In Silico Analysis: HLA-DRB1 Gene's Variants and Their Clinical Impact

Mohamed M. Hassan[1] [iD]**, Mohamed A. Hussain[2],
Sababil S. Ali[1], and Mohammed A. Mahdi[1]**

Abstract

The *HLA-DRB1* gene encodes a protein that is essential for the immune system. This gene is important in organ transplant rejection and acceptance, as well as multiple sclerosis, systemic lupus erythematosus, Addison's disease, rheumatoid arthritis, caries susceptibility, and Aspirin-exacerbated respiratory disease. The following *Homo sapiens* variants were investigated: single-nucleotide variants (SNVs), multi-nucleotide variants (MNVs), and small insertions–deletions (Indels) in the *HLA-DRB1* gene via coding and untranslated regions. The current study sought to identify functional variants that could affect gene expression and protein product function/structure. ALL target variants available until April 14, 2022, were obtained from the Single Nucleotide Polymorphism database (dbSNP). Out of all the variants in the coding region, 91 nsSNVs were considered highly deleterious by seven prediction tools and instability index; 25 of them are evolutionary conserved and located in domain regions. Furthermore, 31 indels were predicted as harmful, potentially affecting a few amino acids or even the entire protein. Last, within the coding sequence (CDS), 23 stop-gain variants (SNVs/indels) were predicted as high impact. High impact refers to the assumption that the variant will have a significant (disruptive) effect on the protein, likely leading to protein truncation or loss of function. For untranslated regions, functional 55 single-nucleotide polymorphisms (SNPs), and 16 indels located within microRNA binding sites, furthermore, 10 functionally verified SNPs were predicted at transcription factor-binding sites. The findings demonstrate that employing *in silico* methods in biomedical research is extremely successful and has a major influence on the capacity to identify the source of genetic variation in diverse disorders. In conclusion, these previously functional identified variants could lead to gene alteration, which may directly or indirectly contribute to the occurrence of many diseases. The study's results could be an important guide in the research of potential diagnostic and therapeutic interventions that require experimental mutational validation and large-scale clinical trials.

Keywords

in silico analysis, *HLA-DRB1* variants, SNVs, SNPs and Indels, *HLA-DRB1* gene, single-nucleotide variants, organ transplantation, inflammatory and autoimmune diseases

Introduction

The human leukocyte antigen (HLA) system is the name of the major histocompatibility complex (MHC) in human, generally inherited from parents as a set name haplotype. *HLA* genes are located on chromosome: 6p (short arm) in the distal portion of the 21.3 band[1]. The HLA system spans a 4 Megabyte (4×10^6 nucleotides) region of the human genome, one of the most polymorphic and gene-dense regions[2]. HLA genes have an important contribution to the immune system and contain several alleles that differ substantially among human populations. The HLA locus has been a focal point of genomic research and clinical practice for several reasons: (1) It is linked to several inflammatory and autoimmune diseases; (2) it is extremely suitable for human genetic diversity studies; and (3) it is critical in tissue and organ transplantation donor–recipient matches[3]. The *HLA* complex genes and

their protein products have been divided into three classes on the basis of their tissue distribution, structure, and function. MHC class II antigens encoded by genes *HLA-DM, HLA-DO, HLA-DP, HLA-DQ, HLA-DR* loci, and their products are involved in list of the immunoglobulin supergene family[4,5]. The *HLA-DR* gene encodes two distinct subunits, DRA (alpha chain) and DRB (beta chain). *HLA-DRB1* is a

[1]Faculty of Medical Laboratory Sciences, National University, Khartoum, Sudan
[2]Faculty of Pharmacy, International University of Africa, Khartoum, Sudan

Corresponding Author:
Mohamed M. Hassan, Faculty of Medical Laboratory Sciences, National University, AlRaqi District, Khartoum 11111, Sudan.
Email: mhassan0210@gmail.com

protein-coding gene that belongs to the HLA class II beta chain (approximately 26–28 kDa) paralogs, and it is found on the cell surface[2,6].

The *HLA-DRB1* gene is located in GRCh38 (Genome Reference Consortium Human Build 38) coordinates 32,578,775 to 32,589,848, has five introns, and is encoded by six exons. Exon 1 encodes the leader peptide; exons 2 and 3 encode the two extracellular domains; exon 4 encodes the transmembrane domain; and exon 5 encodes the cytoplasmic tail[7] (https://www.ncbi.nlm.nih.gov/gene/3123). Compared with its paralogs *DRB3, DRB4*, and *DRB5, DRB1* is expressed at a level that is five times higher[8]. *HLA* genes region is the most polymorphic in the human genome, and the *HLA-DRB1* gene is the most polymorphic in class II of this system[9,10]. The *HLA-DRB1* locus had 3,196 alleles in May 2022, according to the IPD-IMGT/HLA database[11] (https://www.ebi.ac.uk/ipd/imgt/hla/about/statistics/). Many *HLA-DRB1* alleles (a gene's variant forms) have been associated with various diseases. HLA- DRB1*1501[12,13], DRB1*03[14], DRB1*0404[15], DRB1*04:05[16], DRB1*13[17], and DRB1*04[18,19] alleles have been associated with multiple sclerosis[12,13], systemic lupus erythematosus[14], Addison's disease[15], rheumatoid arthritis[16], caries susceptibility[17], graft survival in organ transplant recipients[18], and Aspirin-exacerbated respiratory disease[19]. The 1,000 genome project revealed that single-nucleotide polymorphisms (SNPs) account for the majority of human genetic variation[20].

SNPs are single-nucleotide variants (SNVs) in DNA sequence with a population allele frequency of 1% or higher. It normally occurs throughout the genome with the frequency of about one of each 600 to 1,000 nucleotide, which is considered the simplest and common type of genetic marker leading to DNA variation among individuals[21,22]. Nonsynonymous SNPs (nsSNP) are a type of SNP that represents amino acid substitutions and protein variations in humans. Previous research indicates that nsSNPs account for roughly half of the mutations involved in various genetic diseases[23]. Other important types of genomic variation are indels, which are insertions or deletions of one or more nucleotides in the DNA sequence[24].

The SNP Database (dbSNP) is one of the NCBI's subdivided databases that contain human single-nucleotide variations, microsatellites, and small-scale insertions and deletions. SNP database contains 1,076,992,604 *Homo sapiens* variants as of May 28, 2022. There were 957,193,110 SNPs, SNVs, or MNVs (multi-nucleotide variants) among the total number of variants, and 29,620,962 Indels (single or small length insertions–deletions). (https://www.ncbi.nlm.nih.gov/snp/). Functional variants within coding regions may affect protein structure and function, whereas non-coding variants may have an impact on protein expression[25,26]. Pathological non-coding variants could have an alteration role in various regulatory functions within the genome, such as interacting with transcription factors (TFs), and microRNA

(miRNA)[27]. Identification of variants responsible for phenotypic changes is considered difficult, as it necessitates multiple tests for different variants in candidate genes[8,27,28]. One possible solution would be to prioritize variants based on their structural and functional significance using various bioinformatics prediction tools. The use of computational methods for gaining biological insight is well established[29–33]. Thus, the current study aimed to *in silico* analyze all human SNVs, MNVs, and short Indels in the *HLA-DRB1* gene's coding and untranslated regions to significantly predict functional variants that could affect gene expression and protein product function/structure.

Materials and Methods

Variants Dataset

HLA-DRB1 gene variants were discovered using the NCBI SNP database (https://www.ncbi.nlm.nih.gov/SNP/) on April 14, 2022. The *HLA-DRB1* variants (SNPs, SNVs, MNVs, and INDELs) were retrieved from the SNP database build 155 and mapped on genome assembly GRCh38 using Variation Viewer (https://www.ncbi.nlm.nih.gov/variation/view/). Variants in coding and 3′/5′ untranslated regions have been identified for computational analysis of their effect(s). Several tools have been used to improve the accuracy and reliability of identifying pathogenic variants and their effects on the structure, function, and expression of *HLA-DRB1* (Fig. 1).

Coding Variants Analysis (nsSNPs/nsSNVs, Indels, Stop Gain, and MNVs)

To identify the most deleterious missense or nsSNVs, seven distinct bioinformatics tools, namely, SIFT (Sorting Intolerant From Tolerant), PolyPhen, PredictSNP, Panther (Protein Analysis Through Evolutionary Relationships), SNP&GO (Gene Ontology), PROVEAN (Protein Variation Effect Analyzer), and SNAP2, have been used[34–40]. All nsSNVs identified as harmful by the previous seven tools and predicted as instabilities by the I-mutant server are categorized as high risk (Table 2)[41]. Among the total high-risk variants, nsSNVs with high evolutionary conservation and located in domain sites were chosen (Table 4). InterPro database and the Consurf server were used to identify domains and high evolutionary conservation (grade ≥ 6) amino acids (Table 3 and Fig. 2)[42,43]. To understand the effect of nsSNVs on protein structure, HOPE tool using sequence and missense-3D server using structure model were used (Tables 5 and 6 and Fig. 3)[44,45]. The related protein sequence was (accession number: P01911) obtained from Uniprot database (http://www.uniprot.org). Phyre2 and Swiss-Model servers were used to predict the protein models[46,47]. To select the high-quality model, two evaluation tools [PSICA (Protein
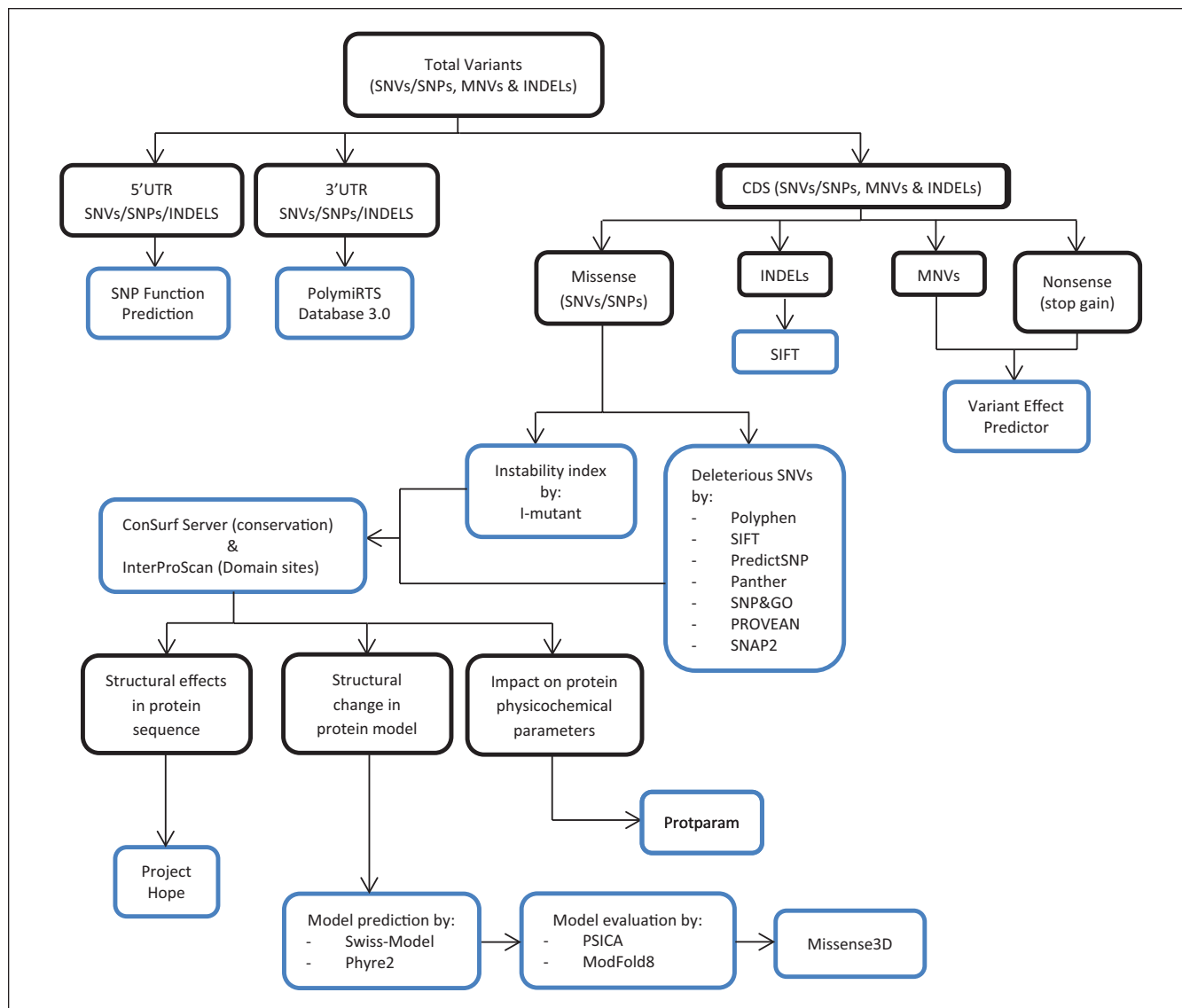
Figure 1. Flowchart for the *in silico* analysis of variants in the *HLA-DRB1* gene and their biological consequences. The black shapes represent the type of data, while the blue shapes represent the names of the prediction tools. SNP: single-nucleotide polymorphism; SNV: single-nucleotide variant; MNV: multi-nucleotide variant; INDEL: insertion–deletion; SIFT: Sorting Intolerant From Tolerant; PANTHER: Protein Analysis Through Evolutionary Relationships; GO: Gene Ontology; PROVEAN: Protein Variation Effect Analyzer.

Structural Information Conformity Analysis) and ModFOLD8] were used (Figs. 4 and 5)[48,49]. For more investigation in the coding regions, indel was entered into the SIFT algorithm to anticipate their functional effect (Table 7). Furthermore, SNVs/indels that result in a premature stop codon (stop gain) and MNVs were submitted to Variant Effect Predictor to assess the impact of this change (Table 8)[50]; https://www.ensembl.org/info/docs/tools/vep/index. html. The ProtParam server was then used to assess the impact of conserved and domain-located nsSNVs on protein physicochemical parameters (Table 9)[51].

