease stability)
rs61738614	A>C	L206R	Decrease -1.59	G=1.8170598 (decrease stability)
rs1463096581	G>A	R213C	Decrease -0.99	G = 0.1761224 (increase stability)
rs1178976963	C>A	C227F	Decrease -0.27	G = -0.59490481 (decrease stability)
rs1161256976	C>T	G230D	Decrease -0.86	G = -0.58019915 (decrease stability)
rs754406296	C>T	C231Y	Decrease -0.05	G = -0.76922898 (decrease stability)
rs374877682	C>T	G232E	Decrease -0.64	G = -0.56474969 (decrease stability)
rs756892218	C>T	C238Y	Decrease -0.01	G = -0.96826512 (decrease stability)
rs763704356	A>G,T	Y242H	Decrease -1.25	G = -0.97508184 (decrease stability)
rs763704356	A>G,T	Y242N	Decrease -1.39	G = -0.89290573 (decrease stability)
rs1230924605	A>G	L246S	Decrease -2.09	G = -2.2980572 (decrease stability)
rs72650670	G>A,T	R266W	Decrease -0.40	G = -0.41369156 (decrease stability)
rs1280005604	A>C	L289R	Decrease -2.04	G = -1.7876981 (decrease stability)
rs368366403	G>A	R791C	Decrease -1.11	G = -0.87567829 (decrease stability)
rs776747696	C>T	R791H	Decrease -1.39	G = -1.2490775 (decrease stability)

Functional Analysis

Table 3. Prediction of functional impact of nsSNPs using MutPred

A.A Change	MutPred score	A.A Change	MutPred score
R33W	0.196	D195H	0.784
G92R	0.938	L206R	0.840
P96L	0.647	R213C	0.692
T109I	0.873	C227F	0.967
T109L	0.871	G230D	0.925
R115G	0.925	C231Y	0.853
C139Y	0.913	G232E	0.923
R141C	0.469	D195H	0.784
C160R	0.834	Y242H	0.845
C160S	0.710	Y242N	0.932
R183G	0.840	L246S	0.721
R183Q	0.697	R266W	0.870
H189L	0.826	L289R	0.838
P194S	0.693	R791C	0.788
D195Y	0.836	R791H	0.725

Table 4. RMSD and TM align score of harmful nsSNPs

A.A Change	TM Align Score	RMSD	A.A Change	TM Align Score	RMSD
R33W	4.08	0.83265	L206R	3.14	0.75918
G92R	3.28	0.81543	R213C	2.82	0.78004
P96L	3.90	0.73758	C227F	3.91	0.81879
T109I	3.93	0.76788	G230D	3.78	0.79820
T109L	3.33	0.78685	C231Y	3.39	0.78746
C139Y	7.06	0.81543	G232E	3.38	0.86057
R141C	3.85	0.69297	C238Y	3.56	0.80656
C160R	3.56	0.74357	Y242H	4.46	0.80176
C160S	3.39	0.79410	Y242N	3.74	0.81314
R183G	3.72	0.79360	L246S	2.91	0.79931
R183Q	3.69	0.79333	R266W	3.68	0.79166
H189L	3.08	0.76921	L289R	3.63	0.79791
P194S	2.83	0.77411	R791C	3.89	0.76935
D195Y	3.05	0.78802	R791H	3.70	0.72214
D195H	3.56	0.72915			

Post-Transcriptional Modifications

Table 5. Prediction of Methylation sites in PTPN22 protein via GPS3.0 and NetPhos3.1

