Screening for Deleterious Non-synonymous SNPs in Human *CCL21* Gene using In-Silico Analysis

Yasir Ali[1], Mehran Akhtar[2], Kainat Khan[1], Nadia Farooqi[1], Shahla Gohar[1], Syed Ishfaq Ahmad[3], Madeeha Ayaz[1], Zia ul Islam[1], Maria Arshad[4] and Fazal Jalil[1*]

[1] Department of Biotechnology, Abdul Wali Khan University Mardan, KPK, Pakistan.
[2] Department of Biotechnology, COMSATS University Islamabad, Abbottabad, Pakistan.
[3]Department of Biotechnology and Genetic Engineering, Hazara University Mansehra Pakistan.
[4] Department of Research & Innovation, Shalamar Institute of Health Sciences, Lahore, Pakistan.

***Corresponding Author's email:** fazaljalil@awkum.edu.pk

ABSTRACT

Rheumatoid arthritis (RA) is a chronic, systematic, and progressive inflammatory disorder, causing severe damage to joints and hence increase mortality. The *Chemokine (C-C motif) ligand 21* (*CCL21*), a member cytokines family, is involved in immuno-inflammatory and regulatory processes. Therefore, identifying the important SNPs (single nucleotide polymorphisms) in the *CCL21* gene is of key importance to evaluate their structural and functional significance and to discover novel therapeutic targets for immune-related diseases, including RA. In this study, we used in silico approaches for identifying the most damaging non-synonymous SNPs (nsSNPs), playing a significant structural and functional role in CCL21 protein. The primary tools used for this study included PROVEAN, SNPs&GO, SIFT and PolyPhen2. Other tools, its stability, Structure and functional effect as well as the conservation profile, were verified using I-Mutant, MutPred, and ConSurf. The site of post-translational modification also predicted. The 3-D modeling of proteins was carried out using I-TASSER which were then visualized in Chimera v1.11. Furthermore, the gene-gene interactions were predicted using STRING and gene MANIA. It was observed that the nsSNPs D30Y (*rs753133670*), I62N (*rs1170851787*), R75C (*rs759733358*), R75S (*rs776954599*) and A83V (*rs776954599*) were the most damaging nsSNPs in the *CCL21* gene. These nsSNPs might have a significant role in CCL2 protein's malfunctioning and possibly causing different autoimmune diseases including RA. Our study concluded that, to study the correlation of the *CCL21* gene with certain autoimmune disorders, i.e. Crohn's Disease (CD), RA and other immune-associated diseases, these SNPs could be the most important ones. In addition, these SNPs need to be studied in animal models and cell cultures in association with certain diseases, to identify if they could be of use for the gene therapy and pharmacogenomics.

Keywords: CCl21; nsSNPs; Polymorphisms; Gene-gene interactions; *In silico*.

Article History
Received: 11 January 2022
Accepted: 23 February 2022
Published: 25 February 2022

INTRODUCTION

Rheumatoid arthritis (RA) leads to inflammation in joints and articular cartilage coupled with synovial hyperplasia, thereby characterized as an autoimmune disease, and causes consistent pain and permanent disability of the patients' physical activities in normal life (Lindler, et al.,2020; Aundhia, et al., 2020). The pathogenesis of RA is still unclear but the incidence and prevalence of RA are said to be the result of different environmental and genetic risk factors (van der Woude, et al., 2018). On the basis of proposed data by different studies (family aggregation and twin consonance), RA was found heritable in 60% of patients. These variations indicate the role of genetic factors in the pathogenesis of RA (Zamanpoor, et al ., 2019, Jalil, et al., 20170). The roles of SNPs have been detected in the non-MHC genes, such as PTPN22 and MHC genes like HLA-DRB-1 which are potent and can drive inflammatory response in RA. Many studies revealed >150 SNPs in RA located at more than 70 gene loci (Jalil, et al., 2013-Issilbayeva, et al., 2021).

CCL21 gene belongs to the Chemokine family having a C-C motif and is located on chromosome 9p13.3. *CCL21* is the chemokine that binds to CCR7 and plays an important role by modulating the process of circulation in the lymphoid and peripheral organs of T cells as well as dendritic cells (Van Raemdonck, et al., 2020). In addition, defective movement of dendritic cells and lymphocytes into T zones has been demonstrated in CCR7 deficient mice (Chauveau, et al., 2020). Previous studies have shown that the endothelial cell growth factor CCL21 is mediated/expressed through the endothelial cell lymph node, which is related to tertiary lymphoid tissue development (Simmons, et al.,2019, Zhang, et al., 2020). Recent research on RA patients revealed an elevated expression of both the CCR7 and CCL21 in fibroblasts and macrophages of endothelial sub-lining cells and synovial tissues. Macrophages in synovial fluids of RA patients produce a high number of CCL21 than normal in vitro differentiated peripheral blood macro-phages (Rana, et al., 2018, Umar, et al., 2021).

In RA pathogenesis, the up-regulation of the *CCL21* gene in synovial tissue linings is dependent on the activation of the fibroblast and macrophages, which in turn produces proangiogenic factors, i.e. vascular endothelial growth factor (VEGF), interleukin-8 (IL-8), and Angiogenin1. The high expression of CCL21 gene has been observed in the sub-lining of endothelial cells than the peripheral ordinary blood cells (Yuan, L.-H., et al., 2017). Currently, research on RA-associated polymorphisms in *CCL21* gene is not only rare but also has diverse outcomes. These differences can be attributed to the small sample sizes and variations in genetic make-up (Wengner, A.M., et al., 2007).

Many studies have reported that nsSNPs in various genes like TAGAP, TOX3 and CCR6 may negatively impact protein structure and function and might be correlated to diseases like Breast cancer, RA, and others (Arshad, et al., 2018; Akhtar, et al., 2021).

METHODOLOGY

Recruiting nsSNPs

Data regarding the location, global minor allele frequencies (MAFs) and residual changes of SNPs in the *CCL21* gene were taken from NCBI dbSNP (Bhagwat, et al., 2010). After analyzing the data, 916 SNPs were searched, out of them, 96 nsSNPs were selected for further processing.

Finding the most Damaging CCL21 nsSNPs

Four bioinformatics tools, including (a) SIFT (Sorting Intolerant from Tolerant) (Kumar, et al., 2009), (b) PROVEAN (Protein Variation Effect Analyzer) (Choi, et al., 2012), (c) SNPs& (Capriotti, et al., 2011) and (d) PolyPhen2 (Polymorphism Phenotyping2) were applied to identify the deleterious nsSNPs (Adzhubei, et al., 2010). Further screening was done on the nsSNPs predicted by all the tools as likely deleterious or intolerant.

