ScienSage

Journal of Advanced Scientific Research

Available online through http://www.sciensage.info

ISSN
0976-9595
Research Article

IDENTIFICATION OF THE POTENTIAL MOLECULAR BIOMARKERS OF COVID-19 AND THE ASSOCIATED DRUGS: A BIOINFORMATICS APPROACH

Veena Puri*[1], Akanksha Arora[1], Sanjeev Puri[2]

[1]*Centre for Systems Biology and Bioinformatics, Panjab University, Chandigarh, India*

[2]*Biotechnology Branch, University Institute of Engineering and Technology (UIET), Panjab University, Chandigarh, India*

**Corresponding author: vpuri1825@pu.ac.in*

ABSTRACT

Coronavirus disease 2019 (COVID-19) has overwhelmed some of the best healthcare systems of the world and poses a major threat to the health of millions of people.

This research aims to discuss the potential molecular biomarkers of COVID-19 and relatedness of these biomarkers with the available battery of drugs duly approved by the Food and Drug Administration (FDA). We retrieved the gene expression and single cell RNA sequencing data of COVID-19 patients from the Array Express Database. The data was analyzed to find the dysregulated genes in COVID-19 infection as well as the cell-types in which they are differentially expressed. It was further assessed to find relatedness to other known diseases and the FDA-approved drugs associated with them using disgenet2r and rDGIdb packages in R respectively. The dysregulated genes were found to be highly linked with Lupus Erythematosus, Leukemia, Malaria, and Rheumatoid Arthritis associated pathways. The key hub genes identified in this study include CD1A, CD40, CXCL10, HLA-DQB1, HLA-DRA, IFNG, IL1RAP, IL2RA, IL4R, IL6R, S100A8, and TLR9. These genes were then further studied using scRNA data to understand the immune profile of COVID-19 patients. Drug-gene interactions studies revealed that these potential biomarkers are related to the FDA approved therapeutic agents namely Triamcinolone, Dacetuzumab, Zidovudine, Methylprednisolone, Ritonavir, Nevirapine, Olsalazine, Daclizumab, Dupilumab, Siltuximab, Tocilizumab, Zinc Chloride, and Hydroxychloroquine. Their interactions with the key hub genes make these drugs potentially repurposable candidates for the possible intervention of coronavirus infections.

Keywords: SARS-CoV-2, Coronavirus, Malaria, Rheumatoid Arthritis, Single-cell RNA sequencing, Drug repurposing.

1. INTRODUCTION

On March 11th, 2020, Coronavirus Disease 2019 (COVID-19) was officially declared as a global pandemic by the World Health Organization (WHO) [1]. As of December 21st, 2020, it is responsible for infecting about 76.8 million people and causing more than 1.69 million deaths worldwide. COVID-19 is characterized by fever, dry cough, myalgia, dyspnoea along with lymphocytopenia. In more severe cases (up to 15-20% of the infected patients), complications like acute respiratory distress syndrome (ARDS), acute heart injury, and secondary infections are observed, which can potentially lead to death of a patient [2, 3].

Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2), the causative agent of COVID-19, belongs to the family coronaviridae is further classified within the order Nidovirales. The members of the coronaviridae family are characterized by a single-stranded and an unusually large RNA genome. They have a distinctive replication strategy and are known to cause various mild to severe enteric, respiratory, or systemic diseases in animals and potentially fatal infections in humans [4-6]. Before Severe Acute Respiratory Syndrome (SARS), they were merely known to cause mild respiratory infections in humans [7].

In the current outbreak, an alarming rise in the number of infections and the related deaths has necessitated a holistic approach towards finding an early cure for this dreaded infection. COVID-19 is affecting people's lives around the world directly as well as indirectly. Thousands of people are dying due to the infection while others are struggling to survive as they are losing their jobs. It is thus crucial to find a cure in a short amount of time to not only save lives but also to protect the economies all over the world from crashing. The vaccine development, which is important for the future

onslaught, though is rigorously underway, will take its regular course. The immediate intervention is equally imperative and currently seems to rely upon repurposing of the already available drugs [8]. This repurposing if proves beneficial, besides drastically lessening the pain and suffering immediately would also reduce the grave monetary burden and time-related costs. Moreover, it will also alleviate the safety risk associated with new drugs or vaccines [9]. In order to repurpose already approved drugs, it is important to find relatedness of COVID-19 with other diseases that already have a therapeutic strategy available. As we know that the development of a reliable vaccine is a time-consuming process, it is necessary to opt for the fast-computational methods and bioinformatics approaches to find a potential treatment for COVID-19 in a significantly lesser amount of time. Presently, the COVID-19 related data is being generated in large amounts and is openly available. This data can be analyzed using various bioinformatics tools to extract the relevant information which can be utilized to find a possible link with a disease with previously known treatment or to discover drugs that can be potentially repurposed to save the patients suffering from COVID-19 [10-12].

Microarray analysis is one of the most popular methods for molecular profiling for the study of complex diseases. In addition to microarray analysis this study aims to provide a deeper insight into the immune cell composition of COVID-19 patients using single cell RNA sequencing data analysis. Bioinformatics approaches pave a significant role in such studies. Although the approach used in our study has been previously applied to the study of numerous diseases, it will still be amongst the pioneering studies for COVID-19. This approach will help to discover the drugs that can potentially save lives of the patients infected with SARS-CoV-2 and on the hind side this will also provide a caution, should there be any possibility of an adverse drug interaction. In this study, we have applied the Bioinformatics approach on the daily transcription profiling data and single cell RNA sequencing data of Peripheral Blood Mononuclear Cells (PBMC) of the COVID-19 patients to identify the genes that are dysregulated during the SARS-CoV-2 infection. Gene-disease and gene-drug association studies have also been performed using different packages in R to obtain a list of potential molecular biomarkers of COVID-19 and the FDA-approved drugs associated with them [13]. These drugs can potentially be repurposed to treat SARS-CoV-2 infections. However, they are definitely required to be validated through clinical trials before being used.

2. METHODS

The complete work flow strategy for the data analysis has been presented in Fig. 1

2.1. Data source and retrieval

2.1.1. Gene-expression data

The gene-expression data set used in this study was obtained from the Array Express database with the accession ID: E-MTAB-8871. It contains the data of daily transcriptional profiling conducted on the whole blood of 10 uninfected human subjects and three SARS-CoV-2 infected patients on days 4 to 13 and day 18 in case control and time series design. The data of the uninfected human subjects was not gender biased. To analyze the data obtained from Array Express, we divided the data into 12 groups (A-L). The strategy used for Bioinformatics based data analysis has been explained in Fig. 1.

A.) Gene expression data of uninfected human subjects (n = 10)
B.) Gene expression data of COVID-19 patients on day 4 (n = 1, r = 2)
C.) Gene expression data of COVID-19 patients on day 5 (n = 1)
D.) Gene expression data of COVID-19 patients on day 6 (n = 2)
E.) Gene expression data of COVID-19 patients on day 7 (n = 2)
F.) Gene expression data of COVID-19 patients on day 8 (n = 2)
G.) Gene expression data of COVID-19 patients on day 9 (n = 3)
H.) Gene expression data of COVID-19 patients on day 10 (n = 2)
I.) Gene expression data of COVID-19 patients on day 11 (n = 3)
J.) Gene expression data of COVID-19 patients on day 12 (n = 3)
K.) Gene expression data of COVID-19 patient on day 13 (n = 1)
L.) Gene expression data of COVID-19 patients on day 18 (n = 1)

Here, n = number of subjects in the group and r = number of replicates

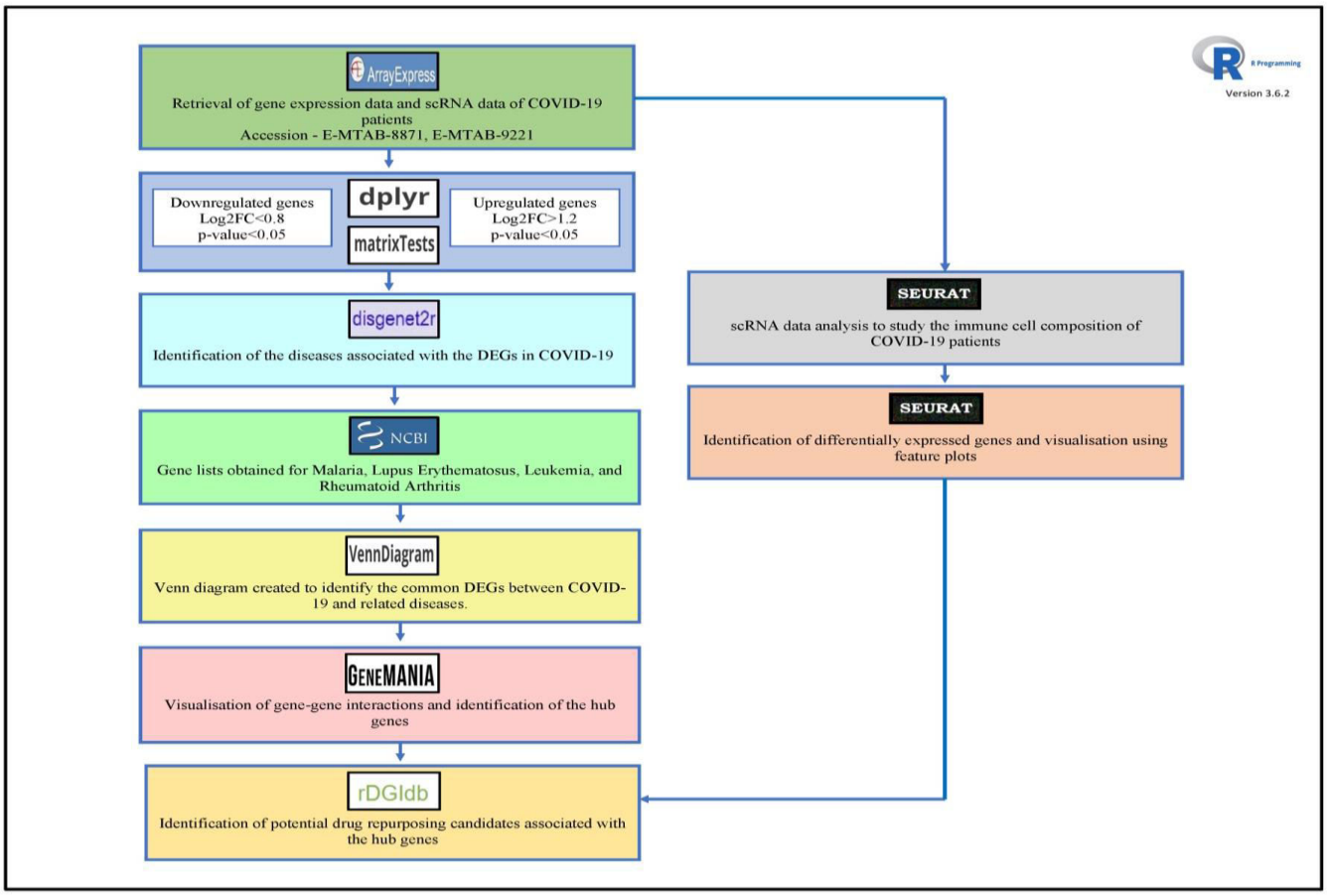


Fig. 1: Methodology- The complete work flow strategy for the data analysis

2.1.2. sc-RNA sequencing data

The single cell RNA sequencing dataset for COVID-19 patients analyzed in this study was obtained from the Array Express database having the accession ID-E-MTAB- 9221. It contains the following data:

A.) PBMC data of uninfected human subjects (n=3)
B.) PBMC data of mildly affected COVID-19 patient
  i) Mildly affected patient on Day 1 of infection (n=1)
  ii) Mildly affected patient on Day 10 of infection (n=1)
C.) PBMC data of recovered COVID-19 patients (n=1)
D.) PBMC data of severely affected COVID-19 patients
  i) Severely infected patients on Day 1 (n=2)
  ii) Severely infected patients on Day 10 (n=2)

Here, n = number of subjects in the group

The recovered patient mentioned in this data is a severe patient that recovered from COVID-19 on day 17 of the infection. These groups were analyzed to see how immune cell compositions and gene expression profiles differ in control human subjects, patients with minor infection, patients with severe infection, and recovered patients.

