J Genet Resour2021;7(2): 227-245 *Homepage: http://sc.journals.umz.ac.ir/*

RESEARCH ARTICLE *DOI: 10.22080/jgr.2021.21373.1252*

J|G|R

Evaluating Functional and Structural Consequences of the most Deleterious Single Nucleotide Polymorphisms of Human C-X-C Motif Chemokine 10 (*CXCL10*) Using *in silico* Analyses

Mozhdeh Riahi[1, 2] and Modjtaba Emadi-Baygi[1, 2*]

[1]Department of Genetics, Faculty of Basic Sciences, Shahrekord University, Shahrekord, Iran
[2]Biotechnology Research Institute, Shahrekord University, Shahrekord, Iran

ARTICLE INFO

Article history:
Received 27 March 2021
Accepted 25 May 2021
Available online 07 Jun 2021

Keywords:
CXCL10
Deleterious substitution
Kaplan-Meier
nsSNP
SIFT

Corresponding authors:
✉ M. Emadi-Baygi
emadi-m@sku.ac.ir

p-ISSN 2423-4257
e-ISSN 2588-2589

ABSTRACT

Single Nucleotide Polymorphisms, especially non-synonymous single-nucleotide polymorphisms (nsSNPs), which are the cause of various diseases, are a major issue in genetics. NsSNPs in protein-coding genes can cause functional and structural variations in the altered protein. The human *CXCL10* gene, localized on chromosome 4q21, is a pro-inflammatory cytokine and plays a role in diverse and critical biological mechanisms. Despite its significance, there is not any document about the impact of variations mapped to this protein. Accordingly, we gathered data about SNPs on the CXCL10 protein and examined the diverse effects of deleterious ones on the function and structure of the protein using various web-based tools. Our analyses indicated that 9 most deleterious nsSNPs (identified by SIFT, PROVEAN, PolyPhen-2, SNPs&GO, PhD-SNP, SNAP2, and PMut) in the conserved region of the CXCL10 affect the molecular function and stability of the protein. By utilizing RMSD values, we concluded that these substitutions in the native structure cause several changes in the protein, including in the N-terminal end, which is vital for binding to the receptor, and finally results in altered regulation, expression, function, and consequently leads to different diseases. Furthermore, some SNPs on the 3′ UTR site showed pattern alterations in the upstream open reading frames (uORFs) and BRD-BOX; moreover, SNPs in this area result in significant changes in miRNA binding sites consequently. Finally, by some analyses, we identified that the CXCL10 deregulation might be a proper prognostic marker in gastric and ovarian cancer. These types of studies help scientists determine whether SNPs are worth following for additional experimental studies to maximize the outcome while studying human health.

Please cite this paper as: Riahi M, Emadi-Baygi M. 2021. Evaluation of functional and structural consequences of the most deleterious single nucleotide polymorphisms of human C-X-C motif chemokine 10 (*CXCL10*) using *in silico* analyses. *J Genet Resour* 7(2): 227-245. doi: 10.22080/jgr.2021.21373.1252

Introduction

The most prevalent alteration in the genome of humans is Single Nucleotide Polymorphisms (SNPs), occurring in the coding and non-coding segments of genes. Protein-coding ones, which are called non-synonymous SNPs (nsSNPs), certainly can cause functional and structural alterations in the mutated protein. While some nsSNPs are phenotypically neutral, others are associated with a particular disease. Several probabilistic-based web servers that can recognize neutral SNPs from disease-related ones have been created (Akhoundi *et al.*, 2016).
A subclass of chemokines called CXC consists of a changeable amino acid among the two first residues, which are greatly conserved cysteine residues. According to the existence or absence of the Glu-Leu-Arg domain, this class is subdivided into two ELR positive and negative

groups, respectively. The ELR positive chemokines have angiogenesis properties, while the ELR negative ones, including the CXCL10, have angiostatic features (Antonelli *et al.*, 2014).

CXCL10 gene is localized on chromosome 4q21 and is related to acute monocytic/B-lymphocyte lineage leukemia (Liu *et al*., 2011). Having four exons, *CXCL10* encodes a 10,000 Daltons protein possessing 98-amino acids. According to PRINTS (Protein Motif fingerprint database) (accessible at [http://www.bioinf.manchester.ac.uk/dbbrowser/PRINTS/]) CXCL10 protein consists of three critical regions 1) N-terminal or triggering domain (^{3}LSRTVR^{8}), 2) GAG binding domain (^{46}KKKGEKR^{52}), and 3) ^{24} LEKLEIIPASQFCPRV EIIATM^{45} constructing β1 strand and a small region of the N-loop and 30s loop (Yang *et al*., 2004). Modification in each of these vital parts might cause functional, structural, and biological changes in the protein.

CXCL10 commences its biological activity by binding to a transmembrane receptor named CXCR3. This procedure is dependent on two vital domains; docking domain (N-loop) and triggering domain placed on the N-terminus of the chemokine. Since these regions are essential for the complete function of the chemokine, polymorphisms in these regions affect receptor binding and activation (Booth *et al*., 2002). According to the previous studies, scientists figured out that *CXCL10* is dysregulated in major human disorders such as cancer, infection, inflammation, and autoimmunity (Liu *et al.*, 2011). It has been proved that the expression of CXCL10 and CXCR3 has increased in a variety of autoimmune diseases as it is a critical factor in the leukocyte homing and causes inflammation and devastation of the target tissue (Young *et al*., 2009). Several studies proved that the up-regulation of CXCL10 in the *central nervous system* results in various neurological diseases (Vinet *et al*., 2010). It has been proved that CXCL10 is a critical interferon-induced chemokine in the anti-viral responses, specifically in respiratory tract infections (Hayney *et al*., 2017).

In the recent pandemic of COVID-19, cytokine storm, a potentially fatal immune reaction, is induced by SARS-Cov-2 infection. Notably, CXCL10 is a key gene related to the cytokine storm of COVID-19 infection (Coperchini *et al*., 2020).

Most of the genetic studies conducted to date have just focused on the polymorphisms located on the promoter part of this gene. For instance, SNP at position -135 (rs56061981) in the promoter of the CXCL10 gene is responsible for the substitution of adenine into guanine. This polymorphism affects the NF-κB binding site in the promoter of the CXCL10 and leads to the expression of various levels of this gene in different individuals (Deng *et al*., 2008).

Research conducted on the patients having colorectal cancer showed that the expression of the CXCL10 gene is elevated both in the plasma and colorectal tissue of the patient group compared to the control group. In addition, they observed obvious differences in genotype distribution and allelic frequencies of a single-nucleotide polymorphism (rs8878) between the patient and normal group (Dimberg *et al*., 2014).

To the best of our knowledge, there is no study on the outcome of non-synonymous SNPs on the structure and function of the *CXCL10 gene*. The existence of polymorphisms in the CXCL10 may change protein stability, binding properties, and post-translational modifications, and, as a consequence, may influence the immune response (Colobran *et al*., 2007). In this research, for the first time, we applied several web-based tools to predict the disease-related nsSNPs, mapped on *CXCL10,* to determine their structural and functional effects on *in silico approaches.* Figure 1 shows different stepts and procedures taken in this study.

Materials and Methods

Datasets

All the data about SNPs (such as ID, protein accession number, the position of substitution, and mutated residue) mapped on the CXCL10 gene was retrieved from the NCBI dbSNP database [http://www.ncbi.nlm.nih.gov/SNP] (Sayers *et al*., 2019) in February 2018.

Recognition of the most deleterious nsSNPs

Five diverse bioinformatics web servers (homology-based and SVM-based classifier) such as SIFT, PROVEAN, PolyPhen-2, PhD-SNP, and SNPs & GO were used to identify the most predicted deleterious nsSNPs (Arshad *et*

al., 2018). Most of the above-mentioned computational tools just need protein sequence and amino acid changes as input.

SIFT (sorting tolerant from intolerant) distinguishes amino acid substitutions having damaging effects on the protein function from neutral ones based on sequence homology and the physical properties. The SIFT score is ranged from 0.0 to 1.0, and substitutions are considered damaging if the SIFT score is ≤ 0.05. This tool is available at [http://sift.bii.a-star.edu.sg/index.html] (Kumar *et al.*, 2009).

