Volume 1 • Issue 1 • 2019

Research Article

Journal of Biotechnology and Immunology

Machine Learning and Insilico Analysis of Single Nucleotide Polymorphism (SNPs) in Human HLA-DRB1Gene

Mona A. M.Khaier[1]*, Lina M. T. Fadl[2], Afra M. Al Bkrye[1] and Hind A. Elnasri[1]

[1]Department of Molecular Biology and Bioinformatics, College of Veterinary Medicine, University of Bahri, Khartoum, Sudan.

[2]Department of Pathology, Faculty of Medicine, University of Khartoum, Khartoum, Sudan.

***Corresponding author:** Mona A. M. Khaier, Department of Molecular Biology and Bioinformatics, College of Veterinary Medicine, University of Bahri, Khartoum, Sudan.

Received: October 10, 2019; **Published:** October 31, 2019

Abstract

Background: This study aimed to investigate the association between non synonymous SNPs found in HLA-DRB1Gene and autoimmune diseases using insilco approach. The HLA complex helps the immune system distinguish the body's own proteins from proteins made by foreign invaders such as viruses and bacteria. The HLA-DRB1 gene provides instructions for making a protein that plays a critical role in the immune system. The HLA-DRB1 gene is part of a family of genes called the human leukocyte antigen (HLA) complex. The human leukocyte antigen (HLA) locus has the major genetic contribution to autoimmune diseases like Rheumatoid Arthritis (RA), with HLA-DRB1 alleles showing the strongest association within the locus.

Methods: Different nsSNPs and protein related sequences were obtained from dbSNP/NCBI and ExPASY database during 2019. The interaction of HLA-DRB1 gene with other genes was studied using GenMANIA software. The nsSNPs with damaging effect were chosen using SIFT software. They were furtheranalyzed for their effect using Provean, Polyphen- 2, and SNPs & GO. Protein stability was investigated using I-Mutant and MUpro software. The structural and functional impacts of point mutations were predicted using Project Hope software.

Results: HLA-DRB1gene was found to have an association with 20 genes most of them are HLA genes like HLA-DRA which have the strongest association. A total of 169 SNPs were retrieved from NCBI, 139 were non synonymous. Using SIFTsoftware 60 nsSNPS were found to be deleterious. By Provean software 84 nsSNPS were found to be deleterious. The double positive deleterious nsSNPS (46) were taken as an input for Polyphen-2 software. This resulted in 24 probably damaging nsSNPS. Using I-Mutant and MUpro software 19 nsSNPS were found to decrease the protein stability. SNP&GO and PHD-SNPs showed only 14 nsSNP as disease related. The effect of the amino acid substitution on the protein structure, charge and hydrophobicity revealed a truncated protein using Project Hope software.

Conclusions: After analysis using the different software, a total of 14nsSNPs were found to be disease related.The 14 nsSNPs detected in this study resulted in different proteins.

Key words: *Autoimmune diseases; HLA-DRB1gene; Insilico analysis; Non synonymous SNPs.*

Citation: Mona A. M.Khaier, Lina M. T. Fadl, Afra M. Al Bkrye and Hind A. Elnasri. (2019). Machine Learning and Insilico Analysis of Single Nucleotide Polymorphism (SNPs) in Human HLA-DRB1Gene. *Journal of Biotechnology and Immunology* 1(1).

Introduction

The HLA complex helps the immune system distinguish the body's own proteins from proteins made by foreign invaders such as viruses and bacteria. (Parkes, et al., 2013). HLA-DRB1 gene, is located at 6p21.32, which is the short (p) arm of chromosome 6 at position 21.32

Molecular Location: base pairs 32,578,769 to 32,589,836 on chromosome 6 (Homo sapiens Annotation Release 109, GRCh38.p12) (NCBI)

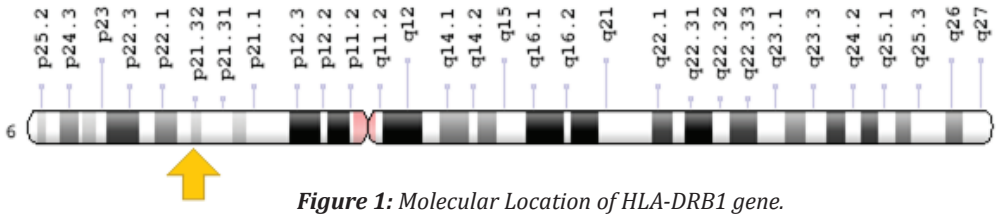


Figure 1: *Molecular Location of HLA-DRB1 gene.*

The HLA complex is the human version of the major histocompatibility complex (MHC), a gene family that occurs in many species. The HLA-DRB1 gene belongs to a group of MHC genes called MHC class II. MHC class II genes provide instructions for making proteins that are present on the surface of certain immune system cells. These proteins attach to protein fragments (peptides) outside the cell. MHC class II proteins display these peptides to the immune system. If the immune system recognizes the peptides as foreign (such as viral or bacterial peptides), it triggers a response to attack the invading viruses or bacteria (Genetic Home Reference, 2018).

Each MHC class II gene has many possible variations, allowing the immune system to react to a wide range of foreign invaders. Researchers have identified hundreds of different versions (alleles) of the HLA-DRB1 gene, each of which is given a particular number, such as HLA-DRB1*04:01 (Kerlanet al., 2001, Poland, et al., 2007; Stahl et al., 2010).

Several common variations of the HLA-DRB1 gene are associated with a risk of developing rheumatoid arthritis. This disease causes chronic abnormal inflammation that primarily affects the joints. HLA-DRB1 is one of several genes in the HLA complex that have been associated with rheumatoid arthritis. Variations in this gene are the most significant known genetic risk factor for the disease. The HLA-DRB1 gene variations associated with an increased risk of rheumatoid arthritis affect single protein building blocks in the beta chain. These changes occur near the antigen-recognizing binding groove, which is the part of the protein that attaches (binds) to viral or bacterial peptides (Hegab, 2016). This binding triggers the immune response that attacks foreign invaders. Although the mechanism by which HLA-DRB1 gene variations increase the risk of rheumatoid arthritis is unclear, researchers suspect it is related to changes in peptide binding that stimulate an abnormal immune response. However, many other genetic and environmental factors also contribute to a person's overall risk of developing rheumatoid arthritis. A few variations of the HLA-DRB1 gene appear to decrease the risk of developing rheumatoid arthritis. It is unclear why these particular changes may be protective (Elias, et al., 2016 and Xiao, et al., 2016).

