

## **A Meta-analysis of comorbidities in COVID-19: Which diseases increase the susceptibility of SARS-CoV-2 infection?**

Manoj Kumar Singh<sup>1, 2, 4</sup>, Ahmed Mobeen<sup>1, 2, 4</sup>, Amit Chandra<sup>1, 2, 4</sup>, Sweta Joshi<sup>3</sup>, Srinivasan Ramachandran<sup>1, 2, 5</sup>

<sup>1</sup>CSIR-Institute of Genomics and Integrative Biology, New Delhi 110007, India

<sup>2</sup>Academy of Scientific and Innovative Research (AcSIR), Ghaziabad-201002, India

<sup>3</sup>Jamia Hamdard, Hamdard Nagar, New Delhi 110062

<sup>4</sup>Equally contributed

<sup>5</sup>**Corresponding author:** Srinivasan Ramachandran; e-mail: [ramu@igib.res.in](mailto:ramu@igib.res.in); Phone number: +91-11-29879130

## Abstract

**Background:** Comorbidities have been frequently reported in COVID-19 patients, which often lead to more severe outcomes. The underlying molecular mechanisms behind these clinical observations have not yet been explained. Herein, we investigated the disease-specific gene expression signatures that may induce susceptibility to SARS-CoV-2 infection.

**Methods:** We studied 30 frequently occurring acute, chronic, or infectious diseases of recent times that have shown comorbidity in one or another respiratory disease(s) caused by pathogenic human infecting coronaviruses, especially SARS-CoV-2. We retrieved array-based gene expression data for each disease and control from relevant datasets. Subsequently, all the datasets were quantile normalized, and log-2 transformed data was used for analysis.

**Results** The expression of *ACE2* receptor and host proteases, namely *FURIN* and *TMPRSS2* that are essential for cellular entry of SARS-CoV-2, was upregulated in all six studied subtypes of leukemia (hereafter, referred as leukemia). The expression of *ACE2* was also increased in psoriasis, lung cancer, Non-alcoholic fatty liver disease (NAFLD), breast cancer, and pulmonary arterial hypertension patients. The expression of *FURIN* was higher in psoriasis, NAFLD, lung cancer, and in type II diabetic liver, whereas it was lowered in breast cancer. Similarly, the expression of *TMPRSS2* was increased during lung cancer and type II diabetes; it was decreased during psoriasis, NAFLD, lung cancer, breast cancer, and cervical cancer.

Furthermore, a heightened expression of genes that are involved in immune response was observed in leukemia patients, as shown by the higher expression of *IFNA2*, *IFNA8*, *IFNA10*, *IFNA14*, *IFNA16*, *IFNA21*, *IFNB1*, *CXCL10*, and *IL6*. The expression of *JAK1*, *STAT1*, *IL6*, and *CXCL10* was higher in NAFLD. Besides, *JAK1* and *STAT1* were upregulated in type II diabetic muscles. In addition, most of the upregulated genes in COVID-19 patients showed a similar trend in leukemia, NAFLD, and psoriasis. Furthermore, SARS-CoV-2, SARS-CoV and MERS CoV, were found to commonly alter two genes, namely, CARBONIC ANHYDRASE 11 and CLUSTERIN.

**Conclusions:** The genes that may confer susceptibility to SARS-CoV-2 infection are mostly upregulated in leukemia patients; hence, leukemia patients are relatively more susceptible to develop COVID-19, followed by other chronic disorders, such as, NAFLD, type II diabetes, psoriasis, and hypertension. This study identifies key genes that are altered in the studied diseases types, which may aid in the infection of SARS-CoV-2 and underlie COVID-19 associated comorbidities.

**Keywords:** COVID-19, Comorbidity, SARS-CoV-2, leukemia, NAFLD, psoriasis, cancer, type II diabetes

## Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has caused the most dreaded pandemic of recent times, i.e., coronavirus disease (COVID-19). As per the global data released on September 09, 2020, by COVID-19 dashboard of Centre of Systems Science and Engineering (CSSE) at Johns Hopkins University, SARS-CoV-2 has infected over 27.57 million people of which more than 897,000 people have died. A significant proportion of the COVID-19 patients have been reported to suffer from other pathophysiological conditions as well. For instance, in a cohort of 1590 COVID-19 patients from China, Guan et al. (2020) reported that 399 patients (25.1%) were having at least one comorbidity, while 130 patients (8.2%) had two or more comorbidities<sup>1</sup>. The researchers reported hypertension, diabetes, cardiovascular diseases, and chronic kidney diseases among the most frequent comorbidities, which occurred in 16.9%, 8.2%, 3.7%, and 1.3% of all COVID-19 patients, respectively. Besides, they identified COPD and malignancy as critical risk factors associated with severe COVID-19 conditions. Another study by Chen et al. (2020), in a cohort of 99 COVID-19 patients in China, 50 patients (51% of all COVID-19 patients) were found to suffer from chronic medical illnesses<sup>2</sup>. Of these, the most prominent comorbid diseases were cardiovascular or cerebrovascular diseases, diabetes, digestive system disease, and malignant tumor that were identified in 40, 12, 11, and 1 patient, respectively. Similarly, others have also reported cancer of lungs<sup>3</sup> and blood<sup>4</sup>, NAFLD<sup>5</sup>, and HIV infections<sup>6</sup>, among others, as frequently occurring comorbidities that often worsen the outcome and increase the risk of mortality in COVID-19 patients.

Thus, it is imperative to understand that comorbidities are widely associated with COVID-19 that hamper the recovery of patients. However, the genetic factors that may induce susceptibility for these deadlier associations remain largely unknown until now. Logically, the specific gene expression patterns of an acute or chronic disorder may alter the immune response, and hence, may promote viral invasion. Also, if the disease physiology increases the availability of host proteins that acts as a viral receptor, viz. ACE2, the susceptibility of infection by SARS-CoV-2, can increase<sup>7</sup>. Thus, an understanding of the gene expression signatures in various diseases can be helpful in revealing the relative abundance of risk factors that predisposes a patient to COVID-19. Furthermore, the accurate information of such gene expression signatures, which can aid in the entry or pathogenesis of SARS-CoV-2, will be effective in studying pre-existing or novel drugs for counteracting COVID-19. This will be effective in reducing the spread and, thus, the doubling time of SARS-CoV-2 infection that currently stands at 6.4 days, as well as in preventing the fatalities in COVID-19 associated comorbidities<sup>8</sup>.

As mentioned earlier, several authors have reported one or other pathophysiological conditions associated with COVID-19; however, none of the available studies has so far dissected the molecular mechanisms behind COVID-19 associated comorbidities. The altered gene expression patterns in different diseases have not yet been correlated with such comorbidities. The aim of this study was to decipher such diseases and their associated genes that may induce susceptibility in the patients to subsequent SARS-CoV-2 infection. We performed a meta-analysis with gene expression datasets in 30 widely prevalent acute, chronic, and infectious diseases to identify the gene expression signatures underlying

COVID-19 associated comorbidities. We retrieved the gene expression profile of relevant control and disease samples and compared the expression of such genes that are implicated in the pathogenesis of COVID-19. We found that the pattern of gene expression during leukemia is most conducive to the development of SARS-CoV-2. This is followed by such chronic diseases, namely NAFLD, psoriasis, and type II diabetes, whose gene expression pattern is conducive to SARS-CoV-2 infection. Our study is the first report that links the differential gene expression in various diseases with COVID-19 associated comorbidities.

## Methods

### Data retrieval

Published and unpublished datasets of human acute, chronic, infectious diseases, and various types of cancer were retrieved from NCBI's GEO database. Largely, we retrieved datasets of patients and controls for the following conditions: Asthma (GSE64913), Chronic Obstructive Pulmonary Disorder (COPD, GSE112811), Cardiovascular diseases (GSE109048), Hypertension (GSE113439), Non-Alcoholic Fatty Liver Disease (NAFLD, GSE49541, GSE107037), Atherosclerosis (Athero, GSE28829), Type 2 Diabetes (T2D, GSE15653, GSE25462, GSE38642, GSE27949), Polycystic Ovary Syndrome (PCOS, GSE124226), Multiple Sclerosis (MS, GSE21942), Psoriasis (GSE78097), Blood Cancer or Leukemia (GSE51082, GSE9476), Breast Cancer (GSE65194), Cervical Cancer (GSE63514), Multiple Myeloma (MM, GSE85837), Lung Cancer (GSE136043), Lung adenocarcinoma or Non-small cell lung cancer (NSCLC, GSE118370), Liver Cancer (GSE88839), Pancreatic Ductal Adenocarcinoma (PDAC, GSE101448), AIDS (GSE73968), Tuberculosis (TB, GSE139825), Malaria (GSE119150), Acute Kidney Injury (AKI, GSE30718), and COVID-19 (GSE150316). Furthermore, six subtypes of leukemia were studied in GSE51082 and GSE9476 datasets; namely, Acute myeloid leukemia (AML), B-cell chronic lymphocytic leukemia (BCLL), Chronic myelogenous leukemia (CML), Myelodysplastic syndrome (MDS), B-acute lymphoblastic leukemia (BALL), and T-cell acute lymphoblastic leukemia (TALL). Similarly, the dataset on cardiovascular diseases included an equal number of patients suffering from coronary artery disease (CAD) and acute myocardial infarction (AMI), and the dataset included for studying breast cancer included the samples of the canonical breast cancer as well as triple-negative breast cancer tissues (TNBC). Also, the 29 samples that were reported in the GSE28829 dataset pertaining to atherosclerosis (Athero), 16 samples were from advanced atherosclerotic plaque (ATHERO-Adv), while 13 samples were taken from early atherosclerotic plaque (ATHERO-Early) regions. We did not get a common expression profile of diseased and control samples for leukemia and NAFLD. Therefore, we used a data-integration strategy as previously explained by Hamid et al. (2009) to analyze in leukemia and NAFLD condition using different datasets<sup>9,10</sup>. Besides, other relevant information about each dataset, namely, the platform of microarray experiment, type of tissue sample, sample size, experimental design, and data processing strategy, has been summarized in Table 1.