Effect of nsSNPs in CCl21 protein

To predict the structural as well as functional effects of the given deleterious nsSNPs on the protein product MutPred tool was used (Li, et al., 2009). An online web-based MutPred tool identifies alteration in the sequence of amino acid and predicts the root cause of the disease as well. This tool is mostly used for screening the physical and functional characteristics like phosphorylation site loss or propensity gains. Deleterious nsSNPs were selected from the FASTA sequence of the *CCL21* gene with $p < 0.05$ being deliberated as assured and $p < 0.01$ being considered as very assured.

The I-Mutant 2.0, a web based tool was applied to test the stability of protein. It shows stability variations in the altered protein and gives us predictions about RI starting from 0 to10. Where 0 is the lowest and 10 is the highest possible RI values. Further analysis was carried out on those nsSNPs which were found to reduce CCL21 protein stability. Each amino acid was identified by the use of ConSurf tool) based on evolutionary conservation. It also illustrates the phylogenetic relationships between homologous sequences. Further analysis was performed only for those sequences which were found highly conserved and which also showed more similarity with lethal nsSNPs.

Protein Modeling

In this study, I-TASSER was to modeled 3D structures of wild type and mutant CCL21 (Yang, et al., 2015). We used Chimera 1.11 to visualize the resultant protein structures and study molecular features. TM-align was used to analyze the structural differences, when the wild-type CCL21 protein was compared to the mutant CCL21 proteins. TM-align calculates root mean square deviation (RMSD), structural superposition and template modelling (TM) score. The structural similarity was predicted by TM score that ranges from 0 to1, where 1 shows higher resemblance. The variation depends upon the RMSD value, greater the RMSD value, higher will be the variation between wild and mutant structures (Zhang, et al., 2005).

Prediction of Possible Post Translational Modification (PTM) Sites

Different tools were applied for identification of possible PTM sites in CCL21 protein. GPS-MSP 3.0, an online tool, was applied to identify the methylation sites in CCL21 protein (Deng, et al., 2017) whereas, possible sites for phosphorylation in the protein were anticipated through NetPhos 3.1 and GPS 3.0. The phosphorylation sites were predicted at the threonine, tyrosine and serine positions of the CCL21 protein. NetPhos 3.1 predicted less specific results having a lower phosphorylation potential than GPS 3.1 (Xue, et al., 2008). In addition, ubiquitylation sites were predicted with the help of BDM-PUB and UbPred) tools (Radivojac, et al., 2010).

Gene-gene Interaction and Effect of Regulatory Region SNPs

The interaction and association of CCL21 with other proteins and its nsSNPs' effects on other proteins were studied by utilisng two in silico tools i,e., GeneMANIA and STRING (Warde-Farley, et al., 2010; Gasteiger, et al., 2003). The GeneMANIA

finds out the interaction between genes on the basis of protein interaction, protein domain similarity, co-expression, co-localization and pathways. STRING calculations are based on top 10 strongly interacting proteins for which the limitation includes co-expression, gene fusion, co-occurrence, biochemical and experimental data. It predicts collective score of each protein, which interacts with the target gene and ranges from 0 to 1, where 1 and 0 are the highest and lowest interactions, respectively. *CCL21* gene was submitted and interactions between genes were predicted.

MicroSNiPer and PolymiRTS Database were used to study whether these nsSNPs in *CCL21* gene has a role in gene regulation. MicroSNiPer particularly show whether or not the target SNPs containing region would impact miRNA. PolymiRTS database, which is a web-based server ensuring that variants in the UTR regions as well as in miRNA seed are affected by deleterious SNPs.

RESULTS

nsSNPs Recruitment

CCL21 gene has 916 SNPs; 96 and 23 are located in the 5'UTR and in the 3'UTR, respectively. In addition, other types of SNPs identified as splice sites, intronic, synonymous, and uncategorized. (Figure 1). Of the 96 nsSNPs selected for further analysis, 7 nsSNPs resulted in stop codons, thus affecting the protein structure directly. Seven of these nsSNPs can result in truncated protein that can potentially lead to disease.

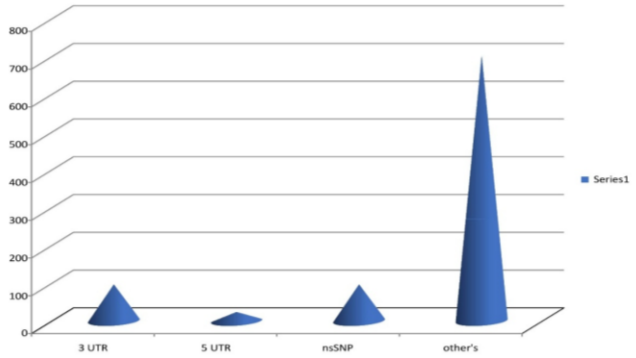

Figure 1: All SNPs in *CCL21* gene.

Characterization of Deleterious nsSNPs in CCL21 Gene

Four bioinformatics tools (PROVEAN, SIFT, SNPs&GO and polyPhen-2) were used in the analysis of 96 nsSNPs recruited from dbSNPs to identify the possible hazardous structural and functional outcomes of these nsSNPs on CCL21 protein. According to the PROVEAN, 25 nsSNPs were predicted deleterious effects while SIFT predicted total 39 nsSNPs as intolerable. Moreover, the SNPs&GO tool evaluated 37 nsSNPs in total as disease causing. Of these, 12 nsSNPs were declared damaging, to the final protein, by all three tools. These shortlisted nsSNPs were later subjected to PolyPhen2, which display its results in three categories i.e. benign, possibly and probably damaging. It predicted 7 nsSNPs as potentially damaging. It also gives a count between 0 and 1 where 1 is considered as the most damaging hence all further investigations were carried out for 7 nsSNPs having score 1 (Table 1).

Table 1: Common Deleterious nsSNPs in *CCL21* gene.