Position	Peptide	Met-Types	Score
799	FSKPKGPRNPPPTWN	R.mono	17.69

Table 6. NetPhos and GPS 3.0 predictions

NetPhos 3.1	Position	Score	Kinase	GPS 3.0 PEPTIDE	KINASE	SCORE
Serine	16	0.968	unsp	KFLDEAQSKKITKEE	AGC	0.6959
Serine	35	0.994	unsp	FLKLKRQSTKYKADK	AGC	0.5828
Serine	69	0.499	cdc2	DILPYDYSRVELSLI	AGC	0.3874
Serine	74	0.704	Unsp	DYSRVELSLITSDED	AGC	0.5638
Serine	78	0.993	unsp	VELSLITSDEDSSYI	AGC	0.2808
Serine	107	0.973	unsp	IATQGPLSTTLLDFW	AGC	0.1709
Serine	121	0.468	CaM-II	WRMIWEYSVLIIVMA	AGC	0.3832

NetPhos 3.1	Position	Score	Kinase	GPS 3.0 PEPTIDE	KINASE	SCORE
Serine	157	0.606	unsp	QLEFGPFSVSCEAEK	AGC	0.3031
Serine	159	0.896	unsp	EFGPFSVSCEAEKRK	AGC	0.3254
Serine	167	0.994	unsp	CEAEKRKSDYIIRTL	AGC	0.4887
Serine	180	0.958	unsp	TLKVKFNSETRTIYQ	AGC	0.4062
Serine	200	0.987	unsp	RCYQEDDSVPICIHC	AGC	0.3601
Serine	220	0.469	Cdc2	VPICIHCSAGCGRTG	AGC	0.4839
Serine	228	0.464	GSK3	GIIPENFSVFSLIRE	AGC	0.4477
Serine	257	0.448	GSK3	PENFSVFSLIREMRT	AGC	0.3908
Serine	260	0.590	CaM-II	EMRTQRPSLVQTQQ	AGC	0.3786
Serine	271	0.981	unsp	DVIRDKHSGTESQAK	AGC	0.4798
Serine	302	0.996	unsp	DKHSGTESQAKHCIP	AGC	0.3410
Serine	306	0.649	DNAPK	TKMEIKESSSFDFRT	AGC	0.5332
Serine	351	0.813	unsp	KMEIKESSSFDFRTS	AGC	0.5496
Serine	352	0.986	unsp	MEIKESSSFDFRTSE	AGC	0.5465
Serine	353	0.607	unsp	SSFDFRTSEISAKEE	AGC	0.6019
Serine	359	0.994	unsp	DFRTSEISAKEELVL	AGC	0.4943
Serine	362	0.997	unsp	DFLELNYSFDKNADT	AGC	0.2688
Serine	386	0.473	cdc2	SLLFEGCSNSKPVNA	AGC	0.2168
Serine	426	0.531	cdc2	LFEGCSNSKPVNAAG	AGC	0.4091
Serine	428	0.567	cdc2	AAGRYFNSKVPITRT	AGC	0.4439
Serine	440	0.455	cdc2	VPITRTKSTPFELIQ	AGC	0.4644
Serine	449	0.972	unsp	RETKEVDSKENFSYL	AGC	0.3689
Serine	465	0.996	unsp	QKVMHVSSAELNYSL	AGC	0.1540
Serine	494	0.448	GSK3	SSAELNYSLPYDSKH	AGC	0.4225
Serine	500	0.459	DNAPK	NYSLPYDSKHQIRNA	AGC	0.4110
Serine	505	0.800	unsp	KHQIRNASNVKHHDS	AGC	0.4814
Serine	513	0.512	cdc2	SNVKHHDSSALGVYS	AGC	0.5439
Serine	520	0.627	PKA	NVKHHDSSALGVYSY	AGC	0.4285
Serine	600	0.494	cdc2	SSLLNQESAVLATAP	AGC	0.3486
Serine	643	0.578	PKC	PNVPKSLSSAVKVKI	AGC	0.1609
Serine	644	0.696	PKC	NVPKSLSSAVKVKIG	AGC	0.2731
Serine	653	0.854	unsp	VKVKIGTSLEWGGTS	AGC	0.1762
Serine	668	0.613	unsp	EPKKFDDSVILRPSK	AGC	0.6106
Serine	674	0.598	PKC	DSVILRPSKSVKLRS	AGC	0.4659
