

In Silico Genetics: Identification of pathogenic nsSNPs in human *STAT3* gene associated with Job's syndrome

Mujahed I. Mustafa^{1,2*}, Abdelrahman H. Abdelmoneim², Nafisa M. Elfadol², Thwayba A. Mahmoud², Mohamed A. Hassan²

1-Department of Biochemistry, University of Bahri, Sudan

2-Department of Biotechnology, Africa city of Technology, Sudan

*Corresponding author: Mujahed I. Mustafa: email: mujahedimustafa@gmail.com

Abstract:

Background: Autosomal dominant hyper-IgE syndrome (AD-HIES) or Job's syndrome is a rare immunodeficiency disease that classically presents in early childhood, characterized by eczematoid dermatitis, characteristic facies, pneumatoceles, hyperextensibility of joints, multiple bone fractures, scoliosis, atopic dermatitis and elevated levels of serum IgE (>2000 IU/ml). The term Autosomal dominant hyper-IgE syndrome has primarily been associated with mutations in *STAT3* gene, Located in human chromosome 17q21. **Methods:** The human *STAT3* gene was investigated in dbSNP/NCBI, 962 SNPs were Homo sapiens; of which 255 were missense SNPs. This selected for in silico analysis by multiple in silico tools to investigate the effect of SNPs on *STAT3* protein's structure and function. **Result:** Eleven novel mutations out of 255 nsSNPs that are found to be deleterious effect on the *STAT3* structure and function. **Conclusion:** A total of eleven novel nsSNPs were predicted to be responsible for the structural and functional modifications of *STAT3* protein. The newly recognized genetic cause of the hyper-IgE syndrome affects complex, compartmentalized somatic and immune regulation. This study will open new doors to facilitate the development of novel diagnostic markers for associated diseases.

Keywords: Autosomal dominant hyper-IgE syndrome (AD-HIES), Job's syndrome, *STAT3* gene, in silico analysis, diagnostic markers.

Introduction:

Autosomal dominant hyper-IgE syndrome (AD-HIES) or Job's syndrome is a rare immunodeficiency disease that classically presents in early childhood, characterized by eczematoid dermatitis, characteristic facies, pneumatocoles, hyperextensibility of joints, multiple bone fractures, scoliosis, atopic dermatitis and elevated levels of serum IgE (>2000 IU/ml).[1-11] Job's syndrome was first described by Davis et al. in 1966[12] Since then, patients from different countries are being increasingly recognized. [2, 6, 13-18]

The treatment for Hyper-IgE syndromes is mainly to control infection, skin care with Cyclosporine A, if necessary, should be done as early as possible hematopoietic stem cell transplantation.[3, 10, 19] The term Autosomal dominant hyper-IgE syndrome has primarily been associated with mutations in *STAT3* gene. [5, 8] Located in human chromosome 17q21, the other genes that have been implicated in HIES include *TYK2*[20, 21] and *DOCK8*[22-24] *STAT3* plays a vital role in signal transduction induced by many cytokines (IL-6, IL-10, IL-17, IL-21, and IL-22).[25, 26] Mutations lead to disruption of *STAT3*-dependent pathways, which are crucial for signaling of many cytokines, including IL-6 and IL-10.[27]

STAT3 also plays important roles in multiple aspects of cancer aggressiveness including migration, invasion, survival, self-renewal, angiogenesis, and tumor cell immune evasion by regulating the expression of multiple downstream target genes.[28, 29] Some studies show that loss-of-function *STAT3* alleles were shown to be dominant-negative, led to a more detailed description of the immunologic phenotype of this primary immunodeficiency, with the demonstration of a deficiency of IL-17A- and IL-22-producing T cells.[5, 30, 31] While other study reveals a pattern of *STAT3*-associated gene expression specific to basal-like breast cancers in human tumors.[32] [28, 29] therefore, the treatment for Hyper-IgE syndromes is crucial to prevent a serious complication in AD-HIES patients as cystic lung disease[33] However, specific immunological abnormalities that can explain the unique susceptibility to particular infections seen in AD-HIES have not yet been clarified.[34, 35]

Single-nucleotide polymorphism (SNPs) refers to single base differences in DNA among individuals. One of the interests in association studies is the association between SNPs and disease development.[36] The aim of this study, to identify functional SNPs within dbSNP located in coding regions of *STAT3* gene. This is the first study which covers an extensive in silico analysis of nsSNPs *STAT3* protein. This study will open new doors to facilitate the development of novel diagnostic markers for associated diseases.[37-39]

2. Materials and Methods:

Data mining:

The data on human *STAT3* gene was collected from National Center for Biological Information (NCBI) web site.[40] The SNP information (protein accession number and SNP ID) of the *PRSS1* gene was retrieved from the NCBI dbSNP (<http://www.ncbi.nlm.nih.gov/snp/>) and the protein sequence was collected from NCBI protein database (NCBI Reference Sequence: XP_024306664.1) (<https://www.ncbi.nlm.nih.gov/protein/>).

SIFT:

We used SIFT to observe the effect of A.A. substitution on protein function. SIFT predicts damaging SNPs on the basis of the degree of conserved amino A.A. residues in aligned sequences to the closely related sequences, gathered through PSI-BLAST.[41] It's available at (<http://sift.jcvi.org/>).

PolyPhen:

PolyPhen (version 2) stands for polymorphism phenotyping version 2. We used PolyPhen to study probable impacts of A.A. substitution on structural and functional properties of the protein by considering physical and comparative approaches.[42] It's available at (<http://genetics.bwh.harvard.edu/pp2>).

Provean:

Provean is an online tool that predicts whether an amino acid substitution has an impact on the biological function of a protein grounded on the alignment-based score. The score measures the change in sequence similarity of a query sequence to a protein sequence homolog between without and with an amino acid variation of the query sequence. If the PROVEAN score ≤ -2.5 , the protein variant is predicted to have a “deleterious” effect, while if the PROVEAN score is > -2.5 , the variant is predicted to have a “neutral” effect.[43] It is available at (<https://roslab.org/services/snap2web/>).

SNAP2:

Functional effects of mutations are predicted with SNAP2 (29). SNAP2 is a trained classifier that is based on a machine learning device called "neural network". It distinguishes between effect and neutral variants/non-synonymous SNPs by taking a variety of sequence and variant features into account. The most important input signal for the prediction is the evolutionary information taken from an automatically generated multiple sequence alignment. Also structural features such as predicted secondary structure and solvent accessibility are

considered. If available also annotation (i.e. known functional residues, pattern, regions) of the sequence or close homologs are pulled in. In a cross-validation over 100,000 experimentally annotated variants, SNAP2 reached sustained two-state accuracy (effect/neutral) of 82% (at anAUC of 0.9). In our hands this constitutes an important and significant improvement over other methods.[44] It is available at (<https://roslab.org/services/snap2web/>).

SNPs&GO:

Single Nucleotide Polymorphism Database (SNPs) & Gene Ontology (GO) is a support vector machine (SVM) based on the method to accurately predict the disease related mutations from protein sequence. FASTA sequence of whole protein is considered to be an input option and output will be the 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 on the parent protein function.[45] it's available at (<https://roslab.org/services/snap2web/>).