2.2. Data pre-processing

2.2.1. Gene expression data

To obtain a list of dysregulated genes for COVID-19, the processed gene expression data was used which already contained the data as log2 values for the gene expressions. The mean value in each group is the average expression value in that entire group. The term fold change |Log2FC| is the ratio of patient vs control group data. Log2FC of patient data groups with respect to Group A (Gene expression data of uninfected human subjects) was calculated. Welch's t-test was applied to the data using the Matrix Tests package on R. This test was chosen for its well-known reliability for comparing the two samples which are unequal in the size. Matrix Tests package provides the fast application and a detailed output of the complex data. The differentially expressed genes (DEGs) were identified by setting a two-fold cut-off for the |Log2FC| values. These genes

were furthered filtered for the p-value<0.05. The data manipulation tasks of importing and filtering the data were performed using the dplyr package on R [14-17].

2.2.2. sc-RNA sequencing data

The gene count data was obtained from the Array Express database with the feature file and barcode matrix. These files were integrated using the Seurat Package on R. The data was then filtered by applying criteria for gene number >250, nUMI (number of unique molecular identifiers) >500, and mitochondria ratio <0.20. The data was log normalized before performing the clustering and dimensionality reduction. For quality control, PCA analysis was used to eliminate mitochondria (MT) genes from the data. The differential markers were identified using Find All Markers() function in Seurat package on R [18-20]. Composition of immune cells and expression of hub genes in immune cells were studied by constructing feature plots in R.

2.3. Gene-Disease association study

The dysregulated genes from each group were analyzed to identify the relationship of these genes with other known diseases using the disgenet2r package on R [21]. This package contains a set of functions to retrieve and visualize the data on DisGeNET repository. It is one of the largest repositories that encompass information about human diseases and their co-morbidities at the molecular level. The DisGeNET scores vary from 0 to 1 and denote the evidence supporting the association of a gene with a disease. We selected the threshold value for scores at 0.5 to find the gene-disease interactions. The disease-gene association heat maps and networks were built for the in-depth study of the gene-disease relationships (Fig.2 and Fig. 3, respectively). The list of genes associated with these diseases was obtained from the National Centre for Biotechnology Information (NCBI) GENE database using the disease names as keywords, *e.g.*, "Lupus Erythematosus", "Leukemia", "Malaria", and "Rheumatoid Arthritis" [22, 23]. Finally, the Venn diagram package on R was used to discover the common genes between these diseases and COVID-19 [24].

2.4. Gene-Gene interaction study

Hub-genes were identified using the GeneMANIA application on Cytoscape version 3.7.2. It uses a database of organism-specific weighted networks to build composite networks. This composite network is organized into groups such as co-expression, physical interactions, predicted, pathway, shared protein domains, co-localization, and genetic interactions. The dysregulated genes were analyzed to obtain a composite network through GeneMANIA which was then studied to identify the hub-genes [25]. These genes were further filtered by selecting the genes with degree>25 and betweenness centrality>0.0065. The data for betweenness centrality parameter was extracted for each hub gene as nodes with a high betweenness centrality are important in signaling pathways and can be important drug targets [26]. The hub genes identified through this method were then studied in the single cell RNA sequencing data clusters of control human subjects, mildly affected individuals, severely infected individuals, and recovered individuals to study the pattern of expression of these hub genes in different types of immune cells.

2.5. Gene-Drug association study:

We used the rDGIdb package on R to recognize the drugs associated with the COVID-19 dysregulated genes. The rDGIdb package provides a wrapper for the Drug Gene Interaction Database (DGIdb). It provides information about the gene-drug interactions, their PMIDs, interaction types, and the source database from where the information was obtained. It also provides the query score and interaction score values to rank the results of the gene-drug interactions [27]. The query score given by DGIdb is relative to the search set and considers the overlap of interactions in the result set. It depends on the evidence scores, number of genes from the search query set that interact with a specific drug, and the ratio of all known gene interactors for all drugs and with the known gene interactors of a specific drug [10]. The interaction score is a score based on evidence of interaction acquired from publications and sources. It does not change from query to query and remains constant for the interactions.

3. RESULTS

3.1. Identification of dysregulated genes in COVID-19 patients

A total of 253 genes were found to be dysregulated in all the groups. The down and up regulated genes in each group are described in Table 1. A total of 25 dysregulated genes were found to be common in COVID-19, Rheumatoid Arthritis, Lupus Erythematosus, Leukemia and Malaria. Out of these 7 were observed to be down regulated, 18 were up regulated. Some of the dysregu-

lated genes common in these groups include IF12B, IFNG, IFNA2, C4BPA, CXCL10, HLA-DQA1, NOS2, CD19, and CD40.

3.2. Diseases associated with dysregulated genes in COVID-19

Three diseases viz. Lupus Erythematosus, Leukemia, Malaria, and Rheumatoid Arthritis were observed to be common in disease-gene networks and heatmaps of the dysregulated genes from each group (fig.2 and fig. 3, respectively). Genes like IRF5, CIITA, HLA-DQA1, IL6R, CD40, STAT3 and FCGR2B have been found to be associated with Lupus Erythematosus, Leukemia, Malaria, and Rheumatoid Arthritis. The scores for each Gene-disease interaction are given in Table 2. The lists of genes associated with these diseases were retrieved from the NCBI GENE database. The number of genes associated with these diseases found in common with COVID-19 was 125, 148, 42, and 137 for Lupus Erythematosus, Leukemia, Malaria, and Rheumatoid Arthritis respectively (fig. 4). Out of these 25 dysregulated genes were found to be involved in the pathophysiology of all four diseases as well as COVID-19. These were further analyzed to identify the potential drug candidates, which can be used as targets in SARS-CoV-2 infection.

Table 1: Down and Up regulated genes in each group mentioned in section 3.1 (B to L)

Group	Down/ Up regulated	No. of genes	Gene symbols
B (day 4)	Down regulated	86	ARG1, BCL3, BCL6, BST1, C1QA, C1QB, C2, CASP1, CCL2, CCL8, CCR1, CCR5, CCRL2, CD14, CD274, CD36, CD58, CDKN1A, CEACAM1, CLEC4E, CLEC5A, CR1, CX3CR1, CXCL1, CXCL10, CXCL12, FAS, FCER1G, GBP1, GBP5, GNLY, GP1BB, GZMB, HLA-A, IFI16, IFIT2, IFITM1, IL18, IL1R1, IL2RG, IRAK2, IRF1, IRF5, ITGA2B, ITGAM, ITGB2, KCNJ2, KIR_Activating_Subgroup_1, KIR_Activating_Subgroup_2, KLRD1, LAG3, LAMP3, LILRA5, LILRA6, LY96, MAPK14, MX1, NLRP3, PDCD1LG2, PML, PRF1, PSMB9, S100A8, S100A9, SELL, SLAMF7, SOCS1, SOCS3, STAT1, STAT2, STAT3, TAP1, TAP2, TBX21, TLR2, TLR3, TLR4, TLR5, TNFAIP6, TNFSF10, TNFSF13B, TNFSF15, TRAF6, ZBTB16, FCGR1A/B, FCGR3A/B
	Up regulated	33	BCL2, C4BPA, CASP10, CCBP2, CD1A, CD22, CD4, CD79A, CIITA, CSF1R, CTSG, FCER1A, FCGR2B, HLA-DMB, HLA-DOB, HLA-DRA, HLA-DRB3, ICOSLG, IKZF2, IL12B, IL1RAP, IL23A, IL2RA, IL6R, ITLN1, KLRB1, LEF1, MS4A1, NT5E, SPP1, TCF7, TNFRSF13C, TNFRSF4
C (day 5)	Down regulated	92	ARG1, BCL3, BCL6, BST1, C2, CASP1, CCL4, CCND3, CCR1, CCR5, CCRL2, CD14, CD274, CD36, CD58, CDKN1A, CEACAM1, CLEC4E, CLEC5A, CLEC7A, CR1, CTNNB1, CX3CR1, CXCL10, CXCL12, FADD, FCGR1A/B, FCGR3A/B, FKBP5, GBP1, GNLY, GP1BB, GZMB, HLA-A, IFI16, IFIT2, IFITM1, IL1A, IL1R1, IL1R2, IL2RG, IL4R, IRAK2, IRAK3, IRF5, ITGA2B, ITGAM, ITGAX, ITGB2, KCNJ2, KIR_Activating_Subgroup_1, KIR_Activating_Subgroup_2, KLRD1, LAG3, LAMP3, LILRA5, LILRA6, LILRB3, LTBR, LY96, MAPK14, MX1, NFKBIA, NOD2, PDGFRB, PML, PPARG, PRF1, PSMB9, RUNX1, S100A8, S100A9, SELL, SOCS1, SOCS3, STAT1, STAT2, STAT3, TAP1, TAP2, TBX21, TLR2, TLR3, TLR4, TLR5, TLR8, TNFAIP3, TNFAIP6, TNFSF10, TNFSF13B, TNFSF15, ZBTB16
	Up regulated	43	C4BPA, CCR7, CCR8, CD1A, CD22, CD4, CD74, CD79A, CD83, CD8A, CIITA, CR2, DPP4, EOMES, FCER1A, FCGR2B, GPR183, GZMK, HLA-DMA, HLA-DMB, HLA-DOB, HLA-DPB1, HLA-DRA, HLA-DRB3, ICOSLG, IKZF2, IL12B, IL1RL2, IL23A, IL2RA, IL6R, ITLN1, KLRB1, LEF1, MS4A1, NT5E, SLAMF1, SPP1, TCF7, TFRC, TNFRSF13C, TRAF1, VCAM1