### Data normalization

The datasets used in this study were all quantile normalized, and log2 transformed. Briefly, raw expression values were quantile normalized using the `normalizeQuantiles` function of LIMMA package in R, irrespective of their normalization status to maintain the uniformity. After that, the average expression value of all probes for each gene in all the disease and

control samples was obtained using collapseRows function of WGCNA R package. Further, normalized values were subsequently log<sub>2</sub>-transformed, provided the dataset was not already log<sub>2</sub>-transformed.

## Data analysis

PRCOMP function in R base package was used for PCA to analyze the homogeneity and segregation of datasets based on linear co-relation and variance in gene expression for all subjects in each disease. Literature mining was done to identify 78 genes, including receptors, host proteases, and several essential genes that are implicated in the replication and pathogenesis of one or another human infecting coronavirus, including SARS-CoV-2. The fold change in expression of these genes was plotted to generate a clustered heatmap using pheatmap R package. In addition, another heatmap was prepared using 182 differentially regulated genes (at fold change  $>\pm 2$  and  $p < 0.05$ ) in COVID-19 human patients as compared to healthy controls. The graphical representation of gene expression values was obtained using GraphPad Prism. We used student's *t*-test to analyze the gene expression data from SARS-CoV-2 infected human lung tissue, and  $p < 0.05$  was used as the level of significance. The data has been plotted as mean  $\pm$  standard error from mean, and each dot represents individual reading.

## Results

### Human genes that are implicated in the pathogenesis of COVID-19 are upregulated in leukemia, psoriasis, NAFLD, and Type II diabetes conditions

We investigated the differential gene expression in a set of chronic, acute, and infectious disease conditions to identify the one(s) that may induce susceptibility to SARS-CoV-2 infection. The overall methodology of this study has been depicted in figure 1. At first, we retrieved the data of differential gene expression in 30 disease conditions along with their respective controls from GEO database. In some disease conditions, there were more than one datasets, and the sample size of individual datasets was large. Therefore, for differentiating the gene expression profiles and clustering of various datasets in each disease, we performed a principle component analysis (PCA). We observed that the gene expression profile of datasets in a given disease correlated with each other, and each disease produced distinctly different gene expression signature (Figure 2A). Several aspects of SARS-CoV-2 infection and spread are still under active research, and many aspects have been reported as identical or similar to that of pathogenic human infecting coronaviruses (HCoVs), namely, SARS-CoV and MERS-CoV. Because gene expression pattern is characteristically correlated with the pathogenesis of diseases, we asked that what are the expression pattern of genes that are implicated in the replication and pathogenesis of SARS-CoV-2 or other HCoVs, in different disease conditions? To this, we performed a literature mining exercise and identified 78 genes that are reported to have important implications in the entry and pathogenesis of HCoVs. These genes are enlisted in Supplementary Table 1. Some of these genes have identified roles in promoting the pathogenesis of SARS-CoV-2 (viz. ACE2 receptor, and furin proteases), while others may find implications subject to the findings of future studies. To identify the expression pattern of these genes, we prepared their expression heatmap in all 30 diseases and in the COVID-19 condition (Figure 2B). We observed that many of these genes are upregulated during SARS-CoV-2 infection and in such diseases, namely, all the studied subtypes of leukemia (hereafter, referred as leukemia), NAFLD, psoriasis, NSCLC1,

breast cancer, cervical cancer, and Type II diabetes. Of note, the gene expression pattern was particularly pronounced in leukemia, and NAFLD, wherein most of the genes were highly upregulated.

### **The expression of *ACE2*, *FURIN*, and *TMPRSS2* is increased in leukemia, NAFLD, and psoriasis patients**

The earliest steps in establishing COVID-19 includes cellular entry of SARS-CoV-2 that is critically dependent on the host's ACE2 receptors, and FURIN and TMPRSS2 serine proteases. Whereas the ACE2 acts as a viral receptor and interacts with viral spike, the FURIN, and TMPRSS2 proteases are essential for processing the spike to facilitate viral entry into the cells<sup>11,12</sup>. The gene expression signatures of such diseases that can promote the cellular entry of SARS-CoV-2 may be a critical factor associated with the observed comorbidities in COVID-19. We, therefore, investigated the expression pattern of ACE2, FURIN, and TMPRSS2 in the patients of breast cancer, cervical cancer, leukemia, NAFLD, NSCLC1, psoriasis, and Type II diabetes. The expression of *ACE2* was significantly upregulated in leukemia, psoriasis, NAFLD, lung cancer, breast cancer, and cervical cancer patients (Figure 3A). Following the interaction of viral spike with ACE2 receptor, the host's FURIN protease cleaves the spike protein at the interface of two subunits of the trimeric spike. Thus, the protease activity of FURIN is critical in promoting spike mediated entry of SARS-CoV-2, and it is also known to be crucial for protein processing in other infectious diseases and cancer as well<sup>13</sup>. Interestingly the expression of *FURIN* was upregulated during leukemia, psoriasis, NAFLD, lung cancer, and in type II diabetic liver whereas it was downregulated during breast cancer (Figure 3B). Similar to FURIN, the proteolytic cleavage of spike protein by TMPRSS2 is critical for its fusogenic activity. We observed that *TMPRSS2* was upregulated during leukemia, lung cancer, and type II diabetes, and it was downregulated during psoriasis, NAFLD, lung cancer, breast cancer, and cervical cancer (Figure 3C).

### **Disease-associated dysregulation of innate and adaptive immune response exposes the patients to SARS-CoV-2 infection**

Following entry, the presence of viral RNA in cellular milieu evokes an immune response in the host. Apart from the receptors and proteases, the heatmap in Figure 2A also shows the differential expression of important cellular proteins, which are involved in the innate and acquired immune response to the SARS-CoV-2 invasion in all 30 diseases. Interferon-alpha and -beta (i.e., IFNA2, IFNA8, IFNA10, IFNA14, IFNA16, IFNA21, and IFNB1, respectively) are the initial response elements of the innate immune signaling pathway that activate several interferon-stimulated genes (ISGs) via JAK1/STAT1 pathway, which leads to early clearance of the viral load<sup>14</sup>. We prepared a heatmap showing the expression of those interferons that were differentially regulated in one or another disease ( $p < 0.05$ , Figure 4A). Besides, we checked the expression of essential genes encoding interferons and cytokines mediated anti-viral immune response, namely, *IL6*, *CXCL10*, *JAK1*, and *STAT1*, (Figure 4B-E, respectively). *IFNA2*, *IFNA8*, *IFNA10*, *IFNA14*, *IFNA16*, *IFNA21* and *IFNB1*, were upregulated in leukemia, and NAFLD, while both were downregulated in type II diabetic liver. The expression of *JAK1* was slightly decreased in leukemia, except BCLL, whereas *STAT1* was slightly decreased in TALL and remained unchanged in other types of leukemia. Both *JAK1* and *STAT1* were increased in NAFLD, and type II diabetic muscles. The initial



interferon-mediated response is followed by a specific cell-mediated adaptive immune response to clear viral invasion. The cytokines, mainly IL6 and CXCL10, are produced by helper T cells and macrophages that promote the migration of immune cells at the site of infection, and they are also associated with the cytokine storm observed in COVID-19 associated mortalities. We found that the expression of *IL6* and *CXCL10* was upregulated in leukemia. Besides, the expression of *IL6* was decreased in NAFLD, whereas *CXCL10* was increased in NAFLD and PDAC.

### **The pattern of differential gene expression is similar in SARS-CoV-2, leukemia, and NAFLD**

In addition to the plausible or reported genes, derived through literature mining, which may have implications in the pathogenesis of SARS-CoV-2, we also sought to analyze the observed differential gene expression after SARS-CoV-2 infection in our studied disease types. From GSE150316 dataset, we retrieved 193 differentially expressed genes from 16 COVID-19 patients (at  $\geq 2$  fold change;  $p < 0.05$ ). Given the upregulation of genes implicated in HCoV's pathogenesis in some disease types, it would be worthwhile to identify the expression pattern of these genes in all other disease types that will provide further insight into similar pathogenic mechanism. Out of these 193 differentially expressed genes, we found that the expression for 182 genes was available in datasets of all disease types included in our study. Therefore, we generated a clustered heatmap depicting the expression of these 182 differentially expressed genes in 30 diseases and in COVID-19 condition (Figure 5A). It is evident from the heatmap that the pattern of gene expression is similar in COVID-19 and PDAC, which segregated together in the heatmap with a small cluster of lung cancer and lung invasive adenocarcinoma (NSCLC). For most of the upregulated genes in COVID-19, we observed similar expression in leukemia, NAFLD, and psoriasis. We analyzed the pathways enriched by these 181 genes using DAVID functional annotation tool v6.8<sup>15,16</sup>. All the observed pathways were mainly implicated in the host's immune response to viral infection or infection-related immune disorders (Figure 5B). Furthermore, we inferred the expression of 10 topmost significantly altered up- and down-regulated genes in COVID-19 patient's to those diseases that showed similar pattern of gene expression, namely, Liver in T2D, NAFLD, Psoriasis, Leukaemia, PDAC, Lung cancer, NSCLC, TNBC, breast cancer, and cervical cancer (Supplementary Figure 1). Interestingly, we found that three genes, namely, RAMP3 (Receptor Activity Modifying Protein 3), S100A2 (S100 Calcium Binding Protein A2), and CLCA2 (Chloride Channel Accessory 2) that are implicated in calcium signalling, were prominently altered in at least 7 studied disease types.