PHD-SNP:

An online Support Vector Machine (SVM) based classifier, is optimized to predict if a given single point protein mutation can be classified as disease-related or as a neutral polymorphism. It's available at: (<http://http://snps.biofold.org/phd-snp/phdsnp.html>)

I-Mutant 3.0:

Change in protein stability disturbs both protein structure and protein function. I-Mutant is a suite of support vector machine, based predictors integrated in a unique web server. It offers the opportunity to predict the protein stability changes upon single-site mutations. From the protein structure or sequence. The FASTA sequence of protein retrieved from UniProt is used as an input to predict the mutational effect on protein and stability RI value (reliability index) computed.[46] It's available at (<http://gpcr2.biocomp.unibo.it/cgi/predictors/I-Mutant3.0/I-Mutant3.0.cgi>).

MUpro:

MUpro is a support vector machine-based tool for the prediction of protein stability changes upon nonsynonymous SNPs. The value of the energy change is predicted, and a confidence score between -1 and 1 for measuring the confidence of the prediction is calculated. A score <0 means the variant decreases the protein stability; conversely, a score >0 means the variant increases the protein stability.[47] It's available at (<http://mupro.proteomics.ics.uci.edu/>).

GeneMANIA:

We submitted genes and selected from a list of data sets that they wish to query. GeneMANIA approach to know protein function prediction integrate multiple genomics and proteomics data sources to make inferences about the function of unknown proteins.[48] It is available at (<http://www.genemania.org/>)

Structural Analysis:

Detection of nsSNPs Location in Protein Structure:

Mutation3D is a functional prediction and visualization tool for studying the spatial arrangement of amino acid substitutions on protein models and structures. Mutation3D is able to separate functional from nonfunctional mutations by analyzing a combination of 8,869 known inherited disease mutations and 2,004 SNPs overlaid together upon the same sets of crystal structures and homology models. Further, it presents a systematic analysis of wholegenome and whole-exome cancer datasets to demonstrate that mutation3D identifies many known cancer genes as well as previously underexplored target genes.[49] It is available at (<http://mutation3d.org>).

Developing 3D structure of mutant *STAT3* gene:

The 3D structure of human Signal transducer and activator of transcription 3 (*STAT3*) protein is not available in the Protein Data Bank. Hence, we used RaptorX to generate a 3D structural model for wild-type *STAT3*. RaptorX is a web server predicting structure property of a protein sequence without using any templates..[50] It is available at (<http://raptorx.uchicago.edu/>).

Modeling Amino Acid Substitution:

UCSF Chimera is a highly extensible program for interactive visualization and analysis of molecular structures and related data, including density maps, supramolecular assemblies, sequence alignments, docking results, conformational analysis[51] Chimera (version 1.8) available at (<http://www.cgl.ucsf.edu/chimera/>).



Figure 1: Diagrammatic representation for *STAT3* gene in coding region in silico work flow.

Results:

Table1: Damaging or Deleterious nsSNPs associated variations predicted by various softwares:

dbSNP rs#	Sub	SIFT Prediction	Score	Polyphen Prediction	Score	PROVEAN Prediction	Score	SNAP2 Prediction	Score
rs869312892	G716M	DAMAGING	0	possibly damaging	0.518	Deleterious	-7.28	effect	56
rs1173682056	P701R	DAMAGING	0	probably damaging	0.98	Deleterious	-8.119	effect	28
rs1425974175	I697G	DAMAGING	0	probably damaging	1	Deleterious	-5.12	effect	38
rs1064796922	E648K	DAMAGING	0	probably damaging	1	Deleterious	-2.741	effect	20
rs1466712110	F638K	DAMAGING	0	possibly damaging	0.665	Deleterious	-8.666	effect	19
rs193922720	W594K	DAMAGING	0	possibly damaging	0.715	Deleterious	-12.625	effect	74
rs1064796762	C582K	DAMAGING	0.02	possibly damaging	0.714	Deleterious	-7.764	effect	25
rs781353111	W542Y	DAMAGING	0	probably damaging	0.996	Deleterious	-8.32	effect	19
rs776053336	F524K	DAMAGING	0	probably damaging	0.995	Deleterious	-7.846	effect	29
rs1281575473	L510R	DAMAGING	0	possibly damaging	0.818	Deleterious	-4.957	effect	38
rs145786768	A507F	DAMAGING	0	possibly damaging	0.649	Deleterious	-5.355	effect	18
rs763944667	P503H	DAMAGING	0.01	possibly damaging	0.855	Deleterious	-5.908	effect	6
rs146620441	N498V	DAMAGING	0	probably damaging	0.982	Deleterious	-7.597	effect	34
rs1442343135	E407A	DAMAGING	0.17	probably damaging	0.979	Deleterious	-4.395	effect	14
rs1401418993	K392E	DAMAGING	0.01	probably damaging	0.997	Deleterious	-3.426	effect	29
rs113994136	R382Q	DAMAGING	0	possibly damaging	0.939	Deleterious	-3.355	effect	19
-	R382L	DAMAGING	0	probably damaging	0.975	Deleterious	-5.941	effect	41
rs113994135	R382W	DAMAGING	0	probably damaging	0.999	Deleterious	-6.631	effect	68
rs755524497	R350M	DAMAGING	0.23	possibly damaging	0.871	Deleterious	-5.179	effect	47
rs776115471	R335Q	DAMAGING	0	probably damaging	0.999	Deleterious	-2.797	effect	3
rs1427784630	P330R	DAMAGING	0	probably damaging	0.968	Deleterious	-6.368	effect	12
rs757347742	Q248P	DAMAGING	0.18	probably damaging	0.979	Deleterious	-4.461	effect	2
rs964892419	R114C	DAMAGING	0.05	possibly damaging	0.508	Deleterious	-4.577	effect	14
rs780604324	C108Y	DAMAGING	0.05	possibly damaging	0.92	Deleterious	-3.748	effect	46

Table2: Disease effect nsSNPs associated variations predicted by SNPs&GO and PHD-SNP softwares:

dbSNP rs#	Sub	SNPs&GO			PHD-SNP		
		Prediction	RI	Probability	Prediction	RI	Probability
rs1173682056	P701R	Disease	6	0.812	Disease	7	0.832
rs1425974175	I697G	Disease	4	0.682	Disease	7	0.863
rs193922720	W594K	Disease	8	0.923	Disease	9	0.964
rs1064796762	C582K	Disease	6	0.801	Disease	7	0.846
rs781353111	W542Y	Disease	4	0.687	Disease	4	0.686
rs776053336	F524K	Disease	7	0.841	Disease	8	0.875
rs1281575473	L510R	Disease	4	0.678	Disease	7	0.85
rs145786768	A507F	Disease	6	0.797	Disease	8	0.901
rs763944667	P503H	Disease	3	0.665	Disease	4	0.718
rs146620441	N498V	Disease	6	0.777	Disease	5	0.768
rs780604324	C108Y	Disease	0	0.522	Disease	6	0.824

***RI: Reliability Index**

Table3: Stability analysis predicted by I-Mutant version 3.0 and MUPro (also Show the 11 novel mutations):

dbSNP rs#	Amino Acid change	SVM2		DDG		Mupro	Mupro Score
		Prediction	RI	Value	Prediction		
rs1173682056	P701R	Decrease	8	-1.17	Decrease		-1.286101
rs1425974175	I697G	Decrease	9	-3.25	Decrease		-2.6272515
rs193922720	W594K	Decrease	10	-1.78	Decrease		-1.4377345
rs1064796762	C582K	Decrease	6	-0.4	Decrease		-1.5723999
rs781353111	W542Y	Decrease	8	-0.9	Decrease		-0.80874247
rs776053336	F524K	Decrease	7	-1.49	Decrease		-2.4196993
rs1281575473	L510R	Decrease	6	-1.37	Decrease		-1.6096433
rs145786768	A507F	Decrease	7	-0.32	Decrease		-0.47460484
rs763944667	P503H	Decrease	8	-1.52	Decrease		-0.73333962
rs146620441	N498V	Increase	3	0.2	Decrease		-0.12137383
rs780604324	C108Y	Decrease	2	-0.22	Decrease		-0.92859819