D (day 6)	Down regulated	27	C2, CASP1, CCR1, CD163, CEACAM1, FKBP5, GBP1, IFI16, IFIH1, IFIT2, IFITM1, IL18, IRF7, JAK2, LAG3, LILRB2, LTBR, MX1, PML, PSMB9, SERPING1, STAT1, STAT2, TNFAIP6, TNFSF10, TNFSF13B, ZBTB16
	Up regulated	12	C4BPA, CCBP2, EOMES, FCER1A, HLA-DOB, HLA-DQA1, HLA-DQB1, ITLN1, MS4A1, PAX5, SPP1, TNFRSF13C
E (day 7)	Down regulated	7	CD274, CEACAM1, IL18RAP, IL1RN, KLRC2, TNFAIP6, TNFRSF17
	Up regulated	19	C4BPA, CD1A, CD22, CD79A, CD97, CR2, EOMES, HLA-C, HLA-DOB, HLA-DQA1, HLA-DQB1, IL12B, IL1RAP, LILRA3, MS4A1, NT5E, PLA2G2E, TFRC, TNFRSF13C
F (day 8)	Down regulated	3	CCR5, IL18RAP, TNFRSF17
	Up regulated	23	AICDA, C4BPA, CCBP2, CD19, CD22, CD79A, DEFB1, FCER1A, HLA-DOB, HLA-DQA1, HLA-DQB1, IFNA2, IL12B, IL1RAP, IL1RL2, ITLN1, KLRAP1, LILRA3, MASP2, MS4A1, PAX5, TNFRSF13C, TNFSF11
G (day 9)	Down regulated	3	CCR1, CCR5, PDCD1
	Up regulated	29	AICDA, C4BPA, CCBP2, CD19, CD209, CD22, CD55, CD79A, CD97, CEBPB, CSF2RB, EGR2, HLA-DQA1, HLA-DQB1, IFNA2, IL12B, IL13RA1, IL1RAP, IL1RL2, IL6R, MAPKAPK2, MBP, MS4A1, PAX5, PDCD1LG2, TFRC, TNFRSF10C, TNFRSF13C, XCL1
H (day 10)	Down regulated	5	CCR5, KIR_Activating_Subgroup_2, KIR_Inhibiting_Subgroup_1, PDGFRB, ZAP70
	Up regulated	39	ARG1, C4BPA, CAMP, CD19, CD22, CD24, CD55, CD79A, CEACAM6, CEBPB, CSF2RB, CSF3R, CXCL11, CXCL12, FCGRT, HLA-DQA1, HLA-DQB1, ICAM4, IFNAR1, IFNG, IL1RAP, IL4R, IL6R, IL7, ITGA2B, ITLN1, LILRA2, LY96, MAPKAPK2, MME, MS4A1, PAX5, S100A8, TAL1, TNFRSF10C, TNFRSF13C, TNFRSF9
I (day 11)	Down regulated	5	CCR5, GNLY, KIR_Activating_Subgroup_1, KIR_Activating_Subgroup_2, SERPING1
	Up regulated	57	B2M, BATF3, C4BPA, CCBP2, CCR6, CD19, CD1A, CD22, CD45RB, CD48, CD79A, CD82, CD97, CEBPB, CLEC4A, CLEC6A, CSF2RB, CXCL10, CXCL11, CXCL12, DEFB1, EGR2, FCER1A, FCGR2B, GPR183, GZMK, HLA-DPA1, HLA-DQA1, HLA-DQB1, HLA-DRA, ICAM2, ICAM4, IDO1, IFNA2, IFNAR1, IFNG, IL1B, IL1RAP, IL1RL2, IL23A, IL32, IL4R, IL6R, KLRB1, LGALS3, LY96, MAPKAPK2, MS4A1, NFIL3, NT5E, PAX5, SOCS3, TFRC, TNFAIP6, TNFRSF10C, TNFRSF13C, XCL1
J (day 12)	Down regulated	2	CCR5, ZBTB16
	Up regulated	75	B2M, BATF3, BCL3, C4BPA, C6, CCBP2, CD19, CD1A, CD22, CD274, CD40, CD45R0, CD45RB, CD48, CD55, CD79A, CD79B, CD80, CD82, CD86, CD97, CEBPB, CLEC6A, CSF2RB, CTSS, CXCL10, CXCL11, CXCL12, CXCR1, FCER1A, FCGR2A, GPR183, GZMA, GZMK, HLA-DMB, HLA-DPA1, HLA-DQA1, HLA-DQB1, HLA-DRA, ICAM2, ICOSLG, IDO1, IFNAR1, IL13RA1, IL1B, IL1RAP, IL32, IL4R, IL6R, KCNJ2, KLRAP1, KLRB1, LGALS3, LILRA2, LILRA5, LILRB3, LY96, MAPK11, MAPKAPK2, MS4A1, NFIL3, PAX5, PDCD1LG2, PTAFR, RELB, SLAMF6, SOCS3, TFRC, TLR9, TNFAIP6, TNFRSF10C, TNFRSF13C, TNFRSF17, TNFSF13B, XCL1
K (day 13)	Down regulated	2	CCR7, PDCD1
	Up regulated	85	B2M, B3GAT1, BATF3, BCL6, BST1, C4A/B, C4BPA, CASP3, CCBP2, CD19, CD1A, CD22, CD3D, CD45R0, CD45RB, CD46, CD48, CD55, CD82, CD86, CD97, CEBPB, CFD, CLEC4A,

			CLEC6A, CSF2RB, CSF3R, CTSS, CXCL10, CXCL11, CXCL12, CXCR1, CXCR2, EGR2, FCER1A, FCER1G, FCGR2A, FCGR2A/C, FCGR2B, GPR183, GZMA, GZMK, HLA-C, HLA-DQA1, HLA-DRA, ICAM4, IFITM1, IFNAR1, IFNG, IL1B, IL1R2, IL1RAP, IL32, IL4R, IL6R, IL7, ITGA2B, KCNJ2, KLRB1, LGALS3, LILRA2, LILRA5, LILRB3, LY96, MAPK11, MAPKAPK2, MME, MS4A1, MSR1, NFIL3, NOD2, PDCD1LG2, PPBP, PTAFR, S100A8, S100A9, SOCS3, TAL1, TFRC, TNFAIP6, TNFRSF10C, TNFRSF13C, TNFSF13B, TNFSF15, XCL1
L (day 18)	Down regulated	14	CASP10, CCL23, CD24, CEACAM6, CEACAM8, CTSG, GNLY, GP1BB, IL1RL1, ITLN1, KIR_Activating_Subgroup_1, KIR_Activating_Subgroup_2, ZAP70, ZBTB16
	Up regulated	45	B2M, BCL2, C4BPA, CD19, CD22, CD45RB, CD48, CD79A, CD82, CD86, CLEC4A, CLEC6A, CTSS, CXCL10, CXCL11, CXCL12, CXCR4, DUSP4, FCGR2B, FKBP5, HLA-C, HLA-DMB, HLA-DOB, HLA-DRA, ICOSLG, IFNAR1, IFNGR1, IL23A, IL32, IL6R, KIR3DL1, KIR_Inhibiting_Subgroup_2, KLRAP1, KLRB1, LEF1, LILRA5, LY96, MARCO, MS4A1, S100A9, TCF7, TIRAP, TNFAIP6, TNFRSF13C, TNFSF13B

Table 2: disgenet2r results for the gene-disease interactions of dysregulated genes in COVID-19 patients in group B to L

Group	DEG	Disease	Score[b]	Down/Up regulated
Day 4 (Group B)	IRF5	Lupus Erythematosus	0.70	Down regulated
	ITGAM	Lupus Erythematosus	0.70	Down regulated
	C1QA	Lupus Erythematosus	0.70	Down regulated
	IRF5	Rheumatoid Arthritis	0.50	Down regulated
	TRAF6	Rheumatoid Arthritis	0.50	Down regulated
	PML	Acute Promyelocytic Leukemia	0.60	Down regulated
	ZBTB16	Acute Promyelocytic Leukemia	0.60	Down regulated
	STAT3	T-Cell Large Granular Lymphocyte Leukemia	0.60	Down regulated
	FCGR2B	Malaria	0.53	Up regulated
	IL2RA	Rheumatoid Arthritis	0.50	Up regulated
	IL6R	Rheumatoid Arthritis	0.50	Up regulated
	CIITA	Rheumatoid Arthritis	0.50	Up regulated
Day 5 (Group C)	ITGAM	Lupus Erythematosus	0.70	Down regulated
	IRF5	Lupus Erythematosus	0.70	Down regulated
	TNFAIP3	Lupus Erythematosus	0.70	Down regulated
	TNFAIP3	Rheumatoid Arthritis	0.50	Down regulated
	IRF5	Rheumatoid Arthritis	0.50	Down regulated
	RUNX3	Acute Myelocytic Leukemia	0.80	Down regulated
	PML	Acute Promyelocytic Leukemia	0.60	Down regulated
	STAT3	T-Cell Large Granular Lymphocyte Leukemia	0.60	Down regulated
	ZBTB16	Acute Promyelocytic Leukemia	0.60	Down regulated
	FCGR2B	Malaria	0.53	Up regulated
	HLA-DPB1	Rheumatoid Arthritis	0.50	Up regulated
	IL2RA	Rheumatoid Arthritis	0.50	Up regulated
	IL6R	Rheumatoid Arthritis	0.50	Up regulated
	CIITA	Rheumatoid Arthritis	0.50	Up regulated
	TRAF1	Rheumatoid Arthritis	0.50	Up regulated
	LEF1	Chronic Lymphocytic Leukemia	0.50	Up regulated
Day 6 (Group D)	IFIH1	Lupus Erythematosus	0.70	Down regulated
	JAK2	Acute Myelocytic Leukemia	0.80	Down regulated
	PML	Acute Promyelocytic Leukemia	0.60	Down regulated
	ZBTB16	Acute Promyelocytic Leukemia	0.60	Down regulated

	JAK2	Chronic Myeloid Leukemia	0.50	Down regulated
	HLA-DQA1	Lupus Erythematosus	0.50	Up regulated
	SPP1	Lupus Erythematosus	0.50	Up regulated
Day 7 (Group E)	HLA-DQA1	Lupus Erythematosus	0.50	Up regulated
	CR2	Lupus Erythematosus	0.69	Up regulated
Day 8 (Group F)	HLA-DQA1	Lupus Erythematosus	0.50	Up regulated
Day 9 (Group G)	PDCD1	Lupus Erythematosus	0.80	Down regulated
	HLA-DQA1	Lupus Erythematosus	0.50	Up regulated
	IL6R	Rheumatoid Arthritis	0.50	Up regulated
Day 10 (Group H)	HLA-DQA1	Lupus Erythematosus	0.50	Up regulated
	IL6R	Rheumatoid Arthritis	0.50	Up regulated
	CSF3R	Chronic Neutrophilic Leukemia	0.60	Up regulated
	CSF3R	Acute/ Chronic Myeloid Leukemia	0.59	Up regulated
Day 11 (Group I)	HLA-DQA1	Lupus Erythematosus	0.50	Up regulated
	FCGR2B	Lupus Erythematosus	0.90	Up regulated
	CCR6	Rheumatoid Arthritis	0.50	Up regulated
	IL6R	Rheumatoid Arthritis	0.50	Up regulated
	FCGR2B	Malaria	0.53	Up regulated
Day 12 (Group J)	HLA-DQA1	Lupus Erythematosus	0.50	Up regulated
	FCGR2A	Lupus Erythematosus	0.50	Up regulated
	IL6R	Rheumatoid Arthritis	0.50	Up regulated
	CD40	Rheumatoid Arthritis	0.50	Up regulated
	FCGR2A	Rheumatoid Arthritis	0.50	Up regulated
Day 13 (Group K)	PDCD1	Lupus Erythematosus	0.80	Down Regulated
	HLA-DQA1	Lupus Erythematosus	0.50	Up regulated
	FCGR2B	Lupus Erythematosus	0.90	Up regulated
	FCGR2A	Lupus Erythematosus	0.50	Up regulated
	FCGR2A	Rheumatoid Arthritis	0.50	Up regulated
	IL6R	Rheumatoid Arthritis	0.50	Up regulated
	CSF3R	Chronic Neutrophilic Leukemia	0.60	Up regulated
	CSF3R	Acute/ Chronic Myeloid Leukemia	0.59	Up regulated
	FCGR2B	Malaria	0.53	Up regulated
Day 18 (Group L)	ZBTB16	Acute Promyelocytic Leukemia	0.60	Down regulated
	FCGR2B	Lupus Erythematosus	0.90	Up regulated
	IL6R	Rheumatoid Arthritis	0.50	Up regulated
	BCL2	Chronic Lymphocytic Leukemia	0.50	Up regulated
	LEF1	Chronic Lymphocytic Leukemia	0.50	Up regulated
	FCGR2B	Malaria	0.53	Up regulated