SNP ID	Amino acid position of SNP	Provean	SNP & Go		SIFT	Polypen-2
		Score	Probability	RI	Score	Score
rs779706400	L7P	-4.530	0.902	8	0.04	1
rs753133670	D30Y	-7.122	0.917	8	0.02	1
rs1453433779	R46C	-3.858	0.843	7	0.01	1
rs1170851787	I62N	-5.576	0.733	5	0.01	0.999
s759733358	C75R	-11.422	0.998	10	0.02	1
rs776954599	C75S	-9.736	0.997	10	0.02	1
rs1182863895	V83A	-3.911	0.977	10	0.00	0.999

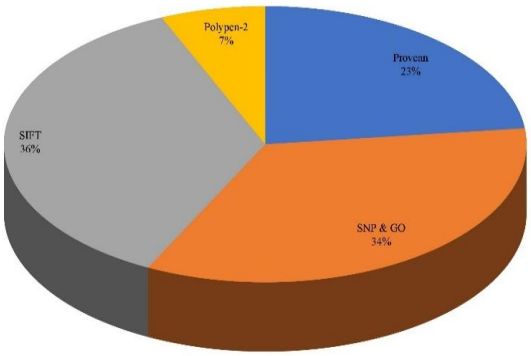

Figure 2: Percent Prediction of nsSNPs by SIFT, Provean, SNP&GO and Polypen-2

Effect of nsSNPs in CCL21 protein

The damaging effect of the finalized nsSNPs on the structure or function of CCL21 protein was predicted through the MutPred server. The results are given in Table 2. I-Mutant predicted the influence of the selected 7 nsSNPs on CCL21 protein stability. Each nsSNP was submitted separately for RI calculations (ranging from 0 to 10) to determine whether stability should be decreased/increased results are given in Table 3. Out of 7 shortlisted nsSNPs, 2 were shown to increase the protein's stability i.e., substitution with L7P (rs779706400) and R46C (rs1453433779) and hence were omitted for further study whereas rest of the 5 nsSNPs showed a decrease in stability of CCL21 protein (Table 3). These 5 nsSNPs were chosen for further investigation.

The conservation profile of these nsSNPs was predicted through the ConSurf tool which depicted C57R, C75S and V83A as highly conserved, exposed structural residues. The amino acid D30 was predicted to be highly conserved, exposed and functional residue, while the amino acid I62 was predicted to be buried. Retention scores for all the selected nsSNPs are depicted in Table 4. The results indicated that the nsSNPs with high conservation were most damaging for structure and function of CCL21 protein.

CCL21 and its Mutants protein 3D-modelling

The 5 nsSNPs which were found to be decreasing the stability of CCL21 protein were further selected for protein modeling. CCL21 proteins PDB file was generated by I-TASSER using the protein sequences of

Table 2. Effects of nsSNPs on structural & functional properties of *CCL21* by MutPred server

Mutation	Probability of deleterious mutation	Top Features	Affected PROSITE and ELM Motifs
L7P	0.862	Altered Signal peptide (p=1.5e-05)	ELME000053
		Altered Stability (p=0.04)	ELME000248
D30Y	0.839	Altered Transmembrane protein (P = 2.1e-04)	
		Loss of Disulfide linkage at C32 (P = 1.2e-03)	ELME000120
		Loss of Allosteric Site at Y84 (P = 0.04)	ELME000182
		Loss of catalytic site at C32 (P = 0.04)	
		Gain of Pyrrolidone carboxylic acid at Q29(P=0.02)	
		Altered signal peptide (P = 4.3e.03)	
R46C	0.219	-	None
I62N	0.894	Altered Stability (P=1.6e-03)	
		Altered Transmembrane protein (p=2.6e-04)	
		Altered Ordered interface (p=0.03)	ELME000276
		Gain of Disulfide linkage at C57 (p=9.9e-04)	
C75R	0.950	Altered Metal binding (p=1.7e-03)	
		Loss of Disulfide linkage at C57 (p=4.1e-04)	
		Altered Disordered interface (p=0.01)	
		Gain Of intrinsic (p=0.02)	None
		Gain of Catalytic site at C75 (p=0.4e-03)	
		Altered Transmembrane protein (p=0.01)	
		Altered Transmembrane protein (p=0.02)	
C75S	0.906	Altered Metal binding (p=1.3e-03)	
		Loss of Disulfide linkage at C57 (p=4.1e-04)	
		Altered Disordered interface (p=0.02)	ELME000085
		Gain of Catalytic site at C75 (p=0.01)	
		Altered transmembrane protein (p=0.01)	ELME000147
V83A	0.564	Gain of intrinsic disorder (p=0.417.2e-03)	
		Altered Disordered interface (p=0.320.02)	
		Altered transmembrane protein (p=0.01)	ELME000106
		Altered Coiled coil (P=0.04)	

Table 3. I-Mutant Results for the nsSNPs

SNP ID	SNP position	Stability	RI	SNP ID	Amino acid Change	Stability	RI
rs779706400	L7P	Increased		s759733358	C75R	Decreased	
rs753133670	D30Y	Decreased		rs776954599	C75S	Decreased	
rs1453433779	R46C	Increased		rs1182863895	V83A	Decreased	
rs1170851787	I62N	Decreased					

Table 4. Conservation profiling of the target amino acids, where the selected nsSNPs are located

SNP ID	Amino acid positions	Conservation Score	Prediction
rs753133670	D30	8	Exposed and Highly Conserved (f)
rs1170851787	I62	8	Buried
s759733358	C75	9	Exposed and Highly Conserved (s)
rs776954599	C75	9	Exposed and Highly Conserved (s)
rs1182863895	V83	9	Burried and Highly Conserved (s)

CCL21 protein (wild-type and mutants). I-TASER used templates 2co9 and 2nbiA (83% identity, 85% coverage of thread alignment). The PDB files for each mutant model along with their TM-scores and RMSD values are given in Table 5. Besides, the structural and molecular characterizations of the protein were analyzed using Chimera 1.11 (Figure2).

Predicted Post Translational Modifications Site (PTMs)

The structures and functions of proteins are controlled by PTMs, the key events in biological systems such as cell signaling and protein-protein interactions. In this study, we predicted the effects of selected nsSNPs on the PTMs of the CCL21 protein. Methylation allows PTM to occur in certain proteins, such as when lysine residues are methylated, it impacts the binding of the protein to DNA, which also influences gene expression. GPS-MSP 3.0 was used to determine CCL21 sites that shouldn't be methylated.

Table 5. RMSD values and TM score of 5 selected nsSNPs

SNP ID	Residual Change	TM Score	RMSD Values	SNP ID	Residual Change	TM Score	RMSD Values
rs753133670	D30Y	0.79644	2.9	rs776954599	C75S	0.82035	2.51
rs1170851787	I62N	0.78680	2.81	rs1182863895	V83A	0.83601	2.22
s759733358	C75R	0.8898	2.38				

Phosphorylation sites in the CCL21 protein were identified through the GPS 3.0 and NetPhos 3.1. In the GPS 3.0 21 residues (Thr: 26%, serine 63%, and Tyr: 11%) could be phosphorylated. While in Netphose 3.1 16 residues (Thr: 04, Ser: 10, Tyr: 02) could be phosphorylated. For comparison and selection of the normal residue based on GPS 3.0 and NetPhos 3.1 (Table 6).