SIFT server. Make an alignment between an order sequence with a large number of homologous sequences to predict if an amino acid substitution will have a phenotypic effect. The Residual's score ranges from zero to one. If the score is less than or equal to 0.05, the amino acid substitution is predicted to be harmful; if the score is greater than 0.05, the substitution is tolerated[34]; https://sift.bii.a-star.edu.sg/.

PolyPhen-2 (Polymorphism Phenotyping v2) server. A tool uses simple physical and comparative considerations to predict the impact of an amino acid substitution on the structure and
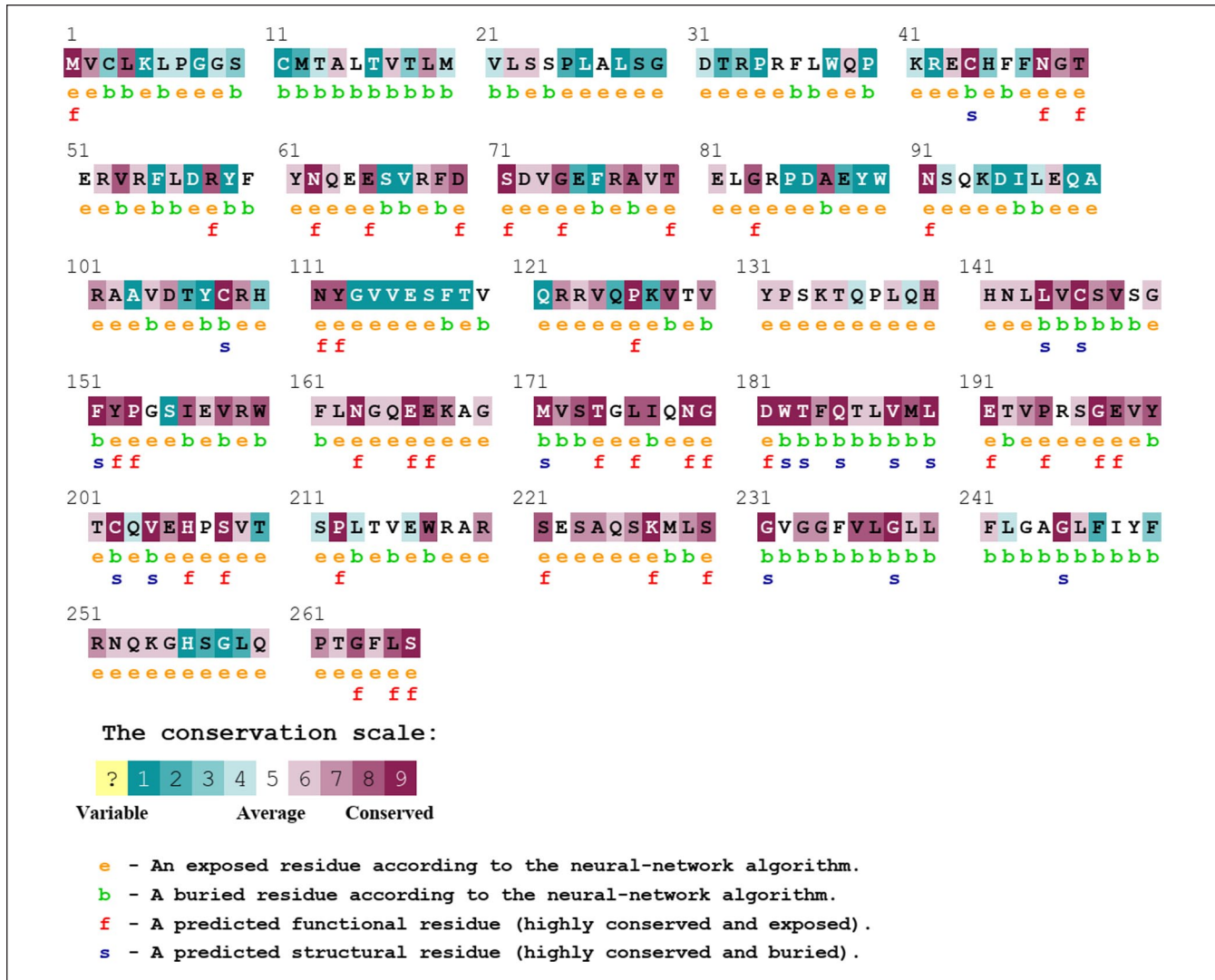
Figure 2. Evolutionary conservancy of *HLA-DRB1* produced by Consurf server.

function of a human protein. A mutation is classified qualitatively, as benign, possibly damaging, or probably damaging[35]. http://genetics.bwh.harvard.edu/pph2/.

PredictSNP tool. The server was developed by combining six disease-related mutation prediction programs. The predicted effect is color-coded: Neutral mutations are green, while deleterious mutations are red[36]; https://loschmidt.chemi.muni.cz/predictsnp/.

PANTHER (Protein Analysis Through Evolutionary Relationships). This classification system was designed to classify proteins (and their genes) to facilitate high-throughput analysis. Proteins have been classified according to family/subfamily, molecular function, and biological process. The tool assesses the functional effects of nsSNPs, with three possible outcomes: probably benign, possibly damaging, and

probably damaging[37]. PANTHER computes the length of time (in millions of years) that a given amino acid has been preserved in the lineage that led to the protein of interest. The longer the preservation time, the more likely it is that functional impact will occur. The method is known as PANTHER-PSEP (position-specific evolutionary preservation). The preservation time outputs are classified as >450, between 200 and 450, and <200 million years, corresponding to probably damaging, possibly damaging, and probably benign[37]; http://www.pantherdb.org/tools/csnpScoreForm.jsp

SNP&GO (Gene Ontology). The server is based on Support Vector Machines (SVM) and has been optimized to predict if a given single-point protein variation can be classified as disease-associated or neutral[38]; https://snps.biofold.org/snps-and-go/snps-and-go.html
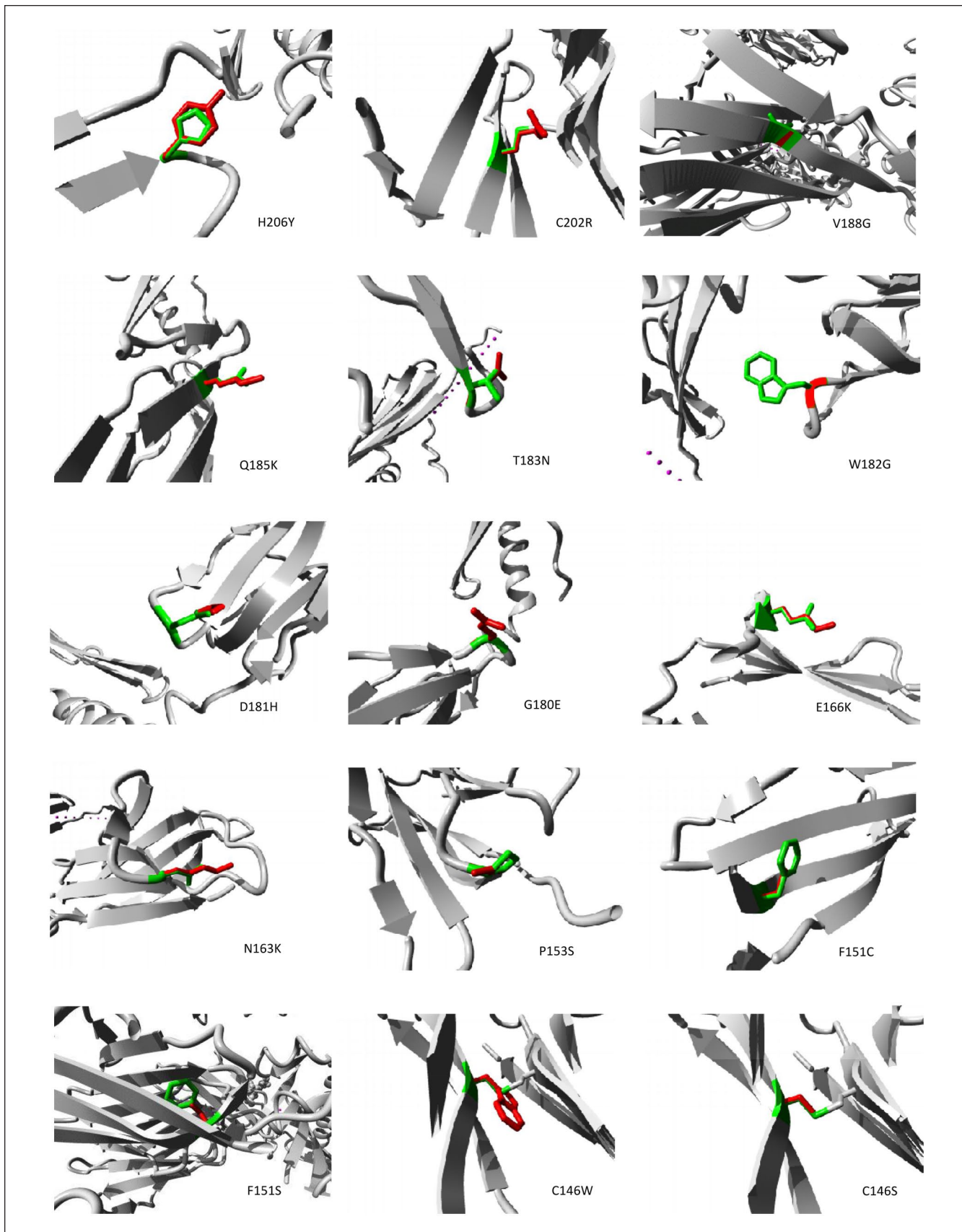
Figure 3. (continued)
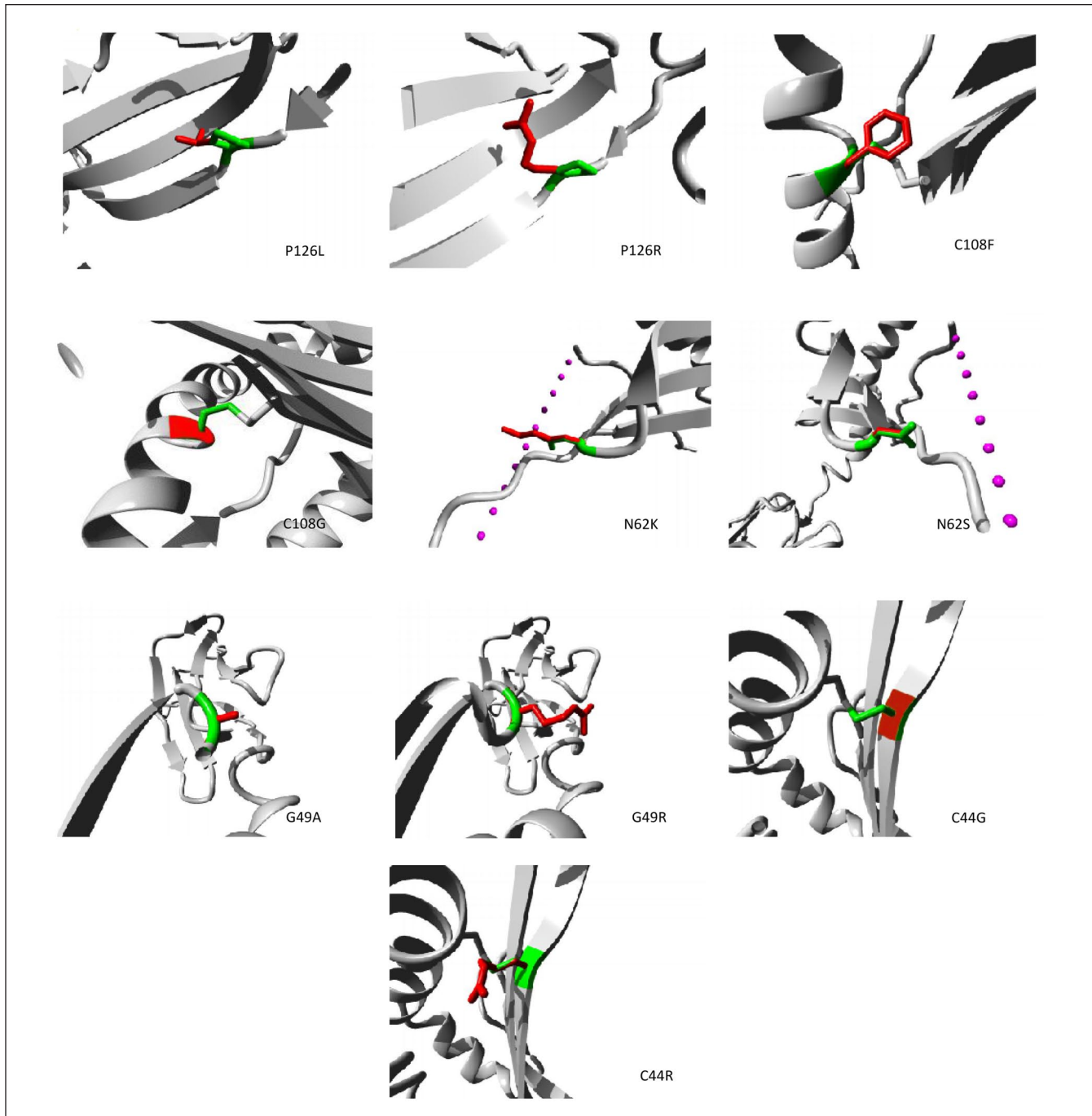
Figure 3. Structural alteration by HOPE server. The protein is shown in gray, the wild type residue in green, and the mutant residue in red.