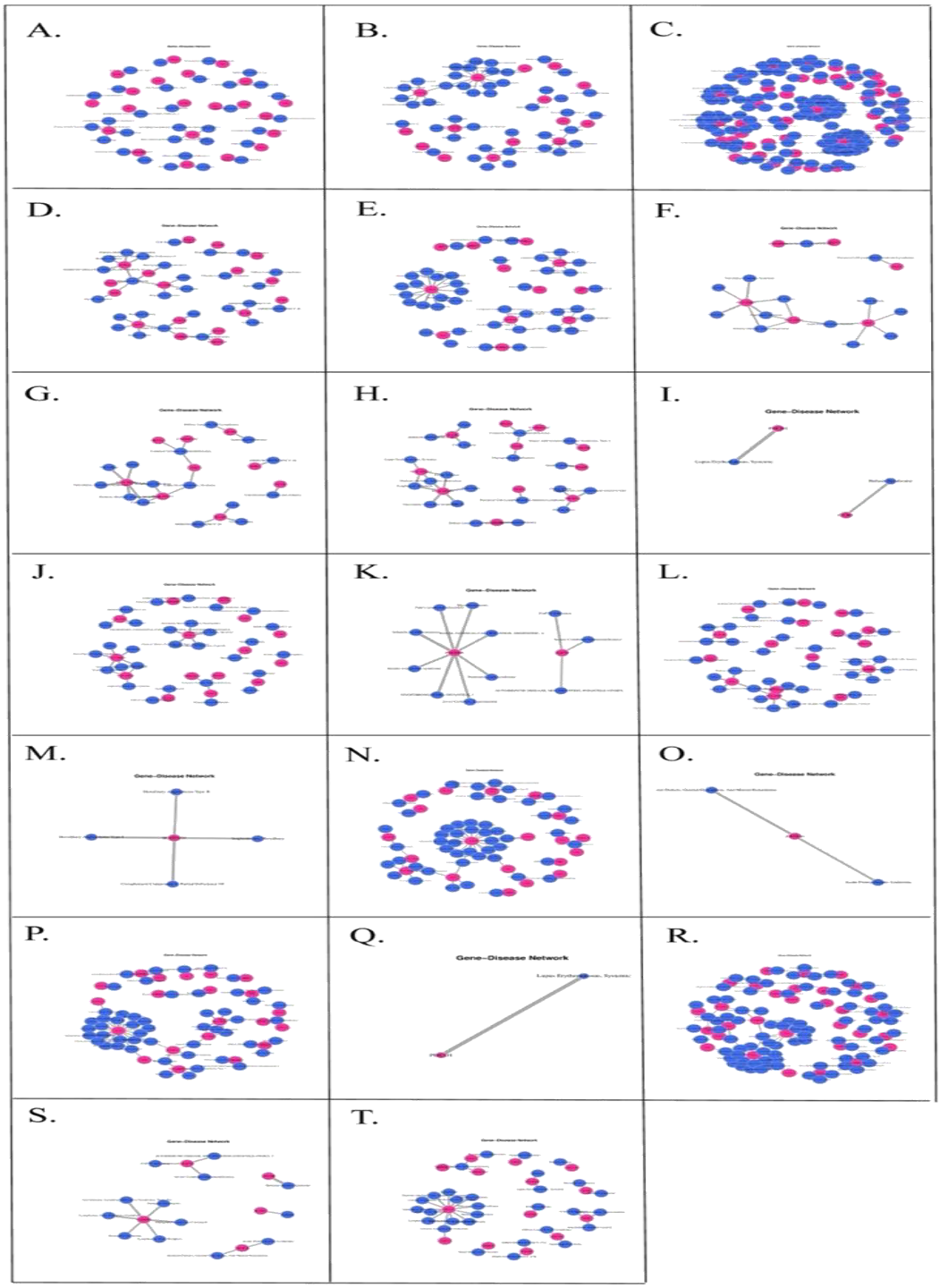
[b] *The score denotes the evidence supporting the association of a gene with a disease*

3.3. Identification of the hub genes

In addition to the 25 common dysregulated genes in COVID-19 associated with Lupus Erythematosus, Leukemia, Malaria, and Rheumatoid Arthritis; we short-listed the dysregulated genes common in more than 7 groups (days as mentioned in methods). The 23 genes common in more than 7 groups were added to the list resulting in a total of 48 genes. A composite network of the 48 dysregulated genes (fig. 5) was made using the GeneMANIA application on Cytoscape version 3.7.2. The two parameters i.e, Betweenness Centrality and the number of neighbors (Degree) were taken into the account to identify the hub genes. The details for the Degree and betweenness centrality along with the annotation of each hub gene are given in Table 3. CD1A, IL4R, and HLA-DQB1 have the highest values for betweenness centrality which makes them the most suitable drug targets for the treatment of COVID-19 [26]. The 12 hub genes identified for the Corona-virus Diseases 2019 are listed as follows. All of the hub genes identified in this study were found to be up regulated in COVID-19 patients.

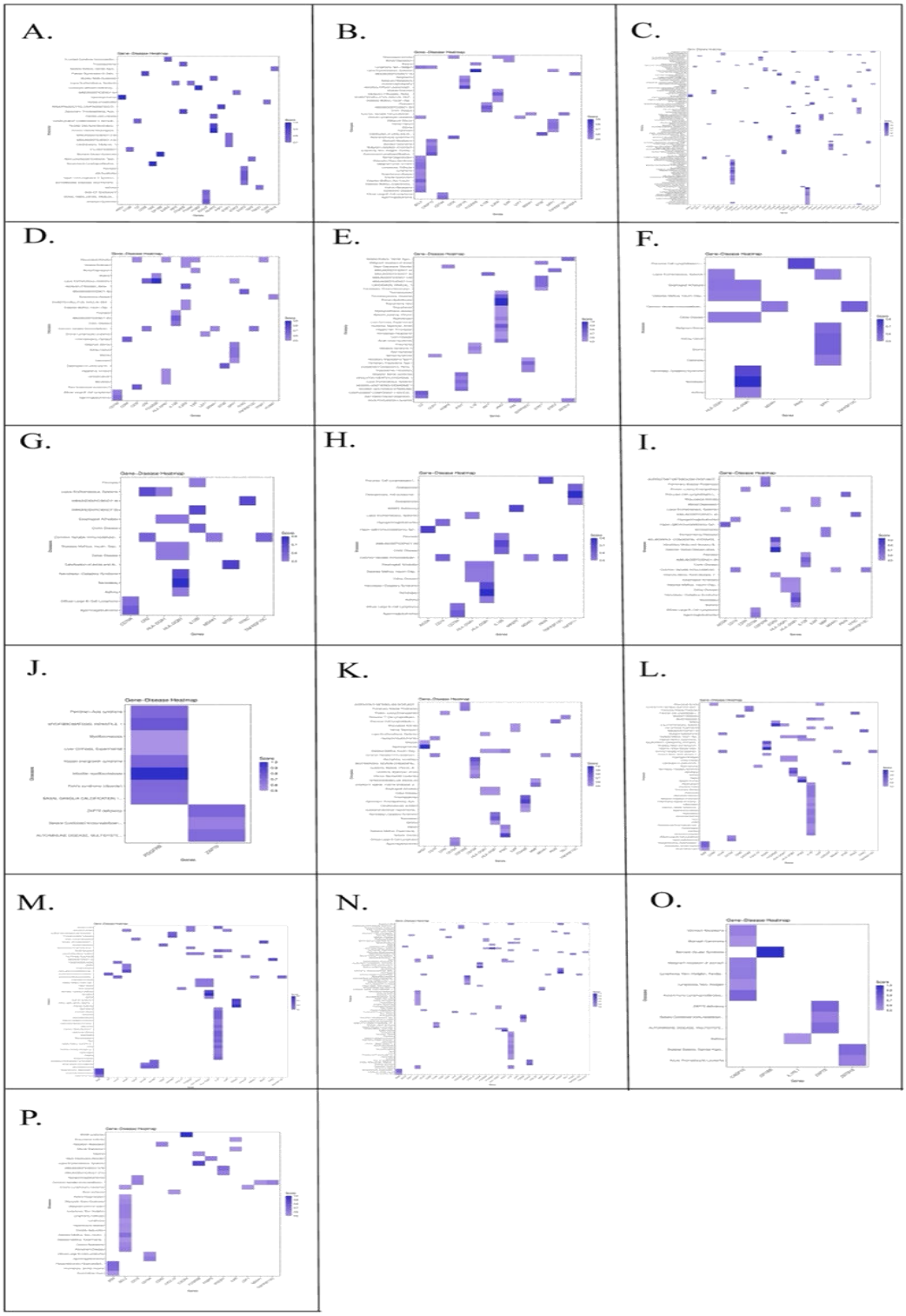
3.3.1. Cluster of Differentiation 1A (CD1A)

CD1A gene encodes for the CD1a protein which belongs to the family of transmembrane glycoproteins. It mediates the lipid and glycolipid antigen presentation to the T-cells. CD1A is involved in the mechanism of inflammatory skin diseases and amplifies TH17 mediated inflammatory response [28].



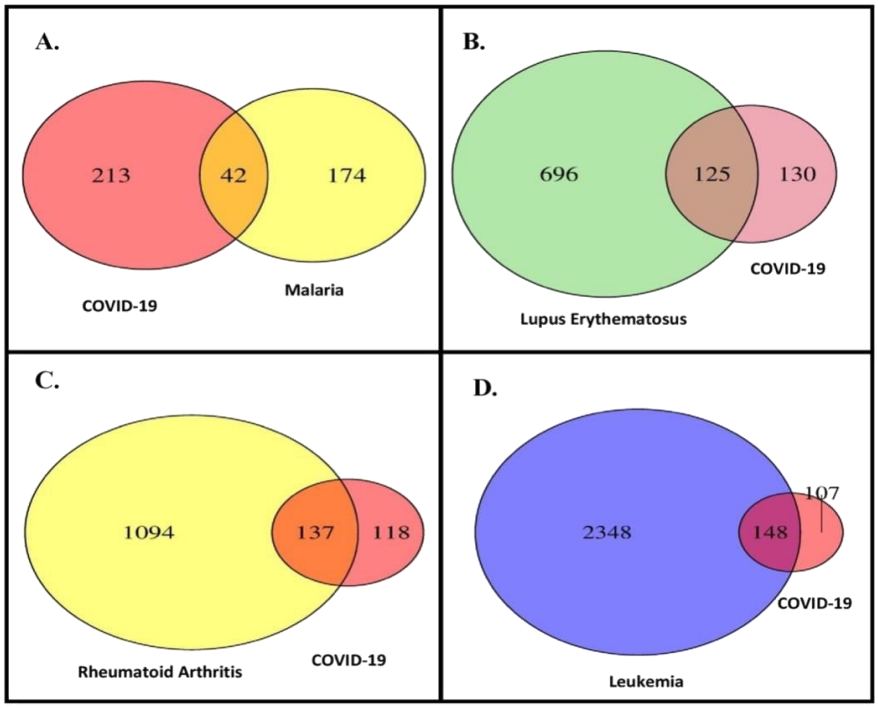
(A) down regulated genes on day 4 of SARS-CoV-2 infection; (B) up regulated genes on day 4 of SARS-CoV-2 infection; (C) down regulated genes on day 5 of SARS-CoV-2 infection; (D) up regulated genes on day 5 of SARS-CoV-2 infection; (E) down regulated genes on day 6 of SARS-CoV-2 infection; (F) up regulated genes on day 6 of SARS-CoV-2 infection; (G) up regulated genes on day 7 of SARS-CoV-2 infection; (H) up regulated genes on day 8 of SARS-CoV-2 infection; (I) down regulated genes on day 9 of SARS-CoV-2 infection; (J) up regulated genes on day 9 of SARS-CoV-2 infection; (K) down regulated genes on day 10 of SARS-CoV-2 infection; (L) up regulated genes on day 10 of SARS-CoV-2 infection; (M) down regulated genes on day 11 of SARS-CoV-2 infection; (N) up regulated genes on day 11 of SARS-CoV-2 infection; (O) down regulated genes on day 12 of SARS-CoV-2 infection; (P) up regulated genes on day 12 of SARS-CoV-2 infection; (Q) down regulated genes on day 13 of SARS-CoV-2 infection; (R) up regulated genes on day 13 of SARS-CoV-2 infection; (S) down regulated genes on day 18 of SARS-CoV-2 infection; (T) up regulated genes on day 18 of SARS-CoV-2 infection [21]