The prediction of CCL21 ubiquitylation was performed by UbPred and BDM-PUB. CCL21 protein was predicted to have 15 possible sites for ubiquitylation, as detected by BDM-PUB, but UbPred identified just a single potential site of ubiquitylation. The results of ubiquitylation prediction in CCL21 protein are listed in the Table 7

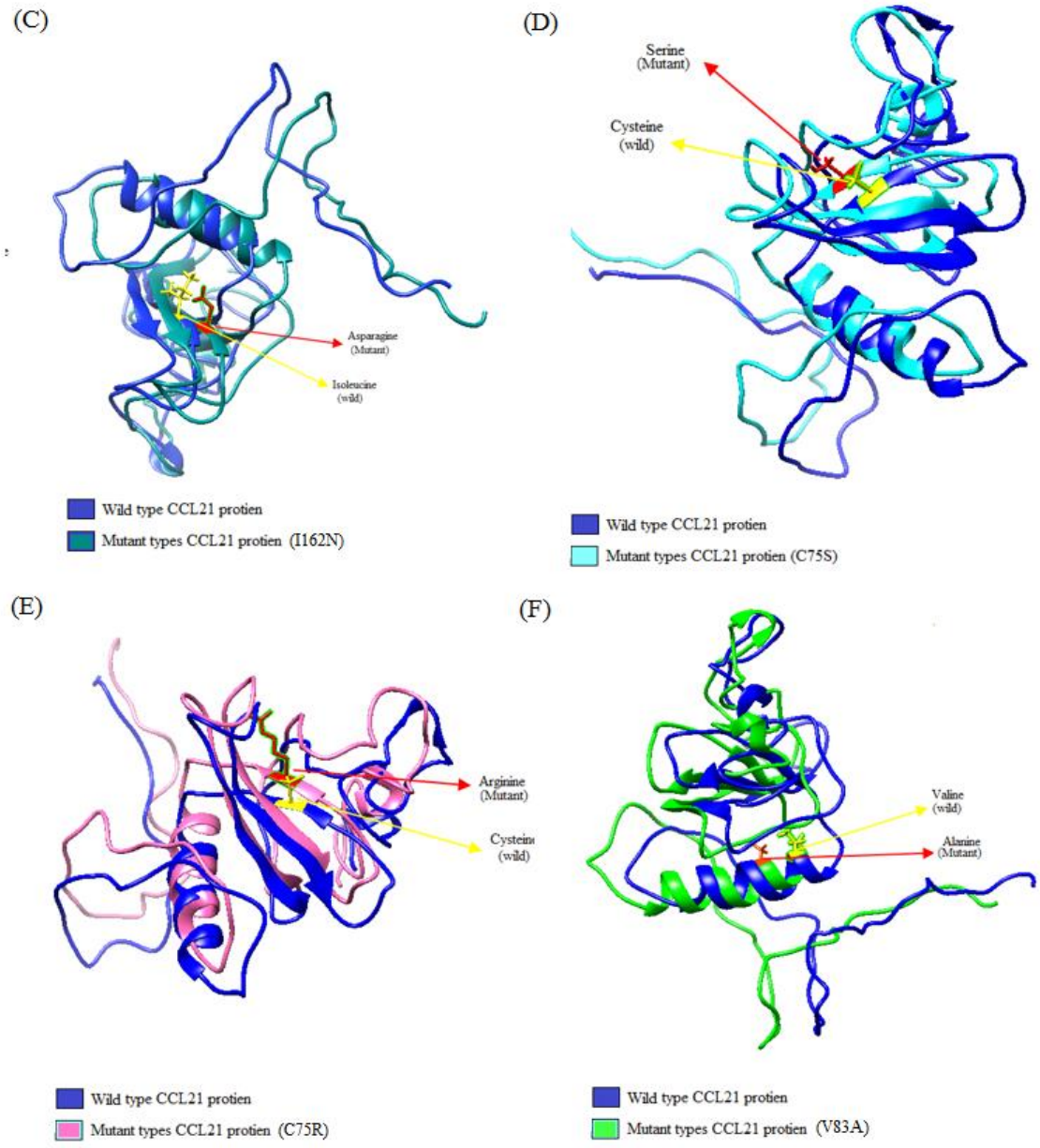

Figure. 3 A). Protein structure of CCL21 wild type, B) Superimposed Structure of D30Y mutant, C) I62N mutant, D) C75S mutant, E) C75SR mutant, and F) V83A mutant.

Gene-gene Interaction and regulatory SNPs in CCL21

The in-silico tool GeneMANIA depicted that the physical interactions of CCL21 are with CCR7, ACKR4, CCl14, IGFBP7, CXCL9 and CXCL13 and it also showed that the CCL21 protein is co-expressed with CCL14, TNFSF11, MS4A1, CXCL9, CXCR2, CCR7, CCL19, CXCR6 and CXCL9. In addition, it also indicated that CCL21 is co-localized with CCR6, CCL19, CXCL9 and ACKR4. The pathway of this protein has relation with CCR4, CXCR6, CXCR2, CCR10, CCR7, CCR8, CRR9,

Table 6. Phosphorylation prediction CCl21 protein

Position	Peptide	GPS3.0			NetPhos3.1	
		Kinase	score	cutoff	kinase	score
24	GIPRTQGSDGGAQDC	CK1/VRK/VRK2	49.038	48.397	DNAPK	.498
128	GCKRTERSQTPKGP*	AGC/GRK/BARK	186.523	65.626	PKC	0.723

Table 7. Ubiquitylation sites prediction of CCL21 protein

Peptides	Position	Score of BDM-PUB (Threshold 0.3)	Scorefrom UbPred (Threshold 0.62)
GAQDCCLKYSQRKIP	34	0.88	Not Ubiquitylated
CLKYSQRKIPAKVVR	39	0.36	Not Ubiquitylated
SQRKIPAKVVRSYRK	43	2.48	Not Ubiquitylated
KVVRSYRKQEPSLGC	50	0.90	Not Ubiquitylated
AILFLPRKRSQAELC	68	2.47	Not Ubiquitylated
QLMQHLDKTPSPQKP	92	0.51	0.77
DKTPSPQKPAQGCRK	98	1.09	Not Ubiquitylated
KPAQGCRKDRGASKT	105	2.84	Not Ubiquitylated
RKDRGASKTGKKGKG	111	3.41	Not Ubiquitylated

CRR3 and C5AR. Moreover, CCL21 was found to have no genetic interactions with other genes. In STRING Predictions, CCL19 is the most interactive gene with CCL21. The presence of CCL21, CCL20, CCL6 and CCR7 are associated with disease in many diseases such as Rheumatoid arthritis, (Viatte et. al., 2013) depicting its significance. Th prediction of gene-gene interactions by STRING and GeneMANIA are presented in Figure 6 and 7, respectively.SNPs situated in UTR of the gene not only can impact the miRNA binding sites but also can affect the half-life of RNA and translation of mRNA. Six SNPs in the CCL21 UTR region were identified by MicroSNiPER that might affect the binding of miRNA, whereas PolymiRTS predicted another SNP to affect miRNA binding (total 7 SNPs).