PROVEAN (Protein Variation Effect Analyzer). It is a software tool that predicts whether an amino acid substitution or indel will affect a protein's biological function. PROVEAN can be used to filter sequence variants to identify non-synonymous or indel variants that are predicted to be functionally important[39]. The PROVEAN prediction score classifies the

substitution as having a deleterious or neutral effect on protein function; http://provean.jcvi.org/index.php

SNAP2 server. A trained classification algorithm based on a machine learning device known as a neural network. SNAP2 predicts the impact (effect) of single amino acid substitutions

GlobalScore=0.6489   GlobalScore=0.6598
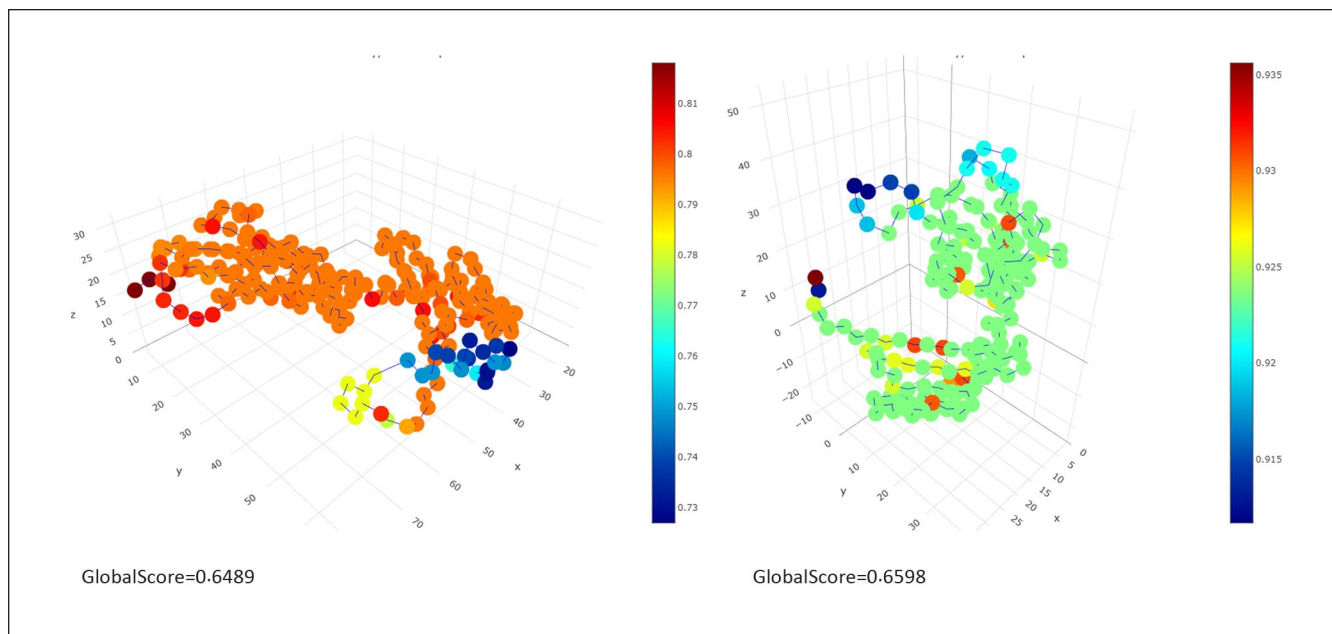
Figure 4. Protein models evaluation using PSICA server. The illustration on the left represents the PHYRE2-server model, while the right represents the SWISS-MODEL structure. PSICA: Protein Structural Information Conformity Analysis.
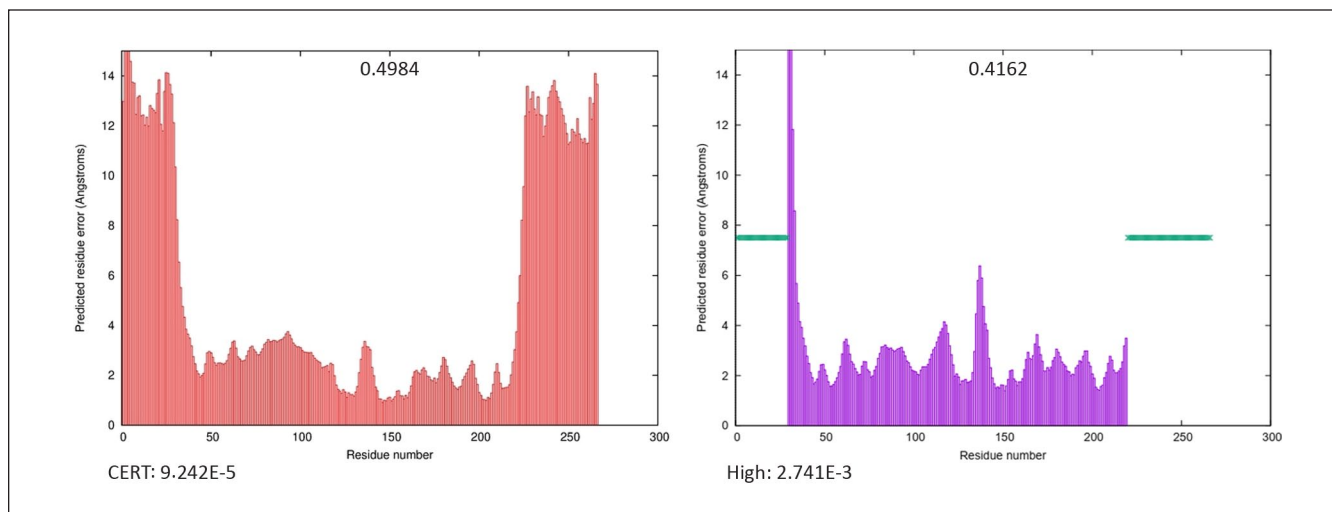
Figure 5. Protein models evaluation using ModFOLD8. The illustration on the left represents the PHYRE2-server model, while the right represents the SWISS-MODEL structure. The upper number represents the global model quality score, while the lower represents the confidence and *P* value.

on protein function. The prediction score ranges from −100 (strong neutral) to +100 (strong effect). According to the findings, the prediction score is to some extent correlated to the severity of effect[40]; https://www.rostlab.org/services/snap/

I-mutant server. I-Mutant v3.0 is a suite of SVM-based predictors integrated in a unique web server. It offers the

opportunity to predict the protein stability changes upon single-site variations from the protein structure or sequence. The I-mutant result is either decrease/increase stability or neutral[41]; http://gpcr2.biocomp.unibo.it/cgi/predictors/I-Mutant3.0/I-Mutant3.0.cgi

InterPro database. It performs functional protein analysis by categorizing them into families and predicting domains and

key locations. InterPro employs prediction models, known as signatures, offered by several databases to categorize proteins in this manner[42]; https://www.ebi.ac.uk/interpro/

Consurf server. It is a bioinformatics tool that uses phylogenetic relationships between homologous sequences to estimate the evolutionary conservation of amino/nucleic acid positions in a protein/DNA/RNA molecule. Position-specific conservation scores are computed using the empirical Bayesian or ML algorithms. For illustration, the continuous conservation scores are grouped into nine categories, ranging from the most changeable places (grade 1) in turquoise to the most conserved positions (grade 9) in maroon[43]; https://consurf.tau.ac.il/

HOPE server. An automatic mutant analysis server can provide information about a mutation's structural effects. HOPE gathers information from a wide variety of sources. Data are stored in a database and used in a decision scheme to determine the effects of a mutation on the protein's 3D structure and function. HOPE's final report includes discovered data on contacts (metal, DNA, hydrogen bonds, ionic interactions, etc.), structural locations (motifs, domains, transmembrane domains, etc.), non-structural features (post-translational modifications), known variants at that position, and amino acid physicochemical properties (size, charge, and hydrophobicity). HOPE creates an easy-to-use and understandable report with text, figures, and animations[44]; https://www3.cmbi.umcn.nl/hope/

Phyre2 and Swiss-Model tools. Protein structure prediction automated servers. Both algorithms are based on comparative modeling methods. Phyre2 uses the alignment of hidden Markov models via HHsearch to significantly improve accuracy of alignment and detection rate. Phyre2 also could use AB-initio method to determine the tertiary structure of protein in the absence of experimentally solved structure[46,47]; http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index, https://swissmodel.expasy.org/.

PSICA and ModFOLD v.8. Both are protein structure quality assessment servers. PSICA is the official implementation of MUfoldQA_S and MUfoldQA_C methods. It is designed to evaluate how much a tertiary model of a given protein primary sequence conforms to the known protein structures of a similar protein[48]. ModFOLD8 combines the strengths of multiple pure-single and quasi-single model methods to predict global and local quality of 3D protein models. The global model quality scores range between 0 and 1. In general, scores less than 0.2 indicate there may be incorrectly modeled domains and scores greater than 0.4 generally indicate more complete and confident models, which are highly similar to the native structure. Depending on the *P* value, each model is also assigned a score confidence level.

CERT, HIGH, MEDIUM, LOW, and POOR are the confidence levels from best to worst[49]; http://qas.wangwb.com/~wwr34/mufoldqa/index.html, https://www.reading.ac.uk/bioinf/ModFOLD/.

Missense3D tool. It predicts the structurally damaging change in the mutant structure[45]; http://missense3d.bc.ic.ac.uk/~missense3d/

ProtParam server. A program that calculates various physical and chemical parameters for a protein sequence. Manual variants were applied to the reference protein sequence separately and resubmitted to calculate the properties changed by variant to detect the impact of the nsSNVs. The calculated parameters include the molecular weight, theoretical pI, atomic composition, extinction coefficient, instability index, aliphatic index, and grand average of hydropathicity (GRAVY)[51].

Untranslated Regions Variants Analysis (SNPs/SNVs and INDELs)

The PolymiRTS database and the SNP Function Prediction tool were used to predict functional variants based on genetic changes (SNPs/SNVs and INDELS) within 3′/5′ UTRs of the *HLA-DRB1* gene (Table 10). PolymiRTS (Polymorphism in microRNAs and their Target-Sites) is a database of naturally occurring DNA variations in the seed regions and target sites of miRNAs. SNPs and INDELs in miRNAs and their target sites may have an impact on miRNA-mRNA interaction, and thus miRNA-mediated gene repression[52]. SNP Function Prediction (FuncPred) was used to predict the effect of SNVs/indels at transcription factor-binding sites (TFBSs; Table 11). Functional variants in the previous region may affect gene expression level, location, or timing[53]; https://compbio.uthsc.edu/miRSNP/, https://snpinfo.niehs.nih.gov/snpinfo/snpfunc.html.

Gene–Gene and Protein–Protein Interactions

GeneMANIA server employed a vast number of functional association data to build a biological network interaction of the top 20 genes associated with our *HLA-DRB1* target gene. GeneMANIA uses a guilt-by-association approach to identify the most related genes to a query gene set. Protein and genetic interactions, pathways, co-expression, co-localization, and protein domain similarity are all examples of association data[54]. Inbio Discover was utilized to establish high confidence protein–protein interactions (PPIs) network. The inBio-Map, a comprehensive map of human protein biology with over 6 million traceable entries, is used by InBio Discover. The predicted trusted interaction networks are based on experimental evidence, pathways, and other curated resources[55]; https://genemania.org/, https://inbio-discover.com/.

Table 1. Distributions of SNVs/MNVs and INDELs.

Molecular consequence	No. of SNVs/MNVs	No. of Indels	Total	Has publications: Yes/No	In ClinVar: Yes/No
Coding regions					
Missense (non-synonymous)	375/5	—	380	15/365	1/379
Nonsense (Stop gain)	31/0	2	33	Nil	Nil
Frame-shift	—	36	36	Nil	Nil
Non-coding (untranslated) regions					
3′UTR	191/0	28	219	1/218	Nil
5′UTR	77/0	12	89	Nil	Nil

SNV: single-nucleotide variant; MNV: multi-nucleotide variant; INDEL: insertion–deletion.

Results

Within the data retrieval date, the *HLA-DRB1* gene contained a total of 9,648 variants, including 7,159 SNVs and 1,078 indels. Except for one, none of the variants have been registered to be significantly associated with human disease, according to the ClinVar database (https://www.ncbi.nlm.nih.gov/clinvar/). In addition, only 26 variants have related publications. From the total variation data, various variants within coding and untranslated regions were chosen for the current study. Information on selected variants is shown in Table 1.

Seven different tools (SIFT, PolyPhen, PredictSNP, PANTHER, SNP&GO, PROVEAN, and SNAP2) with different prediction algorithms were used to identify nsSNVs with significant deleterious effects that could affect the biological structure and function of HLA-DRB1 protein. Out of 375, 91 nsSNVs were predicted by all previous tools to be functional (deleterious or damaging). The I-mutant server predicted changes in stability for all 91 functional nsSNVs identified. Following all previous analyses, the 91 nsSNVs were classified as "high-risk" (Table 2). Most of the high-risk variants are located in exon 2.

The Consurf server and the InterPro database were used to predict the effects of evolutionarily conserved variants on protein functions. The conservation analysis of the HLA-DRB1 protein predicted that 154 positions (≥6 scores) out of 266 amino acids were conserved, as seen in Fig. 2. Table 3 includes the locations and domain names of the InterPro resource that were found. Among the high-risk variants, 25 nsSNVs were identified as conserved and located in domain regions, and they may disrupt or abolish domain function (Table 4).

The effects of the 25 nsSNVs on protein structure were predicted using two tools. The first is the HOPE server, which predicts structural effects based on protein sequence, and the second is Missense3D, which uses a protein model to predict effects. HOPE outcomes show the change in amino acid physiochemical properties, effects on their location, and may disturb the core structure of the located domain (Table 5 and Fig. 3). SWISS-MODEL and PHYRE tools predicted two HLA-DRB1 protein models. Following evaluation by PSICA and ModFOLD, the PHYRE model was selected (Figs. 4 and 5). Using the Missense3D tool, 13 nsSNVs were predicted to cause structural damage to the protein model. The discovered structural damage is displayed in Table 6.