***RI: Reliability Index**

Table 4: The *STAT3* gene functions and its appearance in network and genome:

Function	FDR	Genes in network	Genes in genome
cellular response to growth hormone stimulus	0.011108	3	31
response to peptide	0.011108	5	260
cellular response to peptide	0.011108	5	247
tumor necrosis factor-mediated signaling pathway	0.011108	3	35
growth hormone receptor signaling pathway	0.011108	3	31
response to growth hormone	0.011108	3	32
response to estradiol	0.011108	3	24
JAK-STAT cascade	0.011108	4	91
response to peptide hormone	0.011108	5	255
JAK-STAT cascade involved in growth hormone signaling pathway	0.011108	3	24
response to organic cyclic compound	0.011108	5	246
cellular response to peptide hormone stimulus	0.011108	5	244
regulation of peptidyl-tyrosine phosphorylation	0.019403	4	138
response to estrogen	0.021834	3	47
positive regulation of response to DNA damage stimulus	0.021834	3	47
regulation of protein complex assembly	0.023675	4	153
response to alcohol	0.063704	3	75
peptidyl-tyrosine phosphorylation	0.063704	4	210
cellular response to epidermal growth factor stimulus	0.063704	2	11
peptidyl-tyrosine modification	0.063704	4	210
cellular response to tumor necrosis factor	0.063704	3	74
positive regulation of ERK1 and ERK2 cascade	0.077484	3	83
temperature homeostasis	0.077484	2	14
astrocyte differentiation	0.077484	2	14
response to steroid hormone	0.077484	3	82
regulation of homotypic cell-cell adhesion	0.079432	2	15
positive regulation of protein complex assembly	0.079432	3	90
response to tumor necrosis factor	0.079432	3	87
response to epidermal growth factor	0.079432	2	15
response to lipid	0.084124	4	249

Table (5) The gene co-expressed, share domain and Interaction with *STAT3* gene network:

Gene 1	Gene 2	Weight	Network group
PRKCD	STAT3	0.015713	Co-expression
PLA2G4A	REG1A	0.014944	Co-expression
STAT5B	STAT3	0.025263	Co-expression
PRKCD	PTK2B	0.010729	Co-expression
TRH	TNFSF11	0.017912	Co-expression
PPP1CA	PRKCD	0.010846	Co-expression
PRKCD	STAT3	0.009714	Co-expression
CRP	STAT3	0.004409	Co-expression
PTK2B	STAT5B	0.01216	Co-expression
SPIDR	NFKBIZ	0.014793	Co-expression
PPP1CA	PRKCD	0.013595	Co-expression
EGFR	STAT3	0.014231	Co-expression
PRKCD	PTK2B	0.011206	Co-expression
STAT5B	STAT3	0.012644	Co-expression
IDE	STAT1	0.011536	Co-expression
NFKBIZ	STAT3	0.019434	Co-localization
TNFSF11	STAT1	0.011453	Co-localization
STAT5B	STAT3	0.17252	Co-localization
EGFR	STAT3	0.069194	Co-localization
STAT1	STAT3	0.073874	Co-localization
NFKBIZ	STAT3	1	Genetic Interactions
PPP1CA	STAT3	0.707107	Genetic Interactions
STAT5B	STAT3	0.003239	Pathway
EGFR	STAT3	0.001922	Pathway
EGFR	STAT5B	0.002917	Pathway
PIAS3	STAT3	0.012249	Pathway
PIAS3	STAT5B	0.018584	Pathway
STAT1	STAT3	0.002703	Pathway
STAT1	STAT5B	0.004101	Pathway
STAT1	EGFR	0.002433	Pathway
PTK2B	STAT3	0.003	Pathway
PTK2B	STAT5B	0.004552	Pathway
PRKCD	STAT3	0.004128	Pathway
PRKCD	STAT1	0.005226	Pathway
PRKCD	PTK2B	0.005802	Pathway
SPIDR	STAT3	0.106879	Pathway

MIA2	STAT3	0.099517	Pathway
TRH	STAT3	0.101665	Pathway
REG1A	STAT3	0.098408	Pathway
REG1A	STAT1	0.124574	Pathway
PLA2G4A	PRKCD	0.024516	Pathway
PIAS3	STAT3	0.043991	Pathway
STAT1	STAT3	0.014229	Pathway
PTK2B	STAT5B	0.362976	Pathway
IL17F	STAT3	0.285891	Pathway
CRP	STAT3	0.285891	Pathway
TNFSF11	STAT3	0.285891	Pathway
STAT5B	STAT3	0.211607	Pathway
STAT1	STAT5B	0.201164	Pathway
PTK2B	STAT3	0.56656	Pathway
HES5	STAT3	0.29654	Pathway
EGFR	STAT3	0.090536	Physical Interactions
STAT1	EGFR	0.090536	Physical Interactions
IDE	STAT1	0.045432	Physical Interactions
IDE	STAT3	0.512312	Physical Interactions
STAT1	STAT3	0.382314	Physical Interactions
KIF27	STAT3	0.519377	Physical Interactions
LRRFIP2	STAT3	0.519377	Physical Interactions
PLA2G4A	STAT3	0.519377	Physical Interactions
EGFR	STAT3	0.090909	Physical Interactions
STAT1	EGFR	0.090909	Physical Interactions
NFKBIZ	STAT3	0.114992	Physical Interactions
EGFR	STAT3	0.001486	Physical Interactions
EGFR	STAT5B	0.005754	Physical Interactions
PIAS3	STAT3	0.008193	Physical Interactions
STAT1	STAT3	0.00544	Physical Interactions
STAT1	EGFR	0.002266	Physical Interactions
PTK2B	EGFR	0.002511	Physical Interactions
HES5	STAT3	0.111646	Physical Interactions
PRKCD	STAT3	0.005481	Physical Interactions
PRKCD	STAT1	0.008357	Physical Interactions
PRKCD	PTK2B	0.009259	Physical Interactions
PPP1CA	PIAS3	0.003493	Physical Interactions
EGFR	STAT3	0.086154	Physical Interactions
PRKCD	STAT3	0.130914	Physical Interactions