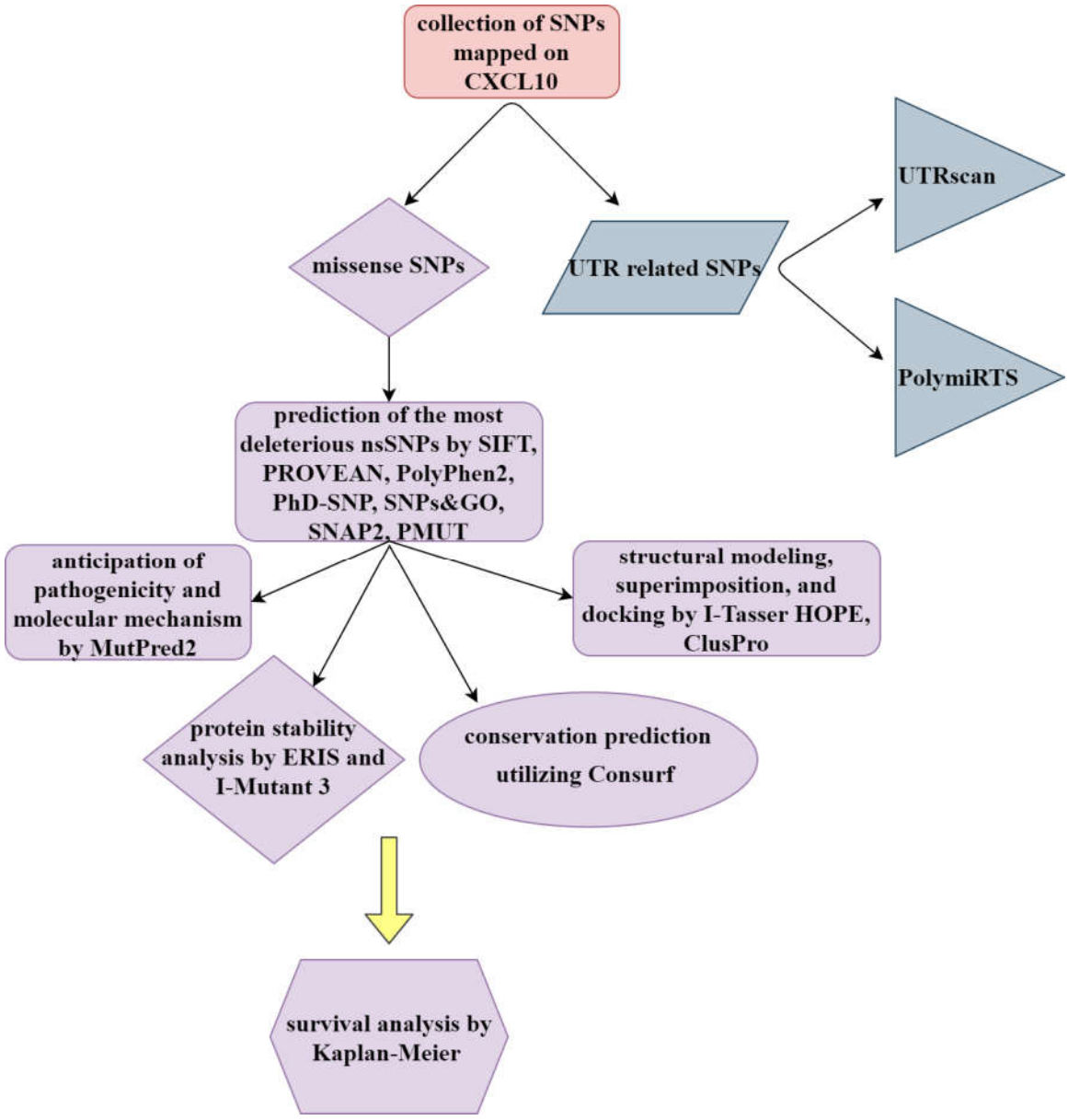

Fig. 1. Schematic flowchart of the steps of the research.

Protein variation effect analyzer/PROVEAN (http://provean.jcvi.org) is another tool for predicting if an nsSNP affects the biological function of a protein. The variants are anticipated as deleterious if the PROVEAN score is ≤ -2.5 and neutral when the score is >-2 (Choi *et al.*, 2015).

Polymorphism phenotyping v2/PolyPhen-2 (http://genetics.bwh.harvard.edu/pph2/) considers the functional and structural impacts of

an amino acid substitution on a protein using evolutionary comparison. The tool provides position-specific independent count (PSIC) scores for each nsSNPs and anticipates variations as probably damaging (more confident), possibly damaging (less confident), and benign. The score ranges from 0 to 1. Amino acid substitutions having a score of 0 to 0.49 are considered benign, possibly damaging if score 0.5 to 0.89, and probably damaging having a score of 0.9 to 1 (Adzhubei *et al.*, 2010).Predictor of human deleterious single nucleotide polymorphisms/PhD-SNP (http://snps.biofold.org/phd-snp/phd-snp.html) is an SVM-based classifier. Its input is protein sequence or Swiss-Prot code, the position of the substitution, and the mutant residue. It provides probability scores range from 0 to 1, and nsSNPs with scores > 0.5 are predicted to be deleterious (Capriotti *et al.*, 2006).

The single nucleotide polymorphism database & gene ontology (SNPs & GO), which It is available at http://snps.biofold.org/snps-and-go/snps-andgo.html, another SVM-based classifier and a GO-integrated predictor and predicts deleterious nsSNPs utilizing sequence or SwissProt code of the protein, GO terms, and a list of considered nsSNPs. The output would be a table of probability scores and reliability indexes (RI) where probability scores > 0.5 are predicted to be deleterious (Capriotti *et al.*, 2013).

Determination of the most deleterious nsSNPs

Screening for non-acceptable polymorphisms/SNAP2 (https://www.rostlab.org/services/SNAP) predicts a score spectrum from -100 (robust neutral prediction) to +100 (robust effect prediction), indicating the probability of a particular SNP to modify the function of the native protein. The input of this tool is protein sequence in FASTA format and forecasts every probable replacement at each position of a protein in a heat map depiction. Dark red displays a high score for effect, while white shows weak signals, and blue illustrates a strong signal for neutral/no effect (Hecht *et al.*, 2015).

Furthermore, we used another web tool named PMut (http://mmb.pcb.ub.es/PMut). It uses protein sequence or its SwissProt/TrEMBL code and merges structural factors with sequence alignment to define deleterious missense substitutions. After choosing the sequence, users can analyze a single mutation or accomplish a complete mutation scan at that position. The result includes a pathogenicity index where indexes >0.5 display pathological mutations and a confidence index ranging from 0 (low) to 9 (high) (Ferrer-Costa *et al.*, 2005).

Prediction of molecular properties

To anticipate the pathogenicity and molecular mechanism of the replacement of each amino acid, MutPred2 (http://mutpred.mutdb.org/) was applied as a web-based tool that effectively examines molecular causes of the diseases, happened after amino acid substitutions (Pejaver *et al.*, 2017). The output of this tool is a general score (g score), molecular mechanisms with P-value <= 0.05, and a P-value and probability specifically for each property. The predicted results are grouped in 3 categories based on both g score and P-value; 1. Very confident hypotheses with *P*-value < 0.01 and g score > 0.75, 2. Confident hypothesis if *P*-value < 0.05 and g > 0.75, and 3. Actionable hypotheses if g > 0.5 and *P*-value < 0.05 (Li *et al.*, 2009).

Prediction of protein stability modifications

To test the stability of the target protein, we carried out I-Mutant 3.0 and ERIS. Analyses with I-Mutant, an support vector machine/ SVM predictor, (http://gpcr2.biocomp.unibo.it/cgi/predictors/I-Mutant3.0/I-Mutant3.0.cgi) anticipates the protein stability changes that occur by a single-site mutation utilizing both the protein structure or the protein sequence. Based on the I-Mutant predictions, the amino acid substitution would largely destabilize the protein if ΔΔG<-0.5 kcal/mol, largely stabilize the protein if ΔΔG>0.5 kcal/mol, or would have a weak effect if -0.5<=ΔΔG<=0.5 kcal/mol (Capriotti *et al.*, 2008). ERIS (https://dokhlab.med.psu.edu/eris/index.php), which similarly predicts the protein stability changes after the occurrence of a mutation, uses Medusa modeling suit to calculate ΔΔG. As an input, the PDB file or code of a protein is uploaded, and the output is free energy changes (ΔΔG). Validation of the results provided by this tool is calculated by comparing

the ΔΔG of a large dataset with the experimental data. If the predicted ΔΔG is greater than 0, there would be a decrease in the stability, and ΔΔG < 0 means an increase in the stability of the protein (Yin *et al.*, 2007).

Analyzing protein evolutionary conservation

The ConSurf server (http://consurf.tau.ac.il/2016/) measures the evolutionary preservation of each amino acid in a protein according to phylogenetic data related to homologous sequences. The tool uses the Bayesian algorithm for conservation and main structural and functional residue prediction (Ashkenazy *et al*., 2010). The output provided by the server is a conservation score (from 1 to 9), which is specified with a color scheme ranging from blue (the most variable residue) to maroon (the most conserved residue). Moreover, if a residue is predicted to be conserved and exposed, it is assumed to be functional, while it would be structural if anticipated as conserved and buried.

Structural analysis

The effect of the nsSNPs on the structural level was examined using the 3D structure of the protein. The native full-length 3D structure of CXCL10 and the mutant proteins were modeled utilizing I-Tasser because the 3D structure available at the PDB bank (https://www.rcsb.org/structure/1LV9) has a mutation. The I-TASSER server (http://zhanglab.ccmb.med.umich.edu/I-TASSER/) first uses a meta-threading technique built to use template proteins. Next, replica-exchange Monte Carlo simulations reassembles the templates and finally builds the unaligned region by *ab initio* modeling, and low free-energy templates are signified by SPICKER. The confidence of models is identified by C-score, which ranges between -5 to 2, and the highest C-score belongs to the best model (Yang *et al*., 2014). All the modeled proteins went under energy minimization using Swiss-PdbViewer v4.1 (Johansson *et al.*, 2012). Mutated proteins were superimposed onto CXCL10 wild protein utilizing SuperPose v1.0 (http://superpose.wishartlab.com) (Maiti *et al*., 2004), and root-mean-square deviation/RMSD values were evaluated. The SuperPose server accepts the PDB accession number of the protein, aligns them using the Needleman-Wunsch pairwise alignment algorithm, superpose structures, and finally calculates RMSD values. The RMSD value is provided for alpha carbons, backbone atoms, heavy atoms, and all atoms. Finally, we visualized superimposed models by Chimera 1.14 (https://www.cgl.ucsf.edu/chimera/) (Pettersen *et al.*, 2004). This tool has a range of applications, such as molecular structure analyses. The superposed structures provided by the SuperPose tool were visualized, colored, and labeled utilizing Chimera.

