Association of SNPs with Rheumatoid Arthritis Susceptibility
Shakeel Safdar*1, Mahmood Kayani1, Hina Iqbal1, and Sohail Safdar2

1Department of Biosciences, COMSATS Institute of Information Technology, Islamabad, Pakistan
2Ahlia University, Bahrain

*For correspondence: shakeel.safdar1984@hotmail.com

ABSTRACT
Rheumatoid Arthritis is one of the types of Arthritis in which the joints swell up. It has been found that there are approximately 340 genes involved in causing the Rheumatoid Arthritis. Moreover, it has also been observed that SNPs of these genes are associated with the disease. In this study, four genes were analyzed for their possible association with Rheumatoid Arthritis using bioinformatics tools. The results showed how SNPs are associated with this disease. Some of the SNPs brought some structural changes in particular proteins while others SNPs produced no change in the structure. SNPs that made the proteins abnormal and brought some structural changes in the protein can be considered as SNPs associated with Rheumatoid Arthritis, as by change of normal amino acid, the orientations of proteins can be disturbed and protein are not involved in specific pathways thus being pathogenic. In some cases, the structural change leads to disruption of the protein structure while sometimes, the structure remains unchanged i.e. due to synonymous mutations. This is due to a reason that sometimes, due to SNPs the abnormal amino acid is coded, which can either disrupt or mutate the whole structure. This fact can be supported by an evidence that the hydrophobic and hydrophilic amino acids provide specific orientations to proteins and whenever some polar amino acid is replaced by non-polar amino acid or some non-polar amino acid is replaced by polar amino acid or some cyclic amino acid is replaced by non-cyclic amino acid, the abnormality occurs, which can lead to disruption of protein structure making it abnormal. In this study, the changes in protein structure by SNPs are observed and such SNPs are predicted to be involved in causing Rheumatoid Arthritis.

Keywords: Rheumatoid Arthritis, SNPs, Mutation, Genes, Proteins, Nucleotides, MIA, CST3, CTLA4, HMGB1

INTRODUCTION
The name Rheumatoid Arthritis is derived from the Greek word rheumatos, means "flowing" which initially gave rise to the term "rheumatic fever", an illness that can follow throat infections and includes joint pain. The suffix -oid means "resembling", i.e. resembling rheumatic fever. Arthr means "joint" and the suffix -itis, a "condition involving inflammation". Thus Rheumatoid Arthritis is a
form of joint inflammation, which resembles rheumatic fever. Rheumatoid Arthritis appears to have been described in paintings more than a century before the first detailed medical description of the condition in 1800 (Azaola, 2013).

Rheumatoid arthritis (RA) is believed to be an ‘inflammatory autoimmune disorder’, in which the joints get severely affected (Voll et al., 2008). It is such a condition, in which the inflammation and joint pains destruct the joints, hence resulting in restriction of joint mobility (William et al., 2004). RA is a systemic disease, often affecting ‘extra-articular tissues’ all over the body, which includes blood vessels, skin, lungs, heart, and the muscles (Cojocaru et al., 2010). Rheumatoid Arthritis normally affects the joints but sometimes it affects the organs as well. Normally small joints are affected but larger joints can also be affected by the Rheumatoid Arthritis (Wilke, 2010).

Rheumatoid arthritis affects women three times more than men and it can develop at any age. Women aged 40 to 50 years usually suffer from this disease. Moreover, Rheumatoid Arthritis occurs 4 times more in smokers than the non-smokers (Williams et al., 2004). The research work in Bioinformatics made it evident that certain change at genetic level is one of the major causes of Rheumatoid Arthritis. Significantly, Single Nucleotide Polymorphism (SNP) is the reported to be one of the major causes of Rheumatoid Arthritis (Saad et al, 2016).

The cause of Rheumatoid Arthritis is a very active area of worldwide research. Genes and their SNPs have been found to be involved in Rheumatoid Arthritis, and in addition, the scientists believe that the environmental factors are also involved in causing Rheumatoid Arthritis. Moreover, it is also believed that smoking tobacco also increases the severity of Rheumatoid Arthritis (Baka et al., 2009).