CCL21 SNPs have been associated with autoimmune thyroid disease (AITD) and rheumatoid arthritis (RA) in several studies (Li, et al., 2017; Myrthianou, et al., 2017). Although several nsSNPs are probably neutral and have little functional effects, many of these nsSNPs have been predicted to be deleterious because due to the disruption of functional sites in proteins or effects on protein's folding. Therefore, the effects of *CCL21* gene SNPs and their association with diseases are critical. The present study showed the result using in silico analysis to determine which nsSNPs are deleterious and what impact they have on the CCL21 protein.

DISCUSSION

In this study, the SNPs of *CCL21* gene were analyzed, which might have a critical role in autoimmune diseases, including RA. The

CCL21 gene has only 10% has nsSNPs and 5'UTR SNPs, while the 3'UTR SNPs comprise 3% of the total. In this study we did not include any other SNPs that

Table 8. Prediction of effected SNPs in CCL21 protein regulatory region

SNP	miR ID	Score change*
rs191058724	hsa-miR-331-5p, hsa-miR-4678, hsa-miR-6509-5p → No pattern	-0.379
Rs1801937	hsa-miR-6077, hsa-miR-6830-5p, → hsa-miR-1910-3p hsa-miR-2467-3p, hsa-miR-3714 hsa-miR-4257 hsa-miR-6511a-5p	-0.162
Only by PolymiRTS rs41305333	hsa-miR-342-5p, hsa-miR-4651, hsa-miR-4664-5p, hsa-miR-608, hsa-miR-6134, hsa-miR-6737-5p, hsa-miR-6742-5p, hsa-miR-6812-5p, hsa-miR-6819-5p, hsa-miR-92a-2-5p → hsa-let-7a-5p hsa-let-7b-5p hsa-let-7c-5p hsa-let-7d-5p hsa-let-7e-5p hsa-let-7f-5p hsa-let-7g-5p hsa-let-7i-5p hsa-miR-4458 hsa-miR-4500 hsa-miR-98-5p	-0.179
rs147542625	hsa-miR-1293, hsa-miR-342-5p, hsa-miR-4483, hsa-miR-4651, hsa-miR-4664-5p, hsa-miR-608, hsa-miR-6737-5p, hsa-miR-6747-5p, hsa-miR-6812-5p, hsa-miR-6819-5p, hsa-miR-6890-5p→ hsa-miR 1245b-3p hsa-miR-6134 hsa-miR-7854-3p	-0.263
rs11574916	hsa-miR-1293, hsa-miR-342-5p, hsa-miR-4483, hsa-miR-4651, hsa-miR-4664-5p, hsa-miR-608, hsa-miR-6737-5p, hsa-miR-6747-5p, hsa-miR-6812-5p, hsa-miR-6819-5p, hsa-miR-6890-5p, hsa-miR-876-3, → hsa-miR-122-5p hsa-miR-140-5p hsa-miR-4534 hsa-miR-504-3p hsa-miR-6514-5p hsa-miR-8082	-0.263
rs140038884	hsa-miR-342-5p, hsa-miR-4651, hsa-miR-4664-5p, hsa-miR-608, hsa-miR-6737-5p, hsa-miR-6747-5p, hsa-miR-6812-5p, hsa-miR-6819-5p, hsa-miR-6890-5p → No pattern	-0.15

belonging to other types of SNPs. Thus, very few SNPs affecting the CCL21 protein directly

All the tools, used to identify the most damaging nsSNPs, agreed on six nsSNPs that were predicted to have deleterious effect. PROVEAN predicted C75R with -11.422 as the highest score, while R46C with -3.858 as the lowest score among the selected nsSNPs. SNP&GO predicted C75S as the most damaging SNPS having the score of 0.997 as the highest. SIFT predicted L7P with the highest score of 0.04 and R46C, 162N, and V83A with the lowest score of 0.01. Polyphen-2 predicted five nsSNPs with score of 1 on the scale of 0 to1 and included L7P, D30Y, R46C, C75R, and C75S.

The most damaging nsSNPs, which were selected in this study, were cross-checked several tools in Ensemble genome browser 96, such as Mutation Assessors, CADD, REVEL, and MetalR. CADD, REVEL, and MetalR identified all the selected SNPs (6 nsSNPs) as deleterious and damaging. Mutation Assessor predicted five SNPs as harmful. CADD scores for these nsSNPs were as follows; 18 for L7P and R46S, 26 for I62N, 25 for C75R 25, and 24 for D30Y, C75S, and V83A (CADD score of 30 and 20 mean the SNP is among 0.1% and 1% of the most damaging SNPs in human genome, respectively). In MutPred1.2, several features are predicted including

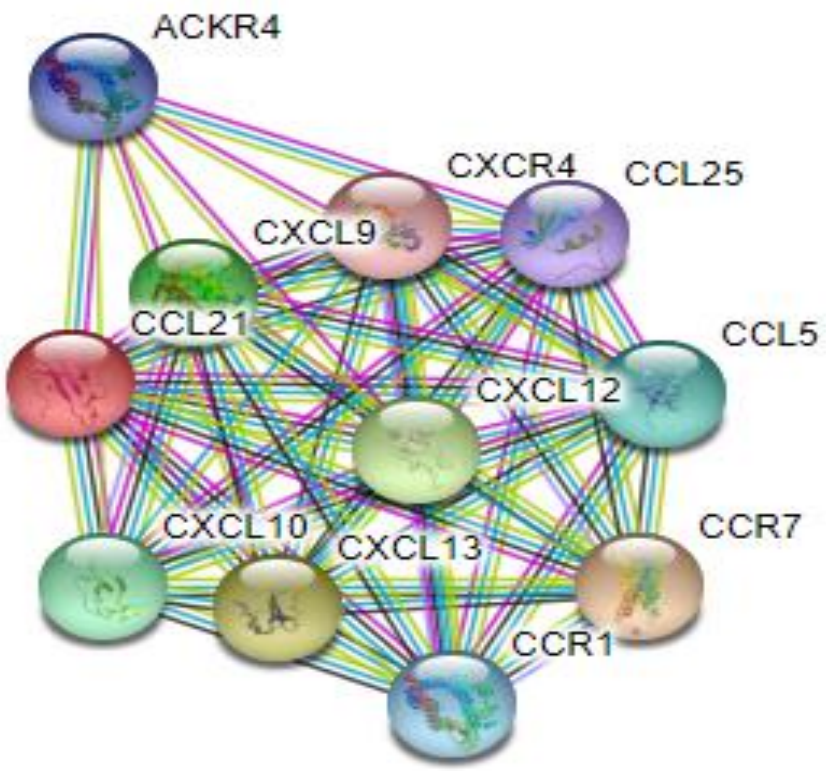

Figure. 5: Gene-gene interactions of *CCl2* by STRING.