Other types of variants (MNVs and indels) were analyzed for further analysis within coding regions to determine whether they might have a harmful effect on protein. All MNVs showed no significance damage appears. In contrast, 31 out of 36 indels were predicted as harmful by SIFT (Table 7). In addition, within the coding sequence (CDS), 23 stop-gain variants (SNVs/INDELs) were predicted as high impact (Table 8). Last, all nsSNVs demonstrated changes in overall protein physicochemical parameters. The properties changed by all 25 conserved and domain-located high-impact nsSNVs were molecular weight, atomic composition, and GRAVY (Table 9).

The purpose of analyzing variants in untranslated regions is to predict the effects of variants in miRNAs and TFBSs. Functional variants (SNVs and indels) within previous regions could affect gene expression. The results of PolymiRTS Database show that 16 indels and 55 SNPs in the 3′UTR have functional effects on various miRNA binding sites (Table 10). Furthermore, no indels and 10 functionally verified SNPs (of 5′UTR variants) were predicted to affect the activity of TFBSs. The findings are summarized in Table 11. GeneMANIA was used to construct the gene–gene interaction network of the *HLA-DRB1* target gene and the closest 20 genes (Fig. 6). Thus, to gain a better understanding, a network of PPIs was constructed using the inBio-Map resource (Fig. 7). The PPIs network that was built predicted 25 interacted proteins and 44 interactions.

Discussion

HLA-DRB1 gene and its product protein are important in several inflammatory diseases, autoimmune diseases, genetic diversity, and tissue or organ transplantation donor–recipient matches[3]. The protein generated by the *HLA-DRB1* gene, known as the beta chain, connects (binds) to another protein

Table 2. High-Risk nsSNPs Identified by Seven *In Silico* Programs and Their Impact on Protein Stability Effect.

Serial no.	SNV ID	Location	Exon no.	A.A change	SIFT	PolyPhen	PredictSNP	PANTHER	SNP&GO	PROVEAN	SNAP2	I-mutant
1	rs769996810	6:32580775	4	G245E	Deleterious	Probably damaging	Deleterious	Probably damaging	Disease	Deleterious	Effect	Increase
2	rs1193189847	6:32580776	4	G245R	Deleterious	Probably damaging	Deleterious	Probably damaging	Disease	Deleterious	Effect	Decrease
3	rs762260834	6:32580784	4	L242P	Deleterious	Probably damaging	Deleterious	Possibly damaging	Disease	Deleterious	Effect	Decrease
4	rs527579312	6:32580787	4	F241S	Deleterious	Probably damaging	Deleterious	Possibly damaging	Disease	Deleterious	Effect	Decrease
5	rs1359742535	6:32580793	4	L239P	Deleterious	Probably damaging	Deleterious	Possibly damaging	Disease	Deleterious	Effect	Decrease
6	rs778903456	6:32580809	4	G234S	Deleterious	Probably damaging	Deleterious	Probably damaging	Disease	Deleterious	Effect	Increase
7	rs531360990	6:32580812	4	G233R	Deleterious	Probably damaging	Deleterious	Possibly damaging	Disease	Deleterious	Effect	Increase
8	rs1489347193	6:32580817	4	G231E	Deleterious	Probably damaging	Deleterious	Probably damaging	Disease	Deleterious	Effect	Increase
9	rs1449348466	6:32580823	4	L229P	Deleterious	Probably damaging	Deleterious	Possibly damaging	Disease	Deleterious	Effect	Decrease
10	rs377738927	6:32580838	4	A224E	Deleterious	Probably damaging	Deleterious	Possibly damaging	Disease	Deleterious	Effect	Decrease
11	rs760231231	6:32580841	4	S223F	Deleterious	Probably damaging	Deleterious	Possibly damaging	Disease	Deleterious	Effect	Increase
12	rs1262495531	6:32581559	3	W217L	Deleterious	Possibly damaging	Deleterious	Probably damaging	Disease	Deleterious	Effect	Decrease
13	rs1261426119	6:32581593	3	H206Y	Deleterious	Probably damaging	Deleterious	Probably damaging	Disease	Deleterious	Effect	Increase
14	rs1204850358	6:32581605	3	C202R	Deleterious	Probably damaging	Deleterious	Probably damaging	Disease	Deleterious	Effect	Decrease
15	rs472398065	6:32581646	3	V188G	Deleterious	Probably damaging	Deleterious	Possibly damaging	Disease	Deleterious	Effect	Decrease
16	rs1219391595	6:32581656	3	Q185K	Deleterious	Possibly damaging	Deleterious	Possibly damaging	Disease	Deleterious	Effect	Increase
17	rs1265251973	6:32581661	3	T183N	Deleterious	Possibly damaging	Deleterious	Probably damaging	Disease	Deleterious	Effect	Increase
18	rs759467362	6:32581665	3	W182G	Deleterious	Possibly damaging	Deleterious	Probably damaging	Disease	Deleterious	Effect	Decrease
19	rs1236785022	6:32581668	3	D181H	Deleterious	Probably damaging	Deleterious	Probably damaging	Disease	Deleterious	Effect	Decrease
20	rs1457558927	6:32581670	3	G180E	Deleterious	Probably damaging	Deleterious	Probably damaging	Disease	Deleterious	Effect	Increase
21	rs1162153385	6:32581679	3	I177T	Deleterious	Possibly damaging	Deleterious	Possibly damaging	Disease	Deleterious	Effect	Decrease
22	rs1561803226	6:32581701	3	G170W	Deleterious	Probably damaging	Deleterious	Possibly damaging	Disease	Deleterious	Effect	Decrease
23	rs1335525050	6:32581713	3	E166K	Deleterious	Probably damaging	Deleterious	Probably damaging	Disease	Deleterious	Effect	Decrease
24	rs879790499	6:32581718	3	G164V	Deleterious	Possibly damaging	Deleterious	Possibly damaging	Disease	Deleterious	Effect	Decrease
25	rs757966595	6:32581719	3	G164R	Deleterious	Probably damaging	Deleterious	Probably damaging	Disease	Deleterious	Effect	Decrease
26	rs2308767	6:32581720	3	N163K	Deleterious	Probably damaging	Deleterious	Probably damaging	Disease	Deleterious	Effect	Decrease
27	rs1256226377	6:32581742	3	I156T	Deleterious	Probably damaging	Deleterious	Probably damaging	Disease	Deleterious	Effect	Decrease
28	rs16822698	6:32581748	3	G154V	Deleterious	Probably damaging	Deleterious	Probably damaging	Disease	Deleterious	Effect	Decrease
29	rs748235111	6:32581752	3	P153S	Deleterious	Probably damaging	Deleterious	Possibly damaging	Disease	Deleterious	Effect	Decrease
30	rs112796209	6:32581754	3	Y152C	Deleterious	Probably damaging	Deleterious	Possibly damaging	Disease	Deleterious	Effect	Decrease
31	rs2308765	6:32581757	3	F151C	Deleterious	Probably damaging	Deleterious	Probably damaging	Disease	Deleterious	Effect	Decrease
32	rs2308765	6:32581757	3	F151S	Deleterious	Probably damaging	Deleterious	Probably damaging	Disease	Deleterious	Effect	Decrease
33	rs1416623764	6:32581769	3	S147F	Deleterious	Probably damaging	Deleterious	Possibly damaging	Disease	Deleterious	Effect	Decrease
34	rs1416623764	6:32581769	3	S147C	Deleterious	Probably damaging	Deleterious	Possibly damaging	Disease	Deleterious	Effect	Decrease
35	rs707941	6:32581771	3	C146W	Deleterious	Probably damaging	Deleterious	Probably damaging	Disease	Deleterious	Effect	Decrease
36	rs1254922824	6:32581772	3	C146S	Deleterious	Probably damaging	Deleterious	Probably damaging	Disease	Deleterious	Effect	Decrease

(continued)

Table 2. (continued)

Serial no.	SNV ID	Location	Exon no.	A.A change	SIFT	PolyPhen	PredictSNP	PANTHER	SNP&GO	PROVEAN	SNAP2	I-mutant
37	rs41557115	6:32581782	3	L143F	Deleterious	Probably damaging	Deleterious	Possibly damaging	Disease	Deleterious	Effect	Decrease
38	rs200516145	6:32581805	3	T135N	Deleterious	Probably damaging	Deleterious	Possibly damaging	Disease	Deleterious	Effect	Decrease
39	rs17433947	6:32581806	3	T135P	Deleterious	Probably damaging	Deleterious	Possibly damaging	Disease	Deleterious	Effect	Decrease
40	rs80190494	6:32581826	3	V128G	Deleterious	Probably damaging	Deleterious	Possibly damaging	Disease	Deleterious	Effect	Decrease
41	rs79706935	6:32581832	3	P126L	Deleterious	Probably damaging	Deleterious	Probably damaging	Disease	Deleterious	Effect	Decrease
42	rs79706935	6:32581832	3	P126R	Deleterious	Probably damaging	Deleterious	Probably damaging	Disease	Deleterious	Effect	Decrease
43	rs17879125	6:32584115	2	R122W	Deleterious	Probably damaging	Deleterious	Possibly damaging	Disease	Deleterious	Effect	Decrease
44	rs751957355	6:32584144	2	Y112F	Deleterious	Possibly damaging	Deleterious	Probably damaging	Disease	Deleterious	Effect	Decrease
45	rs751957355	6:32584144	2	Y112C	Deleterious	Probably damaging	Deleterious	Probably damaging	Disease	Deleterious	Effect	Decrease
46	rs750986830	6:32584152	2	R109S	Deleterious	Possibly damaging	Deleterious	Possibly damaging	Disease	Deleterious	Effect	Decrease
47	rs779577456	6:32584156	2	C108F	Deleterious	Probably damaging	Deleterious	Probably damaging	Disease	Deleterious	Effect	Decrease
48	rs748753529	6:32584157	2	C108G	Deleterious	Probably damaging	Deleterious	Probably damaging	Disease	Deleterious	Effect	Decrease
49	rs767943289	6:32584165	2	D105G	Deleterious	Probably damaging	Deleterious	Probably damaging	Disease	Deleterious	Effect	Decrease
50	rs61759934	6:32584166	2	D105Y	Deleterious	Probably damaging	Deleterious	Possibly damaging	Disease	Deleterious	Effect	Decrease
51	rs61759934	6:32584166	2	D105N	Deleterious	Probably damaging	Deleterious	Possibly damaging	Disease	Deleterious	Effect	Decrease
52	rs17878857	6:32584174	2	A102V	Deleterious	Possibly damaging	Deleterious	Probably damaging	Disease	Deleterious	Effect	Decrease
53	rs17885869	6:32584177	2	R101L	Deleterious	Probably damaging	Deleterious	Probably damaging	Disease	Deleterious	Effect	Decrease
54	rs17885869	6:32584177	2	R101P	Deleterious	Probably damaging	Deleterious	Probably damaging	Disease	Deleterious	Effect	Decrease
55	rs17885869	6:32584177	2	R101Q	Deleterious	Probably damaging	Deleterious	Probably damaging	Disease	Deleterious	Effect	Decrease
56	rs17885222	6:32584178	2	R101W	Deleterious	Probably damaging	Deleterious	Probably damaging	Disease	Deleterious	Effect	Decrease
57	rs17885222	6:32584178	2	R101G	Deleterious	Probably damaging	Deleterious	Probably damaging	Disease	Deleterious	Effect	Decrease
58	rs41308499	6:32584189	2	L97R	Deleterious	Probably damaging	Deleterious	Probably damaging	Disease	Deleterious	Effect	Decrease
59	rs17879230	6:32584216	2	E88G	Deleterious	Possibly damaging	Deleterious	Possibly damaging	Disease	Deleterious	Effect	Decrease
60	rs1059584	6:32584219	2	A87G	Deleterious	Probably damaging	Deleterious	Probably damaging	Disease	Deleterious	Effect	Decrease
61	rs1059584	6:32584219	2	A87D	Deleterious	Probably damaging	Deleterious	Probably damaging	Disease	Deleterious	Effect	Decrease
62	rs41308498	6:32584228	2	R84L	Deleterious	Probably damaging	Deleterious	Possibly damaging	Disease	Deleterious	Effect	Decrease
63	rs72558166	6:32584235	2	L82V	Deleterious	Probably damaging	Deleterious	Possibly damaging	Disease	Deleterious	Effect	Decrease
64	rs780784592	6:32584237	2	E81V	Deleterious	Probably damaging	Deleterious	Possibly damaging	Disease	Deleterious	Effect	Decrease
65	rs17883065	6:32584238	2	E81K	Deleterious	Probably damaging	Deleterious	Possibly damaging	Disease	Deleterious	Effect	Decrease
66	rs1059582	6:32584240	2	T80M	Deleterious	Probably damaging	Deleterious	Probably damaging	Disease	Deleterious	Effect	Decrease
67	rs1059582	6:32584240	2	T80R	Deleterious	Probably damaging	Deleterious	Probably damaging	Disease	Deleterious	Effect	Increase
68	rs17879432	6:32584246	2	A78V	Deleterious	Probably damaging	Deleterious	Possibly damaging	Disease	Deleterious	Effect	Decrease
69	rs17879432	6:32584246	2	A78E	Deleterious	Probably damaging	Deleterious	Possibly damaging	Disease	Deleterious	Effect	Decrease
70	rs17885437	6:32584259	2	G74R	Deleterious	Probably damaging	Deleterious	Probably damaging	Disease	Deleterious	Effect	Decrease
71	rs17882455	6:32584273	2	F69C	Deleterious	Probably damaging	Deleterious	Probably damaging	Disease	Deleterious	Effect	Decrease
72	rs754953589	6:32584277	2	R68C	Deleterious	Probably damaging	Deleterious	Possibly damaging	Disease	Deleterious	Effect	Decrease

(continued)

Table 2. (continued)