Using 54 markers distributed across the entire HLA complex, Jawaheeret al., (2002) performed an extensive haplotype analysis in a set of 469 multiplex families with RA. The frequency of HLA-DRB1*13 alleles was significantly higher in mycetoma patients ($P<0.044$)(Dawiet al., 2013). RA is associated with HLA DRB1 alleles, in Caucasians the frequencies of alleles DRB1 *0401 and DRB1*0404 are increased in patients compared with controls. (Nepomet al., 1986); the frequency of DRB1*0405 is increased in Japanese patients (Watanabe et al., 1989). An excess of DRB1*0101 also has been reported by Shiff et al., 1982. At residues 67-74, the amino acid sequence of DRB1*0404 and DRB1*0405 is identical, whereas the sequence of DRB1 * 0401 differs at only one site; and, at these same sites, the chain encoded by DRB1*0101 is identical to that encoded byDRB1*0404 and DRB1*0405. It therefore has been proposed that RA may be associated primarily with a shared epitope involving residues 67-74 (Gregersenet al., 1987).

Single nucleotide polymorphisms (SNPs) are the most common genetic variations in any population; they occur when a single nucleotide in the genome is altered (Nachman, 2001). They are present in every 200–300 bp in human genome (Lee, et al., 2005). So far, 5,000,000 SNPs have been identified in the coding region of human population responsible for genetic variation diseases, among all SNPs, non-synonymous SNPs (ns SNPs) are present in exonic part of genome, which often leads to change in amino acid residues of gene product(Rajasekaranet al., 2008). Even though many SNP's

have no effect on the biological functions of the cell, some can predispose people to certain diseases, influence their immunological response to drugs and can be considered as biomarkers for disease susceptibility(Kamataniet al., 2004). Importantly, nsSNPs result in changes of the amino acid sequence of proteins and have been reported to be responsible for about 50% of all known genetic variations that are linked to inherited diseases (Krawczaket al., 2000). On the other hand, coding synonymous and those seen outside gene coding or promoter regions may also influence transcription factor binding and gene expression (Prokunina and Alarcón, 2004; Stenson, et al., 2009).

Objectives: This study aimed to investigate the association between non synonymous SNPs found in HLA-DRB1Gene and autoimmune diseases using insilico approach. The effect of the nsSNPs on protein function and stability was also studied

Material and Methods

Single nucleotide polymorphisms (SNPs) for DBR1 gene were obtained from National Center of Biological Information (NCBI) SNPs database. The SNPs and the related Ensembles proteins (ESNP) were obtained from the SNPs database (dbSNPs) for computational analysis from http://www.ncbi.nlm.nih.gov/snp and Uniprot database. The important step in this study was to select SNPs for analysis by computational software. The selection was targeting SNPs in the coding region (exonal SNPs) that are non-synonymous (nsSNPs). Several software used.

GeneMANIA (http://www.genemania.org) is a web interface that helps predicting the function of genes and gene sets. GeneMANIA finds other genes that are related to a set of input genes, using a very large set of functional association data, such as protein and genetic interactions, pathways, co-expression, co-localization and protein domain similarity.

The input was the gene name and the results were shown in a graphical form showing the different relations between the genes (Warde-Farley, et al., 2010).

Sorting Intolerant FromTolerant (SIFT) (http://siftdna.org/www/SIFTdbSNP) It predicts the tolerated and deleterious SNPs and identifies the impact of amino acid substitution on protein function and phenotype alterations, so that users can prioritize substitutions for further studies. The main principle of this program is that it generates alignments with a large number of homologous sequences and assigns scores to each residue ranging from zero to one. The threshold intolerance score for SNPs is 0.05 or less (Ng PC, Henikoff, 2003; González-Pérez et al., 2011).

Protein Variation Effect Analyzer (PROVEAN) http://provean.jcvi.org/index.php) It is a software tool which predicts whether an amino acid substitution or indel has an impact on the biological function of a protein. PROVEAN is useful for filtering sequence variants to identify nonsynonymous or indel variants that are predicted to be functionally important. The performance of PROVEAN is comparable to popular tools such as SIFT or PolyPhen-2. A fast computation approach to obtain pairwise sequence alignment scores enabled the generation of precomputed PROVEAN predictions for 20 single AA substitutions and a single AA deletion at every amino acid position of all protein sequences in human and mouse. The input for this software were a protein sequence in a FASTA format and an amino acid variant. Choi, 2015)

Polymorphism Phenotyping (Polyphen-2), (http://genetics.bwh.harvard.edu/pph2/) is an online bioinformatics program which automatically predicts the consequence of an amino acid change on the structure and function of a protein. This prediction is based on a number of features comprising the sequence, phylogenetic and structural information characterizing substitution. This program basically searches for 3D protein structures, multiple alignments of homologous sequences and amino acid contact information in several protein structure databases, then calculates position-specific independent count scores (PSIC) for each of the two variants and then computes the PSIC scores difference between the two variants. The higher a PSIC score difference, the higher the functional impact a particular amino acid substitution is likely to have. Prediction outcomes could be classified as benign, possibly damaging or probably damaging, according to the posterior probability intervals (0, 0.2), (0.2, 0.85) and (0.85, 1), respectively. nsSNPs that were predicted to be intolerant by SIFT has been submitted to Polyphen-2 as a protein sequence in FASTA format obtained from UniprotKB/Expasy after submitting the relevant ensemble protein (ESNP) there. The position of mutation should be submitted together with the native amino acid and the new substituent for both structural and functional predictions (Venselaaret al., 2010).