In addition, we used have (y) our protein explained/HOPE server (http://www.cmbi.umcn.nl/hope/) that predicts the structural impacts of an nsSNP on the protein sequence using UniProtKB and DAS-servers. The protein sequence and selected substitution are applied to the server comparing the native and mutant structures and providing a short explanation of the consequences each mutation would have on the protein, overlapped schematic structures of the native and the mutant amino acids, and disparities in the properties of native and mutant amino acids (Venselaar *et al.*, 2010).

Consequences of SNPs in UTR sites

The 5′ and 3′ sections are untranslated regions/UTRs that critically regulate post-transcriptional modifications, translation potency, and stability. Functional SNPs were analyzed using a pattern matcher service, UTRScan (http://itbtools.ba.itb.cnr.it/utrscan), which finds UTR functional elements accumulated within the UTRsite of a sequence (Grillo *et al.*, 2009). After insertion of an SNP into the native UTR sequence, if each UTR SNP is determined to change motifs, this SNP is believed to have a functional effect. To carry this out, FASTA format information with the mutation was submitted, and its consequences on the confirmed UTR elements in the unique region (5′and 3′) were predicted

Polymorphism in microRNAs and their target sites

PolymiRTS (http://compbio.uthsc.edu/miRSNP/) is designed for the identification of SNPs in the microRNA regions and their target sites (Bhattacharya *et al*., 2014). MicroRNAs have a vital role in translation and mRNA stability;

thus, SNPs at these sites may cause an impressive effect on the expression and biological pathways (Yuan *et al*., 2018). The outputs provided by PolymiRTS are grouped in 4 classes: class 'D' showing that the allele would disrupt a conserved miRNA site, class 'C' means that the allele would create a new miRNA site, class 'N' represents that the allele disrupts a non-conserved miRNA site, and class 'O' assigned when the ancestral allele cannot be determined. The D and C classes, which respectively may result in loss of normal repression and atypical gene repression control, would have the most obvious functional effects on gene expression. The analysis was performed by submitting the CXCL10 symbol to the tool, and SNPs categorized in the D and C classes were considered to have adverse functional impacts.

Survival analysis

Kaplan-Meier plotter (http://kmplot.com/analysis) is a meta-analyzer to estimate the impact of 54,675 genes on survival using 5,143 breasts, 1,816 ovarian, 2,437 lung, and 1,065 gastric cancer patients (Kishore *et al.*, 2010). Three different sources, including gene expression omnibus/GEO, european genome-phenome archive/EGA, and the cancer genome atlas/TCGA are used by this tool. '204533_at' was used as a probe for *CXCL10*. We performed overall survival analysis on 1402, 1656, 1926, and 876 patients with breast, ovarian, lung, and gastric cancers, respectively. This tool compares the survival rate of two groups of samples according to the expression of a specific gene and measures hazard ratio and log-rank P-value.

Results

Distribution and frequency of CXCL10 SNPs

We characterized a total of 1012 SNPs on *the CXCL10* gene from the tremendous SNPs database, dbSNP. Information about the SNPs (dbSNP Id, allele, positions, and substitution) is shown in Supplementary Table 1 (available upon request). Of the total SNPs annotated on the CXCL10, 68 were missense SNPs, 24 SNPs were located in 5'UTR, 171 SNPs were in the 3'UTR region, and the others belonged to other groups. Moreover, the FASTA sequences were retrieved from the NCBI (NM-001565.4 and NP-001556.2). The whole protein contains 98 amino acids, from which the first 21 amino acids code the signal peptide region. We continued our study with the non-synonymous coding SNPs, 5′, and 3′ UTR region SNPs.

Deleterious nsSNPs identified in CXCL10 using various servers

All the 68 nsSNPs reported in the CXCL10 were analyzed by all five aforementioned *in silico* nsSNP prediction algorithms, and the ones identified to be pathogenic by at least 4 of the algorithms were selected for more analyses (Table 1). Of the total 68 missense SNPs, 14 nsSNPs were predicted deleterious, and G18V located in the signal peptide region of the protein was excluded from the study. Supplementary Table 2 (available upon request) shows the results of the remaining nsSNPs subjected to the five mentioned *in silico* nsSNP prediction algorithms.

Verification of the most deleterious nsSNPs

To verify the pathogenicity of the 13 disease-related nsSNPs, identified by *in silico* analyses, we submitted them to both SNAP2 and PMUT servers. Both servers predicted R29C, C30Y, C30R, C32R, I33F, L45S, A64T, L75P, L75R, and P77S as deleterious nsSNPs. However, PMUT indicated R29H, M66T, and E61Q as neutral. Thus, for further analyses, we excluded R29H, M66T, and E61Q. The heat map provided by SNAP2 and the results produced by PMUT is shown in Fig. 2 and Table 2, respectively.

Identifying molecular properties

The information of the 10 most deleterious nsSNPs, predicted based on the previous analyses, was applied to the MutPred server and the predicted molecular effects on the CXCL10 are shown in Table 3. L75P (g score = 0.905) and L75R (g score = 0.897) were predicted to be highly pathogenic and were anticipated with very confident hypothesis to cause loss of disulfide linkage at C74 having p = 9.1e-04 and p = 8.3e-04, respectively. The mutations R29C (g score = 0.814 and p = 5.1e-03), C30Y (g score = 0.955 and p = 5.2e-03), L75R (g score = 0.897 and p = 4.0e-03), C30R (g score = 0.968 and p = 5.2e-03), and C32R (g score = 0.968 and p = 4.5e-03) were assumed to be highly pathogenic and were anticipated to be the cause of altered

metal binding by a very confident hypothesis. Moreover, the alteration in transmembrane protein was predicted to happen at R29C (g score = 0.814 and p = 4.9e-05), C30Y (g score = 0.955 and p = 1.3e-04), C30R (g score = 0.968 and p = 3.8e-04), and C32R (g score = 0.968 and p = 3.7e-04) with a very confident hypothesis. The gain of strand was predicted to occur due to C30Y (g score = 0.955 and p = 8.3e-05).

Table 1. Outputs for the most deleterious nsSNPs identified by the five in silico nsSNP prediction algorithms.

rsID	Amino acid Change	SIFT		PROVEAN		PolyPhen- 2		PhD-SNP		SNPs & GO	
		Prediction	Score	Prediction	Score	Prediction	Score	Prediction	RI	Prediction	RI
rs11548618	R29C	Damaging	0.000	Deleterious	-7.251	Probably damaging	1.000	Disease	6	Disease	7
rs143493954	G18V	Damaging	0.01	Deleterious	-3.261	Possibly damaging	0.829	Disease	8	Disease	7
rs201830102	C30Y	Damaging	0.000	Deleterious	-10.863	Probably damaging	1.000	Disease	7	Disease	8
rs557248373	I33F	Damaging	0.01	Deleterious	-3.158	Probably damaging	0.998	Disease	1	Disease	7
rs759079472	R29H	Damaging	0	Deleterious	-4.309	Probably damaging	1.000	Disease	1	Disease	7
rs766249571	L75P	Damaging	0	Deleterious	-6.975	Probably damaging	1.000	Neutral	2	Disease	3
	L75R	Damaging	0	Deleterious	-5.991	Probably damaging	1.000	Disease	3	Disease	5
rs776216030	C30R	Damaging	0	Deleterious	-11.821	Probably damaging	1.000	Disease	6	Disease	8
rs1054124819	P77S	Damaging	0.01	Deleterious	-7.936	Probably damaging	0.999	Neutral	6	Disease	4
rs1227385544	C32R	Damaging	0	Deleterious	-11.936	Probably damaging	1.000	Disease	6	Disease	8
rs1340181005	M66T	Damaging	0	Deleterious	-3.979	Probably damaging	0.968	Neutral	2	Disease	4
rs1390990135	L45S	Damaging	0	Deleterious	-4.894	Probably Damaging	0.984	Neutral	3	Disease	0
rs1443390834	A64T	Damaging	0.01	Deleterious	-3.042	Probably Damaging	0.993	Neutral	2	Disease	4
rs1481961916	E61Q	Damaging	0	Deleterious	-3.000	Probably Damaging	1.000	Neutral	5	Disease	6

Table 2. Results of the PMUT server for the most deleterious nsSNPs. Amino acid substitutions with a prediction score of 0 to 0.5 are grouped as neutral, while pathological ones score from 0.5 to 1.

Protein	Mutation	Prediction
P02778	R29C	0.63 Disease
P02778	R29H	0.42 Neutral
P02778	C30Y	0.75 Disease
P02778	C30R	0.75 Disease
P02778	C32R	0.75 Disease
P02778	I33F	0.63 Disease
P02778	L45S	0.64 Disease
P02778	E61Q	0.28 Neutral
P02778	A64T	0.64 Disease
P02778	M66T	0.25 Neutral
P02778	L75P	0.75 Disease
P02778	L75R	0.73 Disease
P02778	P77S	0.62 Disease

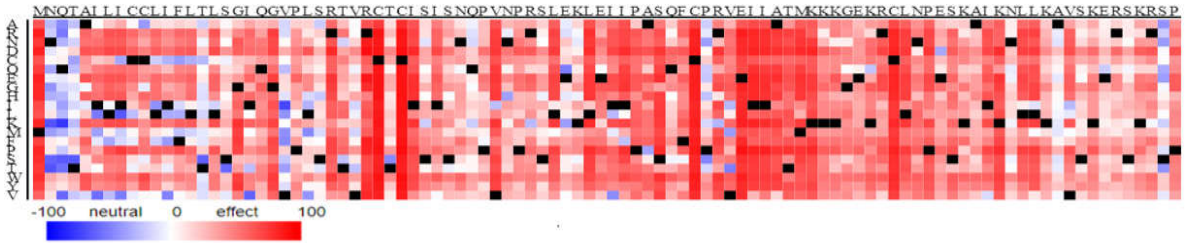

Fig. 2. The heat map provided by the SNAP2 for the most deleterious nsSNPs. The map represents all the possible substitutions in every position. The severity and influence of each substitution are shown by specific colors; dark red means strong effect, white shows weak signals, and blue indicates no effect. Wildtype residues in each position are marked by black. (the color must be used).