Serial no.	SNV ID	Location	Exon no.	A.A change	SIFT	PolyPhen	PredictSNP	PANTHER	SNP&GO	PROVEAN	SNAP2	I-mutant
73	rs754953589	6:32584277	2	R68G	Deleterious	Probably damaging	Deleterious	Possibly damaging	Disease	Deleterious	Effect	Decrease
74	rs754953589	6:32584277	2	R68S	Deleterious	Probably damaging	Deleterious	Possibly damaging	Disease	Deleterious	Effect	Decrease
75	rs1303260918	6:32584285	2	E65G	Deleterious	Probably damaging	Deleterious	Possibly damaging	Disease	Deleterious	Effect	Decrease
76	rs17879242	6:32584293	2	N62K	Deleterious	Probably damaging	Deleterious	Possibly damaging	Disease	Deleterious	Effect	Decrease
77	rs1289742638	6:32584294	2	N62S	Deleterious	Probably damaging	Deleterious	Possibly damaging	Disease	Deleterious	Effect	Decrease
78	rs17878437	6:32584305	2	R58S	Deleterious	Probably damaging	Deleterious	Possibly damaging	Disease	Deleterious	Effect	Decrease
79	rs747606824	6:32584306	2	R58I	Deleterious	Probably damaging	Deleterious	Possibly damaging	Disease	Deleterious	Effect	Decrease
80	rs771765212	6:32584307	2	R58G	Deleterious	Probably damaging	Deleterious	Possibly damaging	Disease	Deleterious	Effect	Decrease
81	rs1335931724	6:32584321	2	V53G	Deleterious	Probably damaging	Deleterious	Possibly damaging	Disease	Deleterious	Effect	Decrease
82	rs1335931724	6:32584321	2	V53E	Deleterious	Probably damaging	Deleterious	Possibly damaging	Disease	Deleterious	Effect	Decrease
83	rs17879469	6:32584333	2	G49A	Deleterious	Probably damaging	Deleterious	Probably damaging	Disease	Deleterious	Effect	Decrease
84	rs61759931	6:32584334	2	G49R	Deleterious	Probably damaging	Deleterious	Probably damaging	Disease	Deleterious	Effect	Decrease
85	rs1561818227	6:32584349	2	C44G	Deleterious	Probably damaging	Deleterious	Probably damaging	Disease	Deleterious	Effect	Decrease
86	rs1561818227	6:32584349	2	C44R	Deleterious	Probably damaging	Deleterious	Probably damaging	Disease	Deleterious	Effect	Decrease
87	rs766505678	6:32584358	2	K41E	Deleterious	Possibly damaging	Deleterious	Possibly damaging	Disease	Deleterious	Effect	Decrease
88	rs9269957	6:32584364	2	Q39K	Deleterious	Probably damaging	Deleterious	Possibly damaging	Disease	Deleterious	Effect	Decrease
89	rs372634318	6:32589652	1	D31Y	Deleterious	Probably damaging	Deleterious	Possibly damaging	Disease	Deleterious	Effect	Increase
90	rs750329958	6:32589678	1	L22Q	Deleterious	Probably damaging	Deleterious	Probably damaging	Disease	Deleterious	Effect	Decrease
91	rs1468275841	6:32589687	1	L19Q	Deleterious	Possibly damaging	Deleterious	Possibly damaging	Disease	Deleterious	Effect	Decrease

SNP: single-nucleotide polymorphism; SNV: single-nucleotide variant; SIFT: Sorting Intolerant From Tolerant; PANTHER: Protein Analysis Through Evolutionary Relationships; GO: Gene Ontology; PROVEAN: Protein Variation Effect Analyzer.

Table 3. Predicting Domains Using InterPro.

Tool	Domains' name and accession numbers	Position
InterPro	MHC_II_b_N (MHC class II, beta chain, N-terminal) IPR000353	42-116
	Ig_C1-set (Immunoglobulin C1-set) IPR003597	128-212
	Ig-like_dom (Immunoglobulin-like domain) IPR007110	126-214
SMART	MHC_II_beta (Class II histocompatibility antigen, beta domain) SM00921	42-116
	IGc1 (Immunoglobulin C-Type) SM00407	141-212
Pfam	MHC_II_beta(Class II histocompatibility antigen, beta domain) PF00969	43-115
	IGc1 (Immunoglobulin C-Type) SM00407	141-212
PROSITE	IG_LIKE (Ig-like domain profile) PS50835	126-214

Table 4. Non-Synonymous SNPs That Are Highly Conserved and Located in Domains' Sites.

Serial no.	Variation ID	Chromosome location	Exon no.	Codons	A.A change	CLIN_SIG
1	rs1261426119	6:32581593	3	Cac/Tac	H206Y	—
2	rs1204850358	6:32581605	3	Tgc/Cgc	C202R	—
3	rs1472398065	6:32581646	3	gTg/gGg	V188G	—
4	rs1219391595	6:32581656	3	Cag/Aag	Q185K	—
5	rs1265251973	6:32581661	3	aCc/aAc	T183N	—
6	rs759467362	6:32581665	3	Tgg/Ggg	W182G	—
7	rs1236785022	6:32581668	3	Gac/Cac	D181H	—
8	rs1457558927	6:32581670	3	gGa/gAa	G180E	—
9	rs1335525050	6:32581713	3	Gaa/Aaa	E166K	—
10	rs2308767	6:32581720	3	aaC/aaA	N163K	—
11	rs748235111	6:32581752	3	Cca/Tca	P153S	—
12	rs2308765	6:32581757	3	tTc/tGc	F151C	—
13	rs2308765	6:32581757	3	tTc/tCc	F151S	—
14	rs707941	6:32581771	3	tgC/tgG	C146W	—
15	rs1254922824	6:32581772	3	tGc/tCc	C146S	—
16	rs79706935	6:32581832	3	cCt/cTt	P126L	—
17	rs79706935	6:32581832	3	cCt/cGt	P126R	—
18	rs779577456	6:32584156	2	tGc/tTc	C108F	—
19	rs748753529	6:32584157	2	Tgc/Ggc	C108G	—
20	rs17879242	6:32584293	2	aaC/aaA	N62K	—
21	rs1289742638	6:32584294	2	aAc/aGc	N62S	—
22	rs17879469	6:32584333	2	gGg/gCg	G49A	—
23	rs61759931	6:32584334	2	Ggg/Cgg	G49R	—
24	rs1561818227	6:32584349	2	Tgt/Ggt	C44G	—
25	rs1561818227	6:32584349	2	Tgt/Cgt	C44R	—

The symbol "—" refers to unavailable data.
CLIN/SIG: clinical significance refers to the ClinVar database, which compiles data on genomic variation and its impact on human health; SNP: single-nucleotide polymorphism.

produced by the *HLA-DRA* gene, known as the alpha chain. They combine to produce the HLA-DR antigen-binding heterodimer, a functional protein complex. This complex presents foreign peptides to the immune system to activate the body's immunological response[6]. Variations in the structural conformation of the HLA-DRB1 protein during bio-molecular interactions are critical for its function. Therefore, determining the effects of harmful HLA-DRB1 variants and their association with various diseases is critical. The purpose of this study was to use computational analysis to identify the most harmful variants (SNVs, MNVs, and INDELS) and their effects on the HLA-DRB1 structure, function, and expression.

In terms of substitution single-variants, several tools predicted that 91 missense (nsSNV) and 22 stop-gain variants within coding regions were functional. The 22 stop-gain

Table 5. HOPE-Based Protein Sequence Predictions (Structural and Function Change).

Serial no.	Variation ID	Amino acids change	Amino acids properties	Location/contacts	Effect of variants on the protein
1.	rs1261426119	H206Y	The M is bigger and more hydrophobic than the W	The W forms a hydrogen bond with Proline at position 153 and Serine at position 208. M is not in the correct position to make the same hydrogen bond as the original W did	M might disturb the core structure of the located domain and abolish its function
2.	rs1204850358	C202R	The M is bigger and more hydrophobic than the W. The W charge is neutral, while the M charge is positive	The W is involved in a cysteine bridge, which is important for stability of the protein. Only Cysteines can make these type of bonds, the mutation causes loss of this interaction and will have a severe effect on the 3D-structure of the protein	M might disturb the core structure of the located domain and abolish its function
3.	rs1472398065	V188G	The M is smaller and less hydrophobic than W	—	M might disturb the core structure of the located domain and abolish its function
4.	rs1219391595	Q185K	The M is bigger and more hydrophobic than the W. The W charge is neutral, while the M charge is positive	W is involved in a multimer contact. The mutation introduces a larger residue at this position, which can disrupt multimeric interactions	M might disturb the core structure of the located domain and abolish its function
5.	rs1265251973	T183N	The M is bigger and less hydrophobic than W	The W forms a hydrogen bond with Asparagine at position 179 and Aspartic Acid at position 181. M is not in the correct position to make the same hydrogen bond as the original W did	M might disturb the core structure of the located domain and abolish its function
6.	rs759467362	W182G	The M is smaller and less hydrophobic than W	M may be too small to form multimer contacts and may also influence hydrogen bond formation	M might disturb the core structure of the located domain and abolish its function
7.	rs1236785022	D181H	The M is Bigger than W. The M charge is NEUTRAL, while the W charge is negative.	The W forms a hydrogen bond with Threonine at position 183. M is not in the correct position to make the same hydrogen bond as the original W did	M might disturb the core structure of the located domain and abolish its function
8.	rs1457558927	G180E	The M is Bigger than W. The W charge is neutral, while the M charge is negative	—	M might disturb the core structure of the located domain and abolish its function
9.	rs1335525050	E166K	The M is Bigger than W. The W charge is negative, while the M charge is positive	The W forms a salt bridge with Arginine at position 159 and Lysine at position 168. The difference in charge will disturb the ionic interaction made by the original W.	M might disturb the core structure of the located domain and abolish its function
10.	rs2308767	N163K	The M is Bigger than W. The W charge is neutral, while the M charge is positive	The W forms a hydrogen bond with Valine at position 199. M is not in the correct position to make the same hydrogen bond as the original W did	M might disturb the core structure of the located domain and abolish its function
11.	rs748235111	P153S	The M is smaller and less hydrophobic than W	—	M might disturb the core structure of the located domain and abolish its function
12.	rs2308765	F151C	The M is smaller than W	—	M might disturb the core structure of the located domain and abolish its function
13.	rs2308765	F151S	The M is smaller and less hydrophobic than the W	—	M might disturb the core structure of the located domain and abolish its function
14.	rs707941	C146W	The M is bigger than the W	The W is involved in a cysteine bridge, which is important for stability of the protein. Only Cysteines can make these type of bonds, the mutation causes loss of this interaction and will have a severe effect on the 3D-structure of the protein	M might disturb the core structure of the located domain and abolish its function
15	rs1254922824	C146S	The W is more hydrophobic than the M	The W is involved in a cysteine bridge, which is important for stability of the protein. Only Cysteines can make these type of bonds, the mutation causes loss of this interaction and will have a severe effect on the 3D-structure of the protein	M might disturb the core structure of the located domain and abolish its function

(continued)

Table 5. (continued)

Serial no.	Variation ID	Amino acids change	Amino acids properties	Location/contacts	Effect of variants on the protein
16	rs79706935	P126L	The M is bigger than the W	—	M might disturb the core structure of the located domain and abolish its function
17	rs79706935	P126R	The M is bigger and less hydrophobic than the W. The W charge is neutral, while the M charge is positive	—	M might disturb the core structure of the located domain and abolish its function
18	rs779577456	C108F	The M is bigger than the W	The W is involved in a cysteine bridge, which is important for stability of the protein. Only Cysteines can make these type of bonds, the mutation causes loss of this interaction and will have a severe effect on the 3D-structure of the protein	M might disturb the core structure of the located domain and abolish its function
19	rs748753529	C108G	The M is smaller and less hydrophobic than the W	The W is involved in a cysteine bridge, which is important for stability of the protein. Glycines are very flexible and can disturb the required rigidity of the protein at this position	M might disturb the core structure of the located domain and abolish its function
20	rs17879242	N62K	M is larger and has a positive charge, whereas W is neutral	The W forms a hydrogen bond with Leucine at position 37. M is not in the correct position to make the same hydrogen bond as the original W did	M might disturb the core structure of the located domain and abolish its function
21	rs1289742638	N62S	The M is smaller and more hydrophobic than the W	The W forms a hydrogen bond with Leucine at position 37. The difference in size and hydrophobicity could affect hydrogen bond formation	M might disturb the core structure of the located domain and abolish its function
22	rs17879469	G49A	The M is bigger and more hydrophobic than the W	—	M might disturb the core structure of the located domain and abolish its function
23	rs61759931	G49R	M is larger and has a positive charge, whereas W is neutral	—	M might disturb the core structure of the located domain and abolish its function
24	rs1561818227	C44G	The M is smaller and less hydrophobic than the W	The W is involved in a cysteine bridge, which is important for stability of the protein. The differences between the old and new residue can cause destabilization of the structure	M might disturb the core structure of the located domain and abolish its function
25	rs1561818227	C44R	The M is bigger and less hydrophobic than the W. The W charge was neutral, while the M charge is positive	The W is involved in a cysteine bridge, which is important for stability of the protein. The differences between the old and new residue can cause destabilization of the structure	M might disturb the core structure of the located domain and abolish its function

The symbol "——" refers to unavailable data.
W: wild type residue; M: mutant type residue.

Table 6. Structural Modifications Brought About by an Amino Acid Substitution Using Missense3D Tool.

Serial no.	Variation ID	Amino acids change	Structural changes predicted
1	rs1261426119	H206Y	Buried charge replaced Buried H-bond breakage
2	rs1204850358	C202R	Disulphide breakage Buried charge introduced Buried hydrophilic introduced Clash
3	rs1457558927	G180E	Disallowed phi/psi angle Gly in a bend
4	rs748235111	P153S	Cis pro replaced
5	rs707941	C146W	Disulphide breakage Clash
6	rs1254922824	C146S	Disulphide breakage
7	rs79706935	P126L	Clash
8	rs79706935	P126R	Clash Buried charge introduced
9	rs779577456	C108F	Disulphide breakage Clash
10	rs748753529	C108G	Disulphide breakage
11	rs17879242	N62K	Buried charge introduced
12	rs1561818227	C44G	Disulphide breakage
13	rs1561818227	C44R	Disulphide breakage

Table 7. SIFT Server Functional Prediction of All Indels in Coding Regions.