I-Mutantversion 3.0 (http://gpcr2.biocomp.unibo.it/cgi/predictors/I-Mutant3.0/I-Mutant3.0.cgi) was used to predict protein stability changes in single-site mutations. I-Mutant basically

can evaluate the stability change of a single site mutation starting from the protein structure or from the protein sequences (Bavaet al., 2004).

Single Nucleotide Polymorphism &Gene Ontology (SNPs & Go): (http://snps.biofold.org/snps-and-go//snps-and-go.html).

It is software that predicts the disease related mutations from protein FASTA sequence. Its output is a prediction results based on the discrimination among disease related and neutral variations of protein sequence. The probability score higher than 0.5 reveals the disease related effect of mutation (Capriotti et al., 2013).

Predictor of human deleterious single nucleotide polymorphisms (PHD-SNP) (http://snps. biofold.org/phd-snp/phd-snp.html). It predicts whether the new phenotype derived from a SNP is disease-related or not (neutral). Protein sequence from Uniprot is submitted to the program after providing position of mutation and the new amino acid residue (Capriotti et al., 2013).

Project HOPE (Available at: http://www.cmbi.ru.nl/hope/home).

It is an easy-to-use web server that analyses the structural effects of intended mutation. HOPE provides the 3D structural visualization of mutated proteins, and gives the results by using Uniprot and DAS prediction servers. Input method of Project HOPE carries the protein sequence and selection of Mutant variants. HOPE server predicts the output in the form of structural variation between mutant and wild type residues (Venselaar et al., 2010).

Reesults and Discussion

This study aimed to investigate the effect of nsSNPs found in DRB1 gene on the protein structure and function. Using GeneMANIA, HLA-DRB1 gene was found to have an association with 20 genes. HLA-DR1 gene was the one which had a strong association with HLA- DRB1 gene, Figure 2.

SNP in the CDS region obtained from dbSNPS were subjected for further analysis using different software. A total of 60 SNPs were detected to be deleterious using SIFT software. Provean revealed 84 SNPs with deleterious effect. Polephen-2 detected 24 probably damaging, Figure 3

SNPs are also known to affect the protein stability which may influence the function of the protein. From the 24 nsSNPS which were found to be probably damaging by Polyphen-2 software, 21 were found to decrease the protein stability using I mutant software but, when using MUpro software, 22 nsSNPs showed to decrease the protein stability, Figure 4

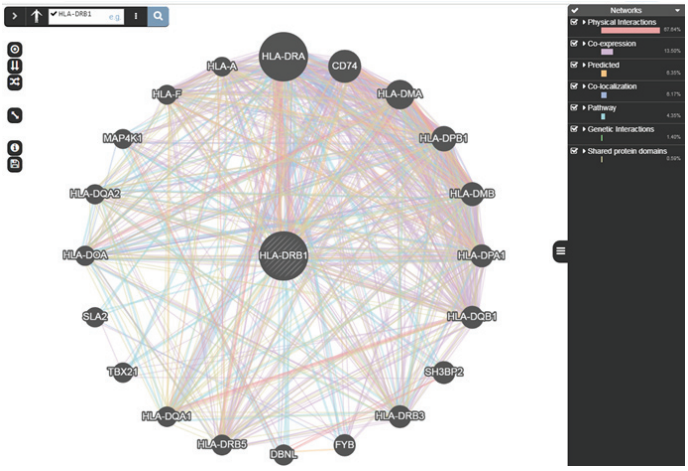


Figure 2: *GenMania showing co=expression.*

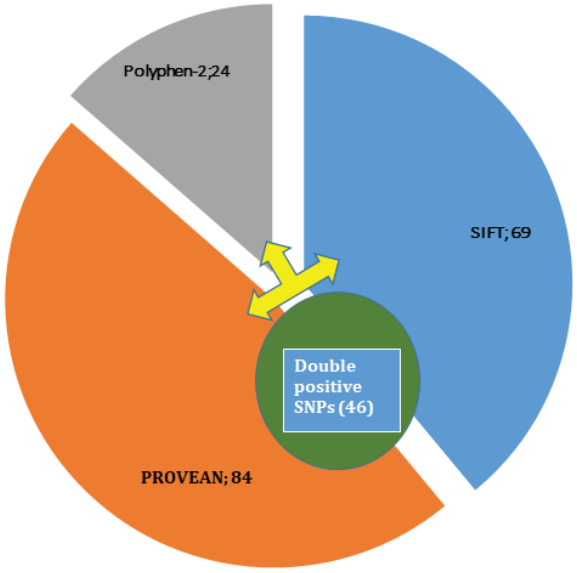


Figure 3: *Results of SIFT, Provean (Deleterious SNPs) and Polyphen -2 (probably damaging) software.*

To evaluate which of the SNPs is disease related (≥0.5) or not disease related (≤0.5) probability, two software were used namely SNPs & GO and PhD –SNP. SNP&GO showed 16 SNPs to be disease related while PhD-SNP showed 15 SNPs to be related, Only 14 SNPs were positive in both Software, Figure 5

To investigate the effect of amino acid substitution on the protein function and structure, Project Hope was used. The figure below shows the schematic structures of the original (left) and the

mutant (right) amino acid. The backbone, which is the same for each amino acid, is colored red. The side chain, unique for each amino acid, is colored black. The results of the 14 nsSNPs were shown on Table 1.