Table 3. The resultant probability scores for the most deleterious nsSNPs as provided by MutPred. All the changes showing a very high confident hypothesis are shown in bold.

Mutation	Mutpred score (g score)	Molecular mechanism with p-value < 0.05	Probability	*P*-value
R29C	0.814	**Altered Metal binding**	0.49	5.1e-03
		Altered Transmembrane protein	0.34	4.9e-05
		Loss of Disulfide linkage at C30	0.12	0.04
C30Y	0.955	**Altered Metal binding**	0.44	5.2e-03
		Gain of Strand	0.36	8.3e-05
		Altered Transmembrane protein	0.31	1.3e-04
		Loss of Disulfide linkage at C30	0.13	0.04
I33F	0.619	Altered Transmembrane protein	0.29	3.5e-04
		Altered Metal binding	0.27	0.02
		Gain of Strand	0.27	0.03
		Gain of Disulfide linkage at C30	0.12	0.04
		Loss of GPI-(anchor amidation at N37)	0.01	0.03
L75P	0.905	**Loss of Disulfide linkage at C74**	0.31	9.1e-04
		Gain of B-factor	0.26	0.02
		Loss of Strand	0.26	0.04
		Loss of Acetylation at K80	0.25	0.01
		Altered Disordered interface	0.23	0.03
		Altered Transmembrane protein	0.16	0.01
L75R	0.897	**Altered Metal binding**	0.49	4.0e-03
		Altered Disordered interface	0.41	3.1e-03
		Loss of Disulfide linkage at C74	0.33	8.3e-04
		Loss of Strand	0.27	0.02
		Gain of Acetylation at K80	0.24	0.01
		Altered Transmembrane protein	0.17	0.01
C30R	0.968	**Altered Metal binding**	0.44	5.2e-03
		Altered Transmembrane protein	0.28	3.8e-04
		Loss of Disulfide linkage at C30	0.13	0.04
P77S	0.440	-		
C32R	0.958	**Altered Metal binding**	0.47	4.5e-03
		Altered Transmembrane protein	0.29	3.7e-04
		Loss of Disulfide linkage at C32	0.12	0.04
		Loss of GPI-anchor amidation at N37	0.01	0.03
L45S	0.587	Gain of Intrinsic disorder	0.41	6.5e-03
		Altered Transmembrane protein	0.26	1.3e-03
		Gain of B-factor	0.26	0.02
		Loss of SUMOylation at K47	0.21	0.03
A64T	0.538	Gain of B-factor	0.25	0.03
		Gain of SUMOylation at K69	0.20	0.03
		Altered Transmembrane protein	0.17	0.01
		Altered Stability	0.14	0.02
		Altered Metal binding	0.05	0.04

Stability changes determined

Most of the mutated proteins that cause diseases display stability changes. Identifying the relation of structure and function of a protein is approachable after the prediction of protein stability. All the 10 most deleterious nsSNPs were submitted to both I-Mutant 3.0 and ERIS servers (utilizing flexible backbone and pre-relaxation settings) to validate the free energy prediction. All the deleterious nsSNPs resulted in protein destabilization (Table 4).

Conservation profile of the most deleterious nsSNPs in *CXCL10*

The evolutionary conservation of each amino acid in every position in the proteins is of paramount importance for the maintenance of functional and structural properties. We used ConSurf to determine the conservation profile of the most deleterious nsSNPs determined in this research (Ashkenazy *et al*., 2016). Results obtained via ConSurf (Fig. 3) represented that R29 and L75 are conserved and functional residues, while C30, C32, L45, and A64 are predicted as structural (buried) residues and conserved. Based on the anticipation done by ConSurf, almost all the deleterious nsSNPs signified in this study are located in the conserved place. Since residue I33 has average conservation in the CXCL10 protein, we excluded it from further investigation.

Table 4. Prediction of the protein stability upon amino acid substitutions by I-Mutant 3.0 and ERIS.

Amino Acid Change	I-Mutant 3.0		ERIS	
	Stability changes	Free energy (ΔΔG) prediction	Mutation category	Score
R29C	Largely Decrease	-1.05	Destabilizing	3.32
C30Y	Largely Decrease	-0.42	Destabilizing	4.11
I33F	Largely Decrease	-1.75	Destabilizing	6.41
L75P	Largely Decrease	-1.39	Destabilizing	4.97
L75R	Largely Decrease	-1.30	Destabilizing	7.92
C30R	Largely Decrease	-0.51	Destabilizing	>10
C32R	Largely Decrease	-0.59	Destabilizing	9.47
P77S	Largely Decrease	-1.35	Destabilizing	5.49
L45S	Largely Decrease	-2.22	Destabilizing	1.05
A64T	Largely Decrease	-0.60	Destabilizing	7.58

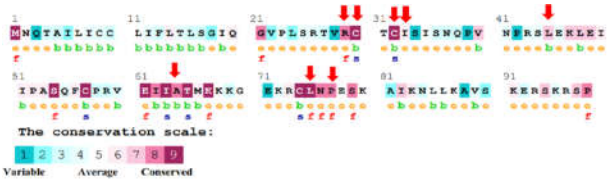

Fig. 3. ConSurf analysis of C-X-C motif chemokine 10 precursor protein (Uniprot ID: P02778). Conservation grades are shown with a range of colors from blue, displaying the most variable residues (grade 1), white the intermediate ones (grade 5), and maroon the most conserved ones (grade 9). The most deleterious nsSNPs anticipated in this research are marked with red arrows. e: an exposed residue, b: a buried residue, f: a predicted functional residue, s: a predicted structural residue. (Color should be used).

Structural modeling and superimposition of the most deleterious nsSNPs

We first modeled the wild and mutant proteins of CXCL10, submitting the protein sequence without the signal peptide fragment to the I-Tasser tool. The first model having the highest C-score and confidence was chosen (Supplementary Table 3 (available upon request)). Evaluating the superimposed structure of the native and mutant protein models (Fig. 4) showed that almost all the monomeric structures exhibited a typical chemokine fold, including a three-stranded β sheet overlaid by an α helix. R29C substitution showed changes in the N-terminal end critical for receptor binding and modification in the site of the N-loop that is called the docking domain and is crucial for recognition of the receptor.

The nsSNPs such as C30R and C30Y, which are vital for the formation of the cysteine bridges, with the high heavy chain and backbone RSMD, further showed alteration in the N-terminal region and some other structural changes. Both the L75P and L75R, showing high RMSD values, entail alteration in the N-terminal end and other conformational and structural modifications. The substitution P77S occurs in the proline residue, which has a significant role in the formation of turns and beta-turns in protein and may result in a significant change in the protein showing a high RMSD value after superimposition. The remaining amino acid substitutions, *i.e.* C32R, L45S, and A64T, had the lowest RMSD values among the others. The substitution R29C illustrated the highest RMSD

value followed by C30R, C30Y, and L75R, respectively. Hope explained about the reaction and physicochemical features of the 3D structures of the wild and mutated CXCL10. The results of all the 9 deleterious nsSNPs and description of changes in the native structure through the new residues are available as follows: We found that the substitutions such as R29C, C30Y, C30R, and C32R in the N-terminal receptor-binding motif ^{8}RCTCISISN^{16} of the CXCL10 protein cause changes in the hydrophobic potential of the protein and destabilize it by damaging cysteine bridges which are of paramount importance for structural stability.

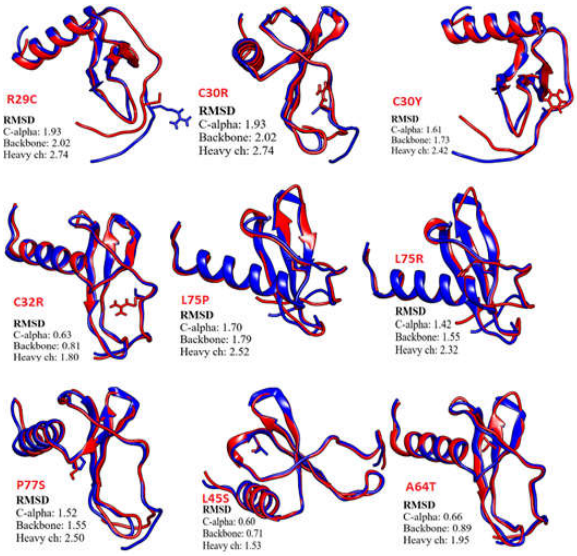

Fig. 4. Superimposition of the 3D structure of the mutant (in red) on the wild-type CXCL10 protein (in blue) using Chimera. Evaluation of the superimposed structure of the native and mutant protein models showed that almost all the monomeric structures exhibited a typical chemokine fold including a three-stranded β sheet overlaid by an α helix (for more information please refer to the main text). (The color must be used). Thus they may inhibit receptor binding of the protein and result in local and structural changes. Substitutions such as L75P and L75R cause protein folding problems because of the replacement of a small residue and positive charge in the core of the protein, respectively. Amino acid substitution P77S may disturb the specific conformation caused by proline residue, which is found in turns. The remaining amino acid substitutions in CXCL10 protein, including L45S (by creating a space in the core of the protein) and A64T (by substituting a larger residue), lead to considerable structural change in the protein. The overlapped structures and the explementary results for all the nine deleterious nsSNPs are presented in Fig. 5.