Serial no.	Variation ID	Amino acid position change	Effect	Confidence score (%)	Causes nonsense mediated decay (NMD)
1	rs1775322739	248-265	Damaging	0.858	No
2	rs1178714115	234-265	Damaging	0.858	No
3	rs140357311	197-266	Damaging	0.858	Yes
4	rs1775509563	195-266	Damaging	0.858	Yes
5	rs1554124346	195-266	Damaging	0.858	Yes
6	rs1775521710	167-266	Damaging	0.858	Yes
7	rs1328066782	174-266	Damaging	0.858	Yes
8	rs35616319	134-266	Damaging	0.858	Yes
9	rs869063545	102-266	Damaging	0.858	Yes
10	rs1554126585	102-266	Damaging	0.858	Yes
11	rs67187877	101-266	Damaging	0.858	Yes
12	rs9281873	100-266	Damaging	0.858	Yes
13	rs752707222	101-266	Damaging	0.858	Yes
14	rs778205073	100-266	Damaging	0.858	Yes
15	rs1561816391	98-266	Damaging	0.858	Yes
16	rs770836206	98-266	Damaging	0.858	Yes
17	rs764153503	98-266	Damaging	0.858	Yes
18	rs17878577	93-266	Damaging	0.858	Yes
19	rs1480365395	66-266	Damaging	0.858	Yes
20	rs796101477	65-266	Damaging	0.858	Yes
21	rs879122917	66-266	Damaging	0.858	Yes
22	rs1554126912	65-266	Damaging	0.858	Yes
23	rs1260282149	51-266	Damaging	0.858	Yes
24	rs1776050880	53-266	Damaging	0.858	Yes
25	rs1776051840	51-266	Damaging	0.858	Yes
26	rs770838956	41-266	Damaging	0.858	Yes
27	rs1554127069	41-266	Damaging	0.858	Yes
28	rs1776067234	38-266	Damaging	0.858	Yes
29	rs772011591	37-266	Damaging	0.858	Yes
30	rs767010367	36-266	Damaging	0.858	Yes
31	rs1581830100	34-266	Damaging	0.858	Yes

SIFT: Sorting Intolerant From Tolerant.

Table 8. High-Impact Stop-Gain SNVs/INDELs Identified by Variant Effect Predictor.

Variant ID	Location	Variant type	Impact	Amino acids	Codons	Strand
rs1218850675	6:32580762	SNP	High	Y/*	taC/taA	−1
rs1207397234	6:32580818	SNP	High	G/*	Gga/Tga	−1
rs1561802650	6:32581602	SNP	High	Q/*	Caa/Taa	−1
rs2308777	6:32581609	SNP	High	Y/*	taC/taG	−1
rs2308777	6:32581609	SNP	High	Y/*	taC/taA	−1
rs754428084	6:32581626	SNP	High	R/*	Cga/Tga	−1
rs1420364217	6:32581677	SNP	High	Q/*	Cag/Tag	−1
rs17405219	6:32581830	SNP	High	K/*	Aag/Tag	−1
rs17882084	6:32581836	SNP	High	Q/*	Caa/Taa	−1
rs1165708016	6:32584112	SNP	High	R/*	Cga/Tga	−1
rs756601075	6:32584155	SNP	High	C/*	tgC/tgA	−1
rs11554463	6:32584158	SNP	High	Y/*	taC/taG	−1
rs11554463	6:32584158	SNP	High	Y/*	taC/taA	−1
rs1207528230	6:32584209	SNP	High	W/*	tgG/tgA	−1
rs769883645	6:32584212	SNP	High	Y/*	taC/taA	−1
rs17883065	6:32584238	SNP	High	E/*	Gag/Tag	−1
rs773064485	6:32584286	SNP	High	E/*	Gag/Tag	−1
rs766505678	6:32584358	SNP	High	K/*	Aag/Tag	−1
rs9269957	6:32584364	SNP	High	Q/*	Cag/Tag	−1
rs9269958	6:32584366	SNP	High	W/*	tGg/tAg	−1
rs9256943	6:32589646	SNP	High	R/*	Cga/Tga	−1
rs1309359000	6:32589730	SNP	High	K/*	Aag/Tag	−1
rs776465322	6:32584351	INDEL	High	E/EV*X	gAg/gAAGTATAAg	−1

The impact for the type of consequence can be High, Moderate, Low, or Modifier. High impact indicates that the variant is assumed to have a high (disruptive) impact on the protein, probably causing protein truncation, loss of function, or triggering nonsense-mediated decay.
SNV: single-nucleotide variant; INDEL: insertion–deletion.

Table 9. The Effect of nsSNVs on HLA-DRB1′ Protein Physicochemical Parameters.

Reference and variants	Molecular weight	Theoretical pI	Atomic composition	Total −ve	Total +ve	Extinction coefficients	Instability index	Aliphatic index	GRAVY
Reference	29,966.14	7.64	$C_{1342}H_{2068}N_{368}O_{389}S_{12}$	25	26	41,285	48.92	77.93	−0.207
H206Y	**29,992.17**	**7.62**	$\mathbf{C_{1345}H_{2070}N_{366}O_{390}S_{12}}$	25	26	**42,775**	**49.78**	77.93	**−0.200**
C202R	**30,019.18**	**8.26**	$\mathbf{C_{1345}H_{2075}N_{371}O_{389}S_{11}}$	25	**27**	**41,160**	**49.93**	77.93	**−0.233**
V188G	**29,924.05**	7.64	$\mathbf{C_{1339}H_{2062}N_{368}O_{389}S_{12}}$	25	26	41,285	48.92	**76.84**	**−0.224**
Q185K	**29,966.18**	**8.20**	$\mathbf{C_{1343}H_{2072}N_{368}O_{388}S_{12}}$	25	**27**	41,285	**48.36**	77.93	**−0.208**
T183N	**29,979.13**	7.64	$C_{1342}H_{2067}\mathbf{N_{369}}O_{389}S_{12}$	25	26	41,285	48.92	77.93	**−0.217**
W182G	**29,836.97**	7.64	$\mathbf{C_{1333}H_{2061}N_{367}}O_{389}S_{12}$	25	26	**35,785**	**49.17**	77.93	**−0.205**
D181H	**29,988.19**	**8.21**	$\mathbf{C_{1344}H_{2070}N_{370}O_{387}}S_{12}$	**24**	26	41,285	**48.81**	77.93	**−0.206**
G180E	**30,038.20**	**7.00**	$\mathbf{C_{1345}H_{2072}}N_{368}\mathbf{O_{391}}S_{12}$	**26**	26	41,285	**50.21**	77.93	**−0.218**
E166K	**29,965.19**	**8.51**	$\mathbf{C_{1343}H_{2073}N_{369}O_{387}}S_{12}$	**24**	**27**	41,285	**46.97**	77.93	**−0.208**
N163K	**29,980.21**	**8.20**	$\mathbf{C_{1344}H_{2074}}N_{368}\mathbf{O_{388}}S_{12}$	25	**27**	41,285	**48.85**	77.93	**−0.208**
P153S	**29,956.10**	7.64	$\mathbf{C_{1340}H_{2066}}N_{368}\mathbf{O_{390}}S_{12}$	25	26	41,285	**48.46**	77.93	**−0.204**
F151C	**29,922.10**	**7.61**	$\mathbf{C_{1336}H_{2064}}N_{368}O_{389}S_{12}$	25	26	41,285	**47.70**	77.93	**−0.208**
F151S	**29,906.04**	7.64	$\mathbf{C_{1336}H_{2064}}N_{368}\mathbf{O_{390}}S_{12}$	25	26	41,285	**47.70**	77.93	**−0.220**
C146W	**30,049.21**	7.67	$\mathbf{C_{1350}H_{2073}N_{369}}O_{389}S_{11}$	25	26	**46,660**	48.92	77.93	**−0.220**
C146S	**30,049.21**	7.67	$\mathbf{C_{1350}H_{2073}N_{369}}O_{389}S_{12}$	25	26	**46,660**	48.92	77.93	**−0.220**
P126L	**29,982.18**	7.64	$\mathbf{C_{1343}H_{2072}}N_{368}O_{389}S_{12}$	25	26	41,285	**47.88**	**79.40**	**−0.186**
P126R	**30,025.21**	**8.20**	$\mathbf{C_{1343}H_{2073}N_{371}}O_{389}S_{12}$	25	**27**	41,285	**48.20**	77.93	**−0.218**
C108F	**30,010.17**	7.67	$\mathbf{C_{1348}H_{2072}}N_{368}O_{389}S_{11}$	25	26	**41,160**	48.92	77.93	**−0.206**
C108G	**29,920.05**	7.67	$\mathbf{C_{1341}H_{2066}}N_{368}O_{389}S_{11}$	25	26	**41,160**	**48.60**	77.93	**−0.218**
N62K	**29,980.21**	**8.20**	$\mathbf{C_{1344}H_{2074}}N_{368}\mathbf{O_{388}}S_{12}$	25	**27**	41,285	**50.10**	77.93	**−0.208**
N62S	**29,939.11**	7.64	$\mathbf{C_{1341}H_{2067}N_{367}}O_{389}S_{12}$	25	26	41,285	**49.93**	77.93	**−0.197**
G49A	**29,980.16**	7.64	$\mathbf{C_{1343}H_{2070}}N_{368}O_{389}S_{12}$	25	26	41,285	**49.81**	**78.31**	**−0.198**
G49R	**30,065.27**	**8.20**	$\mathbf{C_{1346}H_{2077}N_{371}}O_{389}S_{12}$	25	**27**	41,285	**49.81**	77.93	**−0.222**
C44G	**29,920.05**	7.67	$\mathbf{C_{1341}H_{2066}}N_{368}O_{389}S_{11}$	25	26	**41,160**	**46.04**	77.93	**−0.218**
C44R	**30,019.18**	**8.26**	$\mathbf{C_{1345}H_{2075}N_{371}}O_{389}S_{11}$	25	**27**	**41,160**	**46.77**	77.93	**−0.233**

The accession number for the reference sequence is P01911 (https://www.uniprot.org/). Total −ve: total negatively charged residues. Total +ve: total positively charged residues. The parameters that have been changed compared with the reference are highlighted in bold.
SNV: single-nucleotide variant; HLA: human leukocyte antigen; GRAVY: grand average of hydropathicity index.

Table 10. Functional SNPs/Indels in the 3′UTR.

Serial no.	Variant ID	Variant type	Function class	Serial no.	Variant ID	Variant type	Function class
1	rs34839759	SNP	1:C	37	rs1732	SNP	4:D/ 3:C
2	rs114103896	SNP	1:D/ 2:C	38	rs142078339	SNP	3:D/ 6:C
3	rs35136435	INDEL	3:O	39	rs112871130	SNP	3:D/ 4:C
4	rs34266013	SNP	1:D/ 1:C	40	rs148582499	INDEL	8:O
5	rs35413567	SNP	1:C	41	rs35165835	INDEL	6:O
6	rs35513414	SNP	3:C	42	rs34160410	INDEL	2:O
7	rs200428856	INDEL	1:O	43	rs35463048	SNP	2:C
8	rs3205684	SNP	15:C/ 1:D	44	rs34007709	SNP	5:D/ 1:C
9	rs1064717	SNP	2:C/ 1:D	45	rs36084494	SNP	2:D/ 3:C
10	rs185448040	SNP	1:D/ 1:C	46	rs34844328	SNP	5:D/ 4:C
11	rs6920823	SNP	1:D/ 2:C	47	rs3205692	SNP	5:D/ 3:C
12	rs35418460	SNP	1:D/ 1:C	48	rs1060081	SNP	1:D/ 2:C
13	rs34923246	SNP	1:D	49	rs116358897	SNP	1:D
14	rs35263976	SNP	2:D/ 1:C	50	rs182030800	SNP	1:D
15	rs34542752	SNP	4:C	51	rs3200898	SNP	1:D
16	rs199703384	INDEL	5:O	52	rs71864678	INDEL	7:O
17	rs80136018	SNP	2:D/ 6:C	53	rs9269688	SNP	1:D/ 7:C
18	rs34205910	INDEL	5:O	54	rs3180268	SNP	1:D/ 3:C
19	rs1064713	SNP	2:D/ 1:C	55	rs71810699	INDEL	2:O
20	rs34981130	SNP	2:D/ 1:C	56	rs1064699	SNP	1:D/ 1:C
21	rs1064712	SNP	2:D/ 1:C	57	rs35521457	SNP	2:D/ 2:C
22	rs1730	SNP	1:C	58	rs35236441	SNP	3:D/ 1:C
23	rs113493811	INDEL	8:O	59	rs36217730	SNP	2:D/ 4:C
24	rs1060190	SNP	2:D/ 7:C	60	rs35324556	SNP	2:D/ 4:C
25	rs71685135	INDEL	15:O	61	rs146292738	SNP	3:D/ 2:C
26	rs35306263	INDEL	17:O	62	rs36217728	SNP	1:D
27	rs35195677	SNP	6:D/ 6:C	63	rs1064692	SNP	1:D
28	rs1060185	SNP	4:D/ 4:C	64	rs201375698	SNP	2:D
29	rs3208409	SNP	6:D/ 6:C	65	rs202053852	SNP	3:D/ 6:C
30	rs1064710	SNP	6:D/ 2:C	66	rs1064691	SNP	2:D/ 4:C
31	rs1064709	SNP	7:D/ 1:C	67	rs1059920	SNP	1:D
32	rs3200047	SNP	7:D/ 4:C	68	rs41285181	INDEL	5:O
33	rs35000099	SNP	10:D	69	rs71822874	INDEL	5:O
34	rs113804375	SNP	5:D/ 6:C	70	rs68069105	INDEL	2:O
35	rs35716402	SNP	4:D/ 5:C	71	rs9279724	INDEL	2:O
36	rs201099263	SNP	2:C				

Variant ID: related to SNP database. Function class: the number represents the number of miRNAs that have been affected by variants. The letters stand for the following: D: the derived allele disrupts a conserved miRNA site (ancestral allele with support > 2); C: the derived allele creates a new miRNA site; O: the ancestral allele cannot be determined.
SNP: single-nucleotide polymorphism; INDEL: insertion–deletion; miRNA: microRNA.

variants were classified as high impact, implying that the variant will have a significant (disruptive) effect on the protein, most likely resulting in protein truncation or loss of function. The 91 nsSNVs were classified as high risk after the target protein's stability changed. Thirteen of the high-risk nsSNVs (rs9269957, rs17879469, rs17879242, rs17879432, rs17885437, rs17883065, rs41308498, rs1059584, rs17879230, rs41308499, rs17885869, rs61759934, and rs41557115) correspond to pathological variants predicted by Hassan et al.[56] The variants identified as pathological by Hassan's discovery but not in this study could be due to the increased number of tools used in the

current study. The update to the SNP and tool databases may have caused the vice versa to occur. The Consurf server and the InterPro database were used to predict the effects of evolutionarily conserved variants that are located in domains. InterPro resource integrates signatures from the following 13 member databases: CATH, CDD, HAMAP, MobiDB Lite, Panther, Pfam, PIRSF, PRINTS, Prosite, SFLD, SMART, SUPERFAMILY, and TIGRfams. Among the high-risk variants, 25 nsSNVs were identified as conserved and located in domain regions, and they may disrupt or abolish domain function. The effects of the 25 nsSNVs on protein structure were predicted based on sequence and model using HOPE

Table 11. Functionally Verified SNPs at a Transcription Factor-Binding Site.