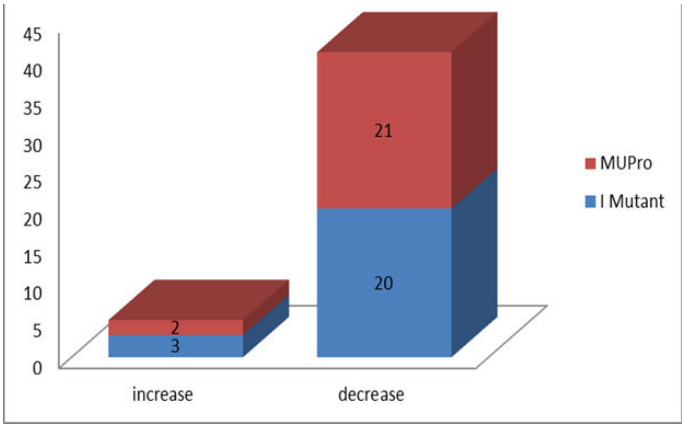


Figure 4: *Effect of SNPs on protein stability using I Mutant and MUPro.*

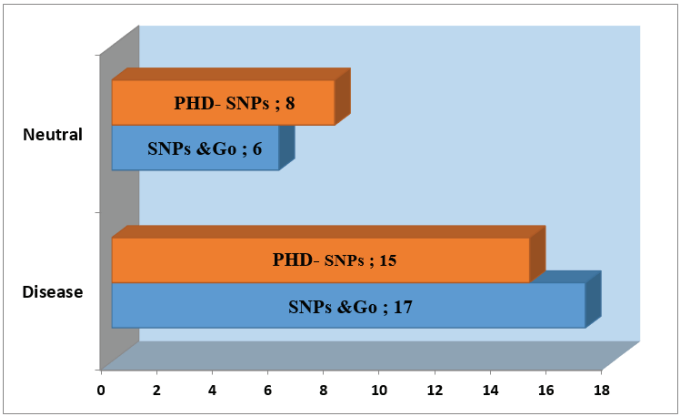
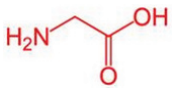
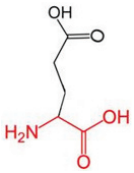
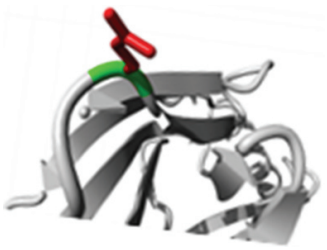
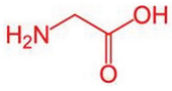
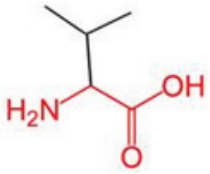
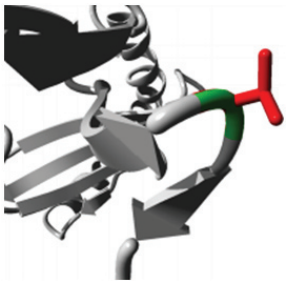
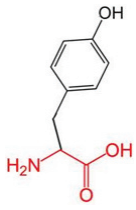
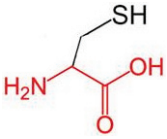
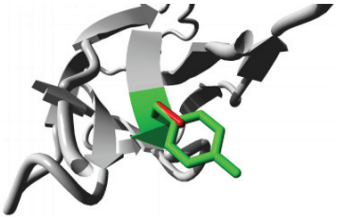
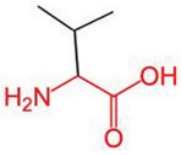
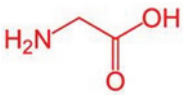
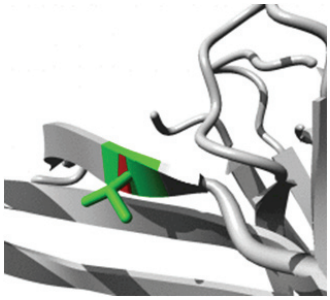
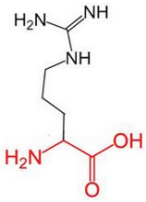
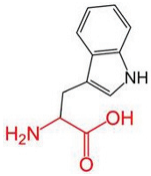
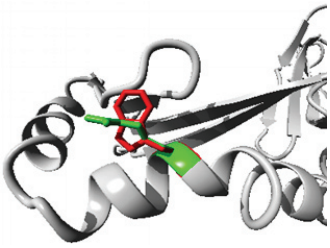
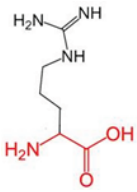
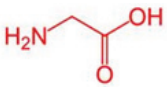
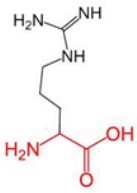
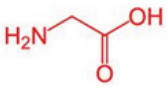
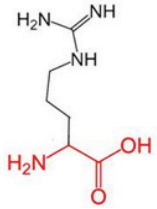
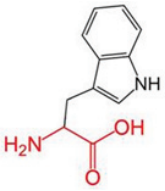
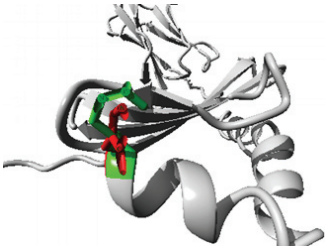
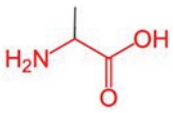
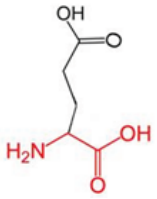
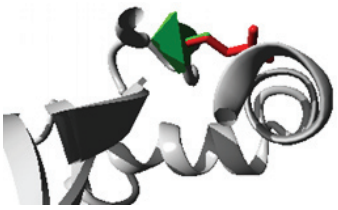
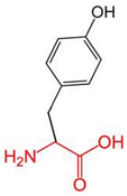
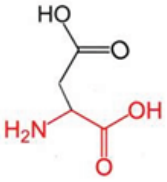
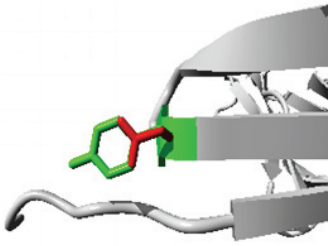
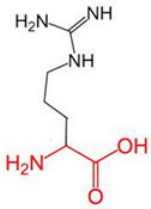
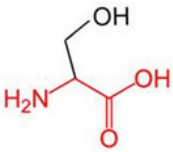