Functional SNPs of UTR regions

We submitted all 195 UTR SNPs to the UTRscan server (Table 6). After the characterization of the functional elements presented in UTR regions, we saw that 2 SNPs, named rs1383287989 and rs1406377228, in the 3′UTR region caused a change in the uORF (Upstream Open Reading Frames) element. The sequences that contain a start codon in the frame and a termination codon placed upstream or downstream of the main AUG are called uORFs. These elements can lessen the performance of translation initiation of the principal downstream ORF in unstressed circumstances (Barbosa *et al.*, 2013).

Another SNP in the 3′UTR region (rs886636633) resulted in the gain of a new element named BRD-BOX (AGCTTTA). BRD-BOX is a conserved motif in the 3'UTR area of Notch pathway target genes, such as members of the basic helix-loop-helix repressor family and the Bearded family, in Drosophila. There is one or several numbers of this element in 3'UTRs and affects post-transcriptional regulation negatively through impacting transcript stability and translation capability. This motif plays a role in the creation of RNA-RNA duplexes with the 5' end of complementary miRNAs (Lai, 2002).

Table 6. SNPs in 3′UTR related to the functional pattern change.

SNP ID	Nucleotide change	UTR position	Functional element change
rs886636633	G/A	3′ UTR	no pattern → BRD-BOX
rs1383287989	A/C	3′ UTR	uORF → no pattern
rs1406377228	T/C	3′ UTR	uORF → no pattern

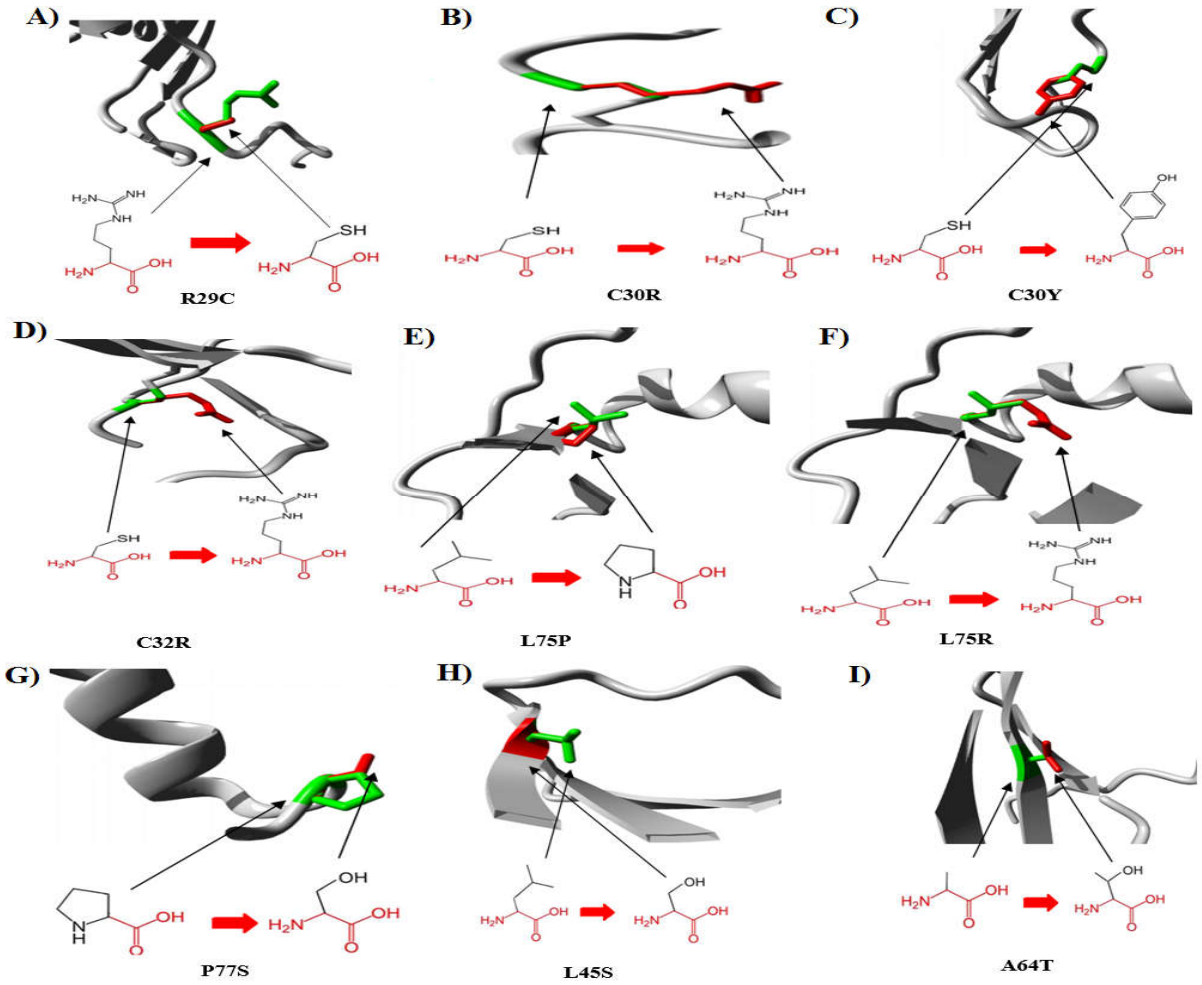

Fig. 5. Overlapped structure of the native and mutant amino acids for the most deleterious nsSNPs obtained from HOPE. The native and mutant amino acids are shown in green and red, respectively. (the color must be used): A) The charge of the native (positive) and substituted (neutral) amino acids are different, and this may result in loss of interaction. There might be a lack of external interactions, for the native residue is bigger than the mutant residue. Moreover, the hydrophobicity of the wild and mutant reside differs (Probably Damaging); B) The arginine residue is big and has a positive charge while the cysteine is small and neutral. The wild-type residue is more hydrophobic. Mutation cause in loss of the cysteine bond (destabilization of the structure and Probably damaging); C) The arginine residue is big and has a positive charge while the cysteine is small and neutral. The wild-type residue is more hydrophobic. Mutation causes loss of the cysteine bond (destabilization of the structure and Probably damaging); D) The wild-type residue charge was neutral; the mutant residue charge is positive. The mutant residue is big. The wild-type residue is more hydrophobic (Damaging and mutation of 100% conserved residue); E) The mutant residue is small (space at the core of the protein). The mutant residue was not observed in other homologous proteins, but residues with some common properties were observed. Sometimes the mutation might occur without damaging impact for the protein; F) The mutant residue makes a positive charge in the buried part of the protein (protein folding problems). The mutant residue charge is positive and big. The mutation will cause loss of hydrophobic interactions (Probably damaging); G) The wild-type residue is more hydrophobic and bigger. The mutation can disturb the special conformation caused by a proline residue (Probably damaging); H) The mutant residue is small (mutation will cause space in the core of the protein). The wild-type residue is more hydrophobic (mutation will cause loss of hydrophobic interactions). Another residue type was observed often at this position in homologous sequences, but it was not similar to this mutant residue (Probably damaging); I) The mutant residue is big and will not fit and also it will cause loss of hydrophobic interactions. Neither this mutant residue nor other residue types with similar properties was not observed at this position in homologous sequences (Probably damaging).

Prediction of SNPs in 3'UTR region

Expression of an mRNA may be up-regulated or down-regulated because of the influences the creation or removal of miRNA target regions may cause the interactions between the mRNA and a miRNA. Of 1012 SNPs characterized in the *CXCL10* gene, 171 were in the 3′ UTR region. Among them, seven functional SNPs were anticipated to disrupt miRNAs conserved sites with a conservation score ≥ 2. Four out of those seven named SNPs might create eight new miRNA target sites. These SNPs may lead to an increase or decrease in the levels of the CXCL10 protein by the creation and deletion of miRNA target sites. The results are presented in Table 7.

Table 7. Seven functional SNPs that disrupt miRNAs conserved site as predicted by PolymiRTS database. The sequence related to each miRNA site is shown in miRSite. Capital letters represent the complementary bases of the seed region, and SNPs are shown in bold.

dbSNP ID	miR ID	Conservation	miRSite	Function class
rs191522507	hsa-miR-219a-2-3p	5	tacatA**C**AATTCc	D
	hsa-miR-297	3	tACATA**C**AAttcc	D
	hsa-miR-3149	3	taCATA**C**AAttcc	D
	hsa-miR-567	3	tACATA**C**Aattcc	D
	hsa-miR-675-3p	3	taCATA**C**AAttcc	D
rs182871280	hsa-miR-548ao-5p	3	tctTTA**C**TTCAtg	D
	hsa-miR-548ax	3	tctTTA**C**TTCAtg	D
	hsa-miR-5585-5p	3	tcttTA**C**TTCAtg	D
	hsa-miR-8060	2	tctttaCTTCATG	D
rs35795399	hsa-miR-302b-5p	3	ttTTAA**A**GAatgc	D
	hsa-miR-302d-5p	3	ttTTAA**A**GAatgc	D
rs58658570	hsa-miR-624-5p	2	gGGTAC**T**Aaggaa	D
	hsa-miR-3121-5p	2	gggtaC**A**AAGGAa	C
rs187517470	hsa-miR-3153	2	ctcacC**T**TTCCCA	D
	hsa-miR-4668-5p	2	ctcacc**T**TTCCCA	D
	hsa-miR-6730-5p	4	ctcACC**T**TTCcca	D
	hsa-miR-6733-5p	2	ctcacC**T**TTCCCA	D
	hsa-miR-6739-5p	2	ctcacC**T**TTCCCA	D
	hsa-miR-3202	4	ctcaCC**C**TTCCca	C
	hsa-miR-4747-5p	2	ctcacC**C**TTCCCA	C
	hsa-miR-5196-5p	2	ctcacC**C**TTCCCA	C
rs147499398	hsa-miR-4511	2	atGTTC**T**TAgtgg	D
	hsa-miR-548b-3p	2	atGTTC**T**TAgtgg	D
	hsa-miR-3914	2	atGTTC**C**TAgtgg	C
rs148141229	hsa-miR-29a-3p	9	aGGTGC**T**Atgttc	D
	hsa-miR-29b-3p	9	aGGTGC**T**Atgttc	D
	hsa-miR-29c-3p	9	aGGTGC**T**Atgttc	D
	hsa-miR-5682	9	aGGTGC**T**Atgttc	D
	hsa-miR-6871-3p	9	aGGTGC**T**Atgttc	D
	hsa-miR-183-5p	3	agGTGC**C**ATgttc	C
	hsa-miR-593-5p	9	aGGTGC**C**Atgttc	C
	hsa-miR-942-3p	3	aggtGC**C**ATGTtc	C