Fig. 2: Disease-gene networks



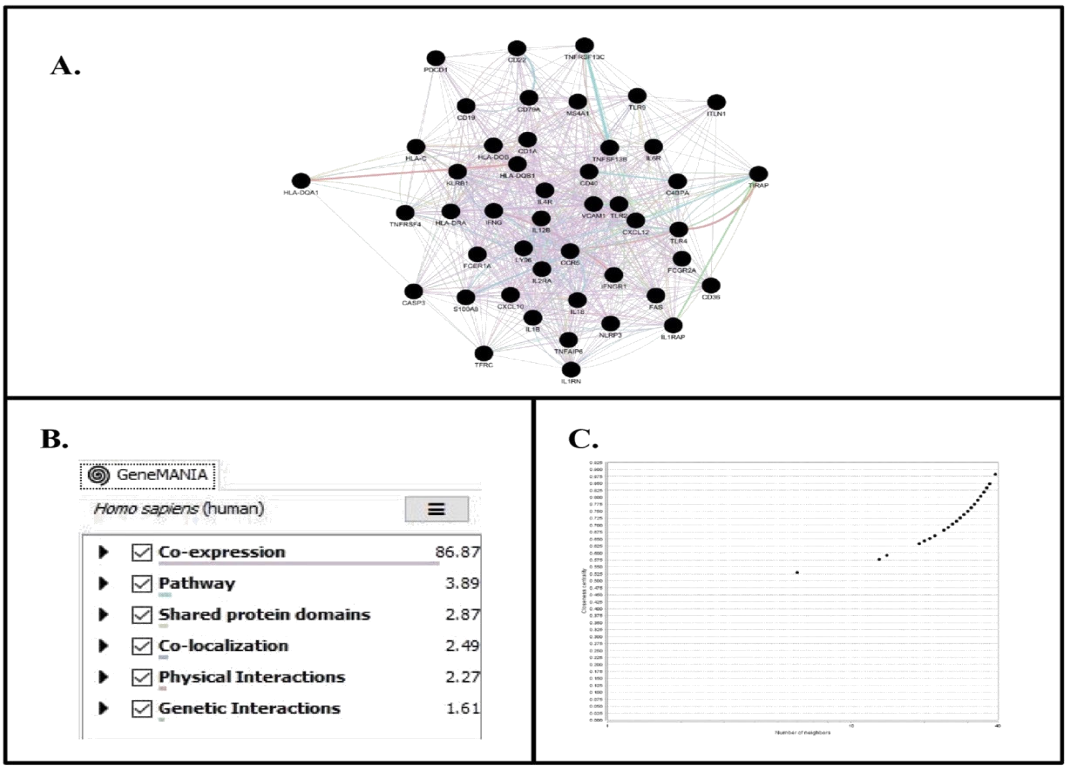
(*A) down regulated genes on day 4 of SARS-CoV-2 infection; (B) up regulated genes on day 4 of SARS-CoV-2 infection; (C) down regulated genes on day 5 of SARS-CoV-2 infection; (D) up regulated genes on day 5 of SARS-CoV-2 infection; (E) down regulated genes on day 6 of SARS-CoV-2 infection; (F) up regulated genes on day 6 of SARS-CoV-2 infection; (G) up regulated genes on day 7 of SARS-CoV-2 infection; (H) up regulated genes on day 8 of SARS-CoV-2 infection; (I) up regulated genes on day 9 of SARS-CoV-2 infection; (J) down regulated genes on day 10 of SARS-CoV-2 infection; (K) up regulated genes on day 10 of SARS-CoV-2 infection; (L) up regulated genes on day 11 of SARS-CoV-2 infection; (M) up regulated genes on day 12 of SARS-CoV-2 infection; (N) up regulated genes on day 13 of SARS-CoV-2 infection; (O) down regulated genes on day 18 of SARS-CoV-2 infection; (P) up regulated genes on day 18 of SARS-CoV-2 infection*

Fig. 3: Disease-gene heatmap



(A) between Malaria and COVID-19; (B) between Lupus Erythematosus and COVID-19; (C) between Rheumatoid Arthritis and COVID-19; (D) between Leukemia and COVID-19 [24]

Fig. 4: Venn diagram for the common dysregulated genes



(A)Composite network of the dysregulated genes; (B) Percentage of genes expressing together, interacting physically and genet ically, sharing protein domains, locations and pathways; (C) Betweenness Centrality vs Neighborhood connectivity chart for the dysregulated genes [25]

Fig. 5: Gene MANIA results for dysregulated genes common in COVID-19, Malaria, Lupus Erythematosus, Rheumatoid Arthritis, and Leukemia

3.3.2. Cluster of differentiation 40 (CD40)

CD40 belongs to the tumor necrosis factor receptor super family and plays an important role in signaling to promote the B cell growth and differentiation and, activation of the dendritic cells [29].

3.3.3. C-X-C motif chemokine 10 (CXCL10)

CXCL10 is a member of the CXC chemokine family which binds to the CXCR3 receptor to mediate immune responses. It is associated with a variety of human diseases including infectious diseases, chronic inflammation, immune dysfunction, and tumor development. CXCL10 can be used as a major biological marker to mediate disease severity [30].

3.3.4. Human Leukocyte Antigen DQ Beta 1 (HLA-DQB1)

HLA-DQB1 gene belongs to the class II of the Major histocompatibility complex (MHC) genes. This gene is associated with several auto-immune disorders like Type I diabetes and Multiple Sclerosis. It has been observed in a recent study that the higher rate of allele HLA-DQB1 is associated with the mortality rate in COVID-19 patients [31].

3.3.5. Human Leukocyte Antigen DR Alpha (HLA -DRA)

HLA-DRA gene also belongs to the class II of the Major histocompatibility complex (MHC) genes. The protein encoded by this gene is expressed on the surface of antigen presenting cells like B cells, dendritic cells and macrophages. In previous studies conducted on COVID-19 patients it was observed that the HLA-DR expression was higher after Day 4 of the infection in COVID-19 patients in comparison to the non-patients [32].

3.3.6. Interferon gamma (IFNG)

It is a pleiotropic cytokine that regulates adaptive and innate immune networks. It is a signature cytokine of the activated T lymphocytes and a powerful activator of the macrophages [33]. Recent studies on COVID-19 have also revealed that plasma levels of COVID-19 patients contain higher levels of IFNG and the elevated levels are associated with higher viral load and lung damage [34].

3.3.7. Interleukin Receptor Accessory Protein (IL1RAP or IL1R3)

This gene encodes for a component of IL1R to activate the IL-1 cytokine. Elevated IL-1 levels have been observed in COVID-19 patients and an early blockage in the production of this cytokine can help prevent complications in the patients [35].

3.3.8. Interleukin 2 Receptor Subunit Alpha (IL 2RA)

The IL2 receptor is composed of IL2R alpha, beta and gamma chains. IL2R is expressed on the surface of immune cells and binds to the IL-2 cytokine. IL2 is crucial for the proliferation and differentiation of the T-cells. The IL2R levels have been observed to be significantly higher in severe cases of COVID-19 [36].

3.3.9. Interleukin 4 Receptor (IL4R)

IL4R is the receptor for the cytokine IL4. IL4 plays a significant role in immune regulation by T helper cells. It is involved in the activation, differentiation and proliferation of B lymphocytes. In several recent studies, a high level of IL4 cytokine is observed in the plasma of COVID-19 patients and is related to the cytokine storm in the severe patients ultimately leading to the multiple organ failure [37].

3.3.10. Interleukin 6 Receptor (IL6R)

IL-6 is a pleiotropic cytokine that contributes to the host defense response against infections and tissue damage. It maintains its multiple physiological functions through its receptor IL-6R. IL-6 plays a crucial role in chronic inflammation and autoimmunity and is the main cytokine that leads to the cytokine storm in the COVID-19 patients leading to complications like ARDS and respiratory failure [38]. In earlier stages of inflammation, IL-6 is produced by the monocytes and macrophages that are stimulated by the TLRs which are found to be up regulated in our study (TLR9). The inhibition of TLRs in the early stages can lead to the lesser production of cytokines which in turn can reduce the mortality rate in COVID-19 patients [39].

3.3.11. Myeloid Related Protein (S100A8 or MRP8)

S100A8 gene belongs to the S100 family and it is a zinc- and calcium-binding protein that plays an important role in inflammation and immune response. In a study conducted earlier, it was observed that severe COVID-19 patients had higher serum levels of S100A8 and possibly associated with the tissue damage and cytokine storm [40].

3.3.12. Toll-like receptor 9 (TLR9)

TLR9 gene encodes for toll-like receptor 9 that plays a role in the activation of transcription factors such as

interferon-regulatory factor-3 (IRF3) and nuclear factor (NF)-κB. It induces the production of interferon-β and inflammatory cytokines. This receptor is known to be expressed in hepatocytes, macrophages, and natural killer (NK) cells [41].

Table 3: GeneMANIA results for the 12 hub genes of COVID-19 identified in this study

Gene Name	Annotation Name	Annotations	p-value	Betweenness Centrality	Degree
HLA-DQB1	intrinsic component of organelle membrane | interferon-gamma-mediated signaling pathway | vacuolar membrane | clathrin-coated vesicle membrane | antigen processing and presentation of exogenous peptide antigen via MHC class II | lymphocyte mediated immunity | humoral immune response	GO:0031300 | GO:0050865 | GO:0030666 | GO:0031301 | GO:0060333 | GO:0005774 | GO:0030665 | GO:0098552	<0.001	0.01528263	36
CXCL10	chemokine activity | leukocyte chemotaxis | cytokine receptor binding	GO:0032388 | GO:0006874 | GO:0006875 | GO:0008009 | GO:0030595 | GO:0060326 | GO:0070098 | GO:0050900	<0.001	0.01053579	34
IL4R	receptor complex | cytokine receptor activity	GO:0043235 | GO:0004896	<0.001	0.01607799	43
CD40	cellular metal ion homeostasis | positive regulation of MAPK cascade | positive regulation of JAK-STAT cascade | mononuclear cell proliferation | B cell activation	GO:0006875 | GO:0043410 | GO:0043406 | GO:0071260 | GO:0043235 | GO:0050730 | GO:0071214 | GO:0042510	<0.001	0.01355904	44
HLA-DRA	intrinsic component of organelle membrane | interferon-gamma-mediated signaling pathway | antigen processing and presentation of exogenous peptide antigen via MHC class II | positive regulation of lymphocyte activation	GO:0031300 | GO:0050865 | GO:0030666 | GO:0031301 | GO:0060333 | GO:0005774 | GO:0030665 | GO:0044440	<0.001	0.02247178	47
IL6R	inflammatory response | positive regulation of leukocyte chemotaxis | acute inflammatory response | positive regulation of MAPK cascade	GO:0006954 | GO:0002690 | GO:0002526 | GO:0043410 | GO:0043235 | GO:0050730 | GO:0050829 | GO:0098542 | GO:0007260 | GO:0032642 | GO:0019955 | GO:0033002	<0.001	0.00709154	28
IL2RA	-	-	<0.001	0.00855696	45
IFNG	interferon-gamma-mediated signaling pathway | leukocyte differentiation | regulation of innate immune response | cytokine activity | humoral immune response | regulation of cytokine-mediated signaling pathway | protein import | vitamin D metabolic process | myeloid leukocyte differentiation	GO:0060333 | GO:0031343 | GO:0050730 | GO:0007260 | GO:0002521 | GO:0006606 | GO:0045088	<0.001	0.01151234	47
S100A8	inflammatory response | cellular metal	GO:0006954 | GO:000687	<0.001	0.00959167	35

	ion homeostasis|leukocyte cell-cell adhesion|leukocyte chemotaxis|activation of cysteine-type endopeptidase activity	5|GO:0007159|GO:0030595|GO:0097202|GO:0050727|GO:2001056			
IL1RAP	inflammatory response	GO:0006954	<0.001	0.00756228	35
CD1A	antigen processing and presentation of exogenous antigen	GO:0019884|GO:0019882|GO:0003823	<0.001	0.01636677	36
TLR9	positive regulation of interleukin-10 production|regulation of interferon-alpha production|protein localization to nucleus|interleukin-8 production|positive regulation of intracellular protein transport|positive regulation of tumor necrosis factor production|tumor necrosis factor superfamily cytokine production|interleukin-6 production|positive regulation of chemokine production|positive regulation of innate immune response	GO:0006954|GO:0032609|GO:0033157|GO:0043410|GO:0042306|GO:0005774|GO:0050729|GO:0070201|GO:0032642|GO:0051223|GO:1900180|GO:0017038|GO:0044437|GO:0051241|GO:0002221	<0.001	0.00705703	31

3.4. Gene-drug interaction results for the hub genes

Gene-drug interactions for the hub genes found in COVID-19 were obtained from the rDGIdb package on R [27]. The detailed results are given in Table 4. We obtained the same resultant drug candidates for genes HLA-DQB1 and HLA-DRA. The drugs interacting with the hub genes are as follows:

3.4.1. CD1A

3.4.1.1. Triamcinolone inhibitor

It is a potent steroid that helps reduce inflammation in the body. It is used to treat allergies and Rheumatic disorders. It decreases the immune reaction by inhibiting the expressions of gene CD1A and HLA-DRA [42].