Many studies have reported the involvement of genes and SNPs in the pathogenesis of the disease, some of which are mentioned here (Mattyasouszky et al., 2006; Kochil et al., 2004; Arlehaq et al., 2005; Vazgiourakis et al., 2007; Shimada et al., 2007). The above-mentioned studies signify the involvement of SNPs of various genes in Rheumatoid Arthritis. We have conducted experiments on MIA, CST3, CTLA 4 and HMGB 1 gene. About CST3, Gauthier (2012) signified the association of CST3 gene with Rheumatoid Arthritis. Gang Li (2014) mentioned in his research work that certain polymorphism in the nucleotides in CTLA4 gene has signified its association with the Rheumatoid Arthritis. Earlier, Gonzalez (1994) also mentioned in his research that how CTLA4 is possibly linked with Rheumatoid Arthritis. Lie (2005) highlighted CTLA4’s association with RA in Chinese population. About HMGB1, Armandis (2010) and Voll (2008) elaborated that they had found certain elevated range of HMGB1 genes in the synovial fluid samples of the patients with Rheumatoid Arthritis (Armandis, 2010). Earlier, Bosserhoff (2002), and then
Yeremenko and his research team, highlighted MIA gene’s association with the Rheumatoid Arthritis (Bosserhoff, 2002; Yeremenko et al, 2013).

**MATERIALS AND METHODS**

In this study, SNPs of Four genes were analyzed namely MIA, CST3, CTLA4, and HMGB1. The main objective of the project was to observe the structural changes in the protein structures, which are mentioned by NCBI to be involved in causing RA. However, in order to attain that objective, there were some minor objectives i.e. searching for SNPs using SNP database, selection of SNPs present in Exon region of DNA sequence, identification of mutation causing SNPs, filtration of synonymous from nonsynonymous SNPs and to find the particular region of SNP in gene sequence.

Various databases and bioinformatics tools were applied in order to attain the minor objectives leading to major objective. The gene sequences, protein sequences and the SNP region sequences were retrieved by NCBI. The tools were applied further to locate the exact SNP region in the gene and protein sequences. The initial step was to align the gene sequence with the SNP region sequence to locate the exact SNP position in the gene sequence. The applied tool was ClustalW®, which is a multiple sequence alignment tool. Higgins (1988) in his research work has elaborated that how multiple sequence pairwise alignments can better initiate a gene/protein similarity study. As we had to do the sequence alignments for analyzing the Single Nucleotide Polymorphism, so we used ClustalW® for better sequential analysis. The next step was to observe the replaced amino acid position in normal protein sequence. As the SNP region sequences were only available in nucleotide sequence form, so they needed to be translated in order to align with normal protein sequence. So for this purpose, TranSeq was applied which can translate nucleotide sequence into protein sequence. After translation, the translated SNP region sequence and normal protein sequence were aligned by using a protein alignment tool Tcoffee. The alignment sometimes gave the signified result, but sometimes the alignments were dispersed or scattered. The reason was that different protein translation tools translates nucleotide sequences by using different frames, so shattered alignments can be observed in such cases. After the application of Tcoffee, the next step was towards the major objective i.e Structural analysis through structure modeling. The tools that were applied for structural analysis were Swiss Model server and Swiss PDB viewer. The Swiss model takes the input sequence of protein and models it. The normal protein sequences were manually mutated according to the mutation mentioned by NCBI and then the structures were modeled. In the case of MIA, when the 17th SNP mutation was manually inserted in the normal protein sequence, the structure was not modeled which indicated towards disruption of protein structure due to the insertion of abnormal amino acid in protein sequence. The Swiss Model server models the protein structure on homology basis i.e. it searches the
relevant or homolog protein in database and designs the homology modeling. As the structure was not designed in case of MIA 17th SNP, so we predict that there is no such protein present in protein database, which resembles with mutated protein having the abnormal amino acid, this means that such abnormal protein does not exists normally in nature, so we predicted the involved SNP as the RA causing SNP. The remaining six proteins were also analysed by using this Swiss Model server but the models were produced. The final step was the application of Swiss PDB viewer. In order to use this tools, the PDB files were needed which were obtained by Swiss Model as well as by Protein Data Bank. The Swiss PDB veiwer modeled the structures for every SNP including the 17th SNP of MIA gene.

RESULTS AND DISCUSSION
The intermediate steps including ClustalW, TranSeq and TCoffee were used to obtain the minor objectives i.e. to move towards the structural analysis. The results that were obtained by the structural analysis are mentioned in this paper. The first analyzed gene was MIA whose total SNPs, mentioned in NCBI, are 20 out of which one SNP was selected for analysis due to its occurrence in Exon region. The SNP change was C to T that results in replacing Thr to Ile at at 113th amino acid position of the protein. Thr is a hydrophillic polar residue, that is why it is usually present on outer region of protein structure but Ile is hydrophobic non-polar residue which tends to move inward of protein structure. When the normal and the abnormal protein structures were overlapped by using Swiss PDB viewer, the result obtained shows that the orientation of the protein structure has been changed. Therefore, we predict that SNP is RA causing SNP as it produces structural change in protein.