CXCL10 deregulation and correlated survival rate of cancerous patients

Here we carried out the Kaplan-Meier plot analysis to figure out the prognostic value of CXCL10 in 4 different cancers. Based on the hazard ratio and log-rank P, the Kaplan Meier plots (Fig. 6) showed an association between ovarian and gastric cancer with CXCL10 gene deregulation. Data showed that the high expression of CXCL10 with a hazard ratio (HR) = 0.81 (0.7-0.92) and log-rank p-value = 0.0019 is related to a higher survival rate in ovarian cancer; furthermore, in the matter of gastric cancer patients, higher expression of CXCL10 having HR = 0.79 (0.67-0.94) and log-rank p-value = 0.0078 decreases the risk (higher survival rate). Therefore, according to this study, the deregulation of CXCL10 seems to be a good prognostic marker for ovarian and gastric cancer, and since SNPs may deregulate the expressed protein, the 9 nsSNPs identified in this research

might have an indirect relation with the occurrence of the above-mentioned cancers.

Discussion

Undoubtedly, genetic differences in non-coding regions can result in phenotypic effects. The reason is that methods for the recognition of disease-related regulatory variants are presently the main focus (Fernald *et al.*, 2011; Shendure *et al.*, 2015). Alteration in the protein sequence affects protein structure, stability, and function. Non-synonymous SNPs can have adverse impacts on the genotype and phenotype of any protein and might be the basis for various diseases, such as cancer (Kulshreshtha *et al.*, 2016).

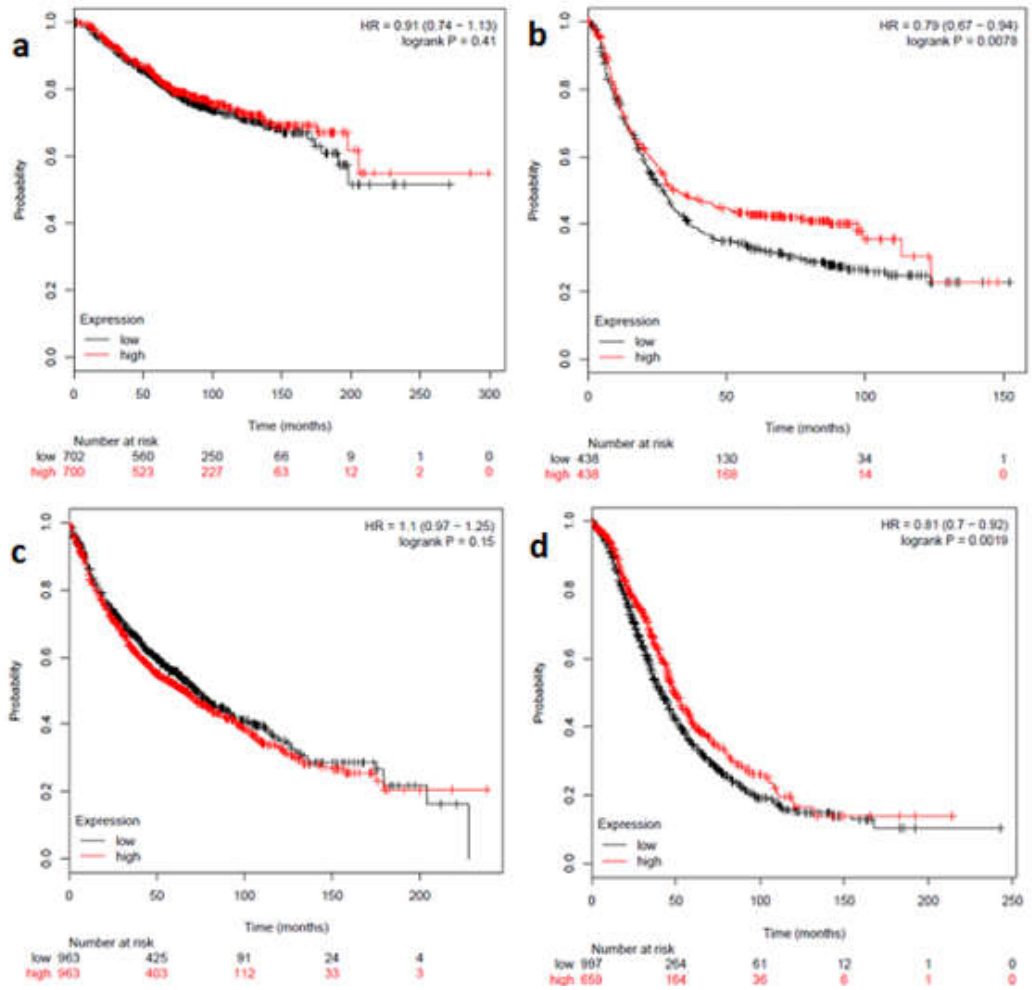

Fig. 6. Kaplan-Meier plotter was utilized to identify the association of deregulation of the CXCL10 gene with the overall survival of patients with different cancer types. a. Breast cancer plot. b. Gastric cancer plot. c. Lung cancer plot. d. Ovarian cancer plot. (Color should be used).

There are several studies on the expression of CXCL10 in various disorders. Of note, IDH1-mutated glioma tumors are less efficiently infiltrated by CD8+ T cells, participating in tumor rescue. In the tumors, the secretion of the chemokines CXCL9 and CXCL10 was limited due to the reduction of the expression of STAT1 (Lucca *et al.*, 2017). Furthermore, the serum CXCL10 concentration was addressed to be related to the count of circulating lymphocytes in head and neck cancer with radiation therapy. Immune cells represent an anti-tumor impact on cancer cells via paracrine CXCL9, -10, -11/ CXCR3 axis in tumor models. Arenberg (2010) *et al.* noted that the application of CXCL10 intratumorally led to superior survival

of mice vaccinated with carcinoma cells of the lung (Arenberg *et al.*, 2010). The application of retroviral CXCL10 gene transduction showed the suppression of tumor growth in models of various tumors (Tokunaga *et al*., 2018). Moreover, CXCL10 overexpression leads to reduced tumor burden and lethal ascites aggregation in the ID8 syngeneic murine model of high-grade serous ovarian cancer/HGSC. Decreased CXCL10 expression in tumors from KD mice leads to increased ascites aggregation and disease progression in contrast to the controls. CXCL10 is a positive identifier of anti-tumor immune responses in HGSC tumor immune microenvironment and disease progression (Au *et al*., 2017). However, polymorphisms located on the CXCL10 gene have not been considered until now. Therefore, in this research, 68 missense SNPs on the CXCL10 gene were retrieved from dbSNP, and several *in silico* tools were used to predict the functional and structural consequences of these missense SNPs on the CXCL10 gene.

Due to the extensive prevalence of SNPs in the human genome, it is more reasonable to utilize bioinformatics tools, sequence or structural based, to screen potentially deleterious SNPs before wet laboratory experiments (Mah *et al*., 2011). Sequence-based algorithms predict all kinds of impacts at the protein level in proteins with known relatives, whereas they cannot identify the related mechanisms responsible for the current phenotype. On the other hand, the structure-based tools can shed light on the underlying mechanisms (Yue *et al*., 2005). However, their main limitation is being unable to work in the lack of 3D structures. In the current research, several prediction algorithms (sequence or structural based) were used to multiply the validity of the results, since every tool has specific advantages and disadvantages, make it different from the others, and apply numbers of various algorithms can help to predict the impact of each SNP with more confidence. Moreover, It was well recognized that the aggregation of sequence homology and structural homology-based servers, including SIFT and PolyPhen-2, obviously correlate damaging/deleterious prediction scores of an SNP in line with experimental/laboratory data derived from site-directed mutagenesis and clinical association analyses (Wang *et al*., 2001; Wang *et al*., 2009; Dong *et al*., 2015; Karbassi *et al*., 2015). Ten out of 68 missense nsSNPs were anticipated to be most deleterious according to the 7 different algorithms (SIFT, PROVEAN, PolyPhen-2, SNPs&GO, PhD-SNP, SNAP2, and PMut). Three motif sequences are present in the CXCL10 protein (the number identified for each residue is given without considering the signal peptidase region). RCTCISISN $^{8\text{-}16}$, which comprises motif 1, encodes the N-terminal region and contains the first two cysteines (Clark-lewis *et al*., 2003). Substitutions rs11548618, rs201830102, rs776216030, rs1227385544, and rs557248373 are located in this region. A monomeric structure of CXCL10, which was found by NMR, illustrated that the N-terminal region, which has been reported to be crucial for receptor binding, forms a hydrophobic cleft with a 30s-loop. Recent studies have shown that residue R8 (rs11548618), locating in the 30s loop, may be important in receptor binding (Booth *et al*., 2002). Moreover, residues C30 and C32 are substituted in rs201830102, rs776216030, and rs1227385544 polymorphisms, respectively. These residues form the C-X-C chemokine motif and disulfide bridges (30:57 and 32:74), which stabilize the 3-D structure of the protein (Swaminathan *et al*., 2003). ^{24}LEKLEIIPAS QFCPRVEIIATM^{45} sequence is the second motif that spans the first two beta-strands and parts of the docking domain, essential for receptor binding (Swaminathan *et al*., 2003), and substitutions rs1390990135, rs1443390834, and rs1481961916 are located in this motif. The last motif (^{47}KKGEKRCLNPESKAIKNLL^{65}) spans a part of the C-terminal helix and is discovered to be vital for binding to the heparin (Proudfoot *et al*., 2001). Two missense SNPs rs1054124819 and rs766249571, are sited in this part of the gene. Campanella et al. indicated that Arg-8, residues 22-26 and 46-47, located in the loop regions of mouse IP-10, are all crucial for receptor binding and signaling (Campanella *et al*., 2003). A number of immunological studies determined that residues 20-36 in the human CXCL10 are involved in CXCR3 binding (Jabeen *et al*., 2008). Furthermore, Yang *et al.* designed various forms of CXCL10 having mutations in the CXCR3 and glycosaminoglycan