3.4.2. CD40

3.4.2.1. Dacetuzumab-inhibitor

It is a humanized monoclonal antibody used for the treatment of CD-40 positive cancers, leukemia and multiple myeloma. It is a potential repurposable drug for COVID-19 due to its inhibiting activity on CD40 gene which is found to be up regulated in COVID-19 patients.

3.4.3. CXCL10

3.4.3.1. Zidovudine -n/a

This drug belongs to the nucleoside reverse transcriptase inhibitors (NRTIs) class of the drugs. It is commonly used in pregnant mothers to prevent the transfer of Human Immunodeficiency Virus (HIV) infection to the unborn baby. Current research on Zidovudine for the treatment of COVID-19 shows that it inhibits the RNA dependent RNA Polymerase of SARS-CoV-2 [43].

3.4.3.2. Ritonavir -n/a

Ritonavir is an HIV protease inhibitor and interferes with the reproductive cycle of HIV. Several studies are being conducted on Ritonavir combined with Lopinavir for the potential treatment of COVID-19. However, the use of Lopinavir-Ritonavir has been discontinued in hospitalized COVID-19 patients because they do not reduce the mortality rate as said by WHO.

3.4.3.3. Methylprednisolone-Inhibitor

It is a synthetic glucocorticoid with immunosuppressive and anti-inflammatory activity. It is five times more potent than cortisol in its action and is extensively used in different fields of medicine. Methylprednisolone has been previously used in the treatment of multiple sclerosis (MS) patients to prevent relapses. It reduces the elevated expression of CXCL10 in the MS patients notably and can be potentially used to control the upregulation of the gene CXCL10 in COVID-19 patients to control the cytokine storm [44, 45].

3.4.4. HLA-DQ1 / HLA-DRA

3.4.4.1. Nevirapine – n/a

It is a benzodiazepine non-nucleoside reverse transcriptase inhibitor (NNRTI) and is known to decrease the

viral load of HIV. It slows down damage to the immune system by increasing the CD4 count. Nevirapine has been previously used to treat SARS patients and can be a potential candidate for the treatment of COVID-19 [46].

3.4.5. IFNG

3.4.5.1. Olsalazine-Inhibitor

Olsalazine (3, 3'-azobis salicylic acid) is an anti-inflammatory drug that has been used to inhibit the development of colorectal cancer in patients. Presently, it is used for treating inflammatory bowel disease (IBD), ulcerative colitis (UC), and other gastrointestinal problems [47, 48]. It is proposed that Olsalazine can also be used as a broad-spectrum anti-cancer agent [49].

3.4.6. IL2RA

3.4.6.1. Daclizumab-inhibitor

It is a monoclonal antibody that inhibits IL2 cytokine. It works by binding to the alpha unit of IL2 cytokine (CD25). It is used for the treatment of patients with relapsing forms of Multiple Sclerosis.

3.4.7. IL4R

3.4.7.1. Dupilumab- inhibitor

It is a monoclonal antibody and antagonist against the cytokine IL4. Since high levels of IL4 cytokines are observed in COVID-19 patients, Dupilumab can act as a potential drug for the treatment of COVID-19.

3.4.8. IL6R

3.4.8.1. Siltuximab-inhibitor

Siltuximab is a monoclonal antibody that neutralizes IL-6 cytokine that is key in COVID-19 related cytokine storm. This drug is already being studied for the treatment of COVID-19 patients suffering from ARDS due to its inhibitory effect on IL-6 cytokine [50].

3.4.8.2. Tocilizumab-inhibitor

It is a monoclonal antibody that inhibits the cytokine IL6 and has been used for the treatment of Rhematoid Arthritis. It is currently being used in the treatment of patients suffering from COVID-19 with a potential risk of cytokine storm [51].

3.4.9. S100A8

3.4.9.1. Zinc Chloride-inhibitor

Zinc chloride acts by inhibiting the apoptosis activity of S100A8 gene. Besides that zinc is capable of enhancing innate and adaptive immunity in the patients. Zinc chloride can be used as a potential supplement in addition with the antiviral drugs to benefit patients suffering from COVID-19 [52].

3.4.10. TLR9

3.4.10.1. Hydroxychloroquine-antagonist

Hydroxychloroquine is an anti-malarial drug that has an inhibitory effect on the TLR signaling. It has also been used to treat autoimmune diseases like Rheumatoid Arthritis, Asthma, and Systemic Lupus Erythematosus (SLE). Hydroxychloroquine has also shown high efficacy in the treatment of pneumonia associated with the SARS-CoV-2 infection [53, 54]. Hydroxychloroquine controls the regulation of TLRs, which is known to contribute to the production of IL-6 in the early stages of inflammation. This could be the reason for its success as a drug for the treatment of COVID-19 patients [39].

3.5. Differential expression of hub genes in immune cells

After analyzing the single cell RNA sequencing data of COVID-19 patients (fig. 7), it was found that CD1A gene is highly expressed in the immune cells of severe and recovered patients whereas high expression of CD40 and HLA-DQB1 is observed in the B cells of patients suffering from mild and severe infections. Another HLA variant HLA-DRA expression increased in B cells, CD14+ monocytes, and FCGR3A+ monocytes in COVID-19 patients. The high expression of receptors of cytokines like IL1, IL4, and IL6 has also been observed in B cells and neutrophils in severe and mild cases on day 10 but the expressions drastically decreased in the data taken on day 10 of the infection. An increase in the expression of S100A8 gene has also been seen in the mild and severe COVID-19 cases in CD14+ monocytes and neutrophils.

Table 4: rDGIdb results for gene-drug interaction of the hub genes

Gene Name	Drug	Interaction Types	Query Score	Interaction Score	Sources	PMIDs
CD1A	Triamcinolone	Inhibitor	0.38	2.47	NCI	11767704
CD40	Hydroquinone	Inhibitor	0.51	0.61	NCI	17118622
CXCL10	Ritonavir	n/a	0.44	0.44	NCI	11141242

	Zidovudine	n/a	0.27	0.27	NCI	11141242
	Methylprednisolone	Inhibitor	1.63	0.55	NCI	17220550
HLA-DQB1	Nevirapine	n/a	0.19	0.21	PharmGKB	-
HLA-DRA	Nevirapine	n/a	0.19	0.21	PharmGKB	-
IFNG	Olsalazine	Inhibitor	3.27	2.24	DrugBank	17034586|9797390
IL1RAP	-	-	-	-	-	
IL2RA	Daclizumab	Antibody, Inhibitor	3.73	1.68	DrugBank|TdgClinicalTrial|ChemblInteractions|TEND|GuideToPharmacology	16908318|19788505|16304857|11752352|8600944|10607689|17564638
IL4R	Dupilumab	Antibody, Antagonist	21.78	56.81	DrugBank|ChemblInteractions|GuideToPharmacology|TTD	23688323
IL6R	Siltuximab	Antagonist, Antibody	6.53	9.08	DrugBank|ChemblInteractions|GuideToPharmacology|TTD	
	Tocilizumab	Antibody, Inhibitor	11.98	17.36	DrugBank|MyCancerGenome|TdgClinicalTrial|ChemblInteractions|TEND|PharmGKB|TTD	16899109|21940820|24978393|27958380
S100A8	Zinc Chloride	Inhibitor	0.09	0.23	DrugBank	17050004
TLR9	Hydroxychloroquine	Antagonist	1.81	1.39	DrugBank|TdgClinicalTrial|GuideToPharmacology	14579285|18220957

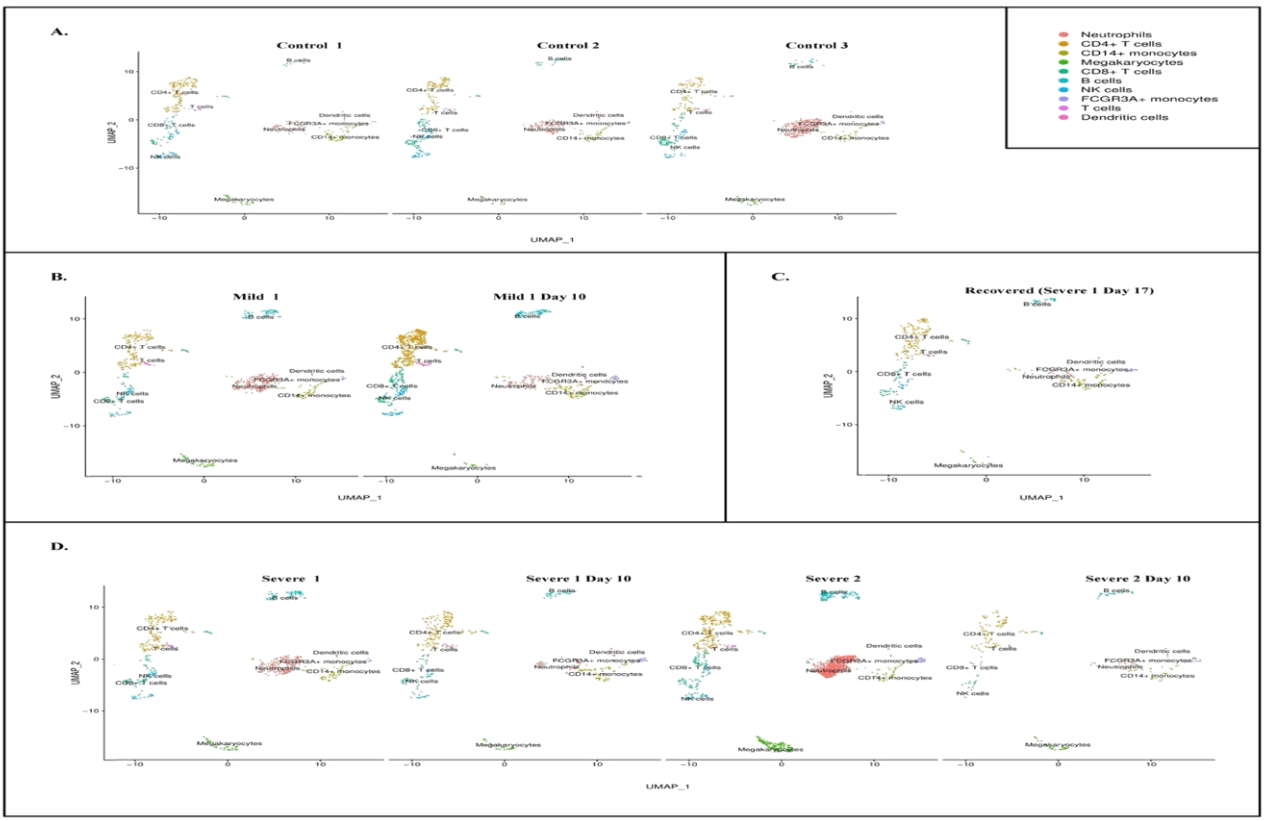


Fig. 6: Immune cell compositions of (A) Uninfected human subjects, (B) Mildly affected COVID-19 patients, (C) Recovered COVID-19 patient, and (D) Severely affected COVID-19 patients [20]