Serine	676	0.830	PKC	VILRPSKSVKLRSPK	AGC	0.4019
Serine	681	0.996	unsp	SKSVKLRSPKSELHQ	AGC	0.3732
Serine	684	0.906	unsp	VKLRSPKSELHQDRS	AGC	0.4351
Serine	691	0.538	PKC	SELHQDRSSPPPPLP	AGC	0.2852
Serine	692	0.987	unsp	ELHQDRSSPPPPLPE	AGC	0.3263
Serine	704	0.517	CKI	LPERTLESFFLADED	AGC	0.3640
Serine	732	0.526	cdc2	YPDTMENSTSSKQTL	AGC	0.1536
Serine	734	0.989	unsp	DTMENSTSSKQTLKT	AGC	0.1623
Serine	735	0.547	CKI	TMENSTSSKQTLKTP	AGC	0.4578
Serine	745	0.984	unsp	TLKTPGKSFTRSKSL	AGC	0.3174
Serine	749	0.626	PKC	PGKSFTRSKSLKILR	AGC	0.3311
Serine	751	0.995	unsp	KSFTRSKSLKILRNM	AGC	0.2999
Serine	765	0.715	PKC	MKKSICNSCPPNKPA	AGC	0.4823
Serine	781	0.459	GSK3	SVQSNNSSSFLNFGF	AGC	0.4045
Serine	782	0.501	cdc2	VQSNNSSSFLNFGFA	AGC	0.2910
Serine	793	0.899	Unsp	FGFANRFSKPKGPRN	AGC	0.4361
Threonine	20	0.600	PKG	EAQSKKITKEEFANE	AGC	0.6997
Threonine	36	0.970	unsp	LKLKRQSTKYKADKT	AGC	0.7209
Threonine	43	0.445	Cdc2	TKYKADKTYPTTVAE	AGC	0.3965
Threonine	77	0.619	CKII	RVELSLITSDEDSSY	AGC	0.3015
Threonine	102	0.610	DNAPK	GPKAYIATQGPLSTT	AGC	0.4343
Threonine	109	0.511	cdc2	TQGPLSTTLLDFWRM	AGC	0.1513
Threonine	173	0.902	PKC	KSDYIIRTLKVKFNS	AGC	0.4708
Threonine	182	0.469	PKG	KVKFNSETRTIYQFH	AGC	0.5079
Threonine	184	0.551	PKC	KFNSETRTIYQFHYK	AGC	0.5756
Threonine	234	0.435	CaM-II	CSAGCGRTGVICAID	AGC	0.4431
Threonine	243	0.452	unsp	VICAIDYTWMLLKDG	AGC	0.1624
Threonine	267	0.981	PKC	SLIREMRTQRPSLVQ	AGC	0.3904
Threonine	275	0.567	unsp	QRPSLVQTQEQYELV	AGC	0.6010
Threonine	304	0.549	PKC	IRDKHSGTESQAKHC	AGC	0.3483
Threonine	332	0.899	PKC	SPNLPKSTTKAAKMM	AGC	0.3759
Threonine	333	0.800	PKC	PNLPKSTTKAAKMMN	AGC	0.1559
Threonine	344	0.635	PKC	KMMNQQRTKMEIKES	AGC	0.4856
Threonine	358	0.438	GSK3	SSSFDFRTSEISAKE	AGC	0.5875
Threonine	376	0.462	cdc2	LHPAKSSTSFDFLEL	AGC	0.2359
Threonine	393	0.522	PKC	SFDKNADTTMKWQK	AGC	0.1757
Threonine	399	0.659	PKC	DTTMKWQTKAFPIVG	AGC	0.1524
Threonine	445	0.604	PKC	FNSKVPITRTKSTPF	AGC	0.4234
Threonine	447	0.459	CaM-II	SKVPITRTKSTPFEL	AGC	0.3722
Threonine	450	0.973	unsp	PITRTKSTPFELIQQ	AGC	0.4024
Threonine	460	0.804	unsp	ELIQQRETKEVDSKE	AGC	0.6583
Threonine	545	0.642	PKC	SSWPPSGTSSKMSD	AGC	0.1581
Threonine	605	0.456	cdc2	QESAVLATAPRIDDE	AGC	0.4996
Threonine	621	0.742	unsp	PPPLPVWTPESFIVV	AGC	0.1608