PIAS3	STAT3	1	Physical Interactions
STAT5B	STAT3	0.026225	Physical Interactions
EGFR	STAT3	0.002278	Physical Interactions
EGFR	STAT5B	0.007592	Physical Interactions
STAT1	STAT3	0.012243	Physical Interactions
STAT1	EGFR	0.003544	Physical Interactions
PPP1CA	STAT3	0.002452	Physical Interactions
PPP1CA	PIAS3	0.007972	Physical Interactions
STAT5B	STAT3	0.010216	Physical Interactions
EGFR	STAT3	0.003073	Physical Interactions
EGFR	STAT5B	0.00669	Physical Interactions
PIAS3	STAT3	0.014844	Physical Interactions
STAT1	STAT3	0.005947	Physical Interactions
STAT1	STAT5B	0.012944	Physical Interactions
STAT1	EGFR	0.003894	Physical Interactions
PTK2B	STAT3	0.006076	Physical Interactions
PTK2B	EGFR	0.003979	Physical Interactions
HES5	STAT3	0.125435	Physical Interactions
PIAS3	STAT3	0.027462	Predicted
STAT1	STAT3	0.006424	Predicted
PTK2B	EGFR	0.013423	Predicted
HES5	STAT3	0.076442	Predicted
PRKCD	STAT3	0.505991	Predicted
PRKCD	STAT3	0.425815	Predicted
STAT1	STAT3	0.607781	Predicted
STAT5B	STAT3	0.069454	Shared protein domains
STAT1	STAT3	0.069464	Shared protein domains
STAT1	STAT5B	0.065597	Shared protein domains
STAT5B	STAT3	0.076689	Shared protein domains
STAT1	STAT3	0.077214	Shared protein domains
STAT1	STAT5B	0.076333	Shared protein domains
PTK2B	EGFR	0.007985	Shared protein domains

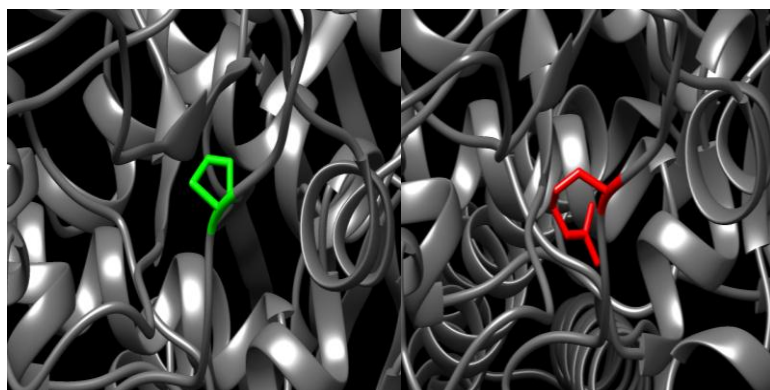


Figure 2: (P701R): change in the amino acid Proline into Arginine at position 701.

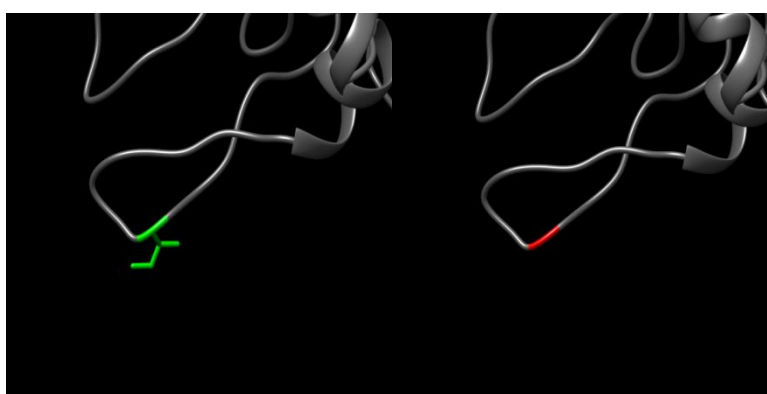


Figure 3: (I697G): change in the amino acid Proline into Arginine at position 701.

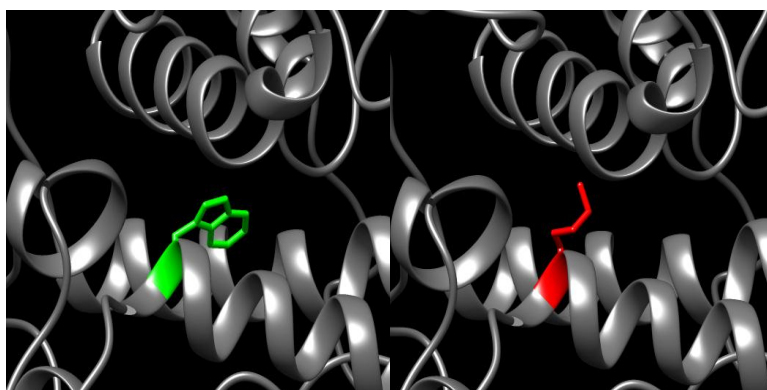


Figure 4: (W594K): change in the amino acid Tryptophan into Lysine at position 594.

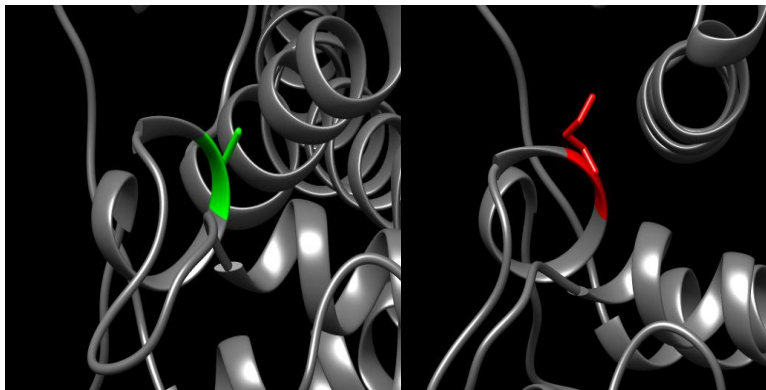


Figure 5: (C582K): change in the amino acid Cysteine into Lysine at position 594.

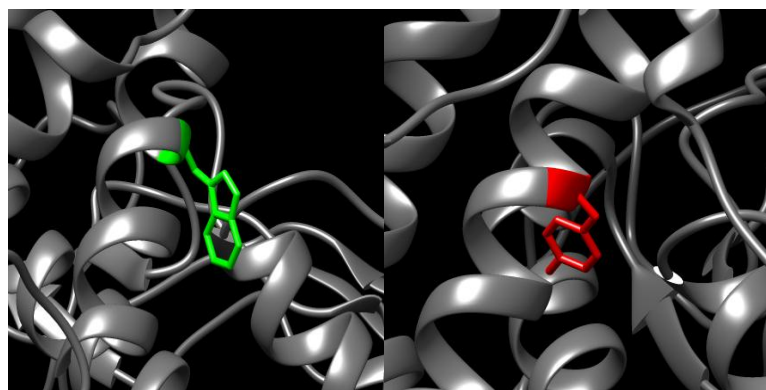


Figure 6: (W542Y): change in the amino acid Tryptophan into Tyrosine at position 542.

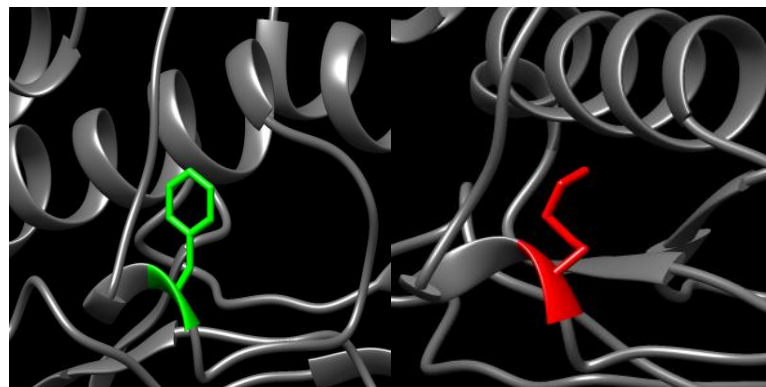


Figure 7: (F524K): change in the amino acid Phenylalanine into Lysine at position 542.