binding sites, and they confirmed that the total lack of receptor binding site but not the glycosaminoglycan binding site inhibits tumor angiostatic activity of the protein (Yang *et al*., 2004). Therefore, the existence of the deleterious nsSNPs studied in this research in the N-terminal region, GAG binding regions, and critical motifs of the protein might inhibit GAG and receptor binding, and finally, the angiostatic activity of the protein.

The stability of the protein structure is essential for its function, and missense SNPs destabilizing protein structure could cause human monogenic disease (Yue *et al*., 2005). Our stability analysis indicated that almost all of the identified deleterious nsSNPs deteriorate protein stability. Under these circumstances, molecular mechanisms predicted by MutPred2 such as altered transmembrane protein (resulted from R29C, C30Y, C30R, and C32R), altered metal binding (cause of R29C, C30Y, L75R, C30R, and C32R), gain of strand (C30Y), altered disordered interface (L75R), and loss of disulfide linkage (L75P and L75R) can be regarded as a consequence of decreased stability.

Moreover, given these points that conserved residues are vital for the function and three-dimensional structure of a protein, thereupon polymorphisms in these residues might result in the impairment of conformation, and consequently affinity and function of the protein (Miller *et al*., 2001). Our results indicate that since all the identified nsSNPs except for one are located in the conserved region of the protein, thereby they may have a major impact on the impairment of molecular mechanisms.

Furthermore, all the identified most deleterious nsSNPs led to local and structural changes based on having high RMSD values. Considering these changes between the native and mutant proteins, it is reasonable to conclude that they can lead to various consequences for the destabilization of protein structure or defective binding to receptor and heparin.

Untranslated regions are responsible for translation regulation and transcript stability control (Mignone *et al*., 2002). Regarding our analysis with UTRscan, we found that two SNPs were related to the functional pattern change of uORF; henceforth, they might be correlated with significantly reduced protein expression levels (Barbosa *et al*., 2013). Furthermore, another SNP in 3′UTR (rs886636633) resulted in the gain of a BRD-BOX element. Notably, this element negatively affects post-transcriptional regulation by changing transcript stability and translation potency (Lai, 2002).

Equally important, we found functional SNPs in the 3′ UTR region of the CXCL10 gene that disrupts/creates miRNAs conserved sites. Since miRNAs affect gene regulation by mRNA degradation (O'Brien *et al*., 2018), both creation and disruption of miRNA target sites can influence the expression level of CXCL10 and thereby cause a range of disorders.

With all things in mind, our approach identified the most deleterious nsSNPs mapped on the CXCL10 gene, perturbing its characteristics at both structural and functional levels. Since structural and functional alterations disturb gene expression, we aim to see if there are any changes in the gene expression in cancerous tissues due to the genetically unstable nature of cancer. Using the Kaplan-Meier plot, we showed that the deregulation of this gene is related to gastric and ovarian cancers. Markedly, the high expression of CXCL10 is correlated with higher survival rates in both mentioned cancers. Qingyuan Meng *et al*. confirms the correlation between the overexpression of CXCL10 and gastric cancer. They revealed that CXCL10 is a potent chemoattractant for T lymphocytes and inhibits tumor growth by suppression of angiogenesis (Meng *et al*., 2020). Besides, Zhao *et al*. (2015) figured out that the overexpression of the CXCL10 gene in a model of rodent cervical cancer makes the tumor more sensitive to radiotherapy (Zhao *et al*., 2015). In the same vein, the correlation between CXCL10 and ovarian cancer revealed that the overexpression of the CXCL10 gene would result in tumor suppression by the recruitment of tumor-infiltrating lymphocytes (Bronger *et al*., 2016). To sum up, it can be postulated that our identified deleterious nsSNPs perturbing the structure and function of the CXCL10 gene have not occurred in gastric and ovarian cancer with more survival rates. Our original implemented in silico approach showed that all identified nsSNPs have possibly major impacts on the structure and function of the CXCL10 protein. Furthermore, the examined SNPs in the 3 UTR

of the CXCL10 transcript affect its post-transcriptional regulation by the emergence of a BRD-BOX element and/or alteration of miRNA target sites. On the whole, these alterations may contribute to the etiology of various diseases, including cancers.

Conclusion

We identified the most deleterious nsSNPs all occurred in the conserved region of the CXCL10 protein and SNPs in the 3 UTR region of its transcript that affects the gene product at both RNA and protein levels. We postulated that under those circumstances, the structure and function of the CXCL10 could alter and lead to various pathophysiological conditions. As we focused on the dried experiments to analyze the SNPs of the CXCL10, our work calls for further wet experimental studies to substantiate the impact of most deleterious substitutions on the function and structure of the CXCL10 gene products to shed light on their role(s) in various pathophysiological conditions.

Conflicts of interest

The authors have no conflicting interests.

References

Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. 2010. A method and server for predicting damaging missense mutations. *Nat Methods* 7: 248-249.

Akhoundi F, Parvaneh N, Modjtaba EB. 2016. In silico analysis of deleterious single nucleotide polymorphisms in human BUB1 mitotic checkpoint serine/threonine kinase B gene. *Meta Gene* 9: 142-150.

Antonelli A, Ferrari SM, Giuggioli D, Ferrannini E, Ferri C, Fallahi P. 2014. Chemokine (C-X-C motif) ligand (CXCL)10 in autoimmune diseases. *Autoimmun Rev* 13(3): 272-280.

Arshad M, Bhatti AJP. 2018. Identification and in silico analysis of functional SNPs of human TAGAP protein: A comprehensive study. *Plos One* 13(1): 1-13.

Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, Pupko T, Ben-Tal N. 2016. ConSurf 2016 : an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res* 44: 344-350.

Ashkenazy H, Erez E, Martz E, Pupko T BN. 2010. ConSurf 2010 : calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res* 38: 529-533.

Au KK, Peterson N, Truesdell P, Reid-Schachter G, Khalaj K, Ren R, Francis JA, Graham CH, Craig AW, Koti M. 2017. CXCL10 alters the tumour immune microenvironment and disease progression in a syngeneic murine model of high-grade serous ovarian cancer. *Gynecol Oncol* 145(3): 436-445.

Arenberg D, Luckhardt TR, Carskadon S, Zhao L, Amin MA, Koch AE. 2010. Macrophage migration inhibitory factor promotes tumor growth in the context of lung injury and repair. *Am J Respir Crit Care Med* 182(8): 1030-1037.

Barbosa C, Peixeiro I, Romão L. 2013. Gene expression regulation by upstream open reading frames and human disease. *PLoS Genet* 9(8): e1003529.

Bhattacharya A, Ziebarth JD, Cui Y. 2014. PolymiRTS database 3.0: linking polymorphisms in microRNAs and their target sites with human diseases and biological pathways. *Nucleic Acids Res* 42: D86-91.

Booth V, Keizer DW, Kamphuis MB, Clark-lewis I, Sykes BD. 2002. The CXCR3 binding chemokine IP-10 / CXCL10 : structure and receptor. *Biochemistry* 41(33): 10418-10425.

Bronger H, Singer J, Windmüller C, Reuning U, Zech D, Delbridge C, Dorn J, Kiechle M, Schmalfeldt B, Schmitt M, Avril S. 2016. CXCL9 and CXCL10 predict survival and are regulated by cyclooxygenase inhibition in advanced serous ovarian cancer. *Br J Cancer* 115(5): 553-563.

Campanella GS, Lee EM, Sun JLA. 2003. CXCR3 and Heparin binding sites of the chemokine ip-10. *J Biol Chem* 278(19): 17066-17074.

Capriotti E, Calabrese R, Casadio R. 2006. Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and

evolutionary information. *Bioinformatics* 22(22): 2729-2734

Capriotti E, Calabrese R, Fariselli P, Martelli P, 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* 14(Suppl 3): S6.

Capriotti E, Fariselli P, Rossi I, Casadio R. 2008. A three-state prediction of single point mutations on protein stability changes. *BMC Bioinformatics* 9(s6): 1-9.

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.

Clark-lewis I, Mattioli I, Gong J, Loetscher P. 2003. Structure-function relationship between the human chemokine receptor CXCR3 and its ligands. *J Biol Chem* 278(1): 289-295.