NetPhos 3.1	Position	Score	Kinase	GPS 3.0 PEPTIDE	KINASE	SCORE
Threonine	652	0.538	PKC	AVKVKIGTSLEWGGT	AGC	0.1632
Threonine	659	0.716	unsp	TSLEWGGTSEPKKFD	AGC	0.4669
Threonine	701	0.528	PKG	PPPLPERTLESFFLA	AGC	0.3111
Threonine	733	0.530	PKC	PDTMENSTSSKQTLK	AGC	0.1630
Threonine	738	0.889	PKC	NSTSSKQTLKTPGKS	AGC	0.6408
Threonine	741	0.608	cdk5	SSKQTLKTPGKSFTR	AGC	0.4392
Threonine	747	0.547	PKC	KTPGKSFTRSKSLKI	AGC	0.4779
Threonine	804	0.481	GSK3	GPRNPPPTWNI****	AGC	0.3687

Table 7. BDM-PUB predictions for Ubiquitination

Peptide	Position	Score	Threshold
FLDEAQSKKITKEEF	17	0.93	0.3
LDEAQSKKITKEEFA	18	0.60	0.3
AQSKKITKEEFANEF	21	1.23	0.3
EFANEFLKLKRQSTK	30	1.16	0.3
ANEFLKLKRQSTKYK	32	1.46	0.3
KLKRQSTKYKADKTY	37	0.94	0.3
STKYKADKTYPTTVA	42	1.13	0.3
YPTTVAEKPKNIKKN	51	0.93	0.3
TTVAEKPKNIKKNRY	53	0.95	0.3
AEKPKNIKKNRYKDI	56	2.31	0.3
EKPKNIKKNRYKDIL	57	1.05	0.3
SVSCEAEKRKSDYII	164	1.55	0.3
SCEAEKRKSDYIIRT	166	0.74	0.3
DYIIRTLKVKFNSET	175	0.63	0.3
IIRTLKVKFNSETRT	177	0.33	0.3
NAVLELFKRQMDVIR	291	0.76	0.3
QMDVIRDKHSGTESQ	300	0.55	0.3
SGTESQAKHCIPEKN	309	0.83	0.3
AKHCIPEKNHTLQAD	315	1.44	0.3
SYSPNLPKSTTKAAK	330	1.69	0.3
NLPKSTTKAAKMMNQ	334	1.17	0.3
KSTTKAAKMMNQQRT	337	2.33	0.3
MMNQQRTKMEIKESS	345	0.97	0.3
ELVLHPAKSSTSFDF	373	0.90	0.3
IVGEPLQKHQSLDLG	411	0.30	0.3
FEGCSNSKPVNAAGR	429	1.00	0.3
LIQQRETKEVDSKEN	461	0.43	0.3
YSLPYDSKHQIRNAS	506	0.40	0.3
IRNASNVKHHDSSAL	516	0.36	0.3
PPSGTSSKMSLDLPE	548	2.77	0.3
EFSPNVPKSLSSAVK	640	0.87	0.3
KSLSSAVKVKIGTSL	647	2.13	0.3
LSSAVKVKIGTSLEW	649	0.68	0.3
WGGTSEPKKFDDSVI	663	0.74	0.3
GGTSEPKKFDDSVIL	664	1.58	0.3
SVILRPSKSVKLRSP	675	3.47	0.3
LRPSKSVKLRSPKSE	678	3.01	0.3
SVKLRSPKSELHQDR	683	2.39	0.3
MENSTSSKQTLKTPG	736	3.14	0.3
TSSKQTLKTPGKSFT	740	1.87	0.3
QTLKTPGKSFTRSKS	744	1.89	0.3
GKSFTRSKSLKILRN	750	1.72	0.3
FTRSKSLKILRNMKK	753	1.64	0.3
LKILRNMKKSICNSC	759	0.99	0.3
GFANRFSKPKGPRNP	794	1.54	0.3
ANRFSKPKGPRNPPP	796	1.23	0.3

