SUMMARY AND CONCLUSION

- Rheumatoid Arthritis (RA), which is a combination of inflammatory and autoimmune disease also, includes complex interplayers among genotype, environmental triggers, and by chance. As till now, there is no permanent cure reported for RA, an attempt was made to identify significant molecular proteins using three different approaches (Gene Interaction Map (GIM), RA- Drugs-target-Protein interactome (RA-DTP) and Microarray data analysis of macrophages and Synovial Fibroblast (SF)) which may bring the accuracy in diagnosis and the therapeutic outcomes.

- In first phase of study, the built Gene Interaction Map (GIM) was built for RA and found to include 1046 molecules as nodes and 4119 edges as interaction. These molecules have been reported to be associated with RA. Twelve genes IRAK1, FADD, MAPK8, MYD88, TRAF6, IKBKB, MAP3K7, REL, RELB, MAP3K14, RPL7, MAPK3 were suggested as candidate targets while studying the four clusters obtained by using vertex weighting algorithm and k-Core application for RA. These genes were validated by three types of statistical measures (Closeness centrality, Cluster Coefficient and Degree). The reported genes were involved in different pathways like Apoptosis, Toll like receptor, ECM receptor interaction signalling, Epithelial signalling, MAPK signalling, T-cell receptor signalling, Focal adhesion, B-cell receptor signalling, WNT signalling, Adherens junction and Adipocytokine signalling pathways.

- In second phase of study, Rheumatoid Arthritis Drug’s-Target-Protein (RA-DTP) interactome was constructed which included 62 seed drugs and their related 103 targets. The interactome model consisted of 1727 nodes and 7954 edges with 20 islands, 55 modules and 123 sub module which also followed the power law distribution. Good interactome coverage of target-protein was detected in Island 2 (Q-Score 0.875) which included
673 target-protein with 20 modules and 68 sub modules. The biological landscape of these modules was examined based on the participation of target-protein in specific cellular localization, molecular function and biological pathway with low p value (≤ 0.05). Strong statistical relationship was identified between different modules and Gene Ontology (GO) terms; GO biological process, GO cellular component and GO molecular function. Functional characterization and diverse biological pathways were also statistically found through KEGG, Reactome and Biocarta with lowest p-values.

- The proteins obtained through traffic value (Betweenness centrality) and GO ontology (Gene Annotation) analysis were reconstructed among each other which gave an undirected network of highly predicted associated proteins related to RA. Traffic values and centrality parameters were applied as the selection criteria for identifying significant proteins from the important clusters. Through this FOS, KNG1, PTGDS, HSP90AA1, REN, POMC, FCER1G, IL6, ICAM1, SGK1, NOS3, PLA2G4A, RENBP, CTNS, NAGK, hROAT1, OPRK1, RENBP, CTNS, NAGK, hROAT1, and OPRK1 were identified as 17 candidate proteins that can serve as drug targets for RA.

- The reported proteins were found to be involved in different pathways like MAPK signalling, Focal adhesion, NGF and mTOR Signalling, adherens junction, ECM-receptor interaction, Adipocytokine signalling, Glucosamine metabolism arachidonic acid pathway.

- In the third phase of the study, Differentially Expressed Genes (DEGs) were studied using statistical measures for each series of macrophages and synovial fibroblast (SF) of RA synovial membrane.

- RMA algorithm was use to normalize the raw signals of all the series which were under Affymetrix and Agilent platform. The multiple raw signals were compared using multiple hypothesis testing, p-value (≤ 0.05) and FC=1.5 (≥ 1.5 and ≤ -1.5).
• The synovial fibroblast series GSE7669 (89 DEGs with 54 up and 35 down-regulated genes), and macrophages series GSE10500 (1485 DEGs with 671 up and 814 down-regulated genes), GSE8286 (3624 DEGs with 2195 up and 1429 down-regulated genes) were differentially expressed.

• Higher correlation coefficient (r ≥ 0.9) with Pearson algorithm was implied in Gene-gene correlation of DEGs which resulted in 403 genes interactome (edges= 81,140) for down regulated and 510 genes interactome (edges = 1, 83,064) for up-regulated genes. The whole interactome gave a good landscape of highly correlated co-expression matrix.

• The proteins identified from the first and second phase were used to map out with the highly differentially expressed genes where it was found that these 12 proteins-PTGDS, CTNS, NAGK, hROAT1, FADD, IRAK1, MAP3K7, RPL7, FOS, KNG1, HSP90AA1 and POMC exhibited the significant differential expression.

• k-Core analysis (k-Core -15) and centrality parameters (Betweenness centrality, cluster coefficient, Degree, Topological coefficient) identified 11 other proteins CHIT1, FN1, AQP9, PLA2G7, CHI3L1, APOE, VCAN, VEGFA, CD69, SPP1 and CASP8.

• These proteins were associated with ECM-receptor interaction pathway, Toll like Receptor signalling pathway, Arachidonic acid metabolism, Osteoclast differentiation, T cell receptor signalling pathway, anti-apoptosis, NF-kappa B signalling pathway, Focal adhesion, VEGF signalling pathway, mTOR signalling pathway, Cell adhesion and Cytokine cytokine receptor interaction pathways.

• To map out the above 40 proteins in relation to identify their GO annotation and pathway relationship STRING and Biointerpreter was used. The 40 proteins were enriched as a network using STRING in which 5 proteins AQP9, RPL7, SGK1, hROAT1 and CTNS didn’t showed any
interaction. The remaining 35 proteins in the network were found to be highly connected among each other

- Among the 35 interconnected proteins three proteins IL6, VEGF and ICAM1 were identified as common proteins, when mapped with RA pathophysiology (KEGG: hsa05323). Hence, a total of 15 proteins network was further studied which included the above three proteins and their interaction.

- Even though IL6, VEGF and ICAM1 are essential proteins and are involved in normal pathways they are used as drug targets for RA but with several side effects. To find out the significant proteins with fewer side effects, database of essential genes was explored to identify essentiality and non-essentiality of any protein that are indispensable to support cellular life.

- Among the 15 proteins 10 were essential proteins and targeting them might lead to many side effects, however the remaining 5 non-essential proteins namely NAGK, CHI3L1, CHIT1, RENBP and SPP1 were found attractive.

- All the four proteins NAGK, CHI3L1, CHIT1, RENBP were involved in N-Acetylglucosamine (GlcNAc) metabolism whereas, SPP1 in inflammation and bone destruction

- The main goal of the study was to establish signature molecules which are characteristic for RA and to identify potential diagnostic as well as therapeutic targets.

- Five proteins NAGK, RENBP, CHI3L1, CHIT1 and SPP1 which were identified as non-essential proteins, involved in three different categories of function specifically in relation to joints that is
  - N-Acetyl Glucosamine metabolism,
  - Fibroblast initiation for pannus formation
  - Th17 initiation for bone degradation.
These three major functions eventually include the involvement of Amino sugar and nucleotide sugar metabolism, ECM-receptor interaction, Osteoclast differentiation, Focal adhesion and glucosamine metabolism pathways through these five proteins.

The above findings hypothesize that these five proteins would be the best possible target proteins in RA and their associated pathways which are consistent among all the above studies, indicative of its key role in RA pathogenesis.

The study can be further extended by experimentally validating the results using in vitro and in vivo approaches.