Colobran R, Pujol-Borrell R, Armengol MP JM. 2007. The chemokine network. II. On how polymorphisms and alternative splicing increase the number of molecular species and configure intricate patterns of disease susceptibility. *Clin Exp Immunol* 150(1): 1-12.

Coperchini F, Chiovato L, Croce L, Magri F, Rotondi M. 2020. The cytokine storm in COVID-19: An overview of the involvement of the chemokine/chemokine-receptor system. Cytokine Growth Factor Rev 53: 25-32.

Deng G, Zhou G, Zhang R, Zhai Y, Zhao W, Yan Z, Deng C, Yuan X, Xu B, Dong X, Zhang X. 2008. Regulatory polymorphisms in the promoter of CXCL10 gene and disease progression in male Hepatitis B virus carriers. *Gastroenterology* 134(3): 716-726.

Dimberg J, Skarstedt M, Löfgren S, Zar N, Matussek A. 2014. Protein expression and gene polymorphism of CXCL10 in patients with colorectal cancer. *Biomed Rep* 2: 340-343.

Dong C, Wei P, Jian X, Gibbs R, Boerwinkle E, Wang K, Liu X. 2015. Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Hum Mol Genet* 24(8): 2125-2137.

Fernald GH, Capriotti E, Daneshjou R, Karczewski KJ, Altman RB. 2011. Bioinformatics challenges for personalized medicine. *Bioinformatics* 27(13): 1741-1748.

Ferrer-Costa C, Gelpí JL, Zamakola L, Parraga I, de la Cruz X, Orozco M. 2005. Structural bioinformatics PMUT: a web-based tool for the annotation of pathological mutations on proteins. *Bioinformatics* 21(14): 3176-3178.

Grillo G, Turi A, Licciulli F, Mignone F, Liuni S, Banfi S, Gennarino VA, Horner DS, Pavesi G, Picardi E, Pesole G. 2009. UTRdb and UTRsite (RELEASE 2010): A collection of sequences and regulatory motifs of the untranslated regions of eukaryotic mRNAs. *Nucleic Acids Res* 38: D75-80.

Hayney MS, Henriquez KM, Barnet JH, Ewers T, Al E. 2017. Serum IFN-gamma-induced protein 10 (IP-10) as a biomarker for severity of acute respiratory infection in healthy adults. *J Clin Virol* 90: 32-37.

Hecht M, Bromberg Y, Rost B. 2015. Better prediction of functional effects for sequence variants. *BMC Genomics* 16(8): S1. https://doi.org/10.1186/1471-2164-16-S8-S1

Jabeen T, Leonard P, Jamaluddin H, Acharya K.R. 2008. Structure of mouse IP-10, a chemokine. *Acta Crystallogr D Biol Crystallogr* 64(6): 611-619.

Johansson MU, Zoete V, Michielin O, Guex N. 2012. Defining and searching for structural motifs using DeepView / Swiss-PdbViewer. *BMC Bioinformatics* 13 (173). https://doi.org/10.1186/1471-2105-13-173

Karbassi I, Maston GA, Love A, DiVincenzo C, Braastad CD, Elzinga CD, Bright AR, Previte D, Zhang K, Rowland CM, McCarthy M. 2015. A standardized DNA variant scoring system for pathogenicity assessments in mendelian disorders. Hum Mutat 37(1): 127-134.

Kishore J, Goel M, Khanna P. 2010. Understanding survival analysis: Kaplan-Meier estimate. *Int J Ayurveda Res* 1(4): 274-278.

Kulshreshtha S, Chaudhary V, Goswami GK, Mathur N. 2016. Computational approaches for predicting mutant protein stability. *J Comput Aided Mol Des* 30(5): 401-412.

Kumar P, Henikoff S, Ng P. 2009. Predicting the effects of coding non-synonymous variants

on protein function using the SIFT algorithm. *Nat Protoc* 4(7): 1073-1082.

Lai EC. 2002. Micro RNAs are complementary to 3′UTR sequence motifs that mediate negative post-transcriptional. Nat Genet 30: 363-364.

Li B, Krishnan VG, Mort ME, Xin F, Kamati KK, Cooper DN, Mooney SD, Radivojac P. 2009. Automated inference of molecular mechanisms of disease from amino acid substitutions. *Bioinformatics* 25(21): 2744-2750.

Liu M, Guo S, Hibbert JM, Jain V, Singh N, Wilson NO SJ. 2011. CXCL10/IP-10 in infectious diseases pathogenesis and potential therapeutic implications. *Cytokine Growth Factor Rev* 22(3): 121-130.

Liu M, Guo S SJ. 2011. The emerging role of CXCL10 in cancer. *Oncol Lett* 2(4): 583-589.

Lucca LE, Hafler DA. 2017. Resisting fatal attraction : a glioma oncometabolite prevents $CD8^+$ T cell recruitment. *J Clin Invest* 127(4): 1218-1220.

Mah JTL, Low ESH, Lee E. 2011. In silico SNP analysis and bioinformatics tools : a review of the state of the art to aid drug discovery. *Drug Discov Today* 16: 800-809.

Maiti R, Domselaar GH Van, Zhang H, Wishart DS. 2004. SuperPose : a simple server for sophisticated structural superposition. *Nucleic Acids Res* 32: 590-594.

Meng Q, Yihong Z, George H.L. 2020. Targeting autophagy facilitates T lymphocyte migration by inducing the expression of CXCL10 in gastric cancer cell lines. *Front Oncol* 10: 886. https://doi.org/10.3389/fonc.2020.00886

Mignone F, Gissi C, Liuni S. 2002. Untranslated regions of mRNAs. *Genome Biol* 3. https://doi.org/10.1186/gb-2002-3-3-reviews 0004

Miller MP, Kumar S. 2001. Understanding human disease mutations through the use of interspecific genetic variation. *Hum Mol Genet* 10(21): 2319-28.

O'Brien J, Hayder H, Zayed Y, Peng C. 2018. Overview of microRNA biogenesis, mechanisms of actions, and circulation. *Front Endocrin* 9: 402. https://doi.org/10.3389/fendo.2018.00402

Pejaver V, Urresti J, Lugo-Martinez J, Pagel KA, Lin GN, Nam HJ, Mort M, Cooper DN, Sebat J, Iakoucheva LM, Mooney SD. 2017. MutPred2 : inferring the molecular and phenotypic impact of amino acid variants. *Nat Commun* 11: 5918. https://doi.org/10.1038/s41467-020-19669-x

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004. UCSF Chimera - a visualization system for exploratory research and analysis. *J Comput Chem* 25(13): 1605-1612.

Proudfoot AE, Fritchley S, Borlat F, Shaw JP, Vilbois F, Zwahlen C, Trkola A, Marchant D, Clapham PR, Wells TN. 2001. The BBXB motif of RANTES is the principal site for heparin binding and controls receptor selectivity. *J Biol Chem* 276(14): 10620-10626.

Sayers EW, Agarwala R, Bolton EE, Brister JR, Canese K, Clark K, Connor R, Fiorini N, Funk K, Hefferon T, Holmes JB. 2019. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 47(D1): D23-D28.

Shendure J, Akey JM. 2015. The origins, determinants, and consequences of human mutations. *Science* 349: 1478-1483.

Swaminathan GJ, Holloway DE, Colvin RA, Campanella GK, Papageorgiou AC, Luster AD, Acharya KR. 2003. Crystal structures of oligomeric forms of the IP-10/CXCL10 chemokine. *Structure* 11(5): 521-532.

Tokunaga R, Zhang WU, Naseem M, Puccini A, Berger MD, Soni S, McSkane M, Baba H, Lenz HJ. 2018. CXCL9, CXCL10, CXCL11/CXCR3 axis for immune activation a target for novel cancer therapy. *Cancer Treat Rev* 63: 40-47.

Venselaar H, Te Beek 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. https://doi.org/10.1186/1471-2105-11-548

Vinet J, De Jong EK, Boddeke HW, Stanulovic V, Brouwer N, Granic I, Eisel UL, Liem RS, Biber K. 2010. Expression of CXCL10 in cultured cortical neurons. *J Neurochem* 112(3): 703-714.

Wang LL, Li Y, Zhou SF. 2009. A bioinformatics approach for the phenotype prediction of nonsynonymous single nucleotide polymorphisms in human cytochromes P450. *Drug Metab Dispos* 37(5): 977-991.

Wang Z, Moult J. 2001. SNPs, protein structure, and disease. *Hum Mut* 17(4): 263-270.

Yang J, Richmond A. 2004. The angiostatic activity of interferon-inducible protein-10/CXCL10 in human melanoma depends on binding to CXCR3 but not to glycosaminoglycan. *Mol Ther* 9(6): 846-855.

Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. 2014. The I-TASSER suite: protein structure and function prediction. *Nat Methods* 12(1): 7-8.

Yin S, Ding F, Dokholyan NV. 2007. Eris: an automated estimator of protein stability. *Nat Methods* 4(6): 466-467.

Young E, Lee Z, Wook Y. 2009. CXCL10 and autoimmune diseases. *Autoimmun Rev* 8(5): 379-383.

Yuan Y, Weidhaas JB. 2018. Functional microRNA binding site variants. *Mol Oncol* 13(1): 4-8.

Yue P, Li Z MJ. 2005. Loss of protein structure stability as a major causative factor in monogenic disease. *J Mol Biol* 353(2): 459-473.

Zhao M, Ma Q, Xu J, Fu S, Chen L, Wang B, Wu J, Yang L. 2015. Combining CXCL10 gene therapy and radiotherapy improved therapeutic efficacy in cervical cancer HeLa cell xenograft tumor models. *Oncol Lett* 2: 768-772.