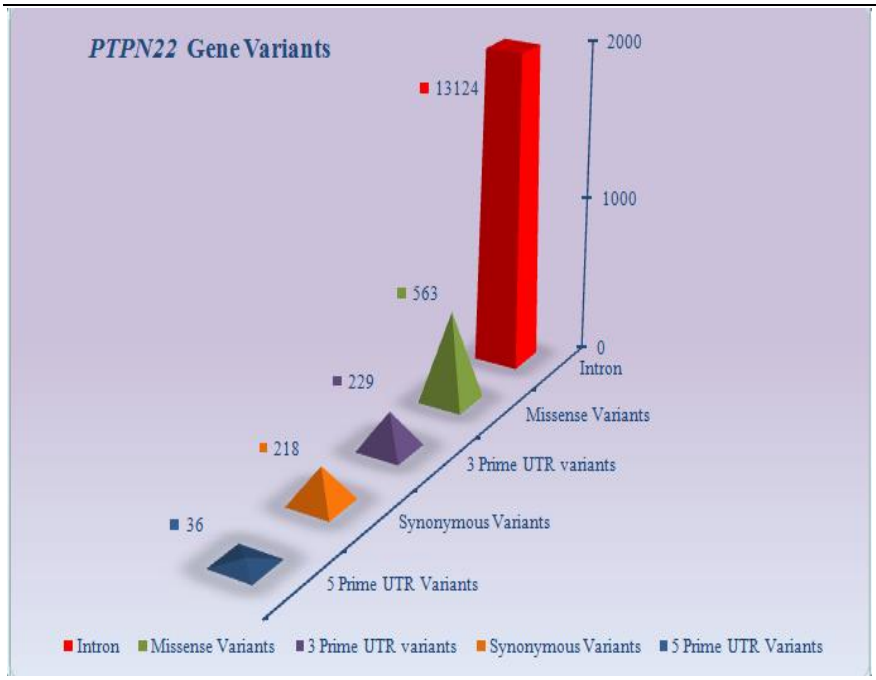

Figure 1: A 3D bar chart illustrating SNPs linked with the PTPN22 gene, showing the distribution across categories: intronic (13,124, red bar), missense variants (563, green bar), 3' prime UTR variants (229, purple bar), synonymous variants (218, orange bar), and 5' prime UTR variants (36, blue bar). The y-axis scales up to 2000 for visual emphasis on larger categories.

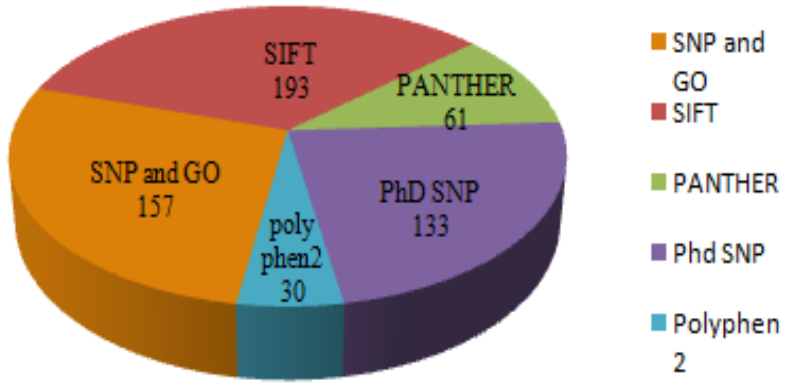

Figure 2: A pie chart showing the prediction of damaging nsSNPs via several tools, with segments labeled: SNP and GO (157, orange), SIFT (193, red), PANTHER (61, green), PhD SNP (133, purple), and PolyPhen 2 (30, blue).