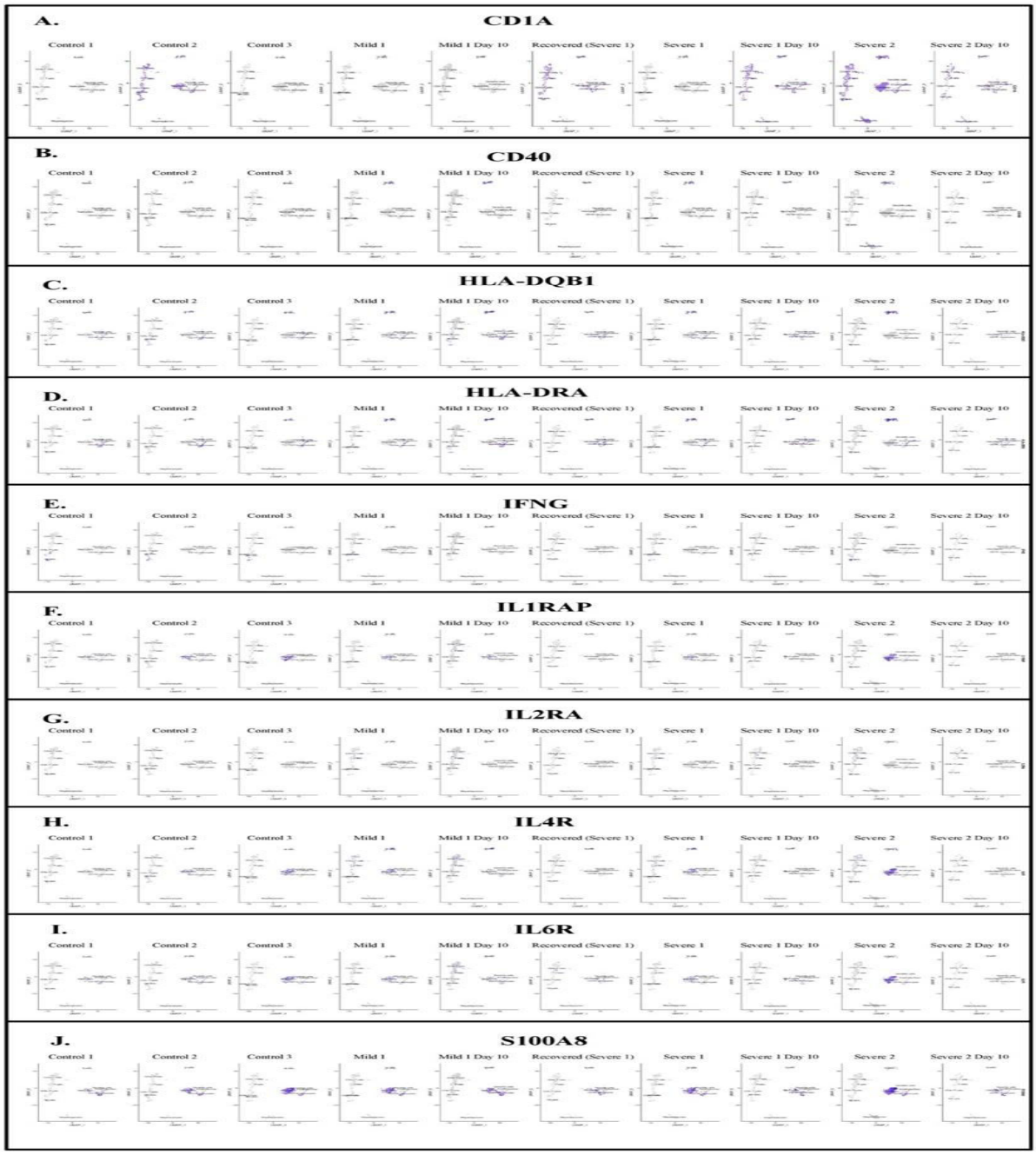


Fig. 7: Feature plots for hub genes (A) CD1A, (B) CD40, (C) HLA-DQB1, (D) HLA-DRA, (E) IFNG, (F) IL1RAP, (G) IL2R, (H) IL4R, (I) IL6R, and (J) S100A8 [20]

4. DISCUSSION

The potential molecular biomarkers of COVID-19 identified in this study include CD1A, CD40, CXCL10, HLA-DQB1, HLA-DRA, IFNG, IL1RAP, IL2RA, IL4R, IL6R, S100A8, and TLR9. These biomarkers were found to be up regulated in the COVID-19 patients in our study. It has been previously studied that elevated levels of IFNG and IL1-RA are associated with the increased viral loads, lung damage and risk of death whereas the increased levels of IL2 receptors in the patient's plasma indicates the severity of disease [3, 55, 56]. Cytokine IL4 plays an important role in the adaptive immunity and regulated the activation of T helper cells and is associated with the severity of respiratory symptoms in COVID-19 patients [57]. The cytokine receptor IL6R observed to be up regulated in our study codes for the cytokine IL6. It is directly related to the severity of the disease and mortality in

COVID-19 patients. These elevated cytokines are responsible for the cytokine patients and work by damaging T cells and affecting the ability of dendritic cells and macrophages to eliminate the virus [58, 59].

Strong activation of TLRs has been seen in various animal coronavirus infections like bovine coronavirus infection and porcine coronavirus infections which can lead to the uncontrollable release of cytokines like IL-6 [60, 61]. Activation of the toll-like receptor (TLR) is one of the host's first defense mechanism against the viral infections [62-64]. In this study, the expression of TLR9 gene has been observed to be up regulated in the COVID-19 patients. The dysregulated TLR signaling can potentially disturb the immune homeostasis by increasing the secretion of cytokines and chemokines, which can lead to inflammation [65].

In this study, chemokine CXCL10 has been found to be up regulated in the COVID-19 patients. Chemokines enable the migration of inflammatory cells to the site of inflammation. In the recent studies it is seen that the patients admitted to the ICU show elevated levels of this chemokine in their blood plasma and it plays a specific role in ARDS in the patients. In a mouse experimental model, it was seen that up regulation of CXCL10 marked the beginning of lung damage [3, 66]. Up regulation of CXCL10 can be a marker of onset of ARDS and therefore helpful in the treatment of critical patients. A nucleotide analogue called Zidovudine has an inhibitory effect on CXCL10 and has also been observed to inhibit the SARS-CoV-2 RNA dependent polymerase (RdRp) making it a potential drug candidate for COVID-19 [43]. Anti retrovirals targeting CXCL10 like Ritonavir has also been found to reduce the viral load of SARS-CoV-2 [67].

S100A8, a genetic marker identified in this study can help identify the COVID-19 patients that may face death ahead of time. It is seen that there is a direct correlation between the increase in the concentrations of this marker with the production of proinflammatory cytokines like IL1RA, IL7, IL17A, and IL10 [57]. Zinc Chloride is a potential drug recognized in this study that inhibits S100A8 gene. It has been observed that zinc supplement can help in increasing the ciliary clearance leading to the reduced risk of infections and improved virus removal [68]. Zinc is also known to decrease SIRT-1 (Sirtuin 1) activity which is crucial for ACE-2 (Angiotensin converting enzyme-2) expression. The reduced ACE-2 expression leads to the blocking of virus from entering the cells [69, 70].

Another genetic marker that determines the severity of this disease is the allele type of HLA gene. The two HLA markers identified in this study are HLA-DQB1 and HLA-DRA. Some types of alleles dampen the immune response in the patients and directly affect their ability to respond to the virus [31]. A non-nucleoside reverse transcriptase inhibitor called Nevirapine controls the expression of HLA-DQB1 and HLA-DRA genes. According to the recent studies, this drug inhibits the replication of SARS-CoV-2 at high concentrations [46].

A probable relationship between the development of Rheumatoid Arthritis and the occurrence of respiratory viral infections has been seen in the previous studies also [71]. A link of parainfluenza and coronavirus with Rheumatoid Arthritis has also been reported in a Korean study [72]. An uncontrolled immune-mediated inflammatory response takes place during the progression of COVID-19. T lymphocytes are the target cells in SARS caused by COVID-19 which triggers the abundant cytokine secretion leading to the exhaustion of immune response. The cascade of the pathways of immunogenic disturbances is similar to the one found in the progression of Rheumatoid Arthritis [73-75]. Some of the potential drug candidates listed in this study like hydroxychloroquine, methylprednisolone, Zidovudine, Ritonavir, Nevirapine, Duplilumab, Siltuximab, Tocilizumab, and Zinc Chloride have been known to show a positive effect in the treatment of coronavirus infections like MERS, SARS-CoV, and SARS-CoV-2 in the previous studies. Increasing understanding about the pathophysiology of the novel coronavirus infection is steering the introduction of drugs commonly used for the treatment of the autoimmune diseases like Rheumatoid Arthritis, Lupus Erythematosus. Drugs like chloroquine, quinine and, hydroxychloroquine are used for the treatment of Rheumatoid Arthritis, Lupus Erythematosus, and Malaria have now been included for the treatment of SARS-CoV-2 infections alongside antiviral drugs [76-78]. Apart from the autoimmune diseases, our study also found a relationship between COVID-19 and Leukemia. A few studies indicate that a class of drugs called tyrosine kinase inhibitors used for the treatment of Leukemia can also show positive results in COVID-19 patients due to their anti-inflammatory properties and inhibitory activity on cytokines like IL4, IL6, and IL2 [79].

The overproduction of the cytokines in the blood plasma of the patients causes a cytokine storm in the body which is fatal. To stop the production of the

responsible cytokines like IL6, IL2, IL4, and IL1, monoclonal antibodies like Tocilizumab and Siltuximab are already in the clinical trials. Tocilizumab can prove to be beneficial for the moderate and severely ill patients with a high level of IL6 in their serum. It reduces the IL6 plasma levels by binding to the IL6R receptor leading to the inhibition of inflammatory activity [51]. The monoclonal antibody Siltuximab is also a potential repurposable drug as it stops the transcription of genes encoding for cytokines such as IL-2 and IL-6 to reduce multiple organ failure in patients due to their overabundance. A few studies also indicate that the use of methylprednisolone reduces the risk of morbidity and mortality specifically in the COVID-19 patients with ARDS. Thus, the administration of methylprednisolone can prove to be advantageous for the treatment of patients who have already developed COVID-19 related ARDS [80]. Another monoclonal antibody Dupilumab that inhibits IL4 cytokine can prove to be a useful drug due to its immune modulating activity and ability to control virus induced asthma [81]. However, not much research has been done on this drug for the treatment of COVID-19. There is no previous evidence or studies to support the usage of drugs Olsalazine and Daclizumab in the treatment of coronavirus infections. Thus, they must undergo proper study and randomized trials before being considered as the drugs to treat SARS-CoV-2 infections.

These hub genes were analyzed using the single cell RNA sequencing data of COVID-19 patients to obtain the information about the immunological profiles of patients and in which type of cells these genes are highly expressed. The hub genes obtained in this study (HLA-DRA, HLA-DQB1, IL6R, IL4R, IL1RAP, CD40, and CD1A) were found to be highly expressed in neutrophils, B cells, CD14+ monocytes, and FCGR3A + monocytes. The number of neutrophils and B cells was also observed to be high in mild and severe patients than the control subjects and recovered patients. Furthermore, an increase in the number of cells in mild and severe patients was only observed on day 1 and not on day 10.

5. CONCLUSION

At this stage, it is extremely essential to contain the COVID-19 pandemic to save the lives of the people around the world. Due to the unavailability of the drugs and vaccines specific for SARS-CoV-2 infection, drug repurposing seems to be the paramount tool to treat the people infected with SARS-CoV-2 worldwide. In this study, we hereby propose about 13 potential drug repurposing candidates involved in various anti-viral, immunosuppressive, anti-inflammatory, and anti-cancer activities. Besides, we also suggest screening additional targets which may potentially be important owing to earlier known existent diseases like diabetes, hypertension, and heart diseases. This of course requires more systematic study but in case of the potential repurposing of the drugs, randomized clinical trials are mandatory before they are used to treat SARS-CoV-2 infection.