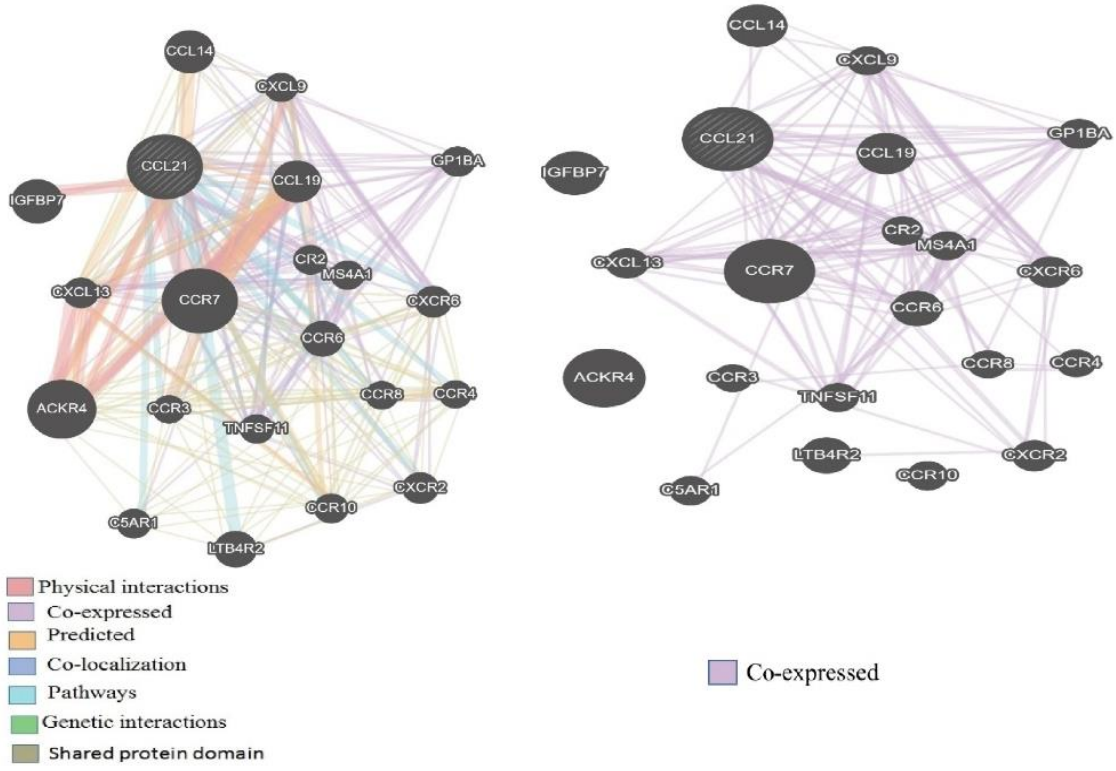

Figure 6: GeneMANIA Predicted *CCl21* Gene-gene interaction with other genes

acetylation gain, Methylation losses, altered interfaces, and intrinsic disorder gain. The C75R and C75S SNPs had the highest *P*-values of 0.950 and 0.906 and followed by L7P, D30Y, and I62N with 0.862, 0.839, and 0.894 respectively. The lowest *P*-values (0.219 and 0.564) were for SNPs R46C and V83A, respectively. The findings suggest that these nsSNPs might affect the CCl21 protein.

For predicting the protein stability, I-Mutant was applied which shows the effects of the nsSNPs. The results are displayed in the form of RI values ranged from 0 to 10, representing the minimum to maximum reliabilities, respectively. The predicted four SNPs that cause lower the protein stability are D30Y, I62N, C75R, and V83A. While the remaining SNPs L7P and R46N were identified to have increased the stability of protein (Table 3). Thus, CUPSAT Server (http:/cupsat.tu-bs.de/) was used to cross-check these results to ensure the reliability of the predictions. The I-Mutant prediction was in closed agreement with UPSAT server predictions. In ConSurf, the highly conserved amino acids are predicted to be critical to the protein structure or function as they are based upon amino acid positions in the protein. The structural or functional importance of amino acids is predicted based on the amino acid role in protein (Berezin, et al., 2004). The more the amino acid are conserved in the protein, the more they are considered to have many important function, such as interactions. As acknowledged by Miller and Kumar, nsSNPs that are located in conserved regions are the most damaging (Miller, et al., 2008). The amino acid positions of the 5 selected nsSNPs were focused on, among which, the D30 was functionally active residue, while the amino acids C75 and V83A were highly conserved, exposed, and structurally active residue and the I62 amino acid was highly conserved and buried. These nsSNPs further confirmed the adverse effect on these selected nsSNPs of CCl21 protein. The wild type and mutated protein structures of CCl21 protein were modeled using I-TASSER, to which FASTA protein sequences were submitted. It used two template sets, 2nbiA and 2co9, which had 85% and 83% similarity. RAMPAGE values were calculated for the protein structures (using RAMPAGE Server) (Lovell, et al., 2003) if they were higher than 80%, then the protein structure was deemed reliable. The RAMPAGE value for the wild type CCl21 protein was 84.8% favored and 15.2% allowed residues. The favored and allowed residues for the mutant D30Y structure were 82.9% and 17.1%, respectively. For the I62N, C73R, C73S, and V83A, the favored residues were 82.6%, 83.2%, 81.2%, and 82.4%, respectively, while the allowed residues were 17.4%, 16.8%, 18.8%, and 16.6% respectively (Morris et al, 1992). All the five selected nsSNPs had high root mean square deviations, indicated that these nsSNPs were the most deleterious SNPs in *CCl21* gene. When the RMSD value is greater than 2Å, the differences between mutant and wild-type proteins are greater.