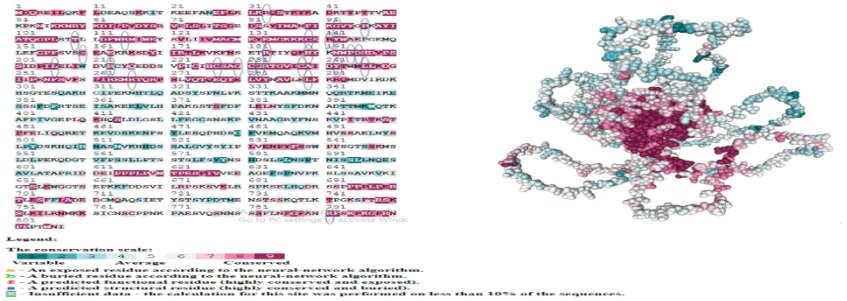

Figure 3: A detailed Consurf visualization for harmful nsSNPs, showing the PTPN22 protein sequence with conservation scale (1-9, variable to conserved) and 3D structure colored by conservation (cyan: variable, magenta: conserved), with annotations for exposed/buried residues and a legend explaining symbols for functional/structural importance.

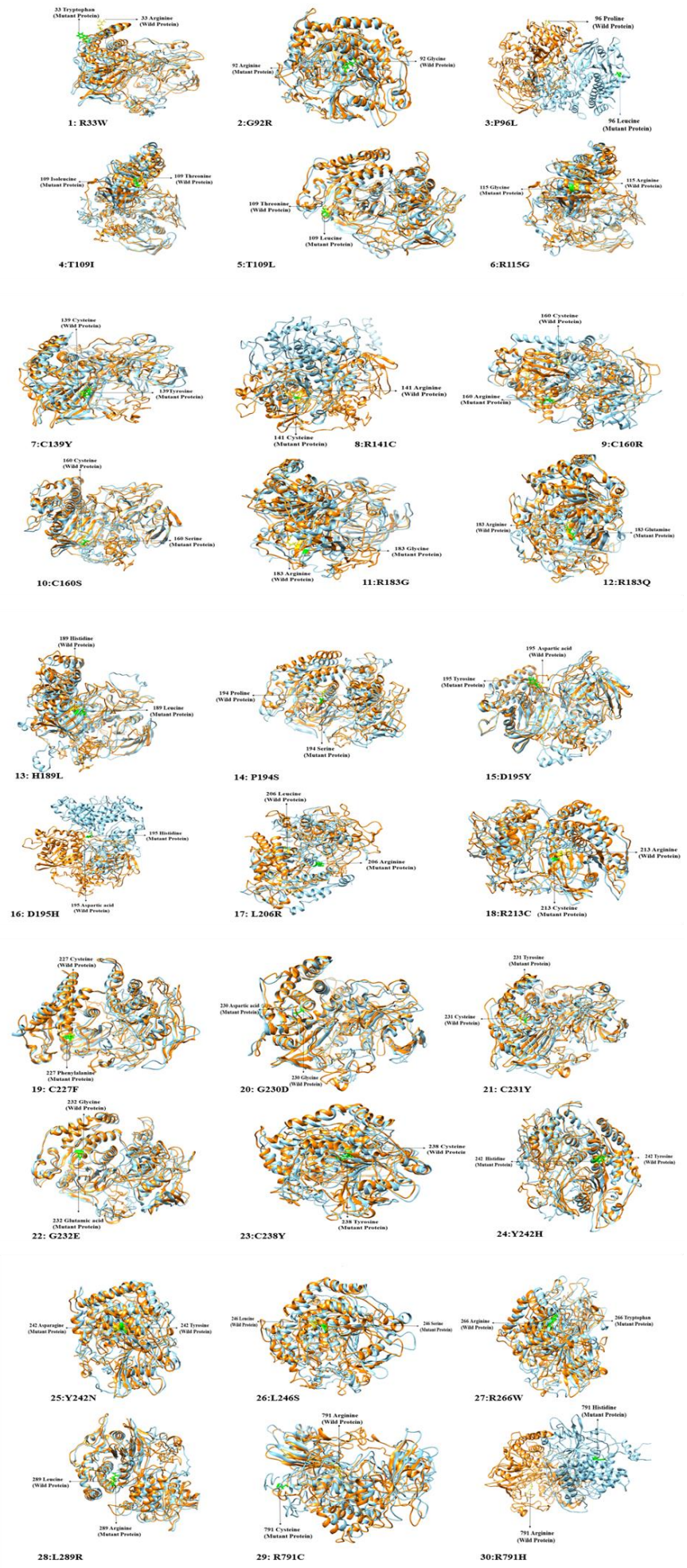
1: R33W
2:G92R
3:P96L
4:T109I
5:T109L
6:R115G
7:C139Y
8:R141C
9:C160R
10:C160S
11:R183G
12:R183Q
13: H189L
14: P194S
15:D195Y
16: D195H
17: L206R
18:R213C
19: C227F
20: G230D
21: C231Y
22: G232E
23:C238Y
24:Y242H
25:Y242N
26:L246S
27:R266W
28:L289R
29: R791C
30:R791H