Serial no.	Variant ID	Allele	Regulatory potential score	Conservation score
1	rs1059546	G/C/A	0.153264	0.000
2	rs17204737	C/T	0.193148	0.000
3	rs17204744	C/G	0.141892	0.001
4	rs17204758	A/C	0.136889	0.003
5	rs17204765	A/G	0.071855	0.000
6	rs17211071	A/G	0.193703	0.000
7	rs17211078	G/T	0.18866	0.000
8	rs17211105	A/G	0.054697	0.000
9	rs28366223	A/G	0.078063	0.000
10	rs9270314	G/T	0.061302	0.000

For more information on regulatory potential and conservation scores, see https://snpinfo.niehs.nih.gov/snpinfo/guide.html#snpfunc.
SNP: single-nucleotide polymorphism.
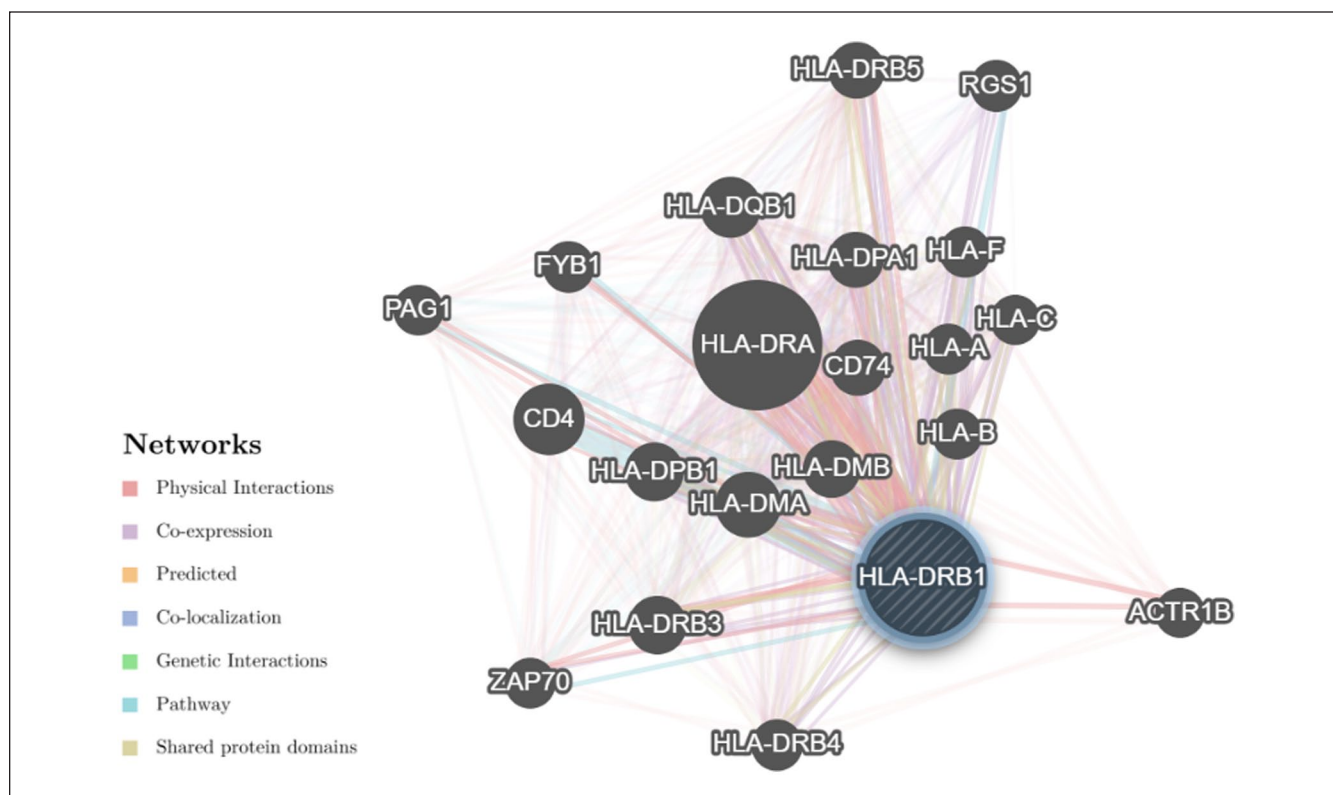
Figure 6. Gene–gene interaction network of the *HLA-DRB1* gene predicted by GeneMANIA.

and Missense3D, respectively. The HOPE results revealed that the amino acid properties of the 25 nsSNVs changed and have the potential to disrupt the domain's core structure. Two algorithms (SWISS-MODEL and PHYRE) were used to predict the HLA-DRB1 models to use the Missense3D tool. Following evaluation by PSICA and ModFOLD, the PHYRE model was selected. Several factors contributed to the selection of the PHYRE model, including its coverage of the entire protein (266 amino acids), higher overall quality

scores, and best confidence value. The Missense3D tool predicted that 13 of the 25 nsSNVs would cause structural damage to the protein model.

Additional types of variants (MNVs and indels) were analyzed for further analysis within coding regions to determine whether they might have a harmful effect on protein function. All MNVs showed nil significance damage appears. In contrast, SIFT predicted 31 indels to be harmful, while the Variant Effect Predictor predicted only one to lead to
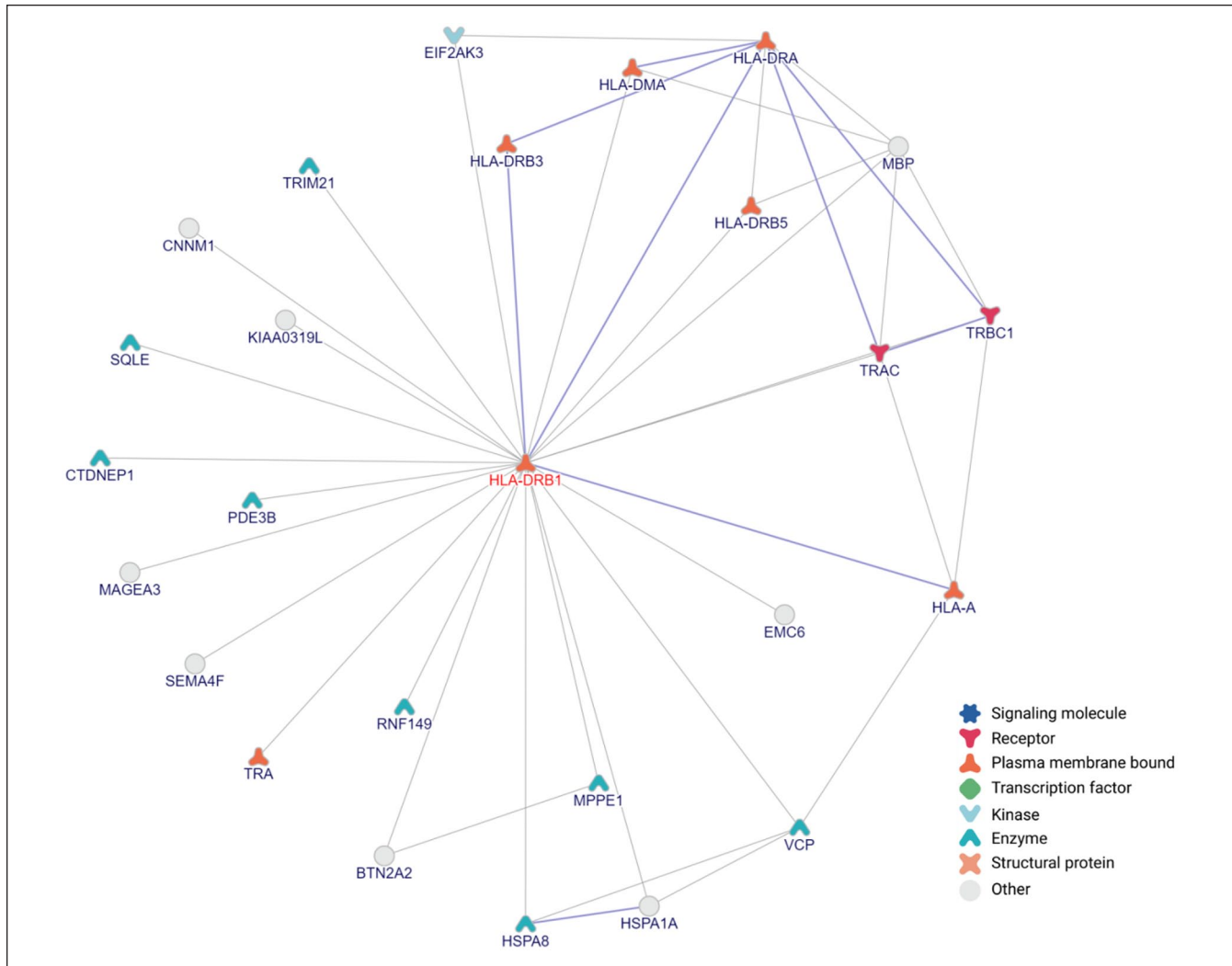
Figure 7. Protein–protein interaction network of the HLA-DRB1 protein predicted by inBio-Discover. Pathway interactions are shown as blue lines, remaining interactions are inBio-Map high-confidence interactions.

premature protein. Functional predicted Indels might affect a few numbers of amino acids and even the complete protein as shown. According to physiochemical properties, the HOPE tool, as previously mentioned, revealed differences in the level of residues between wild and new types, whereas ProtParam indicated that variants caused changes in the entire protein. All 25 conserved and domain-located high-impact nsSNVs agreed to alter the protein's molecular weight, atomic composition, and GRAVY, but there is a divergence in other properties. In general, high-risk nsSNVs affect protein structure, function, and physicochemical properties.

The goal of analyzing variants (SNVs and indels) in 3′/5′ untranslated regions is to predict the effects of variants that may affect the level, location, or timing of gene expression using PolymiRTS and SNP Function Prediction tools[53]. The 3′UTR of the messenger RNAs that serve as their targets is where miRNAs bind[57]. The PolymiRTS Database revealed

that the 16 indels and 55 SNPs have functional effects on various miRNA binding sites. Previous variants disrupted conserved sites of 131 miRNAs and created new binding sites for 149 miRNAs. Furthermore, no indels and 10 functionally verified SNPs (of 5′UTR variants) were predicted to affect transcriptional regulation by influencing the activity of TFBSs.

Genetic interaction is the set of functional association between genes. Gene interactions occur when two or more allelic or non-allelic genes of same genotype influence the outcome of particular phenotypic characters. To understand the molecular basis of this complex biological phenomenon, there is a need of genetic interaction mapping where the effects on one gene are modified by one or several other genes. The gene–gene interaction network of the HLA-DRB1 target gene and the closest 20 genes was built using GeneMANIA. A potent tool for systematically defining gene function and pathways is mapping genetic interactions, accomplished by simultaneously perturbing pairs of genes

that report how genes interact with one another[58]. A case of extreme genetic interaction is synthetic lethality, in which two mutations combine to create a lethal double mutant phenotype even though neither of them would be fatal on their own[59]. Most proteins work consecutively with other proteins in living organisms. Thus, PPI studies give crucial information for comprehending the complicated biological processes that occur in live cells[60]. Thus, to gain a better understanding, a network of PPIs was constructed using the inBio-Map resource. Deleterious variants in the HLA-DRB1 protein could disrupt its interaction with confidence interaction proteins.

Conclusion

HLA-DRB1 gene plays an important role in organ transplantation rejection and many other diseases. The current study shows the *in silico* analysis of genetic variants within the coding region, and 3′/5′ UTRs. Pathological variants may have a direct or indirect impact on the intramolecular/intermolecular interactions of amino acid residues, protein expression, and disease risks. We discovered significant structural and functional changes in HLA-DRB1 proteins by analyzing the conformational changes and interactions of amino acid residues. These changes can explain the activity deviations caused by several variants. This is the first study to predict the effects of coding and 3′/5′ UTR variants (SNVs, MNVs, and indels) in *the HLA-DRB1 gene*. The findings demonstrate that employing *in silico* methods in biomedical research is extremely successful and has a major influence on the capacity to identify the source of genetic variation in diverse disorders. The study's results could be an important guide in the research of potential diagnostic and therapeutic interventions that require experimental mutational validation and large-scale clinical trials.