### **Pathogenic HCoV's differentially regulate the expression of *CARBONIC ANHYDRASE 11* and *CLUSTERIN* genes**

We extended our investigation to identify the gene product(s), which may be commonly altered by the dreaded viruses of recent times. We analyzed the differential gene expression during various viral infections, namely, SARS-CoV (GSE1739), SARS-CoV-2, and MERS-CoV (GSE100496), H1N1 (GSE21802), and other influenza (GSE22319; H7N1, H5N1, H3N2, H5N2). We observed that no gene was commonly altered in all of these viral infections (Figure 6A). However, two genes, namely *CARBONIC ANHYDRASE 11* (CA11) and *CLUSTERIN* (CLU), were significantly altered after the infection by pathogenic HCoV's, namely, SARS-CoV, SARS-CoV-2, and MERS-CoV (Figure 6B). Based on the similarities

in the pattern of differential gene expression during COVID-19 and other 30 diseases, we checked the expression of *CA11* and *CLU* during breast cancer, cervical cancer, leukemia, NAFLD, NSCLC1, psoriasis, and Type II diabetes. The expression of *CA11* was significantly upregulated in COVID-19 and leukemia (Figure 6C), while *CLU* was upregulated in COVID-19 and type II diabetes and downregulated in NAFLD, psoriasis, breast cancer, and cervical cancer patients (Figure 6D).

## Discussion

COVID-19 associated comorbidities have been reported with several acute and chronic diseases that lead to poorer outcomes. For instance, diabetes, cardiovascular disease, renal and pulmonary diseases, etc. have been frequently observed comorbidities that increase the case fatality rate in acute respiratory diseases caused by SARS-CoV<sup>17</sup>, MERS-CoV<sup>18,19</sup>, and SARS-CoV-2. Our study revealed that the characteristic gene expression of diseases; namely, leukemia, NAFLD, psoriasis, and type II diabetes underlie the COVID-19 associated comorbidities. Further, we identified key differentially regulated genes implicated in the pathogenesis of COVID-19 that may induce susceptibility to SARS-CoV-2 infection.

The expression of ACE2 receptors and host proteases such as *FURIN*, *TMPRSS2*, and *TMPRSS4*, along with several components of innate and adaptive immune response was evaluated in a total of 30 chronic, acute, and infectious diseases. Angiotensin I converting enzyme 2, i.e., ACE2, acts as the functional receptor of three human infecting coronaviruses, namely NL-63, SARS-CoV, and SARS-CoV-2. Earlier reports suggest that the expression of ACE2 is highly pronounced in lung alveolar epithelium cells and enterocytes of the small intestine, and to a smaller extent in the vascular endothelial and smooth muscle in several organs including kidney, liver, bone marrow, skin, and brain among others<sup>20</sup>. Through ACE2, these organs may provide an easy port of entry for SARS-CoV-2, and more severe symptoms can develop upon SARS-CoV-2 infection, particularly in the respiratory tract and gut. The degree of ACE2 expression, in association with a 10-20 fold higher binding affinity of SARS-CoV-2 spike, can be an important reason behind the efficient cellular entry and higher infectivity of SARS-CoV-2 as compared to SARS-CoV<sup>21,22</sup>. In normal human physiology, ACE2 acts as a key modulator of the renin-angiotensin system, and clinical conditions such as myocardial infarction is known to increase the expression of ACE2<sup>23</sup>. Upon investigating the gene expression profile in 30 studied disease types, we found that the basal expression of ACE2 was significantly increased in many pathological conditions. To name, we found that the expression of ACE2 was increased in leukemia, cancer of lungs, breasts, and cervix, NAFLD, psoriasis, and pulmonary arterial hypertension. Hence, the increased number of available cellular receptors that facilitate viral entry may explain the increased susceptibility of these diseases to SARS-CoV-2 infection.

Following the initial interaction of viral spike with ACE2 receptors, the host proteases play an important role in spike-receptor interaction and subsequent fusion of viral envelope to the host membrane, which subsequently triggers the initial events of viral replication. The pre-activation of viral spike by cleavage at polybasic S1/S2 site in the spike is mediated by a proprotein convertase *FURIN* that enables a second cleavage by cellular serine protease *TMPRSS2*. Both these proteolytic cleavages are important in facilitating viral entry; and, inactivation of either *FURIN*<sup>12</sup> or *TMPRSS2*<sup>11</sup> has been reported to inhibit cell-cell fusion and entry of SARS-CoV-2 in lung cells. Other cellular proteases such as *TMPRSS4*<sup>24</sup>, and



Cathepsin B and L<sup>22</sup> are also proposed to have a cumulative effect on the *FURIN*-mediated promotion of SARS-CoV-2 entry into enterocytes or liver and lungs cells, respectively. Therefore, we evaluated the expression of several proprotein convertases and cathepsins, including those mentioned above, in 30 diseases and in COVID-19. Barring CLL, BALL, and TALL, wherein *CATHEPSIN A*, *B*, and *D* were downregulated, we found that majority of the host proteases were highly upregulated during leukemia. Although cathepsins have an additive effect, they may not be indispensable for viral entry; however, their increased expression along with that of *FURIN* and *TMPRSS2* can promote the processing of viral spike and enhance cellular entry of SARS-CoV-2<sup>11</sup>. Taken together; these results suggest that the patients of leukemia are highly prone to SARS-CoV-2 infection as the relevant host receptor and proteases are highly upregulated in leukemia. Similarly, the expression of *FURIN* was increased in NAFLD and psoriasis patients, which is conducive to SARS-CoV-2 infection, however, the expression of *TMPRSS2* was found to be decreased. Recently, the abundance of *FURIN* or *TMPRSS2* was shown to be sufficient in promoting the cellular entry of SARS-CoV-2, and protease inhibitors blocked this activity<sup>11</sup>. In addition, proteases other than *TMPRSS2* such as cathepsins or PCSKs may fill in for the reduced expression of *TMPRSS2* in NAFLD or psoriasis patients. Although the relevant information on psoriasis is missing in the literature, Dong et al. (2020) have observed NAFLD comorbidity in COVID-19 patients<sup>5</sup>. Therefore, we propose that the increased expression of *ACE2*, and *FURIN* can be the reason behind COVID-19 associated comorbidities; however, studies aimed at testing the redundancy of proteases will be required to arrive at a definite conclusion.

Following the proteolytic cleavage of viral spike, the next crucial event in the viral entry is the fusion of viral envelope with host membrane and subsequent release of the positive-stranded RNA genome into cell, wherein it takes on the cellular machinery for self-proliferation. The primary defense component against viral invasion is the interferon-mediated innate immune response of the host. The production and binding of type I and type III interferons to their respective cellular receptors culminate into activating JAK1/STAT1 mediated transcription of several anti-viral interferon-stimulated genes. It has been reported that the JAK1 deficient mice exhibit poor lymphoid development, defective response to cytokines, and interferons, and die perinatally<sup>25</sup>. Similarly, mice with disrupted expression of STAT1 had compromised innate immunity<sup>26</sup>, and they were prone to viral infections<sup>27</sup>. Further, somatic mutations and dysregulation of JAK1 mediated signaling has been frequently observed in acute lymphoblastic leukemia<sup>28,29</sup>. In addition, the TYK2 (tyrosine kinase 2) deficiency in humans was also shown to induce cytokine signaling defects and susceptibility to infection<sup>30</sup>. We found that the expression of *IFNA2*, *IFNA8*, *IFNA10*, *IFNA14*, *IFNA16*, *IFNA21* and *IFNB1* were increased, whereas that of *JAK1*, *STAT1*, and *TYK2* did not change significantly in leukemia patients. Thus, the increased interferon response in leukemia patients may involve components other than JAK1, STAT1, and TYK2. In contrast, the increased expression of *JAK1* and *STAT1* corresponds well with the increased *IFNA2*, *IFNA8*, *IFNA10*, *IFNA14*, *IFNA16*, *IFNA21* and *IFNB1* production in NAFLD patients. In contrast, the expression of *JAK1* and *STAT1* decreased in type II diabetic muscle, whereas the interferon production was decreased in type II diabetic liver. Although the phosphorylation status of STAT1 was not deciphered in the datasets of transcriptome profiling that were used in our study, the phosphorylation events are crucial in STAT1 mediated signaling. Thus, the level of p-STAT1 must also be checked in future studies that aim to reveal the correlation between STAT1 expression in leukemia and COVID-19

associated comorbidities. Further, the higher basal levels of interferons may be one of the reasons behind a previously observed report of relatively milder symptomatic COVID-19 in CLL patients<sup>31</sup>. However, SARS-CoV-2 may use several escape or immune-suppression strategies, including the formation of a replication organelle, and 2'-O-methylated capping of viral RNA, to continually proliferate despite increased basal interferon levels in leukemia and NAFLD patients<sup>32</sup>. Thereafter, the specific adaptive immune response comes into effect for curbing the viral invasion. An optimal secretion of cytokines and chemokines (viz. IL6 and CXCL10) from immune cells is essential to adjust the host's immune response against foreign invaders. However, the excess release of cytokines, i.e., cytokine storm, is associated with increased severity of disease and poorer outcomes in SARS-CoV<sup>33,34</sup> and SARS-CoV-2<sup>35</sup> infected patients. Earlier, the inhibition of NF $\kappa$ B mediated production of IL6 was found to increase the survival SARS-CoV infected mice<sup>36</sup>, and IL6 blockade has been thought as a mechanism to manage cytokine storm and save COVID-19 patients<sup>37</sup>. We observed higher expression of *CXCL10* in leukemia, NAFLD, and PDAC patients. Recently Malard et al. (2020) have also reported that patients of hematologic malignancies are at higher risk of developing severe form of COVID-19<sup>38</sup>. Thus, the higher basal level of cytokines may be a causal factor behind the observed cytokine storm and poorer outcomes in COVID-19 associated comorbidities. To the best of our knowledge, this study, for the first time, identifies the signature gene expression behind such observations. We propose that the inhibitors of IL6 and CXCL10 may be examined as precise clinical interventions in leukemia, and NAFLD patients who have been tested positive for SARS-CoV-2.