Figure 5: *Results of SNPs & GO.PHD- SNPs software.*

rs and Amino acid change	Wild and mutant amino acids chemical structure	Project Hope Result Description	Projccrt Hope illustartion of mutated residue (The protein is coloured grey, the side chains of wild-type and the mutant residue are shown and coloured green and red respectively)
rs2308775 Glycine> Glutamic Acid at position 197 G197E	 Glycine  Glutamic Acid	The mutant residue is bigger than the wild-type residue. The residue is located on the surface of the protein, mutation of this residue can disturb interactions with other molecules or other parts of the protein	
rs707954 Glycine > Valine at position 164. G164V	 Glycine  Glutamic Acid	The mutant residue is bigger and more hydrophobic than the wild-type residue. The mutation is located within a domain, annotated in UniProt as:Ig-like C1-type and the mutation can disturb this domain and abolish its function.	

rs112796209 Tyrosine>Cysteine at position 152 Y152C	 Tyrosine  Cysteine	The mutant residue is smaller and more hydrophobic than the wild-type residue. The new residue might be too small to make multimer contacts. Also, any hydrogen bond that could be made by the wild-type residue to other monomers will be lost now and affect the multimeric contacts due to hydrophobicity.	
rs80190494 Valine> Glycine at position 128. V128G	 Valine  Glycine	The mutant residue is smaller than the wild-type residue.and the wild-type residue is more hydrophobic than the mutant residue.The mutation is located within a domain, annotated in UniProt as:Ig-like C1-type The mutation introduces an amino acid with different properties, which can disturb this domain and abolish its function(Glycines are very flexible and can disturb the required rigidity of the protein at this position.)	
rs17885222 Arginine > Tryptophan At position 101 R101W	 Arginine  Tryptophan	The mutant residue is bigger than the wild-type residue. The wild-type residue is positively charged while, the mutant residue charge is neutral The charge of the wild-type residue is lost by this mutation. The mutant residue is more hydrophobic than the wild-type residue This can cause loss of interactions with other molecules	

rs17885222 Arginine> Glycine at position 101 R101G	 Arginine  Glycine	The mutant residue is smaller than the wild-type residue. The wild-type residue charge was POSITIVE, the mutant residue charge is NEUTRAL. The mutant residue is more hydrophobic than the wild-type residue.	
rs1059583 Alanine> Proline at position A87P	 Alanine  Proline	The mutant residue is bigger than the wild-type residue. The wild-type residue is located in a region annotated in UniProt to form an α-helix. The helix will be disturbed and this can have severe effects on the structure of the protein. The wild-type residue was buried in the core of the protein. The mutant residue is bigger and probably will not fit.	
rs17878902 Arginine> Tryptophan at position 84 R84W	 Arginine  Tryptophan	The mutant residue is bigger than the wild-type residue. The wild-type residue charge was POSITIVE, the mutant residue charge is NEUTRAL. The mutant residue is more hydrophobic than the wild-type residue. The differences in amino acid properties can disturb this region and disturb its function.	

rs17879432 Alanine > Glutamic Acid at position 78 A78E	 Alanine  Glutamic Acid	The mutant residue is bigger than the wild-type residue. The wild-type residue charge was NEUTRAL; the mutant residue charge is NEGATIVE. The wild-type residue is more hydrophobic than the mutant residue. The mutation might cause loss of hydrophobic interactions with other molecules on the surface of the protein. The contact with other molecules or domains is still possible and might be affected by this mutation.	
rs1064664 Tyrosine> Aspartic Acid at position 61 Y 61D	 Tyrosine  Aspartic Acid	The mutant residue is smaller than the wild-type residue. The new residue might be too small to make multimer contacts, the new residue is not in the correct position to make the same hydrogen bond as the original wild-type residue did. The wild-type residue charge was NEUTRAL, the mutant residue charge is NEGATIVE. The wild-type residue is more hydrophobic than the mutant residue. This affect hydrogen bond formation. hydrophobicity is important for multimerisation and therefore this mutation could affect the multimer contacts.	
rs17878437 Arginine> Serine at positon 58 R58S	 Arginine  Serine	The mutant residue is smaller than the wild-type residue. This makes the mutant residue not in the correct position to make the same hydrogen bond as the original wild-type residue did. The new residue might be too small to make multimer contacts The wild-type residue charge was POSITIVE, the mutant residue charge is NEUTRAL. This can cause loss of interactions with other molecules. The mutant residue is more hydrophobic than the wild-type residue, the mutation is possibly not damaging to the protein.	

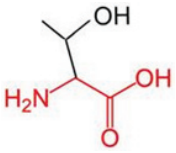
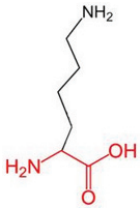
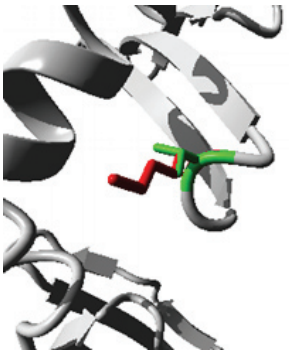
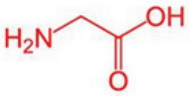
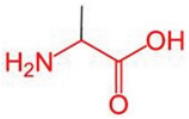
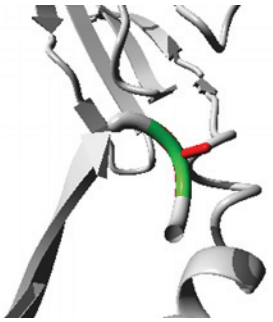
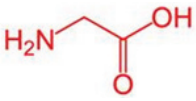
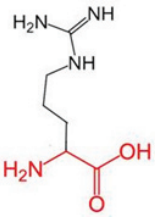
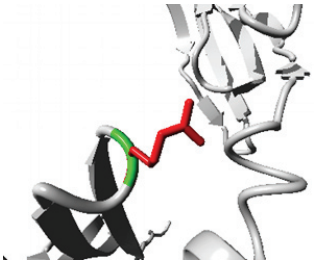
rs61759933 Threonine> Lysine at position 50 T50K	 Threonine  Lysine	The mutant residue is bigger than the wild-type residue. The wild-type residue charge was NEUTRAL; the mutant residue charge is POSITIVE. The wild-type residue is more hydrophobic than the mutant residue. This mutation might cause loss of hydrophobic interactions with other molecules on the surface of the protein; mutation of this residue can disturb interactions with other molecules or other parts of the protein. The hydrophobicity of the wild-type and mutant residue differs.	
rs17879469 Glycine> Alanine at position 49 G49A	 Glycine  Alanine	The mutant residue is bigger than the wild-type residue. The mutant residue is more hydrophobic than the wild-type residue. The flexibility might be necessary for the protein's function. Mutation of this glycine can abolish this function. Mutation of this residue can disturb interactions with other molecules or other parts of the protein.	
rs61759931 Glycine> Arginine at position 49 G 49R	 Glycine  Arginine	The mutant residue is bigger than the wild-type residue. The wild-type residue charge was NEUTRAL, the mutant residue charge is POSITIVE, and this can cause repulsion between the mutant residue and neighboring residues. The wild-type residue is more hydrophobic than the mutant residue. The wild-type residue is a glycine, the most flexible of all residues. This flexibility might be necessary for the protein's function. Mutation of this glycine can abolish this function.	