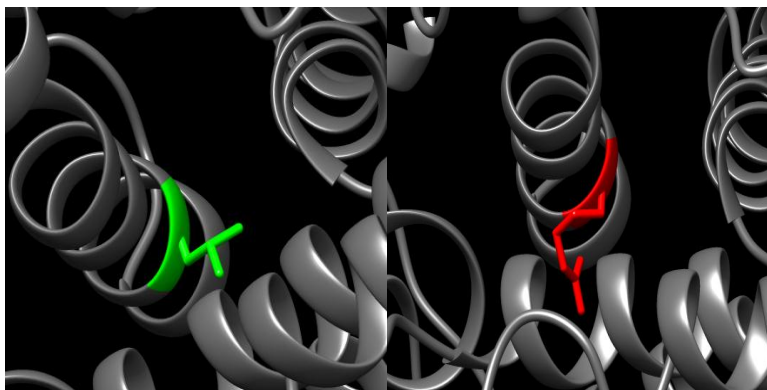


Figure 8: (L510R): change in the amino acid Leucine into Arginine at position 542.

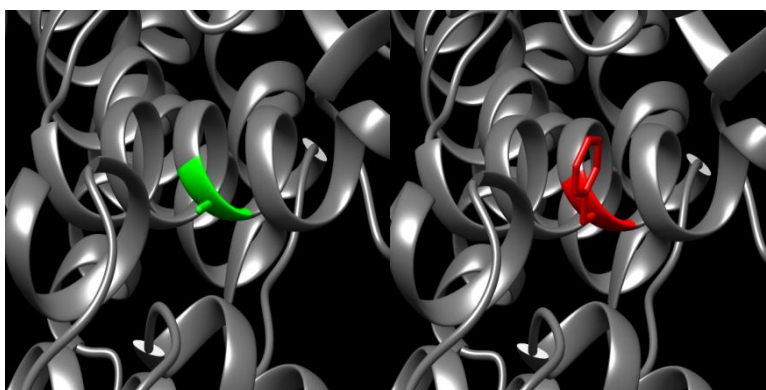


Figure 9: (A507F): change in the amino acid Alanine into Phenylalanine at position 507.

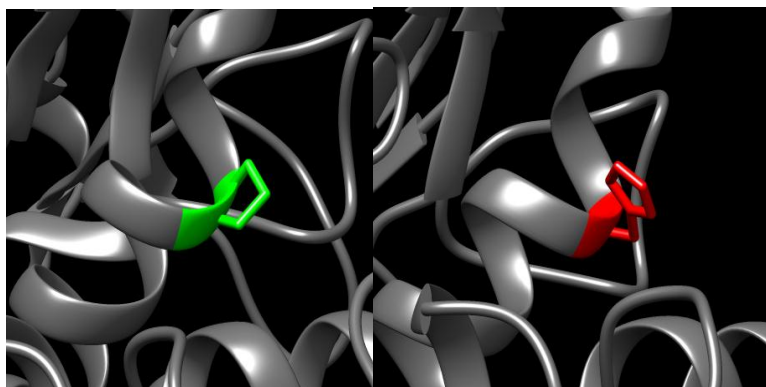


Figure 10: (P503H): change in the amino acid Proline into Histidine at position 542.

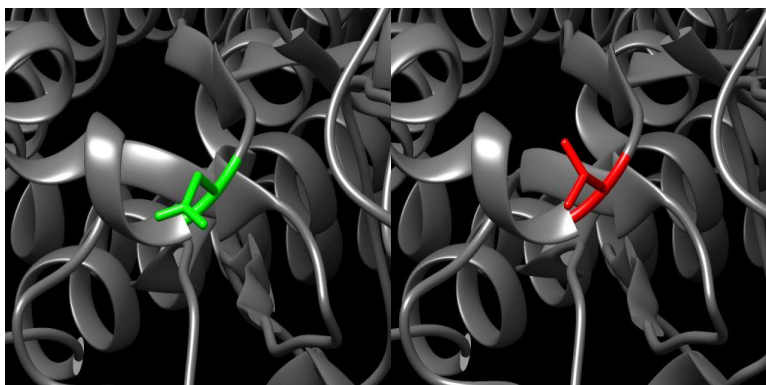


Figure 11: (N498V): change in the amino acid Asparagine into Valine at position 498.

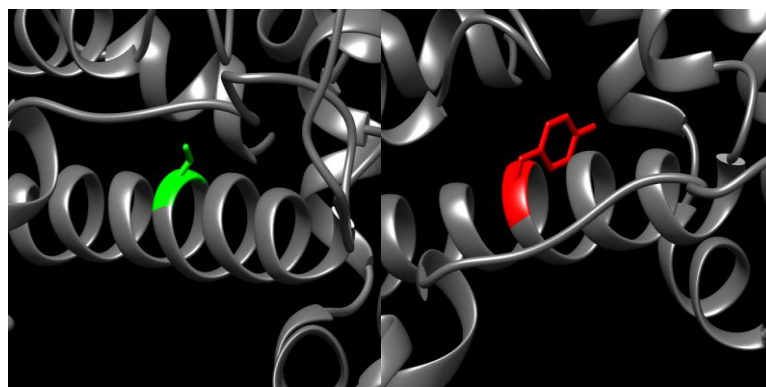


Figure 12: (C108Y): change in the amino acid Cysteine into Tyrosine at position 542.

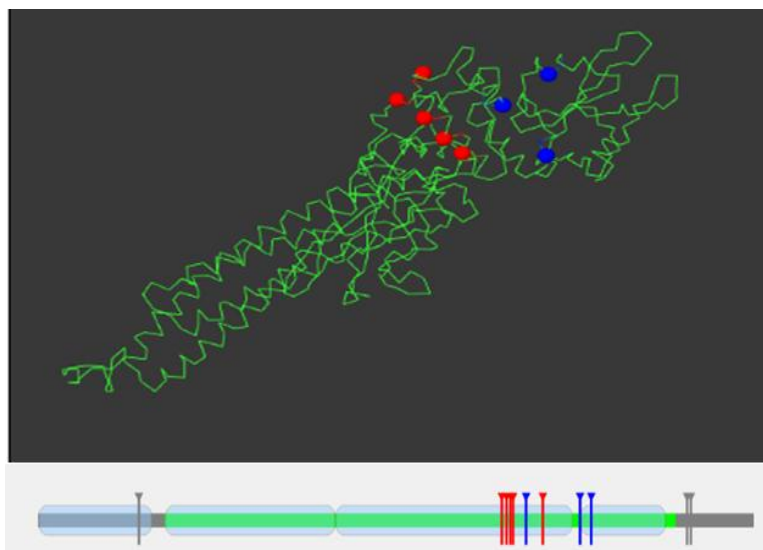


Figure 13: Structural models for wild type *STAT3*, illustrated by Mutation3D.