We also checked the similarities in the pattern of differential gene expression in our studied group of diseases and COVID-19. The pattern of gene expression was remarkably similar in COVID-19 and leukemia patients. Also, we observed that calcium signalling was perturbed in COVID-19 and at least 6 other studied disease types, including leukemia, NAFLD, and psoriasis, as manifested by the altered expression of *RAMP3*, *S100A2*, and *CLCA2* genes. Whereas *RAMP3* is a co-activator that targets calcium-sensing receptor to cell surface<sup>39,40</sup>, *CLCA2* regulates the calcium activated chloride channel currents and enhance the store-operated cellular entry of calcium<sup>41</sup>. *S100A2* encodes a cytoplasmic calcium binding protein and is known to be dysregulated in human cancers<sup>42</sup>. Together, these three genes modulate the cellular calcium level in response to a stimuli, and were distinctly upregulated in leukemia, NAFLD, and psoriasis among (Supplementary Figure 1). Recently, Sun et al. (2020) showed that calcium channel blockers inhibit the replication of SARS-CoV-2 in the cellular milieu, and reduce the COVID-19 associated case-fatality rate<sup>43</sup>. Thus, apart from the expression of SARS-CoV-2 receptor and host proteases, the cellular calcium levels may also play a significant role in inducing the susceptibility to SARS-CoV-2 infection. Further, we found that *CA11* and *CLU* gene were the only two genes that are simultaneously altered by SARS-CoV, SARS-CoV-2, and MERS-CoV. Although *CA11* was upregulated in all the studied leukemia types, no trend was observed in the expression of *CLU* in the studied disease types. As of now, no direct correlation is identified between the expression of *CA11*, or *CLU* with the pathogenesis of SARS-CoV-2, although recently, several authors have identified that COVID-19 may lead to ketosis, ketoacidosis<sup>44</sup>, and altered glucose metabolism associated with the diabetogenic effect of COVID-19<sup>45</sup>. Since *CA11* plays an important role in hepatic gluconeogenesis<sup>46</sup>, it will be interesting to investigate the potential relationship between SARS-CoV-2 infection, differential *CA11* expression, and new-onset of diabetes in COVID-19 patients.

One of the limitations of this study was the inclusion of a single dataset in almost all diseases; although, we believe that the large sample size of each dataset makes our findings widely applicable. A systemic approach of identifying DEGs in several organs in each disease may produce novel findings due to altered cross-talk between different organs in diseased conditions compared to healthy individuals. We conclude that the patients of leukemia are relatively more susceptible to COVID-19 associated comorbidities, followed by NAFLD, psoriasis, type II diabetes, and hypertension, than other diseases examined in this study. We also propose that disease-specific inhibition of IL6, CXCL-10, JAK1, and STAT1, alone or in various combinations, may be helpful in curbing COVID-19 associated comorbidities. Our findings will help the healthcare systems across the globe in devising better management practices for preventing the complications of COVID-19 associated comorbidities.

## Acknowledgement

MKS and AC acknowledge CSIR for the research fellowship, AM acknowledges ICMR for the research fellowship.

## References

1. Guan WJ, Liang WH, Zhao Y, et al. Comorbidity and its impact on 1590 patients with COVID-19 in China: a nationwide analysis. *Eur Respir J* 2020; **55**(5).
2. Chen N, Zhou M, Dong X, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet* 2020; **395**(10223): 507-13.
3. Yang F, Shi S, Zhu J, Shi J, Dai K, Chen X. Clinical characteristics and outcomes of cancer patients with COVID-19. *J Med Virol* 2020.
4. Li W, Wang D, Guo J, et al. COVID-19 in persons with chronic myeloid leukaemia. *Leukemia* 2020.
5. Ji D, Qin E, Xu J, et al. Non-alcoholic fatty liver diseases in patients with COVID-19: A retrospective study. *J Hepatol* 2020.
6. Blanco JL, Ambrosioni J, Garcia F, et al. COVID-19 in patients with HIV: clinical case series. *Lancet HIV* 2020; **7**(5): e314-e6.
7. Arruga F, Gyau BB, Iannello A, Vitale N, Vaisitti T, Deaglio S. Immune Response Dysfunction in Chronic Lymphocytic Leukemia: Dissecting Molecular Mechanisms and Microenvironmental Conditions. *Int J Mol Sci* 2020; **21**(5).
8. Wu JT, Leung K, Leung GM. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *Lancet* 2020; **395**(10225): 689-97.
9. Walsh CJ, Hu P, Batt J, Santos CC. Microarray Meta-Analysis and Cross-Platform Normalization: Integrative Genomics for Robust Biomarker Discovery. *Microarrays (Basel)* 2015; **4**(3): 389-406.
10. Hamid JS, Hu P, Roslin NM, Ling V, Greenwood CM, Beyene J. Data integration in genetics and genomics: methods and challenges. *Hum Genomics Proteomics* 2009; **2009**.
11. Hoffmann M, Kleine-Weber H, Schroeder S, et al. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* 2020; **181**(2): 271-80 e8.

12. Hoffmann M, Kleine-Weber H, Pohlmann S. A Multibasic Cleavage Site in the Spike Protein of SARS-CoV-2 Is Essential for Infection of Human Lung Cells. *Mol Cell* 2020; **78**(4): 779-84 e5.
13. Braun E, Sauter D. Furin-mediated protein processing in infectious diseases and cancer. *Clin Transl Immunology* 2019; **8**(8): e1073.
14. Le Page C, Genin P, Baines MG, Hiscott J. Interferon activation and innate immunity. *Rev Immunogenet* 2000; **2**(3): 374-86.
15. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 2009; **4**(1): 44-57.
16. Huang da W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* 2009; **37**(1): 1-13.
17. Lau EH, Hsiung CA, Cowling BJ, et al. A comparative epidemiologic analysis of SARS in Hong Kong, Beijing and Taiwan. *BMC Infect Dis* 2010; **10**: 50.
18. Kulcsar KA, Coleman CM, Beck SE, Frieman MB. Comorbid diabetes results in immune dysregulation and enhanced disease severity following MERS-CoV infection. *JCI Insight* 2019; **4**(20).
19. Yang YM, Hsu CY, Lai CC, et al. Impact of Comorbidity on Fatality Rate of Patients with Middle East Respiratory Syndrome. *Sci Rep* 2017; **7**(1): 11307.
20. Hamming I, Timens W, Bulthuis ML, Lely AT, Navis G, van Goor H. Tissue distribution of ACE2 protein, the functional receptor for SARS coronavirus. A first step in understanding SARS pathogenesis. *J Pathol* 2004; **203**(2): 631-7.
21. Liu M, Wang T, Zhou Y, Zhao Y, Zhang Y, Li J. Potential Role of ACE2 in Coronavirus Disease 2019 (COVID-19) Prevention and Management. *J Transl Int Med* 2020; **8**(1): 9-19.
22. Shang J, Wan Y, Luo C, et al. Cell entry mechanisms of SARS-CoV-2. *Proc Natl Acad Sci U S A* 2020; **117**(21): 11727-34.
23. Ishiyama Y, Gallagher PE, Averill DB, Tallant EA, Brosnihan KB, Ferrario CM. Upregulation of angiotensin-converting enzyme 2 after myocardial infarction by blockade of angiotensin II receptors. *Hypertension* 2004; **43**(5): 970-6.
24. Zang R, Gomez Castro MF, McCune BT, et al. TMPRSS2 and TMPRSS4 promote SARS-CoV-2 infection of human small intestinal enterocytes. *Sci Immunol* 2020; **5**(47).
25. Igaz P, Toth S, Falus A. Biological and clinical significance of the JAK-STAT pathway; lessons from knockout mice. *Inflamm Res* 2001; **50**(9): 435-41.
26. Durbin JE, Hackenmiller R, Simon MC, Levy DE. Targeted disruption of the mouse Stat1 gene results in compromised innate immunity to viral disease. *Cell* 1996; **84**(3): 443-50.
27. Jung SR, Ashhurst TM, West PK, et al. Contribution of STAT1 to innate and adaptive immunity during type I interferon-mediated lethal virus infection. *PLoS Pathog* 2020; **16**(4): e1008525.
28. Flex E, Petrangeli V, Stella L, et al. Somatically acquired JAK1 mutations in adult acute lymphoblastic leukemia. *J Exp Med* 2008; **205**(4): 751-8.
29. Hornakova T, Staerk J, Royer Y, et al. Acute lymphoblastic leukemia-associated JAK1 mutants activate the Janus kinase/STAT pathway via interleukin-9 receptor alpha homodimers. *J Biol Chem* 2009; **284**(11): 6773-81.
30. Minegishi Y, Saito M, Morio T, et al. Human tyrosine kinase 2 deficiency reveals its requisite roles in multiple cytokine signals involved in innate and acquired immunity. *Immunity* 2006; **25**(5): 745-55.
31. Baumann T, Delgado J, Montserrat E. CLL and COVID-19 at the Hospital Clinic of Barcelona: an interim report. *Leukemia* 2020.