Table 1: *Project Hope Results of different SNPs.*

In human HLA-DRB1 gene rs1059575, D57E which was defined as Benin in kidney transplant by Hassanet al., 2015, were also identified in this study as neutral nsSNPs for HLA-DRB1 gene, while rs17885437, G74R which was defined as deleterious in kidney transplant by Hassanet al., 2015 was found to be deleterious by two software in this study. rs2308775, G197E, rs17885222, R101W, rs17885222, R101G and rs112796209 Y152C which are found to be deleterious, decreasing the protein stability and disease related were also found to associated with cancer development and are reported as CanVAR by Li et al.,2010.

Within exon-2, in HLA-DRB1 gene three non-synonymous mutations were predicted to have deleterious effects on protein function. Moreover, six deletions and three insertions were found in 12% of the cases, resulting in significant loss of amino acid information. V67C, E51K and N48T in HLA-DRB1 gene were found in Omani populations AL-Hudaret al., 2015. The HLA-DRB1 gene is highly damaged in Omani rheumatoid arthritis patients. These mutations are not detected in the present study. This leads to the fact that genetic architecture of RA is different in multiple populations of European, Asian, African, and Middle Eastern ancestries.The CD28 (rs1980422) and PTPN22 (rs2476601) contribute to RA-susceptibility in Egyptians population, seven SNPs [rs2240340, rs1980422, rs5029939, rs10499194, rs6920220, rs2476601 and rs2395175] were detected. (Hegab, et al. 2016). In this study DRB1 gene was associated with different genes including CD74. Aldawi et al., 2013, stated that there is an association between HLA-DRB1 and HLA-DQB1 alleles variations and the occurrence of eumycetoma in Sudanese patients. To conclude HLA-DRB1 gene nsSNPS are affecting the protein structure and function and contribute to many autoimmune diseases.

Conclusions and Recommendations

The 14 nsSNPs detected in this study resulted in different proteins. The amino acids differ in size, structure, function, charge, hydrophobicity and location, resulting in a truncated protein. Further investigation using large-scale studies is needed to maximize the power of genetic association.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors

Appendices

SNP	AMINO ACID CHANGE	SIFT SCORE	SIFT PREDICTION	Provean score	Provean prediction cutoff=-2.5	Polyphen prediction	Polyphen score
rs377738927	A224E	0	Deleterious	-2.794	Deleterious	PROBABLY DAMAGING	1
rs1136846	P212A	0.018	Deleterious	-5.863	Deleterious	BENIGN	0.163
rs2308775	G197A	0.003	Deleterious	-4.367	Deleterious	PROBABLY DAMAGING	0.998
rs2308775	G197E	0.003	Deleterious	-5.742	Deleterious	PROBABLY DAMAGING	0.1
rs707954	G164V	0	Deleterious	-6.414	Deleterious	PROBABLY DAMAGING	0.987
rs112796209	Y152C	0	Deleterious	-7.207	Deleterious	PROBABLY DAMAGING	0.999
rs41557115	L143F	0.007	Deleterious	-3.099	Deleterious	PROBABLY DAMAGING	1
rs200516145	T135N	0.038	Deleterious	-3.543	Deleterious	POSSIBLY DAMAGING	0.926
rs17433947	T135A	0.027	Deleterious	-3.506	Deleterious	BENIGN	0.009
rs1136791	T129P	0.019	Deleterious	-4.146	Deleterious	PROBABLY DAMAGING	1

rs80190494	V128G	0.006	Deleterious	-5.565	Deleterious	PROBABLY DAMAGING	1
rs80190494	V128A	0	Deleterious	-3.181	Deleterious	PROBABLY DAMAGING	1
rs17884043	Y107D	0.04	Deleterious	-5.527	Deleterious	POSSIBLY DAMAGING	0.943
rs61759934	D105N	0.047	Deleterious	-3.892	Deleterious	BENIGN	0.048
rs17885222	R101W	0.015	Deleterious	-7.243	Deleterious	PROBABLY DAMAGING	1
rs17885222	R101G	0	Deleterious	-6.308	Deleterious	PROBABLY DAMAGING	0.998
rs17882450	E98G	0.001	Deleterious	-5.95	Deleterious	BENIGN	0.424
rs41308499	L97R	0.015	Deleterious	-5.019	Deleterious	PROBABLY DAMAGING	0.1
rs17878874	D95N	0.019	Deleterious	-3.875	Deleterious	BENIGN	0.004
rs17879230	E88G	0.026	Deleterious	-5.442	Deleterious	BENIGN	0.014
rs1059583	A87P	0.001	Deleterious	-4.293	Deleterious	PROBABLY DAMAGING	0.999
rs1059583	A87T	0.045	Deleterious	-3.335	Deleterious	POSSIBLY DAMAGING	0.822
rs17887012	P85A	0.011	Deleterious	-5.372	Deleterious	BENIGN	0.049
rs41308498	R84L	0.016	Deleterious	-6.055	Deleterious	BENIGN	0.099
rs17878902	R84W	0	Deleterious	-6.942	Deleterious	PROBABLY DAMAGING	1
rs1059582	T80M	0.002	Deleterious	-4.868	Deleterious	PROBABLY DAMAGING	1
rs1059582	T80R	0	Deleterious	-4.779	Deleterious	POSSIBLY DAMAGING	0.86
rs17879432	A78V	0.008	Deleterious	-3.456	Deleterious	POSSIBLY DAMAGING	1
rs17879432	A78E	0.013	Deleterious	-4.289	Deleterious	PROBABLY DAMAGING	1
rs29029548	F76L	0.006	Deleterious	-4.681	Deleterious	BENIGN	1
rs17885437	G74R	0.012	Deleterious	-7.399	Deleterious	PROBABLY DAMAGING	0.989
rs56158521	D70N	0.021	Deleterious	-4.393	Deleterious	BENIGN	0.132
rs55655909	E64G	0.012	Deleterious	-6.268	Deleterious	BENIGN	0.342
rs17879242	N62K	0.015	Deleterious	-5.383	Deleterious	POSSIBLY DAMAGING	0.781
rs17879995	N62H	0.017	Deleterious	-4.356	Deleterious	POSSIBLY DAMAGING	0.571
rs1064664	Y61D	0.001	Deleterious	-8.125	Deleterious	PROBABLY DAMAGING	1
rs17878437	R58S	0.001	Deleterious		Deleterious	PROBABLY DAMAGING	0.999