Ethical Approval

Ethical approval is not applicable for this article.

Statement of Human and Animal Rights

This article does not contain any studies with human or animal subjects.

Statement of Informed Consent

There are no human subjects in this article and informed consent is not applicable.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

ORCID iD

Mohamed M. Hassan 🆔 https://orcid.org/0000-0003-1544-7932

References

1. Choo SY. The HLA system: genetics, immunology, clinical testing, and clinical implications. Yonsei Med J. 2007;48(1):11–23.
2. Mehra NK, Kaur G. Histocompatibility antigen complex of man. In: Els. John Wiley & Sons; 2016, p. 1–8. doi:10.1002/9780470015902.a0001234.pub4.
3. Fernando MMA, Stevens CR, Walsh EC, De Jager PL, Goyette P, Plenge RM, Vyse TJ, Rioux JD. Defining the role of the MHC in autoimmunity: a review and pooled analysis. PLoS Genet. 2008;4(4):e1000024.
4. Cruz-Tapias P, Castiblanco J, Anaya JM. Major histocompatibility complex: antigen processing and presentation. In: Anaya JM, Shoenfeld Y, Rojas-Villarraga A, Levy RA, Cervera R editors. Autoimmunity: from bench to bedside [Internet]. Bogota (Colombia): El Rosario University Press; 2013. Chapter 10.
5. Maksymowych WP, Van Kerckhove C, Glass DN. Juvenile rheumatoid arthritis, human leukocyte antigen, and other immunoglobulin supergene family polymorphisms. Am J Med. 1988;2385(6A):26–28.
6. Cao HX, Li M, Nie J, Wang W, Zhou SF, Yu XQ. Human leukocyte antigen DRB1 alleles predict risk and disease progression of immunoglobulin A nephropathy in Han Chinese. Am J Nephrol. 2008;28(4):684–91.
7. Sayers EW, Bolton EE, Brister JR, Canese K, Chan J, Comeau DC, Connor R, Funk K, Kelly C, Kim S, Madej T, et al. Database resources of the national center for biotechnology information. Nucleic Acids Res. 2023;51:D29–38.
8. Gouw JW, Jo J, Meulenbroek LAPM, Heijjer TS, Kremer E, Sandalova E, Knulst AC, Jeurink PV, Garssen J, Rijnierse A, Knippels LMJ. Identification of peptides with tolerogenic potential in a hydrolysed whey-based infant formula. Clin Exp Allergy. 2018;48(10):1345–53.
9. Castelli EC, Ramalho J, Porto IO, Lima TH, Felício LP, Sabbagh A, Donadi EA, Mendes-Junior CT. Insights into HLA-G genetics provided by worldwide haplotype diversity. Front Immunol. 2014;5:476.
10. Yan C, Wang R, Li J, Deng Y, Wu D, Zhang H, Zhang H, Wang L, Zhang C, Sun H, Zhang X, et al. HLA-A gene polymorphism defined by high-resolution sequence-based typing in 161 Northern Chinese Han people. Genom Proteom Bioinform. 2003;1(4):304–309.
11. Robinson J, Barker DJ, Georgiou X, Cooper MA, Flicek P, Marsh SGE. IPD-IMGT/HLA database. Nucleic Acids Res. 2020;48(D1):D948–55.
12. Alcina A, Abad-Grau Mdel M, Fedetz M, Izquierdo G, Lucas M, Fernández O, Ndagire D, Catalá-Rabasa A, Ruiz A, Gayán J, Delgado C, et al. Multiple sclerosis risk variant HLA-DRB1*1501 associates with high expression of DRB1 gene in different human populations. PLoS ONE. 2012;7(1):e29819.
13. Creary LE, Mallempati KC, Gangavarapu S, Caillier SJ, Oksenberg JR, Fernández-Viňa MA. Deconstruction of HLA-DRB1*04:01:01 and HLA-DRB1*15:01:01 class II haplotypes using next-generation sequencing in European-Americans with multiple sclerosis. Mult Scler. 2019;25(6):772–82.

14. Hachicha H, Kammoun A, Mahfoudh N, Marzouk S, Feki S, Fakhfakh R, Fourati H, Haddouk S, Frikha F, Gaddour L, Hakim F, et al. Human leukocyte antigens-DRB1*03 is associated with systemic lupus erythematosus and anti-SSB production in South Tunisia. Int J Health Sci (Qassim). 2018;12(1):21–27.

15. Gombos Z, Hermann R, Kiviniemi M, Nejentsev S, Reimand K, Fadeyev V, Peterson P, Uibo R, Ilonen J. Analysis of extended human leukocyte antigen haplotype association with Addison's disease in three populations. Eur J Endocrinol. 2007;157(6):757–61.

16. Wang J, Zhang H, Wang GQ, Quan Y. HLA-DRB1 gene polymorphisms and its associations with rheumatoid arthritis in Chinese Han women of Shaanxi province, northwest of China. Int J Immunogenet. 2016;43(1):25–31.

17. Wang L, Li B, Tie X, Liu T, Zheng S, Liu Y. Association between HLA-DRB1* allele polymorphism and caries susceptibility in Han Chinese children and adolescents in the Xinjiang Uygur Autonomous Region. J Int Med Res. 2020;48(4):300060519893852.

18. Heinold A, Opelz G, Döhler B, Unterrainer C, Scherer S, Ruhenstroth A, Tran TH. Deleterious impact of HLA-DRB1 allele mismatch in sensitized recipients of kidney retransplants. Transplantation. 2013;95(1):137–41.

19. Esmaeilzadeh H, Nabavi M, Amirzargar AA, Aryan Z, Arshi S, Bemanian MH, Fallahpour M, Mortazavi N, Rezaei N. HLA-DRB and HLA-DQ genetic variability in patients with aspirin-exacerbated respiratory disease. Am J Rhinol Allergy. 2015;29(3):e63–19.

20. 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, Abecasis GR. A global reference for human genetic variation. Nature. 2015;526(7571):68–74.

21. Guerra R, Yu Z. Single nucleotide polymorphisms and their applications. In: Zhang W, Shmulevich I, editors. Computational and statistical approaches to genomics. Boston (MA): Springer; 2006, p. 311–349.

22. Zou H, Wu LX, Tan L, Shang FF, Zhou HH. Significance of single-nucleotide variants in long intergenic non-protein coding RNAs. Front Cell Dev Biol. 2020;8:347.

23. Hossain MS, Roy AS, Islam MS. In silico analysis predicting effects of deleterious SNPs of human RASSF5 gene on its structure and functions. Sci Rep. 2020;10(1):14542.

24. Lin M, Whitmire S, Chen J, Farrel A, Shi X, Guo JT. Effects of short indels on protein structure and function in human genomes. Sci Rep. 2017;7(1):9313.

25. Studer RA, Dessailly BH, Orengo CA. Residue mutations and their impact on protein structure and function: detecting beneficial and pathogenic changes. Biochem J. 2013;449(3):581–94.

26. Li MJ, Yan B, Sham PC, Wang J. Exploring the function of genetic variants in the non-coding genomic regions: approaches for identifying human regulatory variants affecting gene expression. Brief Bioinform. 2015;16(3):393–412.

27. Hassan MM, Omer SE, Khalf-Allah RM, Mustafa RY, Ali IS, Mohamed SB. Bioinformatics approach for prediction of functional coding/noncoding simple polymorphisms (SNPs/Indels) in human BRAF gene. Adv Bioinformatics. 2016;2016:2632917.

28. Gagliano SA, Sengupta S, Sidore C, Maschio A, Cucca F, Schlessinger D, Abecasis GR. Relative impact of indels versus SNPs on complex disease. Genet Epidemiol. 2019;43(1):112–17.

29. Karchin R. Next generation tools for the annotation of human SNPs. Brief Bioinform. 2009;10(1):35–52.

30. Lamparter D, Marbach D, Rueedi R, Kutalik Z, Bergmann S. Fast and rigorous computation of gene and pathway scores from SNP-based summary statistics. PLoS Comput Biol. 2016;12(1):e1004714.

31. Wang B, Francis J, Sharma M, Law SM, Predeus AV, Feig M. Long-range signaling in MutS and MSH homologs via switching of dynamic communication pathways. PLoS Comput Biol. 2016;12(10):e1005159.

32. Jamal MS, Parveen S, Beg MA, Suhail M, Chaudhary AG, Damanhouri GA, Abuzenadah AM, Rehan M. Anticancer compound plumbagin and its molecular targets: a structural insight into the inhibitory mechanisms using computational approaches. PLoS ONE. 2014;9(2):e87309.

33. Ahmed F, Kaundal R, Raghava GP. PHDcleav: a SVM based method for predicting human Dicer cleavage sites using sequence and secondary structure of miRNA precursors. BMC Bioinform. 2013;14(Suppl. 14):S9.

34. Sim NL, Kumar P, Hu J, Henikoff S, Schneider G, Ng PC. SIFT web server: predicting effects of amino acid substitutions on proteins. Nucleic Acids Res. 2012;40(Web Server issue):W452–7.

35. 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. Nat Methods. 2010;7(4):248–49.

36. Bendl J, Stourac J, Salanda O, Pavelka A, Wieben ED, Zendulka J, Brezovsky J, Damborsky J. PredictSNP: robust and accurate consensus classifier for prediction of disease-related mutations. Plos Comput Biol. 2014;10(1):e1003440.

37. Tang H, Thomas PD. PANTHER-PSEP: predicting disease-causing genetic variants using position-specific evolutionary preservation. Bioinformatics. 2016;32(14):2230–32.

38. 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 Genom. 2013;14(Suppl. 3):S6.

39. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the functional effect of amino acid substitutions and indels. PLoS ONE. 2012;7(10):e46688.

40. Hecht M, Bromberg Y, Rost B. Better prediction of functional effects for sequence variants. BMC Genom. 2015;16(Suppl. 8):S1.

41. Capriotti E, Fariselli P, Calabrese R, Casadio R. Predicting protein stability changes from sequences using support vector machines. Bioinformatics. 2005; 21(Suppl. 2):ii54–8.

42. Blum M, Chang H, Chuguransky S, Grego T, Kandasaamy S, Mitchell A, Nuka G, Paysan-Lafosse T, Qureshi M, Raj S, Richardson L, et al. The InterPro protein families and domains database: 20 years on. Nucleic Acids Res. 2021;49(D1):D344–54.

43. 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 Res. 2016;44(W1):W344–50.

44. Venselaar H, Te Beek TA, Kuipers RK, Hekkelman ML, Vriend G. Protein structure analysis of mutations causing

inheritable diseases. An e-Science approach with life scientist friendly interfaces. BMC Bioinformatics. 2010;11:548.

45. Ittisoponpisan S, Islam SA, Khanna T, Alhuzimi E, David A, Sternberg MJE. Can predicted protein 3D structures provide reliable insights into whether missense variants are disease associated? J Mol Biol. 2019;431(11):2197–2212.

46. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc. 2015;10(6):845–58.

47. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Rempfer C, Bordoli L, Lepore R, et al. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. 2018;46(W1):W296–303.

48. Wang W, Li Z, Wang J, Xu D, Shang Y. PSICA: a fast and accurate web service for protein model quality analysis. Nucleic Acids Res. 2019;47(W1):W443–50.

49. McGuffin LJ, Aldowsari FMF, Alharbi SMA, Adiyaman R. ModFOLD8: accurate global and local quality estimates for 3D protein models. Nucleic Acids Res. 2021;49(W1):W425–30.

50. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, Flicek P, Cunningham F. The ensembl variant effect predictor. Genome Biol. 2016;17(1):122.

51. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A. Protein identification and analysis tools on the ExPASy server. In: Walker JM, editor. The proteomics protocols handbook. Totowa (NJ): Humana Press; 2005. p. 571–607.

52. Bhattacharya A, Ziebarth JD, Cui Y. PolymiRTS database 3.0: linking polymorphisms in microRNAs and their target sites with human diseases and biological pathways. Nucleic Acids Res. 2014;42(Database issue):D86–91.

53. Xu Z, Taylor JA. SNPinfo: integrating GWAS and candidate gene information into functional SNP selection for genetic association studies. Nucleic Acids Res. 2009; 37(Web Server issue):W600–5.

54. Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, Franz M, Grouios C, Kazi F, Lopes CT, Maitland A, et al. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. Nucleic Acids Res. 2010;38(Web Server issue):W214–20.

55. Li T, Wernersson R, Hansen RB, Horn H, Mercer J, Slodkowicz G, Workman CT, Rigina O, Rapacki K, Stærfeldt HH, Brunak S, et al. A scored human protein-protein interaction network to catalyze genomic interpretation. Nat Methods. 2017;14(1):61–64.

56. Hassan MM, Dowd AA, Mohamed AH, Mahalah SM, Kaheel HH, Mohamed SN, Hassan MA. Computational analysis of deleterious nsSNPs within HLA-DRB1 and HLA-DQB1 genes responsible for Allograft rejection. Int J Comput Bioinform Silico Model. 2014;3(6):562–77.

57. O'Brien J, Hayder H, Zayed Y, Peng C. Overview of MicroRNA biogenesis, mechanisms of actions, and circulation. Front Endocrinol (Lausanne). 2018;9:402.

58. Kaushik S, Kaushik S, Sharma D. Encyclopedia of bioinformatics and computational biology. Cambridge (MA): Academic Press; 2019. Vol. 2, p. 118–33.

59. Fang G, Wang W, Paunic V, Heydari H, Costanzo M, Liu X, Liu X, VanderSluis B, Oately B, Steinbach M, Van Ness B, et al. Discovering genetic interactions bridging pathways in genome-wide association studies. Nat Commun. 2019;10(1):4274.

60. Glass F, Takenaka M. The yeast three-hybrid system for protein interactions. In: Oñate-Sánchez L, editor. Two-hybrid systems. Methods in molecular biology. New York (NY): Humana Press; 2018. p. 195–205.