Posttranscriptional modifications (PTMs) are the major factors that drive proteins to perform important functions, such as PPIs and cell signaling (Dai, et al., 2010; Shiloh, et al., 2013). Among the PTM sites in CCl21 protein, the occurrence of PTM sites at these nsSNP positions was investigated. It was noteworthy that none of the phosphorylation and ubiquityation sites were located at the most damaging nsSNPs sites. Other nsSNPs positions were also investigated using GPS3.0 and NetPhos3.1, which showed 7 phosphorylation sites at the nsSNPs positions (Table 4). Ubiquitylation sites were predicted by BDM-PUB and UbPred. BDM-PUB detected 15 sites and UbPred predicted one ubiquitylation site at 92. The STRING and GeneMANIA predictions indicated that CCL19 was the most interactive protein

with CCl21. However, RA is associated with many proteins, such as CCL20, CCL21, CCR7 and TRAF3IP3, indicating the importance of CCL21 protein in RA (Viatte, Plant &Raychaudhuri 2013). Thus, it can be inferred that any of the most destructive nsSNPs in the gene *CCL21* might ultimately affect and interrupt the normal function of other expressive genes based on their interaction patterns and their co-pression profile with many proteins, such as CCL20, CCL21, CCR7, and TRAF3IP3.

CONCLUSIONS

This study suggests that various nsSNPs can disturb the structure and/or function of CCL21 protein. In native protein of CCL21 gene, five major mutations found were Aspartic acid to Tyrosine at position 30 (rs753133670), Isoleucine to Asparagine at position 62 (rs1170851787), Cysteine to Arginine at position 75 (rs759733358), Cysteine to Serine at position 75 (rs776954599) and Valine to Alanine at position 83 (rs1182863895). These results supported by high MutPred 1.2 and high RMSD P values that also indicate that these nsSNPs have a significant role in the onset of RA and other related diseases. Thus, these SNPs can be considered as critical candidates for the development of diseases associated with CCl21 dysfunction, ultimately helping to discover effective drugs and create precision medicine. Laboratory experiments are required to investigate the effect of these polymorphisms on protein structure and function. Furthermore, the use of various animal for studying CCL21 mutation can be very helpful when studying disease development.

REFERENCES

Adzhubei, I.A., et al., A method and server for predicting damaging missense mutations. Nature methods, 2010. **7**(4): p. 248-249.

Akhtar, M., et al., An in silico approach to characterize nonsynonymous SNPs and regulatory SNPs in human TOX3 gene. Journal of genetics, 2019. **98**(5): p. 1-10.

Akhtar, M., Ali, Y., Islam, Z. U., Arshad, M., Rauf, M., Ali, M., ... & Jalil, F. (2021). Characterization of Rheumatoid Arthritis Risk-Associated SNPs and Identification of Novel Therapeutic Sites Using an In-Silico Approach. *Biology*, *10*(6), 501.

Akhtar, M., Jamal, T., Jamal, H., Din, J. U., Jamal, M., Arif, M., ... & Jalil, F. (2019). Identification of most damaging nsSNPs in human CCR6 gene: In silico analyses. *International journal of immunogenetics*, *46*(6), 459-471.

Akhtar, M., Jamal, T., Hayat, C., Rauf, M., Khan, R. S., Shah, A. A., ... & Jalil, F. (2019). An in silico approach to characterize nonsynonymous SNPs and regulatory SNPs in human TOX3 gene. *Journal of Genetics*, *98*(5), 1-10.

Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., ... & Sunyaev, S. R. (2010). A method and server for predicting damaging missense mutations. *Nature methods*, *7*(4), 248-249.

Ali, Y., Khan, S., Chen, Y., Farooqi, N., Islam, Z. U., Akhtar, M., ... & Jalil, F. (2021). Association of AFF3 Gene Polymorphism rs10865035 with Rheumatoid Arthritis: A Population-Based Case-Control Study on a Pakistani Cohort. *Genetics Research*, *2021*.

Arshad, M., A. Bhatti, and P. John, Identification and in silico analysis of functional SNPs of human TAGAP protein: A comprehensive study. PloS one, 2018. **13**(1): p. e0188143.

Aslam, M. M., John, P., Fan, K. H., Bhatti, A., Aziz, W., Ahmed, B., ... & Kamboh, M. I. (2020). Investigating the GWAS-implicated loci for rheumatoid arthritis in the Pakistani population. *Disease markers*, *2020*.

Aundhia, C., Patel, S., Shah, N., Parmar, G., & Seth, A. (2020). Psychological effects and management of rheumatoid arthritis.

Berezin, C., Glaser, F., Rosenberg, J., Paz, I., Pupko, T., Fariselli, P., ... & Ben-Tal, N. (2004). ConSeq: the identification of functionally and structurally important residues in protein sequences. *Bioinformatics*, *20*(8), 1322-1324.

Bhagwat, M., Searching NCBI's dbSNP database. Current protocols in bioinformatics, 2010. **32**(1): p. 1.19. 1-1.19. 18.

Bhattacharya, A., J.D. Ziebarth, and Y. Cui, PolymiRTS Database 3.0: linking polymorphisms in microRNAs and their target sites with human diseases and biological pathways. Nucleic acids research, 2014. **42**(D1): p. D86-D91.

Capriotti, E. and R.B. Altman, Improving the prediction of disease-related variants using protein three-dimensional structure. BMC bioinformatics, 2011. **12**(4): p. 1-11.

Capriotti, E., P. Fariselli, and R. Casadio, I-Mutant2. 0: predicting stability changes upon mutation from the protein sequence or structure. Nucleic acids research, 2005. **33**(suppl_2): p. W306-W310.

Chauveau, A., Pirgova, G., Cheng, H. W., De Martin, A., Zhou, F. Y., Wideman, S., & Arnon, T. I. (2020). Visualization of T cell migration in the spleen reveals a network of perivascular pathways that guide entry into T zones. *Immunity*, *52*(5), 794-807.

Choi, Y., et al., Predicting the functional effect of amino acid substitutions and indels. 2012.

Dai, C. and W. Gu, p53 post-translational modification: deregulated in tumorigenesis. Trends in molecular medicine, 2010. **16**(11): p. 528-536.

Deng, W., Wang, Y., Ma, L., Zhang, Y., Ullah, S., & Xue, Y. (2017). Computational prediction of methylation types of covalently modified lysine and arginine residues in proteins. *Briefings in Bioinformatics*, *18*(4), 647-658.

Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D., & Bairoch, A. (2003). ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic acids research*, *31*(13), 3784-3788.