rs201128876	F55I	0.006	Deleterious	-3.344	Deleterious	POSSIBLY DAMAGING	0.473
rs17885382	R54L	0.008	Deleterious	-5.56	Deleterious	BENIGN	0.01
rs188617679	R54W	0.038	Deleterious	-5.966	Deleterious	BENIGN	0.036
rs17882378	R52P	0.001	Deleterious	-5.697	Deleterious	BENIGN	0.957
rs74812266	R52G	0.033	Deleterious	-5.719	Deleterious	BENIGN	0.371
rs61759933	T50K	0.014	Deleterious	-4.802	Deleterious	PROBABLY DAMAGING	0.963
rs17879469	G49A	0.033	Deleterious	-5.574	Deleterious	PROBABLY DAMAGING	1
rs61759931	G49R	0	Deleterious	-7.433	Deleterious	PROBABLY DAMAGING	1
rs17882014	E43K	0.036	Deleterious	-3.319	Deleterious	BENIGN	0.354

Table 1: *Results of SIFT, Provean and Polyphen-2 analysis of nsSNPs of HLA-DRB1 gene.*

Amino Acid Change	I-Mutant Svm2 Prediction Effect	DDG Value Prediction	Ri	MUPRO Result
A224E	DECREASE	-1.15	5	DECREASE
V204L	DECREASE	-1.12	6	DECREASE
G197A	DECREASE	-1.24	9	INCREASE
G197E	DECREASE	-1.12	7	DECREASE
G164V	DECREASE	-1.03	8	DECREASE
Y152C	DECREASE	-1.83	9	DECREASE
L143F	INCREASE	0.19	5	DECREASE
T129P	DECREASE	-0.37	1	DECREASE
V128G	DECREASE	-2.31	9	DECREASE
V128A	DECREASE	-1.68	9	DECREASE
R101W	DECREASE	-1.33	5	DECREASE
R101G	DECREASE	-2.91	9	DECREASE
L97R	INCREASE	0.02	3	DECREASE
A87P	DECREASE	-0.72	5	DECREASE
R84W	DECREASE	-0.22	2	DECREASE
T80M	DECREASE	-1.27	9	INCREASE
A78E	DECREASE	-0.61	1	DECREASE
G74R	INCREASE	-0.44	0	DECREASE
Y61D	DECREASE	0.02	2	DECREASE
R58S	DECREASE	-1.47	9	DECREASE
T50K	DECREASE	-1.72	8	DECREASE
G49A	DECREASE	-0.33	1	DECREASE
G49R	DECREASE	0.02	2	DECREASE

Table 2: *Results of I- mutant and MUpro.*

Mutation	Prediction	RI	Probability	Method
G49A	Disease	5	0.774	PhD-SNP
	Disease	4	0.714	SNPs&GO
G49R	Disease	7	0.870	PhD-SNP
	Disease	4	0.721	SNPs&GO
T50K	Disease	8	0.898	PhD-SNP
	Disease	5	0.750	SNPs&GO
R58S	Disease	7	0.862	PhD-SNP
	Disease	6	0.809	SNPs&GO
Y61D	Disease	8	0.889	PhD-SNP
	Disease	3	0.640	SNPs&GO
A78E	Disease	6	0.790	PhD-SNP
	Disease	5	0.736	SNPs&GO
R84W	Disease	7	0.851	PhD-SNP
	Disease	5	0.760	SNPs&GO
A87P	Disease	6	0.813	PhD-SNP
	Disease	4	0.687	SNPs&GO
R101G	Disease	6	0.825	PhD-SNP
	Disease	5	0.730	SNPs&GO
R101W	Disease	8	0.883	PhD-SNP
	Disease	5	0.774	SNPs&GO
V128A	Neutral	2	0.407	PhD-SNP
	Disease	1	0.549	SNPs&GO
T129P	Disease	1	0.549	PhD-SNP
	Neutral	2	0.422	SNPs&GO
Y152C	Disease	7	0.838	PhD-SNP
	Disease	4	0.683	SNPs&GO
G164V	Disease	6	0.796	PhD-SNP
	Disease	5	0.754	SNPs&GO
G197E	Disease	5	0.755	PhD-SNP
	Disease	6	0.794	SNPs&GO
V204L	Neutral	1	0.425	PhD-SNP
	Disease	2	0.587	SNPs&GO
A224E	Neutral	5	0.269	PhD-SNP
	Disease	0	0.502	SNPs&GO