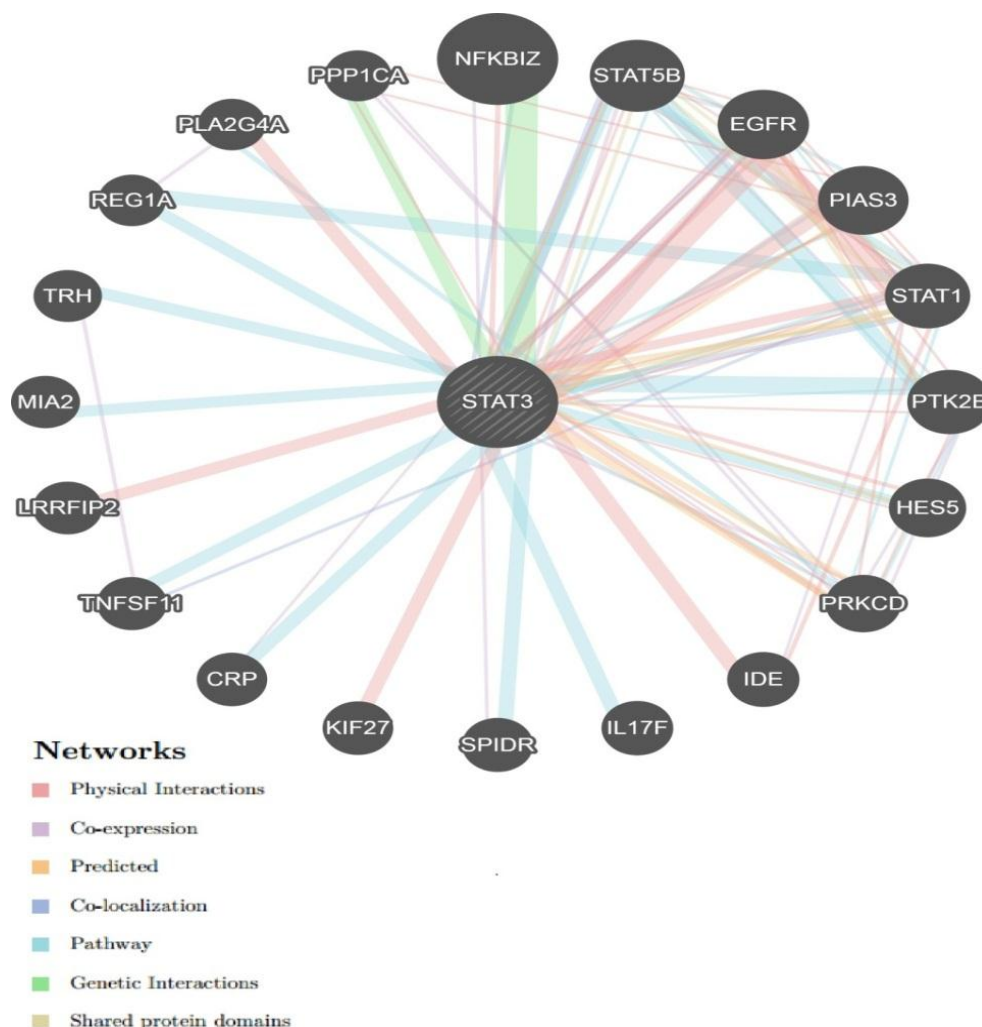


Figure 14: Interaction between *STAT3* and its related genes.

Discussion:

Eleven novel SNPs were found to be deleterious in this study. The in silico methods used to analyze the SNPs were based on different aspects and parameters describing the pathogenicity and provide clues on the molecular level about the effect of mutation on the structure and function of the final protein. (figure 1) 958 SNPs were retrieved from NCBI database, of which 251 were found to be missense. We first analyzed the deleterious effect of different candidate SNPs on the function of the final protein through 4 softwares (SIFT, POLYHEN2, PROVEAN and SNPA2) and we found that the shared deleterious SNPs between them to be 24 SNPs. (table 1) we further analyzed it through (SNPS&GO and PHD-SNP) and we found the double positive result to be 11 SNPs. (table 2)

Stability analysis which predicted by I-mutant 3.0 and MUPro servers revealed that, all SNPs had decreased protein stability by both servers except N498V (rs146620441) had predicted by I-mutant 3.0, increased protein stability. (Table 3) for further study the structural changes we used chimera software (figures 2 to12), In the data extracted from the NCBI database the substitutions (R382W) (rs113994135), V637M (rs113994139) were found to be pathogenic which correspond to what is mentioned in other paper .[53] The substitution R382W (rs113994135) was also found to be deleterious by four soft wares in this study which correspond to previous studies.[53, 54]

GeneMANIA revealed that *STAT3* has many vital functions: blood microparticle, cobalamin metabolic process, extracellular matrix organization, extracellular structure organization, serine hydrolase activity, serine-type endopeptidase activity, serine-type peptidase activity. The genes co-expressed with, share similar protein domain, or participate to achieve similar function were illustrated by GeneMANIA and shown in figure (14) Tables (4 & 5). Additionally, we performed analysis by Mutation3D, our result show that: (N498V, P503H, A507F, L510R and F524K) located in the domain. (Figure 13)

The *STAT3* gene is associated with other disease like inflammatory bowel disease along other genes like *IL23R* and *JAK2* genes.[55] there are also a relations with Autoimmune lymphoproliferative syndrome (ALPS) mainly in p.R278H, p.M394T SNPs,[56] *STAT3* also becomes persistently activated in a high percentage of malignancies (e.g. breast, prostate, ovarian, and colon cancers), thus contributing to malignant transformation and progression which makes *STAT3* an attractive therapeutic target for cancers.[57]there is also an evidence that studying gene expression of *STAT1* , *STAT2* and *STAT3* gene can be useful for evaluating the efficacy of IFN treatment of the MS patients .[58]

Our study is the first in silico analysis which based on functional and structural analysis while all other previous studies based on in vivo analysis, molecular analysis and genome sequencing [53, 59-61] This study revealed eleven Novel Pathological mutations have a potential functional impact and may thus be used as diagnostic markers for Job's syndrome. Finally some appreciations of wet lab techniques are suggested to support our findings.

Conclusion:

A total of eleven novel nsSNPs were predicted to be responsible for the structural and functional modifications of *STAT3* protein .The newly recognized genetic cause of the Autosomal dominant hyper-IgE syndrome affects complex, compartmentalized somatic and immune regulation. This study will opens new doors to facilitate the development of novel diagnostic markers for associated diseases.

Conflict of interest:

The authors have declared that no competing interest exists.

Acknowledgment:

The authors wish to acknowledgment the enthusiastic cooperation of Africa City of Technology

- Sudan.