Issilbayeva, A., Kushugulova, A., Meiramova, A., Kozhakhmetov, S., Akhmetova, Z., Nurgaziyev, M., ... & Ainabekova, B. (2021). Epidemiological Trends of Rheumatoid Arthritis and PADI4, PTPN22, and HLA-DRB9 Genes Distribution in the Kazakhstan Population. *Open Access Macedonian Journal of Medical Sciences*, *9*(B), 747-757.

Jalil, F., Arshad, M., Bhatti, A., Jamal, M., Ahmed, M., Malik, J. M., ... & John, P. (2017). Progression pattern of rheumatoid arthritis: A study of 500 Pakistani patients. *Pakistan journal of pharmaceutical sciences*, *30*(4), 1219-23.

Jalil, S.F., et al., Replication of european rheumatoid arthritis loci in a Pakistani population. The Journal of Rheumatology, 2013. **40**(4): p. 401-407.

Kumar, P., S. Henikoff, and P.C. Ng, Predicting the effects of coding non-synonymous variants on protein function

using the SIFT algorithm. Nature protocols, 2009. **4**(7): p. 1073.

Li, G., Zhao, J., Li, B., Ma, J., Zhao, Q., Wang, X., ... & Liu, J. (2017). Associations between CCL21 gene polymorphisms and susceptibility to rheumatoid arthritis: a meta-analysis. *Rheumatology International*, *37*(10), 1673-1681.

Li, B., Krishnan, V. G., Mort, M. E., Xin, F., Kamati, K. K., Cooper, D. N., ... & Radivojac, P. (2009). Automated inference of molecular mechanisms of disease from amino acid substitutions. *Bioinformatics*, *25*(21), 2744-2750.

Lindler, B. N., Long, K. E., Taylor, N. A., & Lei, W. (2020). Use of herbal medications for treatment of osteoarthritis and rheumatoid arthritis. *Medicines*, *7*(11), 67.

Lohmueller, K. E., Indap, A. R., Schmidt, S., Boyko, A. R., Hernandez, R. D., Hubisz, M. J., ... & Bustamante, C. D. (2008). Proportionally more deleterious genetic variation in European than in African populations. *Nature*, *451*(7181), 994-997.

Lovell, S. C., Davis, I. W., Arendall III, W. B., De Bakker, P. I., Word, J. M., Prisant, M. G., ... & Richardson, D. C. (2003). Structure validation by Cα geometry: ϕ, ψ and Cβ deviation. *Proteins: Structure, Function, and Bioinformatics*, *50*(3), 437-450.

Miller, M.P. and S. Kumar, Understanding human disease mutations through the use of interspecific genetic variation. Human molecular genetics, 2001. **10**(21): p. 2319-2328.

Myrthianou, E., Zervou, M. I., Budu-Aggrey, A., Eliopoulos, E., Kardassis, D., Boumpas, D. T., ... & Goulielmos, G. N. (2017). Investigation of the genetic overlap between rheumatoid arthritis and psoriatic arthritis in a Greek population. *Scandinavian Journal of Rheumatology*, *46*(3), 180-186.

Radivojac, P., Vacic, V., Haynes, C., Cocklin, R. R., Mohan, A., Heyen, J. W., ... & Iakoucheva, L. M. (2010). Identification, analysis, and prediction of protein ubiquitination sites. *Proteins: Structure, Function, and Bioinformatics*, *78*(2), 365-380.

Rana, A. K., Li, Y., Dang, Q., & Yang, F. (2018). Monocytes in rheumatoid arthritis: Circulating precursors of macrophages and osteoclasts and, their heterogeneity and plasticity role in RA pathogenesis. *International immunopharmacology*, *65*, 348-359.

Shiloh, Y. and Y. Ziv, The ATM protein kinase: regulating the cellular response to genotoxic stress, and more. Nature reviews Molecular cell biology, 2013. **14**(4): p. 197-210.

Simmons, S., et al., High-endothelial cell-derived S1P regulates dendritic cell localization and vascular integrity in the lymph node. Elife, 2019. **8**: p. e41239.

Umar, S., Palasiewicz, K., Van Raemdonck, K., Volin, M. V., Romay, B., Ahmad, I., ... & Shahrara, S. (2021). CCL25 and CCR9 is a unique pathway that potentiates pannus formation by remodeling RA macrophages into mature osteoclasts. *European Journal of Immunology*, *51*(4), 903-914.

van der Woude, D. and A.H. van der Helm-van, Update on the epidemiology, risk factors, and disease outcomes of rheumatoid arthritis. Best practice & research Clinical rheumatology, 2018. **32**(2): p. 174-187.

Van Raemdonck, K., S. Umar, and S. Shahrara, The pathogenic importance of

CCL21 and CCR7 in rheumatoid arthritis. Cytokine & Growth Factor Reviews, 2020. **55**: p. 86-93.

Warde-Farley, D., Donaldson, S. L., Comes, O., Zuberi, K., Badrawi, R., Chao, P., ... & Morris, Q. (2010). The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic acids research*, *38*(suppl_2), W214-W220.

Wengner, A. M., Höpken, U. E., Petrow, P. K., Hartmann, S., Schurigt, U., Bräuer, R., & Lipp, M. (2007). CXCR5-and CCR7-dependent lymphoid neogenesis in a murine model of chronic antigen-induced arthritis. *Arthritis & Rheumatism: Official Journal of the American College of Rheumatology*, *56*(10), 3271-3283.

Xue, Y., Ren, J., Gao, X., Jin, C., Wen, L., & Yao, X. (2008). GPS 2.0, a tool to predict kinase-specific phosphorylation sites in hierarchy. *Molecular & cellular proteomics*, *7*(9), 1598-1608.

Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., & Zhang, Y. (2015). The I-TASSER Suite: protein structure and function prediction. *Nature methods*, *12*(1), 7-8.

Yuan, L. H., Chen, X. L., Di, Y., & Liu, M. L. (2017). CCR7/p-ERK1/2/ VEGF signaling promotes retinal neovascular-ization in a mouse model of oxygen-induced retinopathy. *International Journal of Ophthalmology*, *10*(6), 862.

Zamanpoor, M., The genetic pathogenesis, diagnosis and therapeutic insight of rheumatoid arthritis. Clinical genetics, 2019. **95**(5): p. 547-557.

Zhang, S., Wang, H., Xu, Z., Bai, Y., & Xu, L. (2020). Lymphatic Metastasis of NSCLC Involves Chemotaxis Effects of Lymphatic Endothelial Cells through the CCR7–CCL21 Axis Modulated by TNF-α. *Genes*, *11*(11), 1309.

Zhang, Y. and J. Skolnick, TM-align: a protein structure alignment algorithm based on the TM-score. Nucleic acids research, 2005. **33**(7): p. 2302-2309.