Table 3: *SNPs & GO and PhD Result.*

References

1. Al Dawi, Amal F. Maha I. Mustafa, Ahmed H Fahal, Ahmed M EL Hassan, Sahar M. Bakhiet, El Sheikh Magoub, Adil Mergani, Hatim A. Mostafa, Maowia M. Mukhtar(2013). The association of HLA-DRB1 and HLA-DQB1 alleles and theoccurrence of eumycetoma in Sudanese patients. Khartoum Medical Journal Vol. 06, No. 03, pp. 923 – 929
2. Al-Hudar B, Al-Busaidi A, Gupta I, Hassan B, Tamimi Y (2015) HLA–DRB1 Gene is Highly Mutated in Omani Patients Affected with Rheumatoid Arthritis. Rheumatology (Sunnyvale) 5: 170. doi:10.4172/2161-1149.1000170
3. Bava KA, Gromiha MM, Uedaira H, Kitajima K, Sarai A. 2004; ProTherm, version 4.0: thermodynamic database for proteins and mutants. Nucleic AcidsRes. 32: 120-121.
4. Capriotti E, Calabrese R, Fariselli P, Martelli PL, Altman RB, Casadio R (2013). WS-SNPs&GO: a web server for predicting the deleterious effect of human protein variants using functional annotation. BMC Genomics. Suppl 3:S6.
5. Capriotti E, Fariselli P, Calabrese R, Casadio R. 2005; Predicting protein stability changes from sequences using support vector machines. Bioinformatics. 21: 54-58.
6. Choi Y .(2012). A Fast Computation of Pairwise Sequence Alignment Scores Between a Protein and a Set of Single-Locus Variants of Another Protein. In Proceedings of the ACM Conference on Bioinformatics, Computational Biology and Biomedicine (BCB '12). ACM, New York, NY, USA, 414-417.
7. Choi Y, Chan AP (2015). PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. Bioinformatics 31(16): 2745-2747.
8. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP (2012) Predicting the Functional Effect of Amino Acid Substitutions and Indels. PLoS ONE7(10): e46688.
9. FadlallaElshaikh AAE, Elmahdi Ahmed MT, DafAlla TIM, (2016); MogammedElbasheer AS, Ahmed AA, Computational Analysis of Single Nucleotide Polymorphism (SNPs) In Human MYC Gene. J Bioinform, Genomics, Proteomics. 1(3): 1011-1020.
10. González-Pérez A, López-Bigas N. 2011; Improving the assessment of the outcome of nonsynonymous SNPs with a consensus deleteriousness score, Condel. Am J Hum Genet. 88: 440-449.

11. Kamatani N., Sekine A., Kitamoto T., Iida A., Saito S., KogameA. 2004. Large-scale single-nucleotide polymorphism (SNP) and haplotype analyses, using dense SNP Maps, of 199 drug-related genes in 752 subjects: the analysis of the association between uncommon SNPs within haplotype blocks and the haplotypes constructed with haplotype-tagging SNPs. The American Journal of Human Genetics,. 75(2): p. 190–203.
12. Kerlan S,Candon B,Combe R, Vincent J. Clot V, Pinet J, Eliaou F.(2001). HLA-DRB1 gene transcripts in rheumatoid arthritis. Clinical&amp.Experimental Immunology.124(1) 142-149.
13. Krawczak M., Ball E. V., Fenton I., Stenson P. D., Abeysinghe S., Thomas N., (2000), Human gene mutation database—a biomedical information and research resource. Human mutation,. 15(1): p. 45–51.
14. Lee JE, Choi JH, Lee JH, Lee MG. 2005Gene SNPs and mutations in clinical genetic testing: haplotype-based testing and analysis. Mutat Res.; 573: 195-204.
15. Li, J, Dexter T. Duncan, and Bing Zhang(2010). CanProVar: A Human Cancer Proteome Variation Database. Hum Mutat. 31(3):219-28
16. Nachman M.W., 2001.Single nucleotide polymorphisms and recombination rate in humans. TRENDS in Genetics, 17(9): p. 481–485.
17. Ng PC, Henikoff S. (2003).SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res. 2003; 31: 3812-3814.
18. Parkes, M., Cortes, A., Van Heel, D. A. & Brown, M. A. 2013.Genetic insights into common pathways and complex relationships among immune-mediated diseases.Nature Reviews Genetics, 14, 661.
19. Poland, G., Ovsyannikova, I., Jacobson, R. & Smith, D. I. 2007. Heterogeneity in vaccine immune response: the role of immunogenetics and the emerging field of vaccinomics. Clinical Pharmacology & Therapeutics, 82, 653-664.
20. Prokunina L. and Alarcón-Riquelme M.E., (2004) Regulatory SNPs in complex diseases: their identification and functional validation. Expert reviews in molecular medicine,. 6(10): p. 1–15.
21. Rajasekaran R, Doss GP, Sudandiradoss C, Ramanathan K, Rituraj P, Sethumadhavan R, (2008). Computational and structural investigation of deleterious functional SNPs in breast cancer BRCA2 gene. Sheng Wu Gong Cheng XueBao. 24, 851–856.
22. Rajasekaran R, Doss GP, Sudandiradoss C, Ramanathan K, Rituraj P, Sethumadhavan R, 2008; Computational and structural investigation of Stahl, E. A., Raychaudhuri, S., Remmers, E. F., Xie, G., Eyre, S., Thomson, B. P., Li, Y., Kurreeman, F. A., Zhernakova, A. & Hinks, A. 2010. Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. Nature genetics, 42,508-520
23. Stenson P. D., Mort M., Ball E. V., Howells K., Phillips A. D., Thomas N. S., (2008). The human gene mutation database: update. Genome Med, 1(1): p. 13.
24. Venselaar H, TeBeek TA, Kuipers RK, Hekkelman ML, Vriend G. 2010; Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces.BMC Bioinformatics. 11: 548-550.
25. Venselaar H, TeBeek TA, Kuipers RK, Hekkelman ML, Vriend G. 2010; Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. J BMC bioinformatics .11: 548.
26. Warde-Farley D, Donaldson SL,Comes O, Zuberi K, Badrawi R, Chao P, (2010). The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. J. Nucleic Acids Res. Vol.38 Pp 214-220.
27. Weyand, C. M., Hicok, K. C., Conn, D. L. &Goronzy, J. J. (1992). The influence of HLA-DRB1 genes on disease severity in rheumatoid arthritis.Annals of internal medicine, 117, 801-806.