6. REFERENCES

1. Zhang J Jin, Dong X, Cao Y Yuan, Yuan Y dong, Yang Y Bin, Yan Y Qin, et al. *Allergy: European Journal of Allergy and Clinical Immunology*, 2020; **75**:1730-1741.
2. Guan W, Ni Z, Hu Yu, Liang W, Ou C, He J, et al. *New England Journal of Medicine*, 2020; **382**:1708-1720.
3. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. *The Lancet*, 2020; **395**:497-506.
4. Weiss SR, Navas-Martin S. *Microbiology and Molecular Biology Reviews*, 2005; **69**:635-664.
5. Fehr AR, Perlman S. *Methods Protoc.,* 2015, 1-23.
6. Guo YR, Cao QD, Hong ZS, Tan YY, Chen SD, Jin HJ, et al. *Military Medical Research*, 2020; **7**:1-10.
7. Wu Z, McGoogan JM. *JAMA - Journal of the American Medical Association*, 2020; **323**:1239-1242.
8. Yang X, Yu Y, Xu J, Shu H, Xia J, Liu H, et al. *The Lancet Respiratory Medicine*, 2020; **8**:475-481.
9. Cha Y, Erez T, Reynolds IJ, Kumar D, Ross J, Koytiger G, et al. *British Journal of Pharmacology*, 2018; **175**:168-180.
10. Yang W, Zhao X, Han Y, Duan L, Lu X, Wang X, et al. *Cancer Cell International*, 2019; **19:**1-15.
11. Phadke M, Saunik S. *Drug Development Research*, 2020; **81**:541-543.
12. Zhou Y, Hou Y, Shen J, Huang Y, Martin W, Cheng F. *Cell Discovery*, 2020; **6**:1-18.
13. 3.5.1. RDCT. *R Foundation for Statistical Computing*, 2018; **2**:https://www.R-project.org.
14. Wickham H, Francois R. *R Package Version 042*, 2018.
15. Koncevicius, K. *R package version 0.1.8*, 2020.
16. Sayanthooran S, Gunerathne L, Abeysekera TDJ, Magana-Arachchi DN. *International Urology and Nephrology*, 2018; **50**:1667-1677.

17. Collier JM, Chai B, Cole JR, Michalsen MM, Cupples AM. *Applied Microbiology and Biotechnology*, 2019; **103**:7161-7175.
18. Liao J, Yu Z, Chen Y, Bao M, Zou C, Zhang H, et al. *Scientific Data*, 2020; **7**:1-9
19. Wen W, Su W, Tang H, Le W, Zhang X, Zheng Y, et al. *Cell Discovery*, 2020; **6**:1-18.
20. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, et al. *Cell*, 2019; **177**:1888-1902.e21.
21. Piñero J, Bravo Á, Queralt-Rosinach N, Gutiérrez-Sacristán A, Deu-Pons J, Centeno E, et al. *Nucleic Acids Research*, 2017; **45**:D833-839.
22. Gupta A, Puri S, Puri V. *Scientific Reports*, 2019; **9**:161-176.
23. Mittal K, Mani RJ, Katare DP. *Scientific Reports*, 2016; **6**:1-8.
24. Chen H. *R Package*, 2011; **12:**1-7.
25. Montojo J, Zuberi K, Rodriguez H, Kazi F, Wright G, Donaldson SL, et al. *Bioinformatics*, 2010; **26**:2927-2928.
26. Yu H, Kim PM, Sprecher E, Trifonov V, Gerstein M. *PLoS Computational Biology*, 2007; **3**:713-720.
27. Wagner AH, Coffman AC, Ainscough BJ, Spies NC, Skidmore ZL, Campbell KM, et al. *Nucleic Acids Research*, 2016; **44**:D1036-1044.
28. Kim JH, Hu Y, Yongqing T, Kim J, Hughes VA, Le Nours J, et al. *Nature Immunology*,2016;**17**:1159-1166.
29. Schoenberger SP, Toes REM, Van Dervoort EIH, Offringa R, Melief CJM. *Nature*, 1998; **393**:480-483.
30. Liu M, Guo S, Stiles JK. *Oncology Letters*, 2011; **2:**583-589.
31. de Sousa E, Ligeiro D, Lérias JR, Zhang C, Agrati C, Osman M, et al. *International Journal of Infectious Diseases*, 2020; **98**:454-459.
32. Hue S, Beldi-Ferchiou A, Bendib I, Surenaud M, Fourati S, Frapard T, et al. *American Journal of Respiratory and Critical Care Medicine*, 2020; **202**:1509-1519.
33. Kang S, Brown HM, Hwang S. *Immune Network*, 2018; **18**.
34. Costela-Ruiz VJ, Illescas-Montes R, Puerta-Puerta JM, Ruiz C, Melguizo-Rodríguez L. *Cytokine and Growth Factor Reviews*, 2020; **54**:62-75.
35. Van De Veerdonk FL, Netea MG. *Critical Care*, 2020; **24**:1-6
36. Hou H, Zhang B, Huang H, Luo Y, Wu S, Tang G, et al. *Clinical and Experimental Immunology*, 2020; **201**:76-84.
37. Costela-Ruiz VJ, Illescas-Montes R, Puerta-Puerta JM, Ruiz C, Melguizo-Rodríguez L. *Cytokine and Growth Factor Reviews*, 2020; **54**:62-75.
38. Velazquez-Salinas L, Verdugo-Rodriguez A, Rodriguez LL,Borca M V.*Frontiers in Microbiology*, 2019; **10**:1057.
39. Magro G. *Cytokine: X*, 2020; **2**: 100029.
40. Chen L, Long X, Xu Q, Tan J, Wang G, Cao Y, et al. *Cellular and Molecular Immunology*,2020;**17**:992-994.
41. Fischer J, Koukoulioti E, Schott E, Fülöp B, Heyne R, Berg T, et al. *Scientific Reports*, 2018; **8**:1-8
42. Chen D, Bao W, Wang Q. *Chinese Journal of Plastic Surgery*, 2001; **17**:282-284.
43. Jockusch S, Tao C, Li X, Anderson TK, Chien M, Kumar S, et al. *Antiviral Research*, 2020; **180**: 104857.
44. Michałowska-Wender G, Losy J, Szczuciński A, Biernacka-Łukanty J, Wender M. *Pharmacological Reports*, 2006; **58**:920-923.
45. Moreira MA, Tilbery CP, Monteiro LP, Teixeira MM, Teixeira AL. *Acta Neurologica Scandinavica*, 2006; **114**:109-113.
46. De Clercq E. *Nature Reviews Drug Discovery*, 2007; **6**:941.
47. Koelink PJ, Hawinkels LJAC, Wiercinska E, Sier CFM, Dijke P ten, Lamers CBHW, et al. *Cancer Letters*, 2010; **287**:82-90.
48. Velayos FS, Terdiman JP, Walsh JM. *American Journal of Gastroenterology*, 2005; **100**:1345-1353.
49. Cortez-Maya S, Pedro-Hernández LD, Martínez-Klimova E, Ramírez-Ápan T, Martínez-García M. *Biomolecules*, 2019; **9**:360.
50. Palanques-Pastor T, López-Briz E, Poveda Andrés JL. *European Journal of Hospital Pharmacy*, 2020; **27**: 297-298.
51. Luo P, Liu Y, Qiu L, Liu X, Liu D, Li J. *Journal of Medical Virology*, 2020; **92**:814-818.
52. Rahman MT, Idid SZ. *Biological Trace Element Research*, 2020; 1-9.
53. Gao W, Xiong Y, Li Q, Yang H. *Frontiers in Physiology*, 2017; **8**.
54. Fantini J, Di Scala C, Chahinian H, Yahi N. *International Journal of Antimicrobial Agents*, 2020; **55**: 105960.

55. Liu K, Fang YY, Deng Y, Liu W, Wang MF, Ma JP, et al. *Chinese Medical Journal*, 2020; **133**:1025-1031.
56. Yang X, Yu Y, Xu J, Shu H, Xia J, Liu H, et al. *The Lancet Respiratory Medicine*, 2020; **8**:475-481.
57. Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, et al. *The Lancet*, 2020; **395**:507-513.
58. Law HKW, Chung YC, Hoi YN, Sin FS, Yuk OC, Luk W, et al. *Blood*, 2005; **106**:2366-2374.
59. Yoshikawa T, Hill T, Li K, Peters CJ, Tseng C-TK. *Journal of Virology*, 2009; **83**:3039-3048.
60. Xu Z, Zhong H, Huang S, Zhou Q, Du Y, Chen L, et al. *Veterinary Microbiology*, 2019; **228**:226-233.
61. Aich P, Wilson HL, Kaushik RS, Potter AA, Babiuk LA, Griebel P. *Journal of General Virology*, 2007; **88**:2749-2761.
62. Takeda K, Akira S. *International Immunology*, 2005; **17**:1-14.
63. Beutler BA. *Blood*, 2009; **113**:1399-1407.
64. Palm NW, Medzhitov R. *Immunological Reviews*, 2009; **227**:221-233.
65. Gao W, Xiong Y, Li Q, Yang H. *Frontiers in Physiology*, 2017; **8**:508.
66. Neville LF, Abdullah F, McDonnell PM, Young PR, Feuerstein GZ, Rabinovici R. *American Journal of Physiology-Lung Cellular and Molecular Physiology*, 1995; **269**:L884-890.
67. McKee DL, Sternberg A, Stange U, Laufer S, Naujokat C. *Pharmacological Research*, 2020; **157**:1423-1433.
68. Woodworth BA, Zhang S, Tamashiro E, Bhargave G, Palmer JN, Cohen NA. *American Journal of Rhinology and Allergy*, 2010; **24**:6-10.
69. Rosenkranz E, Metz CHD, Maywald M, Hilgers RD, Weßels I, Senff T, et al. *Molecular Nutrition and Food Research*, 2016; **60**:661-671.
70. Patel VB, Zhong JC, Grant MB, Oudit GY. *Circulation Research*, 2016; **118**:1313-1326.
71. Arleevskaya MI, Shafigullina AZ, Filina Y V., Lemerle J, Renaudineau Y. *Frontiers in Immunology*, 2017; **8**:1725.
72. Joo Y Bin, Lim YH, Kim KJ, Park KS, Park YJ. *Arthritis Research and Therapy*, 2019; **21**:1-8.
73. Georgiev T. *Rheumatology International*, 2020; **40**: 825-826.
74. Xu Z, Shi L, Wang Y, Zhang J, Huang L, Zhang C, et al. *The Lancet Respiratory Medicine*, 2020; **8**:420-422.
75. Zou L, Ruan F, Huang M, Liang L, Huang H, Hong Z, et al. *New England Journal of Medicine*, 2020; **382**:1177-1179.
76. Favalli EG, Ingegnoli F, De Lucia O, Cincinelli G, Cimaz R, Caporali R. *Autoimmunity Reviews*, 2020; **19**:102523.
77. World Health Organization. *Who*, 2020:10.
78. Sinha N, Balayla G. *Postgraduate Medical Journal*, 2020; **96**:550-555.
79. Galimberti S, Baldini C, Baratè C, Ricci F, Balducci S, Grassi S, et al. *Pharmacological Research*, 2020; **157**:104866.
80. Wu C, Chen X, Cai Y, Xia J, Zhou Xing, Xu S, et al. *JAMA Internal Medicine*, 2020; **180**:934-943.
81. Thangaraju P, Venkatesan N, Sudha TYS, Venkatesan S, Thangaraju E. *SN Comprehensive Clinical Medicine*, 2020; **2**:2126-2130.