References:

- [1] K. J. Sowerwine, S. M. Holland, and A. F. Freeman, "Hyper-IgE syndrome update," (in eng), *Ann N Y Acad Sci*, vol. 1250, pp. 25-32, Feb 2012.
- [2] E. Heropolitanska-Pliszka, B. Pietrucha, B. Mikoluc, and E. Bernatowska, "[Hyper-IgE syndrome with mutation in STAT3 gene - case report and literature review]," (in pol), *Med Wieku Rozwoj*, vol. 13, no. 1, pp. 19-25, Jan-Mar 2009. Zespół hiper -lge z mutacją w genie stat3- opis przypadku i przegląd piśmiennictwa.
- [3] Y. Y. He, B. Liu, and X. P. Xiao, "[Hyper-IgE syndromes]," (in chi), *Lin Chung Er Bi Yan Hou Tou Jing Wai Ke Za Zhi*, vol. 31, no. 11, pp. 892-896, Jun 5 2017.
- [4] J. Heimall, A. Freeman, and S. M. Holland, "Pathogenesis of hyper IgE syndrome," (in eng), *Clin Rev Allergy Immunol*, vol. 38, no. 1, pp. 32-8, Feb 2010.
- [5] Y. Minegishi and H. Karasuyama, "Genetic origins of hyper-IgE syndrome," (in eng), *Curr Allergy Asthma Rep*, vol. 8, no. 5, pp. 386-91, Sep 2008.
- [6] S. Al Khatib *et al.*, "Defects along the T(H)17 differentiation pathway underlie genetically distinct forms of the hyper IgE syndrome," (in eng), *J Allergy Clin Immunol*, vol. 124, no. 2, pp. 342-8, 348.e1-5, Aug 2009.
- [7] B. Grimbacher *et al.*, "Hyper-IgE syndrome with recurrent infections--an autosomal dominant multisystem disorder," (in eng), *N Engl J Med*, vol. 340, no. 9, pp. 692-702, Mar 4 1999.
- [8] H. Takada and Y. Minegishi, "[Recent advances in the pathogenesis of hyper IgE syndrome]," (in jpn), *Nihon Rinsho Meneki Gakkai Kaishi*, vol. 33, no. 1, pp. 15-9, 2010.
- [9] H. Takada, A. Nomura, and T. Hara, "[Hyper-IgE syndrome]," (in jpn), *Nihon Rinsho Meneki Gakkai Kaishi*, vol. 27, no. 6, pp. 361-6, Dec 2004.
- [10] M. M. Tamas, C. Baican, and S. Rednic, "Hyperimmunoglobulin E syndrome (Job's syndrome)," (in eng), *Rom J Intern Med*, vol. 46, no. 2, pp. 169-71, 2008.
- [11] G. Kurien and T. C. Uwiera, "Tracheal stenosis in hyper immunoglobulin E syndrome: a novel case report," (in eng), *Int J Pediatr Otorhinolaryngol*, vol. 75, no. 6, pp. 868-71, Jun 2011.
- [12] S. D. Davis, J. Schaller, and R. J. Wedgwood, "Job's Syndrome. Recurrent, "cold", staphylococcal abscesses," (in eng), *Lancet*, vol. 1, no. 7445, pp. 1013-5, May 7 1966.
- [13] H. Jiao *et al.*, "Novel and recurrent STAT3 mutations in hyper-IgE syndrome patients from different ethnic groups," (in eng), *Mol Immunol*, vol. 46, no. 1, pp. 202-6, Nov 2008.
- [14] S. Verma and U. Wollina, "Job's syndrome--a case report," (in eng), *J Eur Acad Dermatol Venereol*, vol. 17, no. 6, pp. 711-4, Nov 2003.
- [15] A. C. de Alwis, R. de Silva, S. Gunawardena, D. C. Weerasuriya, and A. H. Jayaweera, "A case of hyper-IgE syndrome," (in eng), *Ceylon Med J*, vol. 51, no. 4, pp. 149-51, Dec 2006.
- [16] M. Moin *et al.*, "The clinical and laboratory survey of Iranian patients with hyper-IgE syndrome," (in eng), *Scand J Infect Dis*, vol. 38, no. 10, pp. 898-903, 2006.
- [17] A. J. Samdani and A. Tania, "Cutaneous manifestations of hyper-IgE syndrome in twins: a case report from Saudi Arabia," (in eng), *J Pak Med Assoc*, vol. 54, no. 8, pp. 438-9, Aug 2004.
- [18] H. Ouair *et al.*, "[Clinical and immunological profile of 15 Moroccan patients with Hyper IgM syndrome]," (in fre), *Pan Afr Med J*, vol. 26, p. 212, 2017. Le profil clinique et immunologique de 15 patients Marocains atteints de syndrome hyper IgM.

- [19] E. Kitai and B. Wolach, "[A case of Job's syndrome]," (in heb), *Harefuah*, vol. 130, no. 4, pp. 239-41; 295, Feb 15 1996.
- [20] Y. Minegishi *et al.*, "Human tyrosine kinase 2 deficiency reveals its requisite roles in multiple cytokine signals involved in innate and acquired immunity," (in eng), *Immunity*, vol. 25, no. 5, pp. 745-55, Nov 2006.
- [21] C. Woellner *et al.*, "The hyper IgE syndrome and mutations in TYK2," (in eng), *Immunity*, vol. 26, no. 5, p. 535; author reply 536, May 2007.
- [22] Q. Zhang *et al.*, "Combined immunodeficiency associated with DOCK8 mutations," (in eng), *N Engl J Med*, vol. 361, no. 21, pp. 2046-55, Nov 19 2009.
- [23] K. R. Engelhardt *et al.*, "Large deletions and point mutations involving the dedicator of cytokinesis 8 (DOCK8) in the autosomal-recessive form of hyper-IgE syndrome," (in eng), *J Allergy Clin Immunol*, vol. 124, no. 6, pp. 1289-302.e4, Dec 2009.
- [24] C. M. Biggs, S. Keles, and T. A. Chatila, "DOCK8 deficiency: Insights into pathophysiology, clinical features and management," (in eng), *Clin Immunol*, vol. 181, pp. 75-82, Aug 2017.
- [25] A. P. Hutchins, S. Poulain, and D. Miranda-Saavedra, "Genome-wide analysis of STAT3 binding in vivo predicts effectors of the anti-inflammatory response in macrophages," (in eng), *Blood*, vol. 119, no. 13, pp. e110-9, Mar 29 2012.
- [26] P. Shrestha, G. Sabharwal, and G. Ghaffari, "A Novel STAT3 Gene Mutation Related Hyper-IgE Syndrome Misdiagnosed as Hidradenitis Suppurativa," vol. 2018, p. 4860902, 2018.
- [27] M. L. Paulson, A. F. Freeman, and S. M. Holland, "Hyper IgE syndrome: an update on clinical aspects and the role of signal transducer and activator of transcription 3," (in eng), *Curr Opin Allergy Clin Immunol*, vol. 8, no. 6, pp. 527-33, Dec 2008.
- [28] W. Huang, Z. Dong, F. Wang, H. Peng, J. Y. Liu, and J. T. Zhang, "A small molecule compound targeting STAT3 DNA-binding domain inhibits cancer cell proliferation, migration, and invasion," (in eng), *ACS Chem Biol*, vol. 9, no. 5, pp. 1188-96, May 16 2014.
- [29] J. Turkson, "STAT proteins as novel targets for cancer drug discovery," (in eng), *Expert Opin Ther Targets*, vol. 8, no. 5, pp. 409-22, Oct 2004.
- [30] M. O. Chandesris *et al.*, "Autosomal dominant STAT3 deficiency and hyper-IgE syndrome: molecular, cellular, and clinical features from a French national survey," (in eng), *Medicine (Baltimore)*, vol. 91, no. 4, pp. e1-19, Jul 2012.
- [31] D. T. Avery *et al.*, "STAT3 is required for IL-21-induced secretion of IgE from human naive B cells," (in eng), *Blood*, vol. 112, no. 5, pp. 1784-93, Sep 1 2008.
- [32] R. W. Tell and C. M. Horvath, "Bioinformatic analysis reveals a pattern of STAT3-associated gene expression specific to basal-like breast cancers in human tumors," (in eng), *Proc Natl Acad Sci U S A*, vol. 111, no. 35, pp. 12787-92, Sep 2 2014.
- [33] A. F. Freeman *et al.*, "Causes of death in hyper-IgE syndrome," (in eng), *J Allergy Clin Immunol*, vol. 119, no. 5, pp. 1234-40, May 2007.
- [34] A. D. Papanastasiou, S. Mantagos, D. A. Papanastasiou, and I. K. Zarkadis, "A novel mutation in the signal transducer and activator of transcription 3 (STAT3) gene, in hyper-IgE syndrome," (in eng), *Mol Immunol*, vol. 47, no. 7-8, pp. 1629-34, Apr 2010.
- [35] M. A. Kashef, S. Kashef, F. Handjani, and M. Karimi, "Hodgkin lymphoma developing in a 4.5-year-old girl with hyper-IgE syndrome," (in eng), *Pediatr Hematol Oncol*, vol. 23, no. 1, pp. 59-63, Jan-Feb 2006.
- [36] G. Shaw, "Polymorphism and single nucleotide polymorphisms (SNPs)," (in eng), *BJU Int*, vol. 112, no. 5, pp. 664-5, Sep 2013.
- [37] P. Katara, "Single nucleotide polymorphism and its dynamics for pharmacogenomics," (in eng), *Interdiscip Sci*, vol. 6, no. 2, pp. 85-92, Jun 2014.