32. Kikkert M. Innate Immune Evasion by Human Respiratory RNA Viruses. *J Innate Immun* 2020; **12**(1): 4-20.
33. Wang W, Ye L, Ye L, et al. Up-regulation of IL6 and TNF-alpha induced by SARS-coronavirus spike protein in murine macrophages via NF-kappaB pathway. *Virus Res* 2007; **128**(1-2): 1-8.
34. Castilletti C, Bordi L, Lalle E, et al. Coordinate induction of IFN-alpha and -gamma by SARS-CoV also in the absence of virus replication. *Virology* 2005; **341**(1): 163-9.
35. Zhang C, Wu Z, Li JW, Zhao H, Wang GQ. Cytokine release syndrome in severe COVID-19: interleukin-6 receptor antagonist tocilizumab may be the key to reduce mortality. *Int J Antimicrob Agents* 2020; **55**(5): 105954.
36. DeDiego ML, Nieto-Torres JL, Regla-Nava JA, et al. Inhibition of NF-kappaB-mediated inflammation in severe acute respiratory syndrome coronavirus-infected mice increases survival. *J Virol* 2014; **88**(2): 913-24.
37. Liu B, Li M, Zhou Z, Guan X, Xiang Y. Can we use interleukin-6 (IL6) blockade for coronavirus disease 2019 (COVID-19)-induced cytokine release syndrome (CRS)? *J Autoimmun* 2020; **111**: 102452.
38. Malard F, Genthon A, Brissot E, et al. COVID-19 outcomes in patients with hematologic disease. *Bone Marrow Transplant* 2020.
39. Bouschet T, Martin S, Henley JM. Receptor-activity-modifying proteins are required for forward trafficking of the calcium-sensing receptor to the plasma membrane. *J Cell Sci* 2005; **118**(Pt 20): 4709-20.
40. D'Souza-Li L. The calcium-sensing receptor and related diseases. *Arq Bras Endocrinol Metabol* 2006; **50**(4): 628-39.
41. Sharma A, Ramena G, Yin Y, Premkumar L, Elble RC. CLCA2 is a positive regulator of store-operated calcium entry and TMEM16A. *PLoS One* 2018; **13**(5): e0196512.
42. Bresnick AR, Weber DJ, Zimmer DB. S100 proteins in cancer. *Nat Rev Cancer* 2015; **15**(2): 96-109.
43. Zhang L, Sun Y, Zeng H-L, et al. Calcium channel blocker amlodipine besylate is associated with reduced case fatality rate of COVID-19 patients with hypertension. *medRxiv* 2020.
44. Li J, Wang X, Chen J, Zuo X, Zhang H, Deng A. COVID-19 infection may cause ketosis and ketoacidosis. *Diabetes Obes Metab* 2020.
45. Rubino F, Amiel SA, Zimmet P, et al. New-Onset Diabetes in Covid-19. *N Engl J Med* 2020.
46. Ismail IS. The Role of Carbonic Anhydrase in Hepatic Glucose Production. *Curr Diabetes Rev* 2018; **14**(2): 108-12.



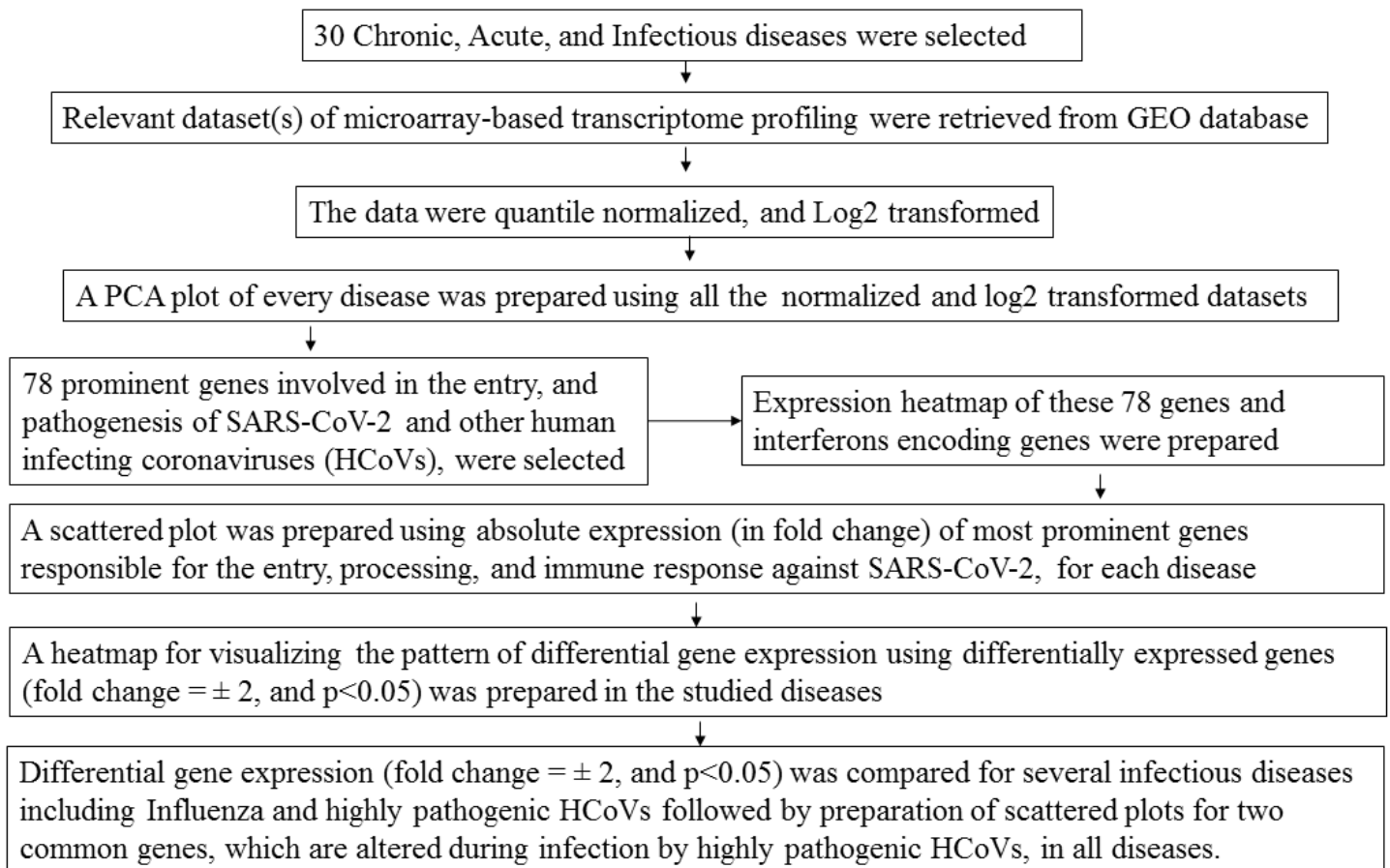
**Table 1: Details of Expression Datasets taken from GEO for the study.**

| <b>Disease Type</b> | <b>Disease/ Condition</b>              | <b>GSE ID</b> | <b>Platform</b>                       | <b>Tissue</b>  | <b>Experimental Design</b>  | <b>Data Processing</b>  |
|---------------------|--|---------------|---------------------------------------|--|---|---|
| Chronic             | Asthma                                 | GSE64913      | GPL570 [HG-U133_Plus_2]<br>Affymetrix | Epithelial brushings from central and peripheral airways | 42 healthy volunteer, 28 asthmatic patients                                   | Preprocessed: Normalization and log2 transformation by GCRMA method |
|                     | Chronic Obstructive Pulmonary Disorder | GSE112811     | GPL570 [HG-U133_Plus_2]<br>Affymetrix | Blood  | 20 COPD patients, 22 healthy volunteer before administration of LPS or saline | Preprocessed: Normalization and log2 transformation by RMA          |
|                     | Cardiovascular                         | GSE109048     | GPL17586 [HTA-2_0]<br>Affymetrix      | Blood platelets  | 19 Healthy donors, 19 CAD patients, 19 AMI patients                           | Preprocessed: SST-RMA normalization and log2 transformation         |
|                     | Hypertension                           | GSE113439     | GPL6244 [HuGene-1_0-st]<br>Affymetrix | Lung   | 15 patients with Pulmonary Arterial Hypertension and 11 normal controls       | Preprocessed: Normalization and log2 transformation by RMA          |
|                     | Non-Alcoholic Fatty Liver Disease      | GSE49541      | GPL570 [HG-U133_Plus_2]<br>Affymetrix | Liver  | 72 patients with NAFLD  | Preprocessed: Normalization and log2 transformation by GCRMA method |
|                     |  | GSE107037     | GPL570 [HG-U133_Plus_2]<br>Affymetrix | Liver  | 33 healthy liver donors   | Preprocessed: Normalization and log2 transformation by RMA          |

|        |                           |           |                                       |                |   |  |
|--------|---------------------------|-----------|---------------------------------------|----------------|---|--|
|        | Atherosclerosis           | GSE28829  | GPL570 [HG-U133_Plus_2]<br>Affymetrix | Carotid artery | Samples from atherosclerotic carotid artery segments of 29 patients | Preprocessed: Normalization and log2 transformation by RMA |
|        | Type 2 Diabetes           | GSE15653  | GPL96 [HG-U133A]<br>Affymetrix        | Liver          | 4 type 2 diabetes and 5 control subjects                            | Preprocessed: MAS5.0 signal intensity.                     |
|        |                           | GSE25462  | GPL570 [HG-U133_Plus_2]<br>Affymetrix | Muscle         | 10 subjects with type 2 diabetes and 15 healthy subjects            | Preprocessed: MAS5.0 signal intensity.                     |
|        |                           | GSE38642  | GPL6244 [HuGene-1_0-st]<br>Affymetrix | Pancreas       | 54 non-diabetic and 9 diabetic cadaver                              | Preprocessed: Normalization and log2 transformation by RMA |
|        |                           | GSE27949  | GPL570 [HG-U133_Plus_2]<br>Affymetrix | Adipose        | 12 Normal and 11 T2D subjects                                       | Preprocessed: Normalization and log2 transformation by RMA |
|        | Polycystic Ovary Syndrome | GSE124226 | GPL570 [HG-U133_Plus_2]<br>Affymetrix | Adipose        | 4 PCOS women and 4 control subjects                                 | Preprocessed: Normalization and log2 transformation by RMA |
|        | Multiple Sclerosis        | GSE21942  | GPL570 [HG-U133_Plus_2]<br>Affymetrix | PBMCs          | 12 MS patients and 15 controls                                      | Preprocessed: Normalization GCRMA method                   |
| Cancer | Blood Cancer (Leukemia)   | GSE78097  | GPL570 [HG-U133_Plus_2]<br>Affymetrix | Skin           | 6 normal skin tissues and 27 psoriatic skin lesion                  | Preprocessed: Normalization GCRMA method                   |
|        |                           | GSE51082  | GPL96 [HG-U133A]<br>Affymetrix        | Bone Marrow    | 37 AML, 41, BCLL1, 22 CML, 10 MDS, 17 B-ALL, 12 T-ALL               | Preprocessed: Normalization and log2 transformation by RMA |
|        |                           | GSE9476   | GPL96 [HG-U133A]<br>Affymetrix        | Bone Marrow    | 38 healthy donors   | Preprocessed: Normalization                                |