The encircled area in the result obtained by the overlapping of normal and abnormal MIA (consist of few residues of SNP region) shows the difference in the orientation of both the proteins. The 2nd analyzed gene was CST3 which has total of 58 SNPs out of which 4 SNPs were selected i.e. SNP number 10, 14, 19 and 20. In case of 10th SNP, the conversion was C to A and the amino acid change was Arg to Ser at position 71. Arg is hydrophillic basic residue whereas Ser is hydrophillic polar residues which are of completely opposite natures. The result obtained by Swiss PDB viewer is mentioned in Fig 2.

The fig 2 shows the change in orientations of both the overlapped proteins. The secondly selected SNP of CST3 was 14th number SNP in which the polymorphism or conversion is G to A and the amino acid change is Arg to His at 71 amino acid position. The polymorphism is at the same position as that of SNP number 10. Arg and His both are hydrophillic basic residues but the change in structure was observed because His has cyclic structure where as Arg does not have cyclic structure. The structural analysis obtained by Swiss PDB viewer is shown in Fig 3.
Fig 1: Result caused by C to T polymorphism

Fig 2: Result caused by C to A polymorphism
Fig 3: Result caused by G to A polymorphism

Fig 4: Result caused by C to G polymorphism
Fig 5: Result caused by G to C polymorphism

Fig 6: Result caused by T to C polymorphism
Fig 7: Result caused by A to G polymorphism

Fig 8: Result caused by C to T polymorphism
**Fig 9:** Result caused by C to T polymorphism
The SNP number 19 has a conversion of C to G and the amino acid change is Arg to Gly at amino acid position 96. Arg is hydrophillic basic residue and Gly is hydrophillic polar residue. The result obtained by Swiss PDB viewer is shown in Fig 4, in which change in protein orientation can be noticed.

The fourth analyzed SNP of CST3 was 20th number SNP in which the conversion is G to C and the amino acid change is Gly to Ala at 38th amino acid position. Gly is hydrophillic polar residue where as Ala is hydrophobic non-polar residue. The overlapping of normal and abnormal protein structures showed the orientation change in protein structure.

The third gene analyzed was CTLA4, which has total 79 SNPs out of which 2 SNPs were selected for analysis. The first analyzed SNP was number 28 in which the nucleotide conversion was T to C and the amino acid change was Met to Thr at amino acid position 91. Met is a hydrophobic non-polar amino acid while Thr is a hydrophillic polar amino acid. Thr tends to arrange itself towards outer region of protein surface while Met being hydrophobic residue, tends to move inward of the protein structure. When the normal and the abnormal protein structures were overlapped, the change in protein orientation was observed which is mentioned in Fig 6.

The second analyzed SNP was 37th SNP in which the nucleotide conversion was A to G and the amino acid change was Thr to Ala at 17th amino acid position. Thr is hydrophillic residue while Ala is hydrophobic non-polar residue. The overlapping of normal and abnormal protein structures shows the change in protein orientation, mentioned in Fig 7.

The fourth gene analyzed was HMGB1 which has a total of 60 SNPs out of which 2 SNPs were selected for analysis. As the purpose was to analyze the effect of polymorphism on the protein creation, so we took only 2 SNPs out of 60. The conversion of nucleotide base and the amino acid change was same in both the SNPs i.e. C to T nucleotide conversion and Pro to Leu amino acid change. The first SNP was 33rd number SNP in which the mutation was at 143rd position and in case of second analyzed SNP i.e. SNP number 47, the mutation was at 98th position. Pro and Leu both are hydrophobic non-polar residues but Pro has cyclic structure while Leu has linear, non-cyclic structure. The overlapping of normal and abnormal protein structures of both the cases i.e. Fig 8 shows the 33rd SNP case and Fig 9 shows 47th SNP case.

CONCLUSION

It is evident that various genes are involved in Rheumatoid Arthritis and it has been observed that various SNPs of different genes are associated with Rheumatoid Arthritis. Our research was based on analyzing the structural changes caused by the SNPs in the respective proteins of above-mentioned genes. In order to analyze the relationship of SNPs with Rheumatoid Arthritis, the SNPs of four genes namely MIA, CST3, CTLA4, and HMGB1 were analyzed and the effects of SNPs were pointed out that how SNP was changing the structure of a protein. The SNPs of the exon regions were only selected for the analysis (excluding the SNPs of intronic region). The
Database used for the data collection was NCBI. There were about 17 SNPs of the above-mentioned genes which were found to be mutation causing. These SNPs replaced the normal amino acids with the mutated ones. When the normal and mutated proteins were overlapped, the proteins were not able to overlap properly at the point of mutation, which indicates that there was a structural difference caused by the SNP and this mutation can become the cause of Rheumatoid Arthritis.

REFERENCES