- [38] J. Wang, G. S. Pang, S. S. Chong, and C. G. Lee, "SNP web resources and their potential applications in personalized medicine," (in eng), *Curr Drug Metab*, vol. 13, no. 7, pp. 978-90, Sep 1 2012.
- [39] D. Gefel, I. Maslovsky, and J. Hillel, "[Application of single nucleotide polymorphisms (SNPs) for the detection of genes involved in the control of complex diseases]," (in heb), *Harefuah*, vol. 147, no. 5, pp. 449-54, 476, May 2008.
- [40] D. A. Benson, I. Karsch-Mizrachi, K. Clark, D. J. Lipman, J. Ostell, and E. W. Sayers, "GenBank," (in eng), *Nucleic Acids Res*, vol. 40, no. Database issue, pp. D48-53, Jan 2012.
- [41] N. L. Sim, P. Kumar, J. Hu, S. Henikoff, G. Schneider, and P. C. Ng, "SIFT web server: predicting effects of amino acid substitutions on proteins," (in eng), *Nucleic Acids Res*, vol. 40, no. Web Server issue, pp. W452-7, Jul 2012.
- [42] E. Capriotti and R. B. Altman, "Improving the prediction of disease-related variants using protein three-dimensional structure," (in eng), *BMC Bioinformatics*, vol. 12 Suppl 4, p. S3, 2011.
- [43] Y. Choi, G. E. Sims, S. Murphy, J. R. Miller, and A. P. Chan, "Predicting the functional effect of amino acid substitutions and indels," (in eng), *PLoS One*, vol. 7, no. 10, p. e46688, 2012.
- [44] M. Hecht, Y. Bromberg, and B. Rost, "Better prediction of functional effects for sequence variants," (in eng), *BMC Genomics*, vol. 16 Suppl 8, p. S1, 2015.
- [45] R. Calabrese, E. Capriotti, P. Fariselli, P. L. Martelli, and R. Casadio, "Functional annotations improve the predictive score of human disease-related mutations in proteins," (in eng), *Hum Mutat*, vol. 30, no. 8, pp. 1237-44, Aug 2009.
- [46] E. Capriotti, P. Fariselli, and R. Casadio, "I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure," (in eng), *Nucleic Acids Res*, vol. 33, no. Web Server issue, pp. W306-10, Jul 1 2005.
- [47] J. Cheng, A. Randall, and P. Baldi, "Prediction of protein stability changes for single-site mutations using support vector machines," (in eng), *Proteins*, vol. 62, no. 4, pp. 1125-32, Mar 1 2006.
- [48] D. Warde-Farley *et al.*, "The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function," (in eng), *Nucleic Acids Res*, vol. 38, no. Web Server issue, pp. W214-20, Jul 2010.
- [49] M. J. Meyer *et al.*, "mutation3D: Cancer Gene Prediction Through Atomic Clustering of Coding Variants in the Structural Proteome," (in eng), *Hum Mutat*, vol. 37, no. 5, pp. 447-56, May 2016.
- [50] S. Wang, W. Li, S. Liu, and J. Xu, "RaptorX-Property: a web server for protein structure property prediction," (in eng), *Nucleic Acids Res*, vol. 44, no. W1, pp. W430-5, Jul 8 2016.
- [51] E. F. Pettersen *et al.*, "UCSF Chimera--a visualization system for exploratory research and analysis," (in eng), *J Comput Chem*, vol. 25, no. 13, pp. 1605-12, Oct 2004.
- [52] S. Wu and Y. Zhang, "MUSTER: Improving protein sequence profile-profile alignments by using multiple sources of structure information," (in eng), *Proteins*, vol. 72, no. 2, pp. 547-56, Aug 2008.
- [53] R. de Silva *et al.*, "Hyper IgE recurrent infection syndrome in South Asia: is there a different outcome?," *Allergy Asthma Clin Immunol*, vol. 14, p. 70, 2018.
- [54] O. Wolach *et al.*, "Variable clinical expressivity of STAT3 mutation in hyperimmunoglobulin E syndrome: genetic and clinical studies of six patients," *J Clin Immunol*, vol. 34, no. 2, pp. 163-70, Feb 2014.
- [55] G. Can *et al.*, "Investigation of IL23R, JAK2, and STAT3 gene polymorphisms and gene-gene interactions in Crohn's disease and ulcerative colitis in a Turkish population," *Turk J Gastroenterol*, vol. 27, no. 6, pp. 525-536, Nov 2016.

- [56] S. Nabhani *et al.*, "STAT3 gain-of-function mutations associated with autoimmune lymphoproliferative syndrome like disease deregulate lymphocyte apoptosis and can be targeted by BH3 mimetic compounds," *Clin Immunol*, vol. 181, pp. 32-42, Aug 2017.
- [57] F. Laudisi, F. Cherubini, G. Monteleone, and C. Stolfi, "STAT3 Interactors as Potential Therapeutic Targets for Cancer Treatment," *Int J Mol Sci*, vol. 19, no. 6, Jun 16 2018.
- [58] S. Manoochehrabadi, S. Arsang-Jang, M. Mazdeh, H. Inoko, A. Sayad, and M. Taheri, "Analysis of STAT1, STAT2 and STAT3 mRNA expression levels in the blood of patients with multiple sclerosis," *Hum Antibodies*, Nov 2 2018.
- [59] B. Saikia *et al.*, "Hyper-IgE syndrome with a novel STAT3 mutation-a single center study from India," *Asian Pac J Allergy Immunol*, vol. 32, no. 4, pp. 321-7, Dec 2014.
- [60] E. D. Renner *et al.*, "Novel signal transducer and activator of transcription 3 (STAT3) mutations, reduced T(H)17 cell numbers, and variably defective STAT3 phosphorylation in hyper-IgE syndrome," (in eng), *J Allergy Clin Immunol*, vol. 122, no. 1, pp. 181-7, Jul 2008.
- [61] K. Kato *et al.*, "Structure and functional analysis of the human STAT3 gene promoter: alteration of chromatin structure as a possible mechanism for the upregulation in cisplatin-resistant cells," (in eng), *Biochim Biophys Acta*, vol. 1493, no. 1-2, pp. 91-100, Sep 7 2000.