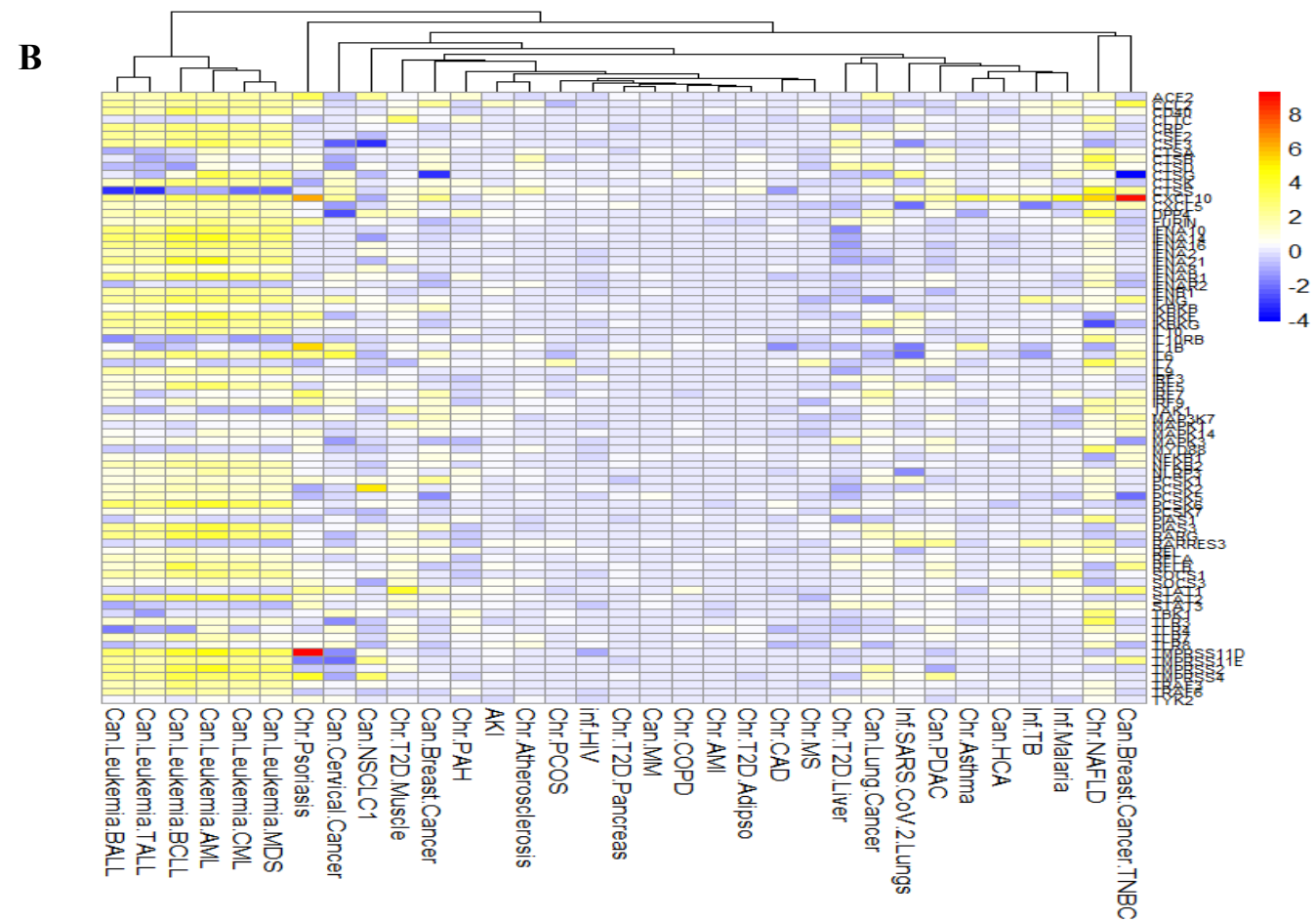
|  |           |                                    |               |   |   |                                |
|--|-----------|------------------------------------|---------------|---|---|--------------------------------|
|  |           |                                    | Affymetrix    |   |   | and log2 transformation by RMA |
| Breast Cancer                                    | GSE65194  | GPL570 [HG-U133_Plus_2] Affymetrix | Breast sample | 11 control breast sample, 98 breast cancer samples, 55 TNBC samples | Preprocessed: Normalization and log2 transformation by GCRMA method                                 |                                |
| Cervical Cancer                                  | GSE63514  | GPL570 [HG-U133_Plus_2] Affymetrix | Cervix        | 24 normal and 28 cancer specimens                                   | Preprocessed: Normalization and log2 transformation by GCRMA method                                 |                                |
| Multiple Myeloma                                 | GSE85837  | GPL10558 Illumina HumanHT-12 V4.0  | Bone Marrow   | 9 control and 9 multiple myeloma patients with bone lesion          | Preprocessed: Robust spline normalization and log2 transformation by lumi R package                 |                                |
| Lung Cancer                                      | GSE136043 | GPL13497 Agilent-026652            | Lung          | 5 lung cancer tissue and 5 lung non-tumour tissues                  | Preprocessed: Normalization by Agilent Feature Extraction Software                                  |                                |
| Lung adenocarcinoma (Non-small cell lung cancer) | GSE118370 | GPL570 [HG-U133_Plus_2] Affymetrix | Lung          | 6 invasive lung adenocarcinoma tissues and 6 normal lung tissues    | Preprocessed: Normalization and log2 transformation by MAS5.0 algorithm                             |                                |
| Liver Cancer                                     | GSE88839  | GPL570 [HG-U133_Plus_2] Affymetrix | Liver         | 35 HCA liver tumours and 3 normal liver samples                     | Preprocessed: Normalization by RMA  |                                |
| Pancreatic Ductal Adenocarcinoma                 | GSE101448 | GPL10558 Illumina HumanHT-12 V4.0  | Pancreas      | 18 with pancreatic tumor and 13 non-tumor pancreatic tissue samples | Preprocessed: Normalization and log2 transformation by Illumina's BeadStudio Data Analysis Software |                                |

|            |                     |           |                                    |                      |  |   |
|------------|---------------------|-----------|------------------------------------|----------------------|--|---|
| Infectious | AIDS                | GSE73968  | GPL6244 [HuGene-1_0-st] Affymetrix | T Cells              | 9 healthy control and 6 HIV positive patients                                  | Preprocessed: Normalization and log2 transformation by RMA            |
|            | Tuberculosis        | GSE139825 | GPL10558 Illumina HumanHT-12 V4.0  | Alveolar Macrophages | Alveolar Macrophages from 5 TB patients and 5 control subjects                 | Preprocessed: Normalization and log2 transformation by lumi R package |
|            | Malaria             | GSE119150 | GPL15207 [PrimeView] Affymetrix    | Blood                | 6 falciparum malaria and 6 normal subjects                                     | Preprocessed: Normalization and log2 transformation by RMA            |
| Acute      | Acute Kidney Injury | GSE30718  | GPL570 [HG-U133_Plus_2] Affymetrix | Kidney               | 28 transplants with AKI to 11 pristine protocol biopsies of stable transplants | Preprocessed: Normalization and log2 transformation by RMA            |
|            | COVID-19            | GSE150316 | GPL18573                           | Lung                 | 16 lung samples with COV2 positive and 5 control lung samples                  | Preprocessed: DEseq2 normalized                                       |



**Figure 1**





### Figure 2

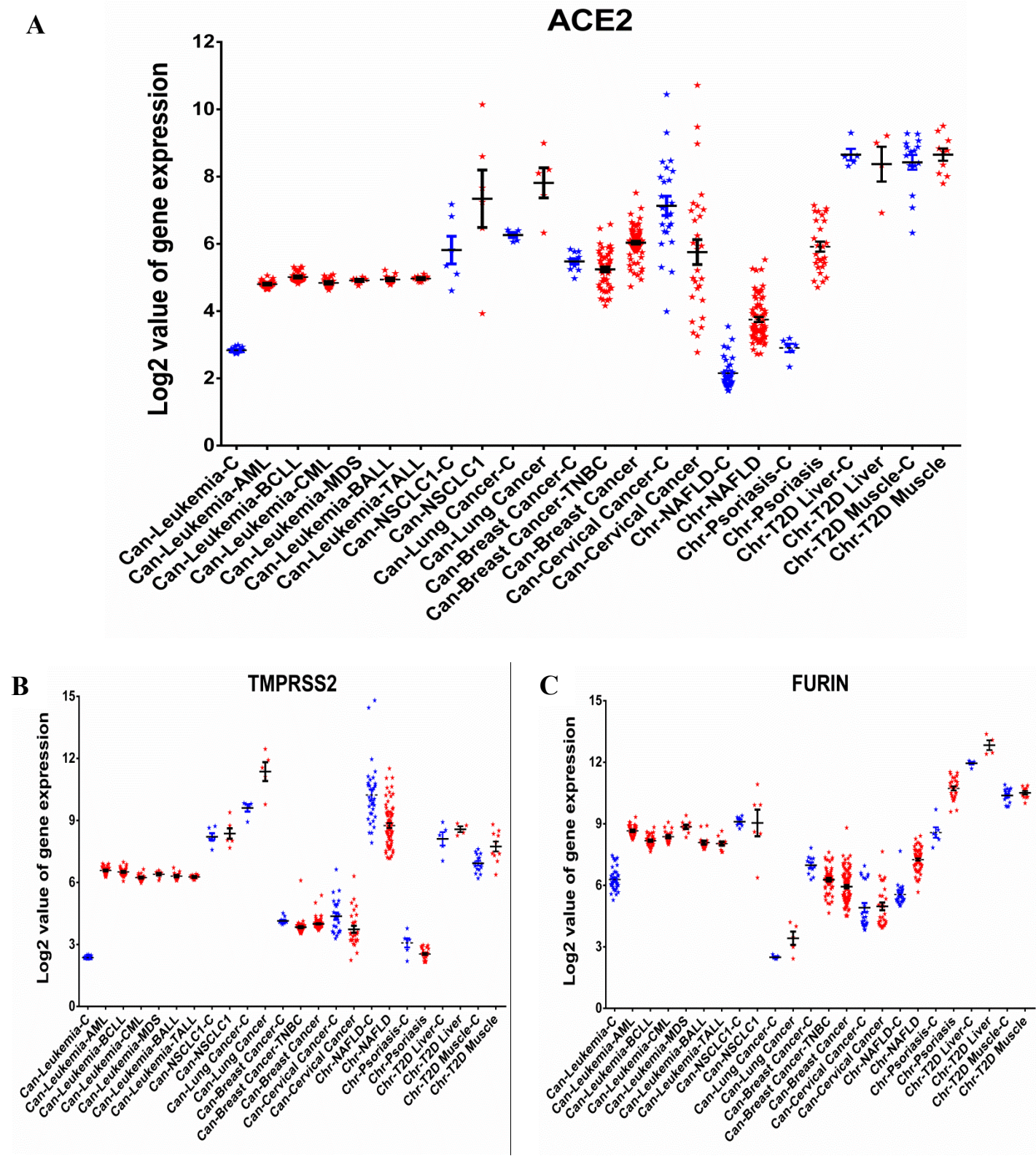


Figure 3



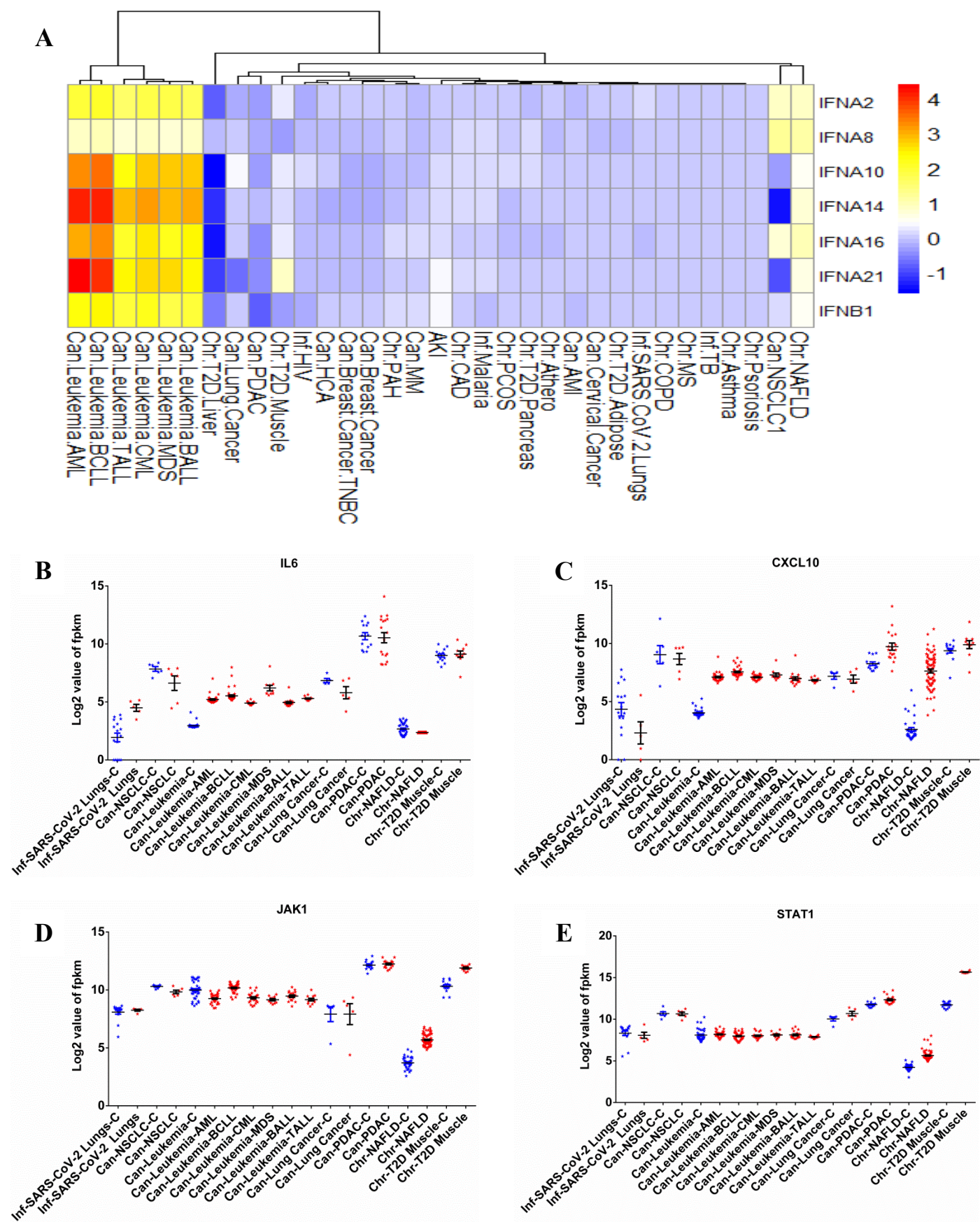
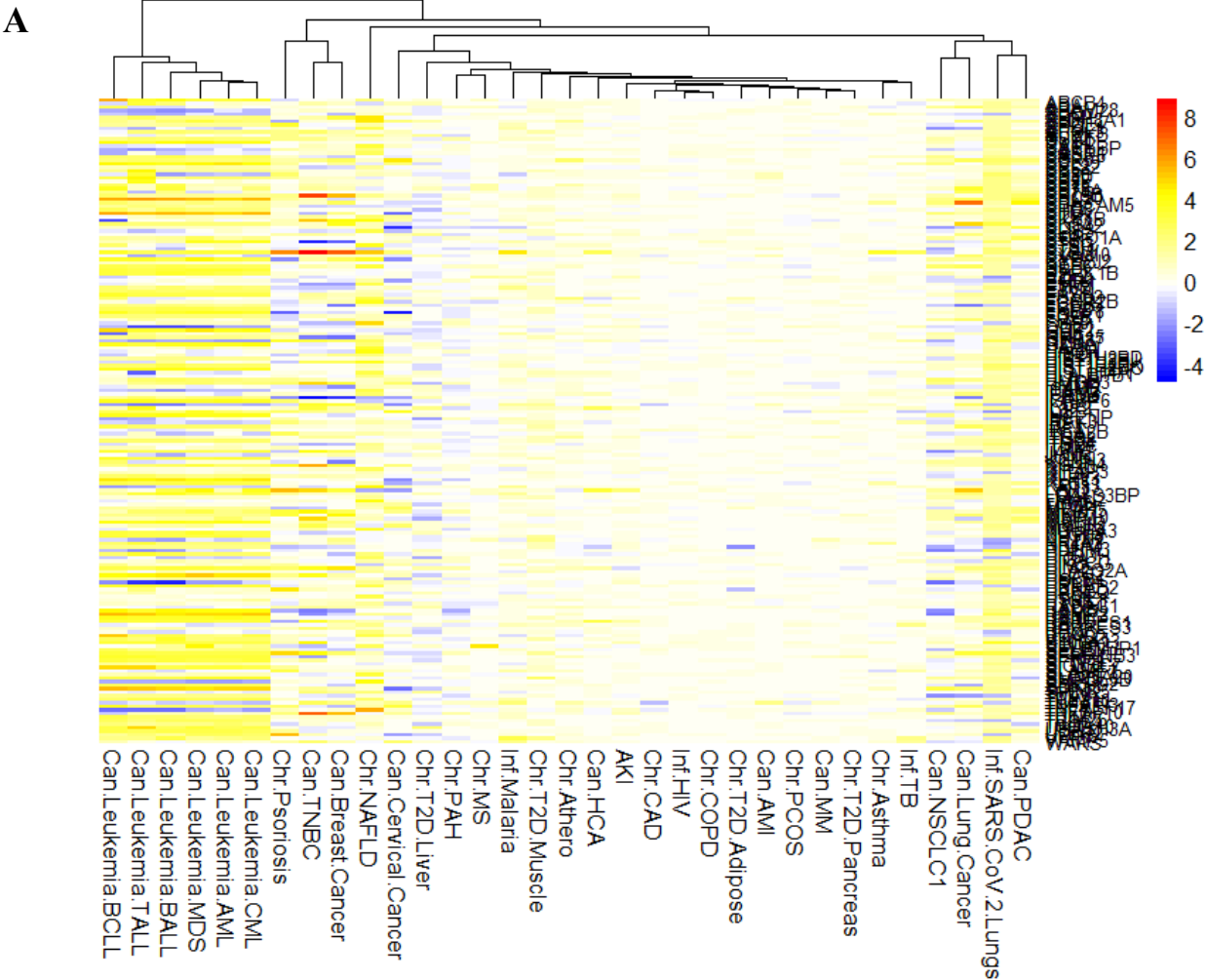


Figure 4



**B**

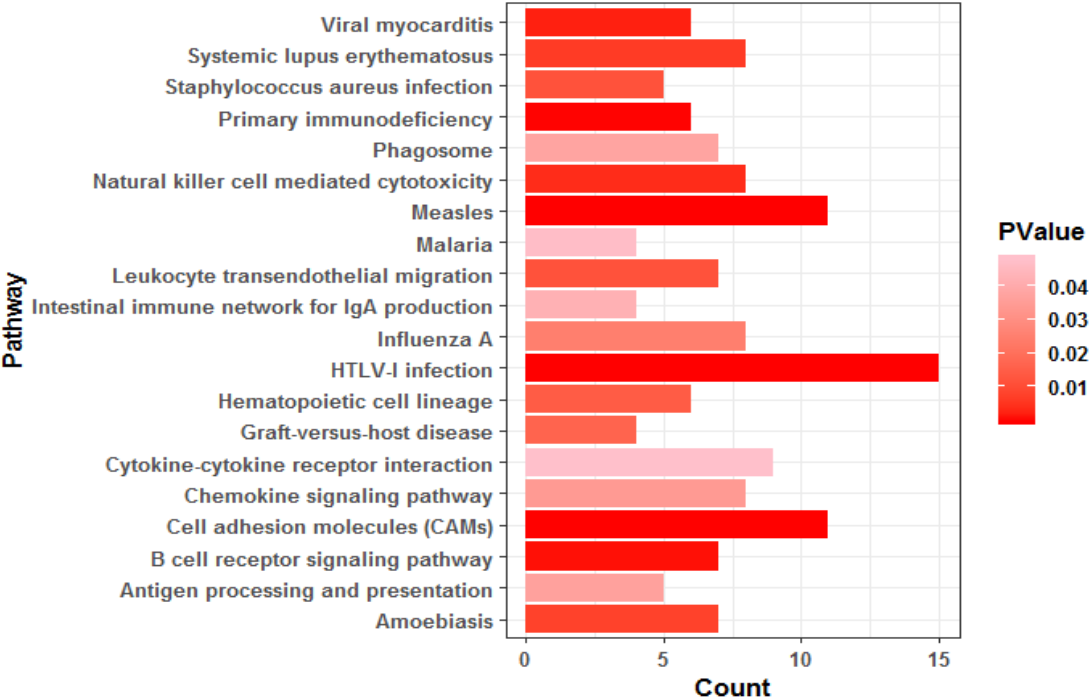


Figure 5

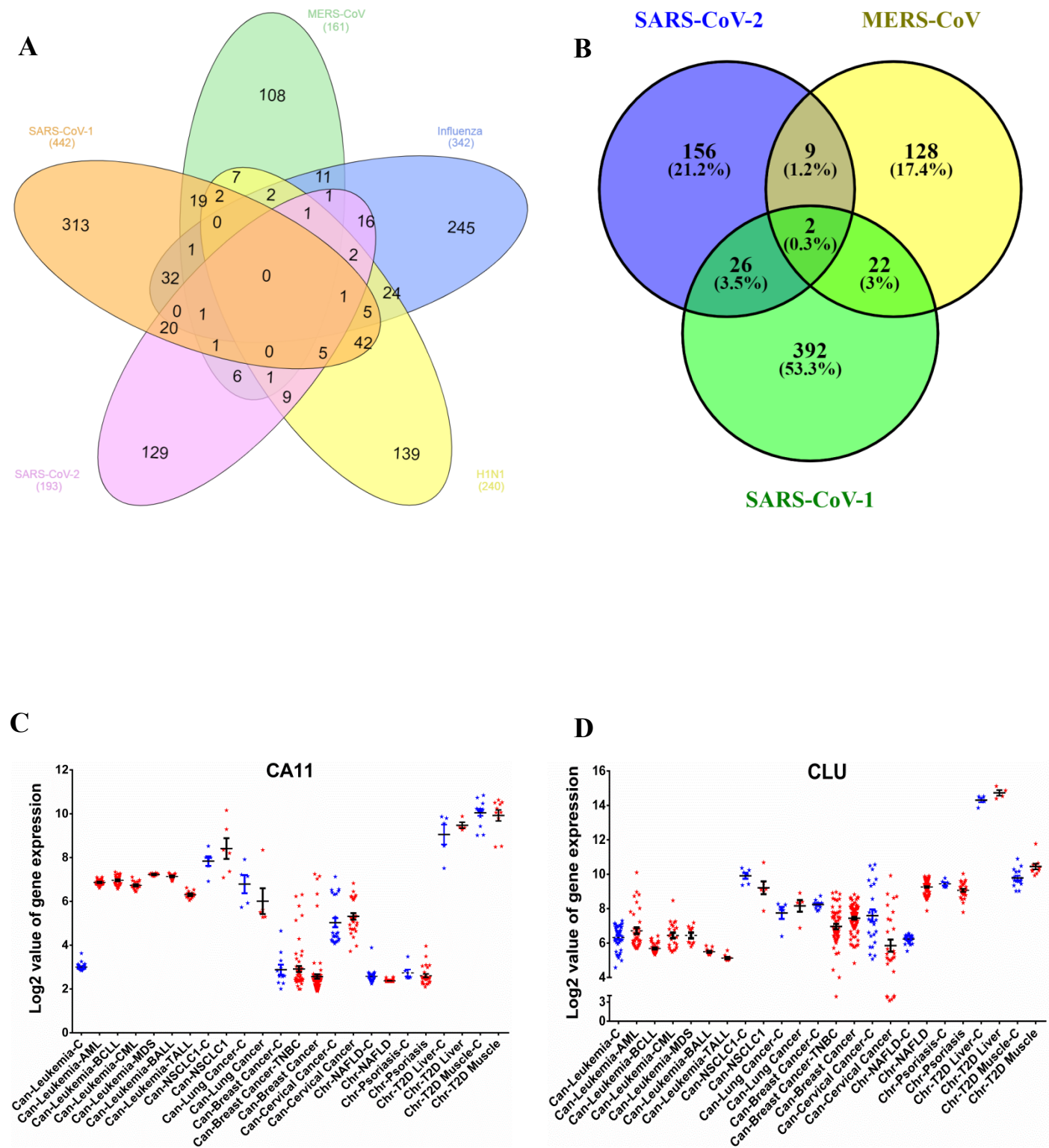


Figure 6



## Figure legends

**Figure 1.** Flow diagram depicting the study design.

**Figure 2.** (A) The PCA of datasets used in this study for 30 disease conditions and their controls. Individual datasets are represented by separate dots. To categorize and differentially color different disease types, we have imparted the prefix “Can,” “Chr,” and “Inf” to identify and cluster various types of cancer, chronic diseases, and infectious diseases together, respectively. (B) Clustered heat map depicting fold change in the expression of 78 genes derived from literature mining, which may have implications in the pathogenesis of SARS-CoV-2, during COVID-19 and 30 studied disease conditions.

**Figure 3.** Quantitative representation of expression of (A) *ACE2*, (B) *TMPRSS2*, and (C) *FURIN* in patients and their respective controls, from 14 selected diseases wherein these genes were differentially regulated. Each dot represents fold change in individual patient, and the bars depict standard error of mean in a given disease.

**Figure 4.** (A) Clustered heat map depicting fold change in the expression of genes encoding various types of interferon- alpha and -beta, which may be crucial in the pathogenesis of SARS-CoV-2, during COVID-19 and 30 other studied disease conditions. Quantification of (B) *IL6*, (C) *CXCL10*, (D) *JAK1*, and (E) *STAT1* expression in patients and their respective controls, from 12 selected diseases including COVID-19, wherein these genes were found to be differentially regulated. Each dot represents fold changes from individual patient, and the bars depict standard error of mean for a given disease.

**Figure 5.** (A) Clustered heat map depicting fold change in the expression of 183 genes that are significantly altered post SARS-CoV-2 infection (at a fold change  $> \pm 2$ ,  $p < 0.05$ ) and in the 30 studied disease conditions. (B) Pathways-enrichment analysis using 183 genes that are simultaneously altered post SARS-CoV-2 infection (at a fold change  $> 2$ ,  $p < 0.05$ ) and in the 30 studied disease conditions.

**Figure 6.** Venn diagram depicting number of genes that are commonly differentially regulated during (A) various viral infections, namely, SARS-CoV, SARS-CoV-2, MERS-CoV, H1N1, and other influenza (namely, H7N1, H5N1, H3N2, and H5N2), and (B) post-infection by only pathogenic human infecting coronaviruses, namely, SARS-CoV, SARS-CoV-2, and MERS-CoV.  $p < 0.05$  was selected as the level of significance. Quantification of (C) *Carbonic anhydrase 11 (CA11)*, and (D) *Clusterin (Clu)* expression in patients and their respective controls, from 14 selected diseases including COVID-19, wherein these genes were found to be differentially regulated. Each dot represents fold changes from individual patients, and the bars depict standard error of mean for a given disease.

**Supplementary Figure 1.** Bubble plot of fold change in 10 topmost significantly altered up- and down-regulated genes in COVID-19 patients, depicting their pattern of expression in 30 other studied disease types. Dark red bubbles represent fold changes that are significant, while the light pink bubbles depict the statistically insignificant changes.  $p < 0.05$  was selected as the level of significance.