Figure 4: A grid of 3D structures for deleterious nsSNPs generated by I-TASSER, displaying individual models for each variant (e.g., R33W, G92R) with wild-type and mutant forms highlighted in different colors, including labels for structural changes and a scale bar for orientation.

Gene-Gene Interaction

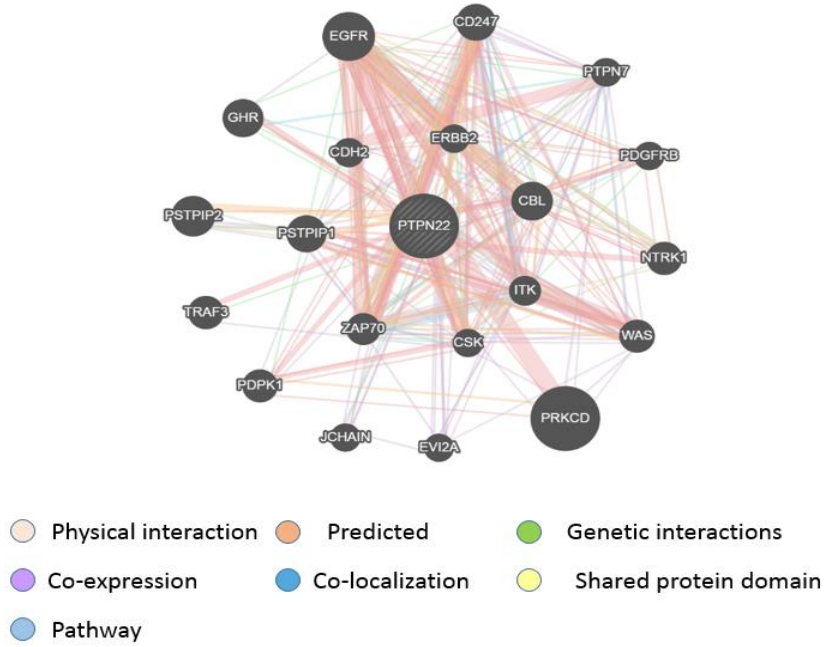

Figure 5: A network diagram of gene-gene interactions for PTPN22 using GeneMANIA, with central PTPN22 node connected to others (e.g., TRAF3, ZAP70) via colored edges: pink (physical), orange (co-expression), blue (co-localization), green (genetic), yellow (shared domain), and purple (pathway).

Protein-Protein Interaction

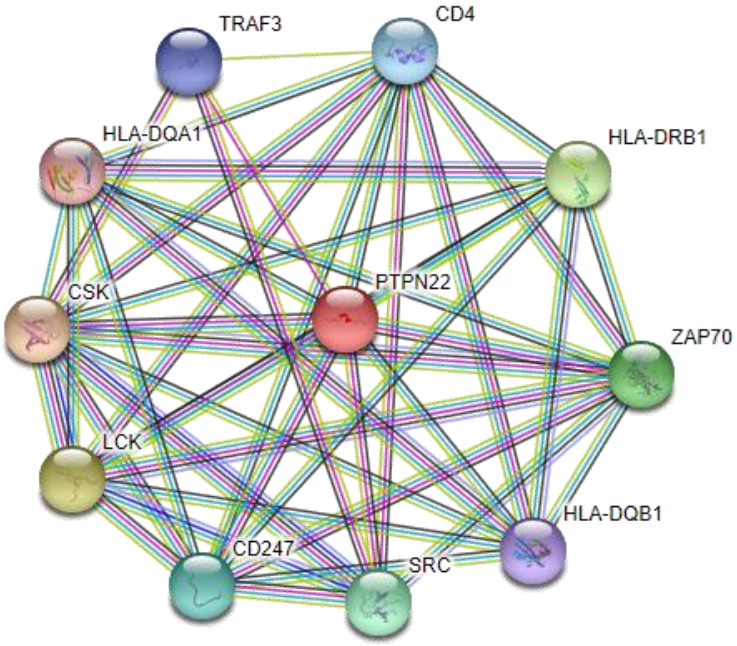

Figure 6 Placeholder: A network diagram of protein-protein interactions for PTPN22 using STRING, with PTPN22 as the central red node linked to proteins like TRAF3, CSK, ZAP70, CD4, and HLA variants via multicolored edges representing interaction types, including a legend for node and edge meanings.

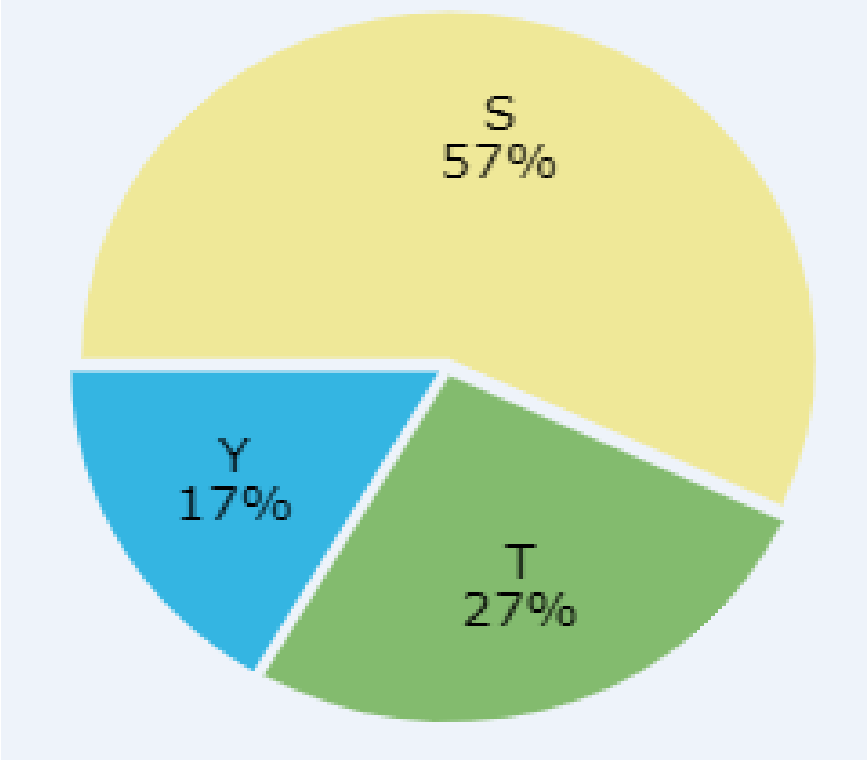

Figure 7: A pie chart illustrating the distribution of S/T/Y p-sites: Serine (57%, yellow segment), Threonine (27%, green segment), and Tyrosine (